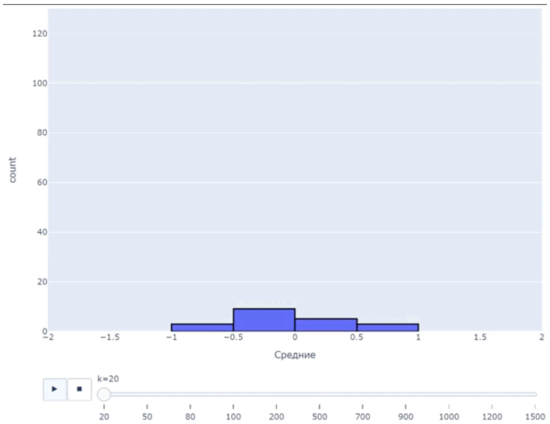
Генерация распределений. Проверка определений известных распределений

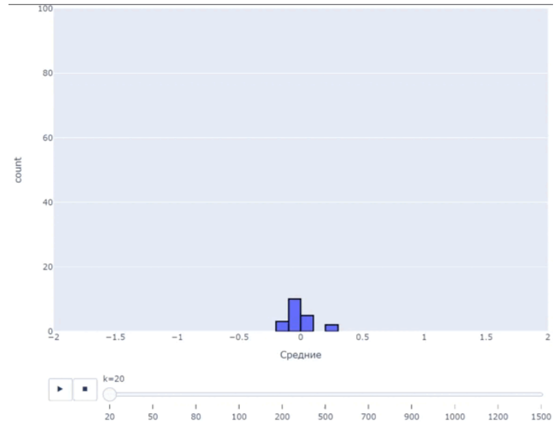
Центральная предельная теорема

k – количество выборок

Среднее выборок из 30 элементов



Среднее выборок из 500 элементов



Постановка задачи

1) Сгенерировать выборку нормального распределения $Y \sim N(\mu, \sigma^2)$ используя определение центральной предельной теоремы. На основе $\mathrm{k} = 20 - 30$ равномерно распределенных реализаций случайных величин образовать новую выборку по определению центральной предельной теоремы.

Если $\mathrm{X}_i \sim U(a_i, b_i), i = 1,2, \ldots, k$, где X_i – равномерно распределенная реализация случайной величины со случайными параметрами $a_i \epsilon \mathbb{R}, b_i \epsilon \mathbb{R}$, то ожидаемая нормально распределенная случайная величина Y будет найдена как:

$$Y = \sum_{i=1}^{k} X_i, i = 1,2, \ldots, k$$

Для генерации выборок рекомендуется пользоваться встроенными в компьютерные статистические пакеты функциями генерации равномерно распределённых случайных величин, которые задаются с помощью параметров границ интервала генерации чисел a и b. (numpy.random.uniform(low=0.0, high=1.0, size=None))

Есть k (20-30 штук) случайных величин X, которые распределены по равномерному закону.
Возьмем k элементов, то есть по одной реализации каждой случайной величины, получим выборку длины k.
Просуммируем все элементы этой выборки и получим новую случайную величину.
То есть новая случайная величина Y является суммой реализаций случайных величин X.

Возьмем k равномерных случайных величин, с количеством реализаций n.
Найдем сумму их реализации.
Тогда мы получим выборку реализаций случайной величины Y длины n.

Если мы построим гистограмму новой полученной случайной величины Y, то ее распределение будет стремится к нормальному распределению при росте длины выборок n.

Постановка задачи

Для получившейся выборки построить гистограмму, визуализировать на гистограмме теоретическую плотность нормального распределения по несмещенным точечным оценкам и провести тест на нормальное распределение с помощью критерия χ^2-Пирсона (степени свободы рассчитывать как $\mathrm{df} = \mathrm{m} - 3$) и методом анаморфоз.

