

Министерство науки и высшего образования Российской Федерации

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ "МЭИ"

Институт информационных и вычислительных технологий

Кафедра математического и компьютерного моделирования

Отчёт по лабораторной работе №1
"Пределные теоремы"

Студент: Симаков А.М.
Преподаватель: Шевченко О.В.

Москва 2023

Цель работы: статистически пронаблюдать существо основных предельных теорем.

1 Теорема Бернулли

Если проводится n независимых испытаний случайного события A , такого что $P(A) = p$, то относительная частота $\frac{\mu}{n}$ появления события A (μ - число появлений A) при большом n приближенно равна вероятности p :

$$\frac{\mu}{n} \approx p$$

Будем писать $\frac{\mu}{n} \rightarrow p$ при $n \rightarrow \infty$, если

$$\forall \varepsilon > 0 \quad P \left\{ \left| \frac{\mu}{n} - p \right| < \varepsilon \right\} \rightarrow 1 \text{ при } n \rightarrow \infty$$

Заметим, что теорема не говорит, что соотношение $\left| \frac{\mu}{n} - p \right| < \varepsilon$ достоверно, однако если n большое, то вероятность его выполнения близка к 1.

Пример. Бросание симметричной монеты

Создадим вектор длины $n = 1850$, заполним его 0 или 1 случайным образом при помощи команды `=trunc(rand(1) + 0.5)` (0 - решка, 1 - орёл).

Определим число появлений орла и его относительную частоту появления в серии из 170 независимых испытаний. Орёл появился 100 раз, относительная частота равна 0.588 $\Rightarrow |0.588 - 0.5| = 0.088 < 0.1$

Определим число появлений орла и его относительную частоту появления в серии из 1850 независимых испытаний. Орёл появился 939 раз, относительная частота равна 0.508 $\Rightarrow |0.508 - 0.5| = 0.008 < 0.03$

SUM case 1-170	100,000
MEAN case 1-170	,588
SUM case 1-1850	939,000
MEAN case 1-1850	,508

2 Закон больших чисел в форме Чебышёва

Говорят, что $\{\xi_i\}_{i=1}^n$ - последовательность случайных величин подчиняется закону больших чисел, если

$$\forall \varepsilon > 0 \quad P \left\{ \left| \frac{1}{n} \sum_{i=1}^n \xi_i - \frac{1}{n} \sum_{i=1}^n M\xi_i \right| < \varepsilon \right\} \rightarrow 1 \text{ при } n \rightarrow \infty$$

Одно из утверждений ЗБЧ состоит в том, что при $M\xi_i = a$ и достаточно большом n $\frac{1}{n} \sum_{i=1}^n \xi_i$ приблизительно равно a :

$$\frac{1}{n} \sum_{i=1}^n \xi_i \approx a$$

Будем писать $\frac{1}{n} \sum_{i=1}^n \xi_i \rightarrow a$ при $n \rightarrow \infty$, если

$$\forall \varepsilon > 0 \quad P \left\{ \left| \frac{1}{n} \sum_{i=1}^n \xi_i - a \right| < \varepsilon \right\} \rightarrow 1 \text{ при } n \rightarrow \infty \quad (1)$$

Теорема Чебышёва

Если $\{\xi_i\}_{i=1}^n$ - последовательность попарно независимых случайных величин, имеющих конечные дисперсии такие, что

$$\exists C : \forall i \in [1, n] \cap \mathbb{N} \quad D\xi_i < C,$$

то ЗБЧ выполняется.

Теорема Маркова

Если $\{\xi_i\}_{i=1}^n$ - последовательность как угодно зависимых случайных величин такова, что

$$\frac{1}{n^2} D \sum_{i=1}^n \xi_i \rightarrow 0 \text{ при } n \rightarrow \infty,$$

то ЗБЧ выполняется.

Теорема Хинчина

Если $\{\xi_i\}_{i=1}^n$ - последовательность независимых случайных величин с одинаковым распределением и существует математическое ожидание, то ЗБЧ выполняется.

Пример. Испытание практически достоверного события.

