

Experiments

RecSys Wrap Up

Где можно использовать рекомендации

- Мобильное приложение/Доставка:
 - «рекомендуемые товары»
 - «с этим товаром покупают»
- Электронная почта / SMS / push
 - устойчивые бандлы товаров
 - рекомендуемые товары (до 3х штук)
 - рекомендуемые категории (1-2-3 категории)

Эвристики вместо перс.реко?

Эвристики вместо перс.реко

- Топ-10/20/30/40 товаров
- Новинки
- Схожие темы
- И т.п.

Алгоритмическое решение

- Ассоциативные правила
- ALS и его вариации:
 - классический ALS (нам доступны явные оценки)
 - implicit ALS (когда работаем с неявными оценками)
 - weighted ALS
- Другие алго – развлечение исследователей 😊

Алгоритмическое решение

- Ассоциативные правила
- ALS и его вариации:
 - классический ALS (нам доступны явные оценки)
 - implicit ALS (когда работаем с неявными оценками)
 - weighted ALS
- Другие алго – развлечение исследователей 😊

Оценка качества?

Позиция	Релевантность
1	1
2	0
3	1
4	1
5	0
6	0
7	1
...	...

$$\text{precision@5} = \frac{3}{5}$$

$$\text{recall@5} = \frac{3}{4} \text{ (если всего 4 релевантных айтема)}$$

Оценка качества?

user	item			
	7	8	10	4
	1	8		
	6		6	
		7		6
	8		3	

Оценка качества

- Если оцениваем бинарный отклик, то:
 - precision@k, recall@k
- Если оцениваем численный отклик, то:
 - MAE@k, MSE@k

Оценка качества

- Покрытие
 - Какая доля товаров рекомендовалась хотя бы раз?
 - Какой доле пользователей хотя бы раз показаны рекомендации?
- Новизна
 - Как много рекомендованных товаров пользователь встречал раньше?
- Прозорливость (serendipity)
 - Способность предлагать товары, которые отличаются от купленных ранее

