

Оглавление

Лекция 1	4
1.1 Предисловие	4
1.2 Конечные вероятностные пространства и равновероятные элементарные исходы. Классическая вероятность	5
1.2.1 Задача о разделе ставки	5
1.2.2 Примеры устойчивости частот	6
1.2.3 Петербургский парадокс	7
Лекция 2	8
2.1 Аксиоматика А. Н. Колмогорова	8
2.1.1 Вероятностное пространство, σ -алгебра событий, вероятность	8
2.1.2 Примеры вероятностных пространств	11
2.1.3 Свойства вероятности	13
Лекция 3	21
3.1 Урновые схемы	21
3.1.1 Выборка с возвращением, биномиальное распределение	21
3.1.2 Выборка без возвращения, гипергеометрическое распределение	22
3.2 Условная вероятность	23
3.3 Независимость множества событий	26
3.3.1 Пример Бернштейна	26
3.4 Формула полной вероятности	27
3.5 Формула Байеса	28
3.6 Дискретные случайные величины, индикаторы событий	29
3.7 Схема Бернулли	31
Лекция 4	33
4.1 Независимость дискретных случайных величин, теорема о независимости двух функций от непересекающихся совокупностей независимых дискретных случайных величин	33
4.2 Математическое ожидание дискретной случайной величины	35
4.2.1 Свойства математического ожидания	35
4.2.2 Пример случайной величины, не имеющей мат. ожидания	38
4.2.3 Мат. ожидание биномиально распределённой случайной величины	39
4.3 Моменты и центральные моменты k -го порядка	39
4.4 Дисперсия и её свойства	40
4.4.1 Дисперсия биномиально распределённой случайной величины	41
4.4.2 Среднеквадратичное отклонение	42

4.5	Ковариация и коэффициент корреляции	42
4.5.1	Свойства коэффициента корреляции	43
4.6	Неравенство А. А. Маркова	44
4.7	Неравенство Чебышева	45
4.7.1	Правило трёх сигм	46
Лекция 5		47
5.1	Закон больших чисел Чебышева	47
5.2	Теорема Бернулли	47
5.3	Пуассоновское распределение	48
5.4	Теорема Пуассона	48
5.4.1	Оценка близости биномиальных вероятностей к пуассоновским	50
5.5	Задача о конкуренции (завязка)	50
5.6	Локальная предельная и интегральная теоремы Муавра-Лапласа . . .	51
5.7	Задача о конкуренции (развязка)	52
Лекция 6		53
6.1	Задача о различении двух гипотез о доле шаров в урне	53
6.1.1	Ошибки первого и второго рода	54
6.1.2	Оценка для числа наблюдений, необходимых для различения гипотез с заданной точностью	55
6.2	Определение вероятностного пространства в общем случае	56
6.3	Сигма-алгебра, порождённая классом множеств	58
6.4	Борелевская сигма-алгебра	58
6.5	Случайные величины как измеримые отображения	60
Лекция 7		62
7.1	Функция распределения случайной величины	62
7.1.1	Свойства функции распределения	62
7.2	Распределение случайной величины	64
7.3	Теорема о единственности продолжения вероятности с алгебры на порождённую ею σ -алгебру	65
7.4	Взаимно однозначное соответствие между функциями распределения и вероятностными распределениями	65
Лекция 8		69
8.1	Неравенство и УЗБЧ Колмогорова	69
8.1.1	Неравенство Колмогорова	69
8.1.2	Усиленный закон больших чисел Колмогорова	70
8.1.3	Применение УЗБЧ: метод Монте-Карло	71
8.2	Сходимость в среднем	72
8.3	Производящие функции	73
8.3.1	Свойства производящих функций	73
Лекция 9		76
9.1	Характеристические функции	76
9.1.1	Свойства характеристических функций	78
9.2	Формула обращения	80

Лекция 10	84
10.1 Слабая сходимос <small>ть</small>	84
10.2 Предельные теоремы для характеристических функций	87
Лекция 11	91
11.1 Метод характеристических функций в доказательстве предельных теорем	91
11.1.1 Центральная предельная теорема	93
Лекция 12	96
12.1 Условное математическое ожидание, условное распределение случайной величины	96
12.1.1 Условное математическое ожидание, условное распределение в случае дискретных случайных величин	96
12.1.2 Условное математическое ожидание, условное распределение в случае абсолютно непрерывных случайных величин	97
12.2 Свойства условного математического ожидания	99
Лекция 13	100
13.1 Введение в математическую статистику	100
13.2 Эмпирическая функция распределения	101
13.3 Эмпирические или выборочные моменты	102
13.4 Порядковые статистики и вариационные ряды	104
13.5 Статистические оценки	105
13.5.1 Свойства несмещенных оценок	106
13.5.2 Свойства состоятельных оценок	106
Лекция 14	107
14.1 Оптимальные оценки	107
14.2 Неравенство Рао - Крамера	108
14.2.1 Функция правдоподобия	108
14.3 Метод моментов	111
14.4 Достаточные статистики	112
Лекция 15	115
15.1 Достаточные и полные статистики	115
15.2 Оценки максимального правдоподобия	116
Лекция 16	119
16.1 Доверительные интервалы и трактовка коэффициента доверия	119
Лекция 17	121
17.1 Методы построения интервальных оценок	121
17.1.1 Метод, основанный на точечных оценках	121
17.1.2 Метод, основанный на центральной статистике	122
17.1.3 Метод, основанный на центральной предельной теореме	124
17.2 Проверка статистических гипотез	126
17.2.1 Общая постановка проверки статистических гипотез	126
17.2.2 Типы гипотез	126

Лекция 18	130
18.1 Лемма Неймана-Пирсона	130
18.2 Равномерно наиболее мощные критерии	132
Лекция 19	134
19.1 Состоятельные критерии	134
19.2 Критерий χ^2 Пирсона	134
19.2.1 Критерий χ^2 Пирсона	134
19.2.2 Асимптотика для критерия Пирсона	135
19.3 Состоятельность критерия Пирсона	136
19.4 Обобщение классического критерия χ^2	137
Лекция 20	138
20.1 Критерий согласия Пирсона χ^2 для абсолютно непрерывных распределений	138
20.2 Критерий независимости χ^2	139
20.3 Критерий независимости χ^2	140
20.4 Проверка гипотез и доверительный интервал	140
20.5 Несмещенные критерии	141

Лекция 1

1.1 Предисловие

Представим себе жидкость, которая с постоянным расходом w перетекает из первого сосуда во второй. Нам точно известно, что через время t из первого сосуда во второй перетечет wt жидкости. Мы точно знаем, что произойдет, наш ответ мы даем с вероятностью один.

Теперь, получив образование и достав где-то N единиц некоторой валюты, положим их в ПИФ "Зияющие высоты". Теперь едва-ли кто-нибудь знает, что там с ними будет через год. Есть несколько вариантов:

- Наши сбережения выросли в полтора раза! Вероятность такого события, положим, 80%.
- Наши сбережения едва-едва не обесценились в результате потрясений в финансовом мире, но остались в сохранности. Вероятность этого события, скажем, 12%.
- И, наконец, наши сбережения исчезли. Вероятность — 8%.

Теперь на вопрос, что же там будет дальше, мы отвечаем, используя слово «вероятность».

Это — одно из обстоятельств, с которым придется иметь дело в этом курсе. Дать точно определенный ответ на вопрос, что же произойдет, нельзя, хотя бы из-за установленных Гейзенбергом фактов квантовой механики. Зато, если повезет, можно точно указать вероятность некоторого события. Чему равна вероятность выпадения в казино на рулетке именно той позиции, на которой некто вечером в пятницу составил башенку из всех фишек, которые смог добыть?

Один из источников теории вероятности — азартные игры, и это не всем нравится. И рулетка, между прочим, возникла как средство борьбы с карточными шулерами!

А владельцы казино тем временем, располагая данными о вероятностях выпадения позиций на рулетке, распределении ставок и выигрышей, посчитали, сколько в среднем выносят из казино и сколько в нем оставляют. Разумеется, выносят немного меньше. Эта игра — несправедливая.

Определение 1.1. *Игра называется справедливой тогда и только тогда, когда средние выигрыши и проигрыши равны между собой.*

Еще один источник теории вероятности — страхование морских грузов. В случае страхования определить требуемые вероятности уже гораздо сложнее. При страховании жизни, например, ставка должна зависеть от очень и очень многих вещей. Взять и посчитать нужные величины, используя только точно известные физические константы и фундаментальные законы не представляется возможным. И здесь появляется математическая статистика, призванная на основе полученных опытным путем данных строить вероятностную модель.

1.2 Конечные вероятностные пространства и равновероятные элементарные исходы. Классическая вероятность

Считается, что теория вероятности родилась летом 1654 года, когда была решена задача о разделе ставки.

1.2.1 Задача о разделе ставки

Задача. Два игрока, точно равных в силах, играют в разные игры и за одну победу получают одно очко. Всего запланировано 11 раундов. Внезапно игры остановились. На момент остановки счет оказался равным 5 : 3. Вопрос, как справедливо делить приз, решается следующим образом: приз делится пропорционально вероятности выигрыша каждого из игроков. Вот только вопрос определения вероятностей выигрышей решить еще осталось! Итак, как делить приз?

Решение. Разберемся с тем, что может произойти за последние три игры. Имеется восемь раскладов для побед в этих играх, все они равновозможны в силу одинаковости игроков в силе. Первый игрок одержал 5 побед, второй — 3. Победу первого игрока будем обозначать «А», победу второго — «В». Итак, варианты исходов последних трёх игр:

AAA	}	при этих раскладах победит первый,
AAB		
ABA		
ABB		
BAA		
BAB		
BBA		
BBB		

BBB — и только в этом случае победит второй.

В семи из восьми случаев в итоге победит первый игрок. И только в одном из них — второй. Все эти случаи равновероятны, равносильны при учете в разделе приза. Поэтому 7/8 получит первый игрок, а 1/8 получит второй. И на самом деле — первому достаточно одержать только одну победу, в то время как второму — целых три. И отношение долей — 7 : 1, а вовсе не 5 : 3.

Теперь следует попытаться формализовать процесс решения задачи. Были рассмотрены все возможные исходы. К счастью, в этой задаче их конечное количество (что не всегда так — вот, например, бросить точку в отрезок от нуля до одного, исходов будет континуум). Все исходы равновероятны (что тоже не всегда так — просто дадим одному из игроков допинг, увеличивающий силу в e раз). Событий было два: побеждает первый игрок или второй. Каждому из событий соответствует свой набор элементарных исходов, и необязательно эти наборы имеют пустое пересечение.

Все дальнейшие определения вводятся в том контексте, что имеется некоторый случайный эксперимент.

Определение 1.2. Пусть $\Omega = \{\omega_1, \dots, \omega_n\}$ — множество всех исходов случайного эксперимента. В этом случае множество Ω называется **пространством элементарных исходов**.

Замечание. Вообще-то, один и тот же эксперимент может быть представлен разными пространствами элементарных исходов. Далее будет нужна равновозможность элементарных исходов и конечность Ω . Остальное пока неважно.

Замечание. С тех пор, как все результаты содержатся в конечном множестве Ω , любое событие A можно считать подмножеством Ω . Строгое определение события появится несколько позже. В задаче о разделе ставки победу первого игрока означало множество из семи элементарных исходов, победу второго игрока — множество из одного элементарного исхода.

Определение 1.3. Пусть есть конечное пространство элементарных исходов Ω , все элементарные исходы равновозможны. Пусть имеется событие A , $A \subseteq \Omega$. Тогда **классическая вероятность** события A определяется как

$$P(A) = \frac{\|\!|A\|\!|}{\|\!|\Omega\|\!|}.$$

Пусть теперь нам удалось построить множество всех событий. Обозначим его F . В задаче о разделе ставки событий было, грубо говоря, два: победа первого игрока и победа второго игрока. Можно говорить о событии «победил один из них» или «не победил никто», можно придумывать события и далее. То есть, речь идет максимум о множестве всех подмножеств Ω . В конечном случае никаких проблем тут не возникает, а вот множество всех подмножеств сегмента от нуля до одного, как выяснится позже, обладает весьма неподходящими для вероятности свойствами.

Определение 1.4. Тройку (Ω, F, P) , где Ω — конечное множество всех элементарных исходов, F — множество всех событий и P — вероятность, определенная для всех элементов из F , будем называть **конечным вероятностным пространством**.

1.2.2 Примеры устойчивости частот

Как на деле установить, равновозможны элементарные исходы или нет — это серьезный вопрос. Можно, например, обратиться к статистике, о которой мы ничего пока не знаем. Так вот выдвинем гипотезу о равновозможности элементарных исходов и оценим, насколько она правдоподобна. Пока ничего не знаем о статистике, эти действия проведем интуитивно и нестрого.

Определение 1.5. Пусть A — некоторое событие. Пусть проведено N экспериментов, причем событие A наступило в N_A экспериментах. Отношение

$$\nu_A = \frac{N_A}{N}$$

будем называть **частотой события A** .

Мысль подсказывает, что если некоторый эксперимент провести много раз, причем так, чтобы эксперименты никак друг на друга не влияли, то «в среднем» отношение экспериментов, в которых событие наступило, к общему числу экспериментов должно дать вероятность этого события. То есть положимся на не всегда верное предположение о том, что

$$P(A) \approx \nu_A.$$

Для примеров из этой лекции оно нормально работает.

Пример. Требуется выяснить, равновозможны ли выпадение герба и решки при бросании идеальной монетки честным экспериментатором. Иными словами, надо разобраться, будут ли элементарные исходы «выпала решка» и «выпал герб» равновозможными. Этим вопросом занимались в разное время разные исследователи:

XVIII в: монету подбрасывал **Бюффон**: на 4040 бросков пришлось 2048 гербов. Частота для выпадения герба составила 0,507;

XIX в: **Морган** подбросил монету 4092 раза и получил столько же гербов. Частота для выпадения герба составила 0,5005;

XX в: **Пирсон** совершил 24000 бросков, герб выпал 12012 раз. Частота для выпадения герба вновь составила 0,5005;

XX в: **Романовский** сделал это 80640 раз, герб выпал 39699 раз. У Романовского частота получилась равной 0,4923.

С точностью до третьего знака все естествоиспытатели получили 0,5. Этот факт демонстрирует устойчивость частоты выпадения герба при подбрасывании правильной монетки (правильным экспериментатором).

Пример. В некотором населенном пункте в 1935 году осуществлялся сбор данных о новорожденных:

в январе родилось 7280 детей, из них 3743 мальчика, частота — 0,514,
в феврале родилось 6957 детей, из них 3550 мальчиков, частота — 0,510.

В демографии вероятность рождения мальчика принимается равной 0,514.

1.2.3 Петербургский парадокс

Пример. Пусть Аня и Боря решили сыграть на деньги. Аня подбрасывает правильную монетку, так что вероятности выпадения герба и решки равны $1/2$ каждая. Боря платит Ане 2^n рублей, если первый герб выпал на n -том шаге.

Проблема заключается в том, что при попытке отыскать средний выигрыш Ани неминуемо получится бесконечность, если нигде не ошибиться. В свое время этот факт не давал покоя некоторым исследователям. Для того, чтобы сделать игру справедливой, Аня должна Боре перед началом подбрасываний монеты предоставить бесконечное количество рублей, этот факт устанавливается на с.38.

Кроме того, в приведённой ситуации пространство элементарных исходов счётно. Классическая вероятность здесь просто неприменима. Так что если попробовать её тут приложить, парадокс непременно возникнет.

Представим себе множество всех натуральных чисел, пусть это будет пространством элементарных исходов $\Omega = \mathbb{N}$. Объявим их равновозможными. Тогда попробуем посчитать вероятность события A , состоящего из одного любого элементарного исхода. Попытка применить тут классическую вероятность приведет в тупик:

$$P(A) = \frac{\|A\|}{\|\Omega\|},$$

если с $\|A\|$ вопросов не возникает, то вот разобраться с тем, что же такое в данном случае $\|\Omega\|$, не получится. Эта формула тут вообще некорректна, потому что Ω вновь счетно, а вовсе не конечно. Кроме того, сама модель не позволяет построить вероятностное пространство. Все станет просто и очевидно после появления более строгих формулировок и определений, но об этом — на следующей лекции.

Лекция 2

2.1 Аксиоматика А. Н. Колмогорова

2.1.1 Вероятностное пространство, σ -алгебра событий, вероятность

Начнём с того, что есть и другие аксиоматики. Даже с отрицательной вероятностью! Однако, в данном курсе вполне достаточно аксиоматики Колмогорова, о которой речь пойдет на этой лекции.

Определение 2.1. Пусть Ω — любое непустое множество. *Событием* будем называть любое подмножество Ω . Событие Ω называется **достоверным** событием, а событие \emptyset — **невозможным**. Само множество Ω называется **пространством элементарных исходов**.

Замечание. Очень скоро под словом «событие» будет пониматься уже не всякое подмножество Ω .

Определение 2.2. События A и B называются **несовместными**, если $A \cap B = \emptyset$.

Определение 2.3. Множество F подмножеств Ω называется **классом событий** $\iff F$ обладает следующими тремя свойствами:

1°. Классу событий принадлежит достоверное событие:

$$\Omega \in F.$$

2°. Класс событий для любого своего события содержит его отрицание:

$$A \in F \implies \bar{A} \in F.$$

3°. Свойство счётной аддитивности:

$$\forall i \in \mathbb{N} A_i \in F \implies \bigcup_{i=1}^{\infty} A_i \in F.$$

Замечание. Здесь важно заметить, что, имея на руках операции объединения и отрицания, можно построить операцию пересечения следующим образом:

$$A \cap B \equiv \overline{\bar{A} \cup \bar{B}}.$$

Определение 2.4. F — вырожденный класс событий $\iff F = \{\emptyset, \Omega\}$.

Замечание. Вырожденный класс событий — это наименьший класс событий, он обладает указанными выше тремя свойствами. Как минимум, Ω и $\bar{\Omega}$ класс событий обязан содержать. Меньше никак нельзя.

Определение 2.5. Множество F , обладающее указанными выше тремя свойствами, называется **сигма-алгеброй**. Разные «счётные» операции за пределы сигма-алгебры не выводят.

Определение 2.6. Событием называется подмножество Ω , являющееся элементом σ -алгебры событий.

Определение 2.7. Пусть P — отображение из класса событий F в \mathbb{R} .

P называется **вероятностью** $\iff P$ обладает следующими тремя свойствами:

1°. Нормировка: вероятность достоверного события равна единице:

$$P(\Omega) = 1.$$

2°. Вероятность неотрицательна:

$$\forall A \in F \quad P(A) \geq 0.$$

3°. Счётная аддитивность вероятности или σ -аддитивность вероятности:

$$\forall i \in \mathbb{N} \quad A_i \in F, \quad \forall i, j \in \mathbb{N} \quad i \neq j \implies A_i A_j = \emptyset$$

\Downarrow

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Утверждение 2.1. Вероятность невозможного события равна нулю.

$$P(\emptyset) \equiv 0.$$

Доказательство. Рассмотрим следующую счётную последовательность событий: $\Omega, \emptyset, \emptyset, \emptyset, \dots$. События из этой последовательности попарно несовместны и объединение их равно Ω . Далее пользуемся свойством счётной аддитивности для этой последовательности:

$$\Omega = \Omega \cup \emptyset \cup \emptyset \cup \emptyset \cup \dots \implies P(\Omega) = P(\Omega) + P(\emptyset) + P(\emptyset) + P(\emptyset) + \dots$$

\Downarrow

$$1 = 1 + P(\emptyset) + P(\emptyset) + \dots \implies 0 = P(\emptyset) + P(\emptyset) + P(\emptyset) + \dots \implies P(\emptyset) = 0. \quad \square$$

Если считать понятие меры известным, то вероятность — это нормированная мера. То есть, вероятность от меры отличается только уловием $P(\Omega) = 1$.

Теперь несколько слов по поводу того, зачем же нужна σ -алгебра. От множества всех подмножеств Ω σ -алгебра в общем случае существенно отличается. Пример тому, описанный наклонным шрифтом на с.12, — подмножество отрезка от нуля до одного, не имеющее меры по Лебегу, такие множества называются **неизмеримыми**. Требуется именно существование меры Лебега подмножеств Ω , с которыми мы будем работать, потому что иначе не существует для них и вероятности. Для таких множеств просто нельзя её ввести, не потеряв при этом какое-нибудь очень нужное из трёх её свойств. И именно членство в правильно построенной σ -алгебре событий строго гарантирует существование вероятности безо всяких нарушений, парадоксов и потерь. Неизмеримые множества из σ -алгебры специально исключаются, и никакими не выводимыми из неё операциями получить их из измеримых множеств нельзя.

А вот для счётного Ω такой проблемы нет.

Утверждение 2.2. Для любой вероятностной модели со счётным Ω в качестве σ -алгебры событий F можно брать множество всех подмножеств Ω , мощность F — континуум.

Доказательство. Итак, пусть у нас имеется счётное количество элементарных исходов:

$$\Omega = \{\omega_1, \omega_2, \omega_3, \dots\}.$$

Покажем, что для множества всех подмножеств Ω , которое обозначим через F , определена вероятность. Выполненность для F свойств σ -алгебры очевидна.

Построим взаимно-однозначное соответствие между элементами F и строками из нулей и единиц счётной длины B . Пусть $A \in F$ — событие. Строку, соответствующую A обозначим B_A , цифру, стоящую в i -той позиции строки B_A обозначим b_A^i . B_A определим следующим образом:

$$b_A^i = \begin{cases} 1, & \text{если } \omega_i \in A, \\ 0, & \text{если } \omega_i \notin A. \end{cases}$$

↓

$$\forall A \in F \exists ! B_A, \forall A_1, A_2 \in F: A_1 \neq A_2 \implies B_{A_1} \neq B_{A_2} \text{ и } \forall B_1 \in B \exists A \in F: B_1 = B_A.$$

Теперь заметим, что множество B эквивалентно множеству всех бесконечных двоичных дробей с нулевой целой частью, то есть, множеству $[0, 1]$. Соответствующую элементу $Q \in B$ бесконечную дробь обозначим как R_Q и просто прямо построим по следующему правилу:

$$R_Q = 0.q_1q_2q_3q_4q_5 \dots, \text{ где } q_i \text{ — } i\text{-тый разряд } Q.$$

Отсюда следует, что B , равно как и F , имеет мощность континуум.

Зададим вероятности событий $\{\omega_i\}$ числами p_i , $i \in \mathbb{N}$, чтобы было верным

$$\sum_{i=1}^{\infty} p_i = 1, \quad p_i \geq 0 \quad \forall i \in \mathbb{N}.$$

События эти плюс пустое попарно несовместны и любое другое событие можно представить как их счётное объединение. Вероятность Ω , которое равно счётному объединению всех $\{\omega_i\}$, равна единице, вероятность \emptyset , не содержащего никого из $\{\omega_i\}$, равна нулю. Любое событие представить как

$$A = \bigcup_{i=1}^{\infty} (\{\omega_i\} \cap I_i) = \bigcup_{i:\omega_i \in A} \{\omega_i\}, \text{ где } I_i = \begin{cases} \Omega, & \text{если } \omega_i \in A, \\ \emptyset, & \text{если } \omega_i \notin A. \end{cases}$$

События $\{\omega_i\} \cap I_i$ по-прежнему попарно несовместны, поскольку $\{\omega_i\} \cap I_i \subseteq \{\omega_i\}$, и можно воспользоваться свойством счётной аддитивности вероятности:

$$P(A) = \sum_{i=1}^{\infty} P(\{\omega_i\} \cap I_i) = \sum_{i=1}^{\infty} p_i b_A^i.$$

Полученный ряд сходится, поскольку сходится ряд из p_i , а $p_i b_A^i \leq p_i$ и оба ряда неотрицательны. Тогда для любого события вероятность определена. \square

Определение 2.8. Пусть Ω — произвольное непустое множество, F — σ -алгебра событий на Ω и P — вероятность, определённая на F . В этих терминах тройка (Ω, F, P) называется **вероятностным пространством**.

2.1.2 Примеры вероятностных пространств

Абстракцию требуется проиллюстрировать примерами.

Конечное число элементарных исходов

Простейший пример — подбрасывание правильного игрального кубика. Он имеет шесть граней, и на каждую из них после броска приземляется с той же вероятностью, что и на любую другую. Элементарных исходов шесть: $\Omega = \{1, 2, 3, 4, 5, 6\}$, где цифрой обозначается число на обращённой вверх после броска грани. Ω — пространство элементарных исходов.

Займемся сигма-алгеброй событий. Выпадение любой грани — это событие. То есть, $A_1 = \{1\}, A_2 = \{2\}, A_3 = \{3\}, A_4 = \{4\}, A_5 = \{5\}, A_6 = \{6\}$ — всё это события. Кроме того, событием будет само Ω . Обратимся к свойствам сигма-алгебры. В ней с любым событием должно содержаться его отрицание. Отрицание Ω — это невозможное событие \emptyset . Отрицание A_j ($j \in \mathbb{Z}, 1 \leq j \leq 6$) можно представить так:

$$\overline{A_j} = \bigcup_{\substack{k=1 \\ k \neq j}}^6 A_k.$$

Обратим внимание на то, что если у нас есть счётная аддитивность и невозможное событие, то можно получить конечную аддитивность, просто положив все хвостовые события невозможными. То есть, если F — σ -алгебра, то в ней содержится невозможное событие (как отрицание достоверного) и есть свойство счётной аддитивности. Значит,

$$A, B \in F \implies A \cup B \cup \emptyset \cup \emptyset \dots \in F \text{ — это счётная аддитивность.}$$

↓

$$A \cup B \in F, \text{ поскольку } A \cup B = A \cup B \cup \emptyset \cup \emptyset \dots$$

Поэтому вышеупомянутые события $A_1, A_2, A_3, A_4, A_5, A_6$ можно как угодно объединять. Полезно заметить, что они попарно несовместны. Дальше то, что получилось, можно как угодно пересекать. Событие «выпало чётное число» представимо в виде $A_2 \cup A_4 \cup A_6$, событие «не единица» представляется как $A_2 \cup A_3 \cup A_4 \cup A_5 \cup A_6$. Нетрудно обнаружить, что σ -алгебра событий в данном случае представляет собой множество всех подмножеств Ω .

Раз кубик правильный, то вероятность выпадения каждой грани равна $1/6$. То есть, $P(A_j) = 1/6$ ($j \in \mathbb{Z}, 1 \leq j \leq 6$). Вероятность вводится как классическая, для любого события её несложно посчитать и тем самым она полностью определена.

Другой пример вероятностного пространства — игра в рулетку. Пространство элементарных исходов — все то, что может выпасть. В честном случае все эти элементарные исходы равновероятны, σ -алгебра событий и вероятность вводятся аналогично случаю с кубиком, но здесь интерес повыше, потому что появляются новые события в духе «красное» и так далее.

Счётное число элементарных исходов

Более сложный пример — это игра Ани и Бори. Речь идет о Петербургском парадоксе. Здесь пространство элементарных исходов сложнее. Выпадение герба будем обозначать «Г», выпадение решки — буквой «Р». Цепочку подбрасываний будем

изображать в виде последовательности этих букв. При этом на выпадении первого герба игра останавливается. Так что герб, в общем-то, всегда стоит в конце цепочки, за ним ничего уже нет. Так что пространство элементарных исходов получается таким:

$$\Omega = \{\Gamma, \text{P}\Gamma, \text{PP}\Gamma, \text{PPP}\Gamma, \text{PPPP}\Gamma, \text{PPPPPP}\Gamma, \text{PPPPPPPP}\Gamma, \dots\}$$

Ω в данном случае счётное, поэтому по утверждению на с.10 σ -алгебру событий можно ввести как множество всех подмножеств Ω .

Обозначим событие «герб выпал при j -том подбрасывании» как A_j . События A_i, A_j ($i, j \in \mathbb{N}$) при $i \neq j$ несовместны. Каждое из них состоит из одного элементарного исхода из Ω . Для симметричной монетки $P(A_1) = 1/2$. Если подбрасывания никак друг от друга не зависят, то $P(A_2) = 1/4$. Это просто получить, если рассмотреть все возможные исходы двух независимых подбрасываний правильной монетки (и отбросить вариант $\Gamma\Gamma$ как соответствующий выпадению герба на первом шаге). Так же получается, что $P(A_j) = 1/2^j$ ($j \in \mathbb{N}$). Любое событие можно представить с помощью операций объединения и отрицания из $\{A_i\}_1^\infty$, всё проводится точно так же, как в доказательстве утверждения на с.10. Значит, вероятность определена для любого события.

Континуум элементарных исходов

Теперь рассмотрим пока что новую ситуацию. Пусть некто бросает точку в отрезок от нуля до одного, причём делает это так, что все точки отрезка равноправны. Некто может попасть как в точку $1/2$, так и в $1/e$ или $1/\pi$. Пространство элементарных исходов теперь имеет мощность континуум. Мы договорились о том, что все точки равноправны. Чему тогда равна вероятность попадания в конкретную точку? Никакой ненулевой конечной величине она не может равняться, иначе сумма вероятностей попаданий в точки последовательности $x_n = 1/n$ (а разных точек в этой последовательности счётное число) даст бесконечность. Но она никак не может превысить единицы. Значит, вероятность попадания в конкретную точку точно равна нулю. Никаких «мало» и «ужасно мало». С этой трудностью призвана бороться построенная особым образом σ -алгебра.

Итак, $\Omega = [0, 1]$.

На с.9 уже говорилось о том, что в множестве всех подмножеств $[0, 1]$ есть неизмеримые множества. Дальше речь пойдет об одном из них. Идея в том, что полуинтервал $[0, 1)$ можно разбить на счётное количество непересекающихся множеств, мера которых совпала бы, если бы существовала. Но предположение о её существовании приводит к противоречию.

Возьмём в множестве $[0, 1)$ любую точку. Выберем из него все её рациональные сдвиги (то есть, все точки, что получаются из неё прибавлением любого рационального числа), которые попали в $[0, 1)$. Изымем выбранные точки из $[0, 1)$ в множество W . Заметим, что так как изъято счётное количество точек, то от $[0, 1)$ остался еще континуум точек. Возьмём из остатка другую точку, выберем все её рациональные сдвиги из $[0, 1)$. Получим множество U . $W \cap U = \emptyset$, поскольку если это не так, то они совпали бы. А мы изначально для U выбрали точку, которой нет в W . В множестве $[0, 1) \setminus W \setminus U$ по-прежнему континуум точек (потому что W и U счётные). По этому принципу будем строить всё дальше и дальше такие же множества. Парно они не пересекаются, все счётные.

Весь полуинтервал $[0, 1)$ разбивается на эти множества (потому что для любой точки из него мы сможем указать множество, построенное по указанному выше способу, в которое эта точка попадёт). Этих множеств — континуум, иначе их объединение было бы счётным либо мощнее континуума, то есть тогда в сумме не получилось бы $[0, 1)$.

Теперь из каждого из полученных множеств возьмём по точке и положим в множество V_0 . Оно целиком содержится в исходном полуинтервале. Все рациональные сдвиги этого множества, лежащие в $[0, 1)$, образуют весь этот полуинтервал (потому что все рациональные сдвиги этого V_0 — это все множества из абзаца выше). Обозначим за V_q множество, полученное из V_0 сдвигом (циклическим, то есть, число, которое «уплыло» при сдвиге за единицу, появляется со стороны нуля) на рациональное число q . Множеств V_q всего счётное число (как рациональных чисел разных). Их объединение даёт весь полуинтервал:

$$\bigcup_{q \in \mathbb{Q}} V_q = [0, 1).$$

Отметим, что $\forall p, q \in \mathbb{Q}: p \neq q \implies V_p \cap V_q = \emptyset$. Теперь вспоминаем свойство счётной аддитивности вероятности. Мера точно так же обладает этим свойством. Значит,

$$P\left(\bigcup_{q \in \mathbb{Q}} V_q\right) = \sum_{q \in \mathbb{Q}} P(V_q).$$

Осталось обнаружить, что $\forall p, q \in \mathbb{Q} P(V_p) = P(V_q)$. Меры двух множеств, одно из которых получено сдвигом другого, равны. Значит, $\forall q \in \mathbb{Q} P(V_q) = P(V_0)$. Последний шаг:

$$P\left(\bigcup_{q \in \mathbb{Q}} V_q\right) = \sum_{q \in \mathbb{Q}} P(V_q) = \sum_{q \in \mathbb{Q}} P(V_0) = P([0, 1)) = 1.$$

Нетрудно заметить, никакое вещественное $P(V_0)$ не может сделать это равенство верным. Мы пришли к противоречию, тянувшемуся с предположения о том, что V_0 имеет меру. Значит, у V_0 меры нет.

Всё это было сделано для того, чтобы проиллюстрировать, что в множестве всех подмножеств $[0, 1]$ водятся неизмеримые типы. Для данного Ω используется σ -алгебра, порождённая пересечением всех интервалов вещественной оси с $[0, 1]$.

Вероятность на такой σ -алгебре вводится как мера Лебега элемента этого множества как множества на прямой. Вот и всё. Подробно об этом речь начинается на с.58.

2.1.3 Свойства вероятности

Все свойства далее обсуждаются в предположении, что мы работаем с каким-то вероятностным пространством. То есть, заданы пространство элементарных исходов Ω , σ -алгебра событий F и вероятность P .

1°. *Вероятность невозможного события равна нулю.*

$$P(\emptyset) \equiv 0.$$

Доказательство. Идентичное утверждение доказано на с.9. □

2° (Свойство конечной аддитивности). Пусть задана σ -алгебра событий F и в ней конечная последовательность $A_1, A_2, A_3, \dots, A_n$, в которой попарно несовместны события-члены. Тогда вероятность объединения всех этих событий — это сумма вероятностей событий-членов, т.е.:

$$\begin{aligned} \{A_i\}_1^n \in F; i, j \in \mathbb{N}, i, j \leq n, i \neq j &\implies A_i A_j = \emptyset \\ &\Downarrow \\ P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i). \end{aligned}$$

Доказательство. Для доказательства просто рассмотрим специально составленную последовательность событий, а затем применим свойство счётной аддитивности вероятности и факт равенства нулю вероятности невозможного события:

$$\begin{aligned} A_1, A_2, A_3, \dots, A_n, \emptyset, \emptyset, \emptyset, \dots &\text{ — есть попарная несовместность,} \\ &\Downarrow \\ P\bigcup_{i=1}^n A_i &= P\left(\bigcup_{i=1}^n A_i \cup \bigcup_{i=n}^{\infty} \emptyset\right) = P(A_1) + P(A_2) + \dots + P(A_n) + P(\emptyset) + P(\emptyset) + \dots \\ &\stackrel{1^\circ}{=} P(A_1) + P(A_2) + P(A_3) + \dots + P(A_n). \end{aligned} \quad \square$$

3°. Вероятность отрицания события — это единица минус вероятность самого события:

$$P(\bar{A}) \equiv 1 - P(A).$$

Доказательство.

$$\begin{aligned} \Omega = A \cup \bar{A}, A \cap \bar{A} = \emptyset &\stackrel{2^\circ}{\implies} P(A) + P(\bar{A}) = P(\Omega) = 1 \\ &\Downarrow \\ P(\bar{A}) &= 1 - P(A). \end{aligned} \quad \square$$

4° (Теорема сложения). Пусть A и B — произвольные события, тогда

$$P(A \cup B) \equiv P(A) + P(B) - P(AB).$$

Доказательство. Разложим A и B следующим образом: $A = AB \cup A\bar{B}$, события AB и $A\bar{B}$ несовместны; $B = AB \cup \bar{A}B$, события AB и $\bar{A}B$ — тоже несовместны. Несовместны и $A\bar{B}$ с $\bar{A}B$. Поэтому верно следующее:

$$P(A) = P(AB) + P(A\bar{B}), P(B) = P(AB) + P(\bar{A}B). \quad (C)$$

$$\text{Заметим, что } A \cup B = AB \cup A\bar{B} \cup \bar{A}B$$

$$\begin{aligned} &\Downarrow \\ P(A \cup B) &= P(AB) + P(\bar{A}B) + P(A\bar{B}) \stackrel{C}{=} P(AB) + P(B) - P(AB) + P(A) - P(AB) \\ &= P(A) + P(B) - P(AB). \end{aligned} \quad \square$$

5° (Формула включения-исключения). В случае более, чем двух событий теорема сложения принимает следующий вид: $\forall A_1, \dots, A_n \in F$

$$P\left(\bigcup_{i=1}^n A_i\right) = (-1)^0 \sum_{i=1}^n P(A_i) + (-1)^1 \sum_{\substack{i,j=1 \\ i < j}}^n P(A_i A_j) + \dots + (-1)^{n-1} P\left(\bigcap_{i=1}^n A_i\right).$$

Доказательство. Нужно применить метод математической индукции. В качестве базы индукции выступает теорема сложения. В качестве утверждения индукции — утверждение теоремы. А вот индукционный переход придётся провести, в переходе * используется теорема сложения: $A_i \in F, 1 \leq i \leq n+1$,

$$\begin{aligned} P\left(\bigcup_{i=1}^{n+1} A_i\right) &= P\left(A_{n+1} \cup \bigcup_{i=1}^n A_i\right) \stackrel{*}{=} P(A_{n+1}) + P\left(\bigcup_{i=1}^n A_i\right) - P\left(A_{n+1} \bigcup_{i=1}^n A_i\right) \\ &= P(A_{n+1}) + P\left(\bigcup_{i=1}^n A_i\right) - P\left(\bigcup_{i=1}^n (A_i A_{n+1})\right) \end{aligned}$$

Выражение для $n+1$ полностью сведено к выражениям для n . Далее предстоит техническая работа — простое раскрытие вероятностей по утверждению индукции и группировка вероятностей пересечений двух событий, трёх и так далее. В результате чего получится утверждение индукции, но уже для $n+1$. Так что переход индукции осуществлён, а утверждение доказано. \square

6° (Монотонность вероятности). Пусть A и B — некоторые события, причём B содержит в себе A . Тогда вероятность A не превышает вероятности B :

$$A \subseteq B \implies P(A) \leq P(B).$$

Доказательство.

$$B \setminus A = B \cap \bar{A}, A \cap (B \cap \bar{A}) = A \cap \bar{A} \cap B = \emptyset \cap B = \emptyset, B = A \cup B \setminus A$$

\Downarrow

$$P(B) = P(A) + P(B \setminus A) \geq P(A) \text{ в силу неотрицательности вероятности. } \square$$

Далее в выкладках если в пересечении или в объединении верхний предел меньше нижнего, то такое пересечение следует считать равным Ω , а объединение — равным \emptyset . То есть,

$$\bigcup_{i=3}^2 = \emptyset, \bigcap_{i=6}^5 = \Omega.$$

7° (Счётная полуаддитивность). Пусть бесконечная последовательность $\{A_i\}$ целиком состоит из событий. Тогда вероятность счётного объединения событий-членов не превышает суммы вероятностей событий (если она вообще есть). То есть,

$$\{A_i\}_1^\infty \in F \implies P\left(\bigcup_{i=1}^\infty A_i\right) \leq \sum_{i=1}^\infty P(A_i).$$

Замечание. Здесь никакой несовместности нету. Тут просто задана последовательность событий, однако и утверждение здесь послабей будет.

Доказательство. Построим новую последовательность попарно несовместных событий на основе данной в условии последовательности, положив предварительно $A_0 = \emptyset$:

$$\begin{aligned}
 D_1 &= A_1, D_2 = A_2 \setminus A_1, D_m = A_m \setminus \bigcup_{i=1}^{m-1} A_i, \quad i, j \in \mathbb{N}, \quad i \neq j \quad (\text{пусть } i > j) \\
 &\Downarrow \\
 D_i D_j &= A_i \cap \bigcup_{k=1}^{i-1} \overline{A_k} \cap A_j \cap \bigcup_{k=1}^{j-1} \overline{A_k} = A_i \cap A_j \cap \bigcap_{k=1}^{i-1} \overline{A_k} \cap \bigcap_{k=1}^{j-1} \overline{A_k} \\
 &= A_i \cap A_j \cap \bigcap_{k=1}^{i-1} \overline{A_k} = A_i \cap A_j \cap \overline{A_j} \cap \bigcap_{k=1}^{j-1} \overline{A_k} \cap \bigcap_{k=j+1}^{i-1} \overline{A_k} = A_i \cap \emptyset \cap \bigcap_{k=1}^{i-1} \overline{A_k} \\
 &= \emptyset \text{ — в } \{D_i\} \text{ события попарно несовместны.}
 \end{aligned}$$

Теперь заметим, что

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} D_i$$

и воспользуемся для $\{D_i\}$ счётной аддитивностью:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{i=1}^{\infty} D_i\right) = \sum_{i=1}^{\infty} P(D_i) \leq \sum_{i=1}^{\infty} P(A_i). \quad \square$$

Определение 2.9. Последовательность событий $\{A_i\}$ — *неубывающая*

$$\begin{aligned} &\Updownarrow \\ A_1 \subseteq A_2 \subseteq A_3 \subseteq A_4 \subseteq A_5 \subseteq \dots \subseteq A_i \subseteq A_{i+1} \subseteq \dots \quad (\forall i \in \mathbb{N} \quad A_i \subseteq A_{i+1}) \end{aligned}$$

Определение 2.10. Последовательность событий $\{A_i\}$ — *невозрастающая*

$$\begin{aligned} &\Updownarrow \\ A_1 \supseteq A_2 \supseteq A_3 \supseteq A_4 \supseteq A_5 \supseteq \dots \supseteq A_i \supseteq A_{i+1} \supseteq \dots \quad (\forall i \in \mathbb{N} \quad A_i \supseteq A_{i+1}) \end{aligned}$$

Определение 2.11. Событие Q называется *пределом неубывающей последовательности* $\{A_i\}$ (обозначается $Q = \lim_{i \rightarrow \infty} A_i$)

$$\begin{aligned} &\Updownarrow \\ Q &= \bigcup_{i=1}^{\infty} A_i. \end{aligned}$$

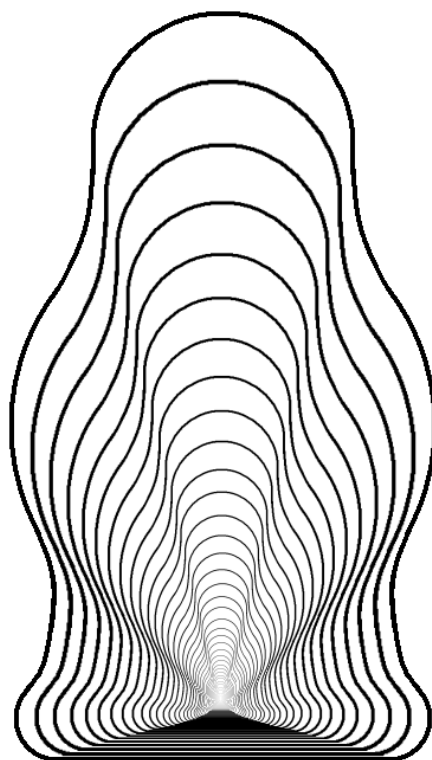
Определение 2.12. Событие Q называется *пределом невозрастающей последовательности* $\{A_i\}$ (обозначается $Q = \lim_{i \rightarrow \infty} A_i$)

$$\begin{aligned} &\Updownarrow \\ Q &= \bigcap_{i=1}^{\infty} A_i. \end{aligned}$$

Определение 2.13. Последовательность событий $\{A_i\}$ — *монотонная*, если она является невозрастающей либо неубывающей.

В обоих случаях монотонности тот факт, что Q — событие, следует из свойства счётной аддитивности σ -алгебры. Ещё в обоих случаях монотонности наглядной иллюстрацией может послужить цепь матрёшек! Каждая точка вложенной матрёшки принадлежит множеству точек объемлющей (точки внутри полой фигуры здесь тоже считаются принадлежащими ей). Так что ничего мистического в этих последовательностях на самом деле нет.

Вообще, здесь идёт речь о простых вещах длинным научным языком. Для облегчения понимания рекомендуется пользоваться сначала кругами Эйлера, а затем аналитическим аппаратом теории множеств. На самом деле, всё то, что пойдёт дальше, весьма и весьма просто и наглядно, если всякое неясное место иллюстрировать себе таким способом.



8° (Непрерывность по монотонной последовательности). Пусть $\{A_i\}$ — монотонная последовательность событий. Тогда вероятность предела $\{A_i\}$ равна пределу вероятности события A_i . Иными словами,

$\{A_i\}$ — монотонная последовательность событий

↓

$$P\left(\lim_{i \rightarrow \infty} A_i\right) = \lim_{i \rightarrow \infty} P(A_i)$$

Доказательство. Для неубывающей последовательности событий построим последовательность несовместных событий на базе данной в условии последовательности и для неё воспользуемся свойством счётной аддитивности:

$$D_n = A_n \setminus \bigcup_{i=1}^{n-1} A_i = A_n \setminus A_{n-1}, \text{ что следует из неубывания } \{A_i\}$$

$$\bigcup_{i=1}^n A_i = \bigcup_{i=1}^n D_i \quad \forall n \in \mathbb{N} \text{ — это наглядно и почти очевидно,}$$

↓

$$P\left(\lim_{n \rightarrow \infty} A_n\right) = \sum_{n=1}^{\infty} P(D_n) = \lim_{n \rightarrow \infty} \sum_{i=1}^n P(D_i) = \lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n D_i\right) = \lim_{n \rightarrow \infty} P(A_n).$$

Факт попарной несовместности событий-членов $\{D_i\}$ обоснован в процессе доказательства свойства счётной полуаддитивности.

Для невозрастающей последовательности заметим, что если $\{B_i\}$ — невозрастающая последовательность, то $\{\overline{B_i}\}$ — неубывающая. А поскольку известен простой и очень важный факт о том, что

$$\bigcap_{i=1}^n B_i \equiv \overline{\bigcup_{i=1}^n \overline{B_i}},$$

то справедливо соотношение

$$Q_B = \lim_{n \rightarrow \infty} B_n = \bigcap_{n=1}^{\infty} B_n = \overline{\bigcup_{n=1}^{\infty} \overline{B_n}} = \overline{\lim_{n \rightarrow \infty} \overline{B_n}} = \overline{Q_{\overline{B}}}$$

Всё это означает, что мы можем с помощью операции отрицания из невозрастающей последовательности событий $\{B_n\}$ сделать неубывающую $\{\overline{B_n}\}$, а для неё уже доказано, что

$$P\left(\lim_{n \rightarrow \infty} \overline{B_n}\right) = \lim_{n \rightarrow \infty} P(\overline{B_n}).$$

Остаётся провести маневр с отрицанием:

$$1 - \lim_{n \rightarrow \infty} P(B_n) = \lim_{n \rightarrow \infty} P(\overline{B_n}) = P\left(\lim_{n \rightarrow \infty} \overline{B_n}\right) = P\left(\overline{\lim_{n \rightarrow \infty} B_n}\right) = 1 - P\left(\lim_{n \rightarrow \infty} B_n\right)$$

↓

$$\lim_{n \rightarrow \infty} P(B_n) = P\left(\lim_{n \rightarrow \infty} B_n\right) \text{ теперь уже для невозрастающей последовательности. } \square$$

Определение 2.14. *Невозрастающая последовательность событий $\{A_n\}$ имеет нулевой предел $\iff \lim_{n \rightarrow \infty} A_n = \emptyset$.*

Замечание. Вместо аксиомы о счётной аддитивности вероятности можно требовать выполнения конечной аддитивности и непрерывности по неубывающей последовательности либо непрерывности по невозрастающей последовательности или даже по неубывающей последовательности с нулевым пределом. Поскольку выше мы показали, что из счётной аддитивности можно получить непрерывность по монотонным последовательностям событий, то доказав ниже обратное (плюс случай невозрастающей последовательности с нулевым пределом) сможем утверждать, что все эти вещи равносильны.

Утверждение 2.3. *Из конечной аддитивности и непрерывности по неубывающей последовательности следует счётная аддитивность.*

Доказательство. Пусть задана бесконечная последовательность событий $\{B_i\}$, причём события в ней попарно несовместны. Запишем условие конечной аддитивности:

$$i, j \in \mathbb{N}, i \neq j \implies B_i B_j = \emptyset \text{ — несовместность}$$

↓

$$P\left(\bigcup_{i=1}^n B_i\right) = \sum_{i=1}^n P(B_i) \quad \forall n \in \mathbb{N}$$

Построим неубывающую последовательность $\{A_n\}$ следующим образом:

$$A_n = \bigcup_{i=1}^n B_i.$$

Для этой последовательности работает непрерывность и, кроме того,

$$\bigcup_{i=1}^n B_i = \bigcup_{i=1}^n A_i \quad \forall n \in \mathbb{N}.$$

Всё, почва для выкладок подготовлена. Выкладки:

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(\lim_{n \rightarrow \infty} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{P}(B_i) = \sum_{i=1}^{\infty} \mathbb{P}(B_i). \quad \square$$

Утверждение 2.4. При наличии конечной аддитивности из непрерывности по невозрастающей последовательности следует непрерывность по неубывающей последовательности и наоборот.

