1 Вероятностное пространство

Начнем с такого вопроса, чтобы проверить вашу вероятностную интуицию:

Вопрос 1.1. Какое из событий более вероятно при бросании симметричной монеты:
(a) выпадет орел, решка, орел, решка, орел, решка;
(b) выпадет орел, орел, орел, решка, решка, решка;
(c) выпадет решка, решка, решка;
(d) (a) и (b) вместе.

1.1 Введение

Если спросить у вас что такое вероятность какого-либо события, то, скорее всего, вы дадите эмпирическое определение — вероятность события есть доля случаев, в которых оно произойдет, если много раз его повторять.

Это определение полностью соответствует нашим интуитивным представлениям о вероятности. Однако, с точки зрения науки оно не слишком удачно. Вот лишь два из множества вопросов, на которые достаточно затруднительно ответить с точки зрения эмпирического определения.

1. Что такое повторение?
   Допустим, я вытащил карту наугад из колоды карт. Теперь я планирую перевернуть верхнюю карту и хочу узнать какая вероятность того, что это дама пик.
   С точки зрения эмпирического определения я должен повторить эксперимент много раз и посмотреть на долю успехов. Но ведь верхняя карта уже есть какая-то, если я буду повторять эксперимент (то есть брать верхнюю карту, смотреть на нее и класть ее обратно), то я всегда буду получать один и тот же результат.
   Нет, скажете вы, повторять эксперимент нужно более полно — каждый раз заново раскладывать колоду в случайном порядке и брать верхнюю карту.
   Однако, попробуйте объяснить в сколько-то общей постановке задачи насколько сильно мне нужно возвращаться в прошлое, чтобы повторить опыт. Нужно ли мне каждый раз покупать новую колоду? Или может быть я должен заново начать эволюцию жизни на Земле, дождаться появления человека, создания печатной промышленности и так далее?
   Мы интуитивно понимаем о каком именно повторении идет речь (хотя и это не всегда), но включить это в определение достаточно затруднительно.
2. Как быть с событиями, которые нельзя повторить?
   Что такое вероятность победы "Барселоны" в завтрашнем противостоянии с "Реал Мадрид"? Мы не сможем повторить этот эксперимент, завтрашний матч случится только один раз. А любой последующий матч уже не будет повторением завтрашнего, это будет другой матч с другими начальными условиями.

В этом случае математикам остается одно — отказаться от эмпирического определения вероятности, а считать, что вероятность — это что-то данное нам изначально. В этой модели, в которой вероятности нам даны, мы можем получить какие-то законы и правила. Одним из таким законов будет, как мы выясним позже, закон больших чисел (теорема 15.2), который и покажет, что вероятность должна соответствовать эмпирическому определению.

1.2 Вероятностное пространство

1.2.1 Пространство элементарных исходов

Итак, пусть у нас есть некоторый опыт, у которого есть несколько (для простоты конечное число N) возможных исходов. Множество возможных исходов мы будем обозначать $\Omega = \{\omega_1, \ldots, \omega_N\}$ и они будут как-то соответствовать итогам исходного физического опыта.

Определение 1.1. Такое пространство называется пространством элементарных исходов, а его элементы — элементарными исходами.

С математической точки зрения таким пространством может быть произвольное множество (пока мы рассматриваем только конечные Ω, но позже введем в рассмотрение множества произвольной структуры).

Пример 1.1. Для броска монеты мы можем выбрать $\Omega = \{O, P\}$, где O соответствует орлу, а P — решке или, скажем, $\Omega = \{0, 1\}$, где 1 соответствует орлу, а 0 — решке.

Для двух бросков монеты — $\Omega = \{0, 1, 2, 3\}$, где $0, 1, 2, 3$ как-то соответствуют четырем возможным элементарным исходам или, скажем, $\Omega = \{(0,0),(0,1),(1,0),(1,1)\}$, где первая цифра соответствует первому броску, а вторая — второму.

Описать тот же эксперимент можно и с помощью $\Omega = \{\{0,0\},\{0,1\},\{1,1\}\}$, если человек, наблюдающий за опытом, не знает какая из подоброшенных монет первая, а видит лишь результат броска. В этом случае у него будет один исход $\{0,1\}$, описывающий итог эксперимента "один раз выпал орел, а один раз решка".

Может быть и гораздо более экзотическое описание экспримента, например, мы можем наблюдать за взаимным положением молекул монеты в пространстве в наших двух случаях. Тогда у нас будет четыре принципиально разных возможных расположения, соответствующих тому как выпадали монеты. Это пространство покажется нам неестественным, но тоже имеет право на существование.

Как мы видим — в выборе пространства элементарных исходов для описания данного эксперимента есть изрядный произвол. Главное — включить в него все ситуации, которые меня интересуют. Так, если я хочу заниматься исследованием реальной монеты с высокой точностью, то неплохо бы добавить к возможным исходам броска случай "монета встала на ребро" (соответственно, для двух бросков возможных исходов вместо 4 станет 9), а то и добавить исход "монета потерялась и эксперимент прекратился".

Вопрос 1.2. Сколько в этом случае станет возможных элементарных исходов у эксперимента?

1.2.2 Алгебра событий

Чаще всего нас интересуют не отдельные исходы, а целые группы исходов. Иначе говоря, мы рассматриваем подмножества Ω. В нашем простом случае можно рассматривать множество всех подмножеств 2^Ω (почему это плохо в более общей постановке мы поговорим в разделе 8). Итак, подмножества Ω будем называть событиями, а их множество будем обозначать $\mathcal{F}$.

Пример 1.2. Для одного подбрасывания монеты с $\Omega = \{0, 1\}$ мы получим

$$\mathcal{F} = \{\emptyset, \Omega, \{0\}, \{1\}\}.$$

Событие $\emptyset$ называется невозможным, а Ω — достоверным. Можно интерпретировать наши 4 события как "монета выпала ни на что" (такого не бывает), "монета на что-то выпала" (такое происходит всегда), "монета выпала на орла", "монета выпала на решку".

Для двух подбрасываний монеты будет уже 16 событий

$$\mathcal{F} = \{\emptyset, \Omega, \{(0,0)\}, \{(1,0)\}, \{(0,1)\}, \{(1,1)\}, \{(0,0),(0,1)\}, \{(0,0),(1,0)\}, \{(0,0),(1,1)\}, \{(0,1),(1,0)\}, \{(0,1),(1,1)\},$$
$$\{(1,0),(1,1)\}, \{(0,0),(1,0),(0,1)\}, \{(0,0),(0,1),(1,1)\}, \{(0,0),(1,0),(1,1)\}, \{(0,1),(1,0),(1,1)\}\}.$$

Скажем, событие $\{(0,0),(0,1)\}$ можно назвать "первый бросок был решкой", а событие $\{(0,0),(1,0),(0,1)\}$ — "выпала хотя бы одна решка".

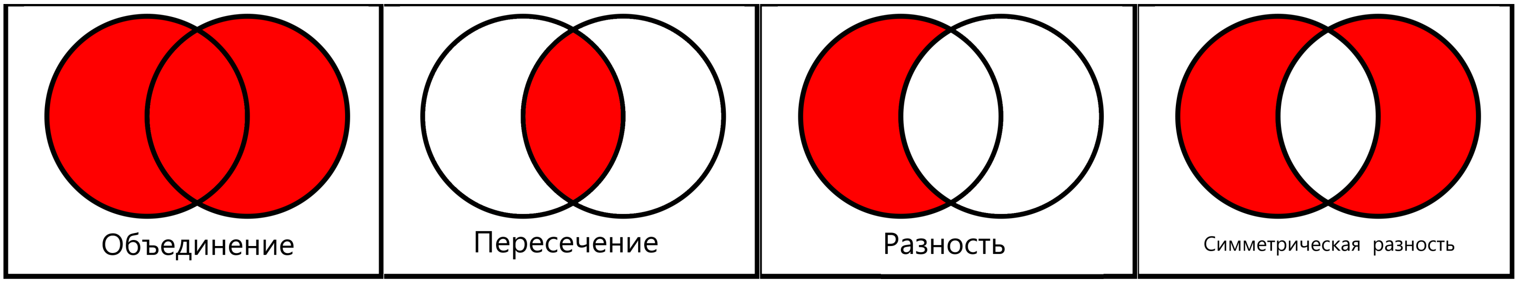
Для общего случая $\mathcal{F}$ будет состоять из 2^N событий.

Над событиями можно производить те же операции, что и над любыми другими множествами:

- дополнение $\overline{A}$ (событие не выполнено);
- объединение $A \cup B$ (выполнено хотя бы одно из событий);
- пересечение $A \cap B$ (выполнены оба события);
- разность $A \setminus B$ (выполнено первое событие, но не второе);
- симметрическая разность ("выполнено ровно одно из событий")

$$A \Delta B = (A \setminus B) \bigcup (B \setminus A).$$

Рис. 1: Операции над множествами



Мы будем для краткости использовать обозначение AB для пересечения событий $A \cap B$.

Нам также будет полезно обозначение "сумма событий". Мы говорим про сумму событий $A+B$, подразумевая под этим $A \bigcup B$, причем с дополнительным предположением, что $AB = \emptyset$.

Вообще говоря, мы можем рассматривать в качестве множества событий $\mathcal{F}$ не все возможные подмножества Ω, а только их часть, но при этом естественно требовать в случае конечного пространства следующие условия

- $\Omega \in \mathcal{F}$;
- Если $A \in \mathcal{F}$, то и $\overline{A} \in \mathcal{F}$;
- Если $A, B \in \mathcal{F}$, то и $A \cup B \in \mathcal{F}$.

Такое множество называется алгеброй подмножеств Ω. В силу соотношений

$$A \cap B = \overline{\overline{A} \cup \overline{B}}, \quad A \setminus B = A \cap \overline{B}, \quad A \Delta B = (A \cap \overline{B}) \cup (B \cap \overline{A})$$

алгебра множеств содержит вместе с каждыми двумя множествами их пересечение, разность и симметрическую разность.

Пример 1.3. Множество $\{\emptyset, \Omega\}$ — алгебра подмножеств Ω. Это самая простая (так называемая тривиальная) алгебра событий.

Множество $\{\emptyset, A, \overline{A}, \Omega\}$ также будет алгеброй. Эта алгебра называется алгеброй, порожденной A.

1.2.3 Вероятность

Каждому событию A сопоставим вероятность $\mathbf{P}(A)$.

Определение 1.2. Вероятностной мерой (или просто вероятностью) на конечном пространстве Ω с алгеброй событий $\mathcal{F}$ называется отображение $\mathbf{P} : \mathcal{F} \to [0, 1]$, удовлетворяющее свойству

1. $\mathbf{P}(\Omega) = 1$.
2. $\mathbf{P}(A + B) = \mathbf{P}(A) + \mathbf{P}(B)$.

Второе свойство называется аддитивностью. Важно заметить, что в более общем случае бесконечных пространств элементарных исходов нам понадобится еще одно (очень важное) свойство вероятности, но пока мы рассматриваем только конечные пространства, нам хватит и этих двух.

Из определения вероятности можно вывести следующие свойства

- $\mathbf{P}(\emptyset) = 0$.
- $\mathbf{P}(\overline{A}) = 1 - \mathbf{P}(A)$.
- $\mathbf{P}(A_1 + \dots + A_n) = \mathbf{P}(A_1) + \dots + \mathbf{P}(A_n)$.
- $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(AB)$.

- $\mathbf{P}(A \setminus B) = \mathbf{P}(A) - \mathbf{P}(AB)$.
- $\mathbf{P}(A \Delta B) = \mathbf{P}(A) + \mathbf{P}(B) - 2\mathbf{P}(AB)$.

Попробуйте доказать их самостоятельно.

Вопрос 1.3. А верно ли в общем случае соотношение $\mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B)$?

Приведенная формула для $\mathbf{P}(A \cup B)$ допускает следующее обобщение:

$$\mathbf{P}\left(\bigcup_{i\le n} A_i\right) = \sum_{i\le n}\mathbf{P}(A_i) - \sum_{i_1<i_2}\mathbf{P}(A_{i_1}A_{i_2}) + \dots - (-1)^n\mathbf{P}(A_1\cdots A_n).$$

Эта формула называется формулой включений-исключений. Докажем её.

Доказательство. Будем действовать по индукции. Для $n = 2$ она представляет указанное ранее свойство.

Пусть для $n = k$ формула доказана. Докажем ее для $n = k + 1$. Пользуясь ей при $n = 2$, получаем

$$\mathbf{P}\left(\bigcup_{i\le k+1} A_i\right) = \mathbf{P}(A_1 \cup A_2 \dots \cup A_k) + \mathbf{P}(A_{k+1}) - \mathbf{P}((A_1A_{k+1}) \cup \dots \cup (A_kA_{k+1})).$$

Применяя предположение индукции, перепишем это выражение в виде

$$\sum_{i\le k}\mathbf{P}(A_i) - \sum_{i_1<i_2\le k}\mathbf{P}(A_{i_1}A_{i_2}) + \dots - (-1)^k\mathbf{P}(A_1\cdots A_k) + \mathbf{P}(A_{k+1}) - \sum_{i\le k}\mathbf{P}(A_iA_{k+1}) +$$
$$\sum_{i_1<i_2\le k}\mathbf{P}(A_{i_1}A_{i_2}A_{k+1}) - \dots - (-1)^{k+1}\mathbf{P}(A_1\cdots A_{k+1}).$$

Перегруппировывая слагаемые, получаем требуемое. □

Поскольку среди событий есть события $\{\omega_1\}$, ..., $\{\omega_N\}$, то мы можем рассмотреть их вероятности $p_1, \dots, p_N$. Для любого события A, тем самым, мы получаем

$$\mathbf{P}(A) = \sum_{i:\omega_i\in A} p_i = \sum_{i:\omega_i\in A}\mathbf{P}(\{\omega_i\}).$$

Поэтому нам достаточно задать вероятности элементарных исходов p_i, вероятности всех остальных событий будут определяться этими числами. Вообще говоря, на роль p_i подойдут любые неотрицательные числа, в сумме дающие 1.

Определение 1.3. Итак, *вероятностным пространством* называется тройка $(\Omega, \mathcal{F}, \mathbf{P})$ из пространства элементарных исходов, алгебры событий и вероятностной меры.

В нашем простом случае мы, чаще всего, будем задавать Ω, в качестве множества событий брать множество всех подмножеств Ω, а вероятность на них будем задавать исходя из вероятностей элементарных исходов $p_1, \dots, p_N$.

1.2.4 Как выбрать вероятностное пространство

Возникает важный вопрос — как сопоставить эксперименту вероятностное пространство?

С Ω особых проблем не возникает — мы просто должны сопоставить эксперименту подходящее описание в каком-то виде, закодировать итоги эксперимента. Хотя, как мы увидим позже, удачный выбор Ω иногда фактически решает задачу.

С $\mathcal{F}$ в текущей постановке все совсем просто — обычно это будет просто множество всех подмножеств Ω.

А вот с $\mathbf{P}$ возникают проблемы. Мы можем сопоставить элементарным исходам любые вероятности, лишь бы их сумма была равна 1. Но, вообще говоря, нам бы хотелось, чтобы теория соответствовала эксперименту. Пока мы плохо понимаем как именно, но каждому с детства понятен анекдот про блондинку, которая сопоставляет исходам "встретить динозавра на улице" и "не встретить динозавра на улице" равные вероятности. Это сопоставление вероятностей элементарным исходам не соответствует нашему представлению о реальных частотах этих событий. Пока мы постараемся обходить эту проблему и считать, что вероятности даны нам свыше. Впрочем, сегодня мы рассмотрим удачный частный случай, в котором понятно как составить вероятности элементарным исходам.

1.3 Классическое вероятностное пространство

1.3.1 Общая постановка

Если мы, например, бросаем игральный кубик, то естественно приписать этому эксперименту $\Omega = \{1, 2, 3, 4, 5, 6\}$ и всех исходам приписать равные вероятности $1/6$, потому что все числа на кубике равноправны.

Определение 1.4. Говорят, что вероятности на пространстве элементарных исходов Ω заданы классически, если $p_i = \mathbf{P}(\omega_i) = 1/N$.

Вспоминая все тот же анекдот, мы видим, что разумно использовать классическое задание вероятностей только если у исходов есть симметрия. У наличия и отсутствия динозавра на улице такой симметрии нет.

Пример 1.4. При двух бросания монеты мы можем рассмотреть $\Omega_1 = \{(0,0), (1,0), (0,1), (1,1)\}$ с $\mathcal{F}_1 = 2^{\Omega_1}$ и $\mathbf{P}_1(\omega_i) = 1/4$.

А можем рассмотреть $\Omega_2 = \{\{0,0\}, \{1,0\}, \{1,1\}\}$ с $\mathcal{F}_2$ и $\mathbf{P}_2(\omega_i) = 1/3$.

При этом пространства находятся в конфликте друг с другом — одно и то же реальное событие "два орла" имеет в одном из них вероятность $1/4$, а во втором $1/3$.

В первом симметрия исходов совершенно разумна — при каждом из бросков орел и решка для нас равнозначны (если, конечно, монетку мы считаем честной). Во втором случае исход "один раз выпала решка, а один орел" совершенно не симметричен исходу "дважды выпал орел". Поэтому в первом случае классическое задание вероятностей соответствует постановке задачи, а во втором нет.

1.3.2 Выбор шаров из урны

Рассмотрим урну с шарами, пронумерованными от 1 до n. Предположим, что мы вытягиваем k шаров и записываем их номера.

1. Упорядоченный выбор с возвращением. Пусть номера записываются в порядке появления, а шары после каждого вытягивания возвращаются в урну. Тогда естественно использовать

   $$\Omega = \{(i_1, \ldots, i_k),\ i_1, \ldots, i_k \in \{1, \ldots, n\}\}, \quad \mathcal{F} = 2^{\Omega}.$$

   При этом классически заданная вероятность соответствует нашим представлениям о задаче — все раскладки шаров равновероятны просто потому, что каждый раз нам совершенно все равно что вытаскивать — 1, 2, ... или k.

   Все возможные исходы представляют собой таблицу $\{1, \ldots, n\} \times \cdots \times \{1, \ldots, n\}$, в которой n^k исходов. Нам более удобно представить их в виде дерева, что легко позволяет убедиться в том, что $|\Omega| = n^k$, где $|\cdot|$ — мощность множества. Действительно, каждая вершина на нижнем слое дерева соответствует одному набору $(i_1, \ldots, i_k)$. Число вершин на нижнем слое при этом в n раз больше, чем на предыдущем, в том — в n раз больше, чем на предыдущем и так далее (см. Рис 2).

2. Упорядоченный выбор без возвращения. Пусть номера записываются в порядке появления, а шары после каждого вытягивания откладываются. Тогда естественно использовать

   $$\Omega = \{(i_1, \ldots, i_k),\ i_1, \ldots, i_k \in \{1, \ldots, n\},\ i_j \neq i_l, j \neq l\}, \quad \mathcal{F} = 2^{\Omega}.$$

   При этом классически заданная вероятность соответствует нашим представлениям о задаче — все раскладки шаров равновероятны просто потому, что каждый раз нам совершенно все равно какой из оставшихся шаров вытягивать.

   Опять же удобно представить исходы в виде дерева, что легко позволяет убедиться в том, что $|\Omega| = n(n-1)\cdots(n-k+1) = A_n^k$.

Рис. 2: Упорядоченный выбор с возвращением

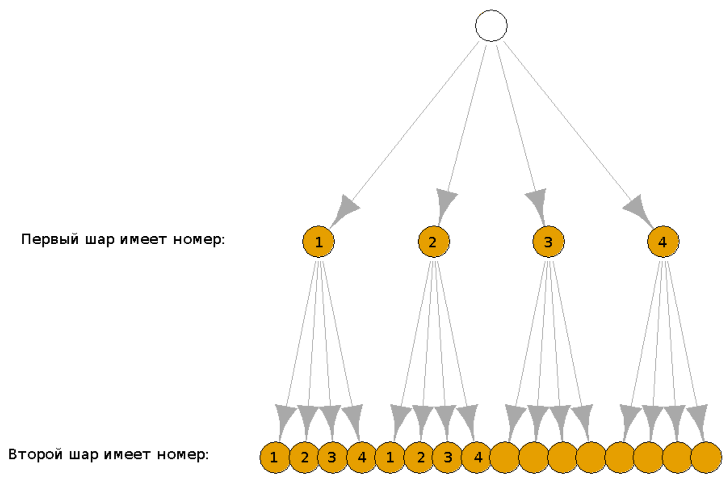
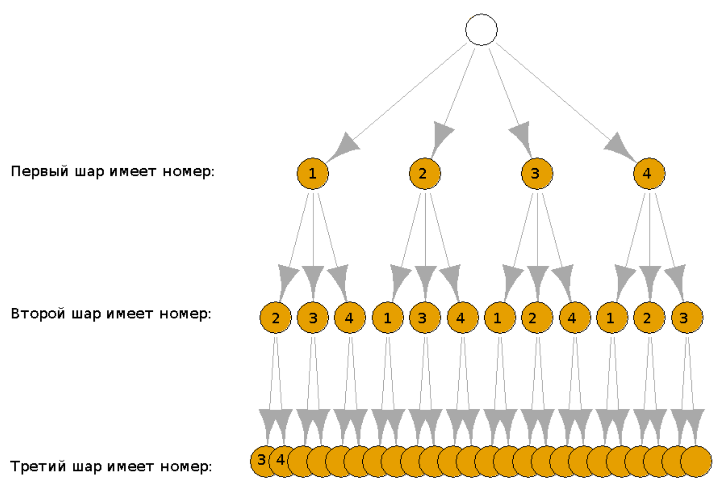


Рис. 3: Упорядоченный выбор без возвращения



Вопрос 1.4. А сколько способов достать пять шаров из семи, если шары не возвращаются в урну, а их порядок неважен?

3. Неупорядоченный выбор без возвращения. Пусть номера записываются без порядка, а шары после каждого вытягивания откладываются. Тогда естественно использовать

$$\Omega = \{\{i_1, \ldots, i_k\},\ i_1, \ldots, i_k \in \{1, \ldots, n\},\ i_j \neq i_l, j \neq l\}, \quad \mathcal{F} = 2^{\Omega}.$$

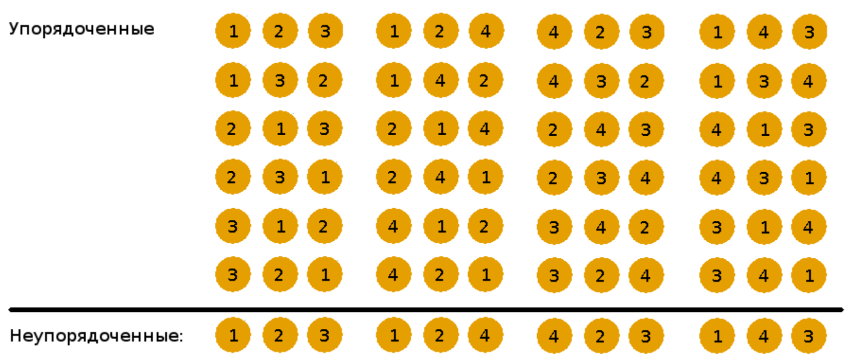
Здесь элементарный исход — это множество из k различных чисел от 1 до n, в котором порядок появления чисел уже не имеет значения. Значит, исходов станет меньше.

Если попробовать сравнить упорядоченные и неупорядоченный исходы, то мы заметим, что каждому неупорядоченному исходу соответствуют k! упорядоченных, которые ему равносильны. Действительно, к неупорядоченному исходу $\{1, \dots, k\}$ приведут любые упорядоченные исходы, в которых используются те же числа. Способов упорядочить $\{1, \dots, k\}$ как раз k!, откуда неупорядоченных исходов в k! раз меньше. Значит,

$$|\Omega| = C_n^k = \frac{1}{k!} A_n^k = \frac{n!}{k!(n-k)!}.$$

Отметим, что упорядоченный и неупорядоченный выбор с возвращением соответствуют одной физической

Рис. 4: Способы выбора трех предметов из четырех



реальности, одинаковые события в обеих моделях имеют одинаковую вероятность.

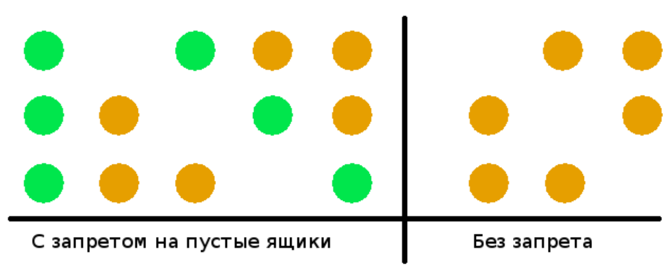
4. Неупорядоченный выбор с возвращением. Пусть номера записываются без порядка, а шары после каждого вытягивания возвращаются в урну. Тогда зададим пространство немного в другом формате

$$\Omega = \{(i_1, \dots, i_n),\ i_1 + \dots + i_n = k\}, \quad \mathcal{F} = 2^\Omega.$$

Здесь элементарный исход — это количество i_1 шаров с цифрой 1, количество i_2 шаров с цифрой 2,..., i_n шаров с цифрой n.

Сколько при этом исходов в Ω? Давайте посмотрим на другую задачу с тем же пространством элементарных исходов: k неразличимых предметов раскладываются по n ящикам. Тогда если i_1 обозначить число предметов в первом ящике,..., i_n — в n-м ящике, то мы получим то же самое множество.

Рис. 5: Связь раскладок 4 шаров по 2 ящикам без пустых ящиков или 2 шаров по 2 ящикам с возможными пустыми. Правые раскладки получаются из левых изъятием зеленых шаров



Чтобы решить эту задачу, решим сперва более простую — пусть мы хотим разложить предметы по ящикам так, чтобы пустых ящиков не было. Это все равно, что поставить между k предметами $n-1$ перегородку (стенку) будущих ящиков. Это все равно что выбрать из $k-1$ места $n-1$ без учета порядка. Мы умеем решать такую задачу, таких способов C_{k-1}^{n-1}.

Теперь вернемся к исходной задаче. Здесь уже нельзя так просто выбрать перегородки — несколько из них могут встать на одну и ту же позицию. Давайте добавим к нашим k предметам еще n штук и заранее положим по одному из добавленных предметов в каждый из ящиков. Теперь уже пустых ящиков не будет. Каждой раскладке k предметов по n ящикам с возможными пустыми ящиками соответствует единственная раскладка $k+n$ предметов по n ящикам без пустых ящиков. То есть у нашей задачи ответ C_{n+k-1}^{n-1}.

Однако, вряд ли мы можем естественным образом выбрать в этом случае классическое задание вероятности, поскольку здесь мы не видим никакой симметрии между исходами. Скажем, при $n = k$ исход ”все шары имеют номер 1” и исход ”все номера встречаются по одному разу” не выглядят равнозначными, второй бывает гораздо чаще. Собственно, пример с двумя монетами, который мы уже рассматривали выше, является частным случаем этого эффекта.

Пример 1.5. Представим себе, что студент Иванов на экзамене знает k билетов из n. Какая вероятность того, что он, стоя в очереди m-м, вытянет знакомый билет?

Этот опыт соответствует модели 2) — упорядоченный выбор без возвращения. Соответственно,

$$\Omega = \{(i_1, \dots, i_m),\ i_j \neq i_l,\ j \neq l\}, \quad |\Omega| = A_m^n, \quad \mathbf{P}(\omega_j) = \frac{1}{A_m^n}.$$

Нас интересует событие $A = \{(i_1, \dots, i_m) \in \Omega : i_m \in \{1, \dots, k\}\}$. Аналогично тому, как мы нашли количество исходов в модели 2), мы можем понять, что

$$|A| = k \cdot (n-1) \cdots (n-m+1),$$

поскольку билет для Иванова выбирается k способами, для первого из студентов — $n-1$ способом и так далее. Итого,

$$\mathbf{P}(A) = \frac{|A|}{|\Omega|} = \frac{k(n-1)\cdots(n-m+1)}{n(n-1)\cdots(n-m+1)} = \frac{k}{n}.$$

Как и следовало ожидать, Иванову совершенно неважно каким тянуть билет, шансы у него те же. В каких-то случаях его шансы вытянуть удачный билет повышаются, в каких-то понижаются, но все эти случаи волшебным образом воплощаются в той же вероятности, что и у первого студента.

Отметим, что здесь мы использовали такой прием — в нашем словесном представлении исходной задачи нам трудно заставить сперва выбрать билет Иванова, а потом всех остальных. А вот в пространстве элементарных исходов мы считаем последовательности и нам совершенно неважно, в каком порядке пронумерованы их члены. Вероятностная модель позволила нам избавиться от лишних условностей, которыми нас связывает физическая постановка задачи.

1.4 Ответы на вопросы

1. (c), потому что два остальных события по существу включаются в в третье.

2. Событий станет 13. К 9 предыдущим события добавятся еще 4 — ”монета потерялась при первом броске”, ”монета выпала на орла, а потом потерялась”, ”монета выпала на решку, а потом потерялась”, ”монета встала на ребро, а потом потерялась”.

3. Нет, неверно. Такие события не так уж часты, они называются независимыми, и мы поговорим о них на следующей лекции. В частности, если $A = B$, то мы придем к равенству только если $\mathbf{P}(A) = \mathbf{P}(A)^2$, то есть $\mathbf{P}(A) \in \{0, 1\}$. Аналогично, если $AB = \emptyset$, то мы придем к равенству только если $\mathbf{P}(A) = 0$ или $\mathbf{P}(B) = 0$.

4. $C_7^5 = 7 \cdot 6/2 = 21$.

2 Условная вероятность. Независимость

2.1 Счетное вероятностное пространство

Сегодня мы будем рассматривать пространства с, вообще говоря, бесконечным числом элементарных исходов, но не более чем счетным.

При этом мы дополнительно требуем от вероятности выполнения свойства счетной аддитивности

$$\mathbf{P}(A_1 + \dots + A_n + \cdots) = \sum_{n=1}^{\infty} \mathbf{P}(A_n).$$

Как и прежде мы будем задавать набор вероятностей p_i элементарных исходов, в сумме дающих единицу:

$$\sum_{i=1}^{\infty} p_i = 1.$$

Отсюда автоматически задается вероятность любого события

$$\mathbf{P}(A) = \sum_{k:\ \omega_k \in A} \mathbf{P}(\omega_k) = \sum_{k:\ \omega_k \in A} p_k,$$

только теперь сумма может быть бесконечной.

2.2 Введение

Наше представление о вероятности события зависит от нашей информированности. Вероятность того, что у моего соперника среди пяти карт будет туз пик равна $5/52$. Если я знаю 5 карт в своей руке, то эта вероятность уже $5/47$ (или 0, конечно же). А если я помню три карты, которые обменял ходом раньше, то уже $5/44$. А если у него за спиной стоит зеркало и я вижу там туза пик, то она равна 1.

В наш век информации мы постоянно получаем новые сведения, которые в силу вышесказанного, меняют нашу вероятность. Сегодня мы поговорим о том, как работать в вероятностной модели в условиях поступающей информации. *Часть до начала раздела 2.3 "Условная вероятность" имеет целью только объяснить происхождение формул, но безболезненно может быть пропущена.*

2.2.1 Классический случай

Пусть $\Omega = \{\omega_1, \dots, \omega_N\}$ — конечное пространство элементарных исходов с классически заданной вероятностью $\mathbf{P}(\omega_i) = 1/N$. Тогда вероятность события A определяется соотношением

$$\mathbf{P}(A) = \frac{|A|}{N}.$$

Предположим, что мы узнали о том, что выполнено событие B. Тогда B становится нашим новым вероятностным пространством, а наше событие A превращается в $A \cap B$. При этом из соображений симметрии все исходы остаются равновероятными, поэтому

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

Можно заметить, что можно записать эту вероятность и в старых терминах.

$$\mathbf{P}_B(A) = \frac{|A \cap B|/N}{|B|/N} = \frac{\mathbf{P}(AB)}{\mathbf{P}(B)}.$$

Итак, вероятность A, "при условии, что случилось B" задается формулой $\mathbf{P}(AB)/\mathbf{P}(B)$.

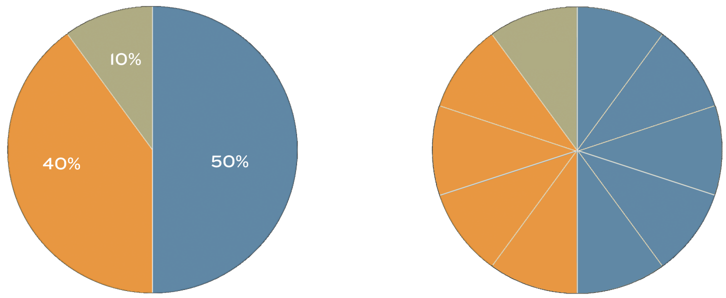
2.2.2 "Рациональный" случай

Пусть $(\Omega, \mathcal{F}, \mathbf{P})$ — вероятностное пространство, где $\Omega = \{\omega_1, \dots, \omega_N\}$ — конечное пространство элементарных исходов, вероятности исходов равны $p_1, \dots, p_N$, соответственно, где $p_1 = a_1/b_1, \dots, p_N = a_N/b_N$ — рациональные числа.

Введем новое пространство $\widetilde{\Omega} = \{\widetilde{\omega}_1, \dots, \widetilde{\omega}_M\}$, где $M = b_1 \cdots b_N$, первые $a_1 b_2 \cdots .b_N$ новых исходов в объединении дают ω_1, вторые $b_1 a_2 b_3 \cdots b_N$ — ω_2 и так далее. Если мы определим вероятность $\widetilde{\mathbf{P}}$ на пространстве $\widetilde{\Omega}$ классическим образом, то $(\Omega, \mathcal{F}, \mathbf{P})$ и $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P})$ окажутся согласованными — любом событие A имеет ту же вероятность $\mathbf{P}(A)$, что и соответствующее ему событие $\widetilde{A}$ из $\widetilde{\mathcal{F}}$. Иначе говоря, мы разбили каждый исход в старом пространстве Ω на части так, чтобы они получились равной вероятности.

Не так важно, как это реализовано физически, для нас существенно, что нашему пространству соответствует пространству с классически определенной вероятностью, на котором соображения симметрии и равноправия позволяют понять как надо определять вероятность. Мы будем прибегать к этому мысленному эксперименту

Рис. 6: Пример разбиения пространства с неклассически заданной вероятностью



и впредь — если мы хотим определить какое-то понятия с физически понятным подтекстом, то мы будем естественным образом определять его на классическом пространстве, а затем переносить его на произвольное.

В пространстве $\widetilde{\Omega}$ мы определяем вероятность события A "при условии, что случилось B" формулой $\widetilde{\mathbf{P}}(AB)/\widetilde{\mathbf{P}}(B)$. Значит в старом пространстве естественно определять ее формулой

$$\frac{\mathbf{P}(AB)}{\mathbf{P}(B)}.$$

2.2.3 Общий конечный случай

Пусть $(\Omega, \mathcal{F}, \mathbf{P})$ — вероятностное пространство, где $\Omega = \{\omega_1, \dots, \omega_N\}$ — конечное пространство элементарных исходов, вероятности исходов равны $p_1, \dots, p_N$, соответственно, где $p_1, \dots, p_N$ не обязательно рациональны. Для любого $\varepsilon > 0$ мы можем выбрать такой набор рациональных вероятностей $\widetilde{p}_1, \dots, \widetilde{p}_N$, что

$$\widetilde{p}_i(1-\varepsilon) < p_i < (1+\varepsilon)\widetilde{p}_i.$$

При этом для любого события $A \in \mathcal{F}$

$$\widetilde{\mathbf{P}}(A) - \varepsilon < \mathbf{P}(A) < \widetilde{\mathbf{P}}(A) + \varepsilon.$$

Итак, пространства очень похожи, все события в них имеют практически те же вероятности. Если мы определяем во втором пространстве вероятность события A "при условии, что случилось B" формулой $\widetilde{\mathbf{P}}(AB)/\widetilde{\mathbf{P}}(B)$, то и в первом естественно определять ее формулой

$$\mathbf{P}(AB)/\mathbf{P}(B).$$

2.2.4 Общий случай

Наконец, если $\Omega = \{\omega_1, \dots, \omega_n, \dots\}$ — бесконечное пространство элементарных исходов с вероятностями $p_1, p_2, \dots$, то для любого положительного ε мы можем найти такое N, что

$$\sum_{i=N+1}^{\infty} p_i < \varepsilon.$$

Возьмем пространство $\widetilde{\Omega} = \{\omega_1, \dots, \omega_N, \widetilde{\omega}\}$, где $\widetilde{\omega} = \cup_{i=N+1}^{\infty}\omega_i$. Попросту говоря, соединим все исходы начиная с $N+1$ в один исход.

Событию $A \subset \Omega$ сопоставим событие $\widetilde{A} \subset \widetilde{\Omega}$, которое совпадает с A при $A \cap \widetilde{\omega} = \emptyset$ и равно $A \cup \widetilde{\omega}$ при $A \cap \widetilde{\omega} \neq \emptyset$. То есть если хотя бы один из исходов после $N+1$ попал в A, то мы весь объединенный кусочек $\widetilde{\omega}$ отнесем в $\widetilde{A}$.

Тогда

$$\widetilde{\mathbf{P}}(\widetilde{A}) - \varepsilon < \mathbf{P}(A) < \widetilde{\mathbf{P}}(\widetilde{A}).$$

Значит вероятности событий A и $\widetilde{A}$ достаточно близки при достаточно малом ε. Пространство $(\widetilde{\Omega}, \widetilde{F}, \widetilde{P})$ подходит под условия прошлого случая, поэтому на нем мы определяем вероятность события $\widetilde{A}$ "при условии, что случилось $\widetilde{B}$" с помощью формулы

$$\frac{\widetilde{\mathbf{P}}(\widetilde{A}\widetilde{B})}{\widetilde{\mathbf{P}}(\widetilde{B})}.$$

Тогда и вероятность события A "при условии, что случилось B" естественно определять как $\mathbf{P}(AB)/\mathbf{P}(B)$.

2.3 Условная вероятность

Итак, мы пришли к следующему определению

Определение 2.1. *Условной вероятностью* события A при условии B называют

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

Здесь предполагается, что $\mathbf{P}(B) > 0$ (то есть не ноль).

Если рассуждения предыдущего раздела вам не понравились, то вы можете просто пропустить их и считать, что это определение и всё тут.

Содержательно мы понимаем под условной вероятностью вероятность события A, когда мы уже знаем, что случилось B.

Вопрос 2.1. При броске двух костей выпали два четных числа. Какая вероятность, что в сумме выпало 8 очков? a) 4/9;
b) 1/3;
c) 2/5;
d) 11/36.

Пример 2.1. Пусть в урне 15 шаров из которых 5 белых и 10 черных шаров, мы выбираем три шара без возвращения. Пусть A — событие "первый шар белый", B — событие "второй шар белый", C — событие "третий шар черный".

Пространство элементарных исходов имеет вид

$$\Omega = \{(i,j,k), i \neq j, i \neq k, j \neq k, i,j,k \in \{1,\dots,15\}\}, \quad A = \{(i,j,k) \in \Omega : i \in \{1,\dots,5\}\},$$
$$B = \{(i,j,k) \in \Omega : j \in \{1,\dots,5\}\}, \quad C = \{(i,j,k) \in \Omega : k \in \{6,\dots,15\}\}.$$

Тогда

$$\mathbf{P}(A) = \frac{5\cdot 14\cdot 13}{15\cdot 14\cdot 13} = \frac{1}{3}, \quad \mathbf{P}(AB) = \frac{5\cdot 4\cdot 13}{15\cdot 14\cdot 13} = \frac{2}{21}, \quad \mathbf{P}(ABC) = \frac{5\cdot 4\cdot 10}{15\cdot 14\cdot 13} = \frac{10}{39}.$$

Следовательно,

$$\mathbf{P}(B|A) = \frac{5\cdot 4\cdot 13}{5\cdot 14\cdot 13} = \frac{4}{14}, \quad \mathbf{P}(C|AB) = \frac{5\cdot 4\cdot 10}{5\cdot 4\cdot 13} = \frac{10}{13}.$$

Ответы совпадают с нашими ожиданиями. Действительно, вероятность B при условии A — это вероятность вытянуть белый шар вторым из урны в мире, в котором первым вытянутым шар был белым. В этом мире в урне перед вторым вытягиванием было 14 шаров, из которых 4 белых. Точно также вероятность C при условии AB — это вероятность вытянуть черный шар из урны, где 10 черных и 3 белых шара.

Зачастую хочется записать что-то в духе двойной условной вероятности $\mathbf{P}((C|B)|A)$ (так писать неправильно) — вероятность, что случится C, если уже случилось B, а перед этим случилось A. Однако, это просто $\mathbf{P}(C|AB)$ как из формального определения, так и из его физической интерпретации.

Как мы видим из примера с шарами, зачастую мы легко можем найти условные вероятности (потому что легко работаем в новой реальности, заданной условием), а вот вероятности пересечения нам получить сложнее. В связи с этим полезно перевернуть определение условной вероятности:

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

Аналогичный результат можно доказать в случае n событий $A_1, \dots, A_n$:

$$\mathbf{P}(A_n \cdots A_1) = \mathbf{P}(A_n|A_{n-1}\cdots A_1)\cdots\mathbf{P}(A_2|A_1)\mathbf{P}(A_1).$$

Этот результат иногда называют теоремой умножения.

Доказательство. Докажем этот факт по индукции. При $n = 2$ он верен.

Пусть при $n = k$ он верен, докажем его при $n = k + 1$. В силу предположения и базы индукции

$$\mathbf{P}(A_{k+1} \cdots A_1) = \mathbf{P}(A_{k+1}|A_k \cdots A_1)\mathbf{P}(A_k \cdots A_1) = \mathbf{P}(A_{k+1}|A_k \cdots A_1)\mathbf{P}(A_k|A_{k-1} \cdots A_1) \cdots \mathbf{P}(A_2|A_1)\mathbf{P}(A_1),$$

что и требовалось доказать. $\square$

Эту формулу удобно использовать для подсчета вероятностей пересечения событий.

Вопрос 2.2. Чему равна вероятность того, что при трех вытягиваниях без возвращения из урны с 8 черными, 5 белыми и 3 красными шарами, мы вытянем черный, белый и снова черный шары.
a) 1/12
b) 5/64
c) 1/27
d) 4/21.

2.4 Формула полной вероятности и формула Байеса

Пример 2.2. Предположим, что мы хотим найти вероятность события B: второй шар, взятый из урны с 5 белыми и 10 черными шарами, будет белым.

Давайте разберемся какой шар выпал первым. Пусть A — событие, заключающееся в том, что первый шар белый. Мы легко можем найти

$$\mathbf{P}(AB) = \mathbf{P}(B|A)\mathbf{P}(A) = \frac{5}{15}\,\frac{4}{14} = \frac{2}{21}, \quad \mathbf{P}(\overline{A}B) = \mathbf{P}(B|\overline{A})\mathbf{P}(\overline{A}) = \frac{10}{15}\,\frac{5}{14} = \frac{5}{21},$$

откуда

$$\mathbf{P}(B) = \mathbf{P}(B|A)\mathbf{P}(A) + \mathbf{P}(B|\overline{A})\mathbf{P}(\overline{A}) = \frac{2}{21} + \frac{5}{21} = \frac{7}{21} = \frac{1}{3}.$$

Мы уже знаем этот ответ с прошлого занятия: вероятность, что второй шар белый та же, что и для первого. Но теперь мы получили новый инструмент получения такой формулы.

В общем случае аналогичная формула также справедлива. Сформулируем ее.

Определение 2.2. События $B_1, \ldots, B_N$ называются разбиением, если

$$\Omega = \sum_{n=1}^{N} B_n,$$

то есть если B_i не пересекаются, и их объединение есть все Ω. Мы предполагаем, что N может быть конечным или бесконечным.

Пусть $B_1, B_2, \ldots$. — события с ненулевой вероятностью, образующие разбиение Ω. Тогда

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

Это соотношение называют *формулой полной вероятности*. Если число событий счетно, то формула остается в силе, но конечная сумма превращается в ряд.

Почему эта формула так полезна? Она позволяет посчитать вероятность в сложно устроенных экспериментах.

Пример 2.3. Пусть мы подбросили кубик, а затем бросили монету с вероятностью орла p столько раз, сколько очков выпало на кубике. Какая вероятность того, что выпала хотя бы одна решка?

Есть некоторые сложности с тем, чтобы построить пространство элементарных исходов. Рассмотрим такие пространство и вероятность на нем:

$$\Omega = \{(n, x),\ n \in \{1, \ldots, 6\},\ x \in \{0, 1\}^n\}, \quad \mathbf{P}((n, x)) = p^{\sum_{i=1}^n x_i}(1-p)^{\sum_{i=1}^n (1-x_i)}.$$

В этом пространстве не слишком удобно работать. Куда удобнее зафиксировать n и после этого уже работать с более простым пространством. Пусть B_n — событие "на кубике выпало n", A — событие "есть хотя бы одна решка". Тогда

$$\mathbf{P}(A) = \sum_{n=1}^{6} \mathbf{P}(A|B_n)\mathbf{P}(B_n) = \frac{1}{6}\sum_{n=1}^{6} \mathbf{P}(A|B_n).$$

При этом

$$\mathbf{P}(A|B_n) = 1 - \mathbf{P}(\overline{A}|B_n) = 1 - p^n,$$

а значит

$$\mathbf{P}(A) = \frac{1}{6}\sum_{n=1}^{6}(1 - p^n) = 1 - \frac{p(1-p^6)}{6(1-p)}.$$

Как мы видим, с помощью формулы полной вероятности можно свести сложную задачу к нескольким простым, не погружаясь в запутанное конструирование общего вероятностного пространства.

Вопрос 2.3. Мы подбрасываем монету с вероятностью орла 1/3. Если она выпадает на решку, то мы вытаскиваем шар из урны с 3 белыми и 7 черными шарами, а если на орла, то из урны с 9 белыми и 1 черным шаром. Какая вероятность того, что мы вытащим белый шар?
a) 3/5
b) 7/10.
c) 2/5.
d) 1/2.

Отметим еще одну формулу, легко вытекающую из формулы полной вероятности. Предположим, что мы знаем $\mathbf{P}(B_1), \dots, \mathbf{P}(B_n)$ для разбиения $B_1, \dots, B_n$. Пусть также известны $\mathbf{P}(A|B_1), \dots, \mathbf{P}(A|B_n)$. Тогда

$$\mathbf{P}(B_1|A) = \frac{\mathbf{P}(A|B_1)\mathbf{P}(B_1)}{\mathbf{P}(A)} = \frac{\mathbf{P}(A|B_1)\mathbf{P}(B_1)}{\sum_{i=1}^{n}\mathbf{P}(A|B_i)\mathbf{P}(B_i)}.$$

Данная формула называется формулой Байеса.

