



Теория вероятностей

Конспекты лекций и семинаров

ЛЕКТОР: Д.А. ШАБАНОВ

Конспекты вели Денис Беляков, Никита Попов и Алексей Хачиянц

НИУ ВШЭ, 2016-2017

Оглавление

1	Лекции	7
1.1	Лекция от 09.09.2016	7
1.1.1	Введение. Принцип устойчивости частоты	7
1.1.2	Вероятностное пространство. Простейшие свойства вероятности	8
1.1.3	Классическая модель. Примеры	10
1.1.4	Условная вероятность. Формула полной вероятности	11
1.2	Лекция от 16.09.2016	12
1.2.1	Классические задачи теории вероятностей	12
1.2.2	Теорема Байеса	15
1.2.3	Независимость событий	15
1.2.4	Случайные величины в дискретных вероятностных пространствах	16
1.3	Лекция от 23.09.2016	16
1.3.1	Распределение случайной величины. Примеры распределений	16
1.3.2	Независимость случайных величин	18
1.3.3	Математическое ожидание. Свойства математического ожидания	18
1.3.4	Дисперсия. Ковариация. Их свойства	20
1.3.5	Неравенства Маркова и Чебышёва	22
1.4	Лекция от 30.09.2016	22
1.4.1	Сходимость случайных величин. Закон больших чисел.	22
1.4.2	Схема испытаний Бернулли	25
1.5	Лекция от 07.10.2016	27
1.5.1	Скорость сходимости закона больших чисел	27
1.5.2	Неравенство Чернова	27
1.5.3	Алгебры событий	29
1.6	Лекция от 14.10.2016	30
1.6.1	Локальная лемма Ловаса: несимметричный и симметричный случаи	30
1.6.2	Задача k-SAT. Теорема Мозеса-Тардоша	32
1.6.3	Геометрические вероятности. Задача о встрече	35
1.7	Лекция от 21.10.2016	35
1.7.1	Понятие вероятностного пространства в общем случае	36
1.7.2	Вероятностные меры на $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.	40
1.8	Лекция от 01.11.2016	41
1.8.1	Взаимно-однозначное соответствие функции распределения и вероятностной меры	41
1.8.2	Классификация функций распределения на \mathbb{R}	43
1.9	Лекция от 08.11.2016	47
1.9.1	Канторова лестница. Продолжение.	47
1.9.2	Случайные величины и векторы	48

1.9.3	Действия над случайными величинами	50
1.9.4	Простые случайные величины. Матожидание в общем случае	51
1.10	Лекция от 15.11.2016	52
1.10.1	Свойства математического ожидания для простых случайных величин.	52
1.10.2	Свойства математического ожидания в общем случае.	54
1.11	Лекция от 22.11.2016	56
1.11.1	Классификация случайных величин. Формулы подсчёта математи- ческого ожидания	56
1.11.2	Независимость случайных величин и векторов	60
1.11.3	Математическое ожидание произведения независимых случайных ве- личин	61
1.12	Лекция от 29.11.2016	62
1.12.1	Многомерный случай	62
1.12.2	Случайные величины в многомерном случае	66
1.12.3	Математическое ожидание в многомерном случае	69
1.13	Лекция от 06.12.2016	70
1.13.1	Полезные формулы	70
1.13.2	Формула замены переменных в кратном интеграле	71
1.13.3	Дисперсия и ковариация.	72
1.14	Лекция от 20.01.2017	75
1.14.1	Дифференцирование характеристических функций	75
1.14.2	Характеристическая функция случайного вектора и её свойства	77
1.14.3	Центральная предельная теорема и примеры её применения	78
1.15	Лекция от 27.01.2017	81
1.15.1	Сходимость случайных векторов	81
1.15.2	Теорема о наследовании сходимости, лемма Слуцкого и их применения	82
1.15.3	Многомерное нормальное распределение	85
1.16	Лекция от 10.02.2017	86
1.16.1	Вычисление условного математического ожидания	86
1.16.2	Свойства условного математического ожидания	86
1.16.3	Условные распределения и плотности	89
2	Семинары	91
2.1	Семинар от 09.09.2016	91
2.2	Семинар от 16.09.2016	96
2.3	Семинар от 23.09.2016	101
2.4	Семинар от 30.09.2016	105
2.5	Семинар от 07.10.2016	110
2.6	Семинар от 21.10.2016	115
2.7	Семинар от 15.11.2016	118
2.8	Семинар от 22.11.2016	119
2.9	Семинар от 29.11.2016	128
3	Дополнительные главы теории вероятностей	131
3.1	Лекция от 24.01.2017	131
3.1.1	Случайные процессы	131
3.1.2	Случайные блуждания	132
3.1.3	Процесс восстановления	132

3.1.4	Простые случайные блуждания	135
3.1.5	Возвращение в ноль в простом случайном блуждании	136
3.2	Лекция от 31.01.2017	137
3.2.1	Числа Каталана. Рекуррентное соотношение. Производящая функция	137
3.2.2	Вероятность возвращения	139
3.2.3	Многомерный случай	139
3.2.4	Числа Каталана через биномиальные коэффициенты	140
3.2.5	Математическое ожидание первого момента возвращения в ноль . . .	141
3.2.6	Среднее время в нуле	141
3.3	Лекция от 07.02.2017	142
3.3.1	Среднее время в нуле	142
3.3.2	Геометрия траекторий. Закон повторного логарифма	144

Глава 1

Лекции

1.1 Лекция от 09.09.2016

1.1.1 Введение. Принцип устойчивости частоты

Чем занимается теория вероятностей? Она изучает *случайные* явления. Допустим, мы провели какой-либо эксперимент. Можем ли мы что-то заранее сказать о результате?

- Если да, то результат называют *детерминированным*. Пример такого эксперимента — выбрасывание кирпича из окна. Очевидно, что кирпич упадёт на землю¹ и результат предопределён. Такие задачи изучают в той же линейной алгебре или где-либо ещё, но не в теории вероятностей.
- А теперь предположим, что заранее сказать, каков будет результат, невозможно. Например, точно сказать, какой стороной упадёт подброшенная монетка, вряд ли получится. Тогда результат называют *недетерминированным*. Именно задачи с недетерминированным результатом и изучаются в теории вероятностей.

Небольшое историческое отступление — вообще говоря, теория вероятностей появилась в связи с изучением азартных игр наподобие рулетки ещё в средних веках. Но тогда она представляла собой скорее набор эмпирических фактов, чем полноценную науку. Теория вероятностей стала такой, какой она является сейчас, лишь в XX веке благодаря трудам А.Н. Колмогорова.

Хорошо, а как изучаются случайные процессы? Ну выпала решка, и что? На самом деле теория вероятностей не о единичных экспериментах, а об *асимптотике*. Это значит следующее: если проводить серию одинаковых экспериментов, то теория вероятностей поможет предсказать частоту, с которой будет появляться какой-либо ответ.

Теория вероятностей держится на крайне важном *принципе устойчивости частоты*. Перед тем, как ввести формальное определение, рассмотрим пару экспериментов, связанных с подбрасыванием монетки:

- В XVIII веке Жорж-Луи Леклерк де Бюффон провёл эксперимент, подбросив монетку 4040 раз. Из них в 2048 бросках выпал герб. В итоге частота составила около 0.506.
- В XIX веке пошли ещё дальше — Карл Пирсон подбросил монетку 24000 раз.² У него получилось так, что герб выпал 12012 раз. В итоге частота составила 0.5005.

¹Если его не запустили с первой космической скоростью, конечно.

²Оставим вопрос о том, как он не поленился проверить это, без ответа.

Отсюда видно, что эксперименты дают частоту, близкую к $1/2$.

Неформально говоря, принцип формулируется так: если мы проводим серию одинаковых экспериментов, то количество появлений одного определённого ответа при делении на число экспериментов сходится к некоторому числу $p \in [0, 1]$. Теперь можно ввести формальное определение.

Принцип устойчивости частоты. Пусть A — некоторое событие, а $v_n(A)$ — число экспериментов, в которых происходит событие A среди первых n . Тогда

$$\lim_{n \rightarrow \infty} \frac{v_n(A)}{n} = p, \quad p \in [0, 1].$$

Получаемое число p называют *вероятностью* события A и обозначают $P(A)$. Например, $P(\text{встретить живого динозавра на улице}) = 0$, так как они все вымерли.

1.1.2 Вероятностное пространство. Простейшие свойства вероятности

Именно после введения этого понятия Колмогоровым теория вероятностей перестала быть прежней. Введём определение для дискретного случая (общий оставим на потом):

Определение 1. Дискретным вероятностным пространством называется пара (Ω, P) , где Ω — множество элементарных исходов, а P — вероятность на Ω .

Множество элементарных исходов Ω — некоторое конечное или счётное множество. Элемент $\omega \in \Omega$ называют *элементарным исходом*. Полагается, что в случайном эксперименте обязательно получается один и только один элементарный исход.

Примеры множеств элементарных исходов:

1. $\Omega = \{O, P\}$ — бросок монеты.
2. $\Omega = \{\omega_1, \dots, \omega_6\}$, где $\omega_i = \{\text{выпало } i \text{ очков}\}$ — бросок игрального кубика.
3. $\Omega = \{\omega_1, \dots, \omega_n, \dots\}$, где $\omega_i = \{\text{на данный момент горит } i \text{ зданий}\}$ — предсказание пожаров в городе.

Определение 2. Подмножество $A \subseteq \Omega$ называется *событием* на вероятностном пространстве (Ω, P) .

Пример события: пусть подбрасывают игральную кость, и $A = \{\text{выпало чётное число очков}\}$. Тогда $A = \{\omega_2, \omega_4, \omega_6\}$.

Определение 3. Отображение $P : \Omega \rightarrow [0, 1]$ называют *вероятностью*³, если $\sum_{\omega \in \Omega} P(\omega) =$

1. В случае счётного множества Ω данный ряд должен сходиться абсолютно.

Пусть у нас есть некоторое событие A на вероятностном пространстве (Ω, P) . Как посчитать его вероятность?

Определение 4. Вероятностью события $A \subseteq \Omega$ называют $P(A) = \sum_{\omega \in A} P(\omega)$.

³В общем случае вероятность ещё могут называть *вероятностной мерой*.

Определение 5. Пусть A — некоторое событие. Тогда *дополнением* к событию A называют событие $\bar{A} = \Omega \setminus A$.

Перед тем, как идти дальше, напомним определение дизъюнктного объединения.

Определение 6. Пусть есть множества A_1, A_2, \dots, A_n . Тогда *дизъюнктным объединением* множеств называют объединение попарно непересекающихся “копий” множеств:

$$\bigsqcup_{i=1}^n A_i = \bigcup_{i=1}^n \{(x, i) \mid x \in A_i\}.$$

В нашем случае полагается, что если пишут дизъюнктное объединение, то множества попарно не пересекаются.

Рассмотрим некоторые свойства вероятности.

Теорема 1 (Простейшие свойства вероятности). *Для любого дискретного вероятностного пространства (Ω, P) выполняется следующее:*

1. $P(\Omega) = 1, P(\emptyset) = 0$.
2. *Конечная аддитивность:* $P\left(\bigsqcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$.
3. $P(A) + P(\bar{A}) = 1$.
4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
5. *Для любого набора событий A_1, A_2, \dots, A_n* $P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i)$.
6. *Счётная аддитивность:* $P\left(\bigsqcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$.

Последнее свойство выполняется только для счётного Ω .

Доказательство. Распишем доказательства для каждого пункта по отдельности:

1. $P(\Omega) = 1$ следует из определения вероятности, а $P(\emptyset) = 0$ следует из определения вероятности события.
2. Случай с конечным множеством Ω очевиден. Тогда положим, что Ω счётно, то есть $\Omega = \{\omega_i \mid i \in \mathbb{N}\}$.

Пусть есть некоторое событие A . Тогда представим его вероятность в удобном для нас виде:

$$P(A) = \sum_{\omega \in A} P(\omega) = \sum_{i: \omega_i \in A} P(\omega_i) = \lim_{N \rightarrow \infty} \sum_{\substack{i: \omega_i \in A \\ i < N}} P(\omega_i).$$

Теперь распишем вероятность дизъюнктного объединения событий A_1, A_2, \dots, A_n :

$$\begin{aligned} P\left(\bigsqcup_{i=1}^n A_i\right) &= \sum_{\omega \in \bigsqcup A_i} P(\omega) = \lim_{N \rightarrow \infty} \left(\sum_{\substack{i: \omega_i \in \bigsqcup A_i \\ i < N}} P(\omega_i) \right) = \lim_{N \rightarrow \infty} \sum_{i=1}^n \sum_{\substack{j: \omega_j \in A_i \\ j < N}} P(\omega_j) = \\ &= \sum_{i=1}^n \lim_{N \rightarrow \infty} \sum_{\substack{j: \omega_j \in A_i \\ j < N}} P(\omega_j) = \sum_{i=1}^n P(A_i). \end{aligned}$$

3. Согласно второму пункту, $1 = P(\Omega) = P(A \cup \bar{A}) = P(A) + P(\bar{A})$.
4. Так как $A \cup B = A \cup (B \setminus A)$ и $A \cap (B \setminus A) = \emptyset$, то $P(A \cup B) = P(A) + P(B \setminus A)$. Далее, заметим, что $B \setminus A = B \setminus (A \cap B)$. Тогда $P(A \cup B) = P(A) + P(B \setminus (A \cap B))$.
Рассмотрим второй член. Заметим, что $P(A \cap B) + P(B \setminus (A \cap B)) = P(B)$. Тогда получаем, что $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
5. Докажем это утверждение по индукции. База была доказана в пункте 4 (так как $P(A \cap B) \geq 0$). Теперь рассмотрим шаг индукции. Пусть утверждение верно для какого-то m . Тогда $P\left(\bigcup_{i=1}^{m+1} A_i\right) \leq P(A_{m+1}) + P\left(\bigcup_{i=1}^m A_i\right) \leq \sum_{i=1}^{m+1} P(A_i)$.
6. За доказательством этого пункта обращайтесь к учебнику матанализа на тему частичных сумм и абсолютной сходимости.⁴

□

1.1.3 Классическая модель. Примеры

Пусть (Ω, P) — некоторое конечное вероятностное пространство, при этом все элементарные исходы равновероятны. Тогда легко посчитать вероятность элементарного исхода: $P(\omega) = \frac{1}{|\Omega|}$ для всех $\omega \in \Omega$. Таковую модель называют *классической*.

Как посчитать вероятность события в классической модели? Очень просто: $P(A) = \frac{|A|}{|\Omega|}$ для всех $A \subseteq \Omega$.

Рассмотрим некоторые примеры классических моделей.

1. Бросок монетки. В таком случае $\Omega = \{O, P\}$ и $P(O) = P(P) = \frac{1}{2}$.
2. Бросок двух монеток. С этой моделью связано одно заблуждение Д'Аламбера. Он рассуждал следующим образом: так как $\Omega = \{OO, PP, OP\}$, то $P(OO) = P(PP) = P(OP) = \frac{1}{3}$. Но это опровергается экспериментами. И как это исправить? Есть два варианта:
 - (а) Можно сказать, что модель не является классической и поправить вероятности: $P(OO) = P(PP) = \frac{1}{4}$, $P(OP) = \frac{1}{2}$.
 - (б) А можно просто изменить множество элементарных исходов. Начнём учитывать порядок выпадения: $\Omega = \{OO, PP, OP, PO\}$. Такая модель уже является классической.

Рассуждая в стиле Д'Аламбера, можно прийти к выводу, что вероятность встретить живого динозавра на улице равна $\frac{1}{2}$, ведь его можно либо встретить, либо не встретить.

3. Бросок n монет. В таком случае вероятностное пространство будет устроено следующим образом: $\Omega = \{\omega = (\omega_1, \omega_2, \dots, \omega_n) \mid \omega_i \in \{O, P\}\}$. Легко понять, что в данной модели 2^n элементарных исходов.

Примечание. Вероятностное пространство такого вида называют *симметрической схемой Бернулли*.

⁴На самом деле я попробую найти доказательство. Когда-нибудь. Но не сейчас. (А.Х.)

Но данная модель является классической только тогда, когда монетки “честные”, то есть которые падают орлом или решкой вверх равновероятно. Если же это не так, то вероятность элементарного исхода $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ задаётся следующей формулой: $P(\omega) = p^{\sum_{i=1}^n \omega_i} (1-p)^{n-\sum_{i=1}^n \omega_i}$.

4. Урновые схемы (размещение частиц по ячейкам). Пусть есть n различных шаров в ящике. Мы случайным образом вынимаем m шаров. Вопрос: каков размер множества элементарных исходов? Сначала приведём ответ, после чего докажем его.

Порядок? Возврат?	Упорядоченный набор	Неупорядоченный набор
С возвратом	n^m	$C_{n+m-1}^m = \frac{(n+m-1)!}{m!(n-1)!}$
Без возврата	$A_n^m = \frac{n!}{(n-m)!}$	$C_n^m = \frac{n!}{m!(n-m)!}$

Доказательство. Будем доказывать утверждения от верхнего левого против часовой стрелки.

- Пусть набор упорядочен и можно возвращать. Тогда любой элемент набора можно получить n способами (так как все элементы можно вернуть). Отсюда получаем n^m .
- Теперь положим, что набор упорядочен, но возвращать нельзя. Тогда первый элемент можно выбрать n способами, второй — $n-1$ способом и так далее до m -го элемента, который можно выбрать $n-m+1$ способом. По правилу умножения получаем $\frac{n!}{(n-m)!} = A_n^m$.
- Рассмотрим случай, когда набор неупорядочен и возвращать нельзя. Тогда необходимо посчитать количество способов выбрать k шаров из m . Достаточно логично, что это равно $\frac{A_n^m}{m!} = \frac{n!}{m!(n-m)!} = C_n^m$, так как в последовательности нам не важен порядок.
- Осталось рассмотреть последний случай — неупорядоченный набор с возвратом. В этом случае нам достаточно указать, сколько раз мы выбрали каждый шар. Как это сделать? Воспользуемся методом точек и перегородок. Пусть есть m точек и нужно распределить их по n группам. Для этого нужно использовать $n-1$ перегородку. Тогда задача сводится к нахождению количества способов выбрать m элементов из $n+m-1$. А это равно $C_{n+m-1}^m = \frac{(n+m-1)!}{m!(n-1)!}$.

□

1.1.4 Условная вероятность. Формула полной вероятности

Пусть (Ω, P) — дискретное вероятностное пространство.

Определение 7. Пусть $A \subseteq \Omega$ — некоторое событие и $B \subseteq \Omega$ — другое событие, причём $P(B) > 0$. Тогда *условной вероятностью события A при условии B* называют

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

Если $P(B) = 0$, то положим, что $P(A | B) = 0$ для любого события $A \subseteq \Omega$.

Условную вероятность можно воспринимать следующим образом: сузим множество элементарных исходов до B и посчитаем вероятность события A на полученном множестве.

Примечание. Если $P(B) > 0$, то $\tilde{P}(A) = P(A | B)$ тоже является вероятностью на Ω .

Определение 8. Пусть B_1, B_2, \dots, B_n — некоторые события на Ω такие, что $\bigsqcup_{i=1}^n B_i = \Omega$. Тогда этот набор событий называется (конечным) *разбиением* Ω .

Теперь докажем важную формулу:

Формула полной вероятности. Пусть B_1, B_2, \dots, B_n — разбиение Ω . Тогда для любого события $A \subseteq \Omega$ верно, что $P(A) = \sum_{i=1}^n P(A | B_i) P(B_i)$.

Доказательство. Так как $A \cap \Omega = A$ и $\bigsqcup_{i=1}^n B_i = \Omega$, то $P(A) = P\left(A \cap \left(\bigsqcup_{i=1}^n B_i\right)\right)$. Заметим, что $A \cap \left(\bigsqcup_{i=1}^n B_i\right) = \bigsqcup_{i=1}^n (A \cap B_i)$. Тогда $P(A) = \sum_{i=1}^n P(A \cap B_i) = \sum_{i=1}^n P(A | B_i) P(B_i)$. \square

Заметим, что формула полной вероятности работает и в случае, когда $P(B_i) = 0$ для какого-то i .

1.2 Лекция от 16.09.2016

1.2.1 Классические задачи теории вероятностей

Задача 1 (Задача о сумасшедшей старушке). *Есть самолёт на n мест, в который садятся n пассажиров. Первой в него заходит безумная старушка, которая садится на случайное место. Каждый следующий пассажир действует по следующему правилу: садится на своё место, если оно свободно, и на случайное, если своё занято. С какой вероятностью*

1. последний пассажир сядет на своё место?
2. предпоследний пассажир сядет на своё место?
3. и последний, и предпоследний пассажир сядут на свои места?

Решение. У первого пункта есть элементарное решение. Пусть при некоторой рассадке пассажиров последний пассажир сел не на свое место (такую рассадку назовем неудачной). Тогда до прихода последнего пассажира его место было занято пассажиром A (A может быть и сумасшедшей старушкой). В момент прихода пассажира A перед ним стоит выбор — какое место занять. В рассматриваемой рассадке он занимает место последнего пассажира. Но с той же вероятностью он мог занять и место старушки, и в дальнейшем все пассажиры, включая последнего, займут свои собственные места. (Конечно, нужно еще пояснить, почему в момент прихода пассажира A старушкино место все еще свободно. Но это действительно так — нетрудно проследить, что пока старушкино место свободно, среди всех еще не вошедших пассажиров есть ровно один, чье место уже занято. Как только

очередной пассажир занимает бабушкино место, все остальные будут садиться только на свои места.) Таким образом, каждой неудачной рассадке соответствует удачная, которая может случиться с той же вероятностью. Это говорит о том, что ровно в половине случаев рассадка будет неудачной.

Теперь рассмотрим формальное решение для первого пункта. Пусть $A = \{\text{последний сядет на своё место}\}$. Если⁵ $n = 2$, то $P(A) = \frac{1}{2}$. Теперь рассмотрим случай $n = 3$. Пусть $B_i = \{\text{бабушка села на } i\text{-е место}\}$. По формуле полной вероятности

$$P(A) = P(A | B_1) P(B_1) + P(A | B_2) P(B_2) + P(A | B_3) P(B_3) = \frac{1}{3} \left(1 + \frac{1}{2} + 0 \right) = \frac{1}{2}.$$

Намечается закономерность. Попробуем сформулировать гипотезу и доказать её:

Гипотеза. Для любого n вероятность того, что последний пассажир сядет на своё место, равна $\frac{1}{2}$.

Доказательство. По индукции. База ($n = 2$) была доказана ранее. Теперь предположим, что утверждение верно для всех $k < n$. Тогда докажем, что утверждение верно для $k = n$. Опять же, распишем вероятность по формуле полной вероятности:

$$P(A) = \sum_{i=1}^n P(A | B_i) P(B_i).$$

Заметим, что $P(B_i) = \frac{1}{n}$. Теперь посмотрим на значения условных вероятностей:

$$P(A | B_i) = \begin{cases} 1, & i = 1 \\ 0, & i = n \\ \frac{1}{2}, & 2 \leq i \leq n - 1 \end{cases}$$

Последнее условие на значение условной вероятности следует из шага индукции, так как i -й пассажир “становится” бабушкой. Тогда $P(A) = \frac{1}{n} \left(1 + \frac{1}{2} + \dots + \frac{1}{2} + 0 \right) = \frac{1}{2}$. \square

Перейдём ко второму пункту задачи. Пусть $C = \{\text{предпоследний пассажир сел на своё место}\}$. Тогда рассмотрим $P(C)$ в случае, когда $n = 3$. Он сможет сесть на своё место тогда и только тогда, когда бабушка не села на его место. Тогда $P(C) = \frac{2}{3}$. Попробуем доказать гипотезу, аналогичную случаю с последним пассажиром.

Гипотеза. Для любого n вероятность того, что предпоследний пассажир сядет на своё место, равна $\frac{2}{3}$.

Доказательство. Доказательство практически такое же, как и для последнего пассажира, с тем отличием, что значения условной вероятности будут несколько другие:

$$P(C | B_i) = \begin{cases} 1, & i = 1, n \\ 0, & i = n - 1 \\ \frac{2}{3}, & 2 \leq i \leq n - 2 \end{cases}$$

Тогда $P(C) = \frac{1}{n} \left(1 + \frac{2}{3} + \dots + \frac{2}{3} + 0 + 1 \right) = \frac{2}{3}$ \square

⁵Это уже вертолёт, скорее. (Д.А. Шабанов)

Аналогичными рассуждениями можно доказать, что вероятность того, что i -й с конца пассажир сядет на своё место, равна $\frac{i}{i+1}$.

Теперь приступим к третьему пункту. Докажем следующее утверждение:

Гипотеза. Вероятность этого события (обозначим его за D) равна $\frac{1}{3}$.

Доказательство. Опять же, по индукции. Базой служит случай $n = 3$. В таком случае условие выполнимо тогда и только тогда, когда бабушка сядет на своё место. Тогда вероятность равна $\frac{1}{3}$ и база верна.

Перейдём к шагу индукции. Рассуждаем абсолютно аналогично: по формуле полной вероятности $P(D) = \sum_{i=1}^n P(D | B_i) P(B_i)$. Значения условных вероятностей равны

$$P(D | B_i) = \begin{cases} 1, & i = 1 \\ 0, & i = n - 1, n \\ \frac{1}{3}, & 2 \leq i \leq n - 2 \end{cases}$$

$$\text{и } P(D) = \frac{1}{n} \left(1 + \frac{1}{3} + \dots + \frac{1}{3} + 0 + 0 \right) = \frac{1}{3}. \quad \square$$

Заметим, что $P(D) = P(A)P(C)$. Это ~~счастливое совпадение~~ неспроста. □

Задача 2 (Задача об удачливом студенте). *Студент начал готовиться к экзамену слишком поздно и выучил только k билетов из n . Студент решил схитрить и выбрать место в очереди такое, чтобы вероятность получить выученный билет была максимальной. Какое место ему выбрать?*

Решение. Пусть $A_s = \{\text{студент вытащил хороший билет, если он встал } s\text{-тым в очереди}\}$. Докажем следующую гипотезу:

Гипотеза. $P(A_s) = \frac{k}{n}$.

Доказательство. Введём разбиение B_0, B_1, \dots, B_k пространства Ω , где $B_i = \{\text{до студента взяли ровно } i \text{ хороших билетов}\}$. Тогда по формуле полной вероятности

$$P(A_s) = \sum_{i=0}^k P(A_s | B_i) P(B_i).$$

Посчитаем $P(B_i)$. Для определения количества успешных исходов нужно выбрать i билетов из k хороших и $s - i - 1$ из $n - k$ плохих. Тогда

$$P(B_i) = \frac{\binom{k}{i} \binom{n-k}{s-i-1}}{\binom{n}{s-1}}.$$

Теперь надо определить значение условной вероятности $P(A_s | B_i)$. Так как есть $n - s + 1$ невыбранный билет и $k - i$ из них изучены, то $P(A_s | B_i) = \frac{k-i}{n-s+1}$. Тогда

$$P(A) = \sum_{i=0}^k \frac{k-i}{n-s+1} \frac{\binom{k}{i} \binom{n-k}{s-i-1}}{\binom{n}{s-1}} = \frac{k}{n} \sum_{i=0}^k \frac{\binom{k-1}{i} \binom{n-k}{s-i-1}}{\binom{n-1}{s-1}}.$$

Теперь заметим, что сумма справа равна 1, так как она соответствует разбиению пространства в случае, когда всего задач $n - 1$. Тогда $P(A_s) = \frac{k}{n}$. □

В итоге как ни вставай — всё равно никакой разницы не будет.⁶ □

⁶Ибо по закону подлости попадётся невыученный билет.

1.2.2 Теорема Байеса

Теорема 2 (Байес). Пусть $\{B_1, B_2, \dots, B_n\}$ — разбиение Ω , причём $P(B_i) > 0$ для всех $i \in \{1, 2, \dots, n\}$. Тогда для любого события A такого, что $P(A) > 0$, выполняется

$$P(B_i | A) = \frac{P(A | B_i) P(B_i)}{\sum_{j=1}^n P(A | B_j) P(B_j)}$$

Доказательство. По определению условной вероятности $P(B_i | A) = \frac{P(A \cap B_i)}{P(A)} = \frac{P(A|B_i) P(B_i)}{P(A)}$. Тогда, применяя формулу полной вероятности для $P(A)$, получаем желаемое. \square

1.2.3 Независимость событий

Определение 9. События A и B на вероятностном пространстве (Ω, P) называются *независимыми*, если $P(A \cap B) = P(A)P(B)$. Иногда используется обозначение $A \perp B$.

Рассмотрим некоторые примеры независимых событий:

1. Задача про сумасшедшую бабушку: События $A = \{\text{последний сядет на своё место}\}$ и $B = \{\text{предпоследний сядет на своё место}\}$ независимы, как было доказано ранее.
2. Бросок игральной кости. Пусть $A = \{\text{выпало чётное число очков}\}$, а $B = \{\text{выпало число очков, кратное 3}\}$. Докажем, что они независимы. Заметим, что $A \cap B = \{\text{число очков кратно 6}\}$. Тогда $P(A \cap B) = \frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3} = P(A)P(B)$.

Данное определение работает только для двух событий. Можно ли как-то его обобщить? Попробуем ввести аналогично:

Определение 10. События A_1, \dots, A_n называются *попарно независимыми*, если для любых i и j таких, что $i \neq j$, A_i независимо от A_j .

Однако обобщение обычно вводят по другому:

Определение 11. События A_1, A_2, \dots, A_n называются *независимыми по совокупности*, если для любого множества $\{i_1, i_2, \dots, i_k\} \subseteq \{1, 2, \dots, n\}$ выполняется, что $P(A_{i_1} \cap \dots \cap A_{i_k}) = \prod_{j=1}^k P(A_{i_j})$.

Примечание. Если говорят про независимость событий и не указывают тип, то обычно подразумевают независимость по совокупности.

Стоит заметить, что попарная независимость — гораздо более слабое условие, чем независимость по совокупности. Приведём пример. Пусть есть тетраэдр, грани которого покрашены следующим образом: первая грань покрашена в красный, вторая — в синий, третья — в зелёный, а четвёртая — во все три цвета сразу. Его подбрасывают. Введём три события: $A_K = \{\text{на нижней грани есть красный цвет}\}$, $A_S = \{\text{на нижней грани есть синий цвет}\}$ и $A_Z = \{\text{на нижней грани есть зелёный цвет}\}$. Очевидно, что вероятность любого события — $\frac{1}{2}$, а любой пары событий — $\frac{1}{4}$. Однако вероятность объединения всех трёх событий равна $\frac{1}{4}$, а не $\frac{1}{8}$. Следовательно, события попарно независимы, но не независимы по совокупности.

Упражнение. Приведите пример n событий таких, что любой набор из $n - 1$ события независим, а все n событий вместе зависимы.

Для независимости выполняются следующие свойства:

1. $A \perp A \iff P(A) = 0$ или $P(A) = 1 \iff$ для любого события B $A \perp B$.
2. $A \perp B \implies \bar{A} \perp B$.
3. Если A_1, A_2, \dots, A_n независимы в совокупности, то любой набор B_1, B_2, \dots, B_n такой, что $B_i = A_i$ или $B_i = \bar{A}_i$, тоже независим.

1.2.4 Случайные величины в дискретных вероятностных пространствах

Определение 12. Пусть (Ω, P) — дискретное вероятностное пространство. Тогда любое отображение $\xi : \Omega \rightarrow \mathbb{R}$ называется *случайной величиной*.

Примечание. В литературе случайную величину часто сокращают до с.в.

Рассмотрим некоторые примеры случайных величин:

1. Широко распространённым примером случайной величины служит индикатор какого-либо события. Введём определение.

Определение 13. Пусть $A \in \Omega$ — событие. Тогда *индикатором* события A называют называется случайная величина, равная

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

2. Бросок игральной кости. Тогда случайной величиной будет число выпавших очков.
3. Схема Бернулли: $\Omega = \{\omega = (\omega_1 \dots \omega_n), \omega_i \in \{0, 1\}\}$. В данном случае вводится случайная величина, равная $\xi(\omega) = \sum_{i=1}^n \omega_i$. Ещё её называют *числом успехов в схеме Бернулли*.

1.3 Лекция от 23.09.2016

1.3.1 Распределение случайной величины. Примеры распределений

Рассмотрим некоторое дискретное пространство (Ω, P) . Тогда случайная величина ξ на этом пространстве принимает не более, чем счетное множество значений. Пусть $X = (x_1, x_2, \dots, x_n, \dots)$ — множество значений случайной величины ξ . Введём событие $A_i = \{\omega : \xi(\omega) = x_i\}$. Его можно интерпретировать, как событие $\{\xi \text{ приняло значение } x_i\}$. Для удобства будем использовать обозначение $A_i := \{\xi = x_i\}$. Также введем обозначение p_i для вероятности события A_i : $p_i = P(A_i) = P(\xi = x_i)$.

Определение 14. Вместе множество значений $X = (x_1, x_2, \dots, x_n, \dots)$ и набор вероятностей $(p_1, p_2, \dots, p_n, \dots)$ образуют то, что называется *распределением случайной величины* ξ .

Понятно, что каждому числу x_i сопоставлено число p_i . Легко заметить, что A_i образуют разбиение пространства Ω . Тогда $\sum_{i=0}^{|X|} p_i = 1$.

Рассмотрим некоторые известные примеры распределений:

1. Распределение Бернулли.

Определение 15. Случайная величина ξ имеет *распределение Бернулли*, если она принимает всего два значения, 1 или 0, с заранее известными вероятностями p и $q \equiv 1 - p$. Обозначение: $\xi \sim \text{Bern}(p)$.

Легко понять, что $X = \{0, 1\}$, $P(\xi = 1) = p$, $P(\xi = 0) = q$. Принято говорить, что событие $\{\xi = 1\}$ соответствует “успеху”, а событие $\{\xi = 0\}$ — “неудаче”. Эти названия достаточно условные, и в зависимости от конкретной задачи могут быть заменены на противоположные.

2. Биномиальное распределение.

Определение 16. *Биномиальное распределение* в теории вероятностей — распределение количества “успехов” ξ в последовательности из n независимых случайных экспериментов, таких, что вероятность “успеха” в каждом из них постоянна и равна p . Обозначение: $\xi \sim \text{Bin}(n, p)$.

Пусть x_1, \dots, x_n — последовательность независимых случайных величин с одинаковым распределением Бернулли с параметром p . Тогда $\xi = \sum_{i=1}^n x_i$. Посчитаем $P(\xi = k)$. Для этого необходимо выбрать k исходов из n и сказать, что они успешны. Тогда $P(\xi = k) = \binom{n}{k} p^k (1 - p)^{n-k}$.

3. Пуассоновское распределение.

Определение 17. Говорят, что случайная величина ξ имеет *распределение Пуассона*, если она принимает любое значение⁷ $k \in \mathbb{N}$ с вероятностью $\frac{\lambda^k}{k!} e^{-\lambda}$, где $\lambda > 0$ — некоторый параметр. Обозначение: $\xi \sim \text{Pois}(\lambda)$.

Почему оно вводится именно так? Об этом будет рассказано позднее.

4. Геометрическое распределение.

Определение 18. Говорят, что случайная величина ξ имеет *геометрическое распределение*, если она принимает любое значение $k \in \mathbb{N}$ с вероятностью $p(1 - p)^{k-1}$. Обозначение: $\xi \sim \text{Geom}(p)$.

⁷Есть путаница, откуда считать — с нуля или же с единицы. Будем считать, что начинаем с единицы.

1.3.2 Независимость случайных величин

Далее множество значений случайной величины ξ на вероятностном пространстве (Ω, \mathcal{P}) будем обозначать как $\xi(\Omega)$.

Определение 19. Пусть ξ, η — случайные величины на вероятностном пространстве (Ω, \mathcal{P}) . Будем говорить, что эти случайные величины *независимы*, если для любых $a \in \xi(\Omega), b \in \eta(\Omega)$ события $\{\xi = a\}$ и $\{\eta = b\}$ независимы.

По определению независимых событий получаем, что

$$\mathcal{P}(\{\xi = a\} \cap \{\eta = b\}) = \mathcal{P}(\xi = a) \mathcal{P}(\eta = b).$$

Примечание. Для простоты левую часть часто обозначают за $\mathcal{P}(\xi = a, \eta = b)$.

Теперь обобщим это понятие на произвольное количество случайных величин:

Определение 20. Пусть ξ_1, \dots, ξ_n — случайные величины на вероятностном пространстве (Ω, \mathcal{P}) и известно, что ξ_i принимает значения $(a_1^{(i)}, a_2^{(i)}, \dots, a_n^{(i)})$. Тогда будем говорить, что случайные величины ξ_1, \dots, ξ_n *независимы (в совокупности)*, если $\forall j_1, \dots, j_n$ выполнено:

$$\mathcal{P}(\xi_1 = a_{j_1}^{(1)}, \dots, \xi_n = a_{j_n}^{(n)}) = \prod_{k=1}^n \mathcal{P}(\xi_k = a_{j_k}^{(k)}).$$

Упражнение. Пусть A, B — некоторые события над (Ω, \mathcal{P}) . Показать, что A и B независимы тогда и только тогда, когда их индикаторы I_A и I_B независимы.

Упражнение. Показать, что случайные величины ξ_1, \dots, ξ_n независимы тогда и только тогда, когда

$$\forall \alpha_1, \alpha_2, \dots, \alpha_n \in \mathbb{R} \quad \mathcal{P}(\xi_1 = \alpha_1, \dots, \xi_n = \alpha_n) = \prod_{i=1}^n \mathcal{P}(\xi_i = \alpha_i)$$

1.3.3 Математическое ожидание. Свойства математического ожидания

Опять же, зафиксируем некоторое дискретное вероятностное пространство (Ω, \mathcal{P}) .

Определение 21. Математическим ожиданием случайной величины ξ называют величину $\mathcal{E}[\xi] = \sum_{\omega \in \Omega} \xi(\omega) \mathcal{P}(\omega)$.

Примечание. Если Ω счётно, то ряд $\sum_{\omega \in \Omega} \xi(\omega) \mathcal{P}(\omega)$ должен сходиться абсолютно; иначе сумма ряда не определена однозначно, так как порядок перебора ω не задан (см. теорему Римана о перестановке членов условно сходящегося ряда).

Смысл у математического ожидания случайной величины простой — его можно понимать как среднее значение этой случайной величины.

Рассмотрим некоторые свойства математического ожидания.

Теорема 3 (Простейшие свойства математического ожидания). Пусть ξ, η — некоторые случайные величины. Тогда выполняются следующие свойства:

1. Матожидание линейно: $\forall a, b \in \mathbb{R} \quad \mathcal{E}[a\xi + b\eta] = a \mathcal{E}[\xi] + b \mathcal{E}[\eta];$

2. Оно сохраняет относительный порядок: если $\xi \leq \eta$ (то есть для любого $\omega \in \Omega$ $\xi(\omega) \leq \eta(\omega)$), то $E[\xi] \leq E[\eta]$;
3. Модуль математического ожидания меньше математического ожидания модуля: $|E[\xi]| \leq E[|\xi|]$;
4. $E[\xi] = \sum_{a \in \xi(\Omega)} a P(\xi = a)$;
5. Для любой функции $\varphi(x)$ выполняется $E[\varphi(\xi)] = \sum_{a \in \xi(\Omega)} \varphi(a) P(\xi = a)$;
6. Если $P(\xi = c) = 1$ для некоторой константы c , то $E[\xi] = c$;
7. Если $\xi \geq 0$, то $E[\xi] \geq 0$;
8. Если $E[\xi] = 0$ и $\xi \geq 0$, то $P(\xi = 0) = 1$;
9. Если ξ и η независимы, то $E[\xi\eta] = E[\xi] E[\eta]$.

Доказательство. Докажем пункты по порядку:

1. $E[a\xi + b\eta] = \sum_{\omega \in \Omega} (a\xi + b\eta)(\omega) P(\omega) = \sum_{\omega \in \Omega} a\xi(\omega) P(\omega) + \sum_{\omega \in \Omega} b\eta(\omega) P(\omega) = a E[\xi] + b E[\eta]$.
2. $E[\xi] = \sum_{\omega \in \Omega} \xi(\omega) P(\omega) \leq \sum_{\omega \in \Omega} \eta(\omega) P(\omega) = E[\eta]$.
3. $E[-|\xi|] \leq E[\xi] \leq E[|\xi|] \implies -E[|\xi|] \leq E[\xi] \leq E[|\xi|] \implies |E[\xi]| \leq E[|\xi|]$
- 4.

$$\begin{aligned} E[\xi] &= \sum_{\omega \in \Omega} \xi(\omega) P(\omega) = \sum_{a \in \xi(\Omega)} \sum_{\omega: \xi(\omega)=a} \xi(\omega) P(\omega) = \sum_{a \in \xi(\Omega)} \left(a \sum_{\omega: \xi(\omega)=a} P(\omega) \right) = \\ &= \sum_{a \in \xi(\Omega)} a P(\xi = a) \end{aligned}$$

5.

$$\begin{aligned} E[\varphi(\xi)] &= \sum_{\omega \in \Omega} \varphi(\xi(\omega)) P(\omega) = \sum_{a \in \xi(\Omega)} \sum_{\omega: \xi(\omega)=a} \varphi(\xi(\omega)) P(\omega) = \\ &= \sum_{a \in \xi(\Omega)} \left(\varphi(a) \sum_{\omega: \xi(\omega)=a} P(\omega) \right) = \sum_{a \in \xi(\Omega)} \varphi(a) P(\xi = a) \end{aligned}$$

6.

$$\begin{aligned} E[\xi] &= \sum_{\omega \in \Omega} \xi(\omega) P(\omega) = \sum_{\omega: \xi(\omega)=c} \xi(\omega) P(\omega) + \sum_{\omega: \xi(\omega) \neq c} \xi(\omega) P(\omega) = c P(\xi = c) + \\ &+ \sum_{\omega: \xi(\omega) \neq c} \xi(\omega) \cdot 0 = c. \end{aligned}$$

7. $\xi \geq 0 \implies E[\xi] \geq E[0] \geq 0$.

8. $E[\xi] = \sum_{\omega \in \Omega} \underbrace{\xi(\omega)}_{\geq 0} \underbrace{P(\omega)}_{\geq 0} = 0 \implies \forall \omega \in \Omega \xi(\omega) P(\omega) = 0$. Тогда если $\xi(\omega) \neq 0$, то $P(\omega) = 0$.

Тогда $P(\xi = 0) = 1$.

Заметим, что если величина может быть и отрицательной, то это свойство не выполняется. Предположим, что ξ равновероятно принимает значения 1 и -1 . Тогда $E[\xi] = 0$, но $P(\xi = 0) = 0$.

9.

$$\begin{aligned} E[\xi\eta] &= \sum_{\omega \in \Omega} \xi(\omega)\eta(\omega) P(\omega) = \left\{ \begin{array}{l} \xi(\Omega) = (a_1, a_2, \dots) \\ \eta(\Omega) = (b_1, b_2, \dots) \end{array} \right\} = \sum_{i,j} \sum_{\substack{\omega: \\ \xi(\omega)=a_i \\ \eta(\omega)=b_j}} \xi(\omega)\eta(\omega) P(\omega) = \\ &= \sum_{i,j} a_i b_j P(\xi = a_i, \eta = b_j) \end{aligned}$$

Так как $\xi \perp \eta$, то $P(\xi = a_i, \eta = b_j) = P(\xi = a_i) P(\eta = b_j)$ и сумма равна

$$\sum_{i,j} a_i b_j P(\xi = a_i) P(\eta = b_j) = \left(\sum_i a_i P(\xi = a_i) \right) \left(\sum_j b_j P(\eta = b_j) \right) = E[\xi] E[\eta]$$

□

Примеры использования математического ожидания:

1. Посчитаем математическое ожидание индикатора события A :
 $E[I_A] = \sum_{\omega \in \Omega} I_A(\omega) P(\omega) = \sum_{\omega \in A} P(\omega) = P(A)$.
2. Если мы рассмотрим классическую модель, т.е. такую модель, где все исходы равновероятны, то $E[\xi] = \frac{1}{|\Omega|} \sum_{\omega \in \Omega} \xi(\omega)$. Иначе говоря, математическое ожидание в классической модели равно среднему арифметическому возможных значений ξ .
3. Пусть $\xi \sim \text{Bin}(n, p)$. Тогда

$$\begin{aligned} E[\xi] &= \sum_{k=0}^n k P(\xi = k) = \sum_{k=1}^n k \binom{n}{k} p^k (1-p)^{n-k} = \\ &= \sum_{k=1}^n n \binom{n-1}{k-1} p^k (1-p)^{n-k} = np \sum_{t=0}^{n-1} \binom{n-1}{t} p^t (1-p)^{n-1-t}. \end{aligned}$$

1.3.4 Дисперсия. Ковариация. Их свойства

Определение 22. Пусть ξ — некоторая случайная величина над (Ω, P) . Тогда *дисперсией* ξ называется $D[\xi] = E[(\xi - E[\xi])^2]$.

Дисперсию случайной величины можно понимать как среднеквадратическое отклонение этой случайной величины от её среднего значения (математического ожидания).

Примечание. Вероятность незначительного отклонения величины от математического ожидания, то есть $P(E[\xi] - 2\sqrt{D[\xi]} \leq \xi \leq E[\xi] + 2\sqrt{D[\xi]})$, всегда велика! Это нам пока не пригодится, но на будущее стоит запомнить.

Определение 23. Пусть ξ и η — две случайные величины. Тогда *ковариацией* этих величин называется $\text{cov}(\xi, \eta) = E[(\xi - E[\xi])(\eta - E[\eta])]$.

Ковариацию стоит воспринимать как меру зависимости двух случайных величин.

Определение 24. Две случайные величины ξ и η называют *некоррелированными*, если $\text{cov}(\xi, \eta) = 0$.

Теорема 4 (Простейшие свойства дисперсии и ковариации). Пусть ξ, η и χ — некоторые случайные величины. Тогда выполняются следующие свойства:

1. Ковариация билинейна: $\text{cov}(\xi, a\eta + b\chi) = a \text{cov}(\xi, \eta) + b \text{cov}(\xi, \chi)$;
2. $D[\xi] = \text{cov}(\xi, \xi)$;
3. Для любого $c \in \mathbb{R}$ верно, что $D[c\xi] = c^2 D[\xi]$ и $D[\xi + c] = D[\xi]$;
4. Дисперсия неотрицательна: $D[\xi] \geq 0$. При этом $D[\xi] = 0 \iff P(\xi = E[\xi]) = 1$;
5. Если ξ и η некоррелированные, то $D[\xi + \eta] = D[\xi] + D[\eta]$;
6. Связь с матожиданием: $D[\xi] = E[\xi^2] - (E[\xi])^2$, $\text{cov}(\xi, \eta) = E[\xi\eta] - E[\xi]E[\eta]$

Доказательство. Докажем все пункты по порядку:

1. По определению ковариации $\text{cov}(\xi, a\eta + b\chi) = E[(\xi - E[\xi])(a\eta + b\chi - E[a\eta + b\chi])]$. Распишем вторую скобку:

$$(a\eta + b\chi) - E[a\eta + b\chi] = (a\eta + b\chi) - aE[\eta] - bE[\chi] = a(\eta - E[\eta]) + b(\chi - E[\chi]).$$

Тогда

$$\begin{aligned} \text{cov}(\xi, a\eta + b\chi) &= E[(\xi - E[\xi])(a(\eta - E[\eta]) + b(\chi - E[\chi]))] \\ &= E[a(\xi - E[\xi])(\eta - E[\eta]) + b(\xi - E[\xi])(\chi - E[\chi])] \\ &= aE[(\xi - E[\xi])(\eta - E[\eta])] + bE[(\xi - E[\xi])(\chi - E[\chi])] \\ &= a \text{cov}(\xi, \eta) + b \text{cov}(\xi, \chi). \end{aligned}$$

2. $\text{cov}(\xi, \xi) = E[(\xi - E[\xi])(\xi - E[\xi])] = E[(\xi - E[\xi])^2] = D[\xi]$.
3. Для произведения $D[c\xi] = \text{cov}(c\xi, c\xi) = c^2 \text{cov}(\xi, \xi) = c^2 D[\xi]$. Для суммы $D[\xi + c] = E[(\xi + c - E[\xi + c])^2] = E[(\xi + c - E[\xi] - c)^2] = E[(\xi - E[\xi])^2] = D[\xi]$.
4. По определению дисперсии, $D[\xi] = E[(\xi - E[\xi])^2]$, а $(\xi - E[\xi])^2 \geq 0$. Тогда $E[(\xi - E[\xi])^2] \geq 0$.
5. Воспользуемся выражением дисперсии через ковариацию и билинейностью ковариации: $D[\xi + \eta] = \text{cov}(\xi + \eta, \xi + \eta) = \text{cov}(\xi, \xi) + \text{cov}(\eta, \eta) + 2 \text{cov}(\xi, \eta)$. Так как ξ и η некоррелированные, то $\text{cov}(\xi, \eta) = 0$ и $D[\xi + \eta] = \text{cov}(\xi, \xi) + \text{cov}(\eta, \eta) = D[\xi] + D[\eta]$.
6. $\text{cov}(\xi, \eta) = E[(\xi - E[\xi])(\eta - E[\eta])] = E[\xi\eta - \xi E[\eta] - \eta E[\xi] + E[\xi]E[\eta]]$. По линейности матожидания $\text{cov}(\xi, \eta) = E[\xi\eta] - 2E[\xi]E[\eta] + E[\xi]E[\eta] = E[\xi\eta] - E[\xi]E[\eta]$. Для дисперсии воспользуемся тем, что $D[\xi] = \text{cov}(\xi, \xi)$.

□

Теорема 5. Если случайные величины независимы, то они некоррелированные. Обратное, вообще говоря, неверно.

Доказательство. Докажем, что из независимости следует некоррелируемость. Пусть ξ и η — независимые случайные величины. Тогда $\text{cov}(\xi, \eta) = E[\xi\eta] - E[\xi]E[\eta]$. Но по свойству математического ожидания из независимости случайных величин следует, что $E[\xi\eta] = E[\xi]E[\eta]$. Тогда $\text{cov}(\xi, \eta) = 0$ и эти величины некоррелированные.

Теперь покажем, что из некоррелированности не обязательно следует независимость. Пусть случайная величина ξ равномерно принимает значения из множества $\{0, 1, -1\}$. Возьмем случайную величину $\eta = \xi^2$. По определению можно проверить, что величины η и ξ некоррелированные: $\text{cov}(\xi, \eta) = E[\xi\eta] - E[\xi]E[\eta] = E[\xi^3] - E[\xi]E[\xi^2] = 0 - 0 = 0$. Но ξ и η не являются независимыми, что проверяется опять же по определению: $P(\xi = 0, \eta = 0) = \frac{1}{3} \neq \frac{1}{9} = P(\xi = 0)P(\eta = 0)$. \square

1.3.5 Неравенства Маркова и Чебышёва

Под конец лекции обсудим два неравенства, которые сами по себе являются весьма полезными.

Теорема 6 (Неравенство Маркова). Пусть $\xi \geq 0$ — неотрицательная случайная величина на (Ω, P) . Тогда для любого положительного a

$$P(\xi \geq a) \leq \frac{E[\xi]}{a}.$$

Доказательство. Как было доказано ранее, математическое ожидание индикатора события равно вероятности события. Тогда $P(\xi \geq a) = E[I_{\xi \geq a}]$. Далее, заметим, что $I_{\xi \geq a} \leq \frac{\xi}{a} I_{\xi \geq a}$. Тогда

$$P(\xi \geq a) \leq E\left[\frac{\xi}{a} I_{\xi \geq a}\right] \leq E\left[\frac{\xi}{a}\right] = \frac{E[\xi]}{a}. \quad \square$$

Теорема 7 (Неравенство Чебышёва). Пусть ξ — случайная величина на (Ω, P) такая, что $D[\xi] < \infty$. Тогда для любого положительного a

$$P(|\xi - E[\xi]| \geq a) \leq \frac{D[\xi]}{a^2}.$$

Доказательство. $P(|\xi - E[\xi]| \geq a) = P((\xi - E[\xi])^2 \geq a^2)$. По неравенству Маркова

$$P((\xi - E[\xi])^2 \geq a^2) \leq \frac{E[(\xi - E[\xi])^2]}{a^2} = \frac{D[\xi]}{a^2}. \quad \square$$

1.4 Лекция от 30.09.2016

1.4.1 Сходимость случайных величин. Закон больших чисел.

Сначала введём пару определений:

Определение 25. Пусть $\{\xi_n\}_{n=1}^{\infty}$ — последовательность случайных величин, заданных на вероятностном пространстве (Ω, P) . Тогда будем говорить, что последовательность

сходится по вероятности к случайной величине ξ , если для любого положительного α выполнено, что

$$\lim_{n \rightarrow \infty} P(|\xi_n - \xi| \geq \alpha) = \lim_{n \rightarrow \infty} P(\{\omega : |\xi_n(\omega) - \xi(\omega)| \geq \alpha\}) = 0.$$

Обозначение: $\xi_n \xrightarrow{P} \xi$.