Доказательство. Тесная связь между этими двумя свойствами уже демонстрировалась. Продемонстрируем её ещё раз несколькими равносильными переходами:

Пусть $\{B_i\}$ — любая невозрастающая последовательность событий и есть непрерывность по невозрастающей последовательности:

$$\forall k \in \mathbb{N} B_k \supseteq B_{k+1}, \mathbb{P}\left(\lim_{n \rightarrow \infty} B_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(B_n);$$

Пусть $\{A_i\}$ — любая неубывающая последовательность. Рассмотрим определение предела для последовательности $\{A_i\}$:

$$\lim_{n \rightarrow \infty} A_n = \bigcup_{i=1}^{\infty} A_i = \overline{\bigcap_{i=1}^{\infty} \overline{A_i}} = \overline{\lim_{n \rightarrow \infty} \overline{A_n}}$$

Последовательность $\{\overline{A_i}\}$ невозрастает, поэтому для неё бесконечное пересечение — это и есть предел, а для невозрастающей последовательности есть непрерывность

$$\mathbb{P}\left(\overline{\bigcap_{i=1}^{\infty} \overline{A_i}}\right) = 1 - \mathbb{P}\left(\bigcap_{i=1}^{\infty} \overline{A_i}\right) = 1 - \lim_{n \rightarrow \infty} \mathbb{P}(\overline{A_n}).$$

Всё, остаётся собрать это всё вместе:

$$\begin{aligned} \mathbb{P}\left(\lim_{n \rightarrow \infty} A_n\right) &= \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(\overline{\bigcap_{i=1}^{\infty} \overline{A_i}}\right) = 1 - \mathbb{P}\left(\bigcap_{i=1}^{\infty} \overline{A_i}\right) = 1 - \lim_{n \rightarrow \infty} \mathbb{P}(\overline{A_n}) \\ &= \lim_{n \rightarrow \infty} (1 - \mathbb{P}(\overline{A_n})) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n). \end{aligned}$$

Доказательство в обратную сторону проводится точно так же. Тут все переходы равносильные, нужно только отрицания иначе расставить. Конечная аддитивность нужна здесь неявно, чтобы было обоснованным равенство $\mathbb{P}(A) + \mathbb{P}(\overline{A}) = 1$. \square

Замечание. Последний установленный факт позволяет заявить, что из непрерывности по невозрастающей последовательности событий и конечной аддитивности следует счётная аддитивность.

Замечание. Поскольку из счётной аддитивности следует непрерывность по невозрастающей последовательности, то непрерывность по невозрастающей последовательности с нулевым пределом следует подавно. Осталось доказать обратное.

Утверждение 2.5. Из конечной аддитивности и непрерывности по невозрастающей последовательности событий с нулевым пределом следует счётная аддитивность.

Доказательство. Пусть есть бесконечная последовательность событий $\{A_i\}$, и события-члены попарно несовместны. Событие

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^n A_i \cup \bigcup_{i=n+1}^{\infty} A_i \quad \forall n \in \mathbb{N}$$

разбили на два события, тоже несовместных. Далее воспользуемся конечной аддитивностью вероятности, которая есть по условию:

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} A_i\right) &= P\left(\bigcup_{i=1}^n A_i\right) + P\left(\bigcup_{i=n+1}^{\infty} A_i\right) \\ &\quad \downarrow \\ P\left(\bigcup_{i=1}^{\infty} A_i\right) - P\left(\bigcup_{i=n+1}^{\infty} A_i\right) &= P\left(\bigcup_{i=1}^n A_i\right) \quad \forall n \in \mathbb{N}, \end{aligned} \tag{A}$$

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) \quad \forall n \in \mathbb{N}, \tag{B}$$

а теперь, воспользовавшись непрерывностью по невозрастающей последовательности событий $\sum_{i=n+1}^{\infty} A_i$ с нулевым пределом (переход помечен *), получим желанный факт:

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} A_i\right) &= P\left(\bigcup_{i=1}^{\infty} A_i\right) - 0 \stackrel{*}{=} P\left(\bigcup_{i=1}^{\infty} A_i\right) - \lim_{n \rightarrow \infty} P\left(\bigcup_{i=n+1}^{\infty} A_i\right) \\ &= \lim_{n \rightarrow \infty} \left(P\left(\bigcup_{i=1}^{\infty} A_i\right) - P\left(\bigcup_{i=n+1}^{\infty} A_i\right) \right) \stackrel{A}{=} \lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n A_i\right) \\ &\stackrel{B}{=} \lim_{n \rightarrow \infty} \sum_{i=1}^n P(A_i) = \sum_{i=1}^{\infty} P(A_i). \end{aligned} \quad \square$$

Так что получается, что если отображение P из σ -алгебры событий F в \mathbb{R} , такое что $\forall A \in F \ P(A) \geq 0$, $P(\Omega) = 1$ и P — конечно-аддитивное, то следующие три свойства эквивалентны:

- P — счётно-аддитивное;
- P — непрерывно по любой монотонной последовательности событий;
- P — непрерывно по невозрастающей последовательности с нулевым пределом.

Лекция 3

3.1 Урновые схемы

Далее будут разобраны два важных примера конечных вероятностных пространств.

3.1.1 Выборка с возвращением, биномиальное распределение

Пусть имеется урна, в которой находится $m_w > 0$ белых шаров и $m_b > 0$ чёрных, $m = m_w + m_b$. Причём урна не простая, а особенная: всякий раз, как из неё берут шар, что-то с ним делают и кладут обратно, в ней происходит процесс, обеспечивающий независимость того, какой шар вынут, от того, какой вынули (и вернули обратно) в любой момент до этого.

Пусть экспериментатор берёт из этой урны один шар, каким-то образом фиксирует его цвет и кладёт обратно. Пусть это действие происходит n раз. Требуется выяснить, какова вероятность того, что белый цвет экспериментатор зафиксирует k раз.

Представим себе все возможные результаты эксперимента как 2^n возможных цепочек результатов (длины n , как легко видеть). В каждой позиции цепочки может стоять либо «Ч», либо «Б», символизирующие соответствующие цвета. Так вот взять все эти цепочки, среди них выделить те, в которых белый шар попался нужное число раз, посчитать их, поделить это на 2^n и выдать за ответ — заведомо неправильно. При таком ходе мыслей неявно используем классическую вероятность, а для её применения требуется равновозможность элементарных исходов. А цепочки результатов могут быть неравновозможными (нигде ведь не сказано, что $m_b = m_w$).

Сделаем все шары уникальными. Проще всего их для этого перенумеровать числами от одного до n . Теперь цепочки, характеризующие результат вытаскиваний шариков, состоят из чисел. Номеров, попавших на белые шары — m_w штук, на чёрные шары попадёт m_b номеров.

Всего цепочек теперь m^n . Есть n позиций, в каждой из которых может находиться любое из m чисел. Теперь цепочки равновероятны! У всех шаров шанс оказаться снаружи урны один и тот же. Вероятность любой конкретной цепочки равна $1/m^n$.

Пересчитаем цепочки, в которых на первых k позициях оказались числа с белых шаров, а на оставшихся — числа с чёрных шаров. На каждой из k «белых» позиций может стоять любое из m_w чисел, точно так же, на каждой из $n - k$ «чёрных» позиций может стоять любое из m_b чисел. Значит, таких цепочек всего $m_w^k m_b^{n-k}$.

Теперь заметим, что цепочек, в которых просто (безо всякого порядка) k позиций с «белыми» номерами и $n - k$ позиций с «чёрными» номерами ровно столько же, сколько цепочек, где k первых номеров — «белые», а оставшиеся — чёрные. Как там между собой расположены «белые» и «чёрные» позиции, совершенно неважно. Важно количество!

Остаётся выяснить, сколькими различными способами можно из n позиций выбрать k «белых» (всё равно, что выбрать $n - k$ «чёрных»). Как известно, ответ на

этот вопрос даёт формулой

$$\frac{n!}{k!(n-k)!} = C_n^k.$$

Просто число сочетаний из n по k .

Теперь осталось вспомнить, что всего у нас исходов — m^n штук. Из них нашему событию подходит $C_n^k m_w^k m_b^{n-k}$. Классическую вероятность можно смело применять, так что в итоге имеем

$$P(\text{Белых} - \text{ровно } k \text{ штук}) = C_n^k \frac{m_w^k m_b^{n-k}}{m^n} = C_n^k \frac{m_w^k}{m^k} \frac{m_b^{n-k}}{m^{n-k}} = C_n^k \left(\frac{m_w}{m}\right)^k \left(1 - \frac{m_w}{m}\right)^{n-k}$$

Теперь обозначим m_w/m за p , $P(\text{Белых} - \text{ровно } k \text{ штук})$ как p_k и получим выражение

$$p_k = C_n^k p^k (1-p)^{n-k}, \quad 0 \leq k \leq n.$$

Определение 3.1. Набор вероятностей (p_0, p_1, \dots, p_n) называется **геометрическим распределением** с параметрами n и p .

3.1.2 Выборка без возвращения, гипергеометрическое распределение

Пусть теперь экспериментатор просто выбирает из урны n шаров, не возвращая их на место. Спрашивается, какова вероятность вытащить ровно k белых шаров, причём считается, что в урне шаров нужного цвета хватит.

Как это ни печально, но если вновь попытаться представить исходы эксперимента как цепочки из «Ч» и «Б», то ничего вновь не получится. Ситуация осложняется ещё и тем, что не всякую цепочку, вообще говоря, удастся получить. Если в прошлом эксперименте шары возвращались в урну и их всегда хватало, то в этот раз шаров нужного цвета ещё и не хватить может. Так что тут этот подход не годится ещё и потому, что его использование грозит излишними трудностями.

Снова перенумеруем все шары и рассмотрим все цепочки длины n из чисел, которые могут получиться в результате эксперимента. Пусть всего шаров снова m , при этом $m \geq n$, иначе задача исчезает. Пусть из этих m шаров опять $m_w > 0$ белых и $m_b > 0$ чёрных. Опять же, $m_w \geq k$, иначе задача вновь исчезнет.

Посчитаем все возможные цепочки: на первом месте может стоять m чисел, на втором — уже $m-1$. На j -том месте — $m-j+1$. Поскольку n -тое место — последнее, то последним в этом произведении окажется число $m-n+1$. Итого в пространстве равновероятных элементарных исходов насчитывается

$$\|\Omega\| = m \cdot (m-1) \cdot \dots \cdot (m-n+1) = \frac{m!}{(m-n)!} \text{ исходов.}$$

Теперь зафиксируем позиции для белых шаров в цепочке, для первого места из «белых» имеется m_w возможностей, для второго — m_w-1 . Для k -го получается m_w-k+1 , на этом числе произведение остановится. Однако, осталось ещё $n-k$ чёрных мест. Точно таким же образом получаем, что для фиксированных позиций получается вариантов

$$m_w \cdot (m_w-1) \cdot \dots \cdot (m_w-k+1) \cdot m_b \cdot (m_b-1) \cdot \dots \cdot (m_b-(n-k)+1) \text{ вариантов.}$$

При этом выбрать k «белых» позиций можно по-прежнему C_n^k способами. Теперь остаётся собрать все формулы в одну. Обозначим вероятность того, что белых шаров выпало ровно k штук как p_k . Получаем

$$\begin{aligned} p_k &= C_n^k \frac{m_w \cdot (m_w - 1) \cdot \dots \cdot (m_w - k + 1) \cdot m_b \cdot (m_b - 1) \cdot \dots \cdot (m_b - (n - k) + 1)}{m!} \\ &= \frac{n! \frac{m_w!}{(m_w - k)!} \frac{m_b!}{(m_b - n + k)!}}{k!(n - k)! \frac{m!}{(m - n)!}} = \frac{m_w!}{k!(m_w - k)!} \frac{m_b!}{(n - k)!(m_b - n + k)!} = \frac{C_{m_w}^k C_{m_b}^{n-k}}{C_m^n} \\ &= \frac{C_{m_w}^k C_{m - m_w}^{n-k}}{C_m^n}. \end{aligned}$$

Нельзя забывать, что $k \leq m_w$ и $k \leq n$. Однако при этих условиях может оказаться, что $n - k > m_b$, то есть надо вытащить чёрных шаров больше, чем их вообще есть в урне. Поэтому для k получаем две границы: снизу $\max(0, n - m + m_w)$ и сверху $\min(m_w, n)$.

Определение 3.2. Набор вероятностей $(p_t, p_{t+1}, \dots, p_l)$, где $l = \min(m_w, n)$ и $t = \max(0, n - m + m_w)$, называется **гипергеометрическим распределением с параметрами n, m и m_w** .

3.2 Условная вероятность

Пусть у нас имеется два события A и B . Пусть событие B вдруг произошло. И тут потребовалось определить вероятность A ! Та вероятность, что мы вычислим при уже произошедшем B , называется условной вероятностью.

Пример. Экспериментатор подбрасывает монету дважды. Событие A — за эти два броска выпал хотя бы один герб. Событие B — выпала хотя бы раз решка. Точно известно, что B произойдёт. Требуется выяснить, чему при таком раскладе равна $P(A)$. Монетка правильная.

Решение. Поскольку точно известно, что за два подбрасывания решка хоть раз да появится, то элементарных исходов, среди которых нужно искать событие A , остаётся три: РГ, РР, ГР. Они равновероятны. Событие A из этих исходов содержит два. Получаем, пользуясь подходом классической вероятности, что $P(A) = 2/3$.

Сама по себе вероятность A (кстати, всё равно что условная относительно Ω вероятность) равна $3/4$, как и вероятность B . Вероятность события AB равна $1/2$. Возможно, это не просто так случилось, что условная вероятность $P(A|B)$ оказалась равной $P(AB)/P(B)$.

Определение 3.3. Условной вероятностью происхождения события A при уже произошедшем (при условии) B (обязательно $P(B) > 0$) называется величина

$$\frac{P(AB)}{P(B)} = P(A|B) \text{ — это её обозначение.}$$

Замечание. Эта формула не используется практически никогда. Используется ценное следствие: $P(AB) = P(A|B)P(B) = P(B|A)P(A)$, которое имеет некий смысл, если потребовать $P(A) \neq 0 \neq P(B)$.

Определение 3.4. События A и B независимы $\iff P(AB) = P(A)P(B)$.

Замечание. Связь со здравым смыслом и наглядностью, кажется, на этом утеряна. Но, если приглядеться, она есть. Посмотрим, что даёт формула для условной вероятности при независимых событиях:

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A), \quad P(B) \neq 0, P(A) \neq 0.$$

Произошло событие B или не произошло — вероятность события A от этого никак не меняется. То есть, не зависит от события B . Точно так же всё это правильно для независимости вероятности события B от того, что там происходит с событием A :

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

Наглядного смысла можно попробовать найти здесь.

Замечание. Событие с нулевой вероятностью независимо с любым другим событием:

$$\begin{aligned} AB \subseteq B \implies P(AB) \leq P(B), \quad P(B) = 0 \\ \Downarrow \\ P(AB) \leq P(B) = 0 \implies P(AB) = 0 = P(B) = P(B)P(A). \end{aligned}$$

Утверждение 3.1. События A и B независимы $\implies A$ и \bar{B} независимы.

Доказательство. Вспомним, что $A = AB \cup A\bar{B}$, причём пересечение их пустое: $AB \cap A\bar{B} = A \cap A \cap (B \cap \bar{B}) = A \cap \emptyset = \emptyset$. Так что мы разложили событие в два несовместных. Поэтому $P(A) = P(AB) + P(A\bar{B})$, а тогда $P(A) - P(AB) = P(A\bar{B})$. Теперь прямо показываем независимость:

$$P(A)P(\bar{B}) = P(A)(1 - P(B)) = P(A) - P(A)P(B) = P(A) - P(AB) = P(A\bar{B}). \quad \square$$

Замечание. Событие с вероятностью единица независимо с любым другим событием. Просто возьмём его отрицание и получим событие с вероятностью ноль. Оно независимо с любым другим, а значит, независимо с любым другим и его отрицание.

Пример. Пусть есть урна с N перенумерованных шаров (числами от 1 до N). Из неё выбирается шар и фиксируется его номер, причём у всех шаров шансов оказаться снаружи поровну. Первый игрок ставит на множество A , второй — на множество B ($A, B \subset \{1, \dots, N\}$) чисел, пытаясь тем самым угадать номер шара, оказавшегося снаружи урны. Вопрос в том, всегда ли существуют A и B , такие что события «номер шара оказался в A » и «номер шара оказался в B » независимы. Примеры в духе $A = \emptyset$ или $A = \{1, \dots, N\}$ интереса не представляют.

Решение. Требуется отыскать подмножества $\Omega = \{1, \dots, N\}$, непустые и не равные Ω , A и B , такие что $P(AB) = P(A)P(B)$. Поскольку все элементарные исходы тут равновозможны, пользуемся подходом классической вероятности:

$$P(A) = \frac{\|A\|}{N}, \quad P(B) = \frac{\|B\|}{N}, \quad P(A)P(B) = \frac{\|A\|\|B\|}{N^2}.$$

Рассмотрим два случая: N — простое и N — составное.

В первом случае в знаменателе $P(AB)$ обязательно будет N , в числителе — число от нуля до $N - 1$ (в силу того, что $\|A\| < N$). Знаменатель простой и никаких общих делителей с числителем не имеет, дробь несократима. А вот в знаменателе $P(A)P(B)$ окажется N^2 , в числителе — произведение чисел, никаких общих делителей с N вновь не имеющих. Дробь опять несократима. Так что ни при каких A и B добиться равенства $P(AB) = P(A)P(B)$ не удастся.

Во втором случае N составное, значит, его можно разложить на два множителя, каждый из которых не меньше двух.

$$N = a \cdot b, \quad a, b \in \mathbb{N}; \quad a, b \geq 2 \implies \min(a, b) \geq 2; \quad a + b \leq 2 \cdot \max(a, b) \leq ab$$

Разобьём всё множество N на два, одно — из a элементов (его обозначим A), второе — из b элементов (обозначим его B). Как показано выше, $a + b \leq ab = N$. Значит, можно сделать это так, чтобы множества эти пересекались только в одном элементе. Более того, делается это неединственным образом (прибавим ко всем числам из A и B единицу, возьмём остаток от деления на N и затем к тем числам, что получим, прибавим единицу к каждому — получатся отличные от предыдущих наборы чисел, так как $N \geq 4$ — составное число). Значит, получаем $\|AB\| = 1$. Теперь независимость показывается прямо:

$$P(A)P(B) = \frac{\|A\|}{N} \frac{\|B\|}{N} = \frac{ab}{N^2} = \frac{1}{N} = \frac{\|AB\|}{N} = P(AB).$$

Пример хорошо иллюстрирует тот факт, что интуиция в вопросах зависимости или независимости событий может подвести. Вообще, ей не следует злоупотреблять в этом курсе.

Замечание. Если два несовместных события независимы, то у одного из них нулевая вероятность: $0 = P(AB) = P(A)P(B)$. Если два события несовместны и имеют ненулевые вероятности, то они зависимы. Это согласуется с наглядным представлением о зависимости: если уж одно из событий «запрещает» другое, то вероятностная связь между ними есть, а значит, они зависимы.

Утверждение 3.2. Пусть (Ω, F, P) — вероятностное пространство и B — любое событие с ненулевой вероятностью. Тогда (Ω, F, P_B) — вероятностное пространство, где $P_B(A) = P(A|B) \forall A \in F$.

Доказательство. Поскольку изменения коснулись из всей тройки только вероятности, то надо только для неё доказать присутствие трёх свойств:

1. $P_B(A) = P(AB)/P(B) \geq 0$ — неотрицательность есть;
2. $P_B(\Omega) = P(B \cap \Omega)/P(B) = P(B)/P(B) = 1$ — нормировка присутствует;
3. счётная аддитивность тоже на месте:

$$\begin{aligned} A_1, A_2, A_3, \dots \in F, \quad \forall i, j \in \mathbb{N}, i \neq j \quad A_i A_j = \emptyset, \quad B A_i \cap B A_j = B \cap A_i A_j = \emptyset \\ P_B \left(\bigcup_{i=1}^{\infty} A_i \right) = \frac{1}{P(B)} P \left(B \cap \bigcup_{i=1}^{\infty} A_i \right) = \frac{1}{P(B)} P \left(\bigcup_{i=1}^{\infty} B A_i \right) = \frac{1}{P(B)} \sum_{i=1}^{\infty} P(B A_i) \\ = \sum_{i=1}^{\infty} \frac{P(B A_i)}{P(B)} = \sum_{i=1}^{\infty} P_B(A_i). \end{aligned} \quad \square$$

3.3 Независимость множества событий

Определение 3.5. Пусть имеется конечное или счётное множество событий $A = \{A_i\}$. События A_i называются **независимыми или независимыми в совокупности**



Для любого конечного подмножества $B \subseteq A$ выполнено

$$P\left(\bigcap_{i: A_i \in B} A_i\right) = \prod_{i: A_i \in B} P(A_i).$$

Определение 3.6. Пусть имеется конечное или счётное множество событий $A = \{A_i\}$. События A_i называются **попарно независимыми**



$$\forall i, j: i \neq j \quad P(A_i A_j) = P(A_i) P(A_j)$$

В случае, когда событий всего два, независимость в совокупности и попарно — это одно и то же. В случае большего количества событий это неправильно. Это иллюстрируется примером Бернштейна.

3.3.1 Пример Бернштейна

Представим тетраэдр, у которого одна грань покрашена в красный цвет, вторая — в зелёный, третья — в синий, а четвёртая содержит рисунок, на котором присутствуют все три цвета одновременно (лучше без смешивания, в полосочку). Тетраэдр подбросим и зафиксируем цвета, присутствующие на грани, на которую он упал. Событие R состоит в том, что на нижней грани присутствует красный, событие G — в том, что присутствует зелёный, событие B — в том, что там есть синий.

Потребуем от тетраэдра, чтобы падения на любую из четырёх граней были равновероятными. Посчитаем вероятности событий RG , GB , RB и RGB :

$$P(RG) = \frac{\|RG\|}{4} = \frac{1}{4}, P(GB) = \frac{\|GB\|}{4} = \frac{1}{4}, P(RB) = \frac{1}{4}, P(RGB) = \frac{\|RGB\|}{4} = \frac{1}{4}.$$

Теперь соответствующие произведения вероятностей:

$$P(R)P(G) = \frac{\|R\|}{4} \frac{\|G\|}{4} = \frac{2}{4} \cdot \frac{2}{4} = \frac{1}{4}, P(G)P(B) = \frac{1}{4}, P(R)P(B) = \frac{1}{4} \text{ аналогично};$$

$$P(R)P(G)P(B) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}.$$

Как видно, события R , G и B попарно независимы, но вот $P(R)P(G)P(B) \neq P(RGB)$.

Тем самым пример демонстрирует, что попарная независимость и независимость в совокупности — это не одно и то же. Но, разумеется, попарная независимость следует из независимости случайных величин в совокупности. Далее независимость событий возникнет при построении понятия независимости случайных величин (на с.33 и ??). Пока придётся его оставить.

3.4 Формула полной вероятности

Пусть есть 2 студента, которые на момент начала экзамена знают 20 из 25 билетов, причём одни и те же. Требуется установить, у кого больше шансы вытянуть «счастливый» билет (один из 20 выученных): у первого или у второго. У всех билетов шансы оказаться в руках студента одинаковые.

Вероятность вытянуть «счастливый» билет для первого считается очень просто: $P(\text{счастливый}) = 20/25 = 4/5$. Со вторым сложнее: если первому повезло, то для второго «счастливых» билетов останется 19 из 24, а если первому не повезло — 20 из 24. Вопрос легко решается с помощью формулы полной вероятности.

Определение 3.7. Множество событий (конечное или счётное) $Q = \{E_i\}$ называется разбиением пространства элементарных исходов $\Omega \iff$ события из Q попарно несовместны, имеют ненулевые вероятности и $\bigcup_{i:E_i \in Q} E_i = \Omega$.

Утверждение 3.3 (Формула полной вероятности). Пусть Q — разбиение Ω , тогда для любого события A

$$P(A) = \sum_{i:E_i \in Q} P(E_i) P(A|E_i).$$

Доказательство. Вспомним, что любое событие A можно представить как объединение по всему разбиению несовместных событий AE_i , то есть

$$\bigcup_{i:E_i \in Q} AE_i = A \bigcup_{i:E_i \in Q} E_i = A\Omega = A; \quad \forall i, j: E_i, E_j \in Q, i \neq j \quad AE_i AE_j = AE_i E_j = A\emptyset = \emptyset.$$

Значит, по свойству счётной либо конечной аддитивности вероятности имеем

$$P(A) = \sum_{i:E_i \in Q} P(AE_i) = \sum_{i:E_i \in Q} P(E_i) P(A|E_i). \quad \square$$

Утверждение 3.4 (провокация). В аудиторию вносят урну с двумя шарами, каждый из которых может быть либо чёрным, либо белым равновероятно и при вытаскивании одного шара шансы оказаться снаружи у обоих одинаковы

↓

принесли урну с чёрным и белым шарами.

Доказательство. Доказательство опирается на тот факт, что если в урне имеется три шара, каждый из которых имеет те же шансы оказаться снаружи, что и другие, и $P(\text{вытащили белый шар}) = 2/3$, то в ней два белых и один чёрный шар. Легко доказывается от противного.

В той урне, что принесли в аудиторию, может быть два чёрных шара — обозначим это событие как $E_1, P(E_1) = 1/4$, там может быть два белых шара — событие $E_3, P(E_3) = 1/4$ и могут быть чёрный и белый шары — событие $E_2, P(E_2) = 1/2$. События E_1, E_2, E_3 образуют разбиение пространства элементарных исходов $\Omega = \{\text{ЧЧ}, \text{ЧБ}, \text{БЧ}, \text{ББ}\}$.

Теперь возьмём эту урну, забросим в неё белый шар и осуществим процесс, в результате которого при вытаскивании одного шара шансы каждого из трёх шаров оказаться снаружи станут равными. Посчитаем, чему равна вероятность вытащить белый шар:

$$P(B) = \sum_{i=1}^3 P(E_i) P(B|E_i) = \frac{1}{4} \cdot 1 + \frac{1}{2} \cdot \frac{2}{3} + \frac{1}{4} \cdot \frac{1}{3} = \frac{3 + 4 + 1}{12} = \frac{2}{3}$$

Мы получили, что для урны с тремя равноправными шарами вероятность вытащить белый равна $2/3$, значит, в ней два белых шара и один чёрный. А поскольку один белый мы туда подложили, то изначально там были два шара разных цветов. Утверждение доказано. \square

Замечание. Утверждение неверное, это очевидно. Для этого достаточно подложить в урну два чёрных шара и осуществить процесс, в результате которого шансы обоих шаров оказаться снаружи сравниваются при вытаскивании одного. Значит, доказательство тоже неправильное. И вот почему.

Неверен переход от «вероятность вытащить белый равна $2/3$ » к «в ней два белых шара и один чёрный». Тонкость в том, что формула полной вероятности учитывает урны со всеми возможными комбинациями сразу с соответствующими вероятностями. Можно представить себе ряд из урн, в которых реализованы все возможные комбинации шаров. Теперь положим во все из них по белому шару. Как из всех этих урн выбрать ту, из которой будет выбираться в итоге шар?

Формула полной вероятности смешивает их все в одну, учитывая вероятности комбинаций шаров внутри них, а затем из этой «средней» урны тянет шар. Кстати, «средней» урной оказалась урна с чёрным и белым шарами. Здесь урна — средняя, составленная из всех возможных.

А тот, кто вносит урну в аудиторию, — не формула полной вероятности. Он выбрал какую-то одну, с конкретной комбинацией шаров. И совершенно необязательно ему попадётся «средняя». Тут она — конкретная, фиксированная урна.

Мы получили результат $P(B) = 2/3$, но для «средней» урны, а вывод сделали так, будто она конкретная и фиксированная.

Можно изменить условие утверждения: потребовать, чтобы в урнах были два шара одного цвета — белого или чёрного, равновероятные варианты. В этом случае ход мыслей доказательства приводит к тому же «принесли урну с чёрным и белым шарами». Тут ярче видно, что происходит при использовании формулы полной вероятности.

Нельзя вот так взять и подменить реальную конкретную урну средней! И в любых других задачах подмена чего-либо на среднее недопустима (конечно, если нет тому обоснования), потому что обязательно приведёт к неверному результату. И хорошо ещё, если этот неверный результат удаётся так легко обнаружить.

3.5 Формула Байеса

Несложные преобразования формулы условной вероятности в комбинации с формулой полной вероятности позволяют получить формулу Байеса:

Утверждение 3.5 (Формула Байеса). Пусть Q — разбиение Ω , A — некоторое событие, $P(A) > 0$. Тогда $\forall i: E_i \in Q$

$$P(E_i|A) = \frac{P(E_i) P(A|E_i)}{\sum_{i: E_i \in Q} P(E_i) P(A|E_i)}$$

Доказательство.

$$P(E_i|A) = \frac{P(AE_i)}{P(A)} = \frac{P(E_i) P(A|E_i)}{\sum_{i: E_i \in Q} P(E_i) P(A|E_i)}$$

Числитель расписан по определению условной вероятности, знаменатель — по формуле полной вероятности. \square

Определение 3.8. В формуле Байеса вероятности $\{P(E_i)\}$ называются *априорными*, до наступления события A , обычно они известны заранее. Вероятности $\{P(E_i|A)\}$ называются *апостериорными*, после наступления события A , заранее обычно неизвестны.

Замечание. Рассмотрим процесс признания человека страдающим либо не страдающим от редкой болезни. Пусть событие H состоит в том, что человек от данной болезни не страдает, S состоит в том, что человек ею болен, P — в том, что его признали носителем этого заболевания и D — в том, что не признали таковым. Проблема начинается с того, что множества H и D не совпадают. А выливается в то, что для редких болезней $P(H|P)$ получается большим. Иллюстрируется это нехитро: просто применим формулу Байеса. Поскольку болезнь редкая, то $P(S)$ обозначим для наглядности как ε — очень маленькое число, тогда $P(H) = 1 - \varepsilon$ и

$$P(H|P) = \frac{(1 - \varepsilon) P(P|H)}{(1 - \varepsilon) P(P|H) + \varepsilon P(P|S)} = \frac{1}{1 + \frac{\varepsilon}{1 - \varepsilon} \cdot \frac{P(P|S)}{P(P|H)}}.$$

Конечно, если страшно мало $P(P|H)$, то $P(H|P)$ не окажется таким уж большим и $P(S|P)$ таким уж малым, однако положим $\varepsilon = 10^{-6}$, $P(P|S) = 0.999999$ и $P(P|H) = 10^{-5}$, выйдет $P(H|P) = 0.(90) > 0.9$. В реальной жизни попадаются редкие болезни с $P(H|P) > 0.9$. Кстати, вероятность того, что признанный здоровым болен, выражается следующим образом:

$$P(S|D) = \frac{\varepsilon P(D|S)}{(1 - \varepsilon) P(D|H) + \varepsilon P(D|S)} = \frac{1}{1 + \frac{1 - \varepsilon}{\varepsilon} \cdot \frac{P(D|H)}{P(D|S)}}.$$

Производит впечатление маленькой величины. Просто взять предел при $\varepsilon \rightarrow 0$ и посмотреть, где получается ноль, а где один, нельзя, потому что величины $P(D|H)$, $P(D|S)$, $P(P|H)$ и $P(P|S)$ в общем случае зависят от ε . Но вид формул показателен. Роковую роль играет отношение вероятности быть больным к вероятности ошибки $P(P|H)$. Если $P(P|H)$ не превышает ε , то, скорее всего, окажется $P(H|P) \geq 0.5$.

3.6 Дискретные случайные величины, индикаторы событий

Определение случайной величины короткое и простое. Пусть задано пространство элементарных исходов Ω , пока ещё коечное либо счётное:

Определение 3.9. *Случайная величина* — измеримое относительно σ -алгебры событий отображение из Ω в \mathbb{R} .

Замечание. Слово «измеримое» пока не комментируется, потому что в случае конечных или счётных пространств элементарных исходов любое отображение из Ω в \mathbb{R} «измеримо». Это связано с тем, что σ -алгебру событий можно в данном случае брать как множество всех подмножеств Ω . О том, что происходит, если Ω может оказаться мощности континуум можно почитать на с.60.

Замечание. Вообще, если Ω — не более, чем счётное и имеется несколько случайных величин, то можно строить на их базе какие угодно функции (лишь бы соблюдались области определения). Всё равно получится случайная величина. Это обстоятельство неявно используется в разговоре о мат. ожидании, дисперсии и других моментах случайных величин. Там правомерность сложения случайных величин, возведения в степень и смещения на константу считается обоснованной в этом замечании.

Самое простое отображение — это константа. Чем бы ни было Ω , константа — случайная величина.

Определение 3.10. Случайная величина называется *дискретной*, если она принимает не больше счётного числа значений.

При наших условиях на Ω в области определения случайной величины есть только счётное количество элементарных исходов, так что она неминуемо дискретная. Можно это размышление для яркости провести от обратного.

Определение 3.11. Пусть задано множество $A \subseteq \Omega$. *Индикатором I_A множества A называется следующая случайная величина:*

$$I_A(\omega) = \begin{cases} 1, & \text{если } \omega \in A, \\ 0, & \text{если } \omega \notin A. \end{cases}$$

Замечание. Индикатор множества — не всегда случайная величина. Но если множество является событием, то индикатор случайной величиной обязательно будет. Отсюда видно, между прочим, что отображение, принимающее два разных значения, уже может не быть случайной величиной.

С помощью индикаторов можно любую дискретную случайную величину X представить с помощью ряда. Пусть имеется разбиение $\Omega \{A_i\}$, то есть множество попарно несовместных непустых событий (счётное или конечное), объединение которых даёт Ω . Тогда ряд (индексы приведены для счётного разбиения)

$$\sum_{i=1}^{\infty} x_i I_{A_i}(\omega)$$

является дискретной случайной величиной. В силу того, что события из разбиения попарно несовместны, при любом ω во всём ряду все индикаторы кроме одного обратятся в ноль. Так что больше чем счётное количество значений ряд не примет. В случае не более чем счётного Ω любую дискретную случайную величину можно разложить в такой ряд, просто взяв как разбиение события, состоящие из одного элементарного исхода, а как x_i — значения X на соответствующих событиях. В случае более мощного Ω надо брать прообразы множеств $\{x_i\}$. В силу того, что X — случайная величина, они будут событиями, но об этом позже (на с.58).

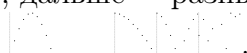
Определение 3.12. *Распределением или распределением вероятности дискретной случайной величины X называется совокупность её значений x_1, x_2, \dots с соответствующим набором вероятностей $p_i = P(X = x_i)$.*

Замечание. Иногда можно считать, что случайная величина принимает ещё континуум каких-нибудь значений, но только с вероятностью ноль. Тогда в определении говорят «совокупность её возможных значений», выбрасывая те, что принимаются с нулевой вероятностью. Строго говоря, проделывая такой манёвр, мы подменяем пространство элементарных исходов, так что не следует этим злоупотреблять. Потеря строгости лучше избегать.

Пример. Пусть осуществляется выбор шаров двух цветов (разных, если угодно) из урны по схеме с возвращением, вероятность вытащить шар цвета 1 равна p . Пусть X — это количество шаров первого цвета 1 в цепочке таких экспериментов, длина которой равна n . Каким бы пространством элементарных исходов ни описывался эксперимент (шары ведь можно занумеровать, что и проделывалось на с.21), оно будет конечным (если оно вменяемое). Так что тут отображение X — случайная величина, дискретная. Распределение дискретной случайной величины удобно представить таблицей:

$$P(X = k) = C_n^k p^k (1 - p)^{n-k}; \quad \begin{matrix} 0 & 1 & \dots & n \\ p_0 & p_1 & \dots & p_n \end{matrix} \text{ — распределение } X.$$

Замечание. Случайные величины, имеющие одинаковые распределение, могут оказаться совершенно непохожими друг на друга. Простой пример — два «противоположных» индикатора, характеризующие результат подбрасывания правильной монетки: индикатор герба и индикатор решки. Пример посложнее: пусть Ω — это целые числа от нуля до девяти, исходы равновозможны. Одна случайная величина — это $X(\omega) = \sin(\omega/3)$, вторая — Y , полученная из X простой перестановкой порядка следования точек. Получится другая случайная величина с тем же распределением. Слева на рисунке изображена X , дальше — разные варианты Y :



Все четыре случайных величины распределены одинаково, но выглядят совершенно по-разному. А вообще, эти десять точек можно переставить $10!$ способами, при этом будет много случайных величин, не равных как функции ни в одной точке.

3.7 Схема Бернулли

Пусть имеется некоторый эксперимент, при проведении которого событие A наступает либо не наступает, обозначим $p = P(A)$. При неизменных условиях эксперимент проводится n раз. Пространство элементарных исходов принимает вид $\Omega = \{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n\}$, где каждый из $\varepsilon_i, i \in \mathbb{Z}, 1 \leq i \leq n$ может быть либо нулём, если событие A в i -том эксперименте не наступило, либо единицей, в противном случае. Причём если $p \neq 1/2$, то равновозможности элементарных исходов нет. Такие эксперименты с двумя возможными результатами называются ещё **испытаниями Бернулли**. Причём обязательно подразумевается их независимость друг от друга, если специально не оговорено что-нибудь другое.

В качестве σ -алгебры можно брать множество всех подмножеств Ω . Вероятность полностью определяется, если задать её для всех событий, состоящих из одного элементарного исхода, что делается следующим образом:

$$P(\{\omega\}) = p^{\sum_{i=1}^n \varepsilon_i} (1 - p)^{n - \sum_{i=1}^n \varepsilon_i}, \text{ где } \omega = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n) \text{ — набор из нулей и единиц.}$$

Объясняется это следующим образом. Пусть цепочка фиксированная, то есть жёстко заданы позиции, на которых нужен ноль и так же жёстко заданы позиции под единицы. Независимо от номера позиции в цепочке (эксперименты ведь проводятся при одинаковых условиях) вероятность того, что в ней оказалась единица, равна p , а вероятность того, что ноль — $1 - p$. Эксперименты между собой независимы в совокупности — значит, вероятность пересечения событий, состоящих в том, что на конкретных позициях стоят нужные цифры, равна произведению этих вероятностей.

То есть, вероятность конкретной цепочки с k единицами и $n - k$ нулями получается равной $p^k(1 - p)^{n-k}$. Между прочим, она не зависит от того, на каких конкретно позициях какие цифры стоят.

Определение 3.13. Вероятностное пространство, отвечающее модели из n одинаковых независимых испытаний, дающих два результата («успех» с вероятностью p и «неуспех» с вероятностью $1 - p$), называется **схемой Бернулли**.

Пусть X — это количество наступлений события A в n испытаниях, случайная величина.

Пусть событие A_i состоит в том, что на i -том шаге A наступило. Тогда X можно представить в виде суммы индикаторов A_i :

$$X = \sum_{i=1}^n I_{A_i}.$$

Эксперименты между собой полностью независимы, тогда события $\{A_i\}$ независимы в совокупности. Вероятность того, что $I_{A_i} = 1$ равна вероятности того, что A_i произошло, то есть p . Опираясь на эти соображения, можно отыскать распределение случайной величины X : всё почти так же, как и в урновой схеме с возвращением, правда, теперь p может оказаться иррациональным.

Разных цепочек длины n с k единиц и $n - k$ нулей можно составить, как известно C_n^k штук (всё равно что C_n^{n-k}). Цепочки с единицами в разных наборах позиций общих между собой не имеют. Поэтому соответствующие им события не пересекаются и вероятность их объединения равна сумме их вероятностей. То есть, вероятность того, что в результате n экспериментов произошло k удач и $n - k$ неудач равна $C_n^k p^k (1 - p)^{n-k}$. Возвращаемся к случайной величине X :

$$\begin{matrix} 0 & 1 & \dots & n-1 & n \\ (1-p)^n & np(1-p)^{n-1} & \dots & np^{n-1}(1-p) & p^n \end{matrix} \text{ — распределение } X \text{ — биномиальное.}$$

Везде в моделях из n независимых одинаковых экспериментов с вероятностью «успеха» p случайная величина, равная общему количеству «успехов» за эти n экспериментов, будет иметь биномиальное с параметрами n и p распределение.

Немного позже выяснится, что среднее X/n равно p (на с.39) и что при $n \rightarrow \infty$ X/n к нему стремится (на с.47). Так что, имея на руках много испытаний Бернулли, можно оценивать вероятность «успеха» в одном испытании.

Лекция 4

4.1 Независимость дискретных случайных величин, теорема о независимости двух функций от пересекающихся совокупностей независимых дискретных случайных величин

В продолжение разговора о независимости событий (который шёл на с. 26) речь пойдёт о независимости случайных величин, пока дискретных.

Пусть задано вероятностное пространство (Ω, F, P) и на нём все упоминаемые ниже случайные величины.

Определение 4.1. Пусть имеется конечный либо счётный набор случайных величин $\{X_i\}$. Случайные величины $\{X_i\}$ независимы или независимы в совокупности



Для любого набора вещественных чисел $\{x_{ij}\}$ независимы события $\{\omega: X_i = x_{ij}\}$.

Снова попарная независимость нескольких случайных величин и их независимость в совокупности — в общем случае не одно и то же. Достаточно взять пример Бернштейна и добавить в него три соответствующих случайных величины.

Замечание. Пусть имеется случайная величина X . Спрашивается, при каких условиях случайные величины из набора $\{X, X\}$ независимы.

Должны быть независимыми события $\{X = x_i\}$ при любых x_i . Пусть для X имеется два различных значения x_i и x_j , причём $P(X = x_i) \neq 0$. Тогда должны выполняться равенства:

$$0 = P(\emptyset) = P(\{X = x_i\} \cap \{X = x_j\}) = P(X = x_i)P(X = x_j),$$
$$P(X = x_i) = P(\{X = x_i\} \cap \{X = x_i\}) = P(X = x_i)^2.$$

Так что одно своё значение эта случайная величина принимает с вероятностью один, а остальные — с вероятностью ноль. Тогда, во-первых, любое количество констант с вероятностью один можно прибавить к независимому набору случайных величин, и зависимым он от этого не станет. Во-вторых, если имеется набор независимых случайных величин, в котором есть одинаковые члены, то они — константы с вероятностью один. Вообще, у константы событие $A(x) = \{\omega: C = x\}$ будет либо пустым, либо равным Ω . Поэтому константа с вероятностью один никогда и ни от кого не зависит в силу того, что \emptyset и Ω всегда независимы с любым событием.

Теорема 4.1. Пусть X_1, X_2, \dots, X_k и Y_1, Y_2, \dots, Y_n — независимые случайные величины. Пусть f и g — измеримые функции (с. 60), причём $f: \mathbb{R}^k \rightarrow \mathbb{R}$ и $g: \mathbb{R}^n \rightarrow \mathbb{R}$. Тогда случайные величины $f(X_1, X_2, \dots, X_k)$ и $g(Y_1, Y_2, \dots, Y_n)$ независимы.

Замечание. Измеримость нужна для того, чтобы $f(X_1, \dots, X_k)$ и $g(Y_1, \dots, Y_n)$ были случайными величинами. Опять же, если Ω конечное либо счётное, то любое отображение из Ω в \mathbb{R} измеримо.

Доказательство. Возьмём произвольные вещественные числа a и b . Построим события A и B следующим образом:

$$A = \{\omega: f(X_1(\omega), X_2(\omega), \dots, X_k(\omega)) = a\} \text{ и } B = \{\omega: g(Y_1(\omega), Y_2(\omega), \dots, Y_n(\omega)) = b\}.$$

Нужно показать независимость этих двух событий.

Заметим, что векторы $\bar{X}(\omega)$ и $\bar{Y}(\omega)$ (равные соответственно $(X_1(\omega), \dots, X_k(\omega))$ и $(Y_1(\omega), \dots, Y_n(\omega))$) составлены из дискретных случайных величин и, следовательно, принимают не более счётного количества значений. Это значит, что любые подмножества множеств значений этих векторов $E_{\bar{X}}$ и $E_{\bar{Y}}$ можно представить как счётные объединения конкретных значений этих векторов. Это используется в выкладках ниже в переходе \spadesuit .

Пусть теперь D и T — некоторые множества из \mathbb{R}^k и \mathbb{R}^n соответственно. Покажем, что события $\{\omega: \bar{X}(\omega) \in D\}$ и $\{\omega: \bar{Y}(\omega) \in T\}$ независимы:

$$\begin{aligned} \mathbb{P}(\bar{X} \in D, \bar{Y} \in T) &= \mathbb{P}(\bar{X} \in D \cap E_{\bar{X}}, \bar{Y} \in T \cap E_{\bar{Y}}) \stackrel{\spadesuit}{=} \mathbb{P}\left(\bigcup_{\substack{\bar{D}_i \in D \cap E_{\bar{X}} \\ \bar{T}_i \in T \cap E_{\bar{Y}}}} \{\bar{X} = \bar{D}_i, \bar{Y} = \bar{T}_i\}\right) \\ &\stackrel{\clubsuit}{=} \sum_{\substack{\bar{D}_i \in D \cap E_{\bar{X}} \\ \bar{T}_i \in T \cap E_{\bar{Y}}}} \mathbb{P}(\bar{X} = \bar{D}_i, \bar{Y} = \bar{T}_i) \stackrel{\clubsuit}{=} \sum_{\substack{\bar{D}_i \in D \cap E_{\bar{X}} \\ \bar{T}_i \in T \cap E_{\bar{Y}}}} \mathbb{P}(\bar{X} = \bar{D}_i) \mathbb{P}(\bar{Y} = \bar{T}_i) \\ &\stackrel{\heartsuit}{=} \left(\sum_{\bar{D}_i \in D \cap E_{\bar{X}}} \mathbb{P}(\{\bar{X} = \bar{D}_i\})\right) \left(\sum_{\bar{T}_i \in T \cap E_{\bar{Y}}} \mathbb{P}(\{\bar{Y} = \bar{T}_i\})\right) \\ &= \mathbb{P}(\bar{X} \in D \cap E_{\bar{X}}) \mathbb{P}(\bar{Y} \in T \cap E_{\bar{Y}}) = \mathbb{P}(\bar{X} \in D) \mathbb{P}(\bar{Y} \in T). \end{aligned}$$

Переход \clubsuit — это независимость случайных величин X_1, X_2, \dots, X_k и Y_1, Y_2, \dots, Y_n . Переход \spadesuit — это не что иное, как счётная или конечная аддитивность вероятности. Как известно, декартово произведение не более, чем счётных множеств не более, чем счётно, поэтому объединение по множеству $\bar{D}_i \in D \cap E_{\bar{X}}, \bar{T}_i \in T \cap E_{\bar{Y}}$ — не более, чем счётное. В \heartsuit произошла перегруппировка слагаемых. Сумма от этого не изменится — ряды, из членов которых составлены члены нового ряда, сходятся абсолютно. Так что полученный ряд тоже абсолютно сходится и суммировать его можно в любом порядке (факт доказывается в курсе мат. анализа и ещё попадётся впереди).

Теперь остаётся в качестве множеств D и T взять множества A и B , предварительно заметив, что они представимы в виде

$$\begin{aligned} A &= \{\omega: (X_1(\omega), X_2(\omega), \dots, X_k(\omega)) \in f^{-1}(a)\}, \\ B &= \{\omega: (Y_1(\omega), Y_2(\omega), \dots, Y_n(\omega)) \in g^{-1}(b)\}. \end{aligned}$$

Всё, теперь остаётся воспользоваться обоснованным выше фактом и получить:

$$\mathbb{P}(AB) = \mathbb{P}(\bar{X} \in f^{-1}(a), \bar{Y} \in g^{-1}(b)) = \mathbb{P}(\bar{X} \in f^{-1}(a)) \mathbb{P}(\bar{Y} \in g^{-1}(b)) = \mathbb{P}(A) \mathbb{P}(B). \quad \square$$

4.2 Математическое ожидание дискретной случайной величины

Пусть пространство элементарных исходов Ω — не более чем счётное, σ -алгебра событий F — множество всех подмножеств Ω , P — вероятность, то есть, задано вероятностное пространство (Ω, F, P) , в котором определена случайная величина $X(\omega)$.

Определение 4.2. *Математическим ожиданием или средним случайной величины $X(\omega)$ называется сумма ряда*

$$\sum_{\omega \in \Omega} X(\omega) P(\{\omega\}),$$

при условии абсолютной его сходимости. Обозначается EX .