Проверим выполнение соотношения (1) для экспоненциально распределённых слагаемых с $M\xi = 1$. Принять $\varepsilon_1 = 0.2, \varepsilon_2 = 0.05$. При произвольном ε и выборе n из условия $n \geq \frac{9D\xi}{\varepsilon^2}$ (2) выполняется с $p = 0.997$. Если же $n \geq \frac{5.4D\xi}{\varepsilon^2}$, то $p = 0.98$. Второе нас устраивает в качестве достоверного события.

$$\varepsilon_1 = 0.2 \implies n_1 = 135; \quad \varepsilon_2 = 0.05 \implies n_2 = 2160; \quad (M\xi = D\xi = 1)$$

Для заполнения вектора используем формулу $-\text{Log}(\text{rnd}(1))$

Среднее для 135 испытаний равно 0.974 $\implies |1 - 0.974| = 0.026 < 0.2$

Среднее для 2160 испытаний равно 0.631 $\implies |1 - 1.012| = 0.012 < 0.05$

Как видно, все условия выполняются.

MEAN case 1-135	,974
MEAN case 1-2160	1,012

Пример. Невыполнение закона больших чисел.

Рассмотрим случайную величину, распределённую по закону Коши с плотностью

$$p(x) = \frac{1}{\pi} \frac{1}{1 + x^2}$$

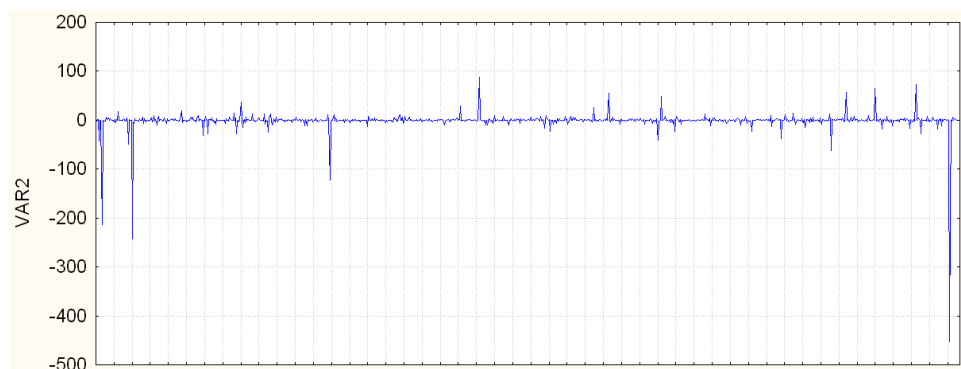
У данного распределения нет ни математического ожидания, ни дисперсии \implies не выполнено условие теоремы Чебышева.

Сделаем 7 выборок объёма $n = 1000$ с распределением Коши по формуле $=V\text{Cauchy}(\text{rnd}(1); 0; 1)$ и определим по каждой среднее значение.

	4,735	,083	4,508	,533	2,379	-1,706	-,249
MEAN case 1-1000	,173	-1,103	621940E7	,857	2,020	-,026	621940E7

Видим, что хотя бы в одной выборке модуль среднего превосходит 1 (например, выборка 2).

Рассмотрим график 2 выборки из распределения Коши. Заметим, что есть редкие наблюдения, отстоящие очень далеко от центра рассеивания - точки 0.

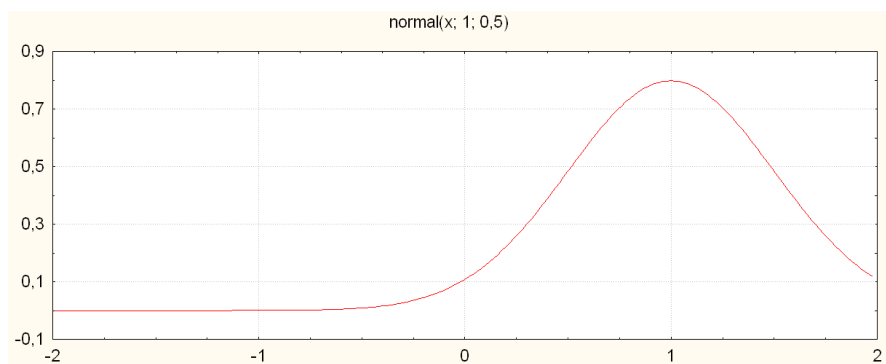
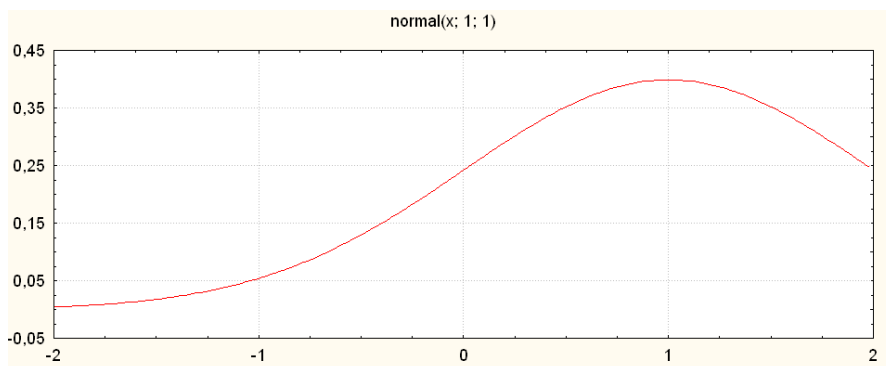


