**Построение графика Биномиального распределения.**

Я провел статистический анализ успешности игр по регионам с помощью биномиального распределения.

Использовал для своих целей Python, поскольку он позволяет достаточно гибко работать с визуализацией графиков, в нем также удобно прописывать основную логику обработки данных из БД и работать с ними.

Для начала я импортировал необходимые мне для дальнейшей работы библиотеки и подключился к базе данных, используя следующий синтаксис:

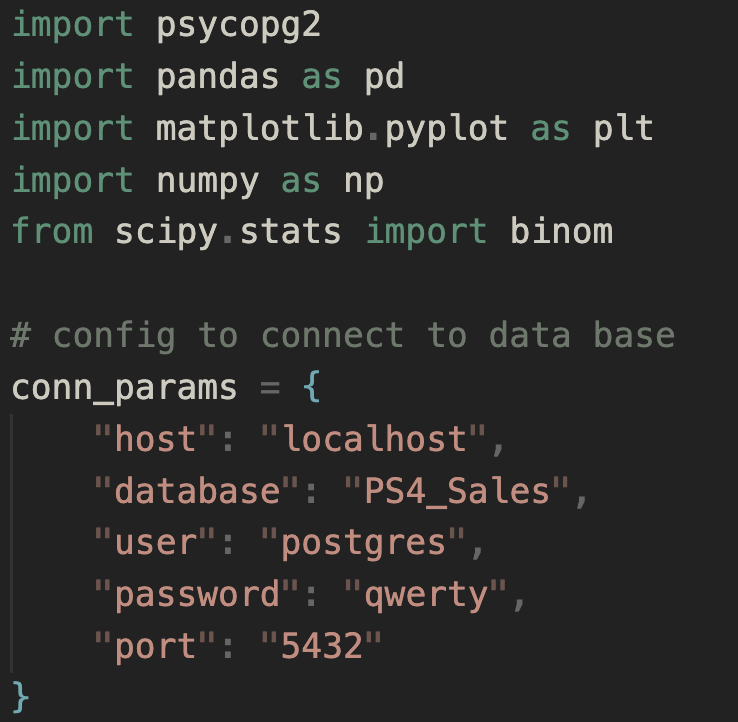


Рис 1 - подключение библиотек и установка параметров для соединения с БД.