Замечание. Абсолютная сходимость понадобится далее в доказательствах некоторых свойств мат. ожидания. Это во-первых. Во-вторых, в каком порядке по Ω надо суммировать — нигде не указано и в общем случае как-то этот порядок не изобрести. Так что если ряд из определения сходится условно, то оно попросту некорректно, потому что по теореме Римана (которая присутствует в курсе мат. анализа) можно так переставить ω в сумме, что ряд сойдётся к любому наперёд заданному вещественному числу. Ну а абсолютно сходящийся ряд имеет единственный предел, так что мат. ожидание для каждой случайной величины единственно.

4.2.1 Свойства математического ожидания

1°. *Мат. ожидание константы ей же и равно. $X(\omega) = C$ (константа) $\implies EX = C$.*

Доказательство. Потребуется всего-то счётная или конечная аддитивность вероятности для суммирования вероятностей попарно несовместных событий вида $\{\omega\}$:

$$EX = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}) = C \cdot \sum_{\omega \in \Omega} P(\{\omega\}) = C \cdot P\left(\bigcup_{\omega \in \Omega} \omega\right) = C \cdot P(\Omega) = C.$$

Хотя утверждение само по себе совершенно очевидное: среднее от константы! \square

Далее возникнет сложение и перемножение случайных величин. Если возникает вопрос, по какому такому праву произведение и сумма случайных величин является случайной величиной, его нужно устранить с помощью замечания на с.30.

2°. *Пусть в нашем вероятностном пространстве имеется две случайных величины X и Y , тогда если мат. ожидание обеих существует, то существует мат. ожидание их суммы и равно оно сумме их мат. ожиданий. То есть,*

$$\exists EX, \exists EY \implies \exists E(X + Y) = EX + EY.$$

Доказательство. Факт попросту повторяет известное свойство абсолютно сходящихся рядов, а переносить сюда выкладки из математического анализа совершенно излишне. Так что доказательство очевидно. \square

3°. *Умножение на константу из-под знака мат. ожидания можно выносить:*

$$E(C \cdot X) = C \cdot EX.$$

Доказательство. Как известно, из-под знака суммирования для абсолютно сходящегося ряда константу можно вынести. \square

Замечание. Раз уж умножение на константу и сложение из-под знака мат. ожидания можно выносить, то можно выносить и вычитание.

4°. Пусть набор значений (x_1, x_2, \dots) с набором вероятностей (p_1, p_2, \dots) — распределение случайной величины X (здесь везде символы стоят для случая, когда X имеет бесконечное количество значений, но для конечного случая всё прodelьвается точно так же). Тогда

$$EX = \sum_{k=1}^{\infty} x_k p_k.$$

Доказательство. Пусть событие A_i состоит в том, что X приняла значение x_i . То есть, $A_i = \{\omega : X(\omega) = x_i\}$. Заметим, что $P(A_i) = p_i$ по определению распределения. Остаётся провести простое преобразование ряда мат. ожидания:

$$\begin{aligned} EX &= \sum_{\omega \in \Omega} X(\omega) P(\omega) \stackrel{*}{=} \sum_{i=1}^{\infty} \sum_{\omega \in A_i} X(\omega) P(\omega) = \sum_{i=1}^{\infty} \sum_{\omega \in A_i} x_i P(\omega) \\ &= \sum_{i=1}^{\infty} x_i \sum_{\omega \in A_i} P(\omega) = \sum_{i=1}^{\infty} x_i P(A_i) = \sum_{i=1}^{\infty} x_i p_i, \end{aligned}$$

помеченный * переход отличен тем, что при нём происходит перегруппировка слагаемых при суммировании. Поскольку ряд сходится абсолютно, на сумму его это никак не влияет. \square

Замечание. Особенность и отличие этого свойства от определения в том, что исходное вероятностное пространство бесследно исчезло. Осталось только распределение. А в определении требуется знать X как функцию ω и P как функцию, заданную на элементах класса событий. Часто речь будет заходить о случайной величине, её распределении и мат. ожидании, но ни слова не будет о вероятностном пространстве, так как распределение случайной величины само по себе может нести достаточное для задачи количество информации.

Замечание. Мат. ожидание индикатора I_A равно вероятности его события $P(A)$:

$$EI_A = 1 \cdot P(A) + 0 \cdot P(\bar{A}) = P(A).$$

5°. Пусть $g: \mathbb{R} \rightarrow \mathbb{R}$ — измеримое отображение (что это значит, объясняется на с.60). Тогда $g(X)$ — случайная величина и

$$Eg(X) = \sum_{i=1}^{\infty} g(x_i) p_i.$$

Опять символы стоят такие, будто X имеет бесконечное количество значений и опять в конечном случае всё точно так же.

Доказательство. Доказательство повторяет доказательство предыдущего свойства с незначительными изменениями. Снова вводим события $A_i = \{\omega : X(\omega) = x_i\}$, снова $P(A_i) = p_i$ и снова распишем мат. ожидание по определению:

$$\begin{aligned} E g(X) &= \sum_{\omega \in \Omega} g(X(\omega)) P(\omega) = \sum_{i=1}^{\infty} \sum_{\omega \in A_i} g(X(\omega)) P(\omega) = \sum_{i=1}^{\infty} \sum_{\omega \in A_i} g(x_i) P(\omega) \\ &= \sum_{i=1}^{\infty} g(x_i) \sum_{\omega \in A_i} P(\omega) = \sum_{i=1}^{\infty} g(x_i) P(A_i) = \sum_{i=1}^{\infty} g(x_i) p_i. \quad \square \end{aligned}$$

6° (Мультипликативность мат. ожидания). Пусть заданы две независимых случайных величины X и Y , и у обеих существует мат. ожидание. Тогда существует мат. ожидание произведения, равное произведению мат. ожиданий исходных случайных величин:

$$X, Y \text{ — независимы и } \exists EX, \exists EY \implies \exists E(XY) = EXEY.$$

Доказательство. Пусть X принимает значения a_1, a_2, a_3, \dots на исходах ω из событий A_1, A_2, A_3, \dots , а Y — значения b_1, b_2, b_3, \dots на событиях B_1, B_2, B_3, \dots . Далее в индексах суммирования будет стоять бесконечность, будто X и Y принимают бесконечное количество значений. Для конечного случая всё точно так же — изменения чисто символические, для случая с континуумом значений все совсем иначе (см. с.??). Итак:

$$\begin{aligned} EXEY &= \sum_{i=1}^{\infty} a_i P(A_i) \sum_{j=1}^{\infty} b_j P(B_j) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_i P(A_i) b_j P(B_j) \stackrel{\clubsuit}{=} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_i b_j P(A_i B_j) \\ &\stackrel{\spadesuit}{=} \sum_{i,j=1}^{\infty} a_i b_j P(A_i B_j) = \sum_{i,j=1}^{\infty} \sum_{\omega \in A_i B_j} X(\omega) Y(\omega) P(\omega) = \sum_{\omega \in \Omega} X(\omega) Y(\omega) P(\omega) = EXY. \end{aligned}$$

Переход \clubsuit — это независимость X и Y . А вот с \spadesuit сложнее: до него суммирование велось сначала по j , а затем по i , после него порядок произвольный. Так поступать можно, потому что каждый член нового ряда — это произведение соответствующих членов двух абсолютно сходящихся рядов. Тогда полученный ряд тоже сходится абсолютно, и в каком порядке его суммировать — совершенно неважно. В курсе мат. анализа этот вопрос разобран под заголовком «Элементарная теория двойных и повторных рядов». \square

Математическое ожидание — это **не** то же самое, что медиана, причём разница очень существенна. Медиана — это значение, слева и справа от которого (на вещественной оси) лежит одинаковое количество значений случайной величины. В США в 1992 году медиана доходов обладателей работающих домашних хозяйств составила примерно 31000\$, в то время как формула мат. ожидания давала 40000\$. 9000\$ — это существенная разница. А вообще, за словом «среднее» может скрываться ещё что-нибудь.

Итак, смысл мат. ожидания — это среднее случайной величины. На практике это станет полезным с появлением Закона больших чисел (а точнее, на с.47) плюс его можно применять как критерий сравнения некоторых сложных объектов. Например, доходности некоторого финансового инструмента. Учебная (не надо воспринимать написанное ниже как руководство к действию, нужно ещё много обсудить) ситуация

такова: на рисунках имеется два графика — один для цены одного финансового инструмента, другой — для цены второго, время разбито по месяцам:



Требуется выяснить, какой из них лучше подходит для вложений.

Вводится такая характеристика, как относительная доходность в момент времени t , обозначается как D_t и определяется так:

$$D_t = \frac{p_t - p_{t-1}}{p_{t-1}}, \text{ где } p_t \text{ — цена в момент времени } t.$$

Предположим, D_t известна в условные моменты времени $t = 0, 5, \dots, 500$. Средняя доходность вычисляется как

$$\frac{D_0 + D_5 + D_{10} + \dots + D_{490} + D_{495} + D_{500}}{101}.$$

Среди D_0, D_5, \dots, D_{500} есть одинаковые — сгруппируем их и получим новый набор значений d_0, d_1, \dots, d_k с набором натуральных чисел n_0, n_1, \dots, n_k , в котором n_i ($0 \leq i \leq k$) — это число вхождений значения d_i в набор D_0, D_5, \dots, D_{500} , так что неминуемо $n_0 + n_1 + \dots + n_k = 101$. Тогда выражение для средней доходности приводится к выражению мат. ожидания D_t :

$$\frac{D_0 + D_5 + \dots + D_{495} + D_{500}}{101} = d_0 \frac{n_0}{101} + d_1 \frac{n_1}{101} + \dots + d_{k-1} \frac{n_{k-1}}{101} + d_k \frac{n_k}{101} = \mathbb{E}D_t.$$

Хотя и не приведено вероятностное пространство, распределение D_t , которого здесь вполне достаточно, присутствует: в наборе n_0, n_1, \dots, n_k все числа остаётся поделить на 101 — и получатся соответствующие вероятности.

Дальше выводы о том, какой из финансовых инструментов лучше всех подходит для вложений, можно делать на основе обычного сравнения между собой вещественных чисел. Например, мат. ожиданий доходностей.

4.2.2 Пример случайной величины, не имеющей мат. ожидания

Пример. Пришло время раскрыть Петербургский Парадокс. Проблема была в том, что сделать игру справедливой посредством предварительных расчётов между игроками нельзя. И вот почему: посчитаем выигрыш подбрасывающего, который будем считать случайной величиной S : он получает 2^n рублей, если первый герб выпадает на n -том шаге. Элементарных исходов здесь счётное число, вероятность того, что первый герб выпадет на n -том шаге, равна 2^{-n} . Вероятностное пространство подробно описано на с.11. Считаем мат. ожидание S :

$$\mathbb{E}S = \sum_{\omega \in \Omega} S(\omega) \mathbb{P}(\{\omega\}) = \sum_{i=1}^{\infty} 2^i 2^{-i} = \sum_{i=1}^{\infty} 1.$$

Этот ряд не сходится абсолютно. Строго говоря, мат. ожидания у этой случайной величины не существует. А вообще, оно кажется невообразимо большим, а точней его не ограничить никаким положительным числом, как и средний выигрыш подбрасывающего. А поскольку передать перед игрой второму игроку неограниченное количество рублей для подбрасывающего не представляется возможным, сделать игру справедливой таким способом нельзя.

Это был пример случайной величины, не имеющей мат. ожидания.

4.2.3 Мат. ожидание биномиально распределённой случайной величины

Пример. Пусть X — случайная величина, имеющая биномиальное распределение с параметрами n и p . Вычислим мат. ожидание X . Попытка прямого вычисления приведёт к небольшим трудностям:

$$\begin{aligned} EX &= \sum_{k=0}^n C_n^k p^k (1-p)^{n-k} = \sum_{k=1}^n \frac{n!}{(n-k)!(k-1)!} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n np \frac{(n-1)!}{((n-1)-(k-1))!(k-1)!} p^{k-1} (1-p)^{(n-1)-(k-1)} \\ &= np \sum_{k=1}^n C_{n-1}^{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} = np \sum_{m=0}^{n-1} C_{n-1}^m p^m (1-p)^{(n-1)-m} \\ &= np(p + (1-p))^{n-1} = np, \end{aligned}$$

а небольшие трудности всегда приятней обойти. Представим лучше X в виде суммы индикаторов успехов на i -том шаге. Иными словами, $X = I_1 + I_2 + I_3 + \dots + I_n$. Почему так можно поступить, объясняется на с.32. Мат. ожидание каждого индикатора равно вероятности того, что произошёл «успех», то есть p . Всего индикаторов n штук и мат. ожидание их суммы равно сумме их мат. ожиданий. Так что легко и просто получаем

$$EX = \sum_{i=1}^n EI_i = \sum_{i=1}^n p = np.$$

4.3 Моменты и центральные моменты k -го порядка

Определение 4.3. *Моментом k -го порядка случайной величины X , $k \in \mathbb{N}$, называется математическое ожидание случайной величины X^k .*

Определение 4.4. *Центральным моментом k -го порядка случайной величины X , где $k \in \mathbb{N}$, называется математическое ожидание случайной величины $(X - EX)^k$.*

Если возникает вопрос о том, с чего вдруг X^k и $(X - EX)^k$ будут случайными величинами, следует обратиться к замечанию на с.30.

Замечание. Центральным моментом первого порядка всегда равен нулю:

$$E(X - EX) = EX - EEX = EX - EX = 0.$$

Нередко бывает удобно чтобы на месте мат. ожидания какой-нибудь случайной величины оказался ноль. И тогда случается так, что вместо исходной случайной величины Y в таких случаях рассматривают случайную величину $Y - EY$, мат. ожидание которой — ноль. Такой ход называется **центрированием**, а случайная величина $Y - EY$ — **центрированной**.

Утверждение 4.1. *Пусть существует момент n -го ($n \in \mathbb{N}$) порядка случайной величины X . Тогда $\forall k \in \mathbb{N}: k < n$ существует момент k -го порядка случайной величины X .*

Доказательство. Поскольку для существования мат. ожидания ряд (из определения мат. ожидания) должен сходиться абсолютно, то существует $E|X|^n$. Тогда в силу факта $\forall y \in \mathbb{R} \quad |y|^k \leq |y|^n + 1$ имеем $|X|^k \leq |X|^n + 1$, а это — оценка членов неотрицательного ряда сверху членами сходящегося. Так что получаем, что мат. ожидание $|X|^k$ тоже существует. \square

4.4 Дисперсия и её свойства

Определение 4.5. *Дисперсией случайной величины X называется её центральный момент второго порядка:*

$$DX = E(X - EX)^2.$$

Дисперсия характеризует разброс случайной величины X относительно своего мат. ожидания. Что замечательно, маленькие отклонения возводятся в квадрат и становятся от того ещё меньше, а большие — наоборот, увеличиваются. В экономических моделях эта конструкция встречается под именем «волатильность».

1°. *Дисперсия константы равна нулю: $DC \equiv 0$.*

Доказательство.

$$DC = E(C - EC)^2 = EC^2 - 2E(CEC) + E(EC)^2 = EC^2 - 2E(C^2) + E(C^2) = 0. \quad \square$$

2°. *Изменение случайной величины на константу не влияет на её дисперсию:*

$$D(X + C) = DX.$$

Доказательство.

$$\begin{aligned} D(X + C) &= E(X + C - E(X + C))^2 = E(X - EX + C - EC)^2 \\ &= E(X - EX + C - C)^2 = E(X - EX)^2 = DX. \end{aligned} \quad \square$$

3°. *Константа выносится из-под знака дисперсии с квадратом:*

$$D(CX) = C^2DX.$$

Доказательство.

$$\begin{aligned} D(CX) &= E(CX - E(CX))^2 = E(CX - CEX)^2 = E(C(X - EX))^2 \\ &= E(C^2(X - EX)^2) = C^2E(X - EX)^2 = C^2DX. \end{aligned} \quad \square$$

4°. *Для любой случайной величины X*

$$DX \geq 0.$$

Доказательство. Дело в том, что ряд из определения дисперсии абсолютно сходится и состоит из неотрицательных членов. \square

5°. *Пусть X и Y — две независимых случайных величины. Тогда если существуют их дисперсии по отдельности, то существует дисперсия суммы, равная сумме дисперсий по отдельности, то есть*

$$X, Y \text{ — независимы и } \exists DX, \exists DY \implies \exists D(X + Y) = DX + DY.$$

Доказательство.

$$\begin{aligned} D(X + Y) &= E(X + Y - E(X + Y))^2 \\ &= E((X - EX)^2 + (Y - EY)^2 + 2E(X - EX)(Y - EY)). \end{aligned}$$

Поскольку $X - EX$ и $Y - EY$ — измеримые функции от независимых случайных величин X и Y , то они — сами независимые случайные величины. Тогда

$$\begin{aligned} E((X - EX)^2 + (Y - EY)^2 + 2E(X - EX)(Y - EY)) &= DX + DY \\ &\quad + 2E(X - EX)E(Y - EY) \\ &= DX + DY, \end{aligned}$$

потому что центральный момент первого порядка всегда нулевой. Кстати, в этом доказательстве не использовалась в явном виде дискретность X и Y , которая скрывается пока что за фактом мультипликативности мат. ожидания. Но когда последний факт будет доказан для более общего случая случайных величин, это доказательство аддитивности дисперсии автоматически перейдёт на этот более общий случай. \square

Следствие 4.1 (Аддитивность дисперсии). Пусть Y_1, Y_2, \dots, Y_n — независимые случайные величины. Тогда если для каждой из них существует дисперсия, то существует дисперсия их суммы, равная сумме их дисперсий. То есть:

$$\begin{aligned} Y_1, Y_2, \dots, Y_n \text{ — независимы и } \exists DY_1, \exists DY_2, \dots, \exists DY_n \\ \Downarrow \\ \exists D(Y_1 + Y_2 + \dots + Y_n) = DY_1 + DY_2 + \dots + DY_n. \end{aligned}$$

Доказательство. Доказательство — привычная мат. индукция. База только что была доказана, предположение имеется в условии следствия, а дальше простой индукционный переход: за X_1 надо взять $Y_1 + Y_2 + \dots + Y_n$, за X_2 — Y_{n+1} . Нужно не забыть заметить, что X_1 и X_2 — независимые случайные величины, что у них существуют дисперсии (у X_1 — по предположению индукции), и применить для них базу. \square

Замечание. В реальных условиях большое количество независимых случайных величин — это ненормально. Как минимум, подозрительно!

Замечание. Часто используется преобразованный вид дисперсии, а именно:

$$\begin{aligned} DX &= E(X - EX)^2 = E(X^2 - 2XEX + (EX)^2) = E(X^2) - 2E(XEX) + E(EX)^2 \\ &= E(X^2) - 2(EX)^2 + (EX)^2 = EX^2 - (EX)^2. \end{aligned}$$

4.4.1 Дисперсия биномиально распределённой случайной величины

Пример. Вычислим дисперсию случайной величины, имеющей биномиальное распределение с параметрами n и p . Мат. ожидание вычислено на с.39 и равно оно np . Попытка прямого вычисления дисперсии приводит к более серьёзным трудностям, которые вновь обходятся с помощью индикаторов. Пусть $I_i, 1 \leq i \leq n$ — индикатор успеха в i -том испытании Бернулли, эти n индикаторов независимы (переход \clubsuit) и одинаково распределены (переход \spadesuit). Заметим, кроме того, что квадрат любого индикатора всегда точно равен самому индикатору (переход \heartsuit). Тогда

$$\begin{aligned} DX &= D(I_1 + I_2 + \dots + I_n) \stackrel{\clubsuit}{=} DI_1 + DI_2 + \dots + DI_n \stackrel{\spadesuit}{=} nDI_1 \\ &= nE(I_1 - EI_1)^2 = n(EI_1^2 - (EI_1)^2) \stackrel{\heartsuit}{=} n(EI_1 - (EI_1)^2) = n(p - p^2). \end{aligned}$$

4.4.2 Среднеквадратичное отклонение

Определение 4.6. *Среднеквадратичным отклонением называется квадратный корень из дисперсии.*

Все свойства среднеквадратичного отклонения вытекают из свойств дисперсии. Оно нелинейно: константу приходится выносить с модулем, отклонение суммы в общем случае не равно сумме отклонений. Зато размерность его совпадает с размерностью исходной случайной величины. Нужно это для того, чтобы экспериментатор после проведения измерений мог вычислить величину отклонения и сравнить её со средней величиной своих результатов. Дальше, получив относительную погрешность, он может принимать решение о том, верить своим результатам или нет.

4.5 Ковариация и коэффициент корреляции

Определение 4.7. *Пусть заданы две случайных величины, имеющих мат. ожидание: X и Y . Ковариацией случайных величин X и Y называется число*

$$\text{cov}(X, Y) = E((X - EX)(Y - EY)).$$

Полезно провести небольшие преобразования выражения в определении:

$$\begin{aligned}\text{cov}(X, Y) &= E((X - EX)(Y - EY)) = E(XY - YEX - XEY + EXEY) \\ &= EXY - EYEX - EXEY + EXEY = EXY - EXEY.\end{aligned}$$

Замечание. Ковариация не изменяется при изменении случайных величин на константы:

$$\begin{aligned}\text{cov}(X + C_X, Y + C_Y) &= E((X + C_X)(Y + C_Y)) - E(X + C_X)E(Y + C_Y) \\ &= E(XY + YC_X + XC_Y + C_XC_Y) - E(X + C_X)E(Y + C_Y) \\ &= EXY - EXEY = \text{cov}(X, Y).\end{aligned}$$

Теперь очевидно, что ковариация независимых случайных величин равна нулю. Обратное, во-первых, почти верно, а во-вторых, в прикладных областях.

Пример. Из равенства нулю ковариации случайных величин не следует их независимость. Пример будет следующий: рассмотрим конечное вероятностное пространство $\Omega = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$ с равновероятными элементарными исходами. Рассмотрим случайные величины

$$X = \sin\left(\frac{\pi\omega}{4}\right) \text{ и } Y = \cos\left(\frac{\pi\omega}{4}\right) :$$

$$\begin{array}{cccccccc} \omega & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ X & 0 & 1 & 1 & 0 & -1 & -1 & 0 & 1 & 1 \\ Y & 1 & 0 & -1 & -1 & 0 & 1 & 1 & 0 & 0 \end{array}$$

Их мат. ожидания равны нулю, выясним мат. ожидание их произведения:

$$EXY = E\left(\sin\left(\frac{\pi\omega}{4}\right)\cos\left(\frac{\pi\omega}{4}\right)\right) = E\frac{1}{2}\sin\left(\frac{\pi\omega}{2}\right) = \frac{1}{18}(0 + \frac{1}{2} + 0 - \frac{1}{2} + 0 + \frac{1}{2} + 0 - \frac{1}{2} + 0) = 0$$

Так что $EXY = EXEY$ и ковариация их равна нулю. Однако события X и Y зависимы: рассмотрим события A и B :

$$A = \{\omega: X = \frac{1}{\sqrt{2}}\}, B = \{\omega: Y = \frac{1}{\sqrt{2}}\}.$$

$$P(A) = \frac{2}{9}, P(B) = \frac{2}{9} \implies P(A)P(B) = \frac{4}{81}, \text{ но } P(AB) = \frac{1}{9} \neq P(A)P(B).$$

Так что равенство нулю ковариации слабее независимости.

Дисперсия суммы двух случайных величин в общем случае имеет следующий вид:

$$\begin{aligned} D(X + Y) &= E((X - EX) + (Y - EY))^2 = DX + DY + 2E((X - EX)(Y - EY)) \\ &= DX + DY + 2 \operatorname{cov}(X, Y). \end{aligned}$$

Выражение для большего количества слагаемых получается точно так же — нужно раскрыть квадрат и собрать нужные слагаемые. Сама дисперсия случайной величины X выражается как $\operatorname{cov}(X, X)$.

Раз из независимости случайных величин следует равенство нулю ковариации, то из неравенства нулю ковариации следует зависимость случайных величин. Это наводит на мысль о том, что величина ковариации характеризует меру зависимости случайных величин. Однако от умножения на константу (не равную нулю) зависимость случайных величин не изменяется никак, а вот ковариацию можно сделать (если она не равна нулю) любым числом, а именно:

$$\operatorname{cov}(CX, Y) = ECXY - ECXEY = C(EXY - EXEY) = C \operatorname{cov}(X, Y).$$

Так что чтобы спасти возникшую мысль, ковариацию необходимо нормировать.

Определение 4.8. Коэффициентом корреляции случайных величин X и Y , имеющих дисперсии, не равные нулю, называется число

$$\rho(X, Y) = \frac{\operatorname{cov}(X, Y)}{\sqrt{DXDY}}.$$

Замечание. Если дисперсия случайной величины равна нулю, то эта случайная величина с вероятностью 1 равна своему мат. ожиданию, т.е. является константой с вероятностью один. Для характеристики независимости случайной величины с такой константой коэффициент корреляции не нужен. На с.33 всё и без него ясно.

Коэффициент корреляции лучше ковариации тем, что от умножения случайных величин на константы его модуль не изменяется:

$$\forall a, b \in \mathbb{R}: ab \neq 0 \quad \rho(aX, bY) = \frac{\operatorname{cov}(aX, bY)}{\sqrt{DaXD bY}} = \frac{ab \operatorname{cov}(X, Y)}{|ab| \sqrt{DXDY}} = \frac{ab}{|ab|} \rho(X, Y).$$

4.5.1 Свойства коэффициента корреляции

1°. Коэффициент корреляции независимых случайных величин равен нулю.

Доказательство. В числителе дроби, которой равен коэффициент корреляции, окажется ноль. В знаменателе нуля быть не должно, это обеспечивается определением. \square

2°. Для любых двух случайных величин (для которых выполнены условия определения) их коэффициент корреляции по модулю не превосходит единицы.

Доказательство. Обозначим эти две случайные величины как X и Y и центрируем: $X_c = X - EX$ и $Y_c = Y - EY$. Недавно показывалось, что $\operatorname{cov}(X, Y) = \operatorname{cov}(X_c, Y_c)$, а что дисперсия случайной величины не меняется от смещения случайной величины на константу — ещё раньше (если это уже забыто, то можно вспомнить о факте $DX = \operatorname{cov}(X, X) = \operatorname{cov}(X + C, X + C) = D(X + C)$). Так что центрирование никак

не повлияет и на коэффициент корреляции. Понадобятся следующие соотношения, оба из которых вытекают из того, что $\mathbf{E}X_c = \mathbf{E}Y_c = 0$:

$$DX_c = \mathbf{E}X_c^2 - (\mathbf{E}X_c)^2 = \mathbf{E}X_c^2, DY_c = \mathbf{E}Y_c^2, \text{cov}(X_c, Y_c) = \mathbf{E}(X_c Y_c) - \mathbf{E}X_c \mathbf{E}Y_c = \mathbf{E}(X_c Y_c).$$

Далее идут те же рассуждения, что часто используются при доказательстве неравенства Коши-Буняковского:

$$\forall a \in \mathbb{R} \quad 0 \leq D(X_c - aY_c) = \mathbf{E}(X_c - aY_c)^2 - (\mathbf{E}(X_c - aY_c))^2 = \mathbf{E}(X_c - aY_c)^2.$$

Полученное неравенство можно рассматривать как квадратное неравенство относительно a , а именно

$$\mathbf{E}(X_c - aY_c)^2 = \mathbf{E}X_c^2 - 2a\mathbf{E}(X_c Y_c) + a^2\mathbf{E}Y_c^2 \geq 0.$$

Поскольку верно это для любого a , то дискриминанту нельзя ни в коем случае быть больше нуля. То есть:

$$(\mathbf{E}(X_c Y_c))^2 - \mathbf{E}X_c^2 \mathbf{E}Y_c^2 \leq 0 \iff |\mathbf{E}(X_c Y_c)| \leq \sqrt{\mathbf{E}X_c^2 \mathbf{E}Y_c^2} \Rightarrow |\text{cov}(X_c, Y_c)| \leq \sqrt{DX_c DY_c}.$$

Поскольку стирание индексов с никаких значений в последнем выражении не изменит, то получено то, что сформулировано в условии. \square

3°. Если $|\rho(X, Y)| = 1$, то с вероятностью один X и Y линейно выражаются друг через друга. То есть,

$$|\rho(X, Y)| = 1 \implies \exists b \neq 0, c \in \mathbb{R}: P(X - bY = c) = 1.$$

Доказательство. Доказательство этого свойства целиком опирается на доказательство предыдущего: если выполнилось равенство $|\text{cov}(X, Y)| = \sqrt{DXDY}$, то квадратное неравенство относительно a может обращаться в равенство при некотором $a = b$. Но это равенство означает, что равна нулю $D(X - bY)$, а это сразу говорит о том, что с вероятностью один $X - bY$ равна константе. Обозначим эту константу за c и получим то, что нужно было доказать. \square

Выражаясь нестрого, чем ближе модуль коэффициента корреляции к единице, тем сильнее зависимость между случайными величинами. Но эта зависимость — не причинная, это важно понимать. О смысле вероятностной зависимости несколько слов было сказано на с.24. Если коэффициент корреляции равен нулю, то о независимость случайных величин ещё не гарантирована, но степень зависимости слаба.

Определение 4.9. Случайные величины называются некоррелированными, если для них существует коэффициент корреляции и равен он нулю.

Для некоррелированных случайных величин дисперсия суммы равна сумме дисперсий — там в доказательстве использовалось соотношение $\mathbf{E}(XY) = \mathbf{E}X\mathbf{E}Y$ безотносительно к его происхождению.

4.6 Неравенство А. А. Маркова

Утверждение 4.2 (Неравенство Маркова). Пусть случайная величина X неотрицательна. Тогда

$$\forall a > 0 \quad P(X \geq a) \leq \frac{\mathbf{E}X}{a}$$

Доказательство. Пусть $I_{X \geq a}$ и $I_{X < a}$ — индикаторы событий $\{\omega: X(\omega) \geq a\}$ и $\{\omega: X(\omega) < a\}$ соответственно. Тогда:

$$X = X \cdot 1 = X \cdot (I_{X \geq a} + I_{X < a}) \geq X \cdot I_{X \geq a}$$

Последнее произведение равно нулю, если $X < a$ и равно X , когда $X \geq a$. Второе обстоятельство даёт право сделать следующую оценку:

$$X \cdot I_{X \geq a} \geq a I_{X \geq a}$$

Теперь, собрав начало и конец этой цепочки вместе и вспомнив, что мат. ожидание индикатора события равно вероятности его события, имеем

$$X \geq a I_{X \geq a} \implies EX \geq a E I_{X \geq a} \implies EX \geq a P(X \geq a) \implies P(X \geq a) \leq \frac{EX}{a}. \quad \square$$

4.7 Неравенство Чебышева

Утверждение 4.3 (Неравенство Чебышева).

$$\exists DX \implies \forall \varepsilon > 0 \quad P(|X - EX| > \varepsilon) \leq \frac{DX}{\varepsilon^2}.$$

Замечание. Пользуясь тем фактом, что $P(\bar{A}) = 1 - P(A)$, условие можно переформулировать так:

$$\exists DX \implies \forall \varepsilon > 0 \quad P(|X - EX| \leq \varepsilon) \geq 1 - \frac{DX}{\varepsilon^2}.$$

Кроме того, вместо $\exists DX$ можно требовать $\exists EX^2$ — тогда существование EX гарантируется утверждением на с.39, а $DX = EX^2 - (EX)^2$, так что эти два требования равносильны.

Доказательство. Случайная величина $|X - EX|^2$ неотрицательна, используем для неё неравенство Маркова:

$$P(|X - EX| > \varepsilon) = P(|X - EX|^2 > \varepsilon^2) \stackrel{*}{\leq} P(|X - EX|^2 \geq \varepsilon^2) \leq \frac{E|X - EX|^2}{\varepsilon^2} = \frac{DX}{\varepsilon^2}. \quad \square$$

Замечание. Начав с перехода *, докажем неравенство в виде

$$P(|X - EX| \geq \varepsilon) \leq \frac{DX}{\varepsilon^2} \quad \text{или} \quad P(|X - EX| < \varepsilon) \geq 1 - \frac{DX}{\varepsilon^2}.$$

В общем-то, изменения незначительные и неважные, однако в разных местах это неравенство может появиться в любой из четырёх упомянутых форм, и вопросов это вызывать не должно.

4.7.1 Правило трёх сигм

В неравенстве Чебышева в качестве ε можно брать любое положительное число. Ничто не мешает для этого использовать корень из дисперсии, например. Но при таком выборе в правой части неравенства (где после вероятности стоит \leq) окажется единица, такая оценка едва ли представляет интерес.

Однако если взять в качестве ε величину 3σ , где σ — стандартное отклонение (то есть именно корень из дисперсии), то получится уже более информативное

$$P(|X - EX| > 3\sigma) \leq \frac{DX}{9DX} = \frac{1}{9} \text{ или то же самое } P(|X - EX| \leq 3\sigma) \geq 1 - \frac{1}{9} = \frac{8}{9}.$$

То есть, вероятность того, что случайная величина находится от своего мат. ожидания не дальше, чем на три стандартных отклонения, не меньше $0,888\dots$, что уже что-то да означает. Полученное соотношение и называется правилом трёх сигм.

Можно попробовать взять два или четыре стандартных отклонения вместо трёх, но тогда в случае двух отклонений получится

$$P(|X - EX| \leq 2\sigma) \geq 1 - \frac{1}{4} = \frac{3}{4},$$

то есть точность попадания случайной величины в интервал уже равна 75%. Для четырёх отклонений имеем

$$P(|X - EX| \leq 4\sigma) \geq 1 - \frac{1}{16} = \frac{15}{16}.$$

Тут точность — гораздо больше, но зато отрезок в два раза больше, чем для двух отклонений. Разумеется, в конкретной ситуации каждое из этих правил может оказаться более подходящим, чем другие. Всё зависит от конкретных условий.

Лекция 5

5.1 Закон больших чисел Чебышева

Закон больших чисел (он же ЗБЧ) будет сформулирован не один раз, от имени не одного и того же исследователя и не с одними и теми же условиями. Поэтому всякий раз рядом с ЗБЧ будет присутствовать указание на то, чей он, если это не будет ясно из контекста.

Теорема 5.1 (ЗБЧ в форме Чебышева). Пусть X_1, X_2, \dots, X_n — независимые случайные величины и дисперсия каждой из них существует и ограничена сверху некоторой константой: $\forall i: 1 \leq i \leq n \exists DX_i \leq C$. Тогда

$$\forall \varepsilon > 0 \quad \mathbb{P} \left(\left| \frac{X_1 + \dots + X_n}{n} - \frac{EX_1 + \dots + EX_n}{n} \right| < \varepsilon \right) \xrightarrow{n \rightarrow \infty} 1.$$

Доказательство. Сконструируем новую случайную величину X и вычислим её мат. ожидание и дисперсию:

$$X = \frac{X_1 + \dots + X_n}{n}, \quad EX = \frac{EX_1 + \dots + EX_n}{n}, \quad DX = D \sum_{i=1}^n \frac{X_i}{n} \stackrel{*}{=} \frac{1}{n^2} \sum_{i=1}^n DX_i.$$

В переходе * используется независимость X_i . Просто так сложение за знак дисперсии выносить нельзя.

Дальше для X воспользуемся неравенством Чебышева и оценим правую часть, используя ограничение на дисперсии X_i :

$$\mathbb{P}(|X - EX| < \varepsilon) \geq 1 - \frac{DX}{\varepsilon^2} = 1 - \frac{DX_1 + \dots + DX_n}{n^2 \varepsilon^2} \geq 1 - \frac{nC}{n^2 \varepsilon^2} = 1 - \frac{C}{n \varepsilon^2}$$

Остаётся вспомнить, что такое X и устремить n к бесконечности:

$$\mathbb{P} \left(\left| \frac{X_1 + \dots + X_n}{n} - \frac{EX_1 + \dots + EX_n}{n} \right| < \varepsilon \right) = \mathbb{P}(|X - EX| < \varepsilon) \geq 1 - \frac{C}{n \varepsilon^2} \xrightarrow{n \rightarrow \infty} 1.$$

Таким образом, искомое утверждение доказано. □

5.2 Теорема Бернулли

Как следствие из только что доказанной теоремы легко получается ЗБЧ Якова Бернулли. В роли независимых случайных величин выступают индикаторы успехов в независимых испытаниях:

Теорема 5.2 (Бернулли). Пусть S_n — число успехов в n испытаниях Бернулли с вероятностью успеха p . Тогда

$$\forall \varepsilon > 0 \quad \mathbb{P} \left(\left| \frac{S_n}{n} - p \right| < \varepsilon \right) \xrightarrow{n \rightarrow \infty} 1.$$

Доказательство. Представим S_n как сумму n индикаторов успеха I_i в i -том испытании:

$$S_n = I_1 + \dots + I_n; I_i = \begin{cases} 1, & \text{если в } i\text{-том испытании произошёл успех,} \\ 0, & \text{в противном случае.} \end{cases}, 1 \leq i \leq n.$$

Как уже отмечалось, $E I_i = p$ и $D I_i = p - p^2$ при любом допустимом i . Дисперсии ограничены сверху константой 0,25, а индикаторы независимы. Остаётся применить ЗБЧ в форме Чебышева:

$$\begin{aligned} P\left(\left|\frac{S_n}{n} - E\frac{S_n}{n}\right| < \varepsilon\right) &= P\left(\left|\frac{S_n}{n} - \frac{E I_1 + \dots + E I_n}{n}\right| < \varepsilon\right) = P\left(\left|\frac{S_n}{n} - \frac{np}{n}\right| < \varepsilon\right) \\ &= P\left(\left|\frac{S_n}{n} - p\right| < \varepsilon\right) \xrightarrow{n \rightarrow \infty} 1. \end{aligned} \quad \square$$

5.3 Пуассоновское распределение

В каком-то смысле распределение Пуассона является продолжением биномиального распределения на бесконечное количество испытаний Бернулли. Причём этот факт доказан (правда, оценка погрешности здесь приводится без обоснования).

Определение 5.1. *Дискретная случайная величина X имеет **распределение Пуассона** с параметром a (считается $a > 0$) тогда и только тогда, когда*

$$P(X = k) = \frac{a^k}{k!} e^{-a}, k = 0, 1, 2, \dots,$$

обозначается это так: $X \sim Poiss(a)$ или $X \sim \Pi(a)$.

Замечание. Тот факт, что случайная величина X имеет биномиальное распределение с параметрами n и p кратко записывается как $X \sim Bi(n, p)$.

Известно, что количество посетителей (закусочной или торговой точки) представляет собой **пуассоновский поток** (но только не в час пик). Это означает следующее: пусть на момент времени t_1 насчиталось $Q(t_1)$ клиентов (счёт ведётся от любого подходящего момента времени), а на момент t_0 — $Q(t_0)$. Тогда при всяких $t_1 > t_0$

$$Q(t_1) - Q(t_0) \sim Poiss(\lambda(t_1 - t_0)),$$

где λ называется интенсивностью процесса. Любопытно, что Q можно представить как возрастающую на единицу в моменты времени $\tau_0, \tau_1, \tau_2 \dots$ величину, где промежутки от τ_{i-1} до τ_i — это независимые случайные величины, имеющие показательное распределение с параметром λ (о нём — на с.??).

5.4 Теорема Пуассона

Теорема 5.3 (Пуассона). *Пусть проводится n обобщённых испытаний Бернулли (то есть, вероятность успеха испытания зависит от n) с вероятностью успеха p_n , S_n — количество успехов в этих испытаниях и $np_n \xrightarrow{n \rightarrow \infty} a$. Тогда*

$$\forall k \in \mathbb{Z}: 0 \leq k \leq n \quad P(S_n = k) \xrightarrow{n \rightarrow \infty} \frac{a^k}{k!} e^{-a}.$$

Доказательство. Тем, кто помнит мат. анализ, доказательство покажется элементарным. Для тех, кто его не помнит, здесь всё расписано вплоть до теоремы о произведении пределов. Как известно, S_n имеет биномиальное распределение с параметрами n и p_n . Требуется перейти к пределу при большом n , больше ничего:

$$\begin{aligned} P(S_n = k) &= C_n^k p_n^k (1 - p_n)^{n-k} = \frac{p_n^k n!}{k!(n-k)!} (1 - p_n)^n (1 - p_n)^{-k} \\ &= \frac{(np_n)^k}{k!} \cdot \frac{n!}{(n-k)!n^k} \cdot \frac{(1 - p_n)^n}{(1 - p_n)^k}. \end{aligned}$$

k — это константа. Поэтому безо всяких неопределённостей

$$\lim_{n \rightarrow \infty} \frac{(np_n)^k}{k!} = \frac{a^k}{k!},$$

После простых преобразований очевиден предел второй дроби:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{n!}{(n-k)!n^k} &= \lim_{n \rightarrow \infty} (1 - \frac{1}{n})(1 - \frac{2}{n}) \cdot \dots \cdot (1 - \frac{k-2}{n})(1 - \frac{k-1}{n}) \\ &= \lim_{n \rightarrow \infty} (1 - \frac{1}{n}) \lim_{n \rightarrow \infty} (1 - \frac{2}{n}) \cdot \dots \cdot \lim_{n \rightarrow \infty} (1 - \frac{k-2}{n}) \lim_{n \rightarrow \infty} (1 - \frac{k-1}{n}) = 1. \end{aligned}$$

При выяснении предела третьей дроби требуется второй замечательный предел и непрерывность показательной функции. Кроме того, отметим, что раз $np_n \xrightarrow[n \rightarrow \infty]{} a$, то $p_n \xrightarrow[n \rightarrow \infty]{} 0$. Так что

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{(1 - p_n)^n}{(1 - p_n)^k} &= \frac{\lim_{n \rightarrow \infty} (1 - p_n)^n}{\lim_{n \rightarrow \infty} (1 - p_n)^k} = \frac{\lim_{n \rightarrow \infty} (1 - p_n)^n}{1} = \lim_{n \rightarrow \infty} (1 - p_n)^{\frac{1}{p_n} np_n} \\ &= \lim_{n \rightarrow \infty} \left((1 - p_n)^{-\frac{1}{p_n}} \right)^{-np_n} = \lim_{n \rightarrow \infty} \left((1 - p_n)^{-\frac{1}{p_n}} \right)^{a - np_n - a} \\ &= \lim_{n \rightarrow \infty} \left((1 - p_n)^{-\frac{1}{p_n}} \right)^{-a} \left((1 - p_n)^{-\frac{1}{p_n}} \right)^{a - np_n} \\ &= \lim_{n \rightarrow \infty} \left((1 - p_n)^{-\frac{1}{p_n}} \right)^{-a} \lim_{n \rightarrow \infty} \left((1 - p_n)^{-\frac{1}{p_n}} \right)^{a - np_n} \end{aligned}$$

Последовательность $a - np_n$ является бесконечно малой, обозначим её как γ_n . Разберёмся отдельно со второй скобкой:

$$\lim_{n \rightarrow \infty} \left((1 - p_n)^{-\frac{1}{p_n}} \right)^{\gamma_n} = \lim_{n \rightarrow \infty} e^{\gamma_n \frac{\ln(1-p_n)}{p_n}} = e^{\lim_{n \rightarrow \infty} \gamma_n \frac{\ln(1-p_n)}{p_n}} = e^{\lim_{n \rightarrow \infty} \gamma_n \lim_{n \rightarrow \infty} \frac{\ln(1-p_n)}{p_n}} = e^{0 \cdot 1} = 1.$$

С первой же скобкой всё гораздо проще:

$$\lim_{n \rightarrow \infty} \left((1 - p_n)^{-\frac{1}{p_n}} \right)^{-a} = \left(\lim_{n \rightarrow \infty} (1 - p_n)^{-\frac{1}{p_n}} \right)^{-a} = e^{-a}.$$

Всё, теперь, собрав всё это вместе, получаем требуемое

$$\begin{aligned} \lim_{n \rightarrow \infty} P(S_n = k) &= \lim_{n \rightarrow \infty} \frac{(np_n)^k}{k!} \cdot \frac{n!}{(n-k)!n^k} \cdot \frac{(1 - p_n)^n}{(1 - p_n)^k} \\ &= \lim_{n \rightarrow \infty} \frac{(np_n)^k}{k!} \cdot \lim_{n \rightarrow \infty} \frac{n!}{(n-k)!n^k} \cdot \lim_{n \rightarrow \infty} \frac{(1 - p_n)^n}{(1 - p_n)^k} = \frac{a^k}{k!} \cdot 1 \cdot e^{-a} = \frac{a^k}{k!} e^{-a}. \square \end{aligned}$$

5.4.1 Оценка близости биномиальных вероятностей к пуассоновским

Насколько близко биномиальное распределение приближается пуассоновским, говорит следующее утверждение, которое здесь не доказано:

Утверждение 5.1. Пусть S_n — число успехов в n (теперь обычных) испытаниях Бернулли. Тогда

$$\forall k \in \mathbb{Z}: n \geq k \geq 0 \quad \left| P(S_n = k) - \frac{a^k}{k!} e^{-a} \right| \leq \frac{a^2}{n} = np^2,$$

где $a = np$.

Замечание. Указанная оценка теряет всякий смысл при «больших» p . Например, при $p = 0,5$ в правой части оценки уже при $n = 2$ наблюдается $1/2$. Но вот при $p = 0,01$ вплоть до $n = 100$ получается не такое уж плохое $0,01$.

5.5 Задача о конкуренции (завязка)

Пример. Пусть каждое утро имеется станция с двумя поездами от разных компаний-перевозчиков и 1000 пассажиров, которым требуется уехать на одном из этих поездов. Каждый пассажир выбирает поезд случайным образом безо всяких пристрастий. Имеется в виду, равновероятно. Для того, чтобы гарантированно перевезти всех пассажиров, в каждом из поездов должно быть по 1000 мест. Требуется выяснить, сколько нужно мест в каждом из поездов, чтобы наверняка перевезти 90% клиентов.

Логично подойти к задаче так: пусть S_n — количество пассажиров, выбравших поезд первой компании-перевозчика. Каждый пассажир — это своего рода испытание Бернулли: с вероятностью $1/2$ он выбирает один из поездов и делает это независимо от остальных. Отсюда видно, что S_n имеет биномиальное распределение с параметрами 1000 и $1/2$. В переводе на язык вероятностей вопрос «сколько нужно мест, чтобы наверняка перевезти 90% клиентов» приобретает вид «найти m , такое что $P(S_n \leq m) \geq 0,9$ ». Причём это m хочется сделать поменьше. Знак \geq стоит в условии вместо $=$ потому, что $P(S_n \leq m)$ вместе с m меняется скачками, и для условия со знаком « \Rightarrow » решения может и не быть вовсе. Попробуем отыскать m :

$$P(S_n \leq m) = P\left(\bigcup_{k=0}^m S_n = k\right) = \sum_{k=0}^m P(S_n = k) = \sum_{k=0}^m C_n^k \left(\frac{1}{2}\right)^n.$$

n у нас, вообще-то, тысяча, так что в итоге приходим к неравенству

$$\sum_{k=0}^m C_{1000}^k \left(\frac{1}{2}\right)^{1000} \geq 0,9 \iff \sum_{k=0}^m C_{1000}^k \geq 2^{1000} \cdot 0,9.$$

Решить его не так, чтобы очень сложно, но и не так, чтобы совсем просто. При наличии некоторой хитрости довольно быстро получается наименьшее из всех возможных m — 520. А вовсе не 900, как хотелось бы предположить. Всякий раз, когда возникают небольшие трудности, приятней их обойти. Немного ниже будет показано, как это делается в данном случае (с.52).

5.6 Локальная предельная и интегральная теоремы Муавра-Лапласа

Первая из теорем приводится без доказательства.

Теорема 5.4 (Локальная предельная теорема Муавра-Лапласа). Пусть S_n — число успехов в n испытаниях Бернулли с вероятностью успеха p . Если $np(1-p) \xrightarrow{n \rightarrow \infty} \infty$, то

$$\forall m \in \mathbb{Z}: 0 \leq m \leq n \quad \mathbb{P}(S_n = m) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2}} \left(1 + \underline{O}\left(\frac{1}{\sigma}\right) \right),$$

где $x = \frac{m - np}{\sigma}$, а $\sigma = \sqrt{DS_n} = \sqrt{np(1-p)}$.

С её помощью несложно доказывается интегральная теорема.

Теорема 5.5 (Интегральная теорема Муавра-Лапласа). Если выполнено условие локальной теоремы и C — произвольная положительная константа, то равномерно по a и b из отрезка $[-C, C]$ (пусть $b \geq a$)

$$\mathbb{P}\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx.$$

Замечание. Вообще, утверждение верно $\forall a, b \in \mathbb{R}$. Здесь излишнее ограничение присутствует для простоты доказательства.

