

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М. В. ЛОМОНОСОВА

МЕХАНИКО МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ

КАФЕДРА ТЕОРИИ ВЕРОЯТНОСТЕЙ

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
(ДИПЛОМНАЯ РАБОТА)
специалиста

О скорости сходимости решающего штампа

Выполнил: студент 6 курса

608 группы:

Котляров Николай Евгеньевич

Подпись студента: _____

Научный руководитель:

Замятин Андрей Андреевич

Подпись научного руководителя: _____

Содержание

1	Введение в задачу	3
2	Вспомогательные утверждения	8
3	Основные результаты	13
4	Альтернативный подход и компьютерное моделирование	20
5	Заключение	33
	Список литературы	34

1 Введение в задачу

В дипломной работе исследуются свойства сходимости разделяющей границы в алгоритме одношагового решающего дерева - решающий штамп. Данный алгоритм предоставляет разделение набора данных на две части так, чтобы в каждой из получившихся частей находились максимально похожие объекты. В данной работе исследуется предельное поведение параметров алгоритма в задаче двухклассовой классификации объектов.

В своей работе автор ставит математические проблемы, исследует способы их решения, доказывает математические теоремы, позволяющие оценить перспективы способа решения, а также предлагает альтернативный подход и проводит эксперименты на различных входных данных. Эта математическая работа даёт понимание в каком направлении стоит продолжать вести будущие исследования и предлагает хорошее приближение для решения поставленной проблемы.

Перейдем к математическому описанию проблемы.

Рассмотрим выборку $(X, l)_n = \{(X_1, l_1), (X_2, l_2), \dots, (X_n, l_n)\}$ - набор независимых, одинаково распределенных случайных величин (н.о.р.с.в.), где

$$X_i \in \mathbf{R}, l_i \in K = \{-1, 1\}$$

Пусть на пространстве $\mathcal{X} = \mathbf{R} \times K$ задана некоторая вероятностная мера P и соответствующая ей Борелевская сигма-алгебра $\mathcal{B}(\mathcal{X})$. Распределение пары (X, l) будем задавать функциями правдоподобия классов F^+, F^- и априорными вероятностями классов: π^+, π^- :

$$\begin{aligned} F^+(t) &= P(X \leq t \mid l = 1) \\ F^-(t) &= P(X \leq t \mid l = -1) \end{aligned} \quad \left| \quad t \in \mathbf{R} \right.$$
$$\pi^+ = P(l = 1)$$
$$\pi^- = P(l = -1) = 1 - \pi^+$$

Общая задача классификации заключается в том, чтобы построить некоторое отображение (классификатор):

$$f : \mathbf{R} \rightarrow K$$

которое для каждого объекта X_i , $i = 1, \dots, n$ из выборки (X_1, \dots, X_n) пытается определить его истинный класс l_i .

Для нашего алгоритма отображение f будет строиться используя одноступенчатый алгоритм решающего дерева. Таким образом отображение f имеет вид:

$$f_t^n(x) = a_n^L I\{x \leq t\} + a_n^R (1 - I\{x \leq t\})$$
$$a_n^L = \begin{cases} 1, & P_n(l = 1 \mid X \leq t) > \frac{1}{2} \\ -1, & P_n(l = 1 \mid X \leq t) \leq \frac{1}{2} \end{cases}$$
$$a_n^R = \begin{cases} 1, & P_n(l = 1 \mid X > t) > \frac{1}{2} \\ -1, & P_n(l = 1 \mid X > t) \leq \frac{1}{2} \end{cases}$$

где P_n - соответствующая эмпирическая мера, задаваемая равенством:

$$P_n(A) = \frac{1}{n} \sum_{i=1}^n I\{(X_i, l_i) \in A\}$$

где $A \in \mathcal{B}(\mathcal{X})$, параметр t - разделяющая граница (splitting point), I - индикатор множества (равен единице, если элемент попал в множество, иначе - нулю).

Таким образом мы будем искать функцию f_t^n из класса $\{f\}_t$ - семейство кусочно-постоянных функций, параметризованных параметром t .

Фактически, это означает, что выбрав оптимально параметр $t \in \mathbf{R}$, наш алгоритм будет разделять действительную ось на два подмножества (меньше t и больше t) так, чтобы объекты из выборки $(X_1, l_1), (X_2, l_2), \dots, (X_n, l_n)$ имели максимально однородное распределение истинных меток класса в каждом из двух новых получившихся подмножеств.

Обучения классификатора составляет поиск оптимальной точки t . На практике такой поиск основан на различных критериях, определяющих понятие однородности. Например, на уменьшении энтропии в каждом подмножестве или какой-нибудь другой функции, которая приводит к однородным частям.

Построения общего алгоритма решающего дерева состоит в том, что оно рекурсивно строит разбиение в каждом из двух новых получившихся подмножеств, причем построение на каждом шаге происходит жадным образом, т.е. алгоритм на каждом шаге производит оптимальный выбор параметра t_i , не задумываясь о следующих шагах.

Завершение построения алгоритма происходит, когда в каждом подмножестве объекты относятся к одному классу, или функционал в каждом листе, отвечающий за однородность выборки становится меньше какого-то заранее заданного порога, или когда достигнута максимальная глубина дерева и т.д. Все такие условия на практике задаются как гиперпараметры модели.

В данной работе будет исследоваться одношаговая модель решающего дерева. Без ограничения общности можно считать, что набор признаков $X \in \mathbf{R}$, (размерность $d = 1$), так как наш алгоритм жадным образом выберет только одну размерность для разделения выборки.

Для того, чтобы искать параметр t , мы будем исследовать точность качества классификации. Чтобы повысить точность классификатора, нужно минимизировать его ошибку. Введем ошибку классификатора по выборке $(X_1, l_1), (X_2, l_2), \dots, (X_n, l_n)$:

$$\begin{aligned} e_n(t) &= \frac{1}{n} |k : f_t^n(X_k) \neq l_k| = \frac{1}{n} |k : f_t^n(X_k) \neq l_k, X_k \leq t| + \frac{1}{n} |k : f_t^n(X_k) \neq l_k, X_k > t| = \\ &= \frac{|k : f_t^n(X_k) \neq l_k, X_k \leq t|}{|k : X_k \leq t|} \frac{|k : X_k \leq t|}{n} + \frac{|k : f_t^n(X_k) \neq l_k, X_k > t|}{|k : X_k > t|} \frac{|k : X_k > t|}{n} = \\ &= \frac{|k : a_n^L \neq l_k, X_k \leq t|}{|k : X_k \leq t|} \frac{|k : X_k \leq t|}{n} + \frac{|k : a_n^R \neq l_k, X_k > t|}{|k : X_k > t|} \frac{|k : X_k > t|}{n} \end{aligned}$$

Эмпирическая ошибка классификации $e_n(t)$ по закону больших чисел Колмогорова сходится почти наверное к своей теоретической вероятности ошибки классификации:

$$\begin{aligned} e(t) &= P(f_t(X) \neq l) = P(f_t(X) \neq l | X \leq t)P(X \leq t) + P(f_t(X) \neq l | X > t)P(X > t) = \\ &= P(l \neq a^L | X \leq t)P(X \leq t) + P(l \neq a^R | X > t)P(X > t) \end{aligned}$$

где соответственно:

$$f_t(x) = a^L I\{x \leq t\} + a^R (1 - I\{x \leq t\})$$

$$a^L = \begin{cases} 1, & P(l = 1 | X \leq t) > \frac{1}{2} \\ -1, & P(l = 1 | X \leq t) \leq \frac{1}{2} \end{cases}$$

$$a^R = \begin{cases} 1, & P(l = 1 | X > t) > \frac{1}{2} \\ -1, & P(l = 1 | X > t) \leq \frac{1}{2} \end{cases}$$

Введем обозначения:

$$q^{\leq}(t) = \begin{cases} P(l = 1|X \leq t), & P(X \leq t) \neq 0 \\ 0, & P(X \leq t) = 0 \end{cases}$$

$$q^>(t) = \begin{cases} P(l = 1|X > t), & P(X > t) \neq 0 \\ 0, & P(X > t) = 0 \end{cases}$$

Распишем вероятность ошибки по частям:

$$P(f_t(X) \neq l|X \leq t) = P(a^L \neq l|X \leq t) = \begin{cases} P(l \neq 1|X \leq t), & q^{\leq}(t) > \frac{1}{2} \\ P(l \neq -1|X \leq t), & q^{\leq}(t) \leq \frac{1}{2} \end{cases} =$$

$$= \begin{cases} 1 - q^{\leq}(t), & q^{\leq}(t) > \frac{1}{2} \\ q^{\leq}(t), & q^{\leq}(t) \leq \frac{1}{2} \end{cases} = \min\{q^{\leq}(t), 1 - q^{\leq}(t)\}$$

Аналогично получаем, что:

$$P(f_t(X) \neq l|X > t) = P(a^R \neq l|X > t) = \min\{q^>(t), 1 - q^>(t)\}$$

Введём функцию S - это некоторая функция, часто называемая в статьях по машинному обучению как split-функция, определенная на отрезке $[0, 1]$ и удовлетворяющая условиям:

1. $S(0) = S(1) = 0$, $S(\frac{1}{2}) = \frac{1}{2}$
2. Неотрицательна и симметрична относительно точки $\frac{1}{2}$: $S(x) = S(1 - x)$
3. S выпукла вверх
4. $S \in C^2[0, 1]$

Заметим, что из приведённых выше четырёх условий следует, что: $S(x) \geq \min\{x, 1 - x\}$

Используя неравенство $\min\{q^{\leq}(t), 1 - q^{\leq}(t)\} \leq S(q^{\leq}(t))$, получаем:

$$e(t) \leq P(X \leq t)S(q^{\leq}(t)) + P(X > t)S(q^>(t))$$

Введем обозначения для функционала правой части:

$$Z(t) = S(q^{\leq})P(X_k \leq t) + S(q^>)P(X_k > t)$$

и его эмпирический аналог:

$$Z_n(t) = S(q_n^{\leq})P_n(X_k \leq t) + S(q_n^>)P_n(X_k > t)$$

Определим T и T_n :

$$T = \arg \min_t (Z(t)), \quad T_n = \arg \min_t (Z_n(t))$$

Под $\arg \min(f)$ везде понимаем наименьшее значение t , при котором достигается минимум функции $f(t)$ на \mathbf{R} . В дальнейшем мы всегда будем предполагать, что минимум функции $Z(t)$ на \mathbf{R} достигается в единственной точке $t = T$.

Следовательно, минимизируя величины $Z(t), Z_n(t)$, мы уменьшаем теоретическую и эмпирическую ошибки классификации $e(t), e_n(t)$ соответственно.

Обоснование подхода, в котором мы уменьшаем некоторый функционал, который является оценкой сверху над ошибкой классификации, заключается в практическом использовании

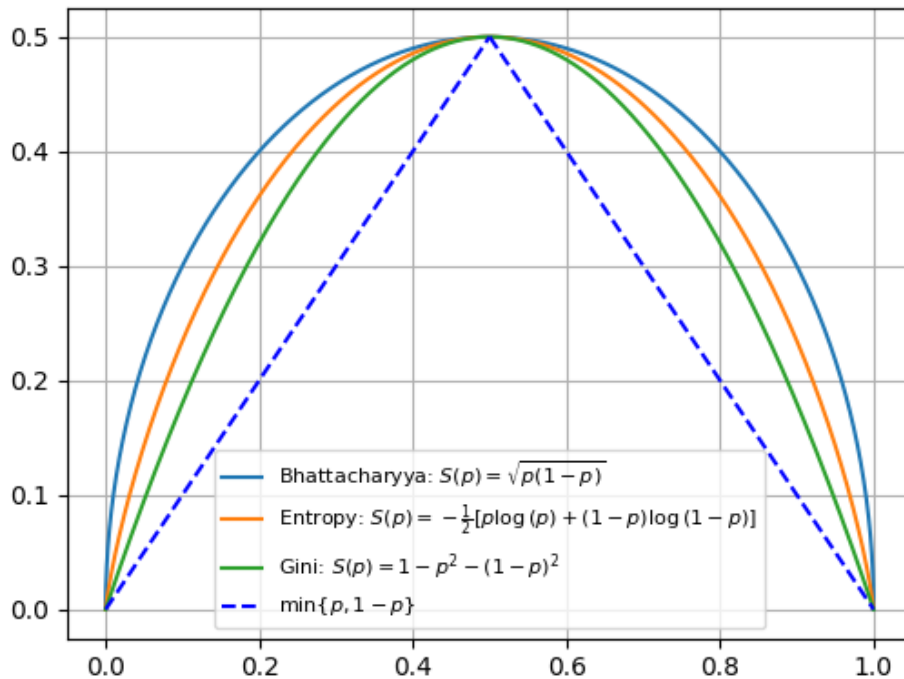
данного алгоритма. В машинном обучении параметры моделей часто настраиваются за счет уменьшения некоторой функции потерь (функционала, зависящего от ответов модели и выборки, на которой он настраивается), и такая эвристика обусловлена двумя основными критериями:

1. Ускорение скорости настройки параметров алгоритма (так как уменьшение непосредственно ошибки является задачей полного перебора)
2. Уменьшение влияния переобучения алгоритма (ситуация, когда алгоритм хорошо работает на выборке, на которой он настраивался и плохо на новых данных)
3. Ошибка классификации не является гладкой функцией, поэтому её трудно исследовать.

Выбор функции S является важной задачей в алгоритме решающего дерева, наиболее распространены такие виды функций:

1. Критерий Бхаттачарья: $S(p) = \sqrt{p(1-p)}$
2. Критерий Джини: $S(p) = 1 - p^2 - (1-p)^2$
3. Энтропийный критерий: $S(p) = -\frac{1}{2}(p \log(p) + (1-p) \log(1-p))$

Ниже представлены графики данных split функций:



Цель данной дипломной работы - изучение предельного поведения случайной величины T_n при $n \rightarrow \infty$.