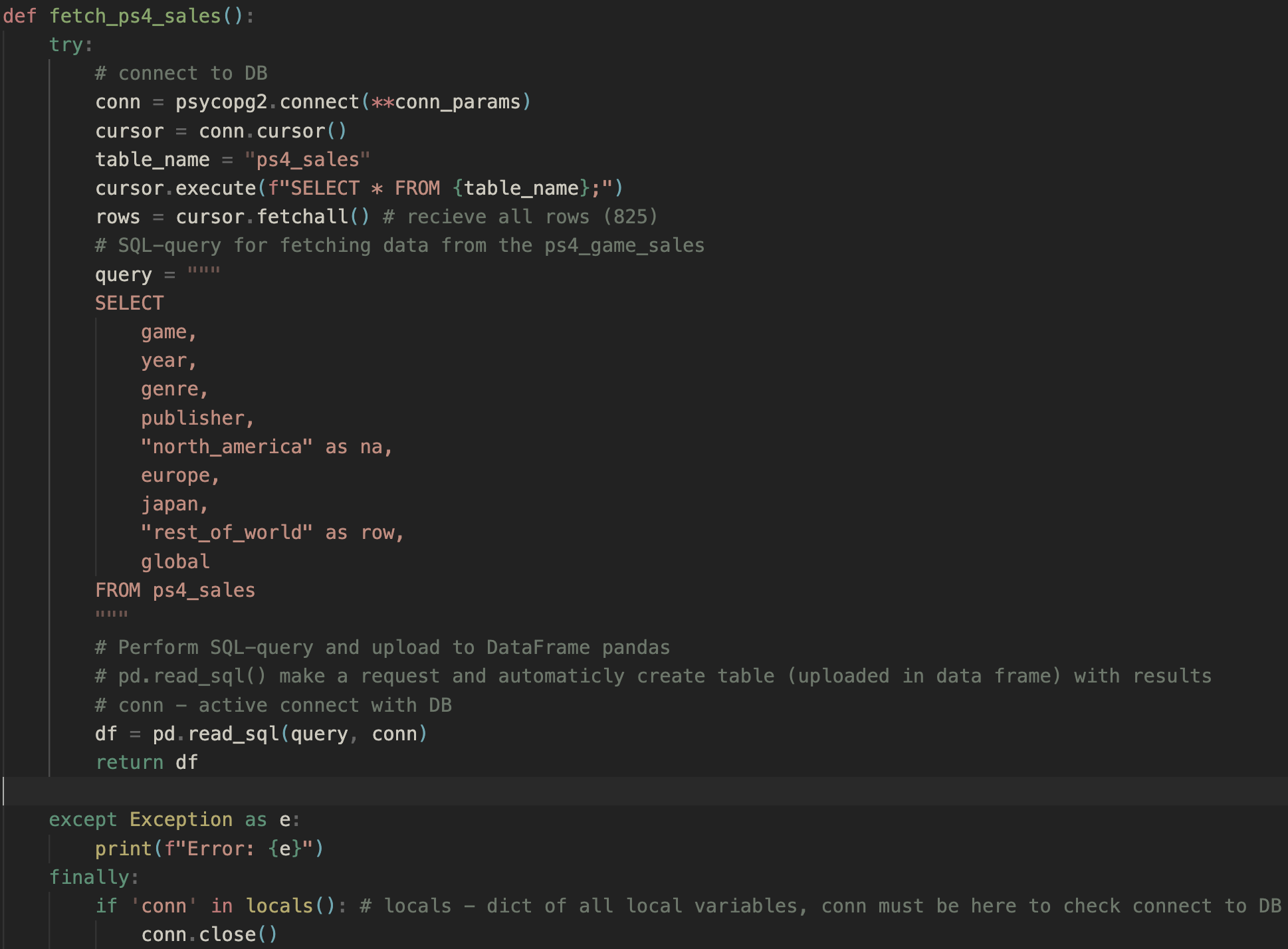


Рис 2 - соединение с БД.

Также установил параметры для определения «успеха» игры в регионах. Поставил везде значение 100 тыс. копий, поскольку для многих (особенно небольших студий разработки) такое количество продаж покрывает затраты на разработку и маркетинг и позволяет получать доход с игры.

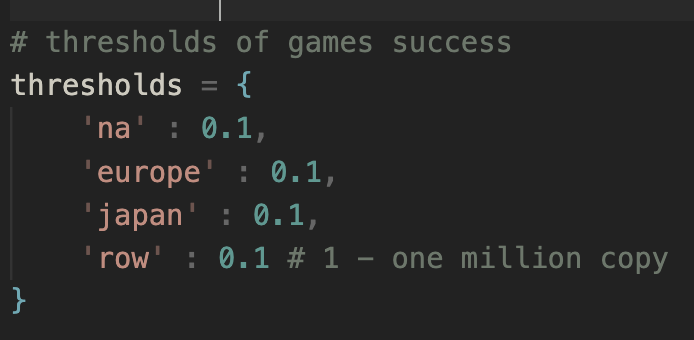


Рис. 3 - определение параметров «успешности».

Теперь мне необходимо было добавить новый столбец в мой Data Frame, чтобы затем применить биномиальное распределение. В этом столбце хранится 1, если продажа игры в регионе составила более установленного порога (true) и 0, если меньше(false).