Определение 26. Случайные величины ξ и η называют *одинаково распределёнными*, если

$$\begin{cases} \xi(\Omega) = \eta(\Omega); \\ \forall x \in \xi(\Omega), P(\xi = x) = P(\eta = x). \end{cases}$$

Говоря простым языком, у этих случайных величин совпадают возможные значения и их вероятности.

Обозначение: $\xi \stackrel{d}{=} \eta$.

Теорема 8 (Закон больших чисел). Пусть есть последовательность попарно независимых и одинаково распределённых случайных величин $\{\xi_n\}_{n=1}^{\infty}$ и существует число C такое, что $D[\xi_n] \leq C$ для всех $n \in \mathbb{N}$.

Введём обозначение $S_n = \xi_1 + \xi_2 + \dots + \xi_n$. Тогда

$$\frac{S_n - E[S_n]}{n} \xrightarrow{P} 0.$$

Доказательство. Для начала заметим, что $E\left[\frac{S_n - E[S_n]}{n}\right] = 0$. Тогда по неравенству Чебышёва выполнено, что:

$$P\left(\left|\frac{S_n - E[S_n]}{n}\right| \geq \varepsilon\right) \leq \frac{D\left[\frac{S_n - E[S_n]}{n}\right]}{\varepsilon^2}.$$

Так как $D\left[\frac{S_n - E[S_n]}{n}\right] = \frac{1}{n^2} D[S_n]$ для фиксированного n , то

$$P\left(\left|\frac{S_n - E[S_n]}{n}\right| \geq \varepsilon\right) \leq \frac{D[S_n]}{n^2 \varepsilon^2}.$$

Распишем дисперсию в виде суммы ковариаций:

$$D[S_n] = \text{cov}(\xi_1 + \dots + \xi_n, \xi_1 + \dots + \xi_n) = \sum_{i,j=1}^n \text{cov}(\xi_i, \xi_j).$$

Если $i \neq j$, то случайные величины ξ_i и ξ_j независимы. Но тогда они некоррелированы и $\text{cov}(\xi_i, \xi_j) = 0$. Следовательно,

$$D[S_n] = \sum_{i=1}^n \text{cov}(\xi_i, \xi_i) \leq Cn.$$

Подставляя выражение дисперсии, получаем, что

$$P\left(\left|\frac{S_n - E[S_n]}{n}\right| \geq \varepsilon\right) \leq \frac{C}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0.$$

□

Примечание. Вообще, в законе больших чисел не обязательна независимость случайных величин — достаточно попарной некоррелированности.

Следствие. Пусть $\{\xi_n\}_{n=1}^\infty$ — последовательность попарно независимых случайных величин. Пусть существуют α и C такие, что $E[\xi_n] = \alpha$ и $D[\xi_n] \leq C$ для всех $n \in \mathbb{N}$. Тогда

$$\frac{\xi_1 + \dots + \xi_n}{n} \xrightarrow{P} \alpha.$$

Доказательство. Из условия независимости случайных величин нам хватит следствия о том, что они являются попарно некоррелированными. Из закона больших чисел следует, что $\frac{S_n - E[S_n]}{n} \xrightarrow{P} 0$. Так как $E[S_n] = n\alpha$, то $\frac{S_n}{n} - \alpha \xrightarrow{P} 0$. Также заметим, что в силу определения сходимости, $\frac{S_n}{n} \xrightarrow{P} \alpha$. \square

Давайте попробуем осознать, в чем заключается смысл Закона больших чисел. Пусть случайная величина задается индикатором происхождения события, $\xi_i = I\{\text{событие } A \text{ произошло в } i\text{-ом эксперименте}\}$, при этом все эксперименты проводятся независимо друг от друга (ξ_n является независимой случайной величиной). Обозначим за $\vartheta(A)$ частоту появления события A в n -м количестве экспериментов, $\vartheta_n(A) = \frac{\xi_1 + \dots + \xi_n}{n}$. Тогда по Закону больших чисел $\vartheta_n(A) \xrightarrow{P} E[\xi_i] = E[I\{\text{событие } A \text{ произошло в } i\text{-ом эксперименте}\}] = P(A)$. Выходит, что Закон больших чисел является теоретическим обоснованием принципа устойчивости частот.

Определение 27. Последовательность случайных величин $\{\xi_n\}_{n=1}^\infty$ сходится к случайной величине ξ с вероятностью 1 (или *почти наверное*⁸), если

$$P\left(\lim_{n \rightarrow \infty} \xi_n = \xi\right) = P\left(\left\{\omega : \lim_{n \rightarrow \infty} \xi_n(\omega) = \xi(\omega)\right\}\right) = 1.$$

Обозначение: $\xi_n \xrightarrow{\text{п.н.}} \xi$ (сходится почти наверное).

Примечание. Данное определение можно записать следующим образом: для любого $\omega \in \Omega$ такого, что $P(\omega) > 0$, выполнено $\xi_n(\omega) \rightarrow \xi(\omega)$.

Теорема 9. В дискретных вероятностных пространствах $\xi_n \xrightarrow{P} \xi \iff \xi_n \xrightarrow{\text{п.н.}} \xi$.

Доказательство. Докажем по очереди в обе стороны:

$[\Rightarrow]$ Допустим, что $\xi_n \xrightarrow{P} \xi$. Докажем, что тогда для любого $\omega \in \Omega$, для которого $P(\omega) > 0$, выполнено $\xi_n(\omega) \rightarrow \xi(\omega)$ (стремится в понимании предела последовательности).

Пусть есть элементарный исход ω_0 такой, что сходимости нет: $\xi_n(\omega_0) \not\rightarrow \xi(\omega_0)$. В таком случае существует положительное α и подпоследовательность $\{\xi_{n_k}(\omega_0)\}_{k=1}^\infty$ такие, что $\forall k \in \mathbb{N} |\xi_{n_k}(\omega_0) - \xi(\omega_0)| \geq \alpha$. Но тогда $P(|\xi_{n_k} - \xi| \geq \alpha) \geq P(\omega_0) > 0$. Но по определению сходимости по вероятности $P(|\xi_{n_k} - \xi| \geq \alpha) \rightarrow 0$. Получаем противоречие. Следовательно, $\xi_n \xrightarrow{\text{п.н.}} \xi$.

$[\Leftarrow]$ Пусть $\xi_n \xrightarrow{\text{п.н.}} \xi$. Согласно определению, $P(\lim_{n \rightarrow \infty} \xi_n = \xi) = 1$. Это равносильно тому, что для любого $\omega \in \Omega$, для которого $P(\omega) > 0$, выполнено $\xi_n(\omega) \rightarrow \xi(\omega)$.

⁸Название “сходимость почти наверное” произошло от применяющегося в англоязычной литературе термина “almost sure convergence”.

Докажем, что выполнено условие сходимости *по вероятности*: $P(|\xi_n - \xi| \geq \alpha) \rightarrow 0$ для любого положительного α . Пусть $\Omega = \{\omega_i \mid i \in \mathbb{N}\}$. Зафиксируем произвольное $\delta > 0$ и выберем номер $N = N(\delta)$ такой, что $\sum_{i>N} P(\omega_i) \leq \delta$. Тогда

$$\begin{aligned} P(|\xi_n - \xi| \geq \alpha) &= \sum_{i: |\xi_n(\omega_i) - \xi(\omega_i)| \geq \alpha} P(\omega_i) \\ &= \sum_{\substack{i: i \leq N \\ |\xi_n(\omega_i) - \xi(\omega_i)| \geq \alpha}} P(\omega_i) + \sum_{\substack{i: i > N \\ |\xi_n(\omega_i) - \xi(\omega_i)| \geq \alpha}} P(\omega_i) \\ &\leq \sum_{\substack{i: i \leq N \\ |\xi_n(\omega_i) - \xi(\omega_i)| \geq \alpha}} P(\omega_i) + \sum_{i>N} P(\omega_i) \\ &\leq \sum_{\substack{i: i \leq N \\ |\xi_n(\omega_i) - \xi(\omega_i)| \geq \alpha}} P(\omega_i) + \delta. \end{aligned}$$

Теперь нужно перейти к пределу при $n \rightarrow \infty$. Но чтобы избежать ситуации, когда предела не существует, перейдем к верхнему пределу для тех же n . Также заметим, что так как $\xi_n(\omega_i) \rightarrow \xi(\omega_i)$, начиная с некоторого n , ω_i перестает входить в сумму. Тогда первая сумма стремится к нулю и

$$\overline{\lim}_{n \rightarrow \infty} P(|\xi_n - \xi| \geq \alpha) \leq 0 + \delta.$$

В силу произвольности $\delta > 0$ можно говорить о том, что этот предел стремится к нулю. \square

Допустим, что $S_n \sim \text{Bin}(n, p)$. Тогда сформулируем два вопроса:

Вопрос 1: Как оценить скорость сходимости $\frac{S_n}{n} - a$ к нулю?

Вопрос 2: Как оценить скорость сходимости $P\left(\left|\frac{S_n}{n} - a\right| \geq \varepsilon\right)$ к нулю?

1.4.2 Схема испытаний Бернулли

Описание: Схема испытаний Бернулли подразумевает много однородных экспериментов, в каждом из которых возможно два исхода — “успех” и “неудача”; при этом нас интересует распределение числа успехов.

Математическая модель: Пусть $\{\xi_i\}$ — независимые *одинаково распределённые* случайные величины (например, последовательные броски монеты); $\xi_i(\Omega) = \{0, 1\}$; пусть также $P(\xi_i = 1) = p$ — вероятность успеха каждого отдельного эксперимента. Введём также обозначение $S_n = \sum_{i=1}^n \xi_i$ — случайную величину, обозначающую количество успехов.

Распределение S_n кажется очевидным: $S_n \sim \text{Bin}(n, p)$, т.е. $P(S_n = k) = C_n^k p^k (1-p)^{n-k}$.

Упражнение. Докажите, что распределение S_n и вправду биномиально.

Упражнение. В качестве упражнения, читателю предлагается оценить, сколько нужно попыток, чтобы $\frac{\text{число успехов}}{\text{число попыток}}$ было примерно равно p , т.е. как быстро наблюдаемая вероятность сходится к реальной. Для этого предлагается оценить следующую вероятность:

$$P\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) = ?$$

Кстати говоря, p совершенно не обязана быть константой и может являться функцией от n : $p(n)$.

Тогда, вообще говоря, есть несколько отдельных случаев:

- $np(n) \xrightarrow{n \rightarrow \infty} \lambda > 0$.

Теорема 10 (Пуассон). Пусть $np(n) \xrightarrow{n \rightarrow \infty} \lambda > 0$. Тогда

$$\forall k \in \mathbb{Z}_+, P(S_n = k) \xrightarrow{n \rightarrow \infty} e^{-\lambda} \frac{\lambda^k}{k!}.$$

Доказательство.

$$\begin{aligned} C_n^k p^k (1-p)^{n-k} &= \frac{n(n-1)(n-2)\dots(n-k+1)}{k!} p(n)^k (1-p(n))^{n-k} \sim \\ &\sim \frac{n^k}{k!} \left(\frac{\lambda}{n} + o\left(\frac{1}{n}\right) \right)^k \left(1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right) \right)^{n-k} = \\ &= \frac{1}{k!} (\lambda + o(1))^k \left(1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right) \right)^{n-k} \xrightarrow{n \rightarrow \infty} \frac{\lambda^k}{k!} e^{-\lambda} \end{aligned}$$

□

Следствие: $P(S_n = k) \xrightarrow{d} P(\xi = k)$, $\xi \sim \text{Pois}(\lambda)$ — сходимость по распределению.

- $np(n) \xrightarrow{n \rightarrow \infty} \infty$.

Сделаем ещё одно отступление в сторону и рассмотрим величину $\frac{S_n - \mathbb{E}[S_n]}{\sqrt{D[S_n]}}$. Пусть ξ_1, \dots, ξ_n — n бернуллиевских случайных величин с $P(\xi_i = 1) = p$, $P(\xi_i = 0) = q$, $p + q = 1$. Тогда для $S_n = \xi_1 + \dots + \xi_n$ находим, что

$$\mathbb{E}[S_n] = np.$$

Несложно также показать, что

$$\begin{aligned} D[S_n] &= \sum D[\xi_i] + \sum \text{cov}[x_i, x_j] = \sum D[\xi_i] = n D[\xi_1] = \\ &= n (\mathbb{E}[\xi_1^2] - (\mathbb{E}[\xi_1])^2) = n(p - p^2) = npq. \end{aligned}$$

Теорема 11 (Муавра-Лапласа). Пусть вероятность p того, что во время эксперимента произойдёт событие, постоянна и находится в интервале $(0, 1)$. Допустим, что было проведено n независимых испытаний. Обозначим за S_n количество испытаний, в которых произошло событие (это равносильно тому, что $S_n \sim \text{Bin}(n, p)$). Введём следующее обозначение:

$$P_n(a, b) := P \left(a < \frac{S_n - \mathbb{E}[S_n]}{\sqrt{D[S_n]}} \leq b \right).$$

Тогда

$$\sup_{a < b} \left| P_n(a, b) - \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \right| \xrightarrow{n \rightarrow \infty} 0.$$

1.5 Лекция от 07.10.2016

1.5.1 Скорость сходимости закона больших чисел

Ранее мы формулировали два вопроса, связанные со скоростью сходимости закона больших чисел. Ответим на них.

1. В первом пункте от нас требуется найти скорость сходимости $\frac{S_n}{n} - p$ к нулю, где $S_n \sim \text{Bin}(n, p)$. Для этого вспомним теорему Муавра-Лапласа:

$$\mathbb{P}\left(a < \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathbb{D}[S_n]}} \leq b\right) \rightarrow \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = c, c \in (0, 1).$$

Преобразуем условие вероятности, пользуясь тем, что $\mathbb{E}[S_n] = np$, а $\mathbb{D}[S_n] = np(1-p)$:

$$a < \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathbb{D}[S_n]}} \leq b \implies \frac{a\sqrt{p(1-p)}}{\sqrt{n}} < \frac{S_n}{n} - p \leq \frac{b\sqrt{p(1-p)}}{\sqrt{n}}.$$

Возьмём $\delta \in (0, 1)$ такое, что $1 - \delta < c$. Тогда при достаточно больших n выполнено (по определению предела):

$$\mathbb{P}\left(\frac{c_1}{\sqrt{n}} < \frac{S_n}{n} - p \leq \frac{c_2}{\sqrt{n}}\right) > 1 - \delta.$$

Увеличивая расстояние между a и b и увеличивая n , получаем следующий результат:

$$\lim_{n, |a-b| \rightarrow \infty} \mathbb{P}\left(\frac{S_n}{n} - p = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)\right) = 1.$$

Отсюда получаем, что $\frac{S_n}{n} = p + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$.

2. Теперь приступим ко второму вопросу. Для него достаточно применить неравенство Чебышёва:

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{\mathbb{D}\left[\frac{S_n}{n}\right]}{\varepsilon^2} = \frac{np(1-p)}{n^2\varepsilon^2} = \mathcal{O}\left(\frac{1}{n}\right).$$

1.5.2 Неравенство Чернова

Следующая теорема показывает то, насколько вероятно отклонение случайной величины от её математического ожидания.

Теорема 12 (Неравенство Чернова). Пусть $S_n \sim \text{Bin}(n, p)$, а $\lambda = \mathbb{E}[S_n] = np$. Тогда для любого $t > 0$ выполнено следующее:

$$\mathbb{P}(S_n \geq \lambda + t) \leq \exp\left(-\frac{t^2}{2(\lambda + \frac{t}{3})}\right), \quad \mathbb{P}(S_n \leq \lambda - t) \leq \exp\left(-\frac{t^2}{2\lambda}\right).$$

Доказательство. Крайний случай для первого неравенства — $t = n - \lambda$ — особого интереса не вызывает. Поэтому положим, что $t < n - \lambda$.

Заметим, что для любого положительного u выполнено следующее:

$$P(S_n \geq \lambda + t) = P(e^{uS_n} \geq e^{u(\lambda+t)}) \leq \{\text{по неравенству Маркова}\} \leq \frac{E[e^{uS_n}]}{e^{u(\lambda+t)}}.$$

Посчитаем $E[e^{uS_n}]$:

$$E[e^{uS_n}] = \sum_{k=1}^n e^{uk} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^n \binom{n}{k} (pe^u)^k (1-p)^{n-k} = (1-p+pe^u)^n.$$

Положим $x = e^u$. Тогда

$$P(S_n \geq \lambda + t) \leq x^{-(\lambda+t)} (1-p+px)^n.$$

Минимизируем это выражение по x . Для этого посчитаем производную и приравняем её к нулю:

$$(x^{-(\lambda+t)} (1-p+px)^n)' = -(\lambda+t)x^{-(\lambda+t+1)} (1-p+px)^n + npx^{-(\lambda+t)} (1-p+px)^{n-1} = 0.$$

Отсюда получаем, что $-(\lambda+t)(1-p+px) + npx = 0$ и $x = \frac{(\lambda+t)(1-p)}{p(n-\lambda-t)}$. Тогда получаем следующую оценку сверху:

$$\begin{aligned} P(S_n \geq \lambda + t) &\leq \left(\frac{p(n-\lambda-t)}{(\lambda+t)(1-p)} \right)^{\lambda+t} \left(1-p+p \frac{(\lambda+t)(1-p)}{p(n-\lambda-t)} \right)^n \\ &= \left(\frac{p(n-\lambda-t)}{(\lambda+t)(1-p)} \right)^{\lambda+t} \left(\frac{n(1-p)}{n-\lambda-t} \right)^n \\ &= \left(\frac{np}{\lambda+t} \right)^{\lambda+t} \left(\frac{n-np}{n-\lambda-t} \right)^{n-\lambda-t} \\ &= \left(\frac{\lambda}{\lambda+t} \right)^{\lambda+t} \left(\frac{n-\lambda}{n-\lambda-t} \right)^{n-\lambda-t} \end{aligned}$$

Представим эту оценку в следующем виде:

$$\left(\frac{\lambda}{\lambda+t} \right)^{\lambda+t} \left(\frac{n-\lambda}{n-\lambda-t} \right)^{n-\lambda-t} = \exp \left\{ -(\lambda+t) \ln \left(1 + \frac{t}{\lambda} \right) - (n-\lambda-t) \ln \left(1 - \frac{t}{n-\lambda} \right) \right\}.$$

Введём функцию $\varphi(x) = (1+x) \ln(1+x) - x$ для $x > -1$. Тогда (проверьте):

$$P(S_n \geq \lambda + t) \leq \exp \left\{ -\lambda \varphi \left(\frac{t}{\lambda} \right) - (n-\lambda) \varphi \left(-\frac{t}{n-\lambda} \right) \right\}.$$

Теперь заметим, что если $x > -1$, то $\varphi(x) > 0$. Тогда будет верно следующее (вторую оценку можно получить абсолютно аналогично первой, только взяв $-u$ вместо u):

$$\begin{aligned} P(S_n \geq \lambda + t) &\leq \exp \left\{ -\lambda \varphi \left(\frac{t}{\lambda} \right) \right\}, \\ P(S_n \leq \lambda - t) &\leq \exp \left\{ -\lambda \varphi \left(\frac{-t}{\lambda} \right) \right\}. \end{aligned}$$

А теперь достаём бубен и начинаем оптимизировать:

[\leq] Заметим, что $\varphi(0) = 0$ и $\varphi'(x) = 1 - 1 + \ln(x+1) = \ln(x+1) \leq x$. Тогда

$$-\varphi(y) = \int_y^0 \varphi'(x) dx \leq \int_y^0 x dx = -\frac{y^2}{2}.$$

Отсюда получаем, что $\varphi(y) \geq \frac{y^2}{2}$ и $P(S_n \leq \lambda - t) \leq \exp \left\{ -\frac{t^2}{2\lambda} \right\}$.

[\geq] Далее идёт полное шаманство (sic!). Заметим, что

$$\varphi''(x) = \frac{1}{1+x} \geq \frac{1}{\left(1 + \frac{x}{3}\right)^3} = \left(\frac{x^2}{2\left(1 + \frac{x}{3}\right)} \right)'' \implies \varphi(x) \geq \frac{x^2}{2\left(1 + \frac{x}{3}\right)}.$$

Отсюда получаем, что $P(S_n \geq \lambda + t) \leq \exp \left\{ -\frac{t^2}{2\left(\lambda + \frac{t}{3}\right)} \right\}$.

□

К счастью, на экзамене будет достаточно понимать идею, и не требуется точное воспроизведение всех вычислений.

Теперь, используя неравенство Чернова, оценим скорость сходимости $P\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right)$ к нулю по-другому:

$$P\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) = P(|S_n - \lambda| \geq n\varepsilon) \leq \exp \left\{ -\frac{n\varepsilon^2}{2p} \right\}.$$

Заметим, что оценка через неравенство Чебышёва давала оценку $\frac{1}{\text{полином}}$, а неравенство Чернова даёт оценку $\frac{1}{\text{экспонента}}$, что сходится гораздо быстрее.

1.5.3 Алгебры событий

Определение 28. Пусть \mathcal{A} — система событий на (Ω, P) . Она называется *алгеброй*, если

1. $\Omega \in \mathcal{A}$,
2. Если $A \in \mathcal{A}$, то и $\bar{A} \in \mathcal{A}$,
3. Если $A \in \mathcal{A}$ и $B \in \mathcal{A}$, то и $A \cap B \in \mathcal{A}$.

Упражнение. Докажите, что алгебра замкнута по основным операциям: \cup , \setminus , Δ .

Примеры:

- $\{\emptyset, \Omega\}$;
- 2^Ω ;
- $\{\emptyset, \Omega, A, \bar{A}\}$ — алгебра, порождённая A ;
- $\alpha(A_1, \dots, A_n)$ — минимальная алгебра, содержащая A_1, \dots, A_n .

1.6 Лекция от 14.10.2016

1.6.1 Локальная лемма Ловаса: несимметричный и симметричный случаи

Предупреждение: в данном случае мы решили несколько изменить доказательство, отойдя от сигма-алгебр к графам зависимости. Идейно эти принципы ничем не отличаются, но на коллоквиуме лучше рассказывать через алгебры событий.

Введём понятие *графа зависимости*.

Определение 29. *Графом зависимости* $D = (V, E)$ для событий A_1, A_2, \dots, A_n называют ориентированный граф такой, что $V = \{1, 2, \dots, n\}$, а для любого $1 \leq i \leq n$ событие A_i независимо от событий A_j , а также от всевозможных объединений и пересечений данных событий, если $(i, j) \notin E$.

Теорема 13 (Несимметричный случай локальной леммы Ловаса). *Пусть A_1, \dots, A_n — события на произвольном вероятностном пространстве (Ω, \mathbf{P}) . Предположим, что*

1. $D = (V, E)$ является орграфом зависимости для определенных выше событий;
2. Существуют действительные числа $x_1, \dots, x_n \in [0, 1)$ такие, что для любого $1 \leq i \leq n$ выполняется

$$\mathbf{P}(A_i) \leq x_i \prod_{(i,j) \in E} (1 - x_j).$$

Тогда вероятность того, что ни одно из событий A_1, \dots, A_n не произойдёт, можно ограничить снизу положительным числом:

$$\mathbf{P}\left(\bigcap_{i=1}^n \overline{A_i}\right) \geq \prod_{i=1}^n (1 - x_i) > 0.$$

Перед тем, как приступить к доказательству этой леммы, докажем следующую формулу:

Теорема 14 (Формула умножения вероятностей). *Пусть B_1, B_2, \dots, B_n — события на вероятностном пространстве (Ω, \mathbf{P}) . Тогда верно следующее:*

$$\mathbf{P}(B_1 \cap \dots \cap B_m) = \mathbf{P}(B_1 | B_2 \cap \dots \cap B_m) \mathbf{P}(B_2 | B_3 \cap \dots \cap B_m) \dots \mathbf{P}(B_{m-1} | B_m) \mathbf{P}(B_m).$$

Доказательство. Будем последовательно применять определение условной вероятности:

$$\begin{aligned} \mathbf{P}(B_1 \cap \dots \cap B_m) &= \mathbf{P}(B_1 | B_2 \cap \dots \cap B_m) \mathbf{P}(B_2 \cap \dots \cap B_m) \\ \mathbf{P}(B_2 \cap \dots \cap B_m) &= \mathbf{P}(B_2 | B_3 \cap \dots \cap B_m) \mathbf{P}(B_3 \cap \dots \cap B_m) \\ &\dots \\ \mathbf{P}(B_{m-1} \cap B_m) &= \mathbf{P}(B_{m-1} | B_m) \mathbf{P}(B_m). \end{aligned}$$

Постепенными подстановками получим желаемое. □

Теперь перейдём к доказательству леммы.

Доказательство. Разобьём доказательство на подпункты.

1. Докажем индукцией по t , что $\forall i = 1, \dots, n$, для любого набора чисел $T \in \{1, \dots, n\} \setminus \{i\}$, $t = |T| < n$, справедливо неравенство $P(A_i \mid \bigcap_{j \in T} \overline{A_j}) \leq x_i$.

Пусть $t = 1$. Тогда неравенство примет следующий вид:

$$P(A_i \mid \overline{A_j}) = \begin{cases} P(A_i), & (i, j) \notin E \\ \frac{P(A_i \cap \overline{A_j})}{P(\overline{A_j})} \leq \frac{P(A_i)}{P(\overline{A_j})} \leq \frac{x_i(1-x_j)}{1-x_j}, & (i, j) \in E \end{cases}$$

Так как по условию $P(A_i) \leq x_i \prod_{(i,j) \in E} (1-x_j) \leq x_i(1-x_j)$, вероятность $P(\overline{A_j}) = 1 - P(A_j) \geq 1 - x_j$.

Также, если $T = \emptyset$, из условия следует, что $P(A_i) \leq x_i$.

Предполагая, что неравенство выполняется для всех $t' < t$, докажем для его для t . Положим $T_1 = \{j \in T, (i, j) \in E\}$, $T_2 = T \setminus T_1$. Тогда:

$$P\left(A_i \mid \bigcap_{j \in T} \overline{A_j}\right) = \frac{P\left(A_i \cap \left(\bigcap_{j \in T_1} \overline{A_j}\right) \mid \bigcap_{k \in T_2} \overline{A_k}\right)}{P\left(\bigcap_{j \in T_1} \overline{A_j} \mid \bigcap_{k \in T_2} \overline{A_k}\right)} \leq \frac{P\left(A_i \mid \bigcap_{k \in T_2} \overline{A_k}\right)}{P\left(\bigcap_{j \in T_1} \overline{A_j} \mid \bigcap_{k \in T_2} \overline{A_k}\right)}.$$

Для оценки числителя заметим, что событие A_i взаимно независимо с событиями $\{A_k : k \in T_2\}$. Следовательно, $P\left(A_i \mid \bigcap_{k \in T_2} \overline{A_k}\right) = P(A_i) \leq x_i \prod_{(i,j) \in E} (1-x_j)$.

Предположим $T_1 = \{j_1, \dots, j_r\}$. При $r = 0$ неравенство $P\left(A_i \mid \bigcap_{j \in T} \overline{A_j}\right) \leq x_i$ заведомо выполняется, так как в нашем графе нет ребер, связывающих вершину i с другими вершинами, а значит, событие A_i не зависит ни от одного события A_j , где $j \in T$.

Для оценки знаменателя применим предположение индукции. В случаях, когда $r > 0$, используя формулу умножения вероятностей, получаем

$$\begin{aligned} P\left(\bigcap_{j \in T_1} \overline{A_j} \mid \bigcap_{k \in T_2} \overline{A_k}\right) &= P\left(\overline{A_{j_1}} \mid \bigcap_{s=2}^r \overline{A_{j_s}} \cap \bigcap_{k \in T_2} \overline{A_k}\right) P\left(\overline{A_{j_2}} \mid \bigcap_{s=3}^r \overline{A_{j_s}} \cap \bigcap_{k \in T_2} \overline{A_k}\right) \dots \\ &\dots P\left(\overline{A_{j_r}} \mid \bigcap_{k \in T_2} \overline{A_k}\right) \geq (1-x_{j_1}) \dots (1-x_{j_r}) \geq \prod_{(i,j) \in E} (1-x_j). \end{aligned}$$

$$\text{Следовательно, } P\left(A_i \mid \bigcap_{j \in T} \overline{A_j}\right) \leq \frac{x_i \prod_{(i,j) \in E} (1-x_j)}{\prod_{(i,j) \in E} (1-x_j)} = x_i.$$

2. Запишем окончательный результат рассуждений выше с использованием формулы полной вероятности и полученной в первом пункте оценки.

$$P\left(\bigcap_{i=1}^n \overline{A_i}\right) = P\left(\overline{A_1} \mid \bigcap_{i=2}^n \overline{A_i}\right) P\left(\overline{A_2} \mid \bigcap_{i=3}^n \overline{A_i}\right) \dots P(\overline{A_n}) \geq (1-x_1)(1-x_2) \dots (1-x_n).$$

□

Упражнение. Сформулируйте и докажите несимметрический случай локальной леммы Ловаса, используя алгебры событий вместо графа зависимости.

Однако на практике гораздо чаще используется более слабая версия леммы.

Теорема 15 (Симметричный случай локальной леммы Ловаса). Пусть A_1, A_2, \dots, A_n — события в произвольном вероятностном пространстве (Ω, \mathcal{P}) . Если выполнены следующие условия:

1. существует $p \in [0, 1)$ такое, что $\mathbb{P}(A_i) \leq p$ для любого $1 \leq i \leq n$;
2. каждое событие A_i взаимно независимо со всеми событиями за исключением не более, чем d событий;
3. $ep(d+1) \leq 1$,

то

$$\mathbb{P}\left(\bigcap_{i=1}^n \overline{A_i}\right) > 0.$$

Доказательство. Для начала рассмотрим экстремальный случай — $d = 0$. В таком случае все события независимы и утверждение теоремы выполнено.

Теперь перейдём к общему случаю. Рассмотрим граф зависимости $D = (V, E)$ для событий A_1, A_2, \dots, A_n , в котором для любого $i \in \{1, 2, \dots, n\}$ выполнено $|\{j \mid (i, j) \in E\}| \leq d$. Далее, для всех i положим $x_i = \frac{1}{d+1}$. Покажем, что в этом случае выполняется требование несимметричного случая локальной леммы Ловаса:

$$x_i \prod_{(i,j) \in E} (1 - x_j) \geq \frac{1}{d+1} \left(1 - \frac{1}{d+1}\right)^d \geq \frac{1}{e(d+1)} \geq p \geq \mathbb{P}(A_i).$$

Как видно, оно выполняется. Тогда по локальной лемме получаем, что

$$\mathbb{P}\left(\bigcap_{i=1}^n \overline{A_i}\right) > 0.$$

□

1.6.2 Задача k-SAT. Теорема Мозеса-Тардоша

Сейчас мы обсудим так называемую задачу k-SAT и её методы решения, основанные на вероятностном подходе. Но начнём мы с формулировки задачи.

Определение 30. Пусть есть некоторая булева формула $f(x_1, x_2, \dots, x_n)$. Тогда k -КНФ называют представление этой формулы в виде

$$f(x_1, x_2, \dots, x_n) = \bigwedge_{i=1}^m (l_{i1} \vee l_{i2} \vee \dots \vee l_{ik}),$$

где $l_{ij} \in \{x_1, \dots, x_n\} \cup \{\overline{x_1}, \overline{x_2}, \dots, \overline{x_n}\}$, причём в дизъюнкте все переменные разные (то есть нет повторов).

Задача 1 (k-SAT). Пусть есть некоторая булева функция $f(x_1, x_2, \dots, x_n)$, представленная в виде k -КНФ. Выполнима ли функция f , то есть существует ли такой набор переменных v_1, v_2, \dots, v_n , что $f(v_1, v_2, \dots, v_n) = 1$?

Докажем следующее утверждение:

Теорема 16. Если k -КНФ для функции $f(x_1, x_2, \dots, x_n)$ содержит меньше, чем 2^k дизъюнктов, то f выполнима.

Доказательство. Рассмотрим ситуацию, когда переменным равновероятно и независимо присваивается либо 0, либо 1. Тогда $P(x_i = 0) = P(x_i = 1) = \frac{1}{2}$ для $i \in \{1, 2, \dots, n\}$.

Чему равна вероятность того, что дизъюнкт $l_{i1} \vee l_{i2} \vee \dots \vee l_{ik}$ будет равен 0? Она равна

$$P(l_{i1} = 0, l_{i2} = 0, \dots, l_{ik} = 0) = P(l_{i1} = 0) P(l_{i2} = 0) \dots P(l_{ik} = 0) = \frac{1}{2^k}.$$

Теперь посмотрим вероятность того, что вся КНФ будет равна 0:

$$\begin{aligned} P(f(x_1, x_2, \dots, x_n) = 0) &= P\left(\bigcup_{i=1}^m \{l_{i1} \vee l_{i2} \vee \dots \vee l_{ik} = 0\}\right) \\ &\leq \sum_{i=1}^m P(l_{i1} \vee l_{i2} \vee \dots \vee l_{ik} = 0) = \frac{m}{2^k} < 1. \end{aligned}$$

Отсюда получаем, что вероятность того, что КНФ будет равна 1, больше нуля. А это и означает выполнимость f . \square

Теперь посмотрим на следующий вопрос. Пусть каждая переменная (или её отрицание) лежит не более, чем в d скобках. При каком d мы можем гарантировать выполнимость? Ответ на этот вопрос даёт следующая теорема:

Теорема 17. Пусть булева формула $f(x_1, x_2, \dots, x_n)$ записана в виде k -КНФ. Также пусть любая переменная или её отрицание входят в не более, чем $\frac{2^{k-1}}{ek}$ дизъюнкций. Тогда f выполнима.

Доказательство. Так же, как и в прошлой теореме, рассмотрим ситуацию, когда переменным равновероятно и независимо присваивается либо 0, либо 1. Тогда $P(x_i = 0) = P(x_i = 1) = \frac{1}{2}$ для $i \in \{1, 2, \dots, n\}$.

Рассмотрим событие $A_i = \{i\text{-й дизъюнкт равен } 0\}$. Исходя из того, что каждый дизъюнкт содержит k литералов, а также того, что все значения переменных равновероятны, мы можем оценить вероятность каждого события следующим образом:

$$P(A_i) = \begin{cases} 0, & \text{в дизъюнкте одновременно есть } y \text{ и } \bar{y}, \\ \frac{1}{2^k}, & \text{иначе.} \end{cases}$$

Отсюда видно, что $P(A_i) \leq \frac{1}{2^k}$ для всех $i \in \{1, 2, \dots, m\}$.

Теперь посмотрим на количество событий, зависящих от A_i . Для этого рассмотрим зависимость на каком-либо примере. Пусть i -й дизъюнкт равен $x_1 \vee \bar{x}_2 \vee \bar{x}_3 \vee \dots \vee \bar{x}_k$. Тогда A_i будет независимо с теми событиями, соответствующие дизъюнкты которых не содержат элементов из $\{x_1, x_2, \dots, x_k\} \cup \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k\}$.

Тогда количество событий, зависящих от A_i равно количеству дизъюнктов, совпадающих хотя бы по одной переменной (с точностью до отрицания). Обозначим его за S_i . Попробуем ограничить его сверху. Выберем произвольную позицию, которая будет общей.

Так как переменная в этой позиции входит в не более, чем $\frac{2^{k-1}}{ek}$ дизъюнкций и переменные должны совпадать с точностью до отрицания, то есть не более, чем $2k \frac{2^{k-1}}{ek} = \frac{2^k}{e}$. Однако мы посчитали и сам дизъюнкт. Поэтому

$$|S_i| \leq \frac{2^k}{e} - 1 = d.$$

Теперь покажем, что выполняется и последнее условие симметричного случая локальной леммы Ловаса: $ep(d+1) \leq \frac{e}{2^k} \frac{2^k}{e} = 1$.

Отсюда по симметричной локальной лемме Ловаса получаем, что $P(\bigcap_{i=1}^n \overline{A_i}) > 0$. Это означает, что вероятность того, что при заданном наборе переменных функция равна 1 больше нуля, что доказывает выполнимость f . \square

Из общего случая локальной леммы можно вывести следующую оценку для задачи выше.

$$P\left(\bigcap_{i=1}^m \overline{A_i}\right) \geq \prod_{i=1}^m (1 - x_i) = \left(\frac{d}{d+1}\right)^m = \left(1 - \frac{e}{2^k}\right)^m = q.$$

Теперь можно задаться вполне резонным вопросом: а как все же искать такое выполняющее означивание? Рассмотрим два подхода к решению этой проблемы.

1. *Наивный подход*: подставляем случайные значения (x_1, x_2, \dots, x_n) . Если $f = 0$, тогда независимо подставим новые значения (y_1, y_2, \dots, y_n) . Так будем повторять до победного конца.

Пусть X — число подстановок новых пар значений переменных (*resampling steps*). Легко понять, что $X \sim \text{Geom}(q)$ и⁹

$$E[X] = \sum_{k=0}^{\infty} kq(1-q)^k = \frac{1}{q} = \mathcal{O}\left(\left(\frac{2^k}{2^k - e}\right)^m\right).$$

2. *Продвинутый подход*: подставляем случайные значения (x_1, x_2, \dots, x_n) . Если $f = 0$, то смотрим, какие именно дизъюнкты равны 0. Если $l_{i1} \vee l_{i2} \vee \dots \vee l_{ik} = 0$, то меняем значения переменных $l_{i1}, l_{i2}, \dots, l_{ik}$ на случайные независимым образом. Данная операция и будет служить *resampling step*. Продолжаем это до тех пор, пока не получим $f = 1$.

Алгоритм в продвинутом подходе крепко связан с локальной леммой, ибо это есть ни что иное, как алгоритмическая версия локальной леммы Ловаса. Данный алгоритм называется *алгоритмом Мозера-Тардоша* и для него верна следующая теорема:¹⁰

Теорема 18. Пусть Y_i — количество *resampling steps* для события $A_i = \{i\text{-й дизъюнкт равен } 0\}$ В условиях задачи $k\text{-SAT}$ для любых $1 \leq i \leq m$, где m — это количество дизъюнктов в формуле, $E[Y_i] \leq \frac{1}{d} = \frac{e}{2^k - e}$.

Пусть $Y = \sum_{i=1}^m Y_i$ — общее число *resampling steps*. Тогда $E[Y] \leq \frac{m}{d} = \mathcal{O}(m)$.

⁹За доказательством этого обращайтесь ко второй задаче 4-го семинара.

¹⁰— А мы не доказываем эту теорему? — Нет. Это тяжело. (Д.А. Шабанов)

1.6.3 Геометрические вероятности. Задача о встрече

Определение 31. Пусть вероятностное пространство $\Omega \in \mathbb{R}^n$ таково, что у Ω определен n -мерный объем, $0 < \mu(\Omega) < \infty$. Тогда в геометрических вероятностях для любого $A \in \Omega$, у которого тоже определен объем $\mu(A)$, полагаем вероятность $P(A) = \frac{\mu(A)}{\mu(\Omega)}$. Это естественное обобщение классической модели на непрерывный случай.

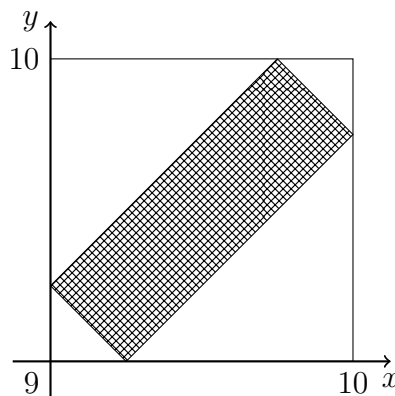
Примечание. Естественное оно по той простой причине, что аналогично дискретно случаю. Да и вообще вполне логично.

Для чего вводится такое определение? Бывают ситуации, когда конечным числом элементарных исходов обойтись нельзя. Ярким примером этому служит классическая *задача о встрече*

Задача 2 (Задача о встрече). Два друга договорились встретиться на автобусной остановке между 9-ю и 10-ю часами утра. Они договорились, что встретятся в течение этого часа, но не условились о конкретном времени встречи. При этом, если, прождав 15 минут с момента своего прихода, один из друзей не видит второго, он уходит, а значит встреча не состоялась. Вопрос: с какой вероятностью друзья встретятся?

Решение. Вообще, то, что конечности вероятностного пространства не хватит, понятно сразу, ведь моменты времени, в которые люди приходят на остановку, конечно же, не дискретны. Отложим на оси x и на оси y по отрезку длины 1, которые будут символизировать час между 9-ю и 10-ю часами, в течение которого на остановку приходит первый и второй человек соответственно. Итак, вероятностное пространство $\Omega = [9, 10] \times [9, 10]$.

Событию A , что встреча произойдет, удовлетворяют все точки получившегося квадрата, координаты которых отличаются не более, чем на четверть. То есть, моменты времени прихода друзей отличаются не более, чем на 15 минут. $A = \{\text{встреча произойдет}\} = \{(x, y) \mid |x - y| \leq \frac{1}{4}\}$. Изобразим это:



Довольно интуитивно хочется посчитать вероятность $P(A)$ как отношение площадей закрашенной области (всех удовлетворяющих исходов) ко всей площади квадрата (множество всех исходов). Это и будет геометрическая вероятность.

$$P(A) = 1 - \left(\frac{3}{4}\right)^2 = \frac{7}{16}.$$

□

1.7 Лекция от 21.10.2016

Будем дальше двигаться в сторону непрерывных вероятностей. Наша задача состоит в определении вероятностного пространства в общем случае.

1.7.1 Понятие вероятностного пространства в общем случае

Определение 32. *Вероятностным пространством (или тройкой Колмогорова)* называют тройку (Ω, \mathcal{F}, P) , где Ω — множество элементарных исходов, \mathcal{F} — множество событий, а P — вероятностная мера.

Разберем по отдельности каждый символ из этой “тройки”.

- Ω — произвольное множество элементарных исходов. Как и в дискретном случае, во время эксперимента возможно получить один и только один элементарный исход.

Например, если событие — это выстрел в мишень, то множество элементарных исходов будет задаваться плоскостью: $\Omega = \mathbb{R}^2$.

- \mathcal{F} — это совокупность подмножеств Ω , называемых *событиями*. В дискретном случае множество событий содержало все подмножества Ω ($\mathcal{F} = 2^\Omega$) и не было никаких проблем. Однако в общем случае возникают некоторые ограничения на \mathcal{F} . Обсудим, какие же. Перед этим напомним определение *алгебры*:

Определение 33. Пусть задано некоторое множество Ω . Система \mathcal{F} подмножеств Ω называется *алгеброй*, если выполняются следующие требования:

1. $\emptyset \in \mathcal{F}$.
2. Если $A \in \mathcal{F}$, то и его дополнение $\bar{A} = \Omega \setminus A \in \mathcal{F}$.
3. Если $A, B \in \mathcal{F}$, то их пересечение и объединение тоже лежит в \mathcal{F} .

Вообще говоря, достаточно требовать либо пересечение, либо объединение, так как по закону де Моргана $A \cup B = \overline{\bar{A} \cap \bar{B}}$.

Пример. Конечные объединения множеств вида $(-\infty, a]$, $(b, c]$, $(d, +\infty)$ образуют алгебру.

Далее, несколько усилим это определение, добавив дополнительное условие:

Определение 34. Пусть задано некоторое множество Ω . Система \mathcal{F} подмножеств Ω называется σ -*алгеброй*, если выполняются следующие требования:

1. \mathcal{F} является алгеброй.
2. Для любой последовательности событий $\{A_n\}_{n=1}^\infty$ их пересечение $\bigcap_{n=1}^\infty A_n$ и объединение $\bigcup_{n=1}^\infty A_n$ лежат в \mathcal{F} .

В данном случае опять же можно обойтись одним из двух — либо пересечением, либо объединением.

В общем случае полагается, что \mathcal{F} обязана быть σ -алгеброй.¹¹

Покажем, что среди σ -алгебр над одним множеством (их, вообще говоря, может быть много) есть минимальная.

¹¹Вообще, это требуется для того, чтобы всё не поломалось: <http://stats.stackexchange.com/questions/199280/why-do-we-need-sigma-algebras-to-define-probability-spaces>

Лемма. Пусть M — некоторая система подмножеств Ω . Тогда существует минимальная (по включению) σ -алгебра, содержащая в себе все подмножества из M .

Идея доказательства. Очевидно, что существуют σ -алгебры, содержащие M (та же 2^Ω). Тогда рассмотрим все σ -алгебры, содержащие M . Их пересечение тоже является σ -алгеброй и, при этом, минимально по включению. \square

Минимальную σ -алгебру, содержащую M , принято обозначать через $\sigma(M)$.

Над \mathbb{R} принято вводить особую σ -алгебру.

Определение 35. Борелевской σ -алгеброй на множестве \mathbb{R} называется минимальная σ -алгебра, содержащая все полуинтервалы вида $(a, b]$, где $a, b \in \mathbb{R}$:

$$\mathcal{B}(\mathbb{R}) = \sigma \{ (a, b] \mid a, b \in \mathbb{R}, a < b \}.$$

Упражнение. $\mathcal{B}(\mathbb{R})$ можно определить через интервалы, лучи, отрезки, открытые и замкнутые множества — это даст одну и ту же σ -алгебру.

Можно показать, что $\mathcal{B}(\mathbb{R}) \neq 2^\mathbb{R}$ (например, множество Витали¹² не входит в $\mathcal{B}(\mathbb{R})$).

Определение 36. Пара (Ω, \mathcal{F}) , где \mathcal{F} — σ -алгебра подмножеств Ω , называется *измеримым пространством*.

- Наконец, мы подошли к рассмотрению последнего члена “тройки Колмогорова”.

Определение 37. Отображение P из \mathcal{F} в $[0; 1]$ ($(P : \mathcal{F} \rightarrow [0; 1])$) называется *вероятностной мерой* на (Ω, \mathcal{F}) , если оно удовлетворяет следующим двум условиям:

1. $P(\Omega) = 1$.
2. *Счётная аддитивность:* для любой последовательности попарно непересекающихся событий $\{A_n\}_{n=1}^\infty$ выполнено, что

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

В дискретных вероятностных пространствах свойство счётной аддитивности доказывалось. Здесь его сразу постулируют.

Лемма (Свойства вероятности). Вероятностные меры обладают следующими свойствами:

1. $P(\emptyset) = 0$.
2. *Конечная аддитивность.* Если $\{A_1, A_2, \dots, A_n\}$ — попарно не пересекающиеся события, то

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

3. Если $A, B \in \mathcal{F}$ и $A \subseteq B$, то $P(A) \leq P(B)$.

¹²https://en.wikipedia.org/wiki/Vitali_set

$$4. P(A) + P(\bar{A}) = 1.$$

$$5. P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

$$6. \text{ Для любого набора событий } A_1, A_2, \dots, A_n \quad P\left(\bigcup_{n=1}^m A_n\right) \leq \sum_{n=1}^m P(A_n).$$

Доказательство. Докажем только первые два свойства, так как остальные доказываются ровно так же, как и в дискретном случае. За этими доказательствами можно обратиться к первой лекции.

1. Рассмотрим последовательность $\{A_n\}_{n=1}^{\infty}$, где для любого $n \in \mathbb{N}$ $A_n = \emptyset$. Тогда $A_i \cap A_j = \emptyset$ для любых $i, j \in \mathbb{N}$ и по свойству счётной аддитивности

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \begin{cases} P\left(\bigcup_{n=1}^{\infty} \emptyset\right) = P(\emptyset) \\ \sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} P(\emptyset) \end{cases}$$

Отсюда получаем, что

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

У этого уравнения есть два решения — 0 и ∞ . Но так как $P(\emptyset) \in [0, 1]$, то $P(\emptyset) = 0$.

2. Положим $A_n = \emptyset$ при $n > m$. Тогда по свойству счётной аддитивности:

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^m P(A_n) + \sum_{n=m+1}^{\infty} P(\emptyset) = \sum_{n=1}^m P(A_n). \quad \square$$

Теорема 19 (О непрерывности вероятностной меры). Пусть P — конечно-аддитивная функция на σ -алгебре событий \mathcal{F} , $P : \mathcal{F} \rightarrow [0, 1]$, $P(\Omega) = 1$. Тогда следующие четыре условия эквивалентны:

(a) P является счётно-аддитивной функцией.

(b) P непрерывна в “нуле”, то есть для любых множеств $A_1, A_2, \dots \in \mathcal{F}$ таких, что $A_{n+1} \subset A_n$, $\bigcap_{n=1}^{\infty} A_n = \emptyset$, выполняется:

$$\lim_n P(A_n) = 0.$$

(c) P непрерывна сверху, то есть для любых множеств $A_1, A_2, \dots \in \mathcal{F}$ таких, что $A_n \subset A_{n+1}$, $\bigcup_{n=1}^{\infty} A_n = A \in \mathcal{F}$, $(A_n \uparrow A)$, выполняется:

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

(d) P непрерывна снизу, то есть для любых множеств $A_1, A_2, \dots \in \mathcal{F}$ таких, что $A_{n+1} \subset A_n$, $\bigcap_{n=1}^{\infty} A_n = A \in \mathcal{F}$, $(A_n \downarrow A)$, выполняется:

$$\lim_n P(A_n) = P\left(\bigcap_{n=1}^{\infty} A_n\right).$$

Доказательство. Докажем эквивалентность всех этих утверждений в несколько шагов:

$[a \Rightarrow c]$ Покажем, что счётная аддитивность влечёт непрерывность сверху. Заметим, что $\bigcup_{n=1}^{\infty} A_n$ можно представить в виде счётного объединения попарно непересекающихся множеств:

$$\bigcup_{n=1}^{\infty} A_n = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \dots$$

Действительно, $A_{i+1} \setminus A_i$ содержит элементы, лежащие только в A_{i+1} . Тогда по счётной аддитивности

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} A_n\right) &= P(A_1) + P(A_2 \setminus A_1) + P(A_3 \setminus A_2) + \dots = \\ &= P(A_1) + P(A_2) - P(A_1) + P(A_3) - P(A_2) + \dots = \lim_{n \rightarrow \infty} P(A_n). \end{aligned}$$

$[c \Rightarrow d]$ Покажем, что непрерывность сверху влечёт непрерывность снизу. Рассмотрим некоторую последовательность $A_n \downarrow A$. Тогда для любого $n \geq 1$ выполнено, что $A_n \subseteq A_1$ и $A_n = A_1 \setminus (A_1 \setminus A_n)$. Следовательно,

$$P(A_n) = P(A_1 \setminus (A_1 \setminus A_n)) = P(A_1) - P(A_1 \setminus A_n).$$

Заметим, что последовательность $\{B_n\}_{n=1}^{\infty}$, где $B_n = A_1 \setminus A_n$ является неубывающей и $B_n \uparrow A_1 \setminus A$. Тогда, согласно пункту c,

$$\lim_{n \rightarrow \infty} P(A_1 \setminus A_n) = P\left(\bigcup_{n=1}^{\infty} (A_1 \setminus A_n)\right).$$

А это означает, что

$$\begin{aligned} \lim_{n \rightarrow \infty} P(A_n) &= P(A_1) - \lim_{n \rightarrow \infty} P(A_1 \setminus A_n) = P(A_1) - P\left(\bigcup_{n=1}^{\infty} (A_1 \setminus A_n)\right) \\ &= P(A_1) - P\left(A_1 \setminus \left(\bigcap_{n=1}^{\infty} A_n\right)\right) = P(A_1) - P(A_1) + P\left(\bigcap_{n=1}^{\infty} A_n\right) \\ &= P\left(\bigcap_{n=1}^{\infty} A_n\right) \end{aligned}$$

$[d \Rightarrow b]$ Достаточно взять последовательность $A_n \downarrow \emptyset$. Тогда из непрерывности снизу будет следовать непрерывность в нуле.

[$b \Rightarrow a$] Покажем, что непрерывность в нуле даёт счётную аддитивность. Рассмотрим некоторую последовательность попарно непересекающихся событий $\{A_n\}_{n=1}^{\infty}$. Заметим, что

$$P\left(\bigsqcup_{n=1}^{\infty} A_n\right) = P\left(\bigsqcup_{i=1}^n A_i\right) + P\left(\bigsqcup_{i=n+1}^{\infty} A_i\right).$$

Осталось только заметить, что $\bigsqcup_{i=n+1}^{\infty} A_i \downarrow \emptyset$ при $n \rightarrow \infty$. Тогда

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

1.7.2 Вероятностные меры на $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Пусть P — некоторая вероятностная мера.

Определение 38. *Функцией распределения* вероятностной меры P называют функцию $F : \mathbb{R} \rightarrow [0, 1]$ такую, что $F(x) = P((-\infty, x])$.

Лемма (Свойства функции распределения).

1. $F(x)$ неубывающая;
2. $\lim_{x \rightarrow +\infty} F(x) = 1$;
3. $\lim_{x \rightarrow -\infty} F(x) = 0$;
4. $F(x)$ непрерывна справа.

Доказательство.

1. Если $y > x$, то в силу аддитивности P , $F(y) - F(x) = P((x, y]) \geq 0$.
2. Пусть $x_n \uparrow +\infty$; тогда $(-\infty, x_n] \uparrow \mathbb{R}$. Значит, $\lim_{x_n \rightarrow +\infty} F(x_n) = \lim_{x_n \rightarrow +\infty} P((-\infty, x_n]) = P(\mathbb{R}) = 1$
3. Аналогично.
4. Пусть $x_n \downarrow x$; тогда, в силу непрерывности вероятностной меры, $F(x_n) \rightarrow F(x)$.