Схема систем

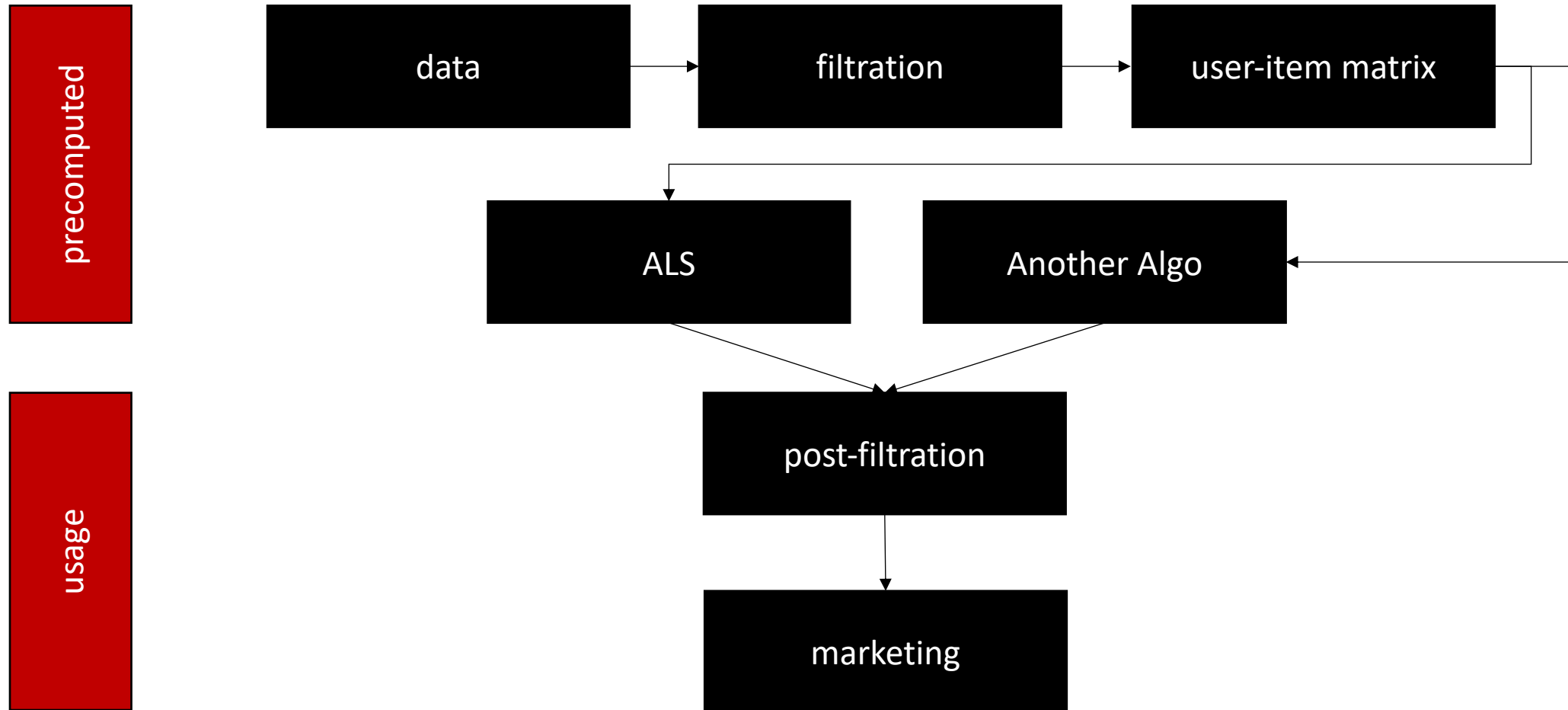
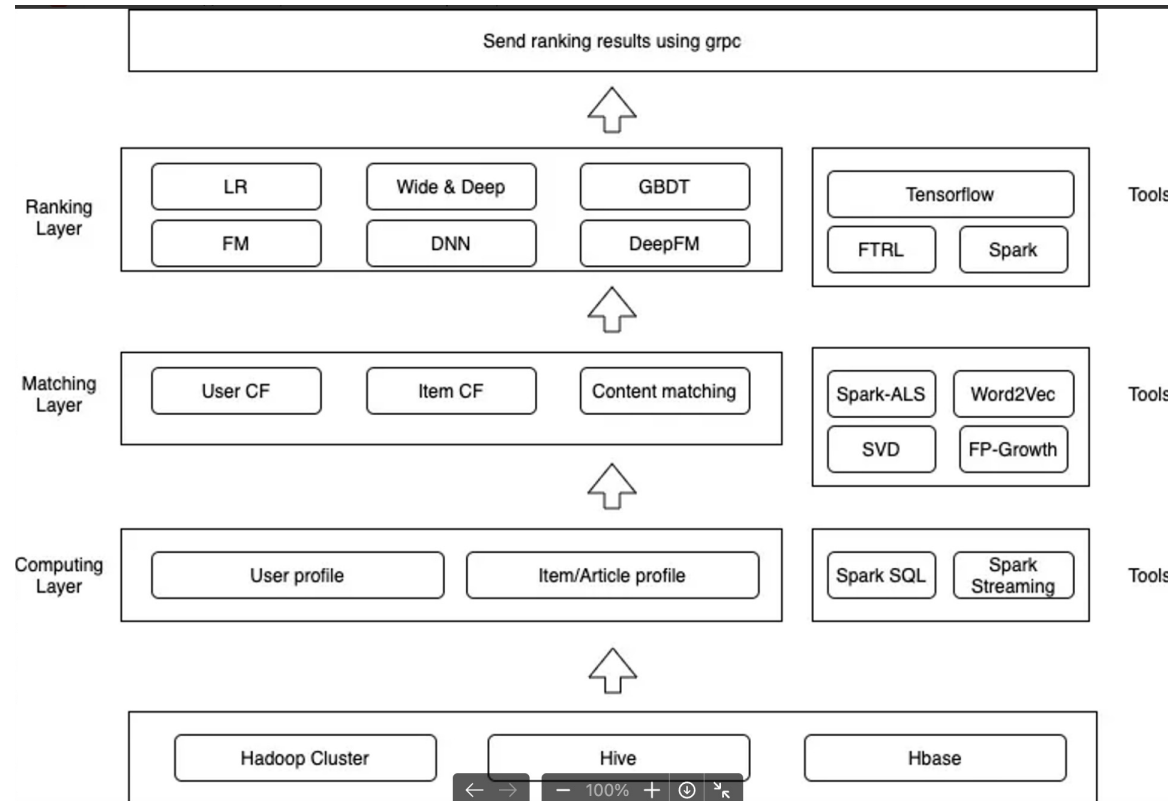


Схема систем

второй уровень

первый уровень



Experiments

Какие экспериментальные дизайны вы знаете?

