

Как разбивать на группы в A/B тесте?

Почему нельзя случайно разбивать?

Пользователь зашел на сайт 1 раз, попал в тестовую группу
Когда он зайдет 2 раз, он должен снова попасть в тестовую группу.

Вывод – надо уметь консистентно разбивать пользователей так, чтобы внутри одного теста не менялись группы

Как разбивать на группы в A/B тесте?

Как разбить случайно, но при этом консистентно?

Наивный способ: взять hash от user id, перевести в число, взять остаток от деления на максимальное число

В чем проблема?

Как разбивать на группы в A/B тесте?

Как разбить случайно, но при этом консистентно?

Наивный способ: взять hash от user id, перевести в число, взять остаток от деления на максимальное число

В чем проблема?

Между разными тестами юзер попадает в одну и ту же группу всегда, так как hash не меняется

Следовательно, тесты влияют друг на друга
Невозможно отделить эффект одного теста от другого

Как это решить?

Как разбивать на группы в A/B тесте?

Как разбить случайно, но при этом консистентно?

Наивный способ: взять hash от user id, перевести в число, взять остаток от деления на максимальное число

В чем проблема?

Между разными тестами юзер попадает в одну и ту же группу всегда, так как hash не меняется

Следовательно, тесты влияют друг на друга

Невозможно отделить эффект одного теста от другого

Как это решить?

Нужно добавить уникальную для эксперимента «соль» (просто строка) к user id

Как разбивать на группы в A/B тесте?

Как это решить?

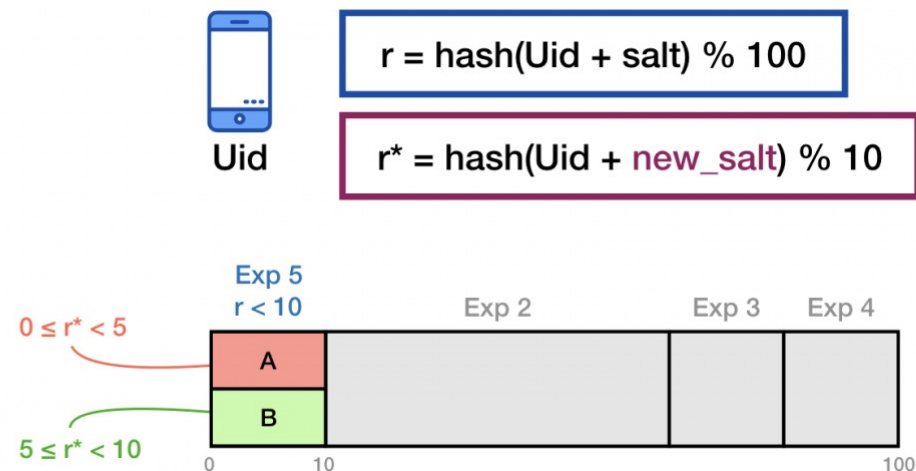
Нужно добавить уникальную для эксперимента «соль»
(просто строка) к user id

Как это решит проблему?

В разных тестах пользователи будут попадать случайно в разные группы.

При этом мы можем запоминать соль для каждого отдельного теста и легко воспроизводить группы в каждом тесте.

Также мы можем запускать много экспериментов одновременно, так как теперь между тестами не будет зависимости



Зачем нужны A/A тесты?

Задача A/A тестов заключается в том, чтобы проверить, работает ли система сплитования корректно

Нам бы хотелось отвергать нулевую гипотезу в одинаковых группах (т.е. совершать O1P) не чаще, чем заданный уровень значимости

Этапы проведения A/A теста

1. Проводим A/A тест. Время на A/A определяется таким образом, чтобы охватить как можно больше факторов влияния на метрику (например, недельная сезонность)
2. Симулируем новые A/A. Тест пересчитывается > 1 тыс. раз при помощи симуляции новых A/A
3. Считаем стат. значимость. В каждом тесте считается p-value при помощи статистического оценщика (бутстрап, t-тест и т.п.)
4. Считаем метрику FPR, False Positive Rate (т.е. O1P)
5. Проверяем, что наш $FPR < \alpha$ (уровень значимости). Если условие сработало корректно, то сплитовалка работает правильно

Как считать FPR

Для проверки качества сплитовалки считаем долю ложных прокрасов FPR на A/A группах:

$$FPR = \frac{\sum_{i=1}^N I\{P \leq \alpha\}}{N} = \frac{FP}{N},$$

I – индикаторная функция

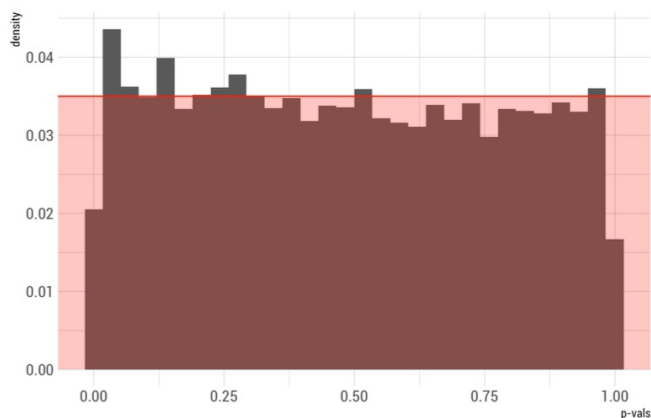
P – p-value

N – количество синтетических тестов

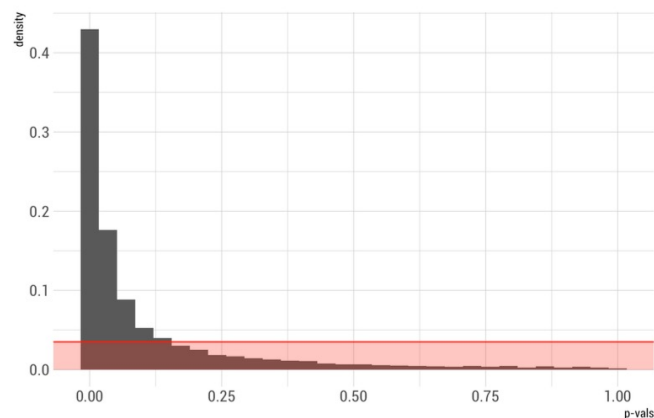
Распределение p-value

При верной нулевой гипотезе p-value распределен равномерно

Корректная сплит-система



Сломанная сплит-система



Проверить можно глазами, а еще лучше – провести тест Колмогорова-Смирнова на равномерное распределение