Определить влияние числа сгенерированных равномерно распределенных величин на итоговое качество генерации нормального распределения при помощи взятия 3 тестовых генераций при разных *k* и проведения теста на распределение

Для расчета P для теоретических частот scipy.stats.norm.cdf()

Постановка задачи

2) Сгенерировать выборку χ^2-распределения $R \sim {\chi^2}_{df}$ используя определение распределения χ^2. На основе Z-оценок нормально распределенных случайных реализаций случайных величин $L_i \sim N(\mu_i, \sigma_i^2)$ образовать новую выборку по определению χ^2-распределения:

$$R = \sum_{i=1}^{k} Z[L_i]^2 \qquad Z[L_i] = \frac{L_i - \mu_{L_i}}{\sigma_{L_i}} \qquad L_i \sim N(\mu_i, \sigma_i^2) \qquad i = 1,2,\dots,k$$

Для получившейся выборки построить гистограмму, визуализировать на гистограмме теоретическую плотность ${\chi^2}_k$ распределения с $\mathrm{df} = \mathrm{k}$ степенями свободы и провести тест на χ^2-распределения с помощью критерия χ^2 -Пирсона ($\mathrm{df} = \mathrm{m} - 2$).

Для расчета P для теоретических частот scipy.stats.chi2.cdf()

Постановка задачи

3) Сгенерировать выборку распределения Фишера на основе определения. На основе двух случайных реализаций Y_1, Y_2 случайных величин, распределенных по χ^2-распределению со степенями свободы d_1, d_2 соответственно. Сгенерировать выборку, распределенную по распределению Фишера $S \sim F(d_1, d_2)$ в соответствии с определением:

$$\mathrm{S} = \frac{Y_1/d_1}{Y_2/d_2}, \qquad S \sim F(d_1, d_2)$$

Для получившейся выборки построить гистограмму, визуализировать на гистограмме теоретическую плотность распределения. Провести тест на распределение Фишера с помощью критерия χ^2-Пирсона ($\mathrm{df} = \mathrm{m} - 3$).

Для расчета P для теоретических частот scipy.stats.f.cdf()

Постановка задачи

4) Сгенерировать выборку t-распределения на основе определения. На основе $k \approx 2 - 9$ случайных реализаций $Y_1, Y_2, \dots, Y_{\mathrm{k}}$ случайных величин, распределенных по стандартному нормальному распределению $Y_{\mathrm{i}} \sim N(0,1), i = 1, 2, \dots, k$, сгенерировать выборку $T \sim t(k)$, распределенную по t-распределению Стьюдента с df = k степенями свободы в соответствии с определением:

$$T = \frac{Y_0}{\sqrt{\frac{1}{k}\sum_{i=1}^{k} Y_i^2}}, \qquad Y_0 \sim N(0,1)$$

Постановка задачи

Для получившейся выборки построить гистограмму, визуализировать на гистограмме теоретическую плотность $t(k)$. Для получившейся выборки провести тест на t-распределение Стьюдента с помощью критерия χ^2-Пирсона.

Для расчета P для теоретических частот scipy.stats.t.cdf()

Для всех заданий количество генерируемых значений выборки n установить равным 500-1000

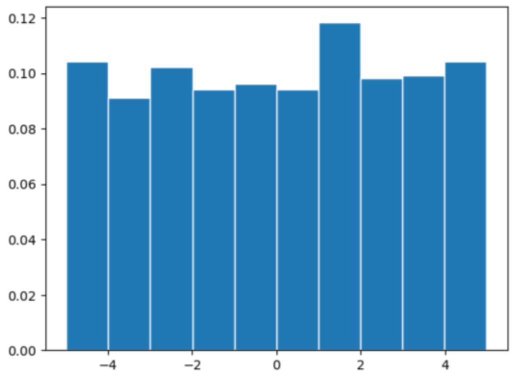
Уровень надежности для критерия χ^2-Пирсона γ = 0.95.

Пример

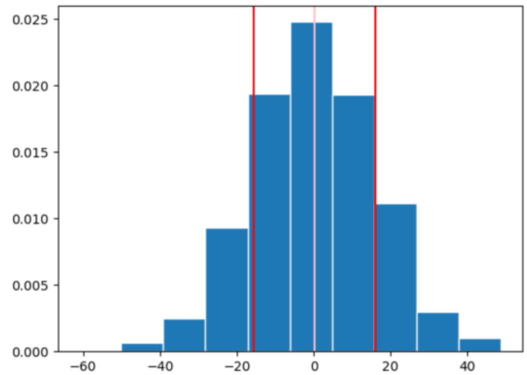
Нормальное распределение

Возьмем 30 реализаций равномерной случайной величины из 1000 элементов
Находим сумму случайных величин, строим гистограмму