Пример. Сжатие распределения с ростом числа слагаемых.

ЗБЧ в форме Чебышёва означает, что распределение случайной величины $\bar{\xi}_n = \frac{1}{n} \sum_{i=1}^n \xi_i$ сжимается с ростом n . Если $M\xi_i = a$, то сжатие происходит в окрестности точки a .

а) Графики плотностей

Сжатие можно, если распределение $\bar{\xi}_n$ легко выписывается. Пусть $\xi_i \sim \mathcal{N}(a, \sigma^2)$, тогда получаем $\bar{\xi}_n \sim \mathcal{N}(a, \sigma^2/n)$. Построим графики для $n = 1, 4$ и $\sigma = 1, a = 1$.



б) Разброс данных

Сгенерируем 20 выборок размером 640 из распределения $\mathcal{R}[0, 1]$. По всем выборкам для $n = 10, 40, 160, 640$ определим средние, а затем вычислим среднеквадратичное отклонение для них.

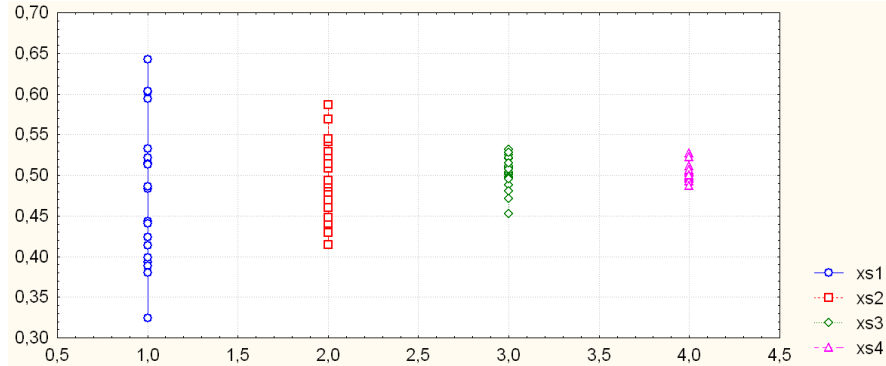
Средние:

MEAN case 1-10	,393	,399	,594	,602	,443	,440	,515	,388	,483	,521	,324	,514	,513	,642	,424	,380	,533	,603	,413	,486
MEAN case 1-40	,465	,476	,529	,482	,521	,429	,541	,441	,489	,493	,415	,586	,568	,545	,475	,469	,460	,509	,448	,514
MEAN case 1-160	,502	,488	,525	,531	,471	,503	,501	,498	,509	,520	,471	,521	,512	,507	,480	,515	,496	,495	,453	,528
MEAN case 1-640	,487	,498	,511	,495	,499	,502	,506	,500	,503	,527	,507	,504	,507	,506	,492	,522	,492	,506	,487	,500

Отклонения, минимальное и максимальное значения соответственно:

,087			,324			,642		
	,046			,415			,586	
		,021			,453			,531
			,010			,487		,527

Видно, что распределение сжимается по мере увеличения n . Изобразим сжатие на графике:



3 Усиленный закон больших чисел

Говорят, что последовательность случайных величин $\{\xi_i\}_{i=1}^n$ подчиняется усиленному ЗБЧ, если

$$P \left\{ \lim_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{i=1}^n \xi_i - \frac{1}{n} \sum_{i=1}^n M\xi_i \right) = 0 \right\} = 1 \quad (2)$$

Или же, если $M\xi_i = a$

$$P \left\{ \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \xi_i = a \right\} = 1$$

Теорема Бореля

Относительная частота $f_n = \frac{\mu_n}{n}$ появления случайного события стремится к истинной вероятности p с вероятностью 1 при росте n независимых испытаний.