Дерево решений в задаче классификации является одним из важнейших методов машинного обучения, который применяется повсеместно за счет своей высокой интерпретируемости, скорости работы (проверяется ограниченное число правил, задаваемых неравенствами). Например, в качестве признака X могут выступать такие параметры, как пол, возраст, температура, симптомы, потраченная сумма на покупки за год и т.д., а в качестве метки l могут быть индикатор

того вернет человек кредит банку или нет, какое заболевание у человека и т.д. Все эти факты порождают особое внимание, которое уделяется изучению различных свойств данного алгоритма.

На базе такого простого алгоритма, как дерево решений работают гораздо более сложные (ансамблирующие) алгоритмы, такие как бустинг над решающими деревьями (где каждое следующее дерево предсказывает градиенты ошибок функции потерь от ответа, полученного от усреднения всех предыдущих деревьев в ансамбле) и случайный лес (где усредняются ответы деревьев, построенных на случайных подвыборках признаков из генеральной совокупности, этот факт уменьшает корреляцию между деревьями, увеличивая обобщающую способность). Например, в статье [8] приводится оценка увеличения обобщающей способности для ансамбля классификаторов, построенных на случайных подвыборках.

Авторы статьи [9] демонстрируют, что в случае шумных наблюдений взвешенное разделение склонно к, так называемым, end-cut разделениям (выбора точки разделения узла на краю диапазона значений признака), которые традиционно считались нежелательными для одиночных деревьев, однако, для случайного леса такие разделения не портят работу алгоритма. В статье приводятся две причины: увеличение размера выборки, которое позволяет дереву восстановиться после плохого разделения, и, если ветка дерева постоянно разделяется по шуму, то, если будет достигнут минимальный размера выборки в листе, плохая ветка завершит свое построение.

Исследований по данной теме большое количество, однако не все из них имеют высокую математическую составляющую и зачастую являются просто компьютерным моделированием и проверкой работ различных эвристик, используемых для улучшения работы данного алгоритма.

Статьи, в которых проводятся исследования асимптотического поведения параметров решающего штампа не так много, однако наиболее близкие к моей работе исследования приводятся в статье [3]. В ней автор исследует сходимость разделяющей границы T_n для ошибки классификатора $L_n = P_n(yf < 0)$, в предположении, что T является медианой распределения $F(l = 1|X \leq t)$. В статье устанавливается, что $|T_n - T|$ сходится к распределению Чернова ($\arg \max_{t \in \mathbf{R}} (W(t) - t^2)$) со скоростью сходимости $n^{\frac{1}{3}}$. Его результаты во многом ссылаются на результаты работы [4], в которой автор исследует сходимость параметра эмпирического процесса (который является усреднение некоторых базовых классификаторов), в которых процесс достигает максимума при условии определенных технических ограничений. В статье [5] автор обобщает некоторые результаты статьи [4] на случай зависимых переменных.

Интересно так же увидеть результаты, полученные на дереве решений с выбранным критерием разделения. Так в работе [6] авторы, используя критерий Gini, исследуют и проводят компьютерные симуляции, того, как ведет себя разделяющая граница в дереве решений в зависимости от исходной выборки. В статье исследуется устойчивость алгоритма к входным данным и показывается, что разность параметров модели, полученных при разных генерациях выборки, стремится к нормальному распределению с параметрами, выводимыми в статье.

2 Вспомогательные утверждения

Напомню введённые ранее обозначения:

$$F^+(t) = P(X \leq t | l = 1)$$

$$F^-(t) = P(X \leq t | l = -1)$$

$$\pi^+ = P(l = 1)$$

$$\pi^- = P(l = -1) = 1 - \pi^+$$

Рассмотрим следующие функционалы и их точки минимума:

$$Z(t) = S(q_p^{\leq})P(X_k \leq t) + S(q_p^>)P(X_k > t)$$

$$Z_n(t) = S(q_n^{\leq})P_n(X_k \leq t) + S(q_n^>)P_n(X_k > t)$$

$$T = \arg \min_t (Z(t)), \quad T_n = \arg \min_t (Z_n(t))$$

И пусть выполнены два условия:

1. Функции $F^{\pm}(t)$ непрерывны
2. Минимум функционала $Z(t)$ достигается в единственной точке T

Теорема 1. $T_n \xrightarrow{n.u.} T$ при $n \rightarrow \infty$.

Лемма 1. Для любого $\epsilon > 0$ существует $\exists \delta > 0$ т.ч. для любого t удовлетворяющего условию: $|Z(T) - Z(t)| < \delta$, выполняется $|T - t| \leq \epsilon$

Доказательство: Докажем от противного: пусть $\exists \epsilon > 0$ т.ч. для $\forall \delta > 0 \exists t$ такое что. $|Z(T) - Z(t)| < \delta$ и $|T - t| > \epsilon$. Можно предполагать, что $t \in [a, b]$ для некоторого отрезка $[a, b]$, так как:

$$\lim_{t \rightarrow +\infty} Z(t) = \lim_{t \rightarrow -\infty} Z(t) = S(\pi^+) = S(\pi^-) = \sup_t Z(t)$$

Возьмем сходящиеся последовательности $t_n \in [a, b]$ и δ_n т.ч:

$$|Z(T) - Z(t_n)| < \delta_n = \frac{1}{n} \text{ и } |T - t_n| > \epsilon$$

Выберем сходящуюся подпоследовательность t_{n_k} , $k = 1, 2, \dots$, $t_{n_k} \rightarrow t^*$ при $k \rightarrow \infty$. Тогда, в силу выбора последовательности t_n , имеем

$$\lim_{k \rightarrow \infty} Z(t_{n_k}) = Z(t^*) = Z(T)$$

С другой стороны, $|T - t^*| > \epsilon$, что противоречит единственности точки, в которой достигается минимум функции $Z(t)$. Тем самым лемма 1 доказана. \square

Доказательство Теоремы 1:

Нетрудно видеть, что $\lim_{n \rightarrow \infty} \sup_t |F_n^{\pm}(t) - F^{\pm}(t)| = 0$, доказательство этого факта полностью повторяет доказательство теоремы Гливленко-Кантелли.

Пусть $P(X \leq t) \neq 0$, тогда распишем вероятности $q^{\leq}(t)$:

$$q^{\leq}(t) = P(l = 1 | X \leq t) = \frac{P(l = 1, X \leq t)}{P(X \leq t)} = \frac{P(X \leq t | l = 1)P(l = 1)}{P(X \leq t)} =$$

$$= \frac{F^+(t)\pi^+}{F^+(t)\pi^+ + F^-(t)\pi^-} =$$

Аналогично получаем, что:

$$q^>(t) = \frac{(1 - F^+(t))\pi^+}{(1 - F^+(t))\pi^+ + (1 - F^-(t))\pi^-}$$

Следовательно, функционал $Z(t)$ можно переписать в виде:

$$Z(t) = H(F^+(t), F^-(t), \pi^+, \pi^-) \quad (1)$$

$$H(F^+, F^-, \pi^+, \pi^-) = (F^+\pi^+ + F^-\pi^-)S\left(\frac{F^+\pi^+}{F^+\pi^+ + F^-\pi^-}\right) + \\ + ((1 - F^+)\pi^+ + (1 - F^-)\pi^-)S\left(\frac{(1 - F^+)\pi^+}{(1 - F^+)\pi^+ + (1 - F^-)\pi^-}\right)$$

где $F^+, F^-, \pi^+, \pi^- \in [0, 1]$.

Таким образом H является гладкой функцией, и используя тот факт, что $\lim_{n \rightarrow \infty} \sup_t |F_n^\pm(t) - F^\pm(t)| = 0$ и ЗБЧ для π_n^\pm , получаем оценку:

$$\lim_{n \rightarrow \infty} \sup_t |Z_n(t) - Z(t)| = 0 \quad (2)$$

Зафиксируем $\epsilon > 0$, тогда в силу доказанной Леммы [1] $\exists \delta > 0$ такое, что для любого t , удовлетворяющего $|Z(T) - Z(t)| < \delta$, выполняется $|t - T| < \epsilon$. Из формулы (2) следует, что с вероятностью единица существует N т.ч. для всех $n > N$:

$$\sup_{t \in R} |Z(t) - Z_n(t)| < \frac{\delta}{2}$$

Имеем неравенства:

$$Z(T) > Z_n(T) - \frac{\delta}{2} > Z_n(T_n) - \frac{\delta}{2} > Z(T_n) - \delta$$

Следовательно, $0 < Z(T_n) - Z(T) < \delta$. Значит $|T_n - T| < \epsilon$.

Тем самым теорема 1 доказана. \square

Идея доказательства теоремы 1 взята из дипломной работы студента Жеглова Д. М.*

Утверждение 1. Пусть $\Delta_n(t) = Z_n(t) - Z(t)$, $I_n = I(T_n \in O_T^\delta)$, где O_T^δ - окрестность точки T радиуса δ , I - индикатор события. Тогда:

$$|T_n - T|^{\frac{3}{2}} I_n \leq \frac{\Delta_n(T_n) - \Delta_n(T)}{\sqrt{|T_n - T|}}$$

Доказательство: Из определения T и T_n следует, что:

$$Z_n(T_n) \leq Z_n(T)$$

Отсюда имеем, что:

$$\Delta_n(T_n) - Z(T) \leq \Delta_n(T) - Z(T_n) \\ 0 \leq Z(T_n) - Z(T) \leq \Delta_n(T) - \Delta_n(T_n)$$

*См. дипломную работу Жеглова Д. М., 2019 год выпуска.

Поскольку T является минимумом функционала $Z(t)$, значит $\exists \delta > 0$ т.ч. существует окрестность точки T радиуса δ , обозначим её как O_T^δ и существует константа α т.ч. для любой точки $t \in O_T^\delta$, выполняется неравенство:

$$\alpha(T - t)^2 \leq Z(t) - Z(T)$$

Следовательно:

$$\begin{aligned} \alpha(T_n - T)^2 I(T_n \in O_T^\delta) &\leq (Z(T_n) - Z(T)) I_n \leq (\Delta_n(T) - \Delta_n(T_n)) I_n \leq \\ &\leq (\Delta_n(T) - \Delta_n(T_n)) \end{aligned}$$

где $I_n = I(T_n \in O_T^\delta)$

Без ограничения общности, будем считать, что $\alpha = 1$, тогда неравенство можно переписать в виде:

$$|T_n - T|^{\frac{3}{2}} I_n \leq \frac{\Delta_n(T_n) - \Delta_n(T)}{\sqrt{|T_n - T|}} \quad (3) \quad \square$$

Определение 1. Последовательность случайных величин $\{X_n\}$, $n = 1, 2, \dots$ называется ограниченной по вероятности, если $\forall \epsilon > 0$ существует $M_\epsilon = M$, $N \in \mathbf{N}$ т.ч. $\forall n > N$ выполнено:

$$P(|X_n| > M) < \epsilon$$

Тогда последовательность обозначается как $X_n = O_P(1)$.

Будем обозначать, $X_n = O_P(a_n)$, если $\frac{X_n}{a_n} = O_P(1)$.

Лемма 2. Пусть имеются две последовательности случайных величин: $\{\xi_n\}, \{\eta_n\}$. Тогда:

1. Если для всех n выполнено неравенство: $|\xi_n| \leq |\eta_n|$, и $\eta_n = O_P(1)$, то $\xi_n = O_P(1)$

Если для всех n выполнено неравенство $|\xi_n| \leq |\eta_n|$, и $\xi_n \neq O_P(1)$, то $\eta_n \neq O_P(1)$

2. Если $\xi_n = O_P(1)$, $\eta_n = O_P(1)$, то $\max(\xi_n, \eta_n) = O_P(1)$

3. Если $\xi_n = O_P(1)$, $\eta_n = O_P(1)$, то $\xi_n + \eta_n = O_P(1)$

4. Если $\xi_n \xrightarrow{d} \xi$, то $\xi_n = O_P(1)$

5. Если $\xi_n \eta_n = O_P(a_n)$, $\eta_n \xrightarrow{n.n.} 1$, то $\xi_n = O_P(a_n)$

Доказательство:

1. Доказательство первого свойства напрямую вытекает из определения $O_P(1)$.

