Теория вероятностей и математическая статистика

pre-α version

И. А. Лимар
ivan.limar95@gmail.com

Санкт-Петербург
2025

Оглавление

Предисловие к *pre-α version*

Летом 2022 года возникла идея сделать расширенную версию конспекта лекций, рассказываемые мной студентам. Прежде всего хотелось все записи собрать в одном месте. Помимо ввиду недостаточности времени доказательства некоторых теорем и утверждений привожу лишь кратко или не затрагиваю вовсе, но все равно хочется их отразить в записях. Плюс некоторые темы при первом знакомстве могут быть сложными для студентов, поэтому некоторой *абстрактностью* приходится жертвовать, но, как правило, находятся хотя бы два-три студента, которым это было бы интересно, поэтому хочется это отразить (например, сейчас в пункте **"Об области определения вероятности*"** говорится о контр-примере, иллюстрирующем необходимость рассмотрения сигма-алгебр; к примеру, на мой взгляд, рассказывать *условные распределения* в первый раз в самом общем виде – не всегда лучшая идея). С другой стороны, есть желание записи сделать максимально подробными и снабдить как, чтобы всем желающим записи были полезны.

До текущего момента удалось сделать почти все достаточные для *скелета* задуманные содержательные моменты, хотя еще нужно доработать все главы, особенно **"Предельные методы теории вероятностей"** и **"Линейные статистические модели"**: добавить содержательные комментарии, перепроверить и доработать определения, добавить больше примеров и упражнений, записать более подробно доказательства, привести все к единому стилю, добавить больше источников (задачи можно перечислять долго). Надеюсь, существенная часть этого будет уже в *α-версии.*

Всего отмеченного выше не получилось бы без добровольцев и помощников, которых считаю нужным отметить. Выражаю благодарность **Матвею Колесову** и **Денису Карпову**, любезно согласившиеся предоставить электронные рукописи, на основе которых набиралась существенная часть текста; **Тимофею Иванову** и **Чулкову Алексею** – за предоставление своих *TeX*-исходников, которые тоже были и будут полезными при работе с текстом. Также хочется поблагодарить **Максима Шехунова**, **Георгия Каданцева** и особенно **Полину Деревицкую**, **Дарью Фирсову** и **Юлию Сандракову**, которые непосредственно помогали набирать текст и без которых работа существенно затянулась бы.

21 *июля* 2023 *года.*

Введение

Тут будет введение.

Глава 1

Введение в теорию вероятностей

В этой главе мы рассмотрим математические основания теории вероятностей – аксиоматику теории вероятностей. Также мы затронем примеры вероятностных моделей, а именно:

- *классическую*, в рамках которой вероятность события определяется согласно широко известному принципу в виде отношения количества благоприятных исходов ко всем;
- *геометрическую* вероятность – отношение длин/площадей/объемов;
- *схему Бернулли* как формальную модель, описывающую подбрасывание монеты.

Кроме того, мы рассмотрим теорему Байеса – в некотором смысле фундаментальное утверждение в философии и науке.

1.1 Вероятностное пространство

В любом разделе математики сначала вводятся фундаментальные понятия, на основе которых строится вся дальнейшая теория: в математическом анализе прежде всего определяется понятие вещественного числа, в линейной алгебре – векторного пространства, в теоретической информатике – модели вычислений. В теории вероятностей фундаментальным понятием является *вероятностное пространство.*

Непосредственно перед определением вероятностного пространства нам понадобятся понятия *алгебры* и *сигма-алгебры* событий.

Определение 1.1. ***Алгебра.***

Пусть Ω – множество, $\mathbf{F}$ – система подмножеств множества Ω. Система множеств $\mathbf{F}$ называется ***алгеброй*** *, если выполняются следующие условия:*

1. $\forall A, B \in \mathbf{F} \Rightarrow A \cup B \in \mathbf{F}$,
2. $\forall A \in \mathbf{F} \Rightarrow \overline{A} \in \mathbf{F}$, *где* $\overline{A} = \Omega \setminus A$.

Другими словами, система подмножеств Ω называется *алгеброй*, если оно замкнуто относительно операций объединения и дополнения. Отметим также несложные соображения в виде замечания.

Замечание. *Алгебра обладает следующими свойствами:*

- *Алгебра замкнута относительно пересечения и разности, так как* $A \cap B = \overline{\overline{A} \cup \overline{B}}$ *и* $A \setminus B = A \cap \overline{B}$ *соответственно. Также алгебра замкнута относительно конечного числа пересечений, объединений, дополнений и разности;*

- $\Omega \in \mathbf{F}$, *так как* $A \cup \overline{A} = \Omega$;

- $\varnothing \in \mathbf{F}$, *ведь* $\varnothing = \overline{\Omega}$.

Далее нам понадобится понятие *сигма-алгебры.*

Определение 1.2. ***Сигма-алгебра.***

Пусть $\mathcal{F}$ *алгебра множеств.* $\mathcal{F}$ *называется* ***сигма-алгеброй****, если она замкнута относительно объединения счётного числа множеств.*

Алгебру, как правило, мы будем обозначать в виде жирных заглавных латинских букв $\mathbf{F}$, $\mathbf{G}$, а *сигма-алгебру* – в виде прописных заглавных латинских букв $\mathcal{F}$, $\mathcal{G}$.

Теперь мы готовы сформулировать определение *вероятностного пространства.*

Определение 1.3. ***Вероятностное пространство.***

Тройка $(\Omega, \mathcal{F}, \mathrm{P})$*, где* $\mathcal{F}$ *– сигма-алгебра подмножеств* Ω*, и* $\mathrm{P} : \mathcal{F} \to \mathbb{R}$ *– функция, удовлетворяющая следующим аксиомам:*

- $\forall A \in \mathcal{F} \Rightarrow \mathrm{P}(A) \geqslant 0$,

- $\mathrm{P}(\Omega) = 1$,

- $\forall (A_j)_{j=1}^{\infty} : A_j \in \mathcal{F}, \forall i, j \in \mathbb{N} \;\; A_i \cap A_j = \varnothing \Rightarrow \;\mathrm{P}\left(\bigcup\limits_{j=1}^{\infty} A_j\right) = \sum\limits_{j=1}^{\infty} \mathrm{P}(A_j)$,

называется ***вероятностным пространством****.*

Множество Ω мы будем интерпретировать как *множество элементарных событий.* В определении на него мы не наложили никаких ограничений, так что у него может быть совершенно произвольная структура: оно может быть конечным, счётным или континуальным, оно может состоять из объектов любого типа.

Сигма-алгебру $\mathcal{F}$ мы будем трактовать как *сигма-алгебру событий*, а его элементы мы будем называть *событиями.* Кроме того, часто вместо термина объединение событий мы будем писать *сумма событий* и на математическом языке писать $A+B$, пересечение событий – *произведение событий*, которое кратко часто будем обозначать как AB, а разность двух событий как $A - B$.

Функцию P мы будем называть *вероятностью*, которая является неотрицательной, *нормированной* и *счетно-аддитивной* функцией.

Замечание. *Знакомый с теорией меры читатель заметит, что вероятность – нормированная мера. Желающие более детально познакомиться с теорией меры могут посмотреть, например, в учебник по вещественному анализу [7] или в монографию [5], где в вводном разделе вероятность определена на языке теории меры.*

Событие $\varnothing$ будем называть *невозможным*, а Ω – *достоверным.* Также мы для обозначения специальных событий или их групп часто будем пользоваться специфическими терминами, которые в дальнейшем изложении будут важны, так что мы оформим их в виде определений.

Определение 1.4. ***Несовместные события.***

События A *и* B *называются* ***несовместными****, если* $AB = \varnothing$.

События из набора $(A_k)_{k=1}^{N}$*,* $N \leqslant \infty$*, называются* ***несовместными****, если любые два события из данного набора являются несовместными.*

Кроме того, иногда, чтобы подчеркнуть, что мы рассматриваем вероятность суммы несовместных событий, мы вместо знака $\cup$ будем использовать символ $\sqcup$ и обозначать термином *дизъюнктное объединение.*

1.1.1 Вероятностное пространство в широком смысле

Для того, чтобы задать *вероятностное пространство*, нам нужно задать *сигма-алгебру* событий, что далеко не всегда является тривиальной задачей, если решать её напрямую. В данном подпункте мы введем *вероятностное пространство в широком смысле*, которое можно расширить до обычного *вероятностного пространства.*

Определение 1.5. ***Вероятностное пространство в широком смысле.***

Тройка $(\Omega, \mathbf{F}, \mathrm{P})$, где $\mathbf{F}$ – алгебра событий, и $\mathrm{P} : \mathbf{F} \to \mathbb{R}$ – функция, удовлетворяющая следующим аксиомам:

- $\forall A \in \mathbf{F} \Rightarrow \mathrm{P}(A) \geqslant 0$,
- $\mathrm{P}(\Omega) = 1$,
- $\forall (A_j)_{j=1}^{\infty} : A_j \in \mathbf{F}, \bigsqcup\limits_{j=1}^{\infty} A_j \in \mathbf{F} \Rightarrow \ \mathrm{P}\left(\bigsqcup\limits_{j=1}^{\infty} A_j\right) = \sum\limits_{j=1}^{\infty} \mathrm{P}(A_j)$,

называется ***вероятностным пространством в широком смысле****.*

Определение *вероятностного пространства в широком смысле* очень похоже на определение *вероятностного пространства*, но здесь мы определяем *вероятность* на *алгебре* событий и требуем, чтобы *дизъюнктное объединение* событий принадлежало *алгебре*, в то время как определение *сигма-алгебры* требует, чтобы объединение принадлежало *сигма-алгебре.*

Далее, пусть F – набор событий. Рассмотрим все возможные *сигма-алгебры* событий $\mathcal{F}$, которые содержат все события из F (такие *сигма-алгебры* существуют, например, множество всех подмножеств Ω), и возьмём их пересечение. Оно тоже будет являться *сигма-алгеброй* событий, которую мы будем называть *минимальной сигма-алгеброй*, содержащей F, и обозначать как $\sigma(F)$. Произвольная *алгебра* или *сигма-алгебра*, содержащая множества из F, называется *порожденной* набором F (простейший пример – замыкание относительно дополнения, пересечения и счётного объединения). Более подробно познакомиться с *минимальными сигма-алгебрами*, в частности с условиями минимальности, можно в книге Ширяева [11].

Введем *борелевскую сигма-алгебру*, которая будет играть большую роль в дальнейшем изложении.

Определение 1.6. ***Борелевская сигма-алгебра.***

Пусть $\Omega = E \subseteq \mathbb{R}$, F – все возможные открытые интервалы вида (a, b) (может быть, с бесконечными концами), содержащиеся в E. Тогда $\mathcal{B} = \sigma(F)$ называется ***борелевской сигма-алгеброй****.*

Замечание. *Борелевская сигма-алгебра $\mathcal{B}^n$ в n-мерном пространстве определяется аналогично. В данном случае в качестве F можно взять открытые параллелепипеды.*

Замечание. *Замкнутые интервалы принадлежат борелевской сигма-алгебре, так как они являются дополнениями к открытым. Кроме того,*

$$\{a\} = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, a + \frac{1}{n}\right), \quad (a, b] = \bigcap_{n=1}^{\infty} \left(a, b + \frac{1}{n}\right),$$

так что одноточечные множества и полуоткрытые интервалы тоже принадлежат борелевской сигма-алгебре.

Замечание. *Борелевскую сигма-алгебру можно точно так же на произвольном топологическом пространствах.*

В заключение данного пункта сформулируем теорему о продолжении меры (Каратеодори), которая показывает, что достаточно задать *вероятностное пространство в широком смысле*, а соответствующее ему *вероятностное пространство* получается с помощью продолжения меры. В частности, для того, чтобы задать вероятность на $\mathbb{R}$, достаточно определить *вероятность* на открытых интервалах.

Теорема 1.1. ***Теорема Каратеодори.***

Пусть $(\Omega, \mathbf{F}, \mathcal{Q})$ *– вероятностное пространство в широком смысле. Тогда существует единственная вероятностная мера* $\mathrm{P} : \sigma(F) \to \mathbb{R}$ *такая, что для любых* $A \in \mathbf{F}$

$$\mathrm{P}(A) = \mathcal{Q}(A).$$

Тем самым вероятностное пространство в широком смысле $(\Omega, \mathbf{F}, \mathcal{Q})$ *автоматически определяет вероятностное пространство* $(\Omega, \sigma(\mathbf{F}), \mathrm{P})$.

Доказательство теоремы мы приводить здесь не будем, желающие ознакомиться с ним могут обратиться к [3], [6] или [10].

1.1.2 Об области определения вероятности

В[1] предыдущем пункте мы затронули *минимальные сигма-алгебры*, однако не проще ли задать *вероятность* на всех подмножествах Ω естественным образом. В общем случае ответ на этот вопрос отрицательный. И данный пункт посвящен обоснованию этого утверждения.

Покажем, что невозможно задать меру μ на всех ограниченных множествах, которая удовлетворяет следующим условиям:

- $\mu[0, 1] = 1$.
- Если $B = A + b = \{a + b : a \in A\}$ и $C = -A = \{-a : a \in A\}$, то $\mu B = \mu A = \mu C$.
- Мера μ счётно-аддитивна.

Пусть $A \subset [-1/2, 1/2]$ – неизмеримое по Лебегу множество (процесс его построения можно посмотреть в [8]), $(r_k)_{k\in\mathbb{N}}$ – последовательность рациональных чисел на интервале $[-1; 1]$ (рациональные числа счётны, то есть их можно занумеровать). Далее положим $A_0 = A$, $A_k = A + r_k$, $k \in \mathbb{N}$. Тогда имеем

$$\left[-\frac{1}{2}; \frac{1}{2}\right] \subset \bigsqcup_{k=0}^{\infty} A_k \subset \left[-\frac{3}{2}; \frac{3}{2}\right].$$

Если мера μ с указанными свойствами существует, то $\mu[-1/2, 1/2] = 1$, $\mu A_k = \Delta$, $k \in \{0, 1, \ldots\}$, $\mu[-3/2, 3/2] = m < \infty$. Тогда мы получим, что

$$1 < \Delta + \Delta + \ldots + \Delta + \ldots < m,$$

что невозможно для любых неотрицательных Δ.

Таким образом, нельзя при $\Omega = [0; 1]$ вероятность определить как длину промежутка в привычном нам смысле и при этом задать её на всех $A \subset [0; 1]$.

[1] данный пункт предназначен для любознательных, при первом прочтении его можно упустить

1.2 Примеры вероятностных пространств

В данном параграфе мы рассмотрим примеры вероятностных пространств.

1.2.1 Классическая вероятность

Прежде всего рассмотрим модель классической вероятности, с которой читатель знаком ещё со школьных времен. Пусть Ω – конечный набор из N элементов $\omega_1, \ldots, \omega_N$, которые мы будем трактовать как элементарные исходы. В качестве сигма-алгебры событий возьмем 2^Ω – множество всех подмножеств Ω, которое очевидно является сигма-алгеброй. Положим $\mathrm{P}(\omega_i) = 1/N$, $i \in \{1, \ldots, N\}$. Тогда вероятность произвольного события $A = \{\omega_{i_1}, \ldots, \omega_{i_M}\}$ вычисляется по формуле

$$\mathrm{P}(A) = \mathrm{P}(\{\omega_{i_1}, \ldots, \omega_{i_M}\}) = \sum_{j=1}^{M} \mathrm{P}(\omega_{i_j}) = M\,\mathrm{P}(\omega_{i_1}) = \frac{M}{N} = \frac{|A|}{|\Omega|},$$

которая известна как отношение числа благоприятных исходов к количеству всех исходов. Числа M, N вычисляются, как правило, комбинаторными способами, которые мы рассмотрим далее.

Некоторые сведения из комбинаторики

Комбинаторика – раздел математики, посвящённый задачам, связанным с выбором, расположением, перебором элементов, как правило, конечного множества. Чаще всего нам нужно будет находить число элементов того или иного множества. Обычно это удается сделать с помощью двух основных принципов комбинаторики – правил суммы и пересечения.

Часто всего интересующее нас множество удаётся разбить на несколько взаимно непересекающихся подмножеств. Тогда общее число элементов искомого множества есть сумма элементов каждого из подмножеств. Этот принцип и называется *правилом суммы.* Можно его сформулировать несколько иначе: если некоторый объект A можно выбрать m способами, объект B – n способами, то выбрать "либо A, либо B" можно $m + n$ способами. Здесь нужно внимательно следить за тем, чтобы данные классы не пересекались. Если разбить искомое множество не получается, тогда стоит применить принцип включений-исключений, который будет разобран ниже.

Часто при составлении комбинации из двух элементов известно, что первый можно выбрать m способами, второй – n способами, причём число независимо от выбранного первого элемента, а число возможных вариантов для второго неизменно. Тогда пару элементов можно составить mn способами. Этот принцип и есть *правило произведения.* Другими словами: если объект A можно выбрать m способами и если после каждого такого выбора объект B можно выбрать n способами, то пару (A, B) можно выбрать mn способами.

Далее введем фундаментальные для комбинаторики понятия: *перестановку*, *размещение* и *сочетание.*

Определение 1.7. ***Перестановка (без повторений).***
Перестановкой (без повторений) *называется упорядоченный набор уникальных элементов множества* $\{1, \ldots, n\}$ *длины* n

Общее число *перестановок* n элементного множества равняется

$$P_n = n \cdot (n-1) \cdot \ldots 2 \cdot 1 = n!,$$

что можно вычислить, например, при помощи принципа произведения.

Определение 1.8. ***Размещение (без повторений).***

Размещением (без повторений) *из n по k, $k \leqslant n$, называется упорядоченный набор из k различных элементов множества $\{1, \ldots, n\}$.*

Общее число *размещений* из n по k равняется

$$A_n^k = n \cdot (n-1) \cdot \ldots \cdot (n-k+1) = \frac{n!}{(n-k)!},$$

что также легко выводится с помощью принципа произведения.

Определение 1.9. ***Сочетание (без повторений).***

Сочетанием (без повторений) *из n по k, $k \leqslant n$, называется неупорядоченный набор из k различных элементов множества $\{1, \ldots, n\}$.*

Общее число *сочетаний* из n по k равняется

$$C_n^k = \frac{n!}{(n-k)!k!}, \tag{1.1}$$

которое в англоязычной литературе традиционно обозначается как $\binom{n}{k}$. Коэффициент C_n^k называется *биномиальным.* Для доказательства формулы (1.1) достаточно заметить, что для подсчета общего числа *сочетаний (без повторений)* из n по k нужно общее число *размещений (без повторений)* A_n^k разделить на число *перестановок (без повторений)* длины k, которое равняется $k!$. Отметим, что наборы $(1, 2)$ и $(2, 1)$ – разные размещения, но с точки зрения сочетаний они представляют один и тот же объект.

Далее рассмотрим *перестановки*, *размещения* и *сочетания с повторениями.*

Определение 1.10. ***Перестановка (с повторениями).***

Перестановкой (с повторениями) *длины $n = n_1 + \ldots + n_m$ множества $\{1, \ldots, m\}$ называется упорядоченный набор, в котором единица встречается n_1 раз, двойка – n_2 раз, $\ldots$, число m – n_m раз.*

Общее число *перестановок (с повторениями)* длины $n = n_1 + \ldots + n_m$ множества $\{1, \ldots, m\}$ равняется

$$P(n, n_1, \ldots, n_m) = \frac{n!}{n_1! \cdot \ldots \cdot n_m!}, \tag{1.2}$$

который называется *мультиномиальным* (или *полиномиальным*) *коэффициентом.* Формулу (1.2) можно вывести и другим способом. Ввиду важности рассуждений при выводе оформим её виде теоремы.

Теорема 1.2. ***Вычисление мультиномиального коэффициента.***

$$P(n, n_1, \ldots, n_m) = C_n^{n_1} C_{n-n_1}^{n_2} \cdot \ldots \cdot C_{n_m}^{n_m}.$$

Доказательство. Нам нужно посчитать общее число упорядоченных наборов длины $n = n_1 + \ldots + n_m$, в котором единица встречается n_1 раз, двойка – n_2 раз, $\ldots$, число m – n_m раз. Сначала выберем позиции в наборе, на которых будут располагаться единицы. Это можно сделать $C_n^{n_1}$ способами. Далее нужно выбрать n_2 позиций для двоек среди оставшихся $n - n_1$ позиций. Это осуществляется $C_{n-n_1}^{n_2}$ способами. В конце нам останется среди оставшихся $n - n_1 - \ldots n_{m-1} = n_m$ позиций выбрать n_m позиций, что можно сделать

$C_{n_m}^{n_m}$ (одним) способом. Искомая формула получается с помощью принципа произведения. Также заметим, что

$$C_n^{n_1} C_{n-n_1}^{n_2} \cdot \ldots \cdot C_{n_m}^{n_m} = \frac{n!(n-n_1)! \cdot \ldots \cdot n_m!}{n_1!(n-n_1)!n_2!(n-n_1-n_2)! \cdot \ldots \cdot n_{m-1}!n_m!n_m!0!}$$
$$= \frac{n!}{n_1! \cdot \ldots \cdot n_m!}.$$

ч.т.д.

Определение 1.11. ***Размещение (с повторениями).***

Размещением (с повторением) *из n по k называется упорядоченный набор из k (необязательно одинаковых) элементов множества $\{1, \ldots, n\}$.*

Ввиду правила произведения очевидно, что общее число *размещений (с повторениями)* равняется n^k.

Определение 1.12. ***Сочетание (с повторением).***

Сочетанием (с повторением) *из n по k называется неупорядоченный набор из k (необязательно одинаковых) элементов множества $\{1, \ldots, n\}$.*

Так вывод формулы подсчёта числа *сочетаний (с повторениями)* также ценен, то оформим это в виде теоремы.

Теорема 1.3. ***Подсчет количества сочетаний с повторениями.***

Общее число сочетаний (с повторениями) из n по k равняется C_{n+k-1}^k (очевидно это же можно записать как C_{n+k-1}^{n-1}).

Доказательство. Рассмотрим уравнение $x_1 + \ldots + x_n = k$ относительно целых неотрицательных переменных $x_1, \ldots x_n$, при этом x_j мы интерпретируем как количество вхождений числа j в конкретный набор. Общее число решений данного уравнения и есть число сочетаний (с повторениями) из n по k. Каждому решению данного уравнения сопоставим последовательность из k шаров и $n-1$ перегородки. Также слева поставим перегородку, которая будет иметь номер ноль, а справа – с номером n. Тогда если между перегородками с номерами j и $j+1$, $j \in \{0, \ldots, n-1\}$, расположены i шаров (перегородки могут стоять рядом друг с другом), то $x_{j+1} = i$. Всего таких расстановок можно осуществить C_{n+k-1}^{m-1} способами. *ч.т.д.*

Далее, если речь идёт о комбинаторных объектах без повторений, мы будем просто писать перестановка, размещение или размещения; комбинаторные объекты с повторениями мы будем писать как перестановка с повторениями, размещение с повторениями или сочетание с повторениями.

Рассмотрим один простой, существенный пример.

Пример 1.1. ***Кости и домино***

Представим, что игра представляет из себя подбрасывание двух кубиков. В качестве элементарного исхода рассмотрим упорядоченную пару (i, j), $i, j \in \{1, 2, 3, 4, 5, 6\}$. Очевидно, что таких исходов 36, откуда вероятность элементарного исхода равняется 1/36.

Но давайте в качестве элементарного исхода возьмем неупорядоченную пару $\{i, j\}$, $i, j \in \{1, 2, 3, 4, 5, 6\}$. Количество таких исходов равняется количеству сочетаний с повторениями из 6 по 2, что есть $C_{6+2-1}^2 = 21$. Получается, нам следует определить вероятность исхода как 1/21? Нет, ведь данные исходы ***не являются равновероятными!***

Как понять, что здесь стоит выбрать именно первую модель? Здесь мы можем представить, что последовательно бросаем два кубика, один за другим, то есть считать исходы $(1,2)$ *и* $(2,1)$ *разными. Когда мы наш исход можем представить как последовательность атомарных событий, предпочтительнее выбрать модель, учитывающую порядок.*

Но когда следует использовать сочетания с повторениями. Например, в домино. Общее количество обычных доминошек равняется количеству сочетаний из 7 по 2, что есть $C^2_{7+2-1} = 28$.

Желающие более детально разобраться в комбинаторике могут обратиться к книге [4], содержащей множество примеров.

Далее в качестве примера рассмотрим задачу, которая известна как *парадокс дней рождения.*

Пример 1.2. ***Парадокс дней рождения.***

Покажем, что в группе, состоящей из 23 или более человек, вероятность совпадения дней рождения хотя бы у двух людей больше 50%, что может не согласовываться с нашей интуицией (поэтому и употребляется слово "парадокс"). Так, вероятность совпадения дней рождения у двух людей равна $1/365 \approx 0,27\%$, *а при умножении этого числа на 23 получится приблизительно 6,7% (формально это рассуждение не является корректным).*

Пусть P_n *– искомая вероятность для группы из n человек. Очевидно, что при* $n > 365$ *(мы предполагаем, что в году 365 дней)* $P_n = 1$. *Для вычисления искомой вероятности найдём вероятность противоположного события (в группе нет людей с совпадающими днями рождения). Обозначим её* $\overline{P}_n$. *В числителе дроби у нас будет число размещений из 365 по n, а в знаменателе – количество размещений с повторениями из 365 по n, то есть*

$$\overline{P}_n = \frac{365!}{(365-n)!365^n} = \frac{365 \cdot 364 \cdot \ldots \cdot (365-n+1)}{365^n}$$
$$= 1 \cdot \left(1 - \frac{1}{365}\right) \cdot \ldots \left(1 - \frac{n-1}{365}\right), \quad n \in \{0, 1, \ldots, 365\}.$$

Для нахождения искомой вероятности от единицы нужно отнять вероятность противоположного события (хоть и простое, но важное соображение, ниже будет сформулировано)

$$P_n = 1 - \overline{P}_n = 1 - 1 \cdot \left(1 - \frac{1}{365}\right) \cdot \ldots \left(1 - \frac{n-1}{365}\right), \quad n \in \{0, 1, \ldots, 365\}.$$

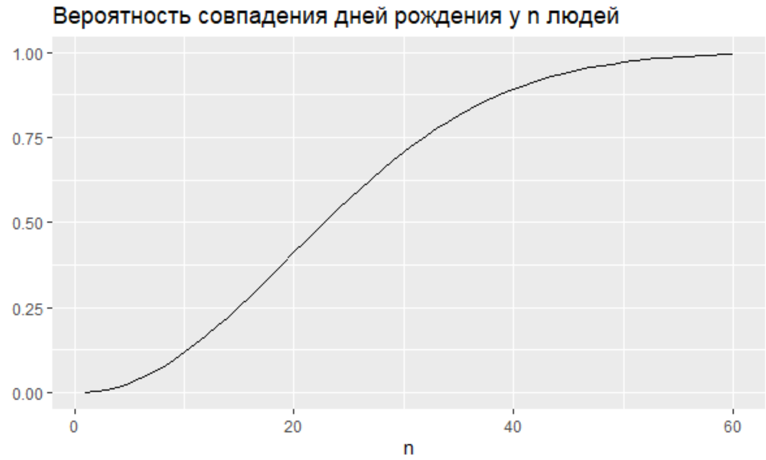
Приведем таблицу, в которой численно вычислены P_n *для различных n:*

n	2	5	10	20	22	23	30	50	100	366
P_n,%	0,27	3	12	41	48	51	71	97	99,999	1

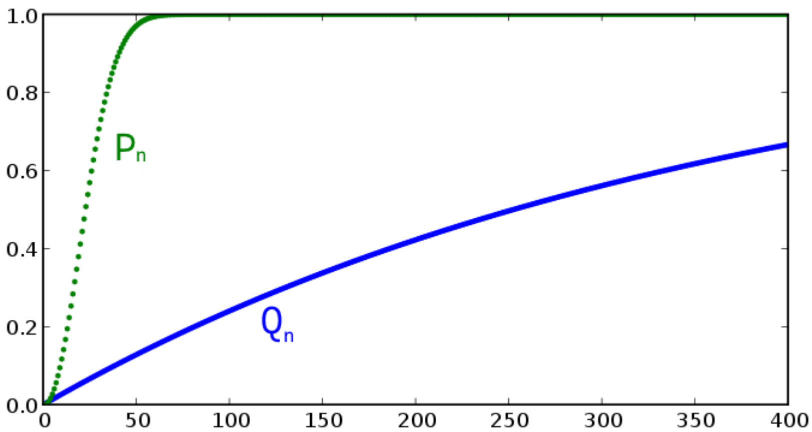
Также график, иллюстрирующий зависимость, имеет вид

Давайте теперь рассмотрим вероятность того, что в группе из n человек день рождения конкретного человека совпадет с днем рождения какого-то другого человека в группе, и сравним ее с P_n.

$$Q_n = 1 - P_n = 1 - \left(\frac{365-1}{365}\right)^n, \quad n \geqslant 0.$$



Сравнивая с P_n, скажем, при $n = 23$ мы обнаружим, что $Q_{23} \approx 0.0612$ (что всего лишь равно 6.12% и уже ближе согласуется с нашей интуицией), а при $n = 366$ мы увидим $Q_{366} \approx 0.6337$ - и это неудивительно, ведь у остальных членов группы дни рождения также могут совпадать между собой. График сравнения P_n и Q_n представлен ниже.



1.2.2 Геометрическая вероятность

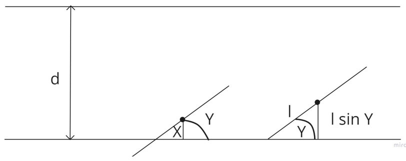
Пусть $\Omega \subset \mathbb{R}^n$ – множество с конечной мерой μ (например, мера Лебега), $\mathcal{F}$ – измеримые подмножества Ω. Тогда вероятность измеримого множества A определим как

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

которую мы будем именовать *геометрической вероятностью*. Другими словами, мы задали *равномерное распределение* на Ω.

Рассмотрим пример, известный как *задача Бюффона о бросании иглы.*

Пример 1.3. ***Задача об игле Бюффона*** *Имеются разлинованный стол, у которого расстояние между линиями равняется d, и игла длины $2l < d$. Найти вероятность пересечения иглой какой-нибудь линии после случайного броска.*



Пусть X – расстояние от центра иглы до ближайшей линии стола, Y – угол между линией и иглой, причем он отсчитывается от линии до иглы против часовой стрелки. Эти величины изменяются в пределах $[0; d/2]$, $[0, \pi]$ соответственно. Таким образом, возможному исходу сопоставляется случайная точка (X, Y) в прямоугольнике $S = [0; d/2] \times [0, \pi]$. Пересечению линии иглой соответствует множество

$$A = \{(x, y) : x \leqslant l \sin y\}.$$

Искомую вероятность найдём как отношение площадей

$$p = \frac{\mu A}{\mu S} = \frac{2\int\limits_0^{\pi} l \sin y \; dy}{\pi d} = \frac{4l}{\pi d}.$$

Пока на этом остановимся, но ниже мы к этому примеру ещё вернёмся.

1.3 Свойства вероятности

В данном параграфе мы сформулируем и докажем свойства *вероятности*, которые будут полезны при решении задач. Так, в *задаче о парадоксах дней рождения* мы уже вычислили искомую величину через вероятность противоположного события.

Простейшие свойства сформулируем в виде следующей теоремы.

Теорема 1.4. ***Простейшие свойства вероятности.***
Справедливы следующие соотношения:

1. $A \subseteq B \Rightarrow \mathrm{P}(A) \leqslant \mathrm{P}(B)$.
2. $\mathrm{P}(A) \leqslant 1$ *для всех событий* A.
3. $\mathrm{P}(A) = 1 - \mathrm{P}(\overline{A})$ *для всех событий* A.
4. $\mathrm{P}(\varnothing) = 0$.
5. $\mathrm{P}(A + B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(AB)$ *для всех событий* A, B.
6. $\mathrm{P}(A + B) \leqslant \mathrm{P}(A) + \mathrm{P}(B)$ *для всех событий* A, B.
7. $\mathrm{P}\left(\bigcup\limits_{k=1}^{N} A_k\right) \leqslant \sum\limits_{k=1}^{N} \mathrm{P}(A_k)$ *для произвольного набора событий* $(A_k)_{k=1}^{N}$, $2 \leqslant N \leqslant \infty$. *Известно как неравенство Буля.*

Доказательство.

1. $A \subseteq B$, поэтому $B = A + (B \setminus A)$. Имеем $\mathrm{P}(B) = \mathrm{P}(A) + \mathrm{P}(B \setminus A) \geqslant \mathrm{P}(A)$.
2. Следует из предыдущего, так так $A \subseteq \Omega$ и $\mathrm{P}(\Omega) = 1$.

3. Следует из представления $\Omega = A + \overline{A}$.

4. Следует из предыдущего.

5. Следует из представлений $A + B = A + (B \setminus AB)$, $B = (B \setminus AB) + AB$.

6. Следует из предыдущего и неотрицательности вероятности.

7. Следует из первого свойства, счётной аддитивности вероятности и представления

$$\bigcup_{k=1}^{N} A_k = \bigcup_{k=1}^{N} A_k B_k, \quad \text{где } B_1 = A_1, B_k = \Omega \setminus \left(\bigcup_{i=1}^{k-1} A_i \right), k \geqslant 2.$$

ч.т.д.

Далее мы сформулируем и докажем свойства, которые равносильны счётной аддитивности *вероятности.*

Теорема 1.5. ***О непрерывности вероятности сверху и снизу.***

Следующие свойства равносильный счётной аддитивности вероятности:

- *Вероятность конечно-аддитивна и для набора событий $(B_n)_{n=1}^{\infty}$ таких, что $B_{n+1} \subset B_n$ и $B = \bigcap\limits_{n=1}^{\infty} B_n$, справедливо $\mathrm{P}(B_n) \to \mathrm{P}(B)$.*
- *Вероятность конечно-аддитивна и для набора событий $(A_n)_{n=1}^{\infty}$ таких, что $A_n \subset A_{n+1}$ и $A = \bigcup\limits_{n=1}^{\infty} A_n$, справедливо $\mathrm{P}(A_n) \to \mathrm{P}(A)$.*

Вторые части первого и второго утверждений мы будем называть непрерывностью сверху и снизу соответственно.

Доказательство. Заметим, что утверждения в формулировке равносильны вследствие законов де Моргана. В самом деле, пусть справедлива непрерывность сверху, то есть для набора событий $(B_n)_{n=1}^{\infty}$ таких, что $B_{n+1} \subset B_n$ и $B = \bigcap\limits_{n=1}^{\infty} B_n$, верно $\mathrm{P}(B_n) \to \mathrm{P}(B)$. Положим $A_n = \overline{B}_n$, $A = \overline{B}$. Тогда $A_n \subset A_{n+1}$ и

$$A = \overline{\bigcap_{n=1}^{\infty} B_n} = \bigcup_{n=1}^{\infty} \overline{B}_n = \bigcup_{n=1}^{\infty} A_n,$$
$$\mathrm{P}(A_n) = \mathrm{P}(\overline{B}_n) = 1 - \mathrm{P}(B_n) \to 1 - \mathrm{P}(B) = \mathrm{P}(\overline{B}) = \mathrm{P}(A).$$

Аналогичным образом доказывается импликация в другую сторону.

Далее покажем равносильность счётной аддитивности и конечной аддитивности вместе с непрерывностью сверху.

Пусть вероятность счётно-аддитивна. Предположим, что нам дана последовательность событий $(B_n)_{n=1}^{\infty}$, для которых $B_{n+1} \subset B_n$ и $B = \bigcap\limits_{n=1}^{\infty} B_n$. Положим $C_k = B_k \overline{B}_{k+1}$, $k \in \mathbb{N}$. Тогда события $B, C_1, C_2, \ldots$ несовместны, $B_n = B + \bigcup\limits_{k=n}^{\infty} C_k$ и мы можем заключить, что

$$\mathrm{P}(B_n) = \mathrm{P}(B) + \sum_{k=n}^{\infty} \mathrm{P}(C_k).$$

В правой части стоит сходящийся ряд, поэтому, перейдя к пределу, имеем $\mathrm{P}(B_n) \to \mathrm{P}(B)$.

Предположим обратное: конечную аддитивность и непрерывность сверху. Пусть $\{A_n\}_{n=1}^{\infty}$ – набор несовместных событий. Тогда

$$\mathrm{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \mathrm{P}\left(\bigcup_{k=1}^{n} A_k\right) + \mathrm{P}\left(\bigcup_{k=n+1}^{\infty} A_k\right).$$

Кроме того, имеет место

$$\sum_{k=1}^{\infty} \mathrm{P}(A_k) = \lim_{n\to\infty} \sum_{k=1}^{n} \mathrm{P}(A_k) = \lim_{n\to\infty} \mathrm{P}\left(\bigcup_{k=1}^{n} A_k\right)$$
$$= \mathrm{P}\left(\bigcup_{k=1}^{\infty} A_k\right) - \lim_{n\to\infty} \mathrm{P}\left(\bigcup_{k=n+1}^{\infty} A_k\right) = \mathrm{P}\left(\bigcup_{k=1}^{\infty} A_k\right).$$

В последнем равенстве мы как раз и воспользовались непрерывностью сверху. *ч.т.д.*

Замечание. *В формулировке и доказательстве мы работали с сигма-алгеброй событий, однако можно модифицировать рассуждения и для алгебры событий: только везде нужно отдельно требовать, чтобы счётное объединение и пересечение принадлежало алгебре.*

Далее мы сформулируем и докажем формулу для вычисления вероятности суммы событий.

Теорема 1.6. ***Формула включений-исключений.***

$$\mathrm{P}(A_1 + \ldots + A_n) = \sum_{k=1}^{n} \mathrm{P}(A_k) - \sum_{1\leqslant k_1<k_2\leqslant n} \mathrm{P}(A_{k_1}A_{k_2}) + \ldots + (-1)^{n+1}\,\mathrm{P}(A_1 \ldots A_n)$$

Доказательство. Будем доказывать индукцией по n. База индукции уже сформулирована и доказана ранее. Предположим, что формула верна при $n-1$ и докажем справедливость формулы для n.

$$\mathrm{P}(A_1 + \ldots + A_{n-1} + A_n) = \mathrm{P}(A_1 + \ldots + A_{n-1}) + \mathrm{P}(A_n) - \mathrm{P}\big((A_1 + \ldots + A_{n-1})A_n\big)$$
$$= \mathrm{P}(A_1 + \ldots + A_{n-1}) + \mathrm{P}(A_n) - \mathrm{P}(A_1A_n + \ldots + A_{n-1}A_n).$$

Применим индукционное предположение к $\mathrm{P}(A_1 + \ldots + A_{n-1})$:

$$\mathrm{P}(A_1 + \ldots + A_{n-1}) = \sum_{k=1}^{n-1} \mathrm{P}(A_k) - \sum_{1\leqslant k_1<k_2\leqslant n-1} \mathrm{P}(A_{k_1}A_{k_2}) + \ldots + (-1)^{n}\,\mathrm{P}(A_1 \ldots A_{n-1}),$$

аналогично к $\mathrm{P}(A_1A_n + \ldots + A_{n-1}A_n)$:

$$\mathrm{P}(A_1A_n + \ldots + A_{n-1}A_n) = \sum_{k=1}^{n-1} \mathrm{P}(A_kA_n) - \sum_{1\leqslant k_1<k_2\leqslant n-1} \mathrm{P}(A_{k_1}A_{k_2}A_n) + (-1)^{n}\,\mathrm{P}(A_1 \cdot \ldots \cdot A_n).$$

Для завершения доказательства осталось подставить полученные выражения для $\mathrm{P}(A_1 + \ldots + A_{n-1})$, $\mathrm{P}(A_1A_n + \ldots + A_{n-1}A_n)$ и сложить их. *ч.т.д.*

Приведем один классический пример, в котором применяется принцип *включений-исключений.*

Пример 1.4. ***Задача о конвертах.***

Имеются n писем и n конвертов. Какова вероятность того, что хотя бы одно письмо попадет в свой конверт?

Пусть A_i, $i \in \{1, \ldots, n\}$, – событие, заключающееся в том, что i-е письмо попало в свой конверт. Тогда $A = A_1 + \ldots + A_n$ – искомое событие и согласно формуле включений-исключений имеем

$$\begin{aligned} \mathrm{P}(A) = \sum_{k=1}^{n} \mathrm{P}(A_k) - \sum_{1 \leqslant k_1 < k_2 \leqslant n} \mathrm{P}(A_{k_1} A_{k_2}) \\ + \sum_{1 \leqslant k_1 < k_2 < k_3 \leqslant n} \mathrm{P}(A_{k_1} A_{k_2} A_{k_3}) + \ldots + (-1)^{n+1} \mathrm{P}(A_1 \ldots A_n) \\ = C_n^1 \frac{1}{n} - C_n^2 \frac{1}{n(n-1)} + C_n^3 \frac{1}{n(n-1)(n-2)} + \ldots + (-1)^{n+1} C_n^n \frac{1}{n!} \\ = 1 - \frac{1}{2!} + \frac{1}{3!} + \ldots + (-1)^{n+1} \frac{1}{n!}. \end{aligned}$$

Несложно заметить, что при больших n данная вероятность приближенно равняется $1 - e^{-1}$.

1.4 Условная вероятность. Независимые события

Рассмотрим пример, подводящий к *условной вероятности* – одному из фундаментальных понятий теории вероятностей. Пусть имеется классическая модель, в которой всего N элементарных исходов, и A, B – события, состоящее из k, l элементарных событий соответственно, при этом событие AB состоит из m элементарных исходов. Найдём вероятность события A, если известно, что B точно произошло. Обозначим эту вероятность $\mathrm{P}(A|B)$. Тогда имеем

$$\mathrm{P}(A|B) = \frac{m}{l} = \frac{m/N}{l/N} = \frac{\mathrm{P}(AB)}{\mathrm{P}(B)}. \tag{1.3}$$

Полученное соотношение можно обобщить на произвольные вероятностные модели и сформулировать определение *условной вероятности.*

Определение 1.13. *Пусть $(\Omega, \mathcal{F}, \mathrm{P})$ – вероятностное пространство и $B \in \mathcal{F}$ – событие с ненулевой вероятностью. Тогда выразим* ***условную вероятность*** *при событии B (вероятность при условии B), которую будем обозначать как $\mathrm{P}(.|B)$, согласно соотношению (1.3). Иногда будем использовать обозначение P_B.*

Замечание. *Условная вероятность P_B удовлетворяет аксиомам, что проверяется непосредственным образом, и тем самым тройка $(\Omega, \mathcal{F}, \mathrm{P}_B)$ тоже является вероятностным пространством.*

Сформулируем и докажем свойство, которое называется *формулой произведения* и будет полезна при решении задач.

Теорема 1.7. ***Формула произведения вероятностей.***

Пусть $\mathrm{P}(A_1 \ldots A_n) > 0$. Тогда справедлива формула

$$\mathrm{P}(A_1 \ldots A_n) = \mathrm{P}(A_1)\, \mathrm{P}(A_2|A_1)\, \mathrm{P}(A_3|A_1 A_2) \cdot \ldots \cdot \mathrm{P}(A_n|A_1 \ldots A_{n-1}).$$

Доказательство. Для доказательства достаточно последовательно воспользоваться определением условной вероятности:

$$\begin{aligned} \mathrm{P}(A_1 \dots A_n) &= \mathrm{P}(A_1 \dots A_{n-1})\,\mathrm{P}(A_n|A_1 \dots A_{n-1}) \\ &= \mathrm{P}(A_1 \dots A_{n-2})\,\mathrm{P}(A_{n-1}|A_1 \dots A_{n-2})\,\mathrm{P}(A_n|A_1 \dots A_{n-1}) \\ &= \dots = \mathrm{P}(A_1)\,\mathrm{P}(A_2|A_1)\,\mathrm{P}(A_3|A_1A_2) \cdot \ldots \cdot \mathrm{P}(A_n|A_1 \dots A_{n-1}) \end{aligned}$$

ч.т.д.

Далее рассмотрим другое фундаментальное понятие теории вероятностей – *независимости.*

Определение 1.14. *События A_1, A_2 называются* ***независимыми****, если*

$$\mathrm{P}(A_1A_2) = \mathrm{P}(A_1)\,\mathrm{P}(A_2).$$

События $A_1, \dots, A_n$ называются ***независимыми*** *(независимыми в совокупности), если для любого поднабора $A_{j_1}, \dots, A_{j_m}$*

$$\mathrm{P}(A_{j_1} \dots A_{j_m}) = \mathrm{P}(A_{j_1}) \cdot \ldots \cdot \mathrm{P}(A_{j_m}).$$

Замечание. *Если $\mathrm{P}(A_2) > 0$, то независимость событий A_1, A_2 равносильна соотношению*

$$\mathrm{P}(A_1|A_2) = \mathrm{P}(A_1),$$

что может служить альтернативным определением независимости.

В определении для *независимости* набора событий мы требовали, чтобы вероятность произведения событий из любого поднабора равнялась произведению вероятностей, но может *независимости* любой пары событий (*попарной независимости*) достаточно? Как мы увидим из следующего примера – нет.

Пример 1.5. ***Контрпример с раскрашенным тетраэдром.***

Пусть имеется правильный тетраэдр, при броске которого вероятность выпадения каждой грани одинакова. Три грани раскрашены в красный, жёлтый и зелёный цвета соответственно, а четвёртая разделена на три части, каждая из которых покрашена в эти цвета.

Формально опишем эту задачу: $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, где ω_k соответствует выпадению k-ой грани, и $\mathrm{P}(\omega_k) = 1/4$, $k \in \{1, 2, 3, 4\}$.

Пусть события $R = \{\omega_1, \omega_4\}$, $Y = \{\omega_2, \omega_4\}$, $G = \{\omega_3, \omega_4\}$ означают, что на грани появился красный, жёлтый или зелёный цвет соответственно. Несложно заметить, что

$$\mathrm{P}(R) = \mathrm{P}(Y) = \mathrm{P}(G) = \frac{1}{2}.$$

Вероятность события RY равняется

$$\mathrm{P}(RY) = \frac{1}{4} = \mathrm{P}(R)\,\mathrm{P}(Y).$$

Аналогично можно расписать и для RG, YG, поэтому события RY, RG, YG попарно независимы, но

$$\mathrm{P}(RYG) = \frac{1}{4} \neq \mathrm{P}(R)\,\mathrm{P}(Y)\,\mathrm{P}(G),$$

то есть события R, Y, G не являются независимыми.

1.4.1 Формула полной вероятности. Теорема Байеса

Далее рассмотрим два простых, но играющих существенную роль в теории вероятностей и приложениях, утверждения – *формулу полной вероятности* и *теорему Байеса.*

Теорема 1.8. ***Формула полной вероятности.***
Пусть $A \subset \bigsqcup B_k$ *и* $\mathrm{P}(B_k) > 0$. *Тогда справедлива формула полной вероятности*

$$\mathrm{P}(A) = \sum \mathrm{P}(A|B_k)\,\mathrm{P}(B_k).$$

Доказательство.

$$\mathrm{P}(A) = \mathrm{P}\left(A \cap \bigsqcup B_k\right) = \sum \mathrm{P}(AB_k) = \sum \mathrm{P}(A|B_k)\,\mathrm{P}(B_k).$$

ч.т.д.

Как правило, $\Omega = \bigsqcup B_k$, поэтому набор событий $\{B_k\}$ называют *полной группой несовместных событий.* Иногда используют термин *гипотезы.*

Теорема 1.9. ***Формула Байеса.***
Пусть $A \subset \bigsqcup B_k$ *и* $\mathrm{P}(B_k) > 0$. *Тогда справедлива формула Байеса*

$$\mathrm{P}(B_k|A) = \frac{\mathrm{P}(A|B_k)\,\mathrm{P}(B_k)}{\sum\limits_j \mathrm{P}(A|B_j)\,\mathrm{P}(B_j)}.$$

Доказательство.

$$\mathrm{P}(B_k|A) = \frac{\mathrm{P}(AB_k)}{\mathrm{P}(A)} = \frac{\mathrm{P}(A|B_k)\,\mathrm{P}(B_k)}{\sum\limits_j \mathrm{P}(A|B_j)\,\mathrm{P}(B_j)}.$$

ч.т.д.

Вероятность события B_k называется *априорной*, вероятность B_k при условии A – *апостериорной*, $\mathrm{P}(A|B_k)$ – likelihood.

Давайте рассмотрим пару примеров, которые дадут противоречащие интуиции результаты.

Пример 1.6. ***Парадокс Монти-Холла.***
Представьте, что вы стали участником игры, в которой вам нужно выбрать одну из трёх дверей. За одной из дверей находится автомобиль, за двумя другими дверями — козы. Вы выбираете одну из дверей, например, номер 1, после этого ведущий, который знает, где находится автомобиль, а где — козы, открывает одну из оставшихся дверей, в которой находится коза, например, номер 3. После этого он спрашивает вас — не желаете ли вы изменить свой выбор и выбрать дверь номер 2? Увеличатся ли ваши шансы выиграть автомобиль, если вы примете предложение ведущего и измените свой выбор?

Распишем все возможные ситуации в виде таблицы (для определенности, пусть игрок выбрал первую дверь):

Вероятность того, что игрок изначально выбрал правильную дверь равна вероятности, что в одной из трех дверей будет автомобиль, т.е. $\frac{1}{3}$, *тогда же как вероятнось*

Дверь 1	Дверь 2	Дверь 3	Ведущий откроет	Результат, если менять выбор	Результат, если не менять выбор
Авто	Коза	Коза	2 или 3	**Коза**	**Авто**
Коза	Авто	Коза	3	**Авто**	**Коза**
Коза	Коза	Авто	2	**Авто**	**Коза**

появления козы равна $\frac{2}{3}$. Меняя дверь, игрок выигрывает, если изначально взял проигрышную, и наоборот, то есть

$$\mathrm{P}(A_{hold}) = 2 \cdot \left(\frac{1}{3} \cdot \frac{1}{2}\right) = \frac{1}{3},$$
$$\mathrm{P}(A_{change}) = 2 \cdot \frac{1}{3} = \frac{2}{3}.$$

Часто при решении этой задачи рассуждают так: ведущий всегда в итоге убирает одну проигрышную дверь, и тогда вероятности появления автомобиля за двумя не открытыми становятся равны $\frac{1}{2}$, вне зависимости от первоначального выбора. Но это неверно: хотя возможностей выбора действительно остаётся две, эти возможности (с учётом предыстории) ***не являются равновероятными****. Это так, поскольку изначально все двери имели равные шансы быть выигрышными, но затем имели разные вероятности быть исключёнными.*

Рассмотрим еще один пример, в некоторой степени являющийся перефразированием предыдущего.

Пример 1.7. ***Пример с заключенным.***

Есть три осужденных, ожидающие исполнения приговора. Глава решил одного из трех осужденных помиловать, причем выбор случаен. Первый заключенный решил узнать у охранника, кто помилован, но тому запрещено называть имя освобожденного. Тогда первый заключенный предложил охраннику сказать следующее:

- *Если освободили второго, то скажи, что приговор будет исполнен в отношении третьего.*
- *Если помилуют третьего, то скажи о "казни" второго.*
- *Если помилуют меня, то случайно скажи о "казни" второго или третьего.*

Пусть I, II, III означает, что помилованы первый, второй или третий соответственно. Охранник сказал, что в отношении второго будет исполнен приговор. Обозначим это за A. Первый заключенный радуется этому, ведь шанс освобождения 50 на 50. Но тут следует посчитать $\mathrm{P}(I|A)$*. По теореме Байеса получаем*

$$\mathrm{P}(I|A) = \frac{\mathrm{P}(A|I)\,\mathrm{P}(I)}{\mathrm{P}(A|I)\,\mathrm{P}(I) + \mathrm{P}(A|II)\,\mathrm{P}(II) + \mathrm{P}(A|III)\,\mathrm{P}(III)}$$
$$= \frac{1/2 \cdot 1/3}{1/2 \cdot 1/3 + 0 \cdot 1/3 + 1 \cdot 1/3} = \frac{1}{3}.$$

Рассмотрим еще один пример. Фигурирующие в нем числа не имеют отношения к реальности.

Пример 1.8. ***Диагностика.***

Пусть имеются заболевание с долей распространения среди населения 0,001 и метод диагностики, который с вероятностью 0,9 выявляет больного, но с вероятностью 0,01 даёт ложноположительный результат (англ. false positive) – "выявление болезни у здорового". Найти вероятность того, что человек здоров, если диагностика показала положительный результат.

Пусть H – человек здоров, I – болен, T – результат диагностики положительный, F – результат диагностики отрицательный. Из условия, что $\mathrm{P}(H) = 0,999$, $\mathrm{P}(I) = 0,001$, $\mathrm{P}(T|I) = 0,9$, $\mathrm{P}(F|I) = 0,1$, $\mathrm{P}(F|H) = 0,99$, $\mathrm{P}(T|H) = 0,01$. *Найдём искомую вероятность*

$$\mathrm{P}(H|T) = \frac{\mathrm{P}(T|H)\,\mathrm{P}(H)}{\mathrm{P}(T|H)\,\mathrm{P}(H) + \mathrm{P}(T|I)\,\mathrm{P}(I)} = \frac{0,01 \cdot 0,999}{0,01 \cdot 0,999 + 0,9 \cdot 0,001} \approx 0,917.$$

Полученный результат может быть протеричив интуиции, ведь казалось бы вероятность ложноположительной диагностики мала. Можно рассмотреть популяцию в 1000 человек. Тогда 1 болен, но плюс-минус у 10 человек будет положительный результат, хотя 9 из них здоровы.

Раз получились такие результаты, то имеет смысл провести вторую проверку, независимую от первой. Давайте посмотрим вероятность того, что человек здоров, если он получил два положительных результата

$$\mathrm{P}(H|T_1T_2) = \frac{\mathrm{P}(T_1T_2|H)\,\mathrm{P}(H)}{\mathrm{P}(T_1T_2)} = \frac{\mathrm{P}(T_1|H)\,\mathrm{P}(T_2|H)\,\mathrm{P}(H)}{\mathrm{P}(T_1|H)\,\mathrm{P}(T_2|H)\,\mathrm{P}(H) + \mathrm{P}(T_1|I)\,\mathrm{P}(T_2|I)\,\mathrm{P}(I)}$$
$$= \frac{0,01^2 \cdot 0,999}{0,01^2 \cdot 0,999 + 0,9^2 \cdot 0,001} \approx 0,11.$$

1.5 Схема независимых испытаний Бернулли

Прежде чем описать схему независимых испытаний Бернулли, или как кратко мы её будем называть *схемой Бернулли*, необходимо формализовать *независимые испытания.*

Определение 1.15. ***Независимость испытаний.***

Пусть два испытания описываются вероятностными пространствами $(\Omega_1, \mathcal{F}_1, \mathrm{P}_1)$, $(\Omega_2, \mathcal{F}_2, \mathrm{P}_2)$. *Будем называть данные испытания* ***независимыми****, если их пара описывается вероятностным пространством* $(\Omega, \mathrm{F}, \mathrm{P})$, *где* $\Omega = \Omega_1 \times \Omega_2$, $\mathrm{F} = \mathcal{F}_1 \times \mathcal{F}_2$ *и для любых* $A_1 \in \mathcal{F}_1$, $A_2 \in \mathcal{F}_2$

$$\mathrm{P}(A_1 \times A_2) = \mathrm{P}_1(A_1)\,\mathrm{P}_2(A_2).$$

Большее число независимых испытаний можно определить по индукции. Так, для серии из n независимых испытания при любых $A_1 \in \mathcal{F}_1, \ldots A_n \in \mathcal{F}_n$ *имеем*

$$\mathrm{P}(A_1 \times \ldots \times A_n) = \mathrm{P}_1(A_1) \cdot \ldots \cdot \mathrm{P}_n(A_n).$$

Замечание. *На самом деле выше мы задали вероятностное пространство в широком смысле, так как* $\mathcal{F}_1 \times \mathcal{F}_2$ *не обязательно является сигма-алгеброй.*

Кроме того, знакомый с теорией меры читатель заметил, что приведенная выше конструкция похожа на произведение мер, о котором подробно можно почитать, например, в [7].

Теперь мы готовы сформулировать определение *схемы Бернулли.*

Определение 1.16. ***Схема Бернулли.***

Схемой *из n испытаний* ***Бернулли*** *будем называть n независимых испытаний, каждое из которых описывается вероятностным пространством $(\Omega, \mathcal{F}, \mathrm{P})$, где*

$$\Omega = \{\omega_s, \omega_u\}, \quad \mathcal{F} = 2^{\Omega},$$
$$\mathrm{P}(\omega_s) = p \in [0,1], \quad \mathrm{P}(\omega_u) = q = 1-p.$$

ω_s мы будем трактовать как успех в одном испытании, ω_u – неудачу. При этом каждой реализации испытаний Бернулли сопоставляется вектор ω из ω_s и ω_u. Тогда, если в векторе ω k компонент равняется ω_s (в n испытаниях было k успехов), то

$$\mathrm{P}(\omega) = p^k q^{n-k}.$$

Отметим почти очевидные утверждения, но в силу важности оформим их в виде теоремы и следствий.

Теорема 1.10. ***Вероятность наступления ровно k успехов.***

Если S_n – число успехов в n испытаниях Бернулли, то имеем

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

Следствие. ***Вероятность наступления хотя бы одного успеха.***

$\mathrm{P}(S_n \geqslant 1) = 1 - q^n$

Следствие. ***О числе испытаний, достаточных хотя бы для одного успеха.***

Пусть $\alpha \in (0,1)$ – некоторый фиксированный уровень доверия, а p – вероятность успеха. Тогда $P(S_n \geqslant 1) \geqslant \alpha$ при $n = \lceil \frac{\ln(1-\alpha)}{\ln(1-p)} \rceil$

Доказательство. По предыдущему следствию:

$$1 - q^n \geqslant \alpha$$
$$q^n \leqslant 1 - \alpha$$
$$n \ln(1-p) \leqslant \ln(1-\alpha)$$
$$n \geqslant \frac{\ln(1-\alpha)}{\ln(1-p)}$$

Тогда, ответом будет $\lceil \frac{\ln(1-\alpha)}{\ln(1-p)} \rceil$ *ч.т.д.*

Далее сформулируем и докажем теорему о наиболее вероятном числе успехов.

