



R для тервера и матстата

Посиделка вторая: варка распределений, Монте-
карло и ЗБЧ

Итоги игры в среднее

- Победили: Лера и Зиля!
- Итоговые две трети от среднего: 14.6
- Многие сказали, что они знают эту игру и знают оптимальную стратегию. И при этом умудрились проиграть... Все, кто сознался в том, что уже играл в эту игру, в табличке зелёные. Все они с точки зрения статистики выбросы.
- Обратите внимание, что есть один выброс в обратную сторону.

Рома	10
Вика Самотканова	54
Настя Петайкина	63
Даша V	1
Надя	10
Марина	10
Зиля	13 +
KLMS (ШТА?)	12
Maxim	30
З.А.	69
Рита	100
Филиппка	0
Реш	10
=')	33,3
Евгений	3,333333333
Михан	12
Лера Лушина	15 +
Кирилл	0
Гвоздева	33
Лена	44
Алёна	0
Саша С.	0
Вета	5
Кирипатьев	0
Саша Дугарова	21
	21,94533333
	14,63022222

Итоги игры в среднее

- Решение игры: поиск равновесия Нэша в этой игре выглядит следующим образом. Давайте отсеивать доминируемые стратегии. Так, числа больше 66 доминируются любым игроком, так как $2/3$ даже от 100 (если вообще все игроки написали на бумажке 100) - меньше 67. Их можно исключить. Как только все игроки использовали эту стратегию, можно выключать числа больше 44, ведь тогда уже никто не запишет цифру больше 66, а $2/3$ от 66 - примерно 44.5. Этот процесс продолжается до тех пор, пока все цифры выше нуля не будут исключены путём итерации алгоритма.
- Игра иллюстрирует отличие между рациональностью игрока и его понятием о рациональности других. Даже абсолютно рациональные игроки не будут называть цифру 0, если только они не знают точно, что остальные игроки абсолютно рациональны.

Итоги игры в среднее

- Все ли игроки будут руководствоваться здравым смыслом?
- Мало кто назовёт ноль. Люди нерациональны.
- Датская газета Politiken: приз 5000 крон, 19196 участников. Результат: 21.6
- Американские Financial Times. Результат: 13
- В аудиториях, где умные люди, обычно итоговое число низкое. Гарвард: 7-8
- [Решение игры от Саватеева](#). На его взгляд игра это тест на то как высоко аудитория оценивает свой интеллектуальный уровень.

**Откуда компьютер
берёт случайности**

Изобретаем велосипед

- Предположим, что мы с вами только что избрели компьютер и нам надо научить его генерировать случайные числа. Как бы вы поступили?

Изобретаем велосипед

- Предположим, что мы с вами только что избрели компьютер и нам надо научить его генерировать случайные числа. Как бы вы поступили?
- **Идея!** Согласно квантовой теории, невозможно узнать наверняка когда произойдёт радиоактивный распад. Давайте положим в компьютер немножечко урана.



Изобретаем велосипед

- Предположим, что мы с вами только что избрели компьютер и нам надо научить его генерировать случайные числа. Как бы вы поступили?
- **Идея!** Согласно квантовой теории, невозможно узнать наверняка когда произойдёт радиоактивный распад. Давайте положим в компьютер немножечко урана.
- **Идея!** Действия человека непредсказуемы. Будем собирать те промежутки времени, которые проходят между нажатиями кнопок на клавиатуре. Это поможет генерировать случайные числа.

Изобретаем велосипед

- Предположим, что мы с вами только что избрели компьютер и нам надо научить его генерировать случайные числа. Как бы вы поступили?
- **Идея!** Согласно квантовой теории, невозможно узнать наверняка когда произойдёт радиоактивный распад. Давайте положим в компьютер немножечко урана.
- **Идея!** Действия человека непредсказуемы. Будем собирать те промежутки времени, которые проходят между нажатиями кнопок на клавиатуре. Это поможет генерировать случайные числа.
- **Идея!** Давайте использовать непредсказуемые шумы в атмосфере. [Пример такого генератора.](#)

Псевдослучайности

- **Идея!** Давайте не будем опираться на природу (это дорого и медленно) и придумаем псевдослучайный алгоритм.

Псевдослучайности

- **Идея!** Давайте не будем опираться на природу (это дорого и медленно) и придумаем псевдослучайный алгоритм.
- Последовательности цифр в числе пи довольно непредсказуема. Давайте окажемся в каком-то месте числа пи и начиная с него начнём генерацию.