Пример 1 реализации случайной величины



Распределение полученных сумм

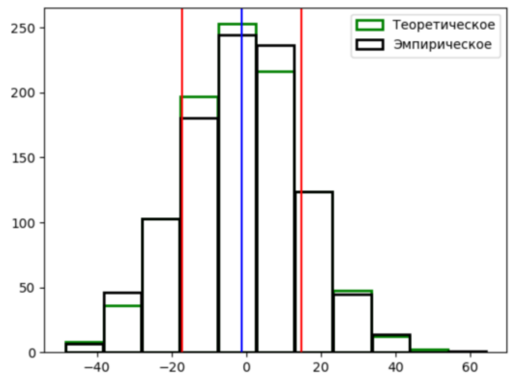


$$\mu_Y = \sum_{i=1}^{k} \mu_{X_i} \qquad {\sigma_Y}^2 = \sum_{i=1}^{k} {\sigma_{X_i}}^2 \qquad \sigma_Y = \sqrt{\sum_{i=1}^{k} {\sigma_{X_i}}^2}$$

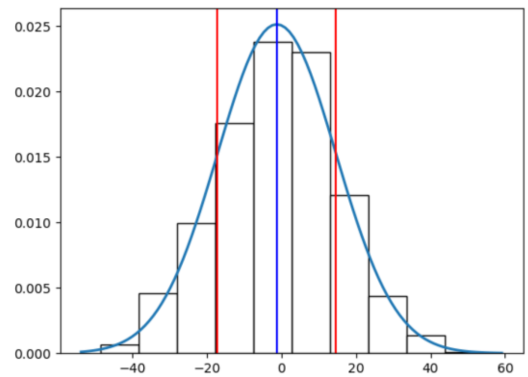
Пример

Нормальное распределение

Для получившейся выборки построить гистограмму, визуализировать на гистограмме теоретическую плотность нормального распределения по несмещенным точечным оценкам.



или



Провести тест на нормальное распределение с помощью критерия χ^2-Пирсона и методом анаморфоз

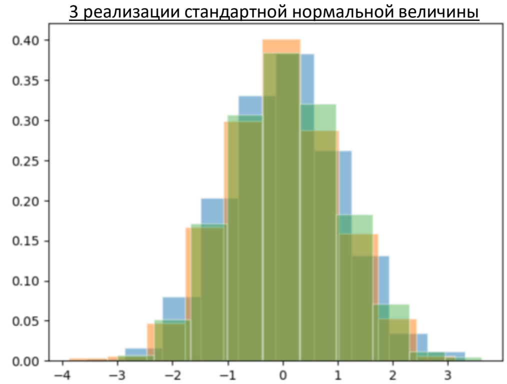
Пример

Распределение Пирсона

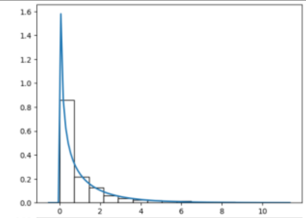
Возьмем k = 1, 3, 10 реализаций стандартной нормальной случайной величины из 1000 элементов

Находим сумму квадратов случайных величин, строим гистограмму

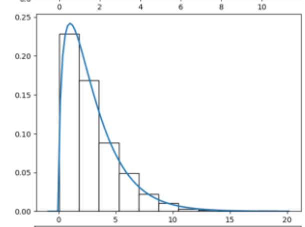
3 реализации стандартной нормальной величины



