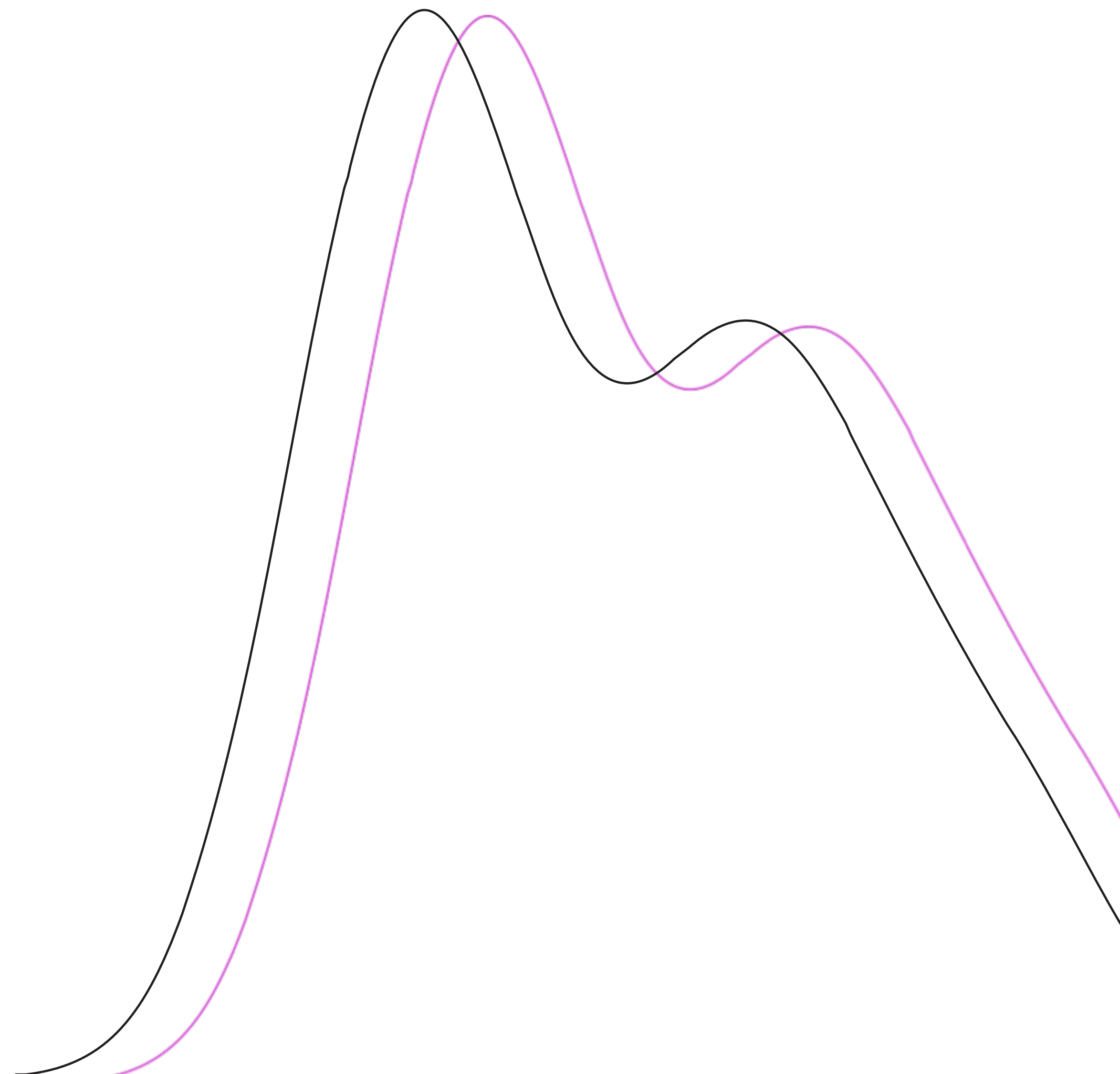


$e^x \text{pf}$

Ускорение экспериментов с помощью Variance-Weighted Estimator

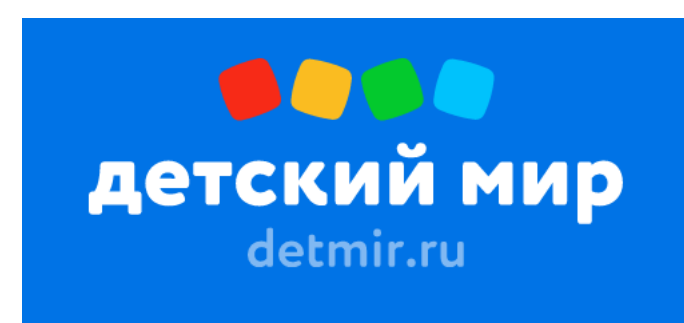
Сёмин Егор, аналитик EXPF



$e^x p f$

Мы консультируем продуктовые команды по анализу онлайн и офлайн экспериментов, начиная от культуры и сбора данных, заканчивая автоматизацией