Теорема Колмогорова (Достаточное условие выполнения (2))

Если последовательность взаимно независимых случайных величин $\{\xi_i\}$ подчиняется условию

$$\frac{1}{n^2} \sum_{n=0}^{\infty} D\xi_n < \infty,$$

то она подчиняется усиленному ЗБЧ.

Теорема

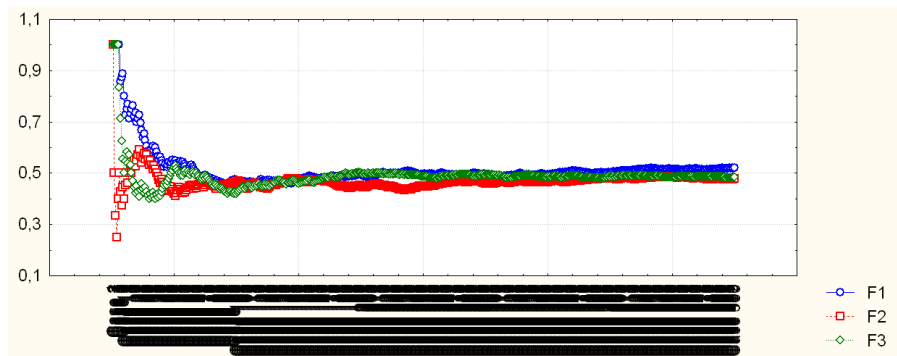
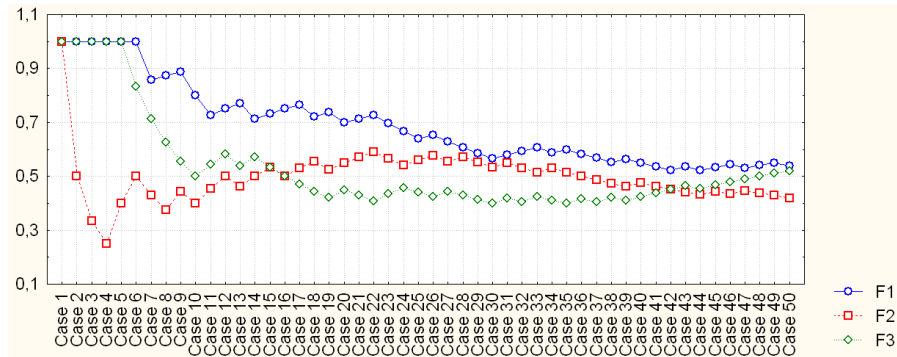
Усиленный ЗБЧ применим к последовательности независимых одинаково распределённых случайных величин тогда и только тогда, когда существует математическое ожидание.

Пример. Эксперименты с монетой.

Сгенерируем 3 последовательности по 500 бросаний монеты в первые 3 столбца таблицы $6\nu \times 500$. Назовём столбцы $x_1, x_2, x_3, f_1, f_2, f_3$, где

$$f_1 = x_1, \quad f_n = \frac{(n-1)f_{n-1} + x_n}{n}, \quad n \in [2, N] \cap \mathbb{N}$$

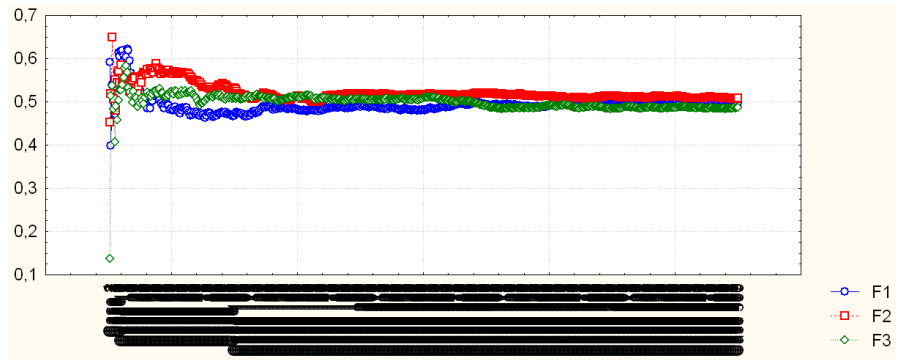
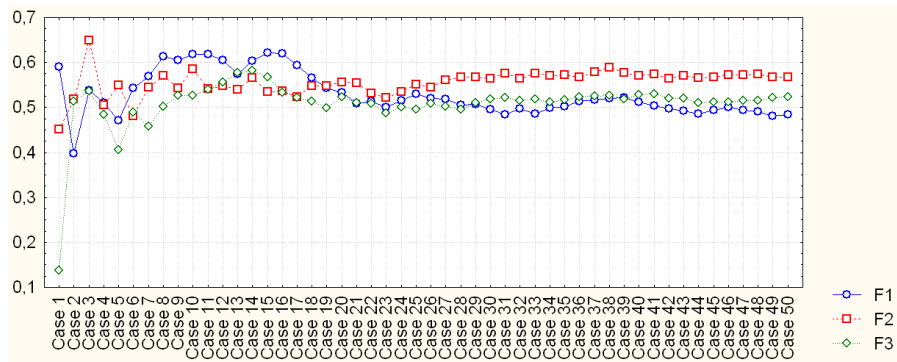
Рассмотрим на графиках зависимость $f_n(n)$ в различных диапазонах: от 1 до 25, до 50, до 100, до 500. Приведены графики для 50 и 500, соответственно.