Псевдослучайности

- **Идея!** Давайте не будем опираться на природу (это дорого и медленно) и придумаем псевдослучайный алгоритм.
- Последовательности цифр в числе пи довольно непредсказуема. Давайте окажемся в каком-то месте числа пи и начиная с него начнём генерацию. Вся псевдослучайность зависит только от начального значения. Это ненадёжный алгоритм.

Псевдослучайности

- **Идея!** Давайте не будем опираться на природу (это дорого и медленно) и придумаем псевдослучайный алгоритм.
- Последовательности цифр в числе пи довольно непредсказуема. Давайте окажемся в каком-то месте числа пи и начиная с него начнём генерацию. Вся псевдослучайность зависит только от начального значения. Это ненадёжный алгоритм.
- Другой пример [Вихрь Мерсена](#), основанный на простых числах Мерсена.
- Можно придумать более надёжный алгоритм. Некоторые алгоритмы держатся под секретом.
- Легче всего научиться генерировать равномерное распределение.

**Варка
распределений**

Правило

- Сейчас мы с вами будем заниматься варкой распределений. В связи с этим называйте меня мистер Хайзенберг, а я вас буду называть Джесси.
- Если кто-то будет обращаться ко мне в течение варки иначе, я буду его игнорировать



Квантильное преобразование

Теорема:

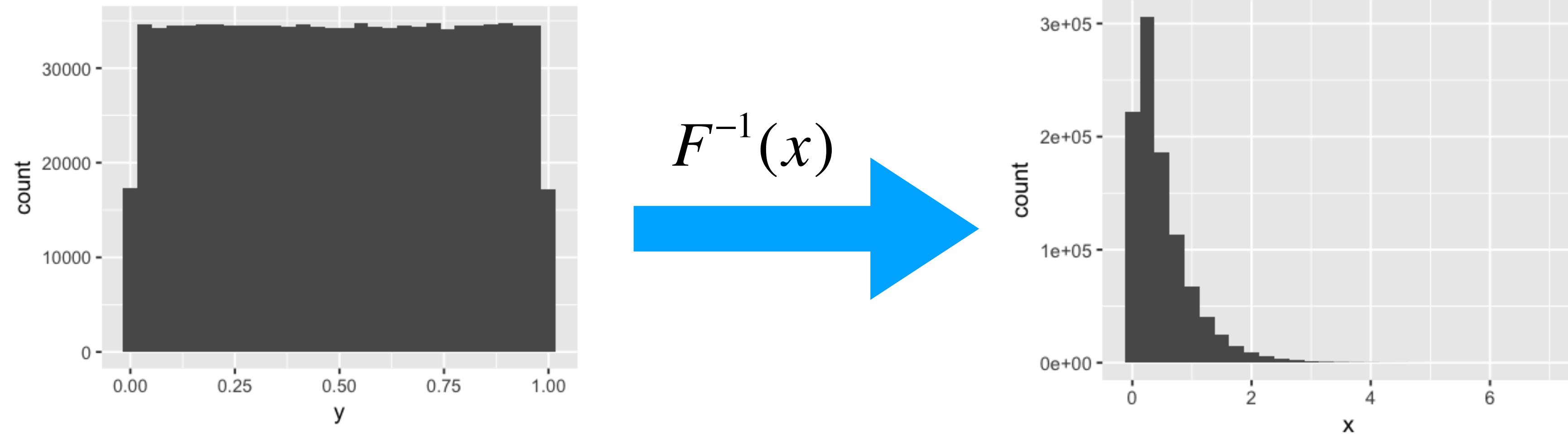
Пусть функция распределения $F_X(x)$ непрерывна. Тогда случайная величина $Y = F(X)$ имеет равномерное распределение на отрезке $[0; 1]$

Следствие:

Пусть $Y \sim U[0; 1]$, а $F(x)$ — произвольная функция распределения. Тогда случайная величина $X = F^{-1}(Y)$ (квантильное преобразование над Y) имеет функцию распределения F