Клиенты



Sigma

$e^x p f$

Платформа для менеджмента экспериментов и фича флагов

Сервис, позволяющий тестировать продуктовые гипотезы на сайте или в приложении

Также сервис позволяет включать и выключать фича флаги приложения или сайта без развертывания нового кода

expf.ru/sigma



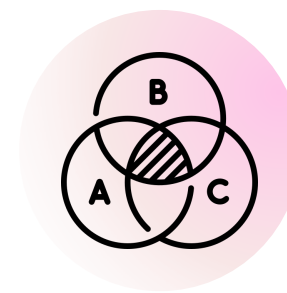
Неограниченный
запуск A/B



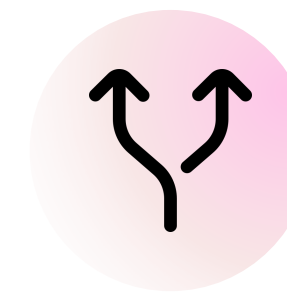
Отправляйте данные
куда угодно



Система слоев



A/B без и с
пересечениями



A/B Split URL



Безопасность ваших
данных



Client + Server
sides SDKs



Сегментация и
таргетирование



A/B по спискам
пользователей

О чем поговорим?

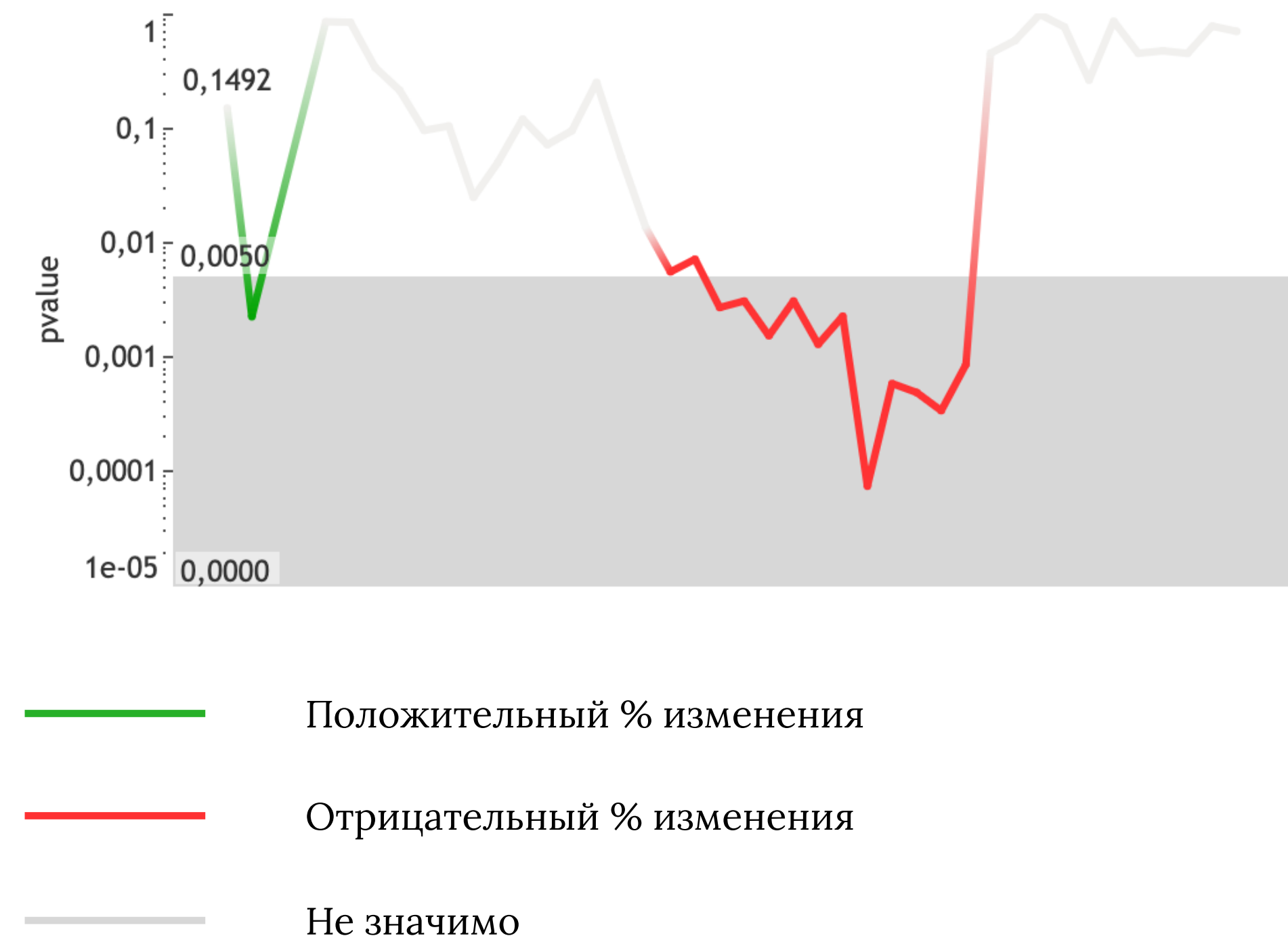
- 01 Как сокращение дисперсии влияет на время проведения эксперимента
- 02 Обзор распространённых способов сокращения дисперсии
- 03 VWE: канон Facebook и бакетизация от EXPF
- 04 Пример реализации