Классификация экспериментов

Обсервационные

Исследователи не вмешиваются активно в события, наблюдают в естественном развитии

- Естественные эксперименты
- Квази-эксперименты
- Когортные исследования
- Поперечные исследования
- Случай-контроль

Контролируемые

Исследователи активно вмешиваются в течение событий, создают контрольные и экспериментальные группы

- Рандомизированные испытания:
 - Контролируемые
 - Слепые
 - Двойные

Что такое АБ-тест?

АБ-тест -

это вид статистического (рандомизированного) эксперимента, в котором обычно фигурирует две группы: одна группа (тестовая) получает изменение, а вторая (контрольная) – нет.

- Примеры изменений:
 - Цвет кнопки на сайте
 - Порядок расположения элементов на сайте
 - Скрипт телефонного звонка
 - Внедрение нового алгоритма рекомендаций
 - Расположение отделов внутри магазинов

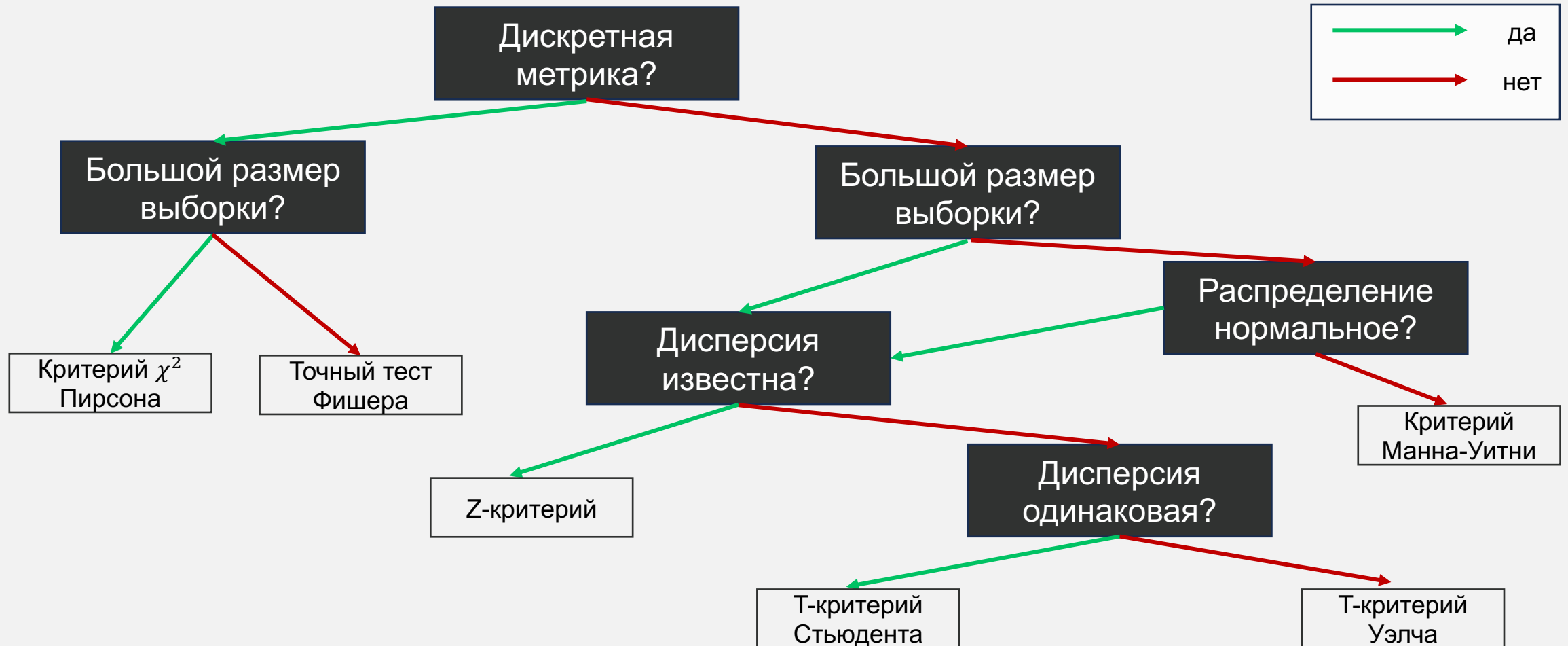
Общая схема АБ-теста



Что стоит помнить

- КОМПОНЕНТЫ ГИПОТЕЗЫ
 - population
 - intervention
 - comparison
 - outcome
 - time
- концепт p-value
- концептуальные формулировки ЗБЧ и ЦПТ
- ошибки 1-2 рода
- основные тесты

Критерии для АВ-тестов



Сильные стороны АБ-тестов

Сильные стороны АБ-тестов?

- понятно, что есть эффект = разница в значении метрики между группами
- большая развитая наука
- можно проводить онлайн и оффлайн
- нивелирует highest paid person opinion

Неудачные для АБ ниши

Неудачные для АБ ниши

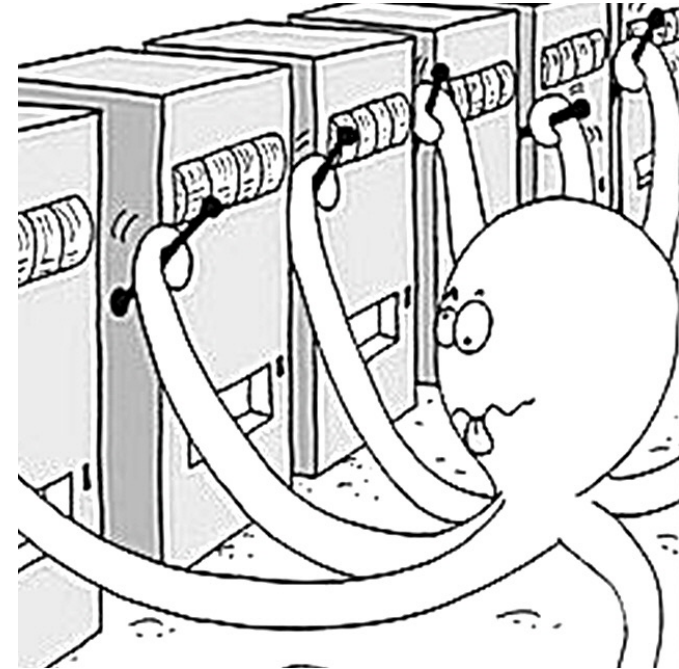
- много вариантов для проверки
- сетевой эффект
- невозможность выделить контрольную группу