Пример 2.4. Представим, что в последнем вопросе мы видим, что шар белый, и хотим восстановить на что упала монета. Тогда если B_1 — монета упала на орла, A — шар белый, то

$$\mathbf{P}(B_1|A) = \frac{\mathbf{P}(A|B_1)\mathbf{P}(B_1)}{\mathbf{P}(A|B_1)\mathbf{P}(B_1) + \mathbf{P}(A|\overline{B_1})\mathbf{P}(\overline{B_1})} = \frac{\frac{1}{3}\cdot\frac{9}{10}}{\frac{1}{3}\cdot\frac{9}{10} + \frac{2}{3}\cdot\frac{3}{10}} = \frac{3}{5}.$$

После вытягивания белого шара вероятность того, что монета упала на орла сильно возросла — от 0.(3) до 0.6, то есть почти вдвое.

2.5 Независимость

Это понятие является ключевым для развития теории вероятностей в отдельности от общей теории меры. Итак, дадим следующее естественное определение.

Определение 2.3. События A и B называются независимыми, если $\mathbf{P}(A|B) = \mathbf{P}(A)$, где предполагается, что $\mathbf{P}(B) > 0$.

Это определение в силу определения условной вероятности переписывается в виде

Определение 2.4. События A и B называются независимыми, если $\mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B)$.

Второе определение чуть более общее (не требует $\mathbf{P}(B) > 0$) и мы будем использовать его.

Вопрос 2.4. Какие из соотношений между событиями невозможны?
(a) $\mathbf{P}(AB) < \mathbf{P}(A)\mathbf{P}(B)$;
(b) $\mathbf{P}(AB) > \mathbf{P}(A)\mathbf{P}(B)$;
(c) невозможны оба неравенства (a) и (b);
(d) возможны оба неравенства (a) и (b).

Начнем с нескольких простых замечаний, которые вам предстоит доказать самостоятельно:

- Если A и B независимы, то A и $\overline{B}$ независимы, B и $\overline{A}$ независимы, $\overline{A}$ и $\overline{B}$ независимы
- События вероятности 0 и 1 независимы от любых событий.
- Событие независимо от самого себя только если оно имеет вероятность 0 или 1.
- Непересекающиеся события независимы только если одно из них имеет вероятность 0.
- Если A не зависит от B и A независит от C, где $BC = \emptyset$, то A независит от $B + C$.

Нам понадобится более общее определение независимости n событий.

Определение 2.5. События $A_1, \ldots, A_n$ — независимы (в совокупности), если для любых $1 \leq i_1 < i_2 < \ldots < i_k \leq n$

$$\mathbf{P}(A_{i_1} \cdots A_{i_k}) = \mathbf{P}(A_{i_1}) \cdots \mathbf{P}(A_{i_k}).$$

Давайте поймем почему нужно такое сложное определение. Неужели не хватит просто независимости каждого события с каждым (такую независимость называют попарной)?

Пример 2.5. Предположим, что мы бросаем симметричную монету дважды. Рассмотрим события A $= \{$первый бросок оказался на орла$\}$, B $= \{$второй бросок оказался на орла$\}$, C $= \{$ровно один бросок оказался на орла$\}$. Тогда

$$\mathbf{P}(A|B) = \mathbf{P}(A) = \frac{1}{2},\ \mathbf{P}(C|A) = \mathbf{P}(C) = \frac{1}{2},\ \mathbf{P}(C|B) = \mathbf{P}(C) = \frac{1}{2}.$$

Поэтому каждые два события независимы. Однако,

$$\mathbf{P}(ABC) = 0 \neq \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C) = \frac{1}{8}.$$

Поэтому каждое из событий не влияет на шансы каждого из других случиться, но два события вместе делают невозможным третье.

Точно также можно построить n событий, любые $n-1$ из которых независимы, а все n вместе зависимы.

Можно привести следующий критерий независимости, обобщающий первое свойство независимости двух событий, указанное выше:

Лемма 2.1. *События $A_1, \ldots, A_n$ независимы тогда и только тогда, когда*

$$\mathbf{P}\left(A_1^{\delta_1} \cdots A_n^{\delta_n}\right) = \mathbf{P}\left(A_1^{\delta_1}\right) \cdots \mathbf{P}\left(A_n^{\delta_n}\right), \tag{1}$$

где $\delta_i \in \{0, 1\}$, $A^1 = A$, $A^0 = \overline{A}$.

Доказательство. 1) Докажем достаточность. Пусть $A_1, \ldots, A_n$ независимы. Покажем большее, чем нам требуется — что

$$\mathbf{P}(A_{i_1}^{\delta_{i_1}} \cdots A_{i_l}^{\delta_{i_l}}) = \mathbf{P}(A_{i_1}^{\delta_{i_1}}) \cdots \mathbf{P}(A_{i_l}^{\delta_{i_l}})$$

при любых l, $1 \leq i_1 < \ldots < i_l \leq n$, $\delta_{i_1}, \ldots, \delta_{i_l} \in \{0, 1\}$. Доказательство проведем индукцией по $k = \delta_{i_1} + \cdots . + \delta_{i_l}$.

База индукции: при $k = 0$ утверждение вытекает из того, что $A_1, \ldots, A_n$ независимы.

Переход индукции: пусть при $k \leq m$ при некотором m утверждение доказано, докажем при $k = m + 1$. Без ограничения общности считаем $\delta_{i_1} = 0$. Тогда

$$\mathbf{P}\left(A_{i_1}^0 A_{i_2}^{\delta_{i_2}} \cdots A_{i_l}^{\delta_{i_l}}\right) = \mathbf{P}\left(A_{i_2}^{\delta_{i_2}} \cdots A_{i_l}^{\delta_{i_l}}\right) - \mathbf{P}\left(A_{i_1}^1 A_{i_2}^{\delta_{i_2}} \cdots A_{i_l}^{\delta_{i_l}}\right) \tag{2}$$

при любых $i_1, \ldots, i_l$. В силу предположения индукции

$$\mathbf{P}\left(A_{i_2}^{\delta_{i_2}} \cdots A_{i_l}^{\delta_{i_l}}\right) = \prod_{i=1}^{l} \mathbf{P}\left(A_{i_2}^{\delta_{i_2}}\right) \cdots \mathbf{P}\left(A_{i_l}^{\delta_{i_l}}\right),\ \mathbf{P}\left(A_{i_1}^1 A_{i_2}^{\delta_{i_2}} \cdots A_{i_l}^{\delta_{i_l}}\right) = \mathbf{P}\left(A_{i_1}^1\right) \mathbf{P}\left(A_{i_2}^{\delta_{i_2}}\right) \cdots \mathbf{P}\left(A_{i_l}^{\delta_{i_l}}\right). \tag{3}$$

Подставляя (3) в (2), получаем

$$\mathbf{P}\left(A_{i_1}^0 A_{i_2}^{\delta_{i_2}} \cdots A_{i_l}^{\delta_{i_l}}\right) = \left(1 - \mathbf{P}\left(A_{i_1}\right)\right) \mathbf{P}\left(A_{i_2}^{\delta_{i_2}}\right) \cdots \mathbf{P}\left(A_{i_l}^{\delta_{i_l}}\right) \mathbf{P}\left(A_{i_1}^0\right) \mathbf{P}\left(A_{i_2}^{\delta_{i_2}}\right) \cdots \mathbf{P}\left(A_{i_l}^{\delta_{i_l}}\right).$$

Достаточность показана.
2) Докажем необходимость. Пусть мы знаем (1). Докажем, что

$$\mathbf{P}\left(A_{i_1} \cdots A_{i_l}\right) = \mathbf{P}\left(A_{i_1}\right) \cdots \mathbf{P}\left(A_{i_l}\right)$$

при всех $l \leq n$. Для удобства будем считать, что $i_1 = 1, \ldots, i_l = l$ (это вопрос нумерации A_i). Тогда

$$\mathbf{P}\left(A_1 \cdots A_l\right) = \sum_{\delta_{l+1}, \ldots, \delta_n \in \{0,1\}} \mathbf{P}\left(A_1 \cdots A_l A_{l+1}^{\delta_{l+1}} \cdots A_n^{\delta_n}\right) = \sum_{\delta_{l+1}, \ldots, \delta_n \in \{0,1\}} \mathbf{P}\left(A_1\right) \cdots \mathbf{P}\left(A_l\right) \mathbf{P}\left(A_{l+1}^{\delta_{l+1}}\right) \cdots \mathbf{P}\left(A_n^{\delta_n}\right) =$$

$$\mathbf{P}\left(A_1\right) \cdots \mathbf{P}\left(A_l\right) \sum_{\delta_{l+1}, \ldots, \delta_n \in \{0,1\}} \mathbf{P}\left(A_{l+1}^{\delta_{l+1}}\right) \cdots \mathbf{P}\left(A_n^{\delta_n}\right).$$

Остается показать, что сумма в правой части равна единице. Для этого заметим, что это в точности

$$\left(\mathbf{P}\left(A_{l+1}^0\right) + \mathbf{P}\left(A_{l+1}^1\right)\right) \cdots \left(\mathbf{P}\left(A_n^0\right) + \mathbf{P}\left(A_n^1\right)\right) = 1 \cdot 1 \cdot \cdots \cdot 1 = 1.$$

Лемма доказана. □

2.6 Ответы на вопросы

1. Наше пространство состоит из пар $(2i, 2j)$, $i, j \in \{1, 2, 3\}$, следовательно в нем 9 исходов. Нам подходят 3 пары $(2, 6)$, $(4, 4)$, $(6, 2)$. Итого $1/3$.
2. В силу теоремы произведения
$$\frac{8}{16} \cdot \frac{5}{15} \cdot \frac{7}{14} = \frac{1}{12}.$$
3. В силу формулы полной вероятности искомая вероятность есть
$$\frac{2}{3} \cdot \frac{3}{10} + \frac{1}{3} \cdot \frac{9}{10} = \frac{5}{10} = \frac{1}{2}.$$
4. Возможны все эти соотношения. Наши неравенства удобнее переписать как соотношения между $\mathbf{P}(A|B)$ и $\mathbf{P}(A)$. В первом примере с шариками мы видели оба вида неравенств:
$$\mathbf{P}(B|A) = \frac{4}{14} < \frac{5}{15} = \mathbf{P}(B),\ \mathbf{P}(C|AB) = \frac{10}{13} > \frac{10}{15} = \mathbf{P}(C).$$
Это вполне логично, случившиеся события могут повышать вероятность других, а могут понижать. Вытаскивая белые шары мы повышаем шансы вытаскивать черные и понижаем шансы выстаскивать белые.

3 Случайные величины

3.1 Последовательность независимых испытаний

Мы научились рассматривать независимые события на одном вероятностном пространстве, но зачастую мы сталкиваемся с другой ситуацией. Мы проводим испытания для каждого из которых вероятностное пространство у нас известно, и хотим построить новое пространство, которое включает их все.

Итак, пусть $(\Omega_1, \mathcal{F}_1, \mathbf{P}_1)$,… с,$(\Omega_n, \mathcal{F}_n, \mathbf{P}_n)$ — вероятностные пространства, каждое из которых описывает свой опыт.

Определение 3.1. Пространство $(\Omega, \mathcal{F}, \mathbf{P})$, где

$$\Omega = \Omega_1 \times \cdots \times \Omega_n, \quad \mathcal{F} = 2^{\Omega}, \quad \mathbf{P}((\omega_{1,i_1}, \ldots, \omega_{n,i_n})) = \mathbf{P}_1(\omega_{1,i_1}) \cdots \mathbf{P}_n(\omega_{n,i_n})$$

называют прямым произведением пространств $(\Omega_1, \mathcal{F}_1, \mathbf{P}_1), \ldots, (\Omega_n, \mathcal{F}_n, \mathbf{P}_n)$.

Прямое произведение будет соответствовать независимому проведению всех этих опытов. Наш новый элементарный исход представляет собой вектор $(\omega_{1,i_1}, \ldots, \omega_{n,i_n})$ из исходов каждого из опытов, а вероятность такого исхода будет равна произведению вероятностей. Мы уже делали это, например, при работе с монетой.

При этом для любых $B_1 \in \mathcal{F}_1, \ldots, B_n \in \mathcal{F}_n$ события

$$A_k = \{(\omega_{1,i_1}, \ldots, \omega_{n,i_n}) : \omega_{k,i_k} \in B_k\}$$

при $k = 1, \ldots, n$ будут независимыми. Это вполне естественно, поскольку события A_k относятся к разным опытам.

Пример 3.1. Представим себе, что все $(\Omega_i, \mathcal{F}_i, P_i)$ одинаковы и соответствуют одному бросанию монеты:

$$\Omega = \{0, 1\}, \quad \mathcal{F} = \{\emptyset, \{0\}, \{1\}, \Omega\}, \quad \mathbf{P}(0) = 1 - p, \ \mathbf{P}(1) = p.$$

Тогда прямое произведение пространств будет представлять собой

$$\Omega = \{(i_1, \ldots, i_n) : i_j \in \{0, 1\}\}, \quad \mathcal{F} = 2^\Omega, \quad \mathbf{P}((i_1, i_2, \ldots, i_n)) = p^{\#\{j : i_j = 1\}}(1 - p)^{\#\{j : i_j = 0\}},$$

где $\#$ — количество элементов в множестве. Это так называемая схема Бернулли, описывающая результат n независимых бросаний несимметричной монеты. Кстати, для вероятностей элементарных исходов можно написать более изящную формулу

$$\mathbf{P}((i_1, i_2, \ldots, i_n)) = p^{i_1 + \cdots + i_n}(1 - p)^{n - i_1 - \cdots - i_n}.$$

3.2 Случайные величины

Люди стараются оперировать с количественными показателями явлений, что упрощает работу с ними. Мы измеряем температуру, время, вес, рост и другие показатели. Также и в теории вероятностей чаще всего нас интересует не сам исход, а некоторая его числовая характеристика — число, выпавшее на кубике, число орлов при n бросаниях монеты и так далее.

Определение 3.2. Пусть $(\Omega, 2^\Omega, \mathbf{P})$ — конечное или счетное вероятностное пространство. Случайной величиной на $(\Omega, 2^\Omega, \mathbf{P})$ называют отображение $X : \Omega \to \mathbb{R}$.

Обратите внимание, что мы определяем так величину только в том случае, если пространство событий есть 2^Ω. О том, как быть, если это не так, мы поговорим позднее. В нашем случае конечных или счетных пространств вполне достаточно рассматривать именно случай $\mathcal{F} = 2^\Omega$.

Пример 3.2. Наиболее простую величину представляет собой I_A — индикатор события A, то есть

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

Определение 3.3. Распределением вероятностей случайной величины X называют отображение $\mathbf{P}_X : 2^\mathbb{R} \to [0, 1]$, заданное соотношением

$$\mathbf{P}_X(B) := \mathbf{P}(\omega : X(\omega) \in B) = \sum_{i : X(\omega_i) \in B} p_i, \quad B \subseteq \mathbb{R}.$$

где $p_i = \mathbf{P}(\omega_i)$. Здесь $2^\mathbb{R}$ — множество всех подмножеств $\mathbb{R}$.

Поскольку распределение вероятностей в нашем случае полностью задается набором вероятностей

$$\mathbf{P}_X(x_i) = \mathbf{P}(\omega : X(\omega) = x_i)$$

для всевозможных x_i из множества $\mathcal{X} = \operatorname{Im} X$ (то есть всевозможных значений, принимаемых случайной величиной), то для задания распределения вероятностей мы будем находить их. Будем называть функцию $\mathbf{P}_X(x) = \mathbf{P}(\omega : X(\omega) = x)$ (вероятностной) функцией масс.

Немного о терминологии. В русскоязычной традиции, к сожалению, функцию масс также принято называть распределением вероятностей случайной величины. Это создает путаницу. Более правильно, вероятно, было бы называть ее функцией распределения вероятностей случайной величины, но этот термин занят другим объектом. В некоторых источниках (например, русскоязычной википедии) ее называют функцией вероятностей случайной величины, однако, насколько я могу судить, это не слишком распространенное обозначение, которое, на мой взгляд, достаточно неудачно.
В английской терминологии все намного четче. Существует probability distribution — распределение вероятностей и множество разных функций: cdf (cumulative distribution function) — то, что в русском языке называют функцией распределения, pmf (probability mass function) — то, что мы назвали функцией масс (а еще есть pdf, pgf и так далее). Поэтому мы позаимствуем этот термин из английского. Учтите, что в других источниках вместо него может использоваться слово распределение.

Функция масс будет равна нулю во всех точках кроме точек из $\mathcal{X}$, поэтому можно охарактеризовать распределение (или функцию масс) с помощью таблицы распределения:

значение	x_1	x_2	$\dots$	x_n
вероятность	p_1	p_2	$\dots$	$p_n.$

Пример 3.3. Рассмотрим схему Бернулли:

$$\Omega = \{(i_1, \dots, i_n) : i_j \in \{0, 1\}\}, \quad \mathcal{F} = 2^\Omega, \quad \mathbf{P}((i_1, \dots, i_n)) = p^{i_1+\dots+i_n}(1-p)^{n-i_1-\dots-i_n}.$$

Естественно рассмотреть случайные величины

$$X_k((i_1, \dots, i_n)) = i_k,$$

характеризующие результаты k-го броска. При этом таблицы распределения величин X_k одинаковые:

значение	0	1
вероятность	1-p	p

соответственно, у них одинаковые функции масс

$$p(x) = \mathbf{P}_X(x) = \begin{cases} p, & x = 1, \\ 1-p, & x = 0, \\ 0, & x \notin \{0, 1\}. \end{cases}$$

и распределения

$$\mathbf{P}_X(B) = \begin{cases} 1, & 0 \in B,\ 1 \in B, \\ 1-p, & 0 \in B,\ 1 \notin B, \\ p, & 0 \notin B,\ 1 \in B, \\ 0, & 0 \notin B,\ 1 \notin B. \end{cases}$$

Для случайной величины $X((i_1, \dots, i_n)) = i_1 + \dots + i_n$ (то есть числа орлов за n бросков монеты) мы также можем вычислить требуемые параметры. Понятно, что $\mathcal{X}$ в данном случае представляет собой $\{0, 1, \dots, n\}$.

Найдем функцию масс: при $k \in \{0, \dots, n\}$

$$\mathbf{P}_X(k) = \mathbf{P}(\omega : X(\omega) = k) = \sum_{(i_1,\dots,i_n): i_1+\dots+i_n=k} \mathbf{P}((i_1, \dots, i_n)) = \sum_{(i_1,\dots,i_n): i_1+\dots+i_n=k} p^k(1-p)^{n-k} =$$
$$p^k(1-p)^{n-k}\#\{(i_1, \dots, i_n) \in \Omega : i_1 + \dots + i_n = k\},$$

где последний множитель представляет собой количество элементарных исходов, в которых ровно k единиц. Число способов выбрать места для k единиц на n местах есть C_n^k, откуда

$$\mathbf{P}_X(k) = C_n^k p^k(1-p)^{n-k}, \quad k \in \{0, \dots, n\}.$$

Обратим внимание, что если мы изучаем случайную величину, то фактически мы создаем отдельное вероятностное пространство

$$(\mathcal{X}, 2^{\mathcal{X}}, \mathbf{P}_X).$$

Как мы видели выше, разные физически величины (например, результат первого броска и результат второго броска монеты) могут задавать одно и то же пространство. Таким образом, есть смысл рассматривать распределения сами по себе, не описывая явно $(\Omega, \mathcal{F}, \mathbf{P})$, а работая с $(\mathcal{X}, 2^{\mathcal{X}}, \mathbf{P}_X)$.

3.3 Классические дискретные распределения

Выпишем несколько полезных дискретных распределений.

1. Бернуллиевское распределение $Bernoulli(p)$. При этом $\mathcal{X} = \{0, 1\}$, $\mathbf{P}_X(0) = 1 - p$, $P_X(1) = p$. Как мы уже говорили — это соотвествует одному бросанию несимметричной монеты.

2. Биномиальное распределение $Binom(n, p)$. При этом $\mathcal{X} = \{0, 1, \ldots, n\}$,
$$\mathbf{P}_X(k) = C_n^k p^k (1-p)^k, \quad k \in \{0, \ldots, n\}.$$
Как мы видели выше — оно описывает число орлов при n бросаниях несимметричной монеты.

3. Дискретное равномерное распределение $R\{1, \ldots, N\}$ (или другое обозначение $U\{1, \ldots, N\}$). При этом $\mathcal{X} = \{1, \ldots, N\}$,
$$\mathbf{P}_X(k) = \frac{1}{N}, \quad k \in \{1, \ldots, N\}.$$
Это распределение описывает номер шара, извлеченного из урны с N шарами.

4. Геометрическое распределение $Geom(p)$. При этом $\mathcal{X} = \mathbb{N} \cup \{0\}$,
$$\mathbf{P}_X(k) = (1-p)^k p.$$
Это распределение, как мы увидим ниже, соответствует количеству решек, выпавших при бросании монеты до первого орла.

5. Пуассоновское распределение $Poiss(\lambda)$. При этом $\mathcal{X} = \mathbb{N} \cup \{0\}$,
$$\mathbf{P}_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$
Это распределение не имеет столь простого описания как предыдущие. Впрочем, как мы увидим на следующих занятиях, у него есть естественная интерпретация — оно хорошо приближает биномиальное распределение $Binom(n, p_n)$ с $p_n \sim \lambda/n$ при больших n.

6. Отрицательное биномиальное распределение $NegBinom(r, p)$. При этом $\mathcal{X} = \mathbb{N} \cup \{0\}$,
$$\mathbf{P}_X(k) = C_{r+k-1}^{r-1} (1-p)^k p^r.$$
Это распределение, как мы увидим ниже, соответствует количеству решек, выпавших при бросании монеты до r-о орла.

7. Гипергеометрическое распределение $HyperGeom(N, M, n)$. При этом $\mathcal{X} = \{\max(0, n-N+M), \ldots, \min(M, n)\}$,
$$P_X(k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}.$$
Это распределение, как мы увидим ниже, соответствует количеству белых шаров среди n извлеченных шаров из урны с N шарами, из которых M белых. Не пугайтесь некрасивого $\mathcal{X}$, наложенные условия просто не дают обнулиться биномиальным коэффициентам.

Вопрос 3.1. Какая из описанных величин является гипергеометрической?

1. Порядковый номер в списке группы наугад выбранного студента группы

2. Студент Фадеев тянет билет. Если билет не первый, то билет замешивается обратно и вытягивается заново. Число неудачных билетов до вытащенного первого.

3. Число студентов ФКИ, знающих доказательство теоремы о неявной функции, среди десяти наугад опрошенных.

4. Каждый день Вероника Евгеньевна выбирает наугад человека из группы, который будет мыть доску в аудитории. Количество девочек, выбранных за месяц.

Пример 3.4. Покажем, что геометрическое распределение действительно соответствуют указанному физическому описанию.

Здесь есть некоторое неудобство — нам сейчас сложно ввести вероятностное пространство для бесконечного числа бросаний монеты, потому что возможными исходами будут все бесконечные последовательности из нулей и единиц, которых несчетное число.

Поэтому мы будем искать именно $\mathbf{P}_X(k) = \mathbf{P}(\omega : X(\omega) = k)$. Для этого нам хватит опыта в виде $k+1$ бросаний монеты, то есть схемы Бернулли:

$$\mathbf{P}((i_1, \ldots, i_{k+1}) : X((i_1, \ldots, i_{k+1})) = k) = \mathbf{P}((0, 0, \ldots, 0, 1)) = (1-p)^k p.$$

Здесь мы пользуемся тем, что нам подходит ровно один исход — k первых бросков выпадала решка, а последний раз орел.

Пример 3.5. Покажем, что отрицательное биномиальное распределение действительно соответствуют указанному физическому описанию.

Опять же будем искать именно $\mathbf{P}_X(k) = \mathbf{P}(\omega : X(\omega) = k)$. Для этого нам хватит опыта в виде $k+r$ бросаний монеты, то есть схемы Бернулли:

$$\mathbf{P}((i_1, \ldots, i_{k+r}) : X((i_1, \ldots, i_{k+1})) = k) = \mathbf{P}((i_1, \ldots, i_{k+r-1}, 1) : i_1 + \cdots + i_{k+r-1} = r-1) = C_{k+r-1}^{r-1} p^r (1-p)^k.$$

Пример 3.6. Покажем, что гипергеометрическое распределение действительно соответствуют указанному физическому описанию. Пространство исходов для опыта вытягивания шаров представляет собой

$$\Omega = \{\{i_1, \ldots, i_n\} : \ i_j \in \{1, \ldots, N\}\}, \ i_j \neq i_l, \ j \neq l\}.$$

Всего в данном множестве C_N^n, которые из симметрии имеют равные вероятности. Подходящие под событие $X = k$ исходы это те, в которых из n чисел ровно k не больше M. Таких исходов $C_M^k C_{N-M}^{n-k}$ штук. Отсюда получаем искомое соотношение

$$\mathbf{P}_X(k) = \mathbf{P}(\omega : X(\omega) = k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}.$$

3.4 Случайные векторы

Определение 3.4. Случайным вектором называют отображение $\vec{X} : \Omega \to \mathbb{R}^d$.

Можно рассматривать случайный вектор $\vec{X}$ как набор величин $X_1, \ldots, X_d$ на одном вероятностном пространстве.

Определение 3.5. Распределением вероятностей случайного вектора $\vec{X}$ называют отображение $\mathbf{P}_X : 2^{\mathbb{R}^d} \to [0, 1]$, заданное соотношением

$$\mathbf{P}_{\vec{X}}(B) := \mathbf{P}(\omega : \vec{X}(\omega) \in B) = \sum_{i : \vec{X}(\omega_i) \in B} p_i, \quad B \subseteq \mathbb{R}^d.$$

где $p_i = \mathbf{P}(\omega_i)$. Здесь $2^{\mathbb{R}^d}$ — множество всех подмножеств $\mathbb{R}^d$.

Опять же распределение вероятностей в нашем случае полностью задается набором вероятностей

$$\mathbf{P}_{\vec{X}}(\vec{x}_i) = \mathbf{P}(\omega : \vec{X}(\omega) = \vec{x}_i)$$

для всевозможных $\vec{x}_i$ из множества $\mathcal{X} = \operatorname{Im} X$ (то есть всевозможных значений, принимаемых случайным вектором). Опять же будем называть функцию

$$\mathbf{P}_{\vec{X}}(\vec{x}) = \mathbf{P}(\omega : \vec{X}(\omega) = \vec{x})$$

(вероятностной) функцией масс. Функцию масс можно задавать таблицей распределения. Так в случае $d = 2$ можно задавать удобную двумерную таблицу

	x_1	x_2	$\dots$	x_n
y_1	$p_{1,1}$	$p_{1,2}$	$\dots$	$p_{1,n}$
y_2	$p_{2,1}$	$p_{2,2}$	$\dots$	$p_{2,n}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
y_m	$p_{m,1}$	$p_{m,2}$	$\dots$	$p_{m,n}$.

Вопрос 3.2. Вектор (X, Y) имеет таблицу распределения, приведенную ниже. Найти $\mathbf{P}(X = Y)$.

	1	2	3
0	0.3	0	0.1
2	0.1	0.05	0.15
3	0.05	0.15	0.1

Пример 3.7. Схема Бернулли задает случайный вектор $\vec{X} = (X_1, \dots, X_n)$, описывающий результаты бросков всех монет

$$\vec{X}(\omega) = \vec{X}((i_1, \dots, i_n)) = (i_1, \dots, i_n).$$

Пример 3.8. Представим себе задачу размещения по n ячейкам N предметов независимо с вероятностями $p_1, \dots, p_n$, соответственно. Этому эксперименту соответствует вероятностное пространство

$$\Omega = \{(i_1, \dots, i_N) : i_j \in \{1, \dots, n\}\}, \quad \mathbf{P}((i_1, \dots, i_N)) = p_{i_1} \cdots p_{i_N} = p_1^{\#\{j : i_j = 1\}} \cdots p_n^{\#\{j : i_j = n\}}.$$

Рассмотрим случайный вектор

$$\nu_1(i_1, \dots, i_N) = \#\{j : i_j = 1\}, \dots, \ \nu_n(i_1, \dots, i_N) = \#\{j : i_j = n\}.$$

Его функция масс задается формулой

$$\mathbf{P}_{\vec{\nu}}(x_1, \dots, x_n) = \mathbf{P}((i_1, \dots, i_N) : \ \nu_1(i_1, \dots, i_N) = x_1, \dots, \nu_n(i_1, \dots, i_N) = x_n) = p_1^{x_1} \cdots p_n^{x_n} M_{x_1, \dots, x_n},$$

где $x_1 + \dots + x_n = N$, $M_{x_1, \dots, x_n}$ — количество наборов $(i_1, \dots, i_N)$, в которых ровно x_1 единиц, x_2 двоек,..., x_n чисел n.

Найдем $M_{x_1, \dots, x_n}$. Представим себе, что все единицы, двойки и так далее индивидуальны и отличаются друг от друга. Тогда число способов расставить наши числа $N!$. Но с точки зрения исходной задачи все возможные перестановки, которые меняют единицы между собой, двойки между собой и так далее, оставляют один исход тем же. Таких перестановок $x_1!_n!$. Следовательно,

$$M_{x_1, \dots, x_n} = \frac{N!}{x_1!_n!}.$$

Значит,

$$\mathbf{P}_{\vec{\nu}}(x_1, \dots, x_n) = \frac{N!}{x_1!_n!} p_1^{x_1} \cdots p_n^{x_n}$$

Распределение с такой функцией масс называют полиномиальным распределением $Polynomial(N, p_1, \dots, p_n)$.

3.5 Независимые алгебры

Нам понадобится понятие независимых случайных величин. Для работы с ними нам будет полезно понятие независимых алгебр.

Напомним определение с первой лекции:

Определение 3.6. Алгеброй подмножеств множества Ω называют такой набор подмножеств $\mathcal{A}$, что

1. $\Omega \in \mathcal{A}$;
2. если $A \in \mathcal{A}$, то $\overline{A} \in \mathcal{A}$;

3. если $A \in \mathcal{A}$, $B \in \mathcal{A}$, то $A \cup B \in \mathcal{A}$.

Как мы говорили в прошлый раз, алгебра также содержит вместе с любыми множествами A, B множества $A \cap B$, $A \setminus B$, $A \Delta B$.

Вопрос 3.3. Какая из названных систем подмножеств множества $\{0, 1, 2, 3, 4\}$ является алгеброй?

1. $\emptyset, \{0, 1\}, \{2, 3, 4\}, \{0, 1, 2\}, \{3, 4\}, \{0, 1, 2, 3, 4\}$;
2. $\emptyset, \{0, 1\}, \{2, 3\}, \{0, 1, 2, 3\}, \{4\}, \{0, 1, 2, 3, 4\}$;
3. $\emptyset, \{0\}, \{1\}, \{0, 1\}, \{2, 3, 4\}, \{0, 2, 3, 4\}, \{1, 2, 3, 4\}, \{0, 1, 2, 3, 4\}$
4. $\{0, 1\}, \{2, 3, 4\}, \{0, 1, 2, 3, 4\}$

Определение 3.7. Алгеброй подмножеств Ω, порожденной разбиением $D_1, \ldots, D_n$, называют систему множеств

$$\{D_{i_1} + \cdots + D_{i_k},\ 1 \leq i_1 < i_2 < \ldots < i_k \leq n,\ k \geq 0\},$$

то есть множество всевозможных объединений подмножеств Ω. Алгебру, порожденную $D_1, \ldots, D_n$ обозначают $\sigma(D_1, \ldots, D_n)$. Здесь n может быть как конечным, так и бесконечным.

Проверьте, что это действительно алгебра, пользуясь определением разбиения. Легко видеть, что в ней в точности 2^n событий (каждое из n порождающих можно либо выбрать, либо не выбрать).

Пример 3.9. Самой простой такой алгеброй будет $\sigma(\Omega) = \{\emptyset, \Omega\}$. Чуть более сложной $\sigma(A, \overline{A}) = \{\emptyset, \Omega, A, \overline{A}\}$.

Определение 3.8. Алгебры $\mathcal{A}_1, \ldots, \mathcal{A}_n$ подмножеств Ω называют независимыми, если $A_1, \ldots, A_n$ независимы для любых $A_i \in \mathcal{A}_i$.

Пример 3.10. Рассмотрим $\Omega = \{1, 2, 3, 4\}$ с классически заданной вероятностью. Рассмотрим алгебры $\mathcal{A}_1 = \{\emptyset, \Omega, \{1, 2\}, \{3, 4\}\}$, $\mathcal{A}_2 = \{\emptyset, \Omega, \{1, 3\}, \{2, 4\}\}$. Покажем, что они независимы. Пусть $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$. Если одно из множеств является Ω или $\emptyset$, то оно автоматически независимо с другим. Тогда $A_1 = B_1^{\delta_1}$, $A_2 = B_2^{\delta_2}$, где $B_1 = \{1, 2\}$, $B_2 = \{1, 3\}$, δ_1, $\delta_2 \in \{0, 1\}$. События B_1, B_2 независимы. Но тогда в силу леммы, доказанной в конце прошлой лекции, независимы $B_1^{\delta_1}$ и $B_2^{\delta_2}$. Значит, $\mathcal{A}_1$ и $\mathcal{A}_2$ независимы.

Теорема 3.1. *Пусть алгебра $\mathcal{A}_i$ порождается разбиением $D_{i,1}, \ldots, D_{i,m_i}$, $i \leq n$. Тогда $\mathcal{A}_1, \ldots, \mathcal{A}_n$ независимы тогда и только тогда, когда $D_{1,i_1}, \ldots, D_{n,i_n}$ независимы при любых $i_1 \leq m_1, \ldots, i_k \leq m_k$.*

Доказательство. Прямо из определения независимости алгебр автоматически следует независимость указанных событий, поэтому доказывать достаточно только в обратную сторону.

Пусть $D_{1,i_1}, \ldots, D_{n,i_n}$ независимы при любых $i_1 \leq m_1, \ldots, i_k \leq m_k$. Покажем, что независимы $\mathcal{A}_1, \ldots, \mathcal{A}_n$.

Мы должны выбрать произвольные события $B_1 \in \mathcal{A}_1, \ldots, B_n \in \mathcal{A}_n$ и проверить их независимость. В силу леммы, доказанной в конце прошлого занятия, для этого достаточно проверить, что

$$\mathbf{P}(B_1^{\delta_1} \cdots B_n^{\delta_n}) = \mathbf{P}(B_1^{\delta_1}) \cdots \mathbf{P}(B_n^{\delta_n})$$

при любых $\delta_1, \ldots, \delta_n \in \{0, 1\}$. Однако, поскольку $B_i^{\delta_i}$ также лежат в $\mathcal{A}_i$, то достаточно доказать, что

$$\mathbf{P}(B_1 \cdots B_n) = \mathbf{P}(B_1) \cdots \mathbf{P}(B_n)$$

для всех $B_i \in \mathcal{A}_i$. При этом из определения $\mathcal{A}_i$ события B_i представляют собой объединения $D_{i,j}$ по каким-то j. Для удобства будем считать, что

$$B_i = D_{i,1} + \cdots + D_{i,l_i}, \quad l_i \leq m_i.$$

Тогда

$$\mathbf{P}(B_1 \cdots B_n) = \sum_{j_1=1}^{l_1} \ldots \sum_{j_n=1}^{l_n} \mathbf{P}(D_{1,j_1} \cdots D_{n,j_n}) = \sum_{j_1=1}^{l_1} \cdots \sum_{j_n=1}^{l_n} \mathbf{P}(D_{1,j_1}) \ldots \mathbf{P}(D_{n,j_n}) =$$
$$\sum_{j_1=1}^{l_1} \mathbf{P}(D_{1,j_1}) \cdots \sum_{j_n=1}^{l_n} \mathbf{P}(D_{n,j_n}) = \mathbf{P}(B_1) \cdots \mathbf{P}(B_n).$$

Теорема доказана. Отметим, что по существу нам не требуется конечность m_i. $\square$

Определение 3.9. Алгеброй, порожденной случайной величиной X, называют алгебру, порожденную разбиением $\{X = x\}$ при всех $x \in \mathcal{X}$. Будем обозначать такую алгебру $\sigma(X)$.

Аналогичным образом определяется алгебра $\sigma(\vec{X})$, порожденная вектором $\vec{X}$.

Определение 3.10. Случайная величина X называется измеримой относительно алгебры $\mathcal{A}$, если

$$\{\omega : X(\omega) = x\} \in \mathcal{A}$$

при всех $x \in \mathcal{X}$. Иначе говоря, порожденная X алгебра должна содержаться в $\mathcal{A}$.

При этом событие $\{\omega : X(\omega) \in B\}$ при любом $B \subset \mathbb{R}$ лежит в $\mathcal{A}$, поскольку является объединением конечного или счетного числа событий $\{\omega : X(\omega) = x\}$ по $x \in B$.

Лемма 3.1. *Пусть f — функция из $\mathcal{X}$ в $\mathbb{R}$. Тогда величина $f(X)$ измерима около $\sigma(X)$.*

Доказательство. Доказательство вытекает из соотношения

$$\{\omega : f(X(\omega)) \in B\} = \{\omega : X(\omega) \in f^{-1}(B)\} \in \sigma(X),$$

где $f^{-1}(B) = \{x : f(x) \in B\}$ — полный прообраз множества. □

3.6 Независимые случайные величины

Определение 3.11. Величины $X_1, \ldots, X_n$, заданные на одном вероятностном пространстве $(\Omega, \mathcal{F}, \mathbf{P})$, называются независимыми, если для любых множеств $B_1, \ldots, B_n$ из $\mathbb{R}$

$$\mathbf{P}(\omega : X_1(\omega) \in B_1, \ldots, X_n(\omega) \in B_n) = \mathbf{P}(\omega : X_1(\omega) \in B_1) \cdots \mathbf{P}(\omega : X_n(\omega) \in B_n)$$

Лемма 3.2. *Следующие три утверждения равносильны:*

1. *Величины $X_1, \ldots, X_n$ независимы;*
2. *Алгебры $\sigma(X_1), \ldots, \sigma(X_n)$ независимы;*
3. *Для любого $\vec{x}$*
$$\mathbf{P}_{\vec{X}}(\vec{x}) = \mathbf{P}_{X_1}(x_1) \cdots \mathbf{P}_{X_n}(x_n).$$

Доказательство. Первое утверждение равносильно второму напрямую в силу определения независимости случайных величин. Третье равносильно второму в силу теоремы 3.1, поскольку алгебры $\sigma(X_i)$ порождаются разбиением $\{X_i = x\}$, $x \in \mathbb{R}$. □

Таким образом, проверять независимость можно на основе функций масс.

Вопрос 3.4. Вектор (X, Y) имеет таблицу распределения, приведенную ниже. Зависимы ли X, Y? X и Y^2?

	1	2
0	0.2	0.6
2	0.05	0.15

Пример 3.11. Бернуллиевские величины $X_1, \ldots, X_n$ являются независимыми, поскольку

$$\mathbf{P}_{\vec{X}}(x_1, \ldots, x_n) = p_1^{x_1 + \cdots + x_n}(1 - p_1)^{n - x_1 - x_2 - \cdots - x_n} = p_1^{x_1}(1 - p_1)^{1 - x_1} \cdots p_n^{x_n}(1 - p_n)^{1 - x_n} = \mathbf{P}_{X_1}(x_1) \cdots \mathbf{P}_{X_n}(x_n).$$

Это вполне естественно, ведь если эксперименты независимы, то и связанные с ними величины должны быть независимы.

Лемма 3.3. *Пусть*

$$\Omega = \Omega_1 \times \cdots \times \Omega_n, \quad \mathcal{F} = 2^{\Omega}, \quad \mathbf{P}((\omega_{1,i_1}, \ldots, \omega_{n,i_n})) = \mathbf{P}_1(\omega_{1,i_1}) \cdots \mathbf{P}_n(\omega_{n,i_n})$$

прямое произведение вероятностных пространств. Пусть набор величин $X_1, \ldots, X_n$ таков, что $X_i(\omega_1, \ldots, \omega_n)$ зависит только от ω_i при любом i. Тогда $X_1, \ldots, X_n$ независимы.

Доказательство. При каждом $i \in \{1, \ldots, n\}$ рассмотрим алгебры $\mathcal{A}_i$, где $\mathcal{A}_i$ порождена разбиением $\Omega_1 \times \cdots \times \Omega_{i-1} \times \{\omega_{i,j}\} \times \Omega_{i+1} \times \cdots \times \Omega_n$ при $j = 1, \ldots, n_i$.

Поскольку в силу определения прямого произведения указанные разбиения независимы, алгебры $\mathcal{A}_i$ также независимы. Но тогда независимы и $\sigma(X_i)$, поскольку $\sigma(X_i) \subset \mathcal{A}_i$ при любом $i \in \{1, \ldots, n\}$. $\square$

Пример 3.12. Компоненты полиномиального вектора $(X_1, \ldots, X_n) \sim Polynomial(N, p_1, \ldots, p_n)$ зависимы, если хотя бы два из $p_1, \ldots, p_n$ положительны. Пусть, скажем, $p_1 > 0$, $p_2 > 0$. Тогда

$\mathbf{P}(\omega : X_1(\omega) = N) = p_1^N$, $\mathbf{P}(\omega : X_2(\omega) = N) = p_2^N$, $\mathbf{P}(\omega : X_1(\omega) = N, X_2(\omega) = N) = 0 \neq \mathbf{P}(\omega : X_1(\omega) = N, X_2(\omega) = N)$.

Лемма 3.4. *Пусть величины $X_1, \ldots, X_n$ независимы, $f_1, \ldots, f_n$ — некоторые функции. Тогда $f_1(X_1), \ldots, f_n(X_n)$ независимы.*

Доказательство. Доказательство вытекает из того, что $\sigma(f_i(X_i))$ содержатся в $\sigma(X_i)$ ввиду леммы 3.1. $\square$

3.7 Независимость случайных векторов

Определение 3.12. Векторы $\vec{X}_1, \ldots, \vec{X}_n$ размерностей $d_1, \ldots, d_n$, соответственно, заданные на одном вероятностном пространстве $(\Omega, \mathcal{F}, \mathbf{P})$, называются независимыми, если для любых множеств $B_1 \in \mathbb{R}^{d_1}, \ldots, B_n \in \mathbb{R}^{d_n}$

$$\mathbf{P}(\omega : \vec{X}_1(\omega) \in B_1, \ldots, \vec{X}_n(\omega) \in B_n) = \mathbf{P}(\omega : \vec{X}_1(\omega) \in B_1) \cdots \mathbf{P}(\omega : \vec{X}_n(\omega) \in B_n)$$

Лемма 3.5. *Следующие три утверждения равносильны:*

1. *Векторы $\vec{X}_1, \ldots, \vec{X}_n$ независимы;*

2. *Алгебры $\sigma(\vec{X}_1), \ldots, \sigma(\vec{X}_n)$ независимы;*

3. *Для любых $\vec{x} = (\vec{x}_1, \ldots, \vec{x}_n)$*
$$\mathbf{P}_{\vec{X}}(\vec{x}) = \mathbf{P}_{\vec{X}_1}(\vec{x}_1) \cdots \mathbf{P}_{\vec{X}_n}(\vec{x}_n).$$

Доказательство совершенно аналогично лемме 3.2.

Пример 3.13. Пусть X_1, X_2 — схема Бернулли. Тогда векторы $(X_1, 1-X_1)$ и $(X_2, 1-X_2)$ независимы, поскольку

$$\mathbf{P}((X_1, 1 - X_1) = (i, 1 - i), (X_2, 1 - X_2) = (j, 1 - j)) = \mathbf{P}(X_1 = i)\mathbf{P}(X_2 = j) = \mathbf{P}((X_1, 1 - X_1) = (i, 1 - i))\mathbf{P}((X_2, 1 - X_2) = (j, 1 - j)).$$

Отметим также аналог леммы 3.4 для векторов.

Лемма 3.6. *Пусть величины $\vec{X}_1, \ldots, \vec{X}_n$ независимы, $f_1, \ldots, f_n$ — некоторые функции. Тогда $f_1(\vec{X}_1), \ldots, f_n(\vec{X}_n)$ независимы.*

3.8 Ответы на вопросы

1. Первая из них равномерная, вторая геометрическая, четвертая биномиальная, а вот третья — гипергеометрическая.

2. Чтобы найти $\mathbf{P}(X = Y)$, сложим $\mathbf{P}(X = Y = 2)$ и $\mathbf{P}(X = Y = 3)$. Итого $\mathbf{P}(X = Y) = 0.05 + 0.1 = 0.15$.

3. $\emptyset, \{0\}, \{1\}, \{0, 1\}, \{2, 3, 4\}, \{0, 2, 3, 4\}, \{1, 2, 3, 4\}, \{0, 1, 2, 3, 4\}$. Первая не содержит $\{2\}$ — разности $\{0, 1, 2\}$ и $\{0, 1\}$, вторая $\{4\}$ — дополнения до $\{0, 1, 2, 3, \}$, четвертая — самого Ω.

4. Прямо по определению X и Y независимы. У X и Y^2 таблица распределения та же самая. Отличие лишь в подписи второй строки, но это не влияет на независимость.

4 Математическое ожидание в дискретном случае

4.1 Определение математического ожидания

Пусть $(\Omega, 2^{\Omega}, \mathbb{P})$ — вероятностное пространство, X — случайная величина на нем. Мы хотим определить среднее значение случайной величины X, заданной на этом пространстве. Опять же, я сделаю это последовательно, демонстрируя, что у нас нет альтернативы данному определению. При желании вы можете пропустить часть 4.1 и сразу перейти к 4.2.

4.1.1 Классический случай

Если на Ω вероятность задана классическим образом, то логично определять среднее значение соотношением

$$\mathbf{E}X = \frac{1}{N}\sum_{k=1}^{N} X(\omega_k) = \sum_{k=1}^{N} X(\omega_k)\mathbf{P}(\omega_k).$$

Физически это центр масс системы точек $X(\omega_1), \dots, X(\omega_N)$ с равными массами.

4.1.2 Случай рациональных вероятностей

Если вероятности исходов Ω рациональны, то аналогично тому, как это было сделано на второй лекции, мы можем ввести новое пространство $\widetilde{\Omega}$ с классически заданной вероятностью $\widetilde{P}$, где каждый исход ω_k будет состоять из нескольких частей $\widetilde{\omega}_{k,1}$,…,$\widetilde{\omega}_{k,n_k}$. Рассмотрим случайную величину $\widetilde{X}$ на $\widetilde{\Omega}$, заданную соотношением

$$\widetilde{X}(\widetilde{\omega}_{k,i}) = X(\omega_k)$$

при всех k, i. Величина $\widetilde{X}$ описывает ту же самую физическую характеристику, что и X, просто заданную на другом вероятностном пространстве. Поскольку мы решили, что

$$\mathbf{E}\widetilde{X} = \sum_{k=1}^{N}\sum_{i=1}^{n_k} \widetilde{X}(\widetilde{\omega}_{k,i})\widetilde{\mathbf{P}}(\widetilde{\omega}_{k,i}) = \sum_{k=1}^{N} X(\omega_k)\sum_{i=1}^{n_k}\widetilde{\mathbf{P}}(\widetilde{\omega}_{k,i}) = \sum_{k=1}^{N} X(\omega_k)\mathbf{P}(\omega_k),$$

то и $\mathbf{E}X$ естественно определять тем же соотношением. Физически это все еще центр масс системы точек $X(\omega_1)$,…,X(ω_N) с разными массами.