□

Примеры:

- Функция распределения константной случайной величины:

$$F(x) = \begin{cases} 1, & x \geq c; \\ 0, & x < c \end{cases}$$

- Функция распределения случайной величины, равномерно распределённой на $[0, 1]$:

$$F(x) = \begin{cases} 1, & x > 1; \\ x, & x \in [0, 1]; \\ 0, & x < 0 \end{cases}$$

Такой функции распределения соответствует вероятностная мера P такая, что $\forall a < b \in [0, 1], P((a, b]) = b - a$. Эту меру также называют *мерой Лебега* ¹³.

1.8 Лекция от 01.11.2016

1.8.1 Взаимно-однозначное соответствие функции распределения и вероятностной меры

Определение 39. Пусть функция $F(x), x \in \mathbb{R}$, удовлетворяет свойствам из леммы, то есть:

1. $F(x)$ неубывающая;
2. $\lim_{x \rightarrow +\infty} F(x) = 1$;
3. $\lim_{x \rightarrow -\infty} F(x) = 0$;
4. $F(x)$ непрерывна справа.

Тогда такую функцию будем называть *функцией распределения* на прямой.

Следующая фундаментальная и очень важная теорема будет введена без доказательства.

Теорема 20 (Теорема Каратеодори о продолжении меры). Пусть Ω — некоторое множество, \mathcal{A} — алгебра подмножеств Ω . Пусть вероятностная мера $P_0 : \mathcal{A} \mapsto [0; 1]$ удовлетворяет следующим свойствам:

1. $P_0(\Omega) = 1$;
2. P_0 — счетно-аддитивна на \mathcal{A} .

Тогда существует (и притом единственна) вероятностная мера P на $\sigma(\mathcal{A})$ такая, что мера P является продолжением меры P_0 , иными словами $\forall A \in \mathcal{A} P_0(A) = P(A)$.

Теорема 21 (Взаимно-однозначное соответствие функций распределения и мер на прямой). Пусть $F(x), x \in \mathbb{R}$ — функция распределения на прямой. Тогда существует (и притом единственна) вероятностная мера P на \mathcal{B} , такая что $F(x)$ — её функция распределения, то есть $\forall x \in \mathbb{R} P((-\infty; x]) = F(x)$.

¹³Говоря простым языком, вероятность попадания в полуинтервал пропорциональна его длине.

Доказательство. Рассмотрим алгебру \mathcal{A} , состоящую из конечных объединений непересекающихся полуинтервалов вида $(a; b]$, $\forall A \in \mathcal{A}$ имеет вид

$$\mathcal{A} = \bigcup_{k=1}^n (a_k; b_k], \quad -\infty \leq a_1 < b_1 < a_2 < b_2 < \dots < b_n \leq +\infty.$$

Зададим на \mathcal{A} меру P_0 по правилу

$$P_0(\mathcal{A}) = \sum_{k=1}^n (F(b_k) - F(a_k)).$$

Тогда $P_0(\mathbb{R}) = F(+\infty) - F(-\infty) = 1 - 0 = 1$ и по построению P_0 будет конечно-аддитивной мерой.

Для того чтобы воспользоваться теоремой Каратеодори, вероятностная мера P_0 должна обладать свойством счетной аддитивности. Но вместо того, чтобы доказывать это свойство напрямую, воспользуемся *теоремой о непрерывности вероятностной меры* из предыдущей лекции и докажем эквивалентное условие. Покажем, что мера P_0 является непрерывной в нуле.

Итак, нужно проверить, выполняется ли

$$\lim_{n \rightarrow \infty} P_0(A_n) = 0$$

для любых множеств $A_1, A_2, \dots \in \mathcal{A}$ таких, что $A_{n+1} \subset A_n$, $\bigcap_{n=1}^{\infty} A_n = \emptyset$

Заметим, что для любого полуинтервала $(a, b]$ и для любого $\delta > 0$ можно взять такое $a' > a$, что:

$$P_0((a; b]) - P_0((a'; b]) \leq \delta$$

$$P_0((a; b]) - P_0((a'; b]) = (F(b) - F(a)) - (F(b) - F(a')) = F(a') - F(a).$$

Зафиксируем произвольное $\varepsilon > 0$. В силу непрерывности справа такое значение a' можно подобрать. Тогда

$$\forall A_n \exists B_n \in \mathcal{A}, P_0(A_n) - P_0(B_n) \leq \varepsilon 2^{-n}, B_n \in A_n \text{ и } [B_n] \in A_n, \text{ где } [B_n] \text{ есть замыкание } B_n.$$

То есть если B_n имеет вид $(a, b]$, то $[B_n]$, являющееся замыканием B_n , имеет вид $[a, b]$. Мы добавили в множество B_n его граничные точки.

Пусть сначала все A_n лежат внутри $[-N; N]$ для некоторого $N \in \mathbb{N}$. Мы знаем, что $\bigcap_{n=1}^{\infty} A_n = \emptyset$. Следовательно, $\bigcap_{n=1}^{\infty} [B_n] = \emptyset$.

По принципу компактности существует n_0 такой, что из $\bigcap_{n=1}^{n_0} [B_n] = \emptyset$ следует $\bigcap_{n=1}^{n_0} B_n = \emptyset$. Тогда:

$$P_0(A_{n_0}) = P_0\left(A_{n_0} \setminus \bigcap_{n=1}^{n_0} B_n\right)$$

Следующий переход основан на том, что если $\omega \in A_n \setminus \bigcap_{n=1}^{n_0} B_n$, то существует такой номер k , что $\omega \notin B_k$, а значит $\omega \in A_{n_0} \setminus B_k$.

$$\begin{aligned}
P_0 \left(A_{n_0} \setminus \bigcap_{n=1}^{n_0} B_n \right) &\leq \sum_{k=1}^{n_0} P_0 (A_{n_0} \setminus B_k) \leq \sum_{k=1}^{n_0} P_0 (A_k \setminus B_k) \leq \\
&\leq \sum_{k=1}^{n_0} (P_0 (A_k) - P_0 (b_k)) \leq \sum_{k=1}^{n_0} \varepsilon 2^{-k} \leq \varepsilon.
\end{aligned}$$

Получили, что $P_0(A_{n_0}) \leq \varepsilon$. Значит,

$$\forall n > n_0 \quad P_0(A_n) \leq \varepsilon \implies \lim_{n \rightarrow \infty} P_0(A_n) = 0.$$

Если в A_1 есть бесконечные полуинтервалы, то выберем полуинтервал $(-N; N]$, что $P_0((-N; N]) \geq 1 - \frac{\varepsilon}{2}$ и рассмотрим $A'_n = A_n \cap (-N; N]$.

По доказанному выше $P_0(A'_n) \leq \frac{\varepsilon}{2}$ при $n > n_0(\varepsilon)$ и

$$P_0(A_n) = P_0(A'_n \cup (A_n \cup (-\infty; -N]) \cup (A_n \cap (N; +\infty))) \leq P_0(A'_n) + P_0(\mathbb{R} \setminus (-N; N]) \leq \varepsilon.$$

По теореме Каратеодори искомая P существует и единственна. \square

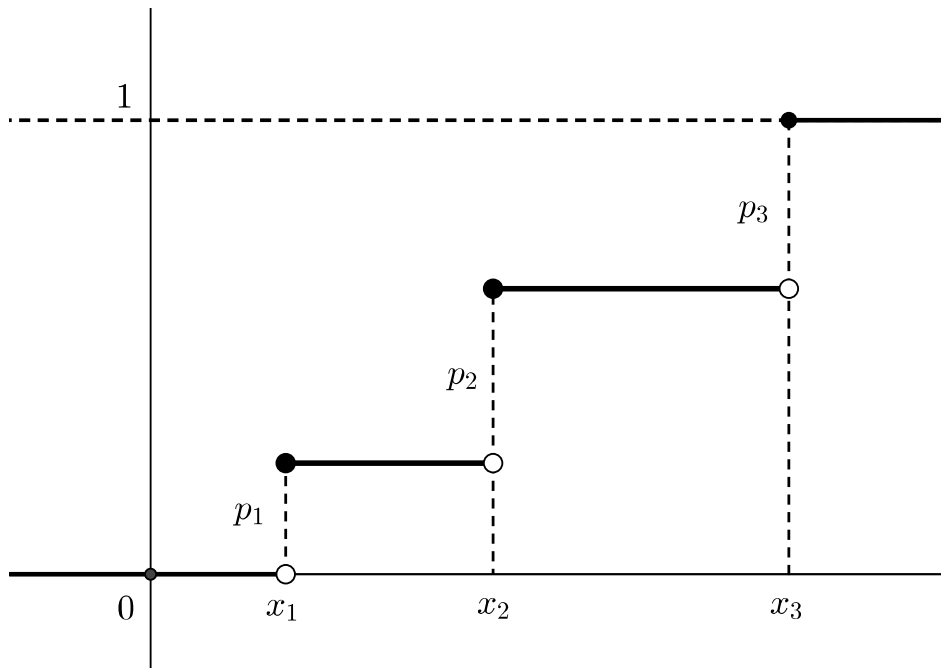
1.8.2 Классификация функций распределения на \mathbb{R}

Дискретные функции распределения

Вероятностная мера P на $\mathcal{B}(\mathbb{R})$ называется дискретной, если $\exists X = \{x_n\}$ — не более чем счётное множество, такое что $P(\mathbb{R} \setminus X) = 0$ и $\forall x_n \in X, P(x_n) > 0$. Между прочим, в этом случае говорят, что P *сосредоточена* в X .

Но нас интересует функция распределения. Она будет выглядеть так:

$$F(x) = P((-\infty, x]) = \sum_{k: x_k \in X} P(x_k)$$



Примеры:

- Равномерное дискретное распределение:

$$x = \{1 \dots n\}, P_k = \frac{1}{n}$$

- Распределение Бернулли:

$$x = \{0, 1\}, P_1 = p, P_0 = 1 - p$$

- Биномиальное распределение:

$$P_k = \binom{n}{k} p^k (1-p)^{n-k}$$

- Пуассоновское распределение:

$$P_k = \frac{\lambda^k}{k!} e^{-\lambda}$$

Абсолютно непрерывные функции распределения

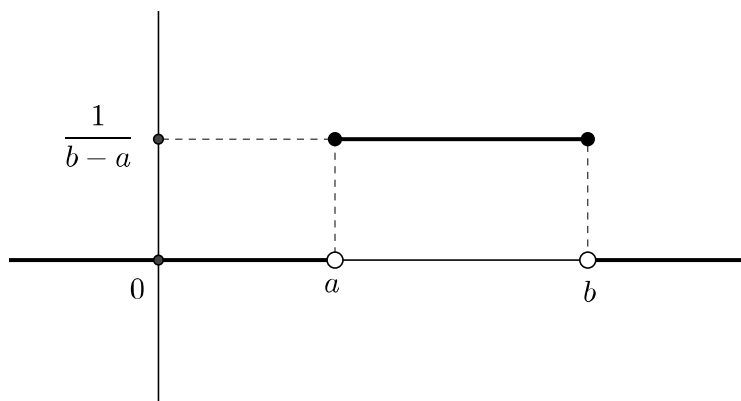
Рассмотрим некоторую функцию $p(t)$, неотрицательную на \mathbb{R} , такую, что $\int_{-\infty}^{\infty} p(t) dt = 1$.

Определение 40. Функция распределения $F(x)$ называется абсолютно непрерывной с плотностью $p(t)$, если $\forall x \in \mathbb{R}, F(x) = \int_{-\infty}^x p(t) dt$

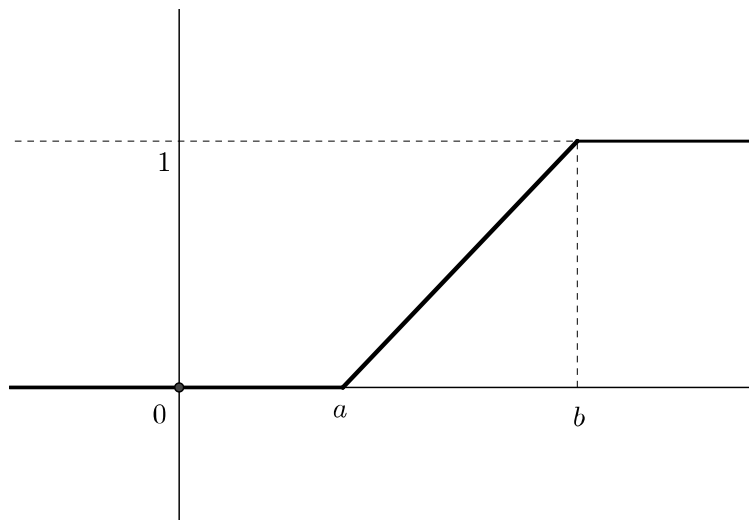
Примеры:

- Равномерное распределение на отрезке:

$$p(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & x \notin [a, b] \end{cases}$$

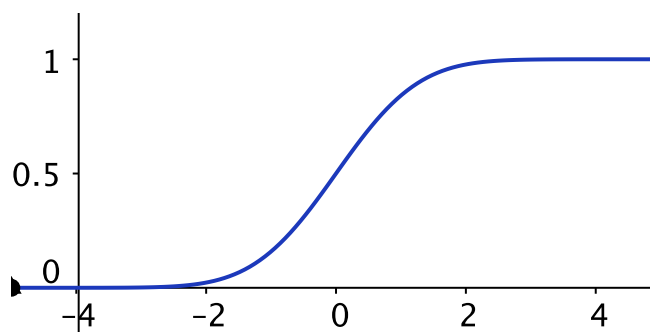


$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & x \in [a, b] \\ 1, & x > b \end{cases}$$

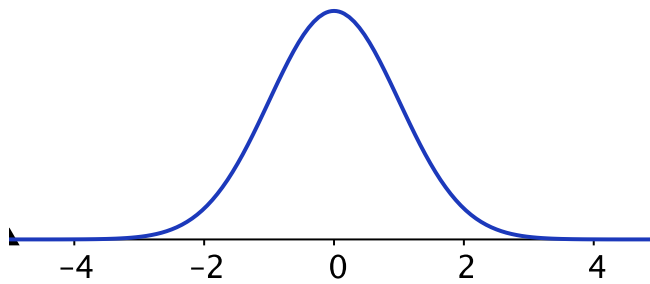


- Нормальное распределение: $N(a, \sigma^2)$. Самое известное и распространённое распределение, встречающееся всюду; например погрешности измерений, как правило, подчиняются этому распределению.

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

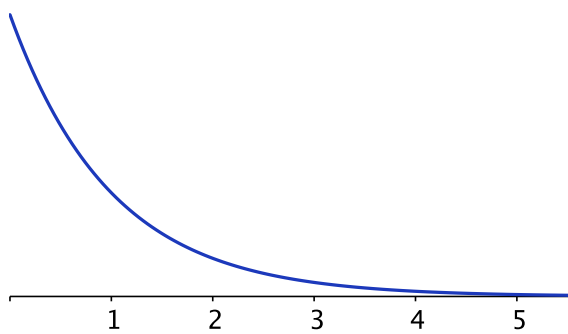


$$F(x) = \Phi\left(\frac{x-a}{\sigma}\right), \text{ где } \Phi(x) \text{ — функция Лапласа}$$

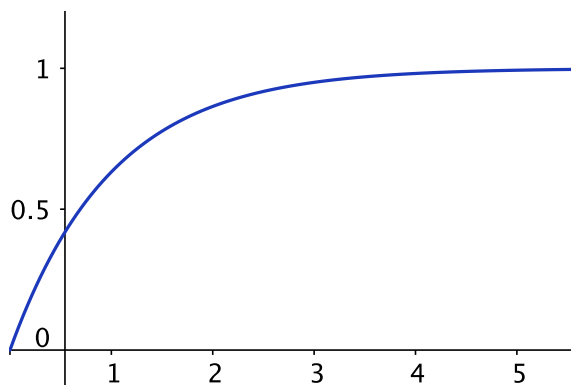


- Экспоненциальное распределение: $\text{Exp}(\alpha)$.

$$p(x) = \begin{cases} \alpha e^{-\alpha x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$



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



Сингулярные функции распределения

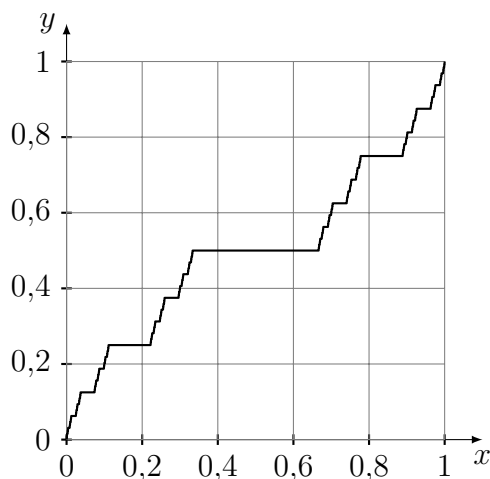
Определение 41. Точка y называется точкой роста функции $F(x)$, $x \in \mathbb{R}$, если $\forall \varepsilon > 0$, $F(y + \varepsilon) - F(y - \varepsilon) > 0$.

Определение 42. $\mathcal{A} \subset \mathbb{R}$ имеет Лебегову меру 0, если $\forall \varepsilon$ существует набор интервалов $(a_n; b_n)$ таких что $\sum (b_k - a_k) < \varepsilon$ и $\mathcal{A} \subset \bigcup (a_k; b_k)$.

Определение 43. Функция распределения $F(x)$ называется *сингулярной*, если она непрерывна и множество её точек роста имеет Лебегову меру 0.

Пример:

- Канторова лестница:



1.9 Лекция от 08.11.2016

1.9.1 Канторова лестница. Продолжение.

Заметим, что канторова лестница — неубывающая непрерывная функция, которая возрастает от 0 до 1. Давайте рассмотрим множество её точек роста:

Легко понять, что точками роста могут быть только точки, не попавшие ни в один заполняемый интервал. Давайте оценим меру Лебега множества точек роста, пользуясь аддитивностью меры и тем, что мера $[0, 1]$ равна 1.

Сумма длин интервалов на i -ом шаге — $\frac{1}{3} \cdot \left(\frac{2}{3}\right)^i$; тогда всего эти интервалы заполняют множество меры $\sum_{i=0}^{\infty} \frac{1}{3} \cdot \left(\frac{2}{3}\right)^i = 1$. Значит, мера множества точек роста равна $1 - 1 = 0$, значит, канторова лестница является сингулярной функцией распределения.

Казалось бы, а зачем вообще нужны сингулярные функции? Оказывается, что дискретные, абсолютно непрерывные и сингулярные функции покрывают все функции распределения.

Теорема 22 (Лебег). Пусть $F(x)$ — функция распределения на \mathbb{R} . Тогда имеет место представление $F(x) = \alpha_1 F_1(x) + \alpha_2 F_2(x) + \alpha_3 F_3(x)$, где

- F_1 — дискретная;
- F_2 — абсолютно непрерывная;
- F_3 — сингулярная.

При этом $\alpha_i \geq 0$, $\alpha_1 + \alpha_2 + \alpha_3 = 1$.

Доказательство. Это хорошая теорема. Доказывать её мы, конечно, не будем. □

1.9.2 Случайные величины и векторы

В дискретном случае мы говорили, что ξ — это произвольное отображение из Ω в \mathbb{R} . Но в дискретном случае никаких сложностей не возникает, так как всегда можно посчитать любую вероятность. В общем случае такой красоты нет. Тогда нужно вводить некоторые ограничения.

Однако, было бы неплохо ввести обобщение так, чтобы дискретный случай был частным случаем общего (как и должно быть). То есть, если на вероятностном пространстве (Ω, \mathcal{F}, P) задана некоторая случайная величина ξ , то мы хотим уметь считать вероятности вида $P(\xi \leq x)$, $P(\xi = x)$, $P(a \leq \xi \leq b)$ и так далее. Поэтому придётся ввести одно ограничение.

Определение 44. Пусть задано вероятностное пространство (Ω, \mathcal{F}, P) . Будем называть случайной величиной отображение $\xi : \Omega \mapsto \mathbb{R}$, если оно удовлетворяет свойству измеримости: для любого $x \in \mathbb{R}$ событие $\{\xi \leq x\} \equiv \{\omega \in \Omega \mid \xi(\omega) \leq x\}$ принадлежит \mathcal{F} .

Смысл измеримости описан ранее и скрыт в названии: если функция измерима, то мы можем её измерить.

Если выбрать случайную точку на той же плоскости, то как её описать? Введём понятие *случайного вектора*:

Определение 45. Случайный вектор — вектор, состоящий из случайных величин.

Казалось бы, почему мы вводим только такие события? Оказывается, этого более, чем достаточно и для случайных величин, и для случайных векторов. Однако для того, чтобы доказать это, нужно ввести ещё одно определение:

Определение 46. Борелевской σ -алгеброй в \mathbb{R}^n называют минимальную σ -алгебру содержащую все $(a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_n, b_n]$:

$$\mathcal{B}(\mathbb{R}^n) = \sigma \{ (a_1, b_1] \times \cdots \times (a_n, b_n] \mid a_1, \dots, a_n, b_1, \dots, b_n \in \mathbb{R}, a_i < b_i \}.$$

Примечание. Можно показать, что в $\mathcal{B}(\mathbb{R}^n)$ лежит много чего: все прямоугольники вида $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$, все открытые и замкнутые множества, элементы вида $(-\infty, b_1] \times (-\infty, b_2] \times \cdots \times (-\infty, b_n]$ и так далее.

Теперь докажем то, что нашего определения достаточно:

Лемма. Пусть $(\Omega, \mathcal{F}, \mathbf{P})$ — вероятностное пространство, и на нём заданы случайная величина ξ и случайный вектор $\eta = (\eta_1, \dots, \eta_n)$. Тогда

1. ξ является случайной величиной тогда и только тогда, когда для любого $B \in \mathcal{B}(\mathbb{R})$ событие $\xi^{-1}(B) = \{\xi \in B\}$ лежит в \mathcal{F} .
2. ξ является случайным вектором тогда и только тогда, когда для любого $B \in \mathcal{B}(\mathbb{R}^n)$ событие $\eta^{-1}(B) = \{\eta \in B\}$ лежит в \mathcal{F} .

Примечание. Если $B \in \mathcal{B}(\mathbb{R}^n)$, то его ещё (неформально) называют борелевским множеством.

Доказательство. Для начала докажем это для случайных величин.

[\Leftarrow] Заметим, что $(-\infty, x] \in \mathcal{B}(\mathbb{R})$ для любого $x \in \mathbb{R}$. Тогда

$$\xi^{-1}((-\infty, x]) = \{\xi \in (-\infty, x]\} \in \mathcal{F}.$$

А это и есть определение измеримости. Следовательно, ξ — случайная величина.

[\Rightarrow] Рассмотрим следующее множество:

$$D = \{B \in \mathcal{B}(\mathbb{R}) \mid \xi^{-1}(B) \in \mathcal{F}\}.$$

Покажем, что оно является σ -алгеброй. Так как полный прообраз сохраняет все теоретико-множественные операции, то в D будут находиться пересечения, дополнения, объединения и счётные объединения. Тогда D действительно является σ -алгеброй.

Однако, согласно определению случайной величины, все лучи вида $(-\infty, x]$ лежат в D . Тогда D есть σ -алгебра, содержащая все лучи. Но тогда она содержит в себе $\mathcal{B}(\mathbb{R})$. Так как по построению $D \subseteq \mathcal{B}(\mathbb{R})$, то $D = \mathcal{B}(\mathbb{R})$.

Идейно доказательство для случайных векторов ничем не отличается.

[\Leftarrow] Заметим, что $(-\infty, x_1] \times \cdots \times (-\infty, x_n] \in \mathcal{B}(\mathbb{R}^n)$ для любых $x_1, x_2, \dots, x_n \in \mathbb{R}$. Тогда

$$\eta^{-1}((-\infty, x_1] \times \cdots \times (-\infty, x_n]) = \{\eta_1 \in (-\infty, x_1]\} \cap \cdots \cap \{\eta_n \in (-\infty, x_n]\} \in \mathcal{F}.$$

Тогда все η_1, \dots, η_n являются случайными величинами и η действительно является случайным вектором.

[\Rightarrow] Рассмотрим следующее множество:

$$D = \{B \in \mathcal{B}(\mathbb{R}) \mid \eta^{-1}(B) \in \mathcal{F}\}.$$

Покажем, что оно является σ -алгеброй. Так как полный прообраз сохраняет все теоретико-множественные операции, то в D будут находиться пересечения, дополнения, объединения и счётные объединения. Тогда D действительно является σ -алгеброй.

Однако, согласно определению случайной величины, все элементы вида $(-\infty, x_1] \times \dots \times (-\infty, x_n]$ лежат в D . Тогда D содержит в себе $\mathcal{B}(\mathbb{R})$. Так как по построению $D \subseteq \mathcal{B}(\mathbb{R})$, то $D = \mathcal{B}(\mathbb{R})$.

□

1.9.3 Действия над случайными величинами

Допустим, у нас есть несколько случайных величин $\xi_1, \xi_2, \dots, \xi_n$, и мы хотим с ними что-то сделать — перемножить, например. Другими словами, мы хотим взять *функцию* от случайных величин $f(\xi_1, \xi_2, \dots, \xi_n)$. Но можно ли сказать, что это действительно будет случайной величиной? Не всегда. Однако есть класс функций, называемых *борелевскими*, для которых это точно верно. Введём определение:

Определение 47. Пусть $f : \mathbb{R}^n \mapsto \mathbb{R}^m$ — некоторая функция. Будем называть её *борелевской*, если выполнено следующее условие:

$$\forall B \in \mathcal{B}(\mathbb{R}^m) \quad f^{-1}(B) \equiv \{x : f(x) \in B\} \in \mathcal{B}(\mathbb{R}^n).$$

Вообще, почти все “обычные” функции являются борелевскими. Например, непрерывные функции являются борелевскими, так как они переводят открытые множества в открытые. Можно так же доказать, что если у функции множество точек разрыва имеет лебегову меру 0, то она тоже является борелевской.¹⁴

Теперь покажем, что борелевские функции действительно ничего не ломают.

Теорема 23. Пусть ξ_1, \dots, ξ_n — случайные величины, а $f : \mathbb{R}^n \mapsto \mathbb{R}^m$ — борелевская функция. Тогда $f(\xi_1, \xi_2, \dots, \xi_n)$ является случайной величиной.

Доказательство. Рассмотрим произвольное борелевское множество $B \in \mathcal{B}(\mathbb{R}^m)$. Тогда

$$\{f(\xi_1, \dots, \xi_n) \in B\} = \{(\xi_1, \xi_2, \dots, \xi_n) \in f^{-1}(B)\}.$$

Так как $f^{-1}(B) \in \mathcal{B}(\mathbb{R}^n)$, то $\{f(\xi_1, \dots, \xi_n) \in B\} \in \mathcal{F}$.

□

Следствие. Если ξ и η — это случайные величины, то $\xi \pm \eta$, $\alpha\xi$ для любого $\alpha \in \mathbb{R}$, $\xi\eta$ и ξ/η при $\eta \neq 0$ тоже являются случайными величинами.

В дальнейшем у нас будут возникать последовательности случайных величин. Хотелось бы понять — а будут ли те же пределы последовательностей случайными величинами? Ответ — да.

¹⁴ Доказательство можно найти по следующей ссылке: <http://math.stackexchange.com/questions/369268/a-function-with-countable-discontinuities-is-borel-measurable>

Теорема 24. Пусть $\{\xi_n\}_{n=1}^{\infty}$ — последовательность случайных величин. Тогда

$$\overline{\lim}_{n \rightarrow \infty} \xi_n, \underline{\lim}_{n \rightarrow \infty} \xi_n, \sup_{n \in \mathbb{N}} \xi_n, \inf_{n \in \mathbb{N}} \xi_n$$

тоже являются случайными величинами.

Доказательство. Покажем, что все они измеримы:

$$\begin{aligned} \left\{ \sup_{n \in \mathbb{N}} \xi_n > x \right\} &= \bigcup_{n=1}^{\infty} \underbrace{\{\xi_n > x\}}_{\in \mathcal{F}} \in \mathcal{F} \\ \left\{ \inf_{n \in \mathbb{N}} \xi_n < x \right\} &= \bigcup_{n=1}^{\infty} \underbrace{\{\xi_n < x\}}_{\in \mathcal{F}} \in \mathcal{F} \\ \overline{\lim}_{n \in \mathbb{N}} \xi_n &= \inf_{n \in \mathbb{N}} \sup_{k \geq n} \xi_k \\ \underline{\lim}_{n \in \mathbb{N}} \xi_n &= \sup_{n \in \mathbb{N}} \inf_{k \geq n} \xi_k \end{aligned}$$

□

1.9.4 Простые случайные величины. Матожидание в общем случае

Напомним определение индикатора:

Определение 48. Пусть $(\Omega, \mathcal{F}, \mathbf{P})$ — вероятностное пространство и $A \in \mathcal{F}$ — некоторое событие. Тогда *индикатором* события A будем называть функцию $\mathbf{I}_A : \Omega \mapsto \{0, 1\}$, устроенную следующим образом:

$$\mathbf{I}_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$$

Иногда индикатор обозначают через $\mathbf{I}\{A\}$.

Покажем, что индикатор является случайной величиной. Действительно,

$$\{\mathbf{I}_A \leq x\} = \begin{cases} \emptyset, & x < 0 \\ \overline{A}, & 0 \leq x < 1 \\ \Omega, & x \geq 1 \end{cases}$$

Так как каждое из трёх подмножеств лежит в \mathcal{F} , то \mathbf{I}_A является случайной величиной, что и требовалось доказать.

Индикатор является один из примеров так называемых *простых* случайных величин. Введём определение:

Определение 49. *Простая случайная величина* — случайная величина, принимающая конечное число значений.

Если x_1, x_2, \dots, x_n — все значения простой случайной величины ξ , то её можно представить в следующем виде:

$$\xi = \sum_{k=1}^n x_k \mathbf{I}\{\xi = x_k\}$$

Теперь о самом матожидании:

- Пусть ξ — простая случайная величина. Тогда, по аналогии с дискретным случаем, $E[\xi]$ будет равно

$$E[\xi] \equiv \sum_{k=1}^n x_k P(\xi = x_k).$$

В дискретном случае похожая формула доказывалась, однако в общем случае она постулируется.

- Теперь допустим, что ξ — это произвольная неотрицательная случайная величина. Как ввести её математическое ожидание? Допустим, что есть последовательность $\{\xi_n\}_{n=1}^{\infty}$ простых случайных величин такая, что $\xi_n(\omega) \uparrow \xi(\omega)$ для любого $\omega \in \Omega$. Тогда положим, что

$$E[\xi] \equiv \lim_{n \rightarrow \infty} E[\xi_n].$$

Однако возникает вопрос: а существует ли она? Докажем, что такая последовательность ξ_n действительно существует.

Рассмотрим последовательность случайных величин $\{\xi_n\}_{n=1}^{\infty}$, устроенную следующим образом:

$$\xi_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \mathbf{I} \left\{ \frac{k-1}{2^n} \leq \xi < \frac{k}{2^n} \right\} + n \mathbf{I} \{ \xi \geq n \}.$$

Непосредственная проверка убеждает в том, что $\xi_n(\omega) \uparrow \xi(\omega)$ для любого $\omega \in \Omega$.

1.10 Лекция от 15.11.2016

1.10.1 Свойства математического ожидания для простых случайных величин.

1. Линейность.

Если ξ, η — простые случайные величины, есть некоторые константы $a, b \in \mathbb{R}$, то

$$E[a\xi + b\eta] = a E[\xi] + b E[\eta].$$

Доказательство. Введем обозначение $\delta = a\xi + b\eta$. Пусть x_1, \dots, x_k и y_1, \dots, y_m — все значения случайных величин ξ и η соответственно, а z_1, \dots, z_n — все значения δ .

Тогда:

$$\begin{aligned}
 E[a\xi + b\eta] &= E[\delta] = \sum_{i=1}^n z_i P(\delta = z_i) = \sum_{i=1}^n z_i P\left(\bigcup_{\substack{t,j: \\ ax_t + by_j = z_i}} \{\xi = x_t, \eta = y_j\}\right) = \\
 &= \sum_{i=1}^n z_i \sum_{\substack{t,j: \\ ax_t + by_j = z_i}} P(\xi = x_t, \eta = y_j) = \\
 &= \sum_{i=1}^n \sum_{\substack{t,j: \\ ax_t + by_j = z_i}} (ax_t + by_j) P(\xi = x_t, \eta = y_j) = \sum_{t=1}^k \sum_{j=1}^m (ax_t + by_j) P(\xi = x_t, \eta = y_j) = \\
 &= \sum_{t=1}^k \sum_{j=1}^m ax_t P(\xi = x_t, \eta = y_j) + \sum_{t=1}^k \sum_{j=1}^m by_j P(\xi = x_t, \eta = y_j) = \\
 &= \sum_{t=1}^k ax_t P(\xi = x_t) + \sum_{j=1}^m by_j P(\eta = y_j) = a E[\xi] + b E[\eta].
 \end{aligned}$$

□

2. Сохранение относительного порядка:

Если ξ, η — простые случайные величины, тогда:

- (a) $\xi \geq 0 \implies E[\xi] \geq 0$
- (b) $\xi \geq \eta \implies E[\xi] \geq E[\eta]$.

Доказательство. Докажем свойства по порядку:

- (a) Если x_1, \dots, x_k — все значения $\xi \geq 0$, то $\forall i = 1, \dots, k$ $x_i \geq 0$. Значит, $E[\xi] = \sum_{i=1}^k x_i P(\xi = x_i) \geq 0$.
- (b) Введем величину $\delta = \xi - \eta$. Так как, $\xi \geq \eta$, $\delta \geq 0$. Тогда по первому пункту $E[\delta] \geq 0 = E[\xi] - E[\eta]$.

□

3.

Лемма. Пусть $\eta, \{\xi_n, n \in \mathbb{N}\}$ — простые случайные величины, причем $\xi_n \uparrow \xi$, $\xi_n \geq 0 \forall n \in \mathbb{N}$, $\eta \leq \xi$, $\eta \geq 0$. Тогда

$$E[\eta] \leq \lim_n E[\xi_n].$$

Доказательство. Возьмем любое $\alpha > 0$ и положим

$$A_n = \{\omega : \xi_n(\omega) \geq \eta(\omega) - \alpha\}.$$

Раз $\xi_n \uparrow \xi$, то $A_n \uparrow \Omega$. По теореме о непрерывности вероятностной меры $P(A_n) \rightarrow 1$. Рассмотрим математическое ожидание случайной величины ξ_n :

$$E[\xi_n] = E[\xi_n I_{A_n}] + E[\xi_n I_{\overline{A_n}}] \geq E[\xi_n I_{A_n}] \geq E[(\eta - \alpha) I_{A_n}] = E[\eta] - E[\eta I_{\overline{A_n}}] - E[\alpha I_{A_n}].$$

Введем обозначение: $c = \max_{\omega \in \Omega} \eta(\omega)$, тогда

$$E[\eta] - E[\eta I_{\overline{A_n}}] - E[\alpha I_{A_n}] \geq E[\eta] - c E[I_{\overline{A_n}}] - \alpha E[I_{A_n}] \geq E[\eta] - c P(\overline{A_n}) - \alpha.$$

Раз для любого n , $\xi_n \leq \xi_{n+1}$, то $E[\xi_n] \leq E[\xi_{n+1}]$. Следовательно, существует $\lim_n E[\xi_n]$. Перейдем к пределу в неравенстве выше:

$$\lim_n E[\xi_n] \geq E[\eta] - \alpha.$$

В силу произвольности $\alpha > 0$, $\lim_n E[\xi_n] \geq E[\eta]$. □

Следствие (Корректность определения матожидания.). *Для неотрицательных случайных величин, матожидание определено корректно.*

Доказательство. Нужно показать, что предел не зависит от выбора последовательности простых случайных величин $0 \leq \xi_n \uparrow \xi$. Пусть есть другая такая последовательность, $0 \leq \eta_m \uparrow \xi$. Тогда по лемме, для любого натурального m $\lim_n E[\xi_n] \geq E[\eta_m]$. Значит, $\lim_n E[\xi_n] \geq \lim_m E[\eta_m]$. С другой стороны, для любых натуральных n ,

$$\lim_m E[\eta_m] \geq E[\xi_n] \implies \lim_m E[\eta_m] \geq \lim_n E[\xi_n].$$

Следовательно, $\lim_m E[\eta_m] = \lim_n E[\xi_n]$. □

1.10.2 Свойства математического ожидания в общем случае.

Пусть ξ — произвольная случайная величина. Введем обозначения: $\xi^+ = \max(\xi, 0)$ и $\xi^- = -\min(\xi, 0)$.

Определение 50. Говорят, что математическое ожидание $E[\xi]$ случайной величины ξ существует, или определено, если по крайней мере одна из величин $E[\xi^+]$ или $E[\xi^-]$ конечна. В этом случае по определению полагают $E[\xi] = E[\xi^+] - E[\xi^-]$. Есть четыре комбинации значений $E[\xi^+]$ и $E[\xi^-]$. Разберем их:

1. Если $E[\xi^+]$ и $E[\xi^-]$ — конечны, то $E[\xi] = E[\xi^+] - E[\xi^-]$.
2. Если $E[\xi^+] = \infty$, $E[\xi^-]$ — конечно, то $E[\xi] = +\infty$.
3. Если $E[\xi^+] — конечно, $E[\xi^-] = \infty$, то $E[\xi] = -\infty$.$
4. Если $E[\xi^+] = \infty$ и $E[\xi^-] = \infty$, то $E[\xi]$ не определено.

Итак, перейдем к рассмотрению свойств математического ожидания случайных величин в общем случае. Но перед этим введем еще одно важное понятие, которое будет использовать для целой группы утверждений в следующей теореме.

Определение 51. Будем говорить, что некоторое свойство A выполнено P — почти наверное, если существует такое множество N с $P(N) = 0$ такое, что это свойство A выполнено для каждой точки $\omega = \Omega \setminus N$. Также можно определить, что событие A происходит почти наверное, если $P(A) = 1$. Обозначение: $A_{п.н.}$.

Теорема 25 (Свойства математического ожидания в общем случае.). Если $E[\xi]$ — математическое ожидание случайной величины ξ , то выполняются следующие свойства:

1. Если $c \in \mathbb{R}$, $E[\xi]$ — конечно, тогда

$$E[c\xi] = cE[\xi].$$

2. Пусть $\xi \leq \eta$ и $E[\xi], E[\eta]$ — конечны. Тогда

$$E[\xi] \leq E[\eta].$$

3. Если $E[\xi]$ и $E[\eta]$ — конечны, то $E[\xi + \eta]$ — конечно и

$$E[\xi + \eta] = E[\eta] + E[\xi].$$

4. Если $E[\xi]$ — существует, тогда

$$|E[\xi]| \leq E[|\xi|].$$

5. (a) Пусть $0 \leq \xi \leq \eta$ и $E[\eta]$ — конечно, тогда $E[\xi]$ также конечно.

(b) Пусть $|\xi| \leq \eta$ и $E[\eta]$ — конечно, тогда $E[\xi]$ также конечно.

6. Если $\xi = 0_{п.н.}$, то $E[\xi] = 0$.

7. Если $\xi = \eta_{п.н.}$, и $E[\eta]$ — конечно, то $E[\xi] = E[\eta]$.

8. Если $\xi \geq 0$ и $E[\xi] = 0$, то $\xi = 0_{п.н.}$.

9. Пусть $\forall A \in \mathcal{F}$, $E[\xi I_A] \leq E[\eta I_A]$ и $E[\xi], E[\eta]$ — конечны. Тогда $\xi \leq \eta_{п.н.}$.

Доказательство. Докажем первую группу свойств:

1. Для простых случайных величин было доказано в первой части лекции. В общем же случае надо рассмотреть представление $\xi = \xi^+ - \xi^-$ и заметить, что для $c \geq 0$, $(c\xi)^+ = c\xi^+$, $(c\xi)^- = c\xi^-$, а для $c < 0$, $(c\xi)^+ = -c\xi^-$, $(c\xi)^- = -c\xi^+$.
2. Для простых случайных величин было доказано в первой части лекции. Рассмотрим общий случай. Пусть теперь $E[\xi] > -\infty$, тогда $E[\xi^-] < \infty$. Если $\xi \leq \eta$, то $\xi^+ \leq \eta^+$, $\xi^- \geq \eta^-$. Тогда $E[\eta^-] \leq E[\xi^-] < \infty$, следовательно, $E[\eta]$ определено и $E[\xi] = E[\xi^+] - E[\xi^-] \leq E[\eta^+] - E[\eta^-] = E[\eta]$. Аналогичным образом рассматривается случай, когда $E[\eta] < \infty$.
3. Случай с простыми случайными величинами разбирался в первой части лекции. В общем же случае, когда $E[|\xi|] < \infty$, $E[|\eta|] < \infty$, сводится к рассмотренному, если воспользоваться тем, что $\xi = \xi^+ - \xi^-$, $\eta = \eta^+ - \eta^-$, $\xi^+ \leq |\xi|$, $\xi^- \leq |\xi|$, $\eta^+ \leq |\eta|$, $\eta^- \leq |\eta|$.
4. Поскольку $-|\xi| \leq \xi \leq |\xi|$, по первым двум свойствам $-E[|\xi|] \leq E[\xi] \leq E[|\xi|]$.

Теперь докажем вторую группу свойств, связанных с понятием “ P – почти нивное”:

6. В самом деле, если ξ – простая случайная величина, тогда она по определению разбивается в следующую сумму: $\xi = \sum x_k I_{A_k}(\omega)$ и $x_k \neq 0$. По условию, $\xi = 0_{n.n.}$, следовательно для любых событий $A_k = \{\xi = x_k\}$, $P(A_k) = 0$, а значит, $E[\xi] = 0$. Если же $\xi \geq 0$ – неотрицательная случайная величина, то для любой последовательности $\{\xi_n\}$, $0 \leq \xi_n \uparrow \xi$,

$$P(\xi_n) \geq P(\xi = 0) = 1 \implies \xi_n = 0_{n.n.}$$

Значит,

$$E[\xi_n] = 0, \forall n \implies E[\xi] = \lim_n E[\xi_n] = 0.$$

Если $\xi = 0_{n.n.}$, то

$$\xi^+ = 0_{n.n.}, \xi^- = 0_{n.n.}, \implies E[\xi^+] = 0_{n.n.}, E[\xi^-] = 0_{n.n.} \implies E[\xi] = E[\xi^+] - E[\xi^-] = 0.$$

7. В самом деле, пусть $N = \{\omega : \xi \neq \eta\}$. Тогда $P(N) = 0$ и $\xi = \xi I_N + \xi I_{\bar{N}}$, $\eta = \eta I_N + \eta I_{\bar{N}} = \eta I_N + \xi I_{\bar{N}}$. По свойствам (3) и (6) $E[\xi] = E[\xi I_N] + E[\xi I_{\bar{N}}] = E[\xi I_N] = E[\eta I_{\bar{N}}]$. Но $E[\eta I_N] = 0$, поэтому по свойству (3) $E[\xi] = E[\eta I_{\bar{N}}] + E[\eta I_N] = E[\eta]$.

8. Пусть ξ – простая случайная величина со значениями x_1, \dots, x_k . Тогда $x_i \geq 0$ и $E[\xi] = \sum_{i=1}^k x_i P(\xi = x_i) = 0$. Следовательно, для любого i , такого что $x_i \neq 0$ выполнено $P(\xi = x_i) = 0 \implies P(\xi = 0) = 1$. Если $\xi \geq 0$, то $\forall n \in \mathbb{N}$ рассмотрим $A_n = \{\xi \geq \frac{1}{n}\}$.

$$P(A_n) = E[I_{A_n}] \leq E[n\xi I_{A_n}] \leq E[n\xi] \leq n E[\xi] = 0.$$

Но $A_n \downarrow \{\xi > 0\} \implies P(\xi > 0) = \lim_n P(A_n) = 0$.

9. В самом деле, пусть $B = \{\omega : \xi(\omega) > \eta(\omega)\}$. Тогда $E[\eta I_B] \leq E[\xi I_M] \leq E[\eta I_B]$ и, значит, $E[\xi I_B] = E[\eta I_B]$. В силу свойства аддитивности (3) $E[(\xi - \eta)I_B] = 0$ и по свойству (8) $(\xi - \eta)I_B = 0_{n.n.}$, откуда $P(B) = 0$.

□

1.11 Лекция от 22.11.2016

1.11.1 Классификация случайных величин. Формулы подсчёта математического ожидания

На предыдущих лекциях мы ввели понятие математического ожидания. Но мы пока что не знаем, как его считать. Разберёмся с этим.

Перед этим введём пару важных определений:

Определение 52. Пусть ξ – случайная величина на вероятностном пространстве (Ω, \mathcal{F}, P) . Тогда функцией распределения случайной величины ξ называют функцию $F_\xi : \mathbb{R} \rightarrow [0, 1]$, определяемую следующим образом:

$$F_\xi(x) = P(\xi \leq x).$$

Определение 53. Пусть ξ — случайная величина на вероятностном пространстве (Ω, \mathcal{F}, P) . Тогда *распределением* случайной величины ξ называется вероятностная мера P_ξ на $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ такая, что для любого $B \in \mathcal{B}(\mathbb{R})$

$$P_\xi(B) = P(\xi \in B).$$

Примечание. P_ξ действительно является вероятностной мерой, а F_ξ действительно является функцией распределения.

Теперь введём следующую классификацию случайных величин по типу функции распределения:

1. Если F_ξ дискретна, то будем называть случайную величину ξ *дискретной*.
2. Если F_ξ абсолютно непрерывна, то будем называть случайную величину ξ *абсолютно непрерывной*. В данном случае можно ввести понятие *плотности распределения*:

Определение 54. Пусть ξ — абсолютно непрерывная случайная величина, и существует неотрицательная функция $p_\xi(x)$ такая, что для любого $x \in \mathbb{R}$ выполнено, что

$$F_\xi(x) = P(\xi \leq x) = \int_{-\infty}^x p_\xi(t) dt.$$

Тогда $p_\xi(x)$ называют *плотностью распределения* случайной величины ξ .

3. Если же $F_\xi(x)$ сингулярна, то ξ называют *сингулярной*.

Начнём разбираться с математическим ожиданием с дискретных случайных величин. Вообще, если ξ дискретна, то множество значений $X = \xi(\Omega)$ не более, чем счётно, причем $P_\xi(X) = 1$.

Теорема 26. Пусть ξ — дискретная случайная величина, а $f(x)$ — некоторая борелевская функция, определённая на множестве значений X . Тогда

$$E[f(\xi)] = \sum_{x \in X} f(x) P(\xi = x).$$

Доказательство. Для начала предположим, что X конечно. Но в этом случае ξ — это простая случайная величина. Следовательно, $f(\xi)$ — это тоже простая случайная величина. Тогда по определению математического ожидания для простой величины:

$$\begin{aligned} E[f(\xi)] &= \sum_{z \in f(X)} f(z) P(f(\xi) = z) = \sum_{z \in f(X)} \left(\sum_{\substack{x \in X \\ f(x) = z}} z P(\xi = x) \right) \\ &= \sum_{z \in f(X)} \left(\sum_{\substack{x \in X \\ f(x) = z}} f(x) P(\xi = x) \right) = \sum_{x \in X} f(x) P(\xi = x). \end{aligned}$$

Теперь рассмотрим случай счётного X . Каким-либо образом занумеруем элементы X : $X = \{x_n\}_{n=1}^\infty$.

Предположим, что $f(x)$ неотрицательна. Тогда введём последовательность случайных величин $\{\eta_n\}_{n=1}^\infty$, устроенную следующим образом:

$$\eta_n = \sum_{i=1}^n f(x_i) I\{\xi = x_i\}.$$

Тогда $\eta_n \uparrow f(\xi)$. Следовательно, $E[\eta_n]$ монотонно возрастает и сходится к $E[f(\xi)]$. Однако, в силу линейности математического ожидания простой случайной величины получаем, что

$$\begin{aligned} E[f(\xi)] &= \lim_{n \rightarrow \infty} E[\eta_n] = \lim_{n \rightarrow \infty} \left(\sum_{i=1}^n f(x_i) E[I\{\xi = x_i\}] \right) \\ &= \lim_{n \rightarrow \infty} \left(\sum_{i=1}^n f(x_i) P(\xi = x_i) \right) = \sum_{i=1}^\infty f(x_i) P(\xi = x_i). \end{aligned}$$

В общем случае же представим $f(x)$ в виде $f^+(x) - f^-(x)$ и пользуемся доказанным ранее. \square

Теперь приступим к абсолютно непрерывным случайным величинам.

Теорема 27. Пусть ξ — абсолютно непрерывная случайная величина с плотностью $p_\xi(x)$. Тогда

$$E[\xi] = \int_{-\infty}^{+\infty} x p_\xi(x) dx.$$

Доказательство. Для начала рассмотрим неотрицательную случайную величину ξ . Введём последовательность простых случайных величин $\{\xi_n\}_{n=1}^\infty$, устроенную следующим образом:

$$\xi_n = \sum_{k=1}^{n \cdot 2^n} \frac{k-1}{2^n} I\left\{ \frac{k-1}{2^n} < \xi \leq \frac{k}{2^n} \right\}.$$

Заметим, что данная последовательность будет сходиться к ξ снизу: $\xi_n \uparrow \xi$. Однако

$$\begin{aligned} E[\xi_n] &= \sum_{k=1}^{n \cdot 2^n} \frac{k-1}{2^n} P\left(\frac{k-1}{2^n} < \xi \leq \frac{k}{2^n} \right) \\ &= \sum_{k=1}^{n \cdot 2^n} \frac{k-1}{2^n} \left(F_\xi\left(\frac{k}{2^n}\right) - F_\xi\left(\frac{k-1}{2^n}\right) \right) = \sum_{k=1}^{n \cdot 2^n} \frac{k-1}{2^n} \int_{\frac{k-1}{2^n}}^{\frac{k}{2^n}} p_\xi(x) dx \end{aligned}$$

Далее заметим, что

$$\int_{\frac{k-1}{2^n}}^{\frac{k}{2^n}} \left(x - \frac{1}{2^n} \right) p_\xi(x) dx \leq \frac{k-1}{2^n} \int_{\frac{k-1}{2^n}}^{\frac{k}{2^n}} p_\xi(x) dx \leq \int_{\frac{k-1}{2^n}}^{\frac{k}{2^n}} x p_\xi(x) dx.$$

Тогда, суммируя все такие элементы и вынося константу в ограничении снизу, получаем, что

$$\int_0^n x p_\xi(x) dx - \frac{1}{2^n} \leq \int_0^n \left(x - \frac{1}{2^n} \right) p_\xi(x) dx \leq E[\xi_n] \leq \int_0^n x p_\xi(x) dx \leq \int_0^\infty x p_\xi(x) dx.$$

Отсюда получаем, что

$$\mathbb{E}[\xi] = \int_0^{\infty} x p_{\xi}(x) dx.$$

Если же случайная величина ξ неположительна, то аналогичными рассуждениями получаем, что

$$\mathbb{E}[\xi] = \int_{-\infty}^0 x p_{\xi}(x) dx.$$

В общем случае представим ξ в виде суммы неотрицательной и неположительной случайных величин. Тогда получим желаемое. \square

Хорошо, мы научились считать матожидание абсолютно непрерывной случайной величины. Но хотелось бы научиться считать матожидание от функции от случайной величины. Следующая теорема говорит нам, как его вычислять.

Теорема 28. Пусть ξ — абсолютно непрерывная случайная величина с плотностью $p(x)$ и $f(x)$ — некоторая борелевская функция. Тогда

$$\mathbb{E}[f(x)] = \int_{-\infty}^{+\infty} f(x) p_{\xi}(x) dx.$$

Набросок доказательства. В случае монотонной функции $f(x)$ рассуждение работает. Теперь предположим более слабое условие — неотрицательность $f(x)$. Опять же, построим последовательность простых случайных величин $\{\eta_n\}_{n=1}^{\infty}$, устроенную следующим образом:

$$\sum_{k=1}^{n \cdot 2^n} \frac{k-1}{2^n} I \left\{ \frac{k-1}{2^n} < f(\xi) \leq \frac{k}{2^n} \right\}.$$

Тогда $\eta_n \uparrow f(\xi)$ и $\mathbb{E}[\eta_n]$ монотонно сходится к $\mathbb{E}[f(\xi)]$. Однако

$$\mathbb{E}[\eta_n] = \sum_{k=1}^{n \cdot 2^n} \frac{k-1}{2^n} \mathbb{P} \left(\frac{k-1}{2^n} < f(\xi) \leq \frac{k}{2^n} \right) = \sum_{k=1}^{n \cdot 2^n} \frac{k-1}{2^n} \mathbb{P} \left(\xi \in f^{-1} \left(\frac{k-1}{2^n}, \frac{k}{2^n} \right] \right)$$

Обозначим $f^{-1} \left(\frac{k-1}{2^n}, \frac{k}{2^n} \right]$ за $B_{k,n}$. Тогда

$$\mathbb{E}[\eta_n] = \sum_{k=1}^{n \cdot 2^n} \frac{k-1}{2^n} \int_{B_{k,n}} p_{\xi}(x) dx.$$

Дальше рассуждаем в стиле того, как мы доказывали формулу для $\mathbb{E}[\xi]$. \square

В принципе, для вычисления математического ожидания функции от случайной величины ξ достаточно знать её функцию распределения.

