

В этом задании вам предстоит проверить работу центральной предельной теоремы, а также поработать с генерацией случайных чисел и построением графиков в Питоне.

Выберите ваше любимое непрерывное распределение (чем меньше оно будет похоже на нормальное, тем интереснее; попробуйте выбрать какое-нибудь распределение из тех, что мы не обсуждали в курсе). Сгенерируйте из него выборку объёма 1000, постройте гистограмму выборки и нарисуйте поверх неё теоретическую плотность распределения вашей случайной величины (чтобы величины были в одном масштабе, не забудьте выставить у гистограммы значение параметра `normed=True`).

Ваша задача — оценить распределение выборочного среднего вашей случайной величины при разных объёмах выборок. Для этого при трёх и более значениях n (например, 5, 10, 50) сгенерируйте 1000 выборок объёма n и постройте гистограммы распределений их выборочных средних. Используя информацию о среднем и дисперсии исходного распределения (её можно без труда найти в википедии), посчитайте значения параметров нормальных распределений, которыми, согласно центральной предельной теореме, приближается распределение выборочных средних. Обратите внимание: для подсчёта значений этих параметров нужно использовать именно теоретические среднее и дисперсию вашей случайной величины, а не их выборочные оценки. Поверх каждой гистограммы нарисуйте плотность соответствующего нормального распределения (будьте внимательны с параметрами функции, она принимает на вход не дисперсию, а стандартное отклонение).

Опишите разницу между полученными распределениями при различных значениях n . Как меняется точность аппроксимации распределения выборочных средних нормальным с ростом n ?

Review criteria

меньше

Решение должно представлять собой IPython-ноутбук, содержащий:

- код, генерирующий выборки и графики;
- краткие описания каждого блока кода, объясняющие, что он делает;
- необходимые графики (убедитесь, что на них подписаны оси);
- выкладки с вычислениями параметров нормальных распределений, аппроксимирующих выборочные средние при различных n ;
- выводы по результатам выполнения задания.

Тема для обсуждения задания с проверкой однокурсниками на тему ЦПТ.

Рекомендации по выполнению

1. Дайте ссылки на статью на Wiki с выбранным вами распределением
2. Укажите формулы для расчёта теоретических мат ожидания и дисперсии исходного распределения (по возможности в LaTeX формате), ниже проведите расчёты в python.
3. Не используйте `numpy.std` и `numpy.mean`! См. пункт 2.
4. Выполните расчёты мат ожидания и дисперсии средневывборочных в отдельной ячейке для всех n . Приведите формулы.
5. Используйте только латинские символы в названии ноутбука. Хорошо - "week4.ipynb", плохо - "ЦПТ.ipynb". Иначе ваше задание будет недоступно для скачивания и вы получите 0 баллов.

Рекомендации по проверке чужих работ

1. Оценивайте чужую работу так, как вы бы хотели, чтобы оценивали вашу.
2. Ставьте оценки только в соответствии с грейдером (не *снижайте оценки за отсутствие ссылок на вики или формул, грейдер не содержит такого требования, их наличие просто знак уважения и помощь проверяющему, но в задании их указывать не требуют*)
3. При невыполнении рекомендаций с форума не снижайте оценку, а дайте соответствующий совет в поле для комментария.
4. При наличии в коде `numpy.std` и `numpy.mean` убедитесь, что они используются для вычисления мат ожидания и дисперсии исходного распределения вместо формул для расчёта теоретических значений. Только после этого снижайте оценку.

FAQ

Question: Я не могу скачать задание однокурсников - открывается какая-то ерунда.

Answer: Сохраните открывшийся файл средствами браузера. Найдите его в Jupiter Notebook. Смените при необходимости расширение с `.txt` на `.ipynb`. Откройте и проверьте.

Question: Я не могу открыть/скачать задание однокурсника. В новом окне просто открывается страница проверки однокурсника, ничего не скачивается.

Answer: Вероятно проверяемый загрузил файл с кириллическими символами в названии (например, "ЦПТ.ipynb"). Coursera не позволяет загружать такие файлы - она их просто "ломает". Файл необходимо перезалить. Напишите об этом в комментарии и поставьте 0 баллов, дайте ссылку на этот FAQ.

Question: Я получил 0 баллов и комментарий, что мой файл не открывается. Что делать?

Answer: Проверьте свою работу, действительно ли её нельзя скачать. Если это так, то причина в кириллических символах (см предыдущий вопрос). Переименуйте файл с работой и перезагрузите задание.