2. Распишем по определению условие 1:

$$\forall \epsilon_1 > 0 \exists M_{\epsilon_1} : \forall M > M_{\epsilon_1} : P(|\xi_n| > M) < \epsilon_1$$

$$\forall \epsilon_2 > 0 \exists M_{\epsilon_2} : \forall M > M_{\epsilon_2} : P(|\eta_n| > M) < \epsilon_2$$

Тогда: $\forall \epsilon > 0 \exists M = \max(M_{\epsilon_1}, M_{\epsilon_2})$:

$$P(|\max(\xi_n, \eta_n)| > M_\epsilon) = P(|\max(\xi_n, \eta_n)| > M_\epsilon \mid |\xi_n| > M_\epsilon) P(|\xi_n| > M_\epsilon) +$$

$$P(|\max(\xi_n, \eta_n)| > M_\epsilon \mid |\xi_n| \leq M_\epsilon)P(|\xi_n| \leq M_\epsilon) = P(|\xi_n| > M_\epsilon) + P(|\eta_n| > M_\epsilon)P(|\xi_n| < M_\epsilon) \leq \\ P(|\xi_n| > M_\epsilon) + P(|\eta_n| > M_\epsilon) = \{\text{пусть } \epsilon_1 = \epsilon_2 = \frac{\epsilon}{2}\} = \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

3. Если $P(A) > 1 - \epsilon$, $P(B) > 1 - \epsilon$, тогда $P(AB) = P(A) + P(B) - P(A \cup B) > 1 - 2\epsilon$

$$\forall \epsilon > 0 \exists M_\epsilon = \frac{M}{2} : \forall M > \frac{M}{2} : P(|\xi_n| > M) < \epsilon_1$$

Применим это неравенство, получим: $P(|\xi_n + \eta_n| < M) > P(|\xi_n| < \frac{M}{2}, |\eta_n| < \frac{M}{2}) > 1 - 2\epsilon$

4. Пусть $\epsilon > 0$, выберем достаточно большое $M_0(\epsilon) = M_0 : P(|\xi| > M_0) < \epsilon$, из сходимости по распределению имеем:

$$P(|\xi_n| > M_0) \rightarrow P(|\xi| > M_0)$$

Таким образом $\exists N_0 : \forall n > N_0$:

$$P(|\xi_n| > M_0) < \epsilon$$

Выберем M_1 т.ч. $P(|X_i| > M_1) < \epsilon$ для $i = 1, 2, \dots, N_0 - 1$. Тогда для $M = \max(M_0, M_1)$ имеем:

$$P(X_n > M) < \epsilon$$

5. Пусть $\epsilon > 0$, тогда $\exists N_1 : \forall n > N_1$:

$$P(|\eta_n| - 1) > \frac{1}{2} < \frac{\epsilon}{2}$$

Также распишем условие на $\xi_n \eta_n = O_P(a_n)$: $\exists M, N_2 : \forall n > N_2$

$$P(|\frac{\xi_n \eta_n}{a_n}| > M) < \frac{\epsilon}{2}$$

Распишем цепочку неравенств, используя тот факт, что $P(A, B) \geq P(A) + P(B) - 1$:

$$P(|\frac{\xi_n \eta_n}{a_n}| > M) \geq P(|\frac{\xi_n \eta_n}{a_n}| > M \mid |\eta_n - 1| \leq \frac{1}{2}) \geq P(|\frac{\xi_n}{a_n}| > 2M \mid |\eta_n - 1| \leq \frac{1}{2}) \geq \\ \geq P(|\frac{\xi_n}{a_n}| > 2M) - \frac{\epsilon}{2} \geq$$

Получаем:

$$P(|\frac{\xi_n}{a_n}| > 2M) \leq \epsilon$$

Что завершает доказательство леммы 2. \square

Лемма 3. Рассмотрим функцию $G : R^d \rightarrow R$, определенную на открытом выпуклом ограниченном множестве O . Пусть $G \in C^2(O)$, тогда существуют две неотрицательные константы c_1, c_2 , такие что $\forall x_1, x_2, y_1, y_2 \in O$ таких, что $x_1 + (y_2 - y_1) \in O$ выполняется неравенство:

$$|((G(y_2) - G(y_1)) - (G(x_2) - G(x_1)))| \leq c_1|(y_2 - y_1) - (x_2 - x_1)| + c_2|y_2 - y_1||y_1 - x_1|$$

где $|\cdot|$ обозначается как стандартная евклидова норма в R^d .

Доказательство: Пусть $t \in R$, введем обозначение :

$$\Delta(t) = (G(x_1 + t(y_1 - x_1) + y_2 - y_1) - G(x_1 + t(y_1 - x_1))) - (G(x_2) - G(x_1))$$

Тогда

$$\Delta(1) = (G(y_2) - G(y_1)) - (G(x_2) - G(x_1))$$

Поскольку O является выпуклым множеством, следовательно $\Delta(t)$ определён для $\forall t \in [0, 1]$.

$$\frac{d\Delta(t)}{dt} = \dot{\Delta}(t) = ([\nabla G(x_1 + t(y_1 - x_1) + y_2 - y_1) - \nabla G(x_1 + t(y_1 - x_1))], y_1 - x_1)$$

где ∇ обозначает градиент, а (\cdot, \cdot) обозначает евклидово скалярное произведение в R^d .

Поскольку G дважды непрерывно дифференцируемая функция, мы имеем:

$$|\nabla G(x_1 + t(y_1 - x_1) + y_2 - y_1) - \nabla G(x_1 + t(y_1 - x_1))| \leq c_1 |y_2 - y_1|$$

для некоторой константы c_1 , не зависящей от x_1, x_2, y_1, y_2 .

Следовательно $\forall t \in [0, 1]$, имеем:

$$\frac{d\Delta(t)}{dt} \leq c_1 |y_2 - y_1| |y_1 - x_1|$$

По формуле Ньютона-Лейбница:

$$\Delta(1) = \Delta(0) + \int_0^1 \dot{\Delta}(t) dt$$

Получаем:

$$|\Delta(1)| \leq |\Delta(0)| + c_1 |y_2 - y_1| |y_1 - x_1|$$

Теперь оценим $|\Delta(0)|$:

$$|\Delta(0)| = |G(x_1 + (y_2 - y_1)) - G(x_2)| \leq c_2 |(y_2 - y_1) - (x_2 - x_1)|$$

Что завершает доказательство леммы 3. \square

3 Основные результаты

Введем функции совместных вероятностей, векторы совместных вероятностей и их эмпирические аналоги:

$$\begin{aligned}\omega^{+L}(t) &= P(X \leq t, l = 1) = F^+(t)\pi^+, & \omega^{+R}(t) &= P(X > t, l = 1) = (1 - F^+(t))\pi^+, \\ \omega^{-L}(t) &= P(X \leq t, l = -1) = F^-(t)\pi^-, & \omega^{-R}(t) &= P(X > t, l = -1) = (1 - F^-(t))\pi^-\end{aligned}$$

Считаем, что функции совместных вероятностей $\omega^{\pm L}(t), \omega^{\pm R}(t) \in C^1(\mathbf{R})$.

Введем, соответствующие им, эмпирические процессы:

$$\omega_n^{+L}(t), \omega_n^{+R}(t), \omega_n^{-L}(t), \omega_n^{-R}(t)$$

Обозначим за $s(t), s_n(t)$ векторы совместных вероятностей и эмпирических совместных вероятностей соответственно:

$$\begin{aligned}s(t) &= (\omega^{+L}(t), \omega^{+R}(t), \omega^{-L}(t), \omega^{-R}(t)) \\ s_n(t) &= (\omega_n^{+L}(t), \omega_n^{+R}(t), \omega_n^{-L}(t), \omega_n^{-R}(t))\end{aligned}$$

Тогда, согласно формуле (1), функционал $Z(t)$ можно переписать в виде:

$$Z(t) = \tilde{H}(s(t)), \quad Z_n(t) = \tilde{H}(s_n(t)),$$

где \tilde{H} некоторая гладкая функция:

Введем обозначения: $\delta_n(t) = s_n(t) - s(t)$.

Следствие 1. *Используя утверждение 1 и лемму 3, получаем:*

$$\begin{aligned}|T_n - T|^{\frac{3}{2}} I_n &\leq \frac{\Delta_n(T) - \Delta_n(T_n)}{\sqrt{|T_n - T|}} = \frac{(Z_n(T) - Z(T)) - (Z_n(T_n) - Z(T_n))}{\sqrt{|T_n - T|}} \leq \\ &C_1 \frac{|(s_n(T) - s(T)) - (s_n(T_n) - s(T_n))|}{\sqrt{|T_n - T|}} + C_2 \frac{|s_n(T) - s(T)||s(T) - s(T_n)|}{\sqrt{|T_n - T|}} = \\ &= C_1 \frac{|\delta_n(T) - \delta_n(T_n)|}{\sqrt{|T_n - T|}} + C_2 |\delta_n(T)| \frac{|s(T) - s(T_n)|}{\sqrt{|T_n - T|}} \leq \\ &\leq C_1 \sup_{t \in O_T^\delta} \frac{|\delta_n(T) - \delta_n(t)|}{\sqrt{|T - t|}} + C_2 |\delta_n(T)| \sup_{t \in O_T^\delta} \frac{|s(T) - s(t)|}{\sqrt{|T - t|}}\end{aligned}$$

Где супремум во втором слагаемом берется от гладкой функцией, зависящей от ограниченных функций (координат вектора совместных вероятностей). Второе слагаемое не содержит случайных величин и обращается в нуль при $t = T$, следовательно оно может быть ограничено некоторой константой. А множитель $|\delta_n(T)| = O_P(n^{\frac{1}{2}})$ в силу теоремы Колмогорова-Смирнова и леммы 2.

Осталось показать, что если $\sqrt{n} \sup_{t \in O_T^\delta} \frac{|\delta_n(T) - \delta_n(t)|}{\sqrt{|T - t|}} = O_P(1)$, то по замечанию 1 следует, что

$$|T_n - T| = O_P(n^{-\frac{1}{3}})$$

Однако, как мы позже выясним, это условие не выполняется. Для этого нам понадобится доказать ряд теорем.

Теорема 2. Пусть для всех четырёх функций распределения, являющиеся координатами вектора $s(t)$, выполнено условие: существует $c \in \mathbf{R}$ т.ч. $|T - F^{-1}(y)| \leq c|T - y|$ *, где $F^{-1}(y) = \{\min z : F(z) = y\}$, $y \in F(O_\delta^T)$, тогда:

$$\sqrt{n} \sup_{t \in O_\delta^T} \frac{|\delta_n(T) - \delta_n(t)|}{\sqrt{|T - t|}} \neq O_P(1) \quad (4)$$

Для доказательства этой теоремы, нужно доказать ряд определенных фактов:

Теорема 3. Пусть X_1, X_2, \dots, X_n - выборка из независимых одинаково распределенных случайных величин с функцией распределения $F(t)$, $\delta_n(t) = F_n(t) - F(t)$.

1. Если выполнено, что $|F^{-1}(y)| \geq c_1|y|$, где c_1 - некоторая положительная константа, положим:

$$f(t) = \begin{cases} \sqrt{t \log \log \frac{1}{t}}, & t < c \\ \sqrt{t \log \log \frac{1}{c}}, & t \geq c \end{cases}$$

где c - некоторая константа, т.ч. $\sqrt{t \log \log \frac{1}{t}}$ монотонно возрастает на $[0, c]$. Тогда:

$$\sqrt{n} \sup_{0 \leq t \leq 1} \frac{|\delta_n(t)|}{f(t)} = O_P(1)$$

2. Если выполнено, что $|F^{-1}(y_1) - F^{-1}(y_2)| \leq c_2|y_1 - y_2|$, где c_2 - некоторая положительная константа, M - некоторая константа, т.ч. $\sqrt{t \log \log \frac{1}{t}}$ монотонно возрастает на $[0, M]$ тогда:

$$\sqrt{n} \sup_{|t-s| < M} \frac{|\delta_n(t) - \delta_n(s)|}{\sqrt{|t-s| \log |t-s|^{-1}}} \neq O_P(1) \quad (6)$$

Доказательство Теоремы 3 пункт 1:

Проведем аналогичную процедуру, как в доказательстве теоремы Колмогорова-Смирнова: рассмотрим другую выборку случайных величин, $Y_j = F(X_j)$. Получим, что случайные величины Y_1, Y_2, \dots, Y_n являются н.о.р.с.в, имеющие равномерное распределение на отрезке $[0, 1]$, тогда:

$$\begin{aligned} \sup_{t < c} \frac{|\delta_n(t)|}{\sqrt{t \log \log \frac{1}{t}}} &= \sup_{t < c} \frac{|\frac{1}{n} \sum_{i=1}^n |I\{X_j < t\} - F(t)||}{\sqrt{t \log \log \frac{1}{t}}} = \sup_{t < c} \frac{|\frac{1}{n} \sum_{i=1}^n |I\{F(X_j) < F(t)\} - F(t)||}{\sqrt{t \log \log \frac{1}{t}}} = \\ &= \sup_{t < c} \frac{|\frac{1}{n} \sum_{i=1}^n |I\{Y_j < y\} - y||}{\sqrt{F^{-1}(y) \log \log \frac{1}{F^{-1}(y)}}} \leq \sup_{t < c} \frac{|\frac{1}{n} \sum_{i=1}^n |I\{Y_j < y\} - y||}{\sqrt{y \log \log \frac{1}{y}}} \end{aligned}$$

Распределение последнего выражения не зависит от $F(x)$, следовательно, без ограничения общности, можно считать, что $X_1, \dots, X_n \sim U[0, 1]$.

Если показать, то выражение в правой части сходится по распределению к некоторой с.в., то по лемме 2.4 и 2.1 мы докажем теорему 3.1.

Обоснование того, что вид функция $f(t)$ имеет такой вид тривиален: так как функция $\sqrt{t \log \log \frac{1}{t}}$ не является монотонной на $[0, 1]$, была введена вторая функция. Это необходимо,

* Данное условие ограничивает сверху константой скорость, при которой функция $F^{-1}(y)$ стремится к точке T , при $y \rightarrow F(T)$. То есть производная функции $F^{-1}(y)$ ограничена сверху константой в окрестности точки $F(O_\delta^T)$.

чтобы при делении кусочно-постоянной функции на данную функцию характер монотонности не менялся.