Доказательство. Обозначим за q величину $1-p$ и проведём небольшие выкладки с использованием локальной предельной теоремы в переходе \diamond . А переход $*$ — это на самом деле перенос знака.

$$\mathbb{P}\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) = \mathbb{P}(np + a\sqrt{npq} \leq S_n \leq np + b\sqrt{npq}) \stackrel{*}{=}$$

Пусть множество M — это все целые числа m , такие что выполнено неравенство в последней вероятности. То есть,

$$\begin{aligned} M &= \{m: np + a\sqrt{npq} \leq m \leq np + b\sqrt{npq}\} \\ &\Downarrow \\ &\stackrel{*}{=} \sum_{m \in M} \mathbb{P}(S_n = m) \stackrel{\diamond}{=} \sum_{m \in M} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x_m^2}{2}} \left(1 + \underline{O}\left(\frac{1}{\sigma}\right) \right), \end{aligned}$$

под x_m понимается $\frac{m - np}{\sigma}$. Обозначим как Δx_m разность $x_m - x_{m-1}$. Присмотримся к этой разности внимательней:

$$\Delta x_m = x_m - x_{m-1} = \frac{m - np}{\sigma} - \frac{m - 1 - np}{\sigma} = \frac{m - np - m + 1 + np}{\sigma} = \frac{1}{\sigma}.$$

Так что после замены в последней сумме $\frac{1}{\sigma}$ на Δx_m становится очевидно, что она — привычная интегральная сумма Римана для интегрируемой на любом отрезке функции $e^{-\frac{x^2}{2}}$ плюс ещё $O\left(\frac{1}{\sigma}\right)$, которое исчезнет и вот почему:

$$\Delta x_m = \frac{1}{\sigma} = \frac{1}{\sqrt{np(1-p)}} \xrightarrow{n \rightarrow \infty} 0 \text{ по условию локальной теоремы.}$$

Так что, во-первых, всё лишнее исчезает, а, во-вторых, диаметр разбиения стремится к нулю. Поэтому на отрезке $[-C, C]$ сумма сходится к интегралу:

$$\sum_{m \in M} \frac{1}{\sqrt{2\pi}} \Delta x_m e^{-\frac{x_m^2}{2}} (1 + O(\Delta x_m)) \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx.$$

Ну а на любом подотрезке, таким образом, тоже неминуемо сойдётся. Причём в силу свойств сумм Римана, разность между суммой и предельным значением будет меньше ε при n не меньше некоторого N_ε , которое для всех допустимых a и b будет одним и тем же — это и есть равномерность.

Если кто забыл, почему это так, рекомендуется вспомнить один из критериев интегрируемости. Нужен тот, в котором речь идёт о пределе разности верхней и нижней сумм Дарбу. Всё, здесь воспоминания об этом разделе мат. анализа заканчиваются.

А утверждение тем самым доказано. \square

5.7 Задача о конкуренции (развязка)

Замечание. При прямом решении задачи о конкуренции возник вопрос расправы с неравенством мало того, что с очень большими числами, так ещё и не очень-то охотно поддающемся каким-то преобразованиям. Это настолько усложняет расчёты, что ручной труд исключается. С помощью последней теоремы (правда, с условием для вещественной оси) и таблиц для функции нормального распределения (с.??) задача решается даже без калькулятора (почти). Сложности приятно обходятся следующим образом:

$$0,9 \leq P(S_n \leq m) = P\left(-\infty < \frac{S_n - np}{\sigma} \leq \frac{m - np}{\sigma}\right) \approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{m - np}{\sigma}} e^{-\frac{x^2}{2}} dx.$$

Здесь n достаточно велико, чтобы пользоваться интегральной теоремой (каков критерий применимости на практике — нужно выяснять на кафедре, в разных источниках встречаются разные советы по этому поводу). Дальше надо брать таблицу значений функции стандартного нормального распределения (этот интеграл с дробью — она и есть) и подбирать подходящее m . Получится, что

$$\frac{m - np}{\sigma} \geq 1,282, \text{ значит, } m \geq np + 1,282\sigma = 500 + 1,282 \cdot \sqrt{250} = 520,27.$$

Значит, надо в качестве m брать 521. Между прочим, от точного ответ отличается только на 1, а это меньше, чем 0,2% разницы. Так что получается весьма точный результат, причём довольно быстро и безболезненно.

Лекция 6

6.1 Задача о различении двух гипотез о доле шаров в урне

На этой лекции произойдёт небольшое вторжение в математическую статистику: речь идёт о проверке статистических гипотез (с.??, если не терпится).

Пусть имеется урна с чёрными и белыми шарами. Она обладает тем свойством, что шансы всех шаров внутри неё оказаться снаружи абсолютно равны, если мы проводим эксперимент по извлечению шара. Пусть есть две гипотезы о том, какова доля белых шаров в урне: первая гласит, что эта доля равна p_0 (саму гипотезу обозначим как H_0), вторая — что доля составляет p_1 (и обозначается соответственно H_1). Разумеется, считаем $p_0 \neq p_1$. Пусть далее для определённости $p_1 > p_0$. Требуется установить, какая из гипотез ближе к реальности.

Проведём выборку с возвращением шаров из урны (конечно, для данной задачи было бы проще высыпать их все и пересчитать по цветам, но природа иногда ставит задачи с такой же моделью, где высыпать все «шары» никак нельзя). Пусть её объём — n . Мы n раз вытащили шар, зафиксировали его цвет и положили шар обратно. Нужно одну из гипотез признать более вероятной, как-то основываясь на полученной выборке. Понятно, что поделив число белых шаров выборки на n точно равного p_0 или p_1 числа мы можем и не получить. Может вообще случиться так, что экспериментатор n раз вытащит один и тот же шар — тут никаких гарантий нет. Другое дело, что вероятность такого «неправильного» события может оказаться крайне малой.

Итак, имеем выборку размера n , то есть последовательность элементов, каждый из которых может принимать два значения. Одно соответствует чёрному шару, второе — белому. Обозначим тот вариант последовательности, что получится после эксперимента, как A — событие такое. Разумно правдоподобие гипотез устанавливать на основе следующих условных вероятностей:

$$h_0 = P(A|H_0) \text{ и } h_1 = P(A|H_1).$$

Если h_0 больше, то «побеждает» H_0 , иначе H_1 лучше. Вопрос в том, что будет при равенстве $h_0 = h_1$ и в каких вообще отношениях состоят h_0 и h_1 .

Рассмотрим случай, когда $A = \{\text{БББ} \dots \text{БББ}\}$ — все шары получились белыми. Вычислим вероятности: $P(A|H_0)$ означает вероятность цепочки A при верной H_0 , а $P(A|H_1)$ — вероятность цепочки A при верной H_1 . То есть,

$$h_0 = P(\{\text{БББ} \dots \text{БББ}\}|H_0) = p_0^n, \quad h_1 = P(\{\text{БББ} \dots \text{БББ}\}|H_1) = p_1^n.$$

Поскольку $p_1 > p_0$, то $h_1 > h_0$ и H_1 правдоподобней. С цепочкой $\{\text{ЧЧЧ} \dots \text{ЧЧЧ}\}$ всё будет точно наоборот — «победит» H_0 . Дальше от одной из этих цепочек (пойдём от белой) будем постепенно двигаться к другой. То есть, по одному будем менять цвета шаров (у нас с белого на чёрный) в цепочке. В зависимости от соотношения между

p_0 и p_1 , на каком-то из шаге этого «перекрашивания» знак неравенства между h_1 и h_0 изменится (в нашем случае с $>$ на $<$). Возможно, он сделает это с переходом через равенство, но в случае такой цепочки указанный подход не позволит различить гипотезы H_0 и H_1 . Так вот пусть в этой цепочке, при переходе через которую знак меняется, b_c чёрных шаров. Это число будем называть **критическим**. Теперь правило, по которому принимается решение о том, какая гипотеза лучше, выглядит так: если чёрных шаров в выборке больше критического, то правдоподобней H_0 , если меньше — H_1 . Если число чёрных шаров равно критическому, то если $h_1 \neq h_0$, решение принять просто: по принципу сравнения h_0 и h_1 . Но может случиться так, что $h_1 = h_0$. В этом случае статистика призывает подбросить монетку (но только не обязательно правильную) — об этом подробно позже, в разделе статистики (если не терпится, то на с.??). Пока предположим, что количество чёрных шаров не равно критическому. Так вот правило, по которому решается судьба гипотез, называется **критерием**.

6.1.1 Ошибки первого и второго рода

Уже говорилось о том, что экспериментатор может вытащить один и тот же шар n раз, получить маловероятную цепочку и сделать неправильные выводы. Гарантировать то, что этого не произойдёт, нельзя. Но какой-то способ контроля таких ошибок должен быть! Для начала следует лучше разобраться с тем, что же мы понимаем под словом «ошибка». Итак, пусть в цепочке, которую получил экспериментатор, чёрных шаров имеется b штук. Ошибка произойдёт, если наше правило даст неверный результат, а именно если $b > b_c$ при на самом деле верной H_1 или $b < b_c$ при на самом деле верной H_0 . Случай $b = b_c$ пока оставим за кадром. Вероятность первой ошибки записывается как $P(b > b_c | H_1)$, но поскольку событие $b > b_c$ означает признание гипотезы H_0 , то эта вероятность будет записываться короче как $P(H_0 | H_1)$. Вероятность второй ошибки — как $P(H_1 | H_0)$.

Теперь проведём небольшую дискриминацию: будем считать, что мы проверяем, верна ли гипотеза H_0 . А гипотезу H_1 будем называть конкурирующей. В такой ситуации несложно понять следующих два определения:

Определение 6.1. При проверке гипотезы H_0 происходит *ошибка первого рода* тогда и только тогда, когда отвергается верная гипотеза H_0 .

Определение 6.2. При проверке гипотезы H_0 происходит *ошибка второго рода* тогда и только тогда, когда принимается неверная гипотеза H_0 .

Что можно сделать строго и точно, так это посчитать вероятности этих ошибок.

Определение 6.3. Вероятность ошибки первого рода называется *значимостью* или *уровнем значимости* критерия и обычно обозначается как α :

$$\alpha = P(H_1 | H_0).$$

С вероятностью ошибки второго рода связано другое понятие. Здесь оно не требуется и договоримся о том, что её будем обозначать как β :

$$\beta = P(H_0 | H_1).$$

Ну а контроль за этими ошибками можно осуществлять посредством условий на эти вероятности. Потребовать, например, чтобы $\alpha < 0,001$ и $\beta < 0,001$.

Замечание. Искать какой-то алгебраической связи между α и β в общем случае лучше и не пытаться. За символами H_0 и H_1 могут оказаться не только гипотезы о биномиальном и гипергеометрическом распределении, но и что-нибудь пострашней. И вот при этих в общем случае заведомо неизвестных гипотезах о распределении надо вычислять вероятности того, что в результате эксперимента получаются те или иные значения параметра, по которому мы принимаем решение. В общем случае это сделать-то не всегда толком получается! Так что вопрос о какой-то простой связи между α и β отбрасывается.

При фиксированных возможностях экспериментатора нельзя сделать α и β сколько угодно малыми. В случае урны с шарами для уменьшения α и β нужно увеличивать длину выборки — сделать ряд из цветов вынутых шаров длинней.

При построении критерия α обычно считается известным, а β пытаются сделать поменьше. Так получают правило для проверки гипотезы на основе результата эксперимента. Причём вероятности ошибок при использовании этого правила известны, их можно цивилизованно учесть.

6.1.2 Оценка для числа наблюдений, необходимых для различения гипотез с заданной точностью

Пусть некто зафиксировал для α и β верхние границы α_0 и β_0 , предоставил две гипотезы насчёт доли белых шаров в урне и потребовал критерий для их различения. Поскольку сделать α и β одновременно сколько угодно малыми при фиксированном объёме выборки заведомо не получится, мы в ответ потребуем сделать выборку достаточно длинной. Некто в ответ спросит, насколько длинной она должна быть. Поскольку на вытаскивание каждого шара, фиксацию его цвета и возвращение в урну тратятся силы, время и финансы, то некто требует минимальную необходимую длину. Выясним, как разобраться с этим вопросом.

Далее вместо «вероятность ошибки i -го рода» будет употребляться «ошибка i -го рода». Будем искать критерий с ошибкой первого рода, равной α_0 и ошибкой второго рода, не превышающей β_0 .

Пусть H_1 гласит, что доля белых шаров в урне равна p_1 , а H_0 — что она составляет p_0 . Не ограничивая общности, будем считать, что $p_1 > p_0$. Пусть в выборке оказалось m белых шаров. Если верна H_1 , то m — случайная величина, имеющая биномиальное распределение с параметрами n и p_1 . Если H_0 — то с параметрами n и p_0 . Представим себе, что критерий построен и мы принимаем H_1 , если $m > m_{\text{крит}}$. Вычислим ошибку первого рода, обозначив $1 - p_0$ как q_0 :

$$P(m > m_{\text{крит}} | H_0) = P\left(\frac{m - np_0}{\sqrt{np_0q_0}} > \frac{m_{\text{крит}} - np_0}{\sqrt{np_0q_0}}\right) = 1 - P\left(\frac{m - np_0}{\sqrt{np_0q_0}} \leq \frac{m_{\text{крит}} - np_0}{\sqrt{np_0q_0}}\right).$$

Интуиция подсказывает, что некто потребует качественный критерий. Это означает, что ошибки должны будут оказаться маленькими. Кроме того, интуиция уверена в том, что малость ошибок влечёт за собой весьма и весьма большую длину выборки. А раз длина выборки «весьма» велика, можно воспользоваться интегральной теоремой Муавра-Лапласа:

$$\alpha = 1 - P\left(\frac{m - np_0}{\sqrt{np_0q_0}} \leq \frac{m_{\text{крит}} - np_0}{\sqrt{np_0q_0}}\right) \approx 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{m_{\text{крит}} - np_0}{\sqrt{np_0q_0}}} e^{-\frac{x^2}{2}} dx = \alpha_0.$$

Подобным образом поступим с ошибкой второго рода ($q_1 = 1 - p_1$):

$$\beta \leq \mathbf{P}(m \leq m_{\text{крит}} | H_1) = \mathbf{P}\left(\frac{m - np_1}{\sqrt{np_1q_1}} \leq \frac{m_{\text{крит}} - np_1}{\sqrt{np_1q_1}}\right) \approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{m_{\text{крит}} - np_1}{\sqrt{np_1q_1}}} e^{-\frac{x^2}{2}} dx \leq \beta_0.$$

Пусть t_{α_0} и t_{β_0} — решения уравнений

$$1 - \Phi(t_{\alpha_0}) = \alpha_0 \text{ и } \Phi(-t_{\beta_0}) = \beta_0, \text{ где } \Phi(t) = \int_{-\infty}^t e^{-\frac{x^2}{2}} dx.$$

Сопоставив это с выражениями для вероятностей ошибок, получаем

$$\begin{aligned} \frac{m_{\text{крит}} - np_0}{\sqrt{np_0q_0}} = t_{\alpha_0} &\implies m_{\text{крит}} = np_0 + t_{\alpha_0}\sqrt{np_0q_0} \quad \text{и} \\ \frac{m_{\text{крит}} - np_1}{\sqrt{np_1q_1}} \leq -t_{\beta_0} &\implies m_{\text{крит}} \leq np_1 - t_{\beta_0}\sqrt{np_1q_1}. \end{aligned}$$

Остаётся исключить переменную $m_{\text{крит}}$ и затем разрешить полученное неравенство $np_0 + t_{\alpha_0}\sqrt{np_0q_0} \leq np_1 - t_{\beta_0}\sqrt{np_1q_1}$ относительно n , после чего окончательно имеем

$$n \geq \left(\frac{t_{\alpha_0}\sqrt{p_0q_0} + t_{\beta_0}\sqrt{p_1q_1}}{p_1 - p_0}\right)^2.$$

Так что чем ближе гипотезы, тем трудней их различить. Знаменатель это наглядно демонстрирует. Посмотрим, что даёт формула на конкретных числах:

$$\begin{aligned} p_0 = 0,5, q_0 = 0,5, p_1 = 0,6, q_1 = 0,4, \alpha_0 = 0,05, \beta_0 = 0,25, t_{\alpha_0} = 1,645, t_{\beta_0} = 0,674 \\ \Downarrow \\ n \geq \left(\frac{0,822 + 0,33}{0,1}\right)^2 > 132 \implies n \geq 133. \end{aligned}$$

При $n = 144$ $m_{\text{крит}} = 81$.

В жизни эта схема наполняется самыми разными способами. Например, старые и новые лекарства можно сравнивать с помощью такого метода.

6.2 Определение вероятностного пространства в общем случае

Пришло время покончить со счётностью пространства элементарных исходов Ω . Чтобы строго и аккуратно построить для Ω мощности континуум класс событий, нужно для начала разработать почву. Определение σ -алгебры уже встречалось (под видом определения класса событий), и оно остаётся основным. Появится ещё немного другой объект, называемый алгеброй. Далее он призрачно возникнет в одном месте: оказывается, заданную на алгебре функцию, удовлетворяющую первым двум требованиям на вероятность, можно продолжить до полноценной вероятности единственным образом при одном условии на элементы алгебры. Об этом ниже (на с.??, правда, без доказательства). Но в целом алгебра присутствует здесь потому, что понадобится однажды в доказательстве одного важного факта.

Алгебра от σ -алгебры отличается только конечной аддитивностью (вместо счётной):

Определение 6.4. Множество \mathcal{A} подмножеств (не всех!) Ω называется σ -алгеброй $\iff \mathcal{A}$ обладает следующими тремя свойствами:

1°. σ -алгебре принадлежит само Ω :

$$\Omega \in \mathcal{A}.$$

2°. Вместе с любым множеством σ -алгебра содержит его дополнение до Ω .

$$B \in \mathcal{A} \implies \bar{B} \in \mathcal{A}.$$

3°. Свойство счётной аддитивности:

$$\forall i \in \mathbb{N} \quad B_i \in \mathcal{A} \implies \bigcup_{i=1}^{\infty} B_i \in \mathcal{A}.$$

Далее под отрицанием будет пониматься дополнение до Ω . Не более, чем счётное объединение и отрицание далее будут называться **допустимыми в σ -алгебре операциями**.

Замечание. Счётные пересечения тоже не выводят за границу σ -алгебры. Это сразу следует из того, что

$$\bigcap_{i=1}^{\infty} B_i \equiv \overline{\bigcup_{i=1}^{\infty} \bar{B}_i}.$$

Этот факт, в свою очередь, устанавливается по определению объединения, пересечения и дополнения.

Определение 6.5. Множество \mathcal{A}_0 подмножеств (не всех!) Ω называется **алгеброй** $\iff \mathcal{A}_0$ обладает следующими тремя свойствами:

1°. Алгебре принадлежит само Ω :

$$\Omega \in \mathcal{A}_0.$$

2°. Алгебра вместе с любым множеством содержит его дополнение до Ω :

$$B \in \mathcal{A}_0 \implies \bar{B} \in \mathcal{A}_0.$$

3°. Свойство конечной аддитивности:

$$B_1, B_2 \in \mathcal{A}_0 \implies B_1 \cup B_2 \in \mathcal{A}_0.$$

Замечание. Свойство конечной аддитивности сформулировано для двух множеств, однако с помощью мат. индукции может быть получено для любого количества слагаемых, лишь бы количество это было конечным.

Раз уж Ω имеет мощность континуум, то и события в общем случае могут иметь мощность континуум. Уже приводилось множество мощности континуум, для которого не существует меры (с.12). В σ -алгебру событий нельзя включать множества, не имеющих меры, потому что на всех элементах класса событий должна быть определена вероятность. Попробуем определиться, какие же множества можно допускать в

кандидаты на события. Начнём с простого: как известно, в случае отрезков и интервалов меру (которая дальше будет обозначаться буквой μ) можно ввести как модуль разности концов множества: $\mu((a, b)) = \mu([a, b)) = \mu((a, b]) = \mu([a, b]) = |b - a|$, где a и b — соответствующие действительные числа. Если вместо a или b окажется $\pm\infty$, то меру такого множества считаем бесконечной. Сначала на основе таких множеств построим алгебру: пусть $\Omega = \mathbb{R}$ и \mathcal{B}_0 — это все множества вида $(-\infty, a)$, $[a, b)$, $[b, +\infty)$ и всевозможные их конечные объединения. Это определение включает в себя отрицания — это обеспечено тремя видами множеств и возможностью их объединять. Нетрудно установить, что \mathcal{B}_0 удовлетворяет всем условиям определения алгебры. Разберёмся с тем, что из себя представляет σ -алгебра, получаемая из этих множеств.

6.3 Сигма-алгебра, порождённая классом множеств

Пусть у нас есть какие-то множества, состоящие из элементов некоторого множества Ω . Предположим, мы хотим изготовить содержащую их σ -алгебру. Бывает так, что можно в какую-нибудь σ -алгебру добавить ещё какое-нибудь не содержащееся там ранее множество и получить всё то же, что мы хотели изготовить. Поэтому отдельно выделяют σ -алгебру, ничего лишнего не содержащую:

Определение 6.6. σ -алгеброй, порождённой классом множеств \mathcal{E} , называется наименьшая σ -алгебра, содержащая все элементы \mathcal{E} и обозначается $\mathcal{S}(\mathcal{E})$.

Замечание. Вероятно, стоит отметить, что если \mathcal{E} окажется σ -алгеброй, то $\mathcal{S}(\mathcal{E}) = \mathcal{E}$. Это напрашивается само по себе: $\mathcal{E} \subseteq \mathcal{S}(\mathcal{E})$ и является σ -алгеброй. Но чтобы сделать σ -алгебру ещё меньше, придётся выбрасывать какие-то её элементы. Но все они принадлежат \mathcal{E} , так что если кого выбросить, то все элементы \mathcal{E} в полученной σ -алгебре присутствовать не будут. А это уже нарушает определение.

Замечание. Подобно тому, как доказывалась конечная аддитивность вероятности, легко показывается, что в σ -алгебре, порождённой классом множеств \mathcal{E} , присутствуют всевозможные конечные объединения множеств из \mathcal{E} — хвост множеств в счётном объединении можно составить из пустых множеств, которые в σ -алгебре есть всегда как $\bar{\Omega}$.

Следующий интуитивно понятный факт приводится без доказательства:

Утверждение 6.1. σ -алгебра, порождённая \mathcal{E} , всегда существует и совпадает с пересечением всех σ -алгебр, содержащих \mathcal{E} .

6.4 Борелевская сигма-алгебра

На этой лекции везде, где речь заходит об алгебре либо σ -алгебре, порождённой числовыми множествами, в качестве Ω подразумевается \mathbb{R} , если не оговорено что-нибудь другое.

Определение 6.7. Борелевская σ -алгебра — это σ -алгебра, порождённая всеми открытыми множествами вещественной прямой. Обозначается символом \mathfrak{B} .

Определение 6.8. Элемент \mathfrak{B} называется борелевским множеством.

Без специальной подготовки представить себе неборелевское множество крайне трудно. С очень большой вероятностью, любое множество, которое студент может себе представить за минуту, окажется борелевским. Неизмеримое множество со с.12 борелевским не является, но и представить его себе не так-то просто.

Утверждение 6.2. σ -алгебра, порождённая \mathcal{B}_0 , совпадает с \mathfrak{B} . Или, в кратких обозначениях

$$\mathcal{S}(\mathcal{B}_0) = \mathfrak{B}.$$

Доказательство. Покажем, что $\mathcal{S}(\mathcal{B}_0) \subseteq \mathfrak{B}$ и $\mathfrak{B} \subseteq \mathcal{S}(\mathcal{B}_0)$.

Первое включение следует из того, что все полуинтервалы и отрезки присутствуют в \mathfrak{B} : счётное (и конечное) пересечение за пределы \mathfrak{B} не выводят, а

$$[b, a) = \bigcap_{n=1}^{\infty} \left(b - \frac{1}{n}, a \right).$$

Это значит, что \mathfrak{B} является в том числе и σ -алгеброй, порождённой всеми множествами из \mathcal{B}_0 . А значит, она включает в себя $\mathcal{S}(\mathcal{B}_0)$.

Второе включение следует из того, что все открытые множества прямой присутствуют в $\mathcal{S}(\mathcal{B}_0)$. Любое открытое множество можно представить как счётное объединение интервалов. Например, взять рациональные точки с нужного размера окрестностями и объединить. А любой интервал можно представить как счётное объединение полуинтервалов:

$$(b, a) = \bigcup_{n=1}^{\infty} \left[b + \frac{1}{n}, a \right).$$

Так что любое открытое множество есть в $\mathcal{S}(\mathcal{B}_0)$. Ну а вместе со всеми открытыми множествами в порождённой σ -алгебре неминуемо появится и \mathfrak{B} . Таким образом, второе включение обосновано. \square

Пример. Борелевскую σ -алгебру можно спокойно использовать в качестве класса событий. В качестве Ω возьмём \mathbb{R} , а вероятность для любого открытого множества определим как интеграл по этому множеству функции

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

Все аксиомы будут выполнены. Кстати, вероятность любого множества, состоящего из одной или нескольких точек, получается равной нулю. Может показаться странным, что любые события состоят из событий с нулевой вероятностью. Тем не менее, «любое» событие может состоять из континуума точек. Как видно, в таком случае точное равенство нулю вероятности составляющих событий не гарантирует равенства нулю вероятности их объединения.

Пример. Кроме \mathfrak{B} в качестве класса событий можно использовать массу кажущихся другими конструкций. В качестве Ω опять возьмём \mathbb{R} , в качестве класса событий — $\mathcal{S}(\mathcal{B}_1)$, где \mathcal{B}_1 — это совокупность всех множеств вида $(-\infty, a)$. В качестве вероятности берём всё тот же интеграл по множеству. В частности, для того же $(-\infty, a)$ получаем

$$P((-\infty, a)) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-\frac{x^2}{2}} dx.$$

Покажем, что $\mathcal{S}(\mathcal{B}_1) = \mathfrak{B}$. Заметим, что $\forall a \in \mathbb{R} \quad \overline{(-\infty, a)} = [a, +\infty)$, так что все множества вида $[a, +\infty)$ в $\mathcal{S}(\mathcal{B}_1)$ тоже присутствуют. При любых действительных a и b , таких что $b > a$, $[a, b) = \overline{(-\infty, a)} \cap [b, +\infty)$, так что все множества вида $[a, b)$ тоже присутствуют в $\mathcal{S}(\mathcal{B}_1)$. Всё вместе это означает, что $\mathcal{S}(\mathcal{B}_1) = \mathcal{S}(\mathcal{B}_0) = \mathfrak{B}$. Поэтому что $\mathcal{S}(\mathcal{B}_1)$, что $\mathcal{S}(\mathcal{B}_0)$, что \mathfrak{B} — это всё равно и все множества одного из этих классов можно получать из множеств другого из этих классов с помощью допустимых в σ -алгебре операций. Этот пример будет уместно вспомнить во время разговора о функции распределения (с.62).

6.5 Случайные величины как измеримые отображения

Здесь, наконец, будет разъяснено, что же такое измеримость. Приводимое определение — это измеримость для частного случая отображений (из σ -алгебры событий в \mathbb{R}). Общий случай здесь не понадобится. Так что мы под измеримостью будем понимать следующее:

Определение 6.9. Пусть задано вероятностное пространство $(\Omega, \mathcal{A}, \mathbb{P})$. Отображение $X: \Omega \rightarrow \mathbb{R}$ называется **\mathcal{A} -измеримым (измеримым относительно \mathcal{A} или для краткости просто измеримым)** тогда и только тогда, когда для любого борелевского множества B прообраз $X^{-1}(B)$ принадлежит классу событий \mathcal{A} :

$$\forall B \in \mathfrak{B} \quad X^{-1}(B) = \{\omega: X(\omega) \in B\} \in \mathcal{A}.$$

Замечание. В общем-то, смысл всей затеи в том, чтобы факт $X \in B$ ($B \in \mathfrak{B}$) в терминах вероятностного пространства был событием. То есть, чтобы взятые вместе ω , для которых это возможно (а это и есть прообраз B), образовывали бы множество из класса событий: $\{\omega: X(\omega) \in B\} \in \mathcal{A}$. В таком случае можно будет говорить о вероятности события, например, в духе «потери не превысят 9000\$», где потери вычисляются как случайная величина. Вот и всё.

Любое измеримое отображение из Ω в \mathbb{R} — это случайная величина. Если быть поостроже, то

Определение 6.10. Пусть задано вероятностное пространство $(\Omega, \mathcal{A}, \mathbb{P})$. Отображение $X: \Omega \rightarrow \mathbb{R}$ называется **случайной величиной** тогда и только тогда, когда оно измеримо относительно \mathcal{A} . То есть,

$$\forall B \in \mathfrak{B} \quad X^{-1}(B) = \{\omega: X(\omega) \in B\} \in \mathcal{A}.$$

В общем-то, определение измеримости всего-то скопировано. А вот то, что будет написано далее — уже не просто копия.

Замечание. Понадобится то обстоятельство, что отрицание прообраза множества — это прообраз отрицания множества и объединение прообразов множеств — это прообраз объединения множеств. То есть, если I — некоторое множество индексов (не более, чем счётное), то

$$\overline{\{\omega: X(\omega) \in B\}} = \{\omega: X(\omega) \in \overline{B}\} \quad \text{и} \quad \bigcup_{i \in I} \{\omega: X(\omega) \in B_i\} = \left\{ \omega: X(\omega) \in \bigcup_{i \in I} B_i \right\}.$$

Это обстоятельство легко проверяется с помощью определений объединения и дополнения.

Утверждение 6.3. Пусть задано вероятностное пространство $(\Omega, \mathcal{A}, \mathbb{P})$ и отображение $X: \Omega \rightarrow \mathbb{R}$. Если все множества вида $\{\omega: X(\omega) < a\}$, где $a \in \mathbb{R}$, принадлежат \mathcal{A} , то X — случайная величина. Или, в кратких обозначениях,

$$\forall a \in \mathbb{R} \quad \{\omega: X(\omega) < a\} \in \mathcal{A} \implies \forall B \in \mathfrak{B} \quad \{\omega: X(\omega) \in B\} \in \mathcal{A}.$$

Доказательство. Из прообразов лучей $(-\infty, a)$ можно с помощью отрицания и объединения (не более, чем счётного) делать прообразы любых борелевских множеств. Это следует из того, что, во-первых, любое борелевское множество можно получить с помощью отрицания и не более, чем счётного объединения из лучей вида $(-\infty, a)$ (доказательство имеется в примере на с.59). А, во-вторых, отрицание и объединение можно выносить за скобки множеств-прообразов (это только что было в замечании). Так что теперь берём любое борелевское множество, получаем его допустимыми операциями из лучей, выносим операции за скобки множеств-прообразов и в общем итоге получаем прообразы лучей, над которыми производятся допустимые в σ -алгебре операции.

Но поскольку у X прообразы множеств лучей лежат в классе событий, то в классе событий окажутся и прообразы всех борелевских множеств, которые получились из событий-прообразов лучей с помощью допустимых в σ -алгебре операций. Так что X измеримо относительно \mathcal{A} . Значит, X является случайной величиной. \square

Пример. Как и в случае менее мощного Ω , константа — всегда случайная величина. Потому что у константы прообразов бывает два: \emptyset и Ω . Пусть $X(\omega) \equiv C$:

$$\forall B \in \mathfrak{B} \quad \{\omega: X(\omega) \in B\} = \begin{cases} \Omega, & \text{если } C \in B, \\ \emptyset, & \text{в противном случае.} \end{cases}$$

Ну а в σ -алгебре событий Ω и $\bar{\Omega}$ присутствуют всегда. Тем самым измеримость доказана.

Замечание. В случае не более, чем счётного Ω , не было множеств, для которых не существовало меры и класс событий строился как множество всех подмножеств. В случае Ω мощности континуум появилось неизмеримое подмножество, из-за которого весь сыр-бор и загорелся. Понятие σ -алгебры согласовано с понятием вероятности — проблемы проистекают от свойства счётной аддитивности, с ним и конфликтует предположение об измеримости описанного на с.12 множества. Это множество — объединение мощности континуум борелевских множеств (точек), которое, как видно, выводит за пределы σ -алгебры.

Все неудобные множества из класса событий исключаются, тем самым устраняются вопросы и обрекаются на провал попытки обрушить построенную теорию каким-нибудь шокирующим примером или рассуждением. Все результаты, полученные для случайных величин ранее, переносятся на случай Ω мощности континуум, их доказательства постепенно появятся.

Ну, а множество всех подмножеств для не более, чем счётного Ω формально является σ -алгеброй, и все рассуждения, приводимые для Ω мощности континуум, можно провести и для не более, чем счётного Ω .

Лекция 7

7.1 Функция распределения случайной величины

Определение 7.1. *Функцией распределения (кумулянтой, кумулятивной функцией распределения) $F_\psi(x)$ случайной величины ψ называется вероятность $P(\psi < x)$, то есть:*

$$F_\psi(x) = P(\psi < x).$$

Определение корректное: все множества вида $(-\infty, x)$ — борелевские, так что если ψ — действительно случайная величина, то прообразы их окажутся в классе событий, и вероятность для них определена.

Замечание. Функция распределения может определяться и иначе: вместо $P(\psi < x)$ может быть $P(\psi \leq x)$. Принципиальных изменений это не вносит, однако существенные различия имеются. При первом варианте определения функция обладает непрерывностью слева, при втором — справа. Иногда это обстоятельство оказывается важным, например условия задач могут подразумевать одно из этих определений и требовать что-нибудь доказать, что для одного из определений неправильно. Чтобы студент в таком случае не оказался в тупике в безысходности, о втором определении надо помнить и быть начеку.

7.1.1 Свойства функции распределения

Везде, где речь идёт о свойствах функции распределения, подразумевается, что ψ — произвольная случайная величина.

1°. *Функция распределения — всегда неубывающая:*

$$\forall x_1, x_2: x_1 \leq x_2 \quad F_\psi(x_1) \leq F_\psi(x_2).$$

Доказательство.

$$\begin{aligned} F_\psi(x_2) - F_\psi(x_1) &= P(\psi < x_2) - P(\psi < x_1) \\ &= P(\psi \in (-\infty, x_1) \cup [x_1, x_2)) - P(\psi \in (-\infty, x_1)) \\ &= P(\psi \in (-\infty, x_1)) + P(\psi \in [x_1, x_2)) - P(\psi \in (-\infty, x_1)) \\ &= P(\psi \in [x_1, x_2)) \geq 0. \end{aligned} \quad \square$$

2°. *Функция распределения непрерывна слева:*

$$\lim_{x \rightarrow x_0 - 0} F_\psi(x) = F_\psi(x_0 - 0) = F_\psi(x_0).$$

Замечание. Вот это свойство и изменяется при определении функции распределения как $P(\psi \leq x)$ на непрерывность справа.

Доказательство. Функция $F_\psi(x)$ монотонна. Так что если доказать, что она стремится к какому-то числу по некоторой последовательности, сходящейся к x слева, то она к нему будет стремиться и по любой другой последовательности, сходящейся к x слева. А это — уже определение предела функции по Гейне, но только опять слева. Итак, рассмотрим разность

$$F_\psi(x) - F_\psi\left(x - \frac{1}{n}\right) = \mathbf{P}(\psi < x) - \mathbf{P}\left(\psi < x - \frac{1}{n}\right) = \mathbf{P}\left(x - \frac{1}{n} \leq \psi < x\right).$$

Имеем невозрастающую (при увеличении n) последовательность событий

$$\left\{\omega : x - \frac{1}{n} \leq \psi(\omega) < x\right\}.$$

Она обладает нулевым пределом:

$$\bigcap_{n=1}^{\infty} \left\{\omega : x - \frac{1}{n} \leq \psi(\omega) < x\right\} = \lim_{n \rightarrow \infty} \left\{\omega : x - \frac{1}{n} \leq \psi(\omega) < x\right\} = \left\{\omega : x \leq \psi(\omega) < x\right\} = \emptyset.$$

Вероятность, как известно, непрерывна по невозрастающей последовательности событий, воспользуемся этим в переходе *:

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(F_\psi(x) - F_\psi\left(x - \frac{1}{n}\right) \right) &= \lim_{n \rightarrow \infty} \mathbf{P}(\psi \in [x - \frac{1}{n}, x)) \\ &\stackrel{*}{=} \mathbf{P}\left(\lim_{n \rightarrow \infty} \left\{\omega : \psi(\omega) \in [x - \frac{1}{n}, x)\right\}\right) = \mathbf{P}(\emptyset) = 0. \end{aligned}$$

Так что, по одной из последовательностей, стремящихся к x слева, F_ψ стремится к своему значению $F_\psi(x)$. \square

3°. Предел функции распределения на $+\infty$ равен одному, на $-\infty$ — нулю.

$$F_\psi(x) \xrightarrow{x \rightarrow +\infty} 1 \text{ и } F_\psi(x) \xrightarrow{x \rightarrow -\infty} 0.$$

Доказательство. Опять нужно воспользоваться непрерывностью вероятности по монотонной последовательности событий: в этот раз последовательности такие:

$$\begin{aligned} B_n &= \left\{\omega : \psi(\omega) \in (-\infty, n)\right\} \implies B_n \subseteq B_{n+1}, \bigcup_{n=1}^{\infty} B_n = \left\{\omega : \psi(\omega) \in \mathbb{R}\right\} = \Omega, \\ C_n &= \left\{\omega : \psi(\omega) \in [-n, +\infty)\right\} \implies C_n \subseteq C_{n+1}, \bigcup_{n=1}^{\infty} C_n = \left\{\omega : \psi(\omega) \in \mathbb{R}\right\} = \Omega. \end{aligned}$$

Обе — неубывающие. Первая из них позволяет отыскать предел $F_\psi(x)$ на $+\infty$:

$$\lim_{n \rightarrow \infty} F_\psi(n) = \lim_{n \rightarrow \infty} \mathbf{P}(B_n) = \mathbf{P}\left(\lim_{n \rightarrow \infty} B_n\right) = \mathbf{P}(\Omega) = 1.$$

Чтобы отыскать предел второй последовательности, функцию распределения отнимем от 1:

$$\begin{aligned} \lim_{n \rightarrow \infty} (1 - F_\psi(-n)) &= \lim_{n \rightarrow \infty} \mathbf{P}(\psi \geq -n) = \lim_{n \rightarrow \infty} \mathbf{P}(C_n) = \mathbf{P}\left(\lim_{n \rightarrow \infty} C_n\right) = \mathbf{P}(\Omega) = 1 \\ &\downarrow \\ \lim_{n \rightarrow \infty} F_\psi(-n) &= 0. \end{aligned}$$

Ну, а поскольку функция распределения всегда монотонна, то, найдя пределы по этим конкретным подпоследовательностям, мы тем самым нашли пределы по всем остальным. Так что вновь по определению Гейне, пределы функции на $+\infty$ и $-\infty$ найдены. \square

7.2 Распределение случайной величины

Далее считается, что задано произвольное вероятностное пространство $(\Omega, \mathcal{A}, \mathbb{P})$ и в нём произвольная случайная величина ψ .

Определение 7.2. *Распределением случайной величины ψ называется отображение $\mathbb{P}_\psi: \mathfrak{B} \rightarrow \mathbb{R}$, при любом $B \in \mathfrak{B}$ равное вероятности $\mathbb{P}(\psi \in B)$. Или, покороче,*

$$\forall B \in \mathfrak{B} \quad \mathbb{P}_\psi(B) = \mathbb{P}(\psi \in B) \text{ — распределение } \psi.$$

Утверждение 7.1. *Тройка $(\mathbb{R}, \mathfrak{B}, \mathbb{P}_\psi)$ образует вероятностное пространство.*

Доказательство. Доказательство целиком заключается в проверке аксиом. Проверить, то, что \mathfrak{B} будет σ -алгеброй на \mathbb{R} — это всё равно что вспомнить определение борелевской σ -алгебры. Так что остаётся убедиться в том, что \mathbb{P}_ψ — вероятность. Для этого нужно проверить три аксиомы вероятности: вероятность Ω — единица, вероятность неотрицательна и счётно-аддитивна. Проверим:

1. $\mathbb{P}_\psi(\mathbb{R}) = \mathbb{P}(\psi \in \mathbb{R}) = 1.$
2. $\forall B \in \mathfrak{B} \quad \mathbb{P}_\psi(B) = \mathbb{P}(\psi \in B) \geq 0.$
3. $\forall \{B_i\}_{i=1}^\infty \in \mathfrak{B}: i, j \in \mathbb{N}, i \neq j \implies B_i B_j = \emptyset$

$$\begin{aligned} \mathbb{P}_\psi\left(\bigcup_{i=1}^\infty B_i\right) &= \mathbb{P}\left(\psi \in \bigcup_{i=1}^\infty B_i\right) = \mathbb{P}\left(\psi^{-1}\left(\bigcup_{i=1}^\infty B_n\right)\right) = \mathbb{P}\left(\bigcup_{i=1}^\infty \psi^{-1}(B_n)\right) \\ &\stackrel{*}{=} \sum_{i=1}^\infty \mathbb{P}(\psi^{-1}(B_n)) = \sum_{i=1}^\infty \mathbb{P}_\psi(B_i). \end{aligned}$$

В переходе * использован тот факт, что прообразы непересекающихся множеств не пересекаются. Предположить противоположное — и получим очевидное противоречие с реальностью $\{\omega: \psi(\omega) \in \emptyset\} \neq \emptyset$. Ну, а тогда можно воспользоваться счётной аддитивностью обычной вероятности.

Таким образом, доказано, что $(\Omega, \mathfrak{B}, \mathbb{P}_\psi)$ — вероятностное пространство. \square

Замечание. В общем-то, дело дошло до того, что можно забыть про исходное вероятностное пространство $(\Omega, \mathcal{A}, \mathbb{P})$ и в нём случайную величину ψ , а оперировать в терминах $(\mathbb{R}, \mathfrak{B}, \mathbb{P}_\psi)$ — с вещественными числами работать удобно и привычно.

Если известно распределение случайной величины, то функция распределения также известна. В каждой точке её можно вычислить через распределение:

$$F_\psi(x) = \mathbb{P}(\psi < x) = \mathbb{P}_\psi((-\infty, x)).$$

Более того, если известна функция распределения, то известно распределение. Но этот факт доказывается уже не так просто. Суть заключается в том, что:

- функция распределения однозначно задаёт вероятность прообразов всех множеств вида $(-\infty, a)$,
- отняв функцию распределения от единицы, получим вероятность прообразов всех множеств вида $[a, +\infty)$,
- отняв от вероятности прообраза $(-\infty, b)$ вероятность прообраза $(-\infty, a)$, где $a < b$, получим вероятность прообраза $[a, b)$.

Появились множества, которые вместе со своими конечными объединениями образуют уже знакомую нам алгебру подмножеств \mathbb{R} \mathcal{B}_0 . А $\mathcal{S}(\mathcal{B}_0) = \mathfrak{B}$. Намёк окончен.

7.3 Теорема о единственности продолжения вероятности с алгебры на порождённую ею σ -алгебру

Теорема 7.1 (О продолжении меры). Пусть \mathcal{F}_0 — алгебра подмножеств Ω , пусть на \mathcal{F}_0 задана функция P и выполнены три условия:

1. $\forall F \in \mathcal{F}_0 \quad P(F) \geq 0$.
2. $P(\Omega) = 1$.
3. Внутри \mathcal{F}_0 P счётно-аддитивна. То есть,

$$\forall \{F_n\}_{n=1}^{\infty} \in \mathcal{F}_0: \forall i, j \in \mathbb{N}: i \neq j \quad F_i F_j = \emptyset \text{ и } \bigcup_{n=1}^{\infty} F_n \in \mathcal{F}_0$$

$$\Downarrow$$

$$P\left(\bigcup_{n=1}^{\infty} F_n\right) = \sum_{n=1}^{\infty} P(F_n).$$

Тогда P единственным образом продолжается до вероятности на $\mathcal{S}(\mathcal{F}_0)$.

7.4 Взаимно однозначное соответствие между функциями распределения и вероятностными распределениями

Теорема 7.2. Функция распределения случайной величины однозначно задаёт её распределение.

Доказательство. Зная функцию распределения, мы можем легко установить вероятности попадания случайной величины (обозначим её как ψ) во все множества следующих видов:

$$P(\psi \in (-\infty, a)) = F_{\psi}(a), \quad P(\psi \in [b, +\infty)) = 1 - F_{\psi}(b) \text{ и } P(\psi \in [a, b)) = F_{\psi}(b) - F_{\psi}(a).$$

Множества этих трёх видов вместе со всевозможными их конечными объединениями образуют алгебру \mathcal{B}_0 , которая подробно обсуждалась на с.58. Там же показано, что $\mathcal{S}(\mathcal{B}_0) = \mathfrak{B}$. Это означает, что $\forall B \in \mathfrak{B}$ вероятность $P(\psi \in B)$ определена и выражается через известные нам вероятности множеств указанных выше видов. А значит, известно распределение. Теорема доказана. \square

Замечание. На самом деле, всё, что произошло в доказательстве, уже встречалось, причём почти в том же контексте. Уверждение на с.61 — практически клон последней теоремы по ходу мыслей. Если что-то осталось неясным, рекомендуется посмотреть туда.

Теперь можно делать вывод о том, что **любое распределение однозначно задаёт функцию распределения и любая функция распределения однозначно задаёт распределение.**

Замечание. Пусть, кто-нибудь задал счётно-аддитивное (внутри \mathcal{B}_0 — о счётных объединениях, оказавшихся вне \mathcal{B}_0 , пока ещё ничего не утверждается) отображение $f: \mathcal{B}_0 \rightarrow [0, 1]$, такое что $f(\Omega) = 1$. Тогда, по теореме о продолжении меры, f единственным образом продолжается до вероятности на $\mathcal{S}(\mathcal{B}_0) = \mathfrak{B}$. Именно единственным, так что для тех счётных объединений, что оказались за границей \mathcal{B}_0 , возможное значение f не только обязательно существует но и единственно.

Утверждение 7.2. Пусть имеется некоторая непрерывная слева неубывающая на всей оси функция f , у которой предел на $+\infty$ равен единице, а на $-\infty$ — нулю. Тогда существуют вероятностное пространство и в нём случайная величина ψ , такие что f — функция распределения ψ .

Замечание. ψ и вероятностное пространство заведомо неединственны. Распределение не фиксирует вид отображения, не говоря уже о том, что о пространстве элементарных исходов ничего сказать не может.

Доказательство. В качестве пространства элементарных исходов проще всего взять \mathbb{R} , в качестве σ -алгебры событий — \mathfrak{B} . Остаётся разобраться с вероятностью и случайной величиной.

Сделаем это следующим образом: будем f трактовать как отображение F из \mathcal{B}_0 в $[0, 1]$, а именно

$$\forall a, b \in \mathbb{R}: b > a \quad F((-\infty, a)) = f(a), \quad F([a, +\infty)) = 1 - f(a) \quad \text{и} \quad F([a, b)) = f(b) - f(a),$$

а для всевозможных конечных объединений множеств этих трёх видов F доопределяем с помощью формулы конечной аддитивности. Для любого конечного объединения множеств из \mathcal{B}_0 она трансформируется в формулу включения-исключения (с.15), и в итоге F будет задано на всей \mathcal{B}_0 :

$$\forall \{A_i\}_{i=1}^n \in \mathcal{B}_0: 1 \leq i, j \leq n, i \neq j \Rightarrow A_i A_j = \emptyset \quad F\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n F(A_i).$$

Если присмотреться, то попросту фабрикуется заданное на всех элементах \mathcal{B}_0 распределение ψ .

Покажем, что F — счётно аддитивно. Для этого докажем пару неравенств: пусть $\{K_n\}_{n=1}^\infty \in \mathcal{B}_0$, $K = \bigcup_{n=1}^\infty K_n$, $K \in \mathcal{B}_0$ и $\forall i, j \in \mathbb{N}: i \neq j \quad K_i K_j = \emptyset$,

$$F(K) \geq \sum_{i=1}^\infty F(K_i) \quad \text{и} \quad F(K) \leq \sum_{i=1}^\infty F(K_i).$$

Первое неравенство. Чем бы на самом деле ни были K_i , они лежат в алгебре \mathcal{B}_0 . Значит, они представимы с помощью конечных объединений из множеств вида $(-\infty, a)$, $[a, b)$ и $[b, +\infty)$ (отрицания не нужны — они тоже выражаются как конечные объединения из множеств этих трёх видов). «Развернём» все K_i в эти конечные объединения — получим новую последовательность попарно непересекающихся множеств L_i . Точно так же поступим и с K . Несмотря на то, что K — счётное объединение непересекающихся множеств указанного вида, оно всё же из \mathcal{B}_0 , а потому раскладывается в их конечное объединение. Конечную последовательность непересекающихся множеств, на которые разбивается K , обозначим как $\{M_j\}_{j=1}^k$. Последовательность из K_i распалась на несколько последовательностей (не исключено,

правда, что она чем была, тем и осталась), объединение каждой из которых сходится к одному из M_j , при этом все K_i и M_j — множества одного из трёх хорошо известных видов.