- Доказательство ищите в блокноте с кодом в R

Квантильное преобразование

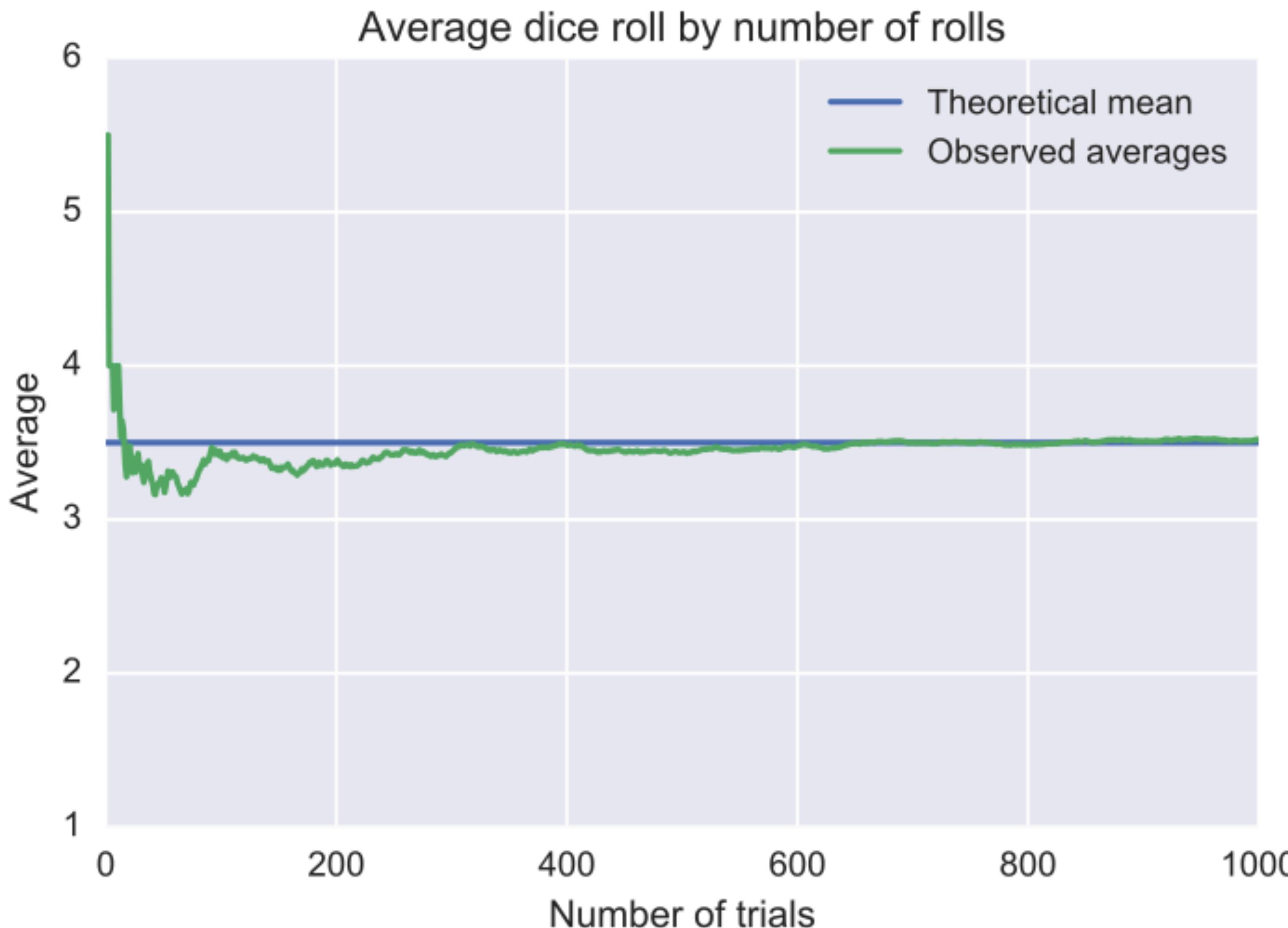


- Позволяет варить из равномерно распределения другие
- Применимо не всегда, например, для нормального распределения используют другие алгоритмы. Почему?

**С варкой всё! Можно
вернуться к нормальным
именам :)**

или нет...

И снова ЗБЧ



Вопрос про больницы

- Есть две больницы: большая и маленькая;



- В обеих принимают роды. Выяснилось, что в одной из больниц оценка вероятности появления мальчика составила 0.7;
- В какой больнице это произошло и почему?

Вопрос про больницы

- Есть две больницы: большая и маленькая;



- Скорее всего, это произошло в маленькой больнице.
При малых объемах выборки вероятность отклониться
от 0.5 больше. Именно это молвят ЗБЧ.