Так как особенность в знаменателе возникает только в окрестности точки 0, следовательно для простоты рассмотрим все $X_{(k)} \leq c$, без ограничения общности будем считать, что их n штук. Функция $|\delta_n(t)| = |F_n(t) - t|$ кусочно линейна на $[X_{(k-1)}, X_{(k)}]$ следовательно достигаем максимумов на концах отрезков. Если её поделить на функцию $\sqrt{t \log \log(\frac{1}{t})}$, то характер монотонности на отрезке не изменится. Следовательно:

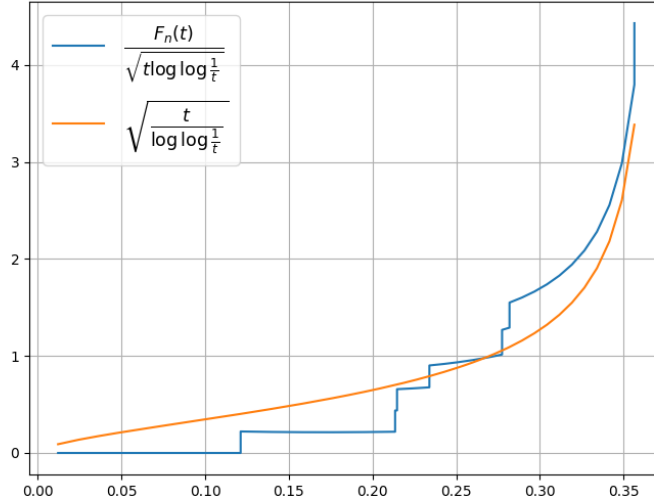


Рис. 1: На рисунке представлен графики функции $\frac{F_n(t)}{\sqrt{t \log \log \frac{1}{t}}}$, $\frac{t}{\sqrt{t \log \log \frac{1}{t}}}$ для $n = 10$

$$\sqrt{n} \sup_{0 \leq t \leq c} \frac{|\delta_n(t)|}{\sqrt{t \log \log \frac{1}{t}}} = \sqrt{n} \max \left(\max_{k=1, \dots, n} \frac{|X_{(k)} - \frac{k}{n}|}{\sqrt{X_{(k)} \log \log \frac{1}{X_{(k)}}}}, \max_{k=1, \dots, n} \frac{|X_{(k)} - \frac{k-1}{n}|}{\sqrt{X_{(k)} \log \log \frac{1}{X_{(k)}}}} \right) = \sqrt{n} \max(G_n, D_n)$$

Рассмотрим G_n : так как $(X_{(1)}, X_{(2)}, \dots, X_{(n)}) \stackrel{d}{=} (\frac{S_1}{S_{n+1}}, \frac{S_1}{S_{n+1}}, \dots, \frac{S_n}{S_{n+1}})$, где $S_i \sim \exp(1)$ - н.о.р.с.в. (см. [1] стр. 468), имеем:

$$\sqrt{n} G_n \stackrel{d}{=} \sqrt{n} \max_{k=1, \dots, n} \frac{|\frac{S_k}{S_{n+1}} - \frac{k}{n}|}{\sqrt{\frac{S_k}{S_{n+1}} \log \log \frac{S_{n+1}}{S_k}}}$$

По теореме Донскера-Прохорова введем кусочно-линейный процесс:

$$Y_t^{(n)} = \frac{S_k - k}{\sqrt{n+1}}(k+1 - (n+1)t) + \frac{S_{k+1} - (k+1)}{\sqrt{n+1}}((n+1)t - k)$$

Где $t \in [\frac{k}{n+1}, \frac{k+1}{n+1}]$, $k = 0, 1, \dots, n$, сходящийся (в смысле случайных процессов) к винеровскому процессу: $Y_t^{(n)} \xrightarrow{D} W(t)$, $t \in [0, 1]$ при $n \rightarrow \infty$. Заметим, что $\frac{S_k - k}{\sqrt{n+1}} = Y^{(n)}(\frac{k}{n+1})$ и $\frac{S_{n+1} - (n+1)}{\sqrt{n+1}} = Y^{(n)}(1)$.

Следовательно $S_{n+1} = Y^{(n)}(1)\sqrt{n+1} + (n+1)$, $S_k = Y^{(n)}(\frac{k}{n+1})\sqrt{n+1} + k$, то есть:

$$\frac{S_k}{S_{n+1}} = \frac{Y^{(n)}(\frac{k}{n+1})\sqrt{n+1} + k}{Y^{(n)}(1)\sqrt{n+1} + (n+1)} = \frac{\frac{Y^{(n)}(\frac{k}{n+1})}{\sqrt{n+1}} + \frac{k}{n+1}}{\frac{Y^{(n)}(1)}{\sqrt{n+1}} + 1}$$

Перепишем теперь исходный процесс G_n :

$$\begin{aligned}
\sqrt{n}G_n &= \sqrt{n} \max_{k=1, \dots, n} \frac{|\frac{S_k}{S_{n+1}} - \frac{k}{n}|}{\sqrt{\frac{S_k}{S_{n+1}} \log \log \frac{S_{n+1}}{S_k}}} \stackrel{n \rightarrow \infty}{=} \frac{n+1}{S_{n+1}} \max_{k=1, \dots, n} \frac{|\frac{S_k - k}{\sqrt{n+1}} - \frac{k}{n+1} \frac{S_{n+1} - (n+1)}{\sqrt{n+1}}|}{\sqrt{\frac{S_k}{S_{n+1}} \log \log \frac{S_{n+1}}{S_k}}} = \\
&= \frac{n+1}{S_{n+1}} \max_{k=1, \dots, n} |Y^{(n)}(\frac{k}{n+1}) - \frac{k}{n+1} Y^{(n)}(1)| \sqrt{\frac{\frac{Y^{(n)}(1)}{\sqrt{n+1}} + 1}{\frac{Y^{(n)}(\frac{k}{n+1})}{\sqrt{n+1}} + \frac{k}{n+1}}} (\log \log \frac{\frac{Y^{(n)}(1)}{\sqrt{n+1}} + 1}{\frac{Y^{(n)}(\frac{k}{n+1})}{\sqrt{n+1}} + \frac{k}{n+1}})^{-1} = \\
&= \frac{n+1}{S_{n+1}} \max_{t \in [0,1]} |Y^{(n)}(t) - tY^{(n)}(1)| \sqrt{\frac{\frac{Y^{(n)}(1)}{\sqrt{n+1}} + 1}{\frac{Y^{(n)}(\frac{[(n+1)t]}{n+1})}{\sqrt{n+1}} + t}} (\log \log \frac{\frac{Y^{(n)}(1)}{\sqrt{n+1}} + 1}{\frac{Y^{(n)}(\frac{[(n+1)t]}{n+1})}{\sqrt{n+1}} + t})^{-1}
\end{aligned}$$

Где в последнем равенстве $k = [(n+1)t]$ и оно выполнено, так как $|Y^{(n)}(t) - tY^{(n)}(1)|$ - кусочно линейная функция на $[X_{(k-1)}, X_{(k)}]$ и квадратный корень помноженный на двойной логарифм не меняют характер монотонности на отрезке. Перейдем к пределу при $n \rightarrow \infty$, получим:

$$\sqrt{n}G_n \xrightarrow{d} \sup_{t \in [0,1]} \frac{|B(t)|}{\sqrt{t \log \log \frac{1}{t}}} \text{ при } n \rightarrow \infty$$

где $B(t)$ - броуновский мост.

Аналогичные рассуждения получаются и для $\sqrt{n}D_n$. Следовательно для отрезка $[0, c]$ сходимость показана.

Для отрезка $[c, 1]$ все рассуждения полностью аналогичны выше, следовательно:

$$\sqrt{n} \sup_{0 \leq t \leq 1} \frac{|\delta_n(t)|}{f(t)} \xrightarrow{d} \sup_{t \in [0,1]} \frac{|B(t)|}{f(t)}, \text{ при } n \rightarrow \infty$$

Теперь покажем, что $\sup_{t \in [0,1]} \frac{|B(t)|}{f(t)} = O_P(1)$.

На отрезке $[c, 1]$: $\sup_{t \in [c,1]} \frac{|B(t)|}{\sqrt{f(t)}} \leq \frac{1}{\sqrt{c \log \log \frac{1}{c}}} \sup_{t \in [c,1]} |B(t)| = O_P(1)$ в силу теоремы Колмогорова-Смирнова.

На отрезке $[0, c]$: напишем закон повторного логарифма: $\limsup_{t \rightarrow 0} \frac{|W(t)|}{\sqrt{t \log \log \frac{1}{t}}} \stackrel{\text{н.н.}}{=} \sqrt{2}$, значит:

$$\forall \epsilon > 0 \text{ для почти всех } \omega \exists T(\omega, \epsilon) = t_0 : \forall 0 \leq t < t_0 : \frac{|W(t, \omega)|}{\sqrt{t \log \log \frac{1}{t}}} \leq \sqrt{2}(1 + \epsilon)$$

Так как $B(t) = W(t) - tW(1)$, следовательно:

$$\sup_{0 < t \leq t_\epsilon} \frac{|B(t)|}{\sqrt{t \log \log \frac{1}{t}}} \leq \sup_{0 < t \leq t_\epsilon} \frac{|W(t)|}{\sqrt{t \log \log \frac{1}{t}}} + \sup_{0 < t \leq t_\epsilon} \frac{|\sqrt{t}W(1)|}{\sqrt{\log \log \frac{1}{t}}}$$

Следовательно, по закону повторного логарифма, последовательность $\sup_{0 < t \leq t_\epsilon} \frac{|W(t)|}{\sqrt{t \log \log \frac{1}{t}}} = O_P(1)$ при $t_\epsilon \rightarrow 0$.

Последовательность $\sup_{0 < t \leq t_\epsilon} \frac{|W(1)t|}{\sqrt{t \log \log \frac{1}{t}}} = \sup_{0 < t \leq t_\epsilon} |W(1)| \sqrt{\frac{t}{\log \log \frac{1}{t}}} \leq |W(1)| \sqrt{\frac{t_\epsilon}{\log \log \frac{1}{t_\epsilon}}} = O_P(1)$, так как корень стремится к 0 при $t_\epsilon \rightarrow 0$. По лемме 2.3: $\sup_{0 < t \leq t_\epsilon} \frac{|B(t)|}{\sqrt{t \log \log \frac{1}{t}}} = O_P(1)$.

Выражение $\sup_{t_\epsilon < t \leq c} \frac{|B(t)|}{\sqrt{t \log \log \frac{1}{t}}} = O_P(1)$, по теореме Колмогорова-Смирнова, так как не имеет особенностей в знаменателе.

Значит по лемме 2.2 $\sup_{t \in [0, c]} \frac{|B(t)|}{f(t)} = O_P(1)$ как максимум из двух ограниченных по вероятности величин. Повторно применим лемму 2.2 и окончательно получаем: $\sup_{t \in [0, 1]} \frac{|B(t)|}{f(t)} = O_P(1)$ Ч.т.д. \square

Доказательство Теоремы 3 пункт 2:

Используя прием, полностью аналогичный доказательству первого пункта теоремы 3, мы можем сделать замену переменных и перейти к другой выборке, имеющей равномерное распределение, затем, используя условие $|F^{-1}(t) - F^{-1}(s)| \leq |t - s|$, мы будем оценивать с.в., ограничивающую снизу исходное выражение и покажем, что данная случайная величина не является ограниченной по вероятности.

Функция $|\delta_n(t) - \delta_n(s)|$ на двумерной сетке при $t > s$ является плоско-линейной, то есть вокруг каждой точки $(X_{(i)}, X_{(j)})$, не лежащей на границе треугольника $t > s$ в плоскости (t, s) сходятся 4 плоскости. Причем, легко показать, что в окрестности каждой такой точки, значения на двух плоскостях будут всегда совпадать. Следовательно, максимумы достигаются в точках $(X_{(k)}, X_{(l)})$, причем в трёх случаях для точек, которые не лежат на границе треугольника в плоскости (s, t) и в двух случаях для граничных точек, лежащих на прямых $s = 0, t = 1, t = s$.

Чтобы понять, как устроена эта поверхность, будем исследовать её по шагам:

1. Пусть $(t, s) = (X_{(k)}, X_{(l)})$ не лежит на границе, тогда в окрестности этой точки максимум может быть в одном из трёх случаев:

$$\sqrt{n} \frac{|\delta_n(X_{(k)}) - \delta_n(X_{(l)})|}{\sqrt{|t - s| \log |t - s|^{-1}}} = \frac{|(X_{(k)} - X_{(l)}) - \frac{k-l}{n}|}{\sqrt{(|X_{(k)} - X_{(l)}| \log \frac{1}{|X_{(k)} - X_{(l)}|})}}$$

Аналогично получаем ещё два значения:

$$\frac{|(X_{(k)} - X_{(l)}) - \frac{k-l-1}{n}|}{\sqrt{(|X_{(k)} - X_{(l)}| \log \frac{1}{|X_{(k)} - X_{(l)}|})}}$$

$$\frac{|(X_{(k)} - X_{(l)}) - \frac{k-l+1}{n}|}{\sqrt{(|X_{(k)} - X_{(l)}| \log \frac{1}{|X_{(k)} - X_{(l)}|})}}$$

2. Рассмотрим граничную область $|t - s| < \epsilon \rightarrow 0$ (гипотенузу треугольника): если $t = s$, то $\delta(t) - \delta(s) = 0$, рассмотрим теперь $t = X_{(k)}, s = X_{(k)} - \epsilon$:

$$\sqrt{n} \frac{|\delta_n(t) - \delta_n(s)|}{\sqrt{|t - s| \log |t - s|^{-1}}} = \left| \frac{\sqrt{n\epsilon}}{\sqrt{\log \frac{1}{\epsilon}}} - \frac{1}{\sqrt{n\epsilon \log \frac{1}{\epsilon}}} \right| = O\left(\frac{1}{\sqrt{n\epsilon \log \frac{1}{\epsilon}}}\right)$$

Аналогично для $t = X_{(k)} + \epsilon, s = X_{(k)}$:

$$\sqrt{n} \frac{|\delta_n(t) - \delta_n(s)|}{\sqrt{|t - s| \log |t - s|^{-1}}} = \sqrt{n} \sqrt{\frac{\epsilon}{\log \frac{1}{\epsilon}}} \rightarrow 0, \epsilon \rightarrow 0$$

Для $t = X_{(k)} + \frac{\epsilon}{2}, s = X_{(k)} - \frac{\epsilon}{2}$ получаем аналогично, как в первом случае:

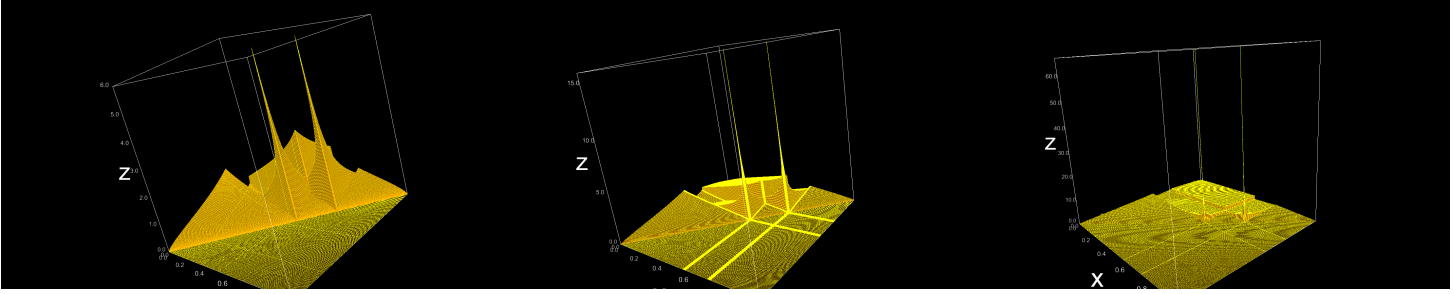
$$\sqrt{n} \frac{|\delta_n(t) - \delta_n(s)|}{\sqrt{|t - s| \log |t - s|^{-1}}} = \left| \frac{\sqrt{n\epsilon}}{\sqrt{\log \frac{1}{\epsilon}}} - \frac{1}{\sqrt{n\epsilon \log \frac{1}{\epsilon}}} \right|$$

Чтобы данное выражение не уходило в бесконечность (в силу произвольно малого выбора ϵ), а было ограниченным сверху, необходимо отделить область, где возникает уход на бесконечность. То есть необходимо рассмотреть область $|t - s| > \epsilon = \epsilon_n$, при условии, что $\epsilon_n \log \frac{1}{\epsilon_n} \geq O(\frac{1}{n})$.