Теорема 29. ¹⁵ Пусть ξ — это случайная величина на вероятностном пространстве (Ω, \mathcal{F}, P) с функцией распределения F_ξ , а $f(x)$ — некая борелевская функция. Тогда математическое ожидание равно

$$E[f(\xi)] = \int_{-\infty}^{+\infty} f(x) dF_\xi(x).$$

Рассмотрим какой-нибудь пример.

Определение 55. Гамма-функцией $\Gamma(\lambda)$ ($\lambda > 0$) называют следующую функцию

$$\Gamma(\lambda) = \int_0^{+\infty} x^{\lambda-1} e^{-x} dx.$$

У гамма-функции есть несколько простых свойств:

- $\Gamma(n) = (n-1)!$ для любого $n \in \mathbb{N}$,
- $\Gamma(\lambda+1) = \lambda\Gamma(\lambda)$ для любого $\lambda > 0$,
- $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ (это не такое простое, ладно).

Определение 56. Пусть ξ — некоторая случайная величина, про которую известно, что

$$p_\xi(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} \text{ при } x \geq 0, \alpha, \beta > 0.$$

В таком случае говорят, что ξ подчиняется *гамма-распределению* с параметрами α и β . Обозначается $\xi \sim \text{Gamma}(\alpha, \beta)$.

Пусть у нас есть случайная величина $\xi \sim \text{Gamma}(\alpha, \beta)$. Найдём $E[\xi]$:

$$E[\xi] = \int_0^{+\infty} \frac{\beta^\alpha x^\alpha e^{-\beta x}}{\Gamma(\alpha)} dx = \left\{ u = x\beta, dx = \frac{du}{\beta} \right\} = \frac{\Gamma(\alpha+1)}{\beta\Gamma(\alpha)} = \frac{\alpha}{\beta}.$$

1.11.2 Независимость случайных величин и векторов

Определение 57. Случайные величины $\xi_1, \xi_2, \dots, \xi_n$ называются *независимыми в совокупности*, если для любых $x_1, x_2, \dots, x_n \in \mathbb{R}$

$$P(\xi_1 \leq x_1, \dots, \xi_n \leq x_n) = \prod_{i=1}^n P(\xi_i \leq x_i).$$

Примечание. Если $\xi_1, \xi_2, \dots, \xi_n$ независимы в совокупности, то любой поднабор тоже независим в совокупности.

Теперь возникает вопрос: а можно ли заменить знак \leq на что-либо другое? Да.

¹⁵ Это вовсе не теорема, а определение математического ожидания. И здесь не интеграл Римана, а интеграл Лебега-Стилтьеса. Попытки избежать теории меры и функциональный анализ проваливаются с треском. (А.Х.)

Теорема 30. *Случайные величины $\xi_1, \xi_2, \dots, \xi_n$ независимы тогда и только тогда, когда для любых множеств $B_1, B_2, \dots, B_n \in \mathcal{B}(\mathbb{R})$ выполнено, что*

$$P(\xi_1 \in B_1, \xi_2 \in B_2, \dots, \xi_n \in B_n) = \prod_{i=1}^n P(\xi_i \in B_i).$$

Пользуясь этой «теоремой», докажем один полезный факт.

Лемма. *Пусть $\xi_1, \xi_2, \dots, \xi_n$ — независимые случайные величины, а f_1, f_2, \dots, f_n — некоторые борелевские функции. Тогда случайные величины $f_1(\xi_1), f_2(\xi_2), \dots, f_n(\xi_n)$ тоже независимы.*

Доказательство. Проверим то, что выполнено определение:

$$P(f_1(\xi_1) \leq x_1, \dots, f_n(\xi_n) \leq x_n) = P(\xi_1 \in f_1^{-1}(-\infty, x_1], \dots, \xi_n \in f_n^{-1}(-\infty, x_n])$$

Заметим, что $f_i^{-1}(-\infty, x_i] \in \mathcal{B}(\mathbb{R})$. Тогда

$$P(\xi_1 \in f_1^{-1}(-\infty, x_1], \dots, \xi_n \in f_n^{-1}(-\infty, x_n]) = \prod_{i=1}^n P(\xi_i \in f_i^{-1}(-\infty, x_i]).$$

Отсюда получаем, что

$$P(f_1(\xi_1) \leq x_1, \dots, f_n(\xi_n) \leq x_n) = \prod_{i=1}^n P(f_i(\xi_i) \leq x_i). \quad \square$$

1.11.3 Математическое ожидание произведения независимых случайных величин

В дискретном вероятностном пространстве было верно следующее утверждение: если случайные величины независимы, то матожидание произведения равно произведению матожиданий. Верно ли оно в общем случае? Докажем, что да.

Теорема 31. *Пусть случайные величины ξ и η независимы, а их матожидания конечны. Тогда матожидание случайной величины $\xi\eta$ тоже конечно и*

$$E[\xi\eta] = E[\xi] E[\eta].$$

Доказательство. Для начала рассмотрим случай простых случайных величин. Тогда их можно представить в следующем виде:

$$\begin{aligned} \xi &= \sum_{i=1}^n x_i I\{\xi = x_i\} \\ \eta &= \sum_{j=1}^m y_j I\{\eta = y_j\} \end{aligned}$$

Теперь распишем матожидание произведения:

$$\begin{aligned} E[\xi\eta] &= \sum_{i=1}^n \sum_{j=1}^m x_i y_j E[I\{\xi = x_i\} I\{\eta = y_j\}] = \sum_{i=1}^n \sum_{j=1}^m x_i y_j P(\xi = x_i, \eta = y_j) \\ &= \sum_{i=1}^n \sum_{j=1}^m x_i y_j P(\xi = x_i) P(\eta = y_j) = \left(\sum_{i=1}^n x_i P(\xi = x_i) \right) \left(\sum_{j=1}^m y_j P(\eta = y_j) \right) \\ &= E[\xi] E[\eta]. \end{aligned}$$

Теперь перейдём к случаю неотрицательных случайных величин. Построим последовательности случайных величин $\{\xi_n\}_{n=1}^\infty$ и $\{\eta_n\}_{n=1}^\infty$ такие, что $\xi_n \uparrow \xi$ и $\eta_n \uparrow \eta$. Тогда $\xi_n \eta_n \uparrow \xi \eta$ и по определению

$$E[\xi \eta] = \lim_{n \rightarrow \infty} E[\xi_n \eta_n].$$

Но нужно следить так, чтобы $E[\xi_n \eta_n] = E[\xi_n] E[\eta_n]$. Для этого скажем, что ξ_n и η_n есть какие-то функции от ξ и η соответственно. Тогда ξ_n и η_n независимы. Следовательно,

$$\lim_{n \rightarrow \infty} E[\xi_n \eta_n] = \lim_{n \rightarrow \infty} E[\xi_n] E[\eta_n] = E[\xi] E[\eta].$$

В общем же случае разобьём обе случайные величины следующим образом:

$$\begin{aligned}\xi &= \xi^+ - \xi^-, \\ \eta &= \eta^+ - \eta^-\end{aligned}$$

Тогда $\xi \eta = (\xi \eta)^+ - (\xi \eta)^-$, где

$$\begin{aligned}(\xi \eta)^+ &= \xi^+ \eta^+ + \xi^- \eta^-, \\ (\xi \eta)^- &= \xi^+ \eta^- + \xi^- \eta^+.\end{aligned}$$

Заметим, что ξ^\pm независимо от η^\pm , так как это функции от независимых случайных величин. Тогда, пользуясь формулой для неотрицательных случайных величин, получаем, что

$$\begin{aligned}E[(\xi \eta)^+] &= E[\xi^+] E[\eta^+] + E[\xi^-] E[\eta^-] \\ E[(\xi \eta)^-] &= E[\xi^+] E[\eta^-] + E[\xi^-] E[\eta^+] \\ E[\xi \eta] &= (E[\xi^+] - E[\xi^-]) (E[\eta^+] - E[\eta^-]) = E[\xi] E[\eta].\end{aligned} \quad \square$$

1.12 Лекция от 29.11.2016

1.12.1 Многомерный случай

Многомерный случай — это одномерный случай.

Д.А. Шабанов

Ранее мы рассматривали только одну случайную величину за раз. Но порой возникают ситуации, когда их нужно рассматривать по несколько одновременно. Введём необходимую теорию для этого.

Определение 58. Пусть P — это некоторая вероятностная мера на \mathbb{R}^n ($n > 1$). Тогда назовём *функцией распределения* $F : \mathbb{R}^n \rightarrow [0, 1]$ такую, что для любых $x_1, x_2, \dots, x_n \in \mathbb{R}$ верно следующее:

$$F(x_1, x_2, \dots, x_n) = P((-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_n]).$$

Примечание. Для удобства это определение можно записать следующим образом (в данном случае $\vec{x} = (x_1, x_2, \dots, x_n)$):

$$F(\vec{x}) = P((-\infty, \vec{x}]).$$

Вообще, запись $(-\infty, \vec{x}]$ не несёт особого смысла, и её стоит понимать, как сокращение для $(-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_n]$.

Как и в одномерном случае, функция распределения обладает несколькими свойствами. Непрерывность справа и предельные значения легко обобщить, а как быть с неубыванием? Для этого это свойство несколько изменяют. В итоге получается следующая лемма.

Лемма (Основные свойства функции распределения). *Функция распределения в многомерном случае обладает следующими свойствами:*

1. Пусть есть некоторая точка \vec{x} и последовательность точек $\{\vec{x}^{(m)}\}_{m=1}^{\infty}$, про которую известно, что $\vec{x}^{(m)} \downarrow \vec{x}$ и $x_i^{(m+1)} \leq x_i^{(m)}$ для всех $i \in \{1, 2, \dots, n\}$. Тогда

$$\lim_{m \rightarrow \infty} F(\vec{x}^{(m)}) = F(\vec{x}).$$

Данное свойство выступает аналогом непрерывности справа в одномерном случае.

2.

$$\lim_{\substack{x_1 \rightarrow +\infty \\ \dots \\ x_n \rightarrow +\infty}} F(x_1, x_2, \dots, x_n) = 1.$$

В данном случае этот предел можно понимать, как повторный.

3. Для любого $i \in \{1, 2, \dots, n\}$ выполнено

$$\lim_{x_i \rightarrow -\infty} F(x_1, x_2, \dots, x_n) = 0.$$

4. Введём оператор $\Delta_{a,b}^i F(x_1, \dots, x_n)$, равный

$$F(x_1, \dots, x_{i-1}, b, x_{i+1}, \dots, x_n) - F(x_1, \dots, x_{i-1}, a, x_{i+1}, \dots, x_n).$$

Тогда для любого набора чисел $a_1, \dots, a_n, b_1, \dots, b_n$ такого, что $a_i \leq b_i$ для любого $i \in \{1, 2, \dots, n\}$, выполнено

$$\Delta_{a_1, b_1}^1 \Delta_{a_2, b_2}^2 \dots \Delta_{a_n, b_n}^n F(x_1, \dots, x_n) \geq 0.$$

Данное свойство обобщает неубывание на многомерное пространство.

Доказательство. В принципе, доказательство первых трёх пунктов идейно ничем не отличается от одномерного случая. Но всё равно докажем их.

1. Если $\{\vec{x}^{(m)}\}_{m=1}^{\infty}$ монотонно сходится к \vec{x} сверху, то (все сходимости монотонные):

$$\left\{ \begin{array}{c} (-\infty, x_1^{(m)}] \downarrow (-\infty, x_1] \\ \dots \\ (-\infty, x_n^{(m)}] \downarrow (-\infty, x_n] \end{array} \right\} \implies (-\infty, \vec{x}^{(m)}] \downarrow (-\infty, \vec{x}].$$

Следовательно, пользуясь непрерывностью меры, получаем, что

$$\lim_{m \rightarrow \infty} F(\vec{x}^{(m)}) = \lim_{m \rightarrow \infty} P((-\infty, \vec{x}^{(m)}]) = P((-\infty, \vec{x}]) = F(\vec{x}).$$

2. Для начала заметим следующее (все сходимости монотонные):

$$\left\{ \begin{array}{l} (-\infty, x_1] \xrightarrow{x_1 \rightarrow +\infty} (-\infty, +\infty) \\ \dots \\ (-\infty, x_n] \xrightarrow{x_n \rightarrow +\infty} (-\infty, +\infty) \end{array} \right\} \implies (-\infty, \vec{x}] \rightarrow \mathbb{R}^n.$$

Тогда по непрерывности меры получаем, что

$$\lim_{\substack{x_1 \rightarrow +\infty \\ \dots \\ x_n \rightarrow +\infty}} F(x_1, x_2, \dots, x_n) = P(\mathbb{R}^n) = 1.$$

3. Если $x_i \rightarrow -\infty$, то $(-\infty, x_i] \downarrow \emptyset$. Тогда

$$(-\infty, \vec{x}] \text{ сходитс} \text{я к } (-\infty, x_1] \times \dots \times \emptyset \times \dots \times (-\infty, x_n] = \emptyset.$$

Тогда по непрерывности меры получаем, что

$$\lim_{x_i \rightarrow -\infty} F(x_1, x_2, \dots, x_n) = P(\emptyset) = 0.$$

4. Следующее свойство достаточно муторно доказывать в общем случае, так что разберём его для двумерного пространства — идея доказательства та же. Распишем это, пользуясь определением оператора $\Delta_{a,b}^i$:

$$\begin{aligned} \Delta_{a_1,b_1}^1 \Delta_{a_2,b_2}^2 F(x_1, x_2) &= \Delta_{a_1,b_1}^1 (F(x_1, b_2) - F(x_1, a_2)) \\ &= F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2). \end{aligned}$$

Теперь воспользуемся определением функции распределения¹⁶:

$$\begin{aligned} F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) &= P((-\infty, b_1] \times (-\infty, b_2]) - \\ &- P((-\infty, a_1] \times (-\infty, b_2]) - P((-\infty, b_1] \times (-\infty, a_2]) + P((-\infty, a_1] \times (-\infty, a_2]) = \\ &= P((a_1, b_1] \times (-\infty, b_2]) - P((a_1, b_1] \times (-\infty, a_2]) = P((a_1, b_1] \times (a_2, b_2]) \geq 0. \end{aligned}$$

Можно заметить, что эта сумма есть ни что иное, как формула включений-исключений. Тогда в общем случае

$$\Delta_{a_1,b_1}^1 \Delta_{a_2,b_2}^2 \dots \Delta_{a_n,b_n}^n F(x_1, \dots, x_n) = P((a_1, b_1] \times \dots \times (a_n, b_n]) \geq 0. \quad \square$$

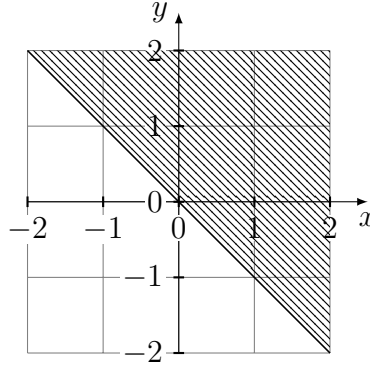
Последнее свойство на первый взгляд кажется странным. Возникает вопрос: а нельзя ли его заменить на неубывание по всем координатам? Увы, но нет.

Пример. Рассмотрим следующую функцию:

$$F(x, y) = \begin{cases} 1, & x + y \geq 0 \\ 0, & \text{иначе} \end{cases}$$

Изобразим её следующим образом: в заштрихованной области значение функции равно 1, в незаштрихованной — 0.

¹⁶На это выражение можно посмотреть и графически, но это роскошь двумерного случая.



Данная функция удовлетворяет свойствам 1-3 и не убывает по каждой координате. Проверим, выполняется ли свойство 4. Рассмотрим следующее значение:

$$\Delta_{-1,1}^1 \Delta_{-1,1}^2 F(x_1, x_2) = F(1, 1) - F(-1, 1) - F(1, -1) + F(-1, -1) = -1 < 0.$$

Как видим, оно не выполняется.

В одномерном случае между функцией распределения и вероятностной мерой есть биекция. Следующая теорема утверждает, что это верно и в многомерном случае:

Теорема 32. Пусть $F(x) : \mathbb{R}^n \rightarrow [0, 1]$ — некоторая функция, удовлетворяющая основным свойствам функции распределения. Тогда существует единственная вероятностная мера P на \mathbb{R}^n такая, что F — это её функция распределения. Другими словами, для любого $\vec{x} \in \mathbb{R}^n$

$$P((-\infty, \vec{x}]) = F(\vec{x}).$$

Теорема хорошая, доказывать мы её, конечно, не будем.

Рассмотрим несколько примеров многомерных распределений.

Самый простое, что может быть — это произведение функций распределения для одномерных случаев. Другими словами, функция распределения определяется следующим образом: пусть есть n одномерных функций распределения F_1, F_2, \dots, F_n . Тогда

$$F(x_1, x_2, \dots, x_n) = \prod_{i=1}^n F_i(x_i).$$

Проверим, что это действительно будет функцией распределения. Легко проверить, что свойства 1 – 3 выполнены. Теперь проверим последнее свойство. Заметим, что

$$\Delta_{a_1, b_1}^1 \Delta_{a_2, b_2}^2 \dots \Delta_{a_n, b_n}^n F(x_1, \dots, x_n) = \prod_{i=1}^n \Delta_{a_i, b_i}^1 F_i(x_i) \geq 0.$$

Теперь рассмотрим несколько другой случай. Пусть есть неотрицательная функция $p(x_1, x_2, \dots, x_n)$ такая, что

$$\int_{\mathbb{R}^n} p(t_1, t_2, \dots, t_n) dt_1 \dots dt_n = 1.$$

Тогда введём следующую функцию:

$$F(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} p(t_1, t_2, \dots, t_n) dt_1 \dots dt_n$$

Покажем, что это функция распределения. Первые три свойства легко проверяются. Теперь проверим четвёртое:

$$\Delta_{a_1, b_1}^1 \Delta_{a_2, b_2}^2 \cdots \Delta_{a_n, b_n}^n F(x_1, \dots, x_n) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} p(t_1, t_2, \dots, t_n) dt_1 \dots dt_n \geq 0.$$

В этом случае $p(x_1, x_2, \dots, x_n)$ принято называть *плотностью* функции распределения F .

Допустим, что есть некоторая функция распределения $F(x_1, x_2, \dots, x_n)$. В каком случае она имеет плотность? Ответ даёт теорема Радона-Никодима¹⁷: функция должна быть абсолютно непрерывной. Что это значит?

Определение 59. Вероятностная мера P называется *абсолютно непрерывной*, если для любого множества $B \in \mathcal{B}(\mathbb{R}^n)$ с лебеговой мерой 0 выполнено $P(B) = 0$.

Если функция F (а значит, и вероятностная мера P) абсолютно непрерывна, то плотность можно получить следующим образом:

$$p(x_1, x_2, \dots, x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \dots \partial x_n} F(x_1, \dots, x_n).$$

1.12.2 Случайные величины в многомерном случае

Для одномерного случая были введены функции распределения и распределения случайных величин. Обобщим это.

Определение 60. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — некоторые случайные величины. Тогда их *совместной функцией распределения* называется следующая функция

$$F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = P(\xi_1 \leq x_1, \dots, \xi_n \leq x_n).$$

Определение 61. *Совместным распределением* случайных величин $\xi_1, \xi_2, \dots, \xi_n$ называется вероятностная мера P_{ξ_1, \dots, ξ_n} на \mathbb{R}^n , заданная следующим правилом: для любого множества $B \in \mathcal{B}(\mathbb{R}^n)$

$$P_{\xi_1, \dots, \xi_n}(B) = P((\xi_1, \dots, \xi_n) \in B).$$

Заметим следующее:

1. F_{ξ_1, \dots, ξ_n} — это настоящая функция распределения.
2. P_{ξ_1, \dots, ξ_n} — это настоящая вероятностная мера.
3. F_{ξ_1, \dots, ξ_n} — это функция распределения для P_{ξ_1, \dots, ξ_n} .

В каких случаях совместная функция распределения хорошо считается? Например, в случае независимых случайных величин.

Теорема 33. Если случайные величины $\xi_1, \xi_2, \dots, \xi_n$ независимы (в совокупности) и функция распределения ξ_i равна F_{ξ_i} для всех $i \in \{1, 2, \dots, n\}$, то

$$F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \prod_{i=1}^n F_{\xi_i}(x_i).$$

¹⁷Эта теорема гласит о том, что если функция абсолютно непрерывна, то она представима в нужном нам виде. За доказательством обращайтесь к учебнику по функциональному анализу.

Доказательство. Распишем совместную функцию распределения по определению:

$$F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = P(\xi_1 \leq x_1, \dots, \xi_n \leq x_n) = \prod_{i=1}^n P(\xi_i \leq x_i) = \prod_{i=1}^n F_{\xi_i}(x_i). \quad \square$$

Теперь, по аналогии со вторым примером функции распределения, введём плотность совместного распределения.

Определение 62. Пусть есть случайные величины ξ_1, \dots, ξ_n с функцией совместного распределения F_{ξ_1, \dots, ξ_n} . Если у F_{ξ_1, \dots, ξ_n} есть плотность p_{ξ_1, \dots, ξ_n} , то p_{ξ_1, \dots, ξ_n} называют *совместной плотностью распределения*. В данном случае

$$F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} p_{\xi_1, \dots, \xi_n}(t_1, t_2, \dots, t_n) dt_1 \dots dt_n.$$

Теперь докажем один весьма полезный факт.

Теорема 34. Если у случайных величин $\xi_1, \xi_2, \dots, \xi_n$ есть совместная плотность, то в каждой из случайных величин есть плотность.

Доказательство. Вспомним определение плотности для одномерной величины: если $p_\xi(x)$ — плотность, то для любого $x \in \mathbb{R}$ выполнено

$$P(\xi \leq x) = \int_{-\infty}^x p_\xi(t) dt.$$

Попробуем получить нечто подобное, используя совместную плотность. По определениям совместной плотности и совместной функции распределения:

$$P(\xi_1 \leq x_1, \dots, \xi_n \leq x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} p_{\xi_1, \dots, \xi_n}(t_1, t_2, \dots, t_n) dt_1 \dots dt_n.$$

Допустим, что мы хотим получить плотность ξ_i . Тогда устремим все $x_k, k \in \{1, 2, \dots, n\} \setminus \{i\}$ к бесконечности. Тогда все условия, кроме i -го будут гарантированно выполнены. Изменяя порядок интегрирования (так как плотность абсолютно непрерывна, то это легально), получаем, что

$$P(\xi_i \leq x_i) = \int_{-\infty}^{x_i} \left(\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} p_{\xi_1, \dots, \xi_n}(t_1, t_2, \dots, t_n) dt_1 \dots dt_n \right) dt_i.$$

Сравнивая это с определением плотности в одномерном случае, получаем, что

$$p_{\xi_i}(x_i) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} p_{\xi_1, \dots, \xi_n}(t_1, t_2, \dots, t_n) dt_1 \dots dt_{i-1} dt_{i+1} \dots dt_n. \quad \square$$

А что можно сказать про плотности независимых случайных величин? Ответ на этот вопрос даёт следующая теорема.

Теорема 35. Пусть есть случайные величины $\xi_1, \xi_2, \dots, \xi_n$ с совместной плотностью распределения p_{ξ_1, \dots, ξ_n} , и для каждой из них существует плотность p_{ξ_i} . Тогда эти случайные величины независимы в совокупности тогда и только тогда, когда

$$p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \prod_{i=1}^n p_{\xi_i}(x_i).$$

Доказательство.

[\Rightarrow] Пусть случайные величины $\xi_1, \xi_2, \dots, \xi_n$ независимы. Тогда

$$F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \prod_{i=1}^n F_{\xi_i}(x_i) = \prod_{i=1}^n \left(\int_{-\infty}^{x_i} p_{\xi_i}(t_i) dt_i \right).$$

Заметим, что произведение интегралов можно представить в следующем виде:

$$\prod_{i=1}^n \left(\int_{-\infty}^{x_i} p_{\xi_i}(t_i) dt_i \right) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} \left(\prod_{i=1}^n p_{\xi_i}(t_i) \right) dt_1 \dots dt_n.$$

Однако

$$F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} p_{\xi_1, \dots, \xi_n}(t_1, t_2, \dots, t_n) dt_1 \dots dt_n.$$

Сравнивая формулы, получаем, что

$$p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \prod_{i=1}^n p_{\xi_i}(x_i).$$

[\Leftarrow] Пусть известно, что

$$p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \prod_{i=1}^n p_{\xi_i}(x_i).$$

Тогда

$$F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} \left(\prod_{i=1}^n p_{\xi_i}(t_i) \right) dt_1 \dots dt_n.$$

Разобьём его в произведение интегралов:

$$\int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} \left(\prod_{i=1}^n p_{\xi_i}(t_i) \right) dt_1 \dots dt_n = \prod_{i=1}^n \left(\int_{-\infty}^{x_i} p_{\xi_i}(t_i) dt_i \right).$$

Отсюда получаем, что

$$F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \prod_{i=1}^n \left(\int_{-\infty}^{x_i} p_{\xi_i}(t_i) dt_i \right) = \prod_{i=1}^n F_{\xi_i}(x_i).$$

А это означает независимость $\xi_1, \xi_2, \dots, \xi_n$.

□

1.12.3 Математическое ожидание в многомерном случае

Простой вопрос: а зачем мы ввели всё это? Ответ тоже прост — для подсчёта математического ожидания от функции многих случайных величин. Поставим вопрос формально.

Задача 1. Пусть есть n случайных величин $\xi_1, \xi_2, \dots, \xi_n$ с совместной плотностью распределения p_{ξ_1, \dots, ξ_n} и некоторая борелевская функция $f(x_1, \dots, x_n)$. Чему равно

$$E[f(\xi_1, \xi_2, \dots, \xi_n)]?$$

Ответ на этот вопрос даёт следующая теорема, которую мы сформулируем без доказательства:

Теорема 36. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — случайные величины с совместной плотностью распределения p_{ξ_1, \dots, ξ_n} , а $f(x_1, x_2, \dots, x_n)$ — некоторая борелевская функция. Тогда

$$E[f(\xi_1, \xi_2, \dots, \xi_n)] = \int_{\mathbb{R}^n} f(x_1, x_2, \dots, x_n) p_{\xi_1, \dots, \xi_n}(x_1, x_2, \dots, x_n) dx_1 \dots dx_n.$$

В данном случае мы пишем теорему без доказательства по той причине, потому что нам не хватает знаний по функциональному анализу и теории меры. Однако мы можем доказать этот результат в некоторых частных случаях. Попробуем доказать формулу для математического ожидания произведения двух случайных величин.

Теорема 37. Пусть ξ и η — некоторые (не обязательно независимые) неотрицательные случайные величины с совместной плотностью распределения $p_{\xi, \eta}$. Тогда

$$E[\xi\eta] = \int_{\mathbb{R}^2} xy p_{\xi, \eta}(x, y) dx dy.$$

Доказательство. Введём последовательность простых случайных величин $\{\delta_n\}_{n=1}^{\infty}$, устроенную следующим образом:

$$\delta_n = \sum_{i=1}^{n2^n} \sum_{j=1}^{n2^n} \frac{i-1}{2^n} \frac{j-1}{2^n} \mathbf{I} \left\{ \frac{i-1}{2^n} < \xi \leq \frac{i}{2^n}, \frac{j-1}{2^n} < \eta \leq \frac{j}{2^n} \right\}.$$

Заметим, что $\delta_n \uparrow \xi\eta$ и $\delta_{n+1} \geq \delta_n$. Тогда по определению математического ожидания

$$E[\xi\eta] = \lim_{n \rightarrow \infty} E[\delta_n].$$

Теперь посчитаем $E[\delta_n]$:

$$\begin{aligned} E[\delta_n] &= \sum_{k=1}^{n2^n} \sum_{j=1}^{n2^n} \frac{k-1}{2^n} \frac{j-1}{2^n} \mathbf{P} \left(\frac{k-1}{2^n} < \xi \leq \frac{k}{2^n}, \frac{j-1}{2^n} < \eta \leq \frac{j}{2^n} \right) \\ &= \sum_{k=1}^{n2^n} \sum_{j=1}^{n2^n} \frac{k-1}{2^n} \frac{j-1}{2^n} \left(\int_{\frac{k-1}{2^n}}^{\frac{k}{2^n}} \int_{\frac{j-1}{2^n}}^{\frac{j}{2^n}} p_{\xi, \eta}(x, y) dx dy \right). \end{aligned}$$

Теперь заметим, что

$$\int_0^n \int_0^n p_{\xi,\eta}(x, y) dx dy - \mathcal{O}\left(\frac{n}{2^n}\right) \leq \mathbb{E}[\delta_n] \leq \int_0^n \int_0^n p_{\xi,\eta}(x, y) dx dy.$$

Отсюда получаем, что

$$\mathbb{E}[\xi\eta] = \int_0^{+\infty} \int_0^{+\infty} p_{\xi,\eta}(x, y) dx dy.$$

□

1.13 Лекция от 06.12.2016

1.13.1 Полезные формулы

Пусть есть некоторый случайный вектор $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ с плотностью распределения $p_\xi(x)$. Докажем следующие три утверждения:

Теорема 38. Для любого $B \in \mathcal{B}(\mathbb{R}^n)$ выполнено, что

$$\mathbb{P}(\xi \in B) = \int \cdots \int_B p_\xi(\mathbf{x}) d\mathbf{x}.$$

Доказательство. Введём функцию $f(\mathbf{x}) = \mathbf{I}_B(\mathbf{x})$. Тогда

$$\mathbb{P}(\xi \in B) = \mathbb{E}[f(\xi)] = \int \cdots \int_{\mathbb{R}^n} \mathbf{I}_B(\mathbf{x}) p_\xi(\mathbf{x}) d\mathbf{x} = \int \cdots \int_B p_\xi(\mathbf{x}) d\mathbf{x}. \quad \square$$

Теорема 39. Для любой борелевской функции $f : \mathbb{R} \rightarrow \mathbb{R}$ и любого $i \in \{1, 2, \dots, n\}$ выполнено

$$\mathbb{E}[f(\xi_i)] = \int \cdots \int_{\mathbb{R}^n} f(x_i) p_\xi(\mathbf{x}) d\mathbf{x}.$$

Доказательство. Введём функцию $g(\mathbf{x}) = f(x_i)$. Тогда

$$\mathbb{E}[f(\xi_i)] = \mathbb{E}[g(\xi)] = \int \cdots \int_{\mathbb{R}^n} g(\mathbf{x}) p_\xi(\mathbf{x}) d\mathbf{x} = \int \cdots \int_{\mathbb{R}^n} f(x_i) p_\xi(\mathbf{x}) d\mathbf{x}. \quad \square$$

Теорема 40. Для любой борелевской функции $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ выполнено, что

$$\mathbb{P}(f(\xi) \in B) = \int \cdots \int_{\mathbf{x}: f(\mathbf{x}) \in B} p_\xi(\mathbf{x}) d\mathbf{x}.$$

Доказательство. Так как функция f борелевская, то для любого $B \in \mathcal{B}(\mathbb{R}^n)$ $f^{-1}(B)$ лежит в $\mathcal{B}(\mathbb{R}^n)$. Тогда, пользуясь рассуждениями, аналогичными теоремам выше, получаем, что

$$\mathbb{P}(f(\xi) \in B) = \mathbb{E}[\mathbf{I}_{f^{-1}(B)}(\xi)] = \int \cdots \int_{f^{-1}(B)} p_\xi(\mathbf{x}) d\mathbf{x} = \int \cdots \int_{\mathbf{x}: f(\mathbf{x}) \in B} p_\xi(\mathbf{x}) d\mathbf{x}. \quad \square$$

1.13.2 Формула замены переменных в кратном интеграле

Допустим, что у нас есть случайная величина ξ . Рассмотрим следующую функцию: $f(\xi) = (f_1(\xi), f_2(\xi), \dots, f_n(\xi))$. Как посчитать плотность полученного вектора?

Для ответа на этот вопрос нужно ввести формулу замены переменной в n -мерном интеграле. Сделаем это.

Определение 63. Матрицей Якоби \mathbf{J}_f для функции $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ матрица частных производных

$$\mathbf{J}_f = \mathbf{J}(f_1, \dots, f_m) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}_{m \times n}$$

Якобианом называют определяют определитель матрицы Якоби (если он существует).

Теорема 41 (Замена переменной). Пусть задано биективное отображение $\psi : \mathbb{R}^n \rightarrow \mathbb{R}^n$, переводящее область D' в D по следующему правилу:

$$\mathbf{x} = \psi(\mathbf{y}) \iff \begin{cases} x_1 = \psi_1(y_1, \dots, y_n) \\ x_2 = \psi_2(y_1, \dots, y_n) \\ \dots \\ x_n = \psi_n(y_1, \dots, y_n) \end{cases}$$

где x_1, \dots, x_n — “старые” координаты, а y_1, \dots, y_n — “новые” координаты. Далее, предположим, что все функции, задающие отображение, гладкие (имеют непрерывные производные первого порядка) на D' , а так же $0 < |\det \mathbf{J}(\psi_1, \dots, \psi_n)| < \infty$. Тогда, если существует интеграл

$$\int_D \cdots \int f(\mathbf{x}) d\mathbf{x} = \int_D \cdots \int f(x_1, \dots, x_n) dx_1 \dots dx_n,$$

то верна следующая формула

$$\int_D \cdots \int f(\mathbf{x}) d\mathbf{x} = \int_{D'} \cdots \int f(\psi(\mathbf{y})) |\det \mathbf{J}(\psi_1, \dots, \psi_n)| d\mathbf{y}.$$

Заметим, что $f^{-1}(f(\mathbf{x})) = \mathbf{x}$, если функция f обратима. Тогда сделаем замену $\mathbf{x} = f^{-1}(\mathbf{t})$. Тогда по формуле замены переменных

$$P(f(\xi) \in B) = \int_{\mathbf{x}: f(\mathbf{x}) \in B} \cdots \int p_\xi(\mathbf{x}) d\mathbf{t} = \int_B \cdots \int p_\xi(f^{-1}(\mathbf{t})) |\det \mathbf{J}(f_1^{-1}, \dots, f_n^{-1})| d\mathbf{t}.$$

Тогда $p_\xi(f^{-1}(\mathbf{t})) |\det \mathbf{J}(f_1^{-1}, \dots, f_n^{-1})|$ будет плотностью функции $f(\xi)$ в точке (t_1, \dots, t_n) . Докажем следующую формулу:

Теорема 42 (Формула свёртки). Пусть ξ и η — независимые случайные величины с плотностями p_ξ и p_η соответственно. Тогда

$$p_{\xi+\eta}(x) = \int_{-\infty}^{+\infty} p_\xi(y)p_\eta(x-y) dy = \int_{-\infty}^{+\infty} p_\xi(x-y)p_\eta(y) dy.$$

Доказательство. Распишем функцию распределения суммы, пользуясь независимостью случайных величин ξ и η :

$$P(\xi + \eta \leq z) = \iint_{(x,y): x+y \leq z} p_\xi(x)p_\eta(y) dx dy = \int_{-\infty}^z p_{\xi+\eta}(y) dy.$$

Сделаем замену переменных:

$$\begin{cases} u = y \\ v = x + y \end{cases} \iff \begin{cases} y = u \\ x = v - u \end{cases}$$

Получаем, что

$$\det \mathbf{J}(x, y) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} -1 & 1 \\ 1 & 0 \end{vmatrix} = 1.$$

Тогда

$$\iint_{(x,y): x+y \leq z} p_\xi(x)p_\eta(y) dx dy = \iint_{(u,v): u \leq z} p_\xi(v-u)p_\eta(u) du dv.$$

Далее, по теореме Фубини

$$\iint_{(u,v): u \leq z} p_\xi(v-u)p_\eta(u) du dv = \int_{-\infty}^z \int_{-\infty}^{+\infty} p_\xi(v-u)p_\eta(u) dv du.$$

Отсюда (по теореме Радона-Никодима) получаем, что почти везде выполнено, что

$$p_{\xi+\eta}(x) = \int_{-\infty}^{+\infty} p_\xi(y)p_\eta(x-y) dy. \quad \square$$

1.13.3 Дисперсия и ковариация.

Пусть (Ω, \mathcal{F}, P) — вероятностное пространство и ξ — случайная величина, для которой определено математическое ожидание $E[\xi]$.

Определение 64. Дисперсией случайной величины ξ называется величина

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

Определение 65. Пусть (ξ, η) — пара случайных величин. Их ковариацией называется

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

Определение 66. Случайные величины ξ, η называются *некоррелированными*, если их ковариация равна нулю: $\text{cov}(\xi, \eta) = 0$.

Определение 67. Если $D[\xi] > 0, D[\eta] > 0$, то величина

$$\rho(\xi, \eta) = \frac{\text{cov}(\xi, \eta)}{\sqrt{D[\xi] D[\eta]}}$$

называется *коэффициентом корреляции* случайных величин ξ и η .

Лемма (Свойства дисперсии и ковариации). Пусть ξ, η, χ — некоторые случайные величины. Тогда верны следующие утверждения:

1. Ковариация билинейна: $\text{cov}(\xi, a\eta + b\chi) = a \text{cov}(\xi, \eta) + b \text{cov}(\xi, \chi)$;
2. $D[\xi] = \text{cov}(\xi, \xi)$;
3. Для любого $c \in \mathbb{R}$ верно, что $D[c\xi] = c^2 D[\xi]$ и $D[\xi + c] = D[\xi]$;
4. Связь с матожиданием: $D[\xi] = E[\xi^2] - (E[\xi])^2$, $\text{cov}(\xi, \eta) = E[\xi\eta] - E[\xi]E[\eta]$;
5. Неравенство Коши-Буняковского

$$E[\xi\eta] \leq \sqrt{E[\xi^2] E[\eta^2]}.$$

6. $|\rho(\xi, \eta)| \leq 1$ и принимает значения ± 1 тогда и только тогда, когда ξ и η являются линейно зависимыми почти наверное.

Доказательство. Первые четыре свойства доказываются точно так же, как и в дискретном случае, так что повторно доказывать мы их не будем. Докажем последние два:

5. Посмотрим на следующее матожидание $E[(\xi + \lambda\eta)^2]$. Заметим, что оно неотрицательно как матожидание от неотрицательной случайной величины. Распишем его, используя свойство линейности:

$$0 \leq E[(\xi + \lambda\eta)^2] = E[\xi^2] + 2\lambda E[\xi\eta] + \lambda^2 E[\eta^2].$$

Теперь рассмотрим его с точки зрения многочлена от λ . Зная, что он неотрицателен, можно сказать, что дискриминант не превосходит нуля:

$$\frac{D}{4} = (E[\xi\eta])^2 - E[\xi^2] E[\eta^2] \leq 0.$$

Откуда и получаем желаемое.

6. $|\rho(\xi, \eta)| \leq 1$ получаем, используя неравенство Коши-Буняковского к случайным величинам $\alpha = \frac{\xi - E[\xi]}{\sqrt{D[\xi]}}$, $\beta = \frac{\eta - E[\eta]}{\sqrt{D[\eta]}}$:

$$\begin{aligned} |\rho(\xi, \eta)| &= \left| \frac{\text{cov}(\xi, \eta)}{\sqrt{D[\xi] D[\eta]}} \right| = \left| E \left[\left(\frac{\xi - E[\xi]}{\sqrt{D[\xi]}} \right) \left(\frac{\eta - E[\eta]}{\sqrt{D[\eta]}} \right) \right] \right| \leq \\ &\leq \sqrt{E \left[\left(\frac{\xi - E[\xi]}{\sqrt{D[\xi]}} \right)^2 \right] E \left[\left(\frac{\eta - E[\eta]}{\sqrt{D[\eta]}} \right)^2 \right]} = 1. \end{aligned}$$

Тогда следующие условия эквивалентны:

- $|\rho(\xi, \eta)| = 1$;
- Дискриминант в неравенстве Коши-Буняковского равен нулю;
- Существует корень уравнения $E[(\alpha + \lambda\beta)^2] = 0$ (обозначим его за λ_0);
- $\alpha + \lambda_0\beta = 0$ почти наверное;
- $\exists a, b \in \mathbb{R} : \xi = a\eta + b$ почти наверное.

□

Следствие. Если случайные величины ξ_1, \dots, ξ_n являются попарно некоррелированными, то

$$D[\xi_1 + \dots + \xi_n] = D[\xi_1] + \dots + D[\xi_n].$$

Доказательство. Распишем дисперсию, как сумму ковариаций:

$$D[\xi_1 + \dots + \xi_n] = \text{cov}(\xi_1 + \dots + \xi_n, \xi_1 + \dots + \xi_n) = \sum_{i,j=1}^n \text{cov}(\xi_i, \xi_j).$$

Так как ξ_i и ξ_j некоррелированы при $i \neq j$, то в полученной сумме ненулевыми могут быть только члены вида $\text{cov}(\xi_i, \xi_i)$. Тогда

$$D[\xi_1 + \dots + \xi_n] = \sum_{i=1}^n \text{cov}(\xi_i, \xi_i) = D[\xi_1] + \dots + D[\xi_n].$$

□

Следовательно, для независимых случайных величин это также выполняется.

Определение 68. Пусть $\xi = (\xi_1, \dots, \xi_n)$ — случайный вектор. Тогда математическим ожиданием случайного вектора ξ будем называть вектор, составленный из математических ожиданий компонент:

$$E[\xi] = (E[\xi_1], \dots, E[\xi_n]).$$

Назовем *матрицей ковариации* вектора матрицу порядка $n \times n$

$$D[\xi] = \|\text{cov}(\xi_i, \xi_j)\|_{i,j=1}^n.$$

Утверждение. Матрица ковариаций любого случайного вектора является *симметрической* и *неотрицательно определенной*.

Доказательство. Исходя из того, что $\text{cov}(\xi_i, \xi_j) = \text{cov}(\xi_j, \xi_i)$, матрица является симметрической.

Теперь докажем второе свойство. Пусть $\vec{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$. Тогда

$$\langle D[\xi]\vec{x}, \vec{x} \rangle = \sum_{i,j=1}^n \text{cov}(\xi_i, \xi_j) x_i x_j = \sum_{i,j=1}^n \text{cov}(\xi_i x_i, \xi_j x_j).$$

Теперь распишем по линейности:

$$\sum_{i,j=1}^n \text{cov}(\xi_i x_i, \xi_j x_j) = \sum_{j=1}^n \text{cov} \left(\sum_{i=1}^n x_i \xi_i, x_j \xi_j \right) = \text{cov} \left(\sum_{i=1}^n x_i \xi_i, \sum_{j=1}^n x_j \xi_j \right) = D \left[\sum_{i=1}^n x_i \xi_i \right] \geq 0.$$

□

1.14 Лекция от 20.01.2017

1.14.1 Дифференцирование характеристических функций

Продолжим обсуждать характеристические функции. Вообще говоря, характеристическая функция — это вполне себе функция.¹⁸ Поэтому хотелось бы уметь, например, брать производные от характеристической функции. Но как они будут устроены? Ответ на это даёт следующая

Теорема 43 (о производных характеристической функции). Пусть ξ — это некоторая случайная величина и $E[|\xi|^n] < +\infty$ для какого-то $n \in \mathbb{N}$. Тогда для всех $r \in \{1, 2, \dots, n\}$ существуют $\varphi_\xi^{(r)}(t)$, причём

$$\varphi_\xi^{(r)}(t) = E[(i\xi)^r e^{it\xi}].$$

Более того,

$$\begin{aligned} \varphi_\xi^{(r)}(t) &= i^r E[\xi^r] \\ \varphi_\xi(t) &= \sum_{k=0}^n \frac{(it)^k}{k!} E[\xi^k] + \frac{(it)^n}{n!} \mathcal{E}_n(t) \end{aligned}$$

где $|\mathcal{E}_n(t)| \leq 3 E[|\xi|^n]$ и $\lim_{t \rightarrow 0} \mathcal{E}_n(t) = 0$.

Доказательство. Докажем первое утверждение только для первой производной, ибо утверждение для старших степеней легко обобщается по индукции аналогичным образом. Рассмотрим производную функции $\varphi_\xi(t)$ в точке t_0 :

$$\varphi'_\xi(t) = \lim_{\Delta t \rightarrow 0} \frac{\varphi_\xi(t_0 + \Delta t) - \varphi_\xi(t_0)}{\Delta t} = \lim_{\Delta t \rightarrow 0} E \left[e^{i\xi t_0} \left(\frac{e^{i\xi \Delta t} - 1}{\Delta t} \right) \right].$$

Теперь заметим, что $\frac{e^{i\xi \Delta t} - 1}{\Delta t} \xrightarrow{\text{п.н.}} i\xi$ (просто поточечная сходимость). Далее,

$$\begin{aligned} |e^{i\xi \Delta t} - 1| &= |\cos(\xi \Delta t) + i \sin(\xi \Delta t) - 1| \\ &\leq |\cos(\xi \Delta t) - 1| + |\sin(\xi \Delta t)| \leq 2|\xi \Delta t| \end{aligned}$$

Следовательно, по теореме Лебега можно считать и предел математического ожидания:

$$\lim_{\Delta t \rightarrow 0} E \left[e^{i\xi t_0} \left(\frac{e^{i\xi \Delta t} - 1}{\Delta t} \right) \right] = E[e^{i\xi t_0} (i\xi)].$$

Если подставить в полученную формулу 0, то получим, что

$$\varphi_\xi^{(r)}(0) = E[(i\xi)^r e^0] = E[(i\xi)^r] = i^r E[\xi^r].$$

Теперь докажем последнее утверждение. Разложим e^{iy} в ряд Тейлора-Маклорена с остаточным членом в форме Лагранжа до $(n-1)$ -го порядка:

$$e^{iy} = \cos(y) + i \sin(y) = \sum_{k=0}^{n-1} \frac{(iy)^k}{k!} + \frac{(iy)^n}{n!} (\cos(\theta_1 y) + i \sin(\theta_2 y)),$$

¹⁸ Да ладно?

где $\theta_1, \theta_2 \in [0, 1]$.

Почему мы берём именно до $(n-1)$ -го порядка? Если взять до n -го порядка, то в остаточном члене будет $(n+1)$ -я степень, а матожидание $E[\xi^{n+1}]$ может быть бесконечным или вообще не существовать.

Теперь возьмём матожидание от этого выражения в точке $y = t\xi$:

$$E[e^{it\xi}] = \sum_{k=0}^{n-1} \frac{(it)^k}{k!} E[\xi^k] + \frac{(it)^n}{n!} E[\xi^n (\cos(\theta_1 t\xi) + i \sin(\theta_2 t\xi))]$$

Теперь добавим и вычтем $\frac{(it)^n}{n!} E[\xi^n]$:

$$E[e^{it\xi}] = \sum_{k=0}^n \frac{(it)^k}{k!} E[\xi^k] + \frac{(it)^n}{n!} E[\xi^n (\cos(\theta_1 t\xi) + i \sin(\theta_2 t\xi) - 1)]$$

Осталось понять, что из себя представляет этот остаточный член. Введём обозначение

$$\mathcal{E}_n(t) = E[\xi^n (\cos(\theta_1 t\xi) + i \sin(\theta_2 t\xi) - 1)].$$

Теперь заметим, что

$$|\mathcal{E}_n(t)| \leq E[|\xi^n (\cos(\theta_1 t\xi) + i \sin(\theta_2 t\xi) - 1)|] \leq 3 E[|\xi|^n]$$

Далее, при $t \rightarrow 0$ выполнена поточечная сходимость:

$$\cos(\theta_1 t\xi) + i \sin(\theta_2 t\xi) - 1 \xrightarrow{\text{п.н.}} 0$$

И, конечно же, выражение внутри матожидания ограничено:

$$|\xi^n (\cos(\theta_1 t\xi) + i \sin(\theta_2 t\xi) - 1)| \leq 3|\xi|^n$$

Тогда, снова по теореме Лебега, можно перейти к пределу матожидания:

$$\lim_{t \rightarrow 0} \mathcal{E}_n(t) = 0.$$

□

Примечание. Заметим, что условие $\lim_{t \rightarrow 0} \mathcal{E}_n(t) = 0$ равносильно тому, что

$$E[\xi^n (\cos(\theta_1 t\xi) + i \sin(\theta_2 t\xi) - 1)] = o(1).$$

Тогда формулу разложения можно записать иным образом:

$$\varphi_\xi(t) = \sum_{k=0}^n \frac{(it)^k}{k!} E[\xi^k] + o\left(\frac{t^n}{n!}\right)$$

Следствие. Если $E[\xi^2] \leq \infty$, то при $t \rightarrow 0$ верна следующая формула

$$\varphi_\xi(t) = 1 + it E[\xi] - \frac{t^2}{2} E[\xi^2] + o(t^2)$$

1.14.2 Харатеристическая функция случайного вектора и её свойства

Ранее мы работали с характеристическими функциями случайных величин. Обобщим это понятие на случайные вектора:

Определение 69. Пусть $\xi \in \mathbb{R}^n$ — случайный вектор. Тогда *характеристической функцией* случайного вектора ξ будем называть функцию

$$\varphi_\xi(\mathbf{t}) = \mathbb{E}[e^{i\langle \xi, \mathbf{t} \rangle}], \mathbf{t} \in \mathbb{R}^n$$

В данном случае $\langle \xi, \mathbf{t} \rangle = \sum_{k=1}^n t_k \xi_k$ — скалярное произведение.

Свойства данной характеристической функции остаются теми же, что и в одномерном случае. Например:

Теорема 44 (единственности). Пусть ξ и η — некоторые случайные векторы в \mathbb{R}^n . Тогда

$$\forall \mathbf{t} \in \mathbb{R}^n \quad \varphi_\xi(\mathbf{t}) = \varphi_\eta(\mathbf{t}) \iff \xi \stackrel{d}{=} \eta.$$

С характеристическими функциями случайных векторов связан ещё один факт. Как вы помните, если случайные величины ξ_1, \dots, ξ_n независимы в совокупности, то функция распределения случайного вектора $\xi = (\xi_1, \dots, \xi_n)$ разбивалась в произведение функций распределения для случайных величин ξ_1, \dots, ξ_n . Оказывается, что похожее утверждение верно и для характеристических функций.

Теорема 45 (критерий независимости для характеристических функций). Случайные величины ξ_1, \dots, ξ_n независимы в совокупности тогда и только тогда, когда характеристическая функция вектора $\xi = (\xi_1, \dots, \xi_n)$ разбивается в произведение характеристических функций ξ_i :

$$\forall \mathbf{t} = (t_1, \dots, t_n) \in \mathbb{R}^n \quad \varphi_\xi(\mathbf{t}) = \prod_{k=1}^n \varphi_{\xi_k}(t_k).$$

Доказательство. Как обычно, будем доказывать по очереди.

[\Rightarrow] Предположим, что случайные величины ξ_1, \dots, ξ_n независимы в совокупности. Тогда рассмотрим характеристическую функцию вектора $\xi = (\xi_1, \dots, \xi_n)$:

$$\varphi_\xi(\mathbf{t}) = \mathbb{E}[e^{i\langle \xi, \mathbf{t} \rangle}] = \mathbb{E}[e^{i \sum_{k=1}^n t_k \xi_k}] = \mathbb{E}\left[\prod_{k=1}^n e^{it_k \xi_k}\right]$$

Но, как известно, функции (борелевские) от независимых случайных величин — это независимые случайные величины, а матожидание произведения независимых случайных величин разбивается в произведение матожиданий. Тогда

$$\varphi_\xi(\mathbf{t}) = \prod_{k=1}^n \mathbb{E}[e^{it_k \xi_k}] = \prod_{k=1}^n \varphi_{\xi_k}(t_k).$$

[\Leftarrow] Теперь предположим, что

$$\forall \mathbf{t} = (t_1, \dots, t_n) \in \mathbb{R}^n \quad \varphi_\xi(\mathbf{t}) = \prod_{k=1}^n \varphi_{\xi_k}(t_k).$$

Покажем, что функция распределения $F_\xi(\mathbf{t})$ аналогичным образом разбивается в произведение. Для этого рассмотрим случайный вектор $\eta = (\eta_1, \dots, \eta_n)$, где случайные величины η_1, \dots, η_n независимы в совокупности и $\eta_i \stackrel{d}{=} \xi_i$ для $i \in \{1, \dots, n\}$. Понятно, что его функция распределения равна

$$F_\eta(\mathbf{t}) = \prod_{k=1}^n F_{\eta_k}(t_k) = \prod_{k=1}^n F_{\xi_k}(t_k)$$

Осталось показать, что $\varphi_\eta(\mathbf{t}) = \varphi_\xi(\mathbf{t})$. Действительно, согласно теореме единственности и первому пункту получаем, что

$$\varphi_\eta(\mathbf{t}) = \prod_{k=1}^n \varphi_{\eta_k}(t_k) = \prod_{k=1}^n \varphi_{\xi_k}(t_k) = \varphi_\eta(\mathbf{t}).$$

Следовательно, по теореме единственности

$$\xi \stackrel{d}{=} \eta \implies F_\xi(\mathbf{t}) = \mathcal{F}_\eta(\mathbf{t}) = \prod_{k=1}^n F_{\xi_k}(t_k).$$

А это означает, что ξ_1, \dots, ξ_n независимы в совокупности. □

1.14.3 Центральная предельная теорема и примеры её применения

Наконец-то мы достигли «жемчужины» нашего курса — центральной предельной теоремы, на которой, можно сказать, держится математическая статистика. Но на самом деле нет. Перед этим сформулируем теорему, необходимую для её доказательства.