$e^x p f$

Зачем сокращать дисперсию

- «Сколько времени нужно на эксперимент?» – самый частый вопрос
- Алгоритм обычно выглядит так:
 - Задаём начальные параметры
 - Считаем MDE (Minimum Detectable Effect)
 - Вычисляем необходимый размер выборки для каждой метрики
 - Вычисляем из этих данных длительность самого эксперимента

Изменение p-value во время A/B



Зачем сокращать дисперсию

Что можно сделать:

- Мощность метрики можно увеличить элементарно: увеличить выборку -> продлить эксперимент на 1/2/... недель

Тогда:

- Страдает Experiment velocity (кол-во активных экспериментов на единицу времени)
- Теряем в часах аналитика, который следит за А/В
- Упускаем в дополнительной выгоде, выраженной в деньгах
- Стопорим очередь, если в компании используются «слои»

Изменение p-value во время А/В

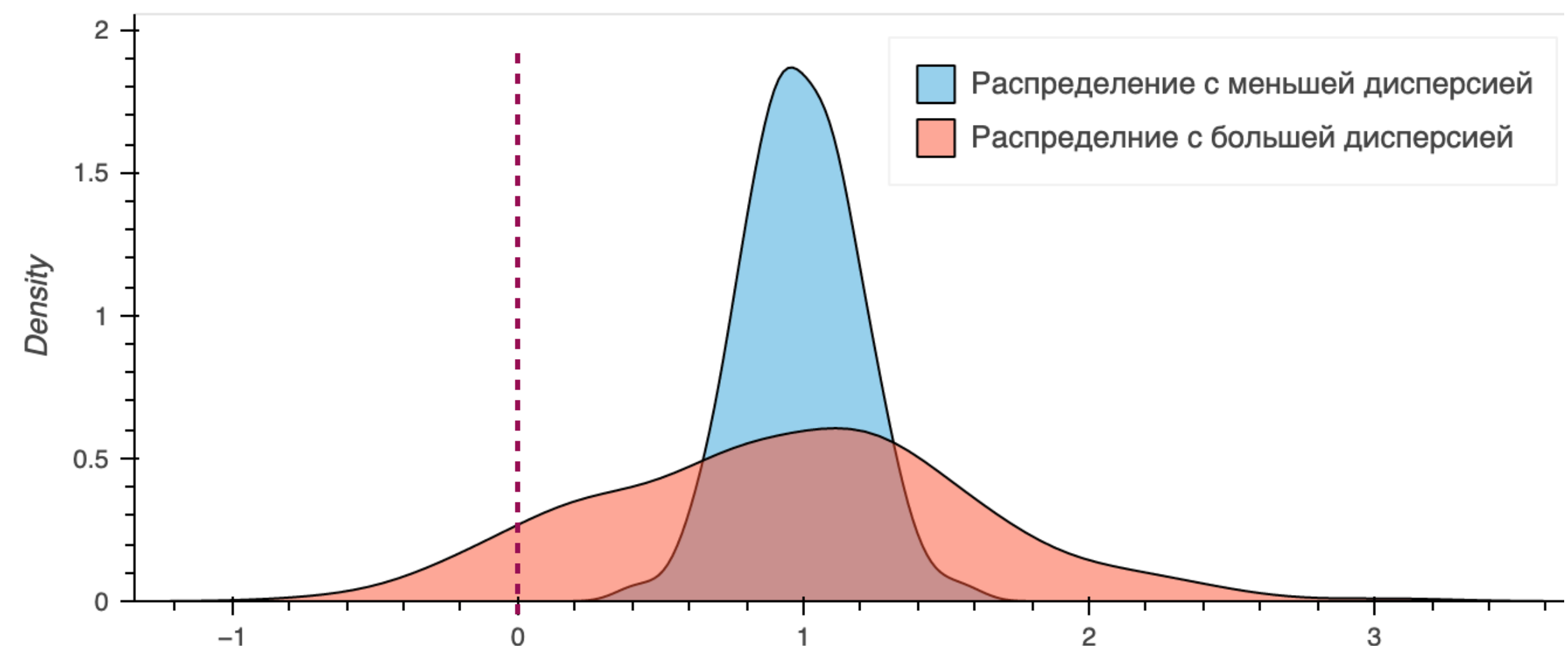


При чём тут дисперсия?