Данное замечание доказывает пункт 2 теоремы 3 (формулу (6)). \square

Разрешив данное неравенство, приходим к оценке (которая является необходимой, но не достаточной для того, чтобы данное выражение было $O_P(1)$): $\epsilon_n \geq \frac{1}{n \log n}$.

На Рисунке ниже приведен график функции $\sqrt{n} \frac{|\delta_n(t) - \delta_n(s)|}{\sqrt{|t-s| \log |t-s|^{-1}}}$ для $n = 2$. В нём отчетливо видны пики, уходящие на бесконечность на границе $|t - s| < \epsilon$ при уменьшении мелкости разбиения сетки (s, t) :



Доказательство Теоремы 2:

Проведя доказательство, полностью аналогично доказательству теоремы 3.1, получаем, что исходное выражение будет сходиться по распределению к:

$$\sup_{t \in O_T^\delta} \frac{\sqrt{(B_1(T) - B_1(t))^2 + (B_2(T) - B_2(t))^2 + (B_3(T) - B_3(t))^2 + (B_4(T) - B_4(t))^2}}{\sqrt{|T - t|}}$$

где B_i - соответствующий i -ой координате вектора $\sqrt{n} \frac{(s_n(t) - s(t))}{\sqrt{T-t}}$ Броуновский мост.

Теорема о непрерывности Леви гласит, что выполнено равенство почти наверное:

$$\lim_{h \rightarrow 0} \sup_{\substack{0 \leq s < t \leq 1 \\ h = t - s}} \frac{|W(t) - W(s)|}{\sqrt{|h| \log |h|^{-1}}} = \sqrt{2}$$

Следовательно по теореме Леви: $\forall \epsilon > 0$ для почти всех $\omega \exists h(\epsilon, \omega) = h_0$ т.ч. $\forall |t - T| < h_0$:

$$\begin{aligned} \frac{|W(T) - W(t)|}{\sqrt{|T - t| \log(\frac{1}{|T - t|})}} &\geq \sqrt{2}(1 - \epsilon) \\ \frac{|W(T) - W(t)|}{\sqrt{|T - t|}} &\geq \sqrt{2}(1 - \epsilon) \log(\frac{1}{|T - t|}) \\ \sup_{|t - T| < h_0} \frac{|W(T) - W(t)|}{\sqrt{|T - t|}} &\geq \sup_{|t - T| < h_0} \sqrt{2}(1 - \epsilon) \log(\frac{1}{|T - t|}) = \infty \end{aligned}$$

Заметим, что для любого Броуновского моста выполнено:

$$\sup_{t \in O_T^\delta} \frac{|B(T) - B(t)|}{\sqrt{|T - t|}} \geq \sup_{t \in O_T^\delta} \frac{|W(T) - W(t)|}{\sqrt{|T - t|}} + O_P(1) \quad (8)$$

Кроме того, для любых случайных величин $\xi_1, \xi_2, \xi_3, \xi_4$, всегда справедлива оценка:

$$\sqrt{\xi_1^2 + \xi_2^2 + \xi_3^2 + \xi_4^2} \geq |\xi_i|, \quad i = 1, 2, 3, 4 \quad (9)$$

Следовательно, если хотя бы одна из $\xi_i \neq O_P(1)$, то и весь корень из суммы квадратов не является ограниченной по вероятности величиной (лемма 2.1).

Неравенства (8), (9) и лемма 2.1 завершает доказательство теоремы 2. \square

Заметим, что условие теоремы 2 можно немного ослабить: если для хотя бы одной функции из четырёх функций распределения вектора $s(t)$ выполнено условие: существует $c \in \mathbf{R}$ т.ч. $|T - F^{-1}(y)| \leq c|T - y|$, то утверждение теоремы 2 остается справедливым. Так как у вектора $\sqrt{n} \frac{(s_n(t) - s(t))}{\sqrt{T-t}}$ соответствующая координата будет неограниченной по вероятности. Следовательно, по неравенству (9), стандартная евклидова норма данного вектора не является ограниченной по вероятности.

К сожалению, данный метод доказательства, не способен дать ответа на вопрос $|T - T_n| = O_P(n^{-\frac{1}{3}})$, так как доказаны невыполнимость условия следствия 1. Это возникает из-за слишком грубых оценок, возникающих в процессе доказательства, а именно: утверждение 1 и леммы 3.

4 Альтернативный подход и компьютерное моделирование

Альтернативным подходом к исследованию проблемы о скорости сходимости является линеаризация исходной задачи. Заметим, что функционал $Z(t)$ является некоторой гладкой функцией от вероятностей правдоподобия классов:

$$Z(t) = Q(F^+(t), F^-(t))$$

$$Z_n(t) = Q(F_n^+(t), F_n^-(t))$$

Разложим по Тейлору функцию $Q(F_n^+(t), F_n^-(t))$ в окрестности точки $(F^+(T), F^-(T))$:

$$\begin{aligned} Q(F_n^+(t), F_n^-(t)) &= Q(F^+(T), F^-(T)) + A(F_n^+(t) - F^+(T)) + B(F_n^-(t) - F^-(T)) + \\ &+ \frac{1}{2} \frac{\partial^2 Q}{\partial F^+ \partial F^-}(F^+(T), F^-(T)) (F_n^+(t) - F^+(T))^2 + \dots \end{aligned}$$

$$A = \frac{\partial Q}{\partial F^+}(F^+(T), F^-(T))$$

$$B = \frac{\partial Q}{\partial F^-}(F^+(T), F^-(T))$$

Обозначим T_n^{Loc} - точку минимума линеаризованного функционала $Q(F_n^+(t), F_n^-(t))$:

$$\begin{aligned} T_n^{Loc} &= \arg \min (Q(F^+(T), F^-(T)) + A(F_n^+(t) - F^+(T)) + B(F_n^-(t) - F^-(T)) = \\ &= \arg \min (AF_n^+(t) + BF_n^-(t)) \end{aligned}$$

Заметим, что T является точкой минимума функционала Q , следовательно выполняется равенство:

$$\left. \frac{dQ(F^+(t), F^-(t))}{dt} \right|_{t=T} = \left[\frac{\partial Q}{\partial F^+} \dot{F}^+(t) + \frac{\partial Q}{\partial F^-} \dot{F}^-(t) \right]_{t=T} = 0 \quad (9)$$

Предположим, что выполнены все достаточные условия, следовательно, точки минимумов функционалов $[AF^+(t) + BF^-(t)]$ и $Q(F^+(t), F^-(t))$ совпадают, значит:

$$T = \arg \min (AF^+(t) + BF^-(t))$$

Также заметим, что из равенства (9) следует, что константы A и B обязаны быть разных знаков.

Гипотеза 1. Гипотеза о том, что $|T - T_n| = O_P(n^{-\frac{1}{3}})$ разбивается на две части:

1. Нужно показать, что

$$|T - T_n^{Loc}| = O_P(n^{-\frac{1}{3}})$$

2. Нужно показать, что

$$|T_n - T_n^{Loc}| = O_P(n^{-\frac{1}{2}})$$

Компьютерная симуляция:

Задавая распределения $F^+(t), F^-(t), \pi^+, \pi^-$, split функцию $S(x)$, будем рассматривать сходимость параметров T, T_n, T_n^{Loc} .

Весь код был написан на языке программирования *Python* с использованием библиотек *numpy, scipy*.

Напомню, что функционал $Z(t)$ имеет вид:

$$Z(t) = S(q^{\leq}(t))P(X_k \leq t) + S(q^>(t))P(X > t)$$

$$q^{\leq}(t) = \begin{cases} P(l = 1 | X \leq t), & P(X \leq t) \neq 0 \\ 0, & P(X \leq t) = 0 \end{cases} \quad q^>(t) = \begin{cases} P(l = 1 | X > t), & P(X > t) \neq 0 \\ 0, & P(X > t) = 0 \end{cases}$$

Основные функции кода из Jupyter Notebook:

```
1 def F_plus(x, type_ = 'R', E = 1.0, D = 1.0):
2     #Функция правдоподобия классов с метками +1
3     if type_ == 'R':
4         return x
5     elif type_ == 'exp':
6         return 1-np.exp(-E*x)
7     elif type_ == 'N':
8         return norm.cdf(x, loc = E, scale = D)
9
10 def F_minus(x, type_ = 'R', E = 1.0, D = 1.0):
11     #Функция правдоподобия классов с метками -1
12     if type_ == 'R':
13         return x
14     elif type_ == 'exp':
15         return 1-np.exp(-E*x)
16     elif type_ == 'N':
17         return norm.cdf(x, loc = E, scale = D)
18
19 def pi(l, p = 1./2):
20     #Априорные вероятности классов
21     q = 1.0 - p
22     if l == 1:
23         return p
24     else:
25         return q
26
27 def S(x, type_ = 'коэффициент Бхаттачарья'):
28     """
29     Split function - сплит функция
30     input x: float64
31     return S(X): float64
32     """
33     if type_ == 'коэффициент Бхаттачарья':
34         x = x*(1-x)
35         return np.sqrt(x)
36     elif type_ == 'gini':
37         return 1 - (x**2 + (1-x)**2)
38     elif type_ == 'entropy':
39         return -(x*np.log2(x) + (1-x)*np.log2(1-x))/2
40
41 def H(x, type_plus, E_plus, D_plus, type_minus, E_minus, D_minus, p_l):
42     """
43     Функция построения теоретического функционала H
44     input: (type_plus, E_plus, D_plus) - тип и параметры для функции прав
45     доподобия класса +1
46     (type_minus, E_minus, D_minus)- тип и параметры для функции прав
47     доподобия класса -1
48     return H(x): float64
49     """
50     F_plus_values = F_plus(x, type_ = type_plus, E = E_plus, D = D_plus)
51     F_minus_values = F_minus(x, type_ = type_minus, E = E_minus, D =
52     D_minus)
53
54     pi_plus, pi_minus = pi(l = 1, p = p_l), pi(l = -1, p = p_l)
```

```

53 P_X_less = F_minus_values*pi_minus + F_plus_values*pi_plus
54 P_X_more = 1. - P_X_less
55 q_less = F_plus_values*pi_plus/P_X_less
56 q_more = (1. - F_plus_values)*pi_plus/P_X_more
57 return S(q_less)*P_X_less + S(q_more)*P_X_more
58
59 def H_n(t, values_plus, values_minus):
60     """
61     Функция построения эмпирического функционала H_n по выборке размера n
62     input: t - отсортированный массив статистик X_k
63             values_plus - отсортированный массив статистик X_k с метками +1
64             values_minus - отсортированный массив статистик X_k с метками -1
65     return: H_n_values: np.array(type: float64)
66     """
67     pi_plus_n = len(values_plus)/(len(values_plus) + len(values_minus))
68     pi_minus_n = 1 - pi_plus_n
69
70     Count_X_less_t_plus = np.searchsorted(values_plus, t, side='right')
71     Count_X_more_t_plus = len(values_plus) - Count_X_less_t_plus
72
73     Count_X_less_t = Count_X_less_t_plus + np.searchsorted(values_minus, t,
74     side='right')
75     Count_X_more_t = len(values_plus) + len(values_minus) - Count_X_less_t
76
77     q_less_n = Count_X_less_t_plus/Count_X_less_t
78     q_more_n = Count_X_more_t_plus/Count_X_more_t
79
80     P_X_less_n = Count_X_less_t/(len(values_plus) + len(values_minus))
81     P_X_more_n = 1.0 - P_X_less_n
82
83     return S(q_less_n)*P_X_less_n + S(q_more_n)*P_X_more_n
84
85 def Argmin_H(args, H, bounds):
86     """
87     Поиск аргминимума функционала H(x)
88     input:
89         args = (param1, param2, ...) - параметры для генерации функционала
90         т.е:
91             (type_plus, E_plus, D_plus, type_minus, E_minus, D_minus,
92             p_l)
93         H - функционал по которому ищем минимум
94     return argmin(H): float64
95     """
96     # Начальное приближение для оптимизации
97     initial_guess = (bounds[0][0] + bounds[0][1])/2
98     # Поиск минимума функции H(x)
99     result = minimize(H, initial_guess, args=args, method='Nelder-Mead',
100     bounds = bounds)
101     # Вывод результата
102     if result.success:
103         print("Минимум функции найден успешно!")
104         print("Аргминимум:", result.x)
105         print("Значение функции в аргминимуме:", result.fun)
106         return result
107     else:
108         print("Не удалось найти минимум функции.")