Теорема 46 (непрерывности). Пусть есть $\{\xi_n \mid n \in \mathbb{N}\}$, ξ — случайные величины и $\varphi_n(t) = \varphi_{\xi_n}(t)$.

1. Пусть $\xi_n \xrightarrow{d} \xi$. Тогда $\varphi_n(t) \rightarrow \varphi_\xi(t)$ для любого $t \in \mathbb{R}$.
2. Пусть существует непрерывная в нуле функция $\varphi(t)$ такая, что

$$\forall t \in \mathbb{R} \lim_{n \rightarrow \infty} \varphi_n(t) = \varphi(t).$$

Тогда $\varphi_\xi(t) = \varphi(t)$ и $\xi_n \xrightarrow{d} \xi$.

Доказательство. Без доказательства.¹⁹ □

Впрочем, делать тут особо нечего: ЦПТ является простым следствием этой теоремы!

Теорема 47 (центральная предельная). Пусть $\{\xi_n \mid n \in \mathbb{N}\}$ — последовательность независимых и одинаково распределённых случайных величин, про которую известно, что $E[\xi] < \infty$ и $0 < D[\xi_1] < +\infty$. Далее, введём обозначения: $S_n = \xi_1 + \dots + \xi_n$, $a = E[\xi_1]$ и $\sigma^2 = D[\xi_1]$. Тогда

$$\frac{S_n - E[S_n]}{\sqrt{D[S_n]}} \xrightarrow{d} \mathcal{N}(0, 1).$$

¹⁹Мне это уже надоело, ей богу. Самое вкусное идёт без доказательства.

Доказательство. Утверждение равносильно (по теореме непрерывности) тому, что характеристическая функция случайной величины

$$\mu_n = \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathbb{D}[S_n]}}$$

стремится к $e^{-t^2/2}$. Покажем это.

Для этого немного преобразуем μ_n , пользуясь независимостью ξ_i :

$$\mu_n = \frac{\sum_{k=1}^n (\xi_k - \mathbb{E}[\xi_k])}{\sqrt{\sum_{k=1}^n \mathbb{D}[\xi_k]}} = \sum_{k=1}^n \frac{\xi_k - a}{\sqrt{n}\sigma} = \left[\eta_k = \frac{\xi_k - a}{\sigma} \right] = \sum_{k=1}^n \frac{\eta_k}{\sqrt{n}}.$$

Заметим, что $\mathbb{E}[\eta_k] = 0$ и $\mathbb{D}[\eta_k] = 1$. Отсюда получаем, что $\mathbb{E}[\eta_k^2] = 1$ и

$$\varphi_{\eta_k}(t) = 1 - \frac{t^2}{2} + o(t^2).$$

Теперь посмотрим на характеристическую функцию μ_n и воспользуемся тем, что η_k независимы в совокупности, как функции от независимых случайных величин:

$$\varphi_{\mu_n}(t) = \mathbb{E} \left[e^{it \sum_{k=1}^n \frac{\eta_k}{\sqrt{n}}} \right] = \prod_{k=1}^n \varphi_{\eta_k} \left(\frac{t}{\sqrt{n}} \right) = \left(1 - \frac{t^2}{2n} + o \left(\frac{t^2}{n^2} \right) \right)^n$$

Но это, как известно, стремится к $e^{-t^2/2}$ при $n \rightarrow \infty$. Тогда $\mu_n \xrightarrow{d} \mathcal{N}(0, 1)$, что и требовалось доказать. \square

Примечание. Утверждение ЦПТ равносильно следующему:

$$\sqrt{n} \left(\frac{S_n}{n} - \mathbb{E}[\xi_1] \right) \xrightarrow{d} \mathcal{N}(0, \mathbb{D}[\xi_1]).$$

Следствие. С помощью ЦПТ можно оценить скорость сходимости в законе больших чисел. Неформально,

$$\frac{S_n}{n} - \mathbb{E}[\xi_1] = O \left(\frac{1}{\sqrt{n}} \right).$$

Теперь рассмотрим пример применения ЦПТ, более приближенный к реальности.

Пример. Допустим, что во время эксперимента мы получили 10^4 чисел, причём каждое из них измерено с точностью 10^{-6} . В каких пределах лежит суммарная ошибка, если все числа получены независимо друг от друга?

Решение. Пусть ξ_i — погрешность в вычислении i -го числа. Будем считать, что $\xi_i \sim U(-10^{-6}, 10^{-6})$. Тогда $\mathbb{E}[\xi_i] = 0$ и $\sigma^2 = \mathbb{D}[\xi_i] = \frac{10^{-12}}{3}$. Далее, введём обозначение $S_n = \xi_1 + \dots + \xi_n$ — суммарная ошибка первых n измерений. Тогда по центральной предельной теореме для достаточно больших n

$$\mathbb{P} \left(\frac{|S_n|}{\sigma\sqrt{n}} \leq x \right) \approx \Phi(x) - \Phi(-x).$$

Теперь возьмём x такое, чтобы $\Phi(x) - \Phi(-x) \geq 0.99$. Можно проверить, что $x \approx 2.58$ вполне подходит. Тогда с вероятностью 0.99 выполнено, что

$$|S_n| \leq x\sigma\sqrt{n} = 258 \cdot \frac{10^{-6}}{3} \approx 1.5 \cdot 10^{-4}.$$

\square

Хорошо, ЦПТ дала нам скорость сходимости для ЗБЧ. А насколько быстро сходится сама ЦПТ? Ответ на этот вопрос даёт следующая

Теорема 48 (Берри, Эсseen). Пусть $\{\xi_n \mid n \in \mathbb{N}\}$ — последовательность независимых и одинаково распределённых случайных величин. Далее, пусть выполнены следующие ограничения: $E[|\xi_1|^3] < +\infty$ и $0 < D[\xi_1] < +\infty$. Введём следующие обозначения:

$$S_n = \sum_{k=1}^n \xi_k$$

$$T_n = \frac{S_n - E[S_n]}{\sqrt{D[S_n]}}$$

Тогда

$$\sup_{x \in \mathbb{R}} |F_{T_n}(x) - \Phi(x)| \leq \frac{C E(\xi_1 - E[\xi_1])^3}{D[\xi_1] \sqrt{n}},$$

где C — это некоторая постоянная, не зависящая от ξ_n .

Доказательство. Без доказательства. □

Что же известно про C в теореме Берри-Эсseenа? Вообще, первоначально было доказано, что $C \leq 100$ и $C \geq \frac{1}{\sqrt{2\pi}}$, причём оценка снизу не улучшаема. Верхнюю же оценку постепенно снижали. В 2013 году доказали, что $C \leq 0.48$.

Как показывает следующий пример, центральную предельную теорему можно использовать для оценки параметров распределения случайной величины.

Пример. Пусть нам известно, что $\xi_1, \dots, \xi_n \sim \mathcal{N}(a, \sigma^2)$ и n достаточно большое, но не известны a и σ . Мы провели n независимых наблюдений и получили значения ξ_1, \dots, ξ_n . Как оценить a и σ ?

Решение. Для начала оценим a . Пусть $\bar{\xi}_n = \frac{\xi_1 + \dots + \xi_n}{n}$. Вообще говоря, хотелось бы сразу сказать, что по ЗБЧ

$$\bar{\xi}_n \approx a.$$

Но это верно с какой-то неизвестной погрешностью — у нас же конечное число измерений. Придётся воспользоваться ЦПТ для оценки погрешности:

$$\sqrt{n} \left(\frac{\bar{\xi}_n - a}{\sigma} \right) \xrightarrow{d} N(0, 1) \implies P \left(\sqrt{n} \left| \frac{\bar{\xi}_n - a}{\sigma} \right| \geq y \right) \approx \Phi(y) - \Phi(-y).$$

Возьмём $y \approx 2.58$. Тогда с вероятностью 0.99 верно, что

$$\bar{\xi}_n - \frac{\sigma y}{\sqrt{n}} \leq a \leq \bar{\xi}_n + \frac{\sigma y}{\sqrt{n}}$$

Проблема — нам всё ещё не известно значение σ . Как его найти? Это мы уже обсудим в следующий раз. □

1.15 Лекция от 27.01.2017

1.15.1 Сходимость случайных векторов

Ранее мы изучали сходимость для случайных величин. Обобщим это для случайных векторов, введя те же типы сходимости.

Определение 70. Пусть $\{\xi_n \mid n \in \mathbb{N}\}, \xi$ — случайные векторы из \mathbb{R}^m . Тогда

1. Будем говорить, что последовательность $\{\xi_n\}$ сходится к ξ *почти наверное* (или с вероятностью 1), если

$$P\left(\lim_{n \rightarrow \infty} \xi_n = \xi\right) = 1.$$

2. Будем говорить, что последовательность $\{\xi_n\}$ сходится к ξ *по вероятности*, если для любого $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(\|\xi_n - \xi\|_2 \geq \varepsilon) = 0$$

В данном случае $\|x\|_2 = \sqrt{x_1^2 + \dots + x_m^2}$.

3. Будем говорить, что последовательность $\{\xi_n\}$ сходится к ξ *по распределению*, если для любой непрерывной ограниченной функции $h : \mathbb{R}^m \mapsto \mathbb{R}$ выполнено, что

$$\lim_{n \rightarrow \infty} E[h(\xi_n)] = E[h(\xi)].$$

Так, стоп. Первые два похожи на определения в одномерном случае, но третье кардинально отличается от одномерного случая — ведь там была сходимость функций распределения. Сразу возникает вопрос: а совпадают ли эти определения для одномерного случая? Ведь случайная величина тоже является случайным вектором. Оказывается, что всё хорошо. Это утверждает следующая

Теорема 49 (об эквивалентности определений сходимости по распределению). Пусть $\{\xi_n \mid n \in \mathbb{N}\}, \xi$ — случайные векторы из \mathbb{R}^m .

1. Пусть $m = 1$. Тогда $\xi_n \xrightarrow{d} \xi$ (в смысле случайных величин), тогда и только тогда, когда для любой непрерывной ограниченной функции $h : \mathbb{R} \mapsto \mathbb{R}$ выполнено, что $\lim_{n \rightarrow \infty} E[h(\xi_n)] = E[h(\xi)]$.
2. Если $m > 1$ и F_ξ непрерывна в \mathbb{R}^m , то $\xi_n \xrightarrow{d} \xi$ равносильно тому, что для любого $x \in \mathbb{R}^m$ $\lim_{n \rightarrow \infty} F_{\xi_n}(x) = F_\xi(x)$.

Доказательство. Как обычно, принимаем без доказательства. □

Следующее упражнение может быть весьма полезно для доказательства некоторых фактов.

Упражнение. Пусть $\{\xi_n \mid n \in \mathbb{N}\}, \xi$ — случайные векторы в \mathbb{R}^m и $\xi_n = (\xi_n^{(1)}, \xi_n^{(2)}, \dots, \xi_n^{(m)})$. Тогда

1. $\xi_n \xrightarrow{\text{п.н.}} \xi \iff \forall i \in \{1, 2, \dots, m\} \xi_n^{(i)} \xrightarrow{\text{п.н.}} \xi^{(i)}$
2. $\xi_n \xrightarrow{P} \xi \iff \forall i \in \{1, 2, \dots, m\} \xi_n^{(i)} \xrightarrow{P} \xi^{(i)}$

3. $\xi_n \xrightarrow{d} \xi \implies \forall i \in \{1, 2, \dots, m\} \xi_n^{(i)} \xrightarrow{d} \xi^{(i)}$, причём в обратную сторону это неверно.

Вспомним, что в одномерном случае существовала «иерархия» сходимостей: почти наверное \implies по вероятности \implies по распределению. Оказывается, что она остаётся верна и в многомерном случае.

Лемма (о взаимосвязи видов сходимости). Пусть $\{\xi_n \mid n \in \mathbb{N}\}$, ξ — случайные векторы в \mathbb{R}^m . Тогда

$$1. \xi_n \xrightarrow{n.n.} \xi \implies \xi_n \xrightarrow{P} \xi$$

$$2. \xi_n \xrightarrow{P} \xi \implies \xi_n \xrightarrow{d} \xi$$

Доказательство. Вообще говоря, тут и доказывать нечего. Первый пункт автоматически следует из утверждения о равносильности общей и покомпонентной сходимости для сходимостей почти наверное и по вероятности. Доказательство второго пункта аналогично одномерному случаю. \square

1.15.2 Теорема о наследовании сходимости, лемма Slutsky и их применения

Сформулируем и докажем две важные теоремы.

Теорема 50 (о наследовании сходимости). Пусть $\{\xi_n \mid n \in \mathbb{N}\}$, ξ — случайные векторы в \mathbb{R}^m , а $h : \mathbb{R}^m \mapsto \mathbb{R}^n$ — это некоторая функция.

1. Если $\xi_n \xrightarrow{n.n.} \xi$ и h непрерывна почти всюду относительно распределения ξ (то есть существует $B \in \mathcal{B}(\mathbb{R}^m)$ такое, что h непрерывна на B и $P(\xi \in B) = 1$), то $h(\xi_n) \xrightarrow{n.n.} h(\xi)$.

2. Аналогично, если $\xi_n \xrightarrow{P} \xi$ и h непрерывна почти всюду относительно распределения ξ , то $h(\xi_n) \xrightarrow{P} h(\xi)$.

3. Если $\xi_n \xrightarrow{d} \xi$ и h непрерывна, то $h(\xi_n) \xrightarrow{d} h(\xi)$.

Доказательство. Вообще говоря, первый и третий пункт по сложности сравнимы с проверкой определений. Самое сложное — это второй пункт.

1. Для начала докажем утверждение для сходимости почти наверное. Рассмотрим вероятность

$$P\left(\lim_{n \rightarrow \infty} h(\xi_n) = h(\xi)\right)$$

Заметим, что для любого $\xi \in B$ это равносильно тому, что $\xi_n \xrightarrow{n.n.} \xi$ (по непрерывности h). Тогда эту вероятность можно ограничить снизу:

$$\begin{aligned} P\left(\lim_{n \rightarrow \infty} h(\xi_n) = h(\xi)\right) &\geq P\left(\lim_{n \rightarrow \infty} \xi_n = \xi, \xi \in B\right) \\ &\geq P\left(\lim_{n \rightarrow \infty} \xi_n = \xi\right) P(\xi \in B) = 1. \end{aligned}$$

Это и означает, что $h(\xi_n) \xrightarrow{n.n.} h(\xi)$.

2. Теперь докажем утверждение для сходимости по вероятности от противного. Допустим, что $h(\xi_n) \not\xrightarrow{P} h(\xi)$. Тогда существуют положительные ε_0 и δ_0 и подпоследовательность $\{\xi_{n_k} \mid k \in \mathbb{N}\}$ такие, что

$$P(\|h(\xi_{n_k}) - h(\xi)\| \geq \varepsilon_0) \geq \delta_0$$

Однако $\xi_{n_k} \xrightarrow{P} \xi$. Тогда есть ещё одна подпоследовательность $\{\xi_{n_{k_s}} \mid s \in \mathbb{N}\}$ такая, что $\xi_{n_{k_s}} \xrightarrow{п.н.} \xi$. Тогда согласно первому пункту $h(\xi_{n_{k_s}}) \xrightarrow{п.н.} h(\xi)$. Но по лемме о взаимосвязи сходимостей $h(\xi_{n_{k_s}}) \xrightarrow{P} h(\xi)$. Получаем противоречие. Следовательно, $h(\xi_n) \xrightarrow{P} h(\xi)$.

3. Предположим, что $\xi_n \xrightarrow{d} \xi$ и $h : \mathbb{R}^n \mapsto \mathbb{R}^k$ — непрерывная функция. Возьмём произвольную ограниченную и непрерывную функцию $f : \mathbb{R}^k \mapsto \mathbb{R}$. Тогда композиция $f \circ h : \mathbb{R}^n \mapsto \mathbb{R}$ — ограниченная и непрерывная функция. Следовательно, так как $\xi_n \xrightarrow{d} \xi$, то

$$E[f(h(\xi_n))] \rightarrow E[f(h(\xi))].$$

Но это означает, что $h(\xi_n) \xrightarrow{d} h(\xi)$.

□

Упражнение. Предположим, что некоторый случайный вектор ξ сходится к константе по распределению. Докажите, что он сходится к ней же по вероятности.

Лемма (Слуцкий). Пусть $\{\xi_n \mid n \in \mathbb{N}\}$ и $\{\eta_n \mid n \in \mathbb{N}\}$ — две последовательности случайных величин. Далее, пусть существуют константа и случайная величина ξ такие, что $\xi_n \xrightarrow{d} \xi$ и $\eta_n \xrightarrow{d} C$. Тогда

$$1. \xi_n + \eta_n \xrightarrow{d} \xi + C,$$

$$2. \xi_n \eta_n \xrightarrow{d} C\xi$$

Доказательство. Докажем это утверждение только для суммы, ибо для произведения всё аналогично.

Пусть t — это точка непрерывности $F_{\xi+C}$. Выберем сколь угодно малый ε такой, что $t \pm \varepsilon$ тоже являются точками непрерывности $F_{\xi+C}$. Теперь рассмотрим $F_{\xi_n+\eta_n}(t) = P(\xi_n + \eta_n \leq t)$. По формуле полной вероятности:

$$F_{\xi_n+\eta_n}(t) = P(\xi_n + \eta_n \leq t, \eta_n > t - \varepsilon) + P(\xi_n + \eta_n \leq t, \eta_n \leq t - \varepsilon).$$

Увеличим вторую вероятность следующим образом: забудем про первое условие, а второе заменим на $|\eta_n - t| \geq \varepsilon$. Тогда

$$F_{\xi_n+\eta_n}(t) \leq P(\xi_n + \eta_n \leq t, \eta_n > t - \varepsilon) + P(|\eta_n - t| \geq \varepsilon).$$

Теперь увеличим первую вероятность. Из условий следует, что $\xi_n + C \leq t + \varepsilon$. Тогда заменим оба условия на это. Следовательно,

$$F_{\xi_n+\eta_n}(t) \leq P(\xi_n + C \leq t + \varepsilon) + P(|\eta_n - t| \geq \varepsilon).$$

Осталось понять, куда стремятся полученные вероятности. Так как $\xi_n \xrightarrow{d} \xi$, то $\xi_n + C \xrightarrow{d} \xi + C$. Учитывая то, что $t + \varepsilon$ есть точка непрерывности, то первая вероятность стремится

к $F_{\xi+C}(t+\varepsilon)$. Вторая же вероятность стремится к нулю, так как $\eta_n \xrightarrow{d} C$ (и, следовательно, $\eta_n \xrightarrow{P} C$).

Отсюда получаем, что

$$\overline{\lim}_{n \rightarrow \infty} F_{\xi_n + \eta_n}(t) \leq F_{\xi+C}(t + \varepsilon).$$

Аналогичными рассуждениями получаем, что

$$P(\xi_n + \eta_n \geq t) \geq P(\xi_n + C \geq t + \varepsilon) + P(|\eta_n - t| \geq \varepsilon).$$

Тогда

$$\overline{\lim}_{n \rightarrow \infty} (1 - F_{\xi_n + \eta_n}(t)) \geq 1 - F_{\xi+C}(t + \varepsilon) \text{ и } \underline{\lim}_{n \rightarrow \infty} F_{\xi_n + \eta_n}(t) \geq F_{\xi+C}(t - \varepsilon).$$

Получаем следующую цепочку ограничений:

$$F_{\xi+C}(t - \varepsilon) \leq \underline{\lim}_{n \rightarrow \infty} F_{\xi_n + \eta_n}(t) \leq \overline{\lim}_{n \rightarrow \infty} F_{\xi_n + \eta_n}(t) \leq F_{\xi+C}(t + \varepsilon).$$

Устремляя ε к нулю и пользуясь тем, что t — это точка непрерывности $F_{\xi+C}$, получаем, что

$$\lim_{n \rightarrow \infty} F_{\xi_n + \eta_n}(t) = F_{\xi+C}(t). \quad \square$$

Теперь посмотрим, как применять их на практике.

Задача 1. Пусть $\{\xi_n \mid n \in \mathbb{N}\}$ — последовательность независимых и одинаково распределённых случайных величин, про которые известно, что $a = E \xi_1 < \infty$ и $0 < \sigma^2 = D \xi_1 < +\infty$. Введём обозначение $S_n = \xi_1 + \dots + \xi_n$. Найдите предел по распределению

$$\sqrt{n} \left(\frac{n}{S_n} - \frac{1}{a} \right)$$

Решение. Преобразуем это выражение следующим образом:

$$\sqrt{n} \left(\frac{n}{S_n} - \frac{1}{a} \right) = -\frac{\sigma}{a} \cdot \frac{n}{S_n} \cdot \sqrt{n} \left(\frac{S_n}{n} - a \right) \frac{1}{\sigma}$$

Осталось заметить следующее:

$$\text{По теореме о наследовании сходимости и уЗБЧ } \frac{S_n}{n} \xrightarrow{\text{п.н.}} a \implies \frac{n}{S_n} \xrightarrow{d} \frac{1}{a}$$

$$\text{По ЦПТ } \sqrt{n} \left(\frac{S_n}{n} - a \right) \frac{1}{\sigma} \xrightarrow{d} \mathcal{N}(0, 1)$$

Следовательно, по лемме Slutsky

$$\sqrt{n} \left(\frac{n}{S_n} - \frac{1}{a} \right) \xrightarrow{d} -\frac{\sigma}{a^2} \mathcal{N}(0, 1) = \mathcal{N}\left(0, \frac{\sigma^2}{a^4}\right). \quad \square$$

Теперь обобщим закон больших чисел:

Теорема 51 (усиленный ЗБЧ для векторов). Пусть $\{\xi_n \mid n \in \mathbb{N}\}$ — последовательность независимых и одинаково распределённых случайных m -мерных векторов. Далее, положим $a = E \xi_1 \in \mathbb{R}^m$. Тогда

$$\frac{\xi_1 + \dots + \xi_n}{n} \xrightarrow{\text{п.н.}} a.$$

Доказательство. Автоматически следует из равносильности покомпонентной и общей сходимости и усиленного ЗБЧ для случайных величин. \square

1.15.3 Многомерное нормальное распределение

Перед тем, как приступать к статистике, нужно ввести понятие многомерного нормального распределения.

Определение 71. Пусть есть случайный вектор $\xi \in \mathbb{R}^m$. Будем говорить, что он подчиняется *многомерному нормальному распределению*, если его характеристическая функция имеет вид

$$\varphi_\xi(\mathbf{t}) = \mathbb{E}[e^{i\langle \mathbf{t}, \xi \rangle}] = \exp \left\{ i \langle \mathbf{a}, \mathbf{t} \rangle - \frac{1}{2} \langle \Sigma \mathbf{t}, \mathbf{t} \rangle \right\},$$

где $\mathbf{a} \in \mathbb{R}^m$ — некоторый фиксированный вектор, а Σ — некоторая симметрическая и неотрицательно определённая матрица $m \times m$. В таком случае пишут, что $\xi \sim \mathcal{N}(\mathbf{a}, \Sigma)$.

Есть одна весьма полезная теорема, которая даёт три эквивалентных определения.

Теорема 52 (о трёх эквивалентных определениях). Пусть $\xi = (\xi_1, \dots, \xi_n)$ — случайный вектор. Тогда следующие утверждения равносильны:

1. Случайный вектор ξ имеет нормальное многомерное распределение.
2. $\xi = \Sigma \eta + \mathbf{a}$ почти наверное, где Σ — матрица размера $n \times n$, $\mathbf{a} \in \mathbb{R}^n$, а $\eta = (\eta_1, \dots, \eta_n)$, где η_i независимы в совокупности и имеют нормальное стандартное распределение.
3. Для любого вектора $\mathbf{x} \in \mathbb{R}^n$ случайная величина $\langle \mathbf{x}, \xi \rangle$ имеет нормальное распределение.

Доказательство. Докажем, что из (2) следует (3). Рассмотрим некоторый вектор $\mathbf{x} \in \mathbb{R}^n$ и случайную величину $\langle \mathbf{x}, \xi \rangle$:

$$\begin{aligned} \langle \mathbf{x}, \xi \rangle &= \langle \mathbf{x}, \Sigma \eta + \mathbf{a} \rangle = \langle \mathbf{x}, \mathbf{a} \rangle + \langle \mathbf{x}, \Sigma \eta \rangle \\ &= \langle \mathbf{x}, \mathbf{a} \rangle + \sum_{k=1}^n (\Sigma^\top \mathbf{x})_k \eta_k \end{aligned}$$

Несложно понять, что это сумма нормальных случайных величин, что есть нормальная случайная величина.

Теперь покажем, что из (3) следует (1). Пусть $\langle \mathbf{x}, \xi \rangle \sim \mathcal{N}(a_{\mathbf{x}}, \sigma_{\mathbf{x}}^2)$. Тогда

$$\begin{aligned} a_{\mathbf{x}} &= \mathbb{E}[\langle \mathbf{x}, \xi \rangle] = \sum_{k=1}^n \mathbb{E}[x_k \xi_k] = \sum_{k=1}^n x_k \mathbb{E}[\xi_k] = \langle \mathbf{x}, \mathbb{E}[\xi] \rangle \\ \sigma_{\mathbf{x}}^2 &= D[\langle \mathbf{x}, \xi \rangle] = D[\mathbf{x}^\top \xi] = \mathbf{x}^\top D[\xi] \mathbf{x} = \langle D[\xi] \mathbf{x}, \mathbf{x} \rangle \end{aligned}$$

Осталось посчитать характеристическую функцию ξ :

$$\begin{aligned} \varphi_\xi(\mathbf{x}) &= \mathbb{E}[e^{i\langle \mathbf{x}, \xi \rangle}] = \varphi_{\langle \mathbf{x}, \xi \rangle}(1) = \exp \left\{ i a_{\mathbf{x}} - \frac{1}{2} \sigma_{\mathbf{x}}^2 \right\} \\ &= \exp \left\{ i \langle \mathbb{E}[\xi], \mathbf{x} \rangle - \frac{1}{2} \langle D[\xi] \mathbf{x}, \mathbf{x} \rangle \right\}. \end{aligned}$$

А это и означает, что ξ имеет многомерное нормальное распределение. □

1.16 Лекция от 10.02.2017

1.16.1 Вычисление условного математического ожидания

Ранее мы ввели условное математическое ожидание. Теперь хотелось бы научиться считать его. Для начала покажем, как это делается в дискретном случае.

Теорема 53. Пусть η имеет дискретное распределение и $X = (x_1, \dots, x_n, \dots)$ — множество значений. Тогда для любой случайной величины ξ выполнено, что

$$E[\xi | \eta] = \sum_{k=1}^{\infty} \frac{E[\xi \mathbf{I}\{\eta = x_k\}]}{P(\eta = x_k)} \mathbf{I}\{\eta = x_k\}$$

Доказательство. Достаточно проверить измеримость и интегральное свойство. Как известно, $\mathbf{I}\{\eta = x\}$ является борелевской функцией от η для любого x . Тогда функция в правой части является борелевской функцией от η .

Теперь проверим интегральное свойство. Для начала возьмём $B = \{x_n\}$, где x_n — это одно из значений η , для которого $P(\eta = x_n) > 0$.

$$\begin{aligned} E[E[\xi | \eta] \mathbf{I}\{\eta = x_n\}] &= E\left[\frac{E[\xi \mathbf{I}\{\eta = x_n\}]}{P(\eta = x_n)} \mathbf{I}\{\eta = x_n\}\right] \\ &= \frac{E[\xi \mathbf{I}\{\eta = x_n\}]}{P(\eta = x_n)} E[\mathbf{I}\{\eta = x_n\}] = E[\xi \mathbf{I}\{\eta = x_n\}] \end{aligned}$$

Теперь обобщим это на произвольное B . Пусть в B есть элементы y_1, \dots, y_n, \dots , которые являются значениям η . Тогда по линейности математического ожидания получаем, что

$$\begin{aligned} E[\xi \mathbf{I}\{\eta \in B\}] &= E\left[\xi \left(\sum_{i=1}^{\infty} \mathbf{I}\{\eta = y_i\}\right)\right] = \sum_{i=1}^{\infty} E[\xi \mathbf{I}\{\eta = y_i\}] \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^n E[\xi \mathbf{I}\{\eta = y_i\}] \end{aligned}$$

Заметим, что $|\xi \mathbf{I}\{\eta = y_i\}| \leq |\xi|$ и $E[\xi \mathbf{I}\{\eta = y_i\}] = E[E[\xi | \eta] \mathbf{I}\{\eta = y_i\}]$. Тогда по теореме Лебега о мажорируемой сходимости

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n E[\xi \mathbf{I}\{\eta = y_i\}] = E\left[E[\xi | \eta] \left(\sum_{i=1}^{\infty} \mathbf{I}\{\eta = y_i\}\right)\right] = E[E[\xi | \eta] \mathbf{I}\{\eta \in B\}]. \quad \square$$

1.16.2 Свойства условного математического ожидания

У условного математического ожидания немало полезных свойств.

Свойство 1. Если $\xi = f(\eta)$, где f — некоторая борелевская функция, то $E[\xi | \eta] = \xi$.

Доказательство. Несложно проверить, что для ξ выполняются свойства измеримости и интегральное свойство. \square

Свойство 2 (Линейность). Для любых случайных величин (или векторов) ξ, η и δ выполнено, что

$$\forall a, b \in \mathbb{R} \quad E[a\xi + b\delta | \eta] = a E[\xi | \eta] + b E[\delta | \eta].$$

Доказательство. Для начала заметим, что и $E[\xi | \eta]$, и $E[\delta | \eta]$ являются борелевскими функциями от η . Тогда $a E[\xi | \eta] + b E[\delta | \eta]$ тоже есть борелевская функция от η .

Осталось проверить интегральное свойство. По линейности матожидания

$$E[(a\xi + b\delta)\mathbf{1}_B(\eta)] = a E[\xi\mathbf{1}_B(\eta)] + b E[\delta\mathbf{1}_B(\eta)].$$

Согласно интегральному свойству:

$$a E[\xi\mathbf{1}_B(\eta)] + b E[\delta\mathbf{1}_B(\eta)] = a E[E[\xi | \eta]\mathbf{1}_B(\eta)] + b E[E[\delta | \eta]\mathbf{1}_B(\eta)].$$

Свернём сумму назад:

$$a E[E[\xi | \eta]\mathbf{1}_B(\eta)] + b E[E[\delta | \eta]\mathbf{1}_B(\eta)] = E[(a E[\xi | \eta] + b E[\delta | \eta])\mathbf{1}_B(\eta)].$$

Тем самым получаем желаемое. \square

Свойство 3 (Формула полной вероятности). «Математическое ожидание убирает условие»: $E[E[\xi | \eta]] = E[\xi]$.

Доказательство. Пусть $\eta \in \mathbb{R}^n$. Воспользуемся интегральным свойством, положив $B = \mathbb{R}^n$:

$$E[\xi] = E[\xi\mathbf{1}_B(\eta)] = E[E[\xi | \eta]\mathbf{1}_B(\eta)] = E[E[\xi | \eta]]. \quad \square$$

Примечание. Название оно получило из-за того, что является обобщением формулы полной вероятности. Покажем это. Пусть $\{B_n | n \in \mathbb{N}\}$ — попарно не пересекающиеся события. Построим η следующим образом: она будет обладать дискретным распределением и $P(\eta = n) = P(B_n)$. Тогда, взяв $\xi = \mathbf{1}_B(x)$, получим, что

$$P(B) = E[E[\mathbf{1}_B | \eta]] = E[P(B | \eta)].$$

А это, по сути, и есть формула полной вероятности.

Свойство 4. Если ξ независимо с η , то $E[\xi | \eta] = E[\xi]$.

Доказательство. Заметим, что $E[\xi]$ — константа. Тогда можно считать, что $E[\xi]$ есть борелевская функция от η . Проверим интегральное свойство:

$$E[\xi\mathbf{1}_B(\eta)] = |\text{по независимости}| = E[\xi] E[\mathbf{1}_B(\eta)] = E[E[\xi] \mathbf{1}_B(\eta)]. \quad \square$$

Свойство 5 (Сохранение отношения порядка). Если $\xi \leq \delta$ почти наверное, то $E[\xi | \eta] \leq E[\delta | \eta]$ почти наверное.

Доказательство. Рассмотрим случайную величину $\zeta = E[\xi | \eta] - E[\delta | \eta]$. Несложно понять, что она является борелевской функцией от η . Теперь рассмотрим $E[\chi\mathbf{1}_{(0,+\infty)}(\chi)]$. Заметим, что $E[\chi\mathbf{1}_{(0,+\infty)}(\chi)] \geq 0$. Теперь скажем, что $\mathbf{1}_{(0,+\infty)}(\chi) = \mathbf{1}_B(\eta)$ для какого-то борелевского множества B .²⁰ Тогда

$$E[\chi\mathbf{1}_{(0,+\infty)}(\chi)] = E[(E[\xi | \eta] - E[\delta | \eta])\mathbf{1}_B(\eta)].$$

Пользуясь линейностью матожидания и интегральным свойством, получаем, что

$$E[\chi\mathbf{1}_{(0,+\infty)}(\chi)] = E[\xi\mathbf{1}_B(\eta)] - E[\delta\mathbf{1}_B(\eta)] \leq 0, \text{ так как } \xi \leq \delta.$$

Тогда получаем, что $\chi\mathbf{1}_{(0,+\infty)}(\chi) = 0$ почти наверное и $\xi \leq \delta$. \square

²⁰На самом деле $B = \chi^{-1}(0, +\infty)$.

Свойство 6 (Телескопическое). Пусть ξ , η и δ — случайные векторы. Тогда

1. $E[E[\xi | \eta] | (\eta, \delta)] = E[\xi | \eta]$,
2. $E[E[\xi | (\eta, \delta)] | \eta] = E[\xi | \eta]$.

Доказательство. Первый пункт почти очевиден. Заметим, что $E[\xi | \eta]$ — борелевская функция от η . Но тогда она является борелевской функцией от (η, δ) . Следовательно, по свойству 1 получаем желаемое.

Теперь докажем второе утверждение. Заметим, что $E[\xi | \eta]$ является борелевской функцией от η . Теперь проверим интегральное свойство. Пусть $\delta \in \mathbb{R}^m$. Тогда по интегральному свойству:

$$\begin{aligned} E[E[\xi | (\eta, \delta)] \mathbf{1}_B(\eta)] &= E[E[\xi | (\eta, \delta)] \mathbf{1}_{B \times \mathbb{R}^m}(\eta, \delta)] = E[E[\xi | (\eta, \delta)] \mathbf{1}_{B \times \mathbb{R}^m}(\eta, \delta)] \\ &= E[\xi \mathbf{1}_{B \times \mathbb{R}^m}(\eta, \delta)] = E[\xi \mathbf{1}_B(\eta)] = E[E[\xi | \eta] \mathbf{1}_B(\eta)]. \end{aligned} \quad \square$$

Свойство 7 (Предельный переход). Пусть $\{\xi_n | n \in \mathbb{N}\}$, ξ , η , δ — случайные величины.

1. Если $0 \leq \xi_n \uparrow \xi$ почти наверное, то $E[\xi_n | \eta] \uparrow E[\xi | \eta]$ почти наверное.
2. Если $\xi_n \xrightarrow{n.n.} \xi$ и $|\xi_n| \leq |\delta|$ для всех натуральных n , то $E[\xi_n | \eta] \xrightarrow{n.n.} E[\xi | \eta]$.

Доказательство. Так как ξ_n монотонно возрастают, то и $E[\xi_n | \eta]$ монотонно возрастают (по свойству 5). Тогда существует (не обязательно конечный) предел $h = \lim_{n \rightarrow \infty} E[\xi_n | \eta]$. Так как для любого натурального n $E[\xi_n | \eta]$ — это борелевская функция от η , то и h является борелевской функцией от η .

Теперь возьмём произвольное борелевское множество B и воспользуемся интегральным свойством:

$$E[\xi_n \mathbf{1}_B(\eta)] = E[E[\xi_n | \eta] \mathbf{1}_B(\eta)].$$

Теперь заметим, что $\xi_n \mathbf{1}_B(\eta) \uparrow \xi \mathbf{1}_B(\eta)$ и $E[\xi_n | \eta] \mathbf{1}_B(\eta) \uparrow h(\eta) \mathbf{1}_B(\eta)$. Тогда по теореме о монотонной сходимости

$$E[\xi \mathbf{1}_B(\eta)] = E[h(\eta) \mathbf{1}_B(\eta)].$$

А это и означает, что $E[\xi_n | \eta] \uparrow E[\xi | \eta]$.

Второй же пункт доказывается по аналогии с доказательством теоремы Лебега для обычного математического ожидания. \square

Свойство 8. Если δ — это борелевская функция от η , то $E[\xi \delta | \eta] = \delta E[\xi | \eta]$.

Доказательство. Пусть $\delta = \mathbf{1}_A(\eta)$, где A — некоторое борелевское множество. Заметим, что $\delta E[\xi | \eta]$ — это борелевская функция от η . Проверим интегральное свойство:

$$\begin{aligned} E[\delta \xi \mathbf{1}_B(\eta)] &= E[\xi \mathbf{1}_A(\eta) \mathbf{1}_B(\eta)] = E[\xi \mathbf{1}_{A \cap B}(\eta)] = E[E[\xi | \eta] \mathbf{1}_{A \cap B}(\eta)] \\ &= E[E[\xi | \eta] \mathbf{1}_A(\eta) \mathbf{1}_B(\eta)] = E[\delta E[\xi | \eta] \mathbf{1}_B(\eta)]. \end{aligned}$$

Пользуясь линейностью математического ожидания, интегральное свойство будет выполнено и для простых функций вида $\delta = \sum c_k \mathbf{1}_{A_k}(\eta)$.

В общем случае для любой $\delta = f(\eta)$ построим последовательность простых функций $\delta_n = f_n(\eta)$ такую, что $\delta_n \xrightarrow{n.n.} \delta$ и $|\delta_n| \leq |\delta|$. Тогда по свойству 7 получаем, что $E[\delta_n \xi | \eta] \xrightarrow{n.n.} E[\delta \xi | \eta]$. Однако

$$E[\delta_n \xi | \eta] = \delta_n E[\xi | \eta] \xrightarrow{n.n.} \delta E[\xi | \eta].$$

Следовательно, $E[\xi \delta | \eta] = \delta E[\xi | \eta]$ (почти наверное, конечно). \square

1.16.3 Условные распределения и плотности

Определение 72. Условным математическим ожиданием при условии $\eta = y$, $\eta, y \in \mathbb{R}^n$ называется $E[\xi \mid \eta = y]$.

Вообще говоря, $E[\xi \mid \eta = y]$ — это борелевская функция $\varphi(y)$ такая, что для любого $B \in \mathcal{B}(\mathbb{R}^n)$

$$E[\xi \mathbf{1}_B(\eta)] = E[\varphi(\eta) \mathbf{1}_B(\eta)].$$

Отсюда можно сделать простой вывод: $E[\xi \mid \eta = y] = \varphi(y) \iff E[\xi \mid \eta] = \varphi(\eta)$.

Свойства у такого условного математического ожидания похожи на свойства для обычного условного: например, выполнены линейность, сохранение относительного порядка и теорема Лебега.

Впрочем, мы не имеем представления о том, как его вычислять. Для этого введём условные распределение и плотность.

Определение 73. Условным распределением случайной величины ξ при условии, что $\eta = y$ назовём функцию $P(\xi \in B \mid \eta = y) \equiv E[\mathbf{1}_B(\xi) \mid \eta = y]$, рассматриваемую, как функцию от $B \in \mathcal{B}(\mathbb{R}^n)$ при фиксированном $y \in \mathbb{R}^k$.

Утверждение. При фиксированном $y \in \mathbb{R}^k$ с вероятностью 1 (по распределению η) условное распределение есть распределение вероятностей на $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Определение 74. Если условное распределение имеет плотность $p_{\xi|\eta}(x \mid y)$, то назовём его *условной плотностью ξ относительно η* . То есть для любого $B \in \mathcal{B}(\mathbb{R}^n)$

$$P(\xi \in B \mid \eta = y) = \int_B p_{\xi|\eta}(x \mid y) dx.$$

Теперь сформулируем две теоремы (пока что без доказательства), которые и позволяют нам считать условное математическое ожидание.

Теорема 54 (о вычислении условного математического ожидания). Пусть ξ — случайная величина, а $f(x)$ — некоторая борелевская функция. Если $E[|f(\xi)|] < +\infty$ и существует плотность $p_{\xi|\eta}(x \mid y)$, то

$$E[f(\xi) \mid \eta = y] = \int_{\mathbb{R}} f(x) p_{\xi|\eta}(x \mid y) dx.$$

Теорема 55. Пусть ξ и η таковы, что есть совместная плотность $p_{\xi,\eta}(x, y)$. Тогда существует условная плотность $p_{\xi|\eta}(x \mid y)$ и она равна

$$p_{\xi|\eta}(x \mid y) = \begin{cases} \frac{p_{\xi,\eta}(x, y)}{p_{\eta}(y)}, & p_{\eta}(y) > 0, \\ 0, & p_{\eta}(y) = 0. \end{cases}$$

Рассмотрим на примере, как их использовать.

Задача 1. Пусть ξ и η — независимые случайные величины, причём $\xi, \eta \sim \text{Exp}(1)$. Найдите $E[\xi^2 \mid \xi + \eta]$.

Решение. Для начала найдём плотность $\xi + \eta$. По формуле свёртки:

$$\begin{aligned} p_{\xi+\eta}(x) &= \int_{-\infty}^{+\infty} e^{-(x-y)} e^{-y} \mathbf{1}_{(0,+\infty)}(x-y) \mathbf{1}_{(0,+\infty)}(y) dy \\ &= \mathbf{1}_{(0,+\infty)}(x) \int_0^x e^{-x} dy = x e^{-x} \mathbf{1}_{(0,+\infty)}(x). \end{aligned}$$

Теперь найдём совместную плотность распределения ξ и $\xi + \eta$. Для этого распишем вероятностную меру:

$$\mathbf{P}((\xi, \xi + \eta) \in B) = \iint_{(x, x+y) \in B} e^{-x} e^{-y} \mathbf{1}_{(0,+\infty)}(x) \mathbf{1}_{(0,+\infty)}(y) dx dy$$

Сделаем замену переменных:

$$\begin{cases} u = x \\ v = x + y \end{cases} \implies \begin{cases} x = u \\ y = u - v \end{cases} \implies \mathbf{J}(u, v) = \begin{vmatrix} 1 & 0 \\ 1 & -1 \end{vmatrix}$$

Тогда $|\det \mathbf{J}(u, v)| = 1$ и

$$\mathbf{P}((\xi, \xi + \eta) \in B) = \iint_{(u, v) \in B} e^{-v} \mathbf{1}_{\{0 < u < v\}} du dv.$$

Следовательно, $p_{\xi, \xi+\eta}(x, y) = e^{-y} \mathbf{1}_{\{0 < x < y\}}$. Теперь посчитаем $p_{\xi|\eta}(x | y)$. Так как $\xi + \eta > 0$ с вероятностью 1, то:

$$p_{\xi|\eta}(x | y) = \frac{e^{-y} \mathbf{1}_{\{0 < x < y\}}}{y e^{-y} \mathbf{1}_{(0,+\infty)}(y)} = \frac{1}{y} \mathbf{1}_{\{0 < x < y\}}.$$

Недолго думая, считаем условное матожидание по формуле:

$$\mathbf{E}[\xi^2 | \xi + \eta = y] = \int_0^y \frac{x^2 dx}{y} = \frac{y^2}{3} \implies \mathbf{E}[\xi^2 | \xi + \eta] = \frac{(\xi + \eta)^2}{3}.$$

□

Глава 2

Семинары

2.1 Семинар от 09.09.2016

Перед тем, как начать решать задачи, кратко опишем вероятностное пространство для броска n -гранного кубика: $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$, $\omega_i = \{\text{выпало число } i\}$, $P(\omega_i) = \frac{1}{n}$ для всех i .

Задача 1. Пусть бросаются n -гранный и m -гранный кубики. Какова вероятность P того, что выпадет одно чётное и одно нечётное число?

Решение. В данной задаче есть два случая:

1. На первом выпало чётное число очков, а на втором — нечётное. Количество чётных чисел от 1 до n равно $\lfloor \frac{n}{2} \rfloor$, а нечётных чисел от 1 до m — $\lfloor \frac{m}{2} \rfloor$. Тогда есть $\lfloor \frac{m}{2} \rfloor \lfloor \frac{n}{2} \rfloor$ успешных исходов.
2. На первом выпало нечётное число очков, на втором — чётное. Аналогичными рассуждениями получаем $\lfloor \frac{m}{2} \rfloor \lfloor \frac{n}{2} \rfloor$ успешных исходов.

Всего же исходов mn . Следовательно,

$$P = \frac{\lfloor \frac{m}{2} \rfloor \lfloor \frac{n}{2} \rfloor + \lfloor \frac{m}{2} \rfloor \lfloor \frac{n}{2} \rfloor}{mn}. \quad \square$$

Задача 2. Пусть бросаются два n -гранных кубика. Какова вероятность $P(i)$ того, что суммарно выпадет $2 \leq i \leq 2n$ очков?

Решение. В данной задаче есть два случая:

1. $i \leq n + 1$. Представим i в следующем виде: $i = k + (i - k)$, где $1 \leq k \leq i - 1$. Такое ограничение сверху на k объясняется тем, что иначе $i - k$ будет меньше 1, а при броске кубика не может выпасть меньше 1 очка. Ограничение снизу объясняется аналогично. Тогда есть $i - 1$ подходящий случай.
2. $n + 2 \leq i \leq 2n$. Опять же, представим i в виде $i = k + (i - k)$. Теперь определим границы для k . Очевидно, что $k \leq n$. Так как $i - k \leq n$, то $k \geq i - n$. Тогда получаем $i - n \leq k \leq n$. Тогда есть $n - (i - n) + 1 = 2n - i + 1$ подходящий случай.

Так как всего есть n^2 разных вариантов того, сколько очков выпадет на кубиках, то получаем, что

$$P(i) = \begin{cases} \frac{i-1}{n^2} & i \leq n + 1, \\ \frac{2n-i+1}{n^2} & n + 2 \leq i \leq 2n. \end{cases} \quad \square$$

Примечание. Если нарисовать график функции $P(i)$, то он будет выглядеть, как треугольник с вершиной в точке $(n+1, \frac{1}{n})$. Такой график называют *треугольным распределением*.

Перейдём от кубиков к монеткам.

Задача 3. Пусть последовательно бросают n монет (полагается, что $\Omega = \{O, P\}$). Какова вероятность P того, что не выпадет последовательно

1. орёл и решка?
2. два орла?

Решение. Как рассказывалось ранее, в такой модели есть 2^n элементарных исходов. Посчитаем количество успешных исходов в каждом случае:

1. В таком случае легко понять, что будут допустимы только последовательности вида $\underbrace{PP \dots P}_{k \text{ раз}} \underbrace{OO \dots O}_{n-k \text{ раз}}$, где $0 \leq k \leq n$. Тогда есть $n+1$ подходящий исход.
2. Пусть f_n — количество последовательностей длины n , в которых нет двух орлов подряд. Как посчитать f_n ? Попробуем выразить рекурсивно. Если при последнем броске выпал орёл, то при предпоследнем обязательно выпала решка. То, что идёт до решки, явно угадать невозможно. Но нам известно, что это последовательность размера $n-2$ и в ней нет двух орлов подряд. Тогда их f_{n-2} вариантов. Если же выпала решка, то есть f_{n-1} вариант. Отсюда получаем, что $f_n = f_{n-1} + f_{n-2}$. Так как $f_1 = 2$, а $f_2 = 3$ (допускаются ОР, РО, РР), то $f_n = F_{n+2}$, где F_n — n -е число Фибоначчи.

Ответ: 1) $P = \frac{n+1}{2^n}$, 2) $P = \frac{F_{n+2}}{2^n}$. □

Перед тем, как идти дальше, сделаем небольшое отступление. Во втором пункте последней задачи нам повезло, что последовательность совпала с последовательностью чисел Фибоначчи. А что делать, если не удаётся угадать последовательность? В таком случае можно воспользоваться общим методом решения. Рассмотрим его на примере из последней задачи:

$$\begin{aligned} f_n &= f_{n-1} + f_{n-2} \\ f_1 &= 2 \\ f_2 &= 3 \end{aligned}$$

Выпишем *характеристическое уравнение* для данного рекуррентного соотношения: $\lambda^2 - \lambda - 1 = 0$. Находим его корни — в данном случае они равны $\frac{1 \pm \sqrt{5}}{2}$. Тогда для каких-то констант a_1 и a_2 верно, что

$$f_n = a_1 \left(\frac{1 + \sqrt{5}}{2} \right)^n + a_2 \left(\frac{1 - \sqrt{5}}{2} \right)^n.$$

Данные константы можно определить по начальным условиям.

Задача 4 (Парадокс дней рождения). В группе 27 студентов. Считаем их дни рождения случайными и равновероятными. Какова вероятность P того, что хотя бы у двух студентов совпадают дни рождения?

Решение. В данной задаче гораздо проще посчитать вероятность дополнения, то есть вероятность того, что у всех 27 студентов будут разные дни рождения. Так как порядок дней рождения важен, то эта вероятность равна $\frac{A_{365}^{27}}{365^{27}}$. В итоге получаем, что

$$P = 1 - \frac{A_{365}^{27}}{365^{27}}.$$

Хорошо, ответ получен. Но по нему сложно сказать, много ли это или мало. Попробуем посчитать его приближенно: $P = 1 - \left(1 - \frac{1}{365}\right) \left(1 - \frac{2}{365}\right) \dots \left(1 - \frac{27}{365}\right)$. Так как $1 + x \approx e^x$, то $P \approx 1 - e^{-\frac{1+2+\dots+27}{365}} \approx 1 - e^{-1.04} \approx 0,66$. Как видно, вероятность достаточно велика. \square

Сделаем небольшое теоретическое отступление. Вспомним формулу включений-исключений: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Попробуем придумать аналогичную формулу для трёх событий:

$$\begin{aligned} P(A \cup B \cup C) &= P(A \cup B) + P(C) - P((A \cup B) \cap C) \\ &= P(A \cup B) + P(C) - P((A \cap C) \cup (B \cap C)) \\ &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \end{aligned}$$

Уже видна некоторая закономерность. Сформулируем обобщение.

Теорема 1 (Общая формула включений-исключений). Пусть A_1, A_2, \dots, A_n — некоторые события на Ω , а $S_k = \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k})$. Тогда

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{k=1}^n (-1)^{k-1} S_k.$$

Доказательство. По индукции. База ($n = 2$) была доказана ранее. Теперь предположим, что утверждение верно для какого-то n . Докажем, что из этого следует, что утверждение верно и для $n + 1$:

$$\begin{aligned} P(A_1 \cup A_2 \cup \dots \cup A_n \cup A_{n+1}) &= P(A_1 \cup A_2 \cup \dots \cup A_n) + \\ &\quad + P(A_{n+1}) - P((A_1 \cup A_2 \cup \dots \cup A_n) \cap A_{n+1}) \end{aligned}$$

Так как $(A_1 \cup A_2 \cup \dots \cup A_n) \cap A_{n+1} = \bigcup_{i=1}^n (A_i \cap A_{n+1})$, то, пользуясь предположением индукции и рассуждениями, аналогичными доказательству для трёх множеств, получаем желаемое. \square

Примечание. Важное следствие из этой формулы: если A_1, A_2, \dots, A_n — некоторые события на Ω , то по закону де Моргана получаем, что $P(\bigcap_{i=1}^n \overline{A_i}) = P(\overline{\bigcup_{i=1}^n A_i}) = 1 - P(\bigcup_{i=1}^n A_i) = 1 - \sum_{k=1}^n (-1)^{k-1} S_k$. Если положить $S_0 = 1$, то эту формулу можно записать в виде

$$P\left(\bigcap_{i=1}^n \overline{A_i}\right) = \sum_{k=0}^n (-1)^k S_k.$$

Задача 5. Пусть мы раскидали n шаров по m ящикам. Какова вероятность P того, что ни один ящик не пуст? Рассмотрите случаи, когда шары различимы и неразличимы.