$$\frac{\text{Разница средних}}{\text{Стандартная ошибка}} = \frac{\bar{\mu}_1 - \bar{\mu}_2}{\sqrt{\frac{S_1^2 \downarrow}{n_1} + \frac{S_2^2 \downarrow}{n_2}}} = \text{t statistic} \uparrow$$

- Чем выше дисперсия в метриках эксперимента, тем больше их функции плотности вероятности будут пересекаться (при одной и той же разнице метрики)
- Соответственно, ниже мощность и выше p-value
- Мы можем сократить дисперсию в знаменателе t-статистики, чтобы повысить чувствительность

Плотность распределения метрики



Способы сокращения дисперсии

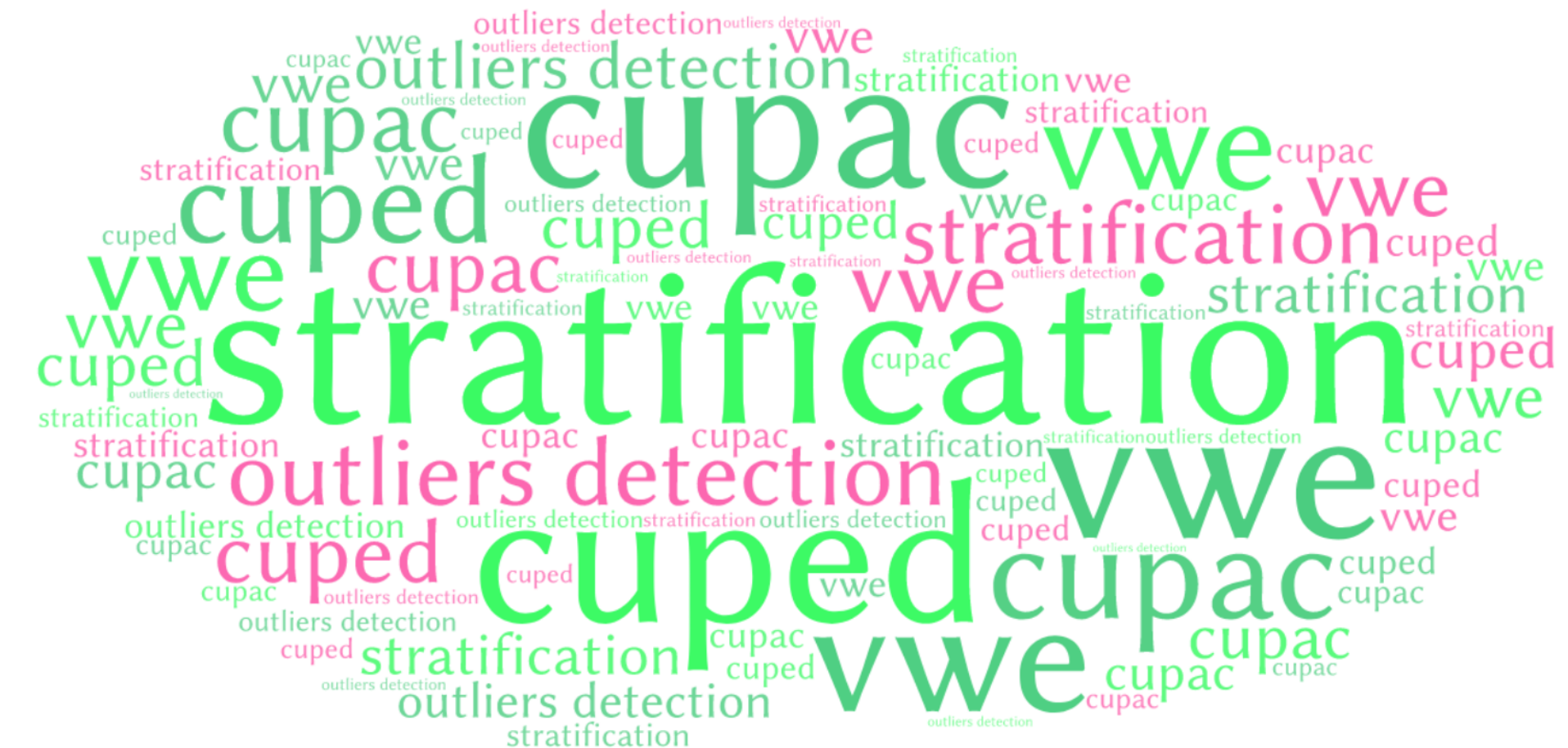
- Это потребует какого-то преобразования исходных данных случайной величины
- Нам нужно подобрать такую функцию:

$$- \begin{cases} y = f(x), \\ \text{var}(y) < \text{var}(x) \end{cases}$$

- При этом не потерять в логике сравнения метрики: не изменить нулевую гипотезу, или предусмотреть обратную трансформацию
- Сохранить направленность (знаки дельт)