Теорема 1.11. ***Наиболее вероятное количество успехов.***

Пусть имеется схема Бернулли из n испытаний. Тогда наиболее вероятное число успехов вычисляется как

$$k_* = \begin{cases} p(n+1) - 1 \text{ и } p(n+1), & \text{если } p(n+1) \in \mathbb{N}, \\ \lceil p(n+1) \rceil, & \text{иначе.} \end{cases}$$

Доказательство. Зафиксируем $k \in \{0, \ldots n-1\}$. Заметим, что

$$\frac{\mathrm{P}(S_n = k+1)}{\mathrm{P}(S_n = k)} = \frac{C_n^{k+1} p^{k+1} q^{n-k-1}}{C_n^k p^k q^{n-k}} = \frac{k!(n-k)!n!}{n!(k+1)!(n-k-1)!} \cdot \frac{p}{1-p} = \frac{(n-k)p}{(k+1)q}.$$

Сравним эту дробь с единицей:

$$\frac{(n-k)p}{(k+1)q} \vee 1 \quad \Leftrightarrow \quad np - kp \vee k + 1 - kp - p \quad \Leftrightarrow \quad (n+1)p - 1 \vee k,$$

Рассмотрим два случая:

1. $p(n+1) \in \mathbb{N}$.

   Тогда существует $k_0 = p(n+1) - 1$, для которого справедливы соотношения:

$$k < k_0 \Rightarrow \mathrm{P}(S_n = k) < \mathrm{P}(S_n = k+1)$$
$$k = k_0 \Rightarrow \mathrm{P}(S_n = k) = \mathrm{P}(S_n = k+1)$$
$$k > k_0 \Rightarrow \mathrm{P}(S_n = k) > \mathrm{P}(S_n = k+1)$$

   Таким образом, $k_0, k_0 + 1$ - наиболее вероятные значения

2. $p(n+1) \notin \mathbb{N}$.

   Тогда существует такой k_1, что

$$\mathrm{P}(S_n = k_1) < \mathrm{P}(S_n = k_1 + 1)$$
$$\mathrm{P}(S_n = k_1 + 1) > \mathrm{P}(S_n = k_1 + 2)$$

   И тогда, наиболее вероятным значением будет $k_1 + 1$.

ч.т.д.

Определение 1.17. ***Полиномиальная схема.***

Рассмотрим n независимых испытаний, m возможных исходов, а также $p = (p_1, ..., p_m)$ – вектор вероятностей исходов, где $\sum\limits_{i=1}^{m} p_i = 1$. Зададим вероятностное пространство $(\Omega^n, \mathcal{F}, \mathrm{P})$ следующим образом:

$$\Omega = \{(i_1, ..., i_m) \in \{0,1\}^m, \sum_{j=1}^{m} i_j = 1\},$$
$$\Omega^n = \{A \in M_{m\times n} : A_{ij} \in \{0,1\}, \sum_i A_{ij} = 1, \sum_{i,j} A_{ij} = n\},$$
$$\mathcal{F} = 2^\Omega, \qquad \mathrm{P}(\omega) = p_1^{\sum A_{1j}} p_2^{\sum A_{2j}} \cdot ... \cdot p_m^{\sum A_{mj}}$$

Теорема 1.12. ***Вычисление вероятностей для полиномиальной схемы.***

Пусть $S_{n,j}$ – количество исходов типа j в n испытаниях. Тогда,

$$\mathrm{P}(S_{n,1} = k_1, S_{n,2} = k_2, ..., S_{n,m} = k_m) = \frac{n!}{k_1!k_2!...k_m!} p_1^{k_1} \cdot ... \cdot p_m^{k_m}, \quad \sum k_j = n$$

Доказательство. Почти очевидно. Подсчет мультиномиального коэффициента напрямую следует из соответствующей теоремы. *ч.т.д.*

1.6 Предельные теоремы, связанные со схемой Бернулли

1.6.1 Закон больших чисел

Скорее всего, читатель ранее слышал, что при большом числе испытаний доля успехов приблизительно равняется вероятности успеха. Это действительно имеет место при определённых предположениях, в частности, для схемы Бернулли. Это утверждение, точнее утверждения, носят название *законов больших чисел.* Приведём их формулировки для схемы Бернулли, а доказывать их мы будем уже ниже в более общих предположениях. Начнём со *слабого закона больших чисел.*

Теорема 1.13. ***Закон больших чисел Бернулли (слабый).***

Пусть, как и прежде, S_n – число успехов в n испытаниях Бернулли и вероятность успеха равняется p. Тогда для любого $\varepsilon > 0$

$$\mathrm{P}\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) \to 0, \quad n \to \infty.$$

Приведём формулировку *усиленного закона больших чисел*

Теорема 1.14. ***Закон больших чисел Бернулли (усиленный).***

$$\mathrm{P}\left(\lim_{n\to\infty} \frac{S_n}{n} = p\right) = 1.$$

Вернёмся к примеру, в котором мы вычислили вероятность пересечения иглы и разлинованного стола при случайном броске. Напомним, что

$$p = \frac{4l}{\pi d},$$

где d – расстояние между линиями, $2l < d$ – длина иглы. Далее рассмотрим уже n независимых бросков иглы на разлинованный стол. Мы можем данную задачу формализовать в рамках схемы Бернулли с указанной нами вероятностью успеха. Благодаря *законам больших чисел*, мы можем считать, что при больших n доля успешных исходов приблизительно равняется вероятности успеха, то есть

$$\frac{S_n}{n} \approx \frac{4l}{\pi d},$$

что можно использовать для приближенного вычисления числа π, так как величины S_n, n, l и d мы можем непосредственно измерить. Это одно из первых решений, иллюстрирующий *метод Монте-Карло.* Он заключается в том, что для приближенного вычисления той или иной величины, запускается многократная симуляция (в нашем примере – бросание иглы), и затем с помощью предельных теорем, в частности *законов больших чисел*, мы получаем соотношения, из которых рассчитывается интересующий показатель.

1.6.2 Локальная предельная теорема Муавра-Лапласа

При больших n вычислить вероятность $\mathrm{P}(S_n = k)$ бывает не всегда просто с вычислительной точки зрения. Да, условный Wolfram сможет вычислить, но реализовать напрямую вычисление этой вероятности в популярных языках программирования не всегда получится, поэтому приходится прибегать к аппроксимациям. Одной из таких аппроксимаций является локальная предельная теорема Муавра-Лапласа.

Начнём со вспомогательных лемм.

Лемма 1.1. *Пусть $p \in (0, 1)$ и положим:*

$$H(x) = x \ln \frac{x}{p} + (1 - x) \ln \frac{1 - x}{1 - p}, \quad p^* = \frac{k}{n}.$$

Тогда при $k \to \infty$, $n - k \to \infty$ (как следствие, $n \to \infty$)

$$\mathrm{P}(S_n = k) \sim \frac{1}{\sqrt{2\pi n p^*(1 - p^*)}} \exp\{-nH(p^*)\}.$$

Доказательство. Воспользуемся формулой Стирлинга, в силу которой $n! \sim \sqrt{2\pi n}n^n e^{-n}$ при $n \to \infty$. Тогда имеем

$$\begin{aligned}
\mathrm{P}(S_n = k) &= \frac{n!}{(n-k)!k!}p^k(1-p)^{n-k} \sim \sqrt{\frac{n}{2\pi k(n-k)}}\frac{n^n}{k^k(n-k)^{n-k}}p^k(1-p)^{n-k} \\
&= \frac{1}{\sqrt{2\pi n p^*(1-p^*)}}\exp\left\{-k\ln\frac{k}{n} - (n-k)\ln\frac{n-k}{n} + k\ln p + (n-k)\ln(1-p)\right\} \\
&= \frac{1}{\sqrt{2\pi n p^*(1-p^*)}}\exp\left\{-nH(p^*)\right\}.
\end{aligned}$$

ч.т.д.

Замечание. *Введенная функция $H(x)$ есть ни что иное как расстояние Кульбака-Лейблера между распределениями Бернулли с параметрами p и x (понятие "распределение" формально вводится в следующей главе). Она показывает величину потери информации от замены "реального" параметра p на x (в нашем случае $x = k/n$). Более подробно с этим понятием можно ознакомиться например в [1], [2].*

Теперь мы уже готовы сформулировать саму локальную предельную теорему Муавра-Лапласа.

Теорема 1.15. ***Локальная предельная теорема Муавра-Лапласа.***
Пусть помимо условий леммы $k - np = o\left(n^{2/3}\right)$. Тогда

$$\mathrm{P}(S_n = k) \sim \frac{1}{\sqrt{2\pi npq}}\exp\left\{-\frac{(k-np)^2}{2npq}\right\}.$$

Доказательство. Заметим, что функция $H(x)$ является аналитической на $(0,1)$ и

$$H'(x) = \ln\frac{x}{p} - \ln\frac{1-x}{1-p}, \quad H''(x) = \frac{1}{x} + \frac{1}{1-x},$$
$$H(p) = H'(p) = 0.$$

Кроме того, так как $k - np = o\left(n^{2/3}\right)$, то $p^* - p = o\left(n^{-1/3}\right)$. Тогда

$$\begin{aligned}
H(p^*) &= \frac{1}{2}\left(\frac{1}{p} + \frac{1}{1-p}\right)(p-p^*)^2 + O\left((p-p^*)^3\right) \\
&= \frac{(p-p^*)^2}{2pq} + o\left(n^{-1}\right).
\end{aligned}$$

В итоге мы получим, что

$$\mathrm{P}(S_n = k) \sim \frac{1}{\sqrt{2\pi npq}}\exp\left\{-\frac{(k-np)^2}{2npq}\right\}.$$

ч.т.д.

Замечание. *Введём функцию $\phi(x)$, которая задаётся следующим образом:*

$$\phi(x) = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2}\right\}.$$

Также заметим, что

$$\mathrm{P}(S_n = k) = \mathrm{P}\left(\frac{S_n - np}{\sqrt{npq}} = \frac{k-np}{\sqrt{npq}}\right).$$

Немного забежим вперёд: np и $\sqrt{npq}$ – математическое ожидание и корень из дисперсии (оно же – стандартное отклонение) биномиального распределения. Положим $x = (k - np)/\sqrt{npq}$. Тогда полученный результат можно переписать как

$$\mathrm{P}\left(\frac{S_n - np}{\sqrt{npq}} = x\right) \sim \frac{1}{\sqrt{npq}}\phi(x). \tag{1.4}$$

1.6.3 Интегральная предельная теорема Муавра-Лапласа

Рассмотрим сумму вида

$$P_n(a,b) = \sum_{k=1}^{m} \mathrm{P}\left(\frac{S_n - np}{\sqrt{npq}} = x_k\right),$$

и предположим, что $x_1 = a$, $x_m = b$. По локальной теореме Муавра-Лапласса и соотношению (1.4) заменим каждое слагаемое аппроксимацией

$$P_n(a,b) \approx \sum_{k=1}^{m} \frac{1}{\sqrt{npq}}\phi(x_k),$$

и заметим, что $x_{k+1} - x_k = 1/\sqrt{npq}$. Тогда правая часть по сути есть интегральная сумма и правдоподобным является соотношение

$$P_n(a,b) \approx \int\limits_a^b \phi(x)dx = \frac{1}{\sqrt{2\pi}}\int\limits_a^b e^{-x^2/2}dx = \Phi(b) - \Phi(a),$$

где $\Phi(x)$ определяется как

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{-t^2/2} = \int\limits_{-\infty}^{x}\phi(t)dt.$$

Забегая вперёд, $\Phi(x)$, $\phi(x)$ – функция распределения и плотность *стандартного нормального закона* соответственно.

Сформулируем интегральную теорему Муавра-Лапласа, но доказывать её не будем, так как в дальнейшем мы разберём более общую теорему.

Теорема 1.16. ***Интегральная предельная теорема Муавра-Лапласа.***
Справедливо соотношение

$$\sup_{-\infty \leqslant a < b \leqslant +\infty}\left|\mathrm{P}\left(a < \frac{S_n - np}{\sqrt{npq}} \leqslant b\right) - (\Phi(b) - \Phi(a))\right| \to 0, \quad n \to \infty.$$

Замечание. *Из теоремы следует, что при $-\infty \leqslant A < B \leqslant +\infty$*

$$\mathrm{P}(A < S_n \leqslant B) - \left(\Phi\left(\frac{B - np}{\sqrt{npq}}\right) - \Phi\left(\frac{A - np}{\sqrt{npq}}\right)\right) \to 0, \quad n \to \infty.$$

Пример 1.9. *Есть село Н., в нем проживает 3200 человек. Каждый день из села в город В. [2] идет поезд, и каждый сельчанин независимо выбирает 6 дней в месяце для*

[2]Под селом Н. и городом В. закодированы реальные населенные пункты, между которыми между которыми ходят, правда, полтора поезда в день.

поездки в город. Считаем, что в месяце 30 дней, все дни равнозначны и из Н. нет никакого иного способа выехать, кроме как на поезде. Требуется найти вместимость поезда M, при которой переполнение случается не чаще 1 раза в 100 дней.

Введем $n = 3200$ *(количество жителей),* $p = \frac{1}{5}$ *– вероятность поездки. Тогда,*

$$np = 640 \quad npq = 512 \quad \sqrt{npq} = 16\sqrt{2}$$
$$\mathrm{P}(S_{3200} > M) \leqslant \frac{1}{100} \quad \Longleftrightarrow \quad \mathrm{P}(S_{3200} \leqslant M) \geqslant \frac{99}{100}$$
$$x_1 = \frac{0 - np}{\sqrt{npq}} \qquad x_2 = \frac{M - np}{\sqrt{npq}}$$

Теперь, по интегральной теореме Муавра-Лапласа.

$$\mathrm{P}\left(\frac{0-np}{\sqrt{npq}} \leqslant \frac{S_{3200}-np}{\sqrt{npq}} \leqslant \frac{M-np}{\sqrt{npq}}\right) \approx \Phi\left(\frac{M-np}{\sqrt{npq}}\right) - \Phi\left(\frac{0-np}{\sqrt{npq}}\right)$$

Однако, значение $\Phi(x_1)$ *пренебрежимо мало, и потому, мы можем его опустить.*

$$\mathrm{P}\left(\frac{S_{3200}-np}{\sqrt{npq}} \leqslant \frac{M-np}{\sqrt{npq}}\right) \approx \Phi\left(\frac{M-np}{\sqrt{npq}}\right)$$
$$\Phi\left(\frac{M-640}{16\sqrt{2}}\right) \geqslant \frac{99}{100}$$

Подберем минимально возможную вместимость, давайте приравняем $\Phi(x_2) = \frac{99}{100}$*. Значит,*

$$M = \lceil q_{\frac{99}{100}} \cdot 16\sqrt{2} + 640 \rceil \quad , \text{ где } \Phi\left(q_{\frac{99}{100}}\right) = \frac{99}{100}$$

Пользуясь следующей таблицей в качестве справочного материала, получим

$$q_{\frac{99}{100}} = 2.31 \quad \Rightarrow \quad M = 693$$

Что и будет ответом на задачу.

1.6.4 Теорема Пуассона

Теорема 1.17. ***Теорема Пуассона.***

Пусть в схеме Бернулли из n *независимых испытаниях* $p_n = \lambda/n + o(1/n)$ *и положим*

$$P_n(k) = \begin{cases} C_n^k p_n^k q_n^{n-k}, & k \in \{0, \ldots, n\}, \\ 0, & \textit{иначе}. \end{cases}$$

Тогда

$$\lim_{n\to\infty} P_n(k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Доказательство. Заметим, что при $k \in \{0, \ldots, n\}$ имеет место

$$P_n(k) = C_n^k p_n^k q_n^{n-k} = \frac{n \cdot \ldots \cdot (n-k+1)}{k!} \left(\lambda/n + o\left(\tfrac{1}{n}\right)\right)^k \left(1 - \lambda/n + o\left(\tfrac{1}{n}\right)\right)^{n-k}$$
$$= \frac{n \cdot \ldots \cdot (n-k+1)}{n^k k!} (\lambda + o(1))^k \frac{(1-\lambda/n + o(\frac{1}{n}))^n}{(1-\lambda/n + o(\frac{1}{n}))^k} \xrightarrow[n\to\infty]{} e^{-\lambda}\frac{\lambda^k}{k!}$$

Этот предельный переход возможен благодаря второму замечательному пределу (коим мы пользуемся в выводе результата).

ч.т.д.

1.6.5 О погрешностях в предельных теоремах

Как видно из формулировки и доказательства локальной теоремы Муавра-Лапласа, её имеет смысл применять при "больших" n и k, отстоящих "недалеко" от np. Кроме того, при маленьких p для приемлемой аппроксимации естественно, чтобы n было "очень большим" (более детально о погрешности в локальной предельной теореме можно почитать в [3]).

Далее определим $F_n(x)$ следующим образом:

$$F_n(x) = \sum_{k=0}^{n} C_n^k p^k q^{n-k} \mathbb{1}(x \leqslant k),$$

где $\mathbb{1}(.)$ – предикаторная функция, возвращающая единицу при истинности её аргумента и ноль – в противном случае. Функция $F_n(x)$ называется *функцией распределения*. Имеет место оценка Берри-Эссена

$$\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| \leqslant C\frac{p^2+q^2}{\sqrt{npq}}, \quad C > 0 \text{ – абсолютная константа,}$$

которая позволяет оценить погрешность в интегральной теореме Муавра-Лапласа. Кроме того, порядок $1/\sqrt{n}$ нельзя улучшить. Видим, что для приемлемой аппроксимации n должно быть "большим", а при маленьких p "ещё больше".

Как раз при малых p имеет смысл использовать теорему Пуассона. Действительно, имеет место оценка Прохорова для соответствующего большинству задач случая$\lambda = np$

$$\left|\sum_{k=0}^{\infty} \mathrm{P}(S_n = k) - e^{-\lambda}\frac{\lambda^k}{k!}\right| \leqslant 2p\min(2, np),$$

откуда видно, что пуассоновская аппроксимация будем хорошо работать при маленьких p и np. Также имеет место оценка, которую мы ниже докажем

$$\left|\mathrm{P}(S_n = k) - e^{\lambda}\frac{\lambda^k}{k!}\right| \leqslant np^2 = \lambda p.$$

Таким образом, можно сформулировать "эвристику": при маленьких p и np (меньше 10) использовать теорему Пуассона, при больших np (больше сотни) и близких к $1/2$ p – нормальную аппроксимацию. В "промежуточных" случаях обе теоремы плюс-минус работают одинаково, но тут все зависит от конкретной задачи. За большим числом примеров можно обратиться, например, к [9].

1.7 Упражнения

1.7.1 Модель классической вероятности

Упражнение 1.1. *Колода из 36 карт хорошо перемешали (т.е ве возможные расположения карт равновероятны). Найти вероятности событий:*

1. *четыре туза расположены рядом,*
2. *места расположения тузов образуют арифметическую прогрессию с шагом 7.*

Упражнение 1.2. *На полке в случайном порядке расставлены 40 книг, среди которых находится трехтомник А.С. Пушкина. Найти вероятность того, что эти тома стоят в порядке возрастания слева направо (но не обязательно рядом).*

Упражнение 1.3. *Брошены три монеты. Предполагая, что элементарные события равновероятны, найти вероятности событий:*

1. *первая монета выпала "гербом"вверх,*
2. *выпали ровно два "герба*
3. *выпало не больше двух "гербов".*

Упражнение 1.4. *Из множества всех последовательностей длины n, состоящих из цифр 0,1,2, случайно выбирается одна. Найти вероятности событий:*

1. *последовательность начинается с 0,*
2. *последовательность содержит ровно $m+2$ нуля, причем 2 из них находятся на концах последовательности,*
3. *последовательность содержит ровно m единиц,*
4. *в последовательности ровно m_0 нулей, m_1 единиц, m_2 двоек.*

Упражнение 1.5. *Какова вероятность того, что четырехзначный номер случайно взятого автомобиля в большом городе*

1. *состоит 4 разных цифр;*
2. *имеет только одинаковые цифры;*
3. *имеет две пары одинаковых цифр;*
4. *имеет только три одинаковые цифры?*

Упражнение 1.6. *Случайно брошены 10 игральных костей. Найти вероятности:*

1. *не выпало ни одной "6";*
2. *выпали ровно три "6";*
3. *выпала хотя бы одна "6";*
4. *выпали хотя бы две "6".*

Упражнение 1.7. *Из множества $\{0, \ldots, N\}$ по схеме равновероятного выбора с возвращением извлекаются числа $X_1, \ldots, X_m$. Пусть $b_{k,N}^{(m)} = \mathrm{P}(X_1 + \ldots + X_m = k)$, $0 \leqslant k \leqslant mN$.*

1. *Доказать, что* $b_{k,N}^{(m)} = b_{mN-k}^{(m)}$;

2. *Доказать, что*

$$\sum_{k=0}^{mN} b_{k,N}^{(m)} z^k = \frac{1}{(N+1)^m} \left(\frac{1 - z^{N+1}}{1 - z} \right)^m,$$

то есть что

$$b_{k,N}^{(m)} = \frac{1}{(N+1)^m} \sum_{j=0}^{\lfloor \frac{k}{N+1} \rfloor} (-1)^j C_{k-j(N+1)+m-1}^{m-1} C_m^j.$$

Упражнение 1.8. *Найти вероятность того, что в номере случайно выбранного в большом городе автомобиля сумма первых двух цифр равна сумме двух последних.*

Упражнение 1.9. *Некоторые горожане считают трамвайный, троллейбусный или автобусный билет "счастливым если сумма первых трех цифр его шестизначного номера совпадает с суммой последних трех цифр. Найти вероятность получить "счастливый" билет.*

Упражнение 1.10. *Из карточек разрезной азбуки составлено слово "СТАТИСТИКА". Затем из этих 10 карточек по схеме случайного выбора без возвращения отобрано 5 карточек. Найти вероятность того, что из отобранных карточек можно составить слово "ТАКСИ".*

Упражнение 1.11. *Из 30 чисел (1, 2 ,..., 29 , 30) случайно отбирается 10 различных чисел. Найти вероятности событий:*

1. *все числа нечетные,*

2. *ровно 5 чисел делится на 3,*

3. *5 чисел четных и 5 нечетных, причем ровно одно число делится на 10.*

Упражнение 1.12. *Из урны, содержащей M_1 шаров с номером 1, M_2 шаров с номером 2, ..., M_N шаров с номером N, случайно без возвращения выбирается n шаров. Найти вероятности событий:*

1. *появилось m_1 шаров с номером 1,m_2 шаров с номером 2, ..., m_N шаров с номером N,*

2. *каждый из N номеров появился хотя бы один раз.*

Упражнение 1.13. *Десять рукописей разложены по 30 папкам (на одну рукопись 3 папки). Найти вероятность того, что в случайно выбранных 6 папках не содержится целиком ни одной рукописи.*

Упражнение 1.14. *За круглый стол рассаживаются в случайном порядке $2n$ гостей. Какова вероятность того, что гостей можно разбить на n непересекающихся пар так, чтобы каждая пара состояла из сидящих рядом мужчины и женщины?*

Упражнение 1.15. *В первом ряду кинотеатра, состоящем из N кресел, сидит n человек. Предполагая, что все возможные размещения этих n человек в первом ряду равновероятны, найти вероятности следующих событий:*

1. *никакие 2 человека не сидят рядом;*

2. *каждый из n человек имеет ровно одного соседа;*

3. *из любых двух кресел, расположенных симметрично относительно середины ряда, хотя бы одно свободно.*

Упражнение 1.16. *В зале кинотеатра в первых двух рядах, каждый из которых состоит из N кресел, сидит n человек. Найти вероятности следующих событий:*

1. *в первом ряду никакие 2 человека не сидят рядом;*

2. *во втором ряду каждый человек имеет ровно одного соседа;*

3. *в первом ряду из любых двух кресел, расположенных симметрично относительно середины ряда, хотя бы одно свободно.*

Упражнение 1.17. *Из всех отображений множества $\{1, 2, \ldots, n\}$ в себя случайно выбирается одно. Найти вероятности событий:*

1. *выбранное отображение каждый из n элементов переводит в 1;*

2. *элемент i имеет ровно k прообразов;*

3. *элемент i переводится в j;*

4. *выбранное отображение элементы $i_1, i_2, \ldots, i_k (1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n)$ переводит в элементы $j_1, j_2, \ldots, j_k$ соответственно.*

Упражнение 1.18. *Из множества перестановок длины n случайно выбирается одна из них. Найти вероятности событий:*

1. *выбрана тождественная перестановка;*

2. *выбранная перестановка, переводящая элементы $i_1, i_2, \ldots, i_k$ $(i_1 < i_2 < \ldots < i_k)$ переводит в элементы $j_1, j_2, \ldots, j_k$ соответственно;*

3. *элемент i в выбранной подстановке образует единичный цикл, т.е $i \to j$ и $j \to i$;*

4. *элементы 1,2,3 образуют цикл длины 3: $1 \to 2 \to 3 \to 1$ или $1 \to 3 \to 2 \to 1$;*

5. *все элементы образуют один цикл.*

Упражнение 1.19. *Найти вероятность P_n того, что в случайно выбранной перестановке длины n найдется хотя бы один цикл единичной длины. Найти $\lim\limits_{n\to\infty} P_n$.*

Упражнение 1.20. *Имеются $z+m$ билетов, из которых n выигрышных. Одновременно приобретаются k билетов. Определить вероятность того, что среди них s выигрышных.*

Упражнение 1.21. *В генуэзской лотерее разыгрываются девяносто номеров, из которых выигрывают пять. По условию можно ставить на любой из девяноста номеров или на любую совокупность двух, трех, четырех или пяти номеров, причем для получения выигрыша должны выиграть все выбранные номера. Какова вероятность выигрыша в каждом из указанных пяти случаев?*

Упражнение 1.22. *Для уменьшения общего количества игр $2n$ команд спортсменов по жребию разбиваются на две подгруппы. Определить вероятность того, что две наиболее сильные команды окажутся:*

1. *в равных подгруппах;*
2. *в одной подгруппе.*

Упражнение 1.23. *В зале, насчитывающем $n + k$ мест, случайным образом занимают места n человек. Определить вероятность того, что будут заняты определенные $m \leqslant n$ мест.*

Упражнение 1.24. *Из колоды карт (52 штуки) наудачу извлекаются три карты. Найти вероятность того, что это будут тройка, семерка и туз:*

1. *независимо от порядка извлечения карт без возвращения;*
2. *в указанном порядке при извлечении карт без возвращения;*
3. *в указанном порядке с возвращением каждой карты в колоду.*

Упражнение 1.25. *Пусть $(i_1, \ldots, i_n)$ случайная перестановка чисел $1, \ldots, n$. Найти:*

1. *вероятность $P_{(m)}$ того, что при перестановках чисел $1, \ldots, n$ в точности m чисел останутся на своих местах (частичный беспорядок);*
2. *вероятность $P_{(\geqslant 1)}$ того, что в результате перестановок по крайней мере одно из чисел $1, \ldots, n$ останется на своем месте;*
3. *вероятность полного беспорядка (когда ни одно из чисел не остается на своем месте).*

Упражнение 1.26. *Пусть имеется n писем и m конвертов. Письма по конвертам раскладываются "случайным образом", иначе говоря, предполагается, что приписывание соответствующих вероятностей осуществляется в соответствии "классическим" способом задания вероятностей. Пусть $P_{(m)}$ - вероятность того, что в точности m писем попадут в "свои" конверты. Как данную вероятность можно приближенно посчитать при фиксированном m и достаточно большом n?*

Упражнение 1.27. *Уходя из детского сада, каждый из n детей "случайным образом" берет один левый и один правый ботинок. Найти:*

1. *вероятность того, что все они уйдут не в своих парах ботинок;*
2. *вероятность того, что каждый из них возьмет не свой левый и не свой правый ботинок.*

Упражнение 1.28. *Сколькими способами n супружеских пар ($n \geqslant 3$) можно разместить за круглым столом так, чтобы мужчины и женщины чередовались, но супруги не сидели рядом?*

Упражнение 1.29. *В урне K красных, L белых и M черных шаров. Из урны с возвращением (без возвращения) извлекаются n шаров. Найти вероятность того, что в выборке будет k красных, l белых и m черных шаров.*

Упражнение 1.30. *30 шаров размещаются по 8 ящикам так, что для каждого шара одинаково возможно попадание в любой ящик. Найти вероятность размещения, при котором будет 3 пустых ящик, 2 ящика – с тремя, 2 ящика – с шестью и 1 ящик – с двенадцатью шарами.*

Упражнение 1.31. *Из урны, содержащей N_1 белых шаров, N_2 чёрных и N_3 красных, последовательно без возвращения извлекают шары до тех пор, пока не появится красный шар. Найти вероятности следующих событий:*

1. *вынуто n_1 белых шаров и n_2 чёрных;*

2. *не появилось ни одного белого шара;*

3. *всего вынуто k шаров.*

1.7.2 Модель геометрической вероятности

Упражнение 1.32. *В точке C, положение которой на телефонной линии AB длины L равновозможно, произошел разрыв. Определить вероятность того, что точка C удалена от точки A на расстояние, не меньшее l.*

Упражнение 1.33. *Параллельно оси абсцисс проведены две линии, ординаты которых равны 8 и 9.5. Определить вероятность того, что круг радиуса 2.5 не будет пересечен осью абсцисс и ни одной из двух параллельных линий, если положение центра круга равновозможно на оси ординат в интервале от 0 до 9.5.*

Упражнение 1.34. *Начерчены пять концентрических окружностей, радиусы которых равны соответственно kr, $(k = 1, 2, 3, 4, 5)$. Круг радиуса r и два кольца с внешними радиусами $3r$ и $5r$ заштрихованы. В круге радиуса $5r$ поставлена точка, положение которой равновозможно в этом круге. Определить вероятность попадания данной точки:*

1. *в круг радиуса $2r$;*

2. *в заштрихованную область.*

Упражнение 1.35. *На отрезке AB длиной i независимо одна от другой поставлены две точки L и M, положение каждой из которых равновозможно на AB. Найти вероятность того, что точка L ближе к точке M, чем к точке A.*

Упражнение 1.36. *На отрезке длиной l независимо одна от другой поставлены две точки, положение каждой из которых равновозможно на этом отрезке. Определить вероятность того, что из трех получившихся частей одного отрезка можно построить треугольник.*

Упражнение 1.37. *К автобусной остановке через каждые четыре минуты подходит автобус линии A и через каждые шесть минут – автобус линии B. Интервал времени между моментами прихода автобуса линии A и ближайшего следующего автобуса линии B равновозможен в пределах от нуля до четырех минут. Определить вероятность того, что:*

1. *первый подошедший автобус окажется автобусом линии A;*

2. *автобус какой-либо линии подойдет в течение двух минут.*

Упражнение 1.38. *Два парохода должны подойти к одному и тому же причалу. Время прихода обоих пароходов независимо и равновозможно в течение данных суток. Определить вероятность того, что одному из пароходов придется ожидать освобождения причала, если время стоянки первого парохода один час, а второго – два часа.*

Упражнение 1.39. *Стержень длиной $l = 200$ мм ломается на части, причем положение каждой точки излома не зависит от положения других точек излома и равновозможно по всей длине стержня. Определить вероятность того, что хотя бы одна часть стержня между точками излома будет не более 10 мм, если точек излома:*

1. *две;*
2. *три.*

Упражнение 1.40. *Определить вероятность того, что корни:*

1. *квадратного $x^2 + 2\xi x + \eta = 0$;*
2. *кубического $x^3 + \xi x + \eta = 0$ уравнений вещественны, если равновозможны значения коэффициентов (ξ, η) в прямоугольнике $[-n, n] \times [-m, m]$;*
3. *квадратного уравнения $x^2 + 2\xi x + \eta = 0$ будут положительными.*

Упражнение 1.41. *Случайная точка A имеет равномерное распределение в квадрате со стороной 1. Найти вероятности следующих событий:*

1. *расстояние от точки A до фиксированной стороны квадрата не превосходит x;*
2. *расстояние от точки A до ближайшей стороны квадрата не превосходит x;*
3. *расстояние от точки A до центра квадрата не превосходит x;*
4. *расстояние от точки A до фиксированной вершины квадрата не превосходит x.*

Упражнение 1.42. *Случайная точка A имеет равномерное распре деление в прямоугольнике со сторонами 1 в 2. Найти вероятности следующих событий:*

1. *расстояние от A до ближайшей стороны прямоугольника не превосходит x;*
2. *расстояние от A до любой стороны прямоугольника не превосходит x;*
3. *расстояние от A до диагоналей прямоугольника не превосходит x;*

Упражнение 1.43. *Случайная точка A имеет равномерное распределение в квадрате со стороной a. Найти вероятность того, что расстояние от A до ближайшей стороны квадрата меньше, чем расстояние от A до ближайшей диагонали квадрата.*

Упражнение 1.44. *Случайная точка X равномерно распределена в правильном треугольнике с вершинами $(a, 0)$, $(-a, 0)$, $(0, a\sqrt{3})$. Найти вероятность того, что квадрат с центром X и сторонами длины b, параллельными осям координат, целиком содержится в этом треугольнике.*

Упражнение 1.45. *В круге радиуса R случайно проводится хорда. Обозначим ξ ее длину. Найти вероятность $Q_x = \{\xi > x\}$, если середина хорды равномерно распределена в круге. Вычислить вероятности Q_R и $Q_{R\sqrt{3}}$, того, что длина хорды больше стороны правильного вписанного шестиугольника и треугольника соответственно. Результат зависит от того, как понимать слово "случайно".*

Упражнение 1.46. *Решить предыдущую задачу, если направление хорды задано, а ее середина равномерно распределена на диаметре, перпендикулярном ее направлению.*

Упражнение 1.47. *Решить предыдущую задачу, если один конец хорды закреплен, а другой равномерно распределен на окружности.*

Упражнение 1.48. *Отрезок длины a_1+a_2, поделен на две части длины a_1, и a_2 соответственно. n точек последовательно бросаются наудачу на отрезок. Найти вероятность того, что ровно m из n точек попадут на часть отрезка длины a_1.*

Упражнение 1.49. *В единичный квадрат наудачу брошена точка. Какова вероятность того, что точка будет удалена от центра квадрата на расстояние меньше, чем 1/3, если известно, что от каждой из сторон квадрата она удалена больше, чем на 1/6?*

1.7.3 Условная вероятность. Независимость. Формула полной вероятности и теорема Байеса

Упражнение 1.50. *Брошено две игральные кости. Предполагается, что все комбинации выпавших очков равновероятны. Найти условную вероятность того, что выпали две пятерки, если известно, что сумма выпавших очков делится па пять.*

Упражнение 1.51. *Среди 25 экзаменационных билетов 5 "хороших". Два студента по очереди берут по одному билету. Найти вероятность того, что:*

1. *первый студент взял "хороший" билет;*
2. *второй студент взял "хороший" билет;*
3. *оба студента взяли "хорошие" билеты.*

Решить задачу для произвольных количества экзаменационных билетов n, количества "хороших" билетов m и j-го студента.

Упражнение 1.52. *Из урны, содержащей 3 белых шара, 5 черных и 2 красных, два игрока поочередно извлекают по одному шару без возвращения. Выигрывает тот, кто первым вынет белый шар. Если появляется красный шар, то объявляется ничья. Пусть $A_1 = \{$выигрывает игрок, начавший игру$\}$, $A_2 = \{$выигрывает второй участник$\}$, $B = \{$игра закончилась вничью$\}$. Найти $\mathrm{P}(A_1)$, $\mathrm{P}(A_2)$, $\mathrm{P}(B)$.*

Упражнение 1.53. *Случайная точка (ξ_1, ξ_2) имеет равномерное распределение в квадрате $\{(\xi_1, \xi_2) : 0 \leqslant x_1, x_2 \leqslant 1\}$ При каких значениях r независимы события $A_r = \{|\xi_1 - \xi_2| \geqslant r\}$ и $B_r = \{|\xi_1 + \xi_2| \leqslant 3r\}$?*

Упражнение 1.54. *События A и B независимы. Являются ли независимыми события:*

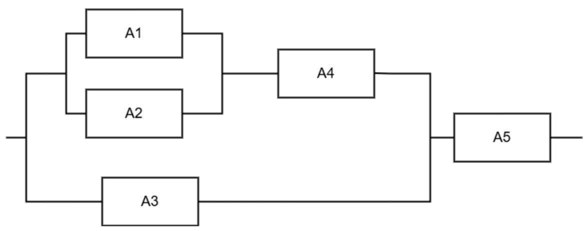
1. *A и $\overline{B}$;*
2. *$\overline{A}$ и $\overline{B}$.*

Упражнение 1.55. *Упрощенная система контроля изделий состоит из двух независимых проверок. В результате k-й проверки $(k = 1, 2)$ изделие, удовлетворяющее стандарту, отбраковывается с вероятностью β_k, а бракованное изделие принимается с вероятностью α_k. Изделие принимается, если оно прошло обе проверки. Найти вероятности событий:*

1. *бракованное изделие будет принято;*

2. *изделие, удовлетворяющее стандарту, будет отбраковано.*

Упражнение 1.56. *Электрическая цепь составлена из элементов A_k, $k = 1, 2, \ldots, 5$, по схеме, приведенной на рисунке. При выходе из строя любого элемента цепь в месте его включения разрывается. Вероятность выхода из строя за данный период элемента A_k равна P_k, $k = 1, \ldots, 5$. Предполагается, что элементы выходят или не выходят из строя независимо друг от друга. Найти вероятность события $C = \{$за рассматриваемый период по цепи может проходить ток$\}$.*



Упражнение 1.57. *В первой урне находятся 1 белый и 9 черных шаров, а во второй – 1 черный и 5 белых шаров. Из каждой урны по схеме случайного выбора без возвращения удалили по одному шару, а оставшиеся шары ссыпали в третью урну. Найти вероятность того, что шар, вынутый из третьей урны, окажется белым.*

Упражнение 1.58. *В пункте проката имеется 10 телевизоров, для которых вероятность исправной работы в течение месяца равна 0.90 и 5 телевизоров с аналогичной вероятностью, равной 0.95. Найти вероятность того, что два телевизора, взятые наудачу в пункте проката, будут работать исправно в течение месяца.*

Упражнение 1.59. *При рентгеновском обследовании вероятность обнаружить заболевание туберкулезом у больного туберкулезом равна $1 - \beta$. Вероятность принять здорового человека за больного равна α. Пусть доля больных туберкулезом по отношению ко всему населению равна γ.*

1. *Найти условную вероятность того, что человек здоров, если он был признан больным при обследовании.*

2. *Вычислить найденную условную вероятность при следующих числовых значениях: $1 - \beta = 0.9, \alpha = 0.01, \gamma = 0.001$.*

Упражнение 1.60. *Отдел технического контроля (ОТК) проводит сортировку выпускаемых заводом приборов. Каждый прибор независимо от остальных имеет дефекты с вероятностью p. При проверке в ОТК наличие дефектов обнаруживается с вероятностью α; кроме того, с вероятностью β исправный прибор при проверке может вести себя как дефектный. Все приборы, у которых при проверке обнаружены отклонения от стандарта, бракуются. Найти вероятность q_0 того, что незабракованный прибор имеет дефекты, и вероятность q_1 того, что забракованный прибор имеет дефекты. При каких условиях $q_0 > q_1$?*

Упражнение 1.61. *Имеется n урн одинакового состава: N белых и M черных шаров. Из первой урны во вторую перекладывается один шар, затем из второй урны в третью перекладывается один шар и т. д. Из последней урны извлекается один шар. Найти вероятность того, что он белый.*

Упражнение 1.62. *Урна содержит один шар, про который известно, что он либо белый, либо черный с одинаковыми вероятностями. В урну кладут один белый шар и затем наудачу извлекают один шар. Он оказался белым. Какова вероятность, что оставшийся в урне шар - белый?*

Упражнение 1.63. *Брошено три игральных кости. Найти вероятность того, что на всех костях выпала шестерка, если известно, что*

1. *на одной кости выпало 6 очков;*
2. *на первой кости выпало 6 очков;*
3. *на двух костях выпали "шестерки";*
4. *по крайней мере на двух костях выпало одинаковое число очков;*
5. *на всех костях выпало одинаковое число очков;*
6. *по крайней мере на одной кости выпало 6 очков.*

Упражнение 1.64. *Группа студентов, сдающая экзамен, состоит из 5 отличников, 10 хороших студентов и 15 слабых студентов; отличник всегда получает оценку "отлично", хороший студент – "отлично" и "хорошо" с равными вероятностями, слабый студент – "хорошо", "удовлетворительно" и "неудовлетворительно" с равными вероятностями. Какова вероятность, что наугад вызванный студент получит оценку*

1. *"отлично";*
2. *"хорошо"?*

1.7.4 Схема Бернулли и связанные с ней предельные теоремы

Упражнение 1.65. *При передаче сообщения вероятность искажения одного знака равна 1/10. Каковы вероятности того, что сообщение из 10 знаков:*

1. *не будет искажено;*
2. *содержит ровно 3 искажения;*
3. *содержит не более трех искажений?*

Упражнение 1.66. *Найти вероятность того, что в $2n$ испытаниях схемы Бернулли с вероятностью успеха p и неудачи $q = 1 - p$ появится $m + n$ успехов и все испытания с четными номерами закончатся успехом.*

Упражнение 1.67. *В одном из матчей на первенство мира по шахматам ничьи не учитывались, и игра шла до тех пор, пока один из участников матча не набирал 6 очков (выигрыш - 1 очко, проигрыш и ничья - 0 очков). Считая участников матча одинаковыми по силе, а результаты отдельных игр независимыми, найти вероятность того, что при таких правилах в момент окончания матча проигравший набирает k очков, $k \in \{0, 1, \dots, 5\}$.*

Упражнение 1.68. *Обрабатываемые на станке детали сортируются по размерам на две группы. Каждая очередная деталь независимо от предыдущих с равными вероятностями попадает в первую или вторую группу. Пусть в начале смены для каждой группы деталей приготовлено по ящику емкости r. Какова вероятность того, что в момент, когда очередную деталь будет некуда класть, в другом ящике будет m деталей?*

Упражнение 1.69. *По каналу связи передается 1000 знаков. Каждый знак может быть искажен независимо от остальных с вероятностью 0.005. Найти приближенное значение вероятности того, что будет искажено не более трех знаков.*

Упражнение 1.70. *В таблице случайных чисел цифры сгруппированы по две. Найти приближенное значение вероятности того, что среди 100 пар набор 09 встретится не менее двух раз.*

Упражнение 1.71. *Найти приближенное значение вероятности того, что число "девяток" среди 10000 случайных чисел заключено между 940 и 1060.*

Упражнение 1.72. *Из таблицы случайных чисел отбирают числа, делящиеся на 3, до тех пор, пока не наберется 1025 таких чисел. Найти приближенное значение вероятности того, что потребуется таблица, содержащая не меньше 2500 чисел.*

Упражнение 1.73. *Театр, вмещающий 1000 человек, имеет два разных входа. Около каждого. из входов имеется свой гардероб. Сколько мест должно быть в каждом из гардеробов для того, чтобы в среднем в 99 случаях из 100 все зрители могли раздеться в гардеробе того входа, через который они вошли? Рассмотреть два случая:*

1. *зрители приходят парами;*
2. *зрители приходят поодиночке.*

Предположить, что входы зрители выбирают с равными вероятностями.

Упражнение 1.74. *Пусть* η_N *- суммарное число появлений "5" и "6" в* N *бросаниях игральной кости. При* $N = 1800$ *найти вероятность того, что* $\eta_N \geqslant 620$

Упражнение 1.75. *Две монеты подбрасывают 4800 раз. Найти приближенное значение вероятности того, что событие "герб-герб" появится меньше 1140 раз.*

Упражнение 1.76. *Вероятность попадания в цель при одном выстреле равна 0,01. Найти приближенное значение вероятности того, что при 100 выстрелах будет не больше трех попаданий.*

Упражнение 1.77. *Из урны, содержащей 1 белый и 4 черных шара, по схеме случайного выбора с возвращением проводят 2500 извлечений шаров. Найти приближенное значение вероятности того, что число появлений белого шара заключено между 480 и 540.*

Упражнение 1.78. *На одной странице 2400 знаков. При типографском наборе вероятность искажения одного знака равна 1/800. Найти приближенное значение вероятности того, что на странице не менее двух опечаток.*

Упражнение 1.79. *При прохождении одного порога байдарка не получает повреждений с вероятностью* p_1*, полностью ломается с вероятностью* p_2*, получает серьезное повреждение с вероятностью* p_3*,* $(p_1 + p_2 + p_3 = 1)$*. Два серьезных повреждения приводят к полной поломке. Найти вероятность того, что при прохождении n порогов байдарка не будет полностью сломана.*

Упражнение 1.80. *Сообщения, передаваемые по каналу связи, составляются из трех знаков* A, B, C*. Из-за помех каждый знак принимается правильно с вероятностью 0.6 и принимается ошибочно за любой из двух других знаков с вероятностью 0.2. Для увеличения вероятности правильного приема каждый знак передается 5 раз. За переданный знак принимается знак, который чаще всего встречается в принятой пятерке знаков. Если наиболее частых знака два, то из них выбирается равновероятно один. Найти вероятность правильного приема знака при указанном способе передачи.*

Упражнение 1.81. *Испытания в полиномиальной схеме с исходами 1, 2, 3, имеющими вероятности* p_1, p_2, p_3 *соответственно, заканчиваются, когда впервые не появится исход 3. Найти вероятность того, что испытания закончатся исходом 1.*

Упражнение 1.82. *Пункт A нужно связать с 10 абонентами пункта B. Вероятность того, что в любой фиксированный момент времени абоненту потребуется линия связи, равна 0.2, причем эта потребность не зависит от потребностей других абонентов. Какое минимальное количество каналов необходимо для того, чтобы можно было в любой момент с вероятностью 0.99 обслужить всех абонентов?*

Упражнение 1.83. *Определить число* n *повторных независимых испытаний, которое нужно произвести для того, чтобы вероятнейшее число появлений события равнялось 20, если вероятность появления этого события при каждом испытании равна 0.8.*

Упражнение 1.84. *Определить вероятность необходимости повторного голосования при выборе* l *человек, если голосуют* n *человек; вероятность быть вычеркнутым для каждого из* k *кандидатов одинакова и равна* p*, а для выбора кандидата необходимо получить большинство голосов. Повторное голосование не производится, если больше половины голосов получает каждый из* l *кандидатов, а любой из оставшихся* $k-1$ *кандидатов - не более половины голосов из* n.

Упражнение 1.85. *В электропоезд, состоящий из шести вагонов, садится двенадцать человек, причем выбор каждым пассажиром вагона равновозможен. Определить вероятность того, что:*

1. *в каждый вагон вошло по два человека;*
2. *в один вагон никто не вошел, в другой – вошел один человек, в два вагона – по два человека, а в оставшиеся два вагона соответственно три и четыре человека.*

Упражнение 1.86. *Урна содержит* l *белых,* m *черных и* n *красных шаров. Производится* $l_1 + m_1 + n_1$ *извлечений шаров по одному с возвращением каждого извлеченного шара. Определить вероятность того, что будет извлечено:*

1. *сначала* l_1 *белых, затем* m_1 *черных и, наконец,* n_1 *красных шаров;*
2. l_1 *белых,* m_1 *черных и* n_1 *красных шаров, причем все шары одного цвета появляются подряд, но последовательность цветов может быть любой;*
3. l_1 *белых,* m_1 *черных и* n_1 *красных шаров в любой последовательности.*

Упражнение 1.87. *Определить вероятность того, что при* n *бросаниях монет герб появится нечетное число раз.*

Глава 2

Случайные величины и их числовые характеристики

2.1 Случайные величины и их распределения. Функция распределения случайной величины

В определении *вероятностного пространства* мы не накладывали никаких ограничений на природу *множества элементарных событий* Ω, однако удобно каждому элементарному событию сопоставлять число (практически все явления в том или ином виде описываются с помощью чисел, в программировании мы в некотором смысле все кодируем в виде чисел). Это сопоставление и есть *случайная величина*. Формализуем это понятие.

Определение 2.1. *Пусть* $(\Omega, \mathcal{F})$ *– множество элементарных событий и сигма-алгебра событий соответственно.* ***Случайная величина*** X *– измеримая функция из* $(\Omega, \mathcal{F})$ *в* $(\mathbb{R}, \mathcal{B})$*, то есть для всякого* $B \in \mathcal{B} \Rightarrow X^{-1}(B) \in \mathcal{F}$ *(*$X^{-1}(B)$ *– прообраз множества* B*), где* $\mathcal{B}$ *– борелевская сигма-алгебра.*

Случайные величины обозначаются последними заглавными латинскими буквами U, V, W, X, Y, Z или греческими ξ, η, ν. Мы будем, как правило, использовать заглавные латинские буквы.

При сопоставлении элементарного события и числа естественным образом переносится вероятность, то есть если числу x соответствует элементарное событие ω, вероятность которого равняется p, то вероятность "выпадения" числа x тоже равняется p. Набор таких вероятностей есть *распределение случайной величины.* Так же формализуем данное понятие.

Определение 2.2. *Пусть* $(\Omega, \mathcal{F})$ *– множество элементарных событий и сигма-алгебра событий соответственно и* X *– случайная величина.* ***Распределением случайной величины*** *называется вероятность* P_X*, заданная на* $\mathcal{B}$ *и определяемая как*

$$\mathrm{P}_X(B) = \mathrm{P}(\{\omega : X(\omega) \in B\}).$$

Последнее равенство мы часто будем кратко писать как $\mathrm{P}(X \in B)$.

Таким образом, распределение случайной величины – вероятность на $\mathbb{R}$, которую достаточно задать на борелевских множествах (на всех возможных интервалах), однако, это оказывается не всегда оказывается удобно. Для устранения данного неудобства оказывается удобной *функция распределения случайной величины*

Определение 2.3. *Пусть $(\Omega, \mathcal{F})$ – множество элементарных событий и сигма- алгебра событий соответственно, и X – случайная величина.* ***Функцией распределения случайной величины*** *называется функция вещественной переменной F_X, которая определяется для каждого $x \in \mathbb{R}$ как*

$$F_X(x) = \mathrm{P}(X \in (-\infty, x]).$$

Последнюю вероятность мы будем обозначать как $\mathrm{P}(X \leqslant x)$.

Замечание. *Ввиду того, что $F_X(x)$ определена как вероятность, то и само распределение P_X является вероятностной мерой и соответствует всем ее свойствам.*

Сформулируем и докажем свойства функции распределения.

Теорема 2.1. *Пусть X – случайная величина и F – её функция распределения. Справедливы следующие свойства:*

1. $F(x) \to 0$ *при* $x \to -\infty$, $F(x) \to 1$ *при* $x \to +\infty$.
2. F *монотонно возрастает (не убывает).*
3. F *непрерывна справа, то есть* $F(x) \to F(x_0)$ *при* $x \to x_0 + 0$.

Доказательство. Доказательство будет опираться на изученные нами ранее свойства вероятности (см. теорему **??**). Пусть $A_n = (-\infty, -x_n]$, $\{x_n\}$ монотонно неограниченно возрастает. Заметим, что $A_{n+1} \subset A_n$ и $\bigcap\limits_n A_n = \varnothing$. Тогда имеем согласно непрерывности сверху

$$F(-x_n) = \mathrm{P}_X(A_n) \to \mathrm{P}_X(\varnothing) = 0, \quad n \to \infty.$$

Положим $B_n = (-\infty, x_n]$. Очевидно, что $B_n \subset B_{n+1}$ и $\bigcup\limits_n B_n = \mathbb{R}$ и в следствие непрерывности снизу имеем

$$F(x_n) = \mathrm{P}_X(B_n) \to \mathrm{P}_X(\mathbb{R}) = 1, \quad n \to \infty.$$

Неубывание следуем из монотонности вероятности. Действительно, при $x_1 \leqslant x_2$

$$F(x_1) = \mathrm{P}_X((-\infty, x_1]) \leqslant \mathrm{P}_X((-\infty, x_2]) = F(x_2).$$

Покажем непрерывность справа. Пусть $\{x_n\}$ монотонно убывает и $x_n \to x_0 + 0$ и $C_n = (-\infty, x_n]$. Ясно, что $\bigcap\limits_n C_n = (-\infty, x_0]$. Тогда имеем

$$F(x_n) \to F(x_0), \quad x_n \to x_0 + 0.$$

ч.т.д.

Замечание. *При нашем определении функция распределения не является непрерывной слева. Действительно, пусть x_n монотонно возрастает и стремится к x_0*

$$F(x_0) - F(x_n) = \mathrm{P}_X((-\infty, x_0]) - \mathrm{P}_X((-\infty, x_n])$$
$$= \mathrm{P}_X((-\infty, x_n]) - \mathrm{P}_X((-\infty, x_n]) + \mathrm{P}_X((x_n, x_0]) = \mathrm{P}_X((x_n, x_0]) \to \mathrm{P}(X = x_0).$$

Замечание. *Кроме того, как видно,* $\mathrm{P}(X \in (a, b]) = F(b) - F(a)$.

Замечание. *Иногда функция распределения определяется как $F(x) = \mathrm{P}(X < x)$. В этом случае непрерывность справа изменяется на непрерывность слева.*

Сформулируем и докажем еще одно свойство функции распределения.

Теорема 2.2. *Пусть F – функция, удовлетворяющая сформулированными в предыдущей теореме. Тогда существует вероятностное пространство $(\Omega, \mathcal{F}, \mathrm{P})$ и заданная на нём случайная величина X такая, что $F_X = F$.*

Доказательство. Сначала построим требуемое вероятностное пространство. Положим $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}$. Осталось задать вероятность. Как мы знаем согласно теореме о продолжении меры достаточно определить вероятность на алгебре $\mathbf{B}$ такой, что $\sigma(\mathbf{B}) = \mathcal{B}$. В нашем случае $\mathbf{B}$ – алгебра, порождённая интервалами вида $(a, b]$. Произвольный элемент данной алгебры представим в виде конечного объединения непересекающихся интервалов вида $(a, b]$. Тогда зададим вероятность как

$$\mathrm{P}(B) = \sum_{j=1}^{n} F(b_j) - F(a_j), \quad B = \bigsqcup_{j=1}^{n} (a_j, b_j],$$

где $a_j \leqslant b_j$ (числа a_j, b_j могут быть бесконечны).

Неотрицательность, конечная аддитивность и нормированность вероятности P следует немедленно из определения и условия теоремы. Осталось доказать счётную аддитивность, непрерывность снизу или непрерывность сверху. Мы покажем непрерывность сверху, то есть из $B_{n+1} \subset B_n$ и $B = \bigcap\limits_n B_n$ должно следовать $\mathrm{P}(B_n) \to \mathrm{P}(B)$ или, что равносильно, $\mathrm{P}(B_n\overline{B}) \to 0$ при $n \to \infty$. Действительно,

$$\mathrm{P}(B_n) = \mathrm{P}(B_n B) + \mathrm{P}(B_n\overline{B}) = \mathrm{P}(B) + \mathrm{P}(B_n\overline{B}).$$

Поэтому, не умаляя общности, будем полагать, что $\bigcap\limits_n B_n = \varnothing$. Сначала предположим, что все множества B_n принадлежат замкнутому интервалу $[-N, N]$. Поскольку B_n состоят из конечного числа сумм непересекающихся интервалов вида $(a, b]$ и в силу непрерывности функции F справа

$$\mathrm{P}((a', b]) = F(b) - F(a') \to F(b) - F(a) = \mathrm{P}((a, b]), \quad a' \to a + 0,$$

то для каждого B_n найдётся $A_n \in \mathbf{B}$ такое, что его замыкание $\mathrm{cl}\ A_n \subset B_n$ (*замыкание* – множество всех предельных точек) и

$$\mathrm{P}(B_n) - \mathrm{P}(A_n) = \mathrm{P}(B_n \setminus A_n) \leqslant \varepsilon 2^{-n},$$

где ε – произвольная положительная константа.

Так как $\bigcap\limits_n B_n = \varnothing$, то $\bigcap\limits_n \mathrm{cl}\ A_n = \varnothing$. Кроме того, имеет место

$$\bigcup_n \{[-N, N] \setminus \mathrm{cl}\ A_n\} = [-N, N] \setminus \bigcap_n \mathrm{cl}\ A_n = [-N, N],$$

то есть набор множеств $\{[-N, N] \setminus \mathrm{cl}\ A_n\}$ – открытое покрытие множества $[-N, N]$ (*открытое* множество – дополнение замкнутого, *замкнутое* множество – совпадающее со своим замыканием множество). Тогда существует n_0 такое, что

$$\bigcup_{n=1}^{n_0} \{[-N, N] \setminus \mathrm{cl}\ A_n\} = [-N, N],$$

что влечёт $\bigcap\limits_{n=1}^{n_0} \mathrm{cl}\, A_n = \varnothing$ (это можно обосновать леммой Гейне-Бореля или тем, что множество $[-N, N]$ – компакт). Далее вследствие вложенности множеств B_n и полученных ранее соотношений имеем

$$\mathrm{P}(B_{n_0}) = \mathrm{P}\left(B_{n_0} \setminus \bigcap_{k=1}^{n_0} A_k\right) + \mathrm{P}\left(\bigcap_{k=1}^{n_0} A_k\right) = \mathrm{P}\left(B_{n_0} \setminus \bigcap_{k=1}^{n_0} A_k\right)$$
$$\leqslant \mathrm{P}\left(\bigcap_{k=1}^{n_0} B_k \setminus A_k\right) \leqslant \sum_{k=1}^{n_0} \mathrm{P}(B_k \setminus A_k) \leqslant \sum_{k=1}^{n_0} 2^{-k}\varepsilon \leqslant \varepsilon.$$

В силу произвольности ε заключаем, что $\mathrm{P}(B_n) \to 0$ при $n \to \infty$. Далее откажемся от предположения ограниченности множеств B_n. В силу поведения функции F на бесконечности и её неубывания для произвольного положительного ε можно выбрать $N > 0$ такое, что

$$\mathrm{P}([-N, N]) > 1 - \varepsilon/2.$$

Далее заметим, что

$$\mathrm{P}(B_n) = \mathrm{P}(B_n \cap [-N, N]) + \mathrm{P}(B_n \cap \overline{[-N, N]}) \leqslant \mathrm{P}(B_n \cap [-N, N]) + \varepsilon/2,$$

а случай $B_n \cap [-N, N]$ уже разобран выше, что и завершает теорему. *ч.т.д.*

Подытожим сказанное. Во-первых, *функция распределения* по сути задаётся *распределением*. Во-вторых, как правило, *случайная величина* нам интересна не природой соответствия между множеством элементарных событий Ω и $\mathbb{R}$, а её *распределением*, поэтому доказанная нами теорема позволяет утверждать, что *функция распределения* задаёт *распределение* и для описания *распределения* достаточно предъявить *функцию распределения*. То, что случайная величина X имеет распределение P_X и функцию распределения F_X, мы будем обозначать как $X \sim \mathrm{P}$ и $X \sim F_X$.