Частота выпадения герба f_n с ростом n приближается к вероятности выпадения герба $p = 0.5$.

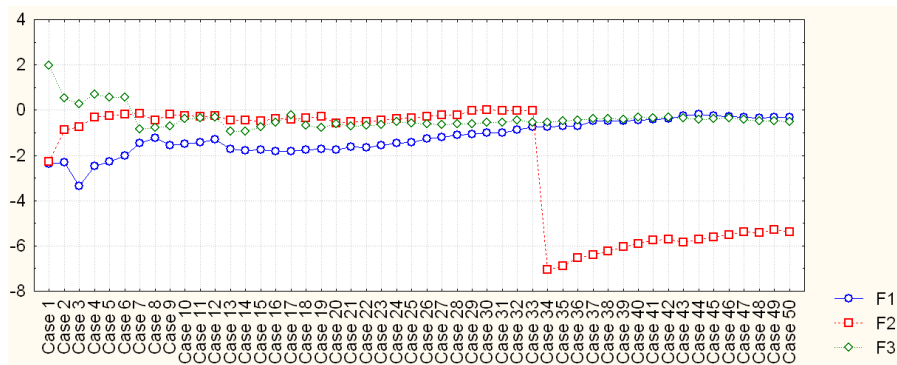
Пример. Эксперименты со случайными числами $\mathcal{R}[0, 1]$.

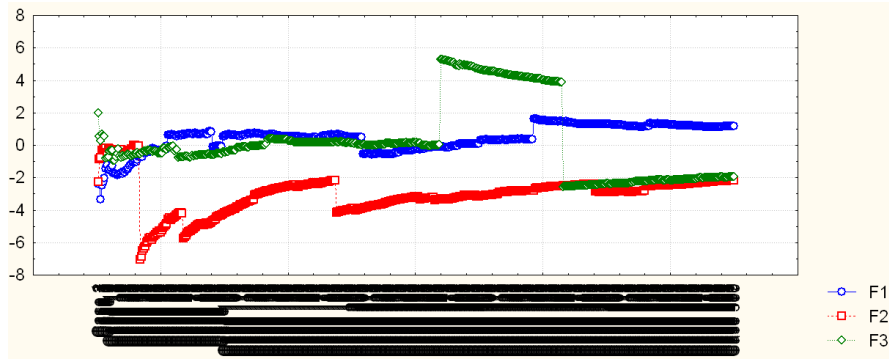
Действия аналогичны предыдущему пункту. Убедимся в том, что последовательность среднеарифметических стремится к 0.5 - математическому ожиданию. Приведу графики для случаев $n = 50$ и $n = 500$.



Пример. Невыполнение закона.

Рассмотрим $\{\xi_i\}$, распределённых по закону Коши. Сделаем 3 выборки. Учтём, что $\tan \pi \xi$ имеет распределение Коши, если $\xi \sim \mathcal{R}[0, 1]$. Выполним преобразования и повторим действия из предыдущих примеров.





Видим, что есть скачки, где значения далеки от 0 - центра распределения.

4 Теорема Гливенко - основная теорема статистики.

$\{x_i\}_{i=1}^n$ - выборка из n независимых наблюдений над случайной величиной X с функцией распределения $F(x)$. Получим вариационный ряд $x_1 \leq x_2 \leq \dots \leq x_n$. Определим функцию эмпирического распределения

$$F_n^*(x) = F_n^*(x; x_1, x_2, \dots, x_n) = \frac{\mu_n(x)}{n},$$

где $\mu_n(x)$ - число тех наблюдений, для которых $x_i < x$. Ясно, что $F_n^*(x)$ - ступенчатая функция; это функция распределения, которое получается, если значениям x_1, \dots, x_n присвоить вероятности, равные $\frac{1}{n}$. $F_n^*(x)$ - случайная, так как зависит от x_1, \dots, x_n .