Способы сокращения дисперсии

- Есть два обширных класса методов уменьшения дисперсии: пред и пост обработка.
 - Пред – использование ретроспективных данных по метрике
 - Пост – то, что было накоплено в процессе эксперимента.
- Самые популярные CUPED/CUPAC и Pre/Post-stratification



Способы сокращения дисперсии

Очистка выбросов

Суть подхода в том, чтобы тем или иным методом пометить аномальные значения и удалить их из анализа

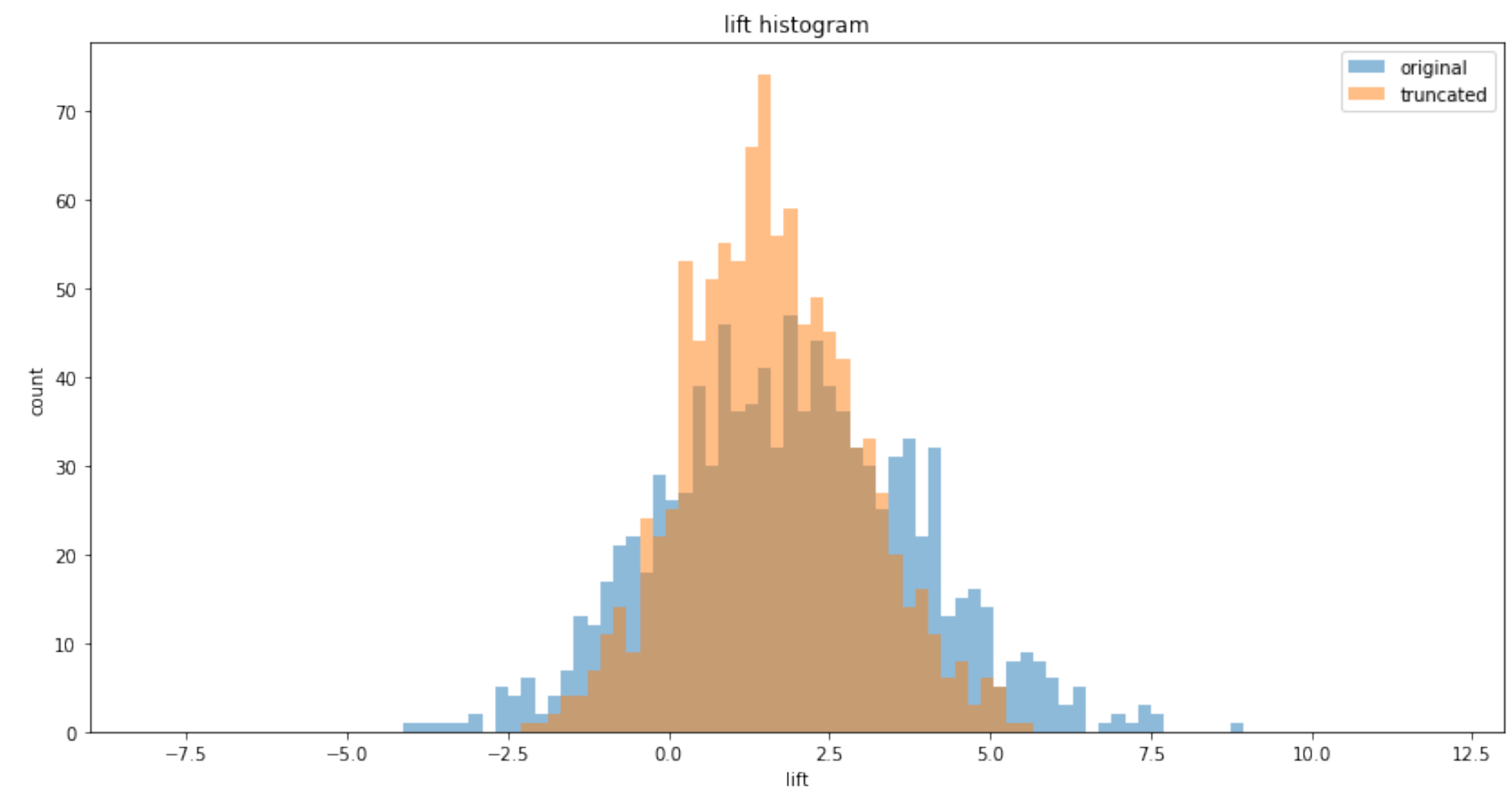
Pros

- Легкая реализация и интерпретация
- Высоко вероятно сократит дисперсию

Cons

- Теряем в мощности
- Можем потерять важные данные

Результат 1000 Монте-Карло симуляций с/без очистки выбросов