4.1.3 Случай общего конечного пространства

Если вероятности не являются рациональными, то фиксируем $\varepsilon > 0$ и рассмотрим пространство $\widetilde{\Omega}$, состоящее из исходов $\widetilde{\omega}_i$, $i \le N+1$, причем

$$\widetilde{\mathbf{P}}(\widetilde{\omega}_i) \le \mathbf{P}(\omega_i) \le \widetilde{\mathbf{P}}(\widetilde{\omega}_i) + \frac{\varepsilon}{N}, \quad \widetilde{\mathbf{P}}(\widetilde{\omega}_i) \in \mathbb{Q},\ i \le N, \quad \widetilde{\mathbf{P}}(\widetilde{\omega}_{N+1}) < \varepsilon$$

Пусть $\min X(\omega_i) = m_-$, $\max X(\omega_i) = m_+$ Рассмотрим две случайных величины

$$\widetilde{X}_1(\widetilde{\omega}_i) = \widetilde{X}_2(\widetilde{\omega}_i) = X_1(\omega_i),\ i \le N, \quad \widetilde{X}_1(\widetilde{\omega}_{N+1}) = m_-, \quad \widetilde{X}_2(\widetilde{\omega}_{N+1}) = m_+.$$

Можно представить себе, что мы отрезали по кусочку от каждого исхода ω_i, чтобы сделать их рациональными, а затем оставшиеся "обрезки"собрали в новый исход. При этом новые величины на обрезанных кусочках мы положили теми же, что и раньше, а на обрезках первую величину сделали минимальным из того, чем была X, а вторую — максимальным. Тогда естественно считать, что величина

$$\mathbf{E}\widetilde{X}_1 = \sum_{k=1}^{N} \widetilde{X}_1(\widetilde{\omega}_k)\mathbf{P}(\widetilde{\omega}_1) + m_-\mathbf{P}(\widetilde{\omega}_{N+1}) = \sum_{k=1}^{N} X(\omega_k)\widetilde{\mathbf{P}}(\widetilde{\omega}_1) + m_-\mathbf{P}(\widetilde{\omega}_{N+1})$$

меньше, чем должно быть $\mathbf{E}X$, а

$$\mathbf{E}\widetilde{X}_2 = \sum_{k=1}^{N} \widetilde{X}_1(\widetilde{\omega}_k)\mathbf{P}(\widetilde{\omega}_1) + m_+\mathbf{P}(\widetilde{\omega}_{N+1}) = \sum_{k=1}^{N} X(\omega_k)\widetilde{\mathbf{P}}(\widetilde{\omega}_k) + m_+\mathbf{P}(\widetilde{\omega}_{N+1})$$

больше. При этом

$$\mathbf{E}\widetilde{X}_1 \le \sum_{k=1}^{N} X(\omega_k)\mathbf{P}(\omega_k) + (\max(0, m_+) - \min(0, m_-))\varepsilon,\ \mathbf{E}\widetilde{X}_2 \ge \sum_{k=1}^{N} X(\omega_k)\mathbf{P}(\omega_k) - (\max(0, m_+) - \min(0, m_-))\varepsilon.$$

Следовательно, $\mathbf{E}X$ должно быть в диапазоне

$$\sum_{k=1}^{N} X(\omega_k)\mathbf{P}(\omega_k) \pm c\varepsilon$$

для $c = \max(0, m_+) - \min(0, m_-)$ и всех $\varepsilon > 0$. В силу произвольности ε мы должны определить среднее как сумму

$$\mathbf{E}X = \sum_{k=1}^{N} X(\omega_k)\mathbf{P}(\omega_k).$$

Это все еще центр масс.

4.1.4 Случай общего пространства и ограниченных величин

В случае, если Ω бесконечно, а X ограничена и принимает значения из некоторого диапазона $[m_-, m_+]$, фиксируем $\varepsilon > 0$ и найдем такое N, что

$$\sum_{k=N+1}^{\infty} \mathbf{P}(\omega_k) < \varepsilon.$$

Тогда рассмотрим $\widetilde{\Omega} = \{\widetilde{\omega}_1, \ldots, \widetilde{\omega}_{N+1}\}$, где

$$\widetilde{\omega}_i = \omega_i,\ i \le N, \quad \widetilde{\omega}_{N+1} = \bigcup_{i=N+1}^{\infty} \omega_i.$$

Введем $\widetilde{X}_1$, $\widetilde{X}_2$ тем же образом, что и в прошлом подпункте. По тем же причинам в этом случае мы должны определять $\mathbf{E}X$ как сумму ряда

$$\mathbf{E}X = \sum_{k=1}^{\infty} X(\omega_k)\mathbf{P}(\omega_k).$$

Отметим, что ряд сходится абсолютно, поскольку мажорируется рядом

$$\sum_{k=1}^{\infty} \max(-m_-, m_+)\mathbf{P}(\omega_k) = \max(-m_-, m_+).$$

4.1.5 Общий случай и неотрицательные величины

Рассмотрим величины

$$X_n(\omega) = \begin{cases} X(\omega), & X(\omega) \in [0, n], \\ n, & X(\omega) \ge n. \end{cases}$$

Тогда $\mathbf{E}X_n$ — монотонно неубывающая последовательность, причем

$$\mathbf{E}X_n \le \sum_{k=1}^{\infty} X(\omega_k)\mathbf{P}(\omega_k).$$

Если ряд
$$\sum_{k=1}^{\infty} X(\omega_k)\mathbf{P}(\omega_k)$$
сходится, то при любом $\varepsilon > 0$ и достаточно больших N
$$\sum_{k=1}^{N} X(\omega_k)\mathbf{P}(\omega_k) > \varepsilon$$
Значит при $n = \max(X(\omega_1), \dots, X(\omega_N))$ получаем
$$\sum_{k=1}^{\infty} X(\omega_k)\mathbf{P}(\omega_k) - \varepsilon \le \mathbf{E}X_n \le \sum_{k=1}^{\infty} X(\omega_k)\mathbf{P}(\omega_k),$$
поскольку все члены ряда с 1 по N вошли в сумму в $\mathbf{E}X_n$. Значит при всех $\varepsilon > 0$
$$\sum_{k=1}^{\infty} X(\omega_k)\mathbf{P}(\omega_k) - \varepsilon \le \lim_{n\to\infty} \mathbf{E}X_n \le \sum_{k=1}^{\infty} X(\omega_k)\mathbf{P}(\omega_k),$$
то есть $\mathbf{E}X_n$ стремится к указанной сумме ряда. Но тогда логично определять $\mathbf{E}X$ как ту же сумму, что и прежде

4.1.6 Общий случай

В общем случае потребуем, чтобы ряд $\sum_{k=1}^{\infty} X(\omega_k)\mathbf{P}(\omega_k)$ сходился абсолютно. Тогда естественно положить
$$\mathbf{E}X = \mathbf{E}X^+ - \mathbf{E}X^-,$$
где
$$X^+(\omega) = \begin{cases} X(\omega), & X(\omega) > 0, \\ 0, & X(\omega) \le 0, \end{cases} \qquad X^-(\omega) = \begin{cases} -X(\omega), & X(\omega) < 0, \\ 0, & X(\omega) \ge 0. \end{cases}$$
Тогда
$$\mathbf{E}X = \sum_{\omega: X(\omega)>0} X(\omega)\mathbf{P}(\omega) + \sum_{\omega: X(\omega)<0} X(\omega)\mathbf{P}(\omega) = \sum_{k} X(\omega_k)\mathbf{P}(\omega_k),$$
где мы переставляем члены ряда в силу абсолютной сходимости.

Если же ряд не сходится абсолютно, то возможны три ситуации:

1. если $\mathbf{E}X^+ = +\infty,\ \mathbf{E}X^- < +\infty$, то логично положить $\mathbf{E}X = +\infty$,
2. если $\mathbf{E}X^+ < +\infty,\ \mathbf{E}X^- = +\infty$, то логично положить $\mathbf{E}X = -\infty$,
3. если $\mathbf{E}X^+ = +\infty,\ \mathbf{E}X^- = +\infty$, то $\mathbf{E}X$ не определено.

4.2 Математическое ожидание случайной величины и функции от случайной величины

Для краткости будем писать $\mathbf{P}(X = x)$ вместо $\mathbf{P}(\omega : X(\omega) = x)$.

Итак, мы приходим к определению:

Определение 4.1. Математическим ожиданием $\mathbf{E}X$ называют сумму ряда
$$\sum_{k=1}^{\infty} X(\omega_k)\mathbf{P}(\omega_k),$$
если ряд сходится абсолютно.

Если ряд не сходится абсолютно, то используют приведенную выше схему 1.–3.

Вопрос 4.1. Какое математическое ожидание имеет число орлов при двух бросаниях монеты с вероятностью орла 1/3?

При работе с отдельной случайной величиной удобнее оперировать с ее распределением напрямую, а не опираться на вероятностное пространство. Пусть X принимает значения x_m с вероятностями p_m. Тогда

$$\mathbf{E}X = \sum_{m=1}^{\infty} x_m p_m = \sum_{m=1}^{\infty} x_m \mathbf{P}(X = x_m).$$

Опять же физически эта величина представляет собой центр системы масс p_i, расположенных в точках x_i.

Отметим, что математическое ожидание в силу приведенной формулы есть параметр распределения (то есть зависит только от распределения верятностей X, но не от вероятностного пространства).

Докажем приведенное соотношение. Оно является частным случае более общей формулы:

Лемма 4.1. *Пусть* $g : \mathbb{R}^m \to \mathbb{R}$*, величины* $X_1, \ldots, X_m$ *принимают значения* $x_{1,1}, \ldots, x_{1,n_1}$, $x_{2,1}, \ldots, x_{2,n_2}, \ldots, x_{m,1}, \ldots, x_{m,n_m}$. *Тогда*

$$\mathbf{E}g(X_1, \ldots, X_m) = \sum_{i_1,\ldots,i_m} g(x_{1,i_1}, \ldots, x_{m,i_m}) \mathbf{P}(X_1 = x_{1,i_1}, \ldots, X_m = x_{m,i_m}),$$

если указанная сумма сходится абсолютно.

Доказательство. Рассмотрим $A_{i_1,\ldots,i_m} = \{\omega : X_1(\omega) = x_{1,i_1}, \ldots, X_m(\omega) = x_{m,i_m}\}$. Тогда

$$\sum_{k=1}^{\infty} g(X_1(\omega_k), \ldots, X_m(\omega_k)) \mathbf{P}(\omega_k) = \sum_{i_1,\ldots,i_m} \sum_{\omega \in A_{i_1,\ldots,i_m}} g(x_{1,i_1}, \ldots, x_{m,i_m}) \mathbf{P}(\omega) = \sum_{i_1,\ldots,i_m} g(x_{1,i_1}, \ldots, x_{m,i_m}) \mathbf{P}(A_{i_1,\ldots,i_m}).$$

□

В частности, для подсчетов математического ожидания $\mathbf{E}g(X)$ не требуется искать распределение величины $g(X)$, а можно работать с исходным распределением.

Вопрос 4.2. Чему равно математическое ожидание $(X + 1)^2 + 2$, где X имеет распределение Бернулли с параметром 1/3?

Отметим еще один удобный способ нахождения математического ожидания в случае целочисленной неотрицательной случайной величины. Для этого заметим, что функция

$$\varphi_X(s) = \sum_{k=0}^{\infty} p_k s^k$$

является суммой степенного ряда. Ряд сходится в $s = 1$, а значит радиус сходимости не меньше единицы. В таком случае если ряд

$$\sum_{k=1}^{\infty} k p_k s^{k-1}$$

сходится при $s = 1$, то это значение совпадает с $\varphi'(1)$. Значит, можно найти функцию φ (ее называют производящей функция распределения вероятностей) и продифференцировать ее в точке $s = 1$ — это и будет математическое ожидание.

4.3 Свойства математического ожидания

4.3.1 Основные свойства

1. $\mathbf{E}c = c$, где c — вещественное число, $\mathbf{E}I_A = \mathbf{P}(A)$, где I_A — индикатор события A.

2. Математическое ожидание линейно. Для любых вещественных a, b и любых X, Y с конечным математическим ожиданием

$$\mathbf{E}(aX + bY) = a\mathbf{E}X + b\mathbf{E}Y.$$

С точки зрения определения через распределение этот факт может показаться удивительным, ведь распределение $X+Y$ вообще не определяется распределениями X и Y. Распределение — нет, а вот математическое ожидание — вполне себе:

$$\mathbf{E}(aX + bY) = \sum_{\omega}(aX(\omega) + bY(\omega))\mathbf{P}(\omega) = a\sum_{\omega} X(\omega)\mathbf{P}(\omega) + b\sum_{\omega} Y(\omega)\mathbf{P}(\omega) = a\mathbf{E}X + b\mathbf{E}Y.$$

3. Если $X \geq 0$ и $\mathbf{E}X = 0$, то $\mathbf{P}(X = 0) = 1$.

4. Если $X \geq 0$, то $\mathbf{E}X \geq 0$.
Доказательство очевидно вытекает из любой из формул для математического ожидания.

5. Если $X \geq Y$, то $\mathbf{E}X \geq \mathbf{E}Y$.
Доказательство получается переходом к $Z = X - Y$ и использованием предыдущего свойства.

6. $|\mathbf{E}X| \leq \mathbf{E}|X|$, причем $\mathbf{E}X$ конечно тогда и только тогда, когда $\mathbf{E}|X|$ конечно.
Само неравенство является следствием предыдущего свойства, поскольку $-|X| \leq X \leq |X|$. Равносильность конечности $\mathbf{E}X$ и $\mathbf{E}|X|$ вытекает из того, что $|X| = X^+ + X^-$.

7. Если X, Y независимы и имеют конечное математическое ожидание, то $\mathbf{E}XY = \mathbf{E}X\mathbf{E}Y$.
Доказательство вытекает из леммы 4.1, откуда

$$\mathbf{E}XY = \sum_{i,j} x_i y_j \mathbf{P}(X = x_i, Y = y_j) = \sum_{i,j} x_i y_j \mathbf{P}(X = x_i)\mathbf{P}(Y = y_j) = \sum_{i} x_i \mathbf{P}(X = x_i) \sum_{j} y_j \mathbf{P}(Y = y_j),$$

где в последней формуле мы воспользовались абсолютной сходимостью обоих рядов, вытекающей из существования математических ожиданий. Свойство доказано.

4.3.2 Неравенства для математических ожиданий

8. Неравенство Йенсена.
Пусть f — выпуклая функция. Тогда $\mathbf{E}f(X) \geq f(\mathbf{E}X)$, если оба математических ожидания существуют.

Доказательство. В силу выпуклости функции f

$$f(x) \geq f(a) + c(x - a)$$

при любом a, некотором c и всех X. Следовательно,

$$f(X) \geq f(\mathbf{E}X) + c(X - \mathbf{E}X)$$

при всех ω. Применяя к обеим частям неравенства математическое ожидание, получаем

$$\mathbf{E}f(X) \geq f(\mathbf{E}X).$$

□

9. Неравенство Ляпунова.
Пусть $p > q > 0$, $\mathbf{E}|X|^p$ конечно. Тогда

$$(\mathbf{E}|X|^q)^{1/q} \leq (\mathbf{E}|X|^p)^{1/p}.$$

В частности, если $\mathbf{E}|X|^p < \infty$, то $\mathbf{E}|X|^q < \infty$ при всех $0 \leq q \leq p$.

Доказательство. Пусть $Y = |X|^q$, $f(x) = x^{p/q}$, тогда в силу неравенства Йенсена

$$(\mathbf{E}|X|^q)^{p/q} = f(\mathbf{E}Y) \leq \mathbf{E}f(Y) = \mathbf{E}|X|^p,$$

что и требовалось доказать. □

10. Неравенство Гельдера.
Пусть $p, q \in (1, \infty)$, $1/p + 1/q = 1$, $\mathbf{E}|X|^p < \infty$, $\mathbf{E}|Y|^q < \infty$. Тогда

$$\mathbf{E}|XY| \le (\mathbf{E}|X|^p)^{1/p} (\mathbf{E}|Y|^q)^{1/q}.$$

Доказательство. Рассмотрим

$$\widetilde{X} = \frac{|X|^p}{\mathbf{E}|X|^p},\ \widetilde{Y} = \frac{|Y|^p}{\mathbf{E}|Y|^p}.$$

При этом $\mathbf{E}\widetilde{X} = \mathbf{E}\widetilde{Y} = 1$.

В силу неравенства Иенсена

$$\frac{1}{p} \cdot \ln x + \frac{1}{q} \cdot \ln y \le \ln\left(\frac{1}{p} \cdot x + \frac{1}{q} \cdot y\right)$$

при любых положительных x, y. Значит,

$$x^{1/p} y^{1/q} \le \frac{x}{p} + \frac{y}{q}.$$

Тогда

$$\frac{\mathbf{E}|XY|}{(\mathbf{E}|X|^p)^{1/p}(\mathbf{E}|Y|^q)^{1/q}} = \mathbf{E}\widetilde{X}^{1/p}\widetilde{Y}^{1/q} \le \mathbf{E}\left(\frac{\widetilde{X}}{p}\right) + \mathbf{E}\left(\frac{\widetilde{Y}}{q}\right) = \frac{1}{p} + \frac{1}{q} = 1,$$

что и требовалось доказать. □

11. Неравенство Коши-Буняковского.
Пусть $\mathbf{E}X^2$, $\mathbf{E}Y^2$ конечны. Тогда

$$\mathbf{E}|XY| \le (\mathbf{E}X^2)^{1/2}(\mathbf{E}Y^2)^{1/2}.$$

Это частный случай неравенства Гельдера при $p = q = 1/2$.

4.4 Дисперсия и ковариация

Математическое ожидание описывает среднее. Логично рассмотреть также дисперсию — величину, характеризующую разброс вокруг среднего:

Определение 4.2. Дисперсией называется величина $\mathbf{E}(X - \mathbf{E}X)^2$.

Можно вывести более удобную формулу для дисперсии

$$\mathbf{E}(X - \mathbf{E}X)^2 = \mathbf{E}(X^2 - 2X\mathbf{E}X + (\mathbf{E}X)^2) = \mathbf{E}X^2 - 2(\mathbf{E}X)^2 + (\mathbf{E}X)^2 = \mathbf{E}X^2 - (\mathbf{E}X)^2.$$

Вспоминая как считать $\mathbf{E}X^2$, мы видим, что

$$\mathbf{D}X = \sum_{k=0}^{\infty} x_k^2 \mathbf{P}(X = x_k) - \left(\sum_{k=0}^{\infty} x_k \mathbf{P}(X = x_k)\right)^2.$$

Для конечности дисперсии требуется условие $\mathbf{E}X^2 < +\infty$. Тогда в силу неравенства Ляпунова $\mathbf{E}X$ тоже конечно. Отметим, что X^2 — неотрицательная величина, поэтому ее математическое ожидание либо конечно, либо равно $+\infty$, но не может не существовать.

В случае целочисленной неотрицательной величины можно найти $\mathbf{D}X$ с помощью производящей функции $\varphi_X(s)$:

$$\varphi''(s) = \sum_{k=2}^{\infty} k(k-1)p_k s^{k-2}, \quad \varphi''(1) = \sum_{k=2}^{\infty} k(k-1)p_k = \mathbf{E}X^2 - \mathbf{E}X, \quad \mathbf{D}X = \varphi''(1) + \varphi'(1) - (\varphi'(1))^2$$

Перед тем как получить свойства дисперсии, введем еще одну удобную величину

Определение 4.3. Ковариацией величин X и Y называют

$$\operatorname{cov}(X,Y) = \mathbf{E}(X - \mathbf{E}X)(Y - \mathbf{E}Y) = \mathbf{E}XY - \mathbf{E}X\mathbf{E}Y.$$

Эквивалентность указанных определений доказывается также, как и для дисперсии.

4.5 Свойства ковариации и дисперсии

1. $\mathbf{D}X = \operatorname{cov}(X,X)$.
2. $\operatorname{cov}(X,Y) = \operatorname{cov}(Y,X)$.
3. $\operatorname{cov}(X,Y+c) = \operatorname{cov}(X,Y), \quad \mathbf{D}(X+c) = \mathbf{D}X$.
4. Если X, Y независимы, то $\operatorname{cov}(X,Y) = 0$ в силу свойства 7 математического ожидания.
5. $\operatorname{cov}(aX+bY,Z) = a\operatorname{cov}(X,Z) + b\operatorname{cov}(Y,Z)$, поскольку

$$\operatorname{cov}(aX+bY,Z) = \mathbf{E}(aX+bY-\mathbf{E}(aX+bY))(Z-\mathbf{E}Z) = a\mathbf{E}(X-\mathbf{E}X)(Z-\mathbf{E}Z) + b\mathbf{E}(Y-\mathbf{E}Y)(Z-\mathbf{E}Z).$$

6. $\mathbf{D}\,(aX) = a^2\mathbf{D}X$.
7. $\mathbf{D}(X+Y) = \mathbf{D}X + 2\operatorname{cov}(X,Y) + \mathbf{D}Y$ в силу свойства 5).
8. $\mathbf{D}(X_1 + \cdots + X_n) = \sum_{i=1}^n \mathbf{D}X_i + 2\sum_{i<j}\operatorname{cov}(X_i,X_j)$.
9. $|cov(X,Y)| \le \sqrt{\mathbf{D}X\mathbf{D}Y}$ в силу неравенства Коши-Буняковского.

Вопрос 4.3. **D**X=2. Чему равна $\mathbf{D}(5-2X)$?

Пример 4.1. Заметим, что в свойстве 4) независимость величин является достаточным, но не необходимым условием для некоррелированности (то есть равенства ковариации нулю). Скажем, у вектора (X,Y) с таблицей распределения

	- 1	0	1
-1	0	0.25	0
0	0.25	0	0.25
1	0	0.25	0

компоненты X и Y имеют математические ожидания 0, а произведение компонент XY и вовсе всегда равно 0 (этот вектор представляет собой выбор одной из вершины квадрата (1,0), (0,1), (-1,0),(0,-1) наугад). Значит, величины X, Y некоррелированы, хотя очевидно зависимы (скажем, если одна 0, то вторая 1).

Заметим, что в силу свойства 9) выполнено соотношение

$$\operatorname{corr}(X,Y) := \frac{\operatorname{cov}(X,Y)}{\sqrt{\mathbf{D}X\mathbf{D}Y}} \in [-1,1].$$

Равенство $\operatorname{corr}(X,Y) = 1$ возможно только в случае $Y - \mathbf{E}Y = c(X - \mathbf{E}X)$, где $c > 0$, то есть $Y = cX + b$ при некотором b. Аналогично $\operatorname{corr}(X,Y) = -1$ только при $Y = cX + b$ при некотором $c < 0$. При этом в силу свойства 4) для независимых величин $\operatorname{corr}(X,Y) = 0$.

Определение 4.4. Коэффициент $\operatorname{corr}(X,Y)$ называют коэффициентом корреляции величин X и Y.

Вопрос 4.4. Чему равен коэффициент корреляции для числа орлов и числа решек, выпавших при n бросаниях монеты?

Этот коэффициент в некотором смысле измеряет зависимость — для линейно зависимых величин он по модулю становится 1, для независимых он равен нулю. Впрочем, для зависимых, как мы видели выше, он также бывает нулевым, а, например, для X и X^2 он почти всегда меньше единицы. Можно сказать, что corr измеряет степень "линейной" зависимости $y = ax + b$.

4.6 Ответы на вопросы

1. $0 \cdot \frac{4}{9} + 1 \cdot 2\frac{2}{3}\frac{1}{3} + 2 \cdot \frac{1}{3^2} = \frac{2}{3}$.
2. $((0+1)^2 + 2) \cdot 2/3 + ((1+1)^2 + 2) \cdot 1/3 = 4$.
3. $\mathbf{D}(5 - 2X) = \mathbf{D}(-2X) = 4\mathbf{D}X = 8$.
4. Поскольку $Y = n - X$, то $\mathrm{corr}(X, Y) = -1$.

Обозначение	Формула для вероятности	Математическое ожидание	Дисперсия
Bern(p)	P(X=0) = 1-p, P(X=1)=p	p	p(1-p)
Binom(n,p)	$P(X = k) = C_n^k p^k (1-p)^{n-k}$	np	$np(1-p)$
Geom(p)	$P(X = k) = (1-p)^k p$	$(1-p)/p$	$(1-p)/p^2$
R$\{0,\ldots,N\}$	$P(X = k) = 1/(N+1)$	$N/2$	$(N+1)(N-1)/12$
Poiss(λ)	$P(X = k) = \lambda^k e^{-\lambda}/k!$	λ	λ
Hypergeom(N,D,n)	$P(X = k) = \frac{C_D^k C_{N-D}^{n-k}}{C_N^n}$	$\frac{nD}{N}$	$\frac{nD}{N}\left(1 - \frac{D}{N}\right)\frac{(N-n)}{(N-1)}$.
NegBinom(r,p)	$P(X = k) = C_{k+r-1}^k p^r (1-p)^k$	$r(1-p)/p$	$r(1-p)/p^2$.

5 Производящие функции распределений

5.1 Определение

Пусть X — неотрицательная целочисленная случайная величина.

Определение 5.1. Производящей функцией (п.ф.) случайной величины X называют функцию

$$\varphi_X(s) = \mathbf{E}s^X = \sum_{k=0}^{\infty} \mathbf{P}(X = k)s^k.$$

Второе равенство в определении п.ф. вытекает из леммы с прошлого семинара, согласно которой

$$\mathbf{E}g(X) = \sum_{k=0}^{\infty} \mathbf{P}(X = k)g(k).$$

Заметим, что п.ф. одна и та же у всех величин с заданным распределением, поэтому можно говорить о производящей функции распределения вероятностей.

Параметр s мы будем рассматривать принадлежащим отрезку $[0, 1]$. Он не несет никакого специального физического смысла, но введение этого параметра позволяет задать распределение целиком, поскольку зная функцию, к которой степенной ряд сходится на отрезке, мы можем восстановить коэффициенты ряда.

Вопрос 5.1. Какую п.ф. имеет бернуллиевское распределение?

- ps
- $1 - p + ps$
- s^p
- Ни один из ответов не является правильным.

Пример 5.1. 1. Константа c имеет п.ф. $\varphi(s) = \mathbf{E}s^c = s^c$.

2. Случайная величина с бернуллиевским распределением имеет п.ф. $\varphi(s) = \mathbf{E}s^X = (1-p) + ps$.
3. Случайная величина с биномиальным распределением имеет п.ф.

$$\phi(s) = \sum_{k=1}^{n} C_n^k p^k (1-p)^{n-k} s^k = (1 - p + ps)^n$$

4. Случайная величина с распределение Пуассона имеет п.ф.

$$\phi(s)=\sum_{k=0}^{\infty}s^k\frac{\lambda^k}{k!}e^{-\lambda}=e^{-\lambda}\sum_{k=0}^{\infty}\frac{(s\lambda)^k}{k!}=e^{\lambda(s-1)}.$$

5.2 Свойства п.ф. как функции

Поскольку

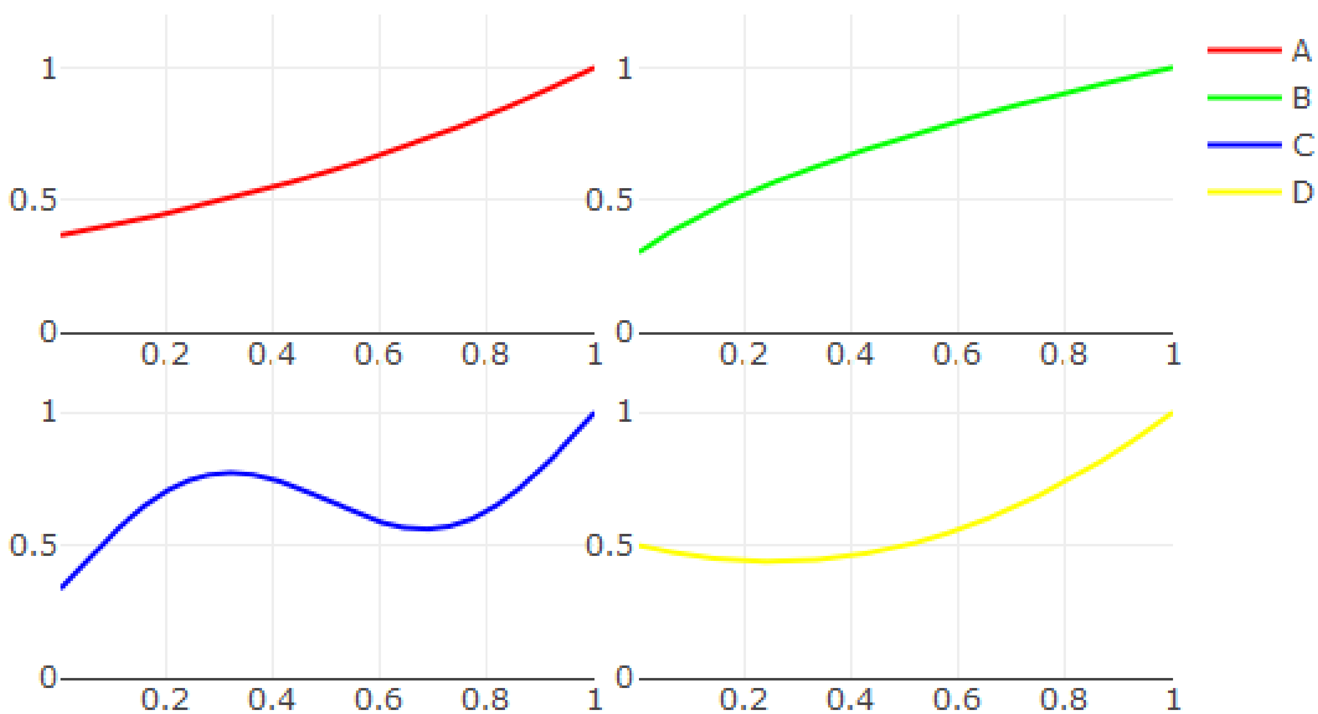
$$\varphi(s)=\sum_{k=0}^{\infty}p_k s^k$$

степенной ряд, абсолютно сходящийся в точке 1, он сходится при $|s|\le 1$. При этом радиус его сходимости может быть и больше 1, например, во всех примерах в прошлом подразделе ряд сходился на всей прямой. Мы будем рассматривать $s\in[0,1]$. Тогда функция обладает следующими свойствами:

1. $\varphi(1)=1,\ \varphi(s)\ge 0,\ s\ge 0.$
2. $0\le\varphi'(s)<\infty,\ s\in[0,1).$
3. $0\le\varphi^{(k)}(s)<\infty,\ k\ge 2,\ s\in[0,1).$

Рис. 7: Какой из графиков изображает п.ф.?

Вопрос 5.2.



Фактически, в свойствах 2-3 мы просто используем теорему о дифференцировании степенного ряда внутри круга сходимости. Отсюда п.ф. представляет собой неотрицательную монотонную выпуклую на $[0,1]$ бесконечно дифференцируемую внутри $(0,1)$ функцию.

5.3 Вероятностные свойства п.ф

Производящая функция не имеет естественного вероятностного смысла, но удобна для характеризации распределения. Мы уже обсуждали ее полезность для подсчета моментов, сформулируем это и несколько новых свойств этой функции.

1. В силу формулы Тейлора
$$\mathbf{P}(X=k)=\frac{\varphi^{(k)}(0)}{k!}.$$
Таким образом, мы можем найти распределение по п.ф.

2. Если $\mathbf{E}X^n$ конечно, то
$$\varphi^{(n)}(1)=\mathbf{E}X(X-1)\cdots(X-n+1).$$
Доказательство. При $s<1$
$$\varphi^{(n)}(s)=\sum_{k=0}^{\infty}\mathbf{P}(X=k)k(k-1)\cdots(k-n+1)s^{k-n}=:s^{-n}\sum_{k=0}^{\infty}a_ks^k$$
в силу теоремы о дифференцировании степенного ряда внутри круга сходимости. Воспользуемся теоремой Абеля

Теорема 5.1. *Пусть* $\sum_{n=0}^{\infty}a_nR^n$ *сходится,* $R>0$. *Тогда*
$$\sum_{n=0}^{\infty}a_nR^n=\lim_{x\to R-0}\sum_{n=0}^{\infty}a_nx^n.$$
Поскольку
$$\sum_{k=0}^{\infty}a_k=\sum_{k=0}^{\infty}k(k-1)\cdots(k-n+1)\mathbf{P}(X=k)<\sum_{k=0}^{\infty}k^n\mathbf{P}(X=k)<\infty,$$
то $\varphi^{(n)}(s)$ сходится при $s\to 1$ к к
$$\sum_{k=0}^{\infty}a_k=\mathbf{E}X(X-1)\cdots(X-n+1).$$
□

3. Если X_i — независимы, $i\le n$, то
$$\varphi_{X_1+\dots+X_n}(s)=\varphi_{X_1}(s)\cdots\varphi_{X_n}(s)$$
при всех $s\in[0,1]$.

Доказательство. Заметим, что
$$\varphi_{X_1+\dots+X_n}(s)=\mathbf{E}s^{X_1+\dots+X_n}=\mathbf{E}s^{X_1}\cdots s^{X_n}.$$
При любом s величины $s^{X_1},\dots,s^{X_n}$ независимы как функции от независимых величин. По свойству 7) математического ожидания
$$\mathbf{E}s^{X_1}\cdots s^{X_n}=\mathbf{E}s^{X_1}\cdots\mathbf{E}s^{X_n}=\varphi_{X_1}(s)\cdots\varphi_{X_n}(s).$$
□

Пример 5.2. Отсюда можно найти п.ф. биномиального распределения более простым путем. Биномиальная величина есть сумма н.о.р. бернуллиевских величин, откуда ее п.ф. есть произведение п.ф. бернуллиевских величин, то есть $\varphi_X(s)^n=(1-p+ps)^n$.

4. Пусть N — целая неотрицательная случайная величина, а распределение Y задано формулой
$$\mathbf{P}(Y=x)=\sum_{k=0}^{\infty}\mathbf{P}(X_1+...+X_k=x)\mathbf{P}(N=k),$$

где X_i — независимые одинаково распределенные величины. Иначе говоря, Y распределена как сумма X_i в случайном количестве N. Тогда
$$\varphi_Y(s) = \varphi_N(\varphi_X(s)).$$

Доказательство. Заметим, что

$$\varphi_Y(s) = \sum_{x=0}^{\infty}\sum_{k=0}^{\infty}\mathbf{P}(X_1+\ldots+X_k=x)\mathbf{P}(N=k)s^x =$$
$$\sum_{k=0}^{\infty}\mathbf{P}(N=k)\sum_{x=0}^{\infty}\mathbf{P}(X_1+\ldots+X_k=x)s^x = \sum_{k=0}^{\infty}\mathbf{P}(N=k)\varphi_{X_1+\ldots+X_k}(s) =$$
$$\sum_{k=0}^{\infty}\mathbf{P}(N=k)\varphi_{X_1}^k(s) = \varphi_N(\varphi_{X_1}(s)).$$

□

5. Теорема непрерывности для п.ф.

Определение 5.2. Будем говорить, что X_n сходится по распределению к X ($X_n \xrightarrow{d} X$), если $\mathbf{P}(X_n = k) \to \mathbf{P}(X = k)$ при всех $k \in \mathbb{N} \cup \emptyset$

Теорема 5.2. *$X_n \xrightarrow{d} X$, $n \to \infty$ тогда и только тогда, когда $\varphi_{X_n}(s) \to \varphi_X(s)$ при всех $s \in [0,1]$, $n \to \infty$.*

Доказательство. Докажем достаточность. Действительно, пусть $p_{n,k} = \mathbf{P}(X_n = k)$, $p_k = \mathbf{P}(X = k)$. При любом N и всех $s \in [0,1)$

$$|\varphi_{X_n}(s) - \varphi_X(s)| \le \sum_{k=0}^{\infty}|p_{n,k} - p_k|s^k \le \sum_{k=0}^{N-1}|p_{n,k} - p_k| + \sum_{k\ge N}s^k.$$

Пусть $\varepsilon > 0$. Выбирая N таким образом, что $\sum_{k>N} s^k < \varepsilon/2$, а M так, что при $n > M$ $|p_{n,k} - p_k| < \varepsilon/(2N)$ при $k < N$, получаем
$$|\varphi_{X_n}(s) - \varphi_X(s)| < \varepsilon$$
при $n > M$. В силу определения предела, это и есть утверждение, что при $s \in [0,1)$ верно

$$\varphi_{X_n}(s) \to \varphi_X(s),\ s \in [0,1).$$

Отметим, что при этом мы не используем то, что p_i в сумме дают 1. А вот при $s = 1$, используя указанный факт, мы видим, что все производящие функции равны единице, а значит утверждение о сходимости также верно.

Докажем необходимость методом от противного. Рассмотрим минимальное $k = l$ при котором $p_{n,k}$ не сходится к p_k. Выберем подпоследовательность $n_0(m)$, по которой $p_{n_0(m),l}$ сходится к $q \ne p_l$. Из $n_0(m)$ выделим подпоследовательность $n_1(m)$, для которой $p_{n_1(m),l+1}$ сходится к какому-то q_{l+1} и так далее. Рассмотрим последовательность $t_m = n_m(m)$. В силу определения при $m \to \infty$ выполнены соотношения $p_{n_m(m),i} \to p_i$ при $i < l$, $p_{n_m(m),l} \to q$, $p_{n_m(m),i} \to q_i$, $i \ge l+1$. Из доказанного в первой части при этом при всех $s \in [0,1)$
$$\sum_{i=0}^{\infty}p_{t_n,i}s^i \to \sum_{i=0}^{l-1}p_is^i + qs^l + \sum_{i=l+1}^{\infty}q_is^i.$$

Отметим, что мы не знаем, правда ли $p_0 + \cdots + p_{l-1} + q + q_{l+1} + \cdots = 1$, поэтому при $s = 1$ мы такого не утверждаем.

При некотором $s \in [0,1)$

$$\sum_{i=0}^{l-1}p_is^i + qs^l + \sum_{i=l+1}^{\infty}q_is^i \ne \sum_{i=0}^{l-1}p_is^i + p_ls^l + \sum_{i=l+1}^{\infty}p_is^i,$$

поскольку

$$q + \sum_{i=1}^{\infty} q_{l+i}s^i \neq p_l + \sum_{i=1}^{\infty} p_{l+i}s^i$$

при достаточно малых s. Но

$$\sum_{i=0}^{\infty} p_{t_n,i}s^i \to \sum_{i=0}^{\infty} p_i s^i.$$

Таким образом, эта последовательность имеет два различных предела, что приводит нас к противоречию. Значит, $p_{n,k}$ сходятся к p_k при всех k. □

5.4 Теорема Пуассона

В качестве применения теоремы непрерывности рассмотрим так называемую теорему Пуассона.

Определение 5.3. Схемой серий назовем набор таких случайных величин

$$X_{1,1},\ X_{2,1},\ X_{2,2}, \ldots, X_{n,1}, \ldots, X_{n,n}, \ldots,$$

что $X_{n,i}$ независимы при $i \leq n$ при любом n.

Теорема 5.3. *Пусть $X_{n,i} \sim Bern(p_{n,i})$ образуют схему серий, $p_n = \sum_{i=1}^{n} p_{n,i}$. Предположим, что $m_n = \max_{i\leq n} p_{n,i} \to 0$, $p_n \to \lambda$, $n \to \infty$. Тогда*

$$\mathbf{P}\left(\sum_{i=1}^{n} X_{n,i} = k\right) \to \mathbf{P}(Z = k),\ n \to \infty,$$

где $k \geq 0$, $Z \sim Poiss(\lambda)$.

Доказательство. Рассмотрим п.ф. $Y_n = X_{n,1} + \cdots + X_{n,n}$:

$$\varphi_{Y_n}(s) = \prod_{i=1}^{n} \varphi_{X_{n,i}}(s) = \prod_{i=1}^{n}(1 - p_{n,i} + p_{n,i}s) = \exp\left(\sum_{i=1}^{n} \ln\left(1 - p_{n,i}(1-s)\right)\right).$$

При этом в силу формулы Тейлора с остаточным членом в форме Лагранжа

$$\ln(1 - p_{n,i}(1-s)) = -p_{n,i}(1-s) - \theta_{n,i}p_{n,i}^2(1-s)^2,$$

где $\theta_{n,i} \in [0, 1/2]$, откуда

$$\sum_{i=1}^{n} \ln(1 - p_{n,i}(1-s)) = -(1-s)\sum_{i=1}^{n} p_{n,i} - (1-s)^2 \sum_{i=1}^{n} \theta_{n,i}p_{n,i}^2,$$

где

$$0 \leq \sum_{i=1}^{n} \theta_{n,i}p_{n,i}^2 \leq \frac{1}{2}m_n \sum_{i=1}^{n} p_{n,i} \to 0,\ n \to \infty.$$

Следовательно,

$$\sum_{i=1}^{n} \ln(1 - p_{n,i}(1-s)) \to -\lambda(1-s),$$

откуда

$$\varphi_{Y_n}(s) \to \exp(-\lambda(1-s)) = \varphi_Z(s),$$

что и требовалось доказать. □

Вопрос 5.3. Носители редкой болезни встречатся с вероятностью 0,0002. Среди 10000 найдется хотя бы один заболевший с вероятностью примерно а) e^{-2}, б) $1 - e^{-2}$, в) $e^{-2} + 2e^{-2}$, г) $1 - 2e^{-2}$?

5.5 Производящие функции случайных векторов

Предположим, что $\vec{X} = (X_1, \ldots, X_k)$ — случайный вектор, компоненты которого являются целочисленными неотрицательными величинами. Тогда мы можем задать распределение такого вектора с помощью

$$\varphi_{\vec{X}}(\vec{s}) = \mathbf{E} s_1^{X_1} \cdots s_k^{X_k}$$

Пример 5.3. Пусть X_1, X_2 — результаты бросков несимметричной монеты с вероятностью успеха p. Тогда

$$\varphi_{X_1,X_2}(s_1, s_2) = \mathbf{E} s_1^{X_1} s_2^{X_2} = p^2 s_1 s_2 + p(1-p)s_1 + p(1-p)s_2 + (1-p)^2 = (ps_1 + 1 - p)(ps_2 + 1 - p).$$

При этом

$$\varphi_{X_1,X_1}(s_1, s_2) = \mathbf{E} s_1^{X_1} s_2^{X_1} = ps_1 s_2 + (1-p),$$

а

$$\varphi_{X_1,1-X_1}(s_1, s_2) = \mathbf{E} s_1^{X_1} s_2^{1-X_1} = ps_1 + (1-p)s_2.$$

В каждом случае компоненты вектора имеют распределение Бернулли, а сами векторы при этом каждый раз распределены по-новому. Поэтому задать распределение вектора производящими функциями каждой из входящих в него величин нельзя, необходимо именно задавать их совместную производящую функцию.

Лемма 5.1. *Компоненты вектора $\vec{X}$ независимы тогда и только тогда, когда при всех $s_1, \ldots, s_k$*

$$\varphi_{\vec{X}}(\vec{s}) = \varphi_{X_1}(s_1) \cdots \varphi_{X_k}(s_k).$$

Доказательство. Достаточность независимости вытекает из того, что $s_1^{X_1} \ldots s_k^{X_k}$ являются независимыми величинами, а значит математическое ожидание их произведение есть произведение их математических ожиданий.

Для доказательства необходимости поглядим на наше тождество:

$$\sum_{l_1,\ldots,l_n} \mathbf{P}(X_1 = l_1, \ldots, X_k = l_k) s_1^{l_1} \ldots s_k^{l_k} = \sum_{l_1} \ldots \sum_{l_n} \mathbf{P}(X_1 = l_1) s_1^{l_1} \cdots \mathbf{P}(X_k = l_k) s_k^{l_k},$$

откуда, приравнивая коэффициенты при каждой степени $s_1^{l_1} \cdots s_k^{l_k}$, получаем

$$\mathbf{P}(X_1 = l_1, \ldots, X_k = l_k) = \mathbf{P}(X_1 = l_1) \cdots \mathbf{P}(X_k = l_k),$$

что и означает независимость. □

Основные свойства п.ф. в векторном случае сохраняются

1. Как и прежде $\varphi_{\vec{X}}(1, \ldots, 1) = 1$. Более того, при подстановке части $s_i = 1$, мы получим производящую функцию вектора с оставшимися компонентами, например,

$$\varphi_{X_1,X_2,X_3}(1, s_1, s_2) = \varphi_{X_2,X_3}(s_1, s_2), \quad \varphi_{X_1,X_2}(s, 1) = \varphi_{X_1}(s).$$

2. Как и прежде производные $\varphi_{\vec{X}}$ в нуле помогают вычислить функцию масс нашего вектора:

$$\mathbf{P}(X_1 = l_1, \ldots, X_k = l_k) = \frac{1}{l_1! \cdots l_k!} \frac{\partial^{l_1} \ldots \partial^{l_k}}{\partial s_1^{l_1} \ldots \partial s_k^{l_k}} \varphi_{\vec{X}}(0, 0, \ldots, 0).$$

Так, например,

$$\mathbf{P}(X_1 = 1, X_2 = 1, X_3 = 0) = \varphi''_{s_1,s_2}(0,0,0), \quad \mathbf{P}(X_1 = 2, X_2 = 1, X_3 = 0) = \frac{1}{2}\varphi'''_{s_1,s_1,s_2}(0,0,0).$$

Доказательство. Давайте посмотрим на дифференцирование φ по одной из переменных (для определенности s_1). Рассмотрим φ как функцию s_1, она является степенным рядом

$$\sum_{m_1=0}^{\infty} s_1^{m_1} b_{m_1}, \quad b_{m_1} = \sum_{m_2=0}^{\infty} \ldots \sum_{m_k=0}^{\infty} \mathbf{P}(X_1 = m_1, \ldots, X_k = m_k) s_2^{m_2} \cdots s_k^{m_k},$$

имеющим радиус сходимости не менее единицы при любых $s_2, \dots s_k \in [0,1]$. Значит, мы можем дифференцировать его под знаком суммы:

$$\left(\sum_{m_1=0}^{\infty} \dots \sum_{m_k=0}^{\infty} \mathbf{P}(X_1 = m_1, \dots, X_k = m_k) s_1^{m_1} \cdots s_k^{m_k}\right)'_{s_1} =$$
$$\sum_{m_1=0}^{\infty} \dots \sum_{m_k=0}^{\infty} \mathbf{P}(X_1 = m_1, \dots, X_k = m_k) m_1 s_1^{m_1-1} s_2^{m_2} \cdots s_k^{m_k}.$$

Аналогичным путем мы можем переставлять порядок суммирования (пользуясь абсолютной сходимостью рассматриваемого ряда) и дифференцировать сумму по следующей переменной и так далее:

$$\frac{\partial^{l_1} \dots \partial^{l_k}}{\partial s_1^{l_1} \dots \partial s_k^{l_k}} \varphi_{\vec{X}}(s_1, \dots, s_k) = \sum_{m_1=l_1}^{\infty} \dots \sum_{m_k=l_k}^{\infty} \mathbf{P}(X_1 = m_1, \dots, X_k = m_k) \times$$
$$m_1 \cdots (m_1 - l_1 + 1) s_1^{m_1 - l_1} \cdots m_k \cdots (m_k - l_k + 1) s_k^{m_k - l_k}.$$

Полагая $(s_1, \dots, s_k) = (0, \dots, 0)$, мы обнулим все слагаемых указанной суммы кроме слагаемого $(m_1, \dots, m_k) = (l_1, \dots, l_k)$, откуда и вытекает требуемое соотношение. $\square$

3. Как и прежде, производные $\varphi_{\vec{X}}$ в единице помогают вычислить моменты наших величин

$$\frac{\partial^{l_1} \dots \partial^{l_k}}{\partial s_1^{l_1} \dots \partial s_k^{l_k}} \varphi_{\vec{X}}(1, \dots, 1) = \mathbf{E} X_1(X_1-1) \cdots (X_1-l_1+1) X_2(X_2-1) \cdots (X_2-l_2+1) \cdots X_m(X_m-1) \cdots (X_m-l_m+1).$$

Такие характеристики называют смешанными факториальными моментами. Доказательство вытекает из представления прошлого раздела и теоремы Абеля.