Теорема Гливенко

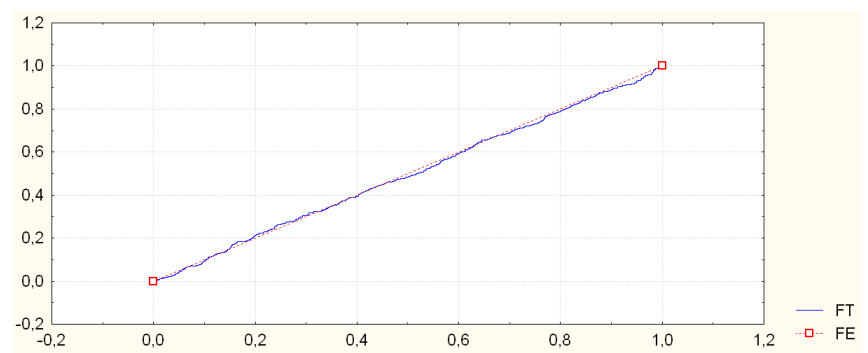
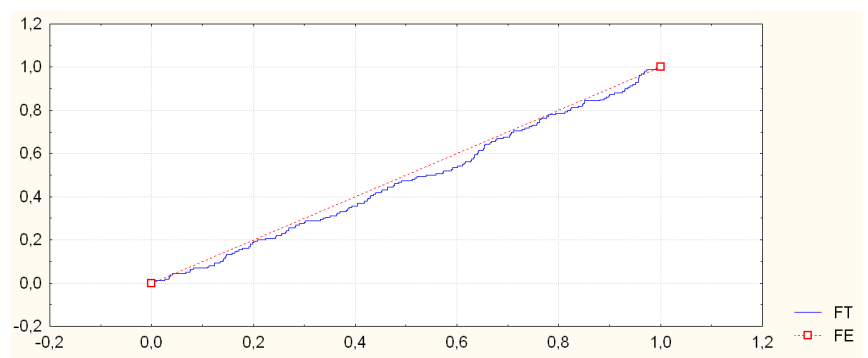
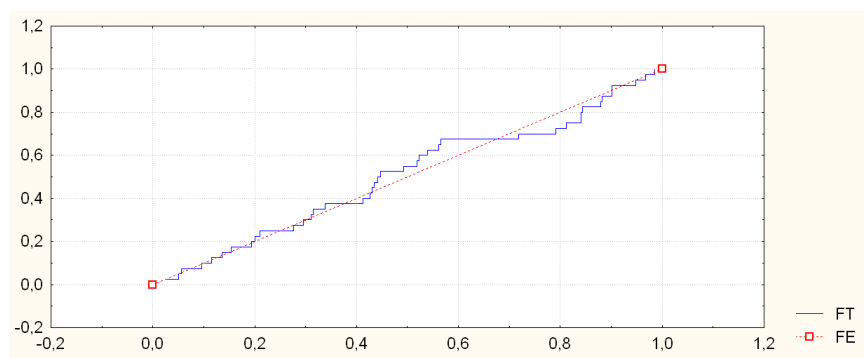
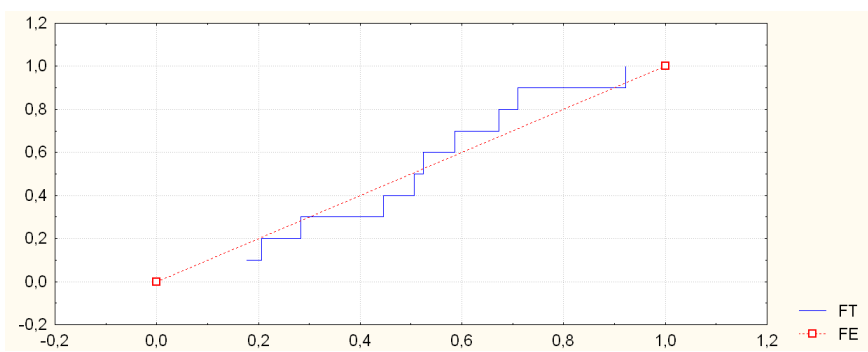
$$\sup_x |F_n(x) - F(x)| \rightarrow 0 \text{ при } n \rightarrow \infty \text{ с вероятностью 1.}$$

Пример. Наблюдение над случайной величиной, имеющей распределение $\mathcal{R}[0, 1]$.