Способы сокращения дисперсии

Стратификация

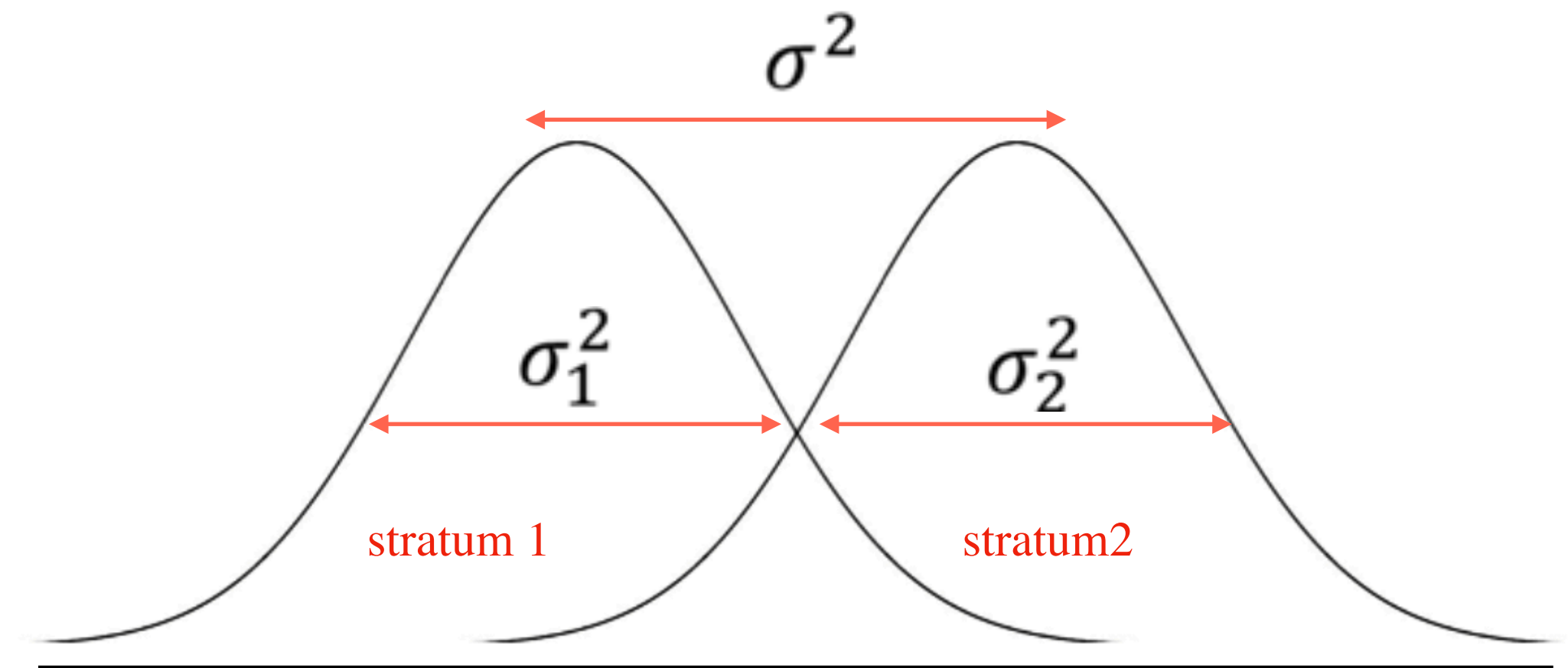
Закljučается в разделении исходной выборки на примерно одинаковые кластеры / страты по какому-то признаку. Например, по типу девайса, версии браузера и пр..

Pros

- В некоторых кейсах показывает высокую эффективность (с малым N)
- Дает гарантию того, что у нас нет перекоса по ключевым группам

Cons

- В канонической реализации требует наличия ретроспективных данных (для pre-stratification)
- При достаточно большой выборке эффект слишком мал (~1%)



$$\text{var}_{strat}(\hat{Y}_{strat}) = \frac{1}{n} \sum_{k=1}^K p_k \sigma_k^2.$$