Вопрос 5.4. Как выразить ковариацию X, Y через п.ф. $\varphi(s,t)$ вектора (X, Y)?

- $\varphi''_{s,t}(1,1) - \varphi'_s(1,1)\varphi'_t(1,1)$;
- $\varphi''_{s,t}(1,1) - \varphi''_{s,s}(1,1)\varphi''_{t,t}(1,1)$;
- $\varphi''_{s,t}(1,1) - \varphi'_s(1,0)\varphi'_t(0,1)$;
- $\varphi''_{s,t}(0,0) - \varphi'_s(0,0)\varphi'_t(0,0)$.

4. Производящая функция суммы независимых векторов $\vec{X}_1, \dots, \vec{X}_n$ есть произведение п.ф. каждого из векторов
$$\varphi_{\vec{X}_1 + \dots + \vec{X}_n}(s_1, \dots, s_k) = \varphi_{\vec{X}_1}(s_1, \dots, s_k) \cdots \varphi_{\vec{X}_n}(s_1, \dots, s_k).$$
Доказательство опять же вытекает из того, что функции
$$s_1^{X_{1,1}} \cdots s_k^{X_{1,k}}, \quad s_1^{X_{2,1}} \cdots s_k^{X_{2,k}}, \quad \dots, \quad s_1^{X_{n,1}} \cdots s_k^{X_{n,k}}$$
независимы как функции независимых векторов, а значит математическое ожидание их произведения (то есть п.ф. суммы наших векторов) есть произведение их математических ожиданий (то есть произведение п.ф. каждого из них).

5. Справедлива и теорема непрерывности, которую мы оставим без доказательства.
Последовательность векторов $\vec{X}_n$ удовлетворяет соотношению
$$\mathbf{P}(\vec{X}_n = \vec{x}) \to \mathbf{P}(\vec{X} = \vec{x}),\ n \to \infty,$$
при всех $\vec{x} \in \mathbb{R}^k$ тогда и только тогда, когда
$$\varphi_{\vec{X}_n}(\vec{s}) \to \varphi_{\vec{X}}(\vec{s}),\ n \to \infty,$$
при всех $\vec{s} \in [0,1]^k$.

Пример 5.4. Рассмотрим $\vec{\nu} = (\nu_1, \ldots, \nu_k)$ — полиномиальный случайный вектор $Polynom(n, p_1, \ldots, p_k)$, то есть вектор из количеств исходов при n независимых испытаниях с k возможными исходами, имеющими вероятности $p_1, \ldots, p_k$.

Мы знаем, что

$$\mathbf{P}(\nu_1 = m_1, \ldots, \nu_k = m_k) = \frac{n!}{m_1! \cdots m_k!} p_1^{m_1} \cdots p_k^{m_k},$$

однако, считать п.ф. данного вектора напрямую достаточно громоздко. Вместо этого отметим, что

$$(\nu_1, \ldots, \nu_k) = \sum_{i=1}^{n} \vec{\eta}_i,$$

где $\vec{\eta}_i$ — вектор, описывающий итог i-го эксперимента, равный $(1, 0, \ldots, 0)$, если опыт закончился первым исходом, $(0, 1, 0, \ldots, 0)$ — если вторым исходом, $(0, \ldots, 0, 1)$ — если k-м исходом. Векторы $\vec{\eta}_i$ независимы, поскольку относятся к независимым испытаниям, и имеют п.ф.

$$f_{\vec{\eta}_i} = p_1 s_1^1 s_2^0 \cdots s_k^0 + p_2 s_1^0 s_2^1 s_3^0 \cdots s_k^0 + \cdots + p_k s_1^0 s_2^0 \cdots s_{k-1}^0 s_k^1 = p_1 s_1 + \cdots + p_k s_k.$$

Значит,

$$\varphi_{(\nu_1, \ldots, \nu_k)}(\vec{s}) = (p_1 s_1 + \cdots + p_k s_k)^n$$

5.6 Ответы на вопросы

1. $\phi(s) = (1-p)s^0 + ps^1 = 1 - p + ps$.
2. B, C невыпуклы, D не монотонна, а вот A — настоящая.
3. В силу теоремы Пуассона вероятность того, что не будет заболевших близка к вероятности того, что пуассоновская величина с параметром $10000 \cdot 0.0002 = 2$ равна 0, то есть к e^{-2}. Значит вероятность того, что заболевшие будут приближенно равна $1 - e^{-2}$.
4. $\mathrm{cov}(X, Y) = \mathbf{E}XY - \mathbf{E}X\mathbf{E}Y = \varphi''_{s,t}(1,1) - \varphi'_s(1,1)\varphi'_t(1,1)$.

Обозначение	Формула для вероятности	Производящая функция
Bern(p)	P(X=0) = 1-p, P(X=1)=p	(1-p)+ps
Binom(n,p)	$P(X = k) = C_n^k p^k (1-p)^{n-k}$	$(1 - p + ps)^n$
Geom(p)	$P(X = k) = (1-p)^k p$	$p/(1 - (1-p)s)$
$R\{0, 1, \ldots, N\}$	$P(X = k) = 1/(N+1)$	$(1 - s^{N+1})/(1 - s)$
Poiss(λ)	$P(X = k) = \lambda^k e^{-\lambda}/k!$	$e^{\lambda(s-1)}$
NegBinom(r,p)	$P(X = k) = C_{k+r-1}^k p^r (1-p)^k$	$p^r/(1 - (1-p)s)^r$.

6 Предельные теоремы в дискретном случае

6.1 Локальные предельные теоремы

Теорема Пуассона позволяет работать с бернуллиевскими величинами с редко случающимися успехами. Не менее, а то и более естественной является ситуация, когда вероятности успеха не являются малыми. Оказывается, что и здесь можно сформулировать аналогичные результаты:

Теорема 6.1 (Локальная теорема Муавра-Лапласа)**.** *Пусть X_i — н.о.р. Bern(p), $p \in (0,1)$. Тогда $S_n = X_1 + \ldots + X_n$ удовлетворяет соотношение*

$$\mathbf{P}(S_n = k) = \frac{1 + o(1)}{\sqrt{2\pi np(1-p)}} \exp\left(-\frac{(k-np)^2}{2np(1-p)}\right),$$

причем $o(1)$ равномерно мало по $|k - np| \le r_n$ при любом $r_n = o(n^{2/3})$.

Таким образом, можно аппроксимировать неудобно выражающиеся биномиальные вероятности (использующие, в частности, факториалы больших чисел и большие степени при больших n) с помощью более простой функции экспоненты. Иллюстрация приближения приведена в ролике по ссылке.

Доказательство. Воспользуемся явной формулой

$$\mathbf{P}(S_n = k) = C_n^k p^k (1-p)^{n-k}.$$

При этом $C_n^k = n!/(k!(n-k)!)$, где

$$n! \sim \sqrt{2\pi n}(n/e)^n, \quad (n-k)! \sim \sqrt{2\pi(n-k)}((n-k)/e)^{n-k}, \quad k! \sim \sqrt{2\pi k}(k/e)^k$$

при $k \to \infty$, $n-k \to \infty$. Следовательно,

$$C_n^k p^k (1-p)^{n-k} = \frac{1}{\sqrt{2\pi n \left(\frac{k}{n}\right)\left(1-\frac{k}{n}\right)}} \left(\frac{n-k}{n(1-p)}\right)^{-n+k} \left(\frac{k}{np}\right)^{-k}.$$

При этом

$$\left(\frac{n(1-p)}{n-k}\right)^{n-k} = \left(1+\frac{k-np}{n-k}\right)^{n-k} = \exp\left((n-k)\ln\left(1+\frac{k-np}{n-k}\right)\right) = \exp\left((k-np) - \frac{(k-np)^2}{2(n-k)} + o\left(\frac{(k-np)^3}{n^2}\right)\right),$$

$$\left(\frac{np}{k}\right)^k = \left(1+\frac{np-k}{k}\right)^k = \exp\left(k\ln\left(1+\frac{np-k}{k}\right)\right) = \exp\left((np-k) - \frac{(k-np)^2}{2k} + o\left(\frac{(k-np)^3}{n^2}\right)\right).$$

Следовательно,

$$\mathbf{P}(S_n = k) = \frac{(1+o(1))}{\sqrt{2\pi p(1-p)n}} \exp\left(-\frac{(k-np)^2}{2np(1-p)}\right), \ n \to \infty,$$

равномерно по $k/n \in [np - r_n, np + r_n]$. $\square$

Пример 6.1. Скажем, какая вероятность, что за 100 бросков симметричной монеты выпадет ровно 50 орлов? Теорема Муавра-Лапласа дает нам приближение

$$\frac{1}{\sqrt{2\pi \cdot 100 \cdot 0.5 \cdot 0.5}} \approx 0.0798.$$

Если же подсчитать вероятность по точной формуле

$$C_{100}^{50} \left(\frac{1}{2}\right)^{50} \left(\frac{1}{2}\right)^{50} \approx 0.0796.$$

Погрешность оказывается достаточно неплохой — около 0.0002.

Нам потребуется следующее определение:

Определение 6.1. Дискретная случайная величина X называется решетчатой с шагом d, если найдется такое a, что множество значений $\mathcal{X}$ нашей величины вложены в $a + d\mathbb{Z}$. Будем предполагать, что d — максимальное такое число.

Узнать шаг решетки достаточно просто — вычтем из всех значений одно из них и найдем самое такое d, что после деления на него полученные значения целые и имеют НОД 1.

Пример 6.2.

- Величина, принимающая значения $\{0, 1, 2\}$, решетчата с шагом 1. Вычтем из всех значений первое, получим $\{1, 2\}$. НОД этих чисел равен 1, откуда $d = 1$.
- Величина, принимающая значения $\{\pi, 3\pi, 5\pi\}$, решетчата с шагом 2π, поскольку $\{2\pi, 4\pi\}$ при делении на 2π переходят в $\{1, 2\}$ с НОД 1.
- Величина, принимающая значения $\{\dots, -5, -2, 1, 4, 7, \dots\}$, решетчата с шагом 3, покольку $\{\dots, -3, 0, 3, 6, 9\dots\}$ при делении на 3 переходят в числа с НОД 1.

- Величина, принимающая значения $\{0, 1, \pi\}$ нерешетчата, поскольку пару $\{1, \pi\}$ нельзя поделить на одно и то же число так, чтобы оба результата получились целыми.

Вопрос 6.1. Какой шаг решетки у величины со значениями $1/3, 1, 13/9$?

Оказывается, что справедлива более общая теорема, чем локальная теорема Муавра-Лапласа, которую мы оставим без доказательства:

Теорема 6.2 (Гнеденко)**.** *Пусть X_i, $i \le n$ — независимые одинаково распределенные решетчатые величины с шагом d и сдвигом a, $\mathbf{E}X_1 = mu$, $\mathbf{D}X_1 = \sigma^2$. Тогда*

$$\mathbf{P}(S_n = k) = \frac{d}{\sqrt{2\pi n}\sigma} \exp\left(-\frac{(k - n\mu)^2}{2n\sigma^2}\right) + \frac{o(1)}{\sqrt{n}},\ n \to \infty,$$

где $o(1)$ равномерно мало по $k \in an + d\mathbb{Z}$. Иначе говоря,

$$\sqrt{n} \sup_{k \in an + d\mathbb{Z}} \left| \mathbf{P}(S_n = k) - \frac{d}{\sqrt{2\pi n}\sigma} \exp\left(-\frac{(k - n\mu)^2}{2n\sigma^2}\right) \right| \to 0,\ n \to \infty.$$

Теорема Гнеденко позволяет существенно обобщить теорему Муавра-Лапласа — какое бы распределение случайных величин мы не взяли, сложив достаточное количество нерешетчатых величин с таким распределением мы получим одно и то же предельное выражение для $\mathbf{P}(S_n = k)$. Больше того, это выражение зависит от распределения только через среднее, дисперсию и шаг решетки, остальные параметры оказываются не столь существенными. Отметим, что хотя теорема сформулирована на всей прямой, по существу она полезна только при $k = \mu n + O(\sqrt{n})$, в противном случае мы просто узнаем, что $\mathbf{P}(S_n = k) = o(n^{-1/2})$.

Пример 6.3. Игрок в рулетку, поставивший три доллара на черное и один доллар на зеро, получает 2$ выигрыша с вероятностью 18/38, 32$ с вероятностью 1/38 и теряет 4$ в оставшемся случае. Как устроен выигрыш за 900 партий?
Найдем

$$\mu = 2 \cdot \frac{18}{38} + 32 \cdot \frac{1}{38} - \frac{4}{2} = \frac{68}{38} - 2 = -\frac{4}{19}.$$

При этом дисперсия будет

$$\sigma^2 = 2^2 \cdot \frac{18}{38} + 32^2 \cdot \frac{1}{38} + \frac{4^2}{2} - \left(\frac{4}{19}\right)^2 \approx 36.8,$$

откуда

$$\mathbf{P}(S_{900} = k) = \frac{1}{\sqrt{2\pi \cdot 900 \cdot 6}} \exp\left(-\frac{(k + 4 \cdot 900/19)^2}{2 \cdot 36 \cdot 900}\right) \approx \frac{1}{450} \exp\left(-\frac{1}{2}\left(\frac{k}{190} + 1\right)^2\right).$$

Таким образом, $\mathbf{P}(S_{900} = k)$ максимально (примерно 1/450) при $k = -190$, а затем убывает до $k = 190$ (и симметрично убывает до -570), где вероятности имеют порядок 1/3000. Таким образом, мы имеем более-менее реальные шансы остаться в плюсе, однако, скорее всего выиграем немного и в этом случае.

Вопрос 6.2. При 1000 подбрасываниях игрального кубика выпало в сумме X очков. Какая из вероятностей согласно локальной предельной теореме больше - что очков выпало 4000 или 3999?

6.2 Интегральная предельная теорема

Как мы увидели в последнем примере, нам было бы полезно просуммировать полученные вероятности по отрезку $[\mu n + a\sigma\sqrt{n}, \mu n + b\sigma\sqrt{n}]$. Оказывается, что справедлива следующая теорема:

Теорема 6.3. *Пусть X_i, $i \le n$ — независимые одинаково распределенные решетчатые величины, $\mathbf{E}X_1 = mu$, $\mathbf{D}X_1 = \sigma^2$. Тогда*

$$\mathbf{P}(S_n - \mu n \in [b\sigma\sqrt{n}, c\sigma\sqrt{n}]) \to \frac{1}{\sqrt{2\pi}} \int_b^c e^{-t^2/2} dt$$

при $n \to \infty$.

Доказательство. Заметим, что если решетка распределения S_n есть $S = an + d\mathbb{Z}$, то

$$\mathbf{P}(S_n - \mu n \in I) = \sum_{k \in S \cap I} \mathbf{P}(S_n = k),$$

где $I = [b\sigma\sqrt{n}, c\sigma\sqrt{n}]$. В силу теоремы Гнеденко

$$\mathbf{P}(S_n = k) = \frac{d(1+o(1))}{\sqrt{2\pi n}\sigma} \exp\left(-\frac{(k-\mu n)^2}{2n\sigma^2}\right),$$

причем $o(1)$ равномерно мало по $k \in S \cap I$. Значит,

$$\mathbf{P}(S_n - \mu n \in I) = \frac{1+o(1)}{\sqrt{2\pi}} \sum_{k \in S\cap I} \frac{d}{\sigma\sqrt{n}} \exp\left(-\frac{(k-\mu n)^2}{2n\sigma^2}\right) \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2} dt,$$

где последний переход вытекает из того, что рассматриваемая сумма является интегральной с шагом $d/(\sigma\sqrt{n})$. $\square$

Теорема 6.4 (Центральная предельная теорема)**.** *Пусть X_i, $i \le n$ — независимые одинаково распределенные решетчатые величины, $\mathbf{E}X_1 = \mu$, $\mathbf{D}X_1 = \sigma^2$. Тогда при всех x*

$$\mathbf{P}(S_n - \mu n \le x\sigma\sqrt{n}) \to \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt \tag{4}$$

при $n \to \infty$.

Данная теорема является частным случаем более общей теоремы 15.4. В действительности, решетчатость величин здесь (в отличие от локальных теорем) не играет никакой роли.

Доказательство. Заметим, что

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2/2} dt = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-s^2} ds = 1,$$

поскольку мы знаем, что интеграл Эйлера-Пуассона равен $\sqrt{\pi}$. Значит, при каждом $\varepsilon > 0$ найдется такое M, что

$$\frac{1}{\sqrt{2\pi}} \int_{-M}^{M} e^{-t^2/2} dt > 1 - \varepsilon, \quad \frac{1}{\sqrt{2\pi}} \int_{|t|>M} e^{-t^2/2} dt \le \varepsilon.$$

Поскольку

$$\mathbf{P}(|S_n - \mu n| \le M\sigma\sqrt{n}) \to \frac{1}{\sqrt{2\pi}} \int_{-M}^{M} e^{-t^2/2} dt > 1 - \varepsilon,$$

то $\mathbf{P}(|S_n - \mu n| > M\sigma\sqrt{n}) \le 2\varepsilon$ при достаточно больших n. При этом

$$\mathbf{P}(S_n - \mu n \le x\sigma\sqrt{n}) = \mathbf{P}(S_n - \mu n < -M\sigma\sqrt{n}) + \mathbf{P}(-M\sigma\sqrt{n} \le S_n - \mu n \le x\sigma\sqrt{n}).$$

Следовательно,

$$\limsup_{n\to\infty} \mathbf{P}(S_n - \mu n \le x\sigma\sqrt{n}) \le 2\varepsilon + \frac{1}{\sqrt{2\pi}} \int_{-M}^{x} e^{-t^2/2} dt \le 2\varepsilon + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt.$$

При этом

$$\liminf_{n\to\infty} \mathbf{P}(S_n - \mu n \le x\sigma\sqrt{n}) \ge \liminf_{n\to\infty} \mathbf{P}(-M\sigma\sqrt{n} \le S_n - \mu n \le x\sigma\sqrt{n}) = \frac{1}{\sqrt{2\pi}} \int_{-M}^{x} e^{-t^2/2} dt \ge \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt - \varepsilon.$$

Итого

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt - \varepsilon \le \liminf_{n\to\infty} \mathbf{P}(S_n - \mu n \le x\sigma\sqrt{n}) \le \limsup_{n\to\infty} \mathbf{P}(S_n - \mu n \le x\sigma\sqrt{n}) \le 2\varepsilon + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt.$$

В силу произвольности $\varepsilon > 0$ имеем требуемую сходимость. $\square$

Таким образом,

$$\mathbf{P}(|S_n - \mu n| > x\sigma\sqrt{n}) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-x}^{x} e^{-t^2/2} dt = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-t^2/2} dt$$

Найдем такое x, что вероятность в правой части меньше 0.05. Функция Φ не выражается в элементарных, но мы можем численно подсчитать интеграл и найти когда же $\Phi(x) = 0.95$. Оказывается, что это происходит приближенно при $x = 2$. Аналогично 3 — такое число, что вероятность правой части приближенно равен 0.0975. Тогда

$$\mathbf{P}(|S_n - \mu n| > 2\sigma\sqrt{n}) \approx 0.95, \quad \mathbf{P}(|S_n - \mu n| > 3\sigma\sqrt{n}) \approx 0.9975$$

Это так называемые правила двух сигм и трех сигм, S_n отклоняется от математического ожидания более чем на $2\sqrt{\mathbf{D}S_n}$ с вероятностью 5%, а более чем на $3\sqrt{\mathbf{D}S_n}$ с вероятностью около 0.25%.

Пример 6.4. Рассмотрим предыдущий пример. Тогда в силу правила двух сигм

$$\mathbf{P}\left(|S_n + 190| > 2 \cdot 6 \cdot 30\right) = \mathbf{P}\left(|S_n + 190| > 360\right) \approx 0.05,$$

то есть выигрыш на рулетке будет лежать в диапазоне $[-550, 170]$ с вероятностью 95%. Точно также по правилу трех сигм выигрыш будет лежать в диапазоне $[-730, 350]$ с вероятностью 99.75%. Можем оценить вероятность того, что мы окажемся в выигрыше

$$\mathbf{P}(S_n > 0) = \mathbf{P}(S_n + 190 > 190) = 1 - \mathbf{P}\left(S_n + 190 \le 6 \cdot 30 \cdot \frac{190}{180}\right) \to \Phi\left(\frac{190}{180}\right) \approx 0.14.$$

Поэтому это событие вполне возможно, но случится лишь раз из 7 случаев.

Вопрос 6.3. Какое количество монет с вероятностью орла 1/3 нужно подбросить, чтобы с вероятностью 50% получить хотя бы 1000 орлов?

6.3 Оценки погрешности приближения в теореме Пуассона и ЦПТ

Дадим без доказательства следующие две теоремы, дополняющие теорему Пуассона и центральную предельную теорему.

Теорема 6.5. *Пусть $X_i \sim Bern(p_i)$ независимы, $S_n = X_1 + \ldots + X_n$, $\lambda = p_1 + \ldots + p_n$, $Z \sim Poiss(\lambda)$. Тогда*

$$|\mathbf{P}(S_n = k) - \mathbf{P}(Z = k)| \le \sum_{i=1}^{n} p_i^2$$

Теорема 6.6 (Неравенство Берри-Эссеена)**.** *Пусть X_i н.о.р. и решетчаты, $S_n = X_1 + \ldots + X_n$, $\mathbf{E}X_1 = \mu$, $\mathbf{D}X_1 = \sigma^2$, $\mathbf{E}|X - \mu|^3 = \rho$. Тогда*

$$\left|\mathbf{P}\left(\frac{S_n - \mu n}{\sigma\sqrt{n}} \le x\right) - \Phi\left(x\right)\right| \le \frac{C\rho}{\sigma^3\sqrt{n}},$$

где $C < 0.5$.

Вопрос 6.4. Какую оценку погрешности приближения в центральной предельной теореме для n подбрасываний симметричной монеты дает неравенство Берри-Эссеена?

Пример 6.5. Таким образом, в примере с рулеткой мы можем оценить погрешность приближения сверху. Прямой подсчет дает нам

$$\frac{\rho}{\sigma^3} \approx \frac{1}{6 \cdot 36}\left(\left(2 + \frac{4}{19}\right)^3 \cdot \frac{18}{38} + \left(32 + \frac{4}{19}\right)^3 \cdot \frac{1}{38} + \frac{1}{2}\left(4 - \frac{4}{19}\right)^3\right) \approx 4,$$

откуда погрешность приближения не превышает $1/15 \approx 0.07$. К сожалению, в данном случае оценка оказалась достаточно большой, поскольку у величины большой параметр ρ.

7 Цепи Маркова

7.1 Определение марковской цепи

Если мы рассматриваем независимые величины, то

$$\mathbf{P}(X_n = x_n, \ldots, X_0 = x_0) = \mathbf{P}(X_n = x_n) \cdots \mathbf{P}(X_0 = x_0).$$

Если мы не накладываем никаких условий на зависимость величин, то

$$\mathbf{P}(X_n = x_n, \dots, X_0 = x_0) = \mathbf{P}(X_n = x_n | X_{n-1} = x_{n-1}, \dots, X_0 = x_0) \cdots \mathbf{P}(X_1 = x_1 | X_0 = x_0)\mathbf{P}(X_0 = x_0).$$

Для независимых величин было очень удобно — совместное распределение задается распределением отдельных величин, для общего случая нам требуется задать условное распределение X_n при условии всех предыдущих величин, что достаточно неудобно. Предположим, что выполнен промежуточный вариант — каждый X_m зависит по существу только от предыдущей X_{m-1}:

$$\mathbf{P}(X_m = x_m | X_{m-1} = x_{m-1}, \dots, X_0 = x_0) = \mathbf{P}(X_m = x_m | X_{m-1} = x_{m-1}), \ m \geq 1 \tag{5}$$

при всех таких $x_0, \dots, x_{m-1}$, что вероятность условия ненулевая.

Определение 7.1. Набор величин $X_0, \dots, X_n$, удовлетворяющий соотношению (5), называется цепью Маркова.

Подчеркнем, что при этом X_m и X_{m-2}, вообще говоря, зависимы, но независимы если фиксировано X_{m-1}. Более явно это будет раскрыто чуть позже.

Для удобства мы не будем требовать чтобы значения X_i были числами, а будет предполагать, что они принадлежат произвольному конечному или счетному множеству S. Мы можем пронумеровать их числами из $\mathbb{N}$ и считать, что $S \subseteq \mathbb{N}$.

Пример 7.1.

- Независимые случайные элементы $X_1, \dots, X_n$ со значениями из какого-то конечного (счетного) множества S будут представлять собой цепь Маркова. Действительно,

$$\mathbf{P}(X_m = x_m | X_{m-1} = x_{m-1}, \dots, X_0 = x_0) = \mathbf{P}(X_m = x_m) = \mathbf{P}(X_m = x_m | X_{m-1} = x_{m-1}.$$

- Сумма $S_n = X_1 + \dots + X_n$, где X_i независимые целочисленные величины, также будет цепью Маркова, поскольку

$$\mathbf{P}(S_n = x_n | S_{n-1} = x_{n-1}, \dots, S_0 = x_0) = \mathbf{P}(X_n = x_n - x_{n-1} | S_{n-1} = x_{n-1}, \dots, S_0 = x_0) =$$
$$\mathbf{P}(X_n = x_n - x_{n-1}) = \mathbf{P}(X_n = x_n - x_{n-1} | S_{n-1} = x_{n-1}) = \mathbf{P}(S_n = x_n | S_{n-1} = x_{n-1}).$$

  Такую последовательность называют случайным блужданием.

- Пусть $S_n = X_1 + \dots + X_n$, где X_i — независимые, $\mathbf{P}(X_i = 1) = \mathbf{P}(X_i = -1) = 1/2$. Предположим также, что если S_n оказывается равным a или $-a$, где $a > 0$ то после этого S_{n+1} обязательно равно, соответственно, $a - 1$ и $-a + 1$. Такую последовательность называют случайным блужданием с отражением. Это также марковская цепь непосредственно из определения.

- Возьмем $a = 3$ и добавим к той же модели добавить задержку: если S_n равно 3, а $S_{n-1} = 2$ (или -3 и -2, соответственно), то S_{n+1} всегда равно 3 (или -3), после чего $S_{n+2} = 2$ (или -2), соответственно. Тогда марковское свойство нарушится. Действительно, теперь

$$\mathbf{P}(S_8 = 3 | S_7 = 3, S_6 = 3, S_5 = 2, S_4 = 3, S_3 = 3, S_2 = 2, S_1 = 1) = 0,$$
$$\mathbf{P}(S_8 = 3 | S_7 = 3, S_6 = 2, S_5 = 1, S_4 = 0, S_3 = 1, S_2 = 0, S_1 = 1) = 1.$$

7.2 Эквивалентные определения

Дадим несколько эквивалентных определений цепи Маркова

Лемма 7.1. *Следующие условия эквивалентны:*

1. X_n *— цепь Маркова;*
2. *Для любых* k, l, $A \in \mathbb{R}^k$, $B \in \mathbb{R}^l$, $i \in \mathbb{N}$, *выполнено соотношение*

$$\mathbf{P}((X_{k+1}, \dots, X_{k+l}) \in B, (X_{k-1}, \dots, X_0) \in A | X_k = i) =$$
$$\mathbf{P}((X_{k+1}, \dots, X_{k+l}) \in B | X_k = i)\mathbf{P}((X_{k-1}, \dots, X_0) \in A | X_k = i).$$

3. *Вероятность* $\mathbf{P}(X_n = x_n | X_{n-1} = x_{n-1}, \dots, X_0 = x_0)$ *зависит только от* x_n, x_{n-1}, n*; Это свойство романтически называют "прошлое не зависит от будущего при условии настоящего".*

Доказательство. 1) Убедимся, что из первого следует второе.

Заметим, что

$$\mathbf{P}(X_n = x_n, \dots, X_0 = x_0) = \mathbf{P}(X_n = x_n | X_{n-1} = x_{n-1}, \dots, X_0 = x_0) \cdots \mathbf{P}(X_1 = x_1 | X_0 = x_0) \mathbf{P}(X_0 = x_0) = p_{x_{n-1}, x_n, n} \cdots p_{x_0, x_1, 1} \mathbf{P}(X_0 = x_0),$$

где $p_{i,j,n} = \mathbf{P}(X_n = j | X_{n-1} = i)$.

Рассмотрим пространство $\widetilde{\Omega} = \{X_k = i\}$, $\widetilde{\mathbf{P}}(\omega) = \mathbf{P}(\omega | X_k = i)$. Тогда мы хотим доказать, что

$$\widetilde{\mathbf{P}}((X_{k+1}, \dots, X_{k+l}) \in B, (X_{k-1}, \dots, X_0) \in A) = \widetilde{\mathbf{P}}((X_{k+1}, \dots, X_{k+l}) \in B)\ \widetilde{\mathbf{P}}((X_{k-1}, \dots, X_0) \in A).$$

Как мы уже знаем, для доказательства независимости случайных векторов достаточно проверить то, что совместная функция масс вектора есть произведение функций масс каждого из них, то есть

$$\widetilde{\mathbf{P}}(X_{k+1} = x_{k+1}, \dots, X_{k+l} = x_{k+l}, X_{k-1} = x_{k-1}, \dots, X_0 = x_0) = \widetilde{\mathbf{P}}(X_{k+1} = x_{k+1}, \dots, X_{k+l} = x_{k+l}) \widetilde{\mathbf{P}}(X_{k-1} = x_{k-1}, \dots, X_0 = x_0),$$

откуда

$$\mathbf{P}(X_{k+l} = x_{k+l}, \dots, X_{k+1} = x_{k+1}, X_{k-1} = x_{k-1}, \dots, X_0 = x_0 | X_k = x_k) = \mathbf{P}(X_{k+l} = x_{k+l}, \dots, X_{k+1} = x_{k+1} | X_k = x_k) \mathbf{P}(X_{k-1} = x_{k-1}, \dots, X_0 = x_0 | X_k = x_k).$$

Левая часть представима в виде дроби

$$\mathbf{P}(X_{k+l} = x_{k+l}, \dots, X_0 = x_0) / \mathbf{P}(X_k = x_k),$$

а правая в виде

$$\mathbf{P}(X_{k+l} = x_{k+l}, \dots, X_{k+1} = x_{k+1}, X_k = x_k) \mathbf{P}(X_k = x_k, \dots, X_0 = x_0) / \mathbf{P}(X_k = x_k)^2.$$

Заметим, что

$$\mathbf{P}(X_{k+l} = x_{k+l}, \dots, X_k = x_k) = \sum_{y_0, \dots y_{k-1} \in S} \mathbf{P}(X_{k+l} = x_{k+l}, \dots, X_k = x_k, X_{k-1} = y_{k-1}, \dots, X_0 = y_0) =$$

$$\sum_{y_0, \dots y_{k-1} \in S} p_{x_{k+l-1}, x_{k+l}, k+l} \cdots p_{x_k, x_{k+1}, k+1} p_{y_{k-1}, x_k, k} \cdots p_{y_0, y_1, 1} \mathbf{P}(X_0 = y_0) = p_{x_{k+l-1}, x_{k+l}, k+l} \cdots p_{x_k, x_{k+1}, k+1} \mathbf{P}(X_k = x_k).$$

Следовательно,

$$\mathbf{P}(X_{k+l} = x_{k+l}, \dots, X_{k+1} = x_{k+1}, X_k = x_k) \mathbf{P}(X_k = x_k, \dots, X_0 = x_0) / \mathbf{P}(X_k = x_k)^2 = p_{x_{k+l-1}, x_{k+l}, k+l} \cdots p_{x_0, x_1, 1} \mathbf{P}(X_0 = x_0) / \mathbf{P}(X_k = x_k).$$

Но

$$\mathbf{P}(X_{k+l} = x_{k+l}, \dots, X_0 = x_0) / \mathbf{P}(X_k = x_k) = p_{x_{k+l-1}, x_{k+l}, k+l} \cdots p_{x_0, x_1, 1} \mathbf{P}(X_0 = x_0) / \mathbf{P}(X_k = x_k),$$

откуда получаем требуемое утверждение.

2) Из условия 3) вытекает

$$\mathbf{P}(X_{k+1} = x_{k+1}, X_{k-1} = x_{k-1}, \dots, X_0 = x_0 | X_k = x_k) = \mathbf{P}(X_{k+1} = x_{k+1} | X_k = x_k) \mathbf{P}(X_{k-1} = x_{k-1}, \dots, X_0 = x_0 | X_k = x_k).$$

Отсюда

$$\mathbf{P}(X_{k+1} = x_{k+1}, \dots, X_0 = x_0) = \mathbf{P}(X_{k+1} = x_{k+1} | X_k = x_k) \mathbf{P}(X_k = x_k, \dots, X_0 = x_0).$$

Следовательно,
$$\mathbf{P}(X_{k+1}=x_{k+1}|X_k=x_k\dots,X_0=x_0)=\mathbf{P}(X_{k+1}=x_{k+1}|X_k=x_k),$$
а значит зависит только от x_k, x_{k+1} и k, что и требовалось доказать.
3) Докажем, что из третьего утверждения следует первое.

Заметим, что
$$\mathbf{P}(X_{k+1}=x_{k+1},X_k=x_k)=\sum_{y_0,\dots,y_{k-1}}\mathbf{P}(X_{k+1}=x_{k+1},X_k=x_k,X_{k-1}=y_{k-1},\dots,X_0=y_0)=$$
$$\sum_{y_0,\dots,y_{k-1}}q_{x_k,x_{k+1},k+1}\cdots q_{y_0,y_1,1}\mathbf{P}(X_0=y_0),$$
где
$$q_{i,j,n}=\mathbf{P}(X_n=j|X_{n-1}=i,X_{n-2}=x_{n-2},\dots,X_0=x_0).$$
Тогда
$$\mathbf{P}(X_{k+1}=x_{k+1},X_k=x_k)=\frac{\sum_{y_0,\dots,y_{k-1}}q_{x_k,x_{k+1},k+1}q_{y_{k-1},x_k,k}\cdots q_{y_0,y_1,1}\mathbf{P}(X_0=y_0)}{\sum_{y_0,\dots,y_{k-1}}q_{y_{k-1},x_k,k}\cdots q_{y_0,y_1,1}\mathbf{P}(X_0=y_0)}=q_{x_k,x_{k+1},k+1}.$$
Лемма доказана □

Следствие 7.1. В частности, из определения 2) вытекает более общая форма марковского свойства (5):
$$\mathbf{P}((X_{k+1},\dots,X_{k+l})\in B|X_k=i,(X_{k-1},\dots,X_0)\in A)=\mathbf{P}((X_{k+1},\dots,X_{k+l})\in B|X_k=i).$$
Крайне важно, что условие $X_k=i$ оставляется при этом в такой локальной форме, иначе свойство перестанет быть верным. В частности, функция $f(X_n)$ от марковской цепи необязательно будет марковской цепью.

7.3 Матрица вероятностей перехода

Определение 7.2. Мы будем рассматривать *однородные* цепи Маркова, то есть такие, что
$$\mathbf{P}(X_m=j|X_{m-1}=i)=p_{i,j}$$
при всех m.

Тогда совместная функция масс $(X_0,\dots,X_n)$ будет задаваться соотношением
$$\mathbf{P}(X_n=x_n,\dots,X_0=x_0)=\mathbf{P}(X_0=x_0)p_{x_0,x_1}\cdots p_{x_{n-1},x_n}.$$
Определение 7.3. Матрица $P=(p_{i,j},\ i,j\in S)$ называется матрицей вероятностей перехода цепи Маркова (в случае счетного S это бесконечная матрица).

Такой матрицей может быть любая стохастическая матрица — матрица с неотрицательными элементами, сумма которых по каждой строке равна 1.

Пример 7.2. Для случайного блуждания и случайного блуждания с отражением м.в.п. будут, соответственно, иметь вид
$$\begin{pmatrix}\dots & & & \dots & & & \dots\\ \dots & 0 & 1/2 & 0 & 1/2 & 0 & \dots\\ \dots & 0 & 0 & 1/2 & 0 & 1/2 & \dots\\ \dots & & & \dots & & & \dots\end{pmatrix},\quad \begin{pmatrix} & 0 & 1 & 0 & 0 & 0 & 0 & \dots\\ \dots & & & & \dots & & & \dots\\ \dots & 0 & 1/2 & 0 & 1/2 & 0 & & \dots\\ \dots & 0 & 0 & 1/2 & 0 & 1/2 & & \dots\\ \dots & & & & \dots & & & \dots\\ \dots & 0 & 0 & 0 & 0 & 1 & 0 & \end{pmatrix}$$
Рассмотрим вектор $\bar{q}^{(n)}=(q_1^{(n)},\dots)=(\mathbf{P}(X_n=1),\dots)$. При этом
$$q_j^{(n)}=\sum_{i\in S}\mathbf{P}(X_n=j|X_{n-1}=i)\mathbf{P}(X_{n-1}=i)=\left(\bar{q}^{(n-1)}P\right)_j.$$

Таким образом,
$$\vec{q}^{(n)} = \vec{q}^{(0)} P^n.$$

Иначе говоря, вероятность обнаружить цепь Маркова в состоянии i в момент n представляет собой i-й элемент вектора $\vec{q}^{(0)} P^n$. Аналогичным образом доказывается так называемое уравнение Колмогорова-Чепмена:

$$P^{(n)} = P^{(m)} P^{(n-m)},\ 0 < n < m,$$

где $P^{(n)} = (p_{i,j}^{(n)},\ i, j \in S)$, $p_{i,j}^{(n)} = \mathbf{P}(X_n = j | X_0 = i)$. Иначе говоря, матрица вероятностей перехода за n шагов $P^{(n)}$ совпадает с P^n, где P — матрица вероятностей перехода.

7.4 Классификация состояний

Заметим, что можно отражать взаимосвязь между состояниями с помощью графа состояний. Расставим состояния цепи в виде вершин нашего ориентированного графа, проведем ребро из состояния i в состояние j, если $p_{i,j} > 0$. Введем несколько видов состояний цепи:

Рис. 8: Граф состояний для произвольной цепи из трех состояний, в которой $p_{2,3} = p_{3,2} = 0$.

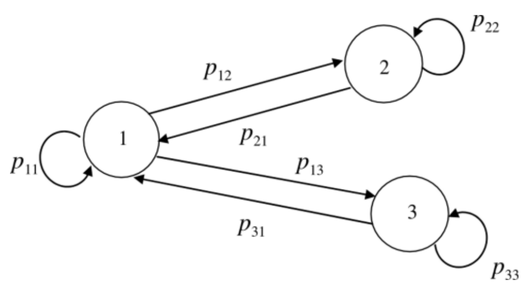
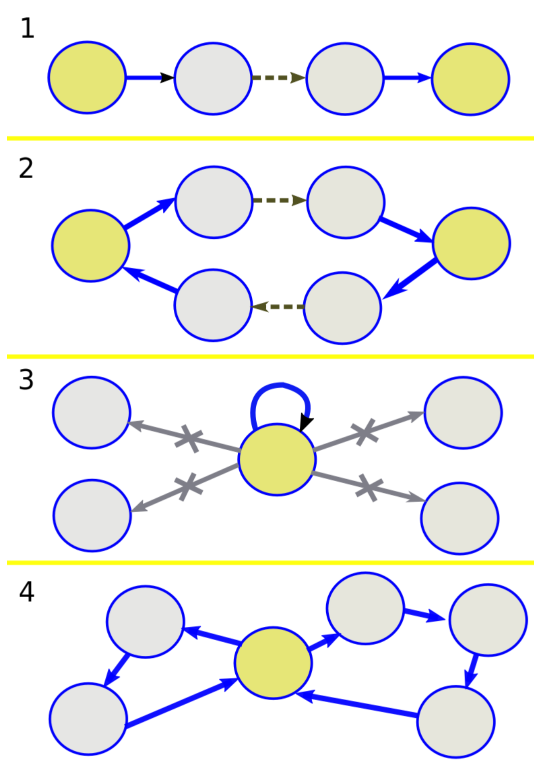


Рис. 9: 1) Из левого состояния следует правое; 2) Левое и правое состояние сообщаются; 3) Желтое состояние поглощающее; 4) Желтое состояние непериодично.



Определение 7.4. 1. Из состояния i следует состояние j, если найдется такое n, что $p_{i,j}(n) > 0$. В терминах графа состояний это означает, что существует путь из i в j.

2. Состояния i, j называются сообщающимися, если из i следует j, а из состояния j следует i. Иначе говоря, существует цикл, проходящий через i, j.

3. Состояние i поглощающее, если из него следует только оно само.

4. Состояние i существенное, если для любого j, которое следует из i, состояния i и j сообщаются. В терминах графа состояний это значит, что любой путь, выходящий из i, можно продлить так, что он вернется в i.

5. Состояние i непериодическое, если оно существенно и существует несколько циклов из i в i, взаимно простой длины. Иначе говоря, НОД длин циклов, проходящих через состояние i, равен единице.

6. Состояние i периодическое с периодом d, если НОД длин циклов, проходящих через состояние i, равен d.

Пример 7.3. Рассмотрим простое случайное блуждание $S_n = X_1 + \cdots + X_n$, где $X_i = \pm 1$ независимы. Тогда любое состояние из $\mathbb{Z}$ существенно, все они сообщаются, поглощающих состояний нет, период у всех состояний одинаковый и равен 2.

Представим себе ветвящийся процесс — в начале есть одна частица, затем она дает случайное число потомков (возможно 0), каждый из потомков (если они есть) независимо от остальных дает новых потомков по тому же закону и так далее. Тогда 0 будет поглощающим состоянием, все остальные состояния будут несущественными.

Цепь называют неразложимой, если любые два состояния в ней сообщаются.

Лемма 7.2. *Состояния любой цепи распадаются на непересекающиеся классы $E_0, E_1, \dots E_k, \dots$, где E_0 — множество всех несущественных состояний, а все состояния класса E_k при любом k сообщаются и все следующие из них состояния лежат в классе E_k.*

Иначе говоря, можно перенумеровать состояния таким образом, что матрица вероятностей перехода станет блочной (кроме несущественных состояний):

$$\left(\begin{array}{c|cc|ccc|cc|c}
 & E_0 & & E_1 & & & E_2 & & \dots \\
\hline
E_0 & p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} & p_{1,5} & p_{1,6} & p_{1,7} & \dots \\
 & p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} & p_{2,5} & p_{2,6} & p_{2,7} & \dots \\
\hline
E_1 & 0 & 0 & p_{3,3} & p_{3,4} & p_{3,5} & 0 & 0 & \dots \\
 & 0 & 0 & p_{4,3} & p_{4,4} & p_{4,5} & 0 & 0 & \dots \\
 & 0 & 0 & p_{5,3} & p_{5,4} & p_{5,5} & 0 & 0 & \dots \\
\hline
E_2 & 0 & 0 & 0 & 0 & 0 & p_{6,6} & p_{6,7} & \dots \\
 & 0 & 0 & 0 & 0 & 0 & p_{7,6} & p_{7,7} & \dots
\end{array}\right).$$

Доказательство. Выделим класс несущественых состояний E_0, рассмотрим одно из существенных состояний (если такое найдется) и все состояния, сообщающиеся с ним. Выберем одно из не вошедших туда существенных состояний (если такое найдется) и рассмотрим все состояния, сообщающиеся с ним. И так далее. Полученное разбиение будет иметь требуемый вид.

Действительно, сообщаемость транзитивна — если i сообщается с j, а j с k, то i сообщается с k. Значит в каждом классе все состояния сообщаются. При этом если i и j из разных классов (не из нулевого), из i следует j, то из j следует i (в силу существенности j), а значит i и j сообщаются. Но тогда сообщаются и порождающие классов i и j, а это противоречит тому, что эти классы разные. □

Таким образом, конечные цепи Маркова устроены следующим образом: возможно цепь стартует из несущественного состояния, спустя какое-то время она попадает в один из неразложимых классов E_j, $j > 0$, а затем совершает переходы между состояниями из этого класса. С момента попадания в неразложимый класс цепь начинает вести себя как неразложимая

Лемма 7.3 (солидарности)**.** *Пусть цепь неразложимая. Тогда все ее состояния имеют одинаковую существенность (все существенны и все несущественны), одинаковую периодичность (все непериодичны или все периодичны с одинаковым периодом).*

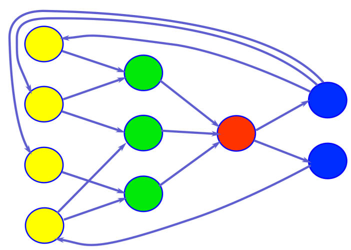
Мы оставим эту лемму для самостоятельного доказательства в качестве несложного упражнения. В силу леммы у неразложимой цепи периоды всех состояний равны одному и тому же числу d. Это число называют периодом цепи d.

Отметим также еще один результат, выступающий в качестве домашней задачи

Лемма 7.4. *Множество состояний S неразложимой цепи с периодом d разбивается на такие классы D_1,...,D_d, что из любого состояния, лежащего в D_1, можно перейти только в состояние из D_2, из любого состояния из D_2 — только в D_3 и так далее, из любого состояния из D_d — только в D_1.*

Соответственно, цепь $X_d, X_{2d}, \ldots$, начинающаяся в состоянии из некоторого класса D_i, уже будет неразложимой и непериодичной цепью.

Рис. 10: Периодическая цепь с периодом 4: из состояний каждого цвета можно попасть только в состояния следующего цвета



7.5 Эргодическая теорема

Мы хотели бы понять как будут у марковской цепи X_n устроены вероятности $\mathbf{P}(X_n = i)$ при больших n.

Пример 7.4. Пусть X_n — цепь с двумя состояниями 1 и 2, $p_{1,1} = p$, $p_{2,2} = q$, причем начальное распределение $(1/3, 2/3)$. При $p = q = 1$ получаем

$$\mathbf{P}(X_n = 1 | X_{n-1} = 2) = \mathbf{P}(X_n = 2 | X_{n-1} = 1) = 1,$$

цепь периодическая и $\mathbf{P}(X_n = i)$ это последовательность $1/3, 2/3, 1/3, 2/3$ и так далее, которая расходится. Поэтому в этом случае предела нет.