Поэтому если доказать, что для таких последовательностей выполнена счётная аддитивность, то она будет доказана для любых последовательностей из \mathcal{B}_0 . Потом просто объединим конечное количество этих последовательностей и для F воспользуемся конечной аддитивностью.

Пусть $K = [a, b]$, где $b > a$ — конечные числа. $K_i = [a_i, b_i]$, попарно непересекающиеся множества. Поскольку множества не бывают «разного знака», то в рядах-объединениях множеств порядок объединения можно свободно менять. Строгое обоснование этого факта можно получить по определению объединения множеств. Зафиксируем произвольное n и множества перенумеруем так, чтобы были выполнены условия $a \leq a_1 < b_1 \leq a_2 < b_2 \leq \dots \leq a_n < b_n \leq b$. Получаем, что

$$\sum_{i=1}^n F(K_i) = \sum_{i=1}^n (f(b_i) - f(a_i)) \stackrel{*}{\leq} f(b) - f(a) = F(K).$$

В переходе $*$ использована монотонность f . Поскольку это верно при любом n , то можно учинить переход к пределу с обеих сторон знака неравенства. И первое неравенство обосновано. Если вдруг a либо b окажется бесконечным, то изменения в ходе мыслей будут число символическими.

Второе неравенство. По тем же простым соображениям, будем работать со случаем $K = [a, b]$ и $K_i = [a_i, b_i]$. В силу непрерывности слева функции f ,

$$\forall \varepsilon > 0, \forall i \in \mathbb{N} \exists a'_i < a_i: f(a'_i) > f(a_i) - \frac{\varepsilon}{2^{i+1}}, \exists b' < b: f(b') > f(b) - \frac{\varepsilon}{2}.$$

$$[a, b'] \subseteq [a, b] = \bigcup_{i=1}^{\infty} [a_i, b_i] \subseteq \bigcup_{i=1}^{\infty} (a'_i, b_i).$$

$[a, b']$ — это компактное множество. Большое объединение справа — это его открытое покрытие. В такой ситуации возникает соблазн выбрать из открытого покрытия конечное подпокрытие. Выберем:

$$\exists m \in \mathbb{N}: [a, b'] \subseteq \bigcup_{i=1}^m (a'_i, b_i).$$

Интервалы (a'_i, b_i) уже могут пересекаться, а объединение их содержит в себе $[a, b']$. Продемонстрируем, что

$$\sum_{i=1}^m (f(b_i) - f(a'_i)) - f(b') + f(a) \geq 0. \quad (**)$$

Факт следует, по сути, из того, что f неубывает. Перенумеруем интервалы из покрытия в том порядке, в котором они встречаются на вещественной оси, если идти слева-направо. Картина будет примерно следующей:

$$a'_1 < a \leq a'_2 < b_1 < a'_3 < b_2 < a'_4 < \dots < b_{m-3} < a'_{m-1} < b_{m-2} < a'_m < b_{m-1} \leq b' < b_m.$$

Теперь перепишем выражение в левой части $**$, так что неравенство станет очевидным: каждая разность в сумме, что идёт после ♣, неотрицательна:

$$\begin{aligned} \sum_{i=1}^n (f(b_i) - f(a'_i)) - f(b') + f(a) &\stackrel{\clubsuit}{=} f(b_m) - f(b') + f(b_{m-1}) - f(a'_m) + f(b_{m-2}) - f(a'_{m-1}) + f(b_{m-3}) - \dots \\ &\dots - f(a'_4) + f(b_2) - f(a'_3) + f(b_1) - f(a'_2) + f(a) - f(a'_1) \geq 0. \end{aligned}$$

Для завершения мысли вспомним, что $f(a'_i) > f(a_i) - \frac{\varepsilon}{2^{i+1}}$ и $f(b') > f(b) - \frac{\varepsilon}{2}$:

$$\begin{aligned}
f(b) - f(a) - \frac{\varepsilon}{2} &< f(b') - f(a) \leq \sum_{i=1}^n (f(b_i) - f(a'_i)) < \sum_{i=1}^{\infty} (f(b_i) - f(a'_i)) \\
&< \sum_{i=1}^{\infty} \left(f(b_i) - f(a_i) + \frac{\varepsilon}{2^{i+1}} \right) \\
&\Downarrow \\
f(b) - f(a) - \frac{\varepsilon}{2} &< \sum_{i=1}^{\infty} \left(f(b_i) - f(a_i) + \frac{\varepsilon}{2^{i+1}} \right) = \sum_{i=1}^{\infty} (f(b_i) - f(a_i)) + \frac{\varepsilon}{2} \\
&\Downarrow \\
\forall \varepsilon > 0 \quad f(b) - f(a) - \varepsilon &< \sum_{i=1}^{\infty} (f(b_i) - f(a_i)).
\end{aligned}$$

В силу произвольности ε в итоге получаем

$$F(K) = f(b) - f(a) \leq \sum_{i=1}^{\infty} (f(b_i) - f(a_i)) = \sum_{i=1}^{\infty} F(K_i).$$

Всё, счётная аддитивность показана. Применяем теорему о продолжении меры и обнаруживаем, что задано отображение из \mathfrak{B} в \mathbb{R} , обладающее всеми свойствами вероятности. Вот её и назовём вероятностью в том пространстве, что мы строим с самого начала доказательства. Тем самым, задано распределение, ну а вместе с ним и функция распределения, которая, если присмотреться, в каждой точке совпадает с f . А случайная величина, кстати, подойдёт тождественная, то есть, $\psi(\omega) = \omega$. Утверждение доказано. \square

Пусть \mathcal{F} — множество всех непрерывных слева неубывающих на \mathbb{R} функций, у которых предел на $+\infty$ равен 1, на $-\infty$ равен 0. Пусть \mathcal{P} — класс всевозможных вероятностных распределений на действительной оси. Последнее утверждение наглядно демонстрирует, что между \mathcal{F} и \mathcal{P} можно построить взаимно-однозначное соответствие, посредством функции распределения. Эти множества эквивалентны. Речь даже заходит об изоморфизме, но для этого сначала нужно ввести групповые операции, а для линейной алгебры места здесь не выделено.

На функцию распределения наложены весьма мягкие условия — она может быть непрерывной на всей оси, может состоять из полуотрезков прямых, параллельных горизонтальной оси (это и наблюдается для дискретных случайных величин), а может просто иметь счётное количество разрывов. Всё это — в рамках наложенных условий.

Лекция 8

8.1 Неравенство и УЗБЧ Колмогорова

8.1.1 Неравенство Колмогорова

Теорема 8.1 (Неравенство Колмогорова). Пусть X_1, X_2, \dots, X_n — независимые случайные величины, $EX_i = 0$, $EX_i^2 < \infty$, $i = 1, \dots, n$. Тогда $\forall a > 0$ справедливо неравенство:

$$P\left(\sup_{1 \leq k \leq n} |X_1 + X_2 + \dots + X_k| > a\right) < \frac{\sum_{i=1}^n EX_i^2}{a^2}.$$

Замечание. Неравенство Чебышева в тех же условиях, таким образом, является частным случаем неравенства Колмогорова, только в неравенстве Чебышева речь идет о сходимости по вероятности, а в неравенстве Колмогорова — почти всюду. Если положить $k = n$, то получится уже знакомое нам неравенство Чебышева.

Доказательство. Положим $S_k = X_1 + X_2 + \dots + X_k$. Вектор из случайных величин (X_1, \dots, X_n) обозначим как Y . Понадобятся следующие четыре факта:

$$EX_i X_j = EX_i \cdot EX_j = 0 \text{ при } i \neq j \implies ES_n^2 = \sum_{i=1}^n EX_i^2; \quad (1)$$

$$E(S_n - S_k) = E \sum_{i=k+1}^n X_i = 0; \quad (2)$$

Введем события $A = \left\{ Y: \sup_{1 \leq k \leq n} |S_k| > a \right\}$ и $A_k = \left\{ Y: \sup_{1 \leq m \leq k-1} |S_m| \leq a, |S_k| > a \right\}$

↓

$$A = \bigcup_{k=1}^n A_k \text{ и } A_i A_j = \emptyset \quad \forall i \neq j; \quad (3)$$

Рассмотрим случайные величины $(S_n - S_k)$ и $S_k I_{A_k}$. Первая из них конструируется как борелевская функция от X_{k+1}, \dots, X_n , вторая — от X_1, \dots, X_k , а случайные величины X_1, \dots, X_n независимы в совокупности. А тогда будут независимыми случайные величины $(S_n - S_k)$ и $S_k I_{A_k}$. А произведение мат. ожиданий независимых случайных величин равно произведению их мат. ожиданий. То есть, верно следующее:

$$\begin{aligned} X_1, \dots, X_n &\text{ — независимые случайные величины} \\ \Downarrow \\ (S_n - S_k) \text{ и } S_k I_{A_k} &\text{ независимы} \\ \Downarrow \\ E((S_n - S_k) \cdot (S_k I_{A_k})) &= E(S_n - S_k) \cdot ES_k I_{A_k}. \end{aligned} \quad (4)$$

Далее происходят небольшие выкладки:

$$\begin{aligned}
\sum_{i=1}^n \mathbb{E}X_i^2 &\stackrel{(1)}{=} \mathbb{E}S_n^2 = \mathbb{E}S_n^2 \cdot 1 \geq \mathbb{E}S_n^2 \cdot I_A = \mathbb{E}S_n^2 \cdot \sum_{k=1}^n I_{A_k} = \sum_{k=1}^n \mathbb{E}(S_k + (S_n - S_k))^2 \cdot I_{A_k} \\
&\geq \sum_{k=1}^n (\mathbb{E}S_k^2 I_{A_k} + 2\mathbb{E}(S_n - S_k)S_k I_{A_k}) \stackrel{(4)}{=} \sum_{k=1}^n (\mathbb{E}S_k^2 I_{A_k} + 2\mathbb{E}(S_n - S_k)\mathbb{E}S_k I_{A_k}) \\
&\stackrel{(2)}{=} \sum_{k=1}^n \mathbb{E}S_k^2 I_{A_k} \stackrel{*}{>} a^2 \sum_{k=1}^n \mathbb{E}I_{A_k} = a^2 \sum_{k=1}^n \mathbb{P}(A_k) \stackrel{(3)}{=} a^2 \mathbb{P}(A).
\end{aligned}$$

Помеченный * переход может вызвать вопросы. Только один из индикаторов в сумме может быть единицей, тогда множитель при нём обязательно превысит a^2 — так определялось событие A_k . Множители же при остальных индикаторах нам вовсе неинтересны, потому что умножаться они всё равно будут на ноль. Сведя конец этой цепочки с ее началом, получим желанное

$$\mathbb{P}\left(\sup_{1 \leq k \leq n} |X_1 + X_2 + \dots + X_k| > a\right) < \frac{\sum_{i=1}^n \mathbb{E}X_i^2}{a^2}. \quad \square$$

8.1.2 Усиленный закон больших чисел Колмогорова

Теорема 8.2 (Усиленный закон больших чисел Колмогорова). Пусть X_1, \dots, X_n — независимые случайные величины

$$\begin{aligned}
&\sum_{n=1}^{\infty} \frac{DX_n}{n^2} < \infty \\
&\quad \Downarrow \\
&\frac{X_1 + X_2 + \dots + X_n}{n} - \frac{\mathbb{E}X_1 + \mathbb{E}X_2 + \dots + \mathbb{E}X_n}{n} \xrightarrow[n \rightarrow \infty]{} 0 \text{ почти всюду.}
\end{aligned}$$

Замечание. В законе больших чисел вместо

$$\sum_{n=1}^{\infty} \frac{DX_n}{n^2} < \infty \text{ было } DX_i \leq C,$$

последнее сильнее первого ограничивает множество подходящих под условие теоремы последовательностей случайных величин.

Доказательство. Центрируем случайные величины X_i :

$$Y_i = X_i - \mathbb{E}X_i; \quad \mathbb{E}Y_i = 0; \quad DY_i = DX_i = \sigma_i^2$$

Обозначим $S_n = Y_1 + Y_2 + \dots + Y_n$ и вектор из величин Y_i как просто Y . Для доказательства теоремы достаточно показать, что

$$\begin{aligned}
\frac{S_n}{n} = \frac{Y_1 + Y_2 + \dots + Y_n}{n} &\rightarrow 0 \text{ почти всюду.} \\
&\Updownarrow \\
\forall \varepsilon > 0 \quad \mathbb{P}\left(\sup_{k \geq n} \frac{|S_k|}{k} > \varepsilon\right) &\xrightarrow[n \rightarrow \infty]{} 0.
\end{aligned}$$

Будем работать с событиями A_k , которые определим следующим образом:

$$A_n = \left\{ Y : \max_{2^{n-1} \leq k < 2^n} \left| \frac{S_k(Y)}{k} \right| > \varepsilon \right\}$$

$$\text{Заметим, что } \mathbb{P}\left(\bigcup_{k=n}^{\infty} A_k\right) \xrightarrow{n \rightarrow \infty} 0 \iff \forall \varepsilon > 0 \quad \mathbb{P}\left(\sup_{k \geq n} \frac{|S_k|}{k} > \varepsilon\right) \xrightarrow{n \rightarrow \infty} 0.$$

Чтобы было верным стоящее слева от знака равносильности утверждение, достаточно показать, что

$$\sum_{k=1}^{\infty} \mathbb{P}(A_k) < \infty.$$

В самом деле:

$$\begin{aligned} \mathbb{P}(A_n) &\leq \mathbb{P}\left(\max_{2^{n-1} \leq k < 2^n} |S_k| > \varepsilon 2^{n-1}\right) \leq \mathbb{P}\left(\max_{k < 2^n} |S_k| > \varepsilon 2^{n-1}\right) \\ &\leq \left\{ \text{По неравенству Колмогорова} \right\} \leq \frac{DS_{2^n}}{\varepsilon^2 2^{2(n-1)}} = 4\varepsilon^{-2} 2^{-2n} \sum_{k \leq 2^n} \sigma_k^2 \end{aligned}$$

Получив оценку сверху для $\mathbb{P}(A_k)$, оценим теперь сверху сумму ряда из этих вероятностей, умноженного на константу:

$$\begin{aligned} \frac{\varepsilon^2}{4} \sum_{n=1}^{\infty} \mathbb{P}(A_n) &\leq \sum_{n=1}^{\infty} 2^{-2n} \sum_{k \leq 2^n} \sigma_k^2 = \sum_{k=1}^{\infty} \sigma_k^2 \sum_{n: 2^n \geq k} 2^{-2n} \\ &\stackrel{*}{\leq} \sum_{k=1}^{\infty} \sigma_k^2 \left(k^{-2} \frac{1}{1 - \frac{1}{4}} \right) = \frac{4}{3} \cdot \sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} < \infty \quad \text{по условию.} \end{aligned}$$

В помеченном * переходе мы поменяли порядок суммирования ряда. Поскольку он знакопостоянный и сходится, то можно менять порядок суммирования, не разрушив сходимость. Остается воспользоваться тем фактом, что остаток сходящегося ряда стремится к нулю:

$$\sum_{k=1}^{\infty} \mathbb{P}(A_k) < \infty \implies \mathbb{P}\left(\bigcup_{k=n}^{\infty} A_k\right) \leq \sum_{k=n}^{\infty} \mathbb{P}(A_k) \xrightarrow{n \rightarrow \infty} 0. \quad \square$$

8.1.3 Применение УЗБЧ: метод Монте-Карло

Пусть дана функция $f(t)$, непрерывная на $[0, 1]$. Не ограничивая общности, будем считать, что $0 \leq f(t) \leq 1$ (в противном случае отнимем от функции ее минимальное на сегменте значение, а потом полученную функцию поделим на ее же максимум, если эта функция — не константа). Окажется, интеграл от этой функции на сегменте $[0; 1]$ можно вычислить с помощью УЗБЧ.

Пусть X и Y — независимые случайные величины, равномерно распределенные на отрезке $[0; 1]$. Введем случайную величину z следующим образом:

$$z = \begin{cases} 1, & \text{если } f(X) > Y \\ 0, & \text{если нет} \end{cases}.$$

Математическое ожидание этой случайной величины оказывается равным интегралу, о котором говорилось выше:

$$Ez = P(f(x) > y) = \int_0^1 f(z) dz.$$

Объясняется это тем, что случайный вектор (X, Y) имеет в квадрате $(0; 0) \times (1; 1)$ на XY -плоскости плотность, равную единице, за пределами квадрата плотность равна нулю. Далее заметим, что из-за выбранного вида величины z её мат. ожидание будет равняться двойному интегралу от единицы по фигуре $\Lambda = \{(x, y): 0 \leq y < f(x), x \in [0; 1]\}$. А поскольку функция на сегменте $[0; 1]$ расположена между нулем и единицей, то фигура Λ — это не что иное, как площадь под графиком $f(x)$. Двойной интеграл от единицы по этой фигуре даст площадь под графиком $f(x)$, то есть тот самый интеграл от $f(x)$ от нуля до одного.

Теперь пусть у нас имеется $2n$ независимых случайных величин, равномерно распределенных на сегменте от нуля до одного: X_1, \dots, X_n и Y_1, \dots, Y_n . Сконструируем из них n случайных величин z_1, \dots, z_n по следующему закону:

$$z_i = \begin{cases} 1, & \text{если } f(X_i) > Y_i \\ 0, & \text{если нет} \end{cases}, \text{ у всех } z_i \text{ одинаковое мат. ожидание: } Ez_i = \int_0^1 f(t) dt.$$

Остается только применить УЗБЧ Колмогорова, предварительно заметив, что дисперсии z_i ограничены, и сами z_i независимы в совокупности. Тогда

$$\frac{z_1 + \dots + z_n}{n} \xrightarrow{n \rightarrow \infty} \int_0^1 f(t) dt \quad \text{почти всюду.}$$

Здесь, однако, имеются серьезные проблемы: во-первых, надо где-то взять очень много независимых равномерно распределенных на сегменте от нуля до одного случайных величин. Это совсем не так просто, как хотелось бы! Кроме того, не приведена оценка погрешности для n -того шага. Так что перед тем, как воспользоваться сим методом, с проблемами этими надо разобраться. Здесь этого не сделано.

8.2 Сходимость в среднем

Определение 8.1. Последовательность X_n *сходится к случайной величине X в среднем порядка k* , где k — натуральное число

$$\begin{aligned} & \Updownarrow \\ E|X_n - X|^k & \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Если $k = 2$, то говорят, что имеется сходимость "в среднем квадратично".

Если $k = 1$, то просто "в среднем".

Утверждение 8.1. Из сходимости в среднем следует сходимость по вероятности:

$$E|X_n - X|^k \xrightarrow{n \rightarrow \infty} 0, k \in \mathbb{N} \implies X_n \xrightarrow{p} X \text{ — стремление по вероятности.}$$

Доказательство.

$$P(|X_n - X|^k > \varepsilon^k) = P(|X_n - X| > \varepsilon) < \frac{1}{\varepsilon^k} E|X_n - X|^k$$

В последнем переходе использовалось неравенство Маркова

$$E|X_n - X|^k \xrightarrow{n \rightarrow \infty} 0 \implies \forall \varepsilon > 0 \quad P(|X_n - X|^k > \varepsilon) \xrightarrow{n \rightarrow \infty} 0. \quad \square$$

Утверждение 8.2. *Обратное неверно.*

Доказательство. Построим последовательность случайных величин, стремящуюся к случайной величине ноль по вероятности, но в среднем расходящуюся:

$$\Omega = [0; 1]; \quad X_n(\omega) = \begin{cases} n, & 0 \leq \omega \leq \frac{1}{n}, \\ 0, & \frac{1}{n} \leq \omega \leq 1. \end{cases}$$

$$E|X_n - 0|^k = \int_0^{1/n} n^k dt = n^{k-1} \geq 1, \quad k \in \mathbb{N}.$$

Стремление построенной последовательности к нулю по вероятности очевидно, а отсутствие сходимости в среднем показано последней формулой. Тем самым, утверждение оправдано. \square

8.3 Производящие функции

Определение 8.2. *Производящей функцией* неотрицательной целочисленной случайной величины x называется функция $\varphi_x(z) = Ez^x$; $|z| \leq 1$; $z \in \mathbb{C}$.

8.3.1 Свойства производящих функций

1°. Указанное выше мат. ожидание всегда существует, поскольку по неравенству треугольника (ну, или свойству интегралов)

$$|Ez^x| \leq E|z|^x \leq 1, \quad \text{потому что } |z| \leq 1.$$

Значит, для неотрицательной целочисленной случайной величины **производящая функция всегда существует.**

Замечание. В открытом круге $|z| < 1$ (речь идет о комплексной плоскости) $\varphi_x(z)$ является равномерно сходящимся рядом. Каждое слагаемое можно дифференцировать бесконечное количество раз. При этом те, кто помнят неравенство Коши-Адамара, заметят, что ряд из производных любого порядка в этом круге равномерно сходится, а значит, производящую функцию можно дифференцировать и производная равна ряду из производных того же порядка слагаемых. А вообще, это — степенной ряд. Этот факт неявно используется при доказательстве следующего утверждения.

2°. Пусть $\varphi_x(z)$ — производящая функция случайной величины x , тогда

$$P(x = k) = \frac{\varphi_x^{(k)}(0)}{k!}, \quad k \in \mathbb{Z}, \quad k \geq 0 \quad \text{— известна функция распределения } x.$$

То есть, по производящей функции однозначно восстанавливается распределение случайной величины.

Доказательство. По определению распишем математическое ожидание случайной величины для дискретного распределения:

$$Ez^x = P(x = 0) + zP(x = 1) + z^2P(x = 2) + \dots = \varphi_x(z).$$

Дальше действуем по индукции либо так: положим $z = 0$ и обнаружим факт $P(x = 0) = \varphi_x(0)$, затем возьмем от производящей функции производную, уже там положим $z = 0$ и обнаружим, что $P(x = 0) = \varphi'_x(0)$. Далее, взяв n -тую производную и положив в ней $z = 0$, получим $P(x = k) = \varphi_x^{(k)}(0)/k!$. Возможно, стоит заметить, что в данном случае 0^0 следует считать единицей. \square

Из последнего факта следует, что между производящими функциями и распределениями целочисленных неотрицательных случайных величин есть взаимнооднозначное соответствие.

З°. Пусть две случайных величины x и y независимы и имеют производящие функции $\varphi_x(z)$ и $\varphi_y(z)$. Тогда

$$\varphi_{x+y}(z) = \varphi_x(z) \cdot \varphi_y(z).$$

Это означает, что **производящая суммы независимых случайных величин равна произведению производящих.**

Доказательство. Доказательство маленькое и симпатичное. Независимость x и y используется в помеченном звездочкой переходе:

$$\varphi_{x+y}(z) = Ez^{x+y} = Ez^x z^y \stackrel{*}{=} Ez^x Ez^y = \varphi_x(z)\varphi_y(z). \quad \square$$

Последнее утверждение - важнейшее свойство производящих функций. По индукции легко доказывается аналогичный факт для суммы n независимых случайных величин. Задача о выяснении распределения суммы целочисленных неотрицательных случайных величин в отдельных случаях сильно упрощается.

Пример. Отыщем производящую функцию для случайной величины x , распределенной по биномиальному закону с параметрами n и p :

Будем действовать по определению:

$$\begin{aligned} \varphi_x(z) &= Ez^x = \sum_{k=0}^n z^k P(x = k) = \sum_{k=0}^n C_n^k z^k p^k (1-p)^{n-k} \\ &= \sum_{k=0}^n C_n^k (pz)^k (1-p)^{n-k} = (pz + 1 - p)^n. \end{aligned}$$

Бинома Ньютона для решения этой задачи вполне достаточно.

Пример. Пусть x_1, \dots, x_n — независимые случайные величины, имеющие распределение Пуассона с параметрами $\lambda_1, \dots, \lambda_n$ соответственно. Требуется выяснить, как распределена их сумма.

Найдем производящую функцию случайной величины x_i :

$$\varphi_{x_i}(z) = Ez^{x_i} = \sum_{k=0}^{\infty} z^k \lambda_i^k \frac{e^{-\lambda_i}}{k!} = e^{-\lambda_i} \sum_{k=0}^{\infty} \frac{(\lambda_i z)^k}{k!} = e^{\lambda_i(z-1)}.$$

Производящую функцию суммы $x_1 + \dots + x_n$ найдем как произведение производящих функций слагаемых:

$$\varphi_{x_1+\dots+x_n}(z) = e^{(\lambda_1+\dots+\lambda_n)(z-1)}.$$

Здесь в глаза бросается тот факт, что производящая функция суммы по виду сильно напоминает производящую функцию одиночной случайной величины. Очевидно, что мы получили производящую функцию случайной величины, распределенной по закону Пуассона с параметром $\lambda_1 + \dots + \lambda_n$. Теперь, пользуясь фактом взаимно-однозначного соответствия между производящими функциями и распределениями целочисленных неотрицательных случайных величин, заключаем, что сумма независимых пуассоновских случайных величин есть пуассоновская случайная величина с параметром, равным сумме параметров распределений слагаемых.

Замечание. Случаи, когда по виду производящей функции суммы можно «угадать» ее распределение, не исчерпываются случаем суммы независимых пуассоновских случайных величин. Такое встретится для некоторых других дискретных распределений.

Лекция 9

9.1 Характеристические функции

Аппарат характеристических функций является одним из основных средств аналитического аппарата теории вероятностей. Наиболее ярко это будет продемонстрировано через 2 лекции при доказательстве некоторых предельных теорем. На этой лекции мы изложим основные свойства и теоремы, связанные с характеристическими функциями.

Наряду с действительными случайными величинами теория характеристических функций требует привлечения комплекснозначных случайных величин.

Многие из определений и свойств, относящихся к действительным случайным величинам переносятся на случай комплекснозначных случайных величин.

Введем теперь определение характеристической функции.

Определение 9.1. *Характеристической функцией* случайной величины X называется функция $\varphi_X(t) = \mathbb{E}e^{itX}$, $t \in \mathbb{R}$, i - мнимая единица.

Замечание. Исходя из определения математического ожидания и интеграла Лебега характеристическую функцию случайной величины X можно переписать в таком виде:

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx} dF(x),$$

где $F(x)$ — функция распределения X .

Если случайная величина X имеет плотность $f(x)$, то ее характеристическая функция имеет вид

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx} f(x) dx.$$

В этом случае характеристическая функция является преобразованием Фурье для функции плотности $f(x)$.

Если случайная величина X дискретна, тогда ее характеристическая функция имеет вид

$$\varphi_X(t) = \sum_{i=1}^{\infty} e^{itx_k} \mathbb{P}(X = x_k).$$

Пример. Вычислим характеристическую функцию пуассоновской случайной величины X с параметром λ ($X \sim \Pi(\lambda)$).

$$\begin{aligned} \varphi_X(t) &= \mathbb{E}e^{itX} = \sum_{i=1}^{\infty} e^{itn} P(x = n) = \sum_{i=1}^{\infty} e^{itn} \frac{e^{-\lambda} \lambda^n}{n!} \\ &= e^{-\lambda} \sum_{i=1}^{\infty} \frac{(\lambda e^{it})^n}{n!} = e^{-\lambda} e^{\lambda e^{it}} = e^{\lambda(e^{it} - 1)}. \end{aligned}$$

Итак, функция $\varphi_X(t) = e^{\lambda(e^{it}-1)}$ является характеристической функцией пуассоновской случайной величины X с параметром λ .

Пример. Вычислим характеристическую функцию стандартной нормальной случайной величины X ($X \sim \mathcal{N}(0, 1)$). Учитывая тот факт, что

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sin tx e^{-\frac{x^2}{2}} dx = 0,$$

мы можем переписать характеристическую функцию в следующем виде

$$\varphi_X(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos tx e^{-\frac{x^2}{2}} dx$$

Продифференцируем характеристическую функцию по t

$$\begin{aligned} \varphi'_X(t) &= \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos tx e^{-\frac{x^2}{2}} dx \right)'_t = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \sin tx e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sin tx d e^{-\frac{x^2}{2}} = -t \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos tx e^{-\frac{x^2}{2}} dx. \end{aligned}$$

Получили дифференциальное уравнение

$$\varphi'_X(t) = -t\varphi_X(t),$$

Решением которого является функция

$$\varphi_X(t) = ce^{-\frac{t^2}{2}}.$$

Вычисляя значение характеристической функции в нуле найдем, что $c = 1$. Итак характеристической функцией стандартной нормальной случайной величины X является функция $\varphi_X(t) = e^{-\frac{t^2}{2}}$.

Сформулируем без доказательства одно свойство математического ожидания и теорему Лебега о мажорируемой сходимости, которые нам потребуются в дальнейшем при доказательстве свойств характеристических функций и теорем с ними связанных.

Теорема 9.1 (свойство математического ожидания). Пусть случайные величины таковы, что $|X| \leq Y$ (н.н) и $EY < \infty$. Тогда существует $E|X|$ и $E|X| < EY$.

Теорема 9.2 (теорема Лебега о мажорируемой сходимости). Пусть случайные величины таковы, что $|X_n| \leq Y$, $EY < \infty$ и $X_n \xrightarrow{n.n.} X$. Тогда $E|X| < \infty$ и

$$EX_n \rightarrow EX, \quad n \rightarrow \infty$$

и

$$E|X_n - X| \rightarrow 0, \quad n \rightarrow \infty.$$

9.1.1 Свойства характеристических функций

1°. $\varphi_X(0) = 1$

Доказательство. Свойство проверяется тривиально

$$\varphi_X(0) = \int_{-\infty}^{\infty} e^0 dF(x) = \int_{-\infty}^{\infty} dF(x) = 1.$$

□

2°. $|\varphi_X(t)| \leq 1$

Доказательство. Очевидно имеем оценки

$$|\varphi_X(t)| = \left| \int_{-\infty}^{\infty} e^{itx} dF(x) \right| \leq \int_{-\infty}^{\infty} |e^{itx}| dF(x) = \int_{-\infty}^{\infty} dF(x) = 1.$$

□

3°. $\varphi_X(t)$ равномерно непрерывна на \mathbb{R}

Доказательство. Свойство следует из оценки

$$|\varphi(t+h) - \varphi(t)| = |\mathbb{E}e^{itx}(e^{ihx} - 1)| \leq \mathbb{E}|e^{ihx} - 1|$$

и теоремы Лебега о мажорируемой сходимости, согласно которой $\mathbb{E}|e^{ihx} - 1| \rightarrow 0$ при $h \rightarrow 0$. □

4°. $\varphi_X(-t) = \overline{\varphi_X(t)}$

Доказательство. Свойство следует из цепочки равенств

$$\varphi_X(-t) = \mathbb{E}e^{i(-t)x} = \mathbb{E}e^{-itx} = \overline{\mathbb{E}e^{itx}}.$$

□

5°. $\varphi_{aX+b}(t) = e^{itb}\varphi_X(at)$

Доказательство. Проводя тривиальные выкладки, убеждаемся в справедливости свойства:

$$\varphi_{aX+b}(t) = \mathbb{E}e^{it(ax+b)} = \mathbb{E}e^{itax}e^{itb} = e^{itb}\varphi_X(at).$$

□

6°. Если для некоторого $n \geq 1$ $\mathbb{E}|X|^n < \infty$, то при всех $r \leq n$ существуют производные $\varphi^{(n)}(t)$ и

$$\varphi^{(n)}(t) = \int_{\mathbb{R}} (ix)^n e^{itx} dF(x), \tag{1}$$

$$\mathbb{E}X^r = \frac{\varphi^{(r)}(0)}{i^r}. \tag{2}$$

Доказательство. Если $\mathbb{E}|X|^n < \infty$, то $\mathbb{E}|X|^r < \infty$ для всех $r < n$. Рассмотрим отношение

$$\frac{\varphi(t+h) - \varphi(t)}{h} = \mathbb{E} e^{itX} \left(\frac{e^{ihX} - 1}{h} \right).$$

Поскольку

$$\left| \frac{e^{ihx} - 1}{h} \right| \leq |x|$$

и $\mathbb{E}|X| < \infty$, то по теореме Лебега о мажорируемой сходимости имеем

$$\lim_{h \rightarrow 0} \mathbb{E} e^{itX} \left(\frac{e^{ihX} - 1}{h} \right) = \mathbb{E} \lim_{h \rightarrow 0} e^{itX} \left(\frac{e^{ihX} - 1}{h} \right) = i\mathbb{E}(X e^{itX}) = i \int_{-\infty}^{\infty} x e^{itx} dF(x).$$

Поэтому существует производная $\varphi'(t)$ и

$$\varphi'(t) = i\mathbb{E}(X e^{itX}) = i \int_{-\infty}^{\infty} x e^{itx} dF(x).$$

Существование $\varphi^{(r)}(t)$, $1 \leq r \leq n$ и справедливость формулы (1) доказывается по индукции. Формула (2) следует непосредственно из (1). \square

7°. $\varphi_X(t)$ является неотрицательно определенной функцией

Доказательство. Неотрицательная определенность функции $f(t)$ означает, что для $\forall n, \forall t_1, \dots, t_n \in \mathbb{R}$ и $\forall z_1, \dots, z_n \in \mathbb{C}$ выполнено

$$\sum_{k,l=1}^n z_k \bar{z}_l f(t_k - t_l) \geq 0.$$

Покажем, что характеристическая функция $\varphi_X(t)$ является неотрицательно определенной.

$$\begin{aligned} \sum_{k,l=1}^n z_k \bar{z}_l \varphi_X(t_k - t_l) &= \sum_{k,l=1}^n z_k \bar{z}_l \mathbb{E} e^{i(t_k - t_l)x} = \sum_{k,l=1}^n \mathbb{E} z_k \bar{z}_l e^{it_k x} e^{-it_l x} \\ &= \sum_{k,l=1}^n \mathbb{E} z_k e^{it_k x} \bar{z}_l e^{-it_l x} = \mathbb{E} \sum_{k,l=1}^n z_k e^{it_k x} \overline{z_l e^{it_l x}} \\ &= \mathbb{E} \sum_{k,l=1}^n w_k \bar{w}_l = \mathbb{E} \sum_{k=1}^n w_k \sum_{l=1}^n \bar{w}_l = \mathbb{E} \sum_{k=1}^n w_k \overline{\sum_{k=1}^n w_k} \geq 0. \end{aligned}$$

Где мы сделали замену $w_k = z_k e^{it_k x}$, $\bar{w}_l = \overline{z_l e^{it_l x}}$. \square

Сформулируем свойство, которое является ключевым при доказательстве предельных теорем для сумм независимых случайных величин методом характеристических функций.

8°. Пусть X_1, \dots, X_n — независимые случайные величины, тогда

$$\varphi_{X_1+\dots+X_n}(t) = \prod_{k=1}^n \varphi_{X_k}(t).$$

Замечание. Обратное, вообще говоря, неверно!

Перейдем к доказательству свойства.

Доказательство. Ранее было доказано, что если X_1, \dots, X_n — независимые случайные величины, и существуют $\mathbf{E}X_k$, где $k = \overline{1, n}$, то существует $\mathbf{E}X_1 \dots X_n$ и

$$\mathbf{E}X_1 \dots X_n = \mathbf{E}X_1 \dots \mathbf{E}X_n.$$

По лемме о группировке, любые комбинации \sin и \cos от независимых случайных величин $X_1 \dots X_n$ (например $\cos tX_1, \sin tX_2, \cos tX_3 \dots \cos tX_n$) будут независимыми. Учитывая это и формулу Эйлера $e^{i\alpha} = \cos \alpha + i \sin \alpha$, получим

$$\begin{aligned} \varphi_{X_1+\dots+X_n}(t) &= \mathbf{E}e^{it(X_1+\dots+X_n)} \\ &= \mathbf{E}(\cos tX_1 + i \sin tX_1) \dots (\cos tX_n + i \sin tX_n) \\ &= (\mathbf{E} \cos tX_1 + i \mathbf{E} \sin tX_1) \dots (\mathbf{E} \cos tX_n + i \mathbf{E} \sin tX_n) = \prod_{k=1}^n \varphi_{X_k}(t). \end{aligned}$$

□

Мы сформулировали некоторые необходимые условия, которым удовлетворяют характеристические функции. Таким образом, если для некоторой функции не выполняется одно из первых условий, то это означает, что рассматриваемая функция не может быть характеристической.

Сложнее обстоит дело с проверкой того, является ли данная функция характеристической. Сформулируем теорему Бохнера-Хинчина, дающую ответ на этот вопрос.

Теорема 9.3 (Теорема Бохнера-Хинчина). *Непрерывная функция $\varphi(t)$, $t \in \mathbb{R}$, такая что $\varphi(0) = 1$, является характеристической функцией некоторой случайной величины X , тогда и только тогда, когда она неотрицательно определена.*

Доказательство. Необходимость нами доказана выше. Достаточность доказывается труднее, и мы примем ее без доказательства. Доказательство можно найти в книге Б.В. Гнеденко "Курс теории вероятностей". □

9.2 Формула обращения

Сформулируем теорему, показывающую, что функция распределения $F = F(x)$ случайной величины X однозначно восстанавливается по своей характеристической функции $\varphi(t)$, и дадим явное представление $F(x)$ через $\varphi(t)$.

Теорема 9.4 (формула обращения). *Пусть $F = F(x)$ — функция распределения и*

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} dF(x)$$

— ее характеристическая функция. Тогда, для любых двух точек $a, b (a < b)$, где функция $F(x)$ непрерывна,

$$F(b) - F(a) = \lim_{c \rightarrow \infty} \frac{1}{2\pi} \int_{-c}^c \frac{e^{-ita} - e^{itb}}{it} \varphi(t) dt.$$

Если $\int_{-\infty}^{\infty} |\varphi(t)| dt < \infty$, то функция распределения $F(x)$ имеет плотность $f(x)$,

$$F(x) = \int_{-\infty}^x f(y) dy,$$

и

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t) dt.$$

Доказательство. Обозначим

$$\Phi_c \equiv \frac{1}{2\pi} \int_{-c}^c \frac{e^{-ita} - e^{itb}}{it} \varphi(t) dt.$$

Преобразуем Φ_c следующим образом:

$$\begin{aligned} \Phi_c &= \frac{1}{2\pi} \int_{-c}^c \frac{e^{-ita} - e^{itb}}{it} \varphi(t) dt = \frac{1}{2\pi} \int_{-c}^c \frac{e^{-ita} - e^{itb}}{it} \left[\int_{-\infty}^{\infty} e^{itx} F(x) dx \right] dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\int_{-c}^c \frac{e^{-ita} - e^{itb}}{it} e^{itx} dt \right] dF(x) = \int_{-\infty}^{\infty} \Psi_c(x) dF(x), \end{aligned}$$

где мы положили

$$\Psi_c(x) = \frac{1}{2\pi} \int_{-c}^c \frac{e^{-ita} - e^{itb}}{it} e^{itx} dt$$

и воспользовались теоремой Фубини (см. замечание в конце теоремы), справедливость которой в данном случае следует из того, что

$$\left| \frac{e^{-ita} - e^{itb}}{it} e^{itx} \right| = \left| \frac{e^{-ita} - e^{itb}}{it} \right| = \left| \int_a^b e^{-itx} dx \right| \leq b - a,$$

$$\int_{-\infty}^{\infty} dF(x) = 1$$

и

$$\int_{-c}^c \int_{-\infty}^{\infty} (b-a) dF(x) dt \leq 2c(b-a) \leq \infty.$$

Далее,

$$\begin{aligned} \Psi_c(x) &= \frac{1}{2\pi} \int_{-c}^c \frac{\sin t(x-a) - \sin t(x-b)}{t} dt \\ &= \frac{1}{2\pi} \int_{-c(x-a)}^{c(x-a)} \frac{\sin v}{v} dv - \frac{1}{2\pi} \int_{-c(x-b)}^{c(x-b)} \frac{\sin v}{v} dv \end{aligned}$$

Функция

$$g(s, t) = \int_s^t \frac{\sin v}{v} dv$$

равномерно непрерывна по s и t и

$$g(s, t) \rightarrow \pi$$

при $t \uparrow \infty$ и $s \downarrow -\infty$. Поэтому существует такая константа C , что для всех c и x $|\Psi_c(x)| < C < \infty$. Отсюда следует, что

$$\Psi_c(x) \rightarrow \Psi(x), \quad c \rightarrow \infty,$$

где

$$\Psi(x) = \begin{cases} 0, & \text{если } x < a \text{ или } x > b; \\ 1/2, & \text{если } x = a \text{ или } x = b; \\ 1, & \text{если } a < x < b. \end{cases}$$

Тогда по теореме Лебега о мажорируемой сходимости возникает предел

$$\int_{-\infty}^{\infty} \Psi(x) dF(x) = Q(a, b) + \frac{1}{2}Q(a) + \frac{1}{2}Q(b) = Q(a(a; b]) = F(b) - F(a).$$

Последнее равенство справедливо только для a и b , являющихся точками непрерывности $F(x)$.

Таким образом, доказана первая часть теоремы.

Вторая часть доказывается проще.

Обозначим

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t) dt.$$

Из теоремы о мажорируемой сходимости следует, что эта функция непрерывна по x и, значит, она интегрируема на $[a, b]$. Применяя теорему Фубини, получим

$$\begin{aligned} \int_a^b f(x) dx &= \frac{1}{2\pi} \int_a^b \left[\int_{-\infty}^{\infty} e^{-itx} \varphi(t) dt \right] dx \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi(t) \left[\int_a^b e^{-itx} dx \right] dt = \lim_{c \rightarrow \infty} \frac{1}{2\pi} \int_{-c}^c \varphi(t) \left[\int_a^b e^{-itx} dx \right] dt \\ &= \lim_{c \rightarrow \infty} \int_c^c \frac{1}{2\pi} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt = F(b) - F(a) \end{aligned}$$

для всех точек a и b непрерывности функции $F(x)$. Отсюда вытекает, что

$$F(x) = \int_{-\infty}^x f(y) dy, \quad x \in \mathbb{R},$$

а так как $f(x)$ - непрерывная, а $F(x)$ - неубывающая функции, то $f(x)$ есть плотность $F(x)$. \square

Замечание. Теорема Фубини будет сформулирована и доказана в курсе функционального анализа, а пока мы ее принимаем на веру. Теорема нужна нам, чтобы показать, что мы имеем право менять пределы интегрирования у интеграла.

Замечание. Формулы

$$\begin{aligned} \varphi(t) &= \int_{-\infty}^{\infty} e^{itx} f(x) dx, \\ f(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t) dt \end{aligned}$$

очень похожи. Первую из них, как отмечалось выше, обычно в анализе называют преобразованием Фурье функции $f(x)$, вторую — обратным преобразованием Фурье. Подробно с этой темой вы познакомитесь в курсах математического и функционального анализов.

Лекция 10

10.1 Слабая сходимоть

Определение 10.1. Последовательность функций распределения

$$F_1(x), F_2(x), \dots, F_n(x), \dots$$

слабо сходится к функции $F(x)$, если при $n \rightarrow \infty$ она сходится к этой последней в каждой ее точке непрерывности.

Замечание. Тот факт, что последовательность функций распределения $\{F_n(x)\}$ слабо сходится к функции $F(x)$ будем обозначать $F_n(x) \Rightarrow F(x)$.

Замечание. Рассмотрим X_1, X_2, \dots — последовательность одинаково распределенных случайных величин с $P(X_i = 1) = p$, $P(X_i = 0) = q$, $p + q = 1$. Используя понятие сходимости по вероятности, запишем закон больших чисел в виде

$$\frac{S_n}{n} \xrightarrow{p} p, \quad n \rightarrow \infty. \quad (1)$$

Обозначим

$$F_n(x) = P \left\{ \frac{S_n}{n} \leq x \right\},$$
$$F(x) = \begin{cases} 1, & x \geq p; \\ 0, & x < p. \end{cases}$$

Из (1) легко вывести, что $F_n(x) \rightarrow F(x)$ для всех точек $x \in \mathbb{R}$, кроме точки $x = p$, где функция $F(x)$ терпит разрыв.

Отсюда следует, что слабая сходимоть не влечет за собой поточечную сходимоть $F_n(x)$ к $F(x)$.

Возникает вопрос, всегда ли будет предельная функция являться функцией распределения. Ответ на этот вопрос отрицательный. Приведем пример поясняющий сказанное.

Пример. Рассмотрим последовательность функций распределения

$$F_n(x) = \begin{cases} 0, & \text{при } x \leq -n; \\ 1/2, & \text{при } -n < x < n; \\ 1, & \text{при } x > n. \end{cases}$$

Очевидно, чтобы предельная функция $F(x) \equiv 1/2$ — не является функцией распределения.

Следовательно, мы должны накладывать некоторые условия, чтобы предельная функция была функцией распределения.

Замечание. Пусть X, X_1, X_2, \dots — случайные величины и $F_{X_n}(x) \Rightarrow F_X(x)$. В этом случае говорят, что случайные величины X_1, X_2, \dots сходятся по распределению к случайной величине X , и записывают $X_n \xrightarrow{d} X$. Эта запись наглядна и поэтому часто в формулировках предельных теорем ее предпочитают записи $F_{X_n}(x) \Rightarrow F_X(x)$.

Утверждение 10.1. Пусть последовательность функций распределения $\{F_n(x)\}$ сходится к функции $F(x)$ на всюду плотном множестве D . Тогда $F_n(x) \Rightarrow F(x)$.

Доказательство. Пусть x — любая точка и x' и x'' — какие-нибудь точки множества D , такие что $x' \leq x \leq x''$. При этом также

$$F_n(x') \leq F_n(x) \leq F_n(x'').$$

Следовательно,

$$\lim_{n \rightarrow \infty} F_n(x') \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \overline{\lim}_{n \rightarrow \infty} F_n(x) \leq \lim_{n \rightarrow \infty} F_n(x'').$$

А так как

$$\lim_{n \rightarrow \infty} F_n(x') = F(x') \quad \text{и} \quad \lim_{n \rightarrow \infty} F_n(x'') = F(x''),$$

то и

$$F(x') \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \overline{\lim}_{n \rightarrow \infty} F_n(x) \leq F(x'').$$

Но средние члены в этих неравенствах не зависят от x' и x'' , поэтому

$$F(x-0) \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \overline{\lim}_{n \rightarrow \infty} F_n(x) = F(x+0).$$

Если функция $F(x)$ в точке x непрерывна, то

$$F(x-0) = F(x) = F(x+0).$$

Следовательно, в точках непрерывности функции $F(x)$

$$\lim_{n \rightarrow \infty} F_n(x) = F(x).$$

□

Далее нам понадобятся две теоремы, принадлежащие Хелли.

Теорема 10.1 (Первая теорема Хелли). Из любой последовательности $\{F_n(x)\}$ функций распределения можно выделить по крайней мере одну слабо сходящуюся подпоследовательность.

Замечание. Идея доказательства теоремы очень похожа на ту, что использовалась при доказательстве теоремы Арцела из курса математического анализа, заключающаяся в том, что из равномерно непрерывной и равномерно ограниченной функциональной последовательности можно извлечь по крайней мере одну подпоследовательность, сходящуюся равномерно на некотором компакте. Так что при желании читатель может провести доказательство самостоятельно.

Доказательство. Рассмотрим какое-нибудь всюду плотное множество D точек x'_1, \dots, x'_n, \dots . Возьмем значения функций последовательности $\{F_n(x)\}$ в точке x'_1

$$F_1(x'_1), F_2(x'_1), \dots, F_n(x'_1), \dots$$

Так как множество этих значений ограничено, то по теореме Больцано-Вейерштрасса из курса математического анализа, оно содержит по меньшей мере одну подпоследовательность

$$F_{11}(x'_1), F_{12}(x'_1), \dots, F_{1n}(x'_1), \dots,$$

сходящуюся к некоторому предельному значению $G(x'_1)$. Рассмотрим теперь множество чисел

$$F_{11}(x'_2), F_{12}(x'_2), \dots, F_{1n}(x'_2), \dots,$$

Так как и это множество ограничено, то и в нем существует подпоследовательность, сходящаяся к некоторому предельному значению $G(x'_2)$. Таким образом из последовательности $\{F_{1n}(x)\}$ мы можем выделить подпоследовательность

$$F_{21}(x), F_{22}(x), \dots, F_{2n}(x), \dots$$

такую, что одновременно $\lim_{n \rightarrow \infty} F_{2n}(x'_1) = G(x'_1)$ и $\lim_{n \rightarrow \infty} F_{2n}(x'_2) = G(x'_2)$. Продолжим такое выделение подпоследовательностей $\{F_{kn}(x)\}$, для которых одновременно для всех $r \leq k$ выполнено $\lim_{n \rightarrow \infty} F_{kn}(x'_r) = G(x'_r)$.