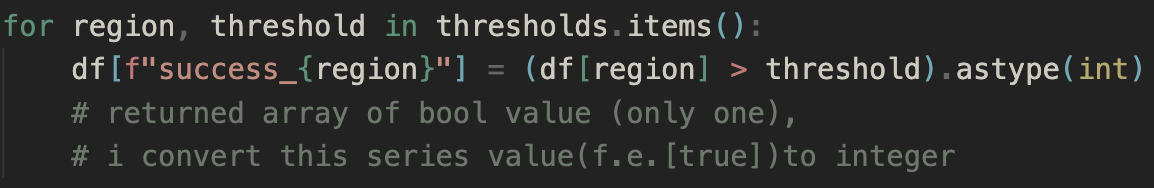


Рис. 4 - добавление столбца в data frame.

Установил значение n (общее количество игр) и создал словарь p\_values с ключами - регионами и значениями - вероятностями успеха игры в этом регионе.

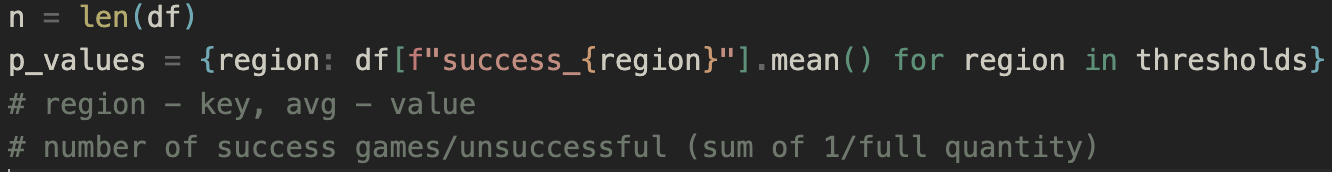


Рис. 5 - добавление n и p\_values.

Теперь же написал функцию для визуализации графиков, использующую метод для вычисления биномиального распределения - binom.pmf(x, n, p), где за n независимых испытаний произойдет ровно x успехов, если вероятность успеха в одном испытании равна p. Пики графиков в точках n\*p, где n - общее число независимых экспериментов, p - вероятность успеха в 1 испытании.

Подробнее о вычислениях:

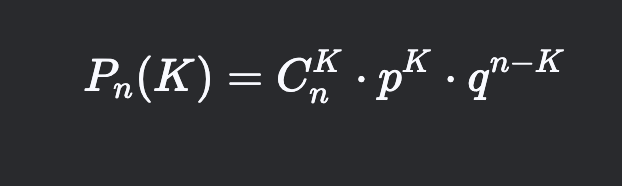


Рис. 6 - Формула Бернулли, использующаяся для вычисления вероятности в биномиальном распределении.

K - вероятность получить K успехов в n испытаниях.

C из n по K - число сочетаний из n по K.

p - вероятность успеха в 1 испытании, q = 1 - p - вероятность неудачи.

Формула работает для независимых испытаний с постоянной вероятностью p.