При $p = 1/2$, $q = 3/4$ можно утверждать, что

$$p^{(n)} = p^{(n-1)} P = (1/3, 2/3),$$

поскольку вектор $(1/3, 2/3)$ при умножении слева на нашу матрицу перейдет в себя. Таким образом, $\mathbf{P}(X_n = 1) = 1/3$, $\mathbf{P}(X_n = 2) = 2/3$ при всех n.

Таким образом, проще всего дело обстоит, если вектор начального распределения $p^{(0)}$ является левым собственным вектором матрицы P с собственным значением 1, то есть $p^{(0)} P = p^{(0)}$. Тогда $p^{(n)} = p^{(0)}$.

Определение 7.5. Такой вектор $\vec{p}$ из вероятностей, что $\vec{p}P = \vec{p}$ называется стационарным распределением вероятностей для цепи Маркова с матрицей вероятностей перехода P.

Теорема 7.1 (Эргодическая теорема о цепи Маркова)**.** *Пусть X_n — неразложимая непериодическая цепь Маркова с конечным числом состояний. Тогда вне зависимости от начального распределения*

$$\mathbf{P}(X_n = j) \to \pi_j, \ n \to \infty,$$

где π — стационарное распределение. При этом стационарное распределение у данной цепи единственно и все вероятности π_j положительны.

Более того, при некоторых $C > 0$, $q \in (0, 1)$ верно неравенство $|\mathbf{P}(X_n = j) - \pi_j| < Cq^n$.

Доказательство. Доказательство может быть проведено различными путями. Мы воспользуемся методом, который называется методом одного вероятностного пространства (coupling). Разобьем доказательство на четыре леммы

Лемма 7.5. *Если цепь неразложима и непериодична, то найдется такое число l_0, что при всех $l > l_0$ матрица P^l состоит только из положительных чисел.*

Доказательство. Рассмотрим состояние i и покажем, что найдется такое m, что $p_{i,i}^{(m+j)} > 0$ при всех $m > 0$. Иначе говоря, существуют циклы, проходящие через i, любой длины более чем m.

В силу непериодичности найдутся циклы таких длин $d_1, \ldots, d_k$, что НОД$(d_1, \ldots, d_k)$=1. Следовательно, найдутся такие целые $a_1, \ldots, a_k$, что

$$a_1 d_1 + \cdots + a_k d_k = 1.$$

Это утверждение является следствием утверждения о том, что для любых двух чисел a, b их НОД можно представить в виде $ac + bd$ для некоторых целых c, d.

Умножим это тождество на $n = 1, 2, \ldots, d_1$ и получим тождества

$$na_1 d_1 + \cdots + na_k d_k = n.$$

Добавим к каждому из этих тождеств $j = d_1 \sum_{i=1}^k |a_i| d_i$ и получим

$$(na_1 + d_1|a_1|)d_1 + \cdots + (na_k + d_1|a_k|)d_k = n + m.$$

Теперь все коэффициенты при $d_1, \ldots, d_k$ неотрицательны. Следовательно, мы можем получить, комбинируя циклы $d_1, \ldots, d_k$ в каких-то количествах, циклы всевозможных длин от $m+1$ до $m+d_1$. С другой стороны, мы можем сверх этого добавить любое количество циклов длины d_1, а, значит, мы можем получить любую длину цикла больше m.

При этом в силу неразложимости найдется такое m_0, что для любых состояний j и k найдется путь длины не больше чем m_0 из j в k. Значит, для любых состояний j и k существует путь любой длины n более чем $m + 1 + 2m_0$ из j в k. Действительно, мы можем перейти не более чем за m_0 шагов из j в i, не более чем за m_0 шагов из i в k, а оставшиеся нам до n шаги блуждать по циклу, проходящему через i (пользуясь тем, что у нас есть циклы любой длины более чем $n - 2m_0$).

Следовательно, P^{m+2m_0+1} состоит только из положительных чисел. Лемма доказана. □

Лемма 7.6. *Если цепь Y_n имеет начальное распределение $\vec{\pi}$, то $\mathbf{P}(Y_n = k) = \pi_k$ при всех k.*

Доказательство. Лемма практически очевидна, ведь мы доказали, что если $q^{(n)} = (\mathbf{P}(Y_n = 1), \ldots)$, то

$$q^{(n)} = \pi P^{(n)}$$

Но

$$\pi P^{(n)} = \pi P^n = (\pi P)P^{n-1} = \pi P^{n-1} = \ldots = \pi.$$

□

Увы, у нашей цепи X_n может быть другое начальное распределение.

Давайте рассмотрим обе наших цепи X_n и Y_n (будем считать их независимыми), рассмотрим случайную величину T — время до первого совпадения наших цепей:

$$\{T = k\} = \{X_i \neq Y_i,\ i < k,\ X_k = Y_k\}.$$

Лемма 7.7. *Найдутся такие $C > 0$, $q \in (0, 1)$, что*

$$\mathbf{P}(T > k) < Cq^k$$

при всех $k \in \mathbb{N}$.

Доказательство.

$$\mathbf{P}(T > m) = \mathbf{P}(X_i \neq Y_i, i \leq m).$$

Выберем такое l, что P^l состоит из положительных элементов и обозначим через p минимальную вероятность в P^l. Тогда

$$\mathbf{P}(T > ml) = \sum_{y_1,\ldots,y_{ml}} \mathbf{P}(X_i \neq y_i, i \leq ml)\mathbf{P}(Y_1 = y_1, \ldots, Y_{ml} = y_{ml}).$$

При этом при любом y_l

$$\mathbf{P}(X_i \neq y_i, i \leq ml) \leq \sum_{j \in S} \mathbf{P}(X_i \neq y_i, i \leq (m-1)l, X_{(m-1)l} = j)\mathbf{P}(X_l \neq y_l | X_0 = j) \leq \mathbf{P}(X_i \neq y_i, i \leq (m-1)l)(1-p).$$

Следовательно,

$$\mathbf{P}(T > ml) \leq (1-p)^m \sum_{y_1,\ldots,y_{ml}} \mathbf{P}(Y_1 = y_1, \ldots, Y_{ml} = y_{ml}) = q^{ml}$$

при некотором $q = 1 - p < 1$. При этом при $ml \leq k \leq (m+1)l$ имеем

$$\mathbf{P}(T > k) \leq \mathbf{P}(T > ml) \leq q^{ml} = q^{k-ml}q^k \leq q^{-l+1}q^k \leq Cq^k,$$

где $C = q^{-l+1}$. Лемма доказана. □

Рассмотрим случайные величины

$$Z_n = \begin{cases} X_n, & n \leq T, \\ Y_n, & n > T. \end{cases}$$

Лемма 7.8. *Последовательность Z_n — цепь Маркова с матрицей вероятностей перехода P.*

Доказательство. Запишем для Z_n

$$\mathbf{P}(Z_n = i_n | Z_{n-1} = i_{n-1}, \ldots, Z_0 = i_0) = \mathbf{P}(Z_n = i_n, \ldots, Z_0 = i_0)/\mathbf{P}(Z_{n-1} = i_{n-1}, \ldots, Z_0 = i_0).$$

При этом

$$\begin{aligned} \mathbf{P}(Z_n = i_n, \ldots, Z_0 = i_0) = \mathbf{P}(T > n, X_n = i_n, \ldots, X_0 = i_0) + \\ \sum_{k=0}^{n} \mathbf{P}(T = k, Y_n = i_n, \ldots, Y_k = i_k, X_{k-1} = i_{k-1}, \ldots, X_0 = i_0) = \\ \mathbf{P}(X_n = i_n, \ldots, X_0 = i_0)\mathbf{P}(Y_n \neq i_n, \ldots, Y_0 \neq i_0) + \sum_{k=0}^{n} \mathbf{P}(Y_n = i_n, \ldots, Y_k = i_k, Y_{k-1} \neq i_{k-1}, \ldots, Y_0 \neq i_0) \\ \times \mathbf{P}(X_k = i_k, \ldots, X_0 = i_0). \end{aligned}$$

Заметим, что

$$\begin{aligned} \mathbf{P}(X_n = i_n, \ldots, X_0 = i_0) = \mathbf{P}(X_0 = i_0)p_{i_0,i_1} \cdots p_{i_{n-1},i_n}, \\ \mathbf{P}(Y_n = i_n, \ldots, Y_k = i_k, Y_{k-1} \neq i_{k-1}, \ldots, Y_0 \neq i_0) = \mathbf{P}(Y_k = i_k, Y_{k-1} \neq i_{k-1}, \ldots, Y_0 \neq i_0)p_{i_k,i_{k+1}} \cdots p_{i_{n-1},i_n}. \end{aligned}$$

Итого,

$$\begin{aligned} \mathbf{P}(Z_n = i_n, \ldots, Z_0 = i_0) = \mathbf{P}(X_0 = i_0)p_{i_0,i_1} \cdots p_{i_{n-1},i_n} \times \\ \left(\mathbf{P}(Y_n \neq i_n, \ldots, Y_0 \neq i_0) + \sum_{k=0}^{n} \mathbf{P}(Y_k = i_k, Y_{k-1} \neq i_{k-1}, \ldots, Y_0 \neq i_0) \right) = \\ \mathbf{P}(X_0 = i_0)p_{i_0,i_1} \cdots p_{i_{n-1},i_n}. \end{aligned}$$

Значит,

$$\mathbf{P}(Z_n = i_n | Z_{n-1} = i_{n-1}, \ldots, Z_0 = i_0) = p_{i_{n-1},i_n},$$

что и требовалось доказать. □

Раз последовательность Z_n — это тоже цепь Маркова с матрицей перехода P, с начальным распределением тем же, что и у X_n, то $\mathbf{P}(Z_n = k) = \mathbf{P}(X_n = k)$ при всех n, k. С другой стороны, $\{Z_n = k\} = \{Y_n = k\}$ при $n > T$. Значит,

$$|\mathbf{P}(Z_n = k) - \mathbf{P}(Y_n = k)| \leq \mathbf{P}(T < n) \leq q^n$$

Остается доказать следующую лемму:

Лемма 7.9. *Неразложимая непериодическая цепь с матрицей перехода P имеет единственное стационарное распределение.*

Доказательство. Пусть k — некоторое состояние, T — время прихода в это состояние, то есть $\{T = i\} = \{X_j \neq k, j < i, X_i = k\}$. Рассмотрим

$$a_j = \sum_{n=0}^{\infty} \mathbf{P}(X_n = j, T > n | X_0 = k),$$

где $a_k = 1$. Тогда утверждается, что $\vec{a}P = \vec{a}$. Действительно, при $i \neq k$ имеем

$$(\vec{a}P)_i = \sum_{j \neq k} \sum_{n=0}^{\infty} \mathbf{P}(X_n = j, T > n | X_0 = k) p_{j,i} + a_k p_{k,i} = \sum_{n=1}^{\infty} \mathbf{P}(X_{n+1} = i, T > n+1 | X_0 = k) + \mathbf{P}(X_1 = i | X_0 = k) =$$
$$\sum_{n=2}^{\infty} \mathbf{P}(X_n = i, T > n | X_0 = k) + \mathbf{P}(X_1 = i | X_0 = k) = a_i.$$

При $i = k$ верно тождество

$$(\vec{a}P)_k = \sum_{j \neq k} \sum_{n=1}^{\infty} \mathbf{P}(X_n = j, T > n | X_0 = k) p_{j,k} + p_{k,k} = \sum_{n=1}^{\infty} \mathbf{P}(T = n+1) + \mathbf{P}(T = 1) = 1.$$

При этом указанные a_j неотрицательны и оцениваются сверху величиной

$$\sum_{n=0}^{\infty} \mathbf{P}(X_n = j, T > n | X_0 = k) \leq \sum_{n=0}^{\infty} \mathbf{P}(T > n) = \mathbf{E}T.$$

Из тех соображений, что и выше, $\mathbf{P}(T > k) < q^k$ при некотором $q \in (0,1)$, откуда, в частности, $\mathbf{E}T$ конечно. Значит, $\vec{a}$, деленный на сумму координат, будет требуемым стационарным распределением.

Единственность стационарного распределения вытекает из рассуждений теоремы. Если бы существовало два стационарных распределения, то у вероятностей $\mathbf{P}(X_n = i)$ было бы два различных предела □

□

Замечание 7.1. В условиях теоремы вектор

$$\pi_i = \frac{1}{\mathbf{E}T_i}$$

будет стационарным распределением, где T_i — время возвращения из состояния i в себя.

Доказательство. Отметим, что при любом i

$$\sum_{j \in S} \sum_{n=0}^{\infty} \mathbf{P}(X_n = j, T_i > n | X_0 = i) = \sum_{n=0}^{\infty} \mathbf{P}(T_i > n | X_0 = i) = \mathbf{E}T_i$$

Из предыдущей леммы любой вектор

$$\left(\frac{1}{\mathbf{E}T_i} \sum_{n=0}^{\infty} \mathbf{P}(X_n = 1, T_i > n | X_0 = i), \ldots, \frac{1}{\mathbf{E}T_i} \sum_{n=0}^{\infty} \mathbf{P}(X_n = |S|, T_i > n | X_0 = i) \right)$$

является стационарным. Но поскольку стационарный вектор единственен, то все такие векторы равны. Но у i вектора i-я компонента равна $1/\mathbf{E}T_i$. В силу произвольности i имеем требуемое. □

Отметим, что подставляя в качестве начальных распределений векторы $(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)$, мы увидим, что для эргодической цепи Маркова

$$\mathbf{P}(X_n = j | X_0 = i) \to \pi_j,\ n \to \infty,$$

при любом i, откуда матрица вероятностей перехода $P^{(n)}$ сходится к матрице с постоянными строками, равными $\vec{\pi}$.

8 Общая вероятностная модель

8.1 Общее вероятностное пространство

Вероятностным пространством $(\Omega, \mathcal{F}, \mathbf{P})$ мы будем называть тройку из:

1. Пространства элементарных исходов Ω, некоторого множества, на которое мы не накладываем никаких ограничений.

2. Сигма-алгебры событий $\mathcal{F}$: множества, содержащего некоторые элементы Ω и являющегося сигма-алгеброй, где

   Определение 8.1. Сигма-алгеброй $\mathcal{A}$ подмножеств множества B называют множество подмножеств B, обладающее следующими свойствами

   (a) $B \in \mathcal{A}$;

   (b) Если $A \in \mathcal{A}$, то $\overline{A} \in \mathcal{A}$, где дополнение рассматривается до B;

   (c) Если $A_1, \ldots, A_n, \ldots$ принадлежат $\mathcal{A}$, то множество

   $$\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}.$$

3. Вероятностной меры $\mathbf{P}$: отображения из $\mathcal{F}$ в $[0, 1]$, являющегося вероятностной мерой, где

   Определение 8.2. Отображение $\mathbf{P}$ из сигма-алгебры $\mathcal{F}$ подмножеств Ω в отрезок $[0, 1]$ называется вероятностной мерой, если выполнены следующие свойства:

   (a) $\mathbf{P}(\Omega) = 1$;

   (b) аддитивность: $\mathbf{P}(A + B) = \mathbf{P}(A) + \mathbf{P}(B)$ при любых непересекающихся A, B из $\mathcal{F}$;

   (c) счетная аддитивность:

   $$\mathbf{P}(A_1 + A_2 + \cdots + A_n + \cdots) = \sum_{n=1}^{\infty} \mathbf{P}(A_n),$$

   где A_i — произвольные непересекающиеся подмножества $\mathcal{F}$.

Пример 8.1. Рассмотрим, например, $\Omega = (0, 1]$, $\mathcal{F} = 2^{(0,1]}$ — множество всех его подмножеств, $\mathbf{P}(A) = k/n$, где k — число точек вида i/n, $i \in \{1, \ldots, n\}$, в множестве A.

Требуемые условия при этом выполнены. Действительно, множество всех подмножеств Ω является сигма-алгеброй, поскольку

- Ω лежит в $\mathcal{F}$;

- Дополнение до любого подмножества A полуинтервала $(0, 1]$ является подмножеством того же интервала;

- Если множества лежат в полуинтервале $(0, 1]$, то все их элементы лежат в $(0, 1]$, а значит их объединение являлется подмножеством $(0, 1]$.

Отображение $\mathbf{P}$ является вероятностной мерой, поскольку

- $\mathbf{P}((0, 1]) = n/n = 1$, поскольку полуинтервал $(0, 1]$ содержит n указанных точек;

- $\mathbf{P}(A+B) = k/n$, где k — количество указанных точек в A и B вместе. Конечно же это число равно сумме количеств указанных точек в A и B, поскольку множества не пересекаются;

- Поскольку точек указанного вида конечное число, то по существу мы можем рассматривать только конечное количество множеств A_n. Тогда условие вытекает из предыдущего.

Пример 8.2. Рассмотрим Ω — множество $(0,1]$, $\mathcal{F}$ — множество всех содержащихся в нем полуинтервалов, $\mathbf{P}$ сопоставляет полуинтервалу его длину. Такая тройка не образует вероятностного пространства, поскольку $\mathcal{F}$ не является сигма-алгеброй. Действительно, объединение полуинтервалов $(0,1/2]$ и $(2/3,1]$ не является полуинтервалом.

Давайте попробуем дополнить множество $\mathcal{F}$ объединениями полуинтервалов, приписывая им вероятность, равную сумме их длин. Однако, свойство 3) сигма-алгебры по-прежнему не будет выполнено. Скажем, объединение полуинтервалов $(0,1-1/n]$ по всем n будет интервалом $(0,1)$, которого в нашем множестве нет. Чуть позже мы поговорим о том, что же еще нужно добавить в $\mathcal{F}$, чтобы получить сигма-алгебру.

8.2 Общие рассуждения

Итак, чтобы описывать более широкий диапазон вероятностных экспериментов мы отказываемся от не более счетных пространств. Теперь множество возможных исходов Ω будет иметь произвольную мощность. При этом подход "определим вероятность на элементарных исходов, а оттуда найдем вероятность любого события" перестает быть разумным — при случайном выборе точки из отрезка естественно считать, что вероятности всех исходов нулевые и определить отсюда вероятность, скажем, половины отрезка не представляется возможным.

Мы могли бы пойти по другому пути — считать, что наша вероятность задана на 2^Ω неведомо каким образом, а считать ее только у интересных нам событий. Однако, оказывается, что непротиворечиво определить вероятность на множестве всех подмножеств отрезка $[0,1]$ уже может быть затруднительным.

Пример 8.3. Представим себе, что мы определили у каждого подмножества A полуинтервала $(0,1]$ вероятность, причем так, что при сдвиге множества A на любое число $b \in (0,1)$ (то есть элементы a из A переводятся в $a+b$, если $a+b \le 1$ и в $a+b-1$, если $a+b > 1$) вероятность множества не меняется. Скажем, обычная длина должна, казалось бы, удовлетворять этому свойству. Однако, оказывается, что такую вероятность определить нельзя.

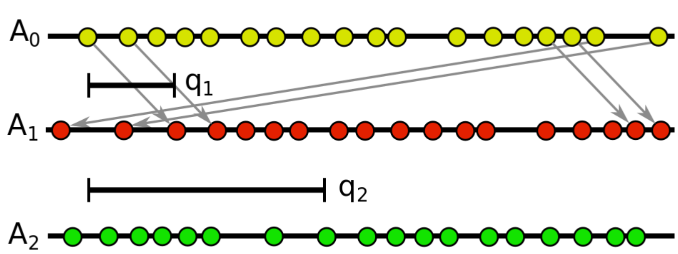
Рассмотрим множество A_0, удовлетворяющее следующим условиям:

- для любого $x \in (0,1]$ найдется такой элемент $a \in A_0$, что $x - a \in \mathbb{Q}$;

- для любых различных $a, b \in A_0$ величина $a - b \notin \mathbb{Q}$.

Множество A не конструктивно, однако, такое множество можно построить. Пронумеруем все числа из $\mathbb{Q} \cap (0,1)$ и рассмотрим A_n — сдвиги A_0 на q_n.

Рис. 11: Красное множество A_1 получается из желтого A_0 сдвигом на q_1 вправо. При этом две оказавшихся правее 1 точки сдвигаются на 1 влево. Аналогично зеленое множество A_2 получается сдвигом на q_2



Тогда A_i не пересекаются, поскольку если $a \in A_i \cap A_j$, то $a_1 = a - q_i \in A_0$, $a_2 = a - q_j \in A_0$, а значит a_1 и a_2 отличаются на рациональное число, что противоречит определению A.

Значит,

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

При этом все $\mathbf{P}(A_i)$ одинаковы, а значит в правой части стоит ряд из одинаковых чисел. Такой ряд может сойтись только если $\mathbf{P}(A_i) = 0$. Следовательно, $\mathbf{P}(\bigcup_{i=1}^{\infty} A_i) = 0$.

Однако, $\bigcup_{n=0}^{\infty} A_n = (0,1]$. Действительно, для любого $x \in (0,1]$ найдется такое $q_n \in \mathbb{Q} \cap (0,1)$, что $x - q_n \in A_0$, а, значит, $x \in A_n$. Таким образом, все элементы полуинтервала лежат в объединении A_n. Но тогда вероятность $\bigcup_{n=0}^{\infty} A_n$ равна 1, что приводит нас к противоречию.

Итак, нельзя построить вероятностную меру на всех подмножествах $(0,1]$ с простым свойством однородности относительно сдвигов. Поэтому нам приходится выделять более маленькие множества $\mathcal{F}$, исключая оттуда некоторые "нехорошие" множества.

8.3 Свойства сигма-алгебр

Заметим, что хотя в определении сигма-алгебры присутствует только замкнутость относительно операций дополнения и счетного объединения (то есть эти операции, примененные к элементам сигма-алгебре, также дают элементы сигма-алгебры), но в действительности спектр таких операций куда шире:

- Если $A_1, \dots, A_n$ лежат в $\mathcal{F}$, то $\bigcup_{i=1}^{n} A_i \in \mathcal{F}$.
  Можно рассмотреть $A_i = \emptyset$, $i > n$, тогда $\bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{\infty} A_i$.
- Если $A_1, \dots, A_n, \dots$ лежат в $\mathcal{F}$, то $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$.
  Действительно,
  $$\bigcap_{i=1}^{\infty} A_i = \overline{\bigcup_{i=1}^{\infty} \overline{A_i}},$$
  где $\overline{A_i}$ лежат в $\mathcal{F}$, их объединение также лежит в $\mathcal{F}$, а дополнение к нему также лежит там же.
- Если A, B лежат в $\mathcal{A}$, то $A \setminus B$ также лежит в $\mathcal{A}$, поскольку $A \setminus B = A \cap \overline{B}$, а как мы видели выше операции пересечения и дополнения не выводят за пределы сигма-алегбры.
- Если A, B лежат в $\mathcal{A}$, то $A \Delta B$ также лежит в $\mathcal{A}$, поскольку $A \Delta B = A \setminus B \bigcup B \setminus A$.

Таким образом, сигма-алгебра замкнута относительно всех основных операций над множества.

8.4 Свойства вероятностных мер

Вероятностная мера также обладает значительно большим количеством свойств, чем указано в ее определении:

- $\mathbf{P}(\overline{A}) = 1 - \mathbf{P}(A)$, поскольку $\mathbf{P}(A) + \mathbf{P}(\overline{A}) = 1$.
- Верна формула включений-исключений
  $$\mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbf{P}(A_i) - \sum_{i_1 < i_2} \mathbf{P}(A_{i_1} A_{i_2}) + \sum_{i_1 < i_2 < i_3} \mathbf{P}(A_{i_1} A_{i_2} A_{i_3}) - \dots + (-1)^n \mathbf{P}(A_1 \dots A_n).$$
- Верно свойство непрерывности меры снизу: если $A_1 \subset A_2 \subset \dots \subset A_n \dots$, то
  $$\mathbf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n \to \infty} \mathbf{P}(A_n).$$
  Действительно, положим
  $$B_n = A_{n+1} \setminus A_n, \ n \geq 1.$$
  Тогда B_i не пересекаются, а значит
  $$\mathbf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathbf{P}\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \mathbf{P}(B_i) = \lim_{n \to \infty} \sum_{i=1}^{n} \mathbf{P}(B_i) = \lim_{n \to \infty} \mathbf{P}(A_{n+1}).$$

- Аналогичным образом верно свойство непрерывности меры сверху: если $A_1 \supset A_2 \supset \ldots \supset A_n \ldots$, то

$$\mathbf{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n\to\infty} \mathbf{P}(A_n).$$

Оказывается, что свойство непрерывности заменяет сигма-аддитивность:

Лемма 8.1. *Пусть отображение $\mathbf{P}$ из $\mathcal{F}$ в $[0,1]$ удовлетворяет свойствам 1) и 2) вероятностной меры и является непрерывным в нуле (то есть если $A_1 \supset A_2 \supset \ldots$), причем $\cap A_n = \emptyset$, то*

$$\mathbf{P}(A_n) \to 0, n \to \infty.$$

Тогда $\mathbf{P}$ — вероятностная мера.

Доказательство. Пусть $B_1, \ldots B_n, \ldots$ — непересекающиеся события. Докажем, что

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

Рассмотрим $A_m = \sum_{n=m}^{\infty} B_n$. Тогда A_m вложены и в пересечении дают пустое множество. В силу свойства непрерывности в нуле

$$\mathbf{P}(A_m) \to 0, m \to \infty.$$

Из свойства 2)

$$\mathbf{P}\left(\sum_{n=1}^{\infty} B_n\right) = \mathbf{P}(B_1) + \cdots + \mathbf{P}(B_{m-1]}) + \mathbf{P}(A_m).$$

Переходя в правой части к пределу по $m \to \infty$, получаем в точности требуемое. □

8.5 Продолжение меры

Давайте определим самую простую вероятностную меру — "длину" подмножеств $(0,1]$.

Начнем с множеств $(a,b]$. Для них положим $\mathbf{P}((a,b]) = b - a$.

Перейдем к множествам вида $\sum_{i=1}^{n} (a_i, b_i]$. Такие множества будем называть простыми. Эта система образует алгебру — объединение счетного числа простых множеств может уже не быть простым, но объединение конечного числа простых множеств будет также простым.

Дальше длина должна распространиться на счетные объединения этих множеств. Потом на счетные пересечения полученных множеств и так далее. Мы сможем доопределять меру на новых множествах до тех пор, пока не получим некоторую сигма-алгебру $\mathcal{F}$, на которой наша мера будет определена. Оказывается, что эта процедура не закончится никогда — каждая следующая итерация будет давать новые множества. Однако, все же можно найти минимальную сигма-алгебру.

Лемма 8.2. *Если $\mathcal{A}$ — некоторое множество подмножеств множества A, то существует минимальная сигма-алгебра $\sigma(\mathcal{A})$, содержащая $\mathcal{A}$, то есть сигма-алгебра, содержащаяся в любой сигма-алгебре, содержащей $\mathcal{A}$.*

Доказательство. Рассмотрим все сигма-алгебры $\mathcal{B}$, содержащие $\mathcal{A}$. Тогда рассмотрим пересечение $\mathcal{C}$ всех таких сигма-алгебр, то есть те подмножества множества A, которые лежат в каждой из сигма-алгебр $\mathcal{B}$. Покажем, что это сигма-алгебра:

- $\Omega \in \mathcal{C}$, поскольку Ω лежит в любой сигма-алгебре $\mathcal{B}$
- Если $C \in \mathcal{C}$, то C лежит в каждой $\mathcal{B}$, $\overline{C}$ лежит в каждой из $\mathcal{B}$, а значит $\overline{C}$ лежит в $\mathcal{C}$.
- Если $C_1, \ldots \in \mathcal{C}$, то C_i лежат в каждой $\mathcal{B}$, значит $C_1 \cup \ldots C_n \ldots$ лежит в $\mathcal{B}$, поскольку $\mathcal{B}$ — сигма-алгебры. Значит объединение лежит в $\mathcal{C}$, что и требовалось доказать.

Значит, $\mathcal{C}$ — сигма-алгебра, содержащаяся во всех содержащих $\mathcal{A}$ сигма-алгебрах $\mathcal{B}$. При этом она содержит $\mathcal{A}$, что и требуется. □

Итак, распространяя меру с исходной алгебры простых множеств, мы должны в итоге распространить ее на минимальную сигма-алгебру $\sigma(\mathcal{A})$. Однако, не вполне понятно почему мы это можем сделать непротиворечиво — вдруг мы получим для какого-то множества два различных представления в виде счетного объединения, причем сумма вероятностей будет различной в первом и втором случае.

Теорема 8.1 (Теорема Каратеодори)**.** *Пусть вероятностная мера* $\mathbf{P}$ *задана на алгебре* $\mathcal{A}$ *и сигма-аддитивна на ней Тогда ее можно продлить на минимальную сигма-алгебру* $\sigma(\mathcal{A})$*, содержащую* $\mathcal{A}$*, причем единственным образом.*

Отметим, что сигма-аддитивная на сигма-алгебре мера определялась следующим образом:
для любых непересекающихся A_i выполнено условие

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

При этом множество $\sum_{i=1}^{\infty} A_i$ лежало в нашей сигма-алгебре в силу условия сигма-аддитивности. Теперь, когда речь идет об алгебре, это уже необязательно так. Но для тех множеств A из $\mathcal{A}$, которые представимы в виде объединения счетного числа непересекающихся множеств A_i из $\mathcal{A}$, мы требуем, чтобы вероятность A была равна сумме ряда из вероятностей A_i.

Мы оставим эту теорему без доказательства, но оценим полезный результат, который она дает: можно задать меру на простых множествах, проверить на них сигма-аддитивность, а отсюда мера автоматически единственным образом продолжится на минимальную сигма-алгебру, содержащую наши простые множества.

Лемма 8.3. *Мера, заданная на полуинтервалах, содержащихся в* $(0, 1]$*, формулой* $\mathbf{P}((a, b]) = b - a$*, единственным образом продолжается на минимальную сигма-алгебру, содержащую такие полуинтервалы.*

Доказательство. Указанная мера, как мы уже обсуждали, однозначно задает вероятности всех множеств, представляющих собой конечные объединения полуинтервалов и дополнения до них. Такие множества образуют алгебру.

В силу теоремы Каратеодори нам достаточно проверить счетную аддитивность нашей меры на этой алгебре. Проверим, пользуясь полученными ранее свойствами, конечную аддитивность и непрерывность меры в нуле.

Конечная аддитивность достаточно очевидна — если полуинтервал $(a, b]$ представляет собой конечное объединение непересекающихся полуинтервалов $(a_k, b_k]$, то один из полуинтервалов (без ограничения общности первый) имеет левый конец a, а правый конец b_1, один из полуинтервалов(без ограничения общности второй) имеет левый конец b_1, а правый b_2 и так далее, последний интервал имеет левый конец b_{k-1}, а правый b. Значит сумма длин полуинтервалов равна $b_1 - a + b_2 - b_1 + \cdots + b - b_n = b - a$, что и требовалось доказать.

Докажем непрерывность нашей меры $\mathbf{P}$ в нуле. Пусть $A_n = \bigcup_{k=1}^{m_n}(a_{n,k}, b_{n,k}]$. Фиксируем $\varepsilon > 0$ и рассмотрим B_n — такие простые множества, что замыкания B_n (будем обозначать их $[B_n]$) лежат в A_n, но при этом $\mathbf{P}(B_n) > \mathbf{P}(A_n) - \varepsilon/2^n$. Например, можно взять $B_n = \bigcup_{k=1}^{m_n}(a_{n,k} + \varepsilon/2^{k+n}, b_{n,k} - \varepsilon/2^{k+n}]$, где $(a_{n,k}, b_{n,k}]$ — полуинтервалы, входящие в состав A_n. Тогда счетное пересечение $[B_n]$ содержится в счетном пересечении A_n, а значит пусто. Но тогда значит и некоторое конечное пересечение $[B_1] \cap \ldots \cap [B_{n_0}]$ пусто. Действительно, множества $C_n = [0, 1] \setminus [B_n]$ образуют открытое покрытие компакта $[0, 1]$. Выделим из него конечное подпокрытие $C_1, \ldots, C_{n_0}$. Тогда $[B_1], \ldots, [B_{n_0}]$ имеют пустое пересечение.

Значит, $\mathbf{P}(B_1 \cap \ldots \cap B_{n_0}) = 0$. Но тогда

$$\mathbf{P}(A_{n_0}) = \mathbf{P}(A_{n_0} \setminus (B_1 \cap \ldots \cap B_{n_0})) + \mathbf{P}(B_1 \cap \ldots \cap B_{n_0}) = \mathbf{P}\left(\bigcup_{i=1}^{n_0}(A_i \setminus B_i)\right) < \sum_{i=1}^{n_0} \frac{\varepsilon}{2^i} < \varepsilon.$$

Следовательно, при любом $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbf{P}(A_n) < \mathbf{P}(A_{n_0}) < \varepsilon.$$

В силу произвольности ε указанный предел равен нулю, что и требовалось доказать. □

Итак, мы построили вероятностную меру, называемую мерой Лебега. Теперь мы можем определить уже невероятностную меру

Определение 8.3. Отображение μ из сигма-алгебры $\mathcal{F}$ подмножеств Ω в $[0, +\infty)$ называется мерой, если выполнены следующие свойства:

1. аддитивность: $\mu(A + B) = \mu(A) + \mu(B)$ при любых непересекающихся A, B из $\mathcal{F}$;

2. счетная аддитивность:
$$\mu(A_1 + A_2 + \dots + A_n + \dots) = \sum_{n=1}^{\infty} \mu(A_n),$$
где A_i — произвольные непересекающиеся подмножества $\mathcal{F}$.

Если $\mu(\Omega) < +\infty$, то мера называется конечной. Выше в теореме Каратеодори мы могли рассматривать любую конечную меру.

Мы задали меру Лебега (будем обозначать ее λ) на $[0, 1]$. На любом другом отрезке $[n, n + 1]$ будет полагать меру той же самой, то есть мера множества A, получаемого сдвигом множества B из отрезка $[0, 1]$, должна быть равна мере множества A.

Наконец, для любого $A \subset \mathbb{R}$ положим
$$\lambda(A) = \sum_{n \in \mathbb{Z}} \lambda(A \cap [n, n + 1]).$$
Это мера Лебега на прямой, то, что мы привыкли называть длиной.

9 Случайные величины в общем случае

9.1 Борелевская сигма-алгебра

На прошлой лекции мы обсудили, что мера Лебега может быть распространена с полуинтервалов $(a, b]$ внутри $(0, 1]$ на минимальную сигма-алгебру, содержащую все полуинтервалы, единственным образом.

Определение 9.1. Минимальная сигма-алгебра $\mathcal{B}(A)$, содержащая все открытые подмножества множества A, называется борелевской.

Лемма 9.1. *Борелевская сигма-алгебра подмножеств* $(0, 1]$ *совпадает с минимальной сигма-алгеброй, содержащей все полуинтервалы.*

Доказательство. 1) Покажем, что сигма-алгебра, содержащая все полуинтервалы, содержит все открытые множества. Для этого заметим, что любой интервал (a, b) должен лежать в нашей сигма-алгебре, поскольку
$$(a, b) = \bigcup_{n=1}^{\infty} \left(a, b - \frac{1}{n}\right].$$
Но тогда и счетные объединения интервалов должны в ней лежать. Значит все открытые множества содержатся в нашей минимальной сигма-алгебре.
2) Покажем, что сигма-алгебра, содержащая все открытые множества, содержит все полуинтервалы. Действительно,
$$(a, b] = \bigcap_{n=1}^{\infty} \left(a, b + \frac{1}{n}\right).$$
Следовательно, любая сигма-алгебра, содержащая все полуинтервалы, содержит все открытые множества, а сигма-алгебра, содержащая все открытые множества, содержит все полуинтервалы. Значит минимальные сигма-алгебры в обоих случаях одинаковые. □

Распространив меру Лебега на всю прямую мы можем найти меру каждого борелевского множества $\mathcal{B}(\mathbb{R})$.

Аналогичным образом, можно определить меру Лебега на плоскости (то есть площадь), начав с прямоугольников вида $(a, b] \times (c, d]$, меру которых положим $(d - c)(b - a)$. Так же можно определить меру Лебега в трехмерном пространстве (объем) или в пространствах большей размерности.

9.2 Геометрическая вероятность

Примером вероятностного пространства является тройка $A \subset \mathbb{R}^k$, $\mathcal{F} = \mathcal{B}(A)$, $\mathbf{P}(B) = V(B)/V(A)$, $B \subseteq A$, где V – объем (или, выражаясь взрослым языком, мера Лебега).

Итак, исходами в нашем случае являются точки A, событиями — борелевские подмножества $\mathcal{F}$, а мерой P — частное объемов наших множеств.

Этот опыт соответствует выбору точки в множестве А наугад. Это аналог классического вероятностного пространства в дискретном случае — только здесь не просто все элементарные исходы имеют равную вероятность (все элементарные исходы имеют вероятностью 0), а все множества одного объема имеют равную вероятность.

Пример 9.1. Вася и Петя договорились встретиться на остановке в промежуток с 2 до 3 часов. Каждый из них приходит в случайное время от 2 до 3 и ждет друга 10 минут. Какая вероятность, что Вася и Петя встретятся? Выбор времени прихода Васи и Пети соответствует выбору точки (x, y) в квадрате $[2, 3]$ наугад, где

Рис. 12: Синим выделено пространство элементарных исходов, желтым — искомое событие

x отражает время прихода Васи, а y — Пети. При этом в силу условия естественно считать нашим вероятностным пространством

$$(\Omega,\ \mathcal{F},\ \mathbf{P}) = ([2,3]^2,\ \mathcal{B}([2,3]^2),\ S),$$

где S — площадь. Нас интересует вероятность события

$$A = \{(x, y) \in [2,3]^2 : |x - y| < 1/6\}.$$

Это дополнение двух треугольников, общей площади 25/36 до квадрата площади 1, а значит множество площади 11/36. Итак,

$$\mathbf{P}(A) = S(A) = 11/36.$$

9.3 Случайные величины

Рассмотрим пространство $(\Omega, \mathcal{F}, \mathbf{P})$. Мы бы хотели определить на нем случайные величины — функции из Ω в $\mathbb{R}$. Можем ли мы, как и прежде, рассматривать произвольные функции?

Будем понимать под $f^{-1}(A)$, где A — некоторое множество, а f — отображение, множество вида $\{x : f(x) \in A\}$.

Для наших величин мы бы хотели считать вероятности попадания их в разные множества A, по-крайней мере в отрезки или полуинтервалы. Но ведь $\mathbf{P}(\omega : X(\omega) \in [a, b]) = \mathbf{P}(X^{-1}([a, b]))$, а $\mathbf{P}$ — отображение из $\mathcal{F}$ в $[0, 1]$, а, значит, множества вида $X^{-1}([a, b])$ должны лежать в $\mathcal{F}$. С другой стороны, если я возьму $X^{-1}(A) \cup X^{-1}(B)$, то это будет $X^{-1}(A \cup B)$, а $X^{-1}(\overline{A})$ — это $\overline{X^{-1}(A)}$. следовательно, пробраз сигма-алгебры при X будет сигма-алгеброй. Следовательно, если прообразы $[a, b]$ лежат в $\mathcal{F}$, то и прообразы множеств из $\mathcal{B}(\mathbb{R})$ (борелевской сигма-алгебры) также лежат в $\mathcal{F}$. Таким образом, мы приходим к следующему естественному определению случайной величины:

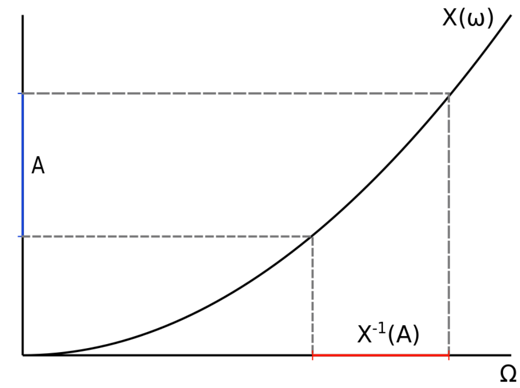
Определение 9.2. *Случайная величина* — это отображение из Ω в $\mathbb{R}$, такое, что прообраз любого борелевского множества $\mathbb{R}$ есть событие из сигма-алгебры $\mathcal{F}$.

Такое отображение называется измеримым.

Пример 9.2. Простейшим примером случайной величины является индикатор события $A \in \mathcal{F}$:

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

Рис. 13: Полный прообраз любого борелевского множества на оси абсцисс должен быть множеством из $\mathcal{F}$



Это действительно случайная величина, поскольку прообраз любого множества из $\mathbb{R}$ является либо A, либо $\overline{A}$, либо Ω, либо $\emptyset$, а все эти множества лежат в $\mathcal{F}$, поскольку там лежит A.

Пример 9.3. Пусть $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$, $X_1(\omega) = \omega$. Тогда $X_1(\omega)$, очевидно, будет случайной величиной, поскольку прообраз $X_1^{-1}([a,b])$ имеет вид $[a,b] \cap [0,1]$.

Вопрос 9.1. Будет ли случайной величиной на том же пространстве $Y(\omega) = \omega^2$? $Z(\omega) = \arcsin\omega$?

Пример 9.4. Пусть $\Omega = [0,1]$, $\mathcal{F} = \{\emptyset, \Omega\}$. Тогда $X_1(\omega) = \omega$ уже не будет случайной величиной, поскольку $X_1^{-1}([0,1/2]) = [0,1/2]$, а $[0,1/2]$ не лежит в нашей сигма-алгебре $\mathcal{F}$. Как мы видим, многое зависит от сигма-алгебры, с которой мы работаем.

Проверять то, что некоторое отображение X — случайная величина, удобнее проверяя то, что $\{\omega : X(\omega) \le x\}$ при любом x лежит в $\mathcal{F}$, поскольку множества $(-\infty, x]$ порождают сигма-алгебру.

9.4 Случайные векторы

Определение 9.3. *Случайный вектор* — это такое отображение из Ω в $\mathbb{R}^n$, что прообраз любого множества из $\mathcal{B}(\mathbb{R}^n)$ есть событие из сигма-алгебры $\mathcal{F}$.

Для проверки того, что отображение $\vec{X} : \Omega \to \mathbb{R}^n$ есть случайный вектор, достаточно убедиться, что

$$\{\omega : \vec{X}(\omega) \in [a_1, b_1] \times \dots \times [a_n, b_n]\} = \bigcap_{i=1}^{n} \{\omega : X_i(\omega) \in [a_i, b_i]\} \in \mathcal{F},$$

где $\vec{X} = (X_1, \dots, X_n)$. Но последнее соотношение равносильно тому, что все X_i являются случайными величинами.

Поэтому случайный вектор – это просто набор случайных величин $X_1, \dots, X_n$.

9.5 Случайные элементы

Если мы захотим рассматривать отображения не в $\mathbb{R}$ или $\mathbb{R}^n$, а в некоторое другое пространство U с сигма-алгеброй $\mathcal{V}$, то определение будет похожим:

Определение 9.4. *Случайный элемент* — это такое отображение из Ω в U, что прообраз любого множества из $\mathcal{V}$ есть событие из сигма-алгебры $\mathcal{F}$.

Определение 9.5. *Борелевской функцией* назовем отображение из $\mathbb{R}^n$ в $\mathbb{R}^m$, измеримое около борелевской сигма-алгебры (то есть прообраз любого борелевского множества борелевский).

Борелевскими функциями из $\mathbb{R}^n$ в $\mathbb{R}$ в частности будут все непрерывные функции, поскольку у них прообраз открытых множеств — открытые, а значит борелевские множества, а значит прообраз любого борелевского борелевский, поскольку открытые множества порождают борелевскую сигма-алгебру.

Если $X_1, ..., X_n$ — случайные величины, то любая борелевская функция из $\mathbb{R}^n$ в $\mathbb{R}$ от них также будет случайной величиной. Это следует из того, что

$$\{\omega : f(X_1, ..., X_n) \in B\} = \{\omega : (X_1, ..., X_n) \in f^{-1}(B)\},$$

а $f^{-1}(B)$ лежит в $\mathcal{B}(\mathbb{R}^n)$, если B было из $\mathcal{B}(\mathbb{R})$, поскольку f — борелевская функция.

В частности, если $X_1, ..., X_n$ — случайные величины, то $X_{(1)} = \min_{i\le n} X_i, X_{(2)} = \max_{i\le n} \min_{i\ne j} X_i, ..., X_{(n)} = \max X_i$ (то есть упорядоченные по возрастанию X_i) тоже будут случайными величинами. Такой ряд величин называется *вариационным рядом.*

Пример 9.5. Пусть $\Omega = [0,1]$, $\mathcal{B}([0,1])$. Положим $X_1(\omega) = \omega,\ X_2(\omega) = 1-\omega,\ X_3(\omega) = 1/4$. Тогда

$$X_{(1)}(\omega) = \begin{cases} \omega, & \omega \in \left[0, \frac{1}{4}\right], \\ \frac{1}{4}, & \omega \in \left(\frac{1}{4}, \frac{3}{4}\right), \\ 1-\omega, & \omega \in \left[\frac{3}{4}, 1\right]. \end{cases} \qquad X_{(2)}(\omega) = \begin{cases} \frac{1}{4}, & \omega \in \left[0, \frac{1}{4}\right], \\ \omega, & \omega \in \left(\frac{1}{4}, \frac{1}{2}\right], \\ 1-\omega, & \omega \in \left(\frac{1}{2}, \frac{3}{4}\right), \\ \frac{1}{4}, & \omega \in \left[\frac{3}{4}, 1\right]. \end{cases} \qquad X_{(3)}(\omega) = \begin{cases} 1-\omega, & \omega \in \left[0, \frac{1}{2}\right], \\ \omega, & \omega \in \left(\frac{1}{2}, 1\right]. \end{cases}$$

9.6 Расширенные случайные величины

Определение 9.6. *Расширенной* (или несобственной) случайной величиной называют измеримое отображение, действующее в $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$.

Иначе говоря, мы искусственно добавляем к нашим значениям $+\infty$ и $-\infty$, а в борелевскую сигма-алгебру вместе с каждым множеством A добавляем $A \cup \{-\infty\}$, $A \cup \{-\infty\}$, $A \cup \{+\infty\} \cup \{-\infty\}$.
Если $X_1, X_2, \ldots$ — случайные величины, то не только $X_1 + \cdots + X_n$, $X_1 \cdot X_2$, $\sup_{i\le n} X_i$, являющиеся борелевскими функциями от $X_1, \ldots, X_n$, но и, скажем, $\limsup X_i$ также будет случайной величиной (возможно, расширенной). Как в этом убедиться?

Пример 9.6. Рассмотрим $X_1, \ldots, X_n, \ldots$ и $Y(\omega) = \limsup_{n\to\infty} X_n(\omega)$. Тогда

$$\{\omega : Y(\omega) < x\} = \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{\omega : X_n(\omega) < x\}.$$

В силу полученного представления и того, что X_n случайные величины, имеем, что и Y – случайная величина. Как получить такое представление? Нужно записать то, что верхний предел меньше x:

$$\exists N :\ \forall n > N\ X_n < x,$$

а затем $\forall$ заменить на пересечение, а $\exists$ на объединение. Подумайте, почему это законно.