2.2 Типы случайных величин и распределений

В этом параграфе мы рассмотрим типы случайных величин и распределений. Также затронем некоторые семейства распределений. Запись $X \sim D$ означает, что случайная величина X имеет некоторое распределение из семейства D.

2.2.1 Дискретные случайные величины и распределения

Определение 2.4. *Случайная величина X (распределение случайной величины) называется* ***дискретной****, если существует такое не более чем счётное множество E, что $\mathrm{P}_X(E) = \mathrm{P}(X \in E) = 1$.*

Соответствующее распределение P_X случайной величины X будем называть ***дискретным****.*

По сути написанное выше означает, что случайная величина X принимает значения $\{x_i\}_{i=1}^N$, $N \leqslant \infty$, (считаем, что $x_1 < x_2 < \ldots$) с соответствующими вероятностями $\{p_i\}_{i=1}^N$, то есть $\mathrm{P}(X = x_i) = p_i$.

Также несложно заметить, что функция распределения дискретного закона кусочно-постоянна и имеет скачки в точках x_i, равные p_i (т.е. представляет собой ступенчатую функцию). Более конкретно имеем

$$F(t) = \sum_k p_k \mathbb{1}(x_k \leqslant t) = \begin{cases} 0, \ t < x_1, \\ p_1, \ x_1 \leqslant t < x_2, \\ p_1 + p_2, \ x_2 \leqslant t < x_3, \\ \ldots, \\ p_1 + p_2 + \ldots + p_i, \ x_i \leqslant t < x_{i+1}, \\ \ldots \end{cases},$$

где $\mathbb{1}(A) = 1$, если условие A истина, $\mathbb{1}(A) = 0$ в противном случае.

Рассмотрим несколько важных дискретных распределений.

Пример 2.1. ***Вырожденное распределение*** *в точке c: Случайная величина X принимает только одно значение c, то есть* $\mathrm{P}(X = c) = 1$. *Будем обозначать $I(c)$ или I_c, $c \in \mathbb{R}$.*

Следующие два распределения встречались в схеме Бернулли.

Пример 2.2. ***Распределение Бернулли***. *Случайная величина X принимает значения 1 и 0 с вероятностями p и $q = 1-p$ соответственно, то есть* $\mathrm{P}(X = 1) = p$, $\mathrm{P}(X = 0) = q$. *Обозначение:* $\mathrm{Bern}(p)$, $p \in (0, 1)$.

Пример 2.3. ***Биномиальное распределение***. $\mathrm{P}(X = k) = C_n^k p^k q^{n-k}$, $k \in \{0, \ldots, n\}$. *Обозначение:* $\mathrm{Bin}(n; p)$, $n \in \mathbb{N}$, $p \in (0, 1)$.

Соответственно распределение Бернулли описывает одно независимое испытание с вероятностью успеха p, а биномиальное – n независимых испытаний с вероятностью успеха p. Далее рассмотрим еще два распределения, связанных со схемой бернуллиевских испытаний.

Пример 2.4. ***Отрицательное биномиальное распределение***. *Пусть $k + r$ – номер испытания, в котором случился r-ый успех, $r \in \mathbb{N}$. Тогда соответствующая случайная величина X равняется k. Более формально $X = \min\{n \in \mathbb{N} : S_n = r\} - r$ (здесь S_n – количество успехов в n испытаниях). Напишем вероятность:*

$$\mathrm{P}(X = k) = pC_{k+r-1}^{r-1} p^{r-1} q^{k+r-1-(r-1)} = C_{k+r-1}^{r-1} p^r q^k = \frac{\Gamma(k+r)}{k!\Gamma(r)} p^r q^k,$$

где $\Gamma(x)$ – гамма-функция. Мы провели рассуждения для натуральных r, но с помощью гамма функции мы можем рассматривать распределение и при неотрицательном r. Будем обозначать $\mathrm{NB}(r, p)$, $r > 0$, $p \in (0, 1)$.

Пример 2.5. ***Геометрическое распределение***. *Частный случай при $r = 1$ отрицательного биномиального распределения будем называть геометрическим и обозначать* $\mathrm{Geom}(p)$, $p \in (0, 1)$. *Соответствующую случайную величину мы можем интерпретировать как количество неудач до первого успеха.*

Замечание. *Возможна и другая интерпретация для геометрического распределения: номер первого успеха, то есть* $\mathrm{P}(X = k) = pq^{k-1}$, $k \in \mathbb{N}$. *Если не оговорено иное, то мы будем придерживаться первой интерпретации.*

Следующее распределение встречалось нам в теореме Пуассона. Оно важно в теории массового обслуживания и в теории случайных процессов пуассоновские процессы также существенны. (*вставить ссылки на учебники/монографии* (ссылок всё ещё нет))

Пример 2.6. ***Распределение Пуассона***. $\mathrm{P}(X=k)=e^{-\lambda}\frac{\lambda^k}{k!}$, $k\in\{0,1,\dots,\}$. *Обозначение:* $\mathrm{Pois}(\lambda)$.

Пример 2.7. ***Гипергеометрическое распределение***. *Обозначим его как $HG(M,N,K)$, и рассмотрим на конкретном примере. Пусть M – общее количество деталей в урне, $N\in[1...M]$ – количество хороших деталей и $K\in[1...M]$ – количество деталей, которые мы вытаскиваем (без возвращения). С какой вероятностью мы вытащим g хороших деталей?*

Пусть X – количество хороших деталей, которые попали в выборку объема K

$$\mathrm{P}(X=j)=\frac{C_N^jC_{M-N}^{K-j}}{C_M^K},$$
$$\begin{cases}0\leqslant j\leqslant N,\\ 0\leqslant K-j\leqslant M-N\end{cases}=\begin{cases}0\leqslant j\leqslant N,\\ N-M+K\leqslant j\leqslant K\end{cases}$$
$$\Leftrightarrow\quad \max(0;N-M+K)\leqslant j\leqslant\min(N;K)$$

Замечание. *Рассмотрим ситуацию, когда $M\to\infty$ и $\frac{N}{M}\to P$. Распишем биномиальные коэффициенты:*

$$\frac{N!(M-N)!K!(M-K)!}{j!(N-j)!(K-j)!(M-N-K+j)!M!}$$
$$=C_K^j\frac{N(N-1)\cdots(N-j+1)\cdot(M-K)(M-K-1)\cdots(M-N-K+j+1)}{M(M-1)\cdots(M-N+1)}$$
$$\longrightarrow C_K^JP^j(1-P)^{K-j}$$

Что, вообще-то говоря, есть биномиальное распределение.

Пример. ***Задача о перекладывании шариков***. *Есть N урн, в каждой урне лежит W белых и B черных шаров. Мы перекладываем из первой урны шарик по вторую, затем из второй в третью, и так далее, пока из последней не вытянем шарик. Какова вероятность, что в конце мы вытащим белый шарик? Ответом будет $\frac{W}{W+B}$.*

А теперь усложним задачу – будем перекладывать по k шариков из урны в урну. Но, внезапно, ответ будет точно таким же! Давайте рассмотрим, почему так, и для начала, введем следующие случайные величины:

Пусть A_n – искомое событие (из урны вытащили белый шар), X_1 – количество белых шаров, переложенных из первой урны во вторую, ..., X_i – аналогично для урн i в $i+1$.

X_1 имеет гипергеометрическое распределение:

$$X_1\approx HG(W+B,W,K)$$
$$X_2\approx HG(W+B+K,W+M_1,K),\ \textit{если } X_1=M_1$$
$$\cdots$$
$$X_i\approx HG(W+B+K,W+M_{i-1},K),\ \textit{если } X_{i-1}=M_{i-1}$$

Что такое $P(A_n)$? Распишем следующую сумму по формуле полной вероятности:

$$P(A_n)=\sum_{cond}\mathrm{P}(A_n|X_{n-1}=M_{n-1},...,X_1=M_1)\cdot\mathrm{P}(X_{n-1}=M_{n-1},...,X_1=M_1),$$
$$cond=M_1\in(\max(0,K-B),\min(W,K)),...,M_i\in(\max(0,M_{i-1}-B),\min(K,W+M_{i-1}))$$

Давайте последовательно разбираться с этими вероятностями:

$$\mathrm{P}(A_n|X_{n-1}=M_{n-1},...,X_1=M_1)=\frac{W+M_{n-1}}{W+B+K}$$

$$\mathrm{P}(X_{n-1}=M_{n-1},...,X_1=M_1)=\mathrm{P}(X_1=M_1)\cdots\mathrm{P}(X_{n-1}=M_{n-1}|X_{n-2}=M_{n-2},...,X_1=M_1)=$$
$$=\frac{C_W^{M_1}C_B^{K-M_1}}{C_{W+B}^K}\cdot\frac{C_{W+M_1}^{M_2}C_{B+K-M_1}^{K-M_2}}{C_{W+B+K}^K}\cdots\frac{C_{W+M_{n-2}}^{M_{n-1}}C_{B+K-M_{n-2}}^{K-M_{n-1}}}{C_{W+B+K}^K}$$

Вернемся к сумме:

$$\sum_{M_1...M_{n-2}}\left((...)\cdot\sum_{M_{n-1}}\mathrm{P}(M_{n-1}|M_{n-2})\frac{W+M_{n-1}}{W+B+K}\right)=\sum_{M_1...M_{n-2}}\left((...)\cdot\frac{W+K\frac{W+M_{n-2}}{W+B+K}}{W+B+K}\right)=$$
$$=\sum_{M_1...M_{n-3}}\left((...)\cdot\frac{W+K\frac{W+K\frac{W+M_{n-3}}{W+B+K}}{W+B+K}}{W+B+K}\right)$$

Получилась цепная дробь. Рассмотрим ее последний этаж:

$$\frac{W+K\frac{W+M_1}{W+B+K}}{W+B+K}\xrightarrow[\sum_{M_1}]{}\frac{W+K\frac{W+K\frac{W}{W+B}}{W+B+K}}{W+B+K}=\frac{W+K\frac{W}{W+B}}{W+B+K}$$

Таким образом, заметим, что эта дробь схлопнулась в дробь из числителя. По итогу, вся цепная дробь поведет себя аналогично, и в конце мы получим:

$$\frac{W+K\frac{W}{W+B}}{W+B+K}=\frac{W}{W+B}$$

2.2.2 Абсолютно непрерывные случайные величины и распределения

Следующий важный класс распределений – абсолютно непрерывные.

Определение 2.5. *Случайная величина X (распределение случайной величины) называется **абсолютно непрерывной**, если существует $p:\mathbb{R}\to[0,+\infty)$ и интегрируемая на $\mathbb{R}$ (относительно меры Лебега), для которой $\mathrm{P}(X\in B)=\int\limits_B p(x)\,dx$. Функцию p будем называть **плотностью**.*

Отметим простейшие свойства, которые вытекают непосредственно из определения, ввиду их важности в виде теоремы.

Теорема 2.3. *Справедливы следующие соотношения:*

- $F(t)=\int\limits_{-\infty}^{t}p(x)\,dx.$
- F *– непрерывная.*

- $\mathrm{P}(X = c) = 0$.

- $F'(x) = p(x)$ *почти всюду.*

Замечание. *Пусть* λ *– мера Лебега и* ν *– некоторая другая мера ,* ν *задана на* $\mathcal{B}$ *(борелевских множествах),* $\nu(\mathbb{R}) < +\infty$*. Тогда,* ν *– абсолютно непрерывна относительно* λ *если выполнена импликация* $\lambda(A) = 0 \Rightarrow \nu(A) = 0$

Для абсолютно непрерывных мер имеет место

Теорема. ***Теорема Радона-Никодима*** *(б/д).*
Пусть ν *абсолютно непрерывна относительно* λ*, тогда*

$$\exists f(x) : \nu(B) = \int\limits_B f(x)\lambda(dx)$$

Замечание. *Об аналогии между функции вероятностей в дискретном случае и плотностью.*

В дискретном случае мы задавали функцию вида $\mathrm{P}(X = x_k) = p_k$*, а в абсолютно непрерывном случае задаем плотность. Рассмотрим вероятность* $\mathrm{P}(X \in [x_0, x_0 + h])$*:*

$$\mathrm{P}(X \in [x_0, x_0 + h]) = F(x_0 + h) - F(x_0) = F'(x_0)h + o(h) = p(x_0)h + o(h), \quad h \to 0.$$

То есть вероятность того, что случайная величина X *принимает значения из окрестности точки* x_0 *примерно равняется* $p(x_0)h$*, то есть плотность можно рассматривать как аналог функции вероятности, хоть она формально может быть больше единицы, главное чтобы* $\int\limits_{\mathbb{R}} p(x)\ dx = 1$*.*

Замечание. *О значениях, которые принимает случайная величина.*

В дискретном случае всё просто: случайная величина принимает значения x_1*,* x_2*, ... В непрерывном случае дело обстоит несколько иначе. Пусть* E *– множество, на котором плотность строго больше нуля. Тогда несложно заметить, что*

$$1 = \int\limits_E p(x)\ dx + \int\limits_{\overline{E}} p(x)\ dx = \int\limits_E p(x)\ dx + \int\limits_{\overline{E}} 0\ dx = \int\limits_E p(x)\ dx,$$

то есть видим, что случайная величина принимает только значения из множества E *и заведомо не может принимать значения к дополнению к нему.*

Далее рассмотрим важнейшие примеры абсолютно непрерывных распределений.

Пример 2.8. ***Непрерывное равномерное распределение****. Будем обозначать* $\mathrm{U}[a, b]$*,* $a < b$*. Плотность имеет вид*

$$p(t) = \frac{1}{b - a}\mathbb{1}(t \in [a, b]),$$

то есть вся вероятностная масса сосредоточена на $[a, b]$ *равномерно и* a*,* b *– левая и правая границы диапазона значений соответственно. Тогда функция распределения имеет вид*

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

Пусть $X \sim \mathrm{U}[a,b]$, $Y = cX + d$. Найдем распределение Y. Для определенности будем считать, что $c > 0$. Напишем функцию распределения для Y. Имеем

$$F_Y(t) = \mathrm{P}(Y \leqslant t) = \mathrm{P}(cX + d \leqslant t) = \mathrm{P}\left(X \leqslant \tfrac{t-d}{c}\right)$$

$$= \begin{cases} 0, \ \frac{t-d}{c} < a, \\ \frac{\frac{t-d}{c} - a}{b-a}, \ \frac{t-d}{c} \in [a,b], \\ 1, \ \frac{t-d}{c} > b, \end{cases} = \begin{cases} 0, \ t < ca + d, \\ \frac{t-d-ca}{cb-ca}, \ t \in [ac+d, cb+d], \\ 1, \ t > cb + d. \end{cases}$$

Видим, что получившаяся функция соответствует равномерному на $[ac+d, cb+d]$ распределению, то есть $Y \sim \mathrm{U}[ac+d, cb+d]$. Случай $c < 0$ рассматривается аналогично. В частности, если $W \sim \mathrm{U}[0,1]$, то $(b-a)W + a \sim \mathrm{U}[a,b]$.

Далее рассмотрим фундаментальное распределение, с которым мы встретились в теоремах Муавра-Лапласа – нормальным.

Пример 2.9. ***Нормальное распределение***. *Плотность нормального распределения $\mathcal{N}(\mu, \sigma^2)$ с параметром сдвига $\mu \in \mathbb{R}$ и масштабирующим параметром $\sigma^2 > 0$ (их вероятностный смысл выясним в дальнейшем) имеет вид:*

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\tfrac{(x-\mu)^2}{2\sigma^2}\right\}.$$

Нормальное распределение $\mathcal{N}(0,1)$ будем называть стандартным, его плотность и функция распределения имеют вид

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\tfrac{x^2}{2}\right\}, \quad \Phi(t) = \int_{-\infty}^{t} \phi(x)\ dx,$$

функция Φ табулирована в литературе и в вероятностно-статистических программах и пакетах имеется её численная реализация.

Замечание. *Помимо Φ существуют еще несколько специальных функций, похожие и связанные с Φ. Например:*

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{t^2}{2}} dt \quad \text{– \textit{функция Лапласа}}$$

$$erf(x) = \frac{2}{\sqrt{\pi}} \int\limits_0^x e^{-t^2} dt \quad \text{– \textit{функция ошибок}}$$

$$\Phi(x) = \frac{1}{2} + \Phi_0(x) = \frac{1}{2}\left(1 + erf\left(\frac{x}{\sqrt{2}}\right)\right), \quad \Phi_0(-x) = -\Phi_0(x)$$

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{t^2}{2}} dt = \left|\sqrt{2}u = t\right| = \frac{1}{\sqrt{\pi}} \int\limits_0^{\frac{x}{\sqrt{2}}} e^{-u^2} du = \frac{1}{2} erf\left(\frac{x}{\sqrt{2}}\right)$$

Пусть $X \sim \mathcal{N}(\mu, \sigma^2)$ и $Y = aX + b$. Для определенности будем считать, что $a > 0$. Найдем функцию распределения для Y. Имеем

$$F_Y(t) = \mathrm{P}(Y \leqslant t) = \mathrm{P}(aX + b \leqslant t) = \mathrm{P}\left(X \leqslant \tfrac{t-b}{a}\right) = \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{\frac{t-b}{a}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}\ dx$$

$$= \begin{bmatrix} x = \frac{y-b}{a} \\ y \in (-\infty, t] \\ dx = \frac{dy}{a} \end{bmatrix} = \frac{1}{\sqrt{2\pi}a\sigma} \int\limits_{-\infty}^{t} \exp\left\{-\frac{(y-b-a\mu)^2}{2a^2\sigma^2}\right\} dy.$$

Несложно заметить, что получили функцию распределения нормального закона с параметром сдвига $a\mu+b$ *и масштабирующим параметром* $a^2\sigma^2$*, то есть* $Y \sim \mathcal{N}(a\mu+b, a^2\sigma^2)$*. В частности, если* $U \sim \mathcal{N}(\mu, \sigma^2)$*, то случайная величина* $\frac{U-\mu}{\sigma}$ *будет иметь стандартное нормальное распределение, и наоборот, если* $W \sim \mathcal{N}(0,1)$*, то* $\sigma W + \mu \sim \mathcal{N}(\mu, \sigma^2)$*.*

Замечание. *Мы используем для нормального распределения параметризацию вида* $\mathcal{N}(\mu, \sigma^2)$*, то есть в качестве второго параметра указываем именно* σ^2*, однако в литературе и ПО может использоваться и иная параметризация: вместо* σ^2 *в качестве второго параметра указывается просто* σ*.*

Пример 2.10. ***Распределение Коши:*** $\mathrm{Cauchy}(x_0, \gamma)$*, где* x_0 *– параметр сдвига,* $\gamma > 0$ *– масштабирующий параметр. Плотность имеет вид*

$$p(x) = \frac{1}{\pi\gamma} \cdot \frac{1}{1 + \left(\frac{x-x_0}{\gamma}\right)^2}.$$

Тогда для функции распределения имеем

$$F(t) = \frac{1}{\pi\gamma}\int\limits_{-\infty}^{t} \frac{dx}{1 + \left(\frac{x-x_0}{\gamma}\right)^2} = \frac{1}{\pi}\mathrm{arctg}\left(\frac{t-x_0}{\gamma}\right) + \frac{1}{2}$$

Пример 2.11. ***Экспоненциальное распределение:*** $\mathrm{Exp}(\lambda)$*,* $\lambda > 0$*. Плотность имеет вид*

$$p(x) = \lambda e^{-\lambda x}\mathbb{1}(x \geqslant 0),$$

и функция распределения

$$F(t) = (1 - e^{-\lambda t})\mathbb{1}(x \geqslant 0).$$

Пример 2.12. ***Гамма-распределение:*** $\Gamma(k, \lambda)$*.* $k, \lambda > 0$*.*

$$p(x) = \frac{\lambda^k x^{k-1}}{\Gamma(k)} e^{-\lambda x}\mathbb{1}(x \geqslant 0)$$

Пример 2.13. ***Распределение Эрланга:*** $\Gamma(k, \lambda)$*. Тоже самое, что и гамма-распределение, однако* $k \in \mathbb{N}$*.*

$$p(x) = \frac{\lambda^k x^{k-1}}{(k-1)!} e^{-\lambda x}\mathbb{1}(x \geqslant 0)$$

Такой знаменатель возникает ввиду выполнения следующего свойства гамма-функции: $\Gamma(k) = (k-1)!$.

Замечание. *В экспоненциальном и гамма-распределением вместо* λ*, которое интерпретируется в приложениях как интенсивность (англ. rate), можно использовать в качестве параметра* λ^{-1}*, интерпретируемое как масштабирующий параметр (англ. scale).*

Стоит отметить, что мы затронули далеко не все абсолютно непрерывные случайные величины, однако для первого знакомства ограничимся этим.

В заключении данного пункта сформулируем и докажем теорему, которая в некоторых случаях может быть полезной для нахождения плотности преобразования случайной величины.

Теорема 2.4. *Пусть X – абсолютно непрерывная случайная величина с плотностью p_X, $Y = g(X)$, где $g : \mathbb{R} \to \mathbb{R}$, $g \in C^1(\mathbb{R})$ и g – строго монотонна. Тогда плотность p_Y может быть вычислена по формулам*

$$p_Y(y) = p_X(g^{-1}(y)) \left|\frac{d(g^{-1}(y))}{dy}\right| = p_X(g^{-1}(y)) \left|\frac{1}{d(g(x))/(dx)}\right|_{x=g^{-1}(y)}.$$

Третье равенство получено из второго с помощью теоремы о производной обратной функции.

Доказательство. Сначала пусть g строго возрастает. Тогда для функции распределения Y ввиду обратимости функции g имеем

$$F_Y(y) = \mathrm{P}(Y \leqslant y) = \mathrm{P}(g(X) \leqslant y) =$$

$$\mathrm{P}(X \leqslant g^{-1}(y)) = F_X(g^{-1}(y)) = \int\limits_{-\infty}^{g^{-1}(y)} p_X(t)\, dt.$$

Для нахождения плотности продифференцируем полученное выражение

$$\frac{d(F_Y(y))}{dy} = p_Y(y) = p_X(g^{-1}(y))\frac{d(g^{-1}(y))}{dy}.$$

Случай строгого убывания рассматривается аналогично, нужно лишь учесть, что при переходе к g^{-1} внутри выражения под вероятностью поменяется знак неравенства и пределы интегрирования будут от $g^{-1}(y)$ до $+\infty$. С учетом этого получится точно такое же выражение, только со знаком минус. Вспоминая, что производная строго возрастающей функции положительна, а строго убывающей – отрицательна, получаем единообразное выражение для двух случаев. *ч.т.д.*

Замечание. *Приведём формулировку этой же теоремы в случае произвольной g.*

$$p_Y(y) = \sum_{k=1}^{n(y)} p_X(g_k(y))|[g_k(y)]'_y|$$

где $n(y)$ – количество решений уравнения $g(x) = y$, а $g_k(y)$ – k-ый корень уравнения.

2.2.3 Сингулярные случайные величины и распределения

Определение 2.6. *Функцию распределения F (и соответствующие ей распределение и случайную величину) будем называть* ***сингулярной****, если $F \in C(\mathbb{R})$ и мера (Лебега) точек роста равняется нулю.*

x_0 – точка роста функции распределения F, если $F(x_0+\varepsilon) - F(x_0-\varepsilon) > 0$ для любого $\varepsilon > 0$.

В качестве примера сингулярной функции распределения можно привести лестницу Кантора. Введем функцию K(x) следующим образом. На первом этапе возьмем интервал (0, 1) и разделим его на 3 равные части. На среднем интервале значение K(x) $= \frac{1}{2}$. На следующем этапе возьмем левый и правый интервалы и повторим процесс разбиения и определения функции на средних интервалах, и так дальше. Рассмотрим получившуюся функцию K. Она определена и непрерывна почти всюду, кроме канторова множества(мера Лебега которого равна 0).

Также сформулируем термин, который будет по сути будет означать множество значений, которые может принимать случайная величина X.

Определение 2.7. ***Носителем распределения*** *будем называть наименьшее по включению замкнутое множество E, для которого* $\mathrm{P}(X \in E) = 1$.

Таким образом, для дискретных распределения носителем является множество вида $\{x_k\}_{k=1}^N$, а для абсолютно непрерывных распределений множество, где плотность строго больше нуля (вместе с границами данного множества).

Замечание. *Внимательный и любознательный читатель заметит, что выше мы сформулировали понятие носителя меры и для абсолютно непрерывного случая носитель вероятностной меры и носитель (как функции) плотности совпадают.*

В заключении параграфа приведем вероятностный вариант теоремы Лебега о разложении функций распределений.

Теорема 2.5. *Пусть $F(x)$ – функция распределения. Тогда* $\exists!\ c_1, c_2, c_3 \in [0, 1]$*, для которых* $c_1 + c_2 + c_3 = 1$ *и*

$$F(x) = c_1F_1(x) + c_2F_2(x) + c_3F_3(x),$$

где $F_1(x)$, $F_2(x)$, $F_3(x)$ – функции распределения дискретной, абсолютно непрерывной и сингулярных случайных величин.

Таким образом, у произвольной функции распределения есть дискретная составляющая, абсолютно непрерывная и сингулярная.

Далее для краткости мы будем *абсолютно непрерывные* случайные величины и распределения называть просто *непрерывными*, так как в дальнейшем изложении сингулярные величины максимум будут только упоминаться.

2.3 Случайные векторы и многомерные распределения

В этом параграфе мы рассмотрим случайные векторы и многомерные распределения.

Определение 2.8. ***Случайным вектором*** *называется $X = (X_1, \ldots, X_n)^T$, если каждая компонента X_i, $i \in \{1, \ldots, n\}$ является случайной величиной.*

Определение 2.9. ***Распределением*** *вектора $X = (X_1, \ldots, X_n)^T$ называется величина* $\mathrm{P}_X(B_1, \ldots, B_n) = \mathrm{P}(X_1 \in B_1, \ldots, X_n \in B_n)$.

Замечание. *Также используется термин совместное распределение случайных величин X_1, …, X_n, которое по сути означает то же самое.*

Аналогично одномерному случаю обобщается и *функция распределения*

Определение 2.10. ***Функцией распределения*** *случайного вектора $X = (X_1, \ldots, X_n)^T$ называется функция* $F(t_1, \ldots, t_n) = \mathrm{P}(X_1 \leqslant t_1, \ldots, X_n \leqslant t_n)$.

Рассмотрим функцию распределения случайного вектора несколько более подробно. В одномерном случае мы показали, что $\mathrm{P}(X \in (a, b]) = F(b) - F(a)$. Посмотрим, что будет в многомерном случае. Рассмотрим векторы a, b, для которых $a_1 \leqslant b_1$, …, $a_n \leqslant b_n$ (в данном параграфе будем кратко писать $a \leqslant b$). Введем разностный оператор Δ_{a_i,b_i}, действующий следующим образом:

$$\Delta_{a_i,b_i}F(x_1, \ldots, x_i, \ldots, x_n) = F(x_1, \ldots, x_{i-1}, b_i, x_{i+1}, \ldots, x_n) - F(x_1, \ldots, x_{i-1}, a_i, x_{i+1}, x_n).$$

Для простоты (в n-мерном случае выкладки будут такими же, только более громоздкими) рассмотрим вероятность попадания двумерного случайного вектора $X = (X_1, X_2)^T$ в множество $(a_1, b_1] \times (a_2, b_2]$. Имеем

$$\begin{gathered}
\mathrm{P}(X_1 \in (a_1, b_1], X_2 \in (a_2, b_2]) = \mathrm{P}(X_1 \in (a_1, b_1], X_2 \leqslant b_2) - \mathrm{P}(X_1 \in (a_1, b_1], X_2 \leqslant a_2) \\
= \mathrm{P}(X_1 \leqslant b_1, X_1 \leqslant b_2) - \mathrm{P}(X_1 \leqslant a_1, X_2 \leqslant b_2) - \mathrm{P}(X_1 \leqslant b_1, X_2 \leqslant a_2) + \mathrm{P}(X_1 \leqslant a_1, X_2 \leqslant a_2) \\
= F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) = \Delta_{a_2,b_2} F(b_1, x_2) - \Delta_{a_2,b_2} F(a_1, x_2) \\
= \Delta_{a_2,b_2}(F(b_1, x_2) - F(a_1, x_2)) = \Delta_{a_1,b_1} \Delta_{a_2,b_2} F(x_1, x_2).
\end{gathered}$$

Соответственно в общем случае имеем

$$\mathrm{P}(X_1 \in (a_1, b_1], \ldots, X_n \in (a_n, b_n]) = \Delta_{a_1,b_1} \ldots \Delta_{a_n,b_n} F(x_1, \ldots, x_n).$$

Для функции распределения случайного вектора справедливы свойства, аналогичные одномерному случаю. Сформулируем их в виде теорем.

Теорема 2.6. *Пусть $F(x_1, \ldots, x_n)$ – функция распределения случайного вектора X. Тогда*

1. *Если $a \leqslant b$, то $\Delta_{a_1,b_1} \ldots \Delta_{a_n,b_n} F(x_1, \ldots, x_n) \geqslant 0$*
2. *$F(+\infty, \ldots, +\infty) = 1$, $F(y) \to 0$ при $y \to \hat{y}$, если хотя бы одна компонента $\hat{y}$ равняется $-\infty$.*
3. *F непрерывна справа.*

Теорема 2.7. *Пусть F – функция, удовлетворяющая свойствам, перечисленным в предыдущей теореме. Тогда существует вероятностное пространство $(\Omega, \mathcal{A}, \mathrm{P})$ и случайный вектор X, для которого F является его функцией распределения.*

Идея доказательств сформулированных выше теорем такая же, как и в одномерном случае, поэтому доказательства мы опустим.

Как и в одномерном случае мы рассматривали разные типы распределений, в многомерном случае мы так же рассмотрим дискретные и непрерывные распределения. Начнем с дискретных

Определение 2.11. *Случайный вектор $X = (X_1, \ldots, X_n)$ и соответствующее ему распределение называются **дискретными**, если существует не более чем счетное множество E, для которого $\mathrm{P}(X \in E) = 1$.*

Рассмотрим пример, связанный с независимыми испытаниями

Пример 2.14. ***Полиномиальное распределение***. *Пусть проведены n независимых испытаний, в каждом из которых возможны m исходов с вероятностями $p_1, \ldots, p_m$ соответственно ($\sum p_i = 1$). Пусть $S_{n,i}$ – количество исходов типа i в n испытаниях. Тогда несложно заметить, что*

$$\mathrm{P}(S_{n,1} = n_1, \ldots, S_{n,m} = n_m) = \frac{n!}{n_1! \cdot \ldots \cdot n_m!} p_1^{n_1} \cdot \ldots \cdot p_m^{n_m},$$
$$n_1 + \ldots + n_m = m.$$

Таким образом, мы описали полиномиальное распределение вектора $S_n = (S_{n,1}, \ldots, S_{n,m})^T$. Будем использовать обозначение $\mathrm{Poly}(n, p)$.

Продолжим с абсолютно непрерывных.

Определение 2.12. *Случайный вектор $X = (X_1, \dots, X_n)^T$ и соответствующее ему распределение будем называть* ***абсолютно непрерывными****, если существует $p : \mathbb{R}^n \to [0, +\infty)$ интегрируемая (по Лебегу) и для которой*

$$\mathrm{P}(X \in B) = \int\limits_B p(x)\ dx,$$

B – борелевское множество в $\mathbb{R}^n$.

Как и в одномерном случае, для краткости *абсолютно* непрерывные распределения будем называть просто *непрерывными*.

Далее рассмотрим *многомерное нормальное распределение*.

Пример 2.15. ***Многомерное нормальное распределение.***

1. *Вектор $X = (X_1, \dots, X_n)$ будем называть* ***стандартным гауссовским****, если его плотность имеет вид*

$$p(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right\} = \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{x^T x}{2}\right\}.$$

В этом случае будем писать $X \sim \mathcal{N}(0, E_n)$, где 0 – n-мерный нулевой вектор, E_n – единичная матрица размера $n \times n$.

2. *Пусть $Y = AX + b$, где X – стандартный гауссовский вектор, A – матрица $m \times n$, $b \in \mathbb{R}^m$. В этом случае будем писать $Y \sim \mathcal{N}(b, AA^T)$.*

3. *Рассмотрим общий случай. Пусть $Y \sim \mathcal{N}(\mu, \Sigma)$, где $\Sigma = \Sigma^T$ и $\Sigma > 0$, то есть $x^T \Sigma x \geqslant 0\ \forall x \in \mathbb{R}^m$ (неотрицательная определенность матрицы Σ).*

Запишем для матрицы Σ спектральное разложение (мы это можем сделать, так как $\Sigma = \Sigma^T$):

$$\Sigma = U \Lambda U^T,$$

где Λ – матрица, на диагонали которой расположены собственные числа матрицы Σ, а на остальных позициях – нули; U – матрица из ортонормированных собственных векторов (поэтому $U^T = U^{-1}$). Пусть $\sqrt{\Lambda}$ – матрица, у которой на диагонали стоят корни из собственных чисел (они неотрицательные, так как $\Sigma \geqslant 0$), а на остальных местах – нули (несложно проверить, что $\sqrt{\Lambda}^2 = \Lambda$). Тогда перепишем полученное равенство следующим образом:

$$\Sigma = (U\sqrt{\Lambda}U^T)(U\sqrt{\Lambda}U^T).$$

Далее пусть $X \sim \mathcal{N}(0, E_n)$. Тогда $Y = (U\sqrt{\Lambda}U^T)X + \mu$. Действительно,

$$(U\sqrt{\Lambda}U^T)(U\sqrt{\Lambda}U^T)^T = U\sqrt{\Lambda}U^T U\sqrt{\Lambda}U^T = U\Lambda U^T = \Sigma,$$

то есть полученное согласуется с написанным выше.

4. *Пусть теперь $\Sigma > 0$. Тогда существует плотность, имеющая вид*

$$p(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma}} \exp\left\{-\tfrac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\}.$$

В заключение пункта приведем формулировку теоремы о плотности функции от непрерывного случайного вектора.

Теорема 2.8. *Пусть* $X = (X_1, \ldots, X_n)$ *случайный вектор с плотностью* $p_X(t_1, \ldots, t_n)$, $g : \mathbb{R}^n \to \mathbb{R}^n$, $Y = g(X)$, *причем* $g \in C^1(\mathbb{R}^n)$ *и обратима. Тогда плотность* p_Y *случайного вектора* Y *имеет вид*

$$p_Y(y_1, \ldots, y_n) = p_X(g^{-1}(y))|\det D \cdot g^{-1}(y)| = p_X(g^{-1}(y))|\det D \cdot g(x)|^{-1}_{x=g^{-1}(y)}$$

где $D(.)$ *– матрица Якоби.*

2.4 Независимые случайные величины и векторы

Вспомним, что события A_1, ..., A_n независимы, если для любого поднабора A_{i_1}, ..., A_{i_m}

$$\mathrm{P}(A_{i_1} \ldots A_{i_m}) = \mathrm{P}(A_{i_1}) \cdot \ldots \cdot \mathrm{P}(A_{i_m}).$$

Очень похожее определение имеет место и случайных величин

Определение 2.13. *Случайные величины* X_1, ..., X_n *-* ***независимые****, если для любых* $B_1, \ldots, B_n \in \mathcal{B}$

$$\mathrm{P}(X_1 \in B_1 \cdots X_n \in B_n) = \mathrm{P}(X_1 \in B_1) \cdot \ldots \cdot \mathrm{P}(X_n \in B_n).$$

Случайные величины $\{X_i\}_{i=1}^{\infty}$ *независимы, если для любого* $n \in \mathbb{N}$ *случайные величины являются независимыми.*

Замечание. *Напомним, что* $\varnothing, \mathbb{R} \in \mathcal{B}$ *являются независимыми, поэтому на самом деле сформулированное определение не противоречит определению независимых событий.*

Определение *независимых случайных векторов* практически точно такое же, только там $B_1, \ldots B_n$ уже n-мерные борелевские множества.

Сформулируем критерий независимости в терминах функций распределения.

Теорема 2.9. $X_1, \ldots, X_n$ *- независимые* $\Leftrightarrow F_{X_1, \cdots, X_n}(t_1, \cdots, t_n) = \prod_{i=1}^{n} F_{X_i}(t_i) \forall t_1, \cdots, t_n \in \mathbb{R}$

Далее рассмотрим критерий независимости дискретных случайных величин.

Теорема 2.10. *Пусть* X_1, ..., X_n *– дискретные.*

$X_1, \ldots, X_n$ *- независимые* $\Leftrightarrow \mathrm{P}(X_1 = x_{1,i_1}, X_2 = x_{2,i_2}, \ldots, X_n = x_{n,i_n}) = \prod_{j=1}^{n} \mathrm{P}(X_j = x_{j,i_j}) \forall x_{i,k}$, *где* $x_{i,k}$ *– значение случайной величины* X_i

Доказательство. $\Rightarrow$

Очевидно из определения, так как одноточечные множества являются борелевскими

$\Leftarrow$

Заметим, что

$$\mathrm{P}(X_1 \in B_1, \cdots, X_n \in B_n) = \mathrm{P}(X_1 \in \{x_{1,i_1}\}, \ldots, X_n \in \{x_{n,i_n}\})$$
$$= \sum_{i_1, \ldots, i_n} \mathrm{P}(X_1 = x_{1,i_1}, \ldots, X_n = x_{n,i_n}) = \sum_{i_1, \ldots, i_n} \mathrm{P}(X_1 = x_{1,i_1}) \cdot \ldots \cdot \mathrm{P}(X_n = x_{n,i_n})$$
$$= \mathrm{P}(X_1 \in B_1) \cdot \ldots \cdot \mathrm{P}(X_n \in B_n).$$

ч.т.д.

Также рассмотрим критерий независимости непрерывных величин.

Теорема 2.11. *X_1, ...,X_n - абсолютно непрерывные. Тогда случайные величины X_1, ...,X_n независимы тогда и только тогда, когда их совместная плотность $p(x_1, \ldots, x_n)$ равняется произведению одномерных плотностей $p_i(x_i)$, то есть*

$$p(x_1, \ldots, x_n) = p_1(x_1) \cdot \ldots \cdot p_n(x_n).$$

Доказательство. Необходимость: Пусть нам известно, что величины X_1, ...,X_n независимы. Тогда их совместная функция распределения равняется произведению одномерных функций распределений. Запишем их сразу как интеграл от плотности.

$$\int\limits_{t_1 \leqslant x_1, \ldots, t_n \leqslant x_n} p(t_1, \ldots, t_n)\ dt_1 \ldots dt_n = \int\limits_{-\infty}^{x_1} p_1(t_1)\ dt_1 \ldots \int\limits_{-\infty}^{x_n} p_n(t_n)\ dt_n.$$

Дифференцируя обе части по x_1, x_2, ..., x_n получаем требуемое. Для доказательства достаточности нужно обе части проинтегрировать по x_1, ..., x_n по промежуткам. *ч.т.д.*

Замечание. *Мы стандартный гауссовский вектор $X = (X_1, \ldots, X_n)^T$ определили через плотность. Заметим, что она имеет вид*

$$p(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{x^T x}{2}} = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_j^2}{2}},$$

откуда следует, что компоненты стандартного гауссовского вектора независимы и распределены согласно стандартному нормальному закону.

2.4.1 Суммы независимых случайных величин

В теории вероятностей и её приложениях часто встречаются суммы независимых случайных величин, поэтому мы рассмотрим некоторое количество примеров, связанных с ними.

Пример 2.16. *Бернуллиевское и биномиальное распределения.*

Пусть $X_1, X_2 \sim \mathrm{Bern}(p)$ и $S_2 = X_1 + X_2$. Очевидно, что S_2 может равняться нулю, единице или двойке. Напишем вероятности для них

$$\mathrm{P}(S_2 = 0) = \mathrm{P}(X_1 = 0, X_2 = 0) = (1-p)^2,$$
$$\mathrm{P}(S_2 = 1) = \mathrm{P}(X_1 = 0, X_2 = 1) + \mathrm{P}(X_1 = 1, X_2 = 0) = 2p(1-p),$$
$$\mathrm{P}(S_2 = 2) = \mathrm{P}(X_1 = 1, X_2 = 1) = p^2.$$

Видим, что $S_2 \sim \mathrm{Bin}(2, p)$.

Покажем теперь, что $S_n = X_1 + \ldots + X_n \sim \mathrm{Bin}(n, p)$, где $X_i \sim \mathrm{Bern}(p)$ и они независимы. Будем показывать по индукции. База очевидна. Далее предположим, что $S_{n-1} \sim \mathrm{Bin}(n-1, p)$ и покажем, что $S_n \sim \mathrm{Bin}(n, p)$. Для этого заметим, что $S_n = S_{n-1} + X_n$. Распишем вероятности $S_n = 0$ и $S_n = n$ с учетом независимости S_{n-1} и X_n:

$$\mathrm{P}(S_n = 0) = \mathrm{P}(S_{n-1} = 0, X_n = 0) = (1-p)^n,$$
$$\mathrm{P}(S_n = n) = \mathrm{P}(S_{n-1} = n-1, X_n = 1) = p^n.$$

При $k \in \{1, \ldots, n-1\}$ *для* $\mathrm{P}(S_n = k)$ *имеем*

$$\mathrm{P}(S_n = k) = \mathrm{P}(S_{n-1} = k-1, X_n = 1) + \mathrm{P}(S_{n-1} = k, X_n = 0)$$
$$= \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1}(1-p)^{n-k} p + \frac{(n-1)!}{k!(n-1-k)!} p^k (1-p)^{n-1-k}(1-p)$$
$$= \frac{(n-1)!}{(k-1)!(n-1-k)!} p^k (1-p)^{n-k} \left(\frac{1}{n-k} + \frac{1}{k} \right) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k},$$

то есть действительно $S_n \sim \mathrm{Bern}(p)$.

Пример 2.17. *Суммирование независимых пуассоновских случайных величин.*

Пусть $X_1 \sim \mathrm{Pois}(\lambda_1), X_2 \sim \mathrm{Pois}(\lambda_2)$ *и они независимы. Несложно заметить, что сумма принимает целые неотрицательные значения. Найдём распределение суммы. Для* $k \in \{0, 1, \ldots\}$ *имеем по формуле полной вероятности*

$$\mathrm{P}(X_1 + X_2 = k) = \sum_{j=0}^{\infty} \mathrm{P}(X_1 + X_2 = k | X_2 = j)\, \mathrm{P}(X_2 = j) = \sum_{j=0}^{k} \mathrm{P}(X_1 = k-j)\, \mathrm{P}(X_2 = j)$$
$$= \sum_{j=0}^{k} e^{-\lambda_1} \frac{\lambda_1^{k-j}}{(k-j)!} e^{-\lambda_2} \frac{\lambda_2^j}{j!} = e^{-(\lambda_1+\lambda_2)} \frac{1}{k!} \sum_{j=0}^{k} \frac{k!}{(k-j)!j!} \lambda_1^{k-j} \lambda_2^j = e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1+\lambda_2)^k}{k!},$$

получаем, что $X_1 + X_2 \sim \mathrm{Pois}(\lambda_1 + \lambda_2)$

Пример 2.18. *Суммирование независимых случайных величин, имеющих геометрическое распределение.*

Пусть $X_1, X_2 \sim \mathrm{Geom}(p)$ *и они независимы. Для распределения суммы независимых случайных величин* $X_1 + X_2$ *имеем*

$$\mathrm{P}(X_1 + X_2 = k) = \sum_{j=0}^{k} \mathrm{P}(X_1 = k-j)\, \mathrm{P}(X_2 = j) = \sum_{j=0}^{k} (1-p)^{k-j} p (1-p)^j p$$
$$= (k+1)(1-p)^k p^2 = C_{k+2-1}^{2-1} p^2 (1-p)^k,$$

то есть получили, что $X_1 + X_2 \sim \mathrm{NB}(2, p)$.

В общем случае можно показать (это мы сделаем несколько ниже), что сумма n *величин имеет* $\mathrm{NB}(n, p)$.

Пример 2.19. *Суммирование непрерывных независимых случайных величин.*

Пусть X_1, X_2 *независимы и имеют плотности соответственно* p_1 *и* p_2. *Найдём функцию распределения суммы* $X_1 + X_2$. *Имеем*

$$F_{X_1+X_2}(t) = \mathrm{P}(X_1 + X_2 \leqslant t) = \iint\limits_{x_1+x_2 \leqslant t} p_{X_1,X_2}(x_1, x_2)\, dx_1 dx_2$$
$$= \int_{-\infty}^{+\infty} p_1(x_1) \left(\int_{-\infty}^{t-x_1} p_2(x_2)\, dx_2 \right) dx_1.$$

Найдем плотность суммы

$$p_{X_1+X_2}(t) = F'_{X_1+X_2}(t) = \int_{-\infty}^{+\infty} p_1(x) p_2(t-x)\, dx.$$

Также заметим, что у нас получилась формула, похожая на дискретный случай. Действительно, если X_1 и X_2 независимы и целочислены, то

$$\mathrm{P}(X_1 + X_2 = k) = \sum_j \mathrm{P}(X_1 = j)\,\mathrm{P}(X_2 = k - j)$$

Полученные формулы называются свёртками.

На основе полученной формулы рассмотрим еще два примера.

Пример 2.20. *Суммирование независимых случайных величин, имеющих экспоненциальное распределение.*

Пусть X, Y – независимы и имеют экспоненциальное распределение с параметром $\lambda > 0$. Тогда плотность p суммы $X + Y$ имеет вид

$$p(t) = \int\limits_{-\infty}^{+\infty} p_1(x)p_2(t-x)\ dx = \int\limits_{0}^{+\infty} \lambda^2 e^{-\lambda x - \lambda(t-x)} \mathbb{1}(t - x \geqslant 0)\ dx$$

$$= \lambda^2 e^{-\lambda t} \int\limits_0^t\ dx = \frac{\lambda^2 t}{1!} e^{-\lambda t}, \quad t > 0.$$

То есть получили, что $X_1 + X_2 \sim \Gamma(2, \lambda)$. В случае n слагаемых получится $\Gamma(n, \lambda)$ или распределение Эрланга порядка n.

Пример 2.21. *Пусть $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$, $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$ и они независимы. Найдём плотность суммы*

$$p(t) = \frac{1}{2\pi\sigma_x\sigma_y} \int\limits_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2}\left(\tfrac{(x-\mu_x)^2}{\sigma_x^2} + \tfrac{(t-x-\mu_y)^2}{\sigma_y^2}\right)\right\}\ dx$$

$$= \begin{bmatrix} u = x - \mu_x, \\ du = dx \end{bmatrix} = \frac{1}{2\pi\sigma_x\sigma_y} \int\limits_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2}\left(\tfrac{u^2}{\sigma_x^2} + \tfrac{(t-u-\mu_x-\mu_y)^2}{\sigma_y^2}\right)\right\}\ dx.$$

Введем обозначение $q = t - \mu_x - \mu_y$ и рассмотрим несколько подробнее выражение внутри экспоненты:

$$\frac{\sigma_y^2 u^2 + \sigma_x^2 (q-u)^2}{\sigma_x^2\sigma_y^2} = \frac{(\sigma_x^2 + \sigma_y^2)u^2 - 2\sigma_x^2 q u + \sigma_x^2 q^2}{\sigma_x^2\sigma_y^2}$$

$$= \frac{\left(\sqrt{\sigma_x^2 + \sigma_y^2}\,u - \frac{\sigma_x^2 q}{\sqrt{\sigma_x^2+\sigma_y^2}}\right)^2 + \sigma_x^2 q^2 - \frac{\sigma_x^4 q^2}{\sigma_x^2+\sigma_y^2}}{\sigma_x^2\sigma_y^2}$$

$$= \frac{\left(\sqrt{\sigma_x^2 + \sigma_y^2}\,u - \frac{\sigma_x^2 q}{\sqrt{\sigma_x^2+\sigma_y^2}}\right)^2}{\sigma_x^2\sigma_y^2} + \frac{q^2}{\sigma_x^2 + \sigma_y^2}.$$

Далее в интеграле положим $\sqrt{\sigma_x^2 + \sigma_y^2}\,u = w$. Следовательно, $du = (\sigma_x^2 + \sigma_y^2)^{-1/2} dw$ и, возвращаясь к исходным обозначениям, мы имеем

$$p(t) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{\sigma_x^2 + \sigma_y^2}} \exp\left\{-\tfrac{(t-\mu_x-\mu_y)^2}{2(\sigma_x^2+\sigma_y^2)}\right\} \int\limits_{-\infty}^{+\infty} \exp\left\{-\frac{\left(w - \frac{\sigma_x^2 q}{\sqrt{\sigma_x^2+\sigma_y^2}}\right)^2}{2\sigma_x^2\sigma_y^2}\right\}\ dw$$

$$= \frac{1}{\sqrt{2\pi(\sigma_x^2 + \sigma_y^2)}} \exp\left\{-\tfrac{(t-\mu_x-\mu_y)^2}{2(\sigma_x^2+\sigma_y^2)}\right\}.$$

Итого получили, что $X+Y \sim \mathcal{N}(\mu_x+\mu_y, \sigma_x^2+\sigma_y^2)$.

2.5 Моделирование случайных величин и распределений

Как правило в высокоуровневых языках программирования функции типа **random** генерируют независимые случайные величины из равномерного распределения (на самом деле генерируется псевдослучайная величина, то есть на самом деле она детерминирована, но поведение функции похоже на "случайную" и даже может пройти соответствующие статистические тесты). Поэтому будем считать, что мы умеем генирировать независимые случайные величины из равномерного на [0; 1] распределения. Будем обозначать вызов данной функции **random**.

Теорема 2.12. *Пусть* $X \sim \mathrm{U}[0,1]$, *функция распределения непрерывной случайной величины, которую мы хотим научиться моделировать -* $F(t)$, $\exists F^{-1}(t)$. *Тогда* $Y = F^{-1}(X)$ *ведёт себя как эта случайная величина.*

Доказательство. Распишем функцию распределения Y:

$$F_Y = \mathrm{P}(Y \leqslant t) = \mathrm{P}(F^{-1}(X) \leqslant t) = \mathrm{P}(X \leqslant F(t)) = F_X(F(t))$$

Функция распределения X имеет вид:

$$F_X(t) = \begin{cases} 0, \ t < 0 \\ t, \ 0 \leqslant t \leqslant 1 \\ 1, \ t > 1 \end{cases}$$

Тогда, поскольку $X \in [0,1]$, верно

$$F_Y(t) = F_X(F(t)) = F(t)$$

То есть функция распределения Y такая же как и функция $F(t)$. Тогда по теореме, Y определяет моделируемую случайную величину. *ч.т.д.*

Пример 2.22. *Генерация бернуллиевской случайной величины.*
$\mathrm{Bern}(p)$*: если* ***random*** $< p$ *вернуть 1, иначе вернуть 0.*

Пример 2.23. *Генерация случайной величины, имеющей биномиальное распределение.*
$\mathrm{Bin}(n,p)$: $\sum_{j=1}^{n} \mathrm{Bern}(p)$.

Пример 2.24. *Генерация случайной величины, имеющей геометрическое распределение.*
$\mathrm{Geom}(p)$: $i=0$; *пока* $\mathrm{Bern}(p) = 0$ $\{i = i+1\}$; *вернуть* i.

Пример 2.25. *Генерация случайной величины, имеющей экспоненциальное распределение.*

Функция распределения $\mathrm{Exp}(\lambda)$ *имеет вид:* $F(t) = (1-e^{-\lambda t})\mathbb{1}(t \geqslant 0)$, *а обратная функция имеет вид:* $F^{-1}(t) = -\frac{\ln(1-x)}{\lambda}$. *По теореме, случайная величина* Y *с функцией распределения* $F_Y = -\frac{\ln(1-x)}{\lambda} \sim \mathrm{Exp}(\lambda)$.

Пример 2.26. *Генерация пуассоновской случайной величины.*

Пусть $X = \max\{n : \sum\limits_{j=1}^{n} Y_j \leqslant \lambda\}$, *где* $Y_j \sim \mathrm{Exp}(1)$. *Тогда* $X \sim \mathrm{Pois}(\lambda)$.

Чтобы это проверить, заметим что событие $\{X \leqslant m\} \Leftrightarrow \{\sum\limits_{j=1}^{m+1} Y_j > \lambda\}$. *Заметим что* $\sum_{j=1}^{m+1} Y_j \sim \Gamma(m+1,1)$

$$\mathrm{P}\left(\sum_{j=1}^{m+1} Y_j > \lambda\right) = \int_{\lambda}^{+\infty} \frac{x^m}{m!}e^{-x}dx = -\frac{x^m}{m!}e^{-x}\Bigg|_{\lambda}^{+\infty} + \int_{\lambda}^{+\infty} \frac{x^{m-1}}{(m-1)!}e^{-x}dx$$
$$= \frac{\lambda^m}{m!}e^{-\lambda} + \int_{\lambda}^{+\infty} \frac{x^{m-1}}{(m-1)!}e^{-x}dx$$

Воспользуемся индукцией и конечная результат пример вид:

$$\mathrm{P}(X \leqslant m) = \mathrm{P}\left(\sum_{j=1}^{m+1} Y_j > \lambda\right) = \sum_{j=0}^{m} \frac{\lambda^j}{j!}e^{-\lambda}$$

И последнее действие

$$\mathrm{P}(X = m) = \mathrm{P}(X \leqslant m) - \mathrm{P}(X \leqslant m-1) = \frac{\lambda^m}{m!}e^{-\lambda}$$

При $m \neq 0$, $\mathrm{P}(m=0) = e^{-\lambda}$. *Получилось распределение такое же как и* $\mathrm{Pois}(\lambda)$

Пример 2.27. *Генерация случайной величины с конечным дискретным распределением.*

Пусть задана функция распределения F. *Напомним, что она кусочно постоянна и имеет вид*

$$F(t) = \sum_{j=1}^{n} p_j \mathbb{1}(x_j \leqslant t) = \begin{cases} 0, t < x_1, \\ p_1, x_1 \leqslant t < x_2, \\ \ldots, \\ 1, t \geqslant x_n. \end{cases}$$

Вызываем функцию ***random****. Если ее значение от 0 до* p_1, *то возвращаем* x_1, *от* p_1 *до* p_1+p_2 *–* x_2, *…, от* $p_1+\ldots+p_{n-1}$ *до 1 –* x_n. *Также заметим, что для ускорения алгоритма можно применить бинарный поиск.*

Пример 2.28. *Генерация вектора, имеющего полиномиальное распределение.*

Чтобы сгенерировать вектор $\sim \mathrm{Poly}(1,p)$, *можно применить схему из предыдущего примера, то есть если результат вызова функции* ***random*** *от 0 до* p_1, *то возвращаем* $(1,0,\ldots,0)^T$, p_1 *до* p_1+p_2 *–* $(0,1,\ldots,0)^T$, *от* $p_1+\ldots+p_{m-1}$ *до 1 –* $(0,0,\ldots,1)^T$.

$\mathrm{Poly}(n,p) : \sum_{j=1}^{n} \mathrm{Poly}(1,p)$.

Пример 2.29. *Генерация нормально распределенных случайных величин.*

Во-первых, можно использовать аппроксимацию с помощью интегральной теоремы Муавра-Лапласа (или более общую центральную предельную теорему), а именно аппроксимация будет выглядеть так:

$$\frac{\mathrm{Bin}(n,p) - np}{\sqrt{npq}}.$$

Кроме того, пусть $x =$ ***random*** *и* $y =$ ***random****. Тогда величины*

$$u = \sqrt{2|\ln x|}\cos(2\pi y),$$
$$w = \sqrt{2|\ln x|}\sin(2\pi y),$$

независимы и имеют стандартное нормальное распределение.

В заключение параграфа проведем еще один алгоритм.

Пример 2.30. *Выборка с отклонением (Rejection sampling)*

Пусть у искомого распределения плотность p. *Также пусть имеется плотность* f, *для которой* $p(x) \leqslant c \cdot f(x)$ $(c > 1)$ *для любого* x *и случайную величину, которой соответствует плотность* f *легко генерировать. Тогда процедура выглядит следующим образом: генерировать пары* (u, y), *где* $u =$ ***random***, *а* y *сгенерирована из распределения с плотностью* f *пока* $u \geqslant \frac{p(y)}{c \cdot f(y)}$, *вернуть* y.

Поясним, почему данный алгоритм корректен. Пусть $U \sim \mathrm{U}[0;1]$, Y *имеет распределение с плотностью* f *и они независимы. Рассмотрим следующую условную вероятность:*

$$\mathrm{P}\left(Y \leqslant t | U \leqslant \frac{p(Y)}{c \cdot f(Y)}\right) = \frac{\mathrm{P}\left(Y \leqslant t,\, U \leqslant \frac{p(Y)}{c \cdot f(Y)}\right)}{\mathrm{P}\left(U \leqslant \frac{p(Y)}{c \cdot f(Y)}\right)}.$$

В числителе имеем

$$\mathrm{P}\left(Y \leqslant t,\, U \leqslant \frac{p(Y)}{c \cdot f(Y)}\right) = \iint\limits_{y \leqslant t, u \leqslant \frac{p(y)}{c \cdot f(y)}} f(y)\; dudy$$

$$= \int\limits_{y \leqslant t} \frac{p(y)}{c \cdot f(y)} f(y)\; dy = \frac{1}{c} F_Y(t).$$

Здесь F_Y *– функция распределения для* Y. *Осталось рассмотреть знаменатель*

$$\mathrm{P}\left(U \leqslant \frac{p(Y)}{c \cdot f(Y)}\right) = \iint\limits_{u \leqslant p(y)/c \cdot f(y)} f(y)\; dudy = \frac{1}{c} \int_{-\infty}^{+\infty} p(y)\; dy = \frac{1}{c}.$$

Таким образом, получаем, что

$$\mathrm{P}\left(Y \leqslant t | U \leqslant \frac{p(Y)}{c \cdot f(Y)}\right) = F_Y(t)$$

2.6 Вероятностные интегралы

В дальнейшем нам понадобятся вероятностные интегралы. В данном параграфе мы не будем заниматься их строгим построением, а рассмотрим их в достаточной для дальнейшего изложения степени.