```

```

106 #Константы, зависящие от split функции. Представлены для  $S = \sqrt{x(1-x)}$ 
107 def A_f(F_plus_T, F_minus_T, pi_plus, pi_minus):
108     """
109     input: (F_plus_T, F_minus_T) - функции в точке минимума T
110            (pi_plus, pi_minus) - вероятности классов
111     return A: float64
112     """
113     I1 = pi_plus*pi_minus*F_minus_T/(2*np.sqrt(pi_plus*pi_minus*F_plus_T*
114     F_minus_T))
115     I2 = pi_plus*pi_minus*(1-F_minus_T)/(2*np.sqrt(pi_plus*pi_minus*(1-
116     F_plus_T)*(1-F_minus_T)))
117     return I1 - I2
118
119 def B_f(F_plus_T, F_minus_T, pi_plus, pi_minus):
120     I1 = pi_plus*pi_minus*F_plus_T/(2*np.sqrt(pi_plus*pi_minus*F_plus_T*
121     F_minus_T))
122     I2 = pi_plus*pi_minus*(1-F_plus_T)/(2*np.sqrt(pi_plus*pi_minus*(1-
123     F_plus_T)*(1-F_minus_T)))
124     return I1 - I2
125
126 #Линеаризованный функционал
127 def H_linear(x, type_plus, E_plus, D_plus, type_minus, E_minus, D_minus,
128 p_l):
129     """
130     Функция построения теоретического линеаризованного функционала H
131     input: (type_plus, E_plus, D_plus) - тип и параметры для функции прав
132            доподобия класса +1
133            (type_minus, E_minus, D_minus)- тип и параметры для функции прав
134            доподобия класса -1
135     return H_linear(x): float64
136     """
137     F_plus_values = F_plus(x, type_ = type_plus, E = E_plus, D = D_plus)
138     F_minus_values = F_minus(x, type_ = type_minus, E = E_minus, D =
139     D_minus)
140
141     pi_plus, pi_minus = pi(l = 1, p = p_l), pi(l = -1, p = p_l)
142
143     return F_plus_values*A + F_minus_values*B
144
145 #Эмпирический линеаризованный функционал
146 def H_n_linear(t, values_plus, values_minus):
147     """
148     Функция построения эмпирического линеаризованного функционала
149     H_n_linear
150     input: t - отсортированный массив статистик X_k
151            values_plus - отсортированный массив статистик X_k с метками +1
152            values_minus - отсортированный массив статистик X_k с метками -1
153     return: H_n_linear_values: np.array(type: float64)
154     """
155     pi_plus_n = len(values_plus)/(len(values_plus) + len(values_minus))
156     pi_minus_n = 1 - pi_plus_n
157
158     Count_X_less_t_plus = np.searchsorted(values_plus, t, side='right')
159     Count_X_less_t_minus = np.searchsorted(values_minus, t, side='right')
160     H_n_linear_values = A*Count_X_less_t_plus/len(values_plus) + B*
161     Count_X_less_t_minus/len(values_minus)
162     return H_n_linear_values

```

```

153 # Функция поиска T_n и T_n_Loc
154 def Search_T_n_T_n_linear(n, T_n_values, T_n_linear_values, simulations =
155     1, type_ = 'N'):
156     """
157     input: n - число объектов в выборке
158             T_n_values: list, T_n_linear_values: list
159                 списки значений, куда будут складываться результаты
160             simulations - число симуляционных запусков
161                 (чем больше этот параметр, тем точнее T_n
162             type_ - F(x) правдоподобие классов
163     return T_n_values, T_n_linear_values
164     """
165     p_l = 0.5
166     if type_ == "N":
167         args = ('N', 0.0, 1., 'N', 1, 2.0, p_l)
168         T = N_T[0]
169     elif type_ == "exp":
170         args = ('exp', 1.0, None, 'exp', 2.0, None, p_l)
171         T = exp_T[0]
172
173     F_plus_T = F_plus(T, type_ = args[0], E = args[1], D = args[2])
174     F_minus_T = F_minus(T, type_ = args[3], E = args[4], D = args[5])
175
176     A = A_f(F_plus_T, F_minus_T,
177         pi_plus=args[6], pi_minus = 1 - args[6]
178     )
179     B = B_f(F_plus_T, F_minus_T,
180         pi_plus=args[6], pi_minus = 1 - args[6]
181     )
182
183     T_n_values_n = []
184     T_n_linear_values_n = []
185
186     for simulation in range(simulations):
187         if simulations!=1:
188             np.random.seed(simulation)
189         else:
190             np.random.seed(42)
191         values_plus = []
192         values_minus = []
193         for i in range(n):
194             p = np.random.rand()
195             if type_ == "N":
196                 if p<=p_l:
197                     values_plus.append(np.random.normal(loc = args[1],
198 scale = args[2]))
199                 else:
200                     values_minus.append(np.random.normal(loc = args[4],
201 scale = args[5]))
202             elif type_ == 'exp':
203                 if p<=p_l:
204                     values_plus.append(np.random.exponential(scale = args
205 [1]))
206                 else:
207                     values_minus.append(np.random.exponential(scale = args
208 [4]))

```



```

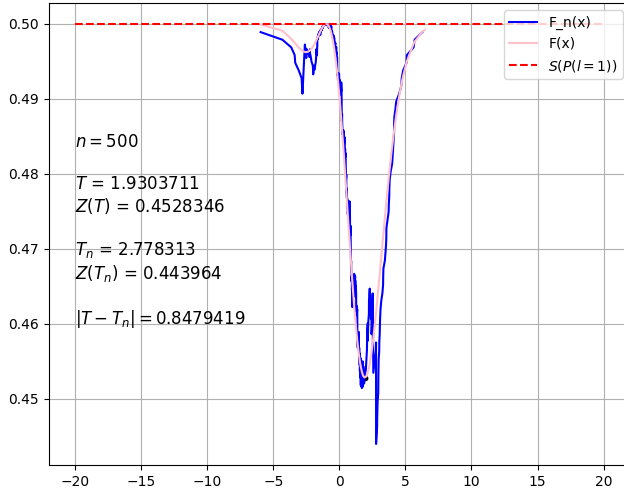
205     values_plus = np.sort(values_plus)
206     values_minus = np.sort(values_minus)
207
208
209     t = np.sort(np.hstack((values_plus, values_minus)))
210     H_n_linear_values = H_n_linear(t, values_plus, values_minus, A, B)
211     nan_indices = np.isnan(H_n_linear_values)
212     T_n_linear = t[np.argmin(H_n_linear_values[~nan_indices])]
213
214
215     H_n_values = H_n(t, values_plus, values_minus)
216     nan_indices = np.isnan(H_n_values)
217     T_n = t[np.argmin(H_n_values[~nan_indices])]
218
219     T_n_values_n.append(T_n)
220     T_n_linear_values_n.append(T_n_linear)
221
222     return np.mean(T_n_values_n), np.mean(T_n_linear_values_n)
223
224 def Search_on_setka_T_n_T_n_linear(setka = None)
225     #Считает T_n, T_n_Linear по сетке (числа объектов выборке)
226     #Функция на следующем шаге догенерирует точки до размера текущего шага
227     #Algorithm complexy: Memory = O(n), Time: O(nlog(n))
228     return: T_n_values: list, T_n_linear_values: list
229     T_n_values = []
230     T_n_linear_values = []
231
232     setka1 = [10*i for i in range(1, 11)]
233     setka2 = [1000*i for i in range(1, 11)]
234     setka3 = [100000*i for i in range(1, 11)]
235     setka4 = [1000000*i for i in range(1, 11)]
236     setka5 = [10000000*i for i in range(1, 11)]
237     setka6 = [10000000*(10 + i) for i in range(1, 11)]
238     if setka is None:
239         setka = setka1 + setka2 + setka3 + setka4 + setka5 + setka6
240
241     for nn in tqdm(setka):
242         T_n, T_n_linear = Search_T_n_T_n_linear(nn, T_n_values,
243         T_n_linear_values, simulations=1, type_ = 'N')
244         T_n_values.append(T_n)
245         T_n_linear_values.append(T_n_linear)
246     return T_n_values, T_n_linear_values

```

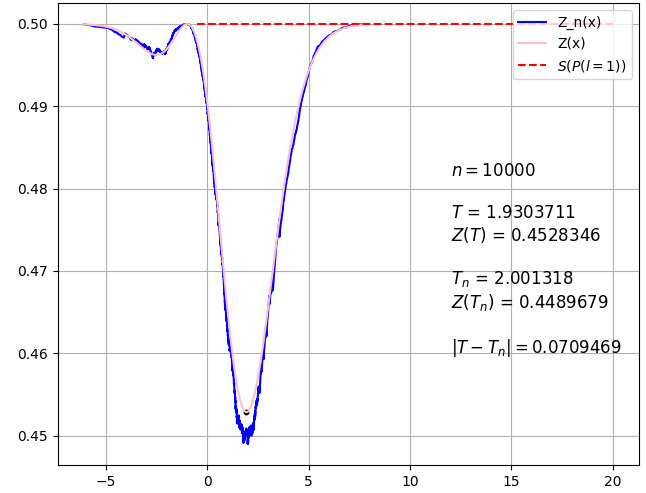
Все примеры отрисовки графиков и вычислений лежат в репозитории на GitHub: [GitHub репозиторий](#)

Ниже приведены графики функционалов $Z(t)$, $Z_n(t)$ и их точек минимумов в зависимости от размера выборки n :

$$F^+(t) \sim N(0, 1), F^-(t) \sim N(1, 2), \pi^+ = \pi^- = \frac{1}{2}, S(x) = \sqrt{x(1-x)}$$

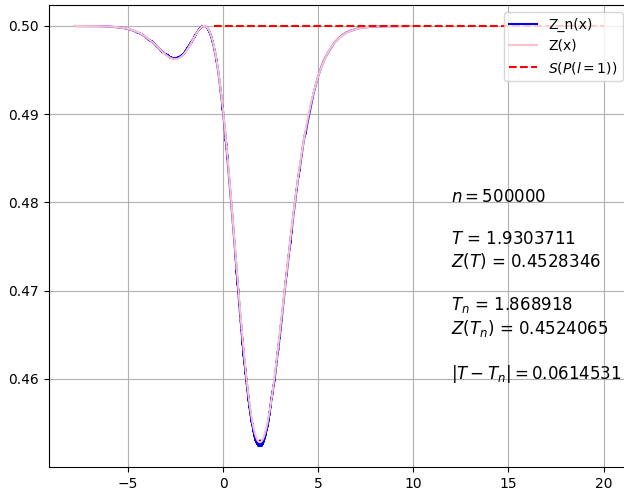


$n = 500$

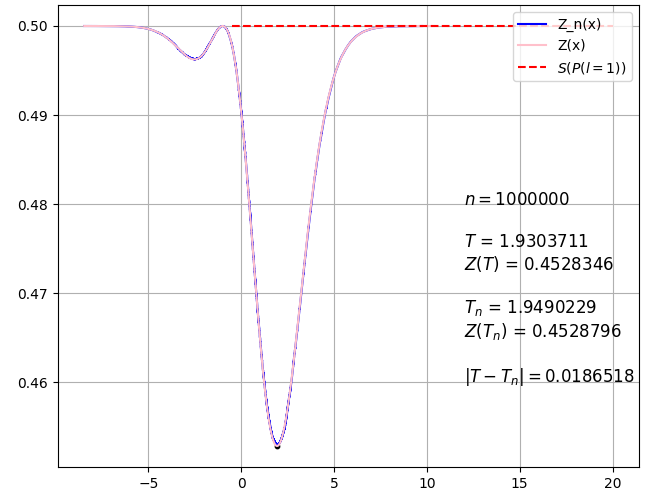


$n = 10000$

Сходимость функционала $Z_n(t)$ к $Z(t)$

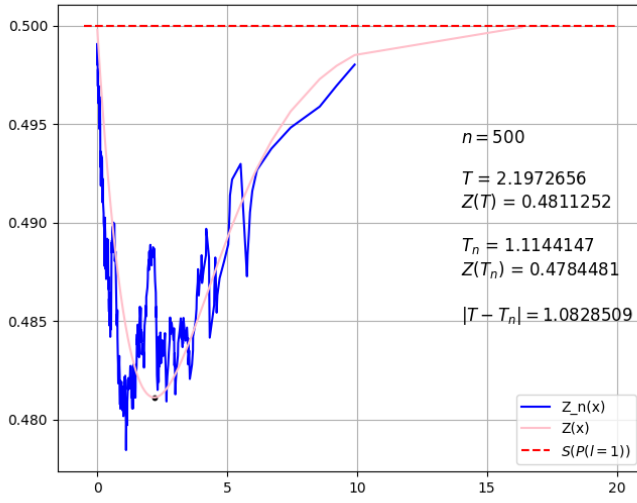


$n = 500000$

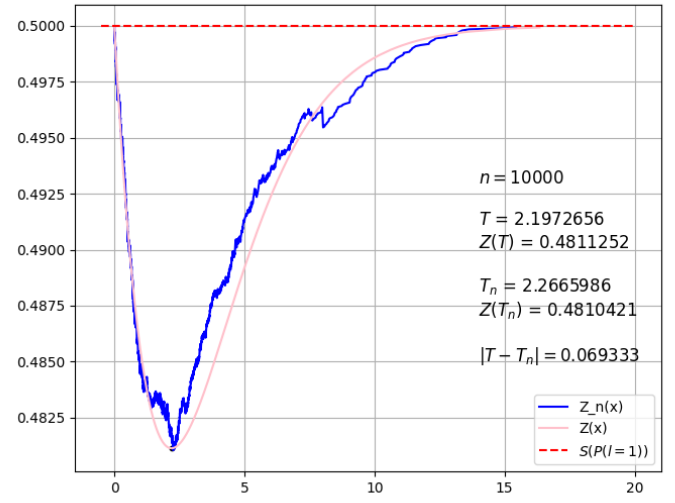


$n = 1000000$

$$F^+(t) \sim \exp(1), F^-(t) \sim \exp(2), \pi^+ = \pi^- = \frac{1}{2}, S(x) = \sqrt{x(1-x)}$$