9.7 Распределение случайной величины

Определение 9.7. Распределением случайной величины X будем называть набор вероятностей $\mathbf{P}(\omega : X(\omega) \in B)$ для $B \in \mathcal{B}(\mathbb{R})$.

Как мы поняли на прошлой лекции, чтобы задать такие меры, достаточно задать вероятности всех полуинтервалов $\mathbf{P}(\omega : X(\omega) \in (a, b])$. Более удобно задавать вероятности лучей $(-\infty, x]$, откуда легко получить вероятности попадания в полуинтервал

$$\mathbf{P}(\omega : X(\omega) \in (a,b]) = \mathbf{P}(\omega : X(\omega) \le b) - \mathbf{P}(\omega : X(\omega) \le a).$$

Определение 9.8. Величина $\mathbf{F}_X(x) = \mathbf{P}(\omega : X(\omega) \le x)$ называется функцией распределения (ф.р.).

Пример 9.7. Для случайной величины $X_1(\omega)$ из примера 9.3 функцией распределения будет $xI_{x\in[0,1]} + 1I_{x>1}$. Такое распределение называется равномерным на $[0,1]$. Для величины $X_2(\omega)$ на том же вероятностном пространстве, равной $X_2(\omega) = 1 - \omega$, функция распределения окажется точно такой же.

Какими свойствами обладает функция распределения?

1. $\mathbf{F}(x) \to 0$ при $x \to -\infty$, $\mathbf{F}(x) \to 1$ при $x \to +\infty$;
2. $\mathbf{F}(x)$ монотонно неубывает;
3. $\mathbf{F}(x)$ непрерывна справа (по свойству непрерывности меры).

При этом слева функция $\mathbf{F}$ непрерывна не будет, поскольку из непрерывности меры

$$\lim_{x_n \to x-0} F_X(x_n) = \mathbf{P}(X \in (-\infty, x)),$$

а вероятность в правой части не будет равна $F_X(x)$, если $\mathbf{P}(X = x) > 0$. Таким образом, ф.р. будет разрывна в тех и только тех точках x, для которых $\mathbf{P}(X = x) > 0$. При этом высота разрыва в каждом случае будет в точности $\mathbf{P}(X = x)$.

Оказывается, что условия 1)–3) необходимы и достаточны для того, чтобы $\mathbf{F}$ была ф.р. некоторой вероятностной меры $\mathbf{P}$.

Лемма 9.2. *Любая функция $\mathbf{F}$, удовлетворяющая 1)–3), соответствует некоторой величине X, то есть существует такая случайная величина X, что $\mathbf{F}$ — ее ф.р. При этом $\mathbf{P}(X \in B)$, $B \in \mathcal{B}$, однозначно будет соответствовать ф.р. $\mathbf{F}(x)$.*

Доказательство. Зададим меру на простых множествах вида $A = \bigcup_{i=1}^{n}(a_i, b_i]$, где $(a_i, b_i]$ не пересекаются, формулой

$$\mathbf{P}(A) = \sum_{i=1}^{n}(\mathbf{F}(b_i) - \mathbf{F}(a_i)).$$

Благодаря монотонности заданная мера будет неотрицательной, а мера прямой будет единицей:

$$\mathbf{P}((-\infty, \infty)) = \lim_{n\to\infty} \mathbf{P}((-n, n)) = \mathbf{F}(n) - \mathbf{F}(-n) \to 1.$$

Проверим, что на этом множестве множеств мера σ-аддитивна. Тогда по теореме Каратеодори ее можно будет продолжить на всю борелевскую сигма-алгебру, причем единственным образом. Доказательство аналогично доказательству сигма-аддитивности на простых множествах меры Лебега.

Действительно, конечная аддитивность вытекает из того, что если простое множество A представляется в виде непересекающегося объединения множеств $A_1, \ldots, A_n$, то полуинтервалы, входящие в A, разбиваются в объединение полуинтервалов, входящих в A_i. При этом как мера самого полуинтервала $(a, b]$, так и сумма мер полуинтервалов, на которые он разбит, задаются одной и той же формулой: $\mathbf{F}(b) - \mathbf{F}(a)$.

Для доказательства счетной аддитивности, как и прежде, покажем непрерывность в нуле нашей меры: если простые множества A_n вложены ($A_n \supset A_{n+1}$) и их пересечение пусто, то $\mathbf{P}(A_n)$ стремится к нулю.

Действительно, пусть A_n лежат в некотором компакте $[-N, N]$. Тогда поскольку $\mathbf{F}$ непрерывна справа, то

$$\mathbf{P}((a, b]) = \mathbf{F}(b) - \mathbf{F}(a) = \lim_{x \to a+0}(\mathbf{F}(b) - \mathbf{F}(x)).$$

Дальше доказательство практически повторяет доказательство леммы 8.3.

Как и прежде найдутся такие вложенные простые множества B_n, что замыкания $[B_n]$ содержатся в A_n и

$$\mathbf{P}(B_n) > \mathbf{P}(A_n) - \varepsilon/2^n.$$

Тогда замыкания $[B_n]$ имеют пустое пересечение.

Если взглянуть на множества $[-N, N] \setminus [B_n]$, то они образуют открытое покрытие $[-N, N]$, а значит из него можно выделить конечное подпокрытие. Но тогда соответствующие ему $[B_n]$ имеют пустое пересечение.

Значит с какого-то момента пересечение первых n_0 множеств B_i пусто. Но тогда $\mathbf{P}(A_n) < \varepsilon/2^n$ при всех $n > n_0$, что и требуется.

Пусть A_n имеют общий вид и не лежат в отрезке $[-N, N]$. Рассмотрим такое N, что $\mathbf{P}((-N, N]) > 1 - \varepsilon$.

Тогда $\mathbf{P}(A_n \cap (-N, N])$ стремятся к нулю в силу доказанного, а $\mathbf{P}(A_n \cap \overline{(-N, N]}) < \varepsilon$. В силу произвольности $\varepsilon > 0$ имеем требуемое. $\square$

9.8 Абсолютно непрерывные распределения

Помимо рассмотренного нами ранее дискретного случая, важным случаем, в котором удобно работать, является случай *абсолютно непрерывного* распределения.

Определение 9.9. Величина с ф.р. $F(x)$ называется абсолютно непрерывной, когда существует функция $f_X(x)$, называемая *плотностью*, такая, что $F(x) = \int_{-\infty}^{x} f_X(t)dt$.

Мы понимаем, что $\int_{-\infty}^{\infty} f_X(x)dx = 1$ и $f_X(x) \geq 0$. При этом для любого $A \in \mathcal{B}(\mathbb{R})$

$$\mathbf{P}(X \in A) = \int_A f_X(x)dx,$$

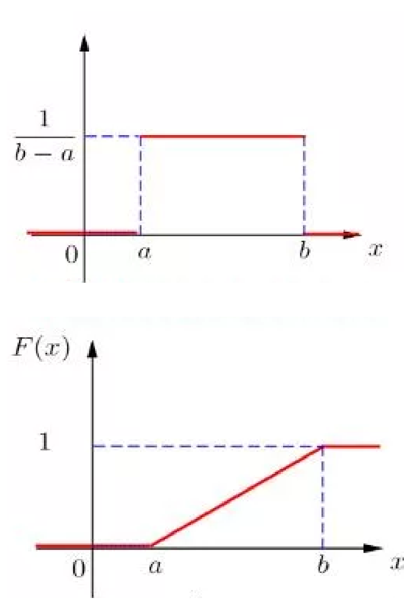
правда, возникает сложный вопрос, что такое интеграл по произвольному множеству, на который мы ответим в следующий раз.

Функция $f_X(x)$ определяется с точностью до множества меры Лебега 0 и равна (с точностью до такого множества) $\mathbf{F}'(x)$.

Пример 9.8. Для примера 2 функцию распределения X можно представить в виде $\int_{-\infty}^{x} f_X(x)dx$, где $f_X(x) = \mathbf{F}'(x) = 1$ при $x \in (0,1)$, $f_X(x) = \mathbf{F}'(x) = 0$ при $x < 0$ и $f_X(x) = \mathbf{F}'(x) = 0$ при $x > 1$. При $x = 0$ и $x = 1$ для определенности положим $f_X(x) = 1$. Мы видим, что плотность будет постоянной на отрезке функцией, равной вне отрезка нулю, что соответствует интуитивному понятию о равномерности.

Соответствующее распределение называется стандартным равномерным распределением и представляет собой частный случай равномерного распределения на отрезке $[a,b]$, описанного ниже.

Рис. 14: Плотность и функция распределения равномерного распределения на отрезке $[a,b]$



Плотность выполняет роль, схожую с $\mathbf{P}(X = x_i)$ для дискретных величин. В силу определения производной

$$f_X(x) = \lim_{u\to\infty} \frac{\mathbf{P}(X \in (x-u, x+u])}{2u},$$

то есть плотность — это 'удельная вероятность' попадания в окрестность точки. Большее значение плотности делает попадание в маленькую окрестность точки более вероятным.

Приведем список основных абсолютно-непрерывных распределений:

1. $f(x) = 1/(b-a),\ x \in [a,b]$ - *равномерное распределение* на отрезке $[a,b]$. Обозначают его $R[a,b]$ или $U[a,b]$;
2. $f(x) = \lambda e^{-\lambda x},\ x > 0, \lambda > 0$ — *экспоненциальное распределение* с параметром λ. Обозначение $\exp(\lambda)$;

3. $f(x) = \lambda/2e^{-\lambda|x|}, \lambda > 0$ — *распределение Лапласа*;

4. $f(x) = (\sqrt{2\pi})^{-1}e^{-x^2/2}$ — *стандартное нормальное распределение*. С ним мы уже встречались, когда сталкивались с центральной предельной теоремой. Обозначение $\mathcal{N}(0,1)$;

5. $f(x) = (\sqrt{2\pi\sigma^2})^{-1}e^{-(x-a)^2/(2\sigma^2)}$ — *нормальное распределение* с параметрами a, σ^2. Это распределение величины $\sigma X + a$, где X — стандартная нормальная величина. Обозначение $\mathcal{N}(a,\sigma^2)$;

6. $f(x) = x^{a-1}e^{-x/b}b^{-a}/\Gamma(a)$, $x > 0$, $b > 0$, $a > 0$ — *гамма-распределение*. Заметим, что экспоненциальное есть $Gamma(1, 1/\lambda)$. Обозначение $Gamma(a,b)$;

7. $f(x) = x^{b-1}(1-x)^{a-1}/B(a,b)$, $0 < x < 1$, $a, b > 0$ — *бета-распределение*. Обозначение $Beta(a,b)$;

8. $f(x) = 1/(\pi(1+x^2))$ — *распределение Коши*. Обозначение $Cauchy$.

10 Математическое ожидание в общем случае

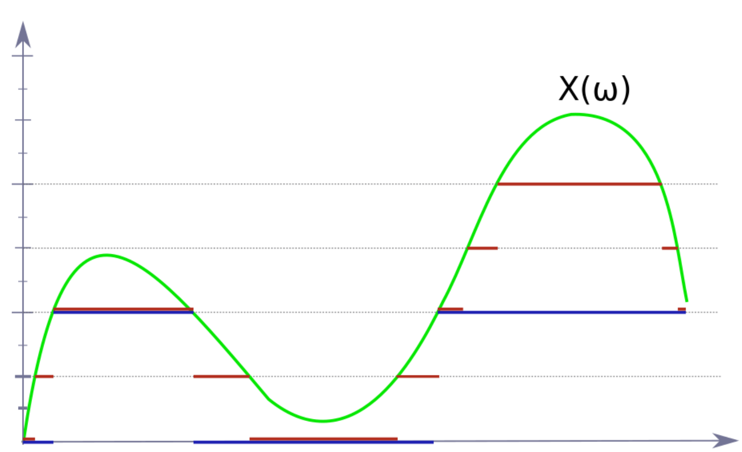
10.1 Общее определение математического ожидания

Сегодня мы обобщим введенное нами для дискретных величин понятие математического ожидания на общий случай. Для этого воспользуемся следующей конструкцией:

1. Для неотрицательной величины X с конечным множеством значений $\{x_0, \ldots, x_n\}$ (такие величины будем называть простыми) положим
$$\mathbf{E}X = \sum_{k=0}^{n} x_k \mathbf{P}(X = x_k),$$
как это было раньше.

2. Для произвольной неотрицательной величины рассмотрим последовательность X_n, монотонно сходящуюся к X ($X_n \uparrow X$) при каждом ω, где X_n принимают конечное число значений. Например, можно взять
$$X_n(\omega) = \sum_{k=0}^{n2^n} (k/2^n) I_{X(\omega)\in(k/2^n, k+1/2^n]} + nI_{X(\omega)>n}.$$

Рис. 15: Случайная величина X (зеленым цветом) и ее дискретные приближения X_1 (синим цветом) и X_2 (красным цветом).



В силу монотонности у последовательности $\mathbf{E}X_n$ есть предел (возможно бесконечный), его и назовем $\mathbf{E}X$.

3. Для произвольной величины представим ее в виде $X = X^+ - X^-$, где $X^+ = \max(X, 0)$, $X^- = -\min(X, 0)$ как и прежде.

   Опять же, в случае $\mathbf{E}X^+ = \mathbf{E}X^- = +\infty$ будет говорить, что математического ожидания у X не существует

   В случае, если $\mathbf{E}X^+ = +\infty$, $\mathbf{E}X^- < +\infty$, будем говорить, что математическое ожидание равно $+\infty$. В случае, если $\mathbf{E}X^+ < +\infty$, $\mathbf{E}X^- = +\infty$, будем говорить, что математическое ожидание равно $-\infty$.

Ключевым моментом, разумеется является пункт 2. Чтобы определение стало легальным, мы должны показать, что для разных последовательностей X_n, сходящихся к X, предел $\mathbf{E}X_n$ будет одним и тем же. Нам понадобится следующая лемма

Лемма 10.1. *Если $X_n \uparrow X$ и $X \geq Y$, где X_n, Y — дискретные неотрицательные величины с конечным числом значений, то $\lim \mathbf{E}X_n \geq \mathbf{E}Y$.*

Доказательство. Пусть $A_n = \{\omega : X_n(\omega) \geq Y(\omega) - \varepsilon\}$. Поскольку $X_n \uparrow X$, а $X \geq Y$, то $A_n \uparrow \Omega$, а значит $\mathbf{P}(A_n) \to 1$ в силу непрерывности вероятностной меры.

С другой стороны,

$$\mathbf{E}X_n \geq \mathbf{E}X_n I_{A_n} \geq \mathbf{E}(Y - \varepsilon) I_{A_n} = \mathbf{E}Y - \mathbf{E}Y I_{\overline{A_n}} - \varepsilon \mathbf{P}(A_n) \geq \mathbf{E}Y - y_m(1 - \mathbf{P}(A_n)) - \varepsilon,$$

где y_m — наибольшее из значений, принимаемых Y с положительной вероятностью. Устремляя n к бесконечности, а ε к нулю, имеем требуемое. $\square$

Отсюда для двух последовательностей $X_n \uparrow X$, $Y_n \uparrow X$ получим, что $\lim_{n\to\infty} \mathbf{E}X_n \geq \mathbf{E}Y_m$ при всех m. В силу произвольности m

$$\lim_{n\to\infty} \mathbf{E}X_n \geq \lim_{m\to\infty} \mathbf{E}Y_m.$$

Из симметрии выполнено обратное неравенство, откуда пределы математическим ожиданий обеих последовательностей одинаковы. Значит, определение $\mathbf{E}X$ во втором пункте не привязано к выбору последовательности X_n, монотонно стремящейся к X.

Таким образом, можно посмотреть на определение математического ожидания (в более общем случае, если P — не обязательно вероятностная мера, оно называется интеграл Лебега) как на предел

$$\lim_{n\to\infty} \left(\sum_{k=1}^{n2^n} \frac{k}{2^n} \mathbf{P}\left(\omega : X(\omega) \in \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right) + n\mathbf{P}(\omega : X(\omega) \geq n) \right).$$

Здесь можно провести аналогию с интегралом Римана, только разбиваем на отрезки мы не область интегрирования, а область значений.

Пример 10.1. Рассмотрим пространство $([0,1], \mathcal{B}([0,1]), \lambda)$, где λ — мера Лебега. Рассмотрим величину $X(\omega) = I_{\omega\in\mathbb{Q}}$. Как функция из $[0,1]$ эта функция не интегрируема по Риману, поскольку верхняя сумма Дарбу для нее всегда равна 1, а нижняя — всегда равна 0.

Однако, интеграл Лебега на $[0,1]$ у нее определен и равен 0, поскольку X — простая случайная величина, принимающая 2 значений: единицу с вероятностью

$$\mathbf{P}(\omega \in \mathbb{Q}) = \lambda([0,1] \cap \mathbb{Q}) = 0,$$

ноль с вероятностью единица.

Будем использовать для $\mathbf{E}X$ также форму записи $\int_\Omega X(\omega) P(d\omega)$.

10.2 Свойства математического ожидания

Определение математического ожидания выглядит довольно сложным. Как же убедиться в том, что для него выполнены простые свойства, справедливые в дискретном случае, например линейность?

1. $\mathbf{E}(aX + bY) = a\mathbf{E}X + b\mathbf{E}Y$.

Доказательство. Разобьем доказательство на несколько шагов.

- $\mathbf{E}(-X) = -\mathbf{E}X$, поскольку $(-X)^+ = X^-$, $(-X)^- = X^+$.
- $\mathbf{E}(aX) = a\mathbf{E}X$ для положительных a.
  - Для дискретных величин мы такое свойство уже проверяли.
  - Для произвольных неотрицательных величин оно прямо следует из того, что если $X_n \uparrow X$, то $aX_n \uparrow aX$, $n \to \infty$.
  - Для величин произвольного знака следует из того что $(aX)^+ = aX^+$, $(aX)^- = aX^-$.
- $\mathbf{E}(X+Y) = \mathbf{E}X + \mathbf{E}Y$
  - Для дискретных величин X, Y мы уже получали это соотношение.
  - Для неотрицательных величин X, Y оно следует из того, что для $X_n \uparrow X$, $Y_n \uparrow Y$, $n \to \infty$, верно соотношение $X_n + Y_n \uparrow X + Y$, $n \to \infty$.
  - Если величина X неотрицательна, а Y — неположительна, а $X+Y$ при этом неотрицательна, то $\mathbf{E}(X+Y) + \mathbf{E}(-Y) = \mathbf{E}X$ в силу аддитивности для положительных величин. Следовательно, $\mathbf{E}(X+Y) = \mathbf{E}X + \mathbf{E}Y$.
    Аналогичный результат получаем для случая неположительной $X+Y$.
  - Если X неотрицательна, а Y неположительна, то по определению

    $$\mathbf{E}(X+Y) = \mathbf{E}(X+Y)I_{X+Y\geq 0} + \mathbf{E}(X+Y)I_{X+Y\leq 0}.$$

    Но величины $XI_{X+Y\geq 0}$, $YI_{X+Y\geq 0}$ удовлетворяют условиям предыдущего пункта. Следовательно,

    $$\mathbf{E}(X+Y)I_{X+Y\geq 0} = \mathbf{E}XI_{X+Y\geq 0} + \mathbf{E}YI_{X+Y\geq 0}.$$

    Аналогичным образом расписывая вторую сумму, получим

    $$\mathbf{E}(X+Y)I_{X+Y\leq 0} = \mathbf{E}XI_{X+Y\leq 0} + \mathbf{E}YI_{X+Y\leq 0}.$$

    Остается сложить полученные выражения и воспользоваться предыдущим пунктом, откуда

    $$\mathbf{E}(X+Y) = \mathbf{E}X + \mathbf{E}Y$$

    и в этом случае.
  - Для произвольных величин нам необходимо доказать, что $\mathbf{E}(X+Y)^+ - \mathbf{E}(X+Y)^- = \mathbf{E}X^+ + \mathbf{E}Y^+ + \mathbf{E}X^- + \mathbf{E}Y^-$.
    Представим $(X+Y)^+$ в виде суммы $(X+Y)I_{A_1} + (X+Y)I_{A_2} + (X+Y)I_{A_3}$, где

    $$A_1 = \{\omega : X \geq 0, Y \geq 0\}, \quad A_2 = \{\omega : X \geq 0, Y < 0, X+Y \geq 0\}, \quad A_3 = \{\omega : X < 0, Y \geq 0, X+Y \geq 0\}.$$

    При этом каждая из величин XI_{A_i}, YI_{A_i} — постоянного знака, откуда для них справедлива аддитивность. Аналогичным образом представим $(X+Y)^-$ с помощью событий $B_1, ..., B_3$, где B_i отличается от A_i тем, что все знаки неравенств заменены на противоположные. Таким образом,

    $$\mathbf{E}(X+Y) = \mathbf{E}X(I_{A_1} + I_{A_3} + I_{B_2}) + \mathbf{E}X(I_{B_1} + I_{B_3} + I_{A_2}) + \mathbf{E}Y(I_{A_1} + I_{A_2} + I_{B_3}) + \mathbf{E}Y(I_{B_1} + I_{B_2} + I_{A_3})$$

    Пользуясь тем, что первые две величины являются неотрицательной и неположительной, мы можем привести $\mathbf{E}X(I_{A_1} + I_{A_3} + I_{B_2}) + \mathbf{E}X(I_{B_1} + I_{B_3} + I_{A_2})$ к виду $\mathbf{E}X$. Аналогично со второй частью. Линейность доказана.

□

2. Для $X \geq Y$ п.н. и $\mathbf{E}X$, $\mathbf{E}Y$ конечны, то $\mathbf{E}X \geq \mathbf{E}Y$.

Доказательство. В силу аддитивности достаточно доказать, что если $X - Y \geq 0$, то $\mathbf{E}(X-Y) \geq 0$. Это так непосредственно по определению математического ожидания неотрицательной величины. □

3. $\mathbf{E}X \le \mathbf{E}|X|$.
   Свойство вытекает из предыдущего, поскольку $-|X| \le X \le |X|$.

4. Если $X = Y$ п.н. (то есть $\mathbf{P}(\omega : X(\omega) = Y(\omega)) = 1$), то и $\mathbf{E}X = \mathbf{E}Y$.

5. (Теорема о монотонной сходимости). Если $X_n(\omega) \uparrow X(\omega)$ (как числовая последовательность при каждом ω) и $\mathbf{E}X_0 > -\infty$, то $\mathbf{E}X_n \to \mathbf{E}X$ (конечному или бесконечному). Аналогичный результат справедлив для монотонно убывающих последовательностей.

   Доказательство. Пусть $X \ge 0$. Тогда рассмотрим последовательности простых функций $Y_{m,n}(\omega)$, монотонно сходящиеся к X_n при $m \to \infty$. Рассмотрим $Y_n(\omega) = \max_{m \le n} Y_{m,n}(\omega)$. Это монотонно сходящаяся последовательность простых случайных величин, причем $Y_n(\omega) \le X_n(\omega) \le X$, поскольку все $Y_{m,n}(\omega) \le X_n(\omega)$. Следовательно, Y_n имеет предел Y при каждом ω, причем

$$\mathbf{E}Y = \lim_{n\to\infty} \mathbf{E}Y_n \le \lim_{n\to\infty} \mathbf{E}X_n(\omega) \le \mathbf{E}X.$$

   С другой стороны, $Y = X$, поскольку а) $Y(\omega) \le X(\omega)$ при всех ω и б) $Y(\omega) \ge X_n(\omega)$ при всех n, а $X_n(\omega) \to X(\omega)$. Значит,

$$\mathbf{E}X \le \lim_{n\to\infty} \mathbf{E}X_n(\omega) \le \mathbf{E}X,$$

   откуда вытекает требуемое рассуждение.

   Для произвольной случайной величины X с $-\infty < \mathbf{E}X_0 < +\infty$ можно рассмотреть последовательность $X_n - X_0$, которая удовлетворяет условиям предыдущей теоремы.

   Наконец для случая $\mathbf{E}X_0 = +\infty$ всех X_n, X также имеют бесконечное математическое ожидание, поскольку их положительные части не меньше чем у X_0, а все отрицательные части не больше.

   Заметим, что без требования монотонности теорема, вообще говоря, не верна.

   Пример 10.2. Если $X_n(\omega) \to X(\omega)$ при $n \to \infty$ при всех ω, то $\mathbf{E}X_n$ может к $\mathbf{E}X$ и не стремиться. Пусть, например, $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$, $\mathbf{P}$ — мера Лебега, $X_n = nI_{(0,1/n)}$. Тогда $X_n(\omega) \to 0$ для всякого ω, но $\mathbf{E}X_n = 1 \not\to 0$. □

6. (Лемма Фату). Если $|X_n(\omega)| \le Y(\omega)$, $\mathbf{E}Y < \infty$, то

$$\mathbf{E} \liminf_{n\to\infty} X_n \le \liminf_{n\to\infty} \mathbf{E}X_n \le \limsup_{n\to\infty} \mathbf{E}X_n \le \mathbf{E} \limsup_{n\to\infty} X_n.$$

   Доказательство. Рассмотрим $Y_n(\omega) = \inf_{m \ge n} X_m(\omega)$. Тогда $Y_n(\omega)$ монотонно возрастает по n и $Y_0(\omega) \ge Y(\omega)$ не может иметь математическое ожидание минус бесконечность. Значит

$$\mathbf{E} \liminf_{n\to\infty} X_n = \mathbf{E} \lim_{n\to\infty} \inf_{m\ge n} X_m = \lim_{n\to\infty} \mathbf{E}Y_n \le \lim_{n\to\infty} \inf_{m\ge n} \mathbf{E}X_m = \liminf_{n\to\infty} \mathbf{E}X_n.$$

   Аналогично доказывается верхняя оценка. □

7. (Теорема о мажорируемой сходимости) Пусть $X_n \to X$ при $n \to \infty$ при каждом ω и $|X_n| \le Y$, где $\mathbf{E}Y$ конечно. Тогда $\mathbf{E}X_n \to \mathbf{E}X$.

   Доказательство. В силу леммы Фату

$$\mathbf{E} \liminf_{n\to\infty} X_n \le \liminf_{n\to\infty} \mathbf{E}X_n \le \limsup_{n\to\infty} \mathbf{E}X_n \le \mathbf{E} \limsup_{n\to\infty} X_n.$$

   Однако, левая и правая части совпадают, откуда

$$\mathbf{E} \lim X_n = \lim \mathbf{E}X_n,$$

   что и требовалось доказать. □

8. Пусть X, Y — независимые величины, тогда $\mathbf{E}XY = \mathbf{E}X\mathbf{E}Y$.

Доказательство. Напомним, что под независимыми величинами подразумеваются такие X, Y, что

$$\mathbf{P}(X \in A, Y \in B) = \mathbf{P}(X \in A)\mathbf{P}(Y \in B).$$

при всех A, B из $\mathcal{B}(\mathbb{R})$.

- Пусть X, Y неотрицательные и простые. Тогда данная формула уже доказывалась в первой части курса.
- Пусть X, Y неотрицательны и необязательно простые. Тогда мы можем рассмотреть последовательности простых X_n, Y_n (например, заданных в пункте 2) определения математического ожидания), которые также будут независимыми, причем $X_n \uparrow X$, $Y_n \uparrow Y$, $X_nY_n \uparrow XY$. Тогда $\mathbf{E}X_nY_n = \mathbf{E}X_n\mathbf{E}Y_n$, причем $\mathbf{E}X_nY_n \to \mathbf{E}XY$, $\mathbf{E}X_n \to \mathbf{E}X$, $\mathbf{E}Y_n \to \mathbf{E}Y$, $n \to \infty$, что и требовалось доказать.
- Если же X, Y произвольны, то разобьем X на $X^+ - X^-$, Y на $Y^+ - Y^-$, XY на $X^+Y^+ + X^-Y^- - X^+Y^- - Y^+X^-$. Применяя к каждому случаю уже известные равенства, получим требуемое.

□

9. (Неравенство Иенсена) Если X — случайная величина, g — выпуклая вниз функция, $\mathbf{E}g(X)$ и $\mathbf{E}X$ существуют, то $\mathbf{E}g(X) \le g(\mathbf{E}X)$.

Доказательство. Доказательство, приведенное в дискретном случае, остается без изменений. □

10. (Неравенство Ляпунова) Если $0 < p < q$, то для любой Y

$$(\mathbf{E}|Y|^p)^{1/p} \le (\mathbf{E}|Y|^q)^{1/q}$$

Доказательство. Доказательство, приведенное в дискретном случае, остается без изменений. □

11. (Неравенство Гельдера) Если $p, q > 0$, $1/p+1/q = 1$ и $\mathbf{E}|X|^p$, $\mathbf{E}|Y|^q$ конечны, то $\mathbf{E}XY \le (\mathbf{E}|X|^p)^{1/p}(\mathbf{E}|Y|^q)^{1/q}$. При этом равенство возможно только если $|X|^p = |Y|^p$ п.н. или одна из X, Y есть 0 п.н.

Доказательство. Доказательство, приведенное в дискретном случае, остается без изменений. □

12. (Неравенство Коши-Буняковского). Пусть $\mathbf{E}X^2 < \infty$, $\mathbf{E}Y^2 < \infty$, тогда

$$\mathbf{E}XY \le (\mathbf{E}X^2)^{1/2}(\mathbf{E}|Y|^2)^{1/2}.$$

13. (Неравенство Минковского). Пусть $p \ge 1$. Тогда

$$(\mathbf{E}|X+Y|^p)^{1/p} \le (\mathbf{E}|X|^p)^{1/p} + (\mathbf{E}|Y|^p)^{1/p}.$$

Оставим это неравенство без доказательства.

10.3 Интеграл в форме Стилтьеса

Определение 10.1. Пусть $g(x)$ — непрерывная функция, обращающаяся в ноль вне $[a,b]$, $F(x)$ — функция распределения. Пусть λ — разбиение $[a,b]$ точками x_k, $0 \le k \le n$ и $\widetilde{x}_k$ — отмеченные точки. Если сушествует предел при диаметре разбиения, стремящееся к нулю, величин

$$\sum_{k=1}^{n} g(\widetilde{x}_k)(F(x_k) - F(x_{k-1})),$$

то этот предел называется интегралом Римана-Стилтьеса функции $g(x)$ и обозначается $\int_a^b g(x)dF(x)$.

Предельным переходом можно доопределить несобственный интегралы Римана-Стильтеса.

Оказывается, что для непрерывных на прямой функций g выполнено соотношение.

$$\mathbf{E}g(X) = \int_{-\infty}^{\infty} g(x)dF(x).$$

По существу эта связь обеспечивается предельным переходом к X от простых величин X_n, описанных в начале построения математического ожидания. Более подробно останавливаться на этом мы не можем, но будем активно использовать полученную формулу.

Наиболее удобная ситуация возникает, если X — абсолютно непрерывна. Тогда $\int g(x)dF(x) = \int g(x)f_X(x)dx$. Таким образом, для абсолютно-непрерывных величин подсчет математического ожидания сводится к подсчету обычного риманова интеграла.

Пример 10.3. Величина, распределенная по закону Коши (то есть с плотностью $f_X(x) = \frac{1}{\pi(1+x^2)}$), математического ожидания не имеет. Действительно, ее матожидания от положительной и отрицательной частей представляют собой один и тот же интеграл $\int\limits_0^\infty \frac{xdx}{\pi(1+x^2)}$, который расходится в плюс бесконечность.

Пример 10.4. Пусть X — равномерно распределена на отрезке $[a, b]$. Тогда ее среднее равно

$$\int_a^b x\frac{1}{b-a}dx = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}.$$

Вопрос 10.1. Математическое ожидание X^2 для равномерной R[0,1] величины равняется?

В более общем случае, если g разрывна, можно представлять математическое ожидание $g(X)$ с помощью интеграла Лебега-Стильтеса $\int_{-\infty}^{\infty} g(x)dF(x)$. В общем случае его конструкция аналогична конструкции интеграла Лебега, мы будем рассматривать это как другую форму записи математического ожидания.

При этом для интегралов Лебега-Стилтьеса справедливы некоторые свойства:

- $$\int\limits_R g(x)d(aF(x)) + \int\limits_R g(x)d(bG(x)) = \int\limits_R g(x)d(aF(x) + bG(x)),$$
- $$\int\limits_R (ag(x) + bh(x))dF(x) = a\int\limits_R g(x)dF(x) + b\int\limits_R h(x)dF(x).$$
- Для функций $g(x)$, являющихся монотонными, справедлива формула интегрирования по частям
$$\int\limits_a^b g(x)dF(x) = g(x)F(x)|_a^b - \int\limits_a^b F(x)dg(x).$$

Пример 10.5. Заметим, что

$$\mathbf{E}X = \int\limits_{-\infty}^{0} xdF(x) - \int\limits_0^{+\infty} xd(1-F(x)) = xF(x)|_{-\infty}^0 - x(1-F(x))|_0^{+\infty} + \int\limits_0^{+\infty} (1 - F(-x) - F(x))dx.$$

Первые два слагаемых обращаются в 0, если математическое ожидание конечно, откуда получается полезная формула

$$\mathbf{E}X = \int_0^\infty (1-F(x))dx + \int_{-\infty}^0 F(x)dx.$$

10.4 Дисперсия и ковариация

Как и прежде мы можем ввести дисперсию

Определение 10.2. Дисперсией называется величина $\mathbf{E}(X - \mathbf{E}X)^2 = \mathbf{E}X^2 - (\mathbf{E}X)^2$.

Для конечности дисперсии требуется условие $\mathbf{E}X^2 < \infty$.

В абсолютно-непрерывном случае справедлива формула

$$\mathbf{E}X^2 = \int_{\mathbb{R}} x^2 dF(x) = \int_{\mathbb{R}} x^2 f_X(x) dx,$$

откуда мы получаем способ искать дисперсию и в этом случае.

Как и прежде введем ковариацию:

Определение 10.3. Ковариацией величин X и Y называют

$$\mathrm{cov}(X, Y) = \mathbf{E}(X - \mathbf{E}X)(Y - \mathbf{E}Y) = \mathbf{E}XY - \mathbf{E}X\mathbf{E}Y.$$

Как и в дискретном случае остаются верными приведенные ниже свойства. Доказательства их остаются теми же, что и в дискретном случае.

1. $\mathbf{D}X = \mathrm{cov}(X, X)$.
2. $\mathrm{cov}(X, Y) = \mathrm{cov}(Y, X)$.
3. $\mathrm{cov}(X, Y + c) = \mathrm{cov}(X, Y)$, $\mathbf{D}(X + c) = \mathbf{D}X$.
4. Если X, Y независимы, то $\mathrm{cov}(X, Y) = 0$.
5. $\mathrm{cov}(aX + bY, Z) = a\,\mathrm{cov}(X, Z) + b\,\mathrm{cov}(Y, Z)$.
6. $\mathbf{D}aX = a^2\mathbf{D}X$.
7. $\mathbf{D}(X + Y) = \mathbf{D}X + 2\,\mathrm{cov}(X, Y) + \mathbf{D}Y$.
8. $\mathbf{D}(X_1 + ... + X_n) = \sum_{i=1}^n \mathbf{D}X_i + 2\sum_{i<j} \mathrm{cov}(X_i, X_j)$.
9. $|cov(X, Y)| \leq \sqrt{\mathbf{D}X\mathbf{D}Y}$ в силу неравенству Коши-Буняковского.

11 Случайные векторы

11.1 Определение

Напомню, что в пространстве $\mathbb{R}^n$ борелевская сигма-алгебра $\mathcal{B}(\mathbb{R}^n)$ — это минимальная сигма-алгебра, содержащая все открытые множества $\mathbb{R}^n$.

Пусть $(\Omega, \mathcal{F}, \mathbf{P})$ — вероятностное пространство.

Случайным вектором мы назвали отображение $\vec{X}$ из Ω в $\mathbb{R}^n$, обладающее свойством измеримости:

$$\{\omega : \vec{X}(\omega) \in B\} \in \mathcal{F}$$

при любом $B \in \mathcal{B}(\mathbb{R}^n)$. Однако, как мы выяснили на позапрошлой лекции, это равносильно тому, что все компоненты вектора $X_1, X_2, \ldots, X_n$ есть случайные величины.

Чтобы задать распределение вектора

$$\mathbf{P}_{\vec{X}}(B) = \mathbf{P}(\vec{X}(\omega) \in B)$$

при всех $B \in \mathcal{B}(\mathbb{R}^n)$, достаточно задать его для некоторых $B \in \mathcal{A}$, где $\mathcal{A}$ — некоторая алгебра, порождающая сигма-алгебру $\mathcal{B}(\mathbb{R}^n)$. В качестве такой алгебры можно рассматривать множество всех простых множеств, то есть множеств, являющихся конечным объединением параллелепипедов $I_1 \times \cdots \times I_n$, где I_j — полуинтервалы прямой $\mathbb{R}$, или дополнением до таких объединений. Для этого достаточно задать для всех $A = (-\infty, x_1] \times \cdots \times (-\infty, x_n]$ вероятность

$$\mathbf{P}(\vec{X} \in A) = \mathbf{P}(X_1 \leq x_1, \ldots, X_n \leq x_n)$$

Определение 11.1. Указанная функция

$$\mathbf{F}_{\vec{X}}(\vec{x}) = \mathbf{P}(X_1 \leq x_1, \ldots, X_n \leq x_n).$$

называется функцией распределения случайного вектора $\vec{X}$.

Положим

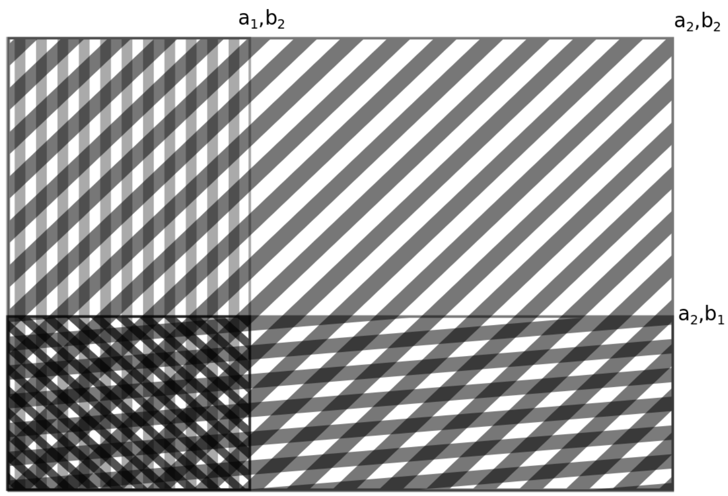
$$\Delta_{i,a_i,b_i}\mathbf{F}(\vec{x}) = \mathbf{F}(x_1, \ldots, x_{i-1}, b_i, x_{i+1}, \ldots, x_n) - \mathbf{F}(x_1, \ldots, x_{i-1}, a_i, x_{i+1}, \ldots, x_n).$$

Тогда

$$\mathbf{P}(X_1 \in (a_1, b_1], \ldots, X_n \in (a_n, b_n]) = \Delta_{1,a_1,b_1} \ldots \Delta_{n,a_n,b_n}\mathbf{F}(x_1, \ldots, x_n).$$

Правая часть довольно громоздко расписывается, но если задуматься, то это сумма значений $\mathbf{F}$ в вершинах

Рис. 16: Для двумерной величины вероятность попадания в прямоугольник будет равна сумме вероятностей попадания в квадранты, выделенные косой штриховкой, из которой вычли вероятности попадания в квадранты с прямой штриховкой



параллелепипеда $(a_1, b_1] \times \cdots (a_n, b_n]$, взятых с разными знаками:

$$\mathbf{F}(b_1, \ldots, b_n) - \mathbf{F}(a_1, b_2, \ldots, b_n) - \mathbf{F}(b_1, a_2, b_3, \ldots, b_n) - \cdots - \mathbf{F}(b_1, \ldots, b_{n-1}, a_n) + \cdots + (-1)^n\mathbf{F}(a_1, ..., a_n).$$

Итак, функция распределения задает вероятности попадания в параллелепипеды, а значит в простые множества, а значит и всю меру $\mathbf{P}_X$.

Функция $\mathbf{F}_{\vec{X}}$ обладает рядом свойств, аналогичных одномерному случаю:

1. при любых $a_i < b_i$, $i \leq n$:
$$\Delta_{1,a_1,b_1} \cdots \Delta_{n,a_n,b_n}\mathbf{F}(x_1, \ldots, x_n) \geq 0; \tag{6}$$

2. функция $\mathbf{F}$ непрерывна справа по совокупности переменных: если вектор $\vec{x}$ стремится к $\vec{x}^0$ по совокупности переменных так, что все $x_i \geq x_i^0 + 0$, то $\mathbf{F}(\vec{x}) \to \mathbf{F}(\vec{x}^0)$;

3. $\mathbf{F}(x_1, \ldots, x_n) \to 1$, $x_1 \to +\infty, \ldots, x_n \to +\infty$ при $n \to \infty$;

4. $\mathbf{F}(\vec{x}) \to 0$, если $\vec{x} \to \vec{x}_0$, где хоть одна из компонент $\vec{x}_0$ равняется $-\infty$.

Аналогично одномерному случаю любая функция, удовлетворяющая указанным свойствам автоматически является функцией распределения некоторого случайного вектора $\vec{X}$. Мы оставим этот факт без доказательства, но идея его та же самая: по функции распределения мы находим вероятности попадания во все простые множества и показываем, что полученная мера будет сигма-аддитивной на простых множеств. Отсюда по теореме Каратеодори она единственным образом продолжается на борелевскую сигма-алгебру.

11.2 Абсолютно-непрерывный случай

Как и прежде наиболее удобные случаи для нас: дискретный, когда

$$\mathbf{P}(\vec{X} \in B) = \sum_{i:\vec{x}_i \in B} \mathbf{P}(\vec{X} = \vec{x}_i),$$

где $\vec{x}_i$ — возможные значения вектора $\vec{X}$, и абсолютно-непрерывный, когда

$$\mathbf{P}(\vec{X} \in B) = \int_B f_X(\vec{x}) d\vec{x}.$$

для любого множества $B \in \mathcal{B}(\mathbb{R}^n)$.

Более аккуратное определение таково.

Определение 11.2. Распределение абсолютно-непрерывно, если найдется такая неотрицательная функция $f_{\vec{X}}(x_1, \ldots, x_n)$, что

$$\mathbf{F}_{\vec{X}}(\vec{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\vec{X}}(u_1, \ldots, u_n) du_1 \cdots du_n$$

при любых $\vec{x}$.

Таким образом,

$$f_{\vec{X}}(\vec{x}) = \frac{\partial}{\partial x_1} \cdots \frac{\partial}{\partial x_1} F_{\vec{X}}(\vec{x}).$$

С другой стороны, справедливо следующее утверждение:

Лемма 11.1. *Любая неотрицательная интегрируемая $f(\vec{x})$, т.ч.*

$$\int_{\mathbb{R}^n} f(\vec{x}) d\vec{x} = 1$$

является плотностью некоторого распределения.

Доказательство. Покажем, что функция

$$F(\vec{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(\vec{t}) dt_1 \ldots dt_n$$

удовлетворяет четырем характеристическим свойствам функции распределения 1)-4).

Свойство 1) выполнено, поскольку

$$\Delta_{1,a_1,b_1} \ldots \Delta_{n,a_n,b_n} \mathbf{F}(\vec{x}) = \int_{[a_1,b_1]} \cdots \int_{[a_n,b_n]} f(\vec{x}) d\vec{x} \geq 0.$$

Свойство 2) выполнено, поскольку F непрерывна как интеграл по верхнему пределу.

Свойства 3) и 4) выполнены из определения несобственного интеграла. □

Рассмотрим меру $\mathbf{P}(\vec{X} \in B)$ на множестве борелевских множеств и меру

$$Q(B) = \int_B f_{\vec{X}}(\vec{x}) d\vec{x},$$

где интеграл в правой части понимается как интеграл Лебега от функции $f_X(\vec{x}) I_B(\vec{x})$ по мере Лебега. На множествах $(-\infty, x_1] \times \cdots \times (-\infty, x_n]$ эти меры совпадают, а значит совпадают и на всех простых множествах. Но в силу теоремы Каратеодори существует единственная сигма-аддитивная мера $\mathbf{P}(\vec{X} \in B)$, продолжающаяся с заданной меры на алгебре простых множеств. Отсюда

$$\mathbf{P}(\vec{X} \in B) = \int_{\mathbb{R}^n} f_X(\vec{x}) I_B(\vec{x}) d\vec{x}.$$

Если не вдаваться в технические детали, то, например, для измеримых по Жордану множеств B полученная формула позволяет вычислять вероятности попадания вектора $\vec{X}$ в различные множества $B \in \mathbb{B}(\mathbb{R}^n)$.

Пример 11.1. Рассмотрим измеримое множество A объема $V(A)$ и вектор $\vec{X}$ с плотностью

$$f_{\vec{X}}(\vec{x}) = \begin{cases} \frac{1}{V(A)}, & x \in A, \\ 0, & x \notin A. \end{cases}$$

Такой вектор называется равномерно распределенным на множестве A. Тогда

$$\mathbf{P}(\vec{X} \in B) = \frac{1}{V(A)} \int_B I_{\vec{x} \in A} d\vec{x} = \frac{V(B \cap A)}{V(A)}$$

при любом $B \in \mathcal{B}(\mathbb{R}^n)$.

Итак, для измеримых множеств B вычисление вероятностей попадания в них абсолютно-непрерывных векторов сводится к подсчету интеграла по этому множеству.

11.3 Плотности подвекторов

Пусть $\vec{X}$ — абсолютно-непрерывный вектор с плотностью $f_{\vec{X}}(\vec{x})$. Тогда:

1. в силу определения плотности
$$\int_{\mathbb{R}^n} f_{\vec{X}}(\vec{x}) d\vec{x} = \mathbf{P}(\vec{X} \in \mathbb{R}^n) = 1;$$
2. при любых непересекающихся $I = \{i_1, \ldots, i_k\}$, $J = \{j_1, \ldots, j_{n-k}\}$, в объединении дающих $\{1, \ldots, n\}$, выполнено
$$\int_{\mathbb{R}^k} f_{\vec{X}}(\vec{x}) dx_{i_1} \ldots dx_{i_k} = f_{X_{j_1}, \ldots, X_{j_{n-k}}}(x_{j_1}, \ldots, x_{j_{n-k}}),$$
поскольку при $B = \{\vec{x} : \vec{x}_J \in A\}$ выполнено соотношение
$$\mathbf{P}(\vec{X}_J \in A) = \mathbf{P}(X \in B) = \int_B f_{\vec{X}}(\vec{x}) d\vec{x} = \int_A \int_{\mathbb{R}^k} f_{\vec{X}}(\vec{x}) dx_{i_1} \ldots dx_{i_k},$$
где $\vec{X}_J = (X_{j_1}, \ldots, X_{j_{n-k}})$.

Плотность подвектора вектора $\vec{X}$ называется маргинальной плотностью.