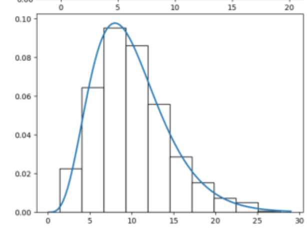
k = 1



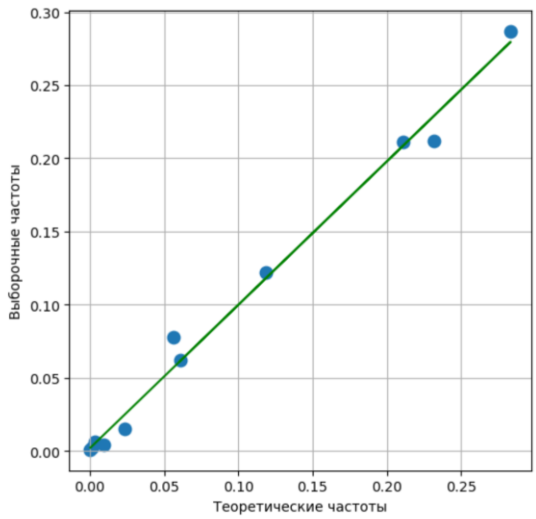
k = 3



k = 10



Графический способ проверки на соответствие распределению

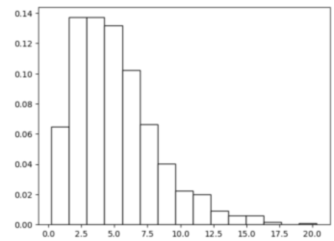


$R^2 = 0.991$

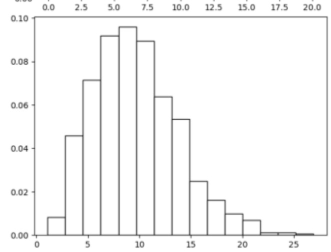
Пример

Распределение Фишера

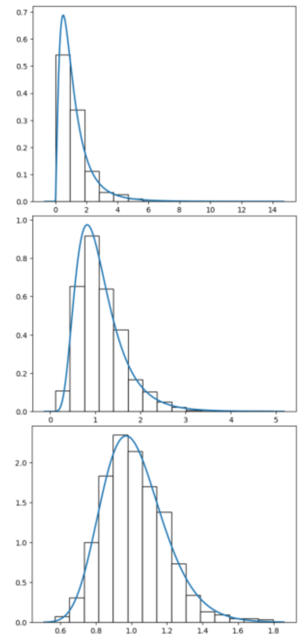
Возьмем 2 реализации случайных величин, распределенных по $\chi 2$-распределению со степенями свободы d1 = 5 и d2 = 10. Длины выборок равны 1000. По определению построим распределение Фишера



d1 = 5



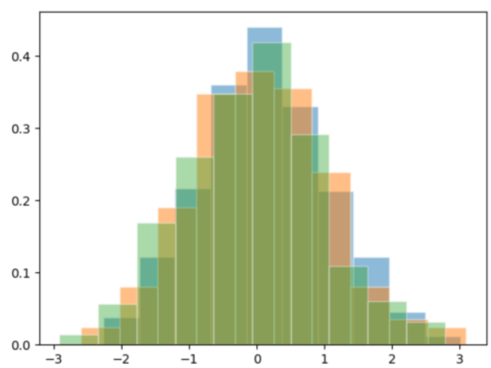
d2 = 10



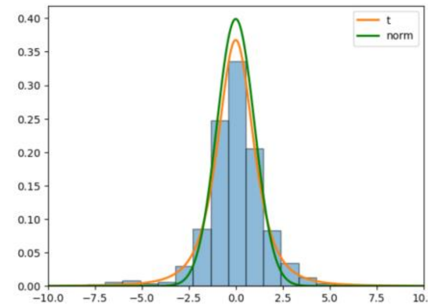
Пример

Распределение Стьюдента

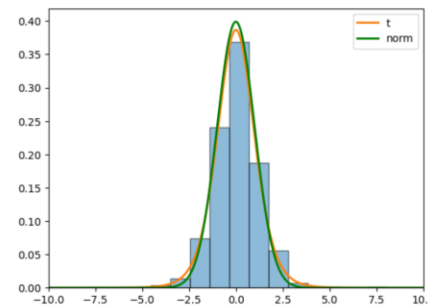
Возьмем 3 реализации случайных величин, распределенных по стандартному нормальному распределению. Длины выборок равны 500. По определению построим распределение Стьюдента



k = 3



k = 8



Структура отчета

- **4.1 Постановка задачи**
- **4.2 Ход выполнения работы**
- **4.2.1 Генерация выборки распределения Гаусса** (определение, формулы, пример сгенерированной выборки, полученные распределения при разных k (1,2,5), сформулированные гипотезы, анаморфоза и критерий Пирсона для каждого)
- **4.2.2 Генерация выборки распределения Пирсона** (определение, формулы, пример сгенерированной выборки, сформулированные гипотезы, критерий Пирсона,также можно использовать для проверки q-q plot и найти коэффициент детерминации)
- **4.2.3 Генерация выборки распределения Фишера** (как в п. 4.2.2.)
- **4.2.4 Генерация выборки распределения Стьюдента** (как в п. 4.2.2.)
- **4.3 Выводы**