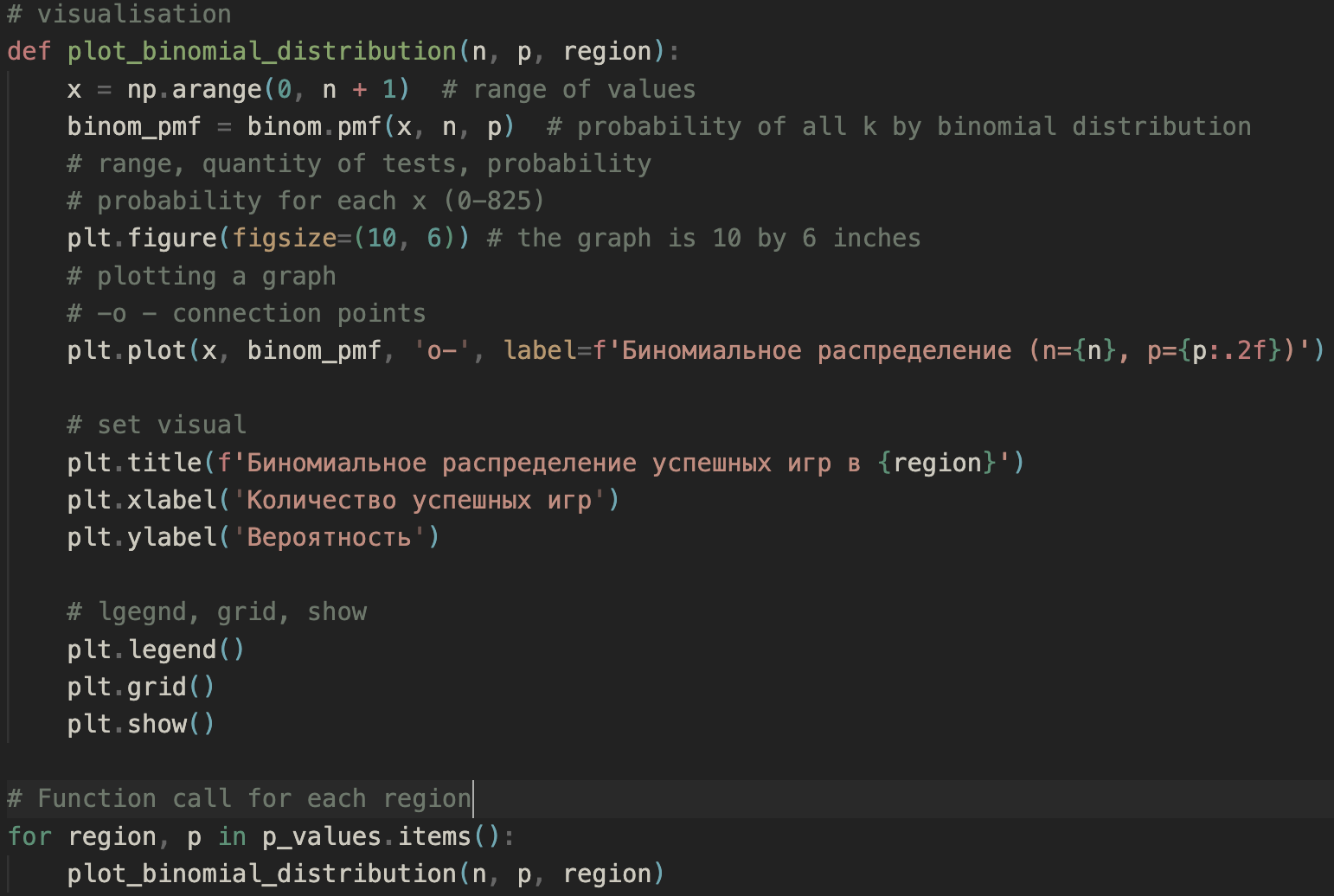


Рис. 7 - функция для визуализации графиков биномиального распределения и ее вызов для каждого региона с собственным параметром вероятности успеха игр.

Далее будет представлен вывод графиков. На них мы можем заметить, сколько «успешных» игр окажется на каждом из региональных рынков с той или иной вероятностью. Можно заметить, что в Европе и Северной Америке ситуация достаточно похожа, немного отличается, но все же приближено к ней и распределение остального мира. А вот у Японии показатели отличаются уже заметно сильнее.

Следовательно, вероятность того, что недавно вышедшая игра окажется успешной выше всего у Северной Америки, ниже всего у Японии. Скорее всего, это связано с региональными и жанровыми предпочтениями игроков.