11.4 Независимость в терминах функций распределения и плотностей

Лемма 11.2. *Пусть $\vec{X}$, $\vec{Y}$ — случайные векторы. Тогда они независимы тогда и только тогда, когда*

$$F_{\vec{X},\vec{Y}}(\vec{x}, \vec{y}) = F_{\vec{X}}(\vec{x}) F_{\vec{Y}}(\vec{y}).$$

Доказательство. Из независимости очевидно вытекает искомое утверждение, поскольку

$$F_{\vec{X},\vec{Y}}(\vec{x}, \vec{y}) = \mathbf{P}(\vec{X} \in (-\infty, x_1] \times \cdots \times (-\infty, x_n], \vec{Y} \in (-\infty, y_1] \times \cdots \times (-\infty, y_m]) =$$
$$\mathbf{P}(\vec{X} \in (-\infty, x_1] \times \cdots \times (-\infty, x_n]) \mathbf{P}(\vec{Y} \in (-\infty, y_1] \times \cdots \times (-\infty, y_m]).$$

С другой стороны, рассмотрим меру, заданную вектором $\vec{U}, \vec{V}$, где $\vec{U}$ имеет то же распределение, что и $\vec{X}$, $\vec{V}$ — то же, что и $\vec{Y}$ и они независимы. Значит

$$F_{\vec{X},\vec{Y}}(\vec{x}, \vec{y}) = F_{\vec{X}}(\vec{x}) F_{\vec{Y}}(\vec{y}) F_{\vec{U},\vec{V}}(\vec{x}, \vec{y}).$$

Опять же отсюда верно равенство мер

$$\mathbf{P}((\vec{X}, \vec{Y}) \in A) = \mathbf{P}((\vec{U}, \vec{V}) \in A)$$

при всех $A \in \mathbb{R}^{n+m}$. В частности, если $A = A_1 \times A_2$, то

$$\mathbf{P}((\vec{X}, \vec{Y}) \in A_1 \times A_2) = \mathbf{P}((\vec{U}, \vec{V}) \in A_1 \times A_2) = \mathbf{P}(\vec{U} \in A_1)\mathbf{P}(\vec{V} \in A_2) = \mathbf{P}(\vec{X} \in A_1)\mathbf{P}(\vec{Y}) \in A_2).$$

Что и требовалось доказать. □

Лемма 11.3. *Пусть $\vec{X}$, $\vec{Y}$ — случайные векторы, имеющие совместную плотность $f_{\vec{X},\vec{Y}}$. Тогда они независимы тогда и только тогда, когда*

$$f_{\vec{X},\vec{Y}}(\vec{x}, \vec{y}) = f_{\vec{X}}(\vec{x}) f_{\vec{Y}}(\vec{y}).$$

Доказательство. Из независимости очевидно вытекает искомое утверждение, поскольку

$$F_{\vec{X},\vec{Y}}(\vec{x}, \vec{y}) = \mathbf{P}(\vec{X} \in (-\infty, x_1] \times \cdots \times (-\infty, x_n])\mathbf{P}(\vec{Y} \in (-\infty, y_1] \times \cdots \times (-\infty, y_m]) =$$
$$\int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} f_{\vec{X}}(\vec{t}) d\vec{t} \int_{-\infty}^{y_1} \ldots \int_{-\infty}^{y_m} f_{\vec{Y}}(\vec{s}) d\vec{s} \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} \int_{-\infty}^{y_1} \ldots \int_{-\infty}^{y_m} f_{\vec{X}}(\vec{t}) f_{\vec{Y}}(\vec{s}) d\vec{t} d\vec{s}.$$

С другой стороны, если плотность распадается в произведение, то

$$F_{\vec{X},\vec{Y}}(\vec{x}, \vec{y}) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} \int_{-\infty}^{y_1} \ldots \int_{-\infty}^{y_m} f_{\vec{X}}(\vec{u}) f_{\vec{Y}}(\vec{v}) d\vec{u} d\vec{v} = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} f_{\vec{X}}(\vec{u}) d\vec{u} \int_{-\infty}^{y_1} \ldots \int_{-\infty}^{y_m} f_{\vec{Y}}(\vec{v}) d\vec{v} = F_{\vec{X}}(\vec{x}) F_{\vec{Y}}(\vec{y}).$$

Что и требовалось доказать. □

Можно слегка усилить предыдущее утверждение:

Лемма 11.4. *Пусть $\vec{X}, \vec{Y}$ — случайные векторы, имеющие совместную плотность $f_{\vec{X},\vec{Y}}(\vec{x}, \vec{y})$, причем $f(\vec{x}, \vec{y}) = g(\vec{x})h(\vec{y})$ для некоторых g, h. Тогда $\vec{X}$, $\vec{Y}$ независимы.*

Доказательство. □

11.5 Преобразование плотности при гладких заменах координат

Пусть $\vec{X}$ — абсолютно-непрерывный вектор с плотностью $f_{\vec{X}}(\vec{x})$, а отображение $g : \mathbb{R}^n \to \mathbb{R}^n$ — обратимое непрерывно-дифференцируемое отображение. Тогда справедлива следующая теорема:

Теорема 11.1.

$$f_{g(\vec{X})}(\vec{y}) = f_{\vec{X}}(g^{-1}(\vec{y}))|J_{g^{-1}}(\vec{y})|,$$

где J — якобиан обратной замены g^{-1}.

Доказательство. Заметим, что для любой области $A \subset \mathbb{R}^n$

$$\int_A f_{g(\vec{X})}(\vec{y}) d\vec{y} = \mathbf{P}(g(\vec{X}) \in A) = \mathbf{P}(\vec{X} \in g^{-1}(A)) = \int_{g^{-1}(A)} f_{\vec{X}}(\vec{x}) d\vec{x} = \int_A f_{\vec{X}}(g^{-1}(\vec{y})|J_{g^{-1}}(\vec{y})| d\vec{y},$$

где в последнем равенстве мы произвели замену под знаком интеграла. Отсюда мы имеем два представления для одной и той же вероятности в виде интеграла. Но тогда

$$f_{g(\vec{X})}(\vec{y}) = f_{\vec{X}}(g^{-1}(\vec{y}))|J_{g^{-1}}(\vec{y})|$$

при п.в. $\vec{y} \in \mathbb{R}^n$. Под п.в. (почти всеми) $\vec{y}$ мы подразумеваем, что множество C всех $\vec{y}$ таких, что наше равенство выполнено, таково, что $\lambda(\overline{C}) = 0$, где λ — мера Лебега. □

11.6 Математические ожидания функций от векторов

Лемма 11.5. *Пусть $g : \mathbb{R}^n \to \mathbb{R}$ — измеримая функция, $\vec{X}$ — случайный вектор с плотностью $f_{\vec{X}}$, причем $\mathbf{E}g(\vec{X})$ существует и конечно. Тогда*

$$\mathbf{E}g(\vec{X}) = \int_{\mathbb{R}^n} g(\vec{x}) f_{\vec{X}}(\vec{x}) d\vec{x}.$$

Доказательство. Заметим, что если $g(x)$ — простая функция (то есть $g(\vec{x}) = \sum_{k=1}^{n} a_k I_{A_k}(\vec{x})$), то формула верна, поскольку

$$\mathbf{E}g(\vec{X}) = \sum_{k=1}^{n} a_k \mathbf{P}(\vec{X} \in A_k) = \sum_{k=1}^{n} a_k \int_{A_k} f_{\vec{X}}(\vec{x}) d\vec{x} = \int_{\mathbb{R}^n} g(\vec{x}) f_{\vec{X}}(\vec{x}) d\vec{x}.$$

Если $g(\vec{x})$ — неотрицательная функция, а $g_n(\vec{x})$ — последовательность сходящихся к ним простых функций, то

$$\mathbf{E}g(\vec{X}) = \lim_{n\to\infty} \mathbf{E}g_n(\vec{X}) = \lim_{n\to\infty} \int_{\mathbb{R}^n} g_n(\vec{x}) f_{\vec{X}}(\vec{x}) d\vec{x} = \int_{\mathbb{R}^n} g(\vec{x}) f_{\vec{X}}(\vec{x}) d\vec{x}.$$

Наконец для произвольной $g(\vec{x})$ верно тождество

$$\mathbf{E}g(\vec{X}) = \mathbf{E}g^+(\vec{X}) - \mathbf{E}g^-(\vec{X}) = \int_{\mathbb{R}} g^+(\vec{x}) f_X(\vec{x}) d\vec{x} - \int_{\mathbb{R}} g^-(\vec{x}) f_X(\vec{x}) d\vec{x},$$

что и требовалось доказать. □

11.7 Нормальные векторы

Определение 11.3. Вектор $\vec{Y} = (Y_1, \ldots, Y_n)$ называют многомерным нормальным вектором, если найдутся такие н.о.р. $\mathcal{N}(0,1)$ величины X_i, $i \le n$, и такие матрица A, вектор $\vec{b}$, что

$$\vec{Y} = A\vec{X} + \vec{b}.$$

Принято задавать нормальный вектор вектором средних $\mathbf{E}Y_i = b_i$ и матрицей ковариаций $\Sigma = (\mathrm{cov}(Y_i, Y_j)) = AA^t$.

Лемма 11.6. *Если A имеет ранг n (то есть Σ невырождена), то $\vec{Y}$ имеет плотность*

$$f_{\vec{Y}}(\vec{y}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det\Sigma}} \exp\left(-\frac{1}{2}(\vec{y}-\vec{b})\Sigma^{-1}(\vec{y}-\vec{b})^t\right)$$

Доказательство. Заметим, что

$$f_{\vec{X}}(\vec{x}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n} x_i^2},$$

поскольку это произведение н.о.р. $\mathcal{N}(0,1)$. Остается применить теорему 11.1:

$$f_{A\vec{X}+\vec{b}}(\vec{x}) = f_{\vec{X}}\left(A^{-1}(\vec{x}-\vec{b})\right)\det(A^{-1}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det\Sigma}} \exp\left(-\frac{1}{2}(\vec{y}-\vec{b})\Sigma^{-1}(\vec{y}-\vec{b})^t\right)$$

□

Зачастую в качестве определения нормального вектора используют свойства 2.1.

Лемма 11.7. *Две компоненты нормального вектора $\vec{X}, \vec{Y}$ независимы тогда и только тогда, когда $\mathrm{cov}(X_i, Y_j) = 0$ при всех допустимых i, j.*

Доказательство. В одну сторону это очевидно — из независимости следует то, что ковариации нулевые.

Если ковариации нулевые, то плотность распадается в произведение плотностей:

$$\frac{1}{(2\pi)^{n/2}\sqrt{\det\Sigma}} \exp\left(-\frac{1}{2}(\vec{y}-\vec{b})\Sigma^{-1}(\vec{y}-\vec{b})^t\right),$$

Поскольку Σ^{-1} — блочная матрица в силу условия, то отсюда получаем, что плотность распадается в произведение двух нормальных плотностей. □

12 Условное математическое ожидание

12.1 Дискретный случай

Пусть задано вероятностное пространство $(\Omega, \mathcal{F}, \mathbf{P})$, $\mathcal{D} = \{D_1, \dots, D_n\}$ — некоторое разбиение Ω, где $n \le \infty$ (под $n = \infty$ будем понимать, что разбиение счетное). Будем считать, что $\mathbf{P}(D_i) > 0$.

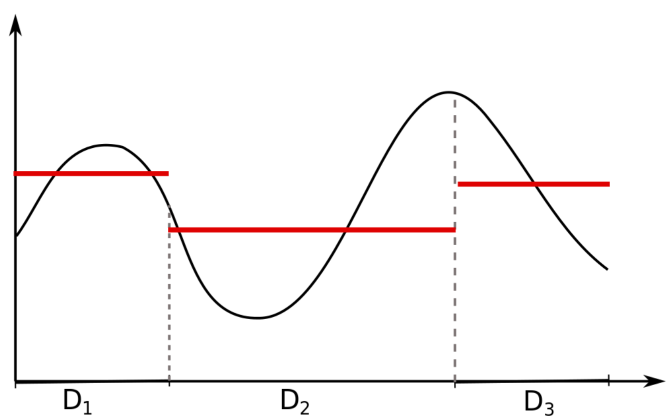
Будем использовать обозначение $\mathbf{E}(X|D_i)$ для математического ожидания величины X по вероятностной мере $\mathbf{P}(\cdot|D_i)$, то есть

$$\mathbf{E}(X|D_i) = \frac{\mathbf{E} X I_{D_i}}{\mathbf{P}(D_i)}.$$

Определение 12.1. Вероятностью события B при условии разбиения $\mathcal{D}$ называется случайная величина (!) $Y(\omega)$, равная $\mathbf{P}(B|D_i)$ при $\omega \in D_i$.

Определение 12.2. Условным математическим ожиданием (УМО) случайной величины X при условии разбиения $\mathcal{D}$ называется случайная величина $Y(\omega)$, равная $\mathbf{E}(X|D_i)$ при $\omega \in D_i$.

Рис. 17: Условное математическое ожидание случайной величины, заданной черным графиком, относительно указанного разбиения оси абсцисс, будет представлять красную кусочно-постоянную случайную величину



Пример 12.1. Пусть X_i — независимые бернуллиевские с параметром $1/2$, а $\mathcal{D}$ — разбиение, заданное $D_k = \{\omega : X_1 + X_2 + X_3 = k\}$, $k \le 3$. Тогда

$$\mathbf{P}(X_1 = 1|\mathcal{D})$$

будет случайной величиной, равной

$$\mathbf{P}(X_1 = 1|D_k) = \frac{\mathbf{P}(X_1 = 1, X_1 + X_2 + X_3 = k)}{\mathbf{P}(X_1 + X_2 + X_3 = k)} = \frac{\mathbf{P}(X_1 = 1)\mathbf{P}(X_2 + X_3 = k - 1)}{\mathbf{P}(X_1 + X_2 + X_3 = k)}$$

при $\omega \in D_k$. При $k = 0$ данная вероятность равна 0, при $k \ge 1$ равна

$$\frac{C_2^{k-1} \cdot p \cdot p^{k-1} \cdot (1-p)^{3-k}}{C_3^k \cdot p^k \cdot (1-p)^{3-k}} = \frac{k!\ 2!}{(k-1)!\ 3!} = \frac{k}{3},$$

Нетрудно заметить, что данная величиина в точности равна $(X_1 + X_2 + X_3)/3$.

Отсюда же

$$\mathbf{E}(X_1|\mathcal{D}) = \frac{X_1 + X_2 + X_3}{3},$$

поскольку при любом k $\mathbf{E}(X_1|D_k) = \mathbf{P}(X_1 = 1|D_k) = k/3$.

При этом для любого множества A, лежащего в $\sigma(\mathcal{D})$, выполнено соотношение

$$\mathbf{E}YI_A = \sum_{i:D_i\subset A}\mathbf{E}YI_{D_i} = \sum_{i:D_i\subset A}\mathbf{E}XI_{D_i} = \mathbf{E}XI_A,$$

где $Y = \mathbf{E}(X|\mathcal{D})$.

Это свойство является также и характеристическим: УМО Y однозначно определяется соотношением

1. $Y(\omega_1) = Y(\omega_2)$ при всех $\omega_1, \omega_2 \in D_i$ при каком-то i;

2. $\mathbf{E}YI_A = \mathbf{E}XI_A$ при всех $A \in \sigma(\mathcal{D})$.

Действительно, пусть $Y(\omega) = y_i$ при $\omega \in D_i$ (в силу свойства 1) она постоянна на D_i). Тогда при $A = D_i$ получаем

$$\mathbf{E}YI_{D_i} = y_i\mathbf{P}(D_i) = \mathbf{E}XI_{D_i},$$

откуда $y_i = \mathbf{E}(X|D_i)$.

12.2 Общее определение

На основе дискретного случая построим общее определение условного математического ожидания.

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

Нам понадобится следующее определение:

Определение 12.3. Величина Y измерима около сигма-алгебры $\mathcal{A} \subseteq \mathcal{F}$, если $\{\omega : Y(\omega) \in B\} \in \mathcal{A}$ при любом $B \in \mathcal{B}(\mathbb{R})$.

В случае, если сигма-алгебра $\mathcal{A}$ порождается разбиением $\mathcal{D}$, это требование означает, что

$$\{Y(\omega) \in B\} = D_{i_1} + \dots + D_{i_k}$$

для некоторых $i_1, \dots, i_k$, а, значит, $Y(\omega) = const$ на любом D_i.

Пусть X — случайная величина, $\mathbf{E}|X| < \infty$. Кроме того, пусть задана сигма-алгебра $\mathcal{A} \subseteq \mathcal{F}$.

Определение 12.4. Условным математическим ожиданием X относительно $\mathcal{A}$ называют случайную величину Y (обозначается она $\mathbf{E}(X|\mathcal{A})$, такую, что:

1. Y — $\mathcal{A}$-измерима;

2. для любого $B \in \mathcal{A}$ $\mathbf{E}XI_B = \mathbf{E}YI_B$.

Лемма 12.1. *Условное математическое ожидание случайной величины существует и единственно п.н. (то есть с точностью до множества вероятности 0: любые две величины Y_1, Y_2, удовлетворяющие определению выше, выполнено свойство $\mathbf{P}(\omega : Y_1(\omega) = Y_2(\omega)) = 1$).*

Доказательство. Мы воспользуемся так называемой теоремой Радона-Никодима, которую оставим без доказательства:

Теорема 12.1. *Пусть ν — конечная мера, μ — σ-конечная мера на некотором S с сигма-алгеброй $\mathcal{S}$ (σ-конечность означает, что S можно разбить на счетное число таких множеств S_1, S_2,..., что $\mu(S_i)$ — конечны). Предположим, что $\nu(A) = 0$ при любом таком A, что $\mu(A) = 0$. Тогда найдется такая измеримая относительно сигма-алгебры $\mathcal{S}$ функция $f(x)$, что*

$$\nu(A) = \int_A f(x)\mu(dx)$$

при любом A. При этом функция f, называемая производной Радона-Никодима меры ν по мере μ, неотрицательна и единственна с точностью до множества μ-меры ноль: если $f_1(x)$, $f_2(x)$ — две таких функции, то $\mu(x : f_1(x) \neq f_2(x)) = 0$.

1) Рассмотрим неотрицательную случайную величину X и введем на $\mathcal{A}$ меру

$$\mathbf{Q}(A) = \mathbf{E}XI_A.$$

Заметим, что это мера. Действительно, она аддитивна: $\mathbf{Q}(A+B) = \mathbf{E}YI_{A+B} = \mathbf{Q}(A) + \mathbf{Q}(B)$. Также она непрерывна из теоремы Лебега о монотонной сходимости, поскольку $XI_{A_1+\dots+A_n}$ монотонно сходится к $XI_{A_1+\dots+A_n+\dots}$, а, значит, их математические ожидания сходятся.

Рассмотрим меру $\widetilde{\mathbf{P}} : \mathcal{A} \to [0,1]$, где $\widetilde{\mathbf{P}}(A) = \mathbf{P}(A)$ (это ограничение меры $\mathbf{P}$ на $\mathcal{A}$, то есть та же мера, но на более маленькой сигма-алгебре).

Если $\widetilde{\mathbf{P}}(A) = 0$, то $\mathbf{Q}(A) = 0$. Действительно, пусть Y_n простые величины, $Y_n \uparrow XI_A$. Тогда $\widetilde{\mathbf{P}}(Y_n > 0) \leq \widetilde{\mathbf{P}}(XI_A > 0) = \widetilde{\mathbf{P}}(A) = 0$, откуда $\mathbf{E}Y_n = 0$, а значит $\mathbf{E}XI_A = 0$.

По теореме Радона-Никодима найдется такая функция $Y(\omega)$, что

$$\mathbf{E}XI_A = \mathbf{Q}(A) = \int_A Y(\omega)\widetilde{\mathbf{P}}(d\omega) = \mathbf{E}YI_A.$$

При этом Y — измеримая функция из пространства $(\Omega, \mathcal{A})$ в $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, а значит

$$\{\omega : Y(\omega) \in B\} \in \mathcal{A}$$

для любого $B \in \mathcal{B}(\mathbb{R})$.

При этом Y единственна п.н. в силу все той же леммы.

2) Пусть X принимает значения обоих знаков. Тогда существование математического ожидания вытекает из того, что X^+ имеет УМО (обозначим его Y^+) и X^- имеет УМО (обозначим Y^-), значит, $X = X^+ - X^-$ также имеет УМО, равное разности $Y^+ - Y^-$.

С другой стороны, покажем, что УМО единственно. Пусть Y_1, Y_2 такие $\mathcal{A}$-измеримые случайные величины, что

$$\mathbf{E}XI_A = \mathbf{E}Y_1I_A = EY_2I_A,$$

Пусть $Y = \mathbf{E}(|X||\mathcal{A})$. Тогда

$$\mathbf{E}(X + |X|)I_A = \mathbf{E}(Y_1 + Y)I_A = \mathbf{E}(Y_2 + Y)I_A.$$

Но тогда $X + |X| = 2X^+$ — неотрицательная величина, а $Y_1 + Y$ и $Y_2 + Y$ — ее УМО около $\mathcal{A}$. Из предыдущего пункта они равны п.н., значит, $Y_1 = Y_2$ п.н. □

12.3 Условное математическое ожидание при условии случайной величины

Зачастую удобнее рассматривать УМО в форме $\mathbf{E}(X|Y) := \mathbf{E}(X|\sigma(Y))$, где $\sigma(Y)$ — сигма-алгебра, порожденная случайной величиной Y, то есть $\{\{\omega : Y \in B\},\ B \in \mathcal{B}(\mathbb{R})\}$.

Рассмотрим к чему это приведет в дискретном случае. Если X, Y — дискретные величины, то $\sigma(Y)$ порождается разбиением $\mathcal{D} = (\{Y = y_1\}, \dots, \{Y = y_k\})$, где $y_1, \dots, y_k$ — возможные значения случайной величины Y (в случае счетного числа значений мы рассматриваем счетное число событий, порождающих $\mathcal{D}$). Следовательно,

$$\mathbf{E}(X|Y) = \begin{cases} \mathbf{E}(X|Y = y_1), & \omega \in \{Y = y_1\}, \\ \dots & \dots \\ \mathbf{E}(X|Y = y_k), & \omega \in \{Y = y_k\}, \end{cases}$$

где

$$\mathbf{E}(X|Y = y_k) = \sum_l x_l \mathbf{P}(X = x_l|Y = y_k).$$

Иначе говоря,

$$\mathbf{E}(X|Y) = g(Y),$$

где $g(y) = \mathbf{E}(X|Y = y)$.

Рассмотрим к чему это приведет в абсолютно-непрерывном случае. Если величины X, Y величины с совместной плотностью $f_{X,Y}$, то положим

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{\int_{\mathbb{R}} f_{X,Y}(u,y)du} = \frac{f_{X,Y}(x,y)}{f_Y(y)},$$

такая функция называется условной плотностью. Положим

$$g(y) = \mathbf{E}(X|Y=y) = \int\limits_{\mathbb{R}} x f_{X|Y}(x|y)dx.$$

Лемма 12.2. *Величина $g(Y)$ будет условным математическим ожидание X при условии Y.*

Доказательство. Проверим, что $g(Y)$ удовлетворяет требуемым условиям. Очевидно, $g(Y)$ является измеримой около $\sigma(Y)$: $\{g(Y) \in B\} = \{Y \in g^{-1}(B)\}$. Остается убедиться во второй части определения: для любого $A \in \sigma(Y)$

$$\mathbf{E}XI_A = \mathbf{E}g(Y)I_A.$$

Но $A = \{\omega : Y(\omega) \in B\}$ в силу определения $\sigma(Y)$, откуда нам нужно доказать, что

$$\int_{\mathbb{R}} \int_B f_{X,Y}(x,y)dxdy = \int_B g(y)f_Y(y)dy.$$

Это утверждение прямо следует из определения $g(y)$. □

Отметим, что в общем случае справедливо следующее утверждение.

Лемма 12.3. *Величина X измерима около сигма-алгебры $\sigma(Y)$ тогда и только тогда, когда $X = g(Y)$ для некоторой измеримой функции g.*

Доказательство. Лемма на лекции оставалась без доказательства и не входит в экзаменационную программу. Приведем доказательство, с которым можно ознакомиться для общего развития.

В одну сторону утверждение очевидно, если $X = g(Y)$, то

$$\{\omega : X(\omega) \in B\} = \{\omega : Y(\omega) \in g^{-1}(B)\}.$$

Если величина X принимает конечное множество значений $x_1, \ldots, x_k$, то множества $\{\omega : X(\omega) = x_k\}$ имеют вид $\{\omega : Y(\omega) \in B_k\}$ для некоторых B_k в силу определения сигма-алгебры, порожденной Y. Тогда $X = g(Y)$, где

$$g(y) = \sum_{i=1}^{k} x_i I_{y \in B_i}.$$

Если X произвольная неотрицательная величина, то рассмотрим X_n — сходящиеся к ней простые величины (например, определенные при введении математического ожидания). Величины X_n измеримы около $\sigma(Y)$, откуда $X_n = g_n(Y)$ в силу предыдущего, где g_n — некоторые измеримые функции. При этом $g_n(Y(\omega)) \uparrow X(\omega)$ при $n \to \infty$. Значит, X есть предел $g_n(Y(\omega))$, то есть функция $Y(\omega)$. Для произвольной величины X разобьем ее на положительную и отрицательную части, каждая из которых опять же измерима около $\sigma(Y)$, то есть являются функциями $Y(\omega)$. □

12.4 Свойства УМО

Изучим ряд свойств у.м.о.:

1. Если X является $\mathcal{A}$-измеримой, то $\mathbf{E}(X|\mathcal{A}) = X$.

   Доказательство. Очевидно, что X удовлетворяет обоим условиям, фигурирующим в определении УМО. □

2. $\mathbf{E}(aX + bY|\mathcal{A}) = a\mathbf{E}(X|\mathcal{A}) + b\mathbf{E}(Y|\mathcal{A})$.

Доказательство. Доказательство достаточно очевидно. Рассмотрим величины $Z_1 = \mathbf{E}(X|\mathcal{A})$, $Z_2 = \mathbf{E}(Y|\mathcal{A})$. Тогда

$$\mathbf{E}Z_1 I_A = \mathbf{E}X I_A, \quad \mathbf{E}Z_2 I_A = \mathbf{E}Y I_A,$$

а значит

$$\mathbf{E}(aZ_1 + bZ_2) I_A = \mathbf{E}(aX + bY) I_A$$

при любом $A \in \mathcal{A}$, что и требовалось доказать. □

3. Если $X \le Y$, то $\mathbf{E}(X|\mathcal{A}) \le \mathbf{E}(Y|\mathcal{A})$.

   Доказательство. В силу прошлой части достаточно доказать, что если $Y - X \ge 0$, то $\mathbf{E}(Y - X|\mathcal{A}) \ge 0$. Это прямо вытекает из конструкции УМО, поскольку производная Радона-Никодима неотрицательна. □

4. Если $\mathcal{A}_1 \subseteq \mathcal{A}_2$, то п.н.

$$\mathbf{E}(\mathbf{E}(X|\mathcal{A}_1)|\mathcal{A}_2) = \mathbf{E}(\mathbf{E}(X|\mathcal{A}_2)|\mathcal{A}_1) = \mathbf{E}(X|\mathcal{A}_1).$$

   В частности, $\mathbf{E}(\mathbf{E}(X|\mathcal{A}_1)) = \mathbf{E}X$.

   Доказательство. Одно из утверждений очевидно: величина $\mathbf{E}(X|\mathcal{A}_1)$ измерима относительно $\mathcal{A}_2$, а значит, в силу свойства 1) имеет требуемое УМО.

   В другую сторону доказательство также не слишком сложно: если $\mathbf{E}(X|\mathcal{A}_2) = Y$, а $\mathbf{E}(Y|\mathcal{A}_1) = Z$, то

$$\mathbf{E}X I_A = \mathbf{E}Y I_A, \quad \mathbf{E}Y I_B = \mathbf{E}Z I_B$$

   при любых $A \in \mathcal{A}_2$, $B \in \mathcal{A}_1$. В частности, если $A = B$ (здесь мы пользуемся тем, что $\mathcal{A}_1 \subseteq \mathcal{A}_2$), то $\mathbf{E}X I_B = \mathbf{E}Z I_B$, где Z — $\mathcal{A}_1$-измерима, значит, Z равно $\mathbf{E}(X|\mathcal{A}_1)$. □

5. Если X, Y независимы, то $\mathbf{E}(X|Y) = \mathbf{E}X$.

   Доказательство. Очевидно, что $\mathbf{E}X$ измерима около любой сигма-алгебры. При этом в силу независимости

$$\mathbf{E}(\mathbf{E}X I_{Y \in B}) = \mathbf{E}X\mathbf{P}(Y \in B) = \mathbf{E}(X I_{Y \in B}),$$

   что и требовалось доказать □

6. Если X измерима около $\mathcal{A}$, то

$$\mathbf{E}(XY|\mathcal{A}) = X\mathbf{E}(Y|\mathcal{A}).$$

   Доказательство. Предположим, что X простая: $X = \sum_{k=1}^m x_k I_{A_k}$, где $A_k \in \mathcal{A}$, пусть $\mathbf{E}(Y|\mathcal{A}) = Z$. Тогда для любого $B \in \mathcal{A}$

$$\mathbf{E}XY I_B = \sum_{k=1}^m x_k \mathbf{E}Y I_{B \cap A_k} = \sum_{k=1}^m x_k \mathbf{E}Z I_{B \cap A_k} = \mathbf{E}ZY I_B,$$

   что и означает, что ZY (измеримая около $\mathcal{A}$ как произведение двух измеримых величин) есть УМО $\mathbf{E}(XY|\mathcal{A})$.

   Если X не простая, то рассмотрим последовательность сходящихся к ней простых $X_n \uparrow X$. Тогда

$$\mathbf{E}X_n Y I_B = \mathbf{E}X_n Z I_B$$

   при любом B по доказанному, значит, по теореме о мажорируемой сходимости

$$\mathbf{E}XY I_B = \mathbf{E}XZ I_B,$$

   что и требовалось доказать. □

7. Пусть $X_n \uparrow X$, $\mathbf{E}|X_n| < \infty$, $\mathbf{E}|X| < \infty$. Тогда

$$\mathbf{E}(X_n|Y) \to \mathbf{E}(X|Y)$$

при п.н. ω.

Доказательство. Если $Z_n = \mathbf{E}(X_n|Y)$. Последовательность $Z_n(\omega)$ монотонна при п.н. ω (в силу свойства 3). Значит $Z_n(\omega)$ сходится к некоторой величине $Z(\omega)$.

Тогда

$$\mathbf{E}Z_n I_{Y\in B} = \mathbf{E}X_n I_{Y\in B}.$$

В силу теоремы о монотонной сходимости

$$\mathbf{E}ZI_{Y\in B} = \lim_{n\to\infty} \mathbf{E}Z_n I_{Y\in B} = \lim_{n\to\infty} \mathbf{E}X_n I_{Y\in B} = \mathbf{E}XI_{Y\in B},$$

что и требовалось доказать. □

12.5 УМО в L^2

Рассмотрим L^2 — множество случайных величин с $\mathbf{E}X^2 < \infty$, причем для удобства мы будем считать равными величины, отличающиеся на множестве вероятности 0. На этом пространстве введем скалярное произведение $\langle X, Y\rangle_{L^2} = \mathbf{E}XY$.

Оно обладает свойствами скалярного произведения: симметрично, билинейно и положительно определено $\langle X, X\rangle \geq 0$, причем равенство нулю возможно только если $X = 0$ п.н. Значит, мы можем рассмотреть L^2 как евклидово пространство.

Рассмотрим в нем подпространство H величин, измеримых около $\mathcal{A}$ (оно действительно замкнуто относительно сложения и умножения на число, то есть является подпространством). Найдем Y в H так, что $\langle X - Y, Z\rangle = 0$ при всех $Z \in H$. Иначе говоря, рассмотрим проекцию X на H.

Лемма 12.4. *Такая проекция существует и единственна (с точностью до множества ω меры ноль), ей является* $\mathbf{E}(X|\mathcal{A})$.

Доказательство. Пусть $Y = \mathbf{E}(X|\mathcal{A})$. Тогда

$$\langle X - Y, Z\rangle = \mathbf{E}(X - Y)Z = \mathbf{E}(\mathbf{E}((X - Y)Z|\mathcal{A})) = \mathbf{E}(Z\mathbf{E}(X - Y|\mathcal{A})) = \mathbf{E}(Z(\mathbf{E}(X|\mathcal{A}) - Y)) = 0,$$

что и требовалось.

При этом для любого такого $Y \in H$, что $\langle X - Y, Z\rangle = 0$ выполнено соотношение

$$\mathbf{E}(X - Y)I_A = 0, \quad \mathbf{E}XI_A = \mathbf{E}YI_A,$$

где $A \in \mathcal{A}$ произвольно. Значит Y является УМО X. □

Иногда это определение более удобно для подсчета условного математического ожидания. Рассмотрим классический пример:

Пример 12.2. Пусть (X, Z) — двумерный нормальный вектор с нулевым вектором средних и матрицей ковариации Σ.
Тогда $\mathbf{E}(X|Z)$ - это такая $g(Z)$, что $X - f(Z)$ ортогонально $h(Z)$ для любой борелевской g.
Заметим, что если $\mathbf{E}(X - f(Z)) = 0$, а $X - f(Z)$ не зависит от Z, то $(X - f(Z), g(Z))_{L^2} = \mathbf{E}(X - f(Z))\mathbf{E}g(Z) = 0$, так что для нахождения $\mathbf{E}(X|Z)$ достаточно найти $f(Z)$, такую что $X - f(Z)$ не зависит от Z и $\mathbf{E}X = \mathbf{E}f(Z)$. Поищем $f(Z)$ в виде $aZ + b$. Тогда $(X - f(Z), Z)$ — нормальный вектор (линейное преобразование нормального вектора) и независимость $X - f(Z)$ и Z равносильна их некореллированности. Отсюда имеем уравнения

$$a\mathbf{E}Z + b = \mathbf{E}X,\ \mathbf{E}XZ - a\mathbf{E}Z^2 - b\mathbf{E}Z = 0,$$

откуда $b = 0$, $a = \operatorname{cov}(X, Z)/DZ$. Следовательно, $\mathbf{E}(X|Z) = Zcov(X, Z)/DZ$.

13 Характеристические функции

13.1 Математическое ожидание комплекснозначной случайной величины

Будут понимать под $\mathbf{E}Z$, где $Z = X + iY$ — комплексная случайная величина, величину $\mathbf{E}X + i\mathbf{E}Y$.

Очевидно, что математическое ожидание в этом случае остается линейным, математическое ожидание произведения независимых величин (то есть таких, у которых пары (действительная часть, мнимая часть) независимы) есть произведение математических ожиданий и так далее. Неочевидными остаются свойства, связанные с модулями, в частности,

$$|\mathbf{E}X| \leq \mathbf{E}|X|.$$

Доказательство. Утверждение равносильно тому, что

$$\sqrt{(\mathbf{E}X)^2 + (\mathbf{E}Y)^2} \leq \mathbf{E}\sqrt{X^2 + Y^2}. \tag{7}$$

Но это неравенство вытекает из неравенства Иенсена

$$\mathbf{E}f(X, Y) \geq f(\mathbf{E}X, \mathbf{E}Y),$$

где f — выпуклая функция. Мы доказывали неравенство Иенсена для функции f одной переменной, однако, и в нашем случае верно то же доказательство: в силу выпуклости

$$f(X, Y) \geq f(\mathbf{E}X, \mathbf{E}Y) + a(X - \mathbf{E}X) + b(Y - \mathbf{E}Y)$$

при некоторых a, b, откуда и вытекает (7). □

Определение 13.1. Говорят, что комплекснозначные величины $Z_1, \dots, Z_n$ независимы, если независимы векторы $(\operatorname{Re} Z_1, \operatorname{Im} Z_1)$, $(\operatorname{Re} Z_2, \operatorname{Im} Z_2)$, ..., $(\operatorname{Re} Z_n, \operatorname{Im} Z_n)$ независимы.

Это определение вытекает из общего определения независимости случайных элементов, но нам будет удобнее использовать его сразу в таком виде.

13.2 Характеристическая функция

Определение 13.2. Характеристической функцией с.в. X называют $\psi_X(t) = \mathbf{E}e^{itX}$, где $t \in R$.

Почему такое матожидание существует? Потому что $e^{itX} = \cos tX + i \sin tX$, и, поскольку, случайные величины $\cos tX$ и $\sin tX$ ограничены, то матожидание всегда конечно. Более того, в силу доказанного $|\mathbf{E}e^{itX}| \leq \mathbf{E}|e^{itX}| = 1$.

Заметим, что для целочисленных величин $\psi(t) = \phi(e^{it})$, поэтому вместо производящей функции мы могли бы использовать характеристическую функцию.

Сформулируем базовые свойства характеристической функции:

1. $\psi(0) = 1$;
2. $\psi(t)$ непрерывная функция, равномерно непрерывная на прямой.

   Доказательство. Заметим, что

   $$|\psi(t + h) - \psi(t)| = \left|\mathbf{E}e^{i(t+h)X} - e^{itX}\right| \leq \mathbf{E}\left|e^{ihX} - 1\right|$$

   Величина под знаком математического ожидания мажорируется 2 и поточечно сходится к нулю при $h \in \mathbb{R}$. В силу теоремы Лебега о мажорируемой сходимости имеем требуемое. □

3. Если величина X^k имеет конечное математическое ожидание, то $\psi(t)$ k раз дифференцируема, причем

   $$\psi^{(k)}(t) = i^k \mathbf{E}X^k e^{itX}, \quad \psi^{(k)}(0) = i^k \mathbf{E}X^k.$$

Доказательство. Доказательство близко к доказательству непрерывности. Докажем индукцией по k. При $k=0$ данная формула верна.

Пусть при $k \le n$ утверждение доказано, докажем его при $k=n+1$. Заметим, что

$$\frac{\psi^{(n)}(t+\Delta t)-\psi^{(n)}(t)}{\Delta t} = \mathbf{E}(iX)^n e^{itX}\left(\frac{e^{i\Delta tX}-1}{\Delta t}\right).$$

Величина под знаком математического ожидания поточечно (при каждом ω) сходится к

$$e^{itX}(iX)^{n+1}$$

при $\Delta t \to 0$. При этом нам пригодится неравенство

$$|e^{it}-1| = \sqrt{(\cos t-1)^2+\sin^2 t} = \sqrt{2(1-\cos t)} = 2|\sin(t/2)| \le |t|,$$

откуда

$$\left|e^{itX}\left(\frac{e^{i\Delta tX}-1}{\Delta t}\right)\right| \le |X|.$$

Отсюда

$$\left|\frac{\psi^{(n)}(t+\Delta t)-\psi^{(n)}(t)}{\Delta t}\right| \le |X|^{n+1},$$

следовательно,

$$\lim_{\Delta t\to 0}\frac{\psi^{(n)}(t+\Delta t)-\psi^{(n)}(t)}{\Delta t} = i^{n+1}\mathbf{E}X^{n+1}e^{itX}$$

по теореме о мажорируемой сходимости. □

4. $\psi_{aX+b}(t) = e^{itb}\psi_X(ta)$.

   Доказательство.

$$\psi_{aX+b}(t) = \mathbf{E}e^{itaX+itb} = e^{itb}\mathbf{E}e^{i(ta)X} = e^{itb}\psi_X(ta).$$

□

5. $\psi_{-X}(t) = \overline{\psi_X(t)}$, где $\overline{z}$ — комплексно сопряженное к z число.

   Доказательство.

$$\psi_{-X}(t) = \mathbf{E}e^{-itX} = \mathbf{E}\cos(tX) - i\mathbf{E}\sin(tX) = \overline{\psi_X(t)}.$$

□

6. ψ_X вещественна при всех t тогда и только тогда, когда $X \overset{d}{=} -X$, т.е. распределение X симметрично.

   Доказательство. Если $X \overset{d}{=} -X$, то $\psi_X(t) = \psi_{-X}(t)$, а значит

$$\psi_X(t) = \frac{1}{2}\left(\psi_X(t)+\psi_X(-t)\right) = \mathbf{E}\cos(tX) \in \mathbb{R}.$$

   Наоборот, если ψ_X вещественна, то

$$\psi_{-X}(t) = \overline{\psi_X(t)} = \psi_X(t).$$

   Значит х.ф. X и $-X$ совпадают. Из теоремы единственности, приведенной ниже, следует что и распределения X и $-X$ совпадают. □

7. Характеристическая функция суммы независимых величин есть произведение х.ф. каждой из них.

Доказательство.
$$\mathbf{E}e^{ti(X_1+\dots+X_n)} = \mathbf{E}e^{itX_1}\cdots e^{itX_n} = \mathbf{E}e^{itX_1}\cdots\mathbf{E}e^{itX_n}.$$

Здесь мы используем то, что если комплекснозначные величины Z_i независимы (то есть, напомним, независимы векторы $(\operatorname{Re} Z_1, \operatorname{Im} Z_1)$, $(\operatorname{Re} Z_2, \operatorname{Im} Z_2)$ и так далее), то $\mathbf{E}Z_1\cdots Z_n = \mathbf{E}Z_1\cdots\mathbf{E}Z_n$.

Действительно, достаточно показать это для двух величин. Но

$$\mathbf{E}Z_1Z_2 = \mathbf{E}\operatorname{Re} Z_1 \operatorname{Re} Z_2 - \mathbf{E}\operatorname{Im} Z_1 \operatorname{Im} Z_2 + i\left(\mathbf{E}\operatorname{Re} Z_1 \operatorname{Im} Z_2 + \mathbf{E}\operatorname{Re} Z_2 \operatorname{Im} Z_1\right).$$

В силу независимости (ReZ_1, ImZ_1) и (ReZ_2, ImZ_2), правая часть преставима в виде

$$\mathbf{E}\operatorname{Re} Z_1\mathbf{E}\operatorname{Re} Z_2 - \mathbf{E}\operatorname{Im} Z_1\mathbf{E}\operatorname{Im} Z_2 + i\left(\mathbf{E}\operatorname{Re} Z_1\mathbf{E}\operatorname{Im} Z_2 + \mathbf{E}\operatorname{Re} Z_2\mathbf{E}\operatorname{Im} Z_1\right) = \left(\mathbf{E}\operatorname{Re} Z_1 + i\mathbf{E}\operatorname{Im} Z_1\right)\left(\mathbf{E}\operatorname{Re} Z_2 + i\mathbf{E}\operatorname{Im} Z_2\right),$$

что и требовалось доказать □

13.3 Примеры характеристических функций

Пример 13.1. Для бернуллиевской случайной величины X

$$\psi_X(t) = \mathbf{E}e^{itX} = pe^{it} + (1-p).$$

Пример 13.2. Для экспоненциальной случайной величины X:

$$\psi_X(t) = \mathbf{E}e^{itX} = \int_0^\infty e^{itx}e^{-\lambda x}\lambda dx = \int_0^\infty \cos(tx)e^{-\lambda x}dx + i\int_0^\infty \sin(tx)dx.$$

Каждый из этих интегралов можно взять по частям:

$$\int_0^\infty \cos(tx)e^{-\lambda x}dx = 1 - (t/\lambda)\int_0^\infty \sin(tx)e^{-\lambda x}dx,$$
$$\int_0^\infty \sin(tx)e^{-\lambda x}dx = (t/\lambda)\int_0^\infty \cos(tx)e^{-\lambda x}dx,$$

откуда

$$\int_0^\infty \cos(tx)e^{-\lambda x}dx = \frac{\lambda^2}{t^2+\lambda^2},\ \int_0^\infty \sin(tx)e^{-\lambda x}dx = \frac{t\lambda}{\lambda^2+t^2},$$

а значит

$$\psi_X(t) = \frac{\lambda(\lambda+it)}{(\lambda-it)(\lambda+it)} = \frac{\lambda}{\lambda-it}.$$

Замечание 13.1. А почему мы просто не взяли интеграл как e^z, сделав замену $z = (t - i\lambda)x$?

Дело в том, что при этом прямая $(-\infty, \infty)$ перешла бы в прямую $tx - i\lambda x$, где $x \in \mathbb{R}$, а интеграл по этой прямой может и не равняться интегралу от той же фукции по вещественной прямой.

Например, $\int_{\mathbb{R}} e^{ix}/(1+x^2)dx$, конечно же сходится, ведь он абсолютно оценивается интегралом от $1/(1+x^2)$, т.е. $\pi/2$. Но если сделать замену $y = ix$, то получим $i\int_{\mathbb{R}} e^y/(1-y^2)dy$, который расходится на бесконечности (да еще и в 1 и в -1). Секрет в том, что на самом деле второй интеграл не по $\mathbb{R}$, а по $i\mathbb{R}$, с которым нужно работать иначе. В связи с этим мы будем осторожно брать интегралы по определению.

13.4 Существование и единственность характеристической функции

А как же, зная характеристическую функцию, найти само распределение и единственно ли распределение с такой характеристической функцией?

На второй вопрос отвечает теорема единственности:

Теорема 13.1. *Пусть F и G — две ф.р., имеющие одинаковые характеристические функции. Тогда $F(x) = G(x)$ при всех x.*

Ответить на первый вопрос для величин с плотностью нам поможет формула обращения:

Теорема 13.2. *1) Пусть х.ф. периодична с максимальным периодом 2π. Тогда величина X целочисленна, причем*

$$\mathbf{P}(X = k) = (2\pi)^{-1} \int_{-\pi}^{\pi} e^{-itk} \psi_X(t) dt.$$

2) Пусть х.ф. абсолютно интегрируема на всей прямой. Тогда X имеет плотность

$$f_X(x) = (2\pi)^{-1} \int_{\mathbb{R}} e^{-itx} \psi_X(t) dt.$$

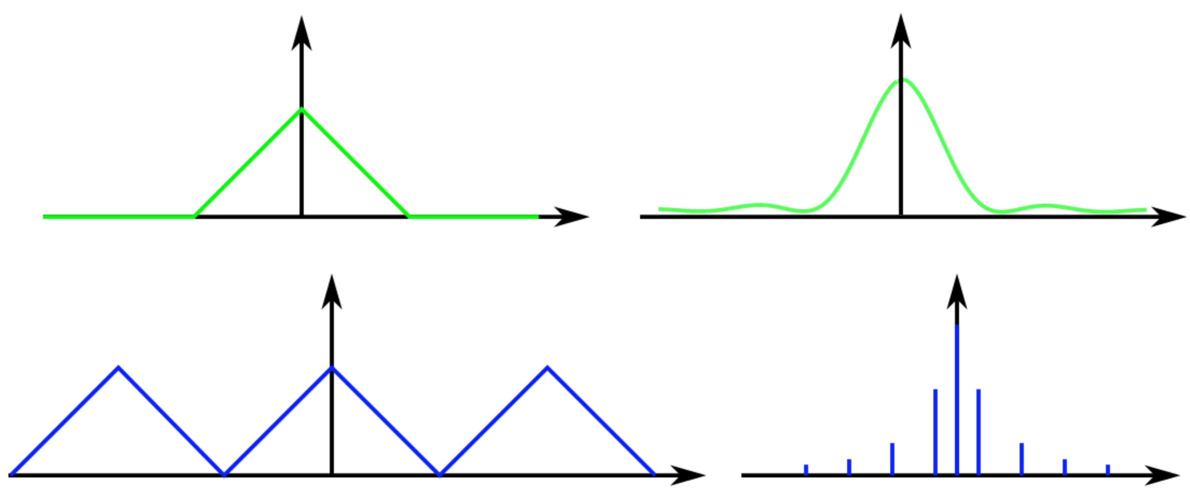
Обе теоремы мы оставим без доказательств.