Пусть $(\Omega, \mathcal{A}, \mathrm{P})$ – вероятностное пространство, X – случайный вектор, g – неотрицательная борелевская функция (прообраз борелевского множества тоже борелевский). Мы специально не пишем $X = (X_1, \ldots, X_n)^T$ и $g(x_1, \ldots, x_n)$, так как написанное ниже будет справедливым для любой натуральной размерности. Рассмотрим следующей интеграл:

$$\int\limits_{\Omega} g(X(\omega))\, \mathrm{P}(d\omega).$$

Этот интеграл мы перепишем иначе (*перейдем от меры* P *к мере* P_X)

$$\int\limits_{\mathbb{R}^n} g(x)\, \mathrm{P}_X(dx).$$

Если X и его распределение дискретны, то под интегралом будем понимать сумму

$$\sum g(x_k)\,\mathrm{P}(X = x_k),$$

в непрерывном случае, то есть наличии плотности p интеграл есть

$$\int\limits_{\mathbb{R}^n} g(x)p(x)dx.$$

Заметим, что интеграл $\int\limits_{\mathbb{R}^n} g(x)\,\mathrm{P}_X(dx)$ от неотрицательной функции g существует, если сходится соответствующий ряд или интеграл. Если ряд или интеграл расходятся, то будем говорить, что данный интеграл не существует.

Замечание. *В данном случае еще можно условиться, что* $\int\limits_{\mathbb{R}^n} g(x)\,\mathrm{P}_X(dx) = +\infty$.

Пусть теперь g – произвольная борелевская функция. Тогда представим g в виде разности неотрицательных функций

$$g = g_+ - g_-,$$
$$g_+ = \max(0, g), \quad g_- = -\min(0, g),$$

то есть g_+ – положительная часть функции g, g_- – отрицательная со знаком "минус". Тогда интеграл $\int\limits_{\mathbb{R}^n} g(x)\,\mathrm{P}_X(dx)$ от произвольной функции g представим в виде разности интегралов от неотрицательных функций

$$\int\limits_{\mathbb{R}^n} g(x)\,\mathrm{P}_X(dx) = \int\limits_{\mathbb{R}^n} g_+(x)\,\mathrm{P}_X(dx) - \int\limits_{\mathbb{R}^n} g_-(x)\,\mathrm{P}_X(dx).$$

Возможны следующие ситуации:

- Оба интеграла конечны. Тогда искомый интеграл есть разность интегралов.
- Если один из интегралов бесконечен, то искомый интеграл не существует.

Замечание. *Второй пункт предыдущего списка можно уточнить:*

- *Если интеграл от g_+ равен $+\infty$, от g_- – конечен, то искомый интеграл равен $+\infty$.*
- *Если интеграл от g_+ конечен, от g_- есть $+\infty$, то искомый интеграл полагается равным $-\infty$.*
- *Если оба интеграла от g_+ и g_- бесконечны, то искомый интеграл не существует.*

Отдельно рассмотрим для размерности $n = 1$ *интеграл Стилтьеса* $\int\limits_{\mathbb{R}} g(x)dF(x)$, где F – функция распределения. Он определяется как предел следующих интегральных сумм:

$$\int\limits_{\mathbb{R}} g(x)dF(x) = \lim_{\max(x_{j+1}-x_j)\to 0} \sum_j g(x_j^*)(F(x_{j+1}) - F(x_j)), \quad x_j^* \in [x_j, x_{j+1}].$$

В дискретном случае функция распределения F кусочно-постоянна и имеет разрывы в точках x_j, для которых $\mathrm{P}(X = x_j) = p_j > 0$, причем величина данного скачка равняется p_j, поэтому интеграл Стилтьеса превращается в сумму:

$$\int\limits_{\mathbb{R}} g(x)dF(x) = \sum_j g(x_j)p_j.$$

В непрерывном случае $F' = p$ почти всюду. Поэтому имеем (*вывод не совсем формальный*):

$$\int\limits_{\mathbb{R}} g(x)dF(x) = \lim_{\max(x_{j+1}-x_j)\to 0} \sum_j g(x_j^*)\frac{(F(x_{j+1}) - F(x_j))}{x_{j+1} - x_j}(x_{j+1} - x_j)$$
$$= \int\limits_{\mathbb{R}} g(x)p(x)dx.$$

В одномерном случае как вероятностный интеграл, так и интеграл Стилтьеса.

Отметим простейшие свойства интеграла:

1. Линейность, то есть $\int(a\cdot f(x) + b\cdot g(x)) = a\int f(x) + b\int g(x)$.
2. Аддитивность по множеству, то есть если $A = B \sqcup C$, то $\int\limits_A = \int\limits_B + \int\limits_C$.
3. $\mathrm{P}(X \in B) = \int\limits_B \mathrm{P}_X(dx)$, в одномерном случае также $\mathrm{P}(X \in B) = \int\limits_B dF(x)$.

В заключение параграфа приведем вероятностную интерпретацию теоремы Фубини и её применение к нахождению функции распределения суммы независимых случайных величин.

Теорема 2.13. ***Теорема Фубини*** *(б/д).*

X, Y - независимые случайные величины, $g : \mathbb{R}^2 \to \mathbb{R}$ - измеримая $\Rightarrow$ $\iint g(x,y)\,\mathrm{P}(dx,dy) = \int[\int g(x,y)dF_x]dF_y(y) = \int[\int g(x,y)dF_y(y)]dF_x(x)$

Пример 2.31. *X, Y-независимые, тогда определим свёртку распределений как*

$$F_{X+Y}(t) = \mathrm{P}(X + Y \leqslant t) = \iint\limits_{x+y\leqslant t} P_{X,Y}(dx,dy) = \iint\limits_{x+y\leqslant t} dF_X(x)dF_Y(y)$$
$$= \int\limits_{\mathbb{R}} \left(\int\limits_{y\leqslant t-x} dF(y)\right) dF_X(x) = \int\limits_{\mathbb{R}} F_Y(t-x)dF_X(x)$$

2.7 Математическое ожидание и дисперсия

2.7.1 Математическое ожидание

Определение 2.14. *Пусть X – случайная величина.* ***Математическим ожиданием*** *случайной величины X будем называть величину $\mathrm{E}\,X = \int\limits_{\mathbb{R}} xdF(x)$ (если данный интеграл существует).*

Замечание. *В дискретном случае математическое ожидание есть сумма:*

$$\mathrm{E}\,X = \sum_j x_j\,\mathrm{P}(X = x_j).$$

В непрерывном случае – интеграл:

$$\mathrm{E}\,X = \int\limits_{\mathbb{R}} xp(x)\;dx,$$

где p – плотность.

Также в русскоязычной литературе часто используется обозначение $\mathrm{M}X$. *Мы будем использовать обозначение* E *(англ. expected value).*

Если нам важно указать, относительно какого распределения берется мат. ожидания мы будем использовать обозначения $\mathrm{E}_{\mathrm{P}_X}$, $\mathrm{E}_{\mathrm{P}(X)}$.

Математическое ожидание также часто называют средним значением случайной величины, это будет видно ниже из одного примера. Сформулируем в виде теоремы свойства математического ожидания.

Теорема 2.14. *Для математического ожидания справедливы следующие свойства (предполагается, что все мат. ожидания существуют):*

1. *Пусть* $X_1, \ldots, X_n$ *– случайные величины и* $g : \mathbb{R}^n \to \mathbb{R}$ *– борелевская функция. Тогда*

$$\mathrm{E}\, g(X_1, \ldots, X_n) = \int\limits_{\mathbb{R}^n} g(x_1, \ldots, x_n)\, \mathrm{P}_{X_1,\ldots,X_n}(dx_1, \ldots, dx_n) = \int\limits_{\mathbb{R}} y dF_{g(X_1,\ldots,X_n)}(y).$$

   Здесь по сути сделана замена переменной. В качестве упражнения можно отдельно расписать для дискретного и непрерывного случаев в одномерном случае.

2. $\mathrm{E}(c_1 X_1 + c_2 X_2) = c_1 \mathrm{E}\, X_1 + c_2 \mathrm{E}\, X_2$. *Следует из линейности интеграла и того, что* $\iint g(x)\, \mathrm{P}_{X,Y}(dx, dy) = \int g(x)\, \mathrm{P}_X(dx)$

3. *Пусть* X, Y *– независимы. Тогда* $\mathrm{E}\, XY = \mathrm{E}\, X \cdot \mathrm{E}\, Y$. *Немедленно следует из теоремы Фубини.* ***Обратное неверно****.*

4. *Пусть* $X \geqslant 0$ *(то есть* $\mathrm{P}(X \geqslant 0) = 1$*), тогда* $\mathrm{E}\, X \geqslant 0$. *Следует из того, что интегрируем по положительной полуоси.*

5. $X \geqslant 0$ *и* $\mathrm{E}\, X = 0$. *Тогда* $X = 0$ *(*$\mathrm{P}(X = 0) = 1$*). Для доказательства воспользуемся тем, что* $\mathrm{P}(X \geqslant c) \leqslant \mathrm{E}\, X/c = 0$ *для любой положительной константы (это мы докажем в следующей теореме), то есть* $\mathrm{P}(X \geqslant c) = 0$. *Действительно*

$$1 = \mathrm{P}(X \geqslant 0) = \mathrm{P}(X < c) + \mathrm{P}(X \geqslant c) = \mathrm{P}(X \leqslant c),$$

   то есть получили, что с вероятностью 1 выполняется соотношение $0 \leqslant X < c$ *для любой положительной константы. По непрерывности вероятностной меры получаем, что* $\mathrm{P}(X = 0) = 1$.

Сформулируем и докажем важное неравенство, названное в честь Маркова.

Теорема 2.15. ***Неравенство Маркова****.*
Пусть $X \geqslant 0$ *и существует* $\mathrm{E}\, X$. *Тогда для любой положительной константы* c

$$\mathrm{P}(X \geqslant c) \leqslant \frac{\mathrm{E}\, X}{c}.$$

Доказательство. Вспомним, что $\mathrm{P}(X \in B) = \int\limits_B dF(x)$ и воспользуемся свойствами интеграла

$$\mathrm{P}(X \geqslant c) = \int\limits_{x \geqslant c} dF(x) \leqslant \int\limits_{x \geqslant c} \frac{x}{c} dF(x) \leqslant \frac{1}{c} \int\limits_{\mathbb{R}} x dF(x) = \frac{\mathrm{E}\, X}{c}.$$

ч.т.д.

2.7.2 Дисперсия

Определение 2.15. ***Дисперсией*** *случайной величины X будем называть число* $\operatorname{Var} X = \mathrm{E}(X - \mathrm{E}\, X)^2$. ***Стандартное отклонение*** *(среднеквадратическое отклонение) – корень из дисперсии. Часто обозначается буквой σ.*

Замечание. *В русскоязычной литературе также используется обозначение* D*. Мы же будем использовать* Var *(англ. variance).*

Дисперсия является мерой отклонения случайной величины от её математического ожидания, что можно увидеть из определения и будет наглядно показано ниже. Сформулируем и докажем свойства дисперсии.

Теорема 2.16. ***Свойства дисперсии***

1. $\operatorname{Var} X \geqslant 0$.
2. $\operatorname{Var} X = \mathrm{E}\, X^2 - (\mathrm{E}\, X)^2$.
3. $\operatorname{Var}(aX + b) = a^2 \operatorname{Var} X$.
4. $\operatorname{Var}(X \pm Y) = \operatorname{Var} X + \operatorname{Var} Y \pm 2(\mathrm{E}(XY) - \mathrm{E}\, X \cdot \mathrm{E}\, Y)$. *В частности, если X, Y – независимы, то* $\operatorname{Var}(X \pm Y) = \operatorname{Var} X + \operatorname{Var} Y$.
5. $\operatorname{Var}(X_1 + \ldots + X_n) = \sum_{j=1}^{n} \operatorname{Var} X_j + 2\sum_{k<j} (\mathrm{E}(X_k X_j) - \mathrm{E}\, X_k \cdot \mathrm{E}\, X_j)$.
6. *Если* $\operatorname{Var} X = 0$, *то* $\mathrm{P}(X = c) = 1$ *для некоторой c.*
7. $\operatorname{Var} X$ *есть минимум функции* $f(a) = \mathrm{E}(X - a)^2$, *причем минимум достигается при* $a = \mathrm{E}\, X$.

Доказательство.

1. Немедленно вытекает из определения и свойств математического ожидания.
2. Действительно, имеем

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

3. В самом деле

$$\operatorname{Var}(aX + b) = \mathrm{E}(aX + b - \mathrm{E}(aX + b))^2 = \mathrm{E}(aX - a\,\mathrm{E}\, X)^2 = a^2 \operatorname{Var} X.$$

4. Распишем цепочку равенств

$$\begin{gathered} \operatorname{Var}(X \pm Y) = \mathrm{E}(X \pm Y)^2 - (\mathrm{E}(X \pm Y))^2 \\ = \mathrm{E}\, X^2 \pm 2\,\mathrm{E}\, XY + \mathrm{E}\, Y^2 - [(\mathrm{E}\, X)^2 \pm 2\,\mathrm{E}\, X\,\mathrm{E}\, Y + (\mathrm{E}\, Y)^2] \\ = \operatorname{Var} X + \operatorname{Var} Y \pm 2(\mathrm{E}\, XY - \mathrm{E}\, X\,\mathrm{E}\, Y) \end{gathered}$$

Если X, Y – независимы, то $\mathrm{E}(XY) = \mathrm{E}\, X \cdot \mathrm{E}\, Y$, откуда и получается, что дисперсия суммы/разности независимых случайных величин равняется сумме дисперсий.

5. Можно доказать индукцией с помощью предыдущих выкладок и свойств мат. ожидания.
6. Применить соответствующее свойство мат. ожидания.

7. $f(a) = a^2 - 2a\,\mathrm{E}\,X + (\mathrm{E}\,X)^2$ – парабола с направленными вверх ветвями, у нее точка минимума $a_* = -\frac{-2\,\mathrm{E}\,X}{2} = \mathrm{E}\,X$. Если подставить $a = a_*$, то получится в точности дисперсия.

ч.т.д.

Также сформулируем и докажем очень важное неравенство, которое мы будем не раз вспоминать по ходу изложения.

Теорема 2.17. ***Неравенство Чебышёва***
X-случайная величина. Тогда для любого $\varepsilon > 0$ справедливо неравенство

$$\mathrm{P}(|X - \mathrm{E}\,X| \geqslant \varepsilon) \leqslant \frac{\mathrm{Var}\,X}{\varepsilon^2}$$

Доказательство. Возведем выражение внутри вероятности в квадрат и применим к полученному неравенство Маркова

$$\mathrm{P}(|X - \mathrm{E}\,X| \geqslant \varepsilon) = \mathrm{P}(|X - \mathrm{E}\,X|^2 \geqslant \varepsilon^2) \leqslant \frac{\mathrm{E}(X - \mathrm{E}\,X)^2}{\varepsilon^2} = \frac{\mathrm{Var}\,X}{\varepsilon^2}.$$

ч.т.д.

Рассмотрим два важных следствия.

Следствие 2.1. ***О стандартном отклонении***
В неравенстве Чебышёва положим $\varepsilon = k\sigma$. Тогда получим

$$\mathrm{P}(|X - \mathrm{E}\,X| \geqslant k\sigma) \leqslant \frac{1}{k^2}.$$

В частности, если $k = 3$, то оценкой сверху является $1/9$ и по сути на этом основывается правило "трёх сигм". Однако это всего лишь оценка сверху и в некоторых случаях она может быть груба, в чем мы убедимся ниже.

Следствие 2.2. ***Закон больших чисел***
Пусть $\{X_j\}_{j=1}^{\infty}$ – последовательность независимых одинаково распределённых случайных величин с математическим ожиданием $\mathrm{E}\,X_1$ и дисперсией $\mathrm{Var}\,X_1$. Пусть $S_n = X_1 + \ldots + X_n$. Тогда для любого $\varepsilon > 0$

$$\lim_{n\to\infty} \mathrm{P}\left(\left|\frac{S_n}{n} - \mathrm{E}\,X_1\right| \geqslant \varepsilon\right) = 0$$

Доказательство. Заметим, что $\mathrm{E}\,S_n/n = \mathrm{E}\,X_1$ и $\mathrm{Var}\,S_n/n = \mathrm{Var}\,X_1/n$. Тогда искомое получается из неравенства Чебышёва. *ч.т.д.*

Замечание. *Закон больших чисел справедлив и без предположения о существовании дисперсии, в чем мы тоже убедимся ниже.*

2.7.3 Вычисление математических ожиданий и дисперсий для некоторых распределений

Пример 2.32. *Вырожденное распределение I_c. Здесь очевидно $\mathrm{E}\,X = c$, $\mathrm{Var}\,X = 0$.*

Пример 2.33. *$X \sim \mathrm{Bern}(p)$. Здесь имеем*

$$\mathrm{E}\,X = 1 \cdot p + 0 \cdot (1 - p) = p,$$
$$\mathrm{E}\,X^2 = 1^2 \cdot p + 0^2 \cdot (1 - p) = p,$$
$$\mathrm{Var}\,X = \mathrm{E}\,X^2 - (\mathrm{E}\,X)^2 = p - p^2 = p(1 - p) = pq.$$

Пример 2.34. $X \sim \mathrm{Bin}(n,p)$. *Пусть* $X = S_n = Y_1 + \ldots + Y_n, \forall i: \ Y_i \sim \mathrm{Bern}(p)$ *и они независимы. Тогда* $S_n \sim \mathrm{Bin}(n,p)$, *и тогда*

$$\begin{gathered} \mathrm{E}\, S_n = \mathrm{E}\, Y_1 + \ldots + \mathrm{E}\, Y_N = np \\ \mathrm{Var}\, S_n = \mathrm{Var}\, Y_1 + \ldots + \mathrm{Var}\, Y_n = npq \end{gathered}$$

Пример 2.35. $X \sim \mathrm{Pois}(\lambda)$. *Считаем мат. ожидание*

$$\mathrm{E}\, X = \sum_{k=0}^{\infty} k \cdot e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}$$

Данная сумма это ряд Маклорена для функции e^x *в точке* λ

$$\mathrm{E}\, X = e^{-\lambda} \lambda e^{\lambda} = \lambda$$

Прежде чем посчитаем дисперсию, установим следующее тождество

$$\begin{gathered} \mathrm{E}\, X(X-1) = \mathrm{E}\, X^2 - \mathrm{E}\, X = \mathrm{E}\, X^2 - (\mathrm{E}\, X)^2 + (\mathrm{E}\, X)^2 - \mathrm{E}\, X \\ = \mathrm{Var}\, X + (\mathrm{E}\, X)^2 - \mathrm{E}\, X \Rightarrow \\ \mathrm{Var}\, X = \mathrm{E}\, X(X-1) - (\mathrm{E}\, X)^2 + \mathrm{E}\, X \end{gathered}$$

Теперь чтобы посчитать дисперсию остаётся посчитать $\mathrm{E}\, X(X-1)$

$$\mathrm{E}\, X(X-1) = \sum_{k=2}^{\infty} k(k-1) e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} = \lambda^2$$

Опять же, сумма сворачивается в e^{λ}. *Дальше считаем дисперсию*

$$\mathrm{Var}\, X = \lambda^2 - \lambda^2 + \lambda = \lambda$$

Пример 2.36. $X \sim \mathrm{Geom(p)}$. *Для геометрического распределения* $\mathrm{P}(X = k) = q^k p, k \in \mathbb{N} \cup \{0\}$. *Мат. ожидание:*

$$\mathrm{E}\, X = \sum_{k=1}^{\infty} k q^k p = pq \sum_{k=1}^{\infty} k q^{k-1} = pq \sum_{k=1}^{\infty} (q^k)'_q$$

Получается сумма производных степенных функций. Как, надеюсь, известно степенной ряд сходится равномерно, а значит сумма производных есть производная суммы

$$\sum_{k=1}^{\infty} (q^k)'_q = \left(\sum_{k=1}^{\infty} q^k \right)'_q = \left(\frac{q}{1-q} \right)'_q = \frac{1-q+q}{(1-q)^2} = \frac{1}{p^2}$$

Итого мат. ожидание: $\mathrm{E}\, X = pq \frac{1}{p^2} = \frac{q}{p}$
Для дисперсии воспользуемся приёмом с подсчётом $\mathrm{E}\, X(X-1)$

$$\mathrm{E}\, X(X-1) = \sum_{k=2}^{\infty} k(k-1) q^k p = pq^2 \sum_{k=2}^{\infty} (q^k)''_q$$

По аналогичной причине от суммы производных переходим к сумме производных

$$\sum_{k=2}^{\infty} (q^k)''_q = \left(\sum_{k=2}^{\infty} q^k \right)''_q = \left(\frac{q^2}{1-q} \right)''_q$$

Честно посчитаем производную

$$\left(\frac{q^2}{1-q}\right)''_q = \left(\frac{2q-q^2}{(1-q)^2}\right)'_q = \frac{2-4q+2q^2+4q-2q^2}{(1-q)^3} = \frac{2}{p^3}$$

И дисперсия будет:

$$\operatorname{Var} X = pq^2\frac{2}{p^3} - \left(\frac{q}{p}\right)^2 + \frac{q}{p} = \left(\frac{q}{p}\right)^2 + \frac{q}{p} = \frac{q(q+p)}{p^2} = \frac{q}{p^2}$$

Пример 2.37. $X \sim \mathrm{U}[a,b]$. *Как было показано ранее, если* $Y \sim \mathrm{U}[0,1] \Rightarrow X = (b-a)Y + a \sim \mathrm{U}[a,b]$, *поэтому мы честно посчитаем мат. ожидание и дисперсию величины* Y *и по свойствам доберёмся до* X

$$\operatorname{E} Y = \int_{-\infty}^{+\infty} y\frac{1}{1}dy = \int_0^1 ydy = \frac{1}{2}$$
$$\operatorname{E} Y^2 = \int_{-\infty}^{+\infty} y^2\frac{1}{1}dy = \int_0^1 y^2dy = \frac{1}{3}$$
$$\operatorname{Var} Y = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$$
$$\operatorname{E} X = \operatorname{E}[(b-a)Y + a] = (b-a)\operatorname{E} Y + a = \frac{b-a}{2} + a = \frac{a+b}{2}$$
$$\operatorname{Var} X = \operatorname{Var}[(b-a)Y + a] = (b-a)^2 \operatorname{Var} Y = \frac{(b-a)^2}{12}$$

Пример 2.38. $X \sim \mathcal{N}(\mu, \sigma^2)$. *Опять же, выразим* X *через более простую величину* $Y \sim \mathcal{N}(0,1)$: $X = \sigma Y + \mu \sim \mathcal{N}(\mu, \sigma^2)$

$$\operatorname{E} Y = \int_{-\infty}^{+\infty} \frac{y}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy = -\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}\Big|_{-\infty}^{+\infty} = 0$$
$$\operatorname{E} Y^2 = \int_{-\infty}^{+\infty} \frac{y^2}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy = -\frac{y}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}\Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy$$
$$= 0 + 1 = 1$$
$$\operatorname{Var} Y = 1 - 0 = 1$$
$$\operatorname{E} X = \operatorname{E}[\sigma Y + \mu] = \sigma \operatorname{E} Y + \mu = 0 + \mu = \mu$$
$$\operatorname{Var} X = \operatorname{Var}[\sigma Y + \mu] = \sigma^2 \operatorname{Var} Y = \sigma^2$$

Пример 2.39. $X \sim \operatorname{Exp}(\lambda)$. *Честно посчитаем интегралы*

$$\operatorname{E} X = \int_{-\infty}^{+\infty} x\lambda e^{-\lambda x}dx\mathbb{1}(X \geqslant 0) = \int_0^{+\infty} x\lambda e^{-\lambda x}dx$$
$$= -xe^{-\lambda x}\Big|_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x}dx = 0 - \frac{1}{\lambda}e^{-\lambda x}\Big|_0^{+\infty} = \frac{1}{\lambda}$$
$$\operatorname{E} X^2 = \int_0^{+\infty} x^2\lambda e^{-\lambda x} = -x^2e^{-\lambda x}\Big|_0^{+\infty} + 2\frac{1}{\lambda}\int_0^{+\infty} x\lambda e^{-\lambda x}dx$$
$$= 0 + 2\frac{1}{\lambda}\operatorname{E} X = \frac{2}{\lambda^2}$$
$$\operatorname{Var} X = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

Пример 2.40. $X \sim \Gamma(n, \lambda)$. *Как известно* $X = Y_1 + \ldots + Y_n \sim \Gamma(n, \lambda)$, *если* $\forall i : Y_i \sim \mathrm{Exp}(\lambda)$ *и независимы. Тогда мат. ожидание и дисперсия легко находятся*

$$\mathrm{E}\,X = \mathrm{E}\,Y_1 + \ldots + \mathrm{E}\,Y_n = \frac{n}{\lambda}$$
$$\mathrm{Var}\,X = \mathrm{Var}\,Y_1 + \ldots + \mathrm{Var}\,Y_n = \frac{n}{\lambda^2}$$

Пример 2.41. $X \sim \mathrm{Cauhy}(\gamma, x_0)$. *Рассмотрим интеграл для мат. ожидания*

$$\frac{1}{\pi\gamma} \int_{-\infty}^{+\infty} \frac{x dx}{1 + \left(\frac{x - x_0}{\gamma}\right)^2}$$

Этот несобственный интеграл, вообще говоря, расходится. А значит мат. ожидание не определено и дисперсию вычислить не можем (иногда говорят, что дисперсия бесконечна).

В завершении параграфа, приведём контрпример, когда $\mathrm{E}\,XY = \mathrm{E}\,X \cdot \mathrm{E}\,Y$, но X и Y зависимы. Пусть $X \sim \mathcal{N}(0,1), Y = X^2$. Посчитаем соответствующие мат. ожидания: $\mathrm{E}\,X = 0$; $\mathrm{E}\,Y = \mathrm{E}\,X^2 = 1$; $\mathrm{E}\,XY = \mathrm{E}\,X^3 = 0$ (упражнение). Тогда действительно верно, что $\mathrm{E}\,XY = \mathrm{E}\,X \cdot \mathrm{E}\,Y$, но X и Y, очевидно, зависимы.

2.8 Другие числовые характеристики

2.8.1 Моменты высших порядков и связанные с ним характеристики

Определение 2.16. ***Моменты***

Начальным моментом *k-го порядка будем называть величину* $\mathrm{E}\,X^k$.

Абсолютным начальным моментом *k-го порядка –* $\mathrm{E}\,|X|^k$.

Центральным моментом *k-го порядка –* $\mathrm{E}(X - \mathrm{E}\,X)^k$.

Центральным абсолютным моментом *k-го порядка –* $\mathrm{E}\,|X - \mathrm{E}\,X|^k$

Далее рассмотрим связанные с моментами высших порядков характеристики

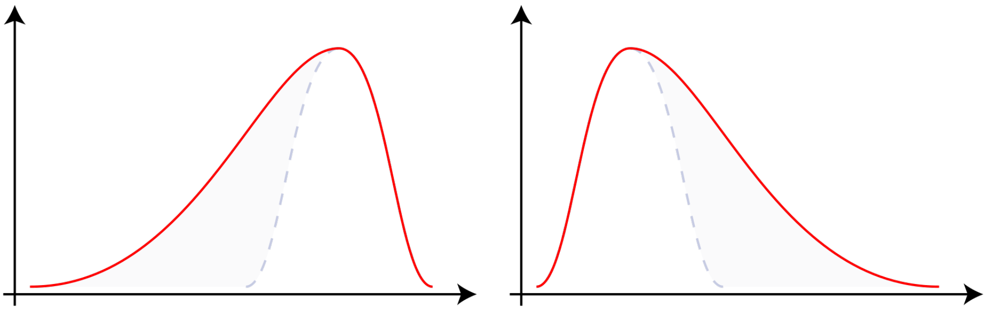
Определение 2.17. ***Коэффициентом асимметрии*** *(англ. skewness) будем называть величину* $\frac{\mathrm{E}(X - \mathrm{E}\,X)^3}{\sigma^3}$

Замечание. *В теории вероятностей и её приложениях используется жаргонизм – хвост распределения. Неформально – поведение распределения на бесконечности. В отдельных случаях различаются правый и левый хвост. Характеризуется длиной (например, у* $\mathrm{U}[0, 1000]$ *правый хвост длиннее чем у* $\mathrm{U}[0, 1]$*) и массой (у распределения Коши хвосты тяжелее чем у нормального)*

Замечание. *Если коэффициент асимметрии положительный, то "правый хвост"длиннее "левого". Если отрицательный, то "левый хвост"длиннее "правого".*

Определение 2.18. ***Коэффициентом эксцесса*** *будем называть величину* $\frac{\mathrm{E}(X - \mathrm{E}\,X)^4}{\sigma^4} - 3$.

Замечание. *"Минус три" добавлено, чтобы у* $\mathcal{N}(0,1)$ *было равным нулю. Если коэффициент эксцесса больше нуля, то "пик" у математического ожидания острый. Если отрицательный – пологий.*



2.8.2 Квантиль. Медиана

Определение 2.19. ***Квантилем*** *порядка* $\alpha \in [0, 1]$ *распределения* P_X *будем называть число* $q_{\mathrm{P}_X,\alpha}$ *число, для которого*

$$\mathrm{P}(X \geqslant q_{\mathrm{P}_X,\alpha}) \geqslant 1 - \alpha, \quad \mathrm{P}(X \leqslant q_{\mathrm{P}_X,\alpha}) \geqslant \alpha.$$

Как правило, распределение, относительно которого вычисляется квантиль, понятно из контекста, поэтому мы в основном будем просто писать q_α.

Замечание. *Если функция распределения строго монотонна, то квантиль однозначно определяется из уравнения* $F(q_\alpha) = \alpha$.

В общем случае согласно сформулированному определению квантиль определяется неоднозначно, поэтому для устранения неопределенности вводится квантильная функция, которая может задавать одним из двух способов.

$$F^{-1}_{\sup}(\alpha) = \sup\{x \in \mathbb{R} : F(x) \leqslant \alpha\},$$
$$F^{-1}_{\inf}(\alpha) = \inf\{x \in \mathbb{R} : F(x) \geqslant \alpha\}.$$

Далее рассмотрим один особый случай.

Определение 2.20. ***Медианой*** *случайной величины* X *(и соответствующего распределения) называется квантиль порядка* 0.5. *Будем обозначать* $\operatorname{med} X$

Интуитивно вероятность того, что случайное значение будет меньше медианы равняется 0.5, случайное значение больше медианы – тоже 0.5. Сформулируем и докажем теорему, характеризующую медиану.

Теорема 2.18. *Пусть существует абсолютный начальный первый момент. Тогда медиана случайной величины есть точка минимума функции* $f(a) = \mathrm{E}\,|X - a|$

Доказательство. Не умаляя общности, будем считать, что у случайной величины медиана равняется нулю, так как мы всегда можем рассмотреть случайную величину $X' = X - \operatorname{med} X$, у которой очевидно медиана равняется нулю. Поэтому нам нужно показать, что $\mathrm{E}\,|X - c| \geqslant \mathrm{E}\,|X|$ для любой константы c.

Сначала рассмотрим случай $c > 0$. Заметим, что

$$|X - c| - |X| = \begin{cases} -c, & X > c, \\ c - 2X, & 0 < X \leqslant c, \\ c, & X \leqslant 0. \end{cases}$$

Отсюда, заключаем, что

$$\begin{aligned}\mathrm{E}(|X-c|-|X|) &= \mathrm{E}(|X-c|-|X|)\mathbb{1}(X>0)+\mathrm{E}(|X-c|-|X|)\mathbb{1}(X\leqslant 0)\\ &\geqslant -c\,\mathrm{E}\,\mathbb{1}(X>0)+c\,\mathrm{E}\,\mathbb{1}(X\leqslant 0)=c(\mathrm{P}(X\leqslant 0)-\mathrm{P}(X>0))\\ &= c(2\,\mathrm{P}(X\leqslant 0)-1)\geqslant 0.\end{aligned}$$

Случай $c<0$ сводится к только что рассмотренному, если подложить $\hat{c}=-c$ и $\hat{X}=-X$. *ч.т.д.*

Пример 2.42. *Пусть X равномерно распределено на множестве $\{x_1,\ldots,x_n\}$ (считаем, что x_i упорядочены по возрастанию). Если $n=2k+1$ нечетно, то $\operatorname{med}X=x_{k+1}$, если $n=2k$, то медианой является любое число между $[x_k,x_{k+1}]$. Кроме того, медиана будем минимизировать функцию*

$$f(y)=\sum_{j=1}^{n}|y-x_j|.$$

2.8.3 Мода

Определение 2.21. *Мы сформулируем определение* ***моды*** *отдельно для дискретного и непрерывного случая:*

- *дискретный: мода – наиболее вероятное значение, то есть x_* для которого $\mathrm{P}(X=x_*)$ наибольшая.*
- *непрерывный: мода – точка глобального максимума плотности.*

Замечание. *Есть понятия, схожие по морфологии с модой, но несколько отличающиеся: унимодальности, двумодальности и многомодальности. Неформально эти понятия означают, что у распределения один, два или больше "пиков" соответственно, в то время как мода только о глобальном максимуме.*

2.9 Вероятностные неравенства

Сперва рассмотрим множество случайных величин с конечным вторым моментом: $L_2=\{X:\mathrm{E}\,X^2<+\infty\}$. Несложно заметить, что L_2 является линейным пространством (линейная комбинация элементов множества принадлежит множеству). Рассмотрим на этом линейном пространстве функцию: $[X,Y]=\mathrm{E}(XY)$. Данная функция обладает свойствами:

1. $[X,Y]=[Y,X]$
2. Линейна по обоим аргументам
3. $[X,X]\geqslant 0;\quad [X,X]=0\Leftrightarrow X=0$

Функция $[X,Y]$ - скалярное произведение в пространстве L_2 и значит можем определить норму в пространстве: $||X||=\sqrt{[X,X]}$. Для скалярного произведения известно неравенство Коши-Буняковского: $|[X,Y]|\leqslant ||X||\cdot||Y||$

Подставив в неравенство наше скалярное произведение неравенство примет вид:

$$\mathrm{E}(XY)\leqslant\sqrt{\mathrm{E}\,X^2}\cdot\sqrt{\mathrm{E}\,Y^2}$$
$$\mathrm{E}^2(XY)\leqslant\mathrm{E}\,X^2\cdot\mathrm{E}\,Y^2$$

В данном пространстве можно выделить другое скалярное произведение: $[X, Y] = \mathrm{E}(X - \mathrm{E}\,X)(Y - \mathrm{E}\,Y)$, и неравенство будет:

$$\mathrm{E}(X - \mathrm{E}\,X)(Y - \mathrm{E}\,Y) \leqslant \sqrt{\mathrm{E}(X - \mathrm{E}\,X)^2} \cdot \sqrt{\mathrm{E}(Y - \mathrm{E}\,Y)^2}$$
$$\mathrm{E}^2(X - \mathrm{E}\,X)(Y - \mathrm{E}\,Y) \leqslant \operatorname{Var} X \cdot \operatorname{Var} Y$$

Замечание. *Данное скалярное произведение будет определено дальше*

Далее рассмотрим несколько вероятностных неравенств.

- Неравенство Маркова: $X \geqslant 0, \exists\, \mathrm{E}\,X \Rightarrow \mathrm{P}(X \geqslant C) \leqslant \frac{\mathrm{E}\,X}{C}, \quad \forall C > 0$

  Доказательство. Предоставлено выше 2.15 *ч.т.д.*

- Неравенство Чебышёва: $\exists \operatorname{Var} X \Rightarrow \mathrm{P}(|X - \mathrm{E}\,X| \geqslant \varepsilon) \leqslant \frac{\operatorname{Var} X}{\varepsilon^2}, \quad \forall \varepsilon > 0$

  Доказательство. Предоставлено выше 2.17 *ч.т.д.*

- Слабый закон больших чисел. $\{X_i\}_{i=1}^n$ - набор независимых и одинаково распределённых случайных величин (в английской литературе сокращённо i.i.d), для которых: $\mathrm{E}\,X_i = \mu, \operatorname{Var} X = \delta^2$; Возьмём $\overline{X_n} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$. Тогда вероятность $\mathrm{P}(|\overline{X_n} - \mu| \geqslant \xi) \leqslant \frac{\delta^2}{n\xi^2}$.

  Доказательство. Несложно понять, что $\mathrm{E}\,\overline{X} = \mu, \operatorname{Var} \overline{X} = \frac{\delta^2}{n}$ из свойств математического ожидания и дисперсии. Запишем неравенство Чебышёва для величины $\overline{X_n}$

  $$\mathrm{P}(|\overline{X} - \mu| \geqslant \xi) \leqslant \frac{\operatorname{Var} \overline{X}}{\xi^2} = \frac{\delta^2}{n\xi^2}$$

  ч.т.д.

- Неравенство Гёльдера:$E|XY| \leqslant (E|X|^p)^{\frac{1}{p}} \cdot (E|Y|^q)^{\frac{1}{q}}, \frac{1}{p} + \frac{1}{q} = 1$

  Доказательство. Сперва докажем неравенство Юнга.

  Лемма. *(Неравенство Юнга)*

  $$\forall a, b \geqslant 0; \forall p, q : \frac{1}{p} + \frac{1}{q} = 1 \Rightarrow ab \leqslant \frac{a^p}{p} + \frac{b^q}{q};$$

  Доказательство. Сначала проверим значения $a = 0, b = 0$ вместе и по отдельности. Неравенство, очевидно, выполняется.
  Как известно, функция $\ln x$ выпукла вверх на $(0, +\infty)$, из чего следует соотношение $\ln(\alpha x_1 + \beta x_2) \geqslant \alpha \ln(x_1) + \beta \ln(x_2)\, \forall \alpha, \beta : \alpha + \beta = 1$. Возьмём $\alpha = \frac{1}{p}, \beta = \frac{1}{q}, x_1 = a^p, x_2 = b^q$ и получим:

  $$\ln(\frac{1}{p}a^p + \frac{1}{q}b^q) \geqslant \frac{1}{p}\ln(a^p) + \frac{1}{q}\ln(b^q) = \ln(a) + \ln(b) = \ln(ab)$$

  Из монотонного возрастания $\ln x$ следует неравенство Юнга. *ч.т.д.*

Подставим в неравенство Юнга $a^p = \frac{|X|^p}{\mathrm{E}\,|X|^p}$, $b^q = \frac{|Y|^q}{\mathrm{E}\,|Y|^q}$.

$$\frac{|X|}{(\mathrm{E}\,|X|^p)^{\frac{1}{p}}}\frac{|Y|}{(\mathrm{E}\,|Y|^q)^{\frac{1}{q}}} \leqslant \frac{1}{p}\cdot\frac{|X|^p}{\mathrm{E}\,|X|^p} + \frac{1}{q}\cdot\frac{|Y|^q}{\mathrm{E}\,|Y|^q}$$

И возьмём математическое ожидание от обоих частей:

$$\mathrm{E}\left(\frac{|X|}{(\mathrm{E}\,|X|^p)^{\frac{1}{p}}}\frac{|Y|}{(\mathrm{E}\,|Y|^q)^{\frac{1}{q}}}\right) \leqslant \mathrm{E}\left(\frac{1}{p}\cdot\frac{|X|^p}{\mathrm{E}\,|X|^p} + \frac{1}{q}\cdot\frac{|Y|^q}{\mathrm{E}\,|Y|^q}\right);$$

$$\frac{\mathrm{E}\,|XY|}{(\mathrm{E}\,|X|^p)^{\frac{1}{p}}\cdot(\mathrm{E}\,|Y|^q)^{\frac{1}{q}}} \leqslant \frac{1}{p}\cdot\frac{\mathrm{E}\,|X|^p}{\mathrm{E}\,|X|^p} + \frac{1}{q}\cdot\frac{\mathrm{E}\,|Y|^q}{\mathrm{E}\,|Y|^q} = \frac{1}{p}+\frac{1}{q} = 1$$

Откуда следует неравенство Гёльдера. *ч.т.д.*

Замечание. *Неравенство Коши-Буняковского есть частный случай неравенства Гёльдера при $p = q = 2$.*

- Неравенство Минковского: $(\mathrm{E}\,|X+Y|^p)^{\frac{1}{p}} \leqslant (\mathrm{E}\,|X|^p)^{\frac{1}{p}} + (\mathrm{E}\,|Y|^p)^{\frac{1}{p}}$

  Доказательство. Сперва напомним, что есть неравенство треугольника: $\forall X, Y\ :\ |X+Y| \leqslant |X| + |Y|$. Рассмотрим $\mathrm{E}\,|X+Y|^p$:

  $$\mathrm{E}\,|X+Y|^p = \mathrm{E}\,|X+Y|^{p-1}|X+Y| \leqslant \mathrm{E}\,|X+Y|^{p-1}|X| + \mathrm{E}\,|X+Y|^{p-1}|Y|;$$

  Возьмём $\frac{1}{q} = 1 - \frac{1}{p}$, откуда $(p-1)q = p$, и дважды воспользуемся неравенством Гёльдера, раскрыв каждое слагаемое в правой части:

  $$\begin{aligned}(\mathrm{E}\,|X+Y|)^1 = \mathrm{E}\,|X+Y|^p \leqslant \mathrm{E}\,|X+Y|^{p-1}|X| + \mathrm{E}\,|X+Y|^{p-1}|Y| \leqslant\\ \leqslant (\mathrm{E}\,|X+Y|^{(p-1)q})^{\frac{1}{q}}\cdot(\mathrm{E}\,|X|^p)^{\frac{1}{p}} + (\mathrm{E}\,|X+Y|^{(p-1)q})^{\frac{1}{q}}\cdot(\mathrm{E}\,|Y|^p)^{\frac{1}{p}} =\\ = (\mathrm{E}\,|X+Y|^p)^{\frac{1}{q}}\cdot((\mathrm{E}\,|X|^p)^{\frac{1}{p}} + (\mathrm{E}\,|Y|^p)^{\frac{1}{p}}) =\\ = (\mathrm{E}\,|X+Y|^p)^{1-\frac{1}{p}}\cdot((\mathrm{E}\,|X|^p)^{\frac{1}{p}} + (\mathrm{E}\,|Y|^p)^{\frac{1}{p}});\end{aligned}$$

  Поделив обе части неравества на $(\mathrm{E}\,|X+Y|^p)^{1-\frac{1}{p}}$, получаем неравенство Минковского. *ч.т.д.*

- Неравенство Йенсена. Пусть дана такая функция g, что функция выпукла вниз при каждом значении некоторой случайной величины $X \Rightarrow g(\mathrm{E}\,X) \leqslant \mathrm{E}\,g(X)$

  Доказательство. То, что g выпукла вниз означает, что касательная к графику в точке $(x_0,\ g(x_0))$ находится под графиком, а значит имеем следующее неравенство:

  $$g'(x_0)(x - x_0) + g(x_0) \leqslant g(x);$$

  Возьмём $x = X, \quad x_0 = \mathrm{E}\,X$:

  $$g'(\mathrm{E}\,X)(X - \mathrm{E}\,X) + g(\mathrm{E}\,X) \leqslant g(X);$$

  И возьмём математическое ожидание от обоих частей неравенства:

  $$\mathrm{E}(g'(\mathrm{E}\,X)(X - \mathrm{E}\,X) + g(\mathrm{E}\,X)) \leqslant \mathrm{E}(g(X));$$
  $$g'(\mathrm{E}\,X)(\mathrm{E}\,X - \mathrm{E}\,X) + \mathrm{E}\,g(\mathrm{E}\,X) = g(\mathrm{E}\,X) \leqslant \mathrm{E}\,g(X)$$

  ч.т.д.

Замечание. *Если g выпукла вверх, то $-g$ выпукла вниз. Если для этой функции применить применить неравенство Йенсена, то получим неравенство Йенсена для случая, где функция выпукла вверх. Единственное отличие заключается в том, что знак сравнения поменяет направление.*

- Неравенство Ляпунова: $(\mathrm{E}\,|X|^p)^{\frac{1}{p}} \leqslant (\mathrm{E}\,|X|^q)^{\frac{1}{q}}, \forall p, q : q > p > 0.$

  Доказательство. Будем пользоваться неравенством Йенсена. Возьмём функцию $g(t) = t^{\frac{q}{p}}, q > p > 0$. Можете убедиться, что $g(t)$ выпукла вниз, рассмотрев вторую производную функции. Рассмотрим неравенство Йенсена для случайной величины $|X|^p$:

  $$(\mathrm{E}\,|X|^p)^{\frac{q}{p}} \leqslant \mathrm{E}(|X|^p)^{\frac{q}{p}} = \mathrm{E}\,|X|^q$$

  Возводя обе части в степень $\frac{1}{q} > 0$, получаем неравенство Ляпунова. *ч.т.д.*

2.10 Числовые характеристики случайных векторов

В данном параграфе мы рассмотрим числовые характеристики случайных векторов.

2.10.1 Ковариация и коэффициент корреляции

Определение 2.22. ***Ковариацией*** *случайных величин X, Y будем называть число*

$$\mathrm{Cov}(X, Y) = \mathrm{E}(X - \mathrm{E}\,X)(Y - \mathrm{E}\,Y)$$

Теорема 2.19. ***Свойства ковариации***

1. $\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}\,X \cdot \mathrm{E}\,Y$
2. $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
3. $\mathrm{Cov}(X, X) = \mathrm{Var}\,X$
4. $\mathrm{Cov}(c_1X_1 + c_2X_2, Y) = c_1\,\mathrm{Cov}(X_1, Y) + c_2\,\mathrm{Cov}(X_2Y)$
5. $\mathrm{Cov}(X, c) = 0$, *в частности,* $\mathrm{Cov}(X + c, Y) = \mathrm{Cov}(X, Y)$
6. X, Y*-независимые* $\Rightarrow \mathrm{Cov}(X, Y) = 0$

Доказательство. 1. Заметим, что

$$\begin{aligned}\mathrm{Cov}(X, Y) &= \mathrm{E}(X - \mathrm{E}\,X)(Y - \mathrm{E}\,Y)\\ &= \mathrm{E}(XY - Y\,\mathrm{E}\,X - X\,\mathrm{E}\,Y + \mathrm{E}\,X\,\mathrm{E}\,Y) = \mathrm{E}(XY) - \mathrm{E}\,X \cdot \mathrm{E}\,Y.\end{aligned}$$

2. Очевидно из определения.

3. Очевидно из определений ковариации и дисперсии

4. Действительно, по линейности математического ожидания имеем

$$\begin{aligned}\mathrm{Cov}(c_1X_1 + c_2X_2, Y) &= \mathrm{E}(c_1X_1 + c_2X_2)Y - \mathrm{E}(c_1X_1 + c_2X_2)\,\mathrm{E}\,Y\\ &= c_1\,\mathrm{E}(X_1Y) + c_2\,\mathrm{E}(X_2Y) - c_1\,\mathrm{E}\,X_1\,\mathrm{E}\,Y - c_2\,\mathrm{E}\,X_2\,\mathrm{E}\,Y\\ &= c_1\,\mathrm{Cov}(X_1, Y) + c_2\,\mathrm{Cov}(X_2, Y).\end{aligned}$$

5. Очевидно из определения и только что доказанного свойства.

6. Следует из того, что для независимых случайных величин $\mathrm{E}\,XY = \mathrm{E}\,X \cdot \mathrm{E}\,Y$.

ч.т.д.

Замечание. *Для дисперсии мы формулировали свойства, связанные с суммами. С учётом введенного понятия ковариации мы можем расписать эти свойства следующим образом:*

$$\operatorname{Var}(X \pm Y) = \operatorname{Var} X + \operatorname{Var} Y \pm 2\operatorname{Cov}(X, Y)$$
$$\operatorname{Var}(X_1 + \ldots + X_n) = \sum_{j=1}^{n} \operatorname{Var}(X_j) + 2 \sum_{1 \leqslant k < j \leqslant n} \operatorname{Cov}(X_k, X_j).$$

Далее рассмотрим нормированную ковариацию.

Определение 2.23. *По абсолютному значению ковариации нельзя судить о степени взаимосвязи случайных величин, так как масштаб зависит от дисперсий.* ***Коэффициент корреляции*** *это, по сути говоря, нормированная ковариация, которая уже отражает суть взаимосвязи*

$$\rho(X, Y) = \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var} X \cdot \operatorname{Var} Y}}$$

Теорема 2.20. ***Свойства коэффициента ковариации***

1. $|\rho(X, Y)| \leqslant 1$

2. $\rho(X, Y) = 1 \Rightarrow X = aY + b, a > 0$

3. $\rho(X, Y) = -1 \Rightarrow X = -aY + b, a > 0$

Доказательство. Во-первых, неравенство верно в силу неравенство Коши-Буняковского. Во-вторых, подумаем вот о чём: найдём такое 'нормирующее' преобразование случайных величин, чтобы их математическое ожидание и дисперсия стали соответственно 0 и 1. Найдём такую величину $\widetilde{X} = aX + b$

$$\begin{cases} 0 = \mathrm{E}\,\widetilde{X} = a\,\mathrm{E}\,X + b \\ 1 = \operatorname{Var} \widetilde{X} = a^2 \operatorname{Var} X \end{cases}$$

Решая систему получаем, что $\widetilde{X} = \frac{X - \mathrm{E}\,X}{\sqrt{\operatorname{Var} X}}$, или $X = \widetilde{X}\sqrt{\operatorname{Var} X} + \mathrm{E}\,X$. Аналогичные выражения для Y. Подставим в формулу коэффициента ковариации данные X, Y

$$\rho(X, Y) = \operatorname{Cov}\left(\frac{X}{\sqrt{\operatorname{Var} X}}, \frac{Y}{\sqrt{\operatorname{Var} Y}}\right) = \operatorname{Cov}\left(\widetilde{X} + \frac{\mathrm{E}\,X}{\sqrt{\operatorname{Var} X}}, \widetilde{Y} + \frac{\mathrm{E}\,Y}{\sqrt{\operatorname{Var} Y}}\right)$$
$$= \operatorname{Cov}(\widetilde{X}, \widetilde{Y})$$

Все преобразования были выполнены по свойствам ковариации. С этим результатом проанализируем дисперсии $\operatorname{Var}(\widetilde{X} + \widetilde{Y})$, $\operatorname{Var}(\widetilde{X} - \widetilde{Y})$

$$\operatorname{Var}(\widetilde{X} + \widetilde{Y}) = \operatorname{Var} \widetilde{X} + \operatorname{Var} \widetilde{Y} + 2\operatorname{Cov}(\widetilde{X}, \widetilde{Y})) = 2 + 2\rho(X, Y) \geqslant 0$$
$$\operatorname{Var}(\widetilde{X} - \widetilde{Y}) = \operatorname{Var} \widetilde{X} + \operatorname{Var} \widetilde{Y} - 2\operatorname{Cov}(\widetilde{X}, \widetilde{Y}) = 2 - 2\rho(X, Y) \geqslant 0$$

Из этих неравенств, тоже, следует первое свойство. Далее рассмотрим случай равенства. С одной стороны, коэффициент ковариации принимает значение ± 1. С другой стороны, дисперсия $\operatorname{Var}(\widetilde{X} \pm \widetilde{Y})$ равна нулю, а значит $\widetilde{X} \pm \widetilde{Y} = const$. Текущие величины линейно зависят от изначальных, а значит возвращаясь к ним, получаем соответствующие знаку коэффициента ковариации линейные зависимости двух величин. *ч.т.д.*

2.10.2 Характеристики случайных векторов

Определение 2.24. *Пусть дан случайный вектор $X = (X_1, \dots, X_n)^T$.* ***Математическое ожидание случайного вектора*** *есть вектор математических ожиданий величин по отдельности* $\mathrm{E}\,X = (\mathrm{E}\,X_1, \dots, \mathrm{E}\,X_n)^T$

Определение 2.25. $\mathrm{Var}\,X$ - ***дисперсия случайного вектора*** *определяется как* $\mathrm{E}(X - \mathrm{E}\,X)(X - \mathrm{E}\,X)^T$. *Несложно понять, что дисперсия - это матрица из ковариаций:* $\mathrm{Var}\,X = \{\mathrm{Cov}(X_i, X_j)\}$, *поэтому дисперсию ещё называют* ***ковариационной матрицей случайного вектора****.*

Определение 2.26. *Случайные вектора X, Y называются* ***некоррелируемыми****, если матрица* $\mathrm{Cov}(X, Y) = \mathrm{E}(X - \mathrm{E}\,X)(Y - \mathrm{E}\,Y)^T$ *является нулевой.*

Замечание. *Можно запутаться и посчитать что некоррелируемость и независимость это одно и тоже, но нет. Из независимости следует некоррелируемость, а обратное (в общем случае) неверно.*

Теорема 2.21. ***Свойства мат.ожидания и дисперсии случайного вектора***

1. $\mathrm{E}(AX_1 + BX_2 + b) = A \cdot \mathrm{E}\,X_1 + B \cdot \mathrm{E}\,X_2 + b;\ \ A, B \in M_{m\times n}(\mathbb{R}), b \in \mathbb{R}^m$
2. $\mathrm{Var}\,X \geqslant 0$ - *неотрицательно определена*
3. $\mathrm{Var}\,C = 0$, *если C - вектор из констант*
4. $\mathrm{Var}(AX) = A\,\mathrm{Var}(X)A^T;\ \ A \in M_{m\times n}(\mathbb{R})$
5. *Если X, Y некоррелируемы, то* $\mathrm{Var}(X + Y) = \mathrm{Var}\,X + \mathrm{Var}\,Y$

Доказательство.
1. Тривиально следует из линейности мат.ожидания
2. Во-первых, очевидно, что $\mathrm{Var}\,X$ - симметрична в силу симметричности ковариации. Далее, возьмём произвольный ненулевой (случай нуля тривиален) вектор $t \in \mathbb{R}^n$. Рассмотрим выражение

   $$t^T\,\mathrm{Var}(X)t = t^T(\mathrm{E}(X - \mathrm{E}\,X)(X - \mathrm{E}\,X)^T)t = (t^T\,\mathrm{E}(X - \mathrm{E}\,X))((X - \mathrm{E}\,X)^T t)$$

   Во второй скобке находится та же матрица что и первой, только транспонирована. Но это матрица 1×1 - число, а значит

   $$t^T\,\mathrm{Var}(X)t = C \cdot C^T = C^2 \geqslant 0$$

   Это неравенство выполняется для произвольного вектора t, а значит матрица $\mathrm{Var}\,X$ - неотрицательно определена
3. Вектор состоит из постоянных величин, следовательно все ковариации равны нулю, следовательно ковариационная матрица нулевая
4. По определению

   $$\begin{aligned}\mathrm{Var}(AX) &= \mathrm{E}[(AX - \mathrm{E}\,AX)(AX - \mathrm{E}\,AX)^T] = \mathrm{E}[(A(X - \mathrm{E}\,X))((X - \mathrm{E}\,X)^T A^T)] \\ &= \mathrm{E}[A(X - \mathrm{E}\,X)(X - \mathrm{E}\,X)^T A^T] = A\,\mathrm{Var}(X)A^T\end{aligned}$$

5. По определению

$$\begin{aligned}\operatorname{Var}(X+Y) &= \mathrm{E}[(X+Y-\mathrm{E}\,X-\mathrm{E}\,Y)(X+Y-\mathrm{E}\,X-\mathrm{E}\,Y)^T]\\ &= \mathrm{E}\left[[(X-\mathrm{E}\,X)+(Y-\mathrm{E}\,Y)][(X-\mathrm{E}\,X)^T+(Y-\mathrm{E}\,Y)^T]\right]\\ &= \mathrm{E}(X-\mathrm{E}\,X)(X-\mathrm{E}\,X)^T+\mathrm{E}(Y-\mathrm{E}\,Y)(Y-\mathrm{E}\,Y)^T\\ &+\mathrm{E}(X-\mathrm{E}\,X)(Y-\mathrm{E}\,Y)^T+\mathrm{E}(Y-\mathrm{E}\,Y)(X-\mathrm{E}\,X)^T\end{aligned}$$

Поскольку X, Y - независимы, следовательно последние два слагаемых равны нулю. Итого

$$\begin{aligned}\operatorname{Var}(X+Y) &= \mathrm{E}(X-\mathrm{E}\,X)(X-\mathrm{E}\,X)^T+\mathrm{E}(Y-\mathrm{E}\,Y)(Y-\mathrm{E}\,Y)^T\\ &= \operatorname{Var}X+\operatorname{Var}Y\end{aligned}$$

ч.т.д.

Пусть случайный вектор Y линейно выражен через вектор X, то есть: $Y = AX + b$, где $A \in Mn \times n(\mathbb{R})$; $b, X, Y \in \mathbb{R}^n$. Известны мат.ожидание $\mathrm{E}\,X$ и дисперсия $\operatorname{Var}X$ вектора X. Тогда, по свойствам, характеристики Y примут вид:

$$\mathrm{E}\,Y = A\,\mathrm{E}\,X + b\operatorname{Var}Y = A\operatorname{Var}(X)A^T$$

На одном примере посчитаем математическое ожидание и дисперсию вектора.