Решение. Начнём со случая различных шаров. В данном случае элементарным исходом будет $\omega = (\omega_1, \omega_2, \dots, \omega_m)$, где ω_i — количество шаров в i -м ящике. В таком случае $|\Omega| = \binom{n+m-1}{m-1}$ (схема выбора неупорядоченных наборов с возвратом). Теперь посчитаем количество подходящих исходов. Так как ни один ящик не пуст, то в каждом из них есть хотя бы по одному шару. Тогда нужно посчитать количество способов раскидать $n-m$ шаров по m ящикам. Это можно сделать $\binom{n-1}{m-1}$ способом. Отсюда получаем, что вероятность равна

$$P = \frac{\binom{n-1}{m-1}}{\binom{n+m-1}{m-1}}.$$

Теперь предположим, что шары неразличимы. Рассмотрим событие $A_i = \{i\text{-й ящик пуст}\}$. Чему равна вероятность такого события? Для каждого из n шаров есть $m-1$ подходящий ящик. Тогда $P(A_i) = \frac{(m-1)^n}{m^n} = \left(1 - \frac{1}{m}\right)^n$. Пересечение $A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}$ означает, что k ящиков с номерами i_1, i_2, \dots, i_k пусты. Тогда $P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \left(1 - \frac{k}{m}\right)^n$. Заметим, что событие “ни один ящик не пуст” равно $\overline{A_1 \cap A_{i_2} \cap \dots \cap A_m}$. Пользуясь формулой включений-исключений, получаем, что вероятность равна

$$P = \sum_{k=0}^m (-1)^k \binom{m}{k} \left(1 - \frac{k}{m}\right)^n.$$

□

Задача 6. Алиса и Боб случайно подбрасывают n монет. Какова вероятность P того, что число орлов у Алисы будет строго больше, чем у Боба? Каков будет ответ на этот вопрос, если Алиса подбросила $n+1$ монету?

Решение. Для начала посмотрим, чему равна вероятность того, что число орлов у Алисы равно числу орлов у Боба. Если у Алисы выпало k орлов, что достигается в $\binom{n}{k}$ случаев, то у Боба тоже должно выпасть k орлов. Тогда достаточно логично, что число успешных исходов равно $\sum_{k=0}^n \binom{n}{k}^2$. Как это упростить? Воспользуемся тем, что $\binom{n}{k} = \binom{n}{n-k}$. Теперь представим себе следующую ситуацию: пусть есть строка, содержащая $2n$ символов. Из первых n нужно выбрать k символов, из вторых n нужно выбрать $(n-k)$. Это можно сделать $\binom{n}{k} \binom{n}{n-k}$ способами. Если просуммировать эти числа по k от 0 до n , то легко заметить, что это то же самое, что и посчитать количество способов выбрать n символов из $2n$. Тогда

$$\sum_{k=0}^n \binom{n}{k}^2 = \binom{2n}{n}.$$

Всего исходов 4^n (по 2^n на Алису и на Боба). Тогда вероятность равна $\frac{\binom{2n}{n}}{4^n}$.

Теперь рассмотрим вероятность из условия. Из-за симметричности она равна вероятности того, что у Алисы будет строго меньше орлов, чем у Боба. Тогда получаем, что $2P + \frac{\binom{2n}{n}}{4^n} = 1$. Отсюда $P = \frac{1}{2} - \frac{\binom{2n}{n}}{2^{2n+1}}$.

Теперь перейдём ко второму пункту. Его мы решим двумя способами — стандартным и “олимпиадным”. Начнём со стандартного. Если у Алисы уже было больше орлов, чем у Боба, то что бы у неё не выпало, то ситуация не изменится. Если же было так, что у неё столько же орлов, сколько у Боба, то ей необходимо, чтобы выпал орёл. Тогда искомая вероятность равна

$$P' = P + \frac{1}{2} \frac{\binom{2n}{n}}{4^n} = \frac{1}{2}.$$

Теперь рассмотрим “олимпиадный” способ решения. Заметим, что вероятность того, что у Алисы будет больше орлов, чем у Боба, равна вероятности того, что у неё будет больше решек (из-за симметрии). При этом вероятность того, что у неё будет больше орлов, равна вероятности того, что у неё будет не больше решек, чем у Боба (пусть у неё на одного орла больше, тогда число решек у них совпадает). Отсюда сразу получаем, что $P' = \frac{1}{2}$. \square

А сейчас мы посмотрим, почему стоит быть осторожным с азартными играми.

Задача 7. Пусть есть 52 карты, и игроку выдают 5 случайных карт. Найдите вероятности получения различных наборов из покера.

Решение. Начнём с того, что заметим, что выбрать 5 карт из 52 мы можем $\binom{52}{5}$ способами. Теперь достаточно найти количество подходящих исходов.

1. Royal Flush — туз, король, дама, валет и десятка одной масти. Есть лишь 4 подходящих исхода.
2. Straight Flush — пять последовательных по достоинству карт одной масти (начиная не с туза). Так как первую карту можно выбрать 8 способами (от пятёрки до короля), то есть $9 \cdot 4 = 36$ успешных исходов.
3. Four Of A Kind — четыре карты одного достоинства. Выберем достоинство (это можно сделать 13) способами и последнюю карту (это можно сделать 48 способами). Тогда есть $13 \cdot 48$ подходящих комбинаций.
4. Full House — три карты одного достоинства и две карты другого достоинства. Выберем первое достоинство (13 вариантов) и выберем 3 карты из 4 подходящих ($\binom{4}{3}$ способов). Теперь выберем второе достоинство (12 вариантов) и 2 карты из 4 ($\binom{4}{2}$ вариантов). Тогда всего есть $13 \cdot \binom{4}{3} \cdot 12 \cdot \binom{4}{2}$ вариантов.
5. Flush — пять карт одной масти. Всего выбрать пять карт одной масти можно $\binom{13}{5}$ способами. Но в таком случае мы ещё учитываем Straight Flush и Royal Flush. Тогда есть $4(\binom{13}{5} - 10)$ вариантов.
6. Straight — пять последовательных карт (не одной масти). Всего пять последовательных карт можно выбрать $10 \cdot 4^5$ способами (сначала выбираем старшую карту, после чего масть для каждой). Но в таком случае учитывается Straight Flush и Royal Flush. Тогда есть $10(4^5 - 4)$ подходящих наборов.
7. Three Of A Kind — три карты одного достоинства. Сначала выберем достоинство (13 вариантов), после чего выберем 3 карты из 4 ($\binom{4}{3}$ вариантов). После чего выберем два разных достоинства (иначе будет Full House), что можно сделать $\binom{12}{2}$ способами, и масти для двух карт (4^2 способа). Итого — $13 \cdot \binom{4}{3} \cdot \binom{12}{2} \cdot 4^2$ варианта.
8. Two Pair — две пары карт одного достоинства. Выберем достоинства и масти для двух пар ($\binom{13}{2}$ варианта для достоинств, по $\binom{4}{2}$ для выбора 2-х карт каждого достоинства). Осталось выбрать последнюю карту — это можно сделать $52 - 8 = 44$ способами. Итого $\binom{13}{2} \binom{4}{2}^2 \cdot 44$ исхода.
9. One Pair — одна пара карт одного достоинства. Выберем достоинство и 2 карты из 4 ($\binom{13}{1} \cdot \binom{4}{2}$ вариантов). Теперь выберем три разных достоинства и масти для карт ($\binom{12}{3} \cdot 4^3$ варианта). Итого $\binom{13}{1} \cdot \binom{4}{2} \cdot \binom{12}{3} \cdot 4^3$ исходов.

10. High Card — ничего из вышеперечисленного. Выберем пять разных достоинств, не идущих подряд ($\binom{13}{5} - 10$ вариантов) и выберем масти для каждой карты так, чтобы они не совпадали ($4^5 - 4$ варианта). Итого $(\binom{13}{5} - 10)(4^5 - 4)$ вариантов.

Теперь приближенно посчитаем вероятность каждого из наборов:

Тип	Вероятность
Royal Flush	$4/2598960 \approx 0,00015\%$
Straight Flush	$36/2598960 \approx 0,0014\%$
Four Of A Kind	$624/2598960 \approx 0,024\%$
Full House	$3744/2598960 \approx 0,15\%$
Flush	$5108/2598960 \approx 0,2\%$
Straight	$10200/2598960 \approx 0,39\%$
Three Of A Kind	$54912/2598960 \approx 2,11\%$
Two Pair	$123552/2598960 \approx 4,75\%$
One Pair	$1098240/2598960 \approx 42,26\%$
High Card	$1302540/2598960 \approx 50,12\%$

Как видно из таблицы, получить что-то лучше, чем одну пару, уже не так просто. \square

2.2 Семинар от 16.09.2016

Начнём с разбора домашнего задания.

Задача 1. n шаров раскладывают по N ящикам. Найдите вероятность того, что для каждого $i = 1, 2, \dots, N$ в i -м ящике лежит n_i шаров, где $n_1 + n_2 + \dots + n_N = n$, если

1. шары различимы,
2. шары неразличимы.

Решение. Начнём со случая различимых шаров. В первый ящик необходимо выбрать n_1 шаров из n , что можно сделать $\binom{n}{n_1}$ способами. Для второго ящика надо выбрать n_2 шаров из $n - n_1$, что даёт $\binom{n-n_1}{n_2}$. Рассуждая аналогично, получаем, что всего есть $\binom{n}{n_1} \binom{n-n_1}{n_2} \dots \binom{n-n_1-\dots-n_{N-1}}{n_N} = \frac{n!}{n_1!(n-1)!} \frac{(n-n_1)!}{n_2!(n-n_1-n_2)!} \dots \frac{(n-n_1-\dots-n_{N-1})!}{n_N!(n-n_1-\dots-n_{N-1}-n_N)!} = \frac{n!}{n_1!n_2!\dots n_N!}$ успешных исходов. Всего же исходов N^n . Тогда искомая вероятность равна

$$P = \frac{n!}{N^n n_1! n_2! \dots n_N!}.$$

Теперь рассмотрим случай, когда шары неразличимы. Заметим, что тогда есть лишь один подходящий случай. Всего же случаев $\binom{n+N-1}{N-1}$. Тогда вероятность равна

$$P = \frac{1}{\binom{n+N-1}{N-1}}. \quad \square$$

Примечание. Заметим, что число успешных исходов в случае различимых шаров, равное $\frac{n!}{n_1!n_2!\dots n_N!}$, принято называть *мультиномиальным коэффициентом*. Его можно получить, рассматривая полином $(x_1 + x_2 + \dots + x_N)^n$.

Задача 2.

- 1) Случайно бросаются три N -гранных кубика, на гранях которых написаны числа от 1 до N . Найдите вероятность события $A_i = \{\text{сумма чисел, выпавших на кубиках, равна } i\}, i \leq N + 2$.
- 2)* Случайно бросаются три N -гранных кубика, на гранях которых написаны числа от 1 до N . Найдите вероятность события $A_i = \{\text{сумма чисел, выпавших на кубиках, равна } i\}, i = 2, \dots, 2N$.
- 3)** Случайно бросаются k различных N -гранных кубиков, на гранях которых написаны от 1 до N . Найдите вероятность события $A_i = \{\text{сумма чисел, выпавших на кубиках, равна } i\}, i = k, \dots, kN$.

Решение пункта (1). Пусть $i = k_1 + k_2 + k_3$, и $1 \leq k_1, k_2, k_3 \leq N$. Как посчитать число подходящих наборов? Воспользуемся методом точек и перегородок. Пусть есть i точек и нужно расставить 2 перегородки по $i - 1$ допустимой позиции. Тогда есть $\binom{i-1}{2}$ допустимых наборов. Отсюда получаем, что искомая вероятность равна

$$P[i] = \frac{(i-1)(i-2)}{2N^3}. \quad \square$$

Примечание. Пункты (2) и (3) на данный момент слишком сложны. Их адекватное решение будет рассказано ближе к концу курса.

Задача 3. Пусть выбирается произвольная перестановка из S_n . Какова вероятность того, что 1 и 2 будут лежать в одном цикле?

Решение. Воспользуемся тем фактом, что каждая перестановка однозначно представима в виде композиции циклов. Пусть цикл, содержащий 1 и 2, состоит из $k + 1$ элемента ($1 \leq k \leq n - 1$). Позицию для 2 можно выбрать k способами. Далее будем заполнять цикл. Есть $(n-2)(n-3) \dots (n-k)$ вариантов его заполнения. Остальное же мы можем заполнять, как хотим. Следовательно, итого есть $k(n-2)(n-3) \dots (n-k)(n-k-1)! = k(n-2)!$ допустимых перестановок с нужным циклом размера k . Тогда, суммируя по k от 1 до $n-1$, получаем $(n-2)!(1+2+\dots+(n-1)) = \frac{1}{2}n!$. Но всего перестановок $n!$. Тогда вероятность равна $1/2$. \square

Задача 4. Пусть в группе 25 студентов. Считаем, что дни рождения равновероятны и случайны. Найдите вероятность того, что найдётся ровно одна пара студентов такая, что

- дни рождения у них совпадают
- у всех других студентов дни рождения не совпадают с днём рождения данной пары студентов

Доказательство. Для начала посчитаем вероятность дополнения к событию. В данном случае дополнением является событие “нет такой пары студентов, что их дни рождения совпадают и у остальных они другие”.

Рассмотрим событие $A_i = \{\text{дни рождения в день } i \text{ совпадают только у двух человек}\}$. Его вероятность равна $P(A_i) = \frac{\binom{25}{2} \cdot 364^{23}}{365^{25}}$. Теперь посчитаем вероятность объединения k событий $A_{i_1}, A_{i_2}, \dots, A_{i_k}$ равна

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \frac{\binom{25}{2} \binom{23}{2} \dots \binom{25-2(k-1)}{2} (365-k)^{25-2k}}{365^{25}}.$$

Теперь посчитаем вероятность дополнения. Она равна $P\left(\bigcap_{i=1}^{365} \overline{A_i}\right)$. Её можно посчитать с помощью формулы включений-исключений. Теперь введём функцию $\alpha(x, y)$, где x — количество студентов, а y — количество дней. Данная функция равна вероятности того, что “среди x студентов нет такой пары студентов, что их дни рождения совпадают, и среди других студентов они другие и находятся среди y выбранных дней”. Она считается аналогично.

Теперь посчитаем вероятность из условия. Для этого выберем двух человек из 25, выберем им день рождения. После чего посчитаем вероятность дополнения для 23 студентов и 364 дней и умножим на 364^{23} . Тогда ответ равен

$$P = \frac{\binom{25}{2} \cdot 365 \cdot \alpha(23, 364) \cdot 364^{23}}{365^{25}}. \quad \square$$

Теперь рассмотрим несколько классических задач на условную вероятность.

Задача 5 (Парадокс Монти-Холла). *Вы участвуете в игре, в которой надо выбрать одну дверь из трёх. За одной из них автомобиль, а за другими — козы. Вы выбрали первую дверь. Ведущий открыл третью дверь, за которой стоит коза. Ведущий предлагает изменить выбор с первой двери на вторую. Стоит ли это делать?*

Решение. Рассмотрим два решения — элементарное и через теорему Байеса. Начнём со второго.

Пусть $C_i = \{\text{машина стоит за } i\text{-й дверью}\}$. Очевидно, что $P(C_i) = 1/3$. Теперь введём событие $H = \{\text{ведущий открывает третью дверь}\}$. Так как ведущий не желает открывать дверь с автомобилем, то условные вероятности будут равны

$$\begin{aligned} P(H \mid C_1) &= 1/2 \\ P(H \mid C_2) &= 1 \\ P(H \mid C_3) &= 0 \end{aligned}$$

Теперь посчитаем вероятность $P(C_2 \mid H)$. По теореме Байеса она равна

$$P(C_2 \mid H) = \frac{P(H \mid C_2) P(C_2)}{P(H \mid C_1) P(C_1) + P(H \mid C_2) P(C_2) + P(H \mid C_3) P(C_3)} = \frac{1}{1/2 + 1 + 0} = \frac{2}{3}$$

Если же не менять дверь, то вероятность не изменится и будет равна $1/3$. Поэтому выгоднее изменить выбор двери.

Теперь рассмотрим элементарное решение. Так как вероятность того, что за дверью будет машина, равна $1/3$, то на вторую и третью дверь вместе приходится $2/3$. Так как после открытия третьей двери оказалось, что за ней стояла коза, то вероятность “переходит” второй двери. Тогда вероятность того, что за первой дверью будет машина, равна $1/3$, а за второй — $2/3$. Выбор очевиден. \square

Задача 6 (Задача о поручике Ржевском). *Поручик Ржевский пришёл в казино и решил поиграть на деньги. Сначала у него есть 8 рублей, и он хочет выйти из казино с 256 рублями. Ему предлагают две тактики:*

- *Каждый раз идти ва-банк.*
- *Каждый раз играть на 1 рубль.*

Какую тактику выбрать поручику, если вероятность выигрыша составляет: (a) $1/4$, (b) $3/4$?

Решение. В случае первой тактики всё просто — он не имеет права проиграть. Поэтому вероятность того, что он уйдёт с желаемой суммой, равна p^5 .

Со второй тактикой дела обстоят интереснее. Введём событие $p_l = \{\text{поручик получил желаемое, изначально имея } l \text{ рублей}\}$. По формуле полной вероятности получим рекурсивную формулу. Вместе с начальными условиями получаем систему

$$\begin{aligned} p_l &= (1-p)p_{l-1} + pp_{l+1} \\ p_0 &= 0 \\ p_{256} &= 1 \end{aligned}$$

Выпишем характеристическое уравнение: $p\lambda^2 - \lambda + (1-p) = 0$. Корни этого уравнения имеют вид

$$\frac{1 \pm \sqrt{1-4p+4p^2}}{2p} = \frac{1 \pm (1-2p)}{2p} = \left[\frac{1-p}{p} \right]$$

Тогда получаем, что $p_l = a_1 + a_2 \left(\frac{1-p}{p}\right)^l$. Теперь определим значение констант. Так как $p_0 = 0$, то $0 = a_1 + a_2$ и $a_1 = -a_2 = -a$. Теперь рассмотрим p_{256} :

$$1 = -a + a \left(\frac{1-p}{p}\right)^{256} \implies a = \frac{1}{\left(\frac{1-p}{p}\right)^{256} - 1}$$

Отсюда получаем, что $p_l = \frac{\left(\frac{1-p}{p}\right)^l - 1}{\left(\frac{1-p}{p}\right)^{256} - 1}$ и вероятность успеха

второй стратегии равна p_8 .

Теперь осталось посчитать. Сначала посчитаем для $p = 1/4$:

$$\begin{aligned} P_I &= \left(\frac{1}{4}\right)^5 = \frac{1}{1024} \\ P_{II} &= \frac{\left(\frac{1-1/4}{1/4}\right)^8 - 1}{\left(\frac{1-1/4}{1/4}\right)^{256} - 1} = \frac{3^8 - 1}{3^{256} - 1} \approx \frac{1}{3^{248}} \end{aligned}$$

В таком случае шанс выйти из казино по своей воле выше, если каждый раз играть ва-банк.

Теперь же посчитаем для $p = 3/4$:

$$\begin{aligned} P_I &= \left(\frac{3}{4}\right)^5 = \frac{243}{1024} \\ P_{II} &= \frac{\left(\frac{1-3/4}{3/4}\right)^8 - 1}{\left(\frac{1-3/4}{3/4}\right)^{256} - 1} = \frac{1 - (1/3)^8}{1 - (1/3)^{256}} \approx \frac{6560}{6561} \end{aligned}$$

В таком же случае гораздо безопаснее каждый раз играть на один рубль. □

Задача 7 (Задача о контрольной работе). Пусть студенты A, B, C пишут контрольную. Студент A решает любую задачу с вероятностью $3/4$, студент B — с вероятностью $1/2$, а студент C — с вероятностью $1/4$. В контрольной работе 4 задачи. Преподаватель получает анонимную работу, в которой решено 3 задачи. Кому эта работа скорее всего принадлежит?

Решение. Пусть $D = \{\text{автор решил 3 задачи из 4}\}$. Теперь введём ещё три события: $D_A = \{\text{автор — студент A}\}$, $D_B = \{\text{автор — студент B}\}$, $D_C = \{\text{автор — студент C}\}$. Очевидно, что $P(D_A) = P(D_B) = P(D_C) = \frac{1}{3}$. Найдём условную вероятность события D для разных условий:

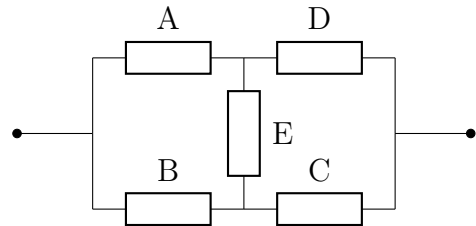
$$\begin{aligned} P(D | D_A) &= \left(\frac{3}{4}\right)^3 \cdot \binom{4}{1} \cdot \frac{1}{4} = \frac{27}{64} \\ P(D | D_B) &= \left(\frac{1}{2}\right)^3 \cdot \binom{4}{1} \cdot \frac{1}{2} = \frac{1}{4} = \frac{16}{64} \\ P(D | D_C) &= \left(\frac{1}{4}\right)^3 \cdot \binom{4}{1} \cdot \frac{3}{4} = \frac{3}{64} \end{aligned}$$

Теперь воспользуемся теоремой Байеса:

$$\begin{aligned} P(D_A | D) &= \frac{P(D | D_A) P(D_A)}{P(D | D_A) P(D_A) + P(D | D_B) P(D_B) + P(D | D_C) P(D_C)} \\ &= \frac{\frac{27}{64}}{\frac{27}{64} + \frac{16}{64} + \frac{3}{64}} = \frac{27}{46} \end{aligned}$$

Аналогично получаем, что $P(D_B | D) = \frac{16}{46}$ и $P(D_C | D) = \frac{3}{46}$. Следовательно, эту работу, скорее всего, сдал студент A . \square

Задача 8. Пусть пять приборов соединены в схему. Каждый из них пропускает ток с вероятностью p . Какова вероятность того, что схема пропускает ток? Какова вероятность того, что есть ток, но при этом E сломан?



Решение. Для начала посмотрим, по каким путям может пройти ток:

- Ток может пойти по AD — тогда вероятность того, что ток будет, равна p^2 .
- Если D сломан, то ток может пойти по AEC . Вероятность такого случая равна $p^3(1-p)$.
- Если A сломан, то ток может пойти по BC . Вероятность этого равна $p^2(1-p)$.
- Если же сломаны и A , и C , то ток пойдёт по BED . Вероятность такого равна $p^3(1-p)^2$.

Тогда итоговая вероятность равна сумме:

$$P(\text{в цепи есть ток}) = p^2 + p^3(1-p) + p^2(1-p) + p^3(1-p)^2.$$

Теперь ответим на второй вопрос. Воспользуемся определением условной вероятности:

$$P(E \text{ не проводит ток} \mid \text{в цепи есть ток}) = \frac{P(E \text{ не проводит ток и в цепи есть ток})}{P(\text{в цепи есть ток})}$$

Вероятность сверху посчитать несложно — достаточно рассмотреть допустимые пути. Тогда ответ равен

$$P(E \text{ не проводит ток} \mid \text{в цепи есть ток}) = \frac{p^2(1-p) + p^2(1-p)^2}{p^2 + p^3(1-p) + p^2(1-p) + p^3(1-p)^2}. \quad \square$$

2.3 Семинар от 23.09.2016

Как обычно, начнём с разбора домашнего задания.

Задача 1. В ящике N различных шаров, из которых ровно M белых. Последовательно вынимают $n \leq N$ шаров. Пусть событие A_k означает, что k -й по счёту вынутый шар — белый, а событие B_m — что всего вынули $m \leq M$ белых шаров.

Найдите $P(A_k | B_m)$, если (а) шары вынимаются без возвращения, (б) с возвращением.

Решение. Начнём со случая, когда нельзя возвращать шары. По определению условной вероятности $P(A_k | B_m) = \frac{P(A_k \cap B_m)}{P(B_m)}$. Для начала посчитаем $P(B_m)$. Как это сделать? Зафиксируем набор из n шаров, в котором первые m шаров белые. Какова вероятность того, что выпадет такой набор? Она равна

$$\frac{M}{N} \cdot \frac{M-1}{N-1} \cdot \dots \cdot \frac{M-m+1}{N-m+1} \cdot \frac{N-M}{N-m} \cdot \frac{N-M-1}{N-m-1} \cdot \dots \cdot \frac{N-M-(n-m)+1}{N-n+1}.$$

Теперь заметим, что если переставить числители местами, то получится вероятность того, что выпадет какой-то другой набор из n шаров, среди которых m белых. Тогда вероятность того, что выпадет хоть какой-то набор, подходящий под это условие, равна

$$P(B_m) = \binom{n}{m} \frac{\frac{M!}{(M-m)!} \frac{(N-M)!}{(N-M-(n-m))!}}{\frac{N!}{(N-n)!}} = \frac{\binom{n}{m} \binom{N-n}{M-m}}{\binom{N}{M}}.$$

Теперь перейдём к числителю. Как посчитать $P(A_k \cap B_m)$? В принципе, точно так же, как и $P(B_m)$. Однако, в данном случае зафиксирована k -я позиция, поэтому нужно лишь выбрать $m-1$ позицию из $n-1$ для белых шаров. Тогда

$$P(A_k \cap B_m) = \frac{\binom{n-1}{m-1} \binom{N-n}{M-m}}{\binom{N}{M}}.$$

Отсюда получаем, что $P(A_k | B_m) = \frac{\binom{n-1}{m-1}}{\binom{n}{m}} = \frac{m}{n}$.

Переходим к случаю (б). Опять же, посчитаем $P[B_m]$ и $P[A_k \cap B_m]$. Рассуждения о перестановке так же имеют место, поэтому:

$$P(B_m) = \binom{n}{m} \frac{M^m (N-M)^{n-m}}{N^n}$$

$$P(A_k \cap B_m) = \binom{n-1}{m-1} \frac{M^m (N-M)^{n-m}}{N^n}$$

Подставляя полученные значения в формулу условной вероятности, получаем, что

$$P(A_k | B_m) = \frac{\binom{n-1}{m-1}}{\binom{n}{m}} = \frac{m}{n}.$$

Ответ: $\frac{m}{n}$ в обоих случаях. □

Задача 2. Ящик содержит a белых и b чёрных шаров (все шары различимы). Наудачу извлекается шар. Он возвращается обратно, и, кроме того, добавляется с шаров одного с ним цвета. Далее, подобная процедура повторяется снова. Пусть событие A_k означает, что на k -м шаге извлечён белый шар. Найдите

- (а) вероятность того, что при первых $n = n_1 + n_2$ извлечениях попало n_1 белых и n_2 чёрных шаров;
- (б) вероятность события A_k ;
- (в) условную вероятность $P(A_m | A_k)$ при $m > k$;
- (г) условную вероятность $P(A_m | A_k)$ при $m < k$;

Решение. Рассмотрим ситуацию, когда последовательно выпало n_1 белых и n_2 чёрных шаров. Какова вероятность такого события? Она равна

$$\frac{a}{a+b} \cdot \frac{a+c}{a+b+c} \cdots \frac{a+(n_1-1)c}{a+b+(n_1-1)c} \cdot \frac{b}{a+b+n_1c} \cdot \frac{b+c}{a+b+(n_1+1)c} \cdots \frac{b+(n_2-1)c}{a+b+(n-1)c}.$$

Теперь переставим числители так, чтобы числители вида $a+x$ и $b+x$ были отсортированы по возрастанию. Тогда эта вероятность будет соответствовать какому-то другому набору из n_1 белых и n_2 чёрных. Пусть $P(a, b, n_1, n_2)$ — вероятность того, что из a белых и b чёрных при первых $n = n_1 + n_2$ извлечениях попало n_1 белых и n_2 чёрных шаров. Тогда

$$P(a, b, n_1, n_2) = \binom{n}{n_1} \frac{a(a+c) \dots (a+(n_1-1)c) b(b+c) \dots (b+(n_2-1)c)}{(a+b)(a+b+c) \dots (a+b+(n-1)c)}.$$

Теперь посчитаем $P(A_k)$. Пусть $C_{ki} = \{\text{до } k\text{-ой процедуры вытащили ровно } i \text{ белых шаров}\}$. Очевидно, что эти события образуют разбиение вероятностного пространства. Тогда по формуле полной вероятности $P(A_k) = \sum_{i=0}^{k-1} P(A_k \cap C_{ki})$. Заметим, что $P(A_k \cap C_{ki})$ совпадает с $P(a, b, i+1, k-i-1)$ с тем лишь отличием, что в данном случае нужно выбрать i позиций из $k-1$:

$$P[A_k \cap C_{ki}] = \binom{k-1}{i} \frac{a(a+c) \dots (a+ic) b(b+c) \dots (b+(k-i)c)}{(a+b)(a+b+c) \dots (a+b+(n-1)c)}.$$

Тогда

$$P(A_k) = \frac{a}{a+b} \sum_{i=0}^{k-1} \binom{k-1}{i} \frac{(a+c) \dots (a+ic) b(b+c) \dots (b+(k-i)c)}{(a+b+c) \dots (a+b+(n-1)c)}.$$

Теперь заметим, что элемент суммы есть ни что иное, как $P(a+c, b, i, k-i-1)$. Но

$$\sum_{i=0}^{k-1} P(a+c, b, i, k-i-1) = 1,$$

так как эта сумма соответствует вероятности вытащить любой набор. Отсюда следует, что

$$P(A_k) = \frac{a}{a+b}.$$

Перейдём к третьему (да и четвёртому тоже) пункту. По определению условной вероятности: $P(A_k | A_m) = \frac{P(A_k \cap A_m)}{P(A_m)}$. Как посчитать числитель? Точно так же, как и во втором случае. Пропустив аналогичные выкладки, выпишем ответ:

$$P(A_k \cap A_m) = \frac{a(a+c)}{(a+b)(a+b+c)}.$$

Тогда получаем, что $P(A_k | A_m) = \frac{a+c}{a+b+c}$. \square

Задача 3. Пусть A, B, C — попарно независимые равновероятные события, причём $A \cap B \cap C = \emptyset$. Найти максимально возможное значение $P(A)$.

Решение. Начнём с того, что заметим следующее: $P(A) \geq P(A \cap (B \cup C)) = P((A \cap B) \cup (A \cap C))$. Так как $A \cap B \cap C = \emptyset$, то $(A \cap B) \cap (A \cap C) = \emptyset$. Следовательно, $P[A] \geq P[A \cap B] + P[A \cap C] = 2(P[A])^2$ и $P(A) \leq 1/2$.

Приведём пример, когда выполняется условие, причём $P(A) = 1/2$. Рассмотрим классическую модель $\Omega = \{1, 2, 3, 4\}$ и события: $A = \{1, 2\}$, $B = \{2, 3\}$, $C = \{3, 4\}$. Легко понять, что данные события удовлетворяют условию и $P(A) = 1/2$. \square

Задача 4. Игроки A и B играют в теннис. При розыгрыше на подаче A игрок A выигрывает с вероятностью p_1 , а при розыгрыше на подаче B — с вероятностью p_2 , все розыгрыши независимы. Игрок A подаёт первым, а выигрывает тот, кто первым наберёт n очков. Существует два варианта правил перехода подачи:

(а) поочерёдная;

(б) игрок подаёт до тех пор, пока не проиграет розыгрыш.

Покажите, что вероятность выигрыша A не зависит от правил перехода подачи, и вычислите её.

Решение. Будем считать, что всего было проведено $2n - 1$ розыгрышей. При таком количестве один игрок гарантированно наберёт не менее n очков, а второй — гарантированно меньше. Теперь опишем вероятностное пространство. Элементарные исходы будут иметь вид $\omega = (a_1, a_2, \dots, a_{2n-1})$, где $a_i \in \{0, 1\}$ (0 соответствует проигрышу, 1 — победе). Согласно этой схеме исход будет подходящим, если в наборе будет не меньше n единиц.

Начнём с пункта (а).

Поймём, как посчитать вероятность какого-либо элементарного случая. Пусть $n = 4$ и мы хотим найти вероятность элементарного исхода 0110110. Она равна $(1 - p_1)p_2p_1(1 - p_2)p_1p_2(1 - p_1) = p_1^2p_2^2(1 - p_1)^2(1 - p_2)^1$. Отсюда получаем закономерность: вероятность элементарного исхода $a_1a_2 \dots a_{2n-1}$ равна

$$p_1^{\sum_{i=0}^{n-1} a_{2i+1}} p_2^{\sum_{i=1}^{n-1} a_{2i}} (1 - p_1)^{n - \sum_{i=0}^{n-1} a_{2i+1}} (1 - p_2)^{n-1 - \sum_{i=1}^{n-1} a_{2i}}.$$

Пусть $k_1 = \sum_{i=0}^{n-1} a_{2i+1}$ — количество единиц на нечётных местах, а $k_2 = \sum_{i=1}^{n-1} a_{2i}$ — на чётных.

Тогда вероятность того, что A выиграл, будет равна

$$P = \sum_{\substack{k_1, k_2 \\ k_1 + k_2 \geq n \\ k_1, k_2 \leq n}} \binom{n}{k_1} \binom{n-1}{k_2} p_1^{k_1} p_2^{k_2} (1 - p_1)^{n-k_1} (1 - p_2)^{n-1-k_2}.$$

Теперь перейдём к пункту (б).

Докажем следующее: между данными методами подачи есть биекция, т.е. игре с поочерёдной подачей можно сопоставить игру с подачей до проигрыша. Рассмотрим частный случай: поочерёдно вышел исход 0110110. Тогда можно “раскидать” партии так:

Подаёт первый: 0110

Подаёт второй: 101

Биекция будет иметь вид 0101101. Как её построить? Разбиваем исход на подачи первого и второго игрока, тем самым получая строки длиной n и $n - 1$. После этого строим по ним новую строку по следующему алгоритму:

Алгоритм 1 Построение исхода в случае подачи до проигрыша по исходу в случае поочерёдной подачи

- 1: Начинаем со строки длины n (строки для первого игрока);
 - 2: Копируем строку посимвольно до тех пор, пока не попадём на 0;
 - 3: Переходим на другую строку;
 - 4: Повторяем два предыдущих шага до тех пор, пока не перенесём все символы.
-

Алгоритм построения исхода при поочерёдной подаче по исходу при подаче до проигрыша будет почти аналогичен. \square

Перейдём к задачам на тему математического ожидания и дисперсии.

Задача 5. Бросили два N -гранных кубика. Пусть ξ — сумма выпавших очков. Найдите $E[\xi]$ и $D[\xi]$.

Решение. Пусть $E[\xi_1]$ — математическое ожидание количество очков, выпавших на первом кубике. Посчитать его несложно: $E[\xi_1] = \sum_{i=1}^N \frac{i}{N} = \frac{N+1}{2}$. Заметим, что $E[\xi] = E[\xi_1 + \xi_2] = E[\xi_1] + E[\xi_2]$. Тогда

$$E[\xi] = \frac{N+1}{2} + \frac{N+1}{2} = N+1.$$

Дисперсию будем считать по следующей формуле: $D[\xi_1] = E[\xi_1^2] - (E[\xi_1])^2$. Посчитаем первый член:

$$E[\xi_1^2] = \sum_{i=1}^N \frac{i^2}{N} = \frac{N(2N+1)(N+1)}{6N} = \frac{(2N+1)(N+1)}{6}.$$

Отсюда получаем, что

$$D[\xi_1] = \frac{N+1}{2} \left(\frac{2N+1}{3} - \frac{N+1}{2} \right) = \frac{(N+1)(N-1)}{12} = \frac{N^2-1}{12}.$$

Заметим, что ξ_1 и ξ_2 независимы (ведь кубики тоже независимы). Тогда $D[\xi] = D[\xi_1] + D[\xi_2]$ и

$$D[\xi] = \frac{N^2-1}{6}. \quad \square$$

Задача 6. Пусть выбрана случайная перестановка $\sigma \in S_n$. Введём случайную величину ξ , равная количеству стационарных точек (чисел i таких, что $\sigma(i) = i$). Найдите $E[\xi]$ и $D[\xi]$.

Решение. Введём событие $A_i = \{\sigma(i) = i\}$. Тогда $\xi = \sum_{i=1}^n I_{A_i}$. По свойству линейности:

$$\mathbb{E}[\xi] = \mathbb{E}\left[\sum_{i=1}^n I_{A_i}\right] = \sum_{i=1}^n \mathbb{E}[I_{A_i}] = \sum_{i=1}^n \mathbb{P}(A_i)$$

Так как $\mathbb{P}(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n}$, то $\mathbb{E}[\xi] = 1$.

Теперь перейдём к подсчёту дисперсии. Распишем дисперсию через ковариации:

$$\mathbb{D}[\xi] = \text{cov}(\xi, \xi) = \text{cov}\left(\sum_{i=1}^n I_{A_i}, \sum_{i=1}^n I_{A_i}\right) = \sum_{i=1}^n \sum_{j=1}^n \text{cov}(I_{A_i}, I_{A_j}).$$

Посчитаем $\text{cov}(I_{A_i}, I_{A_j})$. По свойству ковариации она равна

$$\mathbb{E}[I_{A_i} I_{A_j}] - \mathbb{E}[I_{A_i}] \mathbb{E}[I_{A_j}] = \mathbb{E}[I_{A_i \cap A_j}] - \mathbb{E}[I_{A_i}] \mathbb{E}[I_{A_j}] = \mathbb{P}[A_i \cap A_j] - \mathbb{P}[A_i] \mathbb{P}[A_j].$$

Возникают два случая:

(а) $i = j$. Тогда $\text{cov}(I_{A_i}, I_{A_j}) = \frac{1}{n} - \frac{1}{n^2}$.

(б) $i \neq j$. Тогда $\mathbb{P}[A_i \cap A_j] = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$ и $\text{cov}(I_{A_i}, I_{A_j}) = \frac{1}{n^2(n-1)}$.

Отсюда получаем, что

$$\mathbb{D}[\xi] = \frac{n(n-1)}{2} \frac{1}{n^2(n-1)} + n \left(\frac{1}{n} - \frac{1}{n^2} \right) = 1. \quad \square$$

2.4 Семинар от 30.09.2016

Как обычно, начнём с разбора домашнего задания.

Задача 1. Пусть ξ — некоторая случайная величина. При каком $a \in \mathbb{R}$ достигается минимальное значение функции $f(a) = \mathbb{E}[(\xi - a)^2]$?

Решение. Раскроем матожидание по линейности: $\mathbb{E}[(\xi - a)^2] = \mathbb{E}[\xi^2] - 2a \mathbb{E}[\xi] + a^2$. Теперь добавим и вычтем $(\mathbb{E}[\xi])^2$. Тогда $f(a) = (a - \mathbb{E}[\xi])^2 + \mathbb{E}[\xi^2] - (\mathbb{E}[\xi])^2 = (a - \mathbb{E}[\xi])^2 + \mathbb{D}[\xi]$. Так как $\mathbb{D}[\xi]$ не зависит от a , то минимум достигается при $(a - \mathbb{E}[\xi])^2 = 0 \iff a = \mathbb{E}[\xi]$. \square

Задача 2. Найдите $\mathbb{E}[\xi]$, $\mathbb{D}[\xi]$ и $\mathbb{E}[3\xi]$, если ξ — это а) пуассоновская случайная величина с параметром $\lambda > 0$, б) геометрическая случайная величина с параметром $p \in (0, 1)$.

Решение. Начнём с пуассоновской величины. Как известно, множество её значений равно \mathbb{N} , а $\mathbb{P}(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}$ для любого $k \in \mathbb{N}$. Тогда матожидание ξ равно

$$\mathbb{E}[\xi] = \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \frac{\lambda}{e^{\lambda}} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \frac{\lambda}{e^{\lambda}} e^{\lambda} = \lambda.$$

Перейдём к подсчёту дисперсии:

$$\mathbb{D}[\xi] = \mathbb{E}[\xi^2] - (\mathbb{E}[\xi])^2 = \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} - \lambda^2 = \frac{1}{e^{\lambda}} \sum_{k=1}^{\infty} \frac{k \lambda^k}{(k-1)!} - \lambda^2.$$

Поменяем индекс суммирования, уменьшив его на 1. Тогда дисперсия равна

$$\frac{1}{e^\lambda} \sum_{k=0}^{\infty} \frac{(k+1)\lambda^{k+1}}{k!} - \lambda^2 = \frac{\lambda^2}{e^\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} + \frac{\lambda}{e^\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} - \lambda^2 = \lambda.$$

Теперь перейдём к подсчёту математического ожидания случайной величины 3ξ :

$$\mathbb{E}[3\xi] = \sum_{k=1}^{\infty} 3^k \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{(3\lambda)^k}{k!} = e^{2\lambda}.$$

Перейдём к геометрическому распределению. Напомню, что множество её значений равно \mathbb{N} , а $\mathbb{P}(\xi = k) = p(1-p)^{k-1}$ для любого $k \in \mathbb{N}$. Тогда матожидание равно

$$\mathbb{E}[\xi] = \sum_{k=1}^{\infty} kp(1-p)^{k-1} = p \sum_{k=1}^{\infty} k(1-p)^{k-1}.$$

Как посчитать этот ряд? Возьмём производную от функции $f(x) = \sum_{k=1}^{\infty} x^k, |x| < 1$. Легко увидеть, что она совпадает с этим рядом. Тогда

$$\sum_{k=1}^{\infty} kx^{k-1} = \left(\frac{1}{1-x} \right)' = \frac{1}{(1-x)^2}.$$

Отсюда сразу получаем, что $\mathbb{E}[\xi] = \frac{1}{p}$.

Перейдём к дисперсии: $\mathbb{D}[\xi] = \mathbb{E}[\xi^2] - (\mathbb{E}[\xi])^2 = \mathbb{E}[\xi^2] - \frac{1}{p^2}$. Рассмотрим $\mathbb{E}[\xi^2]$ отдельно:

$$\mathbb{E}[\xi^2] = p \sum_{k=1}^{\infty} k^2(1-p)^{k-1}.$$

С таким рядом простой метод не пройдёт. Но можно добавить и вычесть $\mathbb{E}[\xi]$. Тогда

$$\mathbb{E}[\xi^2] = p \sum_{k=1}^{\infty} k(k-1)(1-p)^{k-1} + \mathbb{E}[\xi] = p(1-p) \sum_{k=1}^{\infty} k(k-1)(1-p)^{k-2} + \frac{1}{p}.$$

Заметим, что ряд равен $\left(\frac{1}{1-x} \right)'' \Big|_{x=1-p} = \frac{2}{(1-x)^3} \Big|_{x=1-p} = \frac{2}{p^3}$. Тогда

$$\mathbb{D}[\xi] = \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

Перейдём к $\mathbb{E}[3\xi]$:

$$\mathbb{E}[3\xi] = 3p \sum_{k=1}^{\infty} 3^{k-1}(1-p)^{k-1}.$$

В данном случае внутри ряда геометрическая прогрессия с знаменателем $3(1-p)$. Для сходимости ряда необходимо, чтобы $3(1-p) < 1 \implies p > 2/3$. Если это не так, то $\mathbb{E}[3\xi] = +\infty$. Иначе же

$$\mathbb{E}[3\xi] = \frac{3p}{1-3(1-p)}.$$

□

Задача 3. На скамейке сидят n человек. Каждый из них независимо бросает игральную шестигранную кость. Случайная величина X равна количеству людей, у которых у хотя бы одного соседа выпало то же число, что и у него самого. Найдите $E[X]$ и $D[X]$.

Решение. Введём событие $A_i = \{\text{у } i\text{-го человека на скамейке число совпало хотя бы с одним из соседей}\}$. Легко понять, что $X = \sum_{i=1}^n I_{A_i}$. Тогда $E[X] = \sum_{i=1}^n P(A_i)$. Чему равно $P(A_i)$? Если $i = 1$ или $i = n$, то число у человека должно совпасть с числом единственного соседа и вероятность равна $1/6$. Если же $1 < i < n$, то покажем, что вероятность равна $11/36$. Пусть мы зафиксировали число у человека посередине. Тогда либо у человека слева совпадает, а у человека справа нет (что даёт 5 вариантов), либо симметрично (ещё 5), либо у обоих сразу (ещё 1). Всего исходов же 36. Отсюда получаем, что

$$E[X] = \frac{11(n-2)}{36} + \frac{1}{3} = \frac{11n-10}{36}.$$

Приступим к подсчёту дисперсии.¹ Распишем дисперсию через сумму ковариаций:

$$D[X] = \text{cov}(X, X) = \sum_{i,j=1}^n \text{cov}(I_{A_i}, I_{A_j}) = \sum_{i,j=1}^n (P(A_i \cap A_j) - P(A_i)P(A_j)).$$

Есть 4 случая:

1. $i = j$. Тогда $\text{cov}(I_{A_i}, I_{A_j}) = \frac{11}{6^2} - \frac{11 \cdot 11}{6^4} = \frac{11 \cdot 25}{6^4}$ для $1 < i < n$ и $\frac{1}{6} - \frac{1}{36} = \frac{5}{36}$ для $i = 1$ или $i = n$.
2. $|i - j| = 1$. Тогда подойдут ситуации $aabb$ и $baac$ (a, b, c — выпавшие числа). Отсюда получаем, что $P(A_i \cap A_j) - P(A_i)P(A_j) = \frac{1}{6} + \frac{6 \cdot 5}{6^4} - \frac{11^2}{36}$ для $1 < i, j < n$. Если, например, $i = 1$, то подойдёт только ситуация aac . Тогда $P(A_i \cap A_j) - P(A_i)P(A_j) = \frac{1}{36} - \frac{11}{6^3}$.
3. $|i - j| = 2$. Тогда подходят ситуации $baaac$ или $bbcaa$. Тогда $P(A_i \cap A_j) = \frac{1}{6^3} + \frac{5 \cdot 4}{6^4}$. В случае $i = 1$ допустимы лишь ситуации $aaac$ или $aabb$. Отсюда получаем, что $P(A_i \cap A_j) = \frac{1}{6^3} + \frac{30}{6^4}$.
4. $|i - j| > 2$. Тогда они никак не влияют на друга и $\text{cov}(I_{A_i}, I_{A_j}) = 0$.

Просуммировав всё это, получаем значение дисперсии. □

Задача 4. Случайный граф $G(n, p)$ получается случайным и независимым удалением ребер из полного графа на n вершинах K_n : любое ребро остается в $G(n, p)$ независимо от других с вероятностью p . Пусть X_n — число треугольников в $G(n, p)$. Вычислите $E[X_n]$ и $D[X_n]$.

Доказательство. Заметим, что всего треугольников в графе на n вершинах можно составить $\binom{n}{3}$. При этом вероятность того, что какие-то три фиксированных ребра будут в графе, равна p^3 . Отсюда сразу получаем, что $E[X_n] = \binom{n}{3} p^3$.

Перейдём к подсчёту дисперсии. Введём событие $A_i = \{i\text{-й треугольник вошёл в случайный граф}\}$ (для этого каким-либо образом пронумеруем все треугольники). Тогда выразим дисперсию через сумму ковариаций индикаторов:

$$D[X_n] = \sum_{i,j=1}^{\binom{n}{3}} \text{cov}(I_{A_i}, I_{A_j}).$$

Есть три случая:

¹Готовьтесь, будет больно.

- \triangle_i и \triangle_j не имеют общих рёбер. Тогда события A_i и A_j независимы и $\text{cov}(I_{A_i}, I_{A_j}) = 0$.
- \triangle_i и \triangle_j имеют одно общее ребро. Тогда $\text{cov}(I_{A_i}, I_{A_j}) = P(A_i \cap A_j) - P(A_i)P(A_j) = p^5 - p^6$.
- \triangle_i и \triangle_j имеют два общих ребра. Тогда эти треугольники совпадают и $\text{cov}(I_{A_i}, I_{A_j}) = P(A_i \cap A_j) - P(A_i)P(A_j) = p^3 - p^6$.

Теперь посчитаем число вторых и третьих случаев. Очевидно, что выбрать 4 вершины из n можно $\binom{n}{4}$ способами. Но при этом эти 4 вершины будут задавать два разных варианта таких треугольников (в зависимости от положения общего ребра). Тогда

$$D[X_n] = 2\binom{n}{4}(p^5 - p^6) + \binom{n}{3}(p^3 - p^6).$$

□

Порой в некоторых задачах нужно доказать, что вероятность того, что случайная величина будет равна нулю, стремится к 0 или к 1. Для упрощения жизни можно доказать два простых и полезных правила:

Метод первого момента: пусть $\{x_n\}_{n=1}^{\infty}$ — последовательность случайных величин с значениями в \mathbb{Z}_+ . Тогда

$$P(x_n > 0) = P(x_n \geq 1) \leq \{\text{по неравенству Маркова}\} \leq E[x_n].$$

Если $E[x_n] \rightarrow 0$ при $n \rightarrow \infty$, то $P(x_n = 0) \rightarrow 1$.

Метод второго момента: пусть $\{x_n\}_{n=1}^{\infty}$ — последовательность случайных величин. Тогда

$$P(x_n = 0) \leq P(x_n \geq 0) = P(|x_n - E[x_n]| \geq E[x_n]) \leq \{\text{по неравенству Чебышёва}\} \leq \frac{D[x_n]}{(E[x_n])^2}.$$

Если $D[x_n] = o((E[x_n])^2)$, то $P(x_n = 0) \rightarrow 0$ при $n \rightarrow \infty$.

Задача 5. Случайный граф $G(n, p)$ получается случайным и независимым удалением ребер из полного графа на n вершинах K_n : любое ребро остается в $G(n, p)$ независимо от других с вероятностью p . Пусть x_n — число треугольников в $G(n, p)$. Докажите, что

$$P(x_n = 0) \rightarrow \begin{cases} 1, & \text{если } np(n) \rightarrow 0, \\ 0, & \text{если } np(n) \rightarrow \infty. \end{cases}$$

Доказательство. Воспользуемся результатами задачи 4. Тогда $E[x_n] = \frac{1}{6}(np)((n-1)p)((n-2)p)$. Пусть $np(n) \rightarrow 0$. Тогда $E[x_n] \rightarrow 0$. Следовательно, по правилу первого момента $P(x_n = 0) \rightarrow 1$.

Теперь предположим, что $np(n) \rightarrow \infty$. Тогда $D[x_n] = O(n^4 p^6)$, а $(E[x_n])^2 = O(n^6 p^6)$. Тогда $D[x_n] = o((E[x_n])^2)$ и по правилу второго момента $P(x_n = 0) \rightarrow 0$ при $n \rightarrow \infty$. □

Примечание. Теперь предположим, что $np(n) \rightarrow c, c > 0$. Тогда легко понять, что $E[x_n] \rightarrow \frac{c^3}{6}$. Вообще, при достаточно больших n x_n будет иметь распределение, близкое к пуассоновскому:

$$E[x_n^k] \rightarrow E[\xi^k], \xi \sim \text{Pois}\left(\frac{c^3}{6}\right).$$

Введём несколько различных определений, которые будут полезны в дальнейшем.

Определение 75. Моментом порядка k для случайной величины ξ называют $E[\xi^k]$.

Определение 76. Центральным моментом порядка k для случайной величины ξ называют $E[(\xi - E[\xi])^k]$.

Определение 77. Факториальным моментом порядка k для случайной величины ξ называют $E[\xi(\xi - 1) \dots (\xi - k + 1)]$.

Примечание. Обычный и факториальный моменты первого порядка совпадают с матожиданием, центральный момент первого порядка всегда равен 0, а второй центральный момент совпадает с дисперсией.

Задача 6. Пусть ξ_1 и ξ_2 — независимые случайные величины, причём $\xi_i \sim \text{Bin}(n_i, p)$. Какое распределение у случайной величины $\xi_1 + \xi_2$?

Решение. Из определения биномиального распределения получаем, что

$$\begin{aligned} P(\xi_1 = k) &= \binom{n_1}{k} p^k (1-p)^{n_1-k}, \\ P(\xi_2 = n - k) &= \binom{n_2}{n-k} p^{n-k} (1-p)^{n_2-n+k}. \end{aligned}$$

Посчитаем $P(\xi_1 + \xi_2 = n)$. Представим n в виде $k + (n - k)$. Теперь поставим ограничения на k :

$$\begin{cases} 0 \leq k \leq n_1 \\ 0 \leq n - k \leq n_2 \end{cases} \implies \begin{cases} 0 \leq k \leq n_1 \\ n - n_2 \leq k \leq n \end{cases}$$

Тогда

$$\begin{aligned} P(\xi_1 + \xi_2 = n) &= \sum_{k=\max(0, n-n_2)}^{\min(n, n_1)} P(\xi_1 = k, \xi_2 = n - k) \\ &= \sum_{k=\max(0, n-n_2)}^{\min(n, n_1)} P(\xi_1 = k) P(\xi_2 = n - k) \\ &= \sum_{k=\max(0, n-n_2)}^{\min(n, n_1)} \binom{n_1}{k} \binom{n_2}{n-k} p^n (1-p)^{n_1+n_2-n} \\ &= p^n (1-p)^{n_1+n_2-n} \sum_{k=\max(0, n-n_2)}^{\min(n, n_1)} \binom{n_1}{k} \binom{n_2}{n-k} \end{aligned}$$

Теперь заметим, что $\sum_{k=\max(0, n-n_2)}^{\min(n, n_1)} \binom{n_1}{k} \binom{n_2}{n-k} = \binom{n_1+n_2}{n}$. Рассуждения для доказательства точно те же, что и для $\sum_{k=0}^n \binom{n}{k}^2 = \binom{2n}{n}$. Тогда

$$P(\xi_1 + \xi_2 = n) = \binom{n_1 + n_2}{n} p^n (1-p)^{n_1+n_2-n} \implies \xi_1 + \xi_2 \sim \text{Bin}(n_1 + n_2, p). \quad \square$$

Примечание. Предположим, что у нас есть n случайных величин, подчиняющихся бернуллиевскому распределению с параметром p . Тогда их сумма будет иметь биномиальное распределение с параметрами n и p . Это ещё раз доказывает, что в случае броска n монет сумма количества выпавших орлов будет подчиняться биномиальному распределению.