Заметим, что если п.ф. определяется значениями в счетном числе точек, то различные х.ф. могут совпадать на целых отрезках. Так, например, х.ф. являются функции

$$\psi_X(t) = \begin{cases} 1 - |t| & |t| \le 1, \\ 0 & |t| \le 1, \end{cases} \qquad \psi_Y(t) = \begin{cases} \dots & \dots \\ 1 - |t|, & -1 \le t \le 1, \\ 1 - |t-2| & 1 \le t \le 3, \\ \dots & \dots . \end{cases}$$

совпадающие на целом отрезке, однако, различные. Более того, первая х.ф. соответствует абсолютно-непрерывному, а вторая — дискретному распределению.

Рис. 18: Первой (зеленой) х.ф. соответствует распределение с зеленой плотностью, второй (синей) – дискретное распределение с синей функцией масс



13.5 Характеристические функции для некоторых распределений

Выпишем некоторые характеристические функции:

1. Бернуллиевская величина — $pe^{it} + q$.
2. Геометрическая величина — $p/(1 - (1-p)e^{it})$.
3. Пуассоновская величина — $e^{\lambda(e^{it}-1)}$.
4. Равномерная на $[a, b]$ величина — $(e^{itb} - e^{ita})/(it(b-a))$. В частности при $a = -b$ получим $\sin(bt)/(bt)$.
5. Экспоненциальная величина — $\frac{\lambda}{\lambda - it}$.
6. Стандартная нормальная случайная величина — $e^{-t^2/2}$.

7. Распределение Коши — $e^{-|t|}$.
8. Гамма-распределение $Gamma(\lambda, m)$ — $(1-\lambda it)^{-m}$.

13.6 Характеристическая функция случайного вектора

Определение 13.3. Характеристической функцией случайного вектора $\vec{X} \in \mathbb{R}^n$ называют функцию

$$\psi_{\vec{X}}(\vec{t}) = \mathbf{E}\exp(i\langle\vec{X}, \vec{t}\rangle).$$

Сформулируем базовые свойства характеристической функции. Теорема единственности сохраняется.

1. $\psi(\vec{0}) = 1$. Более того
$$\psi_{\vec{X}}(0, \ldots, 0, t_{k_1}, 0, \ldots, 0, t_{k_m}, 0 \ldots 0) = \psi_{X_{k_1}, \ldots X_{k_m}}(t_{k_1}, \ldots, t_{k_m});$$
2. $\psi(\vec{t})$ непрерывная функция, равномерно непрерывная на $\mathbb{R}^n$.
3. Если величины X_i^k имеют конечное математическое ожидание, то $\psi(t)$ k раз дифференцируема, причем
$$\frac{\partial}{\partial t_{l_1}} \frac{\partial}{\partial t_{l_2}} \cdots \frac{\partial}{\partial t_{l_k}} \psi(\vec{t}) = \mathbf{E} X_{l_1} \cdots X_{l_k} e^{i\langle\vec{t}, \vec{X}\rangle},$$
где $l_1, \ldots, l_k \in \{1, \ldots, n\}$ могут содержать повторения.
Доказательство полностью аналогично одномерному случаю.
4. Для любых матрицы A размера $m \times n$ и вектора $\vec{b} \in \mathbb{R}^n$ выполнено свойство
$$\psi_{A\vec{X}+\vec{b}}(\vec{t}) = \exp\left(i\langle t, b\rangle\right) \psi_{\vec{X}}(A^T\vec{t}).$$
Доказательство.
$$\psi_{A\vec{X}+\vec{b}}(\vec{t}) = \mathbf{E}\exp\left(i\vec{t}^T A\vec{X} + i\vec{t}^T\vec{b}\right) = \exp\left(i\langle\vec{t}, \vec{b}\rangle\right) \mathbf{E}\exp\left(i\langle A^T\vec{t}, \vec{X}\rangle\right) = \exp\left(i\langle\vec{t}, \vec{b}\rangle\right) \psi_{\vec{X}}(A^T\vec{t}).$$
□
5. $\psi_{-\vec{X}}(\vec{t}) = \overline{\psi_{\vec{X}}(\vec{t})}$.
6. Векторы $\vec{X}_1, \ldots, \vec{X}_n$ независимы тогда и только тогда, когда
$$\psi_{\vec{X}_1, \ldots, \vec{X}_n}(\vec{t}_1, \ldots, \vec{t}_n) = \psi_{\vec{X}_1}(\vec{t}_1) \cdots \psi_{\vec{X}_n}(\vec{t}_n).$$
Иначе говоря, х.ф. вектора, состоящего из нескольких независимых блоков, равна произведению х.ф. блоков.

Доказательство. Из независимости вытекает требуемое свойство, поскольку функции независимых величин независимы. В обратную сторону можно получить ответ используя теорему единственности: как мы уже установили, будь блоки независимыми, х.ф. имела бы тот же самый вид. Но раз она однозначно определяет распределение, то у распределения с такой х.ф. блоки независимыми □
7. Характеристическая функция суммы независимых векторов есть произведение х.ф. каждого из них.

Доказательство.
$$\mathbf{E}e^{i\langle\vec{t}, \vec{X}_1 + \cdots + \vec{X}_n\rangle} = \mathbf{E}e^{i\langle\vec{t}, \vec{X}_1\rangle} \ldots e^{i\langle\vec{t}, \vec{X}_n\rangle} = \mathbf{E}e^{i\langle t, \vec{X}_1\rangle} \ldots \mathbf{E}e^{i\langle t, \vec{X}_n\rangle}.$$
□

13.7 Многомерное нормальное распределение

Теперь мы готовы дать еще одно определение многомерного нормального распределения

Определение 13.4. Вектор $\vec{X}$ имеет многомерное нормальное распределение $\mathcal{N}(\vec{a}, \Sigma)$, если ее х.ф. имеет вид

$$\psi_{\vec{X}}(\vec{t}) = \exp\left(i\langle\vec{a}, \vec{t}\rangle + \frac{1}{2}\vec{t}^T\Sigma\vec{t}\right).$$

Лемма 13.1. *Указанное определение равносильно предыдущему:* $\vec{X} \sim \mathcal{N}(\vec{a}, \Sigma)$ *если* $\vec{X} \stackrel{d}{=} A\vec{Y} + \vec{b}$, *где* $\Sigma = AA^T$, $\vec{Y}$ *— вектор с независимыми* $\mathcal{N}(0,1)$ *компонентами.*

Доказательство. Вектор $\vec{Y}$ имеет х.ф.

$$\psi_{\vec{Y}}(\vec{t}) = \exp(-t_1^2/2 - t_2^2/2 - \dots - t_k^2/2) = \exp\left(-\frac{1}{2}\langle\vec{t}, \vec{t}\rangle\right)$$

поскольку его компоненты независимы. При этом

$$\psi_{A\vec{Y}+\vec{b}}(\vec{t}) = \exp\left(i\langle\vec{t}, \vec{b}\rangle + \frac{1}{2}\vec{t}^T AA^T\vec{t}\right) = \exp\left(i\langle\vec{t}, \vec{b}\rangle + \frac{1}{2}\vec{t}^T\Sigma\vec{t}\right)$$

Значит, $\vec{X} = A\vec{Y} + \vec{b}$. И, напротив, любая симметричная неотрицательно определенная матрица Σ представима в виде AA^t для некоторой матрицы A (поскольку она диагонализируема в собственном базисе, причем с неотрицательными числами на диагонали). Тогда $A\vec{Y} + \vec{b}$ будет иметь распределение с х.ф.

$$\exp\left(i\langle\vec{t}, \vec{b}\rangle + \frac{1}{2}\vec{t}^T\Sigma\vec{t}\right).$$

Значит вектор с такой х.ф. имеет распределение $\mathcal{N}(\vec{b}, \Sigma)$. $\square$

14 Сходимость случайных величин

Пусть $X_1, \dots, X_n, \dots$ — случайные величины, определенные на некотором вероятностном пространстве $(\Omega, \mathcal{F}, \mathbf{P})$. Зададимся вопрос о том, какую сходимость этих величин можно рассматривать.

14.1 Сходимость почти наверное

Первое что приходит в голову — давайте просто рассмотрим поточечную сходимость X_n к некоторой величине X. Потребуем, чтобы при каждом ω $X_n(\omega)$ как обычная числовая последовательность сходилась к $X(\omega)$. Как мы уже обсуждали, если эта величина существует и конечна, то она случайная величина. Слова "при каждом" здесь более удобно заменить на 'при почти каждом'.

Определение 14.1. Величины X_n сходятся *с вероятностью 1* или *почти наверное* к X (обозначение $X_n \to X$ п.н.), если

$$\mathbf{P}(\omega : X_n(\omega) \to X(\omega),\ n \to \infty) = 1.$$

Пример 14.1. Пусть $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$, $\mathbf{P}$ — мера Лебега. Тогда последовательность величин $X_n(\omega) = \omega^n$ сходится п.н. к $X(\omega) = 0$. Фактически это обычная сходиомсть числовой последовательности, зависящей от параметра ω.

Лемма 14.1. *Сходимость* $X_n \to X$ *п.н. выполнена тогда и только тогда, когда при любом* $\varepsilon > 0$.

$$\lim_{N\to\infty} \mathbf{P}(\sup_{n>N} |X_n(\omega) - X(\omega)| > \varepsilon) = 0.$$

Доказательство. Заметим, что можно рассматривать только $\varepsilon = 1/M$ для всех целых M.

По определению при каком-либо ω последовательность $X_n(\omega)$ не сходится к $X(\omega)$ тогда и только тогда, когда

$$\exists M \in \mathbb{N}\ \forall N :\ \exists n > N\ |X_n(\omega) - X_m(\omega)| > 1/M.$$

Отсюда, сходимость п.н. $X_n \to X$ равносильна

$$\mathbf{P}\left(\exists M\ \forall N\ \sup_{n>N} |X_n(\omega) - X(\omega)| > \frac{1}{M}\right) = 0.$$

В свою очередь, указанные вероятности представимы в виде

$$\mathbf{P}\left(\bigcup_{M\in\mathbb{N}} \bigcap_{N\in\mathbb{N}} \bigcup_{n>N} \left\{\omega : |X_n(\omega) - X(\omega)| > \frac{1}{M}\right\}\right),$$

Вероятность объединения равна 0 тогда и только тогда, когда все события имеют нулевую вероятность, то есть

$$\mathbf{P}\left(\bigcap_{N\in\mathbb{N}} \left\{\omega : \sup_{n>N} |X_n(\omega) - X(\omega)| > \frac{1}{M}\right\}\right) = 0.$$

Так как события под знаком пересечения вложены по N, то

$$\mathbf{P}\left(\bigcap_{N\in\mathbb{N}} \left\{\omega : \sup_{n>N} |X_n(\omega) - X(\omega)| > \frac{1}{M}\right\}\right) = \lim_{N\to\infty} \mathbf{P}\left(\left\{\omega : \sup_{n>N} |X_n(\omega) - X(\omega)| > 1/M\right\}\right).$$

Значит, сходимость п.н. равносильна тому, что

$$\lim_{N\to\infty} \mathbf{P}\left(\left\{\omega : \sup_{n>N} |X_n(\omega) - X(\omega)| > \frac{1}{M}\right\}\right) = 0.$$

□

Аналогичным образом мы можем доказать следующий критерий Коши для сходимости почти наверное:

Лемма 14.2. *Последовательность X_n сходится п.н. тогда и только тогда, когда*

$$\mathbf{P}\left(\sum_{n,m>N} |X_n(\omega) - X_m(\omega)| > \varepsilon\right) \to 0,\ N \to \infty.$$

Отметим, что прямо из определения предел п.н. единственен с точностью до множества меры ноль: если $X_n \to X$ п.н. и $X_n \to Y$ п.н., то

$$\mathbf{P}(\omega : X(\omega) = Y(\omega)) = 1.$$

14.2 Сходимость по вероятности

Ослабим сходимость почти наверное, убрав во втором определении супремум:

Определение 14.2. Будем говорить, что если выполнено соотношение

$$\lim_{n\to\infty} \mathbf{P}(|X_n(\omega) - X(\omega)| > \varepsilon) = 0,$$

то последовательность X_n сходится к X *по вероятности*. Обозначать это будем $X_n \xrightarrow{P} X$.

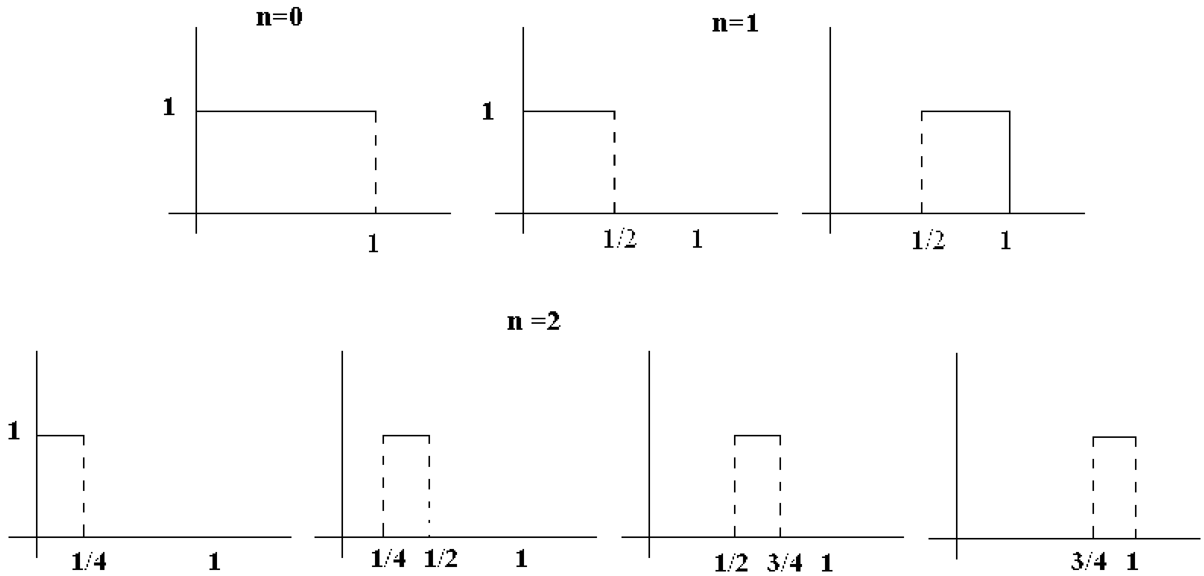
Мы видим, что из сходимости п.н. следует сходимость по вероятности. Обратное неверно, что показывает следующий пример:

Пример 14.2. Рассмотрим то же пространство, что и в примере 1 и положим $X_{n,k} = I_{[k/2^n,(k+1)/2^n)}(\omega)$, где $n \in \mathbb{N}$, $0 \le k \le 2^n - 1$.

Упорядочим полученные величины, положив $Y_k = X_{i,j}$, где $k = 2^i + j$, $j < 2^i$. Иначе говоря, i — это число знаков в двоичной записи k ($[\log_2 k]$), а j — это $k - 2^i$.

Тогда эта последовательность, разумеется, сходится по вероятности к 0, ведь множество тех ω, при которых X_k отличается от 0 есть 2^{-i}, то есть стремится к 0 с ростом k. Но п.н. эта последовательность к 0 не стремится, ведь для каждого ω, кроме 1 при каждом i найдется j, такое, что $X_{i,j}(\omega) = 1$.

Рис. 19: Плавающие ступеньки



Таким образом, сходимость по вероятности — более грубая, чем сходимость почти наверное. Тем не менее, справедливо следующее утверждение:

Лемма 14.3. *1. Пусть $X_n \xrightarrow{P} X$. Тогда существует такая последовательность n_k, что $X_{n_k} \xrightarrow{P} X$, $k \to \infty$.*

2. Пусть $\lim_{n,m\to\infty} \mathbf{P}(|X_n - X_m| > \varepsilon) = 0$. Тогда существует такая последовательность n_k, что

$$\lim_{N\to\infty} \mathbf{P}(\sup_{k,l>N} |X_{n_k} - X_{n_l}| > \varepsilon) = 0.$$

Это утверждение мы оставим без доказательства.

Также как и для сходимости почти наверное можно сформулировать критерий Коши сходимости по вероятности:

Лемма 14.4. *Последовательность X_n сходится по вероятности тогда и только тогда, когда при любом $\varepsilon > 0$*

$$\lim_{n,m\to\infty} \mathbf{P}(\omega : |X_n(\omega) - X_m(\omega)| > \varepsilon) = 0.$$

Доказательство. Пусть $X_n \xrightarrow{P} X$ при $n \to \infty$. Тогда при любых $\delta, \varepsilon > 0$ найдется такое N, что

$$\mathbf{P}(\omega : |X_n(\omega) - X(\omega)| > \varepsilon/2) < \delta/2$$

при всех $n > N$. Но тогда

$$\mathbf{P}(\omega : |X_n(\omega) - X(\omega)| > \varepsilon/2) < \delta/2, \quad \mathbf{P}(\omega : |X_m(\omega) - X(\omega)| > \varepsilon/2) < \delta/2$$

при $n, m > N$. Значит

$$\mathbf{P}(\omega : |X_n(\omega) - X_m(\omega)| > \varepsilon) \le \mathbf{P}(\omega : |X_n(\omega) - X(\omega)| > \varepsilon/2) + \mathbf{P}(\omega : |X_m(\omega) - X(\omega)| > \varepsilon/2) < \delta.$$

В силу произвольности $\delta > 0$ имеем требуемое.

Для доказательства обратной части воспользуемся Леммой 14.3. В силу нее найдется такая последовательность n_k, что

$$\lim_{N\to\infty} \mathbf{P}(\sup_{k,l>N} |X_{n_k} - X_{n_l}| > \varepsilon) = 0.$$

В силу критерия Коши сходимости п.н. X_{n_k} сходится к некой последовательности X п.н. Но тогда

$$\mathbf{P}(|X_m - X| > \varepsilon) \le \mathbf{P}(|X_m - X_{n_k}| > \varepsilon/2) + \mathbf{P}(|X_{n_k} - X| > \varepsilon/2).$$

С ростом m и k первое слагаемое правой части стремится к нулю по условию, а второе — в силу сходимости X_{n_k}. Отсюда имеем требуемое. □

Опять же предел по вероятности п.н. единственен: если $X_n \xrightarrow{P} X$, $X_n \xrightarrow{P} Y$, то

$$\mathbf{P}(\omega : X(\omega) = Y(\omega)) = 1.$$

Действительно, при любых положительных ε, δ выполнены неравенства

$$\mathbf{P}(|X_n - X| > \varepsilon) < \delta,\ \mathbf{P}(|X_n - Y| > \varepsilon) < \delta,$$

откуда

$$\mathbf{P}(|X - Y| > 2\varepsilon) \le \mathbf{P}(|X_n - X| > \varepsilon) + \mathbf{P}(|X_n - Y| > \varepsilon) < 2\delta.$$

В силу произвольности ε и δ

$$\lim_{\varepsilon\to 0} \mathbf{P}(|X - Y| > 2\varepsilon) = 0.$$

Но этот предел есть $\mathbf{P}(|X - Y| > 0)$ в силу непрерывности вероятностной меры. Значит $\mathbf{P}(X \ne Y) = 0$.

14.3 Сходимость в среднем порядка p

Определение 14.3. Пусть $p \ge 1$, $\mathbf{E}|X_n|^p < \infty$ и $\mathbf{E}|X_n - X|^p \to 0$. Тогда говорят, что X_n сходится в L^p к X ($X_n \xrightarrow{L^p} X$).

Такая сходимость называется также сходимостью в среднем порядка p. В частности, при $p = 1$ ее называют сходимостью в среднем, а при $p = 2$ — в среднеквадратичном.

В силу неравенства Ляпунова

$$(\mathbf{E}(X^p))^{1/p} \le (\mathbf{E}(X^q))^{1/q},\ p < q,$$

сходимость в L^q влечет сходимость в L^p при $q > p$.

Заметим, что все тот же пример с плавающей ступенькой показывает, что из этой сходимости не следует сходимость почти наверное. Более того, из сходимости почти наверное также не следует сходимость в L^p.

Пример 14.3. На все том же вероятностном пространстве $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$, $\mathbf{P}$ — мера Лебега, последовательность $X_n = e^n I_{0,1/n}$ сходится почти наверное к 0, но не сходится к нулю в L^p, поскольку

$$\mathbf{E}|X_n - X|^p = e^n \cdot \frac{1}{n} \to \infty.$$

Покажем, что из сходимости в L^p следует сходимость по вероятности. Нам понадобится следующая лемма.

Лемма 14.5. *1. Неравенство Маркова. Для любой неотрицательной величины X и любого $\varepsilon > 0$ выполнено неравенство*

$$\mathbf{P}(X \ge \varepsilon) \le \frac{\mathbf{E}X}{\varepsilon}$$

2. Неравенство Чебышева. Для любой величины X с конечной дисперсией и любого $\varepsilon > 0$ выполнено неравенство

$$\mathbf{P}(|X - \mathbf{E}X| > \varepsilon) \le \frac{\mathbf{D}X}{\varepsilon^2}.$$

3. Для любой величины X с конечным $\mathbf{E}|X|^p$ и любого $\varepsilon > 0$ выполнено неравенство

$$\mathbf{P}(|X - \mathbf{E}X| > \varepsilon) \le \frac{\mathbf{E}|X|^p}{\varepsilon^p}. \tag{8}$$

Доказательство. 1. Заметим, что $X \geq XI_{X\geq\varepsilon}$ в силу неотрицательности величины X. Значит,

$$\mathbf{E}X \geq \mathbf{E}XI_{X\geq\varepsilon} \geq \varepsilon\mathbf{E}I_{X\geq\varepsilon} = \varepsilon\mathbf{P}(X \geq \varepsilon).$$

Имеем требуемое неравенство.

2. Заметим, что если $Y = (X - \mathbf{E}X)^2$, то $\mathbf{E}Y = \mathbf{D}X$, откуда искомое неравенство прямо вытекает из неравенства Маркова.

3. Аналогичным образом при $Y = |X - \mathbf{E}X|^p$ получаем неравенство (8).

□

Из последнего соотношения

$$\mathbf{P}(|X_n - X| \geq \varepsilon) \leq \mathbf{E}|X_n - X|^p/\varepsilon^p,$$

откуда сходимость в L^p влечет сходимость по вероятности.

Опять же предел в L^p единственен п.н. (хотя бы потому, что он и предел по вероятности тоже).

Можно получать сходимость в L^p из сходимости п.н. при дополнительных условиях, пользуясь теоремами о монотонной или мажорируемой сходимостях:

1. Если $X_n \to X$ п.н., X_n п.н. монотонно стремятся к X, $\mathbf{E}|X_n|^p < \infty$, $\mathbf{E}|X|^p < \infty$, то $X_n \xrightarrow{L^p} X$.

2. Если $X_n \to X$ п.н., X_n п.н. стремятся к X, $|X_n| \leq Y$ при всех n и некотором Y: $\mathbf{E}|Y|^p < \infty$, то $X_n \xrightarrow{L^p} X$.

14.4 Сходимость по распределению

Определение 14.4. Последовательность величин X_n (или последовательность распределений F_n) мы назовем *сходящейся по распределению*, если $\mathbf{P}(X_n \leq x) \to \mathbf{P}(X \leq x)$ для всех x, таких, что F_X непрерывна в точке x. Обозначают такую сходимость $X_n \xrightarrow{d} X$.

Эта сходимость вообще не связана с конкретным поведением X_n на Ω, а связана только с ф.р. X_n. Отсюда, например, если $X_n \xrightarrow{d} X$, а Y имеет то же распределение, что и X, то $X_n \xrightarrow{d} Y$. Таким образом, пределом у X_n по распределению можно считать любую из величин с таким распределением, тогда как предел по вероятности определяются с точностью до множества меры 0.

Пример 14.4. Стоит заметить, что условие *в точках непрерывности* X по существу. Например, на все том же вероятностном пространстве величины $X_n = 1/n$ сходятся к 0 почти наверное. Между тем $F_{X_n}(0) = 0$, $F_0(0) = 1$.

Лемма 14.6. *Из сходимости по вероятности следует сходимость по распределению.*

Доказательство. Действительно,

$$\mathbf{P}(X_n \leq x) = \mathbf{P}(|X_n - X| \leq \varepsilon, X_n \leq x) + \mathbf{P}(|X_n - X| > \varepsilon, X_n \leq x) \leq \mathbf{P}(X \leq x + \varepsilon) + o(1),$$

откуда

$$\limsup_{n\to\infty} \mathbf{P}(X_n \leq x) \leq \mathbf{P}(X \leq x + \varepsilon) = F_X(x + \varepsilon).$$

В силу произвольности ε и непрерывности F_X справа имеем оценку $\limsup \mathbf{P}(X_n \leq x) \leq \mathbf{P}(X \leq x)$. Аналогично получаем нижнюю оценку

$$\mathbf{P}(X_n \leq x) \geq \mathbf{P}(X_n \leq x, |X_n - X| \leq \varepsilon) \geq \mathbf{P}(X \leq x - \varepsilon, |X_n - X| \leq \varepsilon) = \mathbf{P}(X \leq x - \varepsilon) - \mathbf{P}(|X_n - X| \leq \varepsilon),$$

откуда

$$\lim_{n\to\infty} \mathbf{P}(X_n \leq x) = \lim_{\varepsilon\to 0} \mathbf{P}(X \leq x - \varepsilon) = \mathbf{P}(X \leq x),$$

где мы пользуемся непрерывностью F_X в точке x слева, поскольку это точка непрерывности. □

Стоит заметить, что если $X_n \xrightarrow{d} C$, где C — константа, то $X_n \xrightarrow{P} C$ (это одна из домашних задач).

Лемма 14.7. *Следующие условия эквивалентны:*

1. $X_n \xrightarrow{d} X$, $n \to \infty$;
2. $\mathbf{E}f(X_n) \to \mathbf{E}f(X)$, $n \to \infty$ *для любой непрерывной ограниченной функции* f;
3. $\psi_{X_n}(t) \to \psi_X(t)$, $n \to \infty$, *при всех* t, *где* ψ *— х.ф.*

Это утверждение достаточно сложно и мы оставим его без доказательства.

Соответственно, можно доказывать сходимость по распределению через сходимость характеристических функций. Это бывает удобно, особенно при работе с суммами случайных величин.

14.5 Функции от сходящихся величин

Если рассмотреть g — непрерывную функцию, а $X_n \xrightarrow{d} X$, то при всех $f \in CB$ (непрерывных ограниченных функциях) выполнено соотношение

$$\mathbf{E}f(X_n) \to \mathbf{E}f(X),\ n \to \infty,$$

в частности,

$$\mathbf{E}f(g(X_n)) \to \mathbf{E}f(g(X)),\ n \to \infty,$$

поскольку непрерывная функция от непрерывной непрерывна и композиция ограниченная функции с любой функцией ограничена. Значит,

$$g(X_n) \xrightarrow{d} g(X),\ n \to \infty.$$

Следовательно, сходимость по распределению сохраняется при действии непрерывных функций.

Сходимость в L^p не обязательно хотя бы потому, что у $g(X)$ может оказаться бесконечным p-й момент.

Что можно сказать про две остальных сходимости?

Лемма 14.8. *Пусть* $X_{n,1} \to X_1$ *п.н. при* $n \to \infty$, $X_{n,2} \to X_2$ *п.н. при* $n \to \infty$, $\ldots$, $X_{n,k} \to X_k$ *п.н. при* $n \to \infty$. *Тогда для любой непрерывной функции* $g : \mathbb{R}^k \to \mathbb{R}$ *п.н.*

$$g(X_{n,1}, \ldots, X_{n,k}) \to g(X_1, \ldots, X_k).$$

Доказательство. Рассмотрим $A_i = \{\omega : X_{n,i}(\omega) \to X_i(\omega)\}$. По условию $\mathbf{P}(A_i) = 1$ при всех i. Значит,

$$\mathbf{P}\left(\bigcap_{i=1}^{k} A_i\right) = 1 - \mathbf{P}\left(\bigcup_{i=1}^{k} \overline{A}_i\right) \geq 1 - \sum_{i=1}^{k} \mathbf{P}(A_i) = 1.$$

При всех $\omega \in \bigcap_{i=1}^k A_i$ верны все k сходимостей $X_{n,i}(\omega) \to X_i$, $i \leq k$. Следовательно,

$$g(X_{n,1}(\omega), \ldots, X_{n,k}(\omega)) \to g(X_1(\omega), \ldots, X_k(\omega)),\ n \to \infty.$$

Следовательно,

$$\mathbf{P}\left(g(X_{n,1}(\omega), \ldots, X_{n,k}(\omega)) \to g(X_1(\omega), \ldots, X_k(\omega))\right) \geq \mathbf{P}\left(\bigcap_{i=1}^{k} A_i\right) \geq 1,$$

что и означает сходимость п.н. □

Как и сходимость п.н. сходимость по вероятности устойчива к действию непрерывных функций:

Лемма 14.9. *Пусть* $X_{n,1} \xrightarrow{P} X_1$ *при* $n \to \infty$, $X_{n,2} \xrightarrow{P} X_2$ *при* $n \to \infty$, $\ldots$, $X_{n,k} \xrightarrow{P} X_k$ *при* $n \to \infty$. *Тогда для любой непрерывной функции* $g : \mathbb{R}^k \to \mathbb{R}$

$$g(X_{n,1}, \ldots, X_{n,k}) \xrightarrow{P} g(X_1, \ldots, X_k),\ n \to \infty.$$

Доказательство. Заметим, что при любом $\varepsilon_1 > 0$ и любом i найдется такое $M > 0$, что

$$\mathbf{P}(X_i \notin [-M, M]) < \varepsilon_1. \tag{9}$$

Это утверждение прямо следует из непрерывности вероятностной меры: $\{\omega : X_i(\omega) \notin [-M, M]\}$ с ростом M сужаются к пустому множеству, а значит их вероятность стремится к нулю. Выберем M так, чтобы выполнялось свойство (9), причем M и $-M$ есть точки непрерывности F_{X_i}: $\mathbf{P}(X_i = -M) = \mathbf{P}(X_i = M) = 0$. Это можно сделать, поскольку M, удовлетворяющих (9), континуум, а точек разрыва ф.р. не более чем счетное.

Тогда при достаточно больших n в силу сходимости

$$F_{X_{n,i}}(-M) < F_{X_i}(-M) + \varepsilon_1/2, \quad F_{X_{n,i}}(M) > F_{X_i}(M) - \varepsilon_1/2,$$

откуда $\mathbf{P}(X_{n,i} \notin [-M, M]) < 2\varepsilon_1$ при всех достаточно больших n и всех i. Значит,

$$\mathbf{P}(\vec{X}_n \notin [-M, M]) < 2k\varepsilon_1, \quad \mathbf{P}(\vec{X} \notin [-M, M]) < 2\varepsilon_1.$$

При этом функция g на компакте $D = [-M, M]^k$ является равномерно непрерывной на нем, а значит для любого $\varepsilon > 0$ найдется $\delta > 0$, для которого из $||\vec{x} - \vec{y}|| < \delta$, $\vec{x}, \vec{y} \in D$, то $|g(\vec{x}) - g(\vec{y})| < \varepsilon$. Выберем n достаточно большим, что

$$\mathbf{P}(||\vec{X}_n - \vec{X}|| \geq \delta) < \sum_{i=1}^{k} \mathbf{P}(|X_{n,i} - X_i| \geq \delta/\sqrt{k}) < \varepsilon_1,$$

пользуясь сходимостью по вероятности. Тогда

$$\mathbf{P}(|g(\vec{X}_n) - g(\vec{X})| > \varepsilon) < \mathbf{P}(||\vec{X}_n - \vec{X})|| \geq \delta) + \mathbf{P}(\vec{X}_n \notin D) + \mathbf{P}(\vec{X} \notin D) + \mathbf{P}(||\vec{X}_n - \vec{X}|| < \delta, \vec{X}_n \in D, \vec{X} \in D).$$

В силу наших условий первые три вероятности в сумме не превосходят $(2k+3)\varepsilon_1$, а последняя равна 0 в силу равномерной непрерывности. В силу произвольности ε_1 имеем требуемое. □

В частности, отсюда сходящиеся п.н. и по вероятности последовательности можно суммировать, вычитать, умножать или делить.

Для сходимости по распределению можно получать какие-то результаты про сходимость сумм или произведений, если одна из величин вырождена:

Лемма 14.10 (Лемма Слуцкого)**.** *Пусть $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{d} c$. Тогда*

1. *$X_n + Y_n \xrightarrow{d} X + c$, $n \to \infty$;*
2. *$X_n Y_n \xrightarrow{d} cX$, $n \to \infty$;*

Пример 14.5. В качестве примера докажем, что если $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{d} 0$, то $X_n Y_n \xrightarrow{d} 0$. Фиксируем $x > 0$

$$\mathbf{P}(X_n Y_n \geq x) = \mathbf{P}(X_n Y_n \geq x, |Y_n| \leq \varepsilon) + \mathbf{P}(X_n Y_n \geq x, |Y_n| \geq \varepsilon) \leq \mathbf{P}(|X_n| \geq x/\varepsilon) + \mathbf{P}(|Y_n| \geq \varepsilon),$$

где ε выбрано так, что x/ε — точка непрерывности $F_X()$ (а точек разрыва у монотонной ограниченной функции не более чем счетное число). Устремляя ε к 0, имеем $\mathbf{P}(X_n Y_n \geq x) \leq \mathbf{P}(|X| \geq x/\varepsilon) + \mathbf{P}(0 \geq \varepsilon)$, $x > 0$. Эта величина засчет выбора ε может быть сделана сколь угодно малой, откуда $\mathbf{P}(X_n Y_n \leq x) \to 1$, $n \to \infty$. Аналогично для отрицательных x $\mathbf{P}(X_n Y_n \leq x) \to 0$. Что и требовалось доказать.

15 Предельные теоремы

Завершим наш курс мы рядом важных результатов, связанных с предельным поведением сумм независимых величин.

15.1 Закон больших чисел

Определение 15.1. Говорят, что последовательность X_n удовлетворяет закону больших чисел (ЗБЧ), если

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{P} a, \ n \to \infty,$$

где a — некоторая константа.

Определение 15.2. Говорят, что последовательность X_n удовлетворяет усиленному закону больших чисел (УЗБЧ), если п.н.

$$\frac{X_1 + \dots + X_n}{n} \to a, \; n \to \infty,$$

где a — некоторая константа.

Поскольку из сходимости п.н. следует сходимость по вероятности, усиленный закон более сильный.

Заметим, что если бы из сходимости по вероятности следовала бы сходимость математических ожиданий, то можно было бы найти a как

$$a = \lim_{n\to\infty} \mathbf{E}\left(\frac{X_1 + \dots + X_n}{n}\right) = \lim_{n\to\infty} \frac{\mathbf{E}X_1 + \dots + \mathbf{E}X_n}{n}.$$

Однако, вообще говоря, это не так, и есть примеры, когда величины X_i не имеют математического ожидания, но удовлетворяют ЗБЧ. Тем не менее именно такое выражение для a бывает во многих типичных ситуациях.

Простейшим примером ЗБЧ является хорошо понятный житейский факт: при большом числе бросаний монеты скорее всего доля выпавших орлов будет близка к вероятности выпадания орла.

Таким образом, если мы покажем, что в схеме Бернулли выполнен ЗБЧ, то фактически обоснуем эмпирическое определение вероятности, о котором мы говорили на первой лекции. Если определять вероятность так, как ее определили мы, то эмпирическое определение также будет верным: если A_i независимые события с равной вероятностью, то доля случившихся за n испытаний событий A_i будет близка к вероятности $\mathbf{P}(A_1)$:

$$\frac{1}{n}\sum_{i=1}^{n} I_{A_i} \xrightarrow{P} \mathbf{P}(A_1), \; n \to \infty.$$

15.1.1 ЗБЧ Чебышева

Теорема 15.1. *Пусть X_i — некоррелированные (то есть $\operatorname{cov}(X_i, X_j) = 0$ при $i \neq j$) величины, $\mathbf{E}X_i = 0$, $\sigma_i^2 = \mathbf{D}X_i$ таковы, что $\sum_{i=1}^n \sigma_i^2 = o(n^2)$, $n \to \infty$. Тогда X_i удовлетворяют ЗБЧ с $a = 0$.*

Доказательство. Фиксируем $\varepsilon > 0$. Воспользуемся неравенством Чебышева, доказанным на прошлой лекции:

$$\mathbf{P}\left(|S_n - \mathbf{E}S_n| > n\varepsilon\right) \le \frac{\mathbf{D}S_n}{n^2\varepsilon^2} = \frac{\sum_{i=1}^n \sigma_i^2}{n^2\varepsilon^2} = o(1),$$

где в предпоследнем равенстве мы воспользовались тем, что $\mathbf{D}S_n = \mathbf{D}X_1 + \dots + \mathbf{D}X_n$ в силу некоррелированности X_i. При этом $\mathbf{E}S_n = 0$, откуда

$$\mathbf{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon\right) \to 0, \; n \to \infty.$$

Теорема доказана. □

Замечание 15.1. Если X_i — некоррелированные величины с ненулевым средним, удовлетворяющие теореме, то переходя к $Y_i = X_i - \mathbf{E}X_i$, получаем

$$\frac{X_1 + \dots + X_n}{n} - \frac{\mathbf{E}X_1 + \dots + \mathbf{E}X_n}{n} \xrightarrow{P} 0, \; n \to \infty.$$

Следствие 15.1. Если X_i — некоррелированные величины с нулевым средним и их дисперсии ограничена в совокупности, то они удовлетворяют ЗБЧ.

15.1.2 ЗБЧ Хинчина

В случае независимых одинаково распределенных величин условие существования дисперсии не требуется:

Теорема 15.2. *Пусть X_i — независимые одинаково распределенные величины с конечным математическим ожиданием $\mathbf{E}X_1$. Тогда X_i удовлетворяют ЗБЧ с $a = \mathbf{E}X_1$.*

Доказательство. Воспользуемся тем, что сходимость по распределению к $\mathbf{E}X_1$ равносильна сходимости по вероятности, поскольку это константа. Значит, достаточно показать, что

$$\psi_{(S_n-na)/n}(t) \to 1,\ n \to \infty,$$

при любом t. При этом

$$\psi_{(S_n-na)/n}(t) = \psi_{S_n-an}(t/n) = \psi^n_{X_1-a}(t/n)$$

в силу свойств х.ф. При этом

$$\psi_{X_1-a}\left(\frac{t}{n}\right) = 1 + \frac{it\mathbf{E}(X_1-a)}{n} + o\left(\frac{1}{n}\right) = 1 + o\left(\frac{1}{n}\right),$$

откуда

$$\psi^n_{X_1-a}(t/n) = \left(1 + o\left(\frac{1}{n}\right)\right)^n \to 1,$$

что и требовалось доказать. □

Оказывается, что в условиях ЗБЧ Хинчина выполнено более сильное утверждение: усиленный закон больших чисел Колмогорова:

Теорема 15.3. *Пусть X_i — независимые одинаково распределенные величины с конечным математическим ожиданием $\mathbf{E}X_1$. Тогда X_i удовлетворяют УЗБЧ с $a = \mathbf{E}X_1$.*

Этот факт мы доказывать не будем.

15.2 Центральные предельные теоремы

Сам факт сходимости в законе больших чисел недостаточен для практических целей. Более точные результаты дает центральная предельная теорема:

Теорема 15.4. *Пусть X_i — н.о.р. сл. в., $\mathbf{E}X_1 = a$, $\mathbf{D}X_1 = \sigma^2 > 0$. Тогда*

$$\mathbf{P}\left(\frac{S_n - na}{\sqrt{n}\sigma} \le x\right) \to \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt,\ n \to \infty.$$

Иначе говоря, $(S_n - na)/(\sqrt{n}\sigma)$ сходится по распределению к $Z \sim \mathcal{N}(0,1)$.

Доказательство. Доказательство ЦПТ также будет основано на характеристических функциях. Покажем, что

$$\psi_{\frac{S_n-na}{\sigma\sqrt{n}}}(t) = \psi_{\frac{S_n-na}{\sigma}}\left(\frac{t}{\sqrt{n}}\right) = \psi^n_{\frac{X_1-a}{\sigma}}\left(\frac{t}{\sqrt{n}}\right)$$

При этом при $s \to 0$

$$\psi_{\frac{X_1-a}{\sigma}}(s) = 1 + \frac{is\mathbf{E}(X_1-a)}{\sigma} + \frac{(is)^2\mathbf{E}(X_1-a)^2}{2\sigma^2} + o(s^2) = 1 - \frac{s^2}{2} + o(s^2).$$

Значит,

$$\psi_{\frac{S_n-na}{\sigma\sqrt{n}}}(t) = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \to e^{-t^2/2},\ n \to \infty.$$

Характеристическая функция в правой части является х.ф. $\mathcal{N}(0,1)$ величины, что и доказывет теорему. □

Мы уже обсуждали пользу этой теоремы в дискретном случае, а теперь только заметим новый аспект понимания этой теоремы: по необъяснимой причине наши центрированные и нормированные суммы

$$\frac{S_n - na}{\sqrt{n}\sigma}$$

при больших n имеют распределение близкое к стандартному нормальному. Тем самым мы можем приближать распределения таких сумм распределением $\mathcal{N}(0,1)$.

Правда, мы по-прежнему можем утверждать только сходимость и не можем, вообще говоря, оценить погрешность такого приближения при заданном n. Оказывается, что в общем случае неравенство Берри-Эссеена, которое мы формулировали для дискретных величин, также справедливо.

15.3 Теорема Ляпунова

Откажемся от условия одинаковой распределенности.

Предположим, что $X_{n,i}$, $i \le n$, при каждом n есть независимые величины, $\mathbf{E}X_{n,i} = 0$, $\mathbf{E}X_{n,i}^2 = \sigma_{n,i}^2$, $\mathbf{E}|X_{n,i}|^3 = \rho_{n,i} < +\infty$, $S_n = X_{n,1} + \cdots + X_{n,n}$.

Таким образом, у нас есть одна величина $X_{1,1}$, две независимых величины $X_{2,1}, X_{2,2}$, три независимых величины $X_{3,1}, X_{3,2}, X_{3,3}$ и так далее. Мы рассматриваем суммы по строкам таких величин и изучаем их распределение. Такая же схема серий изучалась нами для теоермы Пуассона.

Предположим, что $B_n^2 = \sum_{i=1}^n \sigma_{n,i}^2$, $C_n^3 = \sum_{i=1}^n \rho_{n,i}$. Условием Ляпунова назовем условие

$$\frac{C_n}{B_n} \to 0, \quad n \to \infty.$$

Оказывается, что это достаточное условие для того, чтобы распределение S_n/B_n приближалось к нормальному распределению.

Теорема 15.5. *При выполнении условия Ляпунова*

$$\frac{S_n}{B_n} \xrightarrow{d} Z \sim \mathcal{N}(0,1),\ n \to \infty.$$

Доказательство. Доказательство также проведем через характеристические функции

$$\psi_{\frac{S_n}{B_n}}(t) = \psi_{S_n}\left(\frac{t}{B_n}\right) = \prod_{j=1}^n \psi_{X_{n,j}}\left(\frac{t}{B_n}\right).$$

При этом при любом s

$$\left|\psi_{X_{n,j}}(s) - 1 - is\mathbf{E}X_{n,j} - \frac{(is)^2}{2}\mathbf{E}X_{n,j}^2\right| = \left|\mathbf{E}e^{isX_{n,j}} - 1 - is\mathbf{E}X_{n,j} - \frac{(is)^2}{2}\mathbf{E}X_{n,j}^2\right| \le \mathbf{E}\left|e^{isX_{n,j}} - 1 - isX_{n,j} - \frac{(isX_{n,j})^2}{2}\right|.$$

Воспользуемся неравенством

$$\left|e^{i\varphi} - 1 - i\varphi - \frac{(i\varphi)^2}{2}\right| \le \frac{|\varphi|^3}{6},$$

справедливым при всех вещественных φ, которое мы оставим без доказательства. Значит,

$$\left|\psi_{X_{n,j}}(s) - 1 - \frac{(is\sigma_{n,j})^2}{2}\right| \le \frac{|s|^3\rho_{n,j}}{6}.$$

Значит,

$$\psi_{X_{n,j}}(s) = 1 - \frac{\sigma_{n,j}^2 s^2}{2} - \frac{\theta_j(s)\rho_{n,j}|s|^3}{6},$$

где $|\theta_j| \le 1$. Значит,

$$\prod_{j=1}^n \psi_{X_{n,j}}\left(\frac{t}{B_n}\right) = \prod_{j=1}^n \left(1 - \frac{\sigma_{n,j}^2 t^2}{2B_n^2} - \frac{\theta_j(t/B_n)\rho_{n,j}|t|^3}{6B_n^3}\right).$$

При этом в силу неравенства Ляпунова

$$\frac{\sigma_{n,j}^2}{B_n^2} = \mathbf{E}\left(\frac{X_{n,j}}{B_n}\right)^2 \le \left(\frac{\mathbf{E}|X_{n,j}|^3}{B_n^3}\right)^{2/3} = \left(\frac{\rho_{n,j}}{B_n^3}\right)^{2/3} \le \frac{C_n^2}{B_n^2} \to 0,\ n \to \infty.$$

Таким образом, второй и третий член в разложении х.ф. равномерно по i стремятся к нулю, а значит мы можем разложить $\ln\psi$ в окрестности точки 1:

$$\ln\psi_{X_{n,j}}\left(\frac{t}{B_n}\right) = -\frac{\sigma_{n,j}^2 t^2}{2B_n^2} - \frac{\theta_i(t/B_n)\rho n, j|t|^3}{6B_n^3} + o(1)\left(\frac{\sigma_{n,j}^2 t^2}{2B_n^2} + \frac{\theta_i(t/B_n)\rho_{n,j}|t|^3}{6B_n^3}\right),$$

где $o(1)$ равномерно мало по i. Вообще-то с функцией $\ln\psi$ для комплексных ψ могут быть проблемы (в частности, она неоднозначна), но мы рассматриваем ее в окрестности точки 1, в которой для нее справедливо разложение Тейлора. Значит,

$$\sum_{j=1}^{n}\ln\psi_{X_{n,j}}\left(\frac{t}{B_n}\right) = -\frac{(1+o(1))\sum_{j=1}^{n}\sigma_{n,j}^2 t^2}{2B_n^2} - \frac{(1+o(1))\sum_{j=1}^{n}\theta_{n,j}\rho_{n,j}|t|^3}{6B_n^3} = -\frac{1}{2}t^2 + r_n + o(1)(1+|r_n|),$$

где

$$|r_n| \le \frac{\sum_{j=1}^{n}|\theta_{n,j}|\rho_{n,j}|t|^3}{6B_n^3} \le \frac{C_n^3}{6B_n^3} = o(1).$$

Таким образом,

$$\psi_{S_n/B_n}(t) = \prod_{j=1}^{n}\psi_{X_{n,j}}\left(\frac{t}{B_n}\right) \to \exp\left(-\frac{t^2}{2}\right).$$

Что и требовалось доказать. $\square$