Закон больших чисел

- ЗБЧ утверждает, что среднее арифметическое большого числа похожих случайных величин «стабилизируется» с ростом их числа
- Как бы сильно случайные величины не отклонялись от своего среднего значения, эти отклонения взаимно косятся и среднее арифметическое приближается к постоянной величине
- В общем случае ЗБЧ называется любое утверждение, которое говорит, что:

$$\frac{X_1 + \dots + X_n}{n} - \frac{E(X_1) + \dots + E(X_n)}{n} \xrightarrow{p} 0$$

- ЗБЧ сформулировано довольно много: Чебышёва, Бернулли, Хинчина и тд

Слабая форма ЗБЧ (Чебышёв)

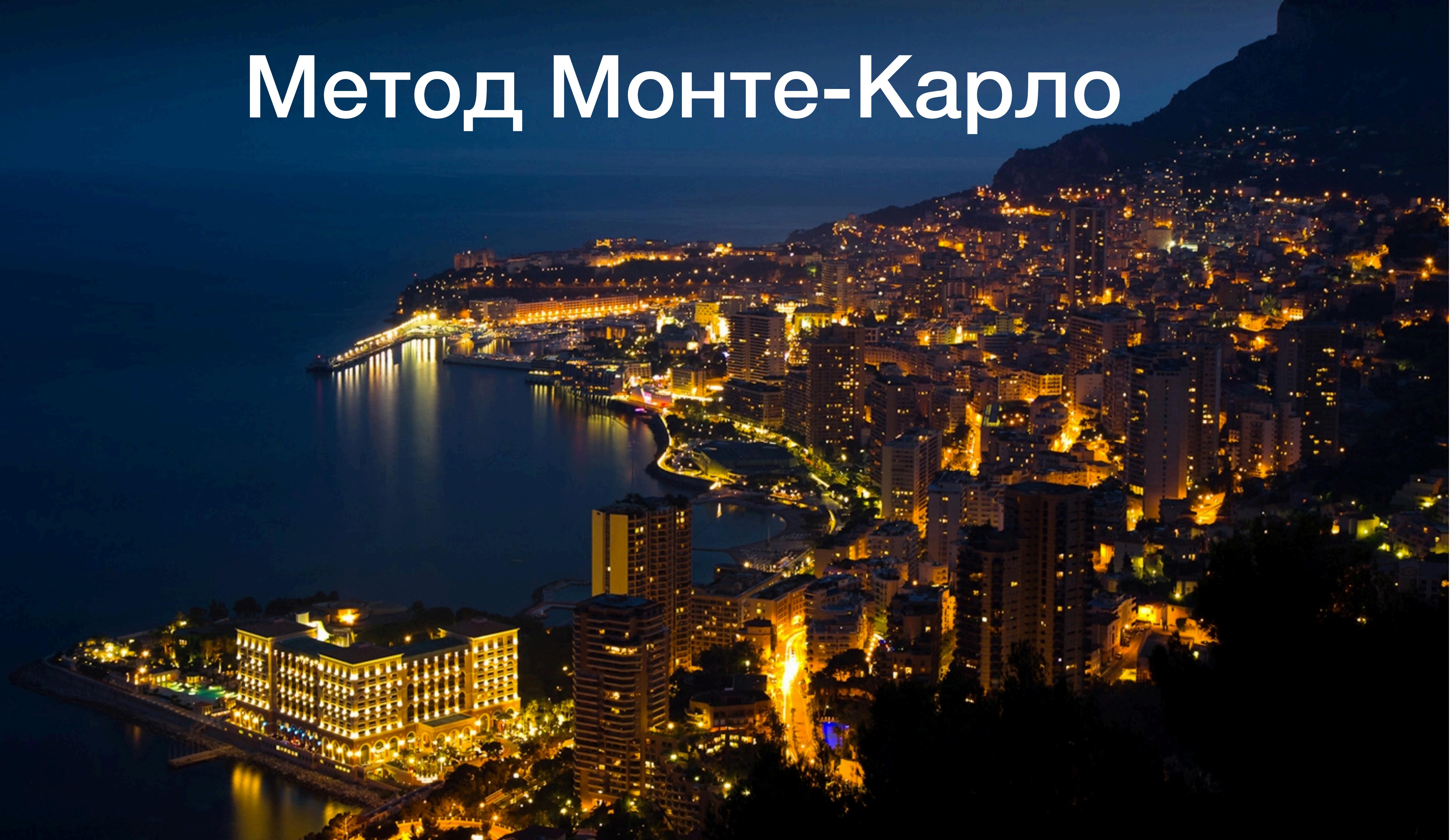
Пусть X_1, \dots, X_n попарно независимые и одинаково распределённые случайные величины с конечным вторым моментом, $E(X_i^2) < \infty$, тогда имеет место сходимость:

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{p} E(X_1)$$

Страховые компании

- ЗБЧ - уникальная теорема. Примерно с 1600-х годов она позволяет зарабатывать деньги на страховках.
- Упражнение: для 25-летней девушки вероятность прожить ещё год составляет 0.9. Страховка в год стоит 1000 рублей при взносе в 110 рублей. Какой будет средняя прибыль компании с одной страховки?