Составим теперь диагональную ("канторовскую") последовательность $\{F_{nn}(x)\}$, сходящуюся по всем x'_k к $G(x)$. Так как все $F_{nn}(x)$ неубывают и ограничены, то $G(x)$ будет неубывающей и ограниченной. В силу доказанного выше утверждения $F_{nn}(x) \Rightarrow G(x)$. \square

Теорема 10.2 (Вторая теорема Хелли). Пусть $g(x)$ — непрерывная функция и $F_n(x) \Rightarrow F(x)$, при этом $F(+\infty) - F(-\infty) = 1$. Тогда

$$\int_{-\infty}^{\infty} g(x) dF_n(x) \rightarrow \int_{-\infty}^{\infty} g(x) dF(x).$$

Доказательство. Сначала докажем, что для некоторого фиксированного $A > 0$

$$\int_{-A}^A g(x) dF_n(x) \rightarrow \int_{-A}^A g(x) dF(x). \quad (2)$$

Фиксируем произвольное $\varepsilon > 0$. Разделим отрезок $[-A, A]$ точками x_0, \dots, x_N : $-A = x_0 < x_1 < \dots < x_N = A$, так что x_i — точки непрерывности $F(x)$ и $|g(x) - g(x_i)| < \varepsilon$ для $\forall x \in [x_{i-1}, x_i]$. Последнее возможно, т.к. $g(x)$ равномерно непрерывна на $[-A, A]$.

Определим ступенчатую функцию $g_\varepsilon(x)$ на $[-A, A]$:

$$g_\varepsilon(x) = g(x_i) \quad \text{для } x \in [x_{i-1}, x_i] \quad i = \overline{1, N} \quad \text{и} \quad g_\varepsilon(-A) = g_\varepsilon(A).$$

Тогда, очевидно, что для $\forall x \in [-A, A]$ $|g_\varepsilon(x) - g(x)| < \varepsilon$. Рассмотрим разность интегралов 1:

$$\begin{aligned} & \left| \int_{-A}^A g(x) dF_n(x) - \int_{-A}^A g(x) dF(x) \right| \leq \int_{-A}^A |g(x) - g_\varepsilon(x)| dF_n(x) \\ & + \left| \int_{-A}^A g_\varepsilon(x) dF_n(x) - \int_{-A}^A g_\varepsilon(x) dF(x) \right| + \int_{-A}^A |g(x) - g_\varepsilon(x)| dF(x) \\ & \leq 2\varepsilon + M \sum_{k=1}^N (|F_n(x_k) - F(x_k)| + |F_n(x_{k-1}) - F(x_{k-1})|), \end{aligned}$$

где $M = \sup_k |g(x)|$.

Так как $F_n(x) \Rightarrow F(x)$, то при $n \rightarrow \infty$ последнее слагаемое стремится к нулю, следовательно (2) доказано.

Фиксируем $\varepsilon > 0$, тогда $\exists A : F(-A) < \varepsilon/4, 1 - F(A) < \varepsilon/4$. Не ограничивая общности, считаем, что $-A$ и A — точки непрерывности F . Тогда, т.к. $F_n(-A) \rightarrow F(-A)$ и $F_n(A) \rightarrow F(A)$, то $\exists n_0 : n \geq n_0$:

$$F_n(-A) < \varepsilon/2 \quad \text{и} \quad 1 - F_n(A) < \varepsilon/2.$$

Таким образом имеем:

$$\begin{aligned} & \left| \int_{-\infty}^{\infty} g(x) dF_n(x) - \int_{-\infty}^{\infty} g(x) dF(x) \right| \leq \left| \int_{-A}^A g(x) dF_n(x) - \int_{-A}^A g(x) dF(x) \right| \\ & + \left| \int_{|x|>A} g(x) dF_n(x) \right| + \left| \int_{|x|<A} g(x) dF_n(x) \right| \leq \left| \int_{-A}^A g(x) dF_n(x) - \int_{-A}^A g(x) dF(x) \right| \\ & + M(F_n(-A) + (1 - F_n(A)) + F(-A) + (1 - F(A))) \\ & \leq \left| \int_{-A}^A g(x) dF_n(x) - \int_{-A}^A g(x) dF(x) \right| + M\varepsilon + M\varepsilon/2. \end{aligned}$$

В силу (2) правая часть стремится к нулю. Таким образом теорема доказана. \square

10.2 Предельные теоремы для характеристических функций

Из прошлой лекции мы знаем, что между функциями распределения и характеристическими функциями существует взаимно однозначное соответствие. Следующие две предельные теоремы показывают, что соответствие не только взаимно однозначно, но и непрерывно.

Теорема 10.3 (Прямая предельная теорема). *Если последовательность $\{F_n(x)\}$ функций распределения слабо сходится к функции распределения $F(x)$, то соответствующая последовательность $\{\varphi_n(t)\}$ характеристических функций сходится к характеристической функции $\varphi(t)$.*

Доказательство. Эта теорема является прямым следствием второй теоремы Хелли. Действительно, из $F_n(x) \Rightarrow F(x)$ вытекает, что

$$\varphi_n(t) = \int_{-\infty}^{\infty} e^{itx} dF_n(x) \rightarrow \int_{-\infty}^{\infty} e^{itx} dF(x) = \varphi(t).$$

□

Теорема 10.4 (Обратная предельная теорема). Пусть $\{\varphi_n(t)\}$ — последовательность характеристических функций, сходящаяся к непрерывной в нуле функции $\varphi(t)$. Тогда последовательность $\{F_n(x)\}$ функций распределения слабо сходится к некоторой функции распределения $F(x)$, и $\varphi_n(t)$ и $\varphi(t)$ — характеристические функции $F_n(x)$ и $F(x)$ соответственно.

Замечание. Возникает вопрос, всегда ли будет характеристической функцией предел $\varphi(t)$ последовательности $\{\varphi_n(t)\}$ характеристических функций.

Ответ на этот вопрос отрицательный. Рассмотрим пример, поясняющий сказанное.

Пример. Рассмотрим последовательность характеристических функций

$$\varphi_n(t) = \begin{cases} 0, & \text{при } t \leq -1/n; \\ nt + 1, & \text{при } -1/n < t \leq 0; \\ -nt + 1, & \text{при } 0 < t < 1/n; \\ 0, & \text{при } x > 1/n. \end{cases}$$

Очевидно, что предельная функция

$$\varphi(t) = \begin{cases} 0, & \text{при } t < 0; \\ 1 & \text{при } t = 0; \\ 0, & \text{при } t > 0 \end{cases}$$

не будет характеристической, т.к. она не является непрерывной.

Доказательству теоремы предположим лемму

Лемма 10.1. Пусть X — случайная величина. Тогда для $\forall \tau > 0$ имеет место неравенство

$$\mathbf{P}(|X| \leq 2/\tau) \geq 2 \left| \frac{1}{2\tau} \int_{-\tau}^{\tau} \varphi(t) dt \right| - 1. \quad (3)$$

Доказательство.

$$\begin{aligned} & \left| \frac{1}{2\tau} \int_{-\tau}^{\tau} \varphi(t) dt \right| = \left| \frac{1}{2\tau} \int_{-\tau}^{\tau} \mathbf{E} e^{itX} dt \right| \\ & = \left| \frac{1}{2\tau} \mathbf{E} \int_{-\tau}^{\tau} e^{itX} dt \right| = \left| \mathbf{E} \frac{\sin \tau X}{\tau X} (I_{|X| \leq 2/\tau} + I_{|X| > 2/\tau}) \right| \\ & \leq \mathbf{P}(|X| \leq 2/\tau) + \frac{1}{2} (1 - \mathbf{P}(|X| \leq 2/\tau)). \end{aligned}$$

□

Теперь мы обладаем всем, чтобы доказать **обратную предельную теорему**.

Доказательство. По первой теореме Хелли из последовательности $F_n(x)$ можно выделить подпоследовательность $F_{nn}(x)$, слабо сходящуюся к некоторой функции $F(x)$.

Докажем, что функция $F(x)$ является функцией распределения. Для этого достаточно доказать, что

$$F(+\infty) - F(-\infty) \geq 1.$$

А так как в силу первой теоремы Хелли $F(x)$ неубывает и $0 \leq F(-\infty) \leq F(+\infty) \leq 1$, то отсюда будет следовать, что $F(x)$ — функция распределения. В силу леммы 2

$$F_{nn}(2/\tau) - F_{nn}(-2/\tau) \geq 2 \left| \frac{1}{2\tau} \int_{-\tau}^{\tau} \varphi_{nn}(t) dt \right| - 1. \quad (4)$$

В неравенстве (4) можно считать, что $-2/\tau, 2/\tau$ — точки непрерывности функции $F(x)$.

Устремив $n \rightarrow \infty$ и применяя теорему Лебега о мажорируемой сходимости, получаем

$$\begin{aligned} F(2/\tau) - F(-2/\tau) &= \lim_{n \rightarrow \infty} F_{nn}(2/\tau) - F_{nn}(-2/\tau) \geq \lim_{n \rightarrow \infty} 2 \left| \frac{1}{2\tau} \int_{-\tau}^{\tau} \varphi_{nn}(t) dt \right| - 1 \\ &= 2 \left| \frac{1}{2\tau} \int_{-\tau}^{\tau} \varphi(t) dt \right| - 1. \end{aligned}$$

Тогда, так при $\tau \rightarrow 0$

$$\frac{1}{2\tau} \int_{-\tau}^{\tau} \varphi(t) dt \rightarrow \varphi(0),$$

то

$$F(+\infty) - F(-\infty) \geq \lim_{\tau \rightarrow 0} 2 \left| \frac{1}{2\tau} \int_{-\tau}^{\tau} \varphi(t) dt \right| - 1 = 1.$$

Покажем теперь, что $F_n(x) \Rightarrow F(x)$. Пусть это не так. Тогда существуют две подпоследовательности $F_{nn}^1(x)$ и $F_{nn}^2(x)$ слабо сходящиеся к $F^1(x)$ и $F^2(x)$. При этом $F^1(x) \neq F^2(x)$.

Пусть $\varphi_{nn}^1(t) \rightarrow \varphi^1(t)$, где $\varphi_{nn}^1(t)$ и $\varphi^1(t)$ — характеристические функции, отвечающие $F_{nn}^1(x)$ и $F^1(x)$ соответственно. И $\varphi_{nn}^2(t) \rightarrow \varphi^2(t)$, где $\varphi_{nn}^2(t)$ и $\varphi^2(t)$ — характеристические функции, отвечающие $F_{nn}^2(x)$ и $F^2(x)$ соответственно. Т.к. $F^1(x) \neq F^2(x)$, то по прямой теореме о непрерывном соответствии $\varphi^1(t) \neq \varphi^2(t)$. Но это противоречит условиям теоремы. Следовательно $F_n(x) \Rightarrow F(x)$.

□

Замечание. Покажем, что требование непрерывности в нуле предельной функции $\varphi(t)$ существенно.

Рассмотрим последовательность нормально распределенных с параметрами 0 и n случайных величин X_n ($X_n \sim \mathcal{N}(0, n)$). Тогда X_n имеет характеристическую функцию $\varphi_n(t) = e^{-nt^2/2}$. Очевидно, что

$$\lim_{n \rightarrow \infty} \varphi_n(t) = \varphi(t), \quad \text{где} \quad \varphi(t) = \begin{cases} 0, & \text{если } t \neq 0, \\ 1, & \text{если } t = 0. \end{cases}$$

Таким образом предельная функция $\varphi(t)$ разрывна при $t = 0$. Следовательно, теорема непрерывности не может иметь место. Это следует из соотношения

$$\begin{aligned} F_n(x) &= \mathbf{P}\{X_n \leq x\} = \mathbf{P}\{n^{-1/2}X_n \leq xn^{-1/2}\} \\ &= \Phi(n^{-1/2}x) \rightarrow \frac{1}{2} \quad n \rightarrow \infty. \end{aligned}$$

Следовательно, предел $\lim_{n \rightarrow \infty} F_n(x) = F(x)$ существует для каждого $x \in \mathbb{R}$, но функция $F(x) = \frac{1}{2}$ не является функцией распределения.

Лекция 11

11.1 Метод характеристических функций в доказательстве предельных теорем

Теперь мы обладаем хорошо развитым аппаратом характеристических функций и можем доказать некоторые простые предельные теоремы.

Сформулируем и докажем закон больших чисел в форме Хинчина.

Теорема 11.1 (закон больших чисел). Пусть X_1, X_2, \dots — последовательность независимых одинаково распределенных случайных величин с $EX_1 < \infty$ и $EX_1 = m$. Тогда $\frac{S_n}{n} \xrightarrow{p} m$, т.е.

$$P \left\{ \left| \frac{S_n}{n} - m \right| \geq \varepsilon \right\} \rightarrow 0, \quad n \rightarrow \infty,$$

где $S_n = X_1 + \dots + X_n$.

Замечание. Напомним, что в законе больших чисел в форме Чебышева было требование существования ограниченных дисперсий. В усиленном законе больших чисел в форме Колмогорова требовалась сходимость некоторого ряда даже в случае не всех ограниченных дисперсий. У Хинчина же требуется только ограниченность первых моментов.

Перед тем как перейти к доказательству ЗБЧ в форме Хинчина, выведем оценку для характеристической функции.

Теорема 11.2. Пусть $\varphi_X(t)$ — характеристическая функция некоторой случайной величины X . Если $E|X|^n < \infty$, то

$$\varphi_X(t) = 1 + itEX + \dots + \frac{(it)^n}{n!} EX^n + R_n(t), \quad (1)$$

где $R_n(t) = o(t^n)$, $t \rightarrow 0$.

Доказательство. Имеем

$$\begin{aligned} \varphi_X(t) &= \int_{-\infty}^{\infty} e^{itx} dF(x) \\ &= \int_{\mathbb{R}} \left(e^{itx} - 1 - itx - \dots - \frac{(itx)^n}{n!} \right) dF(x) + \int_{\mathbb{R}} \left(1 + \dots + \frac{(itx)^n}{n!} \right) dF(x). \quad (2) \end{aligned}$$

Из существования $E|X|^n < \infty$ вытекает существование $E|X|^r < \infty$ для $\forall r \leq n$. Таким образом, мы можем переписать (2) в виде

$$\varphi_X(t) = 1 + itEX + \dots + \frac{(it)^n}{n!} EX^n + R_n(t).$$

Покажем, что

$$\left| e^{iy} - 1 - iy - \dots - \frac{(iy)^n}{n!} \right| \leq \frac{|y|^{n+1}}{(n+1)!}. \quad (3)$$

Для $n = 1$ имеем

$$|e^{iy} - 1| = \left| i \int_0^y e^{iu} du \right| \leq \int_0^{|y|} 1 du = |y|.$$

Воспользуемся индукцией и получим

$$\begin{aligned} \left| e^{iy} - 1 - iy - \dots - \frac{(iy)^n}{n!} \right| &\leq \int_0^y \left(e^{iu} - 1 - \dots - \frac{(iu)^{n-1}}{(n-1)!} \right) du \\ &\leq \int_0^{|y|} \left| e^{iu} - 1 - \dots - \frac{(iu)^{n-1}}{(n-1)!} \right| du \leq \int_0^{|y|} \frac{|u|^n}{n!} du = \frac{|y|^{n+1}}{(n+1)!}. \end{aligned}$$

Выведем еще одну оценку:

$$\left| \underbrace{e^{iy} - 1 - iy - \dots - \frac{(iy)^{n-1}}{(n-1)!}}_{\leq \frac{|y|^n}{n!}} - \frac{(iy)^n}{n!} \right| \leq \frac{|y|^n}{n!} + \frac{|y|^n}{n!} = 2 \frac{|y|^n}{n!}. \quad (4)$$

Перепишем $R_n(t)$ в таком виде

$$R_n(t) = \int_{\mathbb{R}} \left(e^{itx} - 1 - itx - \dots - \frac{(itx)^n}{n!} \right) dF(x) = \int_{|X| \leq a} + \int_{|X| > a}.$$

Оценим $|R_n(t)|$ сверху используя (3) и (4)

$$\begin{aligned} |R_n(t)| &\leq \int_{|X| \leq a} \frac{|tx|^{n+1}}{(n+1)!} dF(x) + \int_{|X| > a} 2 \frac{|tx|^n}{n!} dF(x) \\ &\leq \frac{|t|^{n+1} a^{n+1}}{(n+1)!} + 2 \frac{t^n}{n!} \int_{|X| > a} |x|^n dF(x) \leq |t|^{n+1} e^a + 2 \frac{t^n}{n!} \int_{|X| > a} |x|^n dF(x). \end{aligned}$$

Выберем теперь $a > 0$ так, чтобы

$$\frac{2}{n!} \int_{|X| > a} |x|^n dF(x) < \frac{\varepsilon}{2}.$$

Затем выберем $\delta = \delta(\varepsilon; a)$ так, чтобы

$$|t| e^a < \frac{\varepsilon}{2}.$$

Таким образом, мы получили вторую оценку. □

Перейдем теперь к доказательству закона больших чисел.

Доказательство. Пусть $S_n = X_1 + \dots + X_n$, $\varphi(t) = \mathbb{E}e^{itX_1}$ и $\varphi_{\frac{S_n}{n}}(t) = \mathbb{E}e^{it\frac{S_n}{n}}$. Тогда в силу независимости случайных величин и свойства 8 характеристических функций

$$\varphi_{\frac{S_n}{n}}(t) = \left[\varphi\left(\frac{t}{n}\right) \right]^n.$$

Согласно (1) для $n = 1$

$$\varphi(t) = 1 + itm + \bar{o}(t), \quad t \rightarrow 0.$$

Значит, для всякого $t \in \mathbb{R}$

$$\varphi\left(\frac{t}{n}\right) = 1 + i\frac{t}{n}m + \bar{o}\left(\frac{1}{n}\right), \quad n \rightarrow \infty,$$

и поэтому

$$\varphi_{\frac{S_n}{n}}(t) = \left[1 + i\frac{t}{n}m + \bar{o}\left(\frac{1}{n}\right) \right]^n \rightarrow e^{itm}.$$

Функция $\varphi(t) = e^{itm}$ непрерывна в нуле и является характеристической функцией вырожденного распределения вероятностей, сосредоточенного в точке m . Поэтому

$$\frac{S_n}{n} \xrightarrow{d} m,$$

значит нам осталось доказать, что отсюда следует, что

$$\frac{S_n}{n} \xrightarrow{p} m.$$

По обратной теореме $F_{\frac{S_n}{n}}(x) \Rightarrow F_m(x)$. Фиксируем произвольное $\varepsilon > 0$.

$$\mathbb{P}\left(\left|\frac{S_n}{n} - m\right| < \varepsilon\right) = F_{\frac{S_n}{n}}(m + \varepsilon) - F_{\frac{S_n}{n}}(m - \varepsilon) \rightarrow \underbrace{F_m(m + \varepsilon)}_{=1} - \underbrace{F_m(m - \varepsilon)}_{=0} = 1.$$

Что и требовалось доказать. □

11.1.1 Центральная предельная теорема

Перейдем к центральной предельной теореме. Эта теорема примечательна уже тем, что в ее названии содержится слово «центральная», показывающее, что эта теорема находится не на периферии теории вероятностей, а является центральной.

Предельная теорема Муавра-Лапласа может быть получена как следствие центральной предельной теоремы.

Перейдем к формулировке теоремы.

Теорема 11.3 (Центральная предельная теорема для независимых одинаково распределенных случайных величин). Пусть X_1, X_2, \dots — последовательность независимых одинаково распределенных (невырожденных) случайных величин с $\mathbb{E}X_1^2 < \infty$ и $S_n = X_1 + \dots + X_n$. Тогда при $n \rightarrow \infty$ для $\forall x \in \mathbb{R}$ имеем

$$\mathbb{P}\left\{\frac{S_n - \mathbb{E}S_n}{\sqrt{DS_n}} \leq x\right\} \rightarrow \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{z^2}{2}} dz. \quad (5)$$

Доказательство. Пусть $\mathbf{E}X_1 = m$, $\mathbf{D}X_1 = \sigma^2$ и

$$\varphi(t) = \mathbf{E}e^{it(X_1-m)}.$$

Обозначим

$$\varphi_n(t) = \mathbf{E}e^{it\frac{S_n - \mathbf{E}S_n}{\sqrt{\mathbf{D}S_n}}},$$

то получим, что

$$\varphi_n(t) = \left[\varphi\left(\frac{t}{\sigma\sqrt{n}}\right) \right]^n.$$

В силу формулы (1) для $n = 2$

$$\varphi(t) = 1 - \frac{\sigma^2 t^2}{2} + \bar{o}(t^2), \quad t \rightarrow 0.$$

Поэтому для любого $t \in \mathbb{R}$ и $n \rightarrow \infty$

$$\varphi_n(t) = \left[1 - \frac{\sigma^2 t^2}{2\sigma^2 n} + \bar{o}\left(\frac{1}{n}\right) \right] \rightarrow e^{-\frac{t^2}{2}}.$$

Функция $e^{-\frac{t^2}{2}}$ является характеристической функцией стандартного нормального распределения ($\mathcal{N}(0, 1)$). В силу теорем о непрерывном соответствии между функциями распределения и характеристическими функциями центральная предельная теорема доказана. \square

Замечание. Часто центральную предельную теорему записывают в следующем виде

$$\frac{S_n - \mathbf{E}S_n}{\sqrt{\mathbf{D}S_n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Так как в (5) функция $\Phi(x)$ непрерывна, то на самом деле сходимость равномерная и можно поставить вопрос о скорости сходимости.

Дадим ответ на этот вопрос сначала для случая разнораспределенных независимых случайных величин, а потом приведем соответствующий результат для случая одинаково распределенных случайных величин.

Теорема 11.4 (центральная предельная теорема для разнораспределенных слагаемых). Пусть X_1, X_2, \dots — последовательность независимых случайных величин, для которых существуют $\mathbf{E}X_k = a_k$, $\mathbf{D}X_k = b_k^2$, $c_k = \mathbf{E}|X_k - a_k|^3$. Обозначим

$$\begin{aligned} S_n &= X_1 + \dots + X_n, \\ A_n &= \mathbf{E}S_n = a_1 + \dots + a_n, \\ B_n^2 &= \mathbf{D}S_n = b_1^2 + \dots + b_n^2, \\ C_n &= c_1 + \dots + c_n, \\ T_n &= \frac{S_n - A_n}{B_n}. \end{aligned}$$

Тогда

$$\sup_{x \in \mathbb{R}} |F_{T_n}(x) - \Phi(x)| \leq \alpha_0 \frac{C_n}{B_n^3},$$

где α_0 — абсолютная константа.

В случае независимых одинаково распределенных случайных величин ответ на вопрос о скорости сходимости дает теорема (неравенство) Берри - Эссеена.

Теорема 11.5 (Берри-Эссеен). Пусть X_1, X_2, \dots — последовательность независимых одинаково распределенных случайных величин с $EX_1^3 < \infty$. Тогда в обозначениях предыдущей теоремы имеем:

$$\sup_{x \in \mathbb{R}} |F_{T_n}(x) - \Phi(x)| \leq C \frac{E|X_1 - EX_1|^3}{\sigma^3 \sqrt{n}},$$

где C - универсальная константа, точное значение которой еще не известно, приводятся только неравенства: $\frac{1}{\sqrt{2\pi}} \leq C \leq 0.7655$.

Рассмотрим вопрос применения центральной предельной теоремы.

Пример. При измерении некоторой величины a (длины доски в аудитории) мы получаем приближенное значение X . Сделанная ошибка $\delta = X - a$ может быть представлена в виде суммы двух ошибок

$$\delta = (X - EX) + (EX - a),$$

первая из которых называется случайной ошибкой, а вторая систематической. Пусть результаты измерений лишены систематической ошибки, т.е. $EX = a$. Так бывает не всегда, но для простоты допустим. Случайная ошибка δ имеет нулевое математическое ожидание $E\delta = 0$. Положим $D\delta = \sigma^2$. Для уменьшения ошибки производят n независимых измерений X_1, \dots, X_n . Они представляют собой независимые и одинаково распределенные случайные величины. В качестве приближенного значения a принято брать среднее арифметическое из результатов наблюдений

$$\hat{a} = \frac{X_1 + \dots + X_n}{n}.$$

В силу центральной предельной теоремы

$$P(|\hat{a} - a| < \varepsilon) = P\left(\left|\frac{X_1 + \dots + X_n - na}{\sigma\sqrt{n}}\right| < \frac{\varepsilon\sqrt{n}}{\sigma}\right) \sim \frac{1}{\sqrt{2\pi}} \int_{-\frac{\varepsilon\sqrt{n}}{\sigma}}^{\frac{\varepsilon\sqrt{n}}{\sigma}} e^{-\frac{z^2}{2}} dz.$$

Если мы начнем увеличивать число измерений, наш результат будет становиться все лучше и лучше. В частности, если взять $\frac{\varepsilon\sqrt{n}}{\sigma} = 3$, то $\frac{1}{\sqrt{2\pi}} \int_{-\frac{\varepsilon\sqrt{n}}{\sigma}}^{\frac{\varepsilon\sqrt{n}}{\sigma}} e^{-\frac{z^2}{2}} dz = 0.997$ и следовательно

$$P\left(|\hat{a} - a| < \frac{3\sigma}{\sqrt{n}}\right) \approx 0.997,$$

что уже является неплохой точностью.

Лекция 12

12.1 Условное математическое ожидание, условное распределение случайной величины

12.1.1 Условное математическое ожидание, условное распределение в случае дискретных случайных величин

Определение 12.1. *Условной вероятностью* события A при уже произошедшем (при условии) B (обязательно $P(B) > 0$) называется величина

$$\frac{P(AB)}{P(B)} = P(A|B) \text{ — это её обозначение.}$$

Замечание. Пространство $(\Omega, \mathcal{F}, P(\cdot|B))$ — также вероятностное пространство, на котором $P(A|B)$ как функция от события A есть вероятность.

Рассмотрим дискретные случайные величины X и Y такие, что X принимает значения a_1, \dots, a_k с вероятностями p_1, \dots, p_k , а Y принимает значения b_1, \dots, b_m с вероятностями q_1, \dots, q_m . В качестве события B рассмотрим $B = \{Y = b_i\}$ (b_i фиксировано), $P(B) = q_i > 0$, тогда можем определить математическое ожидание случайной величины X относительно меры $P(\cdot|B)$ как

$$E(X|B) = E(X|Y = b_i) = \sum_{j=1}^k a_j P(X = a_j|B)$$

Определение 12.2. *Условным математическим ожиданием* $E(X|Y)$ случайной величины X относительно случайной величины Y называется **случайная величина**, которая принимает значения $E(X|Y = b_i)$ с вероятностями $P(Y = b_i)$, и определяется как отображение следующего вида

$$\forall \omega \in Y^{-1}(b_i) \in F$$

данное отображение обозначим

$$g(Y(\omega)) = E(X|Y = b_i).$$

Так как Y — случайная величина, то отображение $g(Y)$ также является случайной величиной.

Определение 12.3. *Условным распределением случайной величины X относительно случайной величины Y называется случайная величина*, которая для $\forall A \in \mathcal{F}$ является борелевской σ -алгеброй на прямой \mathbb{R} и для любого $\omega \in Y^{-1}(b_i)$ принимает значение

$$P(x \in A|y = b_i) = E(\mathbb{I}_{x \in A}|Y = b_i)$$

12.1.2 Условное математическое ожидание, условное распределение в случае абсолютно непрерывных случайных величин

Рассмотрим случай абсолютно непрерывной случайной величины Y , для нее (в силу ее абсолютной непрерывности) следует, что $P(Y = y_i) = 0$, где y_i — некоторое фиксированное число. Очевидно, что в данном случае мы не можем воспользоваться рассуждениями для дискретного случая.

Рассмотрим частный случай следующего вида — случайный вектор (X, Y) имеет плотность $f(x, y)$, причем $f(x, y) \in C$. Предположим, что для произвольного достаточно малого $\varepsilon > 0$ выполнено

$$P(-\varepsilon < Y - y_0 < \varepsilon) > 0 \quad (\text{A})$$

Таким образом, если $f_Y(y)$ — плотность случайной величины Y , то

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx,$$

и из формулы (A) следует, что $f_Y(y_0) > 0$.

Рассмотрим $\forall u \in \mathbb{R}$

$$\begin{aligned} P(X < u | Y \in (y_0 - \varepsilon, y_0 + \varepsilon)) &= \frac{P(X < u, Y \in (y_0 - \varepsilon, y_0 + \varepsilon))}{P(y_0 - \varepsilon < Y < y_0 + \varepsilon)} \\ &= \frac{\int_{-\infty}^u \int_{y_0 - \varepsilon}^{y_0 + \varepsilon} f(x, y) dx dy}{\int_{y_0 - \varepsilon}^{y_0 + \varepsilon} f_Y(y) dy} \xrightarrow{\varepsilon \rightarrow 0} \int_{-\infty}^u \frac{f(x, y_0)}{f_Y(y_0)} dx \\ &\stackrel{df}{=} P(X < u | Y = y_0) \end{aligned}$$

Определение 12.4.

$$P(\mathbf{X} < \mathbf{u} | \mathbf{Y} = \mathbf{y}_0) = \int_{-\infty}^u \frac{f(x, y_0)}{f_Y(y_0)} dx$$

Определение 12.5. Условной плотностью случайной величины X при условии $Y = y_0$ называется величина, равная

$$\begin{cases} \frac{f(x, y_0)}{f_Y(y_0)} & , \text{ если } f_Y(y_0) > 0; \\ 0 & , \text{ если } f_Y(y_0) = 0. \end{cases}$$

Замечание. Пусть случайная величина Y имеет равномерное распределение на отрезке $[0; 1]$. Обозначим

$$N_Y = \{z \in \mathbb{R} : f_Y(z) = 0\}$$

тогда

$$P(Y \in N_Y) = \int_{N_Y} f_Y(z) dz = 0.$$

Замечание. Так как $f(z, y)$ и $f_Y(y)$ — плотности, то

$$\frac{f(z, y_0)}{f_Y(y_0)} \geq 0,$$

и

$$\int_{\mathbb{R}} \frac{f(z, y_0)}{f_Y(y_0)} dz = \frac{1}{f_Y(y_0)} \int_{\mathbb{R}} f(z, y_0) dz = \frac{f_Y(y_0)}{f_Y(y_0)} = 1.$$

Следовательно, условная плотность $\frac{f(z, y_0)}{f_Y(y_0)}$ есть плотность.

Определение 12.6. Условным математическим ожиданием случайной величины X при фиксированном значении случайной величины Y называется величина, равная $E(X|Y = y_0) =$

$$\begin{cases} \int_{-\infty}^{+\infty} \frac{xf(x, y_0)}{f_Y(y_0)} dx & , \text{ если } f_Y(y_0) > 0; \\ 0 & , \text{ если } f_Y(y_0) = 0. \end{cases}$$

Определение 12.7. Условным математическим ожиданием случайной величины X относительно случайной величины Y называется случайная величина $g(Y)$:

$$g(Y)(\omega) = g(Y(\omega)) = E(X|Y(\omega)). \quad (B)$$

Утверждение 12.1. Для любой ограниченной борелевской функции $h(y)$

$$Eh(Y)X = Eh(Y)g(Y) \quad (C)$$

где $g(Y)$ из формулы (B)

Доказательство. Распишем правую часть:

$$\begin{aligned} Eh(Y)g(Y) &= \int_{-\infty}^{+\infty} h(y)g(y)f_Y(y)dy = \int_{-\infty}^{+\infty} h(y)f_Y(y)dy \int_{-\infty}^{+\infty} \frac{xf(x, y_0)}{f_Y(y_0)} dx \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \underbrace{h(y)x}_{\text{ф-ия сл. в.}} \underbrace{f(x, y)}_{\text{плотность сл. в.}} dx dy \\ &= Eh(Y)X. \end{aligned}$$

Можно показать, что равенство (C) однозначно определяет $g(Y)$, то есть если (C) выполнено для всех ограниченных борелевских функций $h(y)$ при функциях $g_1(y)$ и $g_2(y)$ (где $g_1(y)$ и $g_2(y)$ считаются измеримыми относительно \mathfrak{B}_Y — борелевской σ -алгебры на прямой, порожденной случайной величиной Y), то

$$g_1(y) = g_2(y) \text{ с вероятностью } 1.$$

Таким образом формулу (C) можно считать определением условного математического ожидания. \square

Замечание. Функция $g_1(y)$ называется измеримой относительно \mathfrak{B}_Y , если

$$\{z : g(z) < a\} \in \mathfrak{B}_Y \quad \forall a \in \mathbb{R}.$$

Замечание. Мы не всегда можем взять в качестве $g(Y)$ случайную величину X , так как она может не быть измеримой относительно \mathfrak{B}_Y .

12.2 Свойства условного математического ожидания

Пусть X и Y — случайные величины, тогда имеют место свойства:

1°. $E(cX|Y) = cE(X|Y)$;

2°. $E(X + Z|Y) = E(X|Y) + E(Z|Y)$;

3°. $E(X|X) = X$ (так как $E(X|X = a_l) = a_l$);

4°. Если X и Y — независимые случайные величины, то

$$E(X|Y) = \sum_{j=1}^{j=k} a_j \frac{P(X = a_j, Y = b_i)}{P(Y = b_i)} = \sum_{j=1}^{j=k} a_j P(X = a_j) \frac{P(Y = b_i)}{P(Y = b_i)} = EX,$$

то есть равно константе.

5°. $EX = E(E(X|Y))$

Лекция 13

13.1 Введение в математическую статистику

Математическая статистика — это математическая дисциплина, родственная теории вероятностей. Она базируется на методах теории вероятностей, но решает свои задачи своими методами.

Математические модели случайных величин в теории вероятностей основываются на понятии вероятностного пространства $(\Omega, \mathcal{A}, \mathbb{P})$. Где Ω — непустое множество, называемое пространством элементарных исходов, \mathcal{A} — класс событий из Ω . Как известно, чтобы \mathcal{A} стал σ -алгеброй надо наложить некоторые требования. \mathbb{P} — вероятность, заданная на событиях из \mathcal{A} . В теории вероятностей \mathbb{P} считается заданной, и задачей теории вероятностей считается разработка методов определения вероятностей более сложных событий.

На практике при изучении некоторого эксперимента редко бывает, что \mathbb{P} полностью известна. Часто можно лишь утверждать, что \mathbb{P} является элементом некоторого класса \mathcal{P} . Этот класс может включать все вероятности, которые можно задать на \mathcal{A} , а может представлять собой более узкое семейство вероятностей. В любом случае \mathcal{P} — это совокупность допустимых вероятностей \mathbb{P} . Таким образом, мы пришли к понятию вероятностно-статистической модели (или просто к статистической модели), понимая под этим набор $(\Omega, \mathcal{A}, \mathcal{P})$.

Пример. Рассмотрим эксперимент, состоящий из n независимых испытаний, в каждом из которых наблюдается 1 в случае успеха и 0 в случае неудачи с вероятностями соответственно p и $q = 1 - p$. Исход эксперимента можно представить вектором $\omega = (x_1, \dots, x_n)$, где $x_i = 0, 1$. Если вероятность p успеха известна, то вероятностной моделью является набор $(\Omega, \mathcal{A}, \mathbb{P})$, где $\Omega = \{\omega : \omega = (x_1, \dots, x_n), x_i = 0, 1\}$, \mathcal{A} — совокупность всех подмножеств Ω и вероятностью \mathbb{P} :

$$\mathbb{P}(\omega) = p^{\sum x_i} q^{n - \sum x_i}.$$

Предположим теперь, что вероятность успеха не известна. Обозначим ее за θ , и мы можем только сказать, что $\theta \in [0, 1]$. Таким образом семейство допустимых вероятностей имеет вид $\mathcal{P} = \{\mathbb{P}_\theta, \theta \in \Theta\}$, где $\mathbb{P}_\theta = \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i}$.

Нашей задачей будет сузить класс \mathcal{P} .

Исходные статистические данные — результат наблюдения конечной совокупности случайных величин $\mathbf{X} = (X_1, \dots, X_n)$. При этом говорят, что эксперимент состоит в проведении n испытаний, в которых результат i -го испытания описывается случайной величиной X_i . Совокупность $\mathbf{X} = (X_1, \dots, X_n)$ называют **выборкой**, величины X_i — элементами выборки, а их число n — объемом выборки.

Реализации выборки \mathbf{X} будем обозначать $\mathbf{x} = (x_1, \dots, x_n)$.

Пусть $\mathcal{X} = \{x\}$ — множество на котором задано распределение случайного вектора \mathbf{X} , т.е. множество всех возможных значений выборки \mathbf{X} . \mathcal{F} — класс допустимых распределений случайной величины \mathbf{X} , заданных на \mathcal{X} . Таким образом под статистической моделью эксперимента будем понимать набор $(\mathcal{X}, \mathcal{F})$.

Следовательно, с каждой выборкой \mathbf{X} свяжем ее функцию распределения $F_{\mathbf{X}}(\mathbf{x})$:

$$F_{\mathbf{X}}(x_1, \dots, x_n) = \mathbb{P}(X_1 < x_1, \dots, X_n < x_n).$$

Мы будем рассматривать случай, когда все компоненты выборки независимы и распределены так же, как и некоторая случайная величина ξ . Такую выборку мы будем называть повторной. Этот случай соответствует эксперименту, в котором проводятся повторные независимые наблюдения над случайной величиной ξ . Тогда $F_{\mathbf{X}}(\mathbf{x}) = F_{\xi}(x_1) \dots F_{\xi}(x_n)$. И говорят, что $\mathbf{X} = (X_1, \dots, X_n)$ — выборка из распределения случайной величины ξ . Иногда множество возможных значений ξ называют генеральной совокупностью, а величину \mathbf{X} — выборкой из этой совокупности.

Если функции распределения из класс \mathcal{F} заданы с точностью до значений некоторого параметра θ с множеством допустимых значений Θ , то такая модель $\mathcal{F} = \{F(x, \theta), \theta \in \Theta\}$ называется параметрической. В этом случае известен тип распределения, но не известен параметр, от которого зависит распределение.

13.2 Эмпирическая функция распределения

Далее нам потребуется понятие **эмпирической функции распределения**.

Определение 13.1. *Эмпирической функцией распределения, построенной по случайной выборке (X_1, \dots, X_n) , называется случайная функция*

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{I}(X_i < x).$$

Для конкретной реализации \mathbf{x} выборки \mathbf{X} и фиксированного числа x величина $F_n(x)$ равна доле тех значений x_i , которые меньше x .

Следующая теорема показывает, что функция $F_n(x)$ является хорошей оценкой для теоретической функции распределения $F(x)$.

Теорема 13.1. *Пусть (X_1, \dots, X_n) — повторная выборка значений случайной величины X , имеющей функцию распределения $F(x)$. Тогда для любого $x \in \mathbb{R}$*

$$\mathbb{P}(\lim_{n \rightarrow \infty} F_n(x) = F(x)) = 1.$$

Доказательство. Обозначим $Y_i = \mathbf{I}(X_i < x)$. Случайные величины Y_1, \dots, Y_n — независимые и одинаково распределенные.

Из определения Y_i следует, что $\mathbb{E}Y_i = \mathbb{P}(X_i < x) = F(x)$ и $\mathbb{D}Y_i = \mathbb{E}Y_i^2 - (\mathbb{E}Y_i)^2 = F(x)(1 - F(x)) < \infty$. Следовательно применим усиленный закон больших чисел и

$$F_n(x) = \frac{Y_1 + \dots + Y_n}{n} \xrightarrow{\text{п.н.}} F(x).$$

□

Весьма замечательно, что имеет место более сильный результат о том, что сходимость в теореме равномерна по x .

Теорема 13.2. *Пусть выполняются условия предыдущей теоремы. Тогда*

$$\mathbb{P}(\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = 0) = 1.$$

Доказательство этого результата приведено например в книге А.Н. Ширяева "Вероятность - 1".

13.3 Эмпирические или выборочные моменты

Пусть (X_1, \dots, X_n) — выборка из распределения $\mathcal{L}(X)$ и $F(x)$ и $F_n(x)$ — соответственно теоретическая и эмпирическая функции распределения. Точно так же, как $F(x)$ ставится в соответствие $F_n(x)$, любой теоретической характеристике можно поставить в соответствие эмпирическую характеристику.

Дадим определение эмпирического момента.

Определение 13.2. Эмпирическим (выборочным) моментом называется момент случайной величины, имеющей эмпирическую функцию распределения в качестве функции распределения.

Обозначим выборочный момент \bar{X} . Из выше сказанного заключаем, что

$$\bar{X} = \int_{\mathbb{R}} x dF_n(x) = \frac{1}{n} \sum_{i=1}^n X_i.$$

Вообще, если $g = \int_{\mathbb{R}} g(x) dF(x)$ — некоторая теоретическая характеристика распределения $F(x)$, то ей можно поставить в соответствие эмпирическую характеристику по формуле

$$G = \int_{\mathbb{R}} g(x) dF_n(x) = \frac{1}{n} \sum_{i=1}^n g(X_i).$$

Если $g(x) = x^k$, то G — выборочный момент k -го порядка и он равен

$$A_k = A_k(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n X_i^k.$$

Выборочным центральным моментом k -го порядка называют случайную величину

$$M_k = M_k(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k.$$

При $k = 2$ величину M_k называют выборочной дисперсией и обозначают символом $S^2 = S^2(\mathbf{X})$:

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Между эмпирическими моментами сохраняются те же соотношения, что и между соответствующими теоретическими моментами.

Так как эмпирическая характеристика G является с своей очередь случайной величиной, то мы можем говорить о ее распределении и различных характеристиках.

Вычислим математическое ожидание и дисперсию выборочного среднего и выборочной дисперсии. Так как все X_i независимы и распределены так же как и наблюдаемая случайная величина X , то для выборочного среднего имеем

$$\begin{aligned} \mathbb{E}\bar{X} &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}X_i = \mathbb{E}X, \\ \mathbb{D}\bar{X} &= \frac{\mathbb{D}X_1 + \dots + \mathbb{D}X_n}{n^2} = \frac{\mathbb{D}X}{n}. \end{aligned}$$

Вычислим теперь математическое ожидание выборочной дисперсии. Заметим, что выборочная дисперсия инварианта относительно сдвига. Действительно

$$\begin{aligned} S^2(X_1 + c, \dots, X_n + c) &= \frac{1}{n} \sum_{i=1}^n \left(X_i + c - \frac{(X_1 + c) + \dots + (X_n + c)}{n} \right)^2 \\ &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = S^2(\bar{X}). \end{aligned}$$

Пусть $EX_1 = m, DX_1 = \sigma^2$. Положим, $Y_i = X_i - m$. В силу инвариантности относительно сдвига имеем $S^2(\mathbf{X}) = S^2(\mathbf{Y})$ и

$$\begin{aligned} S^2 &= \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2 = \frac{1}{n} \sum_{i=1}^n (Y_i^2 - 2\bar{Y}Y_i + \bar{Y}^2) \\ &= \frac{1}{n} \sum_{i=1}^n Y_i^2 - \frac{2n\bar{Y}^2}{n} + \frac{n\bar{Y}^2}{n} = \frac{1}{n} \sum_{i=1}^n Y_i^2 - \bar{Y}^2. \end{aligned} \quad (1)$$

Используя тот факт, что $EY_i = 0$, и то, что математическое ожидание произведения независимых случайных величин равно произведению математических ожиданий, получим

$$\begin{aligned} E\bar{Y}^2 &= \frac{1}{n^2} E \left(\sum_{i=1}^n Y_i \right)^2 = \frac{1}{n^2} \left(E \sum_{i=1}^n Y_i^2 + 2 \sum_{i < j} Y_i Y_j \right) \\ &= \frac{1}{n^2} \left(\sum_{i=1}^n EY_i^2 + 2 \sum_{i < j} EY_i EY_j \right) = \frac{1}{n^2} \sum_{i=1}^n EY_i^2 = \frac{\sigma^2}{n}. \end{aligned}$$

Теперь, применяя (1), получим

$$\begin{aligned} ES^2(\mathbf{X}) &= ES^2(\mathbf{Y}) \\ &= \frac{1}{n} \sum_{i=1}^n EY_i^2 - E\bar{Y}^2 = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n} \sigma^2. \end{aligned}$$

Замечание. Часто используют другое определение выборочной дисперсии:

$$S^2(\mathbf{X}) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Оно лучше тем, что $ES^2(\mathbf{X}) = \sigma^2$. Чтобы отличать от выборочной дисперсии введенной ранее, такую выборочную дисперсию будем обозначать $S_{re}^2 = S_{re}^2(\mathbf{X})$ и называть исправленной выборочной дисперсией.

Введем еще одно определение, которое нам понадобится в будущем.

Определение 13.3. Последовательно случайных величин $\{Y_n\}$ — асимптотически нормальна с параметрами (a_n, b_n) , если для любого z

$$\mathbf{P} \left\{ \frac{Y_n - a_n}{b_n} < z \right\} \rightarrow \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{u^2}{2}} du, \quad n \rightarrow \infty.$$

Теорема 13.3. Последовательность \bar{X}_n — асимптотически нормальна.

Доказательство. Возьмем $a_n = a, b_n = \frac{\sigma}{\sqrt{n}}$. Тогда в силу центральной предельной теоремы

$$\frac{\bar{X}_n - a}{b_n} = \frac{X_1 + \dots + X_n - na}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

□

Замечание. Теорема остается справедливой для выборочных моментов любого порядка k .

13.4 Порядковые статистики и вариационные ряды

Каждой реализации \mathbf{x} выборки \mathbf{X} можно поставить в соответствие упорядоченную последовательность:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}.$$

Обозначим через $X_{(k)}$ случайную величину, которая для каждой реализации \mathbf{x} выборки \mathbf{X} принимает значение $x_{(k)}$. Таким образом мы по выборке $\mathbf{X} = (X_1, \dots, X_n)$ определили новую последовательность случайных величин $X_{(1)}, \dots, X_{(n)}$, называемых порядковыми статистиками. $X_{(k)}$ — k -ая порядковая статистика. $X_{(1)}, X_{(n)}$ — экстремальные значения выборки.

Очевидно, что порядковые статистики удовлетворяют неравенствам

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}. \quad (2)$$

Последовательность (2) называется вариационным рядом.

Интересно поставить вопрос о распределении $X_{(k)}$.

Для $X_{(n)}$ и $X_{(1)}$ имеем

$$P(X_{(n)} < z) = P\left(\bigcap_{i=1}^n \{X_i < z\}\right) = \prod_{i=1}^n P(X_i < z) = F^n(z).$$

$$P(X_{(1)} < z) = 1 - P(X_{(1)} \geq z) = 1 - P\left(\bigcap_{i=1}^n \{X_i \geq z\}\right) = 1 - (1 - F(z))^n.$$

Лемма 13.1.

$$P(X_{(k)} < z) = \sum_{i=k}^n C_n^i F^i(z) (1 - F(z))^{n-i}.$$

Доказательство. Обозначим через $\mu_n(z)$ — число событий $\{j : X_j < z\}$. Если вспомнить определение эмпирической функции распределения, то

$$F_n(z) = \frac{\mu_n(z)}{n}.$$

Из равенства $\{X_k < z\} = \{\mu_n(z) \geq k\}$, доказательство справедливости которого мы предоставляем читателю, вытекает, что

$$P(X_{(k)} < z) = P(\mu_n(z) \geq k) = P\left(\bigcap_{i=k}^n (\mu_n(z) = i)\right).$$

События $\{\mu_n(z) = i\}$ несовместны, и $\mu_n(z) = i$ означает, что из n случайных величин ровно i меньше z , а остальные не меньше z . Следовательно

$$P(X_{(k)} < z) = \sum_{i=k}^n P(\mu_n(z) = i).$$

Так как $\mu_n(z)$ имеет биномиальное распределение с параметрами n и $p = P(X < z) = F(z)$, то доказательство леммы окончено. \square

13.5 Статистические оценки

С этим понятием мы уже сталкивались выше, когда говорили, что эмпирическая функция распределения является оценкой для теоретической функции распределения или выборочная характеристика является оценкой теоретической характеристики. При этом использование понятия "оценка" основалось на том, что при достаточно большом объеме выборки, большие отклонения эмпирической характеристики от теоретической маловероятны.

При применении статистической теории на практике, часто приходится строить оценки теоретических характеристик при ограниченных объемах выборки. Мы изложим некоторые общие методы решения подобных задач.

Пусть у нас есть параметрическая статистическая модель $\mathcal{F} = \{F(x, \theta), \theta \in \Theta\}$ и выборка $\mathbf{X} = (X_1, \dots, X_n)$ из распределения $\mathcal{L}(X) \in \mathcal{F}$.

Априорная информация о наблюдаемой случайной величине X состоит в том, что ее функция распределения $F(x)$ принадлежит некоторому заданному параметрическому семейству \mathcal{F} , т.е. имеет известную функциональную форму, но зависит от некоторого неизвестного параметра θ , который принадлежит параметрическому множеству Θ . Таким образом имеем задачу: по выборке \mathbf{X} сделать статистические выводы об истинном значении неизвестного параметра θ .