Также стоит заметить, что я не анализировал ситуацию по всему миру, так как целью было проведение статистического анализа успешности игр по регионам с помощью биномиального распределения. Кроме того, на мировую ситуацию может сильно

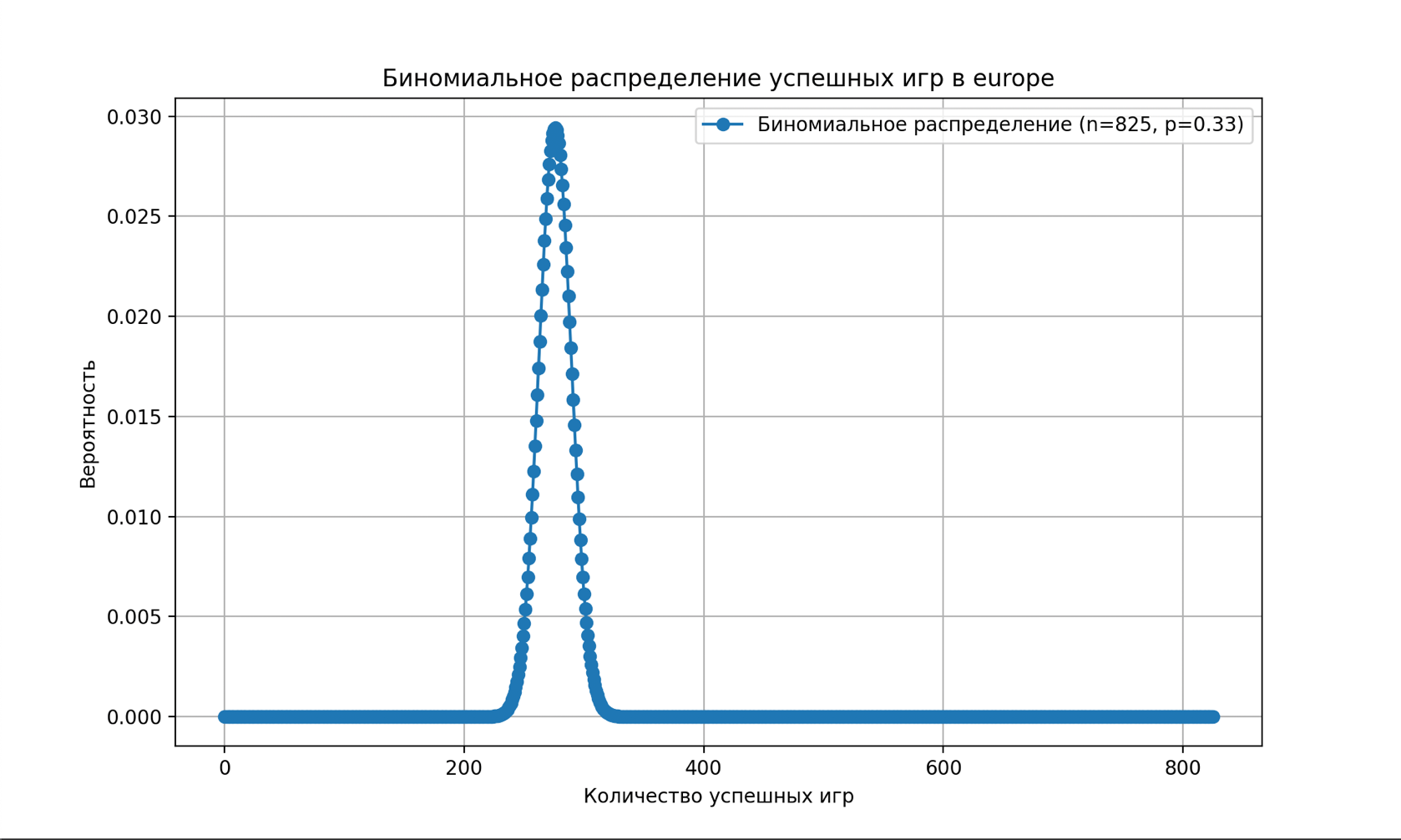


Рис. 8 - график биномиального распределения успешных игр в Европе.