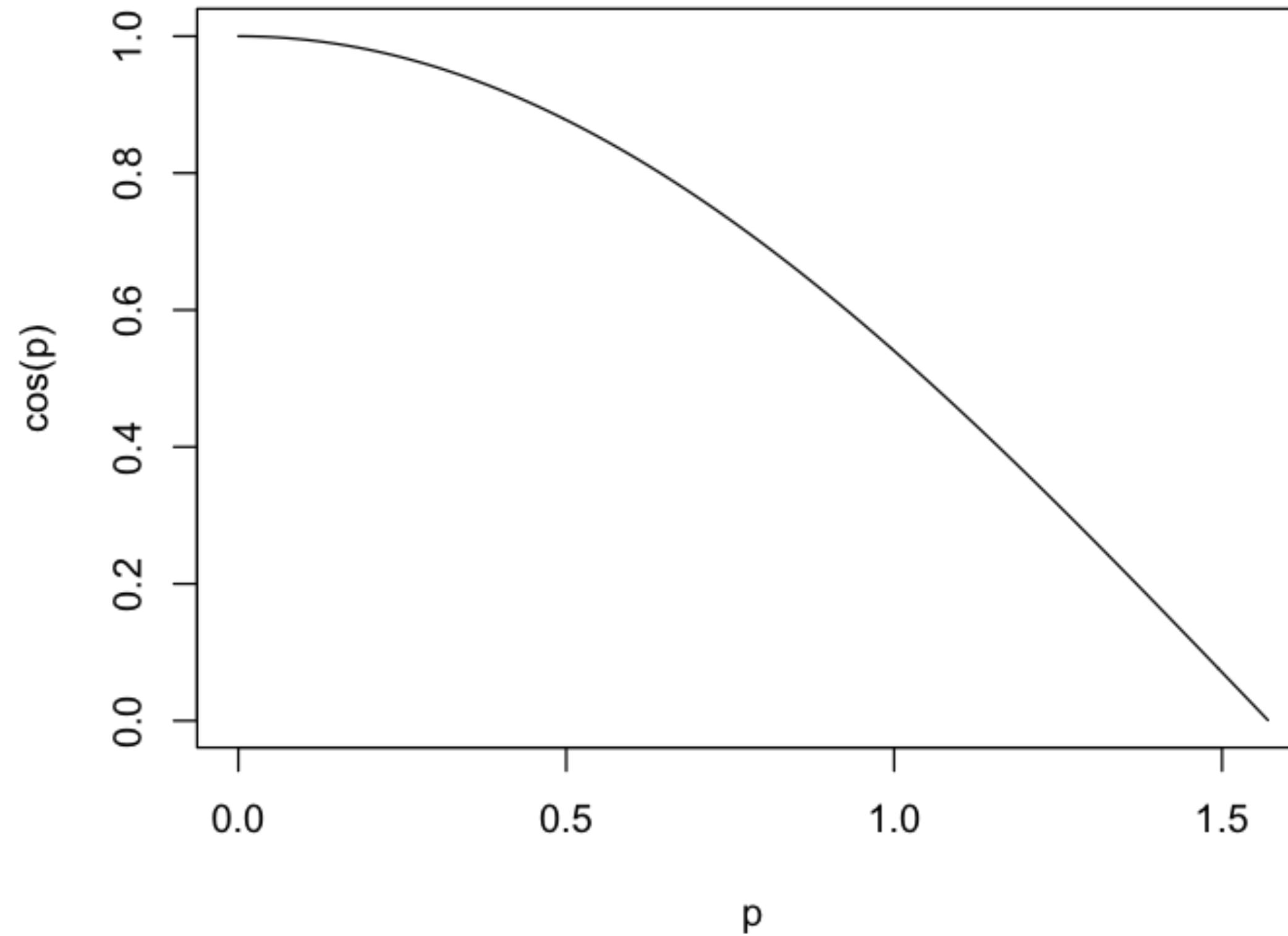
Метод Монте-Карло



Шо за Монте-Карло?