где k – градация страты, p – вероятность попадания в страту

Способы сокращения дисперсии

CUPED

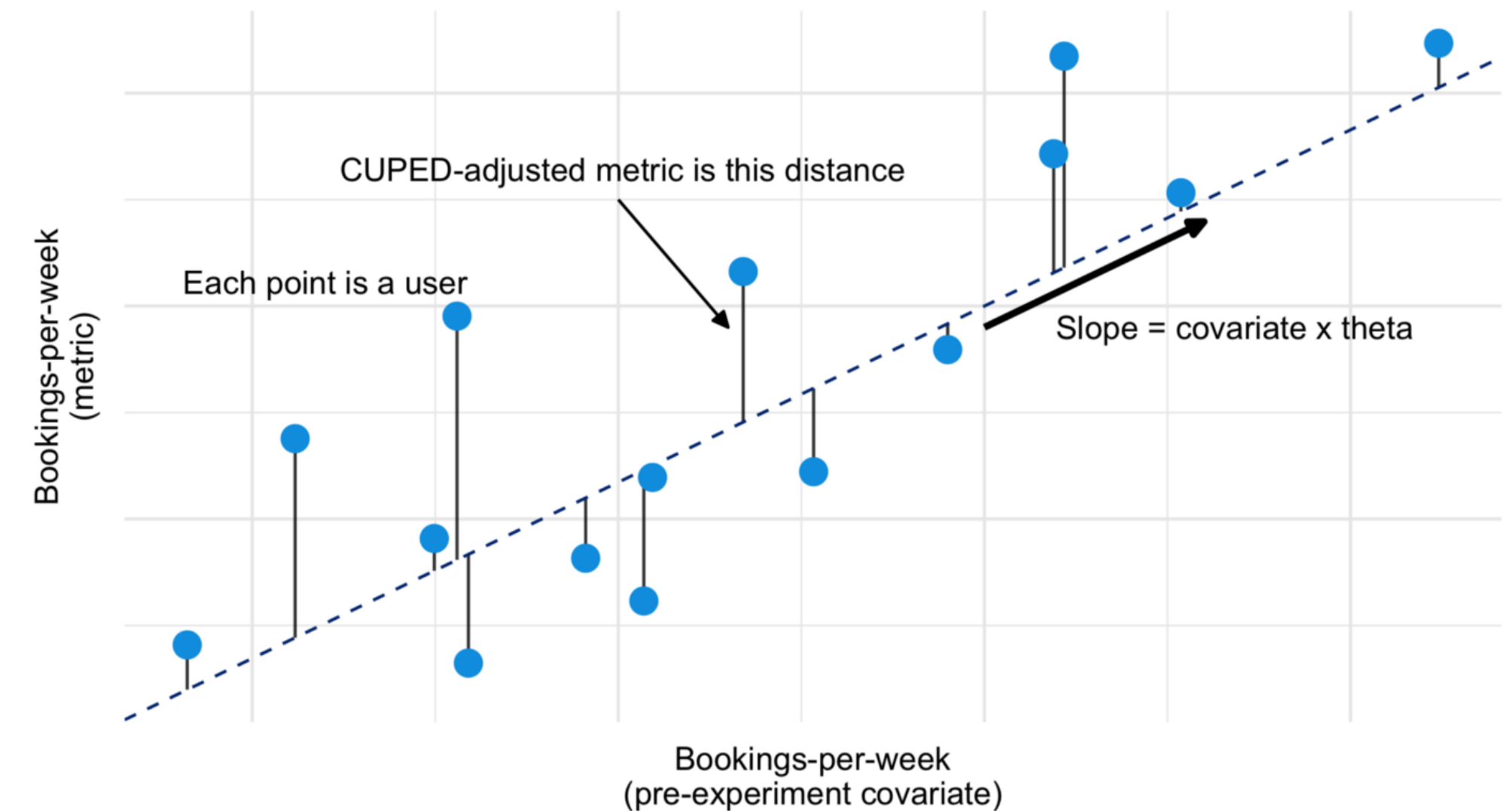
Регрессионные модели на основе исторических данных

Pros

- Существенное сокращение дисперсии
- В большинстве случаев сохраняется среднее
- В случае с хорошей ковариатой(ами) может дать существенное сокращение дисперсии

Cons

- Требует наличия ретроспективных данных
- Для учета сезонности необходимо подбирать ковариаты



$$\text{Var}(\tilde{Y}) = \text{Var}(Y) - 2\theta\text{Cov}(X, Y) + \theta^2\text{Var}(X)$$

Variance-Weighted Estimator

Описание метода

- Стандартный подход предусматривает вариант с использованием ретро-данных.
- Такой подход полезен, когда у нас есть много наблюдений у одного пользователя по X метрике, и с достаточно большим разбросом по всей выборке.
 - Например, количество просмотренных товаров за сеанс.
- Задача состоит в том, чтобы посчитать дисперсию для каждого пользователя и присвоить им обратно пропорциональные веса: у кого дисперсия меньше, у