Пример 2.43. *$X \sim \mathrm{Poly}(n,p), p = (p_1,\ldots,p_m)^T$. Воспользуемся тем, что $\mathrm{Poly}(n,p) \sim \sum_{i=1}^n Y_i$; $Y_i \sim \mathrm{Poly}(1,p)$. Тогда $\mathrm{E}\,X = \sum_{i=1}^n \mathrm{E}\,Y_i$; $\operatorname{Var}X = \sum_{i=1}^n \operatorname{Var}Y_i$. Математическое ожидание Y_i примет вид*

$$\mathrm{E}\,Y_i = (p_1, p_2, \ldots p_m)^T = p$$

Так как каждая величина вектора ведёт себя как $\mathrm{Bern}(p_j)$
Теперь построим ковариационную матрицу. На диагонали расположены дисперсии бернуллиевских величин, т.е. p_jq_j; остальные элементы это $\mathrm{Cov}(y_j,y_k) = \mathrm{E}\,y_jy_k - \mathrm{E}\,y_j\,\mathrm{E}\,y_k$. Поскольку величины y_j, y_k принимают только значения 0 или 1, то их произведение принимает также значения только 0 или 1, но заметим что если произведение равно 1, то каждая из величин должна быть 1, что, вообще говоря, невозможно в контексте вектора Y_i, поскольку сумма всех величин должна быть 1 (а тут получается 2). Значит математическое ожидание произведения величин равно 0, и в итоге $\mathrm{Cov}(y_j,y_k) = 0 - p_jp_k = -p_jp_k$.
Теперь вернёмся к изначальной величине X

$$\mathrm{E}\,X = \sum_{i=1}^n \mathrm{E}\,Y_i = (np_1,\ldots,np_m)^T = np$$

$$\operatorname{Var}X = \sum_{i=1}^n \operatorname{Var}Y_i = \{\sigma_{i,j}\} = \begin{cases} np_i(1-p_i),\ i=j\\ -np_ip_j,\ i\neq j\end{cases}$$

2.10.3 Многомерное нормальное распределение

В этом пункте хочется привести несколько выкладок насчёт многомерного нормального распределения.
Начнём с вычисления математического ожидания и дисперсии. Сперва пусть

$X = (X_1, \dots, X_n)$ - стандартный гауссовский вектор. $\mathrm{E}\, X = 0$, $\mathrm{Var}\, X = E_n$ - единичная матрица.
Если линейно преобразовать вектор: $Y = AX + b$, $A \in M_{m\times n}(\mathbb{R})$, $b \in \mathbb{R}^m \Rightarrow Y \sim \mathcal{N}(b, AA^T)$.
Здесь характеристики соответственно: $\mathrm{E}\, Y = A\,\mathrm{E}\, X + b = b$; $\mathrm{Var}\, Y = A\,\mathrm{Var}(X) A^T = AA^T$.

Теорема 2.22. ***Свойство многомерного нормального распределения*** *Для многомерного нормального распределения, некоррелируемость равносильна независимости.*

Доказательство. То, что из независимости следует некоррелируемость - очевидно, потому что если X, Y - независимы, то их ковариация - нуль.
Пусть есть некоррелируемость, вектор $X \sim \mathcal{N}(\mu, \Sigma)$. Тогда $\mathrm{Var}\, X = \Sigma$ - диагональная матрица и она имеет вид $\mathrm{Var}\, X = diag(\Sigma_1, \dots, \Sigma_n)$. Выразим X через стандартный гауссовский вектор

$$X = \sqrt{\Sigma} \cdot U + \mu, \quad U \sim \mathcal{N}(0, E_n)$$

Поскольку Σ уже диагонализирована, то корень из Σ это просто $diag(\sqrt{\Sigma_1}, \dots, \sqrt{\Sigma_n})$. Тогда запишем матричное выражение

$$\begin{pmatrix} X_1 \\ X_2 \\ \dots \\ X_n \end{pmatrix} = \begin{pmatrix} \sqrt{\Sigma_1} & 0 & \dots & 0 \\ 0 & \sqrt{\Sigma_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \Sigma_n \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \\ \dots \\ U_n \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \\ \dots \\ \mu_n \end{pmatrix} = \begin{pmatrix} \sqrt{\Sigma_1} U_1 + \mu_1 \\ \sqrt{\Sigma_2} U_2 + \mu_2 \\ \dots \\ \sqrt{\Sigma_n} U_n + \mu_n \end{pmatrix}$$

Откуда $X_i = \sqrt{\Sigma_i} U_i + \mu_i$. В стандартном гауссовском векторе все компоненты имеют распределение $\mathcal{N}(0, 1)$. Несложно проверить подстановкой плотностей, что тогда все компоненты независимы. Значит и все X-ые компоненты независимы, поскольку домножение на константу и добавление константы не влияет на независимость.

ч.т.д.

Замечание. *Рассуждения выше, вообще говоря, работают пока $\Sigma_i \neq 0$. Но если это так, то в выражении для X_i компонента стандартного гауссовского вектора пропадает и остаётся константа μ_i. Предоставляется читателю в качестве упражнения проверить что и в этом случае независимость выполняется.*

2.11 Условные распределения

2.11.1 Условные математическое ожидание и дисперсия

Определение 2.27. ***Условным распределением*** *назовём распределение случайных величин при условии, что другие случайные величины принимают некоторые значения. Определим для обоих случаев:*

- $X : \Omega \to \mathbb{R}^k$, $Y : \Omega \to \mathbb{R}^m$ *– случайные вектора состоящие из дискретных случайных величин и* $x \in \mathbb{R}^k$, $y \in \mathbb{R}^m$. *Пусть распределение случайного вектора* $(X, Y)^T$ *задаётся функцией вероятности:* $p_{X,Y}(x, y) = \mathrm{P}(X = x, Y = y)$, *откуда также известна функция вероятности для* Y*:* $p_Y(y)$. *Для* $y_0 \in \mathbb{R}^m : p_Y(y_0) > 0$ *функция условной вероятности будет выглядеть*

$$p_{X|Y}(x|y_0) = \frac{p_{X,Y}(x, y_0)}{p_Y(y_0)}$$

$$\mathrm{P}(X = x | Y = y_0) = \frac{\mathrm{P}(X = x, Y = y_0)}{\mathrm{P}(Y = y_0)}$$

- *X, Y – случайные вектора состоящие из абсолютно непрерывных случайных величин и $x \in \mathbb{R}^k$, $y \in \mathbb{R}^m$. Пусть распределение для вектора $(X, Y)^T$ задаётся плотностью $p_{X,Y}(x, y)$, известна плотность распределение для Y: $p_Y(y)$. Для $y_0 \in \mathbb{R}^m : p_Y(y_0) > 0$ плотность условного распределения будет выглядеть*

$$p_{X|Y}(x|y_0) = \frac{p_{X,Y}(x, y_0)}{p_Y(y_0)}$$

Используя определение условного распределения, теперь может записать аналог формулы полной вероятности для случайной величины.

- X, Y - дискретные случайные величины

$$\mathrm{P}(X = k) = \sum_{i=1}^{n} \mathrm{P}(X = k|Y = y_i)\,\mathrm{P}(Y = y_i)$$

- X, Y - абсолютно непрерывные случайные величины с плотностями $p_X(x), p_Y(y)$ соответственно

$$p_X(x) = \int_{-\infty}^{+\infty} p_{X|Y}(x|y)p_Y(y)dy$$

Определение 2.28. ***Условное математическое ожидание*** *получается суммированием значений распределения по условному распределению:*

- *X, Y - дискретные случайные величины*

$$\mathrm{E}(X|Y = y_0) = \sum_{i=1}^{n} x_i \cdot p_{X|Y}(x_i|y_0)$$

- *X, Y - непрерывные случайные величины*

$$\mathrm{E}(X|Y = y_0) = \int_{-\infty}^{+\infty} x \cdot p_{X|Y}(x|y_0)dx$$

Если из контекста понятно y_0, то равенство Y ему будет опускаться.

Теорема 2.23. ***Свойства условного математического ожидания***

1. $\mathrm{E}(X|X = x_0) = x_0$
2. *X, Y - независимые* $\Rightarrow \mathrm{E}(X|Y) = \mathrm{E}\,X$
3. *Формула полной вероятности:* $\mathrm{E}\,X = \mathrm{E}_Y(\mathrm{E}(X|Y))$

Определение 2.29. ***Условная дисперсия***
$\mathrm{Var}(X|Y) = \mathrm{E}([X - \mathrm{E}(X|Y)]^2|Y) = \mathrm{E}(X^2|Y) - [\mathrm{E}(X|Y)]^2$

Теорема 2.24. $\mathrm{Var}\,Y = \mathrm{E}[\mathrm{Var}(Y|X)] + \mathrm{Var}\,\mathrm{E}(Y|X)$

Доказательство. $\mathrm{Var}\,Y = \mathrm{E}\,Y^2 - (\mathrm{E}\,Y)^2 = \mathrm{E}\,\mathrm{Var}(Y|X) + (\mathrm{E}\,\mathrm{E}^2(Y|X) - \mathrm{E}^2\,\mathrm{E}(Y|X))$
$\mathrm{E}\,Y^2 = \mathrm{E}(\mathrm{E}(Y^2|X)) = \mathrm{E}(\mathrm{Var}(Y|X) + (\mathrm{E}(Y|X))^2)\,\mathrm{E}\,Y = \mathrm{E}\,\mathrm{E}(Y|X)$ *ч.т.д.*

Теорема 2.25.

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \operatorname{Var} X & \operatorname{Cov}(X,Y) \\ \operatorname{Cov}(Y,X) & \operatorname{Var} Y \end{pmatrix}\right)$$
$$\Rightarrow \mathrm{E}(Y|X) = \mu_2 + \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var} X} \cdot (x - \mu_1)$$

Доказательство. Рассмотрим ковариацию следующих двух векторов

$$\operatorname{Cov}\left(Y - \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var} X} \cdot X, X\right) = \operatorname{Cov}(X,Y) - \operatorname{Cov}(X,Y) = 0$$

Поскольку вектор из X, Y имеет многомерное нормальное распределение, то верно что $Y - \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var} X} \cdot X$ независимо от X. Пусть $U = Y - \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var} X} \cdot X$

$$\begin{aligned}
\mathrm{E}(Y|X) &= \mathrm{E}\left(U + \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var} X} \cdot X \Big| X\right) = \mathrm{E}(U|X) + \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var} X} \cdot X \\
&= \mathrm{E}\, U + \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var} X} \cdot X = \mathrm{E}\, Y - \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var} X} \mathrm{E}\, X + \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var} X} \cdot X \\
&= \mu_2 + \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var} X}(X - \mu_1)
\end{aligned}$$

Избавились от условности в математическом ожидании поскольку величины независимы и получили искомое выражение. *ч.т.д.*

2.12 Упражнения

1. Случайная точка B равномерно распределена на окружности $x^2 + (y-a)^2 = r^2$ с центром в точке $O = (0, a)$. Пусть $C = (\xi, 0)$ – точка пересечения прямой OB с осью абсцисс. Найти функцию распределения и плотность случайной величины ξ.

2. Пусть $X \sim \mathrm{Cauchy}(0,1)$. Найти функции распределения и плотности случайных величин $Y_1 = X^2/(1+X^2)$, $Y_2 = 1/(1+X^2)$, $Y_3 = 2X/(1-X^2)$, $Y_4 = 1/X$.

3. Случайная величина X имеет непрерывную функцию распределения $F(t) = \mathrm{P}(X \leqslant t)$. Показать, что $Y = F(X)$ распределена равномерно на [0; 1].

4. Пусть $X \sim U[0;1]$ и определим квантильную функцию следующим образом:

$$F^{-1}(y) = \sup\{x : F(x) \leqslant y\}, \quad 0 \leqslant y \leqslant 1,$$

   где F – функция распределения (необязательно непрерывная) некоторого вероятностного закона. Показать, что случайная величина $Y = F^{-1}(X)$ имеет функцию распределения F.

5. Случайный вектор (X, Y) принимает значения (0; 0), (0; 1), (1; 0), (1; 1), (0; 2), (1; 2), каждое с вероятностью 1/6. Найти распределение случайной величины $X + Y$. Являются ли случайные величины X, Y независимыми?

6. Случайный вектор (X, Y) имеет равномерное распределение в треугольнике с вершинами в точках (-1;0), (0; 1), (1; 0). Найти распределение случайной величины $(X+Y)/2$.

7. Пусть $\xi \sim \mathcal{N}(0;1)$. Найти распределение вектора $\eta = (\ln|\xi|, \operatorname{sign} \xi)$.

8. Совместное распределение X, Y является равномерным в единичном круге. Найти $\mathrm{P}(|X| < 3/4, |Y| < 3/4)$.

9. Пусть $p(x, y) = c(x+y)\mathbb{1}((x, y) \in (0, 1)^2)$ – совместная плотность ξ, η. Найти константу c и плотность случайной величины $\nu = \max(\xi, \eta)$.

10. Неотрицательные случайные величины X_1, X_2 независимы и имеют одну и ту же плотность $p(t)$. Найти плотность совместного распределения $Y_1 = X_1 - X_2$, $Y_2 = \sqrt{X_1^2 + X_2^2}$.

    Подсказка: тут можно применить формулу для вычисления плотности преобразования случайных величин, только для двумерного случая:

    $$p_{Y_1,Y_2}(y_1, y_2) = p_{X_1,X_2}(x_1(y_1, y_2), x_2(y_1, y_2))|\det D(x_1(y_1, y_2), x_2(y_1, y_2))|,$$

    где $D(.)$ – матрица Якоби.

11. К переговорному пункту с двумя кабинами подошли три клиента: первый и второй клиенты заняли кабины №1 и №2 соответственно, а третий остался ждать. Предполагая, что времена τ_1, τ_2, τ_3 разговоров клиентов независимы и одинаково распределены с экспоненциальным законом с параметром λ, найти вероятность того, что третий клиент закончит разговор раньше первого или второго.

12. Пусть X_1, X_2, X_3 – независимые случайные величины, распределенные равномерно на множестве целых чисел от $-n$ до n. Пусть $A(x) = a_0x^2 + a_1x + a_2$, для которого $A(-1) = X_1$, $A(0) = X_2$, $A(1) = X_3$. Найти вероятность P_n, что числа a_0, a_1, a_2 целые?

13. В N ячеек независимо бросают частицы: для каждой частицы вероятность попадания в i-ю ячейку равняется $p_i = 1/N$. Обозначим через $\nu_1 < \ldots < \nu_N$ номера бросков, при которых частицы попадают в пустые ячейки; положим $\tau_1 = \nu_1 = 1$, $\tau_k = \nu_k - \nu_{k-1}$ при $k \geqslant 2$ и обозначим через θ_k номер ячейки, в которую попадает частица при ν_k бросании. Найти:

    - совместное и одномерные распределения величин τ_2, τ_3;
    - совместное и одномерные распределения величин $\tau_2, \ldots, \tau_N$ (являются ли эти величины независимыми?);
    - совместное распределение $\theta_1, \ldots, \theta_N$.

14. Дана последовательность $(X_n)_{n=1}^{\infty}$ независимых случайных величин, равномерно распределённых на $[0; 1]$. Положим

    $$S_n = \sum_{i=2}^{n} |X_{i-1} - X_i|.$$

    Найти $\mathrm{E}S_n$ и $\mathrm{Var}S_n$.

15. По маршруту ходят N автобусов без кондуктора. В каждом автобусе имеется касса, в которой перед выходом в рейс было r билетов. Всего эти автобусы перевезли n пассажиров. Найти математическое ожидание числа X пассажиров, которым не досталось билетов, предполагая, что каждый пассажир независимо от остальных может сесть в любой из автобусов с вероятностью $1/N$.

16. На бесконечный лист клетчатой бумаги (сторона клеточки равняется единице) случайно бросается круг единичного радиуса. Считая, что центр круга равномерно распределен на том единичном квадрате, на который он попал, найти математическое ожидание числа U точек с целочисленными координатами, покрытых этим кругом.

17. Пусть случайные величины $X_1, X_2 \sim \mathcal{N}(0,1)$ и они независимы. Найти совместное распределение случайных величин $Y_1 = aX_1 + bX_2$, $Y_2 = aX_1 - bX_2$ при $a, b \neq 0$. Несколько слов о последней задаче с пары, так как мы с ней немного поспешили. Нам дано, что $X, Y \sim \text{Exp}(\lambda)$ и они независимы. Найти $p_{X|(X+Y=z)}(x)$.

18. Найти $\text{E}(X|X+Y)$ и $\text{Var}(X|X+Y)$ в условиях предыдущей задачи.

19. Пусть $X_1, \ldots X_N$ – независимы случайные величины, распределенные по Пуассону с параметром $\lambda > 0$. Найти $\text{E}(X_1 + \ldots + X_k = m|X_1 + \ldots + X_N = n)$.

20. $X_1, X_2, X_3 \sim \mathcal{N}(0,1)$ и они независимы. Найти распределение случайной величины $Y = \frac{X_1 + X_2 X_3}{\sqrt{1+X_3^2}}$.

21. В схеме Бернулли с вероятностями успеха p и неудачи $q = 1-p$ найти математическое ожидание числа испытаний до первого появления цепочки из трёх единиц. В частности, вычислить указанное математическое ожидание при $p = 1/2$.

22. Пусть точки $A_1, \ldots A_n$ $(n \geqslant 2)$ независимы и имеют равномерное распределение на окружности радиуса r. Пусть $A_{(1)} = A_1$, $A_{(2)}$, $\ldots$, $A_{(n)}$ – точки A_1, $\ldots$, A_n, расположенные в том порядке, в котором они встречаются при обходе окружности по часовой стрелке. Пусть ξ_i – длина дуги $A_{(i)}A_{(i+1)}$ (ξ_n – длина дуги $A_{(n)}A_{(1)}$). Найти совместное распределение величин ξ_1, $\ldots$, ξ_k при $k \leqslant n$.

23. Пусть $X \sim U[0,1]$, то есть $p_X(t) = \mathbb{1}(t \in [0,1])$, и $Y = -\ln(1-X)$. Найти распределение случайной величины Y.

24. Случайный вектор (X, Y) имеет равномерное распределение на единичном круге, то есть плотность $p_{(X,Y)}(x,y) = \pi^{-1}\mathbb{1}(x^2 + y^2 \leqslant 1)$. Найти функцию распределения случайной величины X и $\mathsf{P}(|X| < 3/4, |Y| < 3/4)$.

25. Неотрицательный случайные величины X, Y независимы и имеют одну и ту же плотность $p(x)\mathbb{1}(x \geqslant 0)$. Найти плотность совместного распределения $U = X - Y$, $W = \sqrt{X^2 + Y^2}$.

26. Найти плотность распределения суммы n независимых случайных величин, каждая из которых имеет экспоненциальное распределение с параметром $\lambda > 0$.

27. Пусть $X \sim U[0, 2\pi]$, $U = \cos X$, $W = \sin X$. Найти $\text{E}U$, $\text{E}W$, $\text{E}UW$. Являются ли случайные величины независимыми и почему?

28. Случайные величины X_1 и X_2 независимы и нормально распределены с параметрами $(0,1)$. Являются ли независимыми случайные величины $Y_1 = X_1 + X_2$ и $Y_2 = X_1 - X_2$?

29. В партии из n изделий, каждое из которых независимо от остальных удовлетворяет стандарту с вероятностью p и не удовлетворяет с вероятностью $1-p$. Система контроля качества состоит из двух независимых проверок: на k-ой проверке изделие, удовлетворяющее стандарту, отбраковывается с вероятностью β_k, а бракованное изделие принимается в вероятностью α_k. Изделие принимается, если оно прошло обе проверки. За каждое изделие, удовлетворяющее стандарту и прошедшее проверку,

предприятие получает a рублей, за изделие, прошедшее проверку, но не удовлетворяющее стандарту, уплачивается штраф b рублей, за изделие, не прошедшее проверку, уплачивается штраф c рублей. Найти математическое ожидание прибыли за партию из n изделий.

30. Распределение случайной величины ξ определяется формулами $P\{\xi = k\} = C/k(+ 1)$,$k = 1, 2, \ldots$. Найти:

    (a) постоянную C;

    (b) $\mathrm{P}\{\xi \leqslant 3\}$;

    (c) $\mathrm{P}\{n_1 \leqslant \xi \leqslant n_2\}$

31. Случайные величины ξ и η независимы. Найти $\mathrm{P}\{\xi = \eta\}$, если:

    (a) ξ и η имеют одно и то же дискретное распределение $\mathrm{P}\{\xi = x_k) = \mathrm{P}\{\xi = x_k\} = p_k$, $k = 0, 1, \ldots$.

    (b) функция распределения ξ непрерывна.

32. Случайные величины $\xi_1, \cdots, \xi_n$ независимы и распределены показательно с одинаковым параметром α. Найти плотность распределения величин:

    (a) $\eta_1 = \max\{\xi_1, \cdots, \xi_n\}$

    (b) $\eta_2 = \min\{\xi_1, \cdots, \xi_n\}$.

33. Ввести на сфере в качестве координат широту и долготу, считая их изменяющимися в отрезках $[-\pi/2, \pi/2]$ и $[-\pi, \pi]$ соответственно. Найти плотность $p_\eta(x)$ распределения широты η случайной точки, имеющей равномерное распределение на сфере.

34. Распределение дискретной случайной величины ξ определяется формулами $P\{\xi = k\} = \frac{4}{k(k+1)(k+2)}, k = 1, 2, \ldots$. Найти математическое ожидание случайной величины ξ.

35. Независимые случайные величины $\xi_1, \xi_2, \cdots, \xi_\eta$ положительны и имеют одинаковое невырожденное распределение. Обозначим $\eta_k = \frac{\eta_k}{\eta_1 + \ldots + \eta_n}$. Найти:

    (a) математическое ожидание η_k

    (b) коэффициент корреляции η_k и η_l;

    (c) коэффициент корреляции $\eta_1 + \cdots + \eta_k$ и $\eta_1 + \cdots + \eta_l$

36. Случайные величины $\xi_1, \ldots, \xi_n$ независимы; $\operatorname{Var}\xi_i = \sigma_i^2, i = 1, \ldots, n$. При каких $c_1, \cdots, c_n$, удовлетворяющих условиям $c_k \geqslant 0, c_1 + \ldots + c_n = 1$, случайная величин $\eta_n = c_1\xi_1 + \cdots + c_n\xi_n$ имеет минимальную дисперсию? Найти минимальную дисперсию.

37. Случайная величина ξ имеет непрерывную функцию распределения $F(x)$, случайная величина X принимает только значения 0 и 1: $\mathrm{P}\{X = 1\} = a, \mathrm{P}\{X = 0\} = 1 - a$. Указать совместные распределения ξ и X, при которых достигаются экстремальные значения $\mathrm{E}\,\xi X$, и найти эти экстремальные значения.

38. Случайные величины ξ и η независимы и имеют следующее распределение:

$$\mathrm{P}(\xi = k) = \mathrm{P}(\eta = k) = p(1-p)^{k-1}, \quad k \in \mathbb{N}.$$

    Найти:

(a) $\mathrm{P}\{\xi = \eta\}$;
(b) $\mathrm{P}\{\xi > \eta\}$;
(c) $\mathrm{P}\{\xi < \eta\}$;
(d) $\mathrm{P}\{\xi = k | \xi > \eta\}$;
(e) $\mathrm{P}\{\xi = k | \xi < \eta\}$;
(f) $\mathrm{P}\{\xi = k | \xi = \eta\}$;
(g) $\mathrm{P}\{\xi = k | \xi + \eta = l\}$;
(h) $\mathrm{E}\{\xi | \xi + \eta = l\}, l \geqslant 2$;

39. Найти распределение целочисленной неотрицательной случайной величины ξ, если:

    (a) $\mathrm{P}\{0 < \xi < \infty\} = 1, \mathrm{P}\{\xi = k+1 | \xi > k\} = p, k = 0, 1, \ldots$;
    (b) $\mathrm{P}\{\xi \geqslant 0\} = 1, \mathrm{P}\{\xi = k+1 | \xi \in \{k, k+1\}\} = c < 1/2, k = 0, 1, \ldots$;
    (c) $\mathrm{P}\{\xi \geqslant 0\} = 1, \mathrm{P}\{\xi = k+1 | \xi \in \{k, k+1\}\} = \frac{r}{k+r+1}, r > 0, k = 0, 1, \ldots$.

40. Случайная величина ξ имеет нормальное распределение с нулевым математическим ожиданием единичной дисперсией. Найти $\mathrm{E}\,\xi \cos \xi, \mathrm{E}\,\frac{\xi}{1+\xi^2}, \mathrm{E} \sin \xi$

41. Случайная величина ξ имеет нормальное распределение с математическим ожиданием 0 и дисперсией 1. Найти $\mathrm{E} \cos \xi, \mathrm{Var} \cos \xi$.

42. Случайные величины ξ и η независимы и имеют нормальное распределение с математическим ожиданием 0 и дисперсией 1. Найти

$$\mathrm{P}\{|\xi - \eta| \leqslant 1\}.$$

43. Случайные величины ξ_1, ξ_2, ξ_3 независимы и нормально распределены с параметрами (1, 1), (2, 5), (0, 7) соответственно. Найти:

    (a) $\mathrm{P}\{2\xi_1 - \xi_2 < 0\}$;
    (b) $\mathrm{P}\{-3 < 2\xi_1 - \xi_2 < 5\}$;
    (c) $\mathrm{P}\{1 < 2\xi_1 - \xi_2 + \xi_3 < 4\}$.

44. Случайный вектор (η_1, η_2) имеет нормальное распределение с $\mathrm{E}\,\eta_1 = \mathrm{E}\,\eta_2 = 0$ и матрицей коварияций

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \gamma \\ \gamma & \sigma_2^2 \end{pmatrix}.$$

Найти распределение вектора $(c_1\eta_1, c_2\eta_2)$ при $c_1, c_2 \neq 0$.

45. Случайные величины ξ_1 и ξ_2 независимы и имеют нормальное распределение, $\mathrm{E}\,\xi_1 = \mathrm{E}\,\xi_2 = 0, \mathrm{Var}\,\xi_1 = \mathrm{Var}\,\xi_2 = 1$. Доказать, что случайные величины $\xi_1\xi_2$ и $\frac{1}{2}(\xi_1^2(\sigma_1, \sigma_2 + \gamma) - \xi_2^2(\sigma_1\sigma_2 - \gamma))$ одинаково распределены.

46. Пусть $\xi_1, \ldots, \xi_n$ - независимые одинаково распределенные случайные величины, $\mathrm{P}(\xi_i = 1) = \mathrm{P}(\xi_i = -1) = 1/2, i = 1, 2, ..., n$. Найти распределение случайной величины

$$\eta_n = \prod_{i=1}^{n} \xi_i.$$

47. Можно ли подобрать постоянную c так, чтобы функция cx^{-3} определяла плотность распределения вероятностей на:

 (a) луче $[1, +\infty)$;
 (b) луче $[0, +\infty)$;
 (c) отрезке $[-2, -1]$.

48. Пусть $\xi_1, \xi_2, \dots$ – последовательность независимых одинаково распределенных случайных величин, принимающих значения 0 и 1 с вероятностью $1/2$ каждое. Найти распределение случайной величины

$$\eta_n = \sum_{k=1}^{\infty} \frac{\xi_k}{2^k}$$

49. Пусть ξ - случайная величина, равномерно распределенная на отрезке [0, 1), и $\xi = \frac{\delta_1}{2} + \frac{\delta_2}{2^2} + \frac{\delta_3}{2^3} + ..., \delta_n = 0$ или 1 - двоичное разложение ξ. Доказать, что при любом натуральном n

$$\mathrm{P}(\delta_n = 0) = \mathrm{P}(\delta_n = 1) = \tfrac{1}{2}$$

и случайные величины $\delta_1, \delta_2, \dots$ взаимно независимы.

50. Случайные величины ξ и η независимы и имеют одинаковое распределение с математическим ожиданием a и дисперсией δ^2. Найти коэффициент корреляции случайных величин $\xi_1 = \alpha\xi + \beta\eta$ и $\eta_2 = \alpha\xi - \beta\eta$.

51. Пусть совместное распределение случайных величин ξ и η нормально, причем $\mathrm{E}\,\xi = \mathrm{E}\,\eta = 0$, а коэффициент корреляции ξ и η равен p. Найти коэффициент корреляции случайных величин ξ^2 и η^2

52. Доказать, что если каждая из независимых случайных величин ξ_1 и ξ_2 имеет геометрическое распределение, то случайная величина $\eta = \min\{\xi_1, \xi_2\}$ также имеет геометрическое распределение. Найти параметр этого распределения, если параметры распределений ξ_1 и ξ_2 равны соответственно p_1 и p_2

53. Пусть ξ_1 и ξ_2 – независимые случайные величины, имеющие одинаковое нормальное распределение. Доказать, что случайные величины $\xi_1 + \xi_2$ и $\xi_1 - \xi_2$ независимы.

54. Случайная величина ξ имеет нормальное распределение с математическим ожиданием α и дисперсией σ^2. Найти распределение случайной величины $\operatorname{sign}\xi$

55. Пусть $\xi_0, \xi_1, \xi_2, \dots$ – последовательность независимых одинаково распределенных случайных величин, имеющих равномерное на отрезке [0, 1] распределение. Найти распределение случайной величины

$$\eta_n = \prod_{k=0}^{n} \xi_k$$

56. Имеется n шаров, среди которых k белых и $n - k$ черных. Наудачу выбирается v шаров, где v - случайная величина, принимающая значения от 1 до n с равными вероятностями. Найти математическое ожидание числа белых шаров среди отобранных.

Глава 3

Предельные методы теории вероятностей

3.1 Простейшие приложения центральной предельной теоремы и закона больших чисел

3.1.1 Введение в методы Монте-Карло, ЦПТ, ЗБЧ

Пусть мы хотим вычислить какой-то интеграл:

$$\int\limits_A f(x_1 \dots x_n) dx_1 \dots dx_n = \frac{\mu A}{\mu A} \int\limits_A f(x_1 \dots x_n) dx_1 \dots dx_n = \mu A \cdot E_{U(A)} f(X_1 \dots X_n)$$

$$\Rightarrow I_n = \frac{1}{n} \sum_{i=1}^{n} f(X_{i1}, \dots X_{in}), (X_{i1}, \dots X_{in}) - -i.i.d \sim U(A)$$

Тогда по закону больших чисел $I_n \approx Ef(X_1 \dots X_n)$ Ну и в целом методы, когда мы генерируем много каких-то случайных величин, аггрегируя результат, называются **методами Монте-Карло**.

Замечание 3.1. *У нас уже была теорема Муавра-Лапласа, которая утверждала, что*

$$S_n = \sum_{i=1}^{n} X_i, X_i - -i.i.d, X_i \sim Bern(p)$$

$$\sup_{x \in \mathbb{R}} \left| \mathrm{P}\left(\frac{S_n - np}{\sqrt{npq}} \leqslant x \right) - \Phi(x) \right| \underset{n \to \infty}{\longrightarrow} 0$$

Теорема 3.1. *Центральная предельная теорема. ЦПТ*

$$S_n = \sum_{i=1}^{n} X_i, X_i - -i.i.d, \exists \operatorname{E} X, \operatorname{Var} X \Rightarrow$$

$$\sup_{x \in \mathbb{R}} \left| \mathrm{P}\left(\frac{S_n - n \operatorname{E} X_1}{\sqrt{n \operatorname{Var} X}} \leqslant x \right) - \Phi(x) \right| \underset{n \to \infty}{\longrightarrow} 0$$

Теорема 3.2. *Слабый закон больших чисел. ЗБЧ $X_1, X_2, \dots, X_n$ - случайные величины, независимые в совокупности случайные величины, которые имеют одинаковое распределение и существует $\operatorname{E} X_1 = \mu$. Тогда*

$$S_n = \sum_{k=1}^{n} \Rightarrow \frac{S_n}{n} \xrightarrow{\mathrm{p}} \mu$$

Пример 3.1. *(Игла Бюффона) В первой главе разбиралась задаче об игле Бюффона и настало время воспользоваться этим результатом. Воспроизведём опыт n раз и каждому броску будет сопоставлять случайную величину $X_i = \mathbb{1}$(игла попала на прямую). Все броски, а значит и случайные величины, независимы, и все броски симулируют одинаковую случайную величину. Поскольку мы знаем теоретическую вероятность, то можем посчитать мат.ожидание: $\mathrm{E}\, X_1 = \frac{4l}{\pi d}$. Воспользуемся ЗБЧ - при больших n: $\frac{S_n}{n} \approx \frac{4l}{\pi d}$. Тогда*

$$\pi \approx \frac{n}{S_n}\frac{4l}{d}$$

l, d - даны по условию, $\frac{S_n}{n}$ считаем ручками воспроизводя опыт и получаем примерное значение числа π.

Пример 3.2. *(Объём n-мерной фигуры) Пусть на $\mathbb{R}^n$ аналитически задана ограниченная фигура $\Phi = F(x_1, \ldots, x_n)$. Поскольку фигура ограничена, то это значит, что можно подобрать такой n-мерный куб с длиной c, что фигура полностью в нём содержится. Пусть геометрическое место точек этого куб аналитически задаётся системой*

$$X \in \mathbb{R}^n; X \in Cube_n(c) \Leftrightarrow \begin{cases} a_1 \leqslant x_1 \leqslant b_1 \\ a_2 \leqslant x_2 \leqslant b_2 \\ \ldots \\ a_n \leqslant x_n \leqslant b_n \end{cases}$$

Понятно что $\forall i : b_i - a_i = c$. Будем генерировать координаты $\{x_1, \ldots x_n\}$ точки $X \in \mathbb{R}^n$ равномерно на кубе, то есть

$$\begin{gathered} x_1 \sim \mathrm{U}[a_1, b_1] \\ x_2 \sim \mathrm{U}[a_2, b_2] \\ \ldots \\ x_n \sim \mathrm{U}[a_n, b_n] \end{gathered}$$

Пусть мы сгенерировали m таких точек, будем сопоставлять каждой точке случайную величину $X_i = \mathbb{1}$(точка внутри Φ). По ЗБЧ, при больших m

$$\frac{S_m}{m} \approx \frac{V_\Phi}{V_{Cube_n}} \Rightarrow V_\Phi \approx \frac{S_m}{m} \cdot V_{Cube_n};$$
$$V_{Cube_n} = c^n \Rightarrow V_\Phi \approx \frac{S_m}{m} c^n$$

3.2 Вероятностные сходимости

Зададим на вероятностном пространстве $(\Omega, \mathcal{F}, \mathrm{P})$ случайную величину X, а также последовательность случайных величин $\{X_n\}_{n=1}^{\infty}$. Тогда, при некоторых условиях, можно определить сходимость последовательности случайных величин к случайной величине X, подобно тому как это делается в курсе математического анализа.

Определение 3.1. ***Сходимость почти наверное (almost surely)**: $X_n \xrightarrow{\text{a.s.}} X$, если:*

$$\mathrm{P}(\{\omega | \lim_{n\to\infty} X_n(\omega) \to X(\omega)\}) = 1 \Leftrightarrow \mathrm{P}(\{\omega | \lim_{n\to\infty} X_n(\omega) \not\to X(\omega)\}) = 0$$

Определение 3.2. ***Сходимость по вероятности***: $X_n \xrightarrow{\text{p}} X$, *если*

$$\forall \varepsilon > 0 : \mathrm{P}(|X_n - X| \geqslant \varepsilon) \to 0 \Leftrightarrow \mathrm{P}(|X_n - X| < \varepsilon) \to 1$$

Определение 3.3. ***Сходимость в среднем порядке*** p *(Lp)*: $X_n \xrightarrow{\text{Lp}} X$, *если*

$$\mathrm{E}\,|X_n - X|^p \to 0$$

Определение 3.4. ***Сходимость по распределению***: $X_n \xrightarrow{\text{d}} X$, *если*

$$\forall x \in \{x_0 | F(x_0) - continuous\} : F_{X_n}(x) \to F_X(x)$$

Замечание. *Если предельная функция распределения $F_X(x)$ непрерывна, то можно говорить о её равномерной сходимости последовательности функция F_{X_n} к F_X:*

$$\sup(F_{X_n}(x) - F_X(x)) \to 0 \Leftrightarrow \sup_{x_1 < x_2}(F_{X_n}(x_2) - F_{X_n}(x_1) - [F_X(x_2) - F_X(x_1)]) \to 0$$

В дискретном случае, если множества значений случайных величин (носители вероятности) последовательности совпадает с носителем предельной случайной величины, то есть сходимость функции вероятности:

$$\mathrm{P}(X_n = x_k) = p_n(x_k) \to p(x_k) = \mathrm{P}(X = x_k)$$

Теорема 3.3. *(Свойство сходимости по вероятности) Пусть имеются последовательности случайных величин, которые по отдельности сходятся по вероятности:*

$$X_{1,n} \xrightarrow{\text{p}} X_1$$
$$X_{2,n} \xrightarrow{\text{p}} X_2$$
$$\cdots$$
$$X_{m,n} \xrightarrow{\text{p}} X_m$$

Тогда для функции $g : \mathbb{R}^m \to \mathbb{R}$, непрерывной на $\mathbb{R}^m$ верно:

$$g(X_{1,n}, X_{2,n}, \ldots, X_{m,n}) \xrightarrow{\text{p}} g(X_1, X_2, \ldots, X_m)$$

Доказательство. То, что $g(x_1, \ldots x_m)$ непрерывна в точке $x_0 = (x_{01}, \ldots x_{0m})$ означает что:

$$\forall \varepsilon > 0\ \exists \delta > 0 : \forall x \in \mathbb{R}^m\, ||x - x_0||_{\mathbb{R}^m} < \delta \Rightarrow |g(x) - g(x_0)| < \varepsilon.$$

Заметим что, если $\forall i \in \{1, \ldots, m\} : |x_i - x_{0i}| < \frac{\delta}{\sqrt{m}}$, то тогда $||x - x_0|| < \delta$.
Обозначим вектора $X_n = (X_{1,n}, X_{2,n}, \ldots, X_{m,n}), X = (X_1, X_2, \ldots, X_m)$. Далее воспользуемся тем, что если $A \Rightarrow B$, то $\mathrm{P}(A) \leqslant \mathrm{P}(B)$ (утверждение B 'шире' чем A):

$$\mathrm{P}(|g(X_n) - g(X)| < \varepsilon) \geqslant \mathrm{P}(||X_n - X|| < \delta) \geqslant \mathrm{P}\left(\forall i\ |X_{i,n} - X_i| < \frac{\delta}{\sqrt{m}}\right) =$$
$$= 1 - \mathrm{P}\left(\exists i\ |X_{i,n} - X_i| \geqslant \frac{\delta}{\sqrt{m}}\right) \geqslant 1 - \sum_{i=1}^{m} \mathrm{P}\left(|X_{i,n} - X_i| \geqslant \frac{\delta}{\sqrt{m}}\right) \to 1 - 0 = 1$$

Последнее неравенство было получено из того, что $\mathrm{P}(\sum\limits_{i=1}^{n} A_i) \leqslant \sum\limits_{i=1}^{n} \mathrm{P}(A_i)$.

Получили в итоге:

$$1 \geqslant \mathrm{P}(|g(X_n) - g(X)| < \varepsilon) \geqslant 1 \quad (n \to \infty)$$
$$\mathrm{P}(|g(X_n) - g(X)| < \varepsilon) \to 1$$

Что и означает сходимость по вероятности. *ч.т.д.*

Замечание. *Функцию g можно взять любую и интересными частными случаями являются $g(x_1, x_2) = x_1 + x_2$; $g(x_1, x_2) = x_1 \cdot x_2$; $g(x_1, x_2) = \frac{x_1}{x_2}$ (последнее, конечно же, с нюансом). Это позволяет заключить, что сходимость по вероятности замкнута относительно основных математических операций.*

Теорема 3.4. *(Свойства сходимости по распределению)*

1. *Если $C = const, X_n \xrightarrow{\text{d}} C \Rightarrow X_n \xrightarrow{\text{p}} C$*

2. *Если $C = const, X_n \xrightarrow{\text{d}} X,\ Y_n \xrightarrow{\text{p}} C \Rightarrow X_n + Y_n \xrightarrow{\text{d}} X + C$*

3. *Если $C = const, X_n \xrightarrow{\text{d}} X,\ Y_n \xrightarrow{\text{p}} C \Rightarrow X_n \cdot Y_n \xrightarrow{\text{d}} X \cdot C$*

Доказательство. 1. Рассмотрим сходимость по вероятности:

$$\mathrm{P}(|X_n - C| < \varepsilon) = \mathrm{P}(C - \varepsilon < X_n < C + \varepsilon) \geqslant \mathrm{P}(C - \frac{\varepsilon}{2} < X_n < C + \varepsilon) =$$
$$= F_{X_n}(C + \varepsilon) - F_{X_n}(C - \frac{\varepsilon}{2}) \xrightarrow{\text{d}} F(C + \varepsilon) - F(C - \frac{\varepsilon}{2})$$

Так как $X = C$, то функция распределения имеет вид:

$$F(x) = \begin{cases} 0, & x < C \\ 1, & x \geqslant C \end{cases}$$

А значит

$$\mathrm{P}(|X_n - C| < \varepsilon) \geqslant F_{X_n}(C + \varepsilon) - F_{X_n}(C - \frac{\varepsilon}{2}) \xrightarrow{\text{d}} 1 - 0 = 1$$

Выполняется сходимость по вероятности

2. Сперва сделаем замечание, что если есть сходимость по распределению:

$$X_n \xrightarrow{\text{d}} X \Rightarrow X_n + C \xrightarrow{\text{d}} X + C$$

Значит можем 'отцепить' константу от Y_n к X_n и доказывать для случая $Y_n \xrightarrow{\text{p}} 0$, что и сделаем.
Рассмотрим такую вероятность:

$$\mathrm{P}(X_n + Y_n \leqslant t) = \mathrm{P}(X_n + Y_n \leqslant t,\ |Y_n| < \varepsilon) + \mathrm{P}(X_n + Y_n \leqslant t,\ |Y_n| \geqslant \varepsilon)$$

Рассмотрим каждую вероятность по отдельности:

$$0 \leqslant \mathrm{P}(X_n + Y_n \leqslant t,\ |Y_n| \geqslant \varepsilon) \leqslant \mathrm{P}(|Y_n| \geqslant \varepsilon) \to 0$$

Последнее равенство выполняется так как $Y_n \xrightarrow{\text{p}} 0$. Со второй посложнее. Воспользуемся тем что $|Y_n| < \varepsilon \Leftrightarrow -\varepsilon < Y_n < \varepsilon$ и $X_n + Y_n \leqslant t \Rightarrow X_n \leqslant t + \varepsilon$:

$$\mathrm{P}(X_n + Y_n \leqslant t,\ |Y_n| < \varepsilon) \leqslant \mathrm{P}(X_n \leqslant t + \varepsilon) = F_{X_n}(t + \varepsilon) \xrightarrow{\text{d}} F(t + \varepsilon)$$

Точку t берём чтобы в ней $F(x)$ непрерывна, и можем взять ε настолько маленькое, что в $t + \varepsilon$ функция также непрерывна. Значит последняя сходимость выполняется.

Чтобы ограничить вероятность снизу, воспользуемся тем, что $X_n + \varepsilon \leqslant t, |Y_n| < \varepsilon \Rightarrow X_n + Y_n \leqslant t,\ |Y_n| < \varepsilon$

$$\mathrm{P}(X_n + Y_n \leqslant t,\ |Y_n| < \varepsilon) \geqslant \mathrm{P}(X_n + \varepsilon \leqslant t,\ |Y_n| < \varepsilon) =$$
$$= \mathrm{P}(X_n \leqslant t - \varepsilon) - \mathrm{P}(X_n \leqslant t - \varepsilon,\ |Y_n| \geqslant \varepsilon) \xrightarrow{\mathrm{d}} F(t-\varepsilon) - 0$$

Итого:

$$\forall \varepsilon\ F(t-\varepsilon) \leqslant \lim_{n\to\infty} \mathrm{P}(X_n + Y_n \leqslant t) \leqslant F(t+\varepsilon) \Rightarrow$$
$$\lim_{n\to\infty} \mathrm{P}(X_n + Y_n \leqslant t) \to F(t) = F(t) + 0$$

Что и означает сходимость по распределению

3. Сделаем замечание, что если есть сходимость по распределению:

$$X_n \xrightarrow{\mathrm{d}} X \Rightarrow C \cdot X_n \xrightarrow{\mathrm{d}} C \cdot X$$

Теперь докажем, что свойство верно для $C = 0$. Проверим произведение случайных величин на сходимость по вероятности. Для начала пусть X_n - ограничена.

$$\mathrm{P}(|X_nY_n| \geqslant \varepsilon,\ |X_n| < M) \leqslant \mathrm{P}(M \cdot |Y_n| \geqslant \varepsilon) = \mathrm{P}(|Y_n| \geqslant \frac{\varepsilon}{M}) \to 0$$

Так как $Y_n \xrightarrow{\mathrm{p}} 0$. Теперь покажем для любого X_n

$$\mathrm{P}(|X_nY_n| \geqslant \varepsilon) = \mathrm{P}(|X_nY_n| \geqslant \varepsilon,\ |X_n| < M) + \mathrm{P}(|X_nY_n| \geqslant \varepsilon,\ |X_n| \geqslant M) \leqslant$$
$$\leqslant \mathrm{P}(|X_nY_n| \geqslant \varepsilon,\ |X_n| < M) + \mathrm{P}(|X_n| \geqslant M)$$

Про первую вероятность мы узнали выше. Вторую может уменьшить за счёт выбора M. Тогда устремим вторую вероятность к 0 за счёт M и получим:

$$\mathrm{P}(|X_nY_n| \geqslant \varepsilon) \leqslant \mathrm{P}(|X_nY_n| \geqslant \varepsilon,\ |X_n| < M) + \mathrm{P}(|X_n| \geqslant M) \to 0$$

Что и означает, что $X_nY_n \xrightarrow{\mathrm{p}} 0$.
Пусть теперь $Y_n \xrightarrow{\mathrm{p}} C$, тогда $Y_n - C \xrightarrow{\mathrm{p}} 0$, и вновь воспользуемся вторым свойством сходимости по распределению:

$$X_n \cdot (Y_n - C) \xrightarrow{\mathrm{p}} 0; \quad C \cdot X_n \xrightarrow{\mathrm{d}} C \cdot X$$
$$\Rightarrow X_nY_n = X_n(Y_n - C) + C \cdot X_n \xrightarrow{\mathrm{d}} 0 + C \cdot X = C \cdot X$$

ч.т.д.

Теорема 3.5. *Из сходимости одного типа может следовать другая сходимость:*

1. $X_n \xrightarrow{\mathrm{Lp}} X \Rightarrow X_n \xrightarrow{\mathrm{p}} X$
2. $X_n \xrightarrow{\mathrm{a.s.}} X \Rightarrow X_n \xrightarrow{\mathrm{p}} X$
3. $X_n \xrightarrow{\mathrm{p}} X \Rightarrow X_n \xrightarrow{\mathrm{d}} X$

Доказательство. 1. Рассмотрим вероятность $\mathrm{P}(|X_n - X| \geqslant \varepsilon)$:

$$\mathrm{P}(|X_n - X| \geqslant \varepsilon) = \mathrm{P}(|X_n - X|^p \geqslant \varepsilon^p) \leqslant \frac{\mathrm{E}\,|X_n - X|^p}{\varepsilon^p}$$

По неравенству Маркова. Если устремим обе части неравенства в бесконечность, получим сходимость по вероятности:

$$\mathrm{P}(|X_n - X| \geqslant \varepsilon) \leqslant \frac{\mathrm{E}\,|X_n - X|^p}{\varepsilon^p} \to 0$$

2. Раскроем сходимость почти наверное:

$$\mathrm{P}(X_n \not\to X) = 0$$
$$\mathrm{P}(\{\omega | X_n(\omega) \not\to X(\omega)\}) = 0$$

Множество $\{\omega | X_n(\omega) \not\to X(\omega)\}$ обозначим за B. Определим последовательность множеств $A_n = \bigcup\limits_{m>n} \{|X_n - X| \geqslant \varepsilon\}$. При таком определении верно, что $A_{n+1} \subset A_n$, и пусть $\bigcap A_n = A$. В силу непрерывности вероятностной меры вероятность $\mathrm{P}(A_n) \to \mathrm{P}(A)$. Рассмотрим такой $\omega \notin B$, тогда:

$$\omega \notin B \Leftrightarrow \forall \varepsilon > 0\ \exists N(\varepsilon) : \forall n > N : |X_n - X| < \varepsilon \Rightarrow \omega \notin A_n, n > N$$
$$\omega \notin A$$

То есть, $\forall \omega \in \overline{B} \Rightarrow \omega \in \overline{A}$. В терминах множеств это означает: $\overline{B} \subset \overline{A}$, и инвертируя отношение получаем $A \subset B$. Тогда воспользуемся свойствами вероятностной меры и получим:

$$\mathrm{P}(|X_n - X| \geqslant \varepsilon) \leqslant \mathrm{P}(A_n) \to \mathrm{P}(A) \leqslant \mathrm{P}(B) = 0$$
$$\mathrm{P}(|X_n - X| \geqslant \varepsilon) \to 0$$

Что означает сходимость по вероятности

3. Возьмём следующие случайные величины: $A_n = X$, $B_n = X_n - X$. Очевидно, что $A_n = X \xrightarrow{\mathrm{d}} X$ (не зависит от n), и $B_n = X_n - X \xrightarrow{\mathrm{p}} 0$ ($X_n \xrightarrow{\mathrm{p}} X$). Воспользуемся вторым свойством сходимости по распределению:

$$X_n = A_n + B_n = (X_n - X) + X \xrightarrow{\mathrm{d}} X$$

ч.т.д.

Замечание. *Все другие возможные импликации, в общем случае, неверны.*

Приведём примеры распределений случайных величин, когда обратные импликации не выполняются

Пример 3.3. *($p \not\Rightarrow Lp$)*

Зададим на вероятностном пространстве величины: $Y \sim U[0,1], X = 0$ и последовательность величин $\{X_n\}_{n=1}^{\infty}$:

$$X_n = \begin{cases} e^n, & Y \in [0, 1] \\ 0, & otherwise \end{cases}$$

Рассмотрим сходимость по вероятности. Для некоторого $\varepsilon > 0$:

$$\mathrm{P}(|X_n - X| \geqslant \varepsilon) \leqslant \mathrm{P}(|X_n - X| > 0) = \frac{1}{n} \to 0$$

С другой стороны, проверим сходимость в среднем порядке p. X_n имеет дискретное распределение из значений e^n, 0; с вероятностями $\frac{1}{n}, \frac{n-1}{n}$ соответственно. Тогда

$$\mathrm{E}\,|X_n - X|^p = \mathrm{E}\,|X_n|^p = \frac{e^{np}}{n} \not\to 0$$

Сходимости в среднем порядке p нет.

Замечание. *Можно показать, что здесь есть сходимость a.s и следовательно* $a.s. \not\Rightarrow Lp$

Пример 3.4. *($p \not\Rightarrow a.s.$)*

Мы знаем, что $\forall n \in \mathbb{N},\ \exists (k,p) : k,p \in \mathbb{N}_0 \Rightarrow n = 2^k + p,\ 0 \leqslant p \leqslant 2^k - 1$. *Зададим случайные величины:* $Y \sim U[0,1],\ X = 0$ *и последовательность случайных величин* $\{X_n\}_{n=1}^{\infty} :\ X_n = \mathbb{1}(Y \in [\frac{p}{2^k}, \frac{p+1}{2^k}])$. *Для некоторого* $\varepsilon > 0$*:*

$$\mathrm{P}(|X_n - X| \geqslant \varepsilon) \leqslant \mathrm{P}(|X_n - X| > 0) = \frac{1}{2^k} \to 0$$

Так как $n \to +\infty \Rightarrow k \to +\infty \Rightarrow X_n \xrightarrow{\mathrm{p}} X$. *Теперь сделаем замечание, что для всех возможных значений, которые может принять* Y *существует бесконечно-счётно много номеров* $\{n_1, n_2, n_3, \ldots\}$, *что* $X_{n_i} = 1$. *Это верно, поскольку при неограниченном росте* n *неограниченно растёт* k, *а каждому* k *можно сопоставить соответствующее множество отрезков* $\{[\frac{p}{2^k}, \frac{p+1}{2^k}]\}(p \in [0, 2^k - 1])$, *которые в совокупности составляют отрезок* $[0,1]$, *а значит для каждого* k *случайная величина* Y *попадает хотя бы в один из отрезков заданного выше множества отрезков. И поскольку этих номеров бесконечно много, то нарушается сходимость почти наверное.*

Замечание. *Можно показать, что тут есть сходимость Lp и следовательно* $Lp \not\Rightarrow a.s.$

Пример 3.5. *($d \not\Rightarrow p$)*

Пусть множество элементарных исходов состоит из двух элементов: $\Omega = \{\omega_1, \omega_2\}$, *таких что:* $\mathrm{P}(\omega_1) = \mathrm{P}(\omega_2) = \frac{1}{2}$. *Зададим на этом вероятностном пространстве случайную величину* $X : X(\omega_1) = -1,\ X(\omega_2) = 1$*; и последовательность случайных величин* $\{X_n\}_{n=1}^{\infty} : X_n = (-1)^n \cdot X$. *Заметим что* $\forall n : X_n$ *имеет такое же распределение что и* X. *Тогда:*

$$F_X(t) = F_{X_n}(t) = \begin{cases} 0, & t < -1 \\ \frac{1}{2}, & -1 \leqslant t < 1 \\ 1, & t \geqslant 1 \end{cases}$$

Значит $F_{X_n} \to F_X$, *а значит* $X_n \xrightarrow{\mathrm{d}} X$. *Проверим сходимость по вероятности:*

$$X_n - X = \begin{cases} 0, & n = 2k \\ 2, & n = 2k+1 \end{cases}$$

Тогда $\exists \varepsilon : \mathrm{P}(|X_n - X| \geqslant \varepsilon) \not\to 0$, *то есть нет сходимости по вероятности.*

Отметим еще одно полезное при решении задач свойство сходимости по распределению.

Теорема 3.6. $X_n \xrightarrow{d} X$

F_n-ф.р. $X_n \Rightarrow F_n(X_n) \to F(x)$

F_n-ф.р. X,

$F \in C(\mathbb{R})$

$t_n \to t \in [-\infty; \infty]$

Доказательство. Пусть $t \in \mathbb{R}$ - конечно

Рассмотрим t_n, t как вырожденные случайные величины $\Rightarrow t_n \xrightarrow{p} t$

т.к. $X_n \xrightarrow{d} X$ и $t_n \xrightarrow{p} t \Rightarrow X_n - t_n \xrightarrow{d} X - t$

$F_{X-t}(y) = P(X - t \leqslant y) = P(X \leqslant y + t) = F_x(y + t) \Rightarrow F_{X-t} \in C(\mathbb{R})$

$F_{X_n}(t_n) = P(X_n \leqslant t_n) = P(X_n - t_n \leqslant 0) = F_{X_n - t_n}(0) \to F_{X-t}(0) = F_x(t)$, т.к. $X_n - E_n \xrightarrow{d} X - t_n$ $F_{X-t} \in C(\mathbb{R})$

Пусть $t = -\infty, M > 0 \Rightarrow \exists N : \forall n > N \Rightarrow t_n < -M$

F_{X_n} монотонно $\uparrow \Rightarrow F_{X_n}(t_n) \leqslant F_{X_n}(-) \to F(-M)$

Написанное верно для $\forall > 0 \Rightarrow \lim_{n\to\infty} F_{X_n}(t_n) \leqslant 0 \Rightarrow F_{X_n}(t_n) \to 0 = F(-\infty)$

Случай $t = +\infty$ рассматривается аналогично

ч.т.д.

3.3 Слабая сходимость

Представим, что мы хотим рассмотреть предел по распределению суммы независимых случайных величин. Как мы показали ранее, функция распределения суммы вычитается через свертку, что может быть не самой тривиальной задачей. В этом пункте мы рассмотрим *слабую сходимость*, тесно связанную со *сходимостью по распределению*, и связь этих двух сходимостей.

В этом пункте множество функций распределения будет обозначать как $\mathcal{F}$.

Определение 3.5. *Пусть имеются последовательность $\{F_n\}_{n=1}^{\infty}$, $F_n \in \mathcal{F}$ и $F \in \mathcal{F}$. Будем говорить, что F_n* ***слабо сходятся*** *к F и обозначать это как $F_n \Rightarrow F$, если для любой непрерывной и ограниченной функции f выполняется предельный переход*

$$\lim_{n\to\infty} \int\limits_{\mathbb{R}} f(x)dF_n(x) = \int\limits_{\mathbb{R}} f(x)dF(x).$$

Покажем, что из слабой сходимости следует сходимость по распределению.

Теорема 3.7. *Пусть $\{F_n\}_{n=1}^{\infty}$, $F_n \in \mathcal{F}$, $F \in \mathcal{F}$, $F \Rightarrow F$. Тогда $F_n \xrightarrow{\mathrm{d}} F$.*

Доказательство. Определим функцию f_{ε,x_0}, $\varepsilon > 0$, $x_0 \in \mathbb{R}$, следующим образом:

$$f_{\varepsilon,x_0}(t) = \begin{cases} 1, t \leqslant x_0, \\ \frac{x_0+\varepsilon-t}{\varepsilon}, \quad x \in (x_0, x_0+\varepsilon), \\ 0, \quad t \geqslant x_0 + \varepsilon. \end{cases} \tag{3.1}$$

Очевидно, что функция f_{ε,x_0} является непрерывной и ограниченной.

Далее рассмотрим $x_0 \in C(F)$ и заметим, что для любого $\varepsilon > 0$

$$F_n(x_0) = \int\limits_{-\infty}^{x_0} dF_n(x) = \int\limits_{-\infty}^{x_0} f_{\varepsilon,x_0}(t)dF_n(x)$$
$$\leqslant \int\limits_{-\infty}^{x_0+\varepsilon} f_{\varepsilon,x_0}(x)dF_n(x) = \int\limits_{-\infty}^{+\infty} f_{\varepsilon,x_0}(x)dF_n(x).$$

Отсюда согласно теореме о предельном переходе в неравенстве и слабой сходимости F_n к F получаем

$$\lim_{n\to\infty} F_n(x_0) \leqslant \int\limits_{-\infty}^{+\infty} f_{\varepsilon,x_0}(x)dF(x) = \int\limits_{-\infty}^{x_0+\varepsilon} f_{\varepsilon,x_0}(x)dF(x) \leqslant \int\limits_{-\infty}^{x_0+\varepsilon} dF(x) = F(x_0+\varepsilon).$$

Далее положим $g_{\varepsilon,x_0}(t) = f_{\varepsilon,x_0}(t-\varepsilon)$. Запишем функцию g_{ε,x_0} в явном виде:

$$g_{\varepsilon,x_0}(t) = \begin{cases} 1, & t \leqslant x_0 - \varepsilon, \\ \frac{x_0 - t}{\varepsilon}, & t \in (x_0 - \varepsilon, x_0), \\ 0, & t \geqslant x_0. \end{cases}$$

Тоже очевидно, что g_{ε,x_0} – непрерывная и ограниченная функция.