Если всё открывается хорошо не расстраивайтесь - вероятно 2 других проверяющих смогут оценить вас корректно. А оценка ставится по медиане. Так что 1 некорректная оценка не испортит вам ведомость.

Question: Если я поздно подал задание (до конца сессии осталось x дней), то успею ли я получить оценку? Получу ли я сертификат?

Answer: Если повезёт, то успеете. Если нет, придётся сменить сессию. Рекомендую сдавать задание как можно раньше! При смене сессии есть шанс получить "зачёт автомат", если вы получили хотя бы 1 оценку от однокурсников.

Question: что значит "Сгенерируйте из него выборку объёма 1000"?

Answer: Примером выборки может случить набор из 1000 чисел от 1 до 6, как если бы вы 1000 раз бросили игральную кость и каждый раз записывали результат.

Question: Что значит "сгенерируйте 1000 выборок объёма n "?

Answer: Создайте 1000 наборов по n чисел. Аналогично предыдущему вопросу.

Question: Как нарисовать гистограмму?

Answer: <https://plot.ly/matplotlib/histograms/>

Question: Я сам написал распределение. А теперь не могу посчитать для него мат ожидание и дисперсию. Как быть?

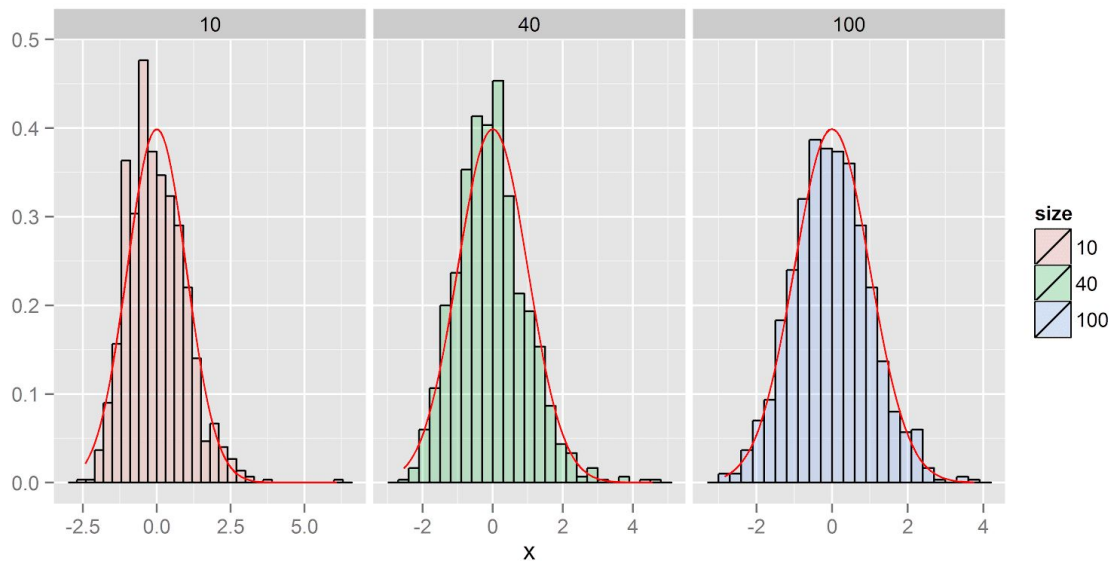
Answer: Придётся начать с начала и выбрать распределения из тех, что есть в модуле `scipy` и для которых есть данные в WIKI.

Question: Пик графика нормального распределения не совпадает с пиком гистограммы.

Answer: Вероятно вы ошиблись в расчёте параметров нормального распределения средневывборочной величины по ЦПТ. Заметьте, что параметры, принимаемые функцией `scipy.stats.norm` - это не совсем то же самое, что мат ожидание и дисперсия.

Question: А если у меня экспоненциальное (другое асимметричное и/или ограниченное с одного бока) распределение?

Answer: Тут может быть некоторое смещение. Однако, с ростом n оно как правило уменьшается. См пример для экспоненциального распределения из [заметки](#):



Question: Высота графика нормального распределения не совпадает с высотой гистограммы. Увеличение n не помогает.

Answer: вероятно вы неверно посчитали дисперсию по ЦПТ. Возможно забыли где-то корень, или наоборот поставили лишний. Заметьте, что параметры, принимаемые функцией `scipy.stats.norm` - это не совсем то же самое, что мат ожидание и дисперсия.