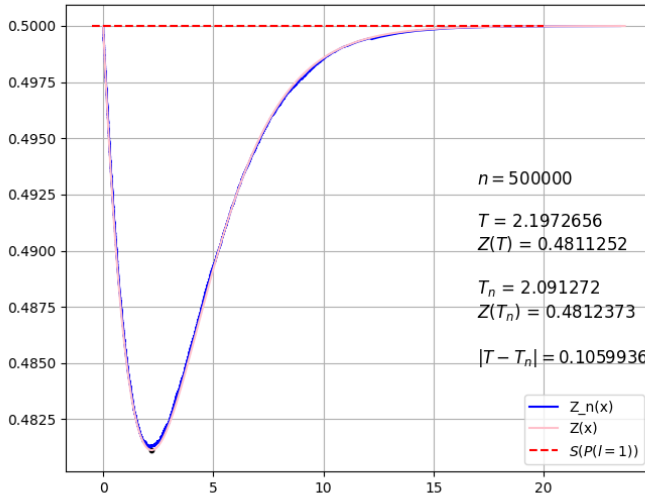


$n = 500$

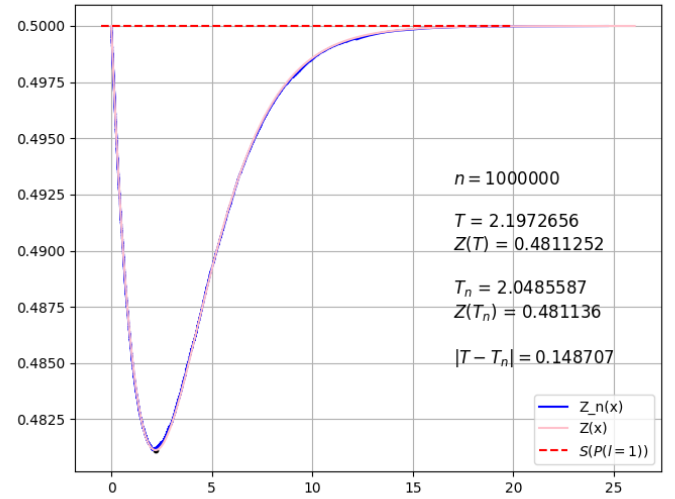


$n = 10000$

Сходимость функционала $Z_n(t)$ к $Z(t)$



$n = 500000$



$n = 1000000$

Построение $|T - T_n|$: Рис (6), Рис (7)

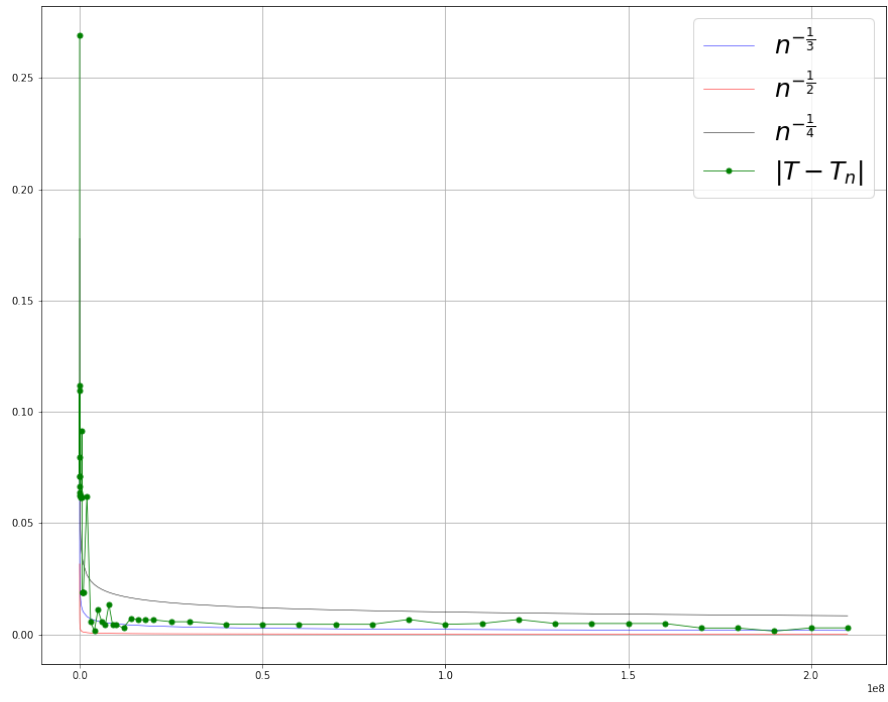


Рис. 6: $[F^+, F^-] \sim [N(0, 1), N(1, 2)]$

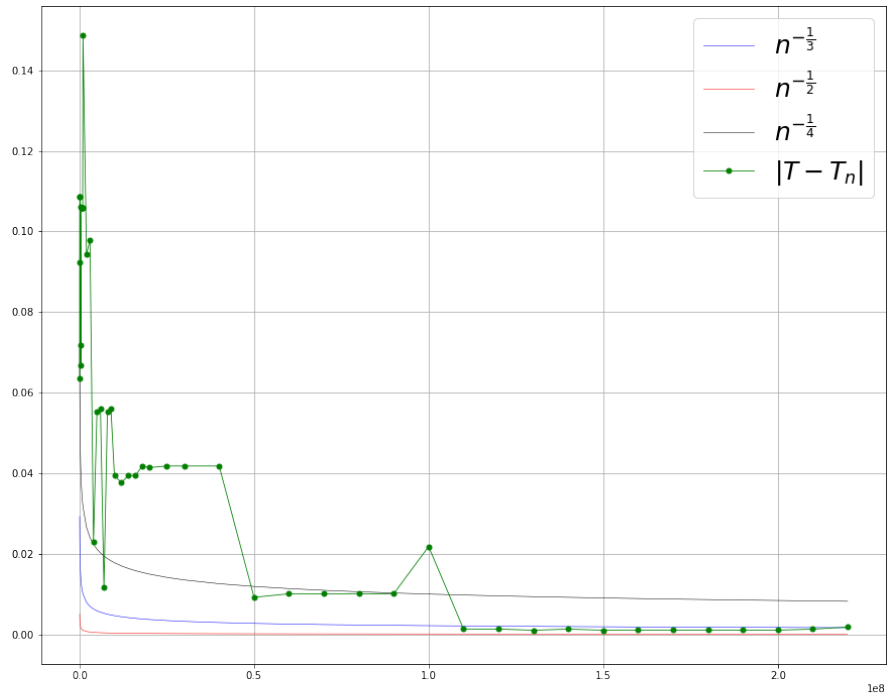


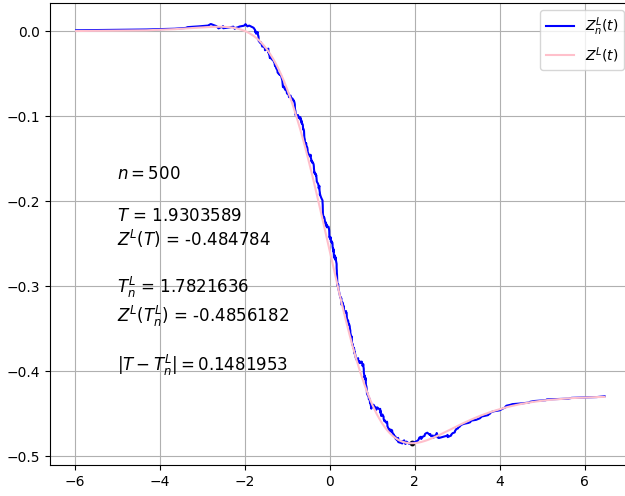
Рис. 7: $[F^+, F^-] \sim [\exp(1), \exp(2)]$

Ниже приведены графики линейризованных функционалов $Z^{Loc}(t)$, $Z_n^{Loc}(t)$ и их точек минимумов в зависимости от размера выборки n :

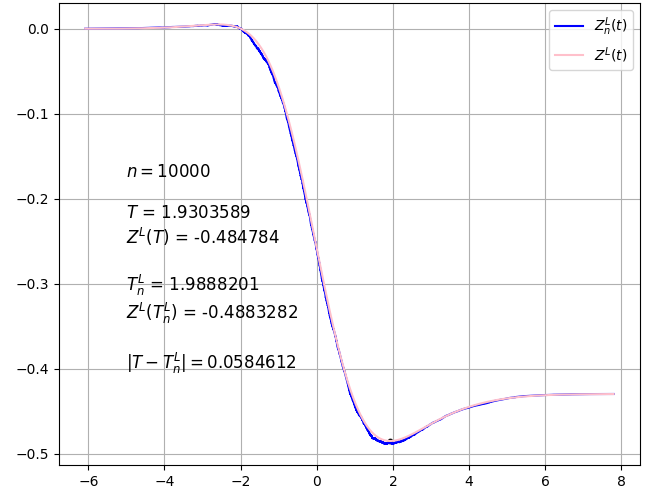
$$A = \frac{\partial Q}{\partial F^+}(F^+(T), F^-(T)) = \left[\frac{1}{2} \sqrt{\frac{\pi^+ F^- \pi^-}{F^+}} - \frac{1}{2} \sqrt{\frac{\pi^+(1-F^-)\pi^-}{(1-F^+)}} \right] \Big|_{t=T}$$

$$B = \frac{\partial Q}{\partial F^-}(F^+(T), F^-(T)) = \left[\frac{1}{2} \sqrt{\frac{\pi^+ F^+ \pi^-}{F^-}} - \frac{1}{2} \sqrt{\frac{\pi^+(1-F^+)\pi^-}{(1-F^-)}} \right] \Big|_{t=T}$$

$$F^+(t) \sim N(0, 1), \quad F^-(t) \sim N(1, 2), \quad \pi^+ = \pi^- = \frac{1}{2}, \quad S(x) = \sqrt{x(1-x)}$$

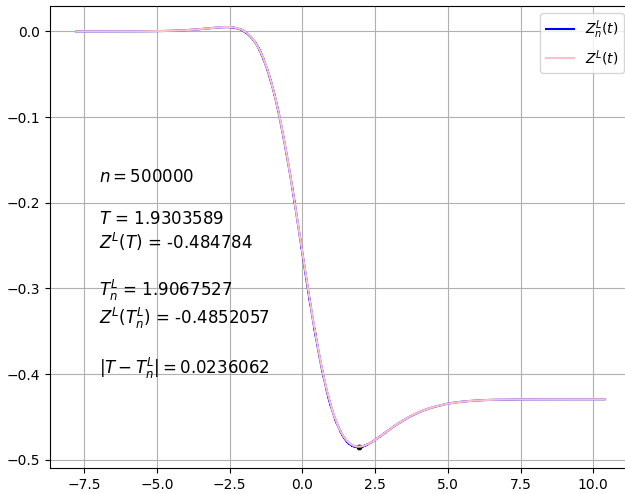


$n = 500$

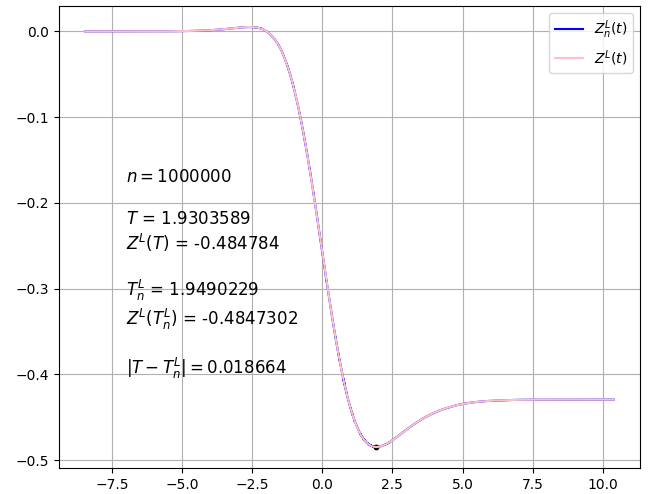


$n = 10000$

Сходимость функционала $Z_n^L(t)$ к $Z^L(t)$
 $(A, B) = (-0.656, 0.227)$

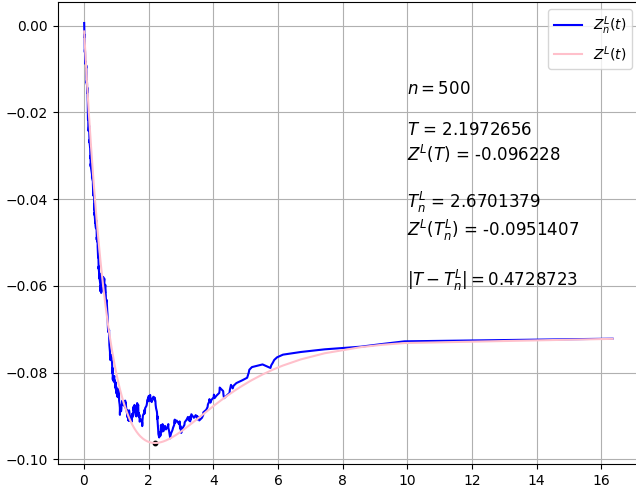


$n = 500000$

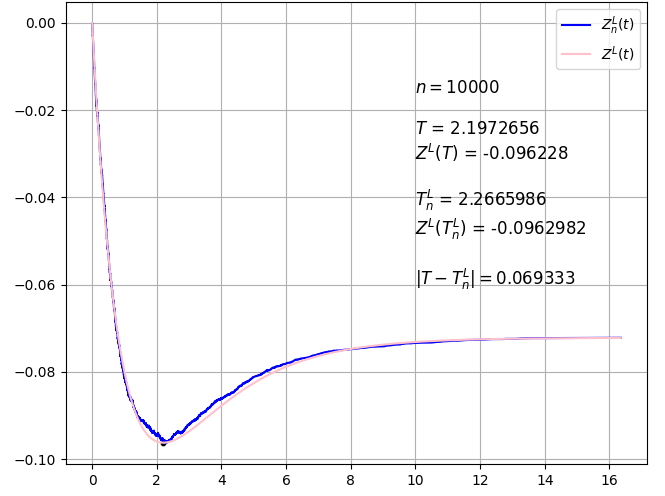


$n = 1000000$

$$F^+(t) \sim \exp(1), F^-(t) \sim \exp(2), \pi^+ = \pi^- = \frac{1}{2}, S(x) = \sqrt{x(1-x)}$$