Рассуждая аналогичным образом, получим нижнюю оценку для $F_n(x_0)$:

$$F_n(x_0) = \int\limits_{-\infty}^{x_0} dF_n(x) \geqslant \int\limits_{-\infty}^{x_0} g_{\varepsilon,x_0}(x)dF_n(x) = \int\limits_{-\infty}^{+\infty} g_{\varepsilon,x_0}(x)dF_n(x).$$

Отсюда согласно теореме о предельном переходе в неравенстве и слабой сходимости F_n к F заключаем, что

$$\lim_{n\to\infty} F_n(x_0) \geqslant \int\limits_{-\infty}^{+\infty} g_{\varepsilon,x_0}(x)dF(x) = \int\limits_{-\infty}^{x_0} g_{\varepsilon,x_0}(x)dF(x)$$

$$\geqslant \int\limits_{-\infty}^{x_0-\varepsilon} g_{\varepsilon,x_0}(x)dF(x) = \int\limits_{-\infty}^{x_0-\varepsilon} dF(x) = F(x_0 - \varepsilon).$$

Таким образом, мы получили оценку

$$F(x_0 - \varepsilon) \leqslant \lim_{n\to\infty} F_n(x_0) \leqslant F(x_0 + \varepsilon).$$

Так как $x_0 \in C(F)$ и в силу произвольности $\varepsilon > 0$ заключаем, что $F_n(x_0) \to F(x_0)$ при $n \to \infty$, что и доказывает требуемую сходимость по распределению. *ч.т.д.*

На самом деле верна и обратная импликация, то есть из сходимости по распределению следует слабая сходимость. Чтобы показать это нам понадобится рассмотреть класс расширенных распределений и изучить некоторые его свойства. Определим класс расширенных распределений $\mathcal{G}$ следующим образом:

$$\mathcal{G} = \{G : \mathbb{R} \to \mathbb{R} \,|\, G \text{ не убывает, } G(t) = G(t_+) \forall t \in \mathbb{R}, G(+\infty) \leqslant 1, G(-\infty) \geqslant 0\}.$$

Расширенным распределениям соответствуют случайные величины, которым разрешается принимать значения $+\infty$ и $-\infty$, а именно $\mathrm{P}(X \in \mathbb{R}) = G(+\infty) - G(-\infty)$, $\mathrm{P}(X = -\infty) = G(-\infty)$, $\mathrm{P}(X = +\infty) = 1 - G(+\infty)$. Рассмотрим следующий пример.

Пример 3.6. *Пусть* $\mathrm{P}(X_n = n) = \mathrm{P}(X_n = -n) = 1/2$. *Согласно здравой логике можно условиться, что если* $n \to \infty$, *то предельная случайная величина принимает значения* $+\infty$ *и* $-\infty$ *с вероятностями 1/2. Также заметим, что*

$$F_n(x) = \begin{cases} 0, & x < -n, \\ 1/2, & x \in [-n, n), \\ 1, & x \geqslant n, \end{cases}$$

и $F_n(x) \to G(x) \equiv 1/2$ *для любого* $x \in \mathbb{R}$, *что согласуется со сказанному выше. Иными словами, последовательность* F_n *сходится к расширенной функции распределения* G *поточечно.*

Замечание 3.2. *1. Сходимости $\xrightarrow{\mathrm{d}}$ и $\Rightarrow$ можно расширить на несобственные (обобщенные) на класс $\mathcal{G}$.*

2. В классе $\mathcal{G}$ тоже из слабой слабой сходимости следует сходимость по распределению, ведь доказательство теоремы 3.7 никак не использует, что $F_n(-\infty) = F(-\infty) - 0$ и $F_n(+\infty) = F(+\infty) = 1$. Однако обратная импликация не верна. В рассмотренном выше примере пусть $f \equiv 1$. Тогда

$$\int_{-\infty}^{+\infty} f(x)dF_n(x) = 1 \not\to \int_{-\infty}^{+\infty} f(x)dG(x) = G(+\infty) - G(-\infty) = 0,$$

то есть $F_n \not\Rightarrow G$.

Класс расширенных распределений $\mathcal{G}$ примечателен тем, что из любой последовательности можно выделить слабо сходящуюся подпоследовательность, в то время как в классе обычных распределений $\mathcal{F}$ – нельзя. Первый факт известен как теорема Хелли. Приведем её формулировку.

Теорема 3.8. *Пусть $\{G_n\}_{n=1}^{\infty}, G_n \in \mathcal{G}$. Тогда существует подпоследовательность $\{G_{n_k}\}_{k=1}^{\infty}$, что $G_{n_k} \Rightarrow G \in \mathcal{G}$.*

Нам также понадобится следствие из этой теоремы.

Следствие 3.1. *Пусть всякая слабо сходящаяся подпоследовательность $\{G_{n_k}\}_{k=1}^{\infty}$ имеет один и тот же предел $G \in \mathcal{G}$. Тогда $G_n \Rightarrow G$.*

Доказательство. Предположим противное, то есть $G_n \not\Rightarrow G$. Тогда существует непрерывная и ограниченная функция f, для которой

$$\lim_{n\to\infty} \int\limits_{-\infty}^{+\infty} f(x)dG_n(x) \neq \int\limits_{-\infty}^{+\infty} f(x)dG(x).$$

Обозначим интеграл под пределом как A_n, интеграл справа – A.

Далее рассмотрим слабо сходящуюся подпоследовательность $\{G_{n_k}\}_{k=1}^{\infty}$. Тогда получаем, что $A_{n_k} \to A$. Утверждается, что все сходящиеся подпоследовательности $\{A_{n_k}\}_{k=1}^{+\infty}$ имеют предел, равный A. Действительно, пусть

$$\lim_{k\to\infty} A_{n_k} = B.$$

Тогда есть две возможности:

1. Последовательность F_{n_k} слабо сходится. По условию следствия слабый предел равняется G. Отсюда немедленно вытекает, что $B = A$.

2. Последовательность F_{n_k} не сходится слабо. Тогда по теореме Хелли существует слабо сходящаяся подпоследовательность $\{F_{n_{k_l}}\}_{l=1}^{+\infty}$, но эта подпопоследовательность исходной последовательности $\{F_n\}_{n=1}^{\infty}$, а по условию слабый предел подпоследовательности равняется G. Откуда также необходимо, чтобы $B = A$.

Таким образом, в самом деле все сходящиеся подпоследовательности $\{A_{n_k}\}_{k=1}^{+\infty}$ имеют предел, равный A. Откуда $A_n \to A$ и мы получили противоречие. *ч.т.д.*

Введем два понятия, с помощью которых мы сформулируем условия существования слабого предела $F \in \mathcal{F}$

Определение 3.6. *Пусть $\{P_n\}_{n=1}^{\infty}$ – последовательность вероятностных распределений на $\mathbb{R}$ и $\{F_n\}_{n=1}^{\infty}$ – последовательность соответствующих функций распределения. Последовательности $\{F_n\}_{n=1}^{\infty}$ и $\{P_n\}_{n=1}^{\infty}$ будем называть* ***плотными*** *, если для любого $\varepsilon > 0$ найдется $M > 0$, для которых*

$$\inf_{n\in\mathbb{N}} P_n([-M, M]) = \inf_{n\in\mathbb{N}} \big(F_n(M) - F_n(-M-0)\big) > 1 - \varepsilon.$$

На бытовом уровне условие плотности можно трактовать, как наличие равномерно маленьких хвостов распределений P_n.

Определение 3.7. *Пусть $\mathbf{L}$ – подмножество непрерывных и ограниченных функций. Будем говорить, что $\mathbf{L}$* ***определяет распределение*** *, если из равенства*

$$\int\limits_{-\infty}^{+\infty} f(x)dF(x) = \int\limits_{-\infty}^{+\infty} f(x)dG(x), \quad F \in \mathcal{F}, G \in \mathcal{G},$$

которое выполнятся для любого $f \in \mathbf{L}$ вытекает, что $F = G$.

Примеры классов функций, определяющих распределение, мы рассмотрим несколько ниже. А сейчас мы сформулируем и докажем критерий существования слабого предела $F \in \mathcal{F}$.

Теорема 3.9. *Пусть имеется последовательность $\{F_n\}_{n=1}^{\infty}$, $F_n \in \mathcal{F}$. Тогда для существования $F \in \mathcal{F}$, к которой слабо сходится последовательность F_n, необходимо и достаточно выполнения двух условий:*

1. *Последовательность $\{F_n\}_{n=1}^{\infty}$ является плотной.*

2. *Для любого $f \in \mathbf{L}$ существует предел $\int f dF_n$.*

Доказательство. Необходимость почти очевидно. Действительно, пусть $F_n \Rightarrow F \in \mathcal{F}$. Тогда для любой ограниченной и непрерывной функции f верно, что

$$\int\limits_{\mathbb{R}} f(x)dF_n(x) \to \int\limits_{\mathbb{R}} f(x)dF(x),$$

а значит это тем более выполняется для всякой $f \in \mathbf{L}$. Далее согласно теореме 3.7 $F_n \xrightarrow{\mathrm{d}} F$, то есть для любой точки $x \in C(F)$ есть сходимость $F_n(x) \to F(x)$. В частности, существует $M > 0$, для которой

$$F_n(M) - F_n(-M) \to F(M) - F(-M) > 1 - \varepsilon,$$

что обеспечивает плотность для последовательности $\{F_n\}_{n=1}^{\infty}$.

Докажем достаточность. По теореме Хелли существует подпоследовательность $\{F_{n_k}\}_{k=1}^{\infty}$, слабо сходящаяся к $G \in \mathcal{G}$. Убедимся, что $G \in \mathcal{G}$. Заметим, что последовательность $\{F_{n_k}\}_{k=1}^{\infty}$ является плотной и согласно теореме 3.7 $F_{n_k} \xrightarrow{\mathrm{d}} G$. Откуда следует, что $G(M) - G(-M) > 1 - \varepsilon$, что влечет $G(+\infty) = 1$, $G(-\infty) = 0$, то есть $G \in \mathcal{F}$.

Далее рассмотрим другую подпоследовательность $\{F_{n_j}\}_{j=1}^{\infty}$, слабо сходящуюся к $F \in \mathcal{F}$. Тогда мы получаем, что для любой непрерывной ограниченной функции f

$$\int\limits_{\mathbb{R}} f(x)dF_{n_k}(x) \to \int\limits_{\mathbb{R}} f(x)dG(x), \quad \int\limits_{\mathbb{R}} f(x)dF_{n_j}(x) \to \int\limits_{\mathbb{R}} f(x)dF(x),$$

в частности, это выполнятся для любой $f \in \mathbf{L}$. С другой стороны по условию нам сказано, что существует предел последовательности $\int\limits_{\mathbb{R}} f(x)dF_n(x)$ для любой $f \in \mathbf{L}$, значит пределы подпоследовательностей $\int\limits_{\mathbb{R}} f(x)dF_{n_k}(x)$ и $\int\limits_{\mathbb{R}} f(x)dF_{n_j}(x)$ должны совпадать, то есть

$$\int\limits_{\mathbb{R}} f(x)dG(x) = \int\limits_{\mathbb{R}} f(x)dF(x),$$

для любой $f \in \mathbf{L}$, значит $F = G$.

Таким образом, любая слабо сходящаяся подпоследовательность имеет предел F, значит по следствию из теоремы Хелли и сами F_n слабо сходятся к F. *ч.т.д.*

В только что доказанной теореме сформулированы условия существования слабого предела из $\mathcal{F}$, однако иногда возникает задача в проверке сходимости F_n к конкретной функции $F \in \mathcal{F}$. Следующая теорема позволяет решить эту задачу.

Теорема 3.10. *Пусть $F_n \in \mathcal{F}$, $F \in \mathcal{G}$, $\mathbf{L}$ определяет распределение и для любой $f \in \mathbf{L}$ выполнено предельное соотношение*

$$\int\limits_{\mathbb{R}} f(x)dF_n(x) \to \int\limits_{\mathbb{R}} f(x)dF(x).$$

Кроме того, предположим, что выполнено хотя бы одно из трех условий:

1. $\{F_n\}_{n=1}^{\infty}$ *плотная.*
2. $F \in \mathcal{F}$.
3. *Функция* $f \equiv 1 \in \mathbf{L}$.

Тогда $F_n \Rightarrow F \in \mathcal{F}$.

Доказательство. Если выполнено условие 1., то мы оказываемся в условиях предыдущей теоремы и получаем, что $F_n \Rightarrow F \in \mathcal{F}$. Если $f \equiv 1 \in \mathbf{L}$, тогда

$$1 = \int\limits_{\mathbb{R}} dF_n(x) \to \int\limits_{\mathbb{R}} dF(x) = F(+\infty) - F(-\infty),$$

а это с учетом того, что $F \in \mathcal{G}$, возможно в том и только в том случае, когда $F(+\infty) = 1$, $F(-\infty) = 0$, то есть $F \in \mathcal{F}$.

Допустим $F \in \mathcal{F}$. Тогда существует по теореме Хелли слабо сходящаяся к $G \in \mathcal{G}$ подпоследовательность $\{F_{n_k}\}_{k=1}^{\infty}$, то есть

$$\int\limits_{\mathbb{R}} f(x)dF_{n_k}(x) \to \int\limits_{\mathbb{R}} f(x)dG(x),$$

и это выполняется, в частности, для любого $f \in \mathbf{L}$. Но с другой стороны единственность предела в $\mathbb{R}$ дает нам, что

$$\int\limits_{\mathbb{R}} f(x)dG(x) = \int\limits_{\mathbb{R}} f(x)dF(x),$$

$F = G$. То есть любая слабо сходящаяся подпоследовательность слабо сходится к F, откуда $F_n \Rightarrow F$. *ч.т.д.*

Далее рассмотрим примеры классов функций, определяющих распределение.

Пример 3.7. *Пусть $L_0 = \{f_{\varepsilon,x_0} : \varepsilon > 0, x_0 \in \mathbb{R}\}$, где функция f_{ε,x_0} определена равенством (3.1). Убедимся, что L_0 определяет распределение. Действительно, пусть*

$$\int\limits_{\mathbb{R}} f_{\varepsilon,x_0}(x)dF(x) = \int\limits_{\mathbb{R}} f_{\varepsilon,x_0}(x)dG(x).$$

Тогда точно так же как и в доказательстве теоремы 3.7 получаем, что $F(x) \leqslant G(x+\varepsilon)$ и $G(x) \leqslant F(x+\varepsilon)$, откуда в силу произвольности ε получаем, что $F = G$.

Пример 3.8. *Пусть L_k – функции из $C_k(\mathbb{R})$, причем с ограниченными производными всех порядков вплоть до k. Убедимся, что L_k определяет распределение. Действительно, пусть для любой $f \in L_k$ верно, что*

$$\int\limits_{\mathbb{R}} f(x)dF(x) = \int\limits_{\mathbb{R}} f(x)dG(x).$$

Далее рассмотрим f_{ε,x_0}. Тогда существует последовательность $\{f_n\}_{n=1}^{\infty}$, $f_n \in L_k$, причем $|f_n(x)| \leqslant M$. Тогда по теореме Лебега о мажорируемой сходимости получаем, что

$$\int\limits_{\mathbb{R}} f_n(x)dF(x) \to \int\limits_{\mathbb{R}} f_{\varepsilon,x_0}(x)dF(x), \quad \int\limits_{\mathbb{R}} f_n(x)dG(x) \to \int\limits_{\mathbb{R}} f_{\varepsilon,x_0}(x)dG(x),$$

Так как предел в $\mathbb{R}$ единственен и интегралы слева равны, значит интегралы справа равны и в силу произвольности x_0 и ε получаем, что $F = G$.

В частности, тождественная единица принадлежит классу L_k, поэтому для доказательства слабой сходимости согласно утверждению 3.10 достаточно проверить условие для любой $f \in L_k$

$$\int\limits_{\mathbb{R}} f(x)dF_n(x) \to \int\limits_{\mathbb{R}} f(x)dF(x).$$

Пример 3.9. *Пусть $L_{k,0}$ – финитные функции из класса L_k, то есть существует для каждой функции $f \in L_{k,0}$ существует $M > 0$, что $f(x) = 0$ при $|x| > M$. Тогда рассуждая так же как в предыдущем примере и учитывая, что всякую $f \in L_k$ можно приблизить последовательностью из $L_{k,0}$, заключаем, что $L_{k,0}$ определяет распределение.*

Теперь мы готовы сформулировать и доказать, что в классе $\mathcal{F}$ из сходимости по распределению следует слабая сходимость и вместе с теоремой 3.7 будет показана равносильность слабой сходимости и сходимости по распределению.

Теорема 3.11. *Пусть $F_n \xrightarrow{\mathrm{d}} F$, $F_n, F \in \mathcal{F}$. Тогда $F_n \Rightarrow F$.*

Доказательство. Рассмотрим класс $L_{k,1}$ – финитные непрерывно-дифференцируемые функции. Тогда согласно утверждению 3.10 для $F_n \Rightarrow F$ достаточно показать, что для любой функции $f \in L_{k,1}$

$$\int\limits_{\mathbb{R}} f(x)dF_n(x) \to \int\limits_{\mathbb{R}} f(x)dF(x).$$

Проинтегрируем оба интеграла по частям, тогда, учитывая, что $f \in L_{k,1}$, убедимся, что

$$\int\limits_{\mathbb{R}} f'(x)F_n(x)dx \to \int\limits_{\mathbb{R}} f'(x)F(x)dx,$$

а это выполняется по теореме Лебега о мажорируемой сходимости, что завершает доказательства теоремы. *ч.т.д.*

Еще один важнейший пример – однопараметрическое семейство функции $\{e^{itx}\}_{t\in\mathbb{R}}$ и оно тесно связано с характеристическими функциями, о которых пойдет речь в следующем пункте.

3.4 Характеристические функции

3.4.1 Определение

Определение 3.8. X - *с.в.* $f(t) = Ee^{itX} = Ecos(tX) + iEsin(tX)$ – ***характеристическая функция случайной величины***, $f_X(t) = \int e^{itx}dF(x)$

3.4.2 Свойства

Свойства 3.4.1. $f(0) = 1$

Свойства 3.4.2. $|f(t)| \leq E|e^{itx}| = 1$

Свойства 3.4.3. $Y = aX + b \Rightarrow f_Y(t) = Ee^{itaX+b} = e^{itb} \cdot f_x(at)$

Свойства 3.4.4. $U = -X \Rightarrow f_U(t) = Ee^{-itX} = \overline{f_X(t)}$ – *комплексное сопряжение*

Свойства 3.4.5. X, Y – *нез* $\Rightarrow f_{X+Y}(t) = Ee^{itX} \cdot Ee^{itY} = f_X(t) \cdot f_Y(t)$

Свойства 3.4.6. f – *равномерно непрерывны на* $\mathbb{R}$:

$$|f(t+h) - f(t)| = |Ee^{itX+ihX} - Ee^{it}| \leq E|e^{ihX} - 1| \xrightarrow{h\to 0} 0$$

$\Rightarrow \sup\limits_{t\in\mathbb{R}}|f(t+h) - f(t)|$ *определен равномерно-непрерывно*
$\Rightarrow e^{ihx} \to 1, |e^{ihx} - 1| \leq 2$

Свойства 3.4.7. $\exists EX^k \Rightarrow \exists f^{(k)}(c)$

Свойства 3.4.8. $f(t) = Ee^{itX}, f'(t) = EiXe^{itX}$

Свойства 3.4.9. $f'0 = iEX$

Свойства 3.4.10. $\forall t_1 \dots t_m \in \mathbb{R}, \lambda_1 \dots \lambda_m \in \mathbb{C} : \sum\limits_{k,j} f(t_k - t_j)\lambda_k\overline{\lambda_j} \geq 0$

Теорема 3.12. *Бохнера-Хинчина*
Пусть $\phi(u)$ – непрерывная функция, $u \in \mathbb{R}^n$ и $\phi(0) = 1$. Для того, чтобы функция $\phi(u)$ была характеристической, необходимо и достаточно, чтобы она была неотрицательно определённой функцией, то есть при каждом целом $m > 0$ для любых вещественных чисел $u_1, u_2, ..., u_m$ и для любых комплексных чисел $z_1, z_2, ..., z_m$ выполняется неравенство

$$\sum_{(i,j=1)}^{m} \phi(u_i - u_j)z_i\overline{z_j} \geqslant 0$$

где $\overline{z_j}$ означает комплексно сопряжённое к z_j число.

Определение 3.9. *Решетчатая случайная величина*
X – *решетчатая* $\Leftrightarrow P(X = a + kh) = 1, \quad k \in \mathbb{Z}$, h - *шаг решетки.*

Теорема 3.13. X – *решетчатая с.в., тогда*

$$\left|f\left(\frac{2\pi h}{n}\right)\right| = 1, \quad k \in \mathbb{Z}, h\text{– шаг решетки}$$

3.4.3 Примеры

Пример 3.10. $I_c : f(t) = e^{itc}$

Пример 3.11. $Bern(p) : f(t) = p + qe^{it}$

Пример 3.12. $Bin(n, p) : f(t) = (p + qe^{it})^n$

Пример 3.13. $Pois(\lambda)$:

$$f(t) = \sum_{k=0}^{\infty} e^{itk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum \frac{(e^{it}\lambda)^k}{k!} = e^{-\lambda} e^{\lambda e^{it}} = e^{\lambda(e^{it}-1)}$$

Пример 3.14. $U[0, 1] : f(t) = \int\limits_0^1 e^{itx} dx = \frac{e^{it}-1}{it}$

Пример 3.15. $U[-1, 1] : f(t) = \frac{1}{2} \cdot \frac{e^{it}-e^{-it}}{it} = \frac{\sin t}{t}$

Пример 3.16. $N(0, 1)$:

$$f(t) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} e^{itx - \frac{x^2}{2}} dx$$

$$f'(t) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} ixe^{itx - \frac{x^2}{2}} dx = -\frac{1}{\sqrt{2\pi}} e^{itx - \frac{x^2}{2}} \Big|_{-\infty}^{+\infty} - t \underbrace{\frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} e^{itx - \frac{x^2}{2}} dx}_{f_t} = -f(t) \cdot t$$

$$\Rightarrow f(t) = e^{\frac{-t^2}{2}}$$

Пример 3.17. $N(\mu, \sigma^2) : Y = \sigma X + \mu \Rightarrow f_Y(t) = e^{it\mu} \cdot e^{-\frac{\sigma^2 t^2}{2}}$

Пример 3.18. $Exp(\lambda) : f(t) = \int\limits_0^{+\infty} \lambda e^{itx - \lambda x} dx = \frac{\lambda}{it-\lambda} e^{itx-\lambda x} \Big|_0^{+\infty} = \frac{-\lambda}{it-\lambda}$

Пример 3.19. $(n, \lambda) : X_1 \dots X_n, X_i \sim Exp(\lambda), S_n = \sum X_i, f_{S_n}(t) = (\frac{-\lambda}{it-\lambda})^n$

Многомерные

Определение 3.10. X- *с.в.* $f_X(t) : \mathbb{R}^n \to \mathbb{C}, f_X(t) = Ee^{i\langle t, X\rangle}$ - *хар фун для с.в.*

Пример 3.20. $N(\mu, \Sigma) : f_X(t) = \exp(i\langle t, \mu\rangle - \frac{1}{2} t^T \Sigma t)$

Доказательство. (С большой вероятностью это дичь, тк отсебятина). Каждая из компонент стандартного гауссовского вектора независима и имеет х.ф. $f(t_i) = \exp\left(-\frac{1}{2}t_i^2\right)$, а х.ф. по сути является матожиданием, но матожидание произведения независимых с.в. является произведением их матожиданий, тогда характеристическая функция стандартного гауссовского вектора размерности n имеет вид:

$$f(\boldsymbol{t}) = \prod_{i=1}^{n} \exp\left(-\frac{1}{2}t_i^2\right) = \exp\left(-\frac{1}{2}\boldsymbol{t}^T\boldsymbol{t}\right)$$

где $\boldsymbol{t} = (t_1, t_2, ..., t_n)^T$ - вектор переменных.

Ну а дальше просто воспользуемся свойствами х.ф. (*откуда у нас свойства многомерной х.ф... :(*):

$$N(\mu, \Sigma^2) : Y = \Sigma X + \mu \Rightarrow f_Y(t) = e^{i\langle t, \mu\rangle} \cdot e^{-\frac{t^T \Sigma t}{2}} = \exp(it^T \mu - \frac{1}{2} t^T \Sigma t)$$

ч.т.д.

Замечание 3.3. X *– многомерное норм распределение* $\Leftrightarrow \sum c_k X_k$ *распределено нормально или вырождено*

Доказательство. Для доказательства этого свойства воспользуемся характеристической функцией. Пусть $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ - многомерное нормально распределенное случайное вектор со средним $\boldsymbol{\mu}$ и ковариационной матрицей $\boldsymbol{\Sigma}$. Рассмотрим линейную комбинацию компонент вектора $\boldsymbol{X}$:

$$Y = \sum_{k=1}^{n} c_k X_k$$

Характеристическая функция случайной величины Y выражается следующим образом:

$$\varphi_Y(t) = \mathbb{E}[e^{itY}] = \mathbb{E}\left[e^{it\sum_{k=1}^{n} c_k X_k}\right] = \mathbb{E}\left[\prod_{k=1}^{n} e^{itc_k X_k}\right] = e^{it\boldsymbol{c}^T\boldsymbol{\mu} - \frac{1}{2}t^2\boldsymbol{c}^T\boldsymbol{\Sigma}\boldsymbol{c}}$$

где $\boldsymbol{c} = (c_1, c_2, ..., c_n)$ - вектор коэффициентов линейной комбинации.

Таким образом, характеристическая функция случайной величины Y имеет вид нормальной плотности вероятности с параметрами $\boldsymbol{\mu}_Y = \boldsymbol{c}^T\boldsymbol{\mu}$ и $\boldsymbol{\Sigma}_Y = \boldsymbol{c}^T\boldsymbol{\Sigma}\boldsymbol{c}$. Если $\boldsymbol{\Sigma}_Y$ вырождена, то распределение Y вырождено. В противном случае, Y имеет многомерное нормальное распределение. *ч.т.д.*

3.4.4 Формула обращения

Теорема 3.14. *Формула обращения* f*- х.ф.* $X \Rightarrow$

$$\forall y > x \in C(F) : F(y) - F(x) = \frac{1}{2\pi} \lim_{\sigma \to 0} \int\limits_{-\infty}^{+\infty} \frac{e^{-itx} - e^{-ity}}{it} \cdot f(t) \cdot e^{-\frac{\sigma^2 t^2}{2}} dt$$

Более того, если $\frac{f(t)}{t}$ *– интегрируема, можно внести предел под интеграл.*

Замечание 3.4. *По другому:* $F(y) - F(x) = \frac{1}{2\pi} \lim_{A\to\infty} \int\limits_{-A}^{+A} \frac{e^{-itx} - e^{-ity}}{it} \cdot f(t) dt$

Доказательство. $f(t) = \int\limits_{\mathbb{R}} e^{itx} p(x) dx$,

$p(z) = \frac{1}{2\pi} \int\limits_{\mathbb{R}} e^{-itz} f(t) dt$

$\Rightarrow F(y) - F(x) = \frac{1}{2\pi} \int\limits_{\mathbb{R}} \frac{e^{-itz}|_x^y}{-it} f(t) dt$

Так через пару действий мы придем к тому, что нужно для X - непр.

$U_\delta = X + Y_\delta, X, Y_\delta$ – нез $Y_\delta \sim N(\mu, \delta^2) \Rightarrow U_\delta$ – непр

$P(U_\delta \le t) = \iint_{X+Y\le t} dP(dx, dy) = \int\limits_{\mathbb{R}} dF_{Y_\delta}(y) \int\limits_{-\infty}^{t-y} dF_X(x) = \int\limits_{\mathbb{R}} \underbrace{F_X(t-y) p_\delta(y)}_{p_{U_\delta}} dy$

Тогда $F_{U_\delta}(y) - F_{U_\delta}(x) = \frac{1}{2\pi}\int\limits_{-\infty}^{+\infty} \frac{e^{it}-e^{-it}}{it} \cdot f(t) \cdot e^{-\frac{\sigma^2 t^2}{2}} dt$– тк.$X \xrightarrow{d} X \Rightarrow U_\delta \xrightarrow{d} X, Y_\delta \xrightarrow{P} 0$

То есть мы сделали предельный переход по δ (в целом произвели "сглаживание"X при помощи Y_δ). *ч.т.д.*

Теорема 3.15. *Леви*
X_n, X *– с.в.*
f_n, f *– х.ф.*

$$X_n \xrightarrow{d} X \Leftrightarrow f_n(t) \to f(t) \forall t$$

Доказательство. Напрямую следует из того, что обсуждалось в начале прошлой лекции. *ч.т.д.*

Лемма 3.1. $P(|x| > \frac{2}{U}) \le \frac{1}{U}\int\limits_{-U}^{U}(1 - f(t))dt$, *где* X *– св,* f *– хф* X*.*

Доказательство.

$$\begin{aligned}\frac{1}{U}\int\limits_{-U}^{U}(1-f(t))dt &= \frac{1}{U}\int\limits_{-U}^{U}\int\limits_{-\infty}^{+\infty}(1-e^{itx})dF(x)dt = \frac{1}{U}\int\limits_{-\infty}^{+\infty}\int\limits_{-U}^{U}(1-e^{itx})dtdF(x)\\ &= \frac{1}{U}\int\limits_{-\infty}^{+\infty}\left(2U - \frac{e^{iUx}-e^{-iUx}}{ix}\right)dF(x) = 2\int\limits_{-\infty}^{+\infty}(1-\frac{\sin Ux}{Ux})dF(x)\\ &\ge 2\cdot\int\limits_{|x|>\frac{2}{U}}\left(1-\frac{|\sin Ux|}{|Ux|}\right)dF(x) \ge 2\cdot\int\limits_{|x|>\frac{2}{U}}\frac{1}{2}dF(x) = P\left(|x|>\frac{2}{U}\right)\end{aligned}$$

ч.т.д.

Теорема 3.16. $(X_n)_{n=1}^{\infty}$ *– с.в,* $(f_n)_{n=1}^{\infty}$ *и* $f_n(t) \to f(t), \forall t \in \mathbb{R}$ *(неизвестно, является ли* f *х.ф.!)* $\Rightarrow$ *(условия равносильны)*

1. f *– х.ф* X
2. $f(0) = 1$ *и* f *непр в 0*
3. (F_n) *– плотная*

Доказательство. $(1 \Rightarrow 2)$ очевидно
$1 \Leftrightarrow 3$ вытекает из доказательств раньше
$2 \Rightarrow 3$
Воспользуемся леммой, и теоремой о среднем, тогда для произвольного U:

$$\lim_{n\to\infty}\int\limits_{|x|>\frac{2}{U}} dF_n(x) \le \lim_{n\to\infty}\frac{1}{U}\int\limits_{-U}^{I}(1-f_n(t_{Un}))dt = \frac{1}{U}\int\limits_{-U}^{I}(1-f_n(t_U))dt = 2(1-f(t_U)) < \varepsilon$$

ч.т.д.

3.4.5 Слабый ЗБЧ, ЦПТ

Утверждение 3.1. *Слабый закон больших чисел*

$$\{X_i\}_{i=1}^{\infty} - -i.i.d, EX_1 = \mu, S_n = \sum_{i=1}^{n} X_i \Rightarrow \frac{S_n}{n} \xrightarrow{P} \mu$$

Доказательство.

$$\frac{S_n}{n} \to P\mu \leftrightarrow \text{для вырожденных} \frac{S_n}{n} \xrightarrow{d} \mu \Leftrightarrow f_{\frac{S_n}{n}}(t) \to e^{it\mu} \Leftrightarrow f_{\frac{S_n}{n}-\mu}(t) \to 1 = e^{it0}$$

$$f_{x_0} = 1 + i\mu t + o(t), t \to 0 \text{ (Тейлор)}$$

$$f_{S_n} = (1 + i\mu t + o(t))^n, t \to 0$$

$$f_{\frac{S_n}{n}}(t) = f_{S_n}(\frac{t}{n}) = (1 + i\mu\frac{t}{n} + o(\frac{t}{n}))^n \xrightarrow{n\to\infty} e^{it\mu}$$

ч.т.д.

Теорема 3.17. *ЦПТ для i.i.d с.в.*
$\{X_i\}_{i=1}^{\infty} - -i.i.d, EX_i = \mu, \operatorname{Var} X_i = \sigma^2 > 0 \Rightarrow$

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} U \sim N(0,1) \Leftrightarrow \sup_{x\in\mathbb{R}} |F_n(x) - \Phi(X)| \xrightarrow{n\to\infty} 0, (F_n(x) = P(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq x))$$

Доказательство. $U_i = \frac{X_i - \mu}{\sigma}, EU_i = 0, \operatorname{Var} U_i = 1, f'(0) = 0$

$$\frac{\sum U_i}{\sqrt{n}} = \frac{S_n - n\mu}{\sqrt{n\sigma^2}}$$

$$f_{U_i}(t) = 1 - \frac{t^2}{2} + o(t^2) \Rightarrow f_{\frac{\sum U_i}{\sqrt{n}}}(t) = (1 - \frac{t^2}{2n} + o(\frac{t^2}{n}))^n \to e^{-\frac{t^2}{2}}$$

ч.т.д.

3.4.6 Ещё свойства

Утверждение 3.2. X *– целочисленная* $\Rightarrow f$*– 2π-периодическая функция*

Утверждение 3.3. *Рассмотрим промежуток* $[-\pi, \pi] : \{e^{ith}\}_{h\in\mathbb{Z}}$ *– полное и ортогональное семейство функций.*

$$\langle f, g\rangle = \int_{-\pi}^{\pi} f(x)\overline{g}(x)dx$$

$$\sum_k p_k e^{itk} = \sum_k c_k e^{itk} \Rightarrow \langle f, e^{ith}\rangle = c_j\langle e^{itj}, e^{itj}\rangle = c_j \cdot 2\pi$$

$$\Rightarrow c_j = \frac{1}{2\pi}\langle f, e^{ith}\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{ith}dt = P(X = j)$$

Оценка погрещности в теорему Пуассона

Лемма 3.2. $\mathbb{R}U < 0 \Rightarrow \begin{cases} |e^U - 1| \leq |U| \\ |e^U - U - 1| \leq \frac{|U|^2}{2} \end{cases}$

Доказательство.

$$|e^U - 1| = |\int\limits_0^U e^t dt| \overset{t=Uv}{=} |\int\limits_0^1 Ue^{Uv} dv| \leq |U| \int\limits_0^1 |e^{Uv}| dv \leq |U|$$

$$|e^U - U - 1| = |\int\limits_0^U (e^t - 1) dt| = \cdots \leq \frac{|U|^2}{2}$$

ч.т.д.

Лемма 3.3. $a_k, b_k, |a_k|, |b_k| \leq 1 \Rightarrow |\overbrace{\prod_{k=1}^n a_k}^{A_n} - \overbrace{\prod_{k=1}^n b_k}^{B_n}| \leq \sum\limits_{k=1}^n |a_k - b_k|$

Доказательство.

$$|A_{n-1}a_n - B_{n-1}b_n| = |A_{n-1}a_n - A_{n-1}b_n + A_{n-1}b_n - B_{n-1}b_n| \leq$$

$$\leq |A_{n-1}| \cdot |a_n - b_n| + |b_n| \cdot |A_{n-1} - B_{n-1}| \leq |a_n - b_n| + |A_{n-1} - B_{n-1}|$$

ч.т.д.

Теорема 3.18. *Об оценке погрешности в теореме Пуассона*
$\{X_i\}_{i=1}^\infty$ X_i– *целочисленны, независимы.* $p_i = P(X_i = 1), 1 - q_i - p_i = P(X_i = 0) \Rightarrow q_i = P(X_i \notin \{0, 1\})$

$$S_n = \sum_{i=1}^n X_i, \lambda = \sum_{i=1}^n p_i (n\text{- фикс}) \Rightarrow |P(S_n = k) - e^\lambda \frac{\lambda^k}{k!}| \leq \sum_{i=1}^n p_i^2 + 2\sum_{i=1}^n q_i$$

Доказательство.

$$f_{X_i} = (1 - p_i - q_i) + p_i e^{it} + q_i \gamma_i(t), \gamma_i(t) \text{ – некоторая х.ф.} = 1 + p_i(e^{it} - 1) + q_i(\gamma_i(t) - 1)$$

$$f_{S_n} = \prod_{i=1}^n f_{X_i}$$

$$\psi_i(t) = e^{p_i(e^{it} - 1)}$$

$$\varphi_i(t) = \prod_{i=1}^n \psi_i(t) = e^{\sum p_i(e^{it} - 1)} \sim Pois(\lambda)$$

$$|f_i(t) - \psi_i(t)| = (1 + p_i(e^{it} - 1)) - e^{p_i(e^{it} - 1)} + q_i(\gamma_i(t) - 1) \underset{\text{по лемме 2}}{\leq} p_i^2 |e^{it} - 1|^2 + 2q_i$$

$$|e^{it} - 1|^2 = (e^i t - 1)(e^{-it} - 1) = (1 - e^{it} - e^{-it} + 1) = 1 - cos(t)$$

$$\frac{1}{2\pi} \int\limits_{-\pi}^{\pi} (\frac{p_i^2}{2} |e^{it} - 1|^2 + 2q_i) dt = 2q_i + \frac{p_i^2}{4\pi} \int\limits_{-\pi}^{\pi} |e^{it} - 1|^2 dt = 2q_i + \frac{p_i^2}{2\pi} \int\limits_{-\pi}^{\pi} (1 - cos(t)) dt = 2q_i + p_i^2$$

Тогда:

$$|P_(S_n = k) - e^{-\lambda}\frac{\lambda^k}{k!}| = \frac{1}{2\pi}|\int\limits_{-\pi}^{\pi}(f_{S_n}(t)e^{-itk} - \varphi_n(t)e^{-itk})dt| \leq \frac{1}{2\pi}|\int\limits_{-\pi}^{\pi}(\prod_{i=1}^{n} f_i(t) - \prod_{i=1}^{n}\psi_i(t))dt|$$

$$\leq \frac{1}{2\pi}\int\limits_{-\pi}^{\pi}\sum_{i=1}^{n}|f_i(t) - \psi_i(t)|dt \leq \frac{1}{2\pi}\sum_{i=1}^{n}\int\limits_{-\pi}^{\pi}|f_i - \psi_i|dt \leq \sum_{i=1}^{n} p_i^2 + 2\sum_{i=1}^{n} q_i$$

ч.т.д.

Следствие 3.2. *Оценка Пуассона*

$$X = np, p_{i,n} = \frac{\lambda}{n}, q_i = 0 \Rightarrow |P(S_n = k) - e^{-np}\frac{(np)^2}{k!}| \leq \sum_{k=1}^{n}\frac{\lambda^2}{n^2} = \frac{\lambda^2}{n}$$

3.4.7 Неравенства

Теорема 3.19. *Неравенство Эссеена*
F, G – функции распределения, f, g – х.ф

$$\sup_{x\in\mathbb{R}}|G'(x)| \leq M \Rightarrow \forall T > 0 : \sup_{x\in\mathbb{R}}|F(x) - G(x)| \leq \frac{2}{\pi}\int\limits_{0}^{T}\frac{|f(t)| - g(t)|}{|t|}dt + \frac{24}{\pi T}\sup|G'(x)|$$

Доказательство. Без доказательства. *ч.т.д.*

Теорема 3.20. *Неравенство Берри-Эссеена*
$(X_i)_{i=1}^{\infty}$ – i.i.d. $S_n = \frac{\sum_{i=1}^{n} X_i - nX_1}{\sqrt{n \operatorname{Var} X_1}}$, $F_n(t) = P(S_n \leq t)$, $E|X_1 - EX_1|^3 = \beta_3$, $\sigma = \sqrt{\operatorname{Var} X_1}$
Тогда:

$$\sup_{x\in\mathbb{R}}|F_n(x) - \Phi(X)| \leq \frac{c\beta_3}{\sigma^3\sqrt{n}}$$

Доказательство. Н.У.О $(\frac{X-EX}{\sigma}) : EX_1 = 0, \operatorname{Var} X_1 = 1 \Rightarrow \beta_3 = E|X_1|^3$
И нам нужно проверить $\sup_{x\in\mathbb{R}}|F_n(x) - \Phi(X)| \leq \frac{c\beta_3}{\sigma^3\sqrt{n}}$
$f_{S_n}(t) = f_{X_1}^n(\frac{t}{\sqrt{n}})$
Возьмем $T = \frac{\sqrt{n}}{5\beta_3}$?
Посчитаем х.ф: $f_{X_1}(t) = 1 - \frac{t^2}{2} + \frac{(it)^3}{6}EX_1^3(\cos(tX\theta_1 + isin(tX\theta_2))), |\theta_1|, |\theta_2| \leq 1$
$f_{X_1}(\frac{t}{\sqrt{n}}) = 1 - \frac{t^2}{2n} + \frac{(it)^3}{n^{\frac{3}{2}}6}EX_1^3(\cos(\frac{t}{\sqrt{n}}X\theta_1 + isin(\frac{t}{\sqrt{n}}X\theta_2)))$
ценим при $t < |T|$:

$$1-|f_{X_1}(\frac{t}{\sqrt{n}}| \leq |1-f_{X_1}(\frac{t}{\sqrt{n}})| = |\frac{t^2}{2n}+\frac{(it)^3}{n^{3/2}6}EX_1^3(\cos(\frac{t}{\sqrt{n}}X\theta+isin(\frac{t}{\sqrt{n}}X\theta)))| \overset{|\cos t|,|\sin t|\leq 1}{\leq} |\frac{t^2}{2n}+\frac{t^3\beta_3}{3n^{3/2}}$$

Таким образом у нас получилось, что х.ф. отделима от нуля $|f_{X_1}(\frac{t}{\sqrt{n}})| \geq \frac{24}{25}$
$f_{S_n}(t) = f_{X_1}^n(\frac{t}{\sqrt{n}}) = \exp(n \ln f_{X_1}(\frac{t}{\sqrt{n}}))$

Замечание 3.5.

$$f(t) = 1 = iEX \cdot t - \frac{t^2}{2}EX^2 - \frac{it^3}{6}EX^3 + \dots$$

$$lnf(t) = S_1 it + \frac{S_2(it)^2}{2} + \frac{S_3(it)^3}{6}$$

S_k называются семинвариантами, причем $S_1 = EX, S_2 = \sigma^2$

$$lnf_{X_1}(\frac{t}{\sqrt{n}}) = -\frac{t^2}{2n} + \frac{(it)^3}{6n^{3/2}}(lnf)'''(\theta\frac{t}{\sqrt{n}})$$

$$ln'''f(s) = \frac{f'''f^2 - 3f''f' + 2(f')^3}{f^3} = \frac{E(iX_i)^3 e^{iX_i s} f^2 - 3E(iX_i)^2 e^{iX_i s} - 3EX_i e^{iX_i s} + 2(EiX_i e^{iX_i s})^3}{f^3(s)}$$

$$\beta_1 \le \beta_2^{1/2} \le \beta_3^{1/3}(\text{по н-ву Ляпунова})$$

$$|f_{X_1}(\frac{t}{\sqrt{n}})| \ge \frac{24}{25}, |t| < \frac{sqrtn}{5\beta_3}, |f(s)| \le 1 \Leftarrow |ln'''(\theta\frac{t}{\sqrt{n}})| \le \frac{\beta_3 + 3\beta_1\beta_2 + 2\beta_1^3}{(\frac{24}{25})^3} \le 7\beta_3$$

Замечание 3.6. *Имеет место неравенство:* $|e^z - 1| \le |z|e^{|z|}$

Оценим:

$$\begin{aligned}
\left|f_{X_1}^n(\frac{t}{\sqrt{n}})) - e^{-t^2/2}\right| &= \left|e^{nlnf_{X_1}(\frac{t}{\sqrt{n}})} - e^{-t^2/2}\right| \\
&= \left|e^{-t^2/2} - e^{-\frac{t^2}{2n} + \frac{(it)^3}{6n^{3/2}}(lnf)'''(\theta\frac{t}{\sqrt{n}})}\right| \\
&\le e^{-t^2/2}\left|\frac{(it)^3}{6\sqrt{n}}(lnf''')(\theta\frac{t}{\sqrt{n}})\right| \exp\left(\left|\frac{t^3}{6\sqrt{n}}lnf'''(\theta\frac{t}{\sqrt{n}})\right|\right) \\
&\le e^{\frac{-t^2}{2}}\frac{t^3}{6\sqrt{n}}7\beta_3 \exp\left(\frac{t^3}{6\sqrt{n}}\right)(lnf''') \\
&\le \frac{t}{6}\frac{\beta_3|t|^3}{\sqrt{n}}e^{\frac{t^2}{4}}
\end{aligned}$$

$$\left|\int_0^T \frac{|f_{S_n}(t) - \Phi(t)}{|t|}dt\right| \le \frac{7}{6}\frac{\beta_3}{\sqrt{n}}\underbrace{\int_0^T |t|^2 e^{-\frac{t^2}{4}}dt}_{\text{заведомо сходится}}$$

ч.т.д.

Следствие 3.3. *Оценка интегральной теоремы Муавра-Лапласа:*

$$\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x| \le c \cdot \frac{p^2+q^2}{\sqrt{pq}\sqrt{n}}$$

Доказательство.

$$E\left|X - EX\right|^3 = E\left|X - p\right|^3 = (1-p)^3 p + p^3(1 - p_= pq(p^2+q^2) \Rightarrow C\frac{(p^2+q^2)pq}{(pq)^{3/2}\sqrt{n}} = C \cdot \frac{p^2+q^2}{\sqrt{pqn}}$$

ч.т.д.

3.5 Упражнения

1. Показать, что случайный вектор $X = (X_1, \ldots, X_n)^T$ является гауссовским тогда и только тогда, когда линейная комбинация его компонент $c_1X_1 + \ldots + c_nX_n$ имеет нормальное или вырожденное распределение.

2. Складывается 10^4 чисел, округленных с точностью до 10^{-m}. Предполагая, что ошибки округления независимы и равномерно распределены в интервале $[-0,5\cdot 10^{-m}, 0,5\cdot 10^{-m}]$, найти пределы, в которых с вероятностью, не меньшей 0,99, будет лежать суммарная ошибка.

3. Случайные величины $\xi_1, \xi_2, \ldots$ независимы и одинаково распределены:

$$\mathrm{P}\{\xi_i = 1,25\} = \mathrm{P}\{\xi_i = 0,75\} = \frac{1}{2}, i \in \mathbb{N}\,.$$

и $\eta_n = \xi_1 \cdot \ldots \cdot \xi_n$.

   (a) Найти $\mathrm{E}\,\eta_{1000}, \mathrm{Var}\,\eta_{1000}, \mathrm{E}\ln\eta_{1000}, \mathrm{Var}\ln\eta_{1000}$;

   (b) Пользуясь асимптотической нормальностью $\ln\eta_n$ при $n \to \infty$, найти приближенные значения

      - $\mathrm{P}\{\eta_{1000} \leqslant 10^{-20}\}$,
      - $\mathrm{P}\{\eta_{1000} < 1,25^501 \cdot 0,75^499\}$,
      - $\mathrm{P}\{\eta_{1000} \leqslant 1,25^501 \cdot 0,75^499\}, \mathrm{P}\{\eta_{1000} \leqslant 10^{-7}\}$.

   (c) Пользуясь формулой Стерлинга, найти

      - $\mathrm{P}\{\eta_{1000} < 1,25^501 \cdot 0,75^499\}$,
      - $\mathrm{P}\{\eta_{1000} \leqslant 1,25^501 \cdot 0,75^499\}$.

4. Случайные величины $\xi_1, \xi_2, \ldots$ независимы и имеют стандартное нормальное распределение. Распределение случайной величины

$$X_n^2 = \xi_1^2 + ... + \xi_n^2.$$

называется *распределением* χ^2 (хи-квадрат) с *n степенями свободы*

   (a) Доказать, что $\lim_{n\to\infty} \mathrm{P}\left\{\left|\frac{X_n^2}{n} - 1\right| > \varepsilon\right\} = 0$ при любом $\varepsilon > 0$;

   (b) Найти $\lim_{n\to\infty} \mathrm{P}\left\{\frac{X_n^2 - M_{X_n^2}}{\sqrt{D_{X_n^2}}} \leqslant x\right\}, x \in \mathbb{R}$.

5. Случайные величины $\xi_1, \xi_2, \ldots$ независимы и имеют стандартное нормальное распределение. Распределение случайной величины

$$\tau_n = \frac{\xi_0}{\sqrt{\xi_1^2 + ... + \xi_n^2/n}}$$

называется *распределением Стьюдента с n степенями свободы.* Найти

$$\lim_{n\to\infty} \mathrm{P}\{\tau_n \leqslant x\}, \quad x \in \mathbb{R}\,.$$

6. Случайная величина ξ_λ распределена по закону Пуассона с параметром λ. Найти

$$\lim_{\lambda\to\infty} \mathrm{P}\left\{\frac{\xi_{\lambda-\lambda}}{\sqrt{\lambda}} \leqslant x\right\}$$

7. Имеется последовательность $\{X_n\}$ независимых случайных величин, равномерно распределенных на $[0, a]$. Положим $\xi_n = n\min(X_1, \ldots, X_n)/a$ и $\eta_n = n(1 - \max(X_1, \ldots, X_n)/a)$. Найти предельные распределения для ξ_n, η_n и показать, что они независимы.

8. *(метод Монте-Карло статистических испытаний)* Вычисление интеграла $I = \int_0^1 f(x)dx = I$можно описать следующим образом. Пусть случайная величина ξ имеет равномерное распределение на отрезке [0, 1]. Тогда

$$\mathrm{E}\, f(\xi) = \int\limits_0^1 f(x)dx = I.$$

Пусть $\xi_1, ..., \xi_n$ взаимно независимы и равномерно распределены на [0, 1]. Рассмотрим $\bar{f}_n = \frac{1}{n}\left[f(\xi_1) + ... + f(\xi_n)\right]$ и предположим, что $\sigma^2 = D\bar{f}_n \leqslant C$ Показать, что $E\bar{f}_n = I$ и $\bar{f} \xrightarrow{p} I, n \to \infty$. Оценить $\mathrm{P}(|\bar{f}_n - I| < \varepsilon)$ для произвольного $\varepsilon > 0$ с помощью центральной предельной теоремы.

9. *(теорема Вейерштрасса)*. Пусть $f(x)$ - непрерывная функция на отрезке (0,1]. Пусть последовательность случайных величин $\xi_1, \ldots, \xi_n$ соответствует схеме Бернулли с вероятностью успеха $\mathrm{P}(\xi_i = 1) = x, 0 < x < 1$, и $S_n = \xi + ... + \xi_n$ Введем многочлены

$$B_n(x) = \mathrm{E}\, f(\frac{S_n}{n}) = \sum_{m=0}^{n} f(\frac{m}{n}) C_n^m x^m (1-x)^{n-m}.$$

Доказать, что при $n \to \infty$

$$\sup_{x \in (0,1)} |f(x) - B_n(x)| \to 0.$$

(Многочлены $B_n(x)$ называются *многочленами Бернштейна.*)

Глава 4

Введение в математическую статистику. Описательная статистика

4.1 Выборка. Эмпирическая функция распределения и свойства.

Чтобы можно было сделать какие-то выводы по выборке, она должна обладать некоторыми свойствами. Во-первых, она должна приходить из какого-то вероятностного пространства (если всё детерменировано, то наши методы теряют смысл). Во-вторых, она должна быть достаточно репрезентативной (большой). Это понятие сложно формализовать, и в разных областях это слово понимается по-своему.

4.1.1 Эмпирическая функция распределения

Пусть $X_1 \dots X_n$ — выборка объёма n, реализация n i.i.d. случайных величин с теоретической функцией распределения $F(x)$: $X_1 \dots X_n \sim F(x)$. Введем пару функций.

$$\mu_n(x) = \sum \mathbb{1}(X_k \leqslant x); \quad F_n(x) = \frac{\mu_n(x)}{n}$$

$\mu_n(x)$ считает количество элементов выборки, которые не превосходят x. Функция F_n называется эмпирической функцией распределения (коротко ЭФР).

Теперь поговорим про свойства эмпирической функции распределения. Заметим, что $\mathbb{1}(X_k \leqslant x) \sim Bern(F(x))$, и тогда $\mu_n(x) = \sum \mathbb{1}(X_k \leqslant x) \sim Bin(n; F(x))$. Это значит, что

$$P\left(F_n(x) = \frac{k}{n}\right) = P(\mu_n(x) = k) = C_n^k F(x)^k (1 - F(x))^{n-k}.$$

Тогда

$$\mathrm{E}\,\mu_n(x) = nF(x),\ \mathrm{E}\,F_n(x) = F(x); \quad \mathrm{Var}\,\mu_n(x) = nF(x)(1 - F(x))$$

Выражение $\mathrm{E}\,F_n(x) = F(x)$ означает *несмещённость* ЭФР.

По ЗБЧ имеет место $\mathrm{P}(|F_n(x) - F(x)| > \varepsilon) \to 0$. В мат. статистике это говорит о *состоятельности* ЭФР. Ещё одно свойство проистекает из ЦПТ:

$$\frac{\mu_n(x) - nF(x)}{\sqrt{nF(x)(1 - F(x))}} \xrightarrow{\mathrm{d}} U \sim N(0, 1)$$

Это свойство означает *асимптотическую нормальность*

Теперь с помощью ЦПТ мы можем получить доверительный интервал для ЭФР. ЦПТ утверждает, что

$$\mathrm{P}\left(t_1 \leqslant \frac{\mu_n(x) - nF(x)}{\sqrt{nF(x)(1-F(x))}} \leqslant t_2\right) \approx \Phi(t_2) - \Phi(t_1).$$

Выражение слева можно переписать как $\mathrm{P}\left(t_1 \leqslant \sqrt{n}\frac{F_n(x) - F(x)}{\sqrt{F(x)(1-F(x))}} \leqslant t_2\right)$. Теперь рассмотрим $\mathrm{P}\left(|\sqrt{n}(F_n(x) - F(x))| < t\right)$:

$$\mathrm{P}\left(|\sqrt{n}(F_n(x) - F(x))| < t\right) = \mathrm{P}\left(-t < \sqrt{n}(F_n(x) - F(x)) < t\right).$$

Поделив это неравенство на $\sqrt{nF(x)(1-F(x))} = \sqrt{\operatorname{Var}\mu_n(x)}$ и применив ЦПТ, получим:

$$\mathrm{P}\left(\frac{-t}{\sqrt{\operatorname{Var}\mu_n(x)}} < F_n(x) - F(x) < \frac{t}{\sqrt{\operatorname{Var}\mu_n(x)}}\right) = \\ \Phi\left(\frac{t}{\sqrt{\operatorname{Var}\mu_n(x)}}\right) - \Phi\left(\frac{-t}{\sqrt{\operatorname{Var}\mu_n(x)}}\right)$$

По свойствам Φ выполнено равенство $\Phi\left(\frac{t}{\sqrt{\operatorname{Var}\mu_n(x)}}\right) - \Phi\left(\frac{-t}{\sqrt{\operatorname{Var}\mu_n(x)}}\right) = 2\Phi\left(\frac{t}{\sqrt{\operatorname{Var}\mu_n(x)}}\right) - 1$. Выражение под корнем можно оценить: при том, что $0 < F(x) < 1$ имеет место неравенство $\frac{1}{\sqrt{F(x)(1-F(x))}} \leqslant 2$, и тогда

$$2\Phi\left(\frac{t}{\sqrt{\operatorname{Var}\mu_n(x)}}\right) - 1 \geqslant 2\Phi(2t) - 1$$

Обозначим $\gamma = 2\Phi(2t) - 1$. Тогда можем сказать, что $2t = q_{\frac{\gamma+1}{2}}$, и $t = \frac{q_{\frac{\gamma+1}{2}}}{2}$. Обозначим просто $q = q_{\frac{\gamma+1}{2}}$. Итак, получается:

$$\mathrm{P}\left(\frac{-q}{2\sqrt{n}} < F_n(x) - F(x) < \frac{q}{2\sqrt{n}}\right) \geqslant \gamma,$$

, где $q = \Phi^{-1}(\frac{\gamma+1}{2})$. Следовательно

$$\mathrm{P}\left(F_n(x) - \frac{q}{2\sqrt{n}} < F(x) < F_n(x) + \frac{q}{2\sqrt{n}}\right) \geqslant \gamma.$$

Такое неравенство называется доверительным интервалом для $F(x)$.