Задача 7. Пусть ξ_1 и ξ_2 — независимые случайные величины, причём $\xi_i \sim \text{Pois}(\lambda_i)$. Какое распределение у случайной величины $\xi_1 + \xi_2$?

Решение. По определению пуассоновского распределения

$$P(\xi_1 = k) = \frac{\lambda_1^k}{k!} e^{-\lambda_1}, \quad P(\xi_2 = n - k) = \frac{\lambda_2^{n-k}}{(n-k)!} e^{-\lambda_2}.$$

Тогда

$$\begin{aligned} P(\xi_1 + \xi_2 = n) &= \sum_{k=0}^n P(\xi_1 = k, \xi_2 = n - k) = \sum_{k=0}^n P(\xi_1 = k) P(\xi_2 = n - k) \\ &= \sum_{k=0}^n \frac{\lambda_1^k}{k!} e^{-\lambda_1} \frac{\lambda_2^{n-k}}{(n-k)!} e^{-\lambda_2} = \frac{e^{-(\lambda_1 + \lambda_2)}}{n!} \sum_{k=0}^n \binom{n}{k} \lambda_1^k \lambda_2^{n-k} \\ &= \frac{(\lambda_1 + \lambda_2)^n}{n!} e^{-(\lambda_1 + \lambda_2)}. \end{aligned}$$

Тогда $\xi_1 + \xi_2 \sim \text{Pois}(\lambda_1 + \lambda_2)$. □

Задача 8. Пусть есть n случайных независимых величин $\xi_1, \xi_2, \dots, \xi_n$, причём $\xi_i \sim \text{Geom}(p)$. Какое распределение у $\xi_1 + \xi_2 + \dots + \xi_n$?

Решение. Напомню, что для геометрического распределения $P(\xi_i = k_i) = p(1-p)^{k_i}$. Тогда

$$P(\xi_1 + \xi_2 + \dots + \xi_n = k) = \sum_{k_1 + k_2 + \dots + k_n = k} p^n (1-p)^k = \binom{n+k-1}{n} p^n (1-p)^k.$$

Данное распределение называется отрицательным биномиальным. □

2.5 Семинар от 07.10.2016

Как обычно, начинаем с разбора домашнего задания.

Задача 1. Пусть случайные величины $\xi_1, \dots, \xi_n, \eta_1, \dots, \eta_m$ независимы в совокупности. Докажите, что для любых функций $f: \mathbb{R}^n \rightarrow \mathbb{R}$ и $g: \mathbb{R}^m \rightarrow \mathbb{R}$ случайные величины $f(\xi_1, \dots, \xi_n)$ и $g(\eta_1, \dots, \eta_m)$ тоже независимы.

Доказательство. Вообще, утверждение данной задачи почти что очевидно на интуитивном уровне. Но интуиция может подводить, так что докажем формально.

Рассмотрим $P(f(\xi_1, \dots, \xi_n) = a, g(\eta_1, \dots, \eta_m) = b)$. Заметим, что это равно следующей сумме:

$$\sum_{\substack{a_1, \dots, a_n, b_1, \dots, b_m: \\ f(a_1, a_2, \dots, a_n) = a \\ g(b_1, b_2, \dots, b_m) = b}} \underbrace{P(\xi_1 = a_1, \dots, \xi_n = a_n, \eta_1 = b_1, \dots, \eta_m = b_m)}_{\text{независимы по условию}}.$$

Тогда $P(\xi_1 = a_1, \dots, \xi_n = a_n, \eta_1 = b_1, \dots, \eta_m = b_m) = P(\xi_1 = a_1) \cdot \dots \cdot P(\xi_n = a_n) \cdot P(\eta_1 = b_1) \cdot \dots \cdot P(\eta_m = b_m)$ и, распределяя слагаемые, получим, что эта сумма равна

$$\left(\sum_{\substack{a_1, \dots, a_n: \\ f(a_1, a_2, \dots, a_n) = a}} P(\xi_1 = a_1) \cdot \dots \cdot P(\xi_n = a_n) \right) \left(\sum_{\substack{b_1, \dots, b_m: \\ g(b_1, b_2, \dots, b_m) = b}} P(\eta_1 = b_1) \cdot \dots \cdot P(\eta_m = b_m) \right).$$

Но это есть ни что иное, как $P(f(\xi_1, \dots, \xi_n) = a) P(g(\eta_1, \dots, \eta_m) = b)$. Отсюда получаем независимость $f(\xi_1, \dots, \xi_n)$ и $g(\eta_1, \dots, \eta_m)$, что и требовалось доказать. \square

Задача 2. *Случайные величины X, Y, Z, W независимы в совокупности и одинаково распределены: каждая равновероятно принимает значения 1 и -1 . Являются ли независимыми в совокупности следующие наборы случайных величин:*

1. XYZ, XYW, XW ;
2. XY, XYW, XW, XZW ;
3. XYZ, XYW, XW, XZW ?

Доказательство. Для начала заметим следующее. Пусть у нас есть две случайных независимые величины ξ и η , равновероятно принимающие значения из $\{-1, 1\}$. Тогда случайная величина $\xi\eta$ тоже равновероятно принимает значения из $\{-1, 1\}$. Покажем это:

$$P(\xi\eta = 1) = P(\xi = 1, \eta = 1) + P(\xi = -1, \eta = -1) = \frac{1}{2}.$$

Это доказывает, что любое произведение X, Y, Z, W , в котором каждая случайная величина участвует не более одного раза, равновероятно принимает значение из $\{-1, 1\}$. Теперь посмотрим на наборы переменных из условия.

1. В данном случае нам надо показать, что для любого набора $(a_1, a_2, a_3), a_i \in \{-1, 1\}$ выполнено

$$P(XYZ = a_1, XYW = a_2, XW = a_3) = \frac{1}{8} = P(XYZ = a_1) P(XYW = a_2) P(XW = a_3).$$

Разложим вероятность слева по формуле полной вероятности для событий $Z = 0$ и $Z = 1$. В первом случае необходимо определить, какие значения X, Y и W должны иметь для выполнения условия $XY = a_1, XYW = a_2, XW = a_3, Z = 1$. Легко проверить, что из этого следует, что $X = \frac{a_1 a_3}{a_2}, Y = \frac{a_2}{a_3}, Z = 1, W = \frac{a_2}{a_1}$. Тогда

$$P(XY = a_1, XYW = a_2, XW = a_3 \mid Z = 1) = \frac{1}{16}.$$

Во втором же случае она считается практически аналогично и равна

$$P\left(X = -\frac{a_1 a_3}{a_2}, Y = \frac{a_2}{a_3}, Z = -1, W = -\frac{a_2}{a_1}\right) = \frac{1}{16}.$$

Суммируя их, получаем, что $P(XYZ = a_1, XYW = a_2, XW = a_3) = \frac{1}{8}$. Тогда случайные величины XYZ, XYW, XW независимы.

2. Решим систему уравнений:

$$\begin{cases} XY = a_1 \\ XYW = a_2 \\ XW = a_3 \\ XZW = a_4 \end{cases} \implies \begin{cases} X = \frac{a_1 a_3}{a_2} \\ Y = \frac{a_2}{a_3} \\ Z = \frac{a_4}{a_3} \\ W = \frac{a_2}{a_1} \end{cases}$$

Отсюда следует, что $P(XY = a_1, XYW = a_2, XW = a_3, XZW = a_4) = \frac{1}{16}$. Отсюда следует независимость.

3. Опять же, решим систему уравнений:

$$\begin{cases} XYZ = a_1 \\ XYW = a_2 \\ XW = a_3 \\ XZW = a_4 \end{cases} \implies \begin{cases} X = \frac{a_1 a_3^2}{a_2 a_4} \\ Y = \frac{a_2}{a_3} \\ Z = \frac{a_4}{a_3} \\ W = \frac{a_2 a_4}{a_1 a_3} \end{cases}$$

Тем самым снова получаем независимость.

□

Задача 3. Докажите, что если последовательность случайных величин $\{\xi_n\}_{n=1}^\infty$ сходится по вероятности к двум случайным величинам ξ, η , $\xi_n \xrightarrow{P} \xi$, $\xi_n \xrightarrow{P} \eta$, то $P(\xi = \eta) = 1$.

Доказательство. Суть у данной задачи следующая: сходимость по вероятности даёт настоящий предел, который единственен.

Из определения сходимости по вероятности следует, что для любого $\varepsilon > 0$

$$\begin{aligned} P\left(|\xi_n - \xi| > \frac{\varepsilon}{2}\right) &\rightarrow 0, \\ P\left(|\xi_n - \eta| > \frac{\varepsilon}{2}\right) &\rightarrow 0. \end{aligned}$$

Тогда рассмотрим $P(|\xi - \eta| > \varepsilon)$:

$$\begin{aligned} P(|\xi - \eta| > \varepsilon) &= P(|\xi - \xi_n + \xi_n - \eta| > \varepsilon) \\ &\leq P(|\xi_n - \xi| + |\xi_n - \eta| > \varepsilon) \\ &\leq P\left(\left\{|\xi_n - \xi| > \frac{\varepsilon}{2}\right\} \cup \left\{|\xi_n - \eta| > \frac{\varepsilon}{2}\right\}\right) \\ &\leq P\left(|\xi_n - \xi| > \frac{\varepsilon}{2}\right) + P\left(|\xi_n - \eta| > \frac{\varepsilon}{2}\right) \rightarrow 0. \end{aligned}$$

Второе неравенство легко понять, если представить его графически. Изобразим первую четверть координатной плоскости с координатами $x = |\xi_n - \xi|$ и $y = |\xi_n - \eta|$. Тогда для выражения справа недопустимая область будет иметь вид прямоугольного треугольника с координатами вершин $(0, 0)$, $(0, 1)$ и $(1, 0)$. Для выражения справа недопустимая область — это квадрат с координатами $(0, 0)$, $(0, 1/2)$, $(1/2, 1/2)$ и $(1/2, 0)$. Но данный квадрат полностью включается в треугольник. □

Задача 4. Случайные величины $\{\xi_n\}_{n=1}^{\infty}$ независимы в совокупности. Обозначим $S_n = \xi_1 + \xi_2 + \dots + \xi_n$. Пусть случайная величина ξ_n

1. принимает три значения $\{-n, 0, n\}$ с вероятностями $(\frac{1}{2n^2}, 1 - \frac{1}{n^2}, \frac{1}{2n^2})$;
2. принимает три значения $\{-2^n, 0, 2^n\}$ с вероятностями $(2^{-n-1}, 1 - 2^{-n}, 2^{-n-1})$.

Выясните, в каком случае выполнен закон больших чисел: $\frac{S_n}{n} \xrightarrow{P} 0$ при $n \rightarrow \infty$?

Доказательство. Для начала заметим, что и в первом, и во втором случае матожидание ξ_n равно 0. Тогда $E[S_n] = 0$ и можно применить неравенство Чебышёва:

$$P\left(\left|\frac{S_n}{n}\right| \geq \alpha\right) \leq \frac{D[S_n]}{\alpha n^2}.$$

Теперь посчитаем $D[\xi_n]$:

1. $D[\xi_n] = \frac{1}{2n^2}n^2 + \frac{1}{2n^2}n^2 = 1$,
2. $D[\xi_n] = \frac{1}{2^{n+1}}2^{2n} + \frac{1}{2^{n+1}}2^{2n} = 2^n$.

Отсюда видно, что в первом случае закон больших чисел точно выполняется. Однако про второй мы ничего сказать не можем. Как быть?

Зафиксируем некоторое $\omega \in \Omega$ и рассмотрим последовательность $\{\xi_n(\omega)\}_{n=1}^{\infty}$. Утверждается следующее: в данной последовательности, начиная с некоторого номера N , будут лишь нули. Докажем это. Зафиксируем некоторое натуральное число m и рассмотрим следующую вероятность:

$$P(\exists n \geq m : \xi_n \neq 0) = P\left(\sum_{n=m}^{\infty} \{\xi_n \neq 0\}\right) \leq \sum_{n=m}^{\infty} P(\xi_n \neq 0) = \sum_{n=m}^{\infty} \frac{1}{2^n} = \frac{1}{2^{m-1}} \xrightarrow{m \rightarrow \infty} 0.$$

Теперь посчитаем вероятность того, что последовательность $\{\xi_n(\omega)\}_{n=1}^{\infty}$ не финитна (то есть содержит бесконечное число ненулевых элементов). Она равна

$$P(\forall m \in \mathbb{N} \exists n \geq m : \xi_n(\omega) \neq 0) = \lim_{m \rightarrow \infty} P(\exists n \geq m : \xi_n \neq 0) = 0.$$

Это и доказывает финитность $\{\xi_n(\omega)\}_{n=1}^{\infty}$. Тогда для любого $\omega \in \Omega$ выполнено

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n \xi_i(\omega)}{n} = 0 \implies \frac{S_n}{n} \xrightarrow{P} 0.$$

Это доказывает то, что и для второго случая тоже выполнен закон больших чисел. \square

Теперь обсудим применения вероятностного метода к решению задач.

Рассмотрим полный граф на n вершинах. Покрасим его ребра в два цвета. Тогда имеет место следующая теорема:

Теорема 2 (Рамсей). В достаточно большом полном графе с произвольной раскраской рёбер в белый и чёрный цвета найдётся чёрный полный подграф размера r и белый полный подграф размера s для наперёд заданных r и s .

Определение 78. Пусть $m, n \in \mathbb{N}$. Число Рамсея $R(m, n)$ — это наименьшее из таких чисел $x \in \mathbb{N}$, что при любой раскраске ребер полного графа на x вершинах в два цвета найдется клика на n вершинах с ребром цвета 1 или клика на m вершинах с ребром цвета 2.

Задача 5 (Спенсер, 1974). *Докажите, что*

$$R(k, k) \geq \frac{\sqrt{2}}{e} k 2^{k/2} (1 + o(1)).$$

Доказательство. Положим, что у нас есть полный граф на n вершинах, раскрашенный в два цвета произвольным образом.

Пусть S — это некоторый подграф на k вершинах. Тогда введём событие $A_S = \{\text{все рёбра в } S \text{ одного цвета}\}$. Легко понять, что $P(A_S) = 2^{1-\binom{k}{2}}$. Заметим, что если для двух подграфов S и T размера k выполнено $|S \cap T| \leq 1$, то события A_S и A_T независимы (так как эти подграфы не содержат общих рёбер). Тогда событие A_S зависит от d событий, где

$$d = |\{T : |T| = k, |S \cap T| > 1\}| = \binom{k}{2} \binom{n-2}{k-2}.$$

Эту оценку можно получить достаточно просто: достаточно выбрать какие-либо две вершины в S и сказать, что эти вершины принадлежат T , после чего добавить $k-2$ каких-либо вершин в T .

Осталось лишь найти n такие, что выполняется требование локальной леммы:

$$ep(d+1) \leq 1 \iff e 2^{1-\binom{k}{2}} \left(\binom{k}{2} \binom{n-2}{k-2} + 1 \right) \leq 1.$$

Преобразуем это неравенство, пользуясь следствием из формулы Стирлинга: $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$

$$\begin{aligned} \binom{k}{2} \binom{n-2}{k-2} &\leq \frac{2^{\binom{k}{2}}}{2e} - 1 \\ \left(\frac{en}{k}\right)^{k-2} &\leq \frac{2^{\binom{k}{2}}}{ek(k-1)} - \frac{1}{k(k-1)} \\ \left(\frac{en}{k}\right)^{k-2} &\leq 2^{\binom{k}{2}} \left(\frac{1}{ek(k-1)} + o(1) \right) \\ \left(\frac{en}{k}\right)^{k-2} &\leq 2^{\binom{k}{2}} (1 + o(1)) \\ n &\leq \frac{1}{e} k 2^{\frac{k}{2} + \frac{k}{2(k-2)}} (1 + o(1)) \\ n &\leq \frac{\sqrt{2}}{e} k 2^{\frac{k}{2}} (1 + o(1)) \end{aligned}$$

Отсюда получаем, что $R(k, k) \geq \frac{\sqrt{2}}{e} k 2^{k/2} (1 + o(1))$.

□

2.6 Семинар от 21.10.2016

Так как на прошлом семинаре была контрольная работа, то начнём с её разбора.

Задача 1. В колоде содержится 52 карты четырёх мастей. Каждому из четырёх игроков раздают по шесть карт. Найдите вероятность того, что хотя бы у одного игрока будут карты одной масти.

Доказательство. Введём событие $A_i = \{\text{у } i\text{-го игрока все карты одной масти}\}$. Тогда вероятность из условия равна $P(A_1 \cup A_2 \cup A_3 \cup A_4)$. Пользуясь формулой включений-исключений, получаем, что

$$P(A_1 \cup A_2 \cup A_3 \cup A_4) = P(A_1) + \dots + P(A_4) - P(A_1 \cap A_2) - \dots - P(A_3 \cap A_4) + P(A_1 \cap A_2 \cap A_3) + \dots + P(A_2 \cap A_3 \cap A_4) - P(A_1 \cap A_2 \cap A_3 \cap A_4).$$

Осталось посчитать эти вероятности:

$$P(A_i) = \frac{4 \cdot \binom{13}{6}}{\binom{52}{6}}, \quad P(A_i \cap A_j) = \frac{4 \cdot \binom{13}{6} \binom{7}{6}}{\binom{52}{6} \binom{46}{6}} + \frac{4 \cdot 3 \cdot \binom{13}{6}^2}{\binom{52}{6} \binom{46}{6}},$$

$$P(A_i \cap A_j \cap A_k) = \frac{4 \cdot 3 \cdot \binom{3}{2} \binom{13}{6}^2 \binom{7}{6}}{\binom{52}{6} \binom{46}{6} \binom{40}{6}} + \frac{4 \cdot 3 \cdot 2 \cdot \binom{13}{6}^3}{\binom{52}{6} \binom{46}{6} \binom{40}{6}},$$

$$P(A_1 \cap A_2 \cap A_3 \cap A_4) = \frac{4! \cdot \binom{13}{6}^4 + 4 \cdot 3 \cdot \binom{4}{2} \binom{13}{6}^2 \binom{7}{6}^2 + 4 \cdot 3 \cdot 2 \cdot \binom{4}{2} \binom{13}{6}^3 \binom{7}{6}}{\binom{52}{6} \binom{46}{6} \binom{40}{6} \binom{34}{6}}. \quad \square$$

Задача 2. Два ящика содержат красные и чёрные шары. Ящик A содержит 2 красных и 1 чёрный шар. Ящик B — 101 красный и 100 черных шаров. Вам предлагается сыграть в следующую игру. В тайне от вас равновероятно выбирается один из ящиков. Ваша задача — угадать, какой ящик был выбран. Для этого из ящика последовательно вынимается два шара. После вытаскивания первого шара и определения его цвета вы решаете, вернуть ли в ящик этот шар перед вторым вытаскиванием или нет. Какой должна быть ваша стратегия, чтобы максимизировать вероятность выигрыша?

Доказательство. Рассмотрим вероятности вытащить определённый набор в зависимости от того, какой ящик выбран:

Ящик Шары	I(возврат)	II(возврат)	I(без возврата)	II(без возврата)
КК	$\frac{4}{9}$	$\frac{101^2}{201^2}$	$\frac{1}{3}$	$\frac{101 \cdot 100}{201 \cdot 200}$
КЧ/ЧК	$\frac{2}{9}$	$\frac{101 \cdot 100}{201^2}$	$\frac{1}{3}$	$\frac{101 \cdot 100}{201 \cdot 200}$
ЧЧ	$\frac{1}{9}$	$\frac{100^2}{201^2}$	0	$\frac{100 \cdot 99}{201 \cdot 200}$

Теперь можно составить таблицу вероятностей того, какой ящик выбран, в зависимости от того, какой набор был вытащен:

Ящик Шары	I(возврат)	II(возврат)	I(без возврата)	II(без возврата)
КК	$\frac{17956}{28157} \approx 0.638$	$\frac{10201}{28157} \approx 0.362$	$\frac{13400}{23500} \approx 0.57$	$\frac{10100}{23500} \approx 0.43$
КЧ/ЧК	$\frac{8978}{19078} \approx 0.471$	$\frac{10100}{19078} \approx 0.529$	$\frac{13400}{23500} \approx 0.57$	$\frac{10100}{23500} \approx 0.43$
ЧЧ	$\frac{4489}{14489} \approx 0.31$	$\frac{10000}{14489} \approx 0.69$	0	1

Допустим, что на первом вытаскивании нам попался красный шар. Что выгоднее: оставить его или же вернуть назад? Если мы оставим шар, то максимальный шанс выигрыша равен 0.57. Если же мы вернём его назад, то минимальный шанс выигрыша будет равен 0.638. Тогда выгоднее вернуть красный шар. Если же мы вынимаем чёрный, то его лучше оставить. При этом указывать на первый ящик стоит только тогда, когда выпало два красных шара или чёрный, а затем красный. \square

Задача 3. Случайные величины $\{\xi_n\}_{n=1}^{\infty}$ независимы в совокупности и принимают значения 3 и -2 с вероятностью p и $1-p$ соответственно. Обозначим $S_n = \xi_1 + \dots + \xi_n$. Найдите условную вероятность

$$P(S_n = x \mid S_k = y, S_m = z) \text{ при } m < n < k, x, y, z \in \mathbb{Z}.$$

Доказательство. По определению условной вероятности

$$P(S_n = x \mid S_k = y, S_m = z) = \frac{P(S_k = y, S_n = x, S_m = z)}{P(S_k = y, S_m = z)}.$$

Преобразуем числитель и знаменатель:

$$\begin{aligned} P(S_k = y, S_n = x, S_m = z) &= P(\underbrace{S_k - S_n = y - x, S_n - S_m = x - z, S_m = z}_{\text{независимы}}) \\ &= P(S_k - S_n = y - x) P(S_n - S_m = x - z) P(S_m = z), \\ P(S_k = y, S_m = z) &= P(S_k - S_m = y - z) P(S_m = z). \end{aligned}$$

Отсюда получаем, что

$$P(S_n = x \mid S_k = y, S_m = z) = \frac{P(S_k - S_n = y - x) P(S_n - S_m = x - z)}{P(S_k - S_m = y - z)}.$$

Осталось посчитать какую-либо из вероятностей. Остальные считаются абсолютно аналогично. Допустим, посчитаем $P(S_k - S_n = y - x)$. Для этого необходимо, чтобы в сумме было l случайных величин, равных 3, и $k - n - l$ случайных величин, равных -2 . Тогда

$$y - x = 3l - 2n + 2k + 2l \implies l = \frac{y - x + 2n - 2k}{5}$$

Отсюда получаем необходимые условия на то, что данная вероятность не равна 0: $y - x + 2n - 2k$ должно быть больше 0 и делиться на 5. Тогда

$$P(S_k - S_n = y - x) = \binom{k-n}{l} p^l (1-p)^{k-n-l}.$$

Посчитав две остальных вероятности и сократив, мы получим ответ. Что самое интересное, он не зависит от p . \square

Задача 4. Каждую целочисленную точку числовой оси независимо от других назовём белой с вероятностью p или чёрной с вероятностью $1 - p$. Пусть S — множество всех таких целочисленных точек x , что расстояние от x до ближайшей черной точки (включая x , если x — чёрная точка) не меньше расстояния от x до начала координат. Найдите математическое ожидание и дисперсию числа элементов в S .

Доказательство. Введём событие $A_i = \{i \in S\}$. Тогда легко понять, что

$$P(A_i) = \begin{cases} 1 & i = 0, \\ p^{2|i|-1} & i \neq 0. \end{cases}$$

Отсюда получаем, что

$$E[S] = 2 \sum_{i=1}^{\infty} p^{2i-1} + 1 = \frac{2p}{1-p^2} + 1.$$

Теперь посчитаем дисперсию. Выразим её через сумму ковариаций:

$$D[S] = \sum_{i,j \in \mathbb{Z}} \text{cov}(I_{A_i}, I_{A_j}) = \sum_{i \in \mathbb{Z}} \text{cov}(I_{A_i}, I_{A_i}) + 2 \sum_{\substack{i,j \in \mathbb{Z} \\ i > j}} \text{cov}(I_{A_i}, I_{A_j}).$$

Так как $\text{cov}(I_{A_i}, I_{A_j}) = P(A_i \cap A_j) - P(A_i)P(A_j)$ и $\text{cov}(I_{A_i}, I_{A_j}) = 0$ при $ij \leq 0$, то дисперсию можно записать в виде

$$D[S] = 2 \sum_{i=1}^{\infty} (p^{2i-1} - p^{4i-2}) + 1 + 2 \sum_{i=1}^{\infty} \sum_{j=i+1}^{\infty} (p^{2j-1} - p^{2(i+j)-2}). \quad \square$$

Теперь вернёмся к традиции разбирать домашнее задание на семинаре.

Задача 5. Задача без условия на ЛЛЛ

Доказательство. содержимое... □

Задача 6. Что-то там про депутатов

Доказательство. содержимое... □

Теперь обсудим задачи на тему общего понятия вероятности события.

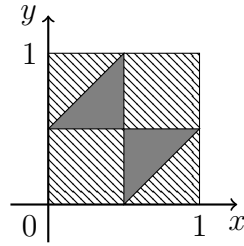
Задача 7. Палочку разламывают в двух произвольных местах. Найдите вероятность того, что из полученных кусков можно сложить треугольник.

Решение. Без ограничения общности положим длину палочки, равную 1.

Пусть палочку разломали в точках x и y , $x < y$. При каких условиях на x и y можно сложить треугольник? Очевидно, если для всех трёх сторон выполнено неравенство треугольника: длина стороны не превосходит суммы длин двух других сторон. Так как стороны равны x , $y - x$ и $1 - y$, то

$$\begin{cases} x < y - x + 1 - y \\ y - x < x + 1 - y \\ 1 - y < x + y - x \end{cases} \implies \begin{cases} x < \frac{1}{2} \\ y < \frac{1}{2} + x \\ y > \frac{1}{2} \end{cases}$$

Изобразим решение этих неравенств и все возможные значения x и y (подходящие значения (x, y) выделены серым):

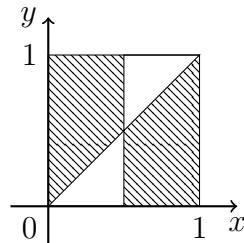


Отсюда получаем, что вероятность этого события равна $\frac{1}{4}$. \square

Немного усложним задачу.

Задача 8. Пусть палочку длины 1 ломают следующим образом: сначала её ломают в произвольном месте, затем более длинный кусок ломают в произвольном месте. Найдите вероятность того, что из полученных кусков можно сложить треугольник.

Решение. На первый взгляд может показаться, что эта задача ничем не отличается от предыдущей. Но нет. В данной задаче отличается множество допустимых пар (x, y) . Изобразим это графически, заштриховав допустимые значения:



При этом подходящие пары (x, y) те же, что и раньше. Отсюда получаем, что вероятность этого события равна $\frac{1}{3}$. \square

2.7 Семинар от 15.11.2016

Начни свой день с разбора домашнего задания.

Задача 1. Найдите математическое ожидание и дисперсию случайной величины ξ , если

1. $\xi \sim U(a, b)$ — равномерное распределение на отрезке $[a, b]$,
2. $\xi \sim \text{Exp}(\lambda)$ — экспоненциальное распределение с параметром $\lambda > 0$.

Доказательство. Идея этой задачи состоит в следующем: нужно найти плотность случайной величины, после чего посчитать первый и второй моменты.

1. В случае равномерного распределения $p_\xi(x) = \frac{1}{b-a} I\{x \in [a, b]\}$. Тогда

$$E[\xi] = \int_{-\infty}^{+\infty} \frac{x}{b-a} I\{x \in [a, b]\} dx = \frac{1}{b-a} \int_a^b x dx = \frac{a+b}{2}.$$

Теперь посчитаем второй момент:

$$E[\xi^2] = \int_{-\infty}^{+\infty} \frac{x^2}{b-a} I\{x \in [a, b]\} dx = \frac{1}{b-a} \int_a^b x^2 dx = \frac{a^2 + ab + b^2}{3}.$$

Тогда

$$D[\xi] = E[\xi^2] - (E[\xi])^2 = \frac{a^2 + ab + b^2}{3} - \frac{a^2 + 2ab + b^2}{4} = \frac{(b-a)^2}{12}.$$

2. Вспомним, что $p_\xi(x) = \lambda e^{-\lambda x} I\{x \geq 0\}$. Тогда

$$\begin{aligned} E[\xi] &= \int_{-\infty}^{+\infty} \lambda x e^{-\lambda x} I\{x \geq 0\} dx = \frac{1}{\lambda} \int_0^{+\infty} t e^{-t} dt \\ &= \frac{1}{\lambda} \left((-te^{-t}) \Big|_0^{+\infty} + \int_0^{+\infty} e^{-t} dt \right) = \frac{1}{\lambda}. \end{aligned}$$

Теперь посчитаем второй момент:

$$\begin{aligned} E[\xi^2] &= \int_{-\infty}^{+\infty} \lambda x^2 e^{-\lambda x} I\{x \geq 0\} dx = \frac{1}{\lambda^2} \int_0^{+\infty} t^2 e^{-t} dt \\ &= \frac{1}{\lambda} \left((-t^2 e^{-t}) \Big|_0^{+\infty} + 2 \int_0^{+\infty} t e^{-t} dt \right) = \frac{2}{\lambda^2}. \end{aligned}$$

Отсюда получаем, что

$$D[\xi] = E[\xi^2] - (E[\xi])^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

□

2.8 Семинар от 22.11.2016

Начнём с разбора домашнего задания.

Задача 1. В треугольнике со сторонами 3, 4 и 5 выбирается случайная точка X , после чего вычисляется случайная величина ξ — сумма длин высот, опущенных из X на стороны треугольника. Вычислите $E[\xi]$.

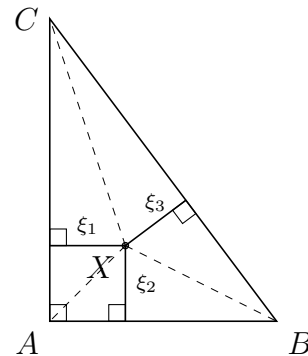
Решение. Для начала заметим, что данный треугольник прямоугольный.

Пусть ξ_1 , ξ_2 и ξ_3 — длины опущенных высот. Тогда $\xi = \xi_1 + \xi_2 + \xi_3$ и $E[\xi] = E[\xi_1 + \xi_2 + \xi_3] = E[\xi_1] + E[\xi_2] + E[\xi_3]$. Теперь заметим следующее:

$$\frac{1}{2} (4\xi_1 + 3\xi_2 + 5\xi_3) = 6.$$

Отсюда получаем, что достаточно вычислить математическое ожидание каких-то двух случайных величин — математическое ожидание третьей будет выражаться через них.

Начнём с подсчёта $E[\xi_1]$. Для начала необходимо посчитать плотность распределения $\rho_{\xi_1}(x)$. Для



этого посчитаем функцию распределения: $F_{\xi_1}(x) = P(\xi_1 \leq x)$.

Как посчитать такую вероятность? Представим, что у нас введена система координат, где AB — это ось OX , а AC — ось OY . Тогда в данном случае вертикальная координата не играет роли. Следовательно, область подходящих точек будет иметь вид прямоугольной трапеции с основанием AC и высотой x . Это равносильно тому, что область неподходящих точек будет иметь вид треугольника, подобного ABC с коэффициентом подобия $\frac{3-x}{3}$. Отсюда получаем, что

$$P(\xi_1 \leq x) = \frac{6 - 6\left(\frac{3-x}{3}\right)^2}{6} = 1 - \left(1 - \frac{x}{3}\right)^2.$$

Отсюда получаем, что если $x \in [0, 3]$, то

$$\rho_{\xi_1}(x) = \frac{dF_{\xi_1}(x)}{dx} = \frac{2}{3} \left(1 - \frac{x}{3}\right) = \frac{2}{9}(3 - x).$$

Теперь можно считать математическое ожидание ξ_1 :

$$E[\xi_1] = \int_0^3 \frac{2}{9}x(3 - x) dx = \frac{2}{9} \int_0^3 (3x - x^2) dx = \frac{2}{9} \left(\frac{27}{2} - \frac{27}{3}\right) = 1.$$

Математическое ожидание ξ_2 считается аналогично. Недолго думая, выпишу ответ:

$$E[\xi_2] = \int_0^4 \frac{1}{8}x(4 - x) dx = \frac{1}{8} \int_0^4 (4x - x^2) dx = \frac{1}{8} \left(\frac{64}{2} - \frac{64}{3}\right) = \frac{4}{3}.$$

Теперь выразим $E[\xi_3]$ через $E[\xi_1]$ и $E[\xi_2]$:

$$4E[\xi_1] + 3E[\xi_2] + 5E[\xi_3] = 12 \iff E[\xi_3] = \frac{4}{5}.$$

В итоге получаем, что

$$E[\xi] = E[\xi_1] + E[\xi_2] + E[\xi_3] = 4 \left(\frac{1}{3} + \frac{1}{4} + \frac{1}{5}\right).$$

Заметим, что точка с координатами $(1, \frac{4}{3})$ есть ни что иное, как *центр масс* (точка пересечения медиан) треугольника. \square

Задача 2. *Случайные величины X и Y независимы, X имеет экспоненциальное распределение с параметром 1, а Y — равномерное на $[0, 1]$. Случайная величина Z равна $Z = \max(X, Y)$. Вычислите $E[Z]$ и $D[Z]$.*

Решение. Для начала вычислим функцию распределения $F_Z(x)$. Так как и X , и Y неотрицательны, то и Z тоже неотрицательно. Следовательно, если $x < 0$, то $F_Z(x) = 0$. Теперь посчитаем её для $x \geq 0$:

$$F_Z(x) = P(Z \leq x) = P(\max(X, Y) \leq x) = P(X \leq x, Y \leq x) = P(X \leq x) P(Y \leq x).$$

Отсюда получаем, что

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

Теперь несложно получить плотность (в данном случае функция распределения гладкая, поэтому плотность не терпит разрывов):

$$\rho_Z(x) = \frac{dF_Z(x)}{dx} = \begin{cases} 0, & x < 0, \\ 1 - e^{-x} + xe^{-x}, & 0 \leq x \leq 1, \\ e^{-x}, & x > 1. \end{cases}$$

Осталось посчитать матожидание:

$$E[Z] = \int_0^1 x(1 - e^{-x} + xe^{-x}) dx + \int_1^{+\infty} xe^{-x} dx.$$

Посчитаем каждый интеграл по отдельности:

1. Разобьём его в сумму трёх интегралов:

$$\int_0^1 x(1 - e^{-x} + xe^{-x}) dx = \int_0^1 x dx - \int_0^1 xe^{-x} dx + \int_0^1 x^2 e^{-x} dx.$$

Теперь начнём считать их. Понятно, что первый интеграл равен

$$\int_0^1 x dx = \left. \frac{x^2}{2} \right|_0^1 = \frac{1}{2}.$$

Второй интеграл возьмём по частям:

$$\int_0^1 xe^{-x} dx = (-xe^{-x}) \Big|_0^1 + \int_0^1 e^{-x} dx = -\frac{1}{e} + (-e^{-x}) \Big|_0^1 = 1 - \frac{2}{e}.$$

Третий интеграл тоже возьмём по частям:

$$\int_0^1 x^2 e^{-x} dx = (-x^2 e^{-x}) \Big|_0^1 + 2 \int_0^1 xe^{-x} dx = -\frac{1}{e} + 2 - \frac{4}{e} = 2 - \frac{5}{e}.$$

Отсюда получаем, что

$$\int_0^1 x(1 - e^{-x} + xe^{-x}) dx = \frac{1}{2} - 1 + \frac{2}{e} + 2 - \frac{5}{e} = \frac{3}{2} - \frac{3}{e}.$$

2. Второй интеграл тоже возьмём по частям:

$$\int_1^{+\infty} x e^{-x} dx = (-x e^{-x}) \Big|_1^{+\infty} + \int_1^{+\infty} e^{-x} dx = \frac{1}{e} + (-e^{-x}) \Big|_1^{+\infty} = \frac{2}{e}.$$

В итоге получаем, что $E[Z] = \frac{3}{2} - \frac{1}{e}$.

Теперь посчитаем дисперсию. Она равна $D[Z] = E[Z^2] - (E[Z])^2$. Тогда нам достаточно посчитать второй момент.

$$E[Z^2] = \int_0^1 x^2(1 - e^{-x} + x e^{-x}) dx + \int_1^{+\infty} x^2 e^{-x} dx.$$

И снова посчитаем каждый интеграл по отдельности:

1. Разобьём его в сумму трёх интегралов:

$$\int_0^1 x^2(1 - e^{-x} + x e^{-x}) dx = \int_0^1 x^2 dx - \int_0^1 x^2 e^{-x} dx + \int_0^1 x^3 e^{-x} dx.$$

Теперь распишем первый и третий интегралы (второй был посчитан ранее):

$$\int_0^1 x^2 dx = \frac{x^3}{3} \Big|_0^1 = \frac{1}{3}.$$

$$\int_0^1 x^3 e^{-x} dx = (-x^3 e^{-x}) \Big|_0^1 + 3 \int_0^1 x^2 e^{-x} dx = -\frac{1}{e} + 6 - \frac{15}{e} = 6 - \frac{16}{e}.$$

Тогда

$$\int_0^1 x^2(1 - e^{-x} + x e^{-x}) dx = \frac{1}{3} - 2 + \frac{5}{e} + 6 - \frac{16}{e} = \frac{13}{3} - \frac{11}{e}.$$

2. Возьмём его по частям:

$$\int_1^{+\infty} x^2 e^{-x} dx = (-x^2 e^{-x}) \Big|_1^{+\infty} + 2 \int_1^{+\infty} x e^{-x} dx = \frac{1}{e} + \frac{4}{e} = \frac{5}{e}.$$

В итоге получаем, что $E[Z^2] = \frac{13}{3} - \frac{6}{e}$.

Отсюда получаем, что

$$D[Z] = \frac{13}{3} - \frac{6}{e} - \left(\frac{3}{2} - \frac{1}{e}\right)^2 = \frac{13}{3} - \frac{6}{e} - \frac{9}{4} - \frac{1}{e^2} + \frac{3}{2e} = \frac{25}{12} - \frac{9}{2e} - \frac{1}{e^2}. \quad \square$$

Задача 3. Случайные величины X_1, X_2, \dots, X_n независимы и равномерно распределены на отрезке $[0, 1]$. Положим Y_k — k -е по порядку значение из набора X_1, X_2, \dots, X_n , $k \in \{1, 2, \dots, n\}$, (т.е. Y_1 — это минимальное значение, а Y_n — максимальное). Вычислите $E[Y_k]$, $k \in \{1, 2, \dots, n\}$.

Решение. По определению функции распределения $F_{Y_k}(x) = P(Y_k \leq x)$. Для начала сразу оговорим, что если $x < 0$, то $F_{Y_k}(x) = 0$, а если $x > 1$, то $F_{Y_k}(x) = 1$. В дальнейшем мы рассматриваем $x \in [0, 1]$.

В каком случае $Y_k \leq x$? В том случае, когда хотя бы k из n случайных величин X_1, X_2, \dots, X_n будут не больше x . Как это записать математически? Для этого введём случайные величины $\xi_k = I\{X_k \leq x\}$ для $k \in \{1, 2, \dots, n\}$. Тогда $P(Y_k \leq x) = P(\sum_{i=0}^n \xi_i \geq k)$.

Теперь вопрос: какое распределение у этой суммы? Для ответа на этот вопрос сперва нужно определить распределение у ξ_i . По определению индикатора ξ_i принимает только два значения — 0 и 1, при этом 1 достигается с вероятностью $P(X_i \leq x) = x$. Тогда $\xi_i \sim \text{Bern}(x)$. Далее заметим, что из независимости X_1, X_2, \dots, X_n следует независимость $\xi_1, \xi_2, \dots, \xi_n$. Тогда $\sum_{i=0}^n \xi_i \sim \text{Bin}(n, x)$.

Отсюда сразу же получаем значение функции распределения:

$$F_{Y_k}(x) = \sum_{i=k}^n \binom{n}{i} x^i (1-x)^{n-i} = \sum_{i=k}^{n-1} \binom{n}{i} x^i (1-x)^{n-i} + x^n.$$

Отсюда получаем плотность:

$$\rho_{Y_k}(x) = \sum_{i=k}^{n-1} \binom{n}{i} (ix^{i-1}(1-x)^{n-i} - (n-i)x^i(1-x)^{n-i-1}) + nx^{n-1}.$$

Заметим, что $i\binom{n}{i} = n\binom{n-1}{i-1}$, а $(n-i)\binom{n}{i} = n\binom{n-1}{i}$. Тогда плотность можно переписать в следующем виде:

$$\rho_{Y_k}(x) = n \sum_{i=k}^{n-1} \binom{n-1}{i-1} x^{i-1} (1-x)^{n-i} - n \sum_{i=k}^{n-1} \binom{n-1}{i} x^i (1-x)^{n-i-1} + nx^{n-1}.$$

Поменяем индекс суммирования в первой сумме на $j = i - 1$. Тогда плотность будет иметь вид:

$$\rho_{Y_k}(x) = n \sum_{j=k-1}^{n-2} \binom{n-1}{j} x^j (1-x)^{n-j-1} - n \sum_{i=k}^{n-1} \binom{n-1}{i} x^i (1-x)^{n-i-1} + nx^{n-1}.$$

Теперь заметим, что $nx^{n-1} = \binom{n-1}{n-1} x^{n-1} (1-x)^{n-1-(n-1)}$. Тогда

$$\rho_{Y_k}(x) = n \sum_{j=k-1}^{n-1} \binom{n-1}{j} x^j (1-x)^{n-j-1} - n \sum_{i=k}^{n-1} \binom{n-1}{i} x^i (1-x)^{n-i-1}.$$

В итоге сократятся все члены, кроме $(k-1)$ -го. Тогда

$$\rho_{Y_k}(x) = n \binom{n-1}{k-1} x^{k-1} (1-x)^{n-k}.$$

Теперь можно считать математическое ожидание:

$$E[Y_k] = \int_0^1 n \binom{n-1}{k-1} x^k (1-x)^{n-k} dx.$$

Внутри интеграла функция, очень близкая к плотности. На что нужно умножить интеграл, чтобы получить плотность?

$$\int_0^1 n \binom{n-1}{k-1} x^k (1-x)^{n-k} dx = c \int_0^1 (n+1) \binom{n}{k} x^k (1-x)^{n-k} dx = c.$$

Тогда

$$\mathbb{E}[Y_k] = c = \frac{n}{n+1} \frac{\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{k}{n+1}. \quad \square$$

Задача 4. *Случайные величины X_1, X_2, \dots, X_n независимы и равномерно распределены на отрезке $[0, a]$. Пусть $Y = \max(X_1, X_2, \dots, X_n)$. Вычислите $\mathbb{E}[(Y - a)^2]$.*

Решение. Воспользуемся результатом предыдущей задачи: плотность случайной величины $\frac{Y}{a}$ будет равна nx^{n-1} .

Преобразуем $\mathbb{E}[(Y - a)^2]$ по линейности:

$$\mathbb{E}[(Y - a)^2] = \mathbb{E}[y^2] - 2a \mathbb{E}[y] + a^2.$$

Осталось посчитать два интеграла:

$$\mathbb{E}[y^2] = a^2 \mathbb{E}\left[\left(\frac{Y}{a}\right)^2\right] = a^2 \int_0^1 nx^{n+1} dx = \frac{na^2}{n+2}.$$

$$\mathbb{E}[y] = a \mathbb{E}\left[\frac{Y}{a}\right] = a \int_0^1 nx^n dx = \frac{na}{n+1}.$$

Отсюда получаем, что

$$\begin{aligned} \mathbb{E}[(Y - a)^2] &= a^2 \left(\frac{n}{n+2} - \frac{2n}{n+1} + 1 \right) \\ &= a^2 \frac{n^2 + n - 2n^2 - 4n + n^2 + 3n + 1}{(n+1)(n+2)} = \frac{2a^2}{(n+1)(n+2)}. \end{aligned} \quad \square$$

Введём ещё одно определение, связанное с распределениями.

Определение 79. Пусть ξ и η — некоторые случайные величины на вероятностном пространстве $(\Omega, \mathcal{F}, \mathbb{P})$. Тогда *совместной функцией распределения* называется следующая функция:

$$F_{\xi, \eta}(x, y) = \mathbb{P}(\xi \leq x, \eta \leq y).$$

Теперь рассмотрим следующую задачу.

Задача 5. *Пусть есть две независимые случайные величины ξ и η с плотностями $p_\xi(x)$ и $p_\eta(x)$ соответственно. Найдите плотность случайной величины $f(\xi, \eta)$, если*

$$f(\xi, \eta) = \xi + \eta.$$

Ответ на эту задачу даёт *формула свёртки*:

Теорема 3 (Формула свёртки).

$$p_{\xi+\eta}(x) = \int_{-\infty}^{+\infty} p_{\xi}(y)p_{\eta}(x-y) dy.$$

Доказательство. ² Для доказательства этого факта нужно ввести понятие *плотности совместного распределения*:

Определение 80. Пусть есть две случайные величины ξ и η . Если существует неотрицательная функция $p_{\xi,\eta}(x_1, x_2)$ такая, что для любого $B \in \mathcal{B}(\mathbb{R}^2)$ выполнено

$$P((\xi, \eta) \in B) = \iint_B p_{\xi,\eta}(x_1, x_2) dx_1 dx_2,$$

то говорят, что случайные величины ξ и η имеют *абсолютно непрерывное совместное распределение*, а функцию $p_{\xi,\eta}(x_1, x_2)$ называют *плотностью совместного распределения*.

Теперь докажем следующее утверждение:

Теорема 4. Если случайные величины ξ и η имеют абсолютно непрерывную плотность совместного распределения $p_{\xi,\eta}(x_1, x_2)$ и независимы, то

$$p_{\xi,\eta}(x_1, x_2) = p_{\xi}(x_1)p_{\eta}(x_2).$$

Доказательство. По определению независимости:

$$F_{\xi,\eta}(x_1, x_2) = F_{\xi}(x_1)F_{\eta}(x_2).$$

Однако:

$$\begin{aligned} F_{\xi,\eta}(x_1, x_2) &= \iint_B p_{\xi,\eta}(x, y) dx dy = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} p_{\xi,\eta}(x, y) dx dy \\ F_{\xi}(x_1)F_{\eta}(x_2) &= \left(\int_{-\infty}^{x_1} p_{\xi}(x) dx \right) \left(\int_{-\infty}^{x_2} p_{\eta}(y) dy \right) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} p_{\xi}(x)p_{\eta}(y) dx dy. \end{aligned}$$

Дифференцируя, получаем желаемое. □

Теперь сформулируем один простой факт:

Теорема 5. Пусть $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ — борелевская функция, а область $D_x \subseteq \mathbb{R}^2$ состоит из точек (x_1, x_2) таких, что $g(x_1, x_2) \leq x$. Далее, пусть есть две случайные величины ξ_1, ξ_2 с абсолютно непрерывной плотностью совместного распределения $p_{\xi,\eta}(x_1, x_2)$. Тогда $\eta = g(\xi_1, \xi_2)$ — это случайная величина, и её функция распределения имеет следующий вид:

$$F_{\eta}(x) = P(g(\xi_1, \xi_2) \leq x) = P((\xi_1, \xi_2) \in D_x) = \iint_{D_x} p_{\xi,\eta}(x_1, x_2) dx_1 dx_2.$$

²На семинаре был дан лишь набросок доказательства. Здесь я попытался дать нечто, похожее на полноценное доказательство.

Для доказательства этого факта достаточно вспомнить определение плотности совместного распределения и то, что борелевская функция от случайной величины есть случайная величина.

Теперь приступим к доказательству. Безусловно, в данном случае нужно требовать абсолютно непрерывную плотность совместного распределения. Вспомним, что $f(x, y) = x + y$ есть борелевская функция. Тогда верна теорема выше:

$$F_{\xi+\eta}(x) = \iint_{x_1+x_2 \leq x} p_{\xi,\eta}(x_1, x_2) dx_1 dx_2 = \int_{-\infty}^x p_{\xi+\eta}(t) dt.$$

По теореме Фубини и независимости случайных величин ξ и η :

$$\iint_{x_1+x_2 \leq x} p_{\xi,\eta}(x_1, x_2) dx_1 dx_2 = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{x-x_1} p_{\xi}(x_1) p_{\eta}(x_2) dx_2 \right) dx_1$$

Теперь сделаем замену $x_2 = t - x_1$. Тогда $t \in (-\infty, x]$ и $dx_2 = dt$. Следовательно (так как мы можем просто поменять местами ξ и η):

$$\int_{-\infty}^{+\infty} \left(\int_{-\infty}^x p_{\xi}(x_1) p_{\eta}(t - x_1) dt \right) dx_1 = \int_{-\infty}^x \left(\int_{-\infty}^{+\infty} p_{\xi}(x_1) p_{\eta}(t - x_1) dx_1 \right) dt.$$

Отсюда получаем желаемое. □

Теперь разберём пару задач на формулу свёртки.

Задача 6. Пусть случайные величины ξ_1 и ξ_2 независимы и обе подчиняются стандартному нормальному распределению. Какова плотность случайной величины $\xi_1 + \xi_2$?

Решение. Воспользуемся формулой свёртки:

$$\begin{aligned} p_{\xi+\eta}(x) &= \int_{-\infty}^{+\infty} p_{\xi}(y) p_{\eta}(x - y) dy = \int_{-\infty}^{+\infty} \frac{1}{2\pi} \exp \left\{ -\frac{y^2}{2} - \frac{(x - y)^2}{2} \right\} dy \\ &= \int_{-\infty}^{+\infty} \frac{1}{2\pi} \exp \left\{ -\frac{x^2}{2} + xy - y^2 \right\} dy = \frac{e^{-\frac{x^2}{2}}}{2\pi} \int_{-\infty}^{+\infty} e^{xy - y^2} dy \\ &= \frac{e^{-\frac{x^2}{4}}}{2\pi} \int_{-\infty}^{+\infty} e^{xy - y^2 - \frac{x^2}{4}} dy = \frac{e^{-\frac{x^2}{4}}}{2\pi} \int_{-\infty}^{+\infty} e^{-(y - \frac{x}{2})^2} dy. \end{aligned}$$

Сделаем замену $t = y - \frac{x}{2}$:

$$\frac{e^{-\frac{x^2}{4}}}{2\pi} \int_{-\infty}^{+\infty} e^{-(y - \frac{x}{2})^2} dy = \frac{e^{-\frac{x^2}{4}}}{2\pi} \int_{-\infty}^{+\infty} e^{-t^2} dt$$

Теперь сделаем замену $t = \frac{u}{\sqrt{2}}$. Тогда $dt = \frac{du}{\sqrt{2}}$ и

$$\frac{e^{-\frac{x^2}{4}}}{2\pi} \int_{-\infty}^{+\infty} e^{-t^2} dt = \frac{e^{-\frac{x^2}{4}}}{2\pi\sqrt{2}} \int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}} du = \frac{\exp \left\{ -\frac{x^2}{4} \right\}}{\sqrt{2\pi} \cdot 2} = p_{\xi+\eta}(x).$$

Отсюда получаем, что $\xi + \eta \sim \mathcal{N}(0, 2)$. □

Упражнение. Пусть теперь $\xi \sim \mathcal{N}(0, a)$, $\eta \sim \mathcal{N}(0, b)$. Каковы будут плотность и распределение суммы?

Задача 7. Пусть случайные величины ξ_1 и ξ_2 независимы и обе подчиняются стандартному экспоненциальному распределению. Какова плотность случайной величины $\xi_1 + \xi_2$?

Решение. Перед тем, как начать пользоваться свёрткой, преобразуем плотности:

$$p_\xi(y) = e^{-y} I\{y \geq 0\},$$

$$p_\eta(x - y) = e^{-(x-y)} I\{x - y \geq 0\}.$$

Воспользуемся формулой свёртки:

$$\begin{aligned} p_{\xi+\eta}(x) &= \int_{-\infty}^{+\infty} p_\xi(y) p_\eta(x - y) dy = \int_{-\infty}^{+\infty} e^{-y} I\{y \geq 0\} e^{-(x-y)} I\{x - y \geq 0\} dy \\ &= e^{-x} \int_{-\infty}^{+\infty} I\{y \geq 0\} I\{x - y \geq 0\} dy = x e^{-x} I\{x \geq 0\}. \end{aligned}$$

В данном случае вышло так, что сумма случайных величин имеет гамма-распределение с параметрами 2 и 1. \square

Возникает разумный вопрос: а есть ли подобные формулы для произведения и для частного? Да, есть.