Определение 13.4. *Статистикой называется любая случайная величина, являющаяся функцией лишь от выборки \mathbf{X} .*

При точечном оценивании ищут статистику $T = T(\mathbf{X})$, которую при заданной реализации \mathbf{x} выборки \mathbf{X} принимают за приближенное значение параметра θ .

Для оценивания θ можно использовать различные оценки. Но чтобы выбрать лучшую из них нужно иметь критерии качества оценок. В разных случаях меры качества различны.

Введем понятие несмещенной и состоятельной оценки.

Определение 13.5. *Оценка $T = T(\mathbf{X})$ является несмещенной оценкой неизвестного параметра θ , если*

$$ET = \theta.$$

Замечание. Иногда требуется найти несмещенную оценку не самого неизвестного параметра θ , а некоторой функции $\tau(\theta)$ от него. В этом случае оценка $T = T(\mathbf{X})$ является несмещенной оценкой $\tau(\theta)$, если

$$ET = \tau(\theta).$$

Определение 13.6. *Оценка $T = T(\mathbf{X})$ является состоятельной оценкой неизвестного параметра θ , если при неограниченном увеличении объема выборки*

$$T \xrightarrow{p} \theta$$

13.5.1 Свойства несмещенных оценок

1°. *Несмещенные оценки не единственны.*

К примеру в качестве несмещенной оценки для математического ожидания EX могут выступать EX_1 или $E\bar{X}$.

2°. *Несмещенные оценки могут не существовать.*

Пример. Пусть $\mathcal{L}(X)$ — пуассоновское распределение с параметром θ . Производится одно наблюдение X над пуассоновской случайной величиной. Требуется оценить параметрическую функцию $\tau(\theta) = \frac{1}{\theta}$. Пусть $T(X)$ — несмещенная оценка для $\tau(\theta)$, т.е.

$$\sum_{x=0}^{\infty} T(x) e^{-\theta} \frac{\theta^x}{x!} = \frac{1}{\theta}, \quad \text{или}$$
$$\sum_{x=0}^{\infty} T(x) e^{-\theta} \frac{\theta^{x+1}}{x!} = e^{\theta} = \sum_{r=0}^{\infty} \frac{\theta^r}{r!}.$$

Очевидно, что не существует функции $T(X)$, удовлетворяющей этому соотношению и не зависящей от θ .

3°. *Несмещенные оценки могут существовать, но быть бессмысленными.*

Пример. Пусть $\mathcal{L}(X)$ — отрицательное биномиальное распределение с параметрами 1 и θ . Производится одно наблюдение X . Требуется оценить параметр θ . Пусть $T(X)$ — несмещенная оценка для θ , т.е.

$$\sum_{x=0}^{\infty} T(x) \theta^x = \frac{\theta}{1-\theta} = \sum_{r=1}^{\infty} \theta^r, \quad \forall \theta \in (0, 1).$$

Приравнивая коэффициенты при одинаковых степенях θ , получаем, что единственной несмещенной оценкой является статистика

$$T(X) = \begin{cases} 0, & \text{при } X = 0, \\ 1, & \text{при } X \geq 1. \end{cases}$$

Но значения этой статистики не принадлежат параметрическому множеству $\Theta = (0, 1)$. Поэтому такая оценка бессмысленна.

4°. *Из того, что $ET(X) = \theta$, вообще говоря, не следует, что $Ef(T(X)) = f(\theta)$.*

13.5.2 Свойства состоятельных оценок

1°. *Состоятельные оценки не единственны.*

Действительно, выборочная дисперсия S^2 и исправленная выборочная дисперсия S_{re}^2 являются состоятельными оценками теоретической дисперсии.

2°. *Состоятельные оценки могут быть смещенными.*

Действительно, как было показано ранее выборочная дисперсия является состоятельной, но смещенной оценкой теоретической дисперсии.

Лекция 14

14.1 Оптимальные оценки

Пусть нам требуется оценить заданную параметрическую функцию $\tau(\theta)$ в модели $\mathcal{F} = \{F(x, \theta), \theta \in \Theta\}$ по выборке $\mathbf{X} = (X_1, \dots, X_n)$ из распределения $\mathcal{L}(X) \in \mathcal{F}$. Пусть статистики $T = T(\mathbf{X})$ являются несмещенными оценками заданной параметрической функции. Обозначим через \mathcal{T}_τ семейство всех несмещенных оценок. Предположим, что дисперсии всех оценок из \mathcal{T}_τ конечны, т.е. $D_\theta T = E_\theta(T - \tau(\theta))^2 < \infty$ для $\forall \theta \in \Theta$.

Рассмотрим теперь две статистики $T_1, T_2 \in \mathcal{T}_\tau$. $E T_1 = E T_2$. Возникает вопрос, какую из оценок выбрать. Рассмотрим дисперсии оценок. Если $D_\theta T_1 < D_\theta T_2$, то берем T_1 , поскольку чем меньше дисперсия, тем меньше разброс среднего. Но оценка должна выполняться для $\forall \theta \in \Theta$.

Таким образом мы пришли к определению оптимальной оценки.

Определение 14.1. Пусть \mathcal{T}_τ семейство несмещенных оценок параметрической функции $\tau(\theta)$. Оценка $T = T(\mathbf{X})$ называется оптимальной, если $D_\theta T < D_\theta T_1$, для любого $\theta \in \Theta$ и $T_1 \in \mathcal{T}_\tau$.

Имеет место следующий результат, показывающий, что если оптимальная оценка существует, то она единственна с точностью до меры нуль.

Теорема 14.1. Пусть статистики $T_1, T_2 \in \mathcal{T}_\tau$. Если T_1 и T_2 — оптимальны, то $T_1 = T_2$ с вероятностью 1.

Доказательство. Так как T_1 и T_2 — оптимальные оценки, то $D T_1 = D T_2$. Положим $D T_1 = D T_2 = \sigma^2$. Определим новую оценку: $T_3 = \frac{T_1 + T_2}{2} \in \mathcal{T}_\tau$. Вычислим дисперсию новой оценки.

$$2T_3 = T_1 + T_2;$$

$$D(2T_3) = D(T_1 + T_2) = D T_1 + D T_2 + 2cov(T_1, T_2) = 2\sigma^2 + 2cov(T_1, T_2).$$

Так как σ^2 — наименьшая дисперсия, то $D T_3 \geq \sigma^2$. Следовательно $cov(T_1, T_2) \geq \sigma^2 = \sqrt{D T_1} \sqrt{D T_2}$. И для коэффициента корреляции имеем

$$\rho = \frac{cov(T_1, T_2)}{\sqrt{D T_1} \sqrt{D T_2}} \geq 1.$$

Но, как известно, коэффициент корреляции $|\rho| \leq 1$. Следовательно $\rho = 1$.

Из свойств коэффициента корреляции известно, что, если он равен 1, то случайные величины линейны зависимы с вероятностью 1. Таким образом $T_1 = a T_2 + b$ с вероятностью 1. Так как T_1 и T_2 несмещенные оценки, то $E T_1 = E T_2 = \theta$. И мы получили, что $\theta = a\theta + b$ или $\theta(1 - a) = b$. Следовательно

$$cov(T_1, T_2) = E(T_1 - \theta)(T_2 - \theta) = E(aT_2 + b - \theta)(T_2 - \theta) = aE(T_2 - \theta)^2 = a\sigma^2.$$

Так как $\rho = 1$, то отсюда находим, что $a = 1, b = 0$, т.е. $T_1 = T_2$. □

14.2 Неравенство Рао - Крамера

14.2.1 Функция правдоподобия

Пусть $\mathbf{X} = (X_1, \dots, X_n)$ — выборка из $\mathcal{L}(X) \in \mathcal{F}$. $\mathbf{x} = (x_1, \dots, x_n)$ — реализация выборки \mathbf{X} .

Если X имеет дискретное распределение, то определим функцию $L_n = L_n(\mathbf{x}, \theta)$ следующим образом

$$L_n = \prod_{i=1}^n P(X_i = x_i),$$

если имеет абсолютно непрерывно распределение, то

$$L_n = \prod_{i=1}^n f(x_i, \theta),$$

где $f(x, \theta)$ - плотность распределения наблюдаемой случайной величины X .

Функция $L_n(\mathbf{x}, \theta)$ называется функцией правдоподобия. В дискретном случае функция правдоподобия является вероятностью того, что выборка \mathbf{X} примет значение (x_1, \dots, x_n) . В непрерывном случае функция правдоподобия является совместной плотностью распределения выборки \mathbf{X} .

Прежде чем сформулировать и доказать неравенство Рао - Крамера докажем вспомогательную лемму.

Лемма 14.1. *Предположим, что для $\forall \theta \in \Theta \subset \mathbb{R}$ существуют $\frac{\partial L_n}{\partial \theta}$, $\frac{\partial^2 L_n}{\partial \theta^2}$, и существуют конечные $E \left| \frac{\partial^2}{\partial \theta^2} \ln L_n \right|$ и $E \left(\frac{\partial}{\partial \theta} \ln L_n \right)^2$. Тогда*

$$E \left(\frac{\partial \ln L_n}{\partial \theta} \right) = 0, \quad (1)$$

$$E \left(\frac{\partial \ln L_n}{\partial \theta} \right)^2 = -E \frac{\partial^2}{\partial \theta^2} \ln L_n. \quad (2)$$

Доказательство. Доказательство проведем для абсолютно непрерывного случая. Для дискретного случае доказательство предлагается провести читателю самостоятельно.

Как отмечалось выше, в абсолютно непрерывном случае функция $L_n(\mathbf{x}, \theta)$ является совместной плотностью выборки \mathbf{X} , следовательно

$$1 = \int_{\mathbb{R}^n} L_n(y, \theta) dy, \quad y = (x_1, \dots, x_n).$$

Продифференцируем это тождество по θ (это мы можем сделать по условию) и получим

$$0 = \int_{\mathbb{R}^n} \frac{\partial L_n}{\partial \theta} dy = \int_{\mathbb{R}^n} \frac{\partial \ln L_n}{\partial \theta} L_n dy = E \frac{\partial}{\partial \theta} \ln L_n(\mathbf{x}, \theta).$$

Тем самым (1) доказано. Следующие выкладки доказывают (2)

$$\begin{aligned}
\mathbb{E} \frac{\partial^2}{\partial \theta^2} \ln L_n(\mathbf{x}, \theta) &= \int_{\mathbb{R}^n} \frac{\partial^2 \ln L_n}{\partial \theta^2} L_n dy \\
&= \int_{\mathbb{R}^n} \frac{L_n \frac{\partial^2 L_n}{\partial \theta^2} - \left(\frac{\partial L_n}{\partial \theta} \right)^2}{L_n^2} L_n dy \\
&= - \int_{\mathbb{R}^n} \left(\frac{\partial}{\partial \theta} \ln L_n \right)^2 L_n dy = -\mathbb{E} \left(\frac{\partial}{\partial \theta} \ln L_n(\mathbf{x}, \theta) \right)^2.
\end{aligned}$$

□

Определение 14.2. *Количеством информации по Фишеру, содержащейся в выборке $\mathbf{X} = (X_1, \dots, X_n)$, называется величина*

$$I_n(\theta) = \mathbb{E} \left(\frac{\partial \ln L_n(\mathbf{x}, \theta)}{\partial \theta} \right)^2.$$

Из только что доказанной леммы следует, что

$$\begin{aligned}
\mathbb{E} \left(\frac{\partial \ln L_n(\mathbf{x}, \theta)}{\partial \theta} \right)^2 &= -\mathbb{E} \frac{\partial^2}{\partial \theta^2} \ln L_n(\mathbf{x}, \theta) = -\mathbb{E} \frac{\partial^2}{\partial \theta^2} \sum_{i=1}^n \ln f(x_i, \theta) \\
&= -n \mathbb{E} \frac{\partial^2}{\partial \theta^2} \ln f(x_1, \theta) = n I_1(\theta).
\end{aligned}$$

Теорема 14.2 (неравенство Рао - Крамера). *Пусть выполнены все условия предыдущей леммы и $\tau(\theta)$ дифференцируемая функция от $\theta \in \Theta$, для которой существует несмещенная оценка $T = T(\mathbf{X})$ и*

$$DT(\mathbf{X}) < \infty, \quad \int_{\mathbb{R}^n} \left| \frac{\partial L_n}{\partial \theta} \right| dy < \infty \quad \text{для } \forall \theta \in \Theta$$

Тогда

$$DT(\mathbf{X}) \geq \frac{(\tau'(\theta))^2}{I_n(\theta)}. \quad (3)$$

При этом равенство имеет место тогда и только тогда, когда

$$\frac{\partial}{\partial \theta} \ln L_n(\mathbf{x}, \theta) = c(\theta)(T(\mathbf{x}) - \tau(\theta)),$$

с некоторой функцией $c(\theta)$ или, что эквивалентно

$$L_n(\mathbf{x}, \theta) = \exp\{\psi_1(\theta) + \psi_2(\theta) + f(\mathbf{x})\}. \quad (4)$$

Доказательство. Так как $T(\mathbf{X})$ — несмещенная оценка, то

$$\int_{\mathbb{R}^n} T(y) L_n(y, \theta) dy = \tau(\theta).$$

Продифференцируем это равенство по θ (это мы можем сделать в силу условия теоремы) и, применяя неравенство Коши - Буняковского - Шварца, получим

$$\begin{aligned} |\tau'(\theta)| &= \left| \int_{\mathbb{R}^n} T(y) \frac{\partial L_n}{\partial \theta} dy \right| = \left| \int_{\mathbb{R}^n} T(y) \frac{\partial \ln L_n}{\partial \theta} L_n dy \right| \\ &= \left| \mathbb{E} \left(T(\mathbf{X}) \frac{\partial \ln L_n(\mathbf{x}, \theta)}{\partial \theta} \right) \right| = \left| \mathbb{E} (T(\mathbf{X}) - \tau(\theta)) \left(\frac{\partial \ln L_n(\mathbf{x}, \theta)}{\partial \theta} - 0 \right) \right| \\ &\leq \left(DT(\mathbf{X}) \cdot D \frac{\partial \ln L_n}{\partial \theta} \right)^{1/2} = (DT(\mathbf{X}) \cdot I_n(\theta))^{1/2}. \end{aligned}$$

Что и доказывает нашу теорему. □

Следствие 14.1. Если $\tau(\theta) = \theta$, то

$$DT(\mathbf{X}) \geq \frac{1}{nI_n(\theta)}.$$

Замечание. Многие часто встречающиеся модели удовлетворяют условию (4). Например модели, в которых закон распределения представляет собой

- нормальное распределение $\mathcal{N}(\theta, \sigma^2)$,
- нормальное распределение $\mathcal{N}(a, \theta)$,
- пуассоновское распределение $\Pi(\theta)$,
- биномиальное распределение $Bi(n, \theta)$,
- гамма - распределение $\Gamma(\theta, \lambda)$,

где θ - неизвестный параметр.

Пример. Пусть имеется выборка $\mathbf{X} = (X_1, \dots, X_n)$ из биномиального распределения с параметрами 1 и θ ($\mathcal{L}(X) \sim Bi(1, \theta)$). Требуется найти оптимальную оценку параметра θ .

Здесь $\mathbb{E}_\theta X_i = \theta$, поэтому $T(\mathbf{X}) = \bar{X}$ является несмещенной оценкой параметра θ . Покажем, что $T(\mathbf{X})$ является оптимальной оценкой.

Вычислим дисперсию $T(\mathbf{X})$

$$DT(\mathbf{X}) = \frac{1}{n^2} \sum_{i=1}^n DX_i = \frac{\theta(1-\theta)}{n}.$$

Вычислим теперь информацию по Фишеру

$$I_1(\theta) = \mathbb{E} \left(\frac{\partial \ln L_1}{\partial \theta} \right)^2 = -\mathbb{E} \frac{\partial^2 \ln L_1}{\partial \theta^2}.$$

Так как для $n = 1$ функция правдоподобия имеет вид $L_1 = \theta^{x_1}(1 - \theta)^{1-x_1}$, то

$$\begin{aligned} \ln L_1 &= x_1 \ln \theta + (1 - x_1) \ln(1 - \theta), \\ \frac{\partial \ln L_1}{\partial \theta} &= \frac{x_1}{\theta} - \frac{(1 - x_1)}{1 - \theta} = \frac{x_1 - \theta}{\theta(1 - \theta)}, \\ E \left(\frac{x_1 - \theta}{\theta(1 - \theta)} \right)^2 &= \frac{\theta(1 - \theta)}{\theta^2(1 - \theta)^2} = \frac{1}{\theta(1 - \theta)}. \end{aligned}$$

Таким образом $DT(\mathbf{X}) = \frac{1}{nI_1(\theta)}$, т.е. оценка $T(\mathbf{X}) = \bar{X}$ имеет наименьшую дисперсию для любого θ . Следовательно является оптимальной.

Сформулируем теперь этот результат в виде теоремы.

Теорема 14.3. *Относительная частота произвольного события в n независимых испытаниях является оптимальной оценкой для вероятности этого события.*

Следствие 14.2. *Для любой фиксированной выборки эмпирическая функция распределения является оптимальной оценкой теоретической функции распределения.*

Определение 14.3. *Эффективностью оценки $T = T(\mathbf{X})$, которая является несмещенной оценкой параметрической функции $\tau(\theta)$, называется величина*

$$e(T) = \frac{(\tau'(\theta))^2}{DT \cdot I_n(\theta)}.$$

Замечание. Из определения и неравенства Рао - Крамера следует, что $0 \leq e(T) \leq 1$.

Определение 14.4. *Несмещенная оценка $T = T(\mathbf{X})$ называется эффективной, если ее эффективность равна 1.*

Замечание. Можно сформулировать это определение иначе: оценка $T = T(\mathbf{X})$ называется эффективной, если в неравенстве Рао - Крамера достигается равенство.

14.3 Метод моментов

Рассмотрим статистическую модель $\mathcal{F} = \{F(x, \theta), \theta \in \Theta\}$, где $\theta = (\theta_1, \dots, \theta_k)$. Пусть $\mathbf{X} = (X_1, \dots, X_n)$ выборка из распределения $\mathcal{L}(X) \in \mathcal{F}$. Предположим, что существует математическое ожидание $EX^k = a_k$, тогда существуют все моменты низших порядков. По выборке \mathbf{X} построим эмпирические моменты

$$m_i = \frac{1}{n}(X_1^i + \dots + X_n^i) = a_i = f_i(\theta_1, \dots, \theta_k).$$

Таким образом мы получаем систему из k уравнений с k неизвестными, которую иногда можно разрешить относительно неизвестных параметров θ_i :

$$\begin{cases} m_1 = a_1 = f_1(\theta_1, \dots, \theta_k) \\ \dots \\ m_k = a_k = f_k(\theta_1, \dots, \theta_k). \end{cases} \quad (5)$$

Определение 14.5. *Оценками по методу моментов называется решение $\theta_1^*, \dots, \theta_k^*$ системы (5).*

Пример. Пусть (X_1, \dots, X_n) — выборка из биномиального распределения $Bi(k, \theta)$, где k, θ — неизвестные параметры. Найдем их оценки по методу моментов.

Прежде всего нам необходимо найти два момента нашего распределения:

$$\begin{aligned} a_1 &= EX = k\theta, \\ a_2 &= EX^2 = DX + (EX)^2 = k\theta(1 - \theta) + (k\theta)^2. \end{aligned}$$

Таким образом система (5) запишется следующим образом

$$\begin{cases} a_1 = k\theta = m_1 \\ a_2 = k\theta(1 - \theta) + (k\theta)^2 = m_2 \end{cases}$$

Решая его, найдем оценки

$$\begin{aligned} \theta^* &= \frac{m_1^2 + m_1 - m_2}{m_1}, \\ k^* &= \frac{m_1^2}{m_1^2 + m_1 - m_2}. \end{aligned}$$

Часто оценки по методу моментов являются состоятельными. Так что имеет место следующая теорема

Теорема 14.4. Пусть $h(x)$ — непрерывная функция и $Y_n \xrightarrow{p} 0$. Тогда

$$h(Y_n + a) \xrightarrow{p} h(a).$$

Доказательство. Из непрерывности функции $h(x)$ следует, что для $\forall \varepsilon > 0 \exists \delta > 0 : |y| < \delta$:

$$|h(a + y) - h(a)| < \varepsilon.$$

Фиксируем $\varepsilon > 0$, тогда

$$\mathbb{P}(\underbrace{|h(a + Y_n) - h(a)| \geq \varepsilon}_A) = \mathbb{P}(\underbrace{A \cap |Y_n| < \delta}_\emptyset) + \mathbb{P}(A \cap |Y_n| \geq \delta) \leq \mathbb{P}(|Y_n| \geq \delta) \rightarrow 0.$$

□

Если оценки, найденные по методу моментов, окажутся непрерывными функциями, то с помощью обобщения теоремы на случай многих переменных, докажем, что они являются состоятельными оценками.

14.4 Достаточные статистики

Определение 14.6. Статистика $T = T(\mathbf{X})$ называется достаточной для модели $\mathcal{F} = \{F(x, \theta), \theta \in \Theta\}$, если распределение выборки $\mathbf{X} = (X_1, \dots, X_n)$ не зависит от θ при условии, что $T(\mathbf{X}) = t$.

Это свойство статистики означает, что она содержит всю информацию о параметре θ , имеющуюся в выборке.

Пример. Пусть $\mathcal{L}(X) = Bi(1, \theta)$. Покажем, что статистика $T(\mathbf{X}) = \sum X_i$ будет достаточной.

$$\begin{aligned} P(X_1 = x_1, \dots, X_n = x_n | T(\mathbf{X}) = t) &= \frac{P(\mathbf{X} = \mathbf{x}, T(\mathbf{X}) = t)}{P(T(\mathbf{X}) = t)} \\ &= \frac{\theta^{\sum X_i} (1 - \theta)^{n - \sum X_i}}{C_n^t \theta^t (1 - \theta)^{n-t}} = \frac{1}{C_n^{\sum X_i}}. \end{aligned}$$

Приведем следующий результат, называемый критерием факторизации, позволяющий определить, существует ли достаточная статистика и установить ее вид.

Теорема 14.5 (Критерий факторизации). $T(\mathbf{X})$ достаточная статистика тогда и только тогда, когда функция правдоподобия может быть представлена в виде

$$L_n(\mathbf{x}, \theta) = g(T(\mathbf{x}, \theta))h(\mathbf{x}).$$

Доказательство. Приведем доказательство для дискретного случая. Для абсолютно непрерывного доказательства предлагается провести читателю самостоятельно.

Пусть $T(\mathbf{X})$ достаточная статистика. Если $T(\mathbf{X}) = t$, то событие $\{\mathbf{X} = \mathbf{x}\} \subseteq \{T(\mathbf{X}) = t\}$. Поэтому

$$\begin{aligned} L(\mathbf{x}, \theta) = P_\theta(\mathbf{X} = \mathbf{x}) &= P_\theta(\mathbf{X} = \mathbf{x} | T(\mathbf{X}) = t) \\ &= \underbrace{P_\theta(T(\mathbf{X}) = t)}_{g(T(\mathbf{x}, \theta))} \underbrace{P_\theta(\mathbf{X} = \mathbf{x} | T(\mathbf{X}) = t)}_{h(\mathbf{x}, t)}. \end{aligned}$$

Пусть теперь функция правдоподобия имеет вид $L_n(\mathbf{x}, \theta) = g(T(\mathbf{x}, \theta))h(\mathbf{x})$. Тогда, если \mathbf{x} таково, что $T(\mathbf{x}) = t$, то

$$\begin{aligned} P(\mathbf{X} = \mathbf{x} | T(\mathbf{X}) = t) &= \frac{P(\mathbf{X} = \mathbf{x}, T(\mathbf{X}) = t)}{P(T(\mathbf{X}) = t)} \\ &= \frac{P(\mathbf{X} = \mathbf{x})}{\sum_{\mathbf{x}': T(\mathbf{x}') = t} P(\mathbf{X} = \mathbf{x}')} = \frac{g(t, \theta)h(\mathbf{x})}{\sum_{\mathbf{x}': T(\mathbf{x}') = t} g(t, \theta)h(\mathbf{x}')} = \frac{h(\mathbf{x})}{\sum_{\mathbf{x}': T(\mathbf{x}') = t} h(\mathbf{x}')}. \end{aligned}$$

Что и требовалось доказать. □

Пример. Рассмотрим общую нормальную модель $\mathcal{N}(\theta_1, \theta_2^2)$, где оба параметра неизвестны. В этом примере функция правдоподобия будет зависеть от векторного параметра $\theta = (\theta_1, \theta_2^2)$. Найдем функцию правдоподобия

$$\begin{aligned} L_n(\mathbf{x}, \theta) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(x_i - \theta_1)^2}{2\theta_2^2}} \\ &= \frac{1}{(\sqrt{2\pi\theta})^n} \exp\left(\frac{1}{2\theta_2^2} \sum_{i=1}^n (x_i - \bar{x})^2 - \frac{n(\bar{x} - \theta_1)^2}{2\theta_2^2}\right) = g(T(\mathbf{x}), \theta)h(\mathbf{x}). \end{aligned}$$

В качестве достаточной статистики можно взять векторную функцию

$$T(\mathbf{X}) = (\bar{x}, \sum_{i=1}^n (x_i - \bar{x})^2).$$

Это не единственный, но классический выбор достаточной статистики.

Пример. Пусть $\mathbf{X} = (X_1, \dots, X_n)$ — выборка из равномерного распределения на $[0, \theta]$, $\theta > 0$. Тогда функция правдоподобия будет иметь вид

$$L_n(\mathbf{x}, \theta) = \frac{1}{\theta^n} \underbrace{\mathbf{I}_{\{x_{(1)} > 0\}}}_{h(\mathbf{x})} \mathbf{I}_{\{x_{(n)} < \theta\}}.$$

Таким образом $T(\mathbf{X}) = X_{(n)}$ достаточная статистика.

Лекция 15

15.1 Достаточные и полные статистики

Роль достаточных статистик в теории оценивания раскрывает следующая теорема.

Теорема 15.1 (теорема Рао - Блекуэлла - Колмогорова). *Если оптимальная оценка существует, то она является функцией от достаточной статистики.*

Доказательство. При доказательстве нам потребуются два свойства условного математического ожидания:

$$E f(x, z) = E(E(f(x, z)|z)), \quad (1)$$

$$E(g(z)|z) = g(z). \quad (2)$$

Эти свойства были доказаны ранее, поэтому мы не будем останавливаться на их доказательстве.

Пусть теперь $T(\mathbf{X})$ — достаточная статистика, и пусть $T_1(\mathbf{X})$ — несмещенная оценка для $\tau(\theta)$, т.е.

$$ET_1(\mathbf{X}) = \tau(\theta).$$

Рассмотрим функцию $H(T) = E(T_1|T)$. Тогда из (1) следует, что

$$EH(T) = E(E(T_1|T)) = ET_1 = \tau(\theta),$$

т.е. $H(T)$ является несмещенной оценкой для $\tau(\theta)$.

Используя равенства (1) и (2) получим

$$\begin{aligned} E((T_1 - H(T))(H(T) - \tau(\theta))) &= E(E((T_1 - H(T))(H(T) - \tau(\theta))|T)) \\ &= E((H(T) - H(T))(H(T) - \tau(\theta))) = 0. \end{aligned}$$

Тогда

$$\begin{aligned} D(T_1) &= E(T_1 - \tau(\theta))^2 = E(T_1 - H(T) + H(T) - \tau(\theta))^2 \\ &= E(T_1 - H(T))^2 + D(H(T)) \geq D(H(T)). \end{aligned}$$

Таким образом $H(T)$ — оптимальная оценка. □

При отыскании явного вида оптимальных оценок важную роль играет свойство полноты достаточной статистики.

Определение 15.1. *Достаточная статистика $T(\mathbf{X})$ называется полной, если для любой функции $\varphi(z)$ из равенства $E_\theta \varphi(T) = 0, \forall \theta \in \Theta$ следует, что $\varphi(T) = 0$ с вероятностью 1.*

Замечание. В определении полноты утверждается, что

$$P_\theta(\mathbf{x} : \varphi(T(\mathbf{x}) \neq 0)) = 0,$$

где P_θ — распределение, порождаемое вектором (x_1, \dots, x_n) на \mathbb{R}^k .

Пример. Пусть $\mathbf{X} = (X_1, \dots, X_n)$ — выборка из равномерного распределения на $[0, \theta]$, $\theta > 0$. На прошлой лекции мы показали, что статистика $T(\mathbf{X}) = X_{(n)}$ является достаточной. Проверим ее полноту. Выпишем плотность распределения $X_{(n)}$:

$$h(z) = \begin{cases} n \frac{z^{n-1}}{\theta^n}, & z \in [0, \theta], \\ 0, & \text{иначе.} \end{cases}$$

Тогда

$$E\varphi(T) = \int_{\mathbb{R}} \varphi(z)h(z) dz = \frac{n}{\theta^n} \int_0^\theta \varphi(z)z^{n-1} dz = 0.$$

Из определения интеграла Лебега сразу следует, что $\varphi(z) = 0$ с вероятностью 1. Таким образом достаточная статистика $T(\mathbf{X}) = X_{(n)}$ является полной.

Теорема 15.2. Если $T(\mathbf{X})$ — полная достаточная статистика, то она является оптимальной оценкой своего математического ожидания.

Доказательство. Докажем, что $T(\mathbf{X})$ является единственной несмещенной оценкой для $ET(\mathbf{X})$. Тогда $T(\mathbf{X})$ будет оптимальной оценкой.

Предположим, что $T_1(\mathbf{X})$ — оптимальная оценка для ET . Из теоремы Рао - Блекуэлла - Колмогорова получаем, что $T_1 = H(T)$ и $ET_1 = ET$. Тогда

$$E \underbrace{(T(\mathbf{X}) - H(T(\mathbf{X})))}_{\varphi(T)} = 0,$$

и из условия полноты $T(\mathbf{X})$ следует, что $\varphi(T) = 0$ с вероятностью 1, т.е. $T = H(T)$ с вероятностью 1. □

15.2 Оценки максимального правдоподобия

Одним из наиболее универсальных методов оценивания параметров распределения является метод максимального правдоподобия. Оценку параметра, получаемую с помощью этого метода, будем обозначать $\hat{\theta} = \hat{\theta}(\mathbf{X})$.

Пусть $\mathbf{X} = (X_1, \dots, X_n)$ — выборка из распределения $\mathcal{L}(X) \in \mathcal{F} = \{F(x, \theta), \theta \in \Theta\}$ и $L_n(\mathbf{x}, \theta)$ — функция правдоподобия.

Определение 15.2. Оценкой максимального правдоподобия $\hat{\theta}$ параметра θ называется такая точка параметрического множества Θ , в которой функция правдоподобия $L_n(\mathbf{x}, \theta)$ при заданном \mathbf{x} достигает максимума, т.е.

$$L_n(\mathbf{x}, \hat{\theta}) = \sup_{\theta \in \Theta} L_n(\mathbf{x}, \theta).$$

Пример. Чтобы пояснить сказанное рассмотрим пример. Пусть $\mathbf{X} = (X_1, X_2)$ — выборка из биномиального распределения $Bi(1, \theta)$. В нашем случае функция правдоподобия будет иметь вид

$$L(\mathbf{x}, \theta) = \theta^{x_1+x_2}(1-\theta)^{2-x_1+x_2}$$

Пусть параметрическое семейство Θ состоит из двух точек: $\Theta = \left\{ \frac{1}{100}, \frac{999}{1000} \right\}$. Проводится эксперимент и наблюдается реализация $(1, 1)$ выборки \mathbf{X} . Функция правдоподобия примет вид

$$L(1, 1, \theta) = \theta^2.$$

Тогда в качестве неизвестного параметра следует, очевидно, взять точку $\left\{ \frac{999}{1000} \right\}$.

Читатель может сам без труда выяснить, какую точку надо взять из множества Θ , если наблюдаются реализации $(0, 0)$, $(1, 0)$ или $(0, 1)$ выборки \mathbf{X} .

Если для каждого \mathbf{x} максимум функции правдоподобия достигается во внутренней точке Θ , и $L_n(\mathbf{x}, \theta)$ дифференцируема по θ , то оценка максимального правдоподобия (о.м.п.) $\hat{\theta}$ удовлетворяет уравнению

$$\frac{\partial \ln L_n(\mathbf{x}, \theta)}{\partial \theta} = 0.$$

Если θ векторный параметр: $\theta = (\theta_1, \dots, \theta_n)$, то это уравнение заменяется системой уравнение

$$\frac{\partial \ln L_n(\mathbf{x}, \theta)}{\partial \theta_i} = 0, \quad i = 1, \dots, n.$$

Эти уравнения называются уравнениями правдоподобия.

Утверждение 15.1. Если существует эффективная оценка $T = T(\mathbf{X})$ скалярного параметра θ , то она совпадает с оценкой максимального правдоподобия.

Доказательство. На прошлых лекциях мы отмечали, что если оценка $T = T(\mathbf{X})$ скалярного параметра θ эффективна, то в неравенстве Рао - Крамера достигается равенство, т.е. имеет место представление

$$\frac{\partial \ln L_n(\mathbf{x}, \theta)}{\partial \theta} = c(\theta)(T(\mathbf{x}) - \theta).$$

Что и доказывает наше утверждение. □

Утверждение 15.2. Если $T = T(\mathbf{X})$ достаточная статистика, а оценка максимального правдоподобия $\hat{\theta}$ существует и единственна, то она является функцией от T .

Доказательство. Из критерия факторизации следует, что если $T = T(\mathbf{X})$ достаточная статистика, то имеет место представление

$$L_n(\mathbf{x}, \theta) = g(T(\mathbf{x}), \theta)h(\mathbf{x}).$$

Таким образом, максимизации $L_n(\mathbf{x}, \theta)$ сводится к максимизации $g(T(\mathbf{x}), \theta)$ по θ . Следовательно $\hat{\theta}$ есть функция от T . □

Пример. Пусть $\mathbf{X} = (X_1, \dots, X_n)$ — выборка из равномерного распределения на $[\theta, \theta + 1]$. Функция правдоподобия имеет вид

$$L_n(\mathbf{x}, \theta) = \mathbf{I}_{\{x_{(1)} > \theta\}} \mathbf{I}_{\{x_{(n)} < \theta + 1\}}.$$

Отсюда следует, что любое $\theta \in [x_{(n)} - 1, x_{(1)}]$ максимизирует функцию правдоподобия. Таким образом о.м.п. не единственна. В качестве решения можно выбрать, например, $\hat{\theta} = (X_{(1)} + X_{(n)} - 1)/2$.

Пример. Рассмотрим так называемую общую нормальную модель $\mathcal{N}(\theta_1, \theta_2^2)$. Максимизация функции правдоподобия в этом случае эквивалентна минимизации по θ функции

$$\psi(\mathbf{x}, \theta) = \frac{(\bar{x} - \theta_1)^2}{2\theta_2^2} + \frac{1}{2} \left(\frac{S^2}{\theta_2^2} - 1 \right) - \ln \frac{S}{\theta_2}.$$

Заметим, что функция

$$f(x) = \frac{1}{2}(x^2 - 1) - \ln x$$

достигает минимума при $x = 1$ и

$$\begin{aligned} f(1) &= 0, \\ f'(x) &< 0 \quad \text{для всех } x \in (0, 1), \\ f'(x) &> 0 \quad \text{для всех } x > 1. \end{aligned}$$

Поэтому

$$\frac{1}{2}(x^2 - 1) \geq \ln x \quad \forall x > 0.$$

Таким образом получили, что $\hat{\theta} = (\bar{X}, S)$ является о.м.п.

Утверждение 15.3 (Принцип инвариантности для оценок максимального правдоподобия). Пусть $f : \Theta \rightarrow \mathcal{F}$ — взаимнооднозначное отображение из Θ в \mathcal{F} . Тогда, если $\hat{\theta}$ есть о.м.п. для θ , то $f(\hat{\theta})$ есть оценка для $f(\theta)$.

Доказательство. Так как $\sup_{\theta \in \Theta} L_n(\mathbf{x}, \theta) = \sup_{z \in \mathcal{F}} L_n(\mathbf{x}, f^{-1}(z))$, то, если \sup левой части достигается при $\theta = \hat{\theta}$, то в правой части при некотором значении \hat{z} , которое удовлетворяет равенству $f^{-1}(\hat{z}) = \hat{\theta}$, т.е. $\hat{z} = f(\hat{\theta})$. \square

Замечание. В заключении отметим, что оценки максимального правдоподобия часто состоятельны, асимптотически несмещенные и асимптотически нормальные.

Лекция 16

16.1 Доверительные интервалы и трактовка коэффициента доверия

Пусть есть выборка X_1, \dots, X_n из $\mathcal{L}(X) \sim N(\theta, 1)$, тогда в качестве несмещенной оптимальной оценки возьмем

$$\theta^* = \bar{X}.$$

(Понятно, что \bar{x}' и \bar{x}'') — это разные вещи.

Пример. Вспомним первую лекцию. Пусть даны цены P_t в моменты времени t , тогда относительная доходность в момент времени t равна

$$X_t = \frac{P_{t+1} - P_t}{P_t}.$$

Для ряда финансовых инструментов поведение X_t близко к нормальному, то есть

$$X_1, X_2, \dots, X_n \text{ из } \mathcal{L}(X) \sim N(\theta_1, \theta_2^2);$$

Естественно, что параметры θ_1, θ_2^2 неизвестны, хотим найти для них оценки. В качестве оценки берем

$$\theta_1^* = \bar{X}.$$

Очевидно, что полученные значения в январе и в феврале не обязаны совпадать, положим, что в январе $\theta_1^* = a$, интересуемся какова вероятность того, что и в феврале получим ту же величину. Понятно, что

$$P(\theta_1^* = a) = 0 \text{ в феврале.}$$

Поскольку вероятность получить конкретное значение равна нулю, будем искать вероятность попадания в небольшой отрезок.

Обозначим γ — **уровень доверия, или коэффициент доверия, или надежность**;

$$\gamma \sim 1.$$

Определение 16.1. Пусть $Y = (X_1, \dots, X_n)$ — выборка из $\mathcal{L}(X)$. **Интервальной оценкой для θ (доверительным интервалом) с коэффициентом доверия γ называется интервал $(T_1(Y), T_2(Y))$ такой, что**

$$P(T_1(Y) < \theta < T_2(Y)) \geq \gamma.$$

$(T_1(Y), T_2(Y))$ — случайные величины.)

Рассмотрим, как это реализовать на практике.
 Пусть X_1, \dots, X_n из $\mathcal{L}(X) \sim N(\theta, 1)$.
 Тогда

$$\theta^* = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim N\left(\theta, \frac{1}{n}\right).$$

Тогда

$$(\bar{X} - \theta)\sqrt{n} \sim N(0, 1)$$

И уже для такой величины, имеющей стандартное нормальное распределение, строим интервальную оценку: то есть находим такое $t_{\gamma/2}$, что

$$P(|(\bar{X} - \theta)\sqrt{n}| < t_{\gamma/2}) = \gamma.$$

Решаем его относительно θ и получаем

$$P\left(\bar{X} - \frac{t_{\gamma/2}}{\sqrt{n}} < \theta < \bar{X} + \frac{t_{\gamma/2}}{\sqrt{n}}\right) = \gamma$$

Возникают два замечания:

Замечание. Почему был взят интервал, симметричный относительно 0? Более информативен, вообще говоря, меньший интервал. То есть наша цель — построить интервал наименьшей длины такой, что площадь под графиком плотности на этом интервале равна γ . Ясно, что в случае стандартного нормального распределения симметрия графика функции относительно оси Oy играет на нас.

Замечание. В конечном итоге получили доверительный интервал

$$\left(\bar{X} - \frac{t_{\gamma/2}}{\sqrt{n}}, \bar{X} + \frac{t_{\gamma/2}}{\sqrt{n}}\right)$$

В этом случае $t_{\gamma/2}$ находится по γ единственным образом, что, вообще говоря, происходит не всегда.

Пример. Вспомним пример с данными о ценах за январь 2006 года. Размер выборки $n = 2000$, и нами получен доверительный интервал $(0, 5\%, 1, 2\%)$ с коэффициентом доверия $\gamma = 0,99$. Как трактовать такой результат?

θ — средняя относительная доходность, это конкретное число, а не случайная величина, то есть вероятности ее нахождения никакой нет: либо параметр θ попал в построенный интервал, либо нет, но и интервал, и параметр уже найдены.

Возьмем N выборок и для каждой выборки посчитаем доверительный интервал.

N_+ — число доверительных интервалов из N построенных, которые действительно содержат θ . Тогда

$\frac{N_+}{N} \sim \gamma \text{ по построению доверительного интервала.}$

Лекция 17

17.1 Методы построения интервальных оценок

17.1.1 Метод, основанный на точечных оценках

Пусть X_1, \dots, X_n из $\mathcal{L}(X)$, $\theta \in \Theta$

$Y = (X_1, \dots, X_n)$, $T(Y)$ — точечная оценка для параметра θ с функцией распределения $G(T, \theta)$. Предположим, что $G(T, \theta)$ монотонна как функция θ .

Пример. Имеем распределение $N(\theta, 1)$ и $T(Y) = \bar{X}$

$$\begin{aligned} P(\theta^* < t) &= P((\theta^* - \theta)\sqrt{n} < (t - \theta)\sqrt{n}) \\ &= \Phi((t - \theta)\sqrt{n}), \end{aligned}$$

где $\Phi(z)$ — функция стандартного нормального распределения.

↓

$$G(t, \theta) = \Phi((t - \theta)\sqrt{n}), \quad \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-u^2/2} du.$$

Следовательно, $G(t, \theta)$ монотонна.

Пусть $\varepsilon : 1/2 < \varepsilon < 1$;

для определенности предположим, что $G(t, \theta)$ монотонна по θ .

Рассмотрим уравнения

$$\begin{cases} G(T(Y), \theta) = \varepsilon, & (1) \\ G(T(Y), \theta) = 1 - \varepsilon. & (2) \end{cases}$$

Пусть $\theta_1^*(Y)$ и $\theta_2^*(Y)$ — решения первого и второго уравнений соответственно.

Предположим, что (θ_1^*, θ_2^*) (или (θ_2^*, θ_1^*)) — интервал такой, что

$$\begin{aligned} P(\theta_1^* < \theta < \theta_2^*) &= \gamma \\ (P(\theta_1^* < \theta < \theta_2^*) &= \gamma), \end{aligned}$$

где $\gamma = 2\varepsilon - 1$.

Пример.

$$\mathcal{L}(X) \sim N(\theta, 1);$$

$$T(Y) = \bar{X} \sim N\left(\theta, \frac{1}{n}\right);$$

$$F_{\bar{X}}(t) = \Phi(\sqrt{n}(t - \theta)), \quad \text{где } \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-u^2/2} du.$$

$\Phi(\sqrt{n}(t - \theta))$ монотонно убывает по θ .

Найдем решения (1) и (2) :

$$\begin{aligned}\Phi(z) = \varepsilon; \quad \Phi(z) \text{ строго возрастает} &\Rightarrow z_\varepsilon = \Phi^{-1}(\varepsilon), \\ \Phi(z) = 1 - \varepsilon; &\Rightarrow z_{1-\varepsilon} = \Phi^{-1}(1 - \varepsilon).\end{aligned}$$

Принимая во внимание вид графика функции стандартного нормального распределения, имеем $z_\varepsilon = -z_{1-\varepsilon}$. Учтем это:

$$\Phi(z_\varepsilon) - \Phi(z_{1-\varepsilon}) = \varepsilon - (1 - \varepsilon) = 2\varepsilon - 1,$$

при этом

$$\begin{aligned}z_\varepsilon &= \sqrt{n}(T(Y) - \theta_1^*), \\ z_{1-\varepsilon} &= \sqrt{n}(T(Y) - \theta_2^*).\end{aligned}$$

Отсюда находим:

$$\begin{aligned}\theta_1^* &= T(Y) - \frac{z_\varepsilon}{\sqrt{n}}, \\ \theta_2^* &= T(Y) + \frac{z_\varepsilon}{\sqrt{n}}\end{aligned}$$

и получаем доверительный интервал

$$\left(\bar{X} - \frac{z_\varepsilon}{\sqrt{n}}, \bar{X} + \frac{z_\varepsilon}{\sqrt{n}} \right).$$

Замечание. Наличие решения зависело от монотонности функции (в случае монотонной функции легко получаем обратную), если же функция плотности ступенчатая, однозначно найти ε не удастся и в результате получим интервал удовлетворяющий неравенству $P(\theta_1^* < \theta < \theta_2^*) > \gamma$.

Замечание. Если случайная величина X имеет функцию распределения $F_X(y)$, то $F_X(X)$ — такая случайная величина, что при определенных дополнительных условиях, она равномерно распределена на $[0; 1]$.

(В частности, если случайная величина X имеет нормальное распределение, то случайная величина $F_X(X) \sim U[0; 1]$).

17.1.2 Метод, основанный на центральной статистике

Пусть $Y = (X_1, \dots, X_n)$ — выборка из распределения $\mathcal{L}(X)$.

Определение 17.1. *Центральная статистика* — это случайная величина $V(Y, \theta)$, удовлетворяющая двум условиям:

1. Распределение V не зависит от θ ;
2. $V(Y, \theta)$ монотонна как функция θ .

Если смогли найти такую функцию, то

Пусть v_1 и v_2 :

$$P(v_1 < V(Y, \theta) < v_2) = \gamma, \quad (3)$$

так как $V(Y, \theta)$ монотонна по θ (для определенности считаем, что монотонно возрастает), то получаем θ_1^* и θ_2^* — решения уравнений

$$V(Y, \theta_i^*) = v_i.$$

Тогда (3) эквивалентно

$$P(\theta_1^* < \theta < \theta_2^*) = \gamma.$$

Проиллюстрируем метод:

Рассмотрим выборку из распределения $\mathcal{L}(X) \sim N(\mu, \theta^2)$ при условии, что μ известно, а θ^2 — нет, и построим доверительный интервал для θ^2 .

Рассмотрим следующую случайную величину — кандидата на звание центральной статистики:

$$V(Y, \theta^2) = \sum_{i=1}^n \left(\frac{X_i - \mu}{\theta} \right)^2 = \frac{1}{\theta^2} \sum_{i=1}^n (X_i - \mu)^2.$$

Проверяем условия:

1. Распределение $V(Y, \theta)$, очевидно, не зависит от θ (как сумма стандартно нормально распределенных случайных величин).

2. $V(Y, \theta)$ монотонно убывает как функция θ^2 .

Следовательно, таким образом заданная функция $V(Y, \theta)$ является центральной статистикой.

Определение 17.2. Случайная величина с распределением, совпадающим с распределением

$$Z_1^2 + \dots + Z_n^2,$$

где Z_i — независимые случайные величины со стандартным нормальным распределением, называется **случайной величиной с χ^2 -распределением с n степенями свободы** с плотностью

$$p_n(y) = \begin{cases} \frac{1}{2^{n/2} \Gamma(n/2)} y^{n/2-1} e^{-y/2}, & \text{при } y > 0 \\ 0, & \text{иначе.} \end{cases}$$

Обозначение — χ_n^2 .

Замечание. Заметим, что при

$n = 1$ функция плотности не ограничена в нуле;

$n = 2$ имеем показательное распределение с параметром $\frac{1}{2}$;

$n = 3$ функция плотности на левой полуоси равна нулю, далее она возрастает, достигает максимума и экспоненциально убывает. Найдем такие v_1, v_2 , что площадь фигуры, ограниченной графиком плотности, двумя прямыми $x = v_1$ и $x = v_2$ и осью абсцисс, в точности равна γ . Естественно, что v_1, v_2 находятся не единственным образом. Для определенности поместим на "хвосты" по половине остаточной массы, т. е. по $(1 - \gamma)/2$. Для нахождения v_1 и v_2 решаем следующие уравнения

$$P(\chi_n^2 < v_1) = (1 - \gamma)/2;$$

$$P(\chi_n^2 > v_2) = (1 - \gamma)/2;$$

Из них v_1 и v_2 находятся однозначно (с помощью таблиц, пакетов).

В результате получаем:

$$P(v_1 < \chi_n^2 < v_2) = \gamma,$$

где χ_n^2 — центральная статистика, монотонно убывающая по θ^2 . Тогда предыдущее равенство эквивалентно

$$P \left(\frac{\sum_{i=1}^n (X_i - \mu)^2}{v_2} < \theta^2 < \frac{\sum_{i=1}^n (X_i - \mu)^2}{v_1} \right) = \gamma.$$

Определение 17.3. *Центральным доверительным интервалом* называется доверительный интервал, при котором на "хвостах" лежит одинаковая масса.