Неудачные для АБ ниши

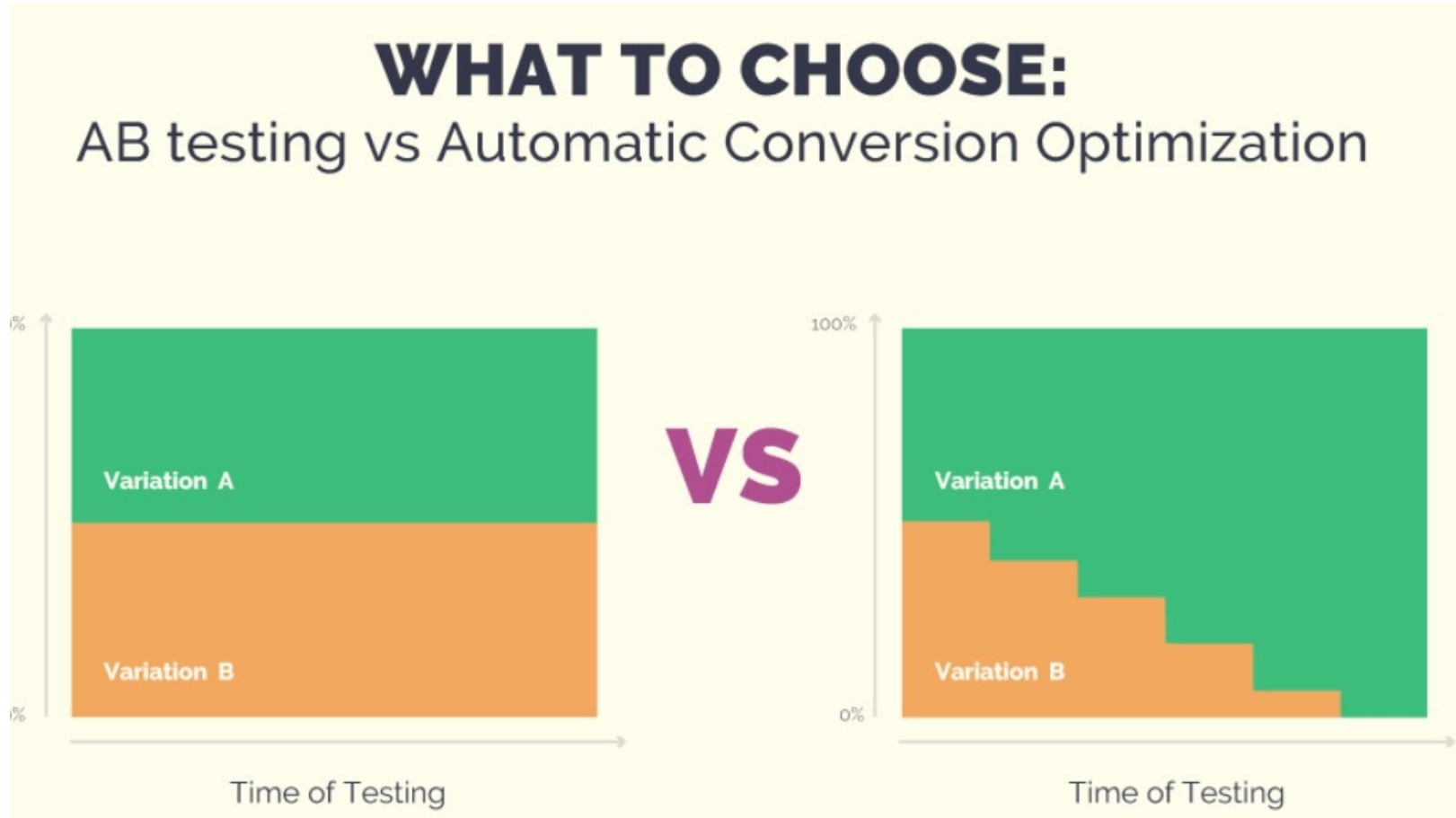
- много вариантов для проверки: A/B/C/D/E... тесты
 - многорукие бандиты
- сетевой эффект
 - свитчбэк-эксперименты
- невозможность выделить контрольную групп
 - синтетический контроль