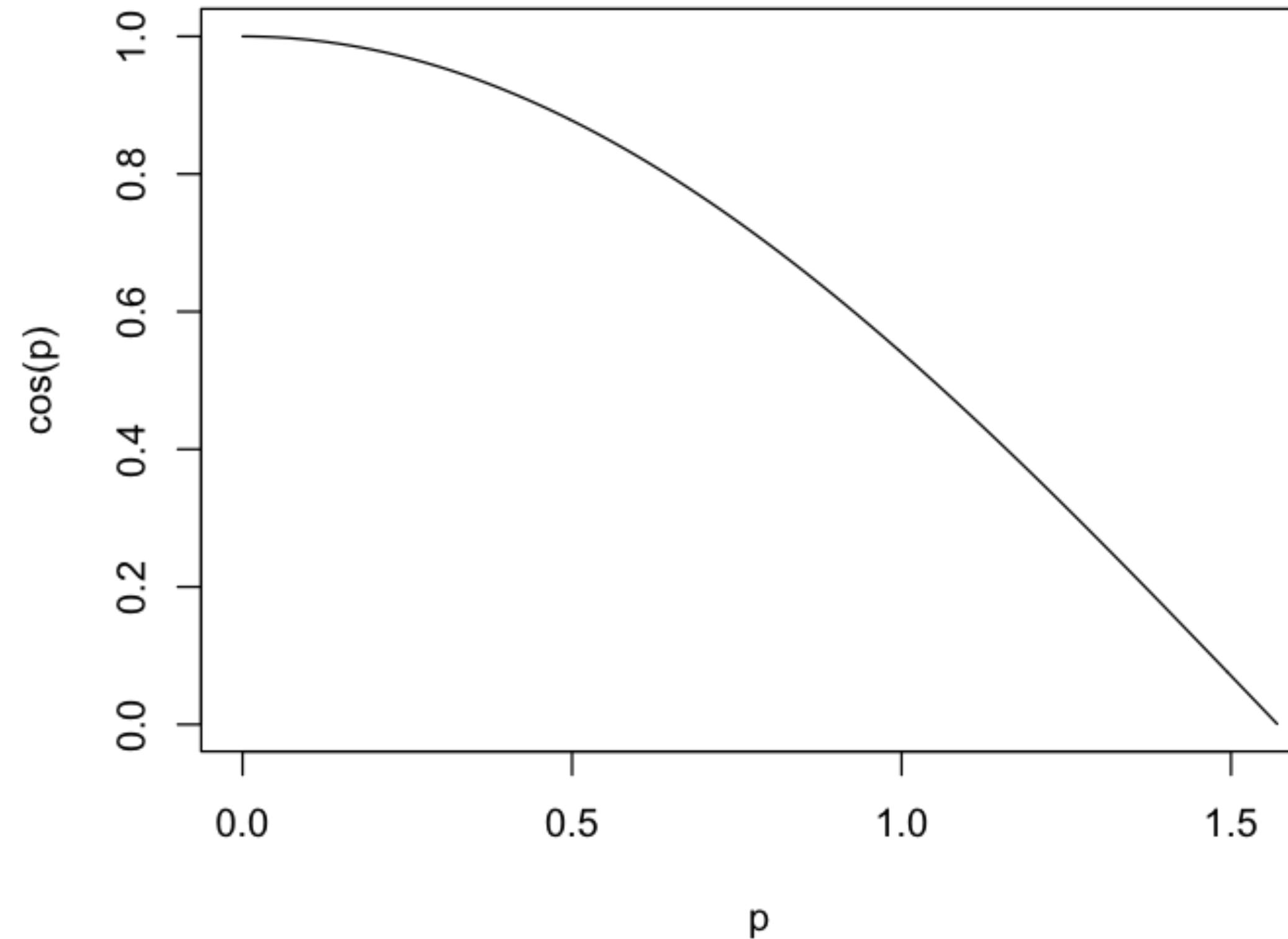
- Монте-Карло – административная территория княжества Монако, крупнейший район страны, расположенный на территории одноименной коммуны в Монако. Город известен своими казино, пляжами и пользуется популярностью у представителей высшего общества. В Монте-Карло начинается и завершается ежегодное ралли «Монте-Карло».
- Так говорит Википедия, но мы будем говорить про метод работы со случайными величинами.

Поиск площади



- Как найти площадь под косинусом?