В первом столбце сделаем выборку объёма n с $\mathcal{R}[0, 1]$ с сортировкой по возрастанию. Во втором столбце получим значения функции эмпирического распределения по формуле $=\nu/n$. В третьем же столбце запишем два значения: 0 и 1, чтобы построить функции теоретического распределения.

Пусть $n = 10, 40, 160, 640$.

Приведу полученные графики, по которым видно, что по мере увеличения n функция эмпирического распределения приближается к теоретической.



5 Центральная предельная теорема.

Рассмотрим сумму случайных величин $S_n = \sum_{i=1}^n \xi_i$. Оказывается, что при широких условиях по мере увеличения числа слагаемых эта сумма приблизительно распределена по нормальному закону

$$S_n = \sum_{i=1}^n \xi_i \sim \mathcal{N}(a = MS_n, \sigma_n^2 = DS_n)$$

Если говорить более точно, то

$$P \left\{ \zeta_n \equiv \frac{\sum_{i=1}^n \xi_i - M \sum_{i=1}^n \xi_i}{\sqrt{D \sum_{i=1}^n \xi_i}} < x \right\} \rightarrow \Phi(x) \equiv \int_{-\infty}^x \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt, \quad n \rightarrow \infty \quad (\text{ЦПТ})$$

Нормировка состоит в вычитании математического ожидания и делении на корень из дисперсии; нормированная сумма ζ_n имеет $M\zeta_n = 0, D\zeta_n = 1$.

Говорят, что к $\{\xi_i\}_{i=1}^n$ применима центральная предельная теорема, если выполняются (ЦПТ).

(ЦПТ) выполняется при различных условиях, например, при следующих:

Теорема.

Если $\{\xi_i\}_{i=1}^n$ - последовательность независимых случайных величин, причем их дисперсии равномерно ограничены с двух сторон, т.е. $0 < c1 \leq D\xi_i \leq c2 < \infty$, то (ЦПТ) выполняется.

Теорема*.

Если $\{\xi_i\}_{i=1}^n$ - последовательность независимых случайных величин, причем $M\xi_i = a, D\xi_i = \sigma^2$, то (ЦПТ) выполняется.

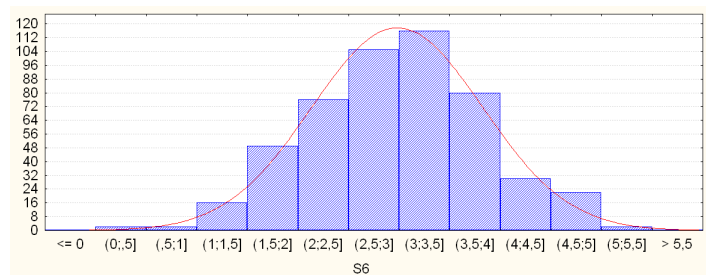
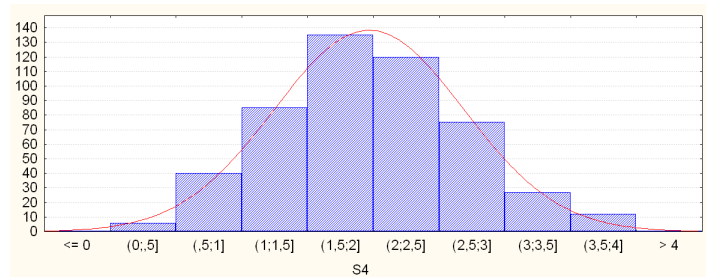
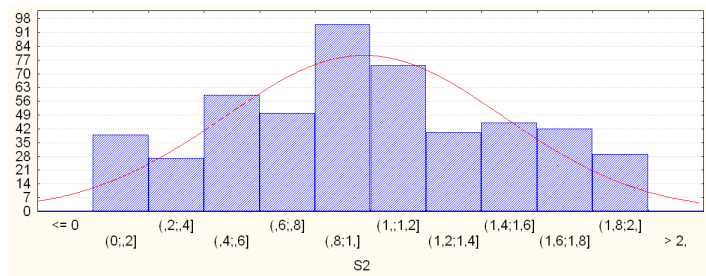
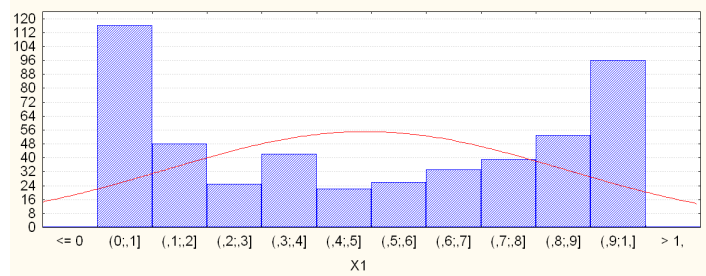
Пример. Одинаково распределенные слагаемые.