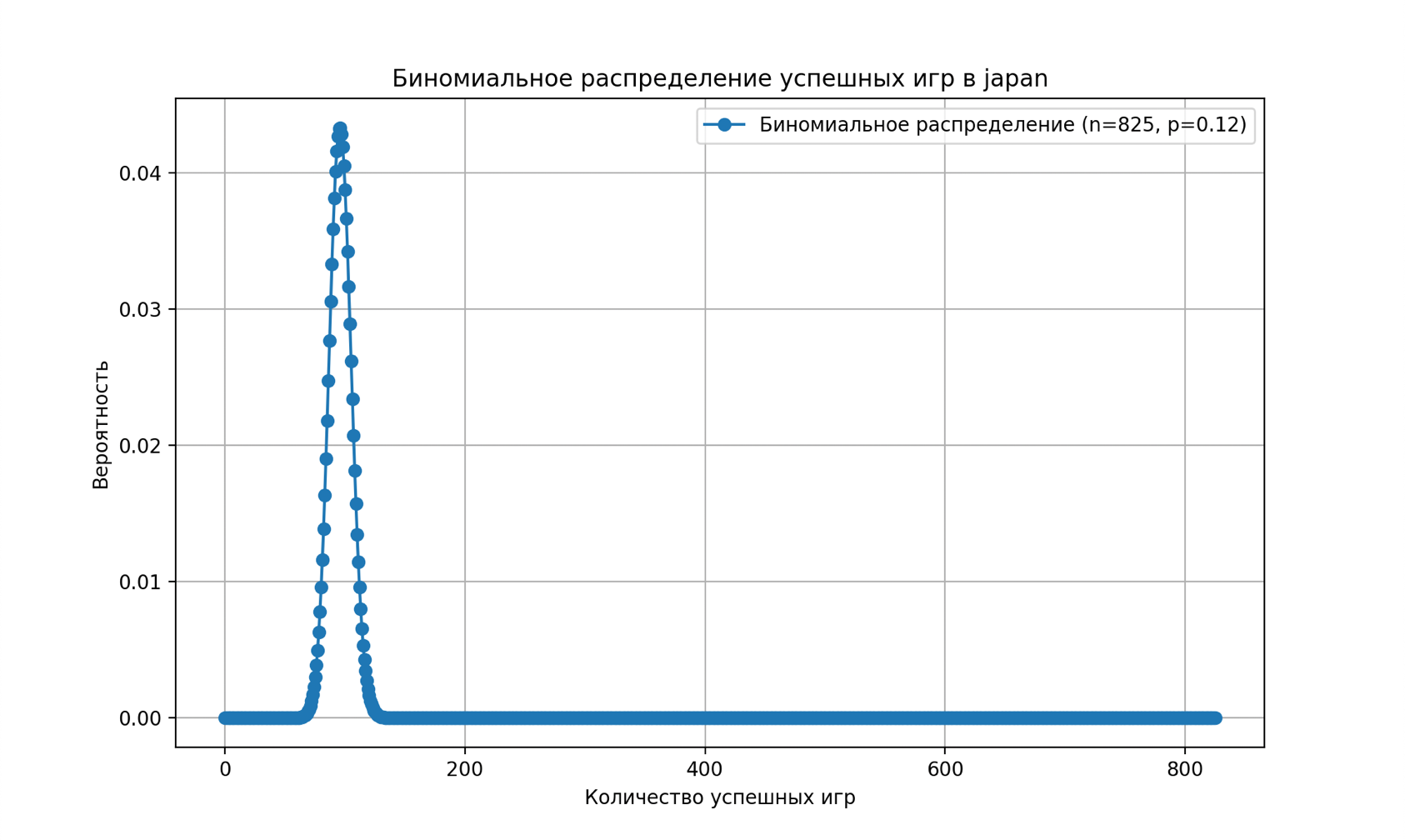


Рис. 9 - график биномиального распределения успешных игр в Японии.