Поиск площади

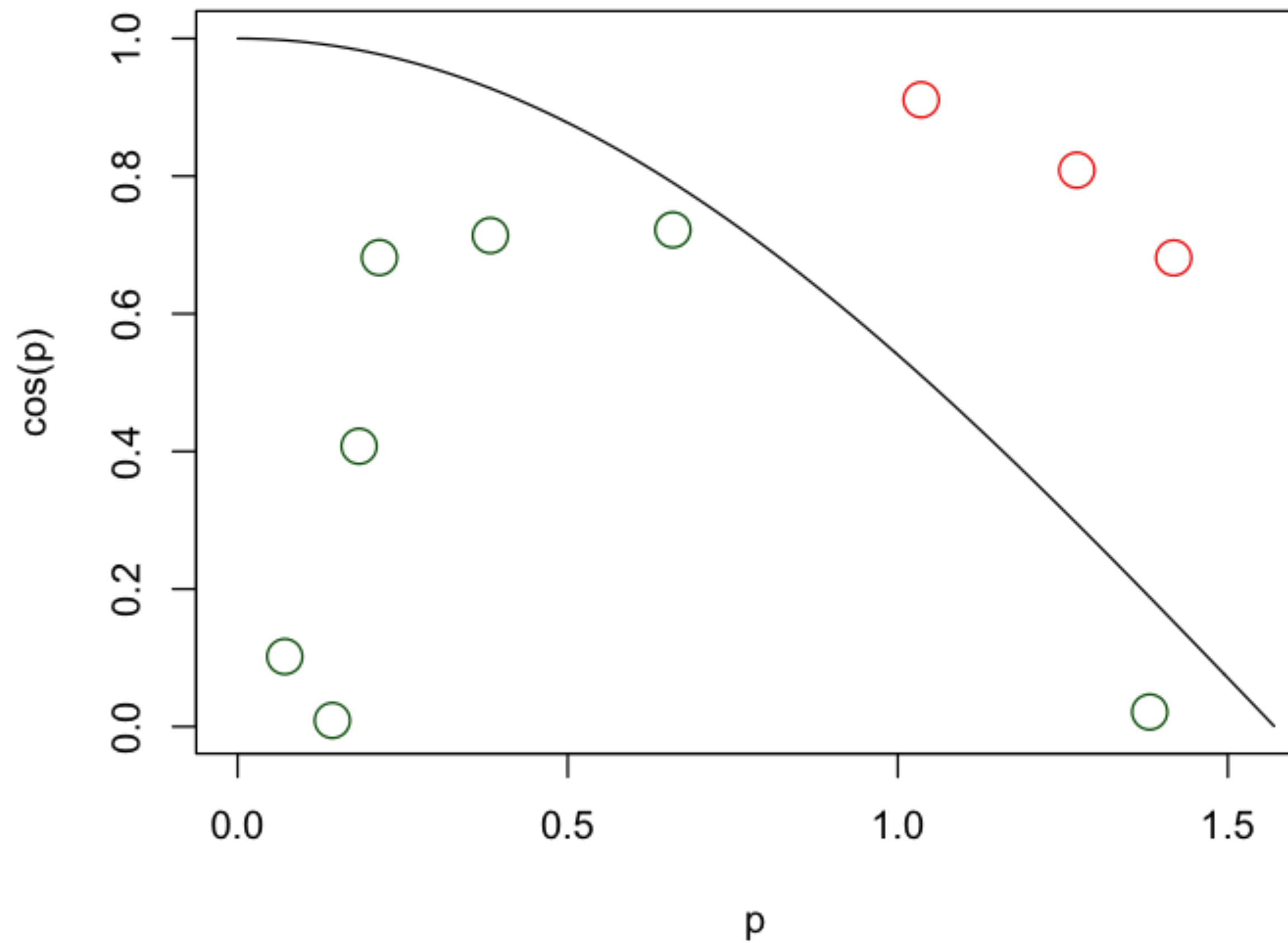


- Проинтегрировать!
- А если под другой, более сложной фигурой?

Идея!

- Будем бомбардировать нашу фигуру точками из равномерного распределения, определённого на прямоугольнике.