Многорукие бандиты

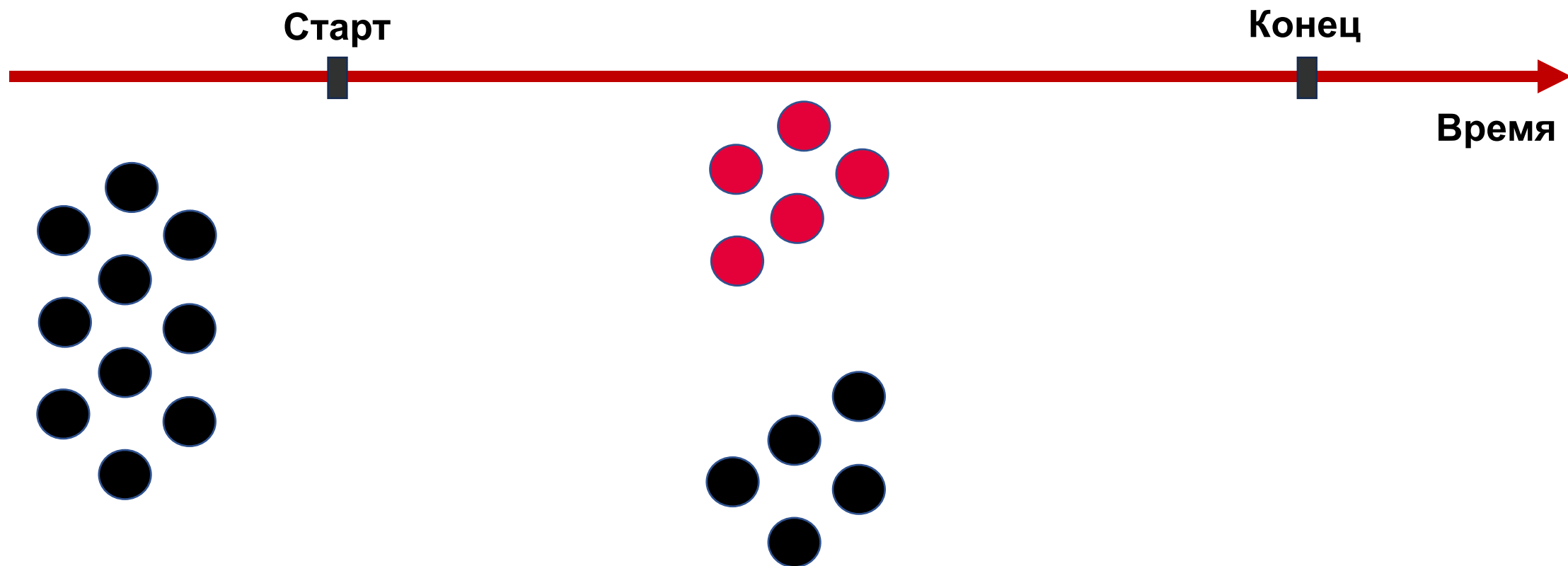
- Есть X автоматов, у каждого ручка
- Хотим оптимизировать суммарный выигрыш и найти автомат, который дает больший выигрыш
- Заранее не знаем, в какой автомат играть
- Играем и смотрим, какой автомат классный
- На нем и играем



Многорукие бандиты



АВ-тесты

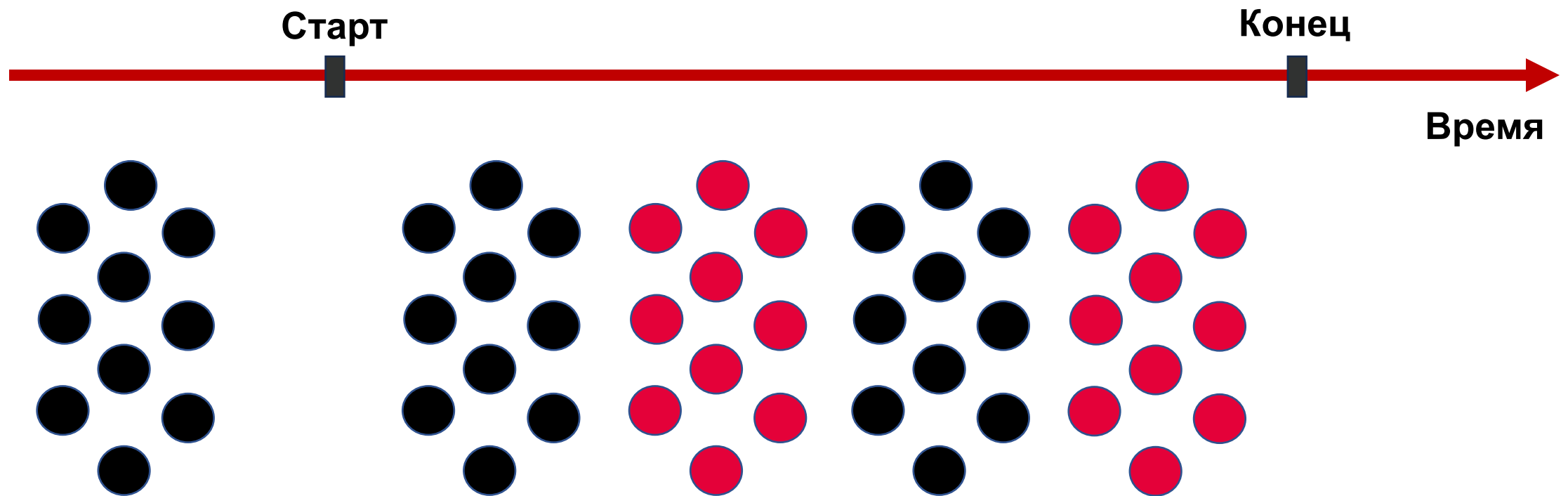


Свитчбэк эксперименты

Мотивация:

- сетевой эффект – не можем сепарировать интервенцию, все объекты будут подвержены шлейфу
- пример:
 - новый алгоритм прайсинга курьеров за доставку заказов
 - новые настройки в системе распределения заказов на сборщиков и курьеров
- много свитчбэков в ride-hailing

Свитчбэк эксперименты



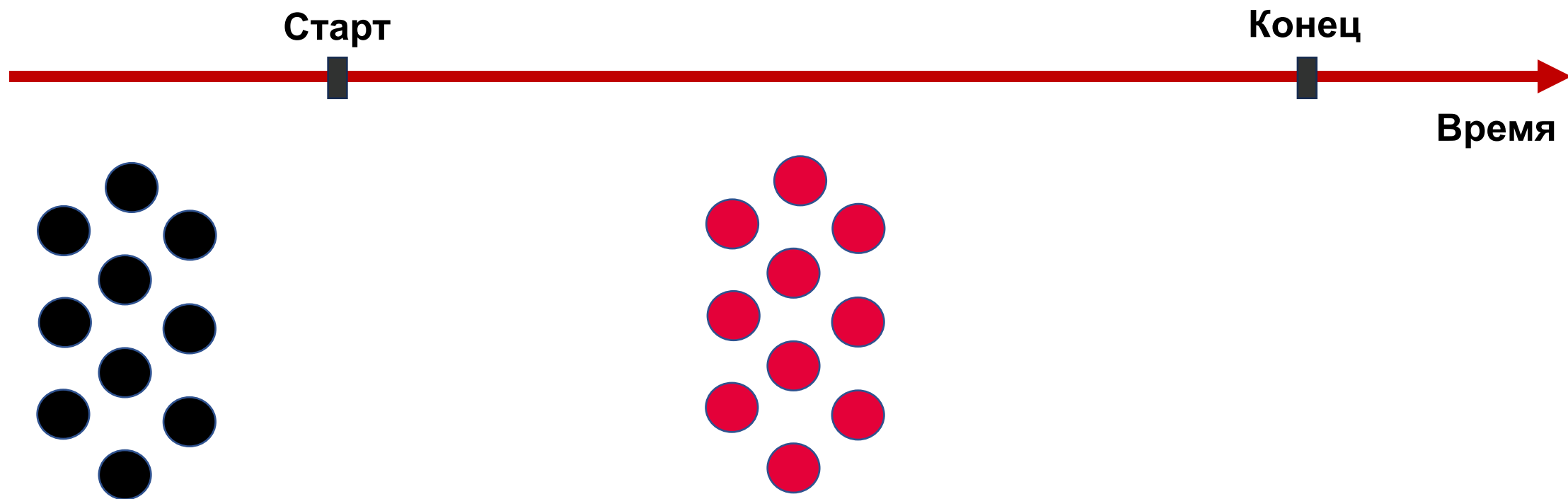
ВЫЗОВЫ СВИТЧБЭКОВ

- Как вообще их считать

ВЫЗОВЫ СВИТЧБЭКОВ

- Как вообще их считать
- Какой длительности интервалы должны быть
- Как должны переключаться между собой
- Как нужно агрегировать статистику для расчетов:
 - поюнитно
 - не агрегировать
- Как нивелировать влияние внешних факторов

Синтетический контроль



Синтетический контроль

- Пример:
 - «Черная пятница»
 - новогодние акции
 - и тп
- Цель: привлечь пользователей, трафик, денежный поток
- Проблема: интервенцию невозможно изолировать от всех объектов

Синтетический контроль

- Пример:
 - «Черная пятница»
 - новогодние акции
 - и тп
- Цель: привлечь пользователей, трафик, денежный поток
- Проблема: интервенцию невозможно изолировать от всех объектов