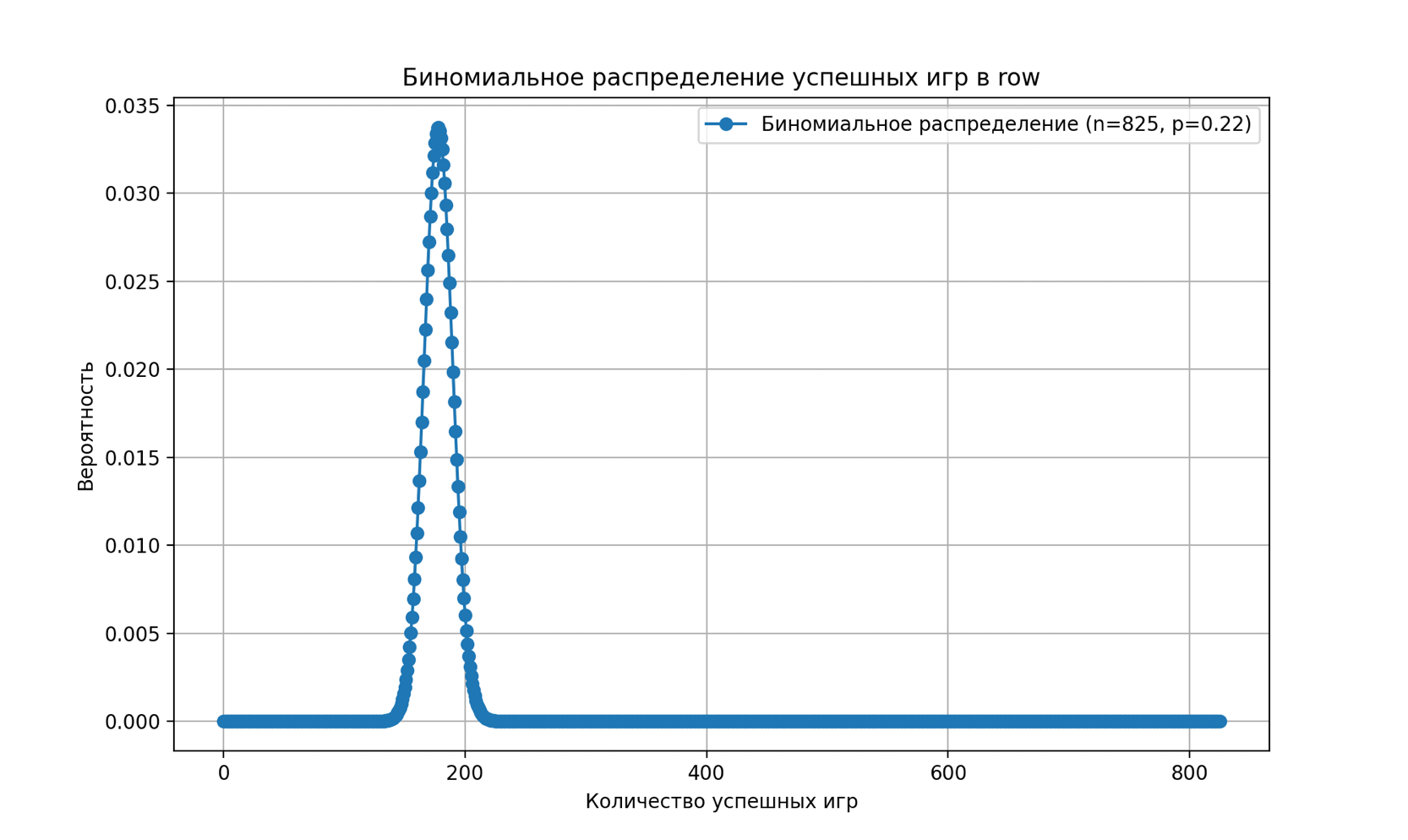


Рис. 10 - график биномиального распределения успешных игр в остальном мире.