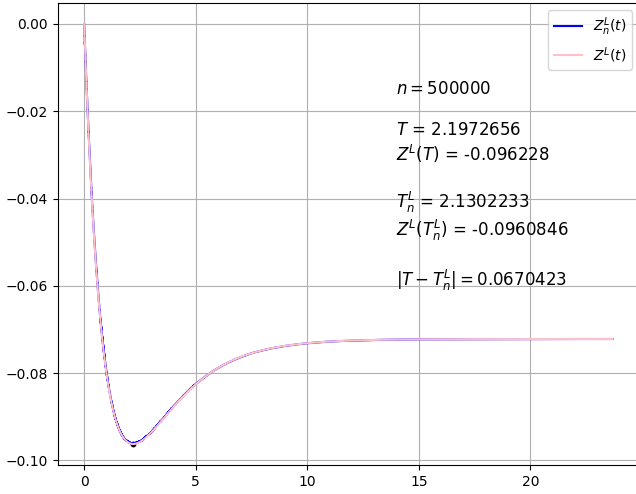


$n = 500$

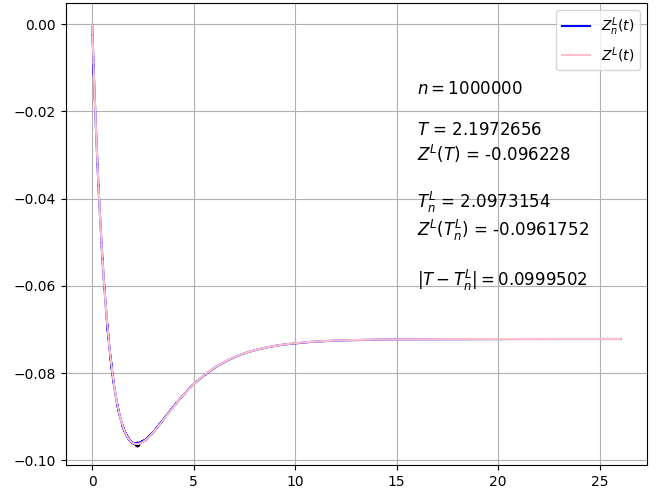


$n = 10000$

Сходимость функционала $Z_n^L(t)$ к $Z^L(t)$
 $(A, B) = (-0.216, 0.144)$



$n = 500000$



$n = 1000000$

Построение $|T - T_n^{Loc}|$: Рис (12), Рис (13)

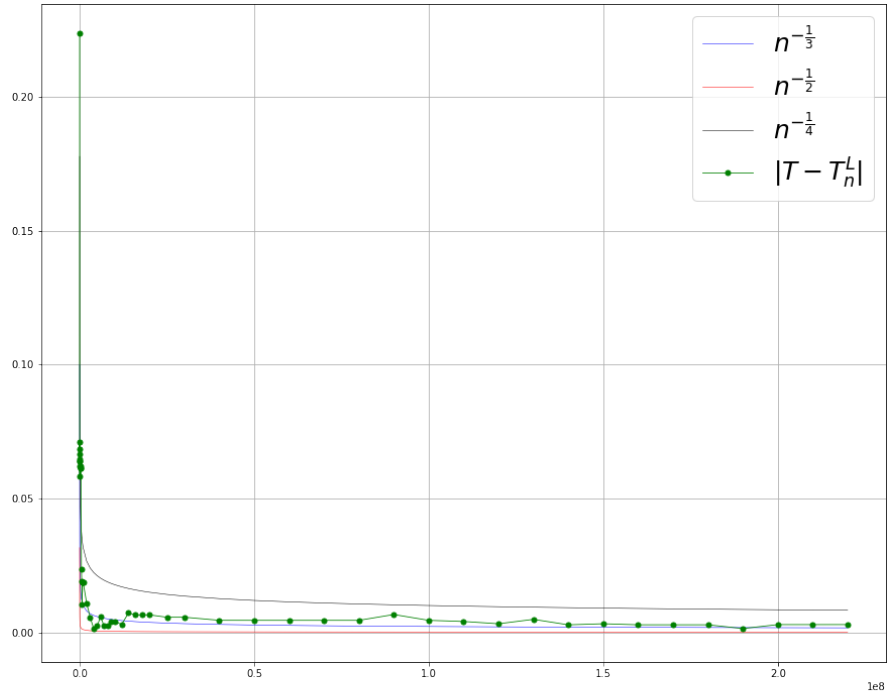


Рис. 12: $[F^+, F^-] \sim [N(0, 1), N(1, 2)]$

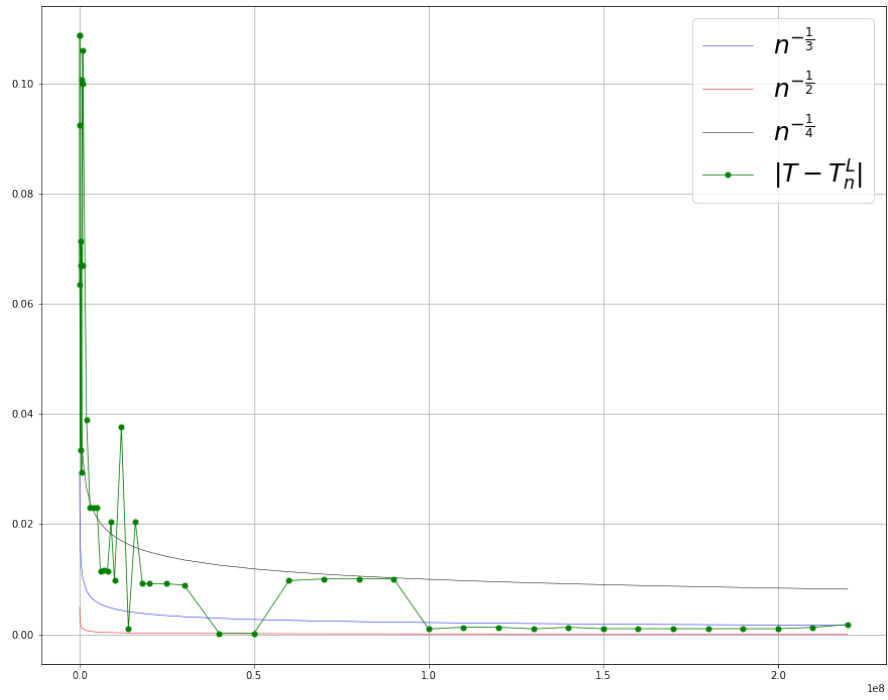


Рис. 13: $[F^+, F^-] \sim [exp(1), exp(2)]$

Построение $T_n - T_n^{Loc}$: Рис (14), Рис (16)

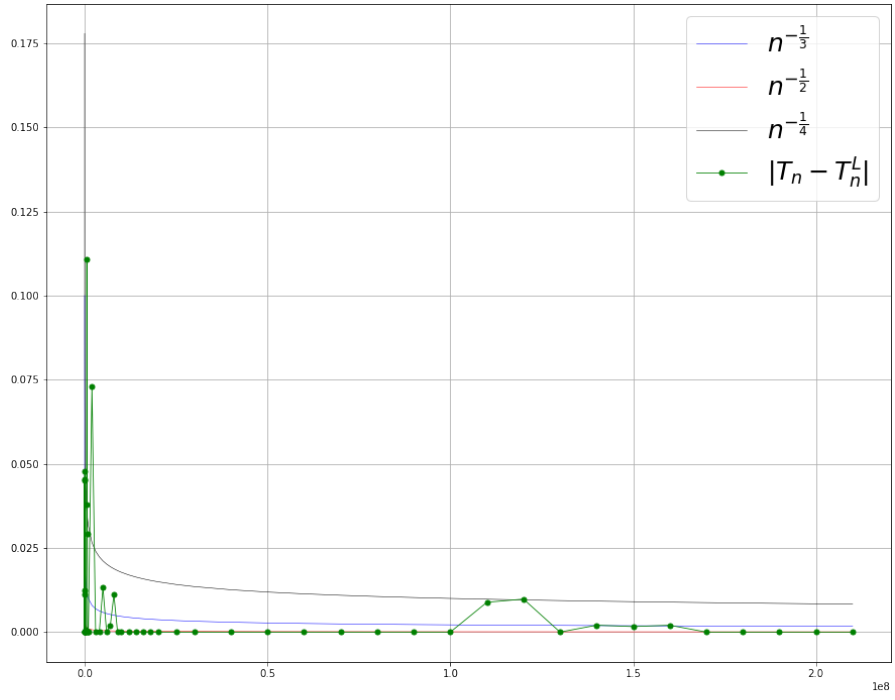


Рис. 14: $[F^+, F^-] \sim [N(0, 1), N(1, 2)]$

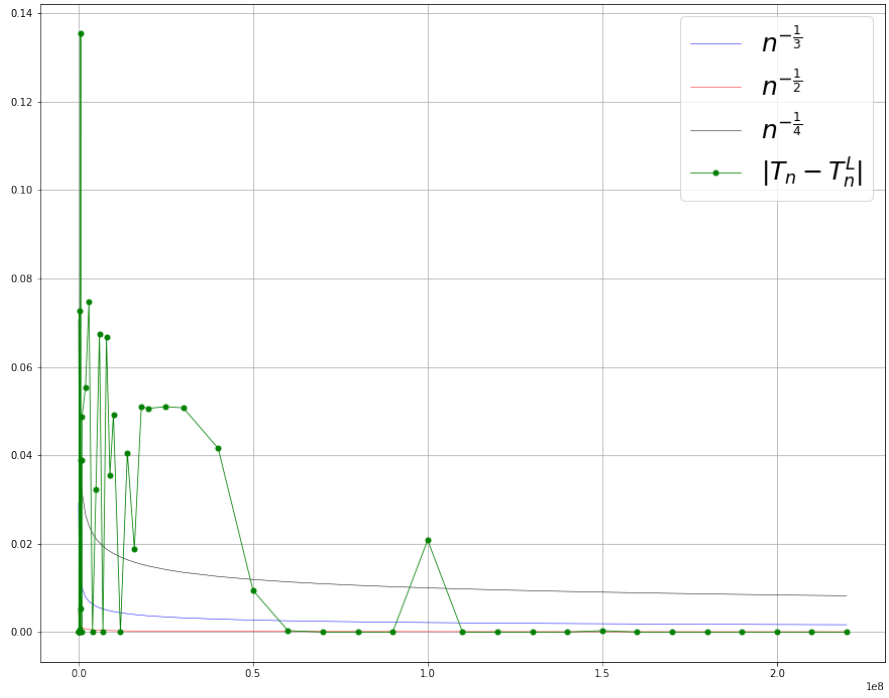


Рис. 15: $[F^+, F^-] \sim [exp(1), exp(2)]$

Данные графики экспериментально подтверждают гипотезу о скорости сходимости $n^{-\frac{1}{3}}$ для $|T - T_n|$ и $|T - T_n^{Loc}|$ и скорости сходимости $n^{-\frac{1}{2}}$ для $|T_n - T_n^{Loc}|$. Константы M_ϵ , возникающие в определении $O_P(1)$ для $|T - T_n|$, при достаточно малом ϵ , необходимым для подтверждения скорости являются величинами порядка 10^8 для данных параметров генерации.

5 Заключение

В дипломной работе был представлен первый подход к исследованию скорости сходимости разделяющей границы T_n - минимума эмпирического функционала Z_n в дереве решений глубины равной 1. Самый первый подход был основан на оценках сверху величины $|T - T_n|$, вытекающих из утверждения 1 и леммы 3. Однако детальное исследование величин, ограничивающих сверху $|T - T_n|$, продемонстрировало их неограниченность по вероятности, что свидетельствует о слишком грубых неравенствах.

Также был предложен новый подход к исследованию задачи, однако в нем пока не было достигнуто значимых теоретических результатов, но его изучение может быть ключом к решению задачи.

Кроме того была написана и проведена компьютерная симуляция оценок величин $|T - T_n|$, $|T - T_n^{Loc}|$, $|T_n - T_n^{Loc}|$, подтверждающая гипотезу о скорости сходимости $n^{-\frac{1}{3}}$ для $|T - T_n|$, $|T - T_n^{Loc}|$, а также $n^{-\frac{1}{2}}$ для величины $|T_n - T_n^{Loc}|$.

Список литературы

- [1] Ширяев А.Н., Вероятность-1, 2. МЦНМО, 2011
- [2] Булинский А.В., Ширяев А.Н., Теория случайных процессов. Физматлит, 2005
- [3] Mohammadi L., van de Geer S. On threshold-based classification rules //Lecture Notes-Monograph Series. – 2003. – С. 261-280.
- [4] Kim J., Pollard D. Cube root asymptotics //The Annals of Statistics. – 1990. – С. 191-219.
- [5] Seo M., Otsu T. Asymptotics for maximum score method under general conditions. – LSE Working Paper, 2015..
- [6] Zhou Y., Zhou Z., Hooker G. Approximation trees: Statistical stability in model distillation //arXiv preprint arXiv:1808.07573. – 2018.
- [7] Peng W., Coleman T., Mentch L. Asymptotic distributions and rates of convergence for random forests via generalized U-statistics //arXiv preprint arXiv:1905.10651. – 2019.
- [8] Konstantin Vorontsov, Andrey Ivahnenko, Tight Combinatorial Generalization Bounds for Threshold Conjunction Rules. 4th International Conference on Pattern Recognition and Machine Intelligence, Moscow, 2011.
- [9] Ishwaran H. The effect of splitting on random forests //Machine learning. – 2015. – Т. 99. – С. 75-118.