Убедимся статистически в том, что сумма нескольких случайных величин распределена приближенно по нормальному закону. Сделаем это на примере суммы $S = \sum_{k=1}^m \xi_i, m = 6$ независимых случайных величин, имеющих $\mathcal{B}(a, b)$ - бета-распределение с плотностью

$$p(x|a, b) = \frac{x^{a-1}(x-1)^{b-1}}{B(a, b)} = \frac{1}{\pi} \frac{1}{\sqrt{x(1-x)}}, \quad B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt$$

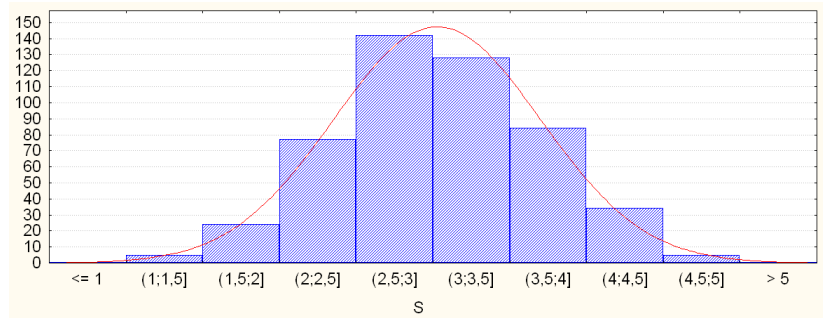
Чтобы статистически оценить закон распределения для S , следует многократно промоделировать суммирование: получим S_1, S_2, \dots, S_N - выборку для суммы. Для неё построим гистограмму и сравним ее с нормальной плотностью. Создадим 6 выборок, имеющих $\mathcal{B}(a, b)$ с параметрами $a = b = 0.5$, и 3 столбца, в которые запишем суммы:

$$S_2 := X_1 + X_2, \quad S_4 := S_2 + X_3 + X_4, \quad S_6 := S_4 + X_5 + X_6$$



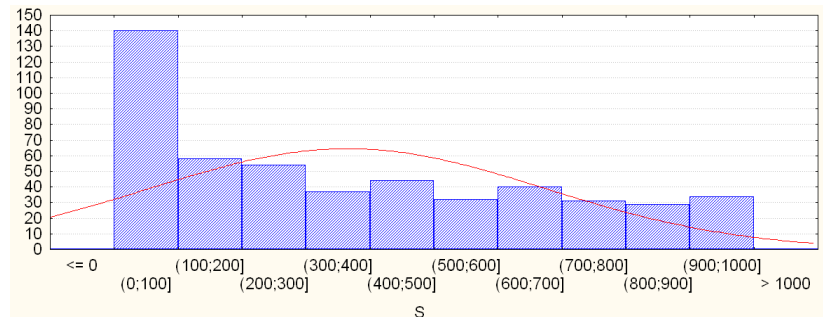
Пример. Различно распределенные слагаемые.

Распределение суммы сходится к нормальному и в том случае, когда слагаемые распределены по различным законам. Оценим экспериментально распределение для $S = \sum_{k=1}^6 \xi_i$, где $\xi_i \sim \mathcal{B}(a_i, b_i)$, где $a_i \in \{1, 0.5, 1, 1, 2, 2\}$, $b_i \in \{0.5, 1, 1, 2, 1, 2\}$. Приведу гистограмму.



Видно, что получается желаемый результат.

Если же добавить к $S = \sum_{k=1}^6 \xi_i$ седьмое слагаемое, дисперсия которого существенно превышает все остальные дисперсии из суммы, то приближённая нормальность не имеет места. Пусть $\xi_7 \sim 1000 * \mathcal{B}(0.5, 1)$. Приведу полученную гистограмму.



Очевидно, что приближённой нормальности нет.