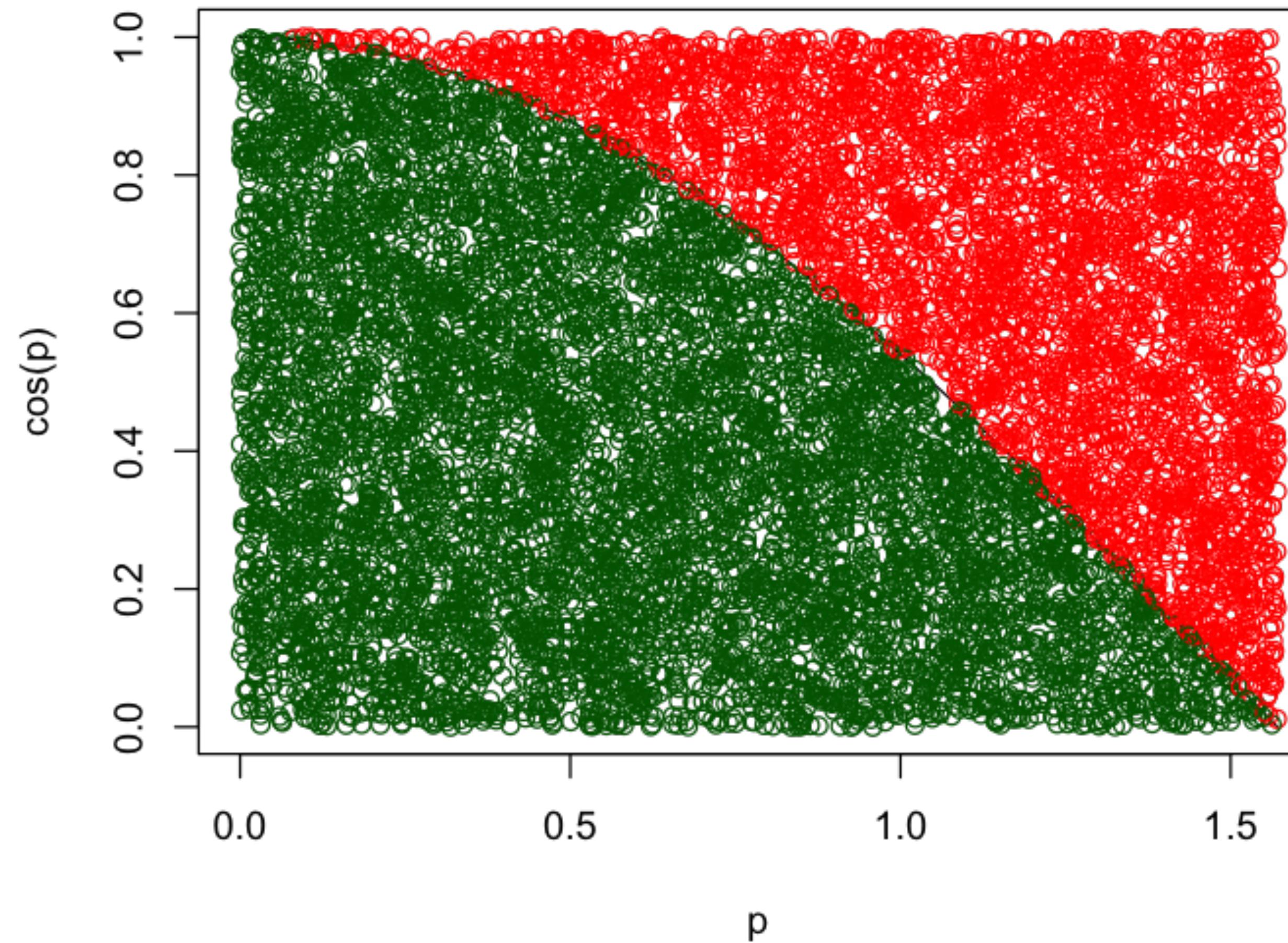
Поиск площади



Поиск площади

- Только часть точек попала под кривую.
- Доля точек, лежащих под кривой, умноженная на площади прямоугольника, стремится к истиной площади под кривой.

Поиск площади



Метод Монте-Карло

- Метод Монте-Карло это общее название группы численных методов, основанных на получении большого числа реализаций случайного процесса, который формируется таким образом, чтобы его вероятностные характеристики совпадали с аналогичными величинами решаемой задач.
- В принципе, всё, что мы делали до этого для оценки математических ожиданий и вероятностей, это и есть метод Монте-Карло.
- Метод можно расширить до строительства доверительных интервалов.

Метод Монте-Карло

- Помните задачку про Мишу, который торгует газетами у себя на районе? Он посмотрел сколько покупателей у него было и собрал выборку

три дня назад	позавчера	вчера	сегодня
25	75	30	70

- После он предположил, что число покупателей имеет распределение Пуассона (выбрал семейство распределений) и оценил его параметры (зафитил распределение).
- На самом деле неплохо было бы проверить гипотезу о том, что выборка действительно из распределения Пуассона.
- После Миша сгенерил из распределения кучу наблюдений. Для спроса. На основе этих наблюдений он может выяснить как распределена прибыль.
- В следующий раз мы посмотрим как такое делается на финансовом рынке.

- Кстати, именно в Монте-Карло состоялась первая схватка Железного Человека с Иваном Ванко (Хлыстом)



Строим доверительные интервалы!

В ответ назовите интервал, в который на ваш взгляд с 90% вероятностью попадает ответ. Если я спрошу количество улиц в Москве, вы должны написать: я на 90% уверен, что их от 2 до 4 тысяч. Понятно, что гуглить нельзя. Иначе эксперимент провалится.

1. Население Никарагуа
2. Вес статуи свободы без постамента
3. Площадь Сахары
4. Стоимость одного грамма иридия
5. Размер Пулитцеровской премии