Продолжим говорить о свойствах ЭФР. Рассмотрим набор чисел из какой-то выборки $x_0 = -\infty < x_1 < \ldots, x_{N-1} < x_N = +\infty$, и связанную с выборкой функцию распределения в этих точках $0 < F(x_1) < \ldots < F(x_{N-1}) < 1$. Рассмотрим следующие величины:

$$p_i = \Delta_i F = F(x_i) - F(x_{i-1}) = P(x_{i-1} < X < x_i),$$

$$\nu_i = \Delta_i \mu_n = \sum_{i=1}^{n} \mathbb{1}[x_{i-1} < X < x_i].$$

ν_i — это частоты, т.е число элементов выборки в интервале между x_{i-1} и x_i. Теперь ЭФР можно записать по другому:

$$F_n(x_i) = \frac{\nu_1 + \nu_2 + \ldots \nu_i}{n}$$

$$\Delta_i F_n(x_i) = F_n(x_i) - F_n(x_{i-1}) = \frac{\nu_i}{n}$$

$$\sqrt{n}(\Delta_i F_n(x) - \Delta_i F(x)) = \sqrt{n}\frac{\nu_i - np_i}{n} = \hat{\nu}_i{}^{(n)}$$

Векторная случайная величина $\hat{\nu}$ имеет полиномиальное распределение:

$$(\nu_1, \ldots, \nu_N)^T \sim Poly(n, (p_1, \ldots, p_N)^T)$$

Тогда по многомерному закону ЦПТ верно

$$\hat{\nu}^{(n)} \xrightarrow{\mathrm{d}} N(0, \Sigma), \quad \Sigma_{i,j} = p_i(\delta_{i,j} - p_j).$$

Здесь Σ — ковариационная матрица

Теорема 4.1. ***(Теорема Гливенко-Кантелли)***
Пусть $\boldsymbol{X} = (X_1, \ldots, X_n)$ - выборка объема n из неизвестного распределения F, $F_n(x)$ - ЭФР, $D_n = \sup\limits_{x\in\mathbb{R}} |F_n(x) - F(x)|$. Тогда D_n сходится к 0 почти наверное:

$$\mathrm{P}(D_n(x) \to 0) = 1$$

Теорема 4.2. ***(Теорема Колмогорова)***
Пусть $\boldsymbol{X} = (X_1, \ldots, X_n)$ – выборка объема n из неизвестного распределения F, где F(x) – непрерывна, $F_n(x)$ – ЭФР, $D_n = \sup\limits_{x\in\mathbb{R}} |F_n(x) - F(x)|$. Тогда:

$$\sqrt{n}D_n \xrightarrow{\mathrm{d}} K(t) \Leftrightarrow \mathrm{P}(\sqrt{n}D_n \leqslant t) \to K(t)$$

где $K(t)$ – распределение Колмогорова с непрерывной функцией распределения:

$$K(t) = \begin{cases} \sum\limits_{j=-\infty}^{\infty} (-1)^j \cdot e^{-2j^2t^2}, & t \geqslant 0 \\ 0, & t < 0 \end{cases}$$

Теорема 4.3. ***(Теорема Смирнова)***
Пусть есть две независимые выборки объемов m и n неизвестного распределения F, где F(x) – непрерывна, F_m, F_n – соответствующие ЭФР. $D_{m,n} = \sup\limits_{x\in\mathbb{R}} |F_m(x) - F_n(x)|$. Тогда:

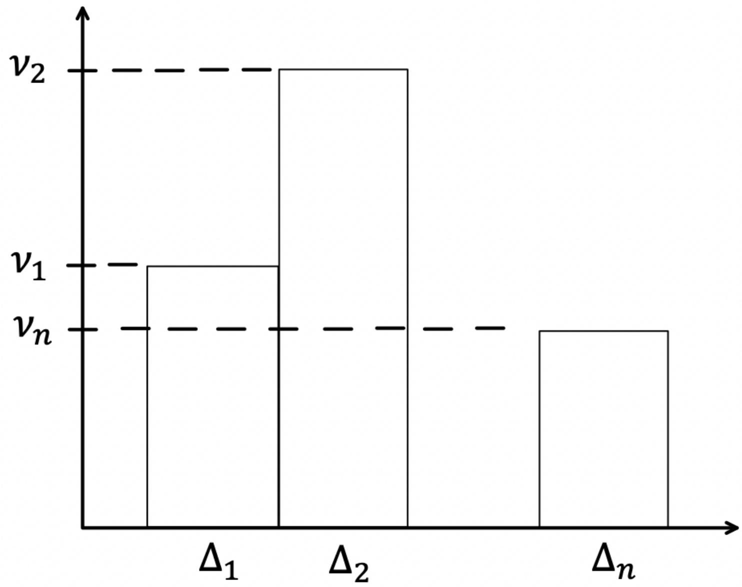
$$\sqrt{\frac{mn}{m+n}}D_{m,n} \xrightarrow{\mathrm{d}} K(t) \Leftrightarrow \mathrm{P}\left(\sqrt{\frac{mn}{m+n}}D_{m,n} \leqslant t\right) \to K(t)$$

4.1.2 Способы визуализации выборки

В заключении, хочется поговорить про способы визуализации выборки. Эти методы полезны в статистике, поскольку предлагают наглядный способ заключения некоторых выводов о выборке.

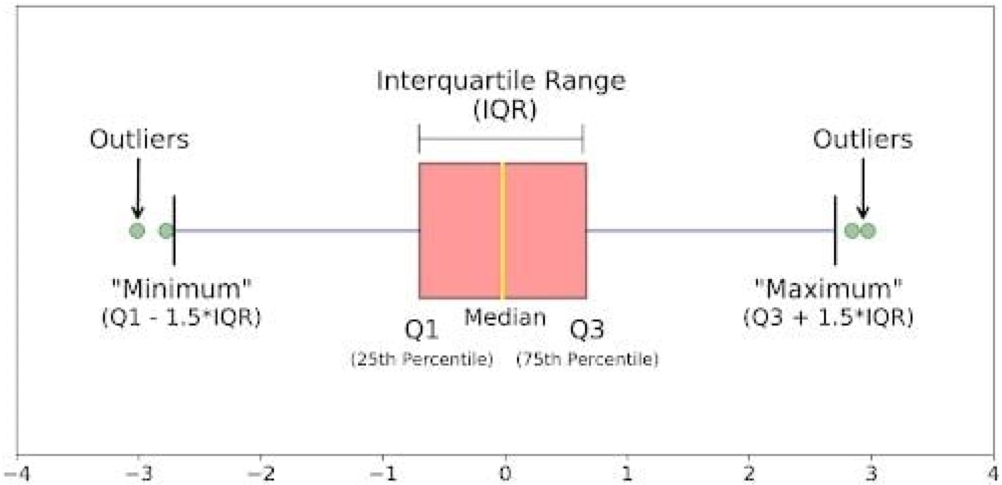
Гистограмма

Одной из характеристик эмпирического распределения является гистограмма. Для заданной выборки $X_1, \ldots, X_n$ разобьем числовую прямую на конечное число промежутков $\Delta_1, \ldots, \Delta_m$. Обозначим через $\nu_k(x)$ число случайный величин среди выборки, попавших в интервал Δ_k. Тогда $\nu_1(x) + \cdots + \nu_m(x) = n$. На каждом из интервалов Δ_k строится прямоугольник, площадь которого пропорциональна ν_k. Общая площадь всех прямоугольников должна равняться единице. Пусть l_k – длина интервала Δ_k. Высота f_k прямоугольника над Δ_k будет равняться: $f_k = \frac{\nu_k}{n\Delta_k}$.



Box-plot

Box-plot, или же ящик с усами компактно отображает одномерную выборку. Для заданной выборки $X_1, \ldots, X_n$ требуется найти его нижний квартиль, медиану и верхний квартиль. Над прямой, отображающей значение случайной величины строится прямоугольник в пределах квартилей и явно отображается медиана. В стороны от прямоугольника строятся промежутки, равные $3/2 \cdot IQR$. Значения из выборки, которые не попадают в прямоугольник или в промежутки являются выбросами и помечаются точкой отдельно.



4.2 Выборочные моменты и их свойства

4.2.1 Базовые выборочные моменты одной выборки

Пусть имеется выборка $X_1, \ldots, X_n - i.i.d. \sim F(X)$. Тогда существуют теоретические моменты. Если теоретические моменты существуют, то в силу закона больших чисел выборочные моменты сходятся по вероятности к своим теоретическим прообразам. Среди выборочных моментов особое место занимают моменты первого и второго порядков.

$$\alpha_k = \mathrm{E}\, X^k - \text{k-ый теоретический момент распределения}$$
$$\beta_k = \mathrm{E}(X - \mathrm{E}\, X)^k - \text{k-ый центральный теоретический момент распределения}$$

Определение 4.1. *К-ый выборочный момент*

$$\hat{\alpha_k} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Определение 4.2. *Выборочное среднее*

$$\hat{\alpha_1} = \overline{X}$$

Определение 4.3. *К-ый центральный выборочный момент*

$$\hat{\beta}_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\alpha_1})^k$$

Определение 4.4. *Выборочная дисперсия*

$$S_*^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \overline{X^2} - (\overline{X})^2 = \hat{\beta}_2$$

4.2.2 Свойства выборочных моментов

1. Несмещённость выборочного момента

$$\mathrm{E}\,\overline{X^k} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\, X_i^k = \frac{1}{n}\cdot n \cdot \mathrm{E}\, X_1^k = \mathrm{E}\, X^k$$

2. Дисперсия выборочного момента

$$\mathrm{Var}\,\overline{X^k} = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}\, X_i^k = \frac{1}{n}\mathrm{Var}\, X_1^k = \frac{\mathrm{Var}\, X^k}{n}$$

3.
$$\mathrm{E}\, S_*^2 = \mathrm{E}\,\overline{X^2} - (\mathrm{E}\,\overline{X})^2 = \alpha_2 - \alpha_1^2 - \frac{\beta_2}{n} = \beta_2 - \frac{\beta_2}{n} = \frac{n-1}{n}\beta_2$$

Т.к. $\mathrm{E}\, S_*^2 \neq \beta_2$, то такую дисперсию называют **смещённой**

$$S^2 = \frac{n}{n-1}S_*^2$$

S^2 - **не смещённая** дисперсия

4. Асимптотическая несмещённость момента

$$\sqrt{n}\frac{\hat{\alpha_k}-\alpha_k}{\sqrt{\alpha_{2k}-(\alpha_k)^2}} \to \sqrt{n}\frac{\hat{\alpha_k}-\alpha_k}{\sqrt{\hat{\alpha_{2k}}-(\hat{\alpha_k})^2}} =$$
$$= \sqrt{n}\sqrt{\frac{\alpha_{2k}-(\alpha_k)^2}{\hat{\alpha_{2k}}-(\hat{\alpha_k})^2}} \cdot \frac{\hat{\alpha_k}-\alpha_k}{\sqrt{\alpha_{2k}-(\alpha_k)^2}} \xrightarrow{\text{p}}$$
$$\xrightarrow{\text{p}} 1 \cdot \sqrt{n}\frac{\hat{\alpha_k}-\alpha_k}{\sqrt{\alpha_{2k}-(\alpha_k)^2}} \xrightarrow{\text{d}} U \sim N(0,1)$$

Здесь, мы сначала перешли к выборочным моментам, поскольку теоретические нам не известны и воспользовались тем, что выборочные моменты сходятся по вероятности к теоретическим.

Теорема 4.4. *Пусть $\hat{\gamma} = P(\hat{\alpha_1}, \ldots, \hat{\alpha_k})$ - непрерывная функция от выборочных моментов. Утверждается, что $\hat{\gamma} \xrightarrow{\text{p}} \gamma$, где $\gamma = P(\alpha_1, \ldots, \alpha_k)$*

Доказательство. Заметим, что $\hat{\alpha_k} \xrightarrow{\text{p}} \alpha_k$. Действительно:

$$\hat{\alpha_k} = \overline{X^k}; \alpha_k = \mathrm{E}\, X^k \xrightarrow{\text{ЗБЧ}} \mathrm{P}(|\hat{\alpha_k} - \alpha_k| \geqslant \varepsilon) \to 0$$

Теперь, ссылаемся на замкнутость сходимости по вероятности относительно математических операций. *ч.т.д.*

В частности, эта теорема показывает, что все моменты - обычные и центральные - являются *состоятельными*.

4.2.3 Прочие выборочные характеристики

Определение 4.5. *Величина*
$$\gamma_1 = \frac{\mathrm{E}(X - \mathrm{E}\, X)^3}{(\operatorname{Var} X)^{3/2}}$$
называется коэффициентом асимметрии и характеризует скошенность распределения по отношению к математическому ожиданию.

Если $\gamma_1 > 0$ *вставить картинку (картинки всё ещё нет)* («длинная часть» кривой распределения расположена справа от математического ожидания)
Если $\gamma_1 < 0$ *вставить картинку (картинки всё ещё нет)* («длинная часть» кривой расположена слева от математического ожидания.)

Определение 4.6. *Величина*
$$\hat{\gamma_1} = \frac{\overline{(X - \overline{X})^3}}{S_*^3}$$
называется выборочным коэффициентом асимметрии

Определение 4.7. *Величина*
$$\gamma_2 = \frac{\mathrm{E}(X - \mathrm{E}\, X)^4}{\operatorname{Var}^2 X} - 3$$
называется коэффициентом эксцесса и характеризует относительную остроконечность или сглаженность распределения по сравнению с нормальным распределением.

«Минус три» в конце формулы введено для того, чтобы коэффициент эксцесса стандартного нормального распределения был равен нулю.
Если $\gamma_2 > 0$ *вставить картинку (ну вы поняли)* (пик распределения около математического ожидания острый
Если $\gamma_2 < 0$ *вставить картинку (ну вы поняли)* (пик распределения около математического ожидания гладкий)

Определение 4.8. *Величина*

$$\hat{\gamma_2} = \frac{\overline{(X - \overline{X})^4}}{S_*^4} - 3$$

называется выборочным коэффициентом эксцесса.

4.2.4 Выборочный моменты для двух выборок

Определение. *Величина*

$$S_*^2 = \overline{(\mathrm{X} - \overline{\mathrm{X}})(\mathrm{Y} - \overline{\mathrm{Y}})} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})$$

или

$$S_*^2 = \overline{XY} - \overline{X} \cdot \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \frac{1}{n^2}\sum_{i=1}^{n} X_i \cdot \sum_{i=1}^{n} Y_i$$

называется выборочным коэффициентом ковариации и численно выражает меру совместной изменчивости двух случайных величин.

Определение. *Величина*

$$\rho_{XY} = \frac{S_*^2(XY)}{S_*(X) \cdot S_*(Y)}$$

называется выборочным коэффициентом корреляции и характеризует степень линейной зависимости между наборами чисел $X_1, \cdots, X_n$ и $Y_1, \cdots, Y_n$.

Обе эти оценки состоятельные и асимптотически нормальные.

4.3 Порядковые статистики

4.3.1 Вариационный ряд. Выборочная квантиль

Определение 4.9. *Вариационный ряд – отсортированная по возрастанию выборка.*

$$X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n)}$$
$$X_{(1)} = \min(X_1, \ldots, X_n)$$
$$X_{(n)} = \max(X_1, \ldots, X_n)$$

Определение 4.10. *Выборочный квантиль порядка $p \in (0, 1)$ вариационного ряда - элемент с номером k: $q_p = X_{(k)}$:*

$$\exists! k \in [1, \ldots, n]: \quad \frac{k-1}{n} < p \leq \frac{k}{n}.$$

Если $p = 0$, то его выборочный квантиль $= q_0 = \min(X_1, \ldots, X_n)$
Если $p = 1$, то его выборочный квантиль $= q_1 = \max(X_1, \ldots, X_n)$

Определение 4.11. *Выборочной медианой выборки называется его выборочный квантиль порядка $\frac{1}{2}$. Он задаётся либо определением выше, либо следующим образом:*

$$med(X_1, ..., X_n) = \begin{cases} \frac{X_{(m)} + X_{(m+1)}}{2}, & n = 2m, \\ X_{(m+1)}, & n = 2m + 1 \end{cases}$$

Определение 4.12. *Выборочный квантиль порядка $p = \frac{1}{4}$ называется нижним квартилем*
Выборочный квантиль порядка $p = \frac{3}{4}$ называется верхним квартилем
IQR - межквартильный размах - разность между нижним и верхним квартилем: $IQR = q_{3/4} - q_{1/4}$. Иногда, в контексте, под межквартильным размахом могут понимать промежуток между квартилями.

4.3.2 Распределение порядковых статистик

Найдём функцию распределения k-й порядковой статистики. Пусть эта функция $F_{(k)}(t)$. Тогда

$$\mathrm{P}(X_{(k)} \leqslant t) = \mathrm{P}(\mu_n(t) \leqslant k) = \sum_{r=k}^{n} C_n^r F^r(t)(1 - F(t))^{n-r}.$$

Здесь мы пользовались тем, что $\mu_n(x) \sim Bin(n, F(x))$. Вообще говоря, этот результат можно выразить через специальные функции. Введём, так называемую, регуляризованную неполную бета-функцию:

$$B(a, b, x) = \frac{(a + b - 1)!}{(a - 1)!(b - 1)!} \int_0^x t^{a-1}(1 - t)^{b-1} dt$$

Можно показать что эта функция с некоторыми параметрами укладывается в ту сумму что представлена выше. Оказывается, что:

$$F_{(k)}(t) = B(k, n - k + 1, F(t))$$

Эта формула оказывается удобной, чтобы посчитать распределение минимума. Функция распределения для максимума $F_{(n)}(t) = F^n(t)$. Функция распределения для минимума $F_{(1)}(t) = 1 - (1 - F(t))^n$.

4.3.3 Асимптотические свойства

Определение 4.13. $X_{(k)}$ *– среднее, если* $\frac{k(n)}{n} \to p \in (0, 1)$
X_k*,* $X_{(k+1-S)}$ *– крайние, если* γ, S *– ограничены*

Теорема 4.5. *(об асимптотике средних членов вариационного ряда)*
Пусть имеется выборка $X = (X_1, \dots, X_n)$ и записан соответствующий ей вариационный ряд. Выборка из непрерывного распределения F(x). Пусть q_p, $p \in (0, 1)$ - его p-квантиль, причем F дифференцируема в точке q_p и $F'(q_p) = f(q_p) > 0$. Тогда:

$$\sqrt{n} \frac{X_{\lceil np \rceil} - q_p}{\sqrt{p(1 - p)}} f(q_p) \xrightarrow{\mathrm{d}} U \sim N(0, 1)$$

Доказательство. (идея доказательства)г

1. Введём функцию $G = \sqrt{n} f(q_p) \frac{t - q_p}{\sqrt{p(1-p)}}$. Заданное выражение мы обозначим за $U_n = G(X_{\lceil np \rceil})$. Дальше мы должны найти плотность $f_{U_n}(t)$

$$f_{U_n}(t) = f_{X_{\lceil np \rceil}}(G^{-1}(t)) \cdot (G^{-1}(t))'$$

2. Нужно показать, что $f_{U_n} \xrightarrow{\mathrm{d}} \dfrac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}$

ч.т.д.

Теорема 4.6. *(об асимптотике крайних членов вариационного ряда)*
Пусть имеется выборка $X = (X_1, \ldots, X_n)$ и записан соответствующий ей вариационный ряд. Выборка из непрерывного распределения F(x). Пусть $r, k \in \mathbb{N}$, тогда имеет место следующий результат:

1. $F(X_{(r)}) \xrightarrow{\mathrm{d}} U_1 \sim \Gamma(r, 1)$ *– гамма-распределение порядка r*

2. $n \cdot (1 - F(X_{n+1-k})) \xrightarrow{\mathrm{d}} U_2 \sim \Gamma(k, 1)$ *– гамма-распределение порядка k*

3. U_1 *и* U_2 *– независимы*

Доказательство. (идея доказательства)

1. Нужно записать совместную плотность U_1, U_2

2. Осуществляем предельный переход $n \to \infty$

ч.т.д.

Глава 5

Оценивание параметров

5.1 Постановка задачи точеченого оценивания параметров

Задача точечного оценивания параметров заключается в том, чтобы оценить неизвестный параметр или параметры распределения на основе имеющихся наблюдений. То есть, в нахождении числовых значений параметров, описывающих распределение случайной величины, на основе выборки из этого распределения. В качестве точечной оценки параметра выбирается одно конкретное значение, которое является наиболее вероятным для данной выборки.

Пусть имеется выборка независимых и одинаково распределенных случайных величин $X_1, X_2, \ldots, X_n$ с неизвестным распределением $F(x; \theta)$, где θ - неизвестный параметр распределения. Задача состоит в том, чтобы на основе выборки $X_1, X_2, \ldots, X_n$ получить оценку параметра θ.

Оценка параметра θ может быть построена по различным критериям, таким как наиболее вероятное значение (Maximum Likelihood Estimator, MLE), метод моментов (Method of Moments, MM) или минимизация квадратичного отклонения (Least Squares Estimator, LSE) и т.д. В зависимости от выбранного метода и типа распределения, полученные оценки могут отличаться между собой.

Цель точечного оценивания параметров состоит в том, чтобы получить наиболее достоверную оценку неизвестного параметра θ на основе имеющихся данных, чтобы можно было использовать эту оценку для принятия решений и делать выводы о характеристиках изучаемой выборки.

Важным моментом при точечном оценивании параметров является оценка точности полученных значений. Для этого используются стандартные ошибки оценок, доверительные интервалы и другие характеристики, которые позволяют оценить разброс полученных значений и определить, насколько они точны.

5.2 Свойства точечных оценок

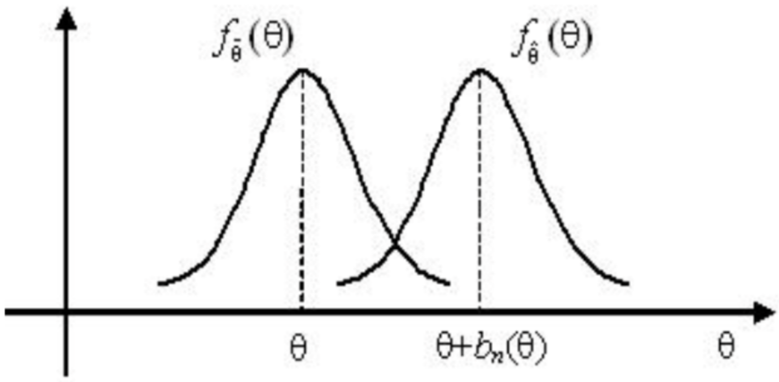
5.2.1 Несмещенность

Оценка называется несмещённой, если матожидание этой оценки равно оцениваемому параметру. Другими словами, среднее значение оценки должно быть равно параметру, который мы пытаемся оценить. То есть, если $\widehat{\theta}$ - несмещённая оценка параметра θ, то

$$E(\widehat{\theta}) = \theta.$$

Разность $b_n(\theta) = E(\widehat{\theta}) - \theta$ называется *смещением* точечной оценки $\widehat{\theta}_n$

Несмещенность оценки $\widehat{\theta}_n$ означает, что реализации этой оценки, рассчитанные для различных реализаций случайной выборки $X_1, \ldots, X_n$ объема n, будут группироваться в среднем около оцениваемого параметра θ. Реализация несмещенной точечной оценки $\widehat{\theta}$ группируются около оцениваемого параметра θ, а реализация смещенной оценки $\widehat{\theta}$ – около величины $\theta + b_n(\theta)$



5.2.2 Состоятельность

Оценка является состоятельной, если с увеличением размера выборки оценка сходится к истинному значению параметра. Другими словами, чем больше данных мы имеем, тем более точной должна становиться наша оценка.

Оценка называется состоятельной по вероятности, если вероятность того, что она отклонится от оцениваемого параметра больше, чем на заданную ширину доверительного интервала, стремится к нулю с ростом размера выборки.

То есть, если $\widehat{\theta}_n$ - оценка параметра θ по выборке размера n, то для $\forall \varepsilon > 0$ при $n \to \infty$:

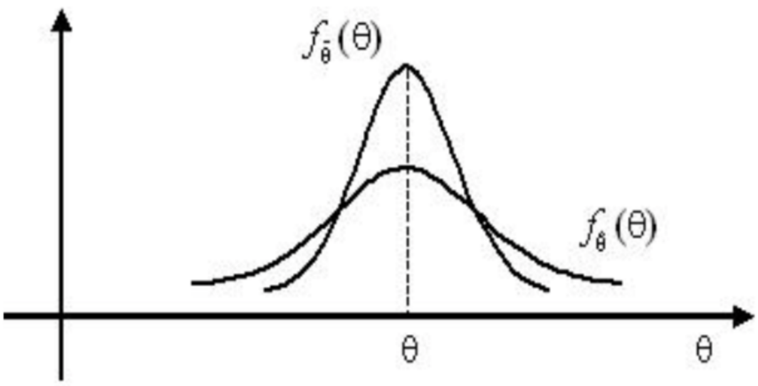
$$\lim_{n\to\infty} \mathbb{P}(|\widehat{\theta}_n - \theta| \geq \varepsilon) = 0$$

$$P(|\widehat{\theta}_n - \theta| < \varepsilon) \to 1$$

5.2.3 Эффективность

Эффективной называется оценка, которая имеет наименьший возможный разброс среди всех несмещенных оценок. Другими словами, оценка является эффективной, если она имеет наименьшую дисперсию среди всех несмещенных оценок. То есть, если $\widehat{\theta}_1$ и $\widehat{\theta}_2$ - две несмещенные оценки параметра θ и $\mathrm{Var}(\widehat{\theta}_1) \leq \mathrm{Var}(\widehat{\theta}_2)$ для любого значения θ, то $\widehat{\theta}_1$ называется более эффективной, чем $\widehat{\theta}_2$. Если оценка $\widehat{\theta}_1$ более эффективна, чем оценка $\widehat{\theta}_2$, то это означает, что реализации оценки $\widehat{\theta}_1$, рассчитанные для различных реализаций случайной выборки $X_1, \ldots, X_n$ объема n, будут иметь меньший разброс около оцениваемого параметра θ, чем реализации менее эффективной оценки $\widehat{\theta}_2$.

Эффективность оценки означает, что достигается минимум среднеквадратической ошибки (mean squared error) $E(hat(theta) - theta)^2$ (для несмещенных оценок MSE совпадает с дисперсией).



5.2.4 Асимптотическая нормальность

Оценка называется асимптотически нормальной, если её распределение приближается к нормальному с ростом размера выборки. То есть, если $\widehat{\theta}_n$ - оценка параметра θ по выборке размера n, то

$$\sqrt{n}(\widehat{\theta}_n - \theta) \to \mathcal{N}(0, \sigma^2) \quad \text{при } n \to \infty$$

.

5.3 Метод моментов

Метод моментов - это статистический метод оценки параметров распределения случайной величины на основе равенства теоретических моментов (средних значений) выборки и их оценок. Более конкретно, этот метод состоит в том, чтобы приравнять теоретические моменты распределения, выраженные через параметры этого распределения, к их оценкам, вычисленным на основе выборки. Затем, решив уравнения относительно параметров распределения, можно получить их оценки.

Конкретный алгоритм метода моментов может отличаться в зависимости от выбранного распределения и его параметров. Однако общий подход заключается в следующих шагах:

1. Выбирается распределение, которое, как предполагается, описывает выборку.
2. Выбирается число моментов, которые будут использоваться для оценки параметров. Обычно используются первый и второй моменты (среднее и дисперсия).
3. Выражаются теоретические моменты через параметры распределения.
4. Вычисляются оценки этих моментов на основе выборки.
5. Путем приравнивания теоретических моментов к их оценкам решаются уравнения относительно параметров распределения.
6. Полученные оценки параметров используются для описания и анализа выборки.

Все оценки, рассчитанные по методу моментов, являются состоятельными, однако их несмещённость и эффективность, так же, как и в случае метода подстановки, не гарантированы.

5.4 Метод максимального правдоподобия

Пусть есть наблюдаемые данные $x_1, x_2, ..., x_n$, которые распределены согласно какому-то параметрическому распределению с неизвестными параметрами θ. Тогда функция правдоподобия (likelihood function) определяется как произведение плотностей вероятности для каждого наблюдения:

$$L(\theta|x_1, x_2, ..., x_n) = f(x_1|\theta) \cdot f(x_2|\theta) \cdot ... \cdot f(x_n|\theta),$$

где $f(x|\theta)$ - это плотность вероятности для наблюдения x при условии параметров θ.

Метод максимального правдоподобия заключается в том, чтобы найти такие значения параметров θ, при которых функция правдоподобия достигает максимума:

$$\hat{\theta} = argmax L(\theta|x_1, x_2, ..., x_n)$$

Таким образом, метод максимального правдоподобия позволяет оценить параметры модели, которые наиболее вероятно приводят к наблюдаемым данным.

5.5 Информация Фишера

Определение 5.1. *Информация Фишера – это мера, которая индицирует, насколько информативна определенная выборка для оценки неизвестных параметров генеральной совокупности. Она выражает, насколько хорошо параметр может быть оценен на основе выборки данных.*

Интуитивно, информация Фишера увеличивается с увеличением количества информации, содержащейся в выборке данных. Используя информацию Фишера, можно оценить дисперсию точечной оценки. Чем больше информация Фишера, тем меньше дисперсия точечной оценки и тем более точной является оценка параметра.

Пусть $L(\theta, x_1, \dots, x_n)$ - функция правдоподобия для данной статистической модели. Тогда определена функция

$$I_n(\theta) = E_\theta\Big(\frac{\partial ln L(\theta, x_1, \dots, x_n)}{\partial \theta}\Big)^2,$$

где E_θ - математическое ожидание при данном θ, то она называтся информация Фишера для данной статистической модели при n независимых испытаниях. Поскольку математическое ожидание функции вклада выборки равно нулю, выписанная величина равна ее дисперсии.

Если выборка состоит из одного элемента, то информация Фишера записывается так:

$$I(\theta) = E_\theta\Big(\frac{\partial ln L(\theta, x)}{\partial \theta}\Big)^2$$

Из того, что в случае независимости случайных величин дисперсия суммы равна сумме дисперсий, следует, что в случае n независимых испытаний $I_n(\theta) = nI(\theta)$.

5.6 Неравенство Рао-Крамера

Неравенство Рао-Крамера устанавлиает ограничение на наименьшую возможную дисперсию для несмещенной оценки $\hat{\theta}$ параметра θ в терминах информации Фишера $I(\theta)$ и ее производной по θ. Формулировка неравенства Рао-Крамера следующая:

$$\mathrm{Var}(\hat{\theta}) \geq \frac{1}{I(\theta)},$$

где $\mathrm{Var}(\hat{\theta})$ - дисперсия оценки $\hat{\theta}$, а $I(\theta)$ - информация Фишера для параметра θ.

Таким образом, оценка $\hat{\theta}$ является более эффективной, если ее дисперсия близка к минимально возможной дисперсии, которую диктует неравенство Рао-Крамера.

Информация Фишера $I(\theta)$ зависит от формы функции правдоподобия, используемой для оценки $\hat{\theta}$. Если функция правдоподобия более крутая (то есть имеет большую производную), то информация Фишера будет выше, и тем самым более эффективные оценки $\hat{\theta}$ будут ближе к минимально возможной дисперсии.

Таким образом, оценка $\hat{\theta}$ будет более эффективной, если она имеет меньшую дисперсию и/или более крутую функцию правдоподобия.

5.7 Свойства оценок метода максимального правдоподобия

5.7.1 Несмещенность

Теорема. *Если существует несмещенная оптимальная оценка в регулярном случае, то она совпадает с оценкой максимального правдоподобия.*

Доказательство. Выразим вклад выборки:

$$\sqsupset \tau(\theta) = \theta$$
$$V(X, \theta) = \frac{1}{a(\theta)}(\hat{\theta} - \theta)$$

где $\hat{\theta}$ – эффективная оценка. Как нам убедится, что эта оценка совпадает с оценкой максимального правдоподобия? Подставим ОМП вместо θ:

$$V(X, \theta_*) = \frac{1}{a(\theta)}(\hat{\theta} - \theta_*)$$
$$0 = \frac{1}{a(\theta)}(\hat{\theta} - \theta_*)$$
$$0 = \hat{\theta} - \theta_*$$
$$\hat{\theta} = \theta_*$$

ч.т.д.

5.7.2 Состоятельность

Теорема. *Пусть θ_0 – реальный параметр, и пусть $\theta \neq \theta_0$. Тогда утверждается, что*

$$P_{\theta_0}(L(X;\theta_0) > L(X;\theta)) \longrightarrow 1$$

Доказательство. Перепишем наше неравенство в вероятности:

$$L(X;\theta_0) > L(X;\theta)$$
$$\frac{L(X;\theta)}{L(X;\theta_0)} < 1$$

Возьмем логарифмы от обеих частей и поделим на $\frac{1}{n}$:

$$\frac{1}{n}\sum_j ln\frac{p(X_j;\theta)}{p(X_j;\theta_0)} < 0$$

При росте $n \to \infty$, по ЗБЧ это сходится к

$$\xrightarrow{d} E_{\theta_0} ln\frac{p(X_j;\theta)}{p(X_j;\theta_0)} \leqslant E_{\theta_0}\left(\frac{p(X_j;\theta)}{p(X_j;\theta_0)} - 1\right)$$

Раскроем матожидание по линейности и получим разность интегралов от плостностей, и ввиду того, что интеграл от плотности равен 1, мы получаем на выходе 0.

$$= \int_X p(X;\theta)dX - \int_X p(X;\theta_0)dX = 0$$

Как отсюда следует состоятельность? Введем три события:

$$S_n = \{X : lnL(X;\theta_0) > lnL(X;\theta_0 - a)\} \cap \{X : lnL(X;\theta_0) > lnL(X;\theta_0 + a)\}$$
$$A_n = \left\{X : \left|\hat{\theta} - \theta\right| < a\right\}$$
$$B_n = \left\{X : \left.\frac{\partial \ln L(x,\theta)}{\partial\theta}\right|_{\theta=\hat{\theta}} = 0\right\}$$

Согласно предыдущему утверждению, $P(S_n)_{\theta_0} \longrightarrow 1$. Заметим, что

$$S_n \subset A_n B_n \subset A_n \Rightarrow P(A_n) \longrightarrow 1$$

ч.т.д.

5.7.3 Асимптотическая нормальность

Теорема. *(Дельта-метод)*
Пусть (ξ_n) – последовательность векторов случайных величин. Будем считать, что имеется следующая сходимость:

$$\sqrt{n}(\xi_n - \mu) \xrightarrow{d} \mathcal{N}(0, \Sigma)$$

Если дана функция $\varphi \in C^1(\mathbb{R}^m)$, $\varphi : \mathbb{R}^m \to \mathbb{R}$, то тогда будет верно

$$\sqrt{n}(\varphi(\xi_n) - \varphi(\mu)) \xrightarrow{d} \mathcal{N}\left(0, \bigtriangledown\phi(\mu)\Sigma(\bigtriangledown\phi(\mu))^T\right)$$

Доказательство. Сначала покажем, что $\xi_n \xrightarrow{p} \mu$.
Иначе это можно записать как $\xi_n - \mu \xrightarrow{p} 0$. Просто домножим и разделим на $\sqrt{n}$ – тогда разность, деленная на $\sqrt{n}$ будет стремится к нормальному многомерному закону, а $\frac{1}{\sqrt{n}}$ – к 0.

$$\frac{(\xi_n - \mu)\sqrt{n}}{\sqrt{n}} = \frac{(\xi_n - \mu)}{\sqrt{n}} \cdot \frac{1}{\sqrt{n}} \xrightarrow{p} 0$$

Теперь запишем разложение ϕ в ряд Тейлора:

$$\phi(\xi_n) = \phi(\mu) + \bigtriangledown\phi(\tilde{\mu})(\xi_n - \mu)$$

Заметим, что если $n \to \infty$, то

$$\phi(\xi_n) - \phi(\mu) \approx \bigtriangledown\phi(\mu)(\xi_n - \mu)$$
$$Var(\phi(\xi_n) - \phi(\mu)) \approx Var(\bigtriangledown\phi(\mu)(\xi_n - \mu))$$

Преобразуем дисперсию, убрав сдвиг:

$$Var(\bigtriangledown\phi(\mu)(\xi_n - \mu)) = Var(\bigtriangledown\phi(\mu)(\xi_n)) = \bigtriangledown\phi(\mu)Var\xi_n(\bigtriangledown\phi(\mu))^T$$

Вернемся теперь к исходному равенству и домножим на $\sqrt{n}$

$$\sqrt{n}(\phi(\xi_n) - \phi(\mu)) \approx \sqrt{n} \bigtriangledown \phi(\mu)(\xi_n - \mu)$$

Заметим, что разность стремится к 0 по предположению из начала доказательства, а все это суммарно стремится к

$$\xrightarrow{d} \mathcal{N}\left(0, \bigtriangledown\phi(\mu)\Sigma(\bigtriangledown\phi(\mu))^T\right)$$

ч.т.д.

Теорема. *Пусть $\Theta \subset \mathbb{R}^d$ – открытое множество, $\theta_0 \in \Theta$ – реальное значение неизвестного параметра. Также рассмотрим $\tau : \Theta \to \mathbb{R}$, причем $\tau \in C^1(\Theta)$.*

Предполагается регулярность модели. Будем полагать, что существует ровно один локальный максимум (а значит и глобальный) по θ функции $L(X;\theta)$ (значит, уравнение $\nabla_\theta \ln L(X, \theta) = 0$ имеет единственное решение $\hat{\theta}$ – оценку максимального правдоподобия). Также предполагаем, что $p(x, \theta)$ трижды дифференцируема по θ и

$$\left|\frac{\partial^3 \ln p(x;\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right| \leqslant M(x), \quad \mathrm{E}\, M(X) < \infty, \quad 1 \leqslant i, j, k \leqslant d.$$

Тогда справедливы следующие утверждения

- *$\hat{\theta}$ является асимптотической нормальной оценкой, то есть*

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{\mathrm{d}} \mathcal{N}(0, i^{-1}(\theta_0)),$$

  где $i^{-1}(\theta)$ – информационная матрица Фишера для одного наблюдения

- *$\tau(\hat{\theta})$ является асимптотически нормальной оценкой $\tau(\theta_0)$, то есть*

$$\sqrt{n}\frac{\tau(\hat{\theta}) - \tau(\theta_0)}{\sigma} \xrightarrow{\mathrm{d}} \mathcal{N}(0, 1),$$

  если $\sigma^2 = \nabla\tau(\theta_0)i^{-1}(\theta_0)\nabla^T\tau(\theta_0) > 0$

- *Справедлива сходимость*

$$\sqrt{n}\frac{\tau(\hat{\theta}) - \tau(\theta_0)}{\hat{\sigma}} \xrightarrow{\mathrm{d}} \mathcal{N}(0, 1),$$

  где $\hat{\sigma} = \sigma(\hat{\theta})$, причем σ – непрерывная по θ функция

Второй и третий пункт немедленно вытекают из первого (см. *дельта-метод*), поэтому мы докажем только первый пункт, причем для наглядности сделаем это при $d = 1$.

Доказательство. Вспомним, что $V(X;\theta) = \frac{\partial L(X;\theta)}{\partial\theta}$. Напишем разложение по формуле Тейлора в окрестности θ_0:

$$V(X;\theta) = V(X;\theta_0) + V'_\theta(X;\theta_0)(\theta-\theta_0) + \frac{V''_\theta(X;\theta^*)}{2}(\theta-\theta_0)^2,$$

где θ^* между θ и θ_0. Подставим вместо θ оценку максимального правдоподобия $\hat\theta$:

$$0 = V(X;\theta_0) + V'_\theta(X;\theta_0)(\hat\theta-\theta_0) + \frac{V''_\theta(X;\theta^*)}{2}(\hat\theta-\theta_0)^2$$

Сделаем несколько наблюдений:

- По ЦПТ

$$\sqrt{n}\cdot\overline{V(X;\theta_0)} \xrightarrow{\text{d}} \mathcal{N}\left(\mathrm{E}\,\frac{\partial\ln p(X;\theta_0)}{\partial\theta}, \mathrm{Var}\,\frac{\partial\ln p(X;\theta_0)}{\partial\theta}\right),$$

  причем ранее мы показали, что $\mathrm{E}\,\frac{\partial\ln p(X;\theta_0)}{\partial\theta} = 0$, а $\mathrm{Var}\,\frac{\partial\ln p(X;\theta_0)}{\partial\theta} = i(\theta_0)$

- По ЗБЧ

$$\overline{V'_\theta} \xrightarrow{\text{p}} \mathrm{E}\,\frac{\partial^2\ln p(X;\theta_0)}{(\partial\theta)^2} = -i(\theta_0)$$

- По ЗБЧ $\overline{|V''_\theta(X;\theta^*)|}$ ограничено , и $\hat\theta \xrightarrow{\text{p}} \theta_0$ (ранее мы показали состоятельность)

Перепишем полученное выше соотношение после подстановки $\hat\theta$ в разложение Тейлора следующим образом:

$$\hat\theta-\theta_0 = -V(X;\theta_0)\left(V'_\theta(X;\theta_0) + \frac{V''_\theta(X;\theta^*)}{2}(\hat\theta-\theta_0)\right)^{-1},$$

$$\sqrt{n}(\hat\theta-\theta_0) = -\sqrt{n}\cdot\overline{V(X;\theta_0)}\left(\overline{V'_\theta(X;\theta_0)} + \frac{\overline{V''_\theta(X;\theta^*)}}{2}(\hat\theta-\theta_0)\right)^{-1}$$

В силу выше сделанных наблюдений и свойств вероятностных сходимостей $(\ldots)^{-1} \xrightarrow{\text{p}} -i^{-1}(\theta_0)$, а $-\sqrt{n}\cdot\overline{V(X;\theta_0)} \xrightarrow{\text{d}} \mathcal{N}(0, i(\theta_0))$, откуда немедлено вытекает искомая сходимость *ч.т.д.*

5.8 Экспоненциальное семейство распределений

Определение. *(Экспоненциальное семейство распределений)*
Пусть модель регулярна. Тогда, распределение относится к экспоненциальному, если его плотность представима в следующей форме:

$$p(X,\theta) = exp(A(\theta)B(X) + C(\theta) + D(X))$$

Пример. $N(), \Gamma(), Pois(), Bin(), NB()$ *относятся к экспоненциальным.*

Давайте попробуем воспользоваться методом максимального правдоподобия.

$$\ln p(X,\theta) = A(\theta)B(X) + C(\theta) + D(X)$$
$$\frac{\partial \ln p(x,\theta)}{\partial\theta} = A'(\theta)B(X) + C'(\theta)$$
$$V(X,\theta) = A'(\theta)\sum_{i=1}^{n} B(X_i) + nC'(\theta)$$
$$V(X,\theta) = n(A'(\theta)\overline{B(X_i)} + C'(\theta))$$
$$\frac{V(X,\theta)}{n} - C'(\theta) = A'(\theta)\overline{B(X)}$$
$$\overline{B(X)} = \frac{V(X,\theta)}{nA'(\theta)} - \frac{C'(\theta)}{A'(\theta)}$$

Из чего следует, что $\overline{B(X)}$ – оптимальная оценка для $-\frac{C'(\theta)}{A'(\theta)}$

5.9 Байесовские оценки

Пусть дана выборка $X_1,\dots,X_n \sim F_\theta$.
В стандартной постановке искомый параметр – некоторое фиксированное неизвестное число. В байесовской же постановке это случайная величина из некого априорного распределения (будем обозначать как $\theta \sim \pi(\theta)$)

Введем $l(\hat\theta,\theta)$ – функция потерь. Обычно в ее качестве используют MSE: $l(\hat\theta,\theta) = (\hat\theta-\theta)^2$
Как мы определяем эффективность оценки? Как ту, которую минимизирует среднеюю ошибку.

Определение. *(риск)* $R(\hat\theta,\theta) = E_{F_\theta} l(\hat\theta,\theta)$

Определение. *(байесовский риск)* $r(\hat\theta,\theta) = E_{\pi_\theta} R(\hat\theta,\theta)$

Определение. *(байесовская оценка)* $\hat{\theta_B} = argmin_{\hat\theta}(\hat\theta)$

Можно считать, что в риске мы считаем матожидание по фиксированной θ, тогда же как в байесовском риске – по всем θ из априорного распределения.
Вспомним теорему Байеса, и запишем ее для оценки:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$
$$P(\theta|X) = \frac{L(X|\theta)\pi(\theta)}{\displaystyle\int L(X|\theta)\pi(\theta)d\theta}$$

где $L()$ – функция правдоподобия.
В задачх машинного обучения иногда используется байесовский подход, и в нем

$$posterior \approx likelihood \times prior$$

Оказывается, что байесовскую оценку можно посчитать как

$$\hat{\theta_B} = argmin_{\hat\theta} E(l(\hat\theta,\theta)|X)$$

Доказательство. Пусть $\theta_* = argmin_{\hat{\theta}} E(l(\hat{\theta}, \theta)|X)$

$$r(\theta_*) = EE[l(\theta_*, \theta)|X] \leqslant EE[l(\hat{\theta}, \theta)|X] = r(\theta)$$

Таким образом, действительно, для подсчета байесовской оценки нужно найти функцию, минимизирующую loss при условии X. *ч.т.д.*

Лемма. *Если функция потерь – квадратическая (т.е. $l(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$), то*

$$\hat{\theta}_B = E[\theta|X]$$

Доказательство.

$$\begin{gathered} E((\hat{\theta} - \theta)^2|X) = \\ E[(\theta - E(\theta|X) - (\hat{\theta} - E(\theta|X)))^2|X] = \\ E[(\theta - E(\theta|X))^2|X] + E[(\hat{\theta} - E(\theta|X))^2|X] - 2E[(\theta - E(\theta|X))(\hat{\theta} - E(\theta|X))|X] = \\ Var(\theta|X) + (\hat{\theta} - E(\theta|X))^2 + const(E(\theta|X) - E(\theta|X)) = \\ Var(\theta|X) + (\hat{\theta} - E(\theta|X))^2 \end{gathered}$$

ч.т.д.

5.10 Минимаксные оценки

Пусть дана функция $m(\hat{\theta}) = \sup_\theta R(\hat{\theta}, \theta)$

Определение. *(минимаксная оценка)* $\hat{\theta}_{WC} = argmin_{\hat{\theta}}(m(\hat{\theta}))$

Теорема. *Если существует априорное распределение $\pi(\theta)$ такое, что $R(\hat{\theta}_B, \theta) \equiv const$, то*

$$\hat{\theta}_{WC} = \hat{\theta}_B$$

Доказательство. Для начала, заметим, что $r(\hat{\theta}) \leqslant m(\hat{\theta})$.
Теперь построим доказательство от противного: предположим, что $\exists \hat{\theta} : m(\hat{\theta}) < m(\hat{\theta}_B)$
Тогда, можем построить следующее неравенство:

$$r(\hat{\theta}) \leqslant m(\hat{\theta}) < m(\hat{\theta}_B) = r(\hat{\theta}_B)$$

Противоречие, так как байесовская оценка по определению является оценкой с наименьшим байесовским риском. *ч.т.д.*

5.11 Доверительные интервалы

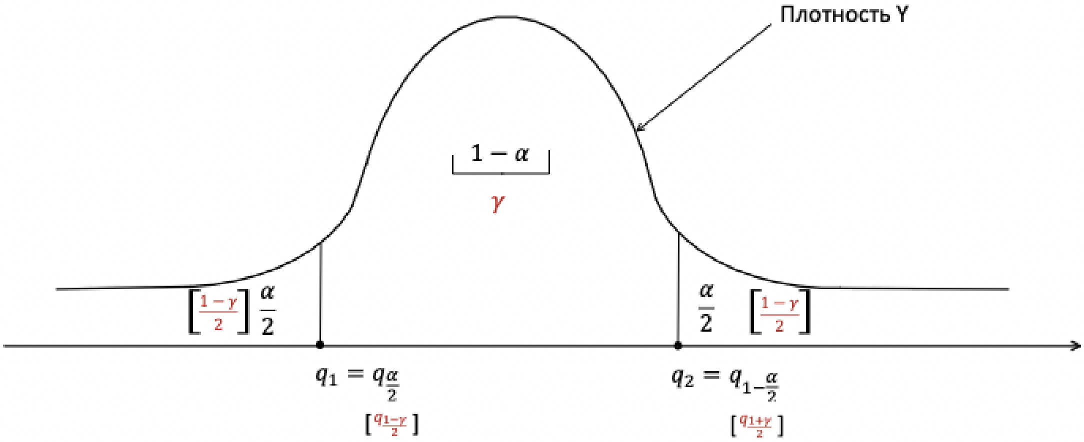
Определение 5.2. *Доверительный интервал*
Пусть есть выборка $X_1, \ldots, X_n$ с распределением F_θ; $\theta \in \Theta \subset \mathbb{R}$.
$T_1(X)$, $T_2(X)$ - статистики.
Будем называть интервал $[T_1(X), T_2(X)]$ доверительным интервалом уровня $1-\alpha$, если:

$$P(\theta \in [T_1(X), T_2(X)]) \geqslant 1 - \alpha$$

Общая схема построения доверительных интервалов:
Пусть $G(X, \theta) \sim Y$ - центральная статистика, при этом распределение Y непрерывно и не зависит от неизвестных параметров, тогда:

$$P(q_1 \leqslant G(X, \theta) \leqslant q_2) = 1 - \alpha$$

Из данной вероятности выражается θ.



Где $q_{\frac{\alpha}{2}}$ и $q_{1-\frac{\alpha}{2}}$ - квантили порядка $\frac{\alpha}{2}$ и $1 - \frac{\alpha}{2}$ соответственно.

Пример 5.1. *Имеется выборка* $X_1, \ldots, X_2 \sim \mathcal{N}(\mu, \sigma^2)$

a. $\mu -?$, σ^2 *- известна*
Возьмем статистику $\sqrt{n}\frac{\overline{X}-\mu}{\sigma} \sim \mathcal{N}(0,1)$. *Она имеет нормальное распределение с матожиданием равным нулю и дисперсией равной единицей в виду того, что исходная выборка имеет нормальное распределение.*

$$P(-q_{1-\frac{\alpha}{2}} \leqslant \sqrt{n}\frac{\overline{X} - \mu}{\sigma} \leqslant q_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

Выразим из данного выражения μ*:*

$$P(\mu \in [\overline{X} - \frac{q_{1-\frac{\alpha}{2}} \cdot \sigma}{\sqrt{n}}; \overline{X} + \frac{q_{1-\frac{\alpha}{2}} \cdot \sigma}{\sqrt{n}}])$$

b. μ *- известно,* $\sigma^2 -?$
Пусть $U_1, \ldots, U_n - i.i.d. \sim \mathcal{N}(0,1)$, *тогда* $U_1^2, \ldots, U_n^2 \sim \chi^2(n)$
Заметим, что $\frac{X_i - \mu}{\sigma}$ *имеет стандартное нормальное распредление. Поэтому, если преобразовать ее следующим образом, она будет иметь распредление* χ^2*:*

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(n)$$

Воспользуемся полученной статистикой для построения доверительного интервала:

$$P(q_{\frac{\alpha}{2}} \leqslant \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 \leqslant q_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

Выразим отсюда σ^2*:*

$$P\left(\frac{\sum\limits_{i=1}^{n}(X_i - \mu)^2}{q_{1-\frac{\alpha}{2}}} \leqslant \sigma^2 \leqslant \frac{\sum\limits_{i=1}^{n}(X_i - \mu)^2}{q_{\frac{\alpha}{2}}}\right)$$

c. *$\mu-?$, $\sigma^2-?$, построить доверительный интервал для σ^2*
Воспользуемся теоремой Фишера и получим статистику:

$$\frac{n \cdot S_*^2}{\sigma^2} \sim \chi^2(n-1),$$

где S_*^2 - выборочная смещённая дисперсия

$$P(q_{\frac{\alpha}{2}} \leqslant \frac{n \cdot S_*^2}{\sigma^2} \leqslant q_{1-\frac{\alpha}{2}}) = 1-\alpha$$

Выразим σ^2:

$$P\Big(\frac{n \cdot S_*^2}{q_{1-\frac{\alpha}{2}}} \leqslant \sigma^2 \leqslant \frac{n \cdot S_*^2}{q_{\frac{\alpha}{2}}}\Big)$$

d. *$\mu-?$, $\sigma^2-?$ построить доверительный интервал для μ*
Пусть есть $U_0, U_1, \ldots, U_n - i.i.d. \sim \mathcal{N}(0,1)$, тогда:

$$T_n = \frac{U_0}{\sqrt{\frac{1}{n}\sum\limits_{i=1}^{n} U_i^2}} \sim T(n),$$

где $T(n)$ - распределение Стьюдента (t-distribution) с n степенями свободы, симметричное относительно 0. Также заметим, что $\sum\limits_{i=1}^{n} U_i^2 \sim \chi^2(n)$.
По теореме Фишера $\overline{X}$ и S_*^2 - независимы. Рассмотрим следующую статистику, где числитель и знаменатель знакомы нам из предыдущих примеров:

$$\frac{\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}}{\sqrt{\frac{1}{n-1} \cdot \frac{n \cdot S_*^2}{\sigma^2}}} \sim T(n-1),$$

где $\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma} \sim \mathcal{N}(0,1)$; $\frac{n \cdot S_*^2}{\sigma^2} \sim \chi^2(n-1)$
Сократим получившееся выражение:

$$\sqrt{n-1}\frac{\overline{X}-\mu}{S_*}$$

Тогда:

$$P(-q_{1-\frac{\alpha}{2}} \leqslant \sqrt{n-1}\frac{\overline{X}-\mu}{S_*} \leqslant q_{1-\frac{\alpha}{2}}) = 1-\alpha$$

$$P\Big(\mu \in \Big[\overline{X} - \frac{q_{1-\frac{\alpha}{2}} \cdot S_*}{\sqrt{n-1}}; \overline{X} + \frac{q_{1-\frac{\alpha}{2}} \cdot S_*}{\sqrt{n-1}}\Big]\Big)$$

e. *Пусть имеются две независимые выборки объемов n и m соответственно: $X_1, \ldots, X_n \sim \mathcal{N}(\mu_1, \sigma_1^2)$ и $Y_1, \ldots, Y_m \sim \mathcal{N}(\mu_2, \sigma_2^2)$. При этом μ_1, μ_2 - известны. Построить доверительный интервал для $\frac{\sigma_2^2}{\sigma_1^2}$*
Воспользуемся распределением Фишера (F распределением):

$$U_1 \sim \chi^2(n_1),$$

$$U_2 \sim \chi^2(n_2),$$

где U_1 и U_2 независимы. Тогда:

$$\frac{U_1/n_1}{U_2/n_2} \sim F(n_1, n_2)$$

Выведем статистику с необходимым нам распределением:

$$\frac{\sum\limits_{i=1}^{n}(X_i-\mu_1)^2 \cdot \sigma_2^2 m}{\sigma_1^2 n \cdot \sum\limits_{j=1}^{m}(Y_i-\mu_2)^2} \sim F(n,m),$$

где $\frac{\sum\limits_{i=1}^{n}(X_i-\mu_1)^2}{\sigma_1^2 n} \sim \frac{\chi^2(n)}{n}$ *и* $\frac{\sigma_2^2 m}{\sum\limits_{j=1}^{m}(Y_i-\mu_2)^2} \sim \frac{m}{\chi^2(m)}$ *Получаем:*

$$P\left(q_{\frac{\alpha}{2}} \leqslant \frac{\sum\limits_{i=1}^{n}(X_i-\mu_1)^2 \cdot m}{n \cdot \sum\limits_{j=1}^{m}(Y_j-\mu_2)^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \leqslant q_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

Выведем $\frac{\sigma_2^2}{\sigma_1^2}$*:*

$$P\left(\frac{\sigma_2^2}{\sigma_1^2} \in \left[\frac{q_{\frac{\alpha}{2}} \cdot n \cdot \sum\limits_{j=1}^{m}(Y_j-\mu_2)^2}{\sum\limits_{i=1}^{n}(X_i-\mu_1)^2 \cdot m}; \frac{q_{1-\frac{\alpha}{2}} \cdot n \cdot \sum\limits_{j=1}^{m}(Y_j-\mu_2)^2}{\sum\limits_{i=1}^{n}(X_i-\mu_1)^2 \cdot m}\right]\right)$$

5.12 Асимптотические доверительные интервалы

Определение 5.3. *Пусть дана выборка $X_1, X_2, ..., X_n$ с неизвестным параметром θ, где каждый X_i является случайной величиной. Тогда $[T_1(X), T_2(X)]$ – асимптотический доверительный интервал для θ с заданным уровнем доверия $1-\alpha$, если:*

$$\lim_{n\to\infty} P\left(T_1(X) \le \theta \le T_2(X)\right) \ge 1-\alpha = \Gamma,$$

где статистики $T_1(X)$ и $T_2(X)$ являются точечными оценками параметра θ и имеют асимптотически нормальное распределение с асимптотической дисперсией σ^2

Статистика для асимптотического доверительного интервала:

$$G(X,\theta) \xrightarrow[n\to\infty]{d} Y,$$

где Y не зависит от θ.

Пример 5.2. *Построить асиптотический доверительный интервал для параметра распределения Бернулли.*
Решение:
Воспользуемся ЦПТ и запишем:

$$\frac{\sum X_i - np}{\sqrt{np(1-p)}} = \sqrt{n}\frac{\bar{X}-p}{\sqrt{p(1-p)}} \xrightarrow{U} \sim N(0,1)$$

Заметим, что

$$P(|S_*^2 - p(1-p)| > \epsilon \xrightarrow{0}$$

$$S_* \xrightarrow{p} \sqrt{p(1-p)}$$

Тогда можем написать:

$$\sqrt{n}\frac{\bar{X}-p}{S_*} \xrightarrow{U} \sim N(0,1)$$

Напишем искомый доверительный интервал:

$$\bar{X} - \frac{q_{1-\frac{\alpha}{2}}S_*}{\sqrt{n}} \le p \le \bar{X} + \frac{q_{1-\frac{\alpha}{2}}S_*}{\sqrt{n}}$$

Замечание: для распределения Бернулли p – вероятность успеха, который соответствует математическому ожиданию.

Обобщая, мы построили асимптотический доверительный интервал для математического ожидания, то есть:

$$\frac{\sqrt{n}(\bar{X}-\mu)}{S_*} \xrightarrow{d} Y \sim N(0,1)$$

Получаем асимптотический доверительный интервал:

$$\left[\bar{X} - \frac{q_{1-\frac{\alpha}{2}}S_*}{\sqrt{n}}; \bar{X} + \frac{q_{1-\frac{\alpha}{2}}S_*}{\sqrt{n}}\right]$$

Построим асимптотический доверительный интервал для дисперсии выборки:

$$\frac{\sqrt{n}(S_*^2-\sigma^2)}{\sqrt{\mu_4 - S_*^4}} \xrightarrow{d} U \sim N(0,1)$$

Получаем:

$$\left[S_*^2 - \frac{q_{1-\frac{\alpha}{2}}\sqrt{\mu_4 - S_*^4}}{\sqrt{n}}; S_*^2 + \frac{q_{1-\frac{\alpha}{2}}\sqrt{\mu_4 - S_*^4}}{\sqrt{n}}\right]$$

Теорема 5.1. *Теорема об асимптотическом поведении крайних членов вариационного ряда:*

Пусть $X_1, X_2, \ldots, X_n$ - выборка из распределения с функцией распределения $F(x)$. Тогда при $n \to \infty$ выполнены следующие сходимости:

$$nF(x_r) \to U_1 \sim \Gamma(r,1)$$

$$n(1 - F(x_{(n+1-s)})) \to U_2 \sim \Gamma(s,1)$$

5.13 Теорема Фишера и примыкающие к ней леммы

Лемма. *(о независимости линейной и квадратичной статистик)*
Пусть даны $X_1, ..., X_n \sim N(\mu, \sigma^2)$ - независимые случайные величины,

- $T = AX$, *где* $X = (X_1, \ldots, X_n)^T$, $A = \mathbb{M}_{m\times n}(\mathbb{R})$
- $Q = X^TBX$, *где* $B \in \mathbb{M}_n(\mathbb{R}), B = B^T$
- $AB = 0$

тогда T и Q – независимые.