Замечание. Критерием того, насколько хорошо выбран доверительный интервал, является его длина. В данном случае она пропорциональна разности $\frac{1}{v_1} - \frac{1}{v_2}$ с ко-

эффициентом пропорциональности $\sum_{i=1}^n (X_i - \mu)^2$. Таким образом, перед нами встает

задача минимизации указанной выше разности при условии, что $\int_{v_1}^{v_2} p_n(y) dy = \gamma$. В явном виде решения получить, вообще говоря, нельзя из-за вида функции плотности, так что для минимизации длины используются таблицы хи-квадрат распределения.

17.1.3 Метод, основанный на центральной предельной теореме

Пусть X_1, \dots, X_n из $\mathcal{L}(X)$, $\theta \in \Theta$, и считаем, что все нужные для дальнейшего рассмотрения производные и мат. ожидания существуют. Функция правдоподобия в данном случае имеет вид:

$$p_n(y, \theta) = \prod_{i=1}^n p(x_i, \theta),$$

где p — плотность случайной величины X . Так как случайные величины X_1, \dots, X_n независимы и одинаково распределены, то, рассмотрев производную по θ натурального логарифма функции правдоподобия, имеем

$$\frac{\partial}{\partial \theta} (\ln p_n(Y, \theta)) = \sum_{i=1}^n \frac{\partial}{\partial \theta} (\ln p(X_i, \theta)) = \sum_{i=1}^n Z_i,$$

где случайные величины Z_1, \dots, Z_n — независимы и одинаково распределены.

По ЦПТ $\forall c, d, c \leq d$ имеем:

$$P \left(c \leq \frac{Z_1 + \dots + Z_n - a}{\sigma \sqrt{n}} \leq d \right) \xrightarrow{n \rightarrow \infty} \Phi(d) - \Phi(c), \quad (4)$$

где $\Phi(z)$ — функция стандартного нормального распределения, $a = EZ_1$, $\sigma^2 = DZ_1$. Ранее было доказано, что

$$E \left(\frac{\partial \ln p(X_1, \theta)}{\partial \theta} \right) = 0; \quad (\text{т. е. } a = 0)$$

$$\sigma^2 = DZ_1 = E \left(\frac{\partial \ln p}{\partial \theta} \right)^2 = -E \left(\frac{\partial^2 \ln p}{\partial \theta^2} \right).$$

Рассмотрим функцию

$$V_n(\theta) = \frac{\sum_{i=1}^n \frac{\partial \ln p(X_i, \theta)}{\partial \theta}}{\left(\mathbb{E} \left(\frac{\partial \ln p(X_i, \theta)}{\partial \theta} \right)^2 \right)^{1/2}}.$$

По ЦПТ распределение $V_n(\theta) \rightarrow N(0, 1)$, т. е. если нам нужно найти z_γ — решение следующего уравнения

$$P(|V_n(\theta)| < z_\gamma) = \gamma, \quad (5)$$

то заменяем уравнение (5) на предельное

$$\Phi(z_j) - \Phi(-z_j) = \gamma, \quad (6)$$

затем по таблицам или из пакетов программ находим из (6) z_γ по заданному γ . Предположим, что неравенство

$$|V_n(\theta)| < z_\gamma \quad (7)$$

разрешается относительно θ в виде интервала $\theta_1^* < \theta < \theta_2^*$, тогда (θ_1^*, θ_2^*) — доверительный интервал с надежностью γ . Стоит заметить, что неравенство (7), вообще говоря, не является однозначно разрешимым или разрешимым относительно θ .

Пример. Рассмотрим выборку из Пуассоновского распределения:

$$\mathcal{L}(X) \sim Pois(\theta), \theta > 0, P(X = k) = \frac{\theta^k}{k!} e^{-\theta}.$$

В данном случае

$$p_n(y, \theta) = \theta^{\sum_{i=1}^n x_i} e^{-n\theta} \prod_{i=1}^n \frac{1}{x_i!} = \theta^{n\bar{X}} e^{-n\theta} \prod_{i=1}^n \frac{1}{x_i!}$$

\Rightarrow

$$\frac{\partial \ln p_n}{\partial \theta} = n\bar{X} \frac{1}{\theta} - n = \frac{n}{\theta} (\bar{X} - \theta)$$

тогда

- В силу неравенства Рао-Крамера \bar{X} — **эфффективная оценка для θ** .
- Для нахождения оценок максимального правдоподобия нужно решить уравнение правдоподобия:

$$\frac{\partial}{\partial \theta} \ln p_n(Y, \theta) = 0, \text{ откуда}$$

$$\theta^* = \bar{X},$$

т. е. \bar{X} — оценка максимального правдоподобия.

Имеем

$$\begin{aligned} DZ_1 &= E \left(\frac{\partial^2 \ln p_n}{\partial \theta^2} \right) = E \left(\frac{n\bar{X}}{\theta^2} \right) = nE \left(\frac{\bar{X}}{\theta^2} \right) = \frac{n}{\theta} \\ &\Rightarrow \\ V_n(Y, \theta) &= \sqrt{\frac{n}{\theta}} (\bar{X} - \theta), \end{aligned}$$

Решаем неравенство (7) в виде интервала относительно θ и, введя обозначение

$$B(\gamma, n) = z_\gamma \sqrt{\frac{\bar{X}}{n} + \frac{z_\gamma^2}{4n^2}},$$

получаем, что доверительный интервал имеет вид

$$\bar{X} + \frac{z_\gamma}{2n} - B(\gamma, n) < \theta < \bar{X} + \frac{z_\gamma}{2n} + B(\gamma, n).$$

17.2 Проверка статистических гипотез

17.2.1 Общая постановка проверки статистических гипотез

Пусть X_1, \dots, X_n из $\mathcal{L}(X) \in \mathcal{P}$,

т. е. есть некоторая статистическая структура $(\mathcal{X}, \mathcal{A}, \mathcal{P})$, где

\mathcal{X} — совокупность всех возможных значений выборки,

\mathcal{A} — σ -алгебра,

\mathcal{P} — семейство вероятностных распределений.

Основная задача — максимально сузить \mathcal{P} , т. е. найти распределение, наиболее соответствующее выборке.

Определение 17.4. *Статистической гипотезой* называется любое предположение о распределении случайной величины X .

Определение 17.5. *Статистическая гипотеза называется простой*, если предполагаемое семейство вероятностных распределений состоит в точности из одного распределения.

Пример. Предположим, что имеется набор значений цен P_t в момент времени t , тогда $X_t = \frac{P_{t+1} - P_t}{P_t}$ — выборка относительных доходностей. Допустим, что известно, что $X_t \sim N(\theta_1, \theta_2^2)$, тогда как только мы фиксируем оба параметра, то получаем простую гипотезу, например, $X_t \sim N(0.1, 0.001)$.

Пусть нулевое предположение $H_0 : \mathcal{L}(X) \in \mathcal{P}_0$, альтернатива $H_1 : \mathcal{L}(X) \in \mathcal{P}_1$, где $\mathcal{P}_0, \mathcal{P}_1$ могут состоять как из одного, так и из больше числа распределений, и \mathcal{P}_0 и \mathcal{P}_1 не пересекаются.

17.2.2 Типы гипотез

- Гипотезы о виде распределения (например, $X \sim N(0, 0.001)$)
- Гипотезы о проверке однородности выборки

Пример. Предположим, что имеется несколько выборок:

$$x_{11}, \dots, x_{1n_1} \sim \mathcal{L}(X),$$

$$x_{21}, \dots, x_{2n_2} \sim \mathcal{L}(Y)$$

тогда $H_0 : \mathcal{L}(X) = \mathcal{L}(Y)$ (гипотеза об однородности)

$$H_1 : \mathcal{L}(X) \neq \mathcal{L}(Y).$$

В качестве содержательного примера можно рассмотреть следующий: имеются два метода лечения, старый и новый. Есть два набора результатов, количество которых n_1 и n_2 для старого и нового методов соответственно. Тогда в качестве H_0 можно взять гипотезу о том, что новый метод не лучше старого.

• **Гипотезы о независимости признаков**

Пусть имеется выборка (набор двумерных векторов), полученная из случайного вектора (X, Y)

$$(x_1, y_1), \dots, (x_n, y_n) \sim \mathcal{L}(X, Y).$$

$H_0 : X, Y$ независимы,

$H_1 : X, Y$ зависимы.

Определение 17.6. *Статистическим критерием (критерием) называется правило, согласно которому гипотеза H_0 принимается или отвергается.*

Пример. Допустим, что семейство вероятностных распределений состоит из двух распределений $\mathcal{P} : N(0, 1), N(2, 1)$,

$$H_0 : \mathcal{L}(X) \sim N(0, 1),$$

$$H_1 : \mathcal{L}(X) \sim N(2, 1),$$

и имеется всего лишь одно наблюдение x_1 . Ясно, что графики функций плотностей, соответствующих каждому из распределений пересекаются в точке $x = 1$ и при $x < 1$ график функции плотности, соответствующей распределению $N(0, 1)$, лежит выше. Аналогично при $x > 1$ выше лежит график функции плотности, соответствующей $N(2, 1)$. В данном случае вполне логично рассмотреть следующий критерий:

Если $x_1 < 1$, то принимается гипотеза H_0 .

Естественно, что при проверке гипотез возможны ошибки.

Определение 17.7. *Ошибка первого рода — отвергается гипотеза H_0 , при условии, что она верна.*

Обозначение — $H_1|H_0$.

Определение 17.8. *Ошибка второго рода — считаем, что гипотеза H_0 не противоречит наблюдаемым данным, при условии, что она не верна.*

Обозначение — $H_0|H_1$.

Определение 17.9. $\alpha = P(H_1|H_0)$ — *уровень значимости критерия.*

Определение 17.10. $1 - P(H_0|H_1) = P(H_1|H_1)$ — *мощность критерия.*

Пример. Рассмотрим предыдущий пример, в котором критерием выбора гипотезы было сравнение значения x_1 с $x_{\text{крит.}} = 1$. Тогда при $x_{\text{крит.}} = 1$ получим

$$\alpha = P(X_1 \geq x_{\text{крит.}} | H_0) = 1 - \Phi(1) \approx 0.17,$$

$$\beta = P(H_0 | H_1) = P(X_1 < x_{\text{крит.}} | H_1) = \Phi(-1) = 1 - \Phi(1) \approx 0.17.$$

Ясно, что с ростом $x_{\text{крит.}}$ α также растет, а β убывает; при уменьшении же $x_{\text{крит.}}$ α уменьшается, зато наблюдается рост β . То есть не возможно одновременно сделать

ошибки обоих родов сколь угодно малыми (при фиксированном объеме выборки). Обычно для фиксированного n (объема выборки) фиксируют α и среди всех возможных критериев с заданным уровнем значимости α выбирают наиболее мощный.

Предположим, что есть выборка X_1, \dots, X_n из $\mathcal{L}(X)$, с параметром $\theta \in \Theta$.
 $H_0 : \theta \in \Theta_0, H_1 : \theta \in \Theta_1. \Theta_0 \cap \Theta_1 = \emptyset$.

Определение 17.11. Пусть \mathcal{X} — совокупность всех возможных значений (X_1, \dots, X_n) ,
 $S \subset \mathcal{X}$ — некоторая область, называемая критической областью.
 Тогда критерий, при котором при условии $(x_1, \dots, x_n) \in S$ гипотеза H_0 отвергается, называется **S-критерием**.

В рассмотренном ранее примере $S = [1; +\infty)$.

Определение 17.12. Пусть $p_n(y, \theta)$ — функция правдоподобия, тогда функция

$$W(S, \theta) = \int_S p_n(y, \theta) dy$$

называется **функцией мощности**.

Содержательный смысл функции мощности:
 Если $\Theta_0 = \theta_0, \Theta_1 = \theta_1$, т. е. H_0, H_1 — простые, тогда

$$W(S, \theta_0) = \int_S p_n(y, \theta_0) dy = P(Y \in S | H_0) = \alpha,$$

$$W(S, \theta_1) = \int_S p_n(y, \theta_1) dy = P(Y \in S | H_1) \text{ — мощность критерия.}$$

Определение 17.13. S-критерий S^* называется **оптимальным критерием с заданным уровнем значимости α** , если

$$\begin{aligned} W(S^*, \theta_0) &= \alpha, \\ W(S^*, \theta_1) &\geq W(S, \theta_1), \end{aligned} \quad (8)$$

для любого S-критерия: $W(S, \theta_0) = \alpha$.

Если мощность $\#\Theta_1 > 1$, то (8) выполняется для всех $\theta_1 \in \Theta_1$, но в дальнейшем будут рассматриваться простые гипотезы.

Определение 17.14. Пусть \mathcal{X} — совокупность всех возможных значений (X_1, \dots, X_n) ,
 и введена функция $\varphi: \mathcal{X} \rightarrow [0; 1]$, тогда, отвергая H_0 с вероятностью $\varphi(x_1, \dots, x_n)$,
 получаем **рандомизированный, или φ -критерий**.

Оптимальный критерий с заданным уровнем значимости α существует не всегда.

Если $\varphi(x_1, \dots, x_n) = \mathbf{1}_S(x_1, \dots, x_n)$, то получаем S-критерий.

Определение 17.15. Функция

$$W(\varphi, \theta) = \int_{\mathbb{R}^n} \varphi(y) p_n(y, \theta) dy, \quad y = (x_1, \dots, x_n)$$

называется **функцией плотности для рандомизированного критерия**.

Аналогично

$W(\varphi, \theta_0)$ – уровень значимости,

$W(\varphi, \theta_1)$ – мощность критерия.

Определение 17.16. φ^* – *оптимальный рандомизированный критерий с уровнем значимости α , если*

$$W(\varphi^*, \theta_0) = \alpha,$$

$$W(\varphi^*, \theta_1) \geq W(\varphi, \theta_1),$$

для любого φ : $W(\varphi, \theta_0) = \alpha$.

Оптимальный рандомизированный критерий существует всегда.

Лекция 18

18.1 Лемма Неймана-Пирсона

Пусть X_1, \dots, X_n — повторная выборка из $\mathcal{L}(X)$, $\theta \in \Theta$;
и относительно значений параметра имеем две простые гипотезы:

$$H_0 : \theta = \theta_0;$$

$$H_1 : \theta = \theta_1;$$

$p_n(y, \theta)$ — функция правдоподобия,
 $p_1(y) = p_n(y, \theta_1), p_0(y) = p_n(y, \theta_0)$;

Тогда отношением правдоподобия называется величина $\frac{p_1(y)}{p_0(y)}$.

Определение 18.1. *Критерии отношения правдоподобия* — это критерии, основанные на отношении $\frac{\mathbf{p}_1(\mathbf{y})}{\mathbf{p}_0(\mathbf{y})}$.

Лемма 18.1. Для $\forall \alpha \in [0; 1] \exists$ постоянные $c > 0$ и $\varepsilon \in [0; 1]$ такие, что φ — критерий

$$\varphi^*(y) = \begin{cases} 1, & p_1(y) > cp_0(y), \\ \varepsilon, & p_1(y) = cp_0(y), \\ 0, & p_1(y) < cp_0(y) \end{cases}$$

является *оптимальным критерием*, т. е. наиболее мощным среди всех φ — критериев с уровнем значимости α .

Замечание. Если рассмотреть пример из прошлой лекции, где X_1 из $\mathcal{L}(X) \sim N(\theta, 1)$, $H_0 : \theta = 0$, $H_1 : \theta = 2$, то критерий отношения правдоподобия в данном случае естественен, т. к. берется та гипотеза, при которой график функции плотности располагается выше. Стоит заметить, что, вообще говоря, нельзя положить c равным 1, так как при выборе c нужно отталкиваться от заданного уровня значимости α .

Замечание. φ^* — критерий воспринимаем следующим образом:

$$\varphi(y) = 1 \Rightarrow H_0 \text{ отвергаем,}$$

$$\varphi(y) = 0 \Rightarrow H_0 \text{ принимаем,}$$

$$\varphi(y) = \varepsilon \Rightarrow H_0 \text{ отвергается с вероятностью } \varepsilon.$$

Рассмотрим крайние случаи для α :

$\alpha = 0 = P(H_1 H_0)$ (т. е. мы никогда не допускаем ошибку 1-го рода), тогда

$$\varphi^*(y) = \begin{cases} 1, & p_1(y) = 0, \text{ (но сюда мы никогда не попадем)} \\ 0, & p_0(y) > 0, \text{ (здесь гипотеза } H_0 \text{ не отрицается)} \end{cases}$$

$\alpha = 1$ (всегда имеет место ошибка 1-го рода), H_0 всегда отвергается, т. е. $\varphi^*(y) = 1$.

Доказательство. Доказательство будет состоять из двух частей: в 1-ой мы конструктивно укажем, как по заданному α найти c и ε , во 2-ой части будет доказана оптимальность и то, что данный критерий обладает заданным уровнем значимости.

1. Определим функцию $g(c) = P(p_1(Y) \geq cp_0(Y) \mid H_0)$ (т. е. случайный вектор Y имеет распределение, соответствующее нулевой гипотезе). Тогда

$$\begin{aligned} 1 - g(c) &= P(p_1(Y) < cp_0(Y) \mid H_0) = P(p_1(Y) < cp_0(Y) \mathbf{1}_{\{p_0(Y) > 0\}} \mid H_0) \\ &= P\left(\frac{p_1(Y)}{p_0(Y) \mathbf{1}_{\{p_0(Y) > 0\}}} < c \mid H_0\right) \\ &= \text{Ф-ия распределения сл. в. } z \text{ как Ф-ия от } c. \end{aligned}$$

Тогда можно выделить следующие свойства функции $g(c)$:

1. $g(c)$ не возрастает;
2. $g(c)$ непрерывна слева;
3. $g(0) = 1; g(+\infty) = 0$;

Далее по заданному α определим c_α следующим образом:

$$g(c_\alpha + 0) = \lim_{c \rightarrow c_\alpha + 0} g(c) < \alpha \leq g(c_\alpha),$$

если c_α — точка разрыва функции $g(c)$, в противном случае

$$c_\alpha : \alpha = g(c_\alpha).$$

Для функции $g(c)$ рассмотрим три возможных случая:

1. Функция $g(c)$ терпит разрыв так, что ее график не пересекается с графиком функции $y = \alpha$. Т. е. в некоторой точке c_0 $g(c)$ разрывна и $g(c_0) > \alpha, g(c_0 - 0) < \alpha$. Тогда $c_\alpha = c_0$.
2. Функция $g(c)$ непрерывна и значение, равное α , достигается в единственной точке c_0 . Тогда, аналогично, $c_\alpha = c_0$.
3. Функция $g(c)$ непрерывна и значение, равное α , достигается на отрезке $[c_1; c_2]$. Тогда в качестве c_α берем произвольную точку из отрезка $[c_1; c_2]$.

Таким образом определенное c_α и есть искомое c из определения функции $\varphi^*(y)$. ε же определяется следующим образом:

во 2-ом и 3-ьем случаях (когда решение существует) $\varepsilon = 0$,

в 1-ом случае положим $\varepsilon = \frac{\alpha - g(c_\alpha + 0)}{g(c_\alpha) - g(c_\alpha + 0)}$.

2. Уровень значимости

$$\begin{aligned}
 W(\varphi, \theta_0) &= \int_{\mathbb{R}^n} \varphi^*(y) p_0(y) dy = \int_{p_1(y) > c_\alpha p_0(y)} p_0(y) dy + \varepsilon \cdot \int_{p_1(y) = c_\alpha p_0(y)} p_0(y) dy \\
 &= g(c_\alpha) + (\varepsilon - 1) \cdot \int_{p_1(y) = c_\alpha p_0(y)} p_0(y) dy \\
 &= g(c_\alpha) + (\varepsilon - 1)(g(c_\alpha) - g(c_\alpha + 0)) \\
 &= \begin{cases} g(c_\alpha) + \alpha - g(c_\alpha) & \text{в 1-ом случае} \\ g(c_\alpha) & \text{во 2-ом и 3-ьем случаях} \end{cases} \\
 &= \alpha \text{ для всех трех случаев.}
 \end{aligned}$$

Таким образом показано, что φ^* – критерий имеет уровень значимости α .

Предположим, что произвольный φ – критерий имеет уровень значимости α , и докажем, что

$$W(\varphi^*, \theta_1) \geq W(\varphi, \theta_1) \quad (1)$$

Рассмотрим интеграл

$$\begin{aligned}
 I &= \int_{\mathbb{R}^n} (\varphi^*(y) - \varphi(y))(p_1(y) - c_\alpha p_0(y)) dy \\
 &= \int_{\varphi^* > \varphi} (\varphi^*(y) - \varphi(y))(p_1(y) - c_\alpha p_0(y)) dy + \int_{\varphi^* < \varphi} (\varphi^*(y) - \varphi(y))(p_1(y) - c_\alpha p_0(y)) dy = \\
 &= I_1 + I_2;
 \end{aligned}$$

Покажем, что I_1 и I_2 неотрицательны.

- 1) на множестве тех y , где $\varphi^*(y) > \varphi(y) \geq 0$, выполнено также и $\varphi^*(y) > 0$, и, следовательно, $p_1(y) > c_\alpha p_0(y) \Rightarrow I_1 \geq 0$
- 2) аналогично на множестве тех y , где $\varphi^*(y) < \varphi(y) \leq 1$, выполнено также и $\varphi^*(y) < 1$, и, следовательно, $p_1(y) \leq c_\alpha p_0(y) \Rightarrow I_2 \geq 0$

Отсюда

$$\begin{aligned}
 0 \leq I &= W(\varphi^*, \theta_1) - W(\varphi, \theta_1) - c_\alpha(W(\varphi^*, \theta_0) - W(\varphi, \theta_0)) \\
 &= \{W(\varphi^*, \theta_0) = W(\varphi, \theta_0) = \alpha\}.
 \end{aligned}$$

$\Rightarrow W(\varphi^*, \theta_1) \geq W(\varphi, \theta_1)$, и (1) доказано. □

18.2 Равномерно наиболее мощные критерии

Пример. Пусть X_1, \dots, X_n из $\mathcal{L}(X) \sim N(\theta, 1)$, $H_0 : \theta = 0$, $H_1 : \theta = \theta_1 > 0$. Тогда

$$\frac{p_1(y)}{p_0(y)} = \exp \left[-\frac{1}{2} \left(-\sum_{i=1}^n 2x_i\theta_1 + n\theta_1^2 \right) \right] = \exp \left[\sum_{i=1}^n x_i\theta_1 - \frac{n}{2}\theta_1^2 \right] > c$$

$$\begin{aligned} & \Updownarrow \\ & \sum_{i=1}^n x_i > c_1 \\ & \Updownarrow \\ & \bar{X} > c_2 \end{aligned}$$

Пусть α задано. Тогда $P(\bar{X} > c_2 \mid H_0) = \alpha$.

Из H_0 следует, что $\bar{X} \sim N(0, \frac{1}{n})$, поэтому $\sqrt{n}\bar{X} \sim N(0, 1)$ и, если u_α определяется из равенства

$$1 - \Phi(u_\alpha) = \alpha, \quad (2)$$

то

$$c_\alpha = \frac{u_\alpha}{\sqrt{n}}.$$

Следовательно, оптимальным критерием является следующее утверждение:

"Гипотеза H_0 отвергается, если $\bar{X} > \frac{u_\alpha}{\sqrt{n}}$."

Замечание. Так как условие $\theta_1 > 0$ не содержится в описании критерия, то критерий остается неизменным для всех $\theta_1 > 0$, т. е. этот критерий является **равномерно наиболее мощным критерием для альтернативы $H_1 : \theta = \theta_1 > 0$.**

Замечание. Если $\theta_1 < 0$, то $\bar{X} < -\frac{u_\alpha}{\sqrt{n}}$, где u_α из (2), и данный критерий также является **равномерно наиболее мощным.**

Замечание. Рассмотрим мощность построенного критерия в случае $\theta_1 > 0$:

$$\begin{aligned} W(\varphi^*, \theta_1) &= P(\bar{X} > \frac{u_\alpha}{\sqrt{n}} \mid H_1) = \{ \bar{X} \sim N(\theta_1, \frac{1}{n}) \} \\ &= 1 - \Phi(u_\alpha - \theta_1\sqrt{n}). \end{aligned}$$

Замечание. Если $\theta_1 \sim 0$, то $W(\varphi^*, \theta_1) \sim \alpha$, т. е. мощность мала. Это можно объяснить тем, что в случае, когда альтернатива H_1 близка к H_0 , вероятность ошибки второго рода довольно велика.

Замечание. Если $n \rightarrow \infty$, то $W(\varphi^*, \theta_1) \rightarrow 1$, т. е. вероятность ошибки второго рода стремится к нулю, и ее можно сделать сколько угодно малой, меняя объем выборки.

Лекция 19

19.1 Состоятельные критерии

Рассмотрим пример из предыдущей лекции. Пусть X_1, \dots, X_n из $\mathcal{L}(X) \sim N(\theta, 1)$, $H_0: \theta = \theta_0 = 0$, $H_1: \theta = \theta_1 > 0$. Используя лемму Неймана-Пирсона, построим н. м. к. (наиболее мощный критерий):

$$\bar{X} > \frac{u_\alpha}{\sqrt{n}},$$

где

$$1 - \Phi(u_\alpha) = \alpha \text{ — уровень значимости.}$$

(Этот критерий является равномерно наиболее мощным.) Несложно видеть, что при фиксированном значении θ_1 и при $n \rightarrow \infty$ $W \rightarrow 1$.

Определение 19.1. Критерий называется *состоятельным*, если его мощность стремится к 1 при неограниченном увеличении объема выборки.

Замечание. Таким образом, построенный выше критерий является состоятельным.

19.2 Критерий χ^2 Пирсона

19.2.1 Критерий χ^2 Пирсона

Пусть X_1, \dots, X_n из $\mathcal{L}(X)$, где $\mathcal{L}(X)$ — дискретное распределение, при котором случайная величина принимает значения

$$a_1, \dots, a_k$$

с вероятностями

$$p_1, \dots, p_k.$$

И рассматриваются две гипотезы:

$$H_0: p_i = p_{i_0} \quad \forall i = \overline{1, k},$$

$$H_1: \sum_{i=1}^k (p_i - p_{i_0})^2 > 0 \quad , \text{ т. е. хотя бы для одного } i \text{ } p_i \neq p_{i_0}.$$

Тогда по критерию χ^2 Пирсона предлагается рассмотреть следующую статистику: пусть x_1, \dots, x_n — реализация выборки, а ν_1, \dots, ν_k — набор чисел, составленный по следующему закону:

ν_i равно количеству встретившихся в данной выборке значений a_i . Если H_0 справедлива, то $\nu_i \sim Bi(n, p_{i_0})$. Тогда

$$\begin{aligned} E\nu_i &= np_{i_0}, \\ D\nu_i &= np_{i_0}(1 - p_{i_0}), \\ \bar{\chi}^2 &= \sum_{i=1}^k \frac{(\nu_i - np_{i_0})^2}{np_{i_0}}. \end{aligned}$$

То есть получена статистика $\bar{\chi}^2$, или $\bar{\chi}^2$ – критерий.

19.2.2 Асимптотика для критерия Пирсона

Пример. Рассмотрим следующие данные: в 2006 году в России насчитывалось 53 миллиардера, из которых 12% родились под знаком Девы. Предположим, что исследуемая случайная величина принимает значения 1 (миллиардер родился под знаком Девы) и 0 (под другим знаком), т. е. $a_1 = 1, a_2 = 0$. Тогда гипотезы H_0 и H_1 можно сформулировать следующим образом:

$$H_0 : p_{10} = \frac{1}{12} \text{ (вероятность родиться под знаком Девы равняется } \frac{1}{12} \text{)}.$$

$$H_1 : p_{11} \neq \frac{1}{12} \text{ (вероятность родиться под знаком Девы не равна } \frac{1}{12} \text{, в данном случае больше, т. к. } \frac{12}{100} > \frac{1}{12} \text{)}.$$

Предположим, что объем выборки $n = 100$ и $\nu_1 = 12$, тогда, используя критерий χ^2 Пирсона, получаем, что $\bar{\chi}^2 = 1.76$. Если же справедлива H_0 , то $\bar{\chi}^2 \sim 0$. Тогда критическая область S имеет вид

$$S = \{ \bar{\chi}^2 : \bar{\chi}^2 > \chi_{\text{крит.}}^2 \},$$

и

$$\alpha = P(\bar{\chi}^2 > \chi_{\text{крит.}}^2 | H_0),$$

причем найти из этого уравнения $\chi_{\text{крит.}}^2$ по заданному α довольно непросто. Используем асимптотический подход:

$$\begin{aligned} \bar{\chi}^2 &= \frac{(\nu_1 - np_{10})^2}{np_{10}} + \frac{(\nu_2 - np_{20})^2}{np_{20}} = \{ \nu_1 + \nu_2 = n, p_{10} + p_{20} = 1 \} = \\ &= \frac{(\nu_1 - np_{10})^2}{np_{10}} + \frac{(n - \nu_1 - n(1 - p_{10}))^2}{np_{20}} = \frac{(\nu_1 - np_{10})^2}{n} \left(\frac{1}{p_{10}} - \frac{1}{1 - p_{10}} \right) = \\ &= \frac{(\nu_1 - np_{10})^2}{np_{10}(1 - p_{10})} = \left(\frac{\nu_1 - np_{10}}{\sqrt{np_{10}(1 - p_{10})}} \right)^2 = \left(\frac{Y_1 + \dots + Y_n - np_{10}}{\sqrt{np_{10}(1 - p_{10})}} \right)^2, \end{aligned}$$

где

$$Y_i = \begin{cases} 1, & \text{если } x_i = a_i, \\ 0, & \text{иначе} \end{cases}.$$

Применяем ЦПТ : выражение, стоящее внутри скобок сходится по распределению к стандартному нормальному, т. е. все выражение сходится по распределению к распределению χ^2 с одной степенью свободы. Таким образом, доказана

Теорема 19.1.

$$\forall \mathbf{z} \in \mathbb{R} \quad P(\bar{\chi}^2 < \mathbf{z}) \rightarrow P(\chi_1^2 < \mathbf{z}).$$

Теорема 19.2.

$$\forall \mathbf{z} \in \mathbb{R} \quad \mathbb{P}(\bar{\chi}^2 < \mathbf{z}) \rightarrow \mathbb{P}(\chi_{k-1}^2 < \mathbf{z}),$$

где k — число значений в дискретном распределении $\mathcal{L}(X)$,
 χ_{k-1}^2 — χ^2 — случайная величина с $(k-1)$ степенями свободы.

Пример. В рассмотренном выше примере о миллиардерах

$$\alpha = \mathbb{P}(\bar{\chi}^2 > \chi_{\text{крит.}}^2 \mid H_0) \sim \mathbb{P}(\chi_1^2 > \chi_{\text{крит.}}^2 \mid H_0),$$

и при $\alpha = 0.1$ получаем $\chi_{\text{крит.}}^2 = 2.71$, т.е. гипотеза H_0 не противоречит данным.

19.3 Состоятельность критерия Пирсона

Теорема 19.3. Критерий χ^2 является *состоятельным*, т. е.

$$\begin{aligned} \mathbb{P}(\bar{\chi}^2 = \sum_{i=1}^k \frac{(\nu_i - np_{i_0})^2}{np_{i_0}} > \chi_{\text{крит.}}^2 \mid H_1) &\rightarrow 1 \text{ при росте } n, \text{ или} \\ \mathbb{P}(\bar{\chi}^2 = \sum_{i=1}^k \frac{(\nu_i - np_{i_0})^2}{np_{i_0}} < \chi_{\text{крит.}}^2 \mid H_1) &\rightarrow 0 \text{ при росте } n. \end{aligned}$$

Доказательство.

$$\begin{aligned} \bar{\chi}^2 &= \sum_{i=1}^k \frac{(\nu_i - np_{i_0})^2}{np_{i_0}} = \sum_{i=1}^k \frac{(\nu_i - np_{i_1} + np_{i_1} - np_{i_0})^2}{np_{i_0}} = \\ &= \{ \text{при условии } H_1 : \nu_i \sim Bi(n, p_{i_1}) \Rightarrow \mathbb{E}\nu_i = np_{i_1} \} = \\ &= \sum_{i=1}^k \frac{(\nu_i - np_{i_1})^2}{np_{i_0}} + 2 \sum_{i=1}^k \frac{(\nu_i - np_{i_1})(np_{i_1} - np_{i_0})}{np_{i_0}} + n \sum_{i=1}^k \frac{(np_{i_1} - np_{i_0})^2}{np_{i_0}} = \\ &= Y_1 + 2Y_2 + nC_1(k), \text{ причем } Y_1 > 0, C_1(k) > 0. \end{aligned}$$

Тогда в терминах введенных случайных величин Y_1, Y_2 надо доказать, что

$$\mathbb{P}(Y_1 + 2Y_2 + nC_1(k) < \chi_{\text{крит.}}^2 \mid H_1) \rightarrow 0.$$

Обозначим $-nC_2(k) = \chi_{\text{крит.}}^2 - nC_1(k)$, причем $C_2(k) > 0$.

И, т. к. $Y_1 > 0$, достаточно показать, что

$$\mathbb{P}(Y_2 < -\frac{nC_2(k)}{2} \mid H_1) \rightarrow 0, \text{ или}$$

$$\mathbb{P}(Y_2 < -nC_3(k) \mid H_1) \rightarrow 0.$$

$$\begin{aligned} \mathbb{P}(Y_2 < -nC_3(k) \mid H_1) &\leq \mathbb{P}(|Y_2| > nC_3(k) \mid H_1) \leq \\ &\leq \{ \text{по н-ву Чебышева} \} \leq \frac{\mathbb{D}Y_2}{n^2C_3^2(k)}. \end{aligned}$$

Для того, чтобы доказать стремление к нулю, достаточно показать, что $\mathbb{D}Y_2 \leq nC_4(k)$, а этот факт будет следовать из того, что $\mathbb{E}Y_2^2 \leq nC_4(k)$. Покажем последнее

неравенство:

$$\begin{aligned}
 EY_2^2 &= E \left(\sum_{i=1}^k \frac{(\nu_i - np_{i_1})(p_{i_1} - p_{i_0})}{p_{i_0}} \right)^2 = \\
 &= E \left(\sum_{i=1}^k (\nu_i - np_{i_1}) \frac{(p_{i_1} - p_{i_0})}{p_{i_0}} \right)^2 \leq \\
 &\leq \{ \text{по н-ву Коши-Буняковского} \} \leq \\
 &\leq C_5(k) E \sum_{i=1}^k (\nu_i - np_{i_1})^2 = \\
 &= nC_5(k) \sum_{i=1}^k p_{i_1}(1 - p_{i_1}) = nC_4(k).
 \end{aligned}$$

□

19.4 Обобщение классического критерия χ^2

Рассмотрим $\mathcal{L}(X)$ — дискретное распределение, при котором случайная величина принимает значения

$$a_1, \dots, a_k$$

с вероятностями

$$p_1, \dots, p_k.$$

И нулевая гипотеза имеет вид:

$$H_0 : p_i = p_{i_0}(\theta) \quad \forall i = \overline{1, k}, \text{ где } \theta = (\theta_1, \dots, \theta_r).$$

Тогда следует рассматривать

$$\bar{\chi}^2 = \sum_{i=1}^k \frac{(\nu_i - np_{i_0}(\hat{\theta}))^2}{np_{i_0}(\hat{\theta})}, \text{ где } \hat{\theta} \text{ — оценка для } \theta. \quad (1)$$

Теорема 19.4. Если функция $p_n(X, \theta)$ достаточно гладкая, и $\hat{\theta}$ — оценка максимального правдоподобия для θ , то статистика, определяемая (1), сходится по распределению к χ_{k-1-r}^2 .

Лекция 20

20.1 Критерий согласия Пирсона χ^2 для абсолютно непрерывных распределений

Пусть $\mathcal{L}(X)$ — абсолютно непрерывное распределение. Разобьем вещественную прямую на интервалы следующим образом:

$$\mathbb{R} = (-\infty; z_1] \cup \left(\bigcup_{i=1}^{k-2} (z_i; z_{i+1}] \right) \cup [z_{k-1}; +\infty).$$

В качестве ν_i берем число x_i , попавших в $C_i = (z_{i-1}; z_i]$, считая $z_0 = -\infty$, $z_k = +\infty$. Сформулируем гипотезу H_0 для функции плотности $f(y)$

$$H_0: f(y) = f_0(y).$$

Положим

$$p_i = P(X \in C_i) = \int_{C_i} f_0(y) dy = p_{i_0}.$$

Далее можем, как и ранее, считать

$$\bar{\chi}^2 = \sum_{i=1}^k \frac{(\nu_i - np_{i_0})^2}{np_{i_0}} \rightarrow \chi_{k-1}^2.$$

Таким образом проведена дискретизация непрерывной задачи.

Тем не менее, остаются нерешенными два вопроса:

1. выбор z_i ,
2. количество точек разбиения z_i .

Пример. Рассмотрим случай $k = 2$. Для любого симметричного распределения получаем $p_{1_0} = p_{2_0} = \frac{1}{2}$, и, как следствие, при разбиении вещественной прямой на две полупрямые мы уже не можем различить два неравных друг другу распределения с функциями плотности, симметричными относительно оси абсцисс. Т. е. необходимо, чтобы дискретная модель максимально отражала непрерывную.

Наиболее правдоподобное приближение непрерывной модели дискретной получается при достаточно большом (k) и при более мелком разбиении прямой для тех участков, где функция плотности $f_0(y)$ принимает большие значения. С другой стороны, нельзя неограниченно увеличивать (k), т. к. при фиксированном объеме выборки (n) мы можем получить ситуацию, при которой многие ν_i будут равняться нулю, что не прибавит точности.

20.2 Критерий независимости χ^2

Пример. Допустим, что проведен опрос (n) выпускников некоторого учебного заведения и имеется следующая выборка — X_1, \dots, X_n , где $X_i = (Z_{1i}, Z_{2i})$,

$$Z_{1i} = \begin{cases} 1, & \text{если } i\text{-тый выпускник имеет красный диплом} \\ 0, & \text{иначе} \end{cases},$$

$$Z_{2i} = \begin{cases} 2, & \text{если } i\text{-тый выпускник доволен работой} \\ 1, & \text{если } i\text{-тый выпускник нейтрально относится к работе} \\ 0, & \text{если } i\text{-тый выпускник не доволен работой} \end{cases}.$$

Т. е. имеются две случайные величины: Z_1 , отражающая распределение красных дипломов, и Z_2 , показывающая отношение выпускников к текущей работе. Тогда можно сформулировать гипотезу

$$H_0: Z_1 \text{ и } Z_2 \text{ независимы,}$$

и по полученной выборке двумерных векторов попытаться определить, насколько она соответствует полученным данным.

Рассмотрим следующую абстрактную модель: имеется выборка из (n) двумерных векторов X_1, \dots, X_n , где $X_i = (Z_{1i}, Z_{2i})$, ν_{ij} — число векторов из данной выборки, у которых первая компонента равна b_i , а вторая — a_j , $i = \overline{1, l}$, $j = \overline{1, k}$. По гипотезе H_0 случайные величины Z_1 и Z_2 независимы, т. е.

$$p_{ij} = P(Z_1 = b_i; Z_2 = a_j) = P(Z_1 = b_i) \cdot P(Z_2 = a_j) = p_{i\cdot} \cdot p_{\cdot j}.$$

Положим

$$\bar{\chi}_{\text{незав.}}^2 = \sum_{i=1}^l \sum_{j=1}^k \frac{(\nu_{ij} - np_{ij})^2}{np_{ij}}.$$

Для мультиномиального распределения функция правдоподобия имеет вид

$$L = c \prod_{i,j} p_{ij}^{\nu_{ij}},$$

и для рассматриваемой нулевой гипотезы получим

$$L = c \prod_{i=1}^l p_{i\cdot}^{\nu_{i\cdot}} \prod_{j=1}^k p_{\cdot j}^{\nu_{\cdot j}}.$$

Можно показать, что оценка максимального правдоподобия для $p_{i\cdot}$ и $p_{\cdot j}$ такова

$$p_{i\cdot}^* = \frac{\nu_{i\cdot}}{n},$$

$$p_{\cdot j}^* = \frac{\nu_{\cdot j}}{n}.$$

Аналогично случаю обобщения классического критерия χ^2

$$\bar{\chi}_{\text{незав.}}^2 = \sum_{i=1}^l \sum_{j=1}^k \frac{n \left(\nu_{ij} - \frac{\nu_{i\cdot} \cdot \nu_{\cdot j}}{n} \right)^2}{\nu_{i\cdot} \cdot \nu_{\cdot j}} \rightarrow \chi_{kl-1-(l-1)-(k-1)}^2 = \chi_{(l-1)(k-1)}^2.$$

Так же, как и ранее, для некоторым образом определенного $\bar{\chi}_{\text{крит.}}^2$, гипотеза H_0 отвергается при $\bar{\chi}_{\text{незав.}}^2 \geq \bar{\chi}_{\text{крит.}}^2$. Поскольку считать вероятность

$P(\bar{\chi}_{\text{незав.}}^2 \geq \bar{\chi}_{\text{крит.}}^2 | H_0)$ довольно сложно, используем то, что

$$\alpha = P(\bar{\chi}_{\text{незав.}}^2 \geq \bar{\chi}_{\text{крит.}}^2 | H_0) \rightarrow P(\chi_{(l-1)(k-1)}^2 \geq \bar{\chi}_{\text{крит.}}^2 | H_0).$$

20.3 Критерий независимости χ^2

Рассмотрим (l) выборок из дискретных распределений $\mathcal{L}(X_1), \dots, \mathcal{L}(X_l)$, набор значений в которых одинаков — a_1, \dots, a_k :

$$\begin{aligned} X_{11}, \dots, X_{1n_1} &\text{ из } \mathcal{L}(X_1), \\ X_{21}, \dots, X_{2n_2} &\text{ из } \mathcal{L}(X_2), \\ &\dots \\ X_{l1}, \dots, X_{ln_l} &\text{ из } \mathcal{L}(X_l). \end{aligned}$$

По гипотезе H_0 предполагается, что $\mathcal{L}(X_i)$ одинаковы, т. е. имеем задачу об однородности выборки. Пусть ν_{ij} — число элементов i -той выборки, совпадающих со значением a_j , тогда H_0 такова, что $p_{ij} = p_j \quad \forall j = \overline{1, k}$ и сразу для всех i .

$$\begin{aligned} \bar{\chi}_{\text{однор.}}^2 &= \sum_{i=1}^l \sum_{j=1}^k \frac{(\nu_{ij} - n_i p_{ij})^2}{n_i p_{ij}} = \\ &= \left\{ p_{ij} = p_j(H_0) \text{ неизвестны, ОМП для них: } p_j^* = \frac{\nu_{\cdot j}}{n}, n = n_1 + \dots + n_l \right\} = \\ &= n \sum_{i=1}^l \sum_{j=1}^k \frac{\left(\nu_{ij} - \frac{\nu_{\cdot j} \cdot n_i}{n} \right)^2}{\nu_{\cdot j} \cdot n_i} \rightarrow \chi_{(l-1)(k-1)}^2 \text{ при } k \rightarrow \infty, \end{aligned}$$

и

$$\alpha = P(\bar{\chi}_{\text{однор.}}^2 \geq \bar{\chi}_{\text{крит.}}^2 \mid H_0) \rightarrow P(\chi_{(l-1)(k-1)}^2 \geq \bar{\chi}_{\text{крит.}}^2 \mid H_0),$$

т. е. H_0 отвергается при $\bar{\chi}_{\text{однор.}}^2 \geq \bar{\chi}_{\text{крит.}}^2$.

20.4 Проверка гипотез и доверительный интервал

Пример. Рассмотрим выборку X_1, \dots, X_n из $\mathcal{L}(X) \sim N(\theta, 1)$,

$$\begin{aligned} H_0: \theta &= 0, \\ H_1: \theta &> 0. \end{aligned}$$

Тогда р. н. м. кр. (равномерно наиболее мощный критерий) таков:

$$\text{при } \bar{X} > \frac{u_\alpha}{\sqrt{n}} \text{ гипотеза } H_0 \text{ отвергается,}$$

где α находится из уравнения $1 - \Phi(u_\alpha) = \alpha$.

Аналогично, если

$$\begin{aligned} H_0: \theta &= \theta_0, \\ H_1: \theta &> \theta_1, \end{aligned}$$

р. н. м. кр. таков:

$$\text{при } \bar{X} > \frac{u_\alpha}{\sqrt{n}} + \theta_0 \text{ гипотеза } H_0 \text{ отвергается,}$$

где $\alpha: 1 - \Phi(u_\alpha) = \alpha$.

Допустим, что имеется нерандомизированный S -критерий ($S = S(\theta_0)$), т. е. если $y = (x_1, \dots, x_n)$ — реализация выборки и $y \in S$, то простая нулевая гипотеза $H_0: \theta = \theta_0$ отвергается. Понятно, что H_0 принимается при $y \in \overline{S}$, причем $P(Y \in \overline{S} | H_0) = 1 - \alpha$.

Рассмотрим случай $n = 1$, тогда множество всех значения выборки $\mathcal{X} = \mathbb{R}$, р. н. м. кр.: $X_1 > \theta_0 + u_\alpha$, т. е.

$$S = S(\theta_0) = \{x: x > \theta_0 + u_\alpha\}.$$

Для $\forall \theta \in \mathbb{R}$ получим

$$\overline{S(\theta)} = \mathcal{X} \setminus S(\theta) \subset \mathcal{X},$$

т. е. имеем однопараметрическое семейство множеств $\{\overline{S(\theta)}\}$. С другой стороны, фиксируя произвольное $x_* \in \mathbb{R}$, определим

$$T(x_*) = \{\theta \in \Theta = \mathbb{R}: x_* \in \overline{S(\theta)}\},$$

тогда получим совокупность множеств $\{T(x)\}$ в пространстве Θ . Предположим, что

$$\begin{aligned} x &\in \overline{S(\theta)} \\ &\Downarrow \\ \theta &\in T(x), \end{aligned} \tag{20.1}$$

и, рассматривая вместо x случайную величину X_1 , будем иметь

$$1 - \alpha = P(X_1 \in \overline{S(\theta)}) = P(\theta \in T(X_1)).$$

Таким образом, область $T(X_1)$ в Θ в данном случае рассматривается нами как **доверительная область** для неизвестного параметра θ с коэффициентом доверия $(1 - \alpha)$.

Можно двигаться и в обратном направлении: пусть $T(X)$ — построенная доверительная область с коэффициентом доверия γ , тогда из (20.1) получим S -критерий с уровнем значимости $(1 - \gamma)$.

20.5 Несмещенные критерии

Определение 20.1. *Статистический критерий называется **несмещенным**, если для любого значения параметра из альтернативы H_1 мощность критерия не меньше уровня его значимости.*

Пусть X_1, \dots, X_n из $\mathcal{L}(X) \sim N(\theta, 1)$,

$$\begin{aligned} H_0: \theta &= \theta_0, \\ H_1: \theta &\neq \theta_0, \end{aligned}$$

α — уровень значимости. Можно применить лемму Неймана-Пирсона для построения критериев для альтернатив $\theta > \theta_0$ и $\theta < \theta_0$:

$$\begin{aligned} \text{для } \theta > \theta_0 \text{ критическая область } \overline{X} &> \theta_0 + \frac{u_{\alpha_1}}{\sqrt{n}}, \\ \text{для } \theta < \theta_0 \text{ критическая область } \overline{X} &< \theta_0 - \frac{u_{\alpha_2}}{\sqrt{n}}, \end{aligned}$$

где $\alpha_1, \alpha_2: \alpha_1 + \alpha_2 = \alpha$. Тогда критическая область

$$S = \left\{ x_1, \dots, x_n: \bar{x} > \theta_0 + \frac{u_{\alpha_1}}{\sqrt{n}} \right\} \cup \left\{ x_1, \dots, x_n: \bar{x} < \theta_0 - \frac{u_{\alpha_2}}{\sqrt{n}} \right\}.$$

Можно показать, что

1. мы получим несмещенный критерий только при $\alpha_1 = \alpha_2$,
2. среди всех несмещенных критериев данный критерий с критической областью S при $\alpha_1 = \alpha_2$ является равномерно наиболее мощным несмещенным критерием.

Так как

$$\bar{S} = \left\{ x_1, \dots, x_n: |\bar{x} - \theta_0| < \frac{u_{\alpha_2}}{\sqrt{n}} \right\},$$

то из (20.1) можно получить доверительный интервал

$$\bar{x} - \frac{u_{\alpha_2}}{\sqrt{n}} < \theta_0 < \bar{x} + \frac{u_{\alpha_2}}{\sqrt{n}},$$

который является доверительным интервалом наименьшей длины. Можно также показать, что если получен доверительный интервал минимальной длины, то соответствующий ему критерий будет равномерно наиболее мощным несмещенным критерием.