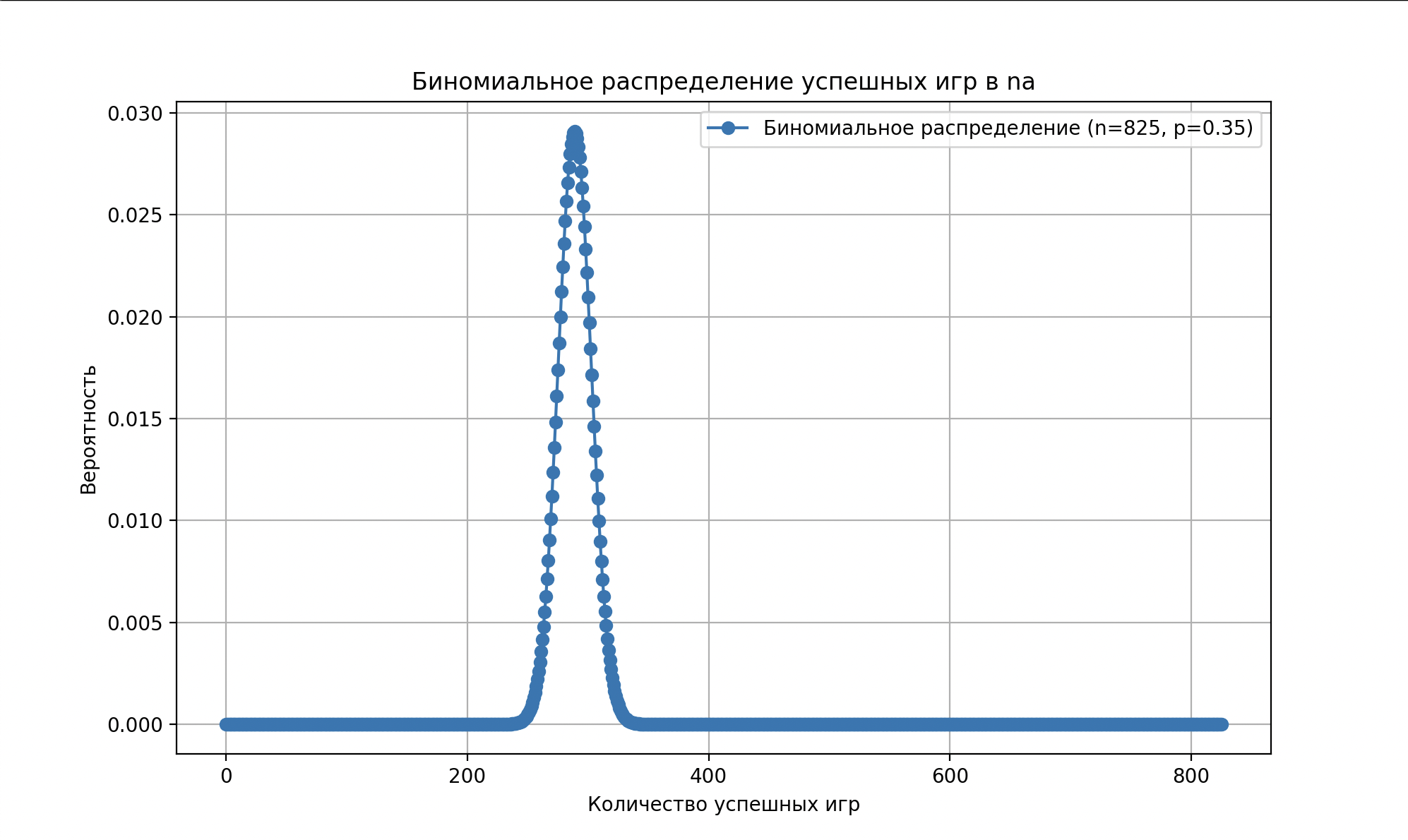


Рис. 11 - график биномиального распределения успешных игр в Северной Америке.