Доказательство. Рассмотрим $\Lambda = U^T B U$, где

- $\Lambda = diag(\lambda_1, \lambda_2, \dots, \lambda_m, 0, \dots, 0)$ и где λ_k – собственные числа, $\lambda_k \neq 0$
- $U = (u_1, \dots, u_m)$ – ортонормированный базис (u_k – собственные вектора)

Соответственно, мы можем выразить B как

$$B = U \Lambda U^T = \sum_{j=1}^{m} \lambda_j u_j u_j^T$$

$$\Rightarrow Q = \sum_{j=1}^{m} \lambda_j (X^T u_j)(u_j^T X) = \sum_{j=1}^{m} \lambda_j (u_j^T X)^2$$

Выпишем равенство $AB = 0$:

$$\Rightarrow A \left(\sum_{j=1}^{m} \lambda_j u_j u_j^T \right) = 0$$

$$\sum_{j=1}^{m} \lambda_j A u_j u_j^T = 0$$

Зафиксируем $1 \leqslant k \leqslant m$ и домножим на u_k справа. У нас не занулится только одно слагаемое:

$$\lambda_k A u_k u_k^T u_k = \lambda_k A u_k = 0$$

$$A u_k = 0 \quad \text{– ввиду того, что } \lambda \neq 0$$

$$\Rightarrow \forall i : A_{[i,*]} u_k = 0$$

Давайте теперь введем вектор

$$(u_1^T X, \dots, u_m^T X, A_{[i,*]} X, \dots, A_{[m,*]} X)^T = (U^T, A)^T X$$

Что можно сказать про распределение этого преобразования? X – гауссовский вектор, и следовательно, все преобразование имеет нормальное распредление.
Давайте проверим, что $\forall i, k : A_{[i,*]} X$ и $u_k^T X$ – независимые, а для этого посчитаем ковариацию:

$$cov(A_{[i,*]} X, u_k^T X) = cov(A_{[i,*]} X, X^T u_k) = A_{[i,*]} cov(X, X^T) u_k$$

При этом, $cov(X, X^T)$ заведомо $\neq 0$, а значит

$$= A_{[i,*]} Var X u_k = \sigma^2 A_{[i,*]} u_k = 0$$

ч.т.д.

Лемма. *(о независимости двух квадратичных статистик)*
Пусть даны $X_1, ..., X_n \sim N(\mu, \sigma^2)$ - независимые случайные величины,

- $Q_1 = X^T B_1 X$*, где* $B_1 \in \mathbb{M}_n(\mathbb{R}), B_1 = B_1^T$
- $Q_2 = X^T B_2 X$*, где* $B_2 \in \mathbb{M}_n(\mathbb{R}), B_2 = B_2^T$

- $B_1 B_2 = B_2 B_1 = 0$

тогда Q_1 и Q_2 – независимые.

Доказательство. Доказывается аналогично предыдущей лемме. *ч.т.д.*

Определение. *(Хи-квадрат)*
Если случайные величины $\xi_1, \xi_2, \dots$ независимы и имеют стандартное нормальное распределение, то распределение случайной величины

$$X_n^2 = \xi_1^2 + ... + \xi_n^2.$$

называется распределением χ^2 (хи-квадрат) с n степенями свободы

Лемма. *(о распределении квдаратичной статистики)*
Пусть даны $X_1, ..., X_n \sim N(0,1)$

- $Q = X^T B X$
- $B = B^2$

тогда $Q \sim \chi^2(r)$, где $r = rg(B) = tr(B)$.

Доказательство. Введем следующую цепочку равенств:

$$\lambda u = Bu = B^2 u = B\lambda u = \lambda^2 u$$

Утверждается, что собственные числа λ_k равны либо 0, либо 1. Пользуясь предыдущей леммой, выведем B как

$$B = \sum_{k=1}^{r} \lambda_k u_k u_k^T$$
$$B = \sum_{k=1}^{r} u_k u_k^T$$

Тогда, мы хотим доказать, что

$$Q = \sum_{k=1}^{r} (u_k^T X)^2 \sim \chi^2(r)$$

1. $u_k^T \sim N(u_k^T EX, u_k^T I_n u_k) = N(0,1)$
2. Докажем независимость.
   $cov(u_k^T X, u_j^T X) = 0$, так как скалярное произведение $\langle u_k, u_j \rangle = 0$
3. Докажем, что $r = tr(B)$.

$$B = U\Lambda U^T$$
$$rgB = rg\Lambda = tr\Lambda \quad \text{– ввиду того, что на диагонали } \Lambda \text{ только 0 и 1}$$
$$trB = tr(U\Lambda U^T)$$

Заметим, что $B_{j,j} = \lambda_j u_j u_j^T = \lambda_j$

ч.т.д.

Теорема 5.2. *(Фишер)*
Пусть даны $X_1, ..., X_n \sim N(\mu, \sigma^2)$, значит верны следующие свойства:

1. $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$
2. $\frac{nS_*^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$
3. $S^2, \overline{X}$ *– независимы*
   $S_*^2, \overline{X}$ *– независимы*

Доказательство. Рассмотрим по пунктам:

1. Выводится из свойств нормального распределения.
2. Рассмотрим $Y_j = \frac{X_j - \mu}{\sigma}$. Тогда,

$$\overline{Y_j} = \frac{1}{\sigma}(\overline{X} - \mu)$$
$$S_*^2(Y) = \frac{1}{n}\sum_{j=1}^{n}(Y - \overline{Y})^2 = \frac{S_*^2(X)}{\sigma^2}$$

   По определению,

$$\overline{Y} = \frac{\sum_j Y_j}{n} = \left(\frac{1}{n} \cdot \frac{1}{n} \dots \frac{1}{n}\right) \cdot (Y_1 \dots Y_n) = bY$$
$$nS_*^2(Y) = (Y - BY)^T(Y - BY) = Y^T(I - B)^T(I - B)Y = Y^T(I - B)Y$$

   Последнее равенство верно ввиду того, что $B = B^2$:

$$I - B + B + B^2 = I - B + 2B = I - B$$

   В итоге, матрица совпала с ее квадаратом. А значит, соблюдаются условия предыдущей леммы, и

$$\Rightarrow \chi^2(tr(I - B)) = \chi^2(n-1), \text{ где}$$
$$tr(I - B) = \sum_{k=1}^{n}(1 - \frac{1}{n}) = n - 1$$

3. $b(I - b) = b - b = 0$, по лемме о независимости линейной и квадратичной статистик

ч.т.д.

Глава 6

Проверка статистических гипотез

6.1 Постановка задачи проверки статистической гипотезы

На входе имеется выборка X объема n, причем выборка может из себя представлять просто набор из n чисел и иметь более нетривиальную структуру.

Имеется некоторая проблема. Утверждение "по умолчанию" для данной проблемы будем называть нулевой гипотезой и обозначать его H_0 (монетка подбрасывается честно, между явлениями нет никакой связи, некоторый показатель не имеет аномалий и т.п.). Наши "подозрения" будем называть альтернативной гипотезой и обозначать его H_1 (подбрасывание монетки нечестно/смещено к орлу, между явлениями есть связь, показатель имеет аномалии и т.д).

6.1.1 Статистический критерий

Опишем принцип работы статистического критерия или статистического теста – функции, по выборке возвращающая H_0 или H_1. Зададим число $\alpha \in (0; 1)$, которое мы будем называть уровень значимости (типичные значения: 0.1 , 0.05, 0.01). Если критерий выбрал H_0, то говорят, что гипотеза H_0 принята (accept), если H_1 – гипотеза H_0 опровергнута (receccted) в пользу альтернативной H_1.

Пусть $T(X)$ – функция от выборки, которая при условии справедливости нулевой гипотезы имеет распределение (или стремится по распределению при $n \to \infty$) с плотностью $p(.)$. Пусть $T_{0,\alpha}$, $T_{1,\alpha}$ – области принятия и опровержения (критическая) области, то есть

$$\mathrm{P}(T(X) \in T_{0,\alpha}) \simeq 1 - \alpha,$$
$$\mathrm{P}(T(X) \in T_{1,\alpha}) \simeq \alpha,$$

где запись $\mathrm{P}(\ldots) \simeq c$ означает, что вероятность $\mathrm{P}(\ldots)$ равняется или стремится при $n \to \infty$ к c. Как правило, $T_{1,\alpha} = \overline{T_{0,\alpha}}$.

Далее рассмотрим три типа критериев и областей $T_{0,\alpha}$ и $T_{1,\alpha}$: двусторонний, левосторонний и правосторонний.

1. Двусторонний случай. $T_{0,\alpha} = [q_{\alpha/2}, q_{1-\alpha/2}]$;
2. Односторонний случай.
   - Правосторонняя альтернатива. $T_{0,\alpha} = (-\infty, q_{1-\alpha}]$;
   - Левосторонняя альтернатива. $T_{0,\alpha} = [q_\alpha, +\infty)$.

Здесь q_c – квантиль порядка c.

p-value

- левосторонний случай: $p_l := p[Y \leqslant T(x)|H_0]$; если $T(x) > q_2$, то $pl > \alpha$;
- правосторонний: $p_2 := p[Y \geqslant T(x)|H_0]$; если $T(x) < q_1 - 2$, то $p_2 > \alpha$, иначе $p_2 \leqslant \alpha$;
- двусторонний случай: $p := 2min(p_l; p_2)$; принимаем $H_0 \Leftrightarrow \leqslant \alpha$.

Состоятельность выборки: при росте объема выборки $\beta \to 0$. $W(F) := p[G(X_n) \in T_1(\alpha)|F]$, где F-распределение; $\alpha = W(F_0), F_0$ соответствует распределению нулевой гипотезы; $\beta = 1 - W(F_1) = P[G(X_n) \in T_0(\alpha)|F_1]$, F_1 - соответствует распределению альтернативной гипотезы; $1 - \beta = W(F_1)$-мощность критерия. На практике фиксируется α и выбирается критерий с наибольшей мощностью.

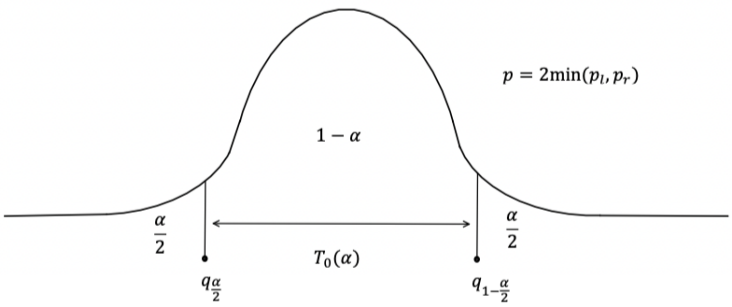
6.2 Проверка гипотез и доверительные интервалы

$X_1, ..., X_n \sim F_\theta$; $G(X, \theta) \underset{\to}{\sim} Y$ - не зависит от θ H_0 $\quad \theta = \theta_0$ 3 альтернативы:

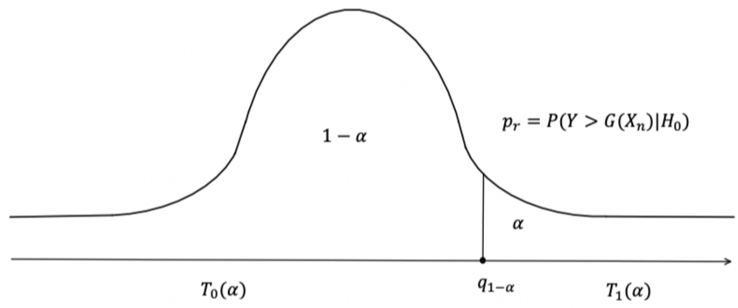
- двусторонняя альтернатива: $\theta \neq \theta_0$;
- правосторонняя альтернатива: $\theta \geqslant \theta_0$;
- левосторонняя альтернатива: $\theta \leqslant \theta_0$.

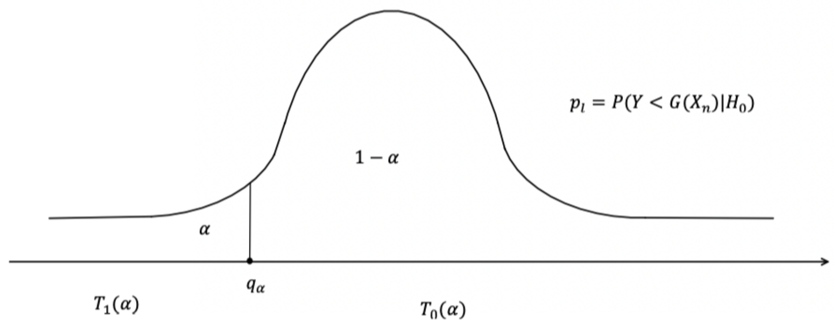
Пример 6.1. *Рассмотрим Пример с монеткой. Пусть робот подбрасывает монетку независимо и одинаково, то есть у нас имеется реализация бернулиевских величин с вероятностью успеха p (условимся считать, что p — вероятность орла). По умолчанию робот кидает монетку честно, то есть $p = \frac{1}{2}$, что будет являться H_0. Возможны три подозрения:*

- *робот нечестно кидает монету, то есть $p! = \frac{1}{2}$ В этом случае $p! = \frac{1}{2}$ и будет H_1*



- *робот кидает монетку так, чтобы скорее выпадал орел, то есть $p > \frac{1}{2}$, что и будет H_1*
- *робот кидает монетку так, чтобы выпала решка, то есть $p < \frac{1}{2}$, что и будет H_1*





6.3 Критерии хи-квадрат

6.3.1 Критерий Колмогорова-Смирнова

Теорема 6.1. *(Критерий Колмогорова)*

- $X_1, \ldots, X_n \sim F$
- $H_0 : F = F_0$ *(F_0 - непрерывна)*
- $H_1 : F \neq F_0$

$D_n = \sqrt{n} \cdot \sum\limits_{x \in \mathbb{R}} |F_n(x) - F_0(x)|$, *$F_n$ - эмпирическая функция распределения*
if $D_n > q_{1-\alpha}$ then reject H_0 else accept H_0

Идея критерия основана на одноименной теореме Колмогорова.

- $n \geq 20$ работает хорошо, при маленьких n есть спец таблицы
- Так же есть приблеженные формулы для D_n
- Есть следующая «сложная» модификация:
  $H_0 : F = F(\theta), H_1 : \neg H_0 \Longrightarrow D_n = \sqrt{n} \underset{x}{sup} |F_n(x) - F_0(x, \theta)|$, где $\theta \to \widehat{\theta}$

Теорема 6.2. *(Критерий Сиирнова)*

- $X_1, \ldots, X_n$ – *независимые*
- $Y_1, \ldots, Y_m$ – *независимые*
- $H_0 : F_X = F_Y (= F_0)$

- $H_1 :\neq H_0$

$D_{n,m} = \sqrt{n} \underset{x}{sup} |F_n(x) - F_m(x)|$
$T_1(\alpha) = (q_{1-\alpha}, +\infty)$

Тут идея тоже основана на одноименной формуле Смирнова.

6.3.2 Критерий согласия Пирсона хи-квадрат

Теорема 6.3. $X_1, \ldots, X_n \sim F(x)$ *- непрерывные.*
Давайте дискритезируем данные:
$\Delta_1 : \nu_1$ *- количество элементов выборки, попадающих в* Δ_1
...
$\Delta_N : \nu_N$
$p_{\Delta_k} = \int_{\delta_k} p(x)dx$, $p(x) = F'(x)$
Рассмотрим $\{1, 2, \ldots, N\}$, $p = (p_1, \ldots, p_N)$ *- настоящий вектор вероятностей*

- $p_0 = (p_{01}, \ldots, p_{0N})$ *- ожидаемый фиксированный вектор вероятностей*
- ν_k *- количество элементов в выборке типа* k
- $H_0 : p = p_0$
- $H_1 : p \neq p_0$
- $n = \sum_{k=1}^{N} \nu_k$

$$\Xi_N^2 = \sum_{k=1}^{N} \frac{(\nu_k - np_{0k})^2}{np_{0k}}$$

Теорема. $\Xi_N^2 \Longrightarrow \Xi^2(N-1)$ *при условии* H_0

Доказательство. Рассмотрим доказательство при $N = 2$:

$$\frac{(\nu_1 - np_{01})^2}{np_{01}} + \frac{(\nu_2 - np_{02})^2}{np_{02}} = \frac{(\nu_1 - np_{01})^2}{np_{01}} + \frac{(n - \nu_1 - n(1 - p_{01}))^2}{n(1 - p_{01})} =$$
$$\frac{(\nu_1 - np_{01})^2}{n} \left(\frac{1}{p_{01}} + \frac{1}{1 - p_{01}} \right) = \frac{(\nu_1 - np_{01})^2}{np_{01}(1 - p_{01})}$$

Это выражение, если рассмотреть его без квадратов, будет стремится по ЦПТ к $N(0, 1)$, но мы можем навесить на него непрерывную функцию возведения в квадрат и получим то, что нам нужно.

ч.т.д.

- Критерий состоятельный.
- Критическая область критерия правосторонняя.
- Оптимальные способы дискритизации существуют.

- Для проверки согласованности с нормальным законом следует использовать специальные тесты (Шапиро, коэффициент эксцесса)

Теорема. *(сложная гипотеза)*

- $H_0 : p = p_0(\theta)$
- $\theta \in \Theta \subset \mathbb{R}^d, d < N - 1$

$$\Xi_N^2 = \sum_{k=1}^{N} \frac{(\nu_k - np_{0k}(\theta))^2}{np_{0k}(\theta)}|_{\theta=\widehat{\theta}}$$

Теорема. *Пусть верны следующие утверждения:*

- $p_0(\theta) > 0, \forall\theta$
- $\frac{\partial p_0}{\partial\theta}, \frac{\partial^2 p_0}{(\partial\theta)^2}$ *- непрерывные*

Тогда:

$$\Xi_N^2 \to X^2(N-1-d)$$
$$rk\left(\frac{\partial p_{0k}}{\partial\theta_j}\right)_{1\le k\le N, 1\le j\le d} = d$$

6.3.3 Критерий однородности хи-квадрат

Теорема 6.4. *(Критерий однородности хи-квадрат)*

- K *независимых выборок, все они из множества* $\{1, 2, \dots, N\}$
- $p^{(j)}$ *- истинный вектор вероятностей для соответствующей выборки*
- $H_0 : p^{(1)} = \dots = p^{(k)}$
- $H_1 :\neq H_0$

Пусть ν_{ij} *- количество элементов типа* j *в* i*-ой выборке, тогда*

$$n_i = \sum_j \nu_{ij} = \nu_{i*} \text{ – объем } i\text{-ой выборки}$$
$$n = n_1 + \dots + n_k$$

Пусть $p^{(1)}, \dots, p^{(k)}$ *- известны*

$$\Xi_{n_1}^2 = \sum_{j=1}^{N} \frac{\left(\nu_{ij} - n_i p_j^{(i)}\right)^2}{n_i p_j^{(i)}}, df = N - 1$$
$$\Xi_{n_1,n_2,\dots,n_k}^2 = \sum_{i=1}^{K} \Xi_{n_i}^2, df = k(N-1)$$

Рассмотрим $p^{(1)}, \dots, p^{(k)}$ *- неизвестны*

$$df = k(N-1) - (N-1) = (N-1)(k-1)$$

Давайте подставим сюда оценку максимального правдоподобия. Она будет равна:

$$\widehat{p_j} = \frac{\nu_{1j} + \dots + \nu_{kj}}{n} = \frac{\nu_{*j}}{n}$$
$$\textit{так как } L(\dots) = p_1^{(\nu_{*1})} \dots p_k^{(\nu_{*k})}, \quad n = \sum_j \nu_{*j}$$

Теорема. *(Случай 2×2) Рассмотрим $m = k = 2$*

$$Z_{n_1,n_2} = \left(\frac{v_{11}}{n_1} - \frac{v_{21}}{n_2}\right) \sqrt{\frac{nn_1n_2}{v_{*1}v_{*2}}}$$
$$\textit{где } p_1 = \frac{v_{11}}{n_1}, p_2 = \frac{v_{21}}{n_2}$$

Утверждается, что

$$Z_{n_1,n_2} = \chi_{n_1,n_2} \longrightarrow N(0,1)$$

Тогда мы можем сформулировать следующий тест:

- $H_0 : p_1 = p_2$
- $H_1 :$
  - $p \neq p_2[-q_{1-\alpha/2}; q_{1-\alpha/2}]$
  - $p_1 \geqslant p_2(-\infty; q_{1-\alpha}]$
  - $p_1 \leqslant p_2[-q_{1-\alpha}; \infty)$

6.3.4 Критерий независимости хи-квадрат

Теорема 6.5. *(Критерий независимости хи-квадрат)*

$$X_1, \dots, X_n : \{1, 2, \dots, N\}$$
$$Y_1, \dots, Y_n : \{1, 2, \dots, M\}$$

Пусть ν_{ij} - количество пар, в которых первая компонента равна i, а вторая j. Это можно представить в виде таблицы сопряженности:

X \ Y	1	· · ·	k	Σ
1	v_{11}	· · ·	v_{1k}	v_{1*}
·	·		·	·
·	·		·	·
·	·		·	·
s	v_{s1}	· · ·	v_{sk}	v_{s*}
Σ	v_{*1}	· · ·	v_{*k}	n

Просуммируем по каждому столбцу и по каждой строчке. Пусть

$$p_{ij} = P(X = i, Y = j)$$
$$p_{xi} = P(X = i)$$
$$p_{yj} = P(Y = j)$$

И введем следующий критерий:

- $H_0 : p_{ij} = p_{xi} p_{yj} \forall i, j$
- $H_1 : \neg H_0$

$$\Xi^2 = \sum_{1 \le i \le N, 1 \le j \le M} \frac{(\nu_{ij} - p_{ij} n)^2}{n p_{ij}}$$

При известных $p_{i,j} : df = MN - 1$: Если же они неизвестны, то

$$df = MN - 1 - (N - 1) - (M - 1) =$$
$$MN - N - M + 1 = N(M - 1) - (M - 1) = (M - 1)(N - 1)$$

Давайте подставим сюда оценку максимального правдоподобия. Она будет равна:

$$\widehat{p_{Xi}} = \frac{\nu_{i*}}{n}, \quad \widehat{p_{Yj}} = \frac{\nu_{j*}}{n}$$

Теорема. *(Случай 2×2) Рассмотрим $m = k = 2$*
Тогда у χ^2 будет всего одна степень свободы:

$$Z_n^2 = \chi^2$$
$$Z_n = \left(\frac{v_{11}}{v_{1*}} - \frac{v_{21}}{v_{2*}} \right) \sqrt{\frac{n v_{1*} v_{2*}}{v_{*1} v_{*2}}} = \sqrt{n} g_n,$$

где g_n – выборочный коэффициент корреляции Пирсона

Пример. *Рассмотрим следующие величины:*

- $X \sim Bern(px)$; *A – успех;* $\bar{A}$ *– неудача*
- $Y \sim Bern(py)$; *B – успех;* $\bar{B}$ *– неудача*

$$\rho(XY) = \frac{cov(X, Y)}{\sqrt{P(A)P(\bar{A})P(B)P(\bar{B})}} =$$
$$cov(X, Y) = EXY - EX \cdot EY = \frac{P(A, B) \cdot P(B) - P(A) \cdot P(B)}{\sqrt{\cdots}} =$$
$$\frac{P(B)}{\sqrt{\cdots}}(P(A|B) - P(A)) = \frac{P(B)}{\sqrt{\cdots}}(P(A|B) - P(A|B) \cdot P(B) - P(A|\hat{B}) \cdot P(\hat{B}) =$$
$$\frac{P(B)}{\sqrt{\cdots}}(P|A|B) \cdot P(\hat{B}) - P(A\ \hat{B}) \cdot P(\hat{B}) = \frac{P(B) \cdot P(\hat{B})}{\sqrt{\cdots}}(P(A|B) - P(A|\hat{B}))$$

$$H_0 : P(A|B) = P(A|\hat{B})$$
$$H_1 : P(A|B) \neq P(A|\hat{B})$$
$$P(A|B) \geqslant P(A|\hat{B})$$
$$P(A|B) \leqslant P(A|\hat{B})$$

6.3.5 Критерий квантилей и знаков

Теорема 6.6. *(Критерий квантилей)*

$$H_0 : \begin{cases} F(q_1) = \alpha_1 \\ F(q_2) = \alpha_2 \\ \vdots \\ F(q_N) = \alpha_N \end{cases}, \quad \text{где } \alpha_0 = 0 \le \alpha_1 < \dots < \alpha_N \le 1 = \alpha_{N+1}$$

$$H_1 : \neg H_0$$

Подумаем, как подогнать под критерий хи-квадрат. Критерий согласия Пирсона работает с дискретными распределениями, значит, давайте дискретизировать:

$$q_0 = \inf suppP < q_1 < \dots < q_N < \sup suppP = q_{N+1}$$

Введем интервалы:

$$\begin{cases} \Delta_1 = [q_0, q_1) \\ \vdots \\ \Delta_{N+1} = [q_N, q_{N+1}) \end{cases} = \begin{cases} P(\Delta_1) = \alpha_1 - \alpha_0 \\ \vdots \\ P(\Delta_N) = \alpha_{N+1} - \alpha_N \end{cases}$$

Далее уже по критерию согласия Пирсона.

Теорема 6.7. *(Критерий знаков)*
Рассмотрим особый случай

$$H_0 : F(q) = \frac{1}{2}$$

К этому критерию обычно сводят следующую задачу:
Пусть даны $(X_1, Y_1)^T, \dots, (X_n, , Y_n)^T$, мы хотим проверить что:

- *выборки независимы*
- *распределения одинаковы*

Доказательство. Если это верно, то $F(x, y) = F(x) \cdot F(y)$.
И как свести это к критерию знаков? А пусть $U = X - Y \Rightarrow \text{med } U = 0$, если X, Y – i.i.d.
ч.т.д.

Пусть ν_1 - количество элементов новой выборки $>$ med, тогда статистика критерия будет равна

$$Z_n = \frac{2}{\sqrt{n}} \left(\nu_1 - \frac{n}{2} \right) \to N(0, 1)$$

6.4 Ранговые критерии

Определение 6.1. *Пусть $X_1, \dots, X_n$ – выборка, тогда*
$R(X_k)$ – номер X_k в вариационном ряде.

6.4.1 Критерий Манна-Уитни-Вилконсона

Пусть даны две независимые выборки:

- $X_1, \ldots, X_n$
- $Y_1, \ldots, Y_m$

Давайте объединим их в одну и рассмотрим $(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$:

Определение. $T = R_1 + R_2 + \ldots + R_n$ *– статистика Вилконсона,*
где R_i *- ранг* X_i *в объединенной выборке.*

Теперь введем вспомогательную индикаторную функцию: $Z_{rs} = \mathbb{1}(X_r < Y_s)$

Определение. $U = \sum\limits_{r=1}^{n} \sum\limits_{s=1}^{m} Z_{rs}$ *- статистика Манна-Уитни*

Эти статистики имеют следующее соотношению между собой:

$$T + U = mn + \frac{n(n+1)}{2}$$

Теперь мы хотим проверить, что распределение X совпадает с Y. Рассмотрим следующий критерий:

- $EU = mnE\mathbb{1}(X < Y) = mnP(X < Y) = \frac{1}{2}$ - при условии H_0
- $H_0 : P(X < Y) = \frac{1}{2}$
- $H_1 :\neq \frac{1}{2}, > \frac{1}{2}, < \frac{1}{2}$

$$U \sim N\left(\frac{mn}{2}, \frac{nm(m+n+1)}{12}\right)$$

6.4.2 Коэффициенты корреляции Пирсона, Спирмена, Кендалла. Статистические тесты, основанные на них

Пусть даны выборки $(X_1, Y_1)^T, \ldots, (X_n, , Y_n)^T$.
Хотим проверить независимость

- R_i - ранг X_i (в своей выборке)
- S_i - ранг Y_i (в своей выборке)

Определение 6.2. *(Коэффициент корреляции Спирмена)*

$$\rho = \frac{12}{n(n^2-1)} \sum \left(R_i - \frac{n+1}{2}\right)\left(S_i - \frac{n+1}{2}\right)$$

где ρ *- выборочный коэффициент корреляции между* R_i *и* S_i

Сформулируем соответствующий тест:

- $H_0 : \rho = 0$

- $H_1 : \rho \neq 0, \rho > 0, \rho < 0$, так как $\sqrt{n}\rho \to N(0,1)$

Определение 6.3. *(Коэффициент корреляции Кендалла)*

$$\tau = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} {}_{j=i+1}^{n} sign(T_j - T_i)$$

H_0 верно $\Rightarrow E\tau = 0, \operatorname{Var}\tau = \frac{2(2n+5)}{9n(n-1)}$ и $\tau \approx N(0, \frac{4}{9n})$

- $H_0 : \tau = 0$
- $H_1 :> 0, < 0, \neq 0$

6.5 Критерии отношения правдоподобия

В данном параграфе мы рассмотрим критерии, в которых тестовая статистика есть отношение правдоподобия. Мы начнем со случая простых гипотез, в рамках которой можно явно построить оптимальный критерий, максимизирующий мощность при заданном уровне значимости.

6.5.1 Простые гипотезы

Пусть имеется выборка $X_1, \ldots, X_n \sim F$. Рассмотрим две простые гипотезы $H_0 : F = F_0$ и $H_1 : F = F_1$, L_0, L_1 – функции правдоподобия для нулевой и альтернативной гипотез, $l(X) = L_1(X)/L_0(X)$ – отношение правдоподобия.

Сначала будем полагать, что F_0 и F_1 соответствуют абсолютно непрерывным распределениям и дополнительно предположим, что носитель обоих распределений есть $\mathbb{R}$. В этом случае критерий отношения правдоподобия выглядит следующим образом

$$\delta(X, H_0, H_1, c) = \begin{cases} 0, & l(X) < c, \\ 1, & l(X) \geqslant c, \end{cases}$$

где $c > 0$, 0 означает, что мы приняли H_0, 1 – опровержение H_0 в пользу H_1.

Зададим функцию

$$\psi(c) = \mathrm{P}(l(X) \geqslant c | H_0) = \int\limits_{X: l(X) \geqslant c} L_0(X) dX, \quad c \geqslant 0.$$

Заметим, что данная функция монотонно убывает по c и помимо этого

$$1 \geqslant \mathrm{P}(l(X) \geqslant c | H_1) = \int\limits_{X: L_1(X) \geqslant c L_0(X)} L_1(X) dx \geqslant c \int\limits_{X: L_1(X) \geqslant c L_0(X)} L_0(X) dx = c\psi(c),$$

откуда $\psi(c) \leqslant 1/c$. Поэтому $\psi(c) \to 0$ при $c \to +\infty$. Кроме того, $\psi(0) = 1$. Предположим, что для любого $\alpha \in (0,1)$ существует c_α, для которого $\psi(c_\alpha) = \alpha$.

Замечание. *Сказанное имеет место, если ψ – непрерывна, и в этом случае $\mathrm{P}(l(X) = c | H_0) = 0$, так как распределение случайной величины $l(X)$ при условии H_0 является непрерывным.*

Теперь мы сформулируем и докажем лемму Неймана-Пирсона

Теорема. ***Лемма Неймана-Пирсона.***

Пусть выполнены сформированные выше предположения и функция ψ является непрерывной. Тогда критерий максимального правдоподобия при $c = c_\alpha$ является наиболее мощным среди всех критериев с вероятностью ошибки первого рода, равной α.

Доказательство. Прежде всего отметим, что по построению критерий отношения правдоподобия при $c = c_\alpha$ имеет вероятность ошибки первого рода, равную α: ведь $\psi(c)$ есть в точности вероятность ошибки первого рода критерия отношения правдоподобия для данного c.

Рассмотрим другой критерий со тестовой статистикой G, у которого вероятность ошибки первого рода равняется α. Обозначим за $T_0(\alpha)$ и $T_1(\alpha)$ его области принятия и опровержения, а за $T_0^*(\alpha)$ и $T_1^*(\alpha)$ – соответствующие области для критерия отношения максимального правдоподобия при $c = c_\alpha$. Заметим, что

$$\begin{gathered}
\mathrm{P}(l(X) \in T_1^*(\alpha) \setminus \{l(X) \in T_1^*(\alpha), G(X) \in T_1(\alpha)\}|H_0) \\
= \mathrm{P}(l(X) \in T_1^*(\alpha)|H_0) - \mathrm{P}(l(X) \in T_1^*(\alpha), G(X) \in T_1(\alpha)|H_0) \\
= \alpha - \mathrm{P}(l(X) \in T_1^*(\alpha), G(X) \in T_1(\alpha)|H_0) \\
= \mathrm{P}(G(X) \in T_1(\alpha)|H_0) - \mathrm{P}(l(X) \in T_1^*(\alpha), G(X) \in T_1(\alpha)|H_0) \\
= \mathrm{P}(G(X) \in T_1(\alpha) \setminus \{l(X) \in T_1^*(\alpha), G(X) \in T_1(\alpha)\}|H_0),
\end{gathered}$$

так как вероятность ошибки первого рода для обоих тестов равняется α. В силу полученного равенства того, что $L_1(X) \geqslant c_\alpha L_0(X)$, если $l(X) \in T_1^*$ и $c_\alpha L_0(X) > L_1(X)$ в противном случае, то мы получаем

$$\begin{gathered}
\mathrm{P}(l(X) \in T_1^*(\alpha) \setminus \{l(X) \in T_1^*(\alpha), G(X) \in T_1(\alpha)\}|H_1) \\
\geqslant c_\alpha \,\mathrm{P}(l(X) \in T_1^*(\alpha) \setminus \{l(X) \in T_1^*(\alpha), G(X) \in T_1(\alpha)\}|H_0) \\
= c_\alpha \,\mathrm{P}(G(X) \in T_1(\alpha) \setminus \{l(X) \in T_1^*(\alpha), G(X) \in T_1(\alpha)\}|H_0) \\
> \mathrm{P}(G(X) \in T_1(\alpha) \setminus \{l(X) \in T_1^*(\alpha), G(X) \in T_1(\alpha)\}|H_1).
\end{gathered}$$

Первый знак неравенства объясняется тем, что для события $\big\{l(X) \in T_1^*(\alpha) \setminus \{l(X) \in T_1^*(\alpha), G(X) \in T_1(\alpha)\}\big\}$ верно $l(X) \in T_1^*$, то есть $L_1(X) \geqslant c_\alpha L_0(X)$, а второй знак неравенства аналогично: для события $\big\{G(X) \in T_1(\alpha) \setminus \{l(X) \in T_1^*(\alpha), G(X) \in T_1(\alpha)\}\big\}$ необходима истинность $\{l(X) \in T_0^*(\alpha)\}$, то есть $c_\alpha L_0(X) > L_1(X)$. Прибавляя к обеим частям полученного неравенства $\mathrm{P}(l(X) \in T_1^*(\alpha), G(X) \in T_1(\alpha)\}|H_1)$ мы заключаем, что мощность критерия отношения правдоподобия при $c = c_\alpha$ действительно максимальна при заданных ограничениях, то есть теорема доказана. *ч.т.д.*

Теперь рассмотрим дискретный случай. В этом случае функция $l(X)$ может принимать не более чем счетное количество значений. Пусть l_k есть данные значения, причем они упорядочены по возрастанию. Тогда для данного $\alpha \in (0, 1)$ мы можем указать номер $k(\alpha)$, для которого

$$\psi(l_{k+1}) = \sum_{X: l(X) \geqslant l_{k(\alpha)+1}} L_0(X) < \alpha \leqslant \sum_{X: l(X) \geqslant l_{k(\alpha)}} L_0(X) = \psi(l_k)$$

Если правое неравенство на самом деле есть равенство, то есть $\psi(l_k) = \alpha$, то мы можем применить рассуждения из только что доказанной леммы и тем самым показать, что в данной ситуации мы критерий отношения правдоподобия является наиболее мощным среди всех критериев с уровнем значимости α. Обычно это не так. Введем обозначения:

$$q_j = \sum_{X: l(X) \geqslant l_{k(\alpha)}+1} L_j(X), \quad j \in \{0, 1\},$$
$$p_j = \sum_{X: l(X) = l_{k(\alpha)}} L_j(X), \quad j \in \{0, 1\}.$$

Построим *рандомизированный* критерий, чтобы все-таки вероятность ошибки первого рода в точности равнялась α. Для этого введем функцию

$$\phi^*(X) = \begin{cases} 1, \quad l(X) > l_{k(\alpha)}, \\ \frac{\alpha - q_0}{p_0}, \quad l(X) = l(X) = l_{k(\alpha)}, \\ 0, \quad l(X) < l_{k(\alpha)}. \end{cases}$$

Для данной выборки $X = x$ значение функции $\phi(x)$ есть в точности вероятность опровергнуть нулевую гипотезу. То есть, если $\phi(x) = 1$ или $\phi(x) = 0$ мы точно опровергаем или принимаем гипотезу H_0 соответственно, а, если $\phi(x) = \frac{\alpha - q_0}{p_0}$, то мы разыгрываем бернулиевскую случайную величину с указанной вероятностью «успеха» – в данном случае опровержения нулевой гипотезы. Убедимся, что вероятность ошибки первого рода есть α:

$$\begin{aligned} \mathrm{P}(reject|H_0) = \mathrm{E}(\phi^*(X)|H_0) = \mathrm{P}(l(X) > l_{k(\alpha)}|H_0) + \frac{\alpha - q_0}{p_0}\mathrm{P}(l(X) = l_{k(\alpha)}|H_0) \\ = q_0 + \frac{\alpha - q_0}{p_0} p_0 = \alpha. \end{aligned}$$

Покажем, что построенный рандомизированный критерий с функцией ϕ_* имеет наибольшую мощность среди всех (рандомизированных) критериев с функций ϕ и вероятностью ошибки первого рода, равной α. Рассмотрим разность $\phi^*(X) - \phi(X)$. Если эта разность строго больше нуля, тогда $\phi^*(X) > \phi(X) \geqslant 0$, то есть $l(X) \geqslant l_{k(\alpha)}$ и $L_1(X) \geqslant l_{k(\alpha)} L_0(X)$. Если же данная разность строго отрицательна, значит $\phi^*(X) < \phi(X) \leqslant 1$, откуда $L_1(X) \leqslant l_{k(\alpha)} L_0(X)$. Тогда можно заключить, что

$$\sum_X \big(\phi^*(X) - \phi(X)\big)\big(L_1(X) - l_{k(\alpha)} L_0(X)\big) = \sum_{X: \phi^*(X) > \phi(X)} \ldots + \sum_{X: \phi^*(X) < \phi(X)} \ldots \geqslant 0.$$

Следовательно, разность мощностей можно оценить как

$$\begin{aligned} \mathrm{E}(\phi^*(X)|H_1) - \mathrm{E}(\phi(X)|H_1) = \sum_X \big(\phi^*(X) - \phi(X)\big) L_1(X) \\ \geqslant l_{k(\alpha)} \sum_X \big(\phi^*(X) - \phi(X)\big) L_0(X) \\ = l_{k(\alpha)} \big(\mathrm{E}(\phi^*(X)|H_0) - \mathrm{E}(\phi(X)|H_0)\big) = l_{k(\alpha)}(\alpha - \alpha) = 0. \end{aligned}$$

Замечание. *В непрерывном случае мы потребовали, чтобы функция $\psi(c)$ была непрерывной. Однако, если это все-таки не так, мы можем по приведенной схеме построить рандомизированный критерий.*

Глава 7

Линейные статистические модели

7.1 Модель линейной регрессии

n - число наблюдений, k - количество переменных (непрерывные); $k < n$;
$Y = Xb + E$;
предполагаем, что

- наблюдения независимы или хотя бы $Cov[\xi_i, \xi_j] = 0 i \neq j$;
- $E[\xi]$;
- $Var[\xi_1] = \dots = Var[\xi_n] = \delta^2$ (гомоскедастичность);

δ^2 - остаточная дисперсия; $Var[\xi] = \delta^2 E_n$;
цель: "оценить"b и δ^2
$A := X^\tau X \in M_{kxk}; A \geqslant 0; A > 0 \Leftrightarrow A = k$; **Метод наименьших квадратов:** $S^2(b) := (Y - Xb)^\tau (Y - Xb) = \sum_{i=1}^n (Y_i - X_{[i,*]} b)^2$;
Оценка наименьших квадратов (ОНК): $\widehat{b} = argmin_b S^2(b)$.

Утверждение 1. $A > 0 \Rightarrow \widehat{b} = A^{-1} X^\tau Y$

Доказательство: $b = \widehat{b} + \delta$;
$S^2(b) = (Y - X\widehat{b} - X\delta)^\tau (Y - X\widehat{b} - X\delta) = S^2(\widehat{b}) + \delta^\tau X^\tau X \delta - (Y - X\widehat{b})^\tau X\delta - \delta^\tau X^\tau (Y - X\widehat{b}) = S^2(\widehat{b}) + \delta^\tau X^\tau X\delta \geqslant S^2(\widehat{b})$;
равенство достигается тогда и только тогда, когда $\delta = 0$
$t = Tb$- линейная функция от b;$\widehat{t} = T\widehat{b}$- ОНК для t.

Утверждение 2. *В рамках тех предположений, которые мы описали в начале:*

- *ОНК - несмещенная;*
- *ОНК обладает наименьшей дисперсией в классе линейных несмещенных оценок:* $Var[\widehat{b}] = \delta^2 T A^{-1} T^{tau}$ *;*

Доказательство:

- $E[\widehat{t}] = T\dot{E}[\widehat{b}] = TA^{-1}X^{-1}Xb = Tb = t$

- $l = LY$ - линейная несмещенная оценка t
  $El = Tb$
  $LEY = LXb \Rightarrow T = LX$
  $VarLY = LVarYL^\tau = \delta^2 LEL^\tau = \delta^2 LL^\tau$
  $LL^\tau = (TA^{-1}X^\tau)(TA^{-1}X^\tau)^\tau + (L - TA^{-1}X^\tau)(L - TA^{-1}X^\tau)^\tau \geqslant 0$
  $L = TA^{-1}X^\tau$-оптимально для $Var\widehat{b} = \theta^2 A^{-1}$ - следствие

$E[S^2(b)] = E[\epsilon^\tau \epsilon] = n\delta^2; E[(\widehat{b} - b)^\tau A(\widehat{b} - b)] = \sum_{i,j} a_{ij} E[(\widehat{b}_i - b_i)^\tau (\widehat{b}_j - b_j)] = \sum_{i,j} a_{ij} cov[\widehat{b}_i, \widehat{b}_j] = \sum_{ij} a_{ij} a_{ij}^{-1} \delta^2 = k\delta^2$
$n\delta^2 = E[S^2(\widehat{b})] + k\delta^2$, т.к $S^2(b) = S^2(\widehat{b}) + (\widehat{b} - b)^\tau A(\widehat{b} - b)$;
несмещенная оценка δ^2:

$$\hat{\sigma^2} = \frac{s^2(\hat{b})}{n-k} = \frac{(X\hat{b} - Y)^T (X\hat{b} - Y)}{n-k} = \frac{\epsilon^T \epsilon}{n-k}$$

$$Y - X\hat{b} = Y - A^{-1}X^T Y = (E - A^{-1}X^T)Y$$

$$B = B^T B, \quad B = n - k \sim \hat{\sigma^2} = \frac{Y^T BY}{n-k}$$

Условная оптимальная несмещенная оценка (ОНК) имеет вид:

$$\hat{b_T} = argmin_{Tb=t_0} S^2(b); \quad S_T^2 = S^2(\hat{b_T}), \quad T \in M_{m,k}, \quad T = m, \quad m \leq k$$

Определение 7.1. $\hat{b_T} = \hat{b} - A^{-1}T^T D^{-1}(T\hat{b} - t_0)$, *где* $D = TA^{-1}T^T$;

$$S^2(b) = S^2(\hat{b_T} - b)^T A(\hat{b_T} - b) \geq S^2(\hat{b_T});$$

$$s_T^2 = s^2(\hat{b}) + (T\hat{b} - t_0)^T D^{-1}(T\hat{b} - t_0);$$

$Q_T \sim \xi^2(m)$, если $\epsilon \sim N(0, \sigma^2 E_n) \Leftrightarrow Y \sim N(Xb, \sigma^2 E_n)$

$$L(Y, b, \sigma^2) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - X_{i*}b)^2\right] = \frac{1}{(\sqrt{2\pi\sigma^2})^n} exp\left[-\frac{S^2(b)}{2\sigma^2}\right]$$

$$L(Y, b, \sigma^2) \xrightarrow[\sigma^2 - fix]{s} up \Leftrightarrow S^2(b) \xrightarrow[\sigma^2 - fix]{i} nf$$

Следовательно, ОНК – оптимальная в классе всех несмещенных оценок $\Rightarrow \hat{\sigma^2} = \frac{S^2(\hat{b})}{n}$ – смещенная оценка σ^2.

Теорема 7.1. *Основная теорема о регрессии.*
$\epsilon \sim N(0, \sigma^2)$, *тогда:*

- $\hat{b}$, $S^2(\hat{b})$ – *независимы;*
- $S^2(b) - S^2(\hat{b})$ – *независимы;*
- $\hat{b} \sim N(b, \sigma^2 A^{-1})$;
- $\frac{S^2(\hat{b})}{\sigma^2} \sim \xi^2(n-k)$;

- $\frac{S^2(\hat{b})-S^2(\hat{b})}{\sigma^2} \sim \xi^2(k)$;

Доказательство. $\epsilon' = \frac{\epsilon}{\sigma}$; $\hat{b} = A^{-1}X^T(Xb + \sigma\epsilon') = b + \sigma A^{-1}X^T\epsilon'$; $\frac{S^2(\hat{b})}{\sigma^2} = \epsilon'^T B\epsilon'$; $BA^{-1}X^T = (E - XA^{-1}X^T)A^{-1}X^T$; $A^{-1}X^T(E - XA^{-1}X^T) = A^{-1}X^T - A^{-1}X^TXA^{-1}X^T = 0 \Rightarrow \hat{b}$ и $S^2(\hat{b})$ – независимы; $S^2(b) - S^2(\hat{b}) = (\hat{b} - b)^T A(\hat{b} - b)$ и $S^2(\hat{b})$ – независимы. *ч.т.д.*

Следствие:

- $\frac{\hat{b_i}-b_i}{\sigma\sqrt{a_{ii}^{-1}}} \sim N(0,1)$;
- $\sqrt{n-k}\frac{\hat{b}_i-b_i}{\sqrt{a_{ii}^{-1}S^2(\hat{b})}} \sim T(n-k)$ – можем построить доверительные интервалы для b или проверить $H_0 : b_i = 0$ (t-критерий);
- $\frac{n-k}{k} \cdot \frac{S^2(b)-S^2(\hat{b})}{S^2(\hat{b})} \sim F(k, n-k)$;
- с помощью $\frac{S^2(\hat{b})}{\sigma^2} \sim \xi^2(n-k)$ можем построить доверительный интервал для σ^2;
- также можем строить доверительные интервалы для Y из области выборки, а также из тестовой X^* – набор тестовых значений; $Y^* = X^*b + \epsilon^*$; предположим, что $Var[\epsilon^*] = \sigma^2 E$ и ϵ^*, ϵ – независимы, $\epsilon^* \sim N(0, \sigma^2 E)$; $\hat{b} \sim N(b, \sigma^2 A^{-1}) \Rightarrow Y^* \sim N(X^*b, X^*\sigma^2 A^{-1}(X^*)^T + \sigma^2 E) \sim N(X^*b, \sigma^2[X^*A^{-1}(X^*)^T + E])$; $Y_i^* \sim N(X_{i*}^*b, \sigma^2[...]_{ii})$; $\sqrt{n-k}\frac{x_{i*}^*\hat{b}-Y_i^*}{\sqrt{S^2(\hat{b})}\sqrt{(...)_{ii}}} \sim T(n-k)$

.

7.1.1 F-критерий

$H_0 : Tb = b_0$, где $T \in M_{m,k}$, где $T = m, m \leqslant k$; если $m = k$, то очень часто смотрят $b = 0$;

$H_1 : Tb \neq b_0$;

$\widehat{b}_T$- ОНК при условии $Tb = b_0$; $S_T^2 = S^2(\widehat{b}_T)$; $S_T^2 = S^2(\widehat{b}) + Q_T, Q_T = (T\widehat{b} - b_0)^T D^{-1}(T\widehat{b} - b_0)$;

по теореме о регрессии: $\frac{S^2(\widehat{b})}{\sigma^2} \sim X^2(n-k), T\widehat{b} \sim N(b_o, \sigma^2 D) \Rightarrow \frac{Q_T}{\sigma^2} \sim X^2(m), Q_T$- не зависит от $S^2(\widehat{b})$;

$F := \frac{n-k}{m} \cdot \frac{Q_T}{S^2(\widehat{b})} \sim F(m, n-k)$, если H_0 - верно; $F = \frac{n-k}{m} \cdot \frac{S_T^2 - S^2(\widehat{b})}{S^2(\widehat{b})}$ -статистика F-критерия;

частный случай $H_0 : b = 0$

$F = \frac{n-k}{k} \cdot \frac{S^2(0)-S^2(\widehat{b})}{S^2(\widehat{b})}$;

$Tb - b_0$ "большое", если H_0 - неверно;

$B - const, B \in M_{n,n}, Y \in \mathbb{R}^n$, компоненты Y независимы;

$E[Y^T BY] = \sum_{i,j} b_{ij} E[Y_i Y_j] = \sum_{i,j} b_{ij}(E[Y_i]E[Y_j] + Cov[Y_i, Y_j]) = \sum_i b_{ii} Var[Y_i] + \sum_{i,j} b_{ij} E[Y_i]E[Y_j] = T_r(BVar[Y]) + (E[Y])^T BE[Y]$;

$\frac{1}{m}E[Q_T] = \frac{1}{m}(m\sigma^2 + \frac{1}{m}(Tb - b_0)^T D^{-1}(Tb - b_0)) \geqslant \sigma^2$;

$E[\frac{S^2(\widehat{b})}{n-k}] = \sigma^2$;

следовательно, при "больших" значениях статистики F H_0 скорее неверно $\Rightarrow [q_{1-\alpha}; +\infty)$ - критическая область.

$R_{k-1} = \frac{\sum_{i=1}^n (Y_i - \overline{Y})(\widehat{Y}_i - \overline{\widehat{Y}})}{\sqrt{\sum_i (Y_i - \overline{Y})^2 \sum_i (\widehat{Y}_i - \overline{\widehat{Y}})^2}}$; $\widehat{Y} = X\widehat{b}$ - линейный кэффициент корреляциuё

R_{k-1}^2-коэффициент детерминации

$0 \leqslant R^2 \leqslant 1$

$S^2(\widehat{b}) = (1 - R_{k-1}^2) \sum i(\widehat{Y}_i - \overline{Y})^2$.

7.2 Однофакторный дисперсионный анализ

Модель: $y_{ij} = \mu_i + \xi_{ij}; 1 \leqslant i \leqslant I, 1 \leqslant j \leqslant J_i; n = \sum_{i=1}^{I} J_i;$

I - количество уровней фактора; μ_i-среднее влияние для фактора, у которого i-й уровень;

ξ_{ij}-независимы, $\xi_{ij} \sim N(0, \sigma^2)$;

$H_0 : \mu_1 = \mu_2 = \cdots = \mu_I;$

$H_1 : \exists i_1, i_2 : \mu_{i1} \neq \mu_{i2};$

$F = \frac{n-I}{I-1} \cdot \frac{S_B^2}{S_W^2}$ - статистика F-критерия; $S_B^2 := \sum_i J_i(\overline{y_{i*}} - \overline{y})^2, \overline{y_{i*}} = \frac{1}{J_i}\sum_{i=1}^{Ji} y_{ij}; S_w^2 := \sum_{i,j}(y_{ij} - \overline{y_i*})^2;$

$\frac{S_B^2}{I-1}$- межгрупповая дисперсия; $\frac{S_w^2}{n-1}$- внутригрупповая дисперсия

7.3 Двухфакторный дисперсионный анализ

Модель : $y_ij = \mu + \alpha_i + b_j + \xi_{ij}; 1 \leqslant i \leqslant I, 1 \leqslant j \leqslant J;$

μ- общее среднее; α_i-среднее влияние фактора 1 на уровне $i; b_j$- среднее влияние фактора 2 на уровне j;

факторы независимы и $\sum_i \alpha_i = \sum_j b_j = 0;$

ξ_{ij}-независимы, $\xi_{ij} \sim N(0, \sigma^2)$;

Самые частые гипотезы:

$H_A : \alpha_i = 0 \forall_i; H_B : b_j = 0 \forall_j; H : \alpha_i = 0, b_j = 0 \forall_{i,j};$

$S_{all}^2 := \sum_{i,j}(y_{ij} - \overline{y_{i*}} - \overline{y_{*j}} - \overline{y})^2;\ F = \frac{(I-1)(J-1)}{I+J-2} \cdot \frac{S_A^2 + S_B^2}{S_{all}^2};$

$S_A^2 := J\sum_i(\overline{y_{i*}} - \overline{y})^2; F_A = \frac{(I-1)(J-1)}{I-1} \cdot \frac{S_A^2}{S_{all}^2};$

$S_B^2 := I\sum_j(\overline{y_{j*}} - \overline{y})^2; F_B = \frac{(I-1)(J-1)}{J-1} \cdot \frac{S_B^2}{S_{all}^2};$

7.4 Ковариационный анализ

Модель: $y_{ij} = \beta_i + \gamma \cdot Z_{ij} + \xi_{ij}; 1 \leqslant i \leqslant I, 1 \leqslant j \leqslant J$

β_1-среднее влияние фактора на уровне i,

ξ_{ij}- независимы, $\xi_{ij} \sim N(0, \sigma^2)$;

Самые частые гипотезы:

$H_B : \beta_1 = \beta_2 = ... = \beta_I; H_\gamma : \gamma = 0;$

Проверка H_γ:

$\widehat{\beta}, \widehat{\gamma}$- ОНК; по теореме о регрессии: $\sqrt{n-I-1}\,\frac{\widehat{\gamma}-\gamma}{\sqrt{a_{ii}^{-1} S^2(\widehat{\beta},\widehat{\gamma})}} \sim T(n-I-1)$; строим ДИ и проверяем, что он содержит 0;

Проверка H_β:

$F = \frac{n-I-1}{I-1} \cdot \frac{\sum_{i,j}(y_{i,j}^2) - n\overline{y}^2 - P}{P}$

$P = \sum_{ij}(y_{ij}^2) - n\overline{y} - \frac{(\sum_{i,j} Z_{ij}y_{ij} - n\overline{z})^2}{\sum_{i,j}(Z_{ij}^2) - n\overline{z}^2}$

7.5 Обобщенные линейные модели

$Y_i = g(X_i^T b)$ - обобщенные линейные модели

7.5.1 Логистическая регрессия (бинарная классификация)

Модель: $y_i \approx \frac{1}{1+e^{-X_i^T b}}$;

$0 \leqslant \frac{1}{1+e^{-t}} \leqslant 1$, при $t = 0$ равна $\frac{1}{2}$; для бинарной классификации просто выбирается порог;

$y_i \sim Bern(\frac{1}{1+e^{-X_i^T b}}), y_i$ - независимы;

$L(Y, b) = (\frac{1}{1+e^{-X_i^T b}})^{\sum y_i}((\frac{e^{-X_i^T b}}{1+e^{-X_i^T b}}))^{n-\sum y_i} - \sum y_i ln(\frac{1}{1+e^{-X_i^T b}}) + (1-y_i)ln(\frac{e^{-X_i^T b}}{1+e^{-X_i^T b}}) \to inf$

$lnL(Y, b) = \sum \left[y_i ln(\frac{1}{1+e^{-X_i^T b}}) + (1-y_i)ln(\frac{e^{-X_i^T b}}{1+e^{-X_i^T b}})\right]$

$H_0 : b = 0$;

используем критерий отношения правдоподобия: $\lambda_n = \frac{L(Y,0)}{sup_{b\in\mathbb{R}^k} L(Y,b)}$;

если H_0 верно, то $-2ln\lambda_n \to U \sim X^2(k)$.

Предметный указатель

Список литературы

[1] S. Kulback and R. Leibler. “On Information and Sufficiency”. English. In: *The Annals of Mathematical Statistics* 22 (1951), pp. 79–86.

[2] S. Kullback. *Information Theory and Statistics*. English. Wiley, 1959.

[3] А.А. Боровков. *Теория вероятностей*. Изд. 5-е, существенно перераб. и доп. М.: URSS, 2009. ISBN: 978-5-397-00582-1.

[4] Н.Я. Виленкин. *Комбинаторика*. М.: Наука, 1969.

[5] Б.В. Гнеденко и А.Н. Колмогоров. *Предельные распределения для сумм независимых случайных величин*. М. и Л.: Гос. изд-во техн.-теорет. лит., 1949.

[6] М. Лоэв. *Теория вероятностей*. Пер. Б.А. Севастьянов. М.: Изд-во иностр. лит., 1962.

[7] Б.М. Макаров и А.Н. Подкорытов. *Лекции по вещественному анализу*. Санкт-Петербург: БХВ-Петербург, 2011. ISBN: 978-5-9775-0631-1.

[8] И.П. Натансон. *Теория функций вещественной переменной*. Изд. 5-е, стер. СПб.: Лань, 2008. ISBN: 978-5-8114-0136-9.

[9] В. Феллер. *Введение в теорию вероятностей и ее приложения*. Пер. Ю.В. Прохоров. Изд. 2-е. Т. 1. М.: URSS, 2009. ISBN: 978-5-397-01035-1.

[10] П. Халмош. *Теория меры*. Пер. Д.А. Васильков. М.: Факториал Пресс, 2003. ISBN: 5-88688-065-8.

[11] А.Н. Ширяев. *Вероятность*. Изд. 6-е, испр. М.: Изд-во МЦНМО, 2016. ISBN: 978-5-4439-1093-2.