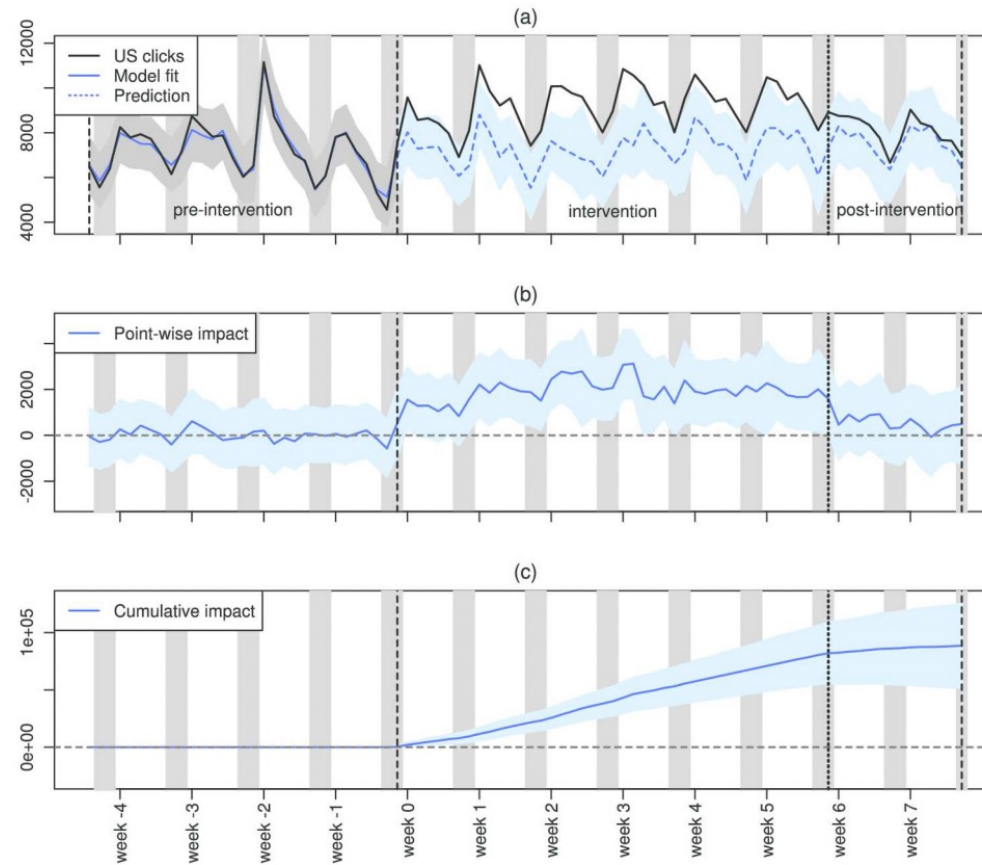
Синтетический контроль

- как правило, вопрос касается каких-то временных рядов:
 - как выглядела бы кривая продаж без интервенции?
 - как менялось бы число пользователей без интервенции?
- идея методов:
 - восстановим синтетическую кривую, а как было бы в отсутствии интервенции
 - в принципе, можно использовать любой подход, чтобы смоделировать кривую
- разница между двумя кривыми = эффект от интервенции

Causal Impact

- Тулза от гугла 6-летней давности
- временные ряды обычно раскладывают три компоненты:
 - тренд, сезонность, ошибки
- CI делает примерно так же:
 - тренд
 - сезонность
 - байесовская регрессия на ковариаты
 - ошибки

Causal Impact



Ссылки

- Рон Кохави «Доверительное АБ-тестирование»
- [Видео](#) Михаила Каменщикова про бандитов в Авито
- [Видео](#) Евгения Комарова про бандитов в Яндекс.Лавке
- Свежая большая [статья](#) про свитчбэки
- [Статья](#) про causal impact
- [Видео](#) про causal impact
- [Как](#) Лента использует Causal Impact

Ссылки по recsys

- ARL-туториал
- <https://habr.com/ru/company/ods/blog/353502/>
- ALS-туториал
<https://medium.com/radon-dev/als-implicit-collaborative-filtering-5ed653ba39fe>
- ALS и BPR-туториал
<https://towardsdatascience.com/recommender-system-using-bayesian-personalized-ranking-d30e98bba0b9>
- Обзор методов:
<https://habr.com/ru/post/486802/>