Лемма (Формула свёртки для произведения). Пусть ξ и η — независимые случайные величины. Тогда

$$p_{\xi\eta}(x) = \int_{-\infty}^{+\infty} \frac{1}{|y|} p_\xi(y) p_\eta\left(\frac{x}{y}\right) dy.$$

Доказательство. Рассуждение идейно почти не отличается от рассуждения для сложения. Опять же, воспользуемся теоремой:

$$\int_{-\infty}^x p_{\xi\eta}(t) dt = F_{\xi\eta}(x) = \iint_{x_1 x_2 \leq x} p_{\xi,\eta}(x_1, x_2) dx_1 dx_2 = \iint_{x_1 x_2 \leq x} p_\xi(x_1) p_\eta(x_2) dx_1 dx_2.$$

Сделаем замену переменных: $u = x_1 x_2$, $v = x_1$. Это равносильно тому, что $x_1 = v$, $x_2 = \frac{u}{v}$. Тогда якобиан будет иметь вид

$$\begin{vmatrix} \frac{\partial x_1}{\partial u} & \frac{\partial x_1}{\partial v} \\ \frac{\partial x_2}{\partial u} & \frac{\partial x_2}{\partial v} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ \frac{1}{v} & -\frac{uv'}{v^2} \end{vmatrix} = -\frac{1}{v}.$$

Следовательно,

$$\iint_{x_1 x_2 \leq x} p_{\xi,\eta}(x_1, x_2) dx_1 dx_2 = \int_{-\infty}^x \int_{-\infty}^{+\infty} \frac{1}{|v|} p_\xi(v) p_\eta\left(\frac{u}{v}\right) dv du.$$

\square

Лемма (Формула свёртки для частного). Пусть ξ и η — независимые случайные величины. Тогда

$$p_{\xi/\eta}(x) = \int_{-\infty}^{+\infty} p_{\xi}(xy)p_{\eta}(y)|y| dy.$$

Доказательство. Опять же, воспользуемся теоремой:

$$\int_{-\infty}^x p_{\xi/\eta}(t) dt = F_{\xi/\eta}(x) = \iint_{\substack{x_1 \leq x \\ x_2 \leq x}} p_{\xi,\eta}(x_1, x_2) dx_1 dx_2 = \iint_{\substack{x_1 \leq x \\ x_2 \leq x}} p_{\xi}(x_1)p_{\eta}(x_2) dx_1 dx_2.$$

Сделаем замену $u = \frac{x_1}{x_2}$, $v = x_2$. Отсюда получаем, что $x_1 = uv$, $x_2 = v$ и якобиан равен

$$\begin{vmatrix} \frac{\partial x_1}{\partial u} & \frac{\partial x_1}{\partial v} \\ \frac{\partial x_2}{\partial u} & \frac{\partial x_2}{\partial v} \end{vmatrix} = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v.$$

Тогда по формуле замены переменных и теореме Фубини:

$$\iint_{\substack{x_1 \leq x \\ x_2 \leq x}} p_{\xi}(x_1)p_{\eta}(x_2) dx_1 dx_2 = \int_{-\infty}^x \int_{-\infty}^{+\infty} |v| p_{\xi}(uv)p_{\eta}(v) dv du. \quad \square$$

2.9 Семинар от 29.11.2016

Разбор домашки ждать не будет.

Задача 1. Случайные величины ξ_1 , ξ_2 независимы и имеют равномерное распределение на отрезке $[0, 1]$. Найдите плотности случайных величин $\xi_1 + \xi_2$, $\xi_1 - \xi_2$ и $\xi_1\xi_2$.

Доказательство. Перед тем, как пользоваться формулами свёртки, найдём распределение $\eta = -\xi_2$:

$$F_{\eta}(x) = P(\eta \leq x) = P(\xi \geq -x) = 1 - P(\xi \leq -x) = 1 - F_{\xi}(-x).$$

$$p_{\eta} = p_{\xi}(-x) = I\{x \in [-1, 0]\}.$$

Напомним формулу свёртки:

$$p_{\xi+\eta} = \int_{-\infty}^{+\infty} p_{\xi}(x-y)p_{\eta}(y) dy.$$

Теперь можем считать:

1. Для начала рассмотрим сумму. Заметим, что $0 \leq \xi_1 + \xi_2 \leq 2$, поэтому при $x \notin [0, 2]$ $p_{\xi_1+\xi_2}(x) = 0$. Иначе же

$$\begin{aligned} p_{\xi_1+\xi_2}(x) &= \int_{-\infty}^{+\infty} I\{(x-y) \in [0, 1]\} I\{y \in [0, 1]\} dy \\ &= \int_{-\infty}^{+\infty} I\{y \in [x-1, x]\} I\{y \in [0, 1]\} dy. \end{aligned}$$

Осталось рассмотреть два случая. Если $x \in [0, 1]$, то $p_{\xi_1+\xi_2}(x) = x$. Если же $x \in [1, 2]$, то $p_{\xi_1+\xi_2}(x) = 1 - x + 1 = 2 - x$.

Запишем ответ:

$$p_{\xi_1+\xi_2}(x) = \begin{cases} x, & x \in [0, 1]; \\ 2 - x, & x \in [1, 2]; \\ 0, & \text{иначе.} \end{cases}$$

2.

□

Глава 3

Дополнительные главы теории вероятностей

3.1 Лекция от 24.01.2017

«На прошлой лекции Дмитрий Александрович запнулся, на этой посмотрел записи. Ну всё, действительно сложный ТеорВер начался.»

Один из слушателей

3.1.1 Случайные процессы

Определение 1. Пусть есть множество T (неформально оно обозначает время). Набор случайных величин $\{x_t, t \in T\}$ будем называть *случайным процессом*.

Примечание. То, что написано в определении, на самом деле, называется *случайной функцией*, но для определенности оставим определение в таком же виде, потому что в основном будем использовать $T \subseteq \mathbb{R}$, что уже действительно является случайным процессом по определению во многих учебниках.

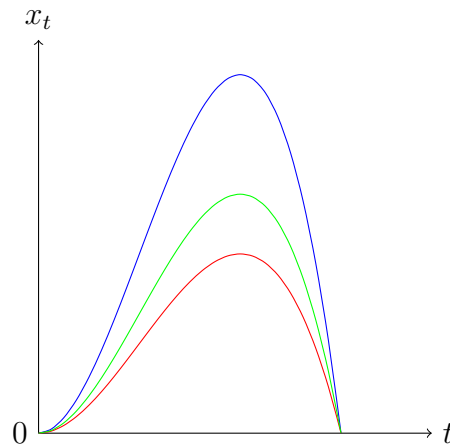
Определение 2. Классифицируем случайные процессы:

- если $T = \mathbb{N}, \mathbb{Z}, \mathbb{Z}_+$ — случайный процесс с дискретным временем.
- если $T = [a, b], \mathbb{R}, \mathbb{R}_+$ — случайный процесс с непрерывным временем.
- если $T \subseteq \mathbb{R}^d, d > 1$, тогда случайный процесс будет случайным полем.

Сейчас уделим больше внимания дискретным процессам. Будем считать, что у нас задана тройка Колмогорова $(\Omega, \mathcal{F}, \mathbb{P})$ и все $x_t : \Omega \rightarrow \mathbb{R}$.

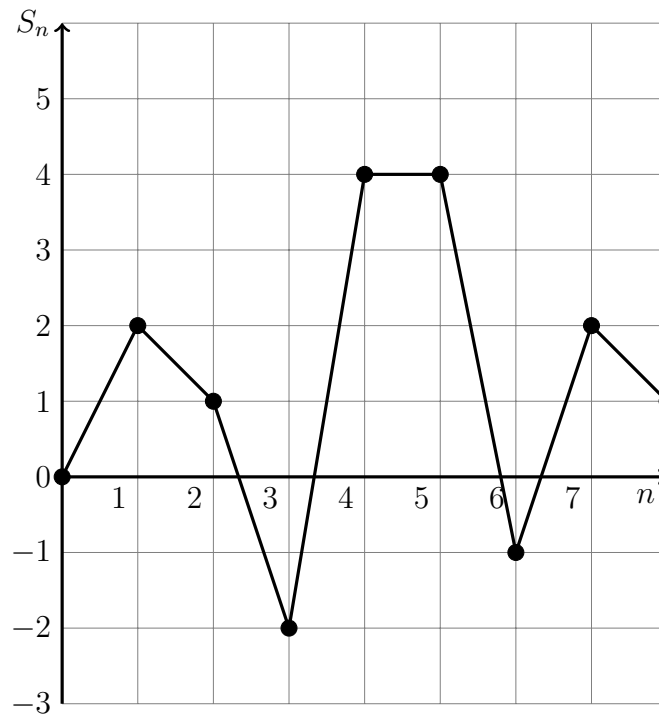
Определение 3. Для фиксированного $\omega_0 \in \Omega$ функция $\tilde{x}_t(t) = x_t(\omega_0), t \in T$ является *траекторией* (или *реализацией*) случайного процесса.

Пример. $x_t = f(t)\xi$, где $f(t)$ — какая-то функция, ξ — случайная величина. Приведем пример траекторий для некоторых ω_0 .



3.1.2 Случайные блуждания

Определение 4. Пусть $\{\xi_n, n \in \mathbb{N}\}$ — независимые случайные величины. Определим $S_0 = 0$, а $S_n = \xi_1 + \dots + \xi_n, n \in \mathbb{N}$. Тогда $\{S_n, n \in \mathbb{Z}_+\}$ называют *случайным блужданием*.



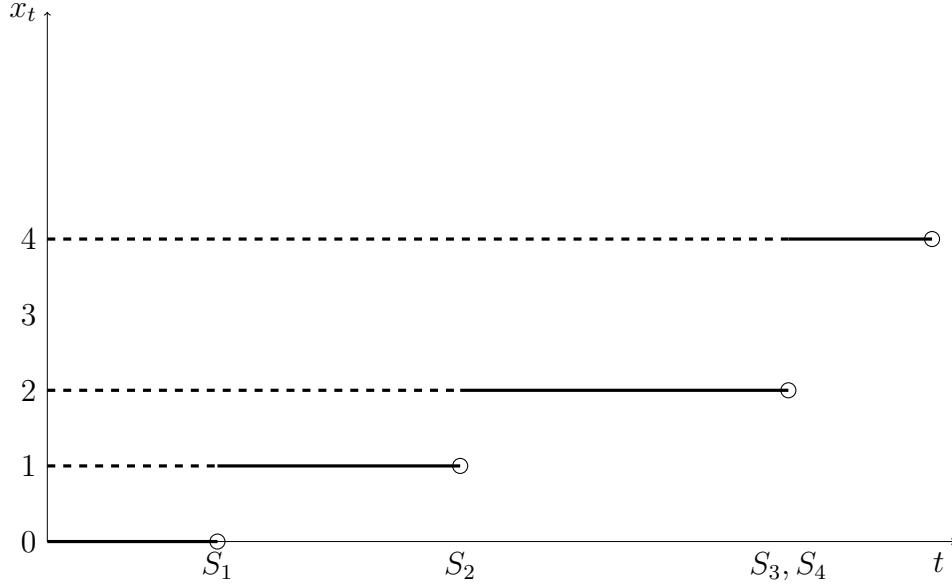
Пример. Физической моделью, соответствующую случайному блужданию, могут являться прыжки кузнечика.

3.1.3 Процесс восстановления

Определение 5. Пусть $\{\xi_n, n \in \mathbb{N}\}$ — одинаково распределенные неотрицательные случайные величины. Положим $S_0 = 0, S_n = \xi_1 + \dots + \xi_n, n \in \mathbb{N}$, а также для каждого $t \geq 0$ рассмотрим такие случайные величины: $x_t = \max\{n : S_n \leq t\}$ (если максимума не существует, положим $x_t = +\infty$).

Процесс $(x_t, t \geq 0)$ называется *процессом восстановления* для случайных величин $\{\xi_n, n \in \mathbb{N}\}$.

График можно описать как-то так для какого-то ω_0 :



Где возможны склеивания, как в S_3, S_4 — это лишь означает, что $\xi_4 = 0$.

Пример. Физической моделью, соответствующую процессу восстановления, может служить процесс горения лампочки, где ξ_n — общее время горения n -ой лампочки, а x_t тогда будет количеством поменянных лампочек до времени t .

Пара бы уже что-то доказать. Действительно, покажем, что мы не можем часто убегать в бесконечность. На самом деле, такие ситуации очень плохи в реальной жизни. Происходит «перенасыщение» чего-то. В примере с лампочкой, мы можем менять бесконечное число лампочек за время ноль. Такая ситуация очень плоха, поэтому, чтобы успокоиться, докажем следующее утверждение:

Теорема 1. *Процесс восстановления конечен с вероятностью один, если $P(\xi_i = 0) < 1$.*

Доказательство. Пусть $t > 0$ фиксировано. $P(x_t = +\infty) = P(\forall n : S_n \leq t)$.

Заметим, что $S_n \leq S_{n+1}$ из-за неотрицательности ξ_{n+1} .

Выпишем тривиальное равенство пределов:

$$\lim_{n \rightarrow +\infty} P(S_n \leq t) = \lim_{n \rightarrow +\infty} P\left(\frac{S_n}{n} \leq \frac{t}{n}\right)$$

Есть 2 случая:

- $E[\xi_i]$ конечно и равно $A > 0$ (больше нуля, так как $P(\xi_i = 0) < 1$). Тогда по закону больших чисел имеем, что $\frac{S_n}{n} \xrightarrow{P} A$ (на самом деле мы немного лукавим, потому что в основном курсе эта теорема была доказана с использованием, что все моменты до четвёртого конечны, но ЗБЧ работает и при конечности средней величины).

Тогда продолжим равенство пределов:

$$= \lim_{n \rightarrow +\infty} P\left(\frac{S_n}{n} \leq \frac{t}{n}\right) \leq \lim_{n \rightarrow +\infty} P\left(\frac{S_n}{n} \leq \frac{A}{2}\right)$$

Действительно, с какого-то момента $\frac{t}{n} < \frac{A}{2}$, так как $A > 0$. Далее, из закона больших чисел получаем:

$$= \lim_{n \rightarrow +\infty} P\left(A \leq \frac{A}{2}\right) = 0.$$

Последнее равенство идёт из-за неотрицательности A .

Поэтому $P(x_t = +\infty) = 0$ в этом случае.

- $E[\xi_i] = +\infty$.

Рассмотрим $\hat{\xi}_i = \min(\xi_i, 1) \leq \xi_i$. Откуда сразу получаем, что $\hat{S}_n = \hat{\xi}_1 + \dots + \hat{\xi}_n \leq S_n$.

Заметим, что $E[\hat{S}_n]$ конечно, так как матожидание каждого из $\hat{\xi}_i$ конечно (так как $\hat{\xi}_i \leq 1$).

А значит, что $P(S_n \leq t) \leq P(\hat{S}_n \leq t)$ (так как $S_n \geq \hat{S}_n$, а значит, что \hat{S}_n принимает меньшие значения).

Но мы уже всё доказали для конечного матожидания $\hat{\xi}_i$, поэтому получаем, что $P(\hat{S}_n \leq t) \rightarrow 0$, откуда и $P(S_n \leq t) \rightarrow 0$, что нам и требуется.

□

Приведём более мощный пример, обобщим эту модель.

Пример. Пусть $\{\xi_n, n \in \mathbb{N}\}$ — независимые и одинаково распределенные случайные величины, $\{\eta_n, n \in \mathbb{N}\}$ — независимые и одинаково распределенные случайные величины, притом независимы с $\{\xi_n, n \in \mathbb{N}\}$.

Пусть $(x_t, t > 0)$ — процесс восстановления, то есть $x_t = \max\{n : \xi_1 + \dots + \xi_n \leq t\}$.

Для $y_0, c > 0 \in \mathbb{R}$ введём

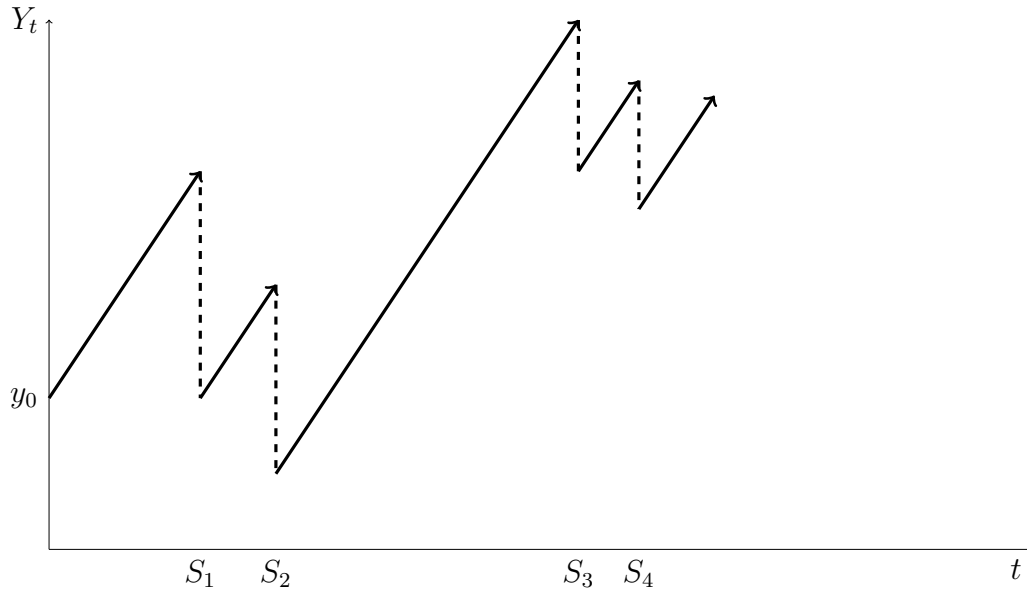
$$Y_t = y_0 + ct - \sum_{k=1}^{x_t} \eta_k$$

Эту модель называют моделью страхования Спарре-Андресена. Поясним, что значит каждая введенная переменная/величина.

- y_0 — начальный капитал.
- c — скорость поступления страховых взносов. Для простоты считают, поступление линейно, что примерно одинаково «бьются» машины в любое время года.
- η_k — размер выплаты с номером k .
- ξ_k — время между $k-1$ -й и k -й выплатой.
- x_t — количество выплат к моменту времени t .
- $\sum_{k=1}^{x_t} \eta_k$ — общий объём выплат по времени.
- И понятно, что тогда Y_t — текущий капитал.

В будущем, когда в нашем курсе мы затронем мартингалы, мы сможем понять и оценить, какова вероятность, что компания разорится.

Ясно, что тогда график капитала от времени будет выглядеть примерно таким образом:



3.1.4 Простые случайные блуждания

Определение 6. Пусть $\{\xi_n, n \in \mathbb{N}\}$ независимые одинаково распределенные случайные величины такие, что для какого-то $p \in [0, 1]$

$$P(\xi_n = 1) = p, P(\xi_n = -1) = 1 - p = q$$

Положим $S_0 = 0, S_n = \xi_1 + \dots + \xi_n, n \in \mathbb{N}$. Тогда $\{S_n, n \in \mathbb{Z}_+\}$ называют *простым случайным блужданием на прямой*.

Смысл этого определения в том, что на каждом шаге мы выбираем с какими-то вероятностями в какую сторону пойти.

Понятное дело, что этим можно не ограничиваться и, например, ходить в 4 разные стороны на плоскости. Но давайте пока разберёмся с одномерным случаем.

Определение 7. Если $p = q = \frac{1}{2}$, то говорят, что случайное блуждание симметрично.

Примечание. Не лишним будет упомянуть, что в данном случае $E[S_n] = (p - q)n$. Действительно, ответ можно легко получить, если раскрыть по линейности, и воспользоваться совершенно ясным фактом, который получается из определения, что $E[\xi_i] = p - q$.

Перед нами возникают достаточно интересные вопросы:

1. Какова вероятность вернуться в ноль после ненулевого количества шагов?
2. Какое среднее время мы проведем в нуле? То есть сколько в среднем раз мы окажемся в нуле при достаточно больших n .
3. Геометрия траекторий. То есть то, как выглядит график, какие существуют зависимости. Нашей кульминацией на этот вопрос будет закон повторного логарифма, который показывает, насколько далеко в среднем мы можем отходить от нуля.

Будем на все эти вопросы отвечать. Каждый из них — отдельная история, поэтому будем отвечать постепенно.

3.1.5 Возвращение в ноль в простом случайном блуждании

Легко понять, что нас интересует $P(\exists n > 0 : S_n = 0)$.

Попробуем найти $P(S_n = x)$. Во-первых, ясно, что n и x должны быть одной четности, так как иначе мы не сможем из нуля прийти в x . Также, нужно, чтобы $n \geq |x|$, иначе мы просто не дойдём до x , но скомпенсируем в будущем это тем, что $\binom{n}{k} = 0$ при $k > n$.

Пусть мы сделали k шагов вправо и $n - k$ шагов влево. Под шагами подразумевается $+1$ и -1 соответственно. Тогда ясно, что $k = \frac{n+x}{2}$, так как должно выполняться равенство $k - n + k = x$.

Поэтому получаем, что нам надо выбрать k ходов, ответ для данной вероятности:

$$P(S_n = x) = \binom{n}{k} p^k q^{n-k} I\{n + x \text{ чётно}\}$$

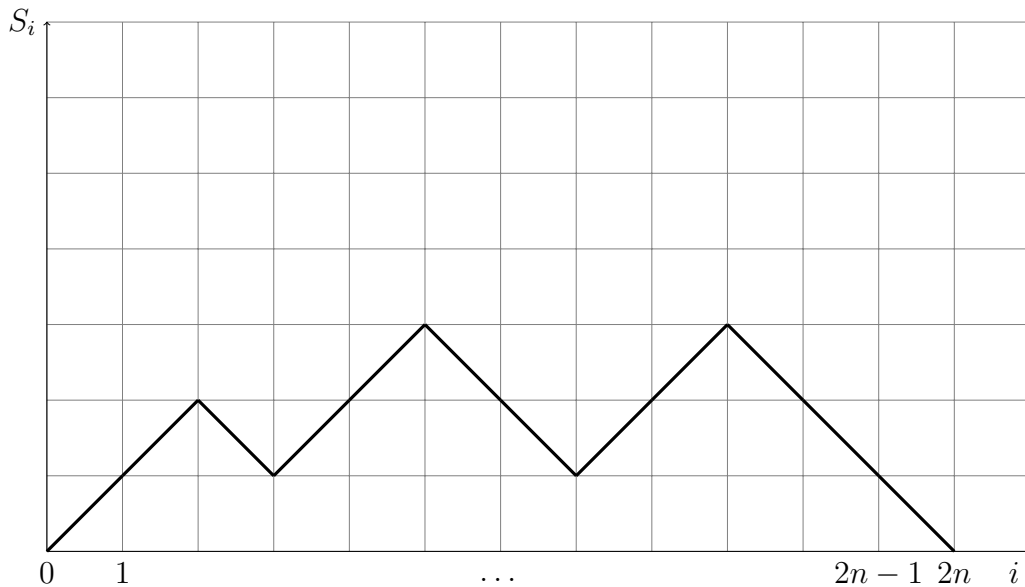
Но этого мало. Мы не сможем как-то легко посчитать вероятность существования n , что $S_n = 0$.

Поэтому давайте поймём, когда впервые достигнем нуля. Ясное дело, что нуля можно достигнуть только на чётных ходах. Поэтому давайте посчитаем такую вероятность:

$$P(S_1 \neq 0, \dots, S_{2n-1} \neq 0, S_{2n} = 0)$$

Заметим, что все S_1, \dots, S_{2n-1} либо одновременно больше нуля, либо все одновременно меньше нуля, так как из-за дискретной непрерывности, если есть S_i, S_j разных знаков, то найдётся между ними ноль, что противоречит тому, что мы ищем. Так как нам надо сделать n ходов вправо и влево, эти случаи ничем не отличаются. Посчитаем вероятность, когда все $S_i > 0, i \in [2n - 1]$.

Посмотрим на траектории, которые у нас могут получиться:



Теперь мы хотим посчитать все такие пути.

Обозначим через ε_i — выбор ± 1 на i -ом шаге. Тогда путь подходит тогда и только тогда, когда $\sum_{i=1}^{2n} \varepsilon_i = 0$ и $\sum_{i=1}^k \varepsilon_i > 0, k \in [2n - 1]$.

Обозначим количество таких путей через \tilde{C}_n .

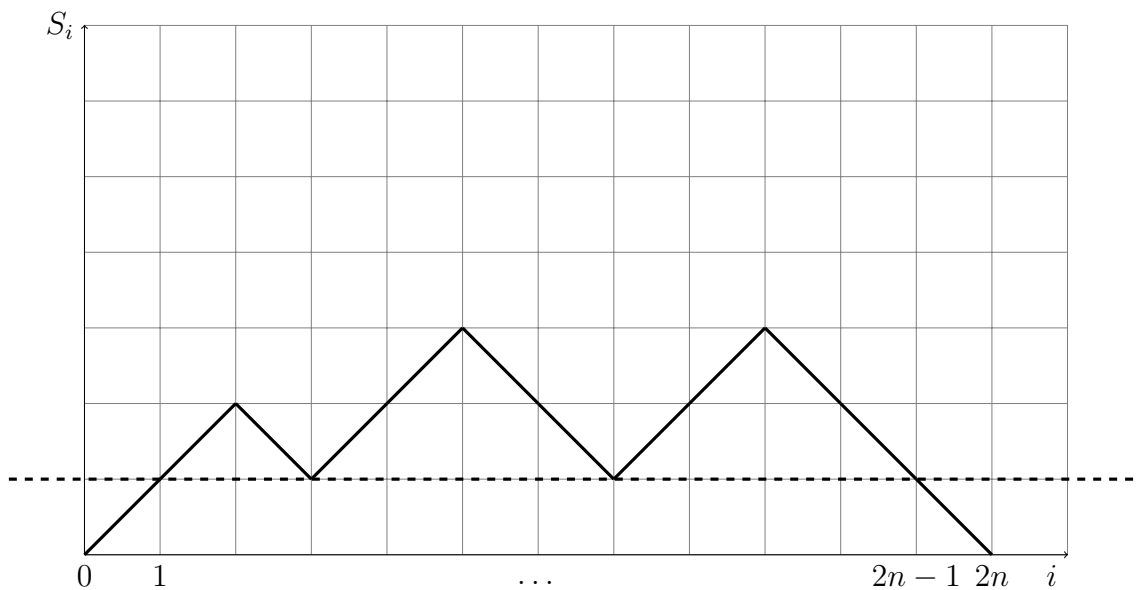
А теперь вспомним, что числа Каталана задаются практически так же. Те, кто ходил на курс дискретной математики, помнят, что есть соответствие между числами Каталана и количеством путей, отвечающим свойствам: $\sum_{i=1}^{2n} \varepsilon_i = 0$ и $\sum_{i=1}^k \varepsilon_i \geq 0, k \in [2n - 1]$.

Обозначим количество таких путей через C_n .

Лемма. $\tilde{C}_{n+1} = C_n$.

Доказательство. Рассмотрим любой путь, соответствующий \tilde{C}_{n+1} . Заметим, что первые и последние шаги обязательно $+1$ и -1 соответственно. Поэтому, при «поднятии» оси Oi на единицу мы получим, что перед нами путь из C_n , действительно, это так, так как префиксные суммы уменьшились на один и не стали отрицательными, а сумма по-прежнему сохранилась нулевой.

В другую сторону доказывается аналогично. См. иллюстрацию.



□

Получается, что

$$P(S_1 > 0, \dots, S_{2n-1} > 0, S_{2n} = 0) = C_{n-1}(pq)^n$$

И соответственно:

$$P(S_1 \neq 0, \dots, S_{2n-1} \neq 0, S_{2n} = 0) = 2C_{n-1}(pq)^n$$

Внимательный читатель заметит, что чтобы посчитать самую исходную вероятность, надо просто сложить все такие выше по всем $n \in \mathbb{N}$. О том, как такие вещи складывать (в частности о характеристических функциях), мы поговорим в следующей лекции.

3.2 Лекция от 31.01.2017

3.2.1 Числа Каталана. Реккурентное соотношение. Производящая функция

Лемма. Обозначим $C_0 = 1$, тогда имеет место следующее равенство:

$$C_n = \sum_{k=0}^{n-1} C_k C_{n-1-k}$$

Доказательство. Будем опять рассуждать в терминах положительной траектории (см. рис с предыдущей лекции).

Пусть $2k > 0$ первый момент, когда наша траектория придёт в ноль. Действительно, такой момент найдётся, потому что в момент времени $2n$ мы придём в ноль.

Тогда от 0 до $2k$ положительная траектория, от $2k$ до $2n$ неотрицательная, поэтому всего таких путей $\tilde{C}_k C_{n-k}$. Чтобы получить все траектории, надо сложить все такие вещи по всем $k = 1, \dots, n$, поэтому мы получим, что имеет место равенство по последней лемме из предыдущей лекции и тем, что $C_0 = 1$:

$$C_n = \sum_{k=1}^n \tilde{C}_k C_{n-k} = \sum_{k=1}^n C_{k-1} C_{n-k} = \text{«замена } j = k - 1\text{»} = \sum_{j=0}^{n-1} C_j C_{n-1-j}$$

□

Определение 8. Для последовательности $\{a_n\}_{n=0}^{+\infty}$ производящей функцией называется $f(x) = \sum_{n=0}^{+\infty} a_n x^n$.

Ряд может расходиться на некоторых x и тогда мы просто рассматриваем ряд формально, на них можно ввести операции сложения и прочие операции.

Заметим, что $C_n \leq 2^{2n}$, потому что всего путей не больше 2^{2n} , на самом деле их меньше аж примерно в $n^{3/2}$ раза, но это нам нужно для того, чтобы производящая функция чисел Каталана была такова, что при $|x| < \frac{1}{4}$ ряд сходиллся. Как мы потом увидим, он будет сходиться и при $|x| = \frac{1}{4}$.

Сейчас мы будем рассматривать только $f(x) = \sum_{n=0}^{+\infty} C_n x^n$ при $|x| < \frac{1}{4}$.

Теорема 2. $f(x) = \frac{1 - \sqrt{1 - 4x}}{2x}$.

Доказательство.

$$f^2(x) = \left(\sum_{n=0}^{+\infty} C_n x^n \right)^2 = \sum_{n=0}^{+\infty} \left(\sum_{k=0}^n C_k C_{n-k} \right) x^n.$$

Внимательный читатель заметит, что мы в скобках в точности получили рекуррентное соотношение для чисел Каталана для C_{n+1} . Поэтому это равно:

$$\sum_{n=0}^{+\infty} \left(\sum_{k=0}^n C_k C_{n-k} \right) x^n = \sum_{n=0}^{+\infty} C_{n+1} x^n = \frac{f(x) - 1}{x}.$$

Где последнее равенство возникает из-за того, что $C_0 = 1$.

Откуда мы получаем квадратное уравнение относительно $f(x)$.

$$x f^2(x) - f(x) + 1 = 0.$$

Решая уравнение, получим, что

$$f(x) = \frac{1 \pm \sqrt{1-4x}}{2x}$$

Но мы знаем, что $f(0) = 1$, поэтому с плюсом не подходит, так как предел будет не тот. Откуда единственный подходящий вариант будет

$$f(x) = \frac{1 - \sqrt{1-4x}}{2x}$$

□

Заметим, что и при $\frac{1}{4}$ мы получим конечное число, поэтому по непрерывности можно сказать, что и в $x = \frac{1}{4}$ ряд сходится.

3.2.2 Вероятность возвращения

Теорема 3. *Случайное блуждание возвратно с вероятностью $1 - |p - q|$.*

Доказательство. Заметим, что подставить pq в производящую функцию можно, так как $pq \leq \frac{1}{4}$.

$$\begin{aligned} P(\exists n : S_n = 0) &= \sum_{n=1}^{+\infty} P(S_1 \neq 0, \dots, S_{2n-1} \neq 0, S_{2n} = 0) = \\ &= \sum_{n=1}^{+\infty} 2C_{n-1}(pq)^n = 2pq \sum_{n=0}^{+\infty} C_n(pq)^n = 2pq \frac{1}{2pq} (1 - \sqrt{1-4pq}) = \text{«так как } 1 = (p+q)^2\text{»} = \\ &= 1 - \sqrt{(p-q)^2} = 1 - |p - q| \end{aligned}$$

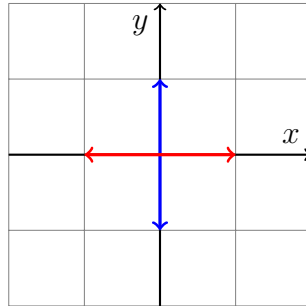
□

Примечание. Если $p = q = \frac{1}{2}$, то случайное блуждание возвратно с вероятностью один.

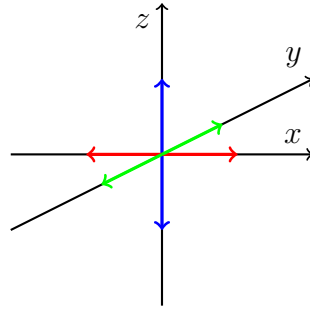
3.2.3 Многомерный случай

Абстрактно поговорим о многомерном случае. То есть мы находимся в $\mathbb{Z}^d, d > 1$.

На плоскости мы можем идти в 4 разные стороны с равными вероятностями. Как ни странно, вероятность возвращения в ноль в данном случае будет тоже равна один.



В трёхмерном случае не всё так однозначно. У нас есть шесть направлений. И в данном случае вероятность возвращения будет строго меньше единицы. С этим разобраться мы предложим читателю в будущих задачах.



На самом деле всё это зависит от $P(S_{2n} = 0)$. При $d = 1$ и $p = q = \frac{1}{2}$ получим, что $P(S_{2n} = 0) = \binom{2n}{n} 4^{-n} \approx \Theta\left(\frac{1}{\sqrt{n}}\right)$, ряд расходится.

При $d = 2$ будет $\Theta\left(\frac{1}{n}\right)$, что тоже расходится, с каждой новой размерностью добавляется этот самый \sqrt{n} в знаменателе, поэтому с $d = 3$ ряд будет сходящимся. Это была подсказка на будущие задачи, ничего тут мы пока не доказывали.

3.2.4 Числа Каталана через биномиальные коэффициенты

Вспомним некоторые факты про ряд Тейлора, а именно, что

$$(1+x)^\alpha = \sum_{n=0}^{+\infty} \binom{\alpha}{n} x^n,$$

где $\binom{\alpha}{n} = \frac{\alpha(\alpha-1)\dots(\alpha-n+1)}{n!}$.

Поэтому давайте преобразуем ряд для $f(x) = \sum_{n=0}^{+\infty} C_n x^n$.

$$\sqrt{1-4x} = \sum_{n=0}^{+\infty} \binom{1/2}{n} x^n = 1 + \sum_{n=1}^{+\infty} \frac{\frac{1}{2}(\frac{1}{2}-1)\dots(\frac{1}{2}-n+1)}{n!} (-4x)^n =$$

Вынесем $1/2$ из каждой дроби, везде поменяем знак, получим, что это домножится на $(-1)^{n-1}$, что совместно с $(-4x)^n$ даст знак минус, в скобках останется $(2n-3)!!$:

$$= 1 - \sum_{n=1}^{+\infty} \frac{4^n x^n (2n-3)!!}{2^n n!} = 1 - \sum_{n=1}^{+\infty} \frac{2^n x^n (2n-3)!!}{n!} =$$

Домножим на $n!$ и разделим на него, а также домножим на $2n-1$ и опять же разделим. Воспользуемся тем, что $(2n)! = (2n-1)!! n! \cdot 2^n$:

$$= 1 - \sum_{n=1}^{+\infty} \frac{(2n)! x^n}{n! n!} \frac{1}{2n-1} = 1 - \sum_{n=1}^{+\infty} \binom{2n}{n} \frac{x^n}{2n-1}$$

Откуда:

$$f(x) = \frac{1}{2x} (1 - \sqrt{1-4x}) = \frac{1}{2} \sum_{n=1}^{+\infty} \binom{2n}{n} \frac{1}{2n-1} x^{n-1} = \frac{1}{2} \sum_{n=0}^{+\infty} \binom{2n+2}{n+1} \frac{1}{2n+1} x^n = \sum_{n=0}^{+\infty} \binom{2n}{n} \frac{1}{n+1} x^n$$

В последнем равенстве можно убедиться непосредственно.

Откуда получаем $C_k = \frac{1}{k+1} \binom{2k}{k}$.

Примечание. $P(S_1 \neq 0, \dots, S_{2n-1} \neq 0, S_{2n} = 0) = 2C_{n-1}(pq)^n = \binom{2n}{n} (pq)^n \frac{1}{2n-1}$ — распределение первого момента в нуле.

3.2.5 Математическое ожидание первого момента возвращения в ноль

Определение 9. $X = \min(2n : S_{2n} = 0)$.

Теорема 4. $E[X] = +\infty$.

Доказательство. Если $p \neq q$, тогда $P(x = +\infty) > 0$, поэтому матожидание уже точно бесконечность.

Если $p = q = \frac{1}{2}$, тогда $E[X] = \sum_{n=1}^{+\infty} P(X = 2n) \cdot (2n) = \sum_{n=1}^{+\infty} \binom{2n}{n} \frac{2n}{2^{n-1} 4^n} = \Theta\left(\frac{1}{\sqrt{n}}\right)$, что расходится \square

Получается, что да, в ноль мы вернемся, но очень не скоро и бесконечно долго будем ходить вне. Вот такой вот парадокс. Перейдём к следующему пункту.

3.2.6 Среднее время в нуле

Оказывается, что в нуле мы будем находиться не так мало. За n шагов примерно \sqrt{n} раз мы будем в нуле. Это по большей мере связано с тем, что нули расположены рядом, то есть есть большая вероятность, что если мы пришли в ноль, то через малое количество шагов окажемся опять там.

Определение 10. Будем рассматривать симметричное простое случайное блуждание. Тогда $L_n = |k : k \in \overline{0, \dots, n}, S_k = 0|$.

Хотим понять, какая асимптотика у $E[L_n]$.

Лемма. $E[L_n] = E[|S_{n+1}|]$

Доказательство. Распишем $|S_{n+1}|$:

$$|S_{n+1}| = \begin{cases} 1, & \text{если } |S_n| = 0 \\ S_n + \xi_{n+1}, & \text{если } S_n > 0 \\ -S_n - \xi_{n+1}, & \text{если } S_n < 0 \end{cases}$$

Действительно, при положительном S_n мы не поменяем знак, поэтому надо лишь добавить то, что мы выбирали на следующем ходу, при отрицательном S_n аналогично, а если $S_n = 0$, то в любом случае модуль равен единице.

Запишем $|S_{n+1}|$ через индикаторы:

$$|S_{n+1}| = I\{|S_n| = 0\} + (S_n + \xi_{n+1})I\{S_n > 0\} - (S_n + \xi_{n+1})I\{S_n < 0\}$$

Заметим, что $S_n I\{S_n > 0\} - S_n I\{S_n < 0\} = |S_n|$ (в этом легко убедиться, проверив несколько случаев). Поэтому, введя функцию знака, можно утверждать, что:

$$|S_{n+1}| = I\{|S_n| = 0\} + |S_n| + \xi_{n+1} \text{sgn}(S_n)$$

Будем раскрывать рекурсивно, откуда получим:

$$|S_{n+1}| = \sum_{k=0}^n I\{|S_k| = 0\} + |S_0| + \sum_{k=0}^n \xi_{k+1} \text{sgn}(S_k)$$

Заметим, что $L_n = \sum_{k=0}^n I\{|S_k| = 0\}$, поэтому давайте запишем под знаком матожидания с использованием, что $E[|S_0|] = 0$.

$$E[|S_{n+1}|] = E[L_n] + \sum_{k=0}^n E[\xi_{k+1} \operatorname{sgn}(S_k)]$$

Но величины ξ_{k+1} и $\operatorname{sgn}(S_k)$ независимы, так как нет пересечений по ходам, поэтому это распадается в произведение матожиданий.

Ну и финальный аккорд состоит в том, что $E[\xi_{k+1}] = 0$, поэтому мы доказали равенство. \square

Про асимптотику поговорим в следующей лекции.

3.3 Лекция от 07.02.2017

3.3.1 Среднее время в нуле

Вспомним, что по ЦПТ и тем фактом, что среднее ξ_i равно нулю и дисперсия равна единице мы имеем следующее:

$$\frac{S_n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1)$$

По теореме о наследовании сходимости можно заключить, что

$$\frac{|S_n|}{\sqrt{n}} \xrightarrow{d} |\mathcal{N}(0, 1)|$$

А теперь мы хотим осознать, а как ведет себя среднее левой части. Заметим, что просто матожидание мы не можем взять, так как функция $f(x) = x$ не ограничена и эквивалентным определением сходимости по распределению напрямую воспользоваться нельзя. Поэтому нужны более сильные знания о поведении случайных величин.

Определение 11. Последовательность $\{\xi_n, n \in \mathbb{N}\}$ называется равномерно интегрируемой, если

$$\lim_{c \rightarrow +\infty} \sup_{n \in \mathbb{N}} E[|\xi_n| I\{|\xi_n| \geq c\}] = 0$$

Поясним определение.

Если у случайной величины есть плотность, то фактически мы говорим, что

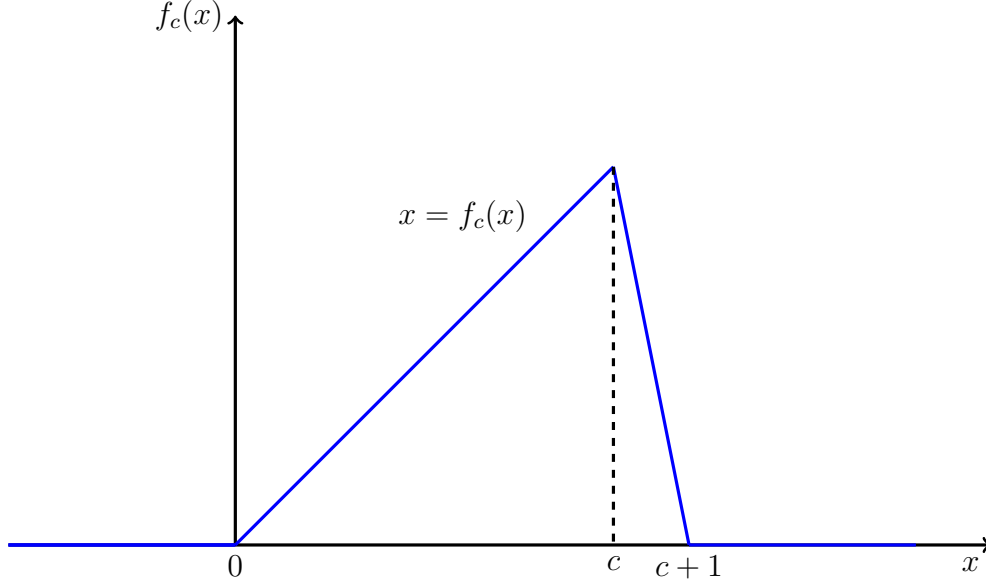
$$\sup_{n \in \mathbb{N}} \int_{|x| \geq c} |x| p_{\xi_n}(x) dx \rightarrow 0$$

То есть модуль среднего на бесконечности в обе стороны близок к нулю.

Теорема 5. Пусть $\xi_n \xrightarrow{d} \xi, \xi_n \geq 0$ — все величины с конечным матожиданием, тогда $E[\xi_n] \rightarrow E[\xi]$ тогда и только тогда, когда последовательность $\{\xi_n, n \in \mathbb{N}\}$ равномерно интегрируема.

Доказательство. Докажем в одну сторону, в другую оставим в качестве домашней задачи.

Пусть последовательность равномерно интегрируема. Тогда для $c > 0$ рассмотрим функцию $f_c(x)$:



То есть до 0 эта функция тождественно ноль, потом ведет себя так же, как и аргумент до c , потом резко убывает и снова становится нулем. Мы вводим такую функцию, чтобы приблизить $f(x) = x$, но ограниченным образом.

Заметим, что $f_c(x)$ непрерывна и ограничена. Оценим $|E[\xi_n] - E[\xi]|$. Добавим и вычтем каждое из двух выражений под модулем $E[f_c(\xi_n)]$, $E[f_c(\xi)]$ и воспользуемся неравенством треугольника:

$$|E[\xi_n] - E[\xi]| \leq |E[\xi_n] - E[f_c(\xi_n)]| + |E[\xi] - E[f_c(\xi)]| + |E[f_c(\xi_n)] - E[f_c(\xi)]|$$

Зафиксируем произвольное $\varepsilon > 0$. Первое слагаемое не больше $E[\xi_n I\{\xi_n \geq c\}]$, так как при $\xi_n \leq c$, величины просто совпадают, от c до $c+1$ будет какой-то остаток, не больший, чем $E[\xi_n I\{c \leq \xi_n \leq c+1\}]$, а дальше уже просто $E[f_c(\xi_n)]$ не даёт никакого вклада. Откуда $E[\xi_n I\{\xi_n \geq c\}] \leq \frac{\varepsilon}{2}$ при всех $c > c_0(\varepsilon)$. Эта оценка работает по условию теоремы при всех n .

Второе слагаемое для всех $c > c_1(\varepsilon)$ тоже не больше, чем $\frac{\varepsilon}{2}$, так как мы приближаем $f_c(x)$ положительную функцию (и при $c \rightarrow +\infty$ $f_c(x) \rightarrow x$) и матожидание конечно, то с какого-то момента разность станет очень маленькой.

Третье слагаемое стремится к нулю с ростом n из-за эквивалентного определения сходимости по распределению.

Поэтому получаем, что

$$\forall \varepsilon > 0 \quad \overline{\lim}_{n \rightarrow +\infty} |E[\xi_n] - E[\xi]| \leq \varepsilon$$

Откуда есть предел и равен он нулю, что и требовалось доказать. \square

Но теперь, чтобы наконец-то понять асимптотику среднего в нуле, давайте поймём, какие условия «полегче» нужно наложить, чтобы последовательность была равномерно интегрируемой. Это можно сделать из следующей леммы:

Лемма. Пусть $\{\xi_n, n \in \mathbb{N}\}$ — последовательность случайных величин. Если для всех $n \in \mathbb{N}$ выполнено, что $E[\xi_n^2] \leq C < +\infty$, тогда последовательность равномерно интегрируема.

Доказательство. Для любого $\varepsilon > 0$ выберем t_0 такое, что $t_0 > \frac{C}{\varepsilon}$. Тогда для любого $t > t_0$ и любого $n \in \mathbb{N}$

$$E[|\xi_n| I\{|\xi_n| \geq t\}] \leq E\left[\frac{\xi_n^2}{t} I\{|\xi_n| \geq t\}\right] \leq \frac{C}{t} \leq \frac{C}{t_0} < \varepsilon$$

Значит предел есть и равен нулю. □

Теперь сформулируем основную теорему:

Теорема 6.

$$E[L_n] \sim \sqrt{\frac{2n}{\pi}}$$

Доказательство. По последней лемме из прошлой лекции имеем $E[L_n] = E[|S_{n+1}|]$.

А также мы выяснили, что

$$\frac{|S_n|}{\sqrt{n}} \xrightarrow{d} |\mathcal{N}(0, 1)|$$

Посмотрим на второй момент левой части и вспомним, что среднее у S_n равно нулю:

$$E\left[\frac{S_n^2}{n}\right] = \frac{D[S_n]}{n} = 1$$

Последнее равенство следует из того, что дисперсия независимых величин равна сумме дисперсий.

Получаем, что второй момент всегда ограничен единицей, значит по теореме 5 мы получаем, что

$$E\left[\frac{|S_n|}{\sqrt{n}}\right] \rightarrow E[|\mathcal{N}(0, 1)|] = \sqrt{\frac{2}{\pi}}$$

Где последнее равенство легко проверяется интегрированием (см одно из ДЗ обычного курса).

Получаем, что $E[L_n] \sim \sqrt{\frac{2n}{\pi}}$ □

3.3.2 Геометрия траекторий. Закон повторного логарифма

Начнём сразу с теоремы, потом будем пояснять её смысл и постепенно доказывать.

Теорема 7 (Закон повторного логарифма). Пусть $\{S_n, n \in \mathbb{Z}_+\}$ — простое симметричное случайное блуждание.

Тогда

$$P\left(\overline{\lim}_{n \rightarrow +\infty} \frac{S_n}{\sqrt{2n \ln \ln n}} = 1\right) = 1$$

Выведем маленькое следствие из этого:

Примечание. Докажем, что из ЗПЛ следует следующий факт:

$$\mathbf{P}\left(\lim_{n \rightarrow +\infty} \frac{S_n}{\sqrt{2n \ln \ln n}} = -1\right) = 1$$

Ну это легко понять, если заменить $S_n = -X_n$ и надо лишь осознать, что $\{X_n, n \in \mathbb{Z}_+\}$ тоже случайное блуждание, но это совершенно ясно из определения.

Давайте будем пояснять смысл.

Закон повторного логарифма занимает промежуточное положение между законом больших чисел и центральной предельной теоремой. Мы знаем, что

$$\frac{S_n}{n} \xrightarrow{\mathbf{P}} 0, \frac{S_n}{\sqrt{n}} \xrightarrow{\mathbf{d}} \mathcal{N}(0, 1)$$

Центральная предельная теорема утверждает, что суммы S_n с делителем \sqrt{n} сходятся к стандартному нормальному распределению, и эта последовательность сумм не сходится к какой-либо конкретной величине ни по вероятности, ни почти наверное, а бесконечно блуждает.

Таким образом величина $S_n/\sqrt{2n \ln \ln n}$ будет к любой точке отрезка $[-1, 1]$ бесконечное число раз приближаться сколь угодно близко почти наверное.

Так же ЗПЛ означает, что с вероятностью один график блуждания лежит между $(1 + \varepsilon)\sqrt{2n \ln \ln n}$ и $-(1 + \varepsilon)\sqrt{2n \ln \ln n}$ для любого $\varepsilon > 0$ и бесконечное число раз выходит за пределы $(1 - \varepsilon)\sqrt{2n \ln \ln n}$ и $-(1 - \varepsilon)\sqrt{2n \ln \ln n}$.

Вспомним лемму Бореля-Кантеля, а сам ЗПЛ докажем на следующей лекции.

Определение 12. Пусть $\{A_n, n \in \mathbb{N}\}$ – события. Тогда событием $\{A_n \text{ беск. число}\}$ (или $\{A_n \text{ б. ч.}\}$) называют $\bigcap_{n=1}^{+\infty} \left(\bigcup_{m \geq n} A_m \right)$. Событие состоит в том, что произошло бесконечное число событий.

Теорема 8 (Лемма Бореля-Кантеля). Пусть $\{A_n, n \in \mathbb{N}\}$.

- 1) Если $\sum_{n=1}^{+\infty} \mathbf{P}(A_n)$ сходится, тогда $\mathbf{P}(\{A_n \text{ б. ч.}\}) = 0$.
- 2) Если $\sum_{n=1}^{+\infty} \mathbf{P}(A_n)$ расходится и A_n независимые величины, тогда $\mathbf{P}(\{A_n \text{ б. ч.}\}) = 1$.

Доказательство. 1)

$$\mathbf{P}(\{A_n \text{ б. ч.}\}) = \mathbf{P}\left(\bigcap_{n=1}^{+\infty} \left(\bigcup_{m \geq n} A_m \right)\right) = \lim_{n \rightarrow +\infty} \mathbf{P}\left(\bigcup_{m \geq n} A_m\right) \leq \lim_{n \rightarrow +\infty} \sum_{m \geq n} \mathbf{P}(A_m) = 0$$

Последнее равенство верно, так как остаток сходящегося ряда стремится к нулю.

2) Тут уже напрямую не получится, надо провести более тонкий анализ:

$$\mathbf{P}(\{A_n \text{ б. ч.}\}) = \lim_{n \rightarrow +\infty} \mathbf{P}\left(\bigcup_{m \geq n} A_m\right) = 1 - \lim_{n \rightarrow +\infty} \mathbf{P}\left(\bigcap_{m \geq n} \overline{A_m}\right) =$$

Хочется воспользоваться независимостью, но в данном случае надо применять один трюк, чтобы это стало возможным. Поставим повторный предел:

$$= 1 - \lim_{n \rightarrow +\infty} \lim_{N \rightarrow +\infty} \mathbf{P} \left(\bigcap_{m=n}^N \overline{A_m} \right) =$$

Вот теперь уже можно пользоваться независимостью событий (а значит и их дополнений).

$$1 - \lim_{n \rightarrow +\infty} \lim_{N \rightarrow +\infty} \prod_{m=n}^N (1 - \mathbf{P}(A_m))$$

Применим неравенство, что $1 - x \leq e^{-x}$ и поэтому можно написать, что

$$1 - \lim_{n \rightarrow +\infty} \lim_{N \rightarrow +\infty} \prod_{m=n}^N (1 - \mathbf{P}(A_m)) \geq 1 - \lim_{n \rightarrow +\infty} \lim_{N \rightarrow +\infty} e^{-\sum_{m=n}^N \mathbf{P}(A_m)}$$

Но остаток расходящегося ряда стремится к $+\infty$, поэтому имеем равенство

$$= 1 - \lim_{n \rightarrow +\infty} e^{-\infty} = 1$$

□