того и вес больше:
$$Var(\hat{\mu}) = \frac{\sum_i w_i^2 \sigma_i^2}{\left(\sum_i w_i\right)^2}$$

user	variance	value
1	861 931	1278
2	1 670 776	1424
3	1 692 338	1258
4	677 831	1324
5	2 046 633	1549
6	458 513	1190
7	568 898	1157
8	7 676 695	1483
9	576 725	1224
10	933 500	1251

$$1278 * \frac{1}{861931} + 1424 * \frac{1}{1670776} + \dots$$

$$Var(value) = 16963 \rightarrow 162$$

Variance-Weighted Estimator

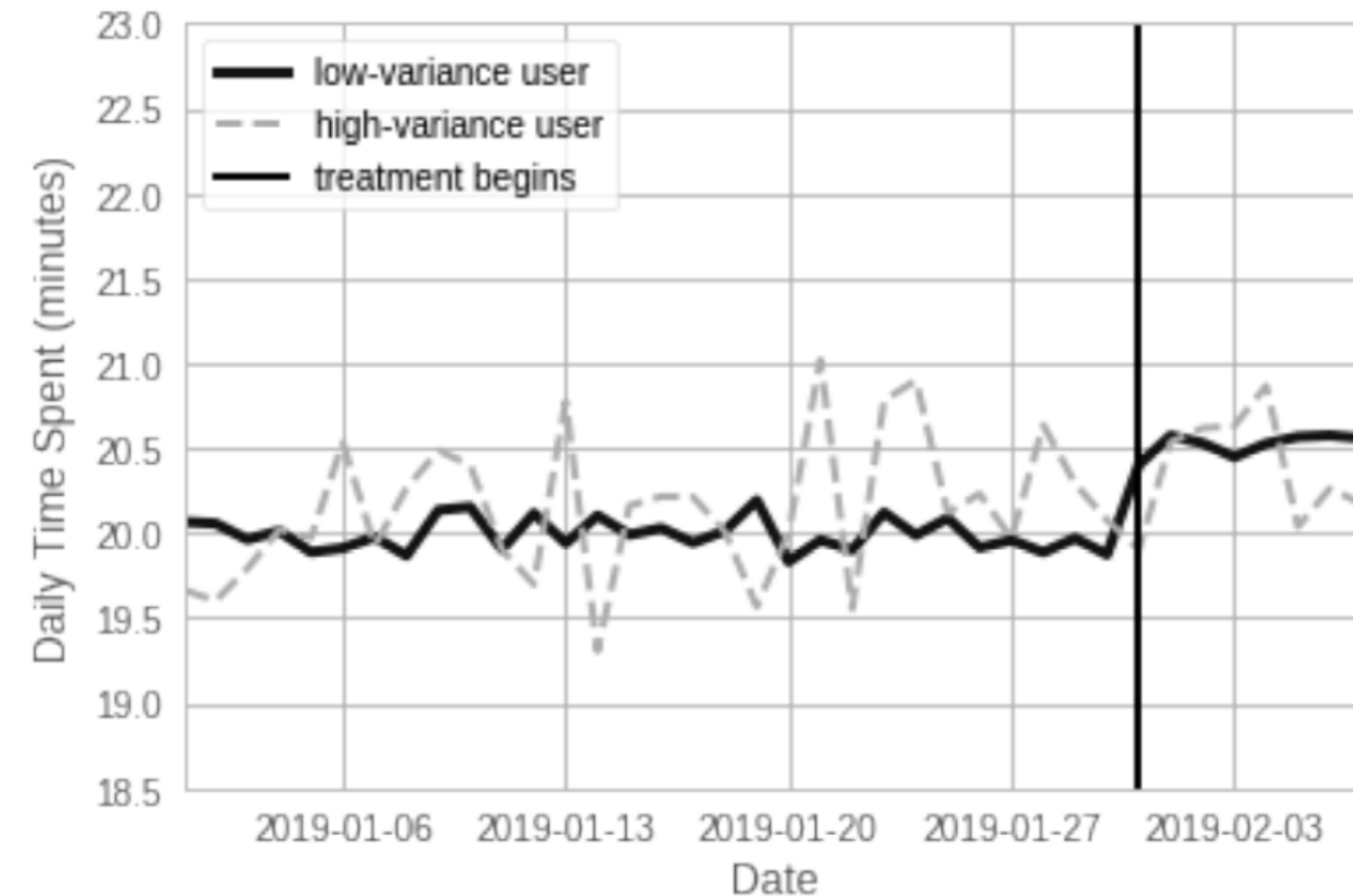
Мы используем его в связке со стратификацией, сохраняя среднее

$$E(\hat{\delta}) = E\left(\frac{\sum_k w_k \delta_k}{\sum_k w_k}\right) = \frac{1}{\sum_k w_k} \sum_k w_k E(\delta_k) = \delta$$

Где δ_k - разница средних двух вариантов для каждой из k страт

А оптимальная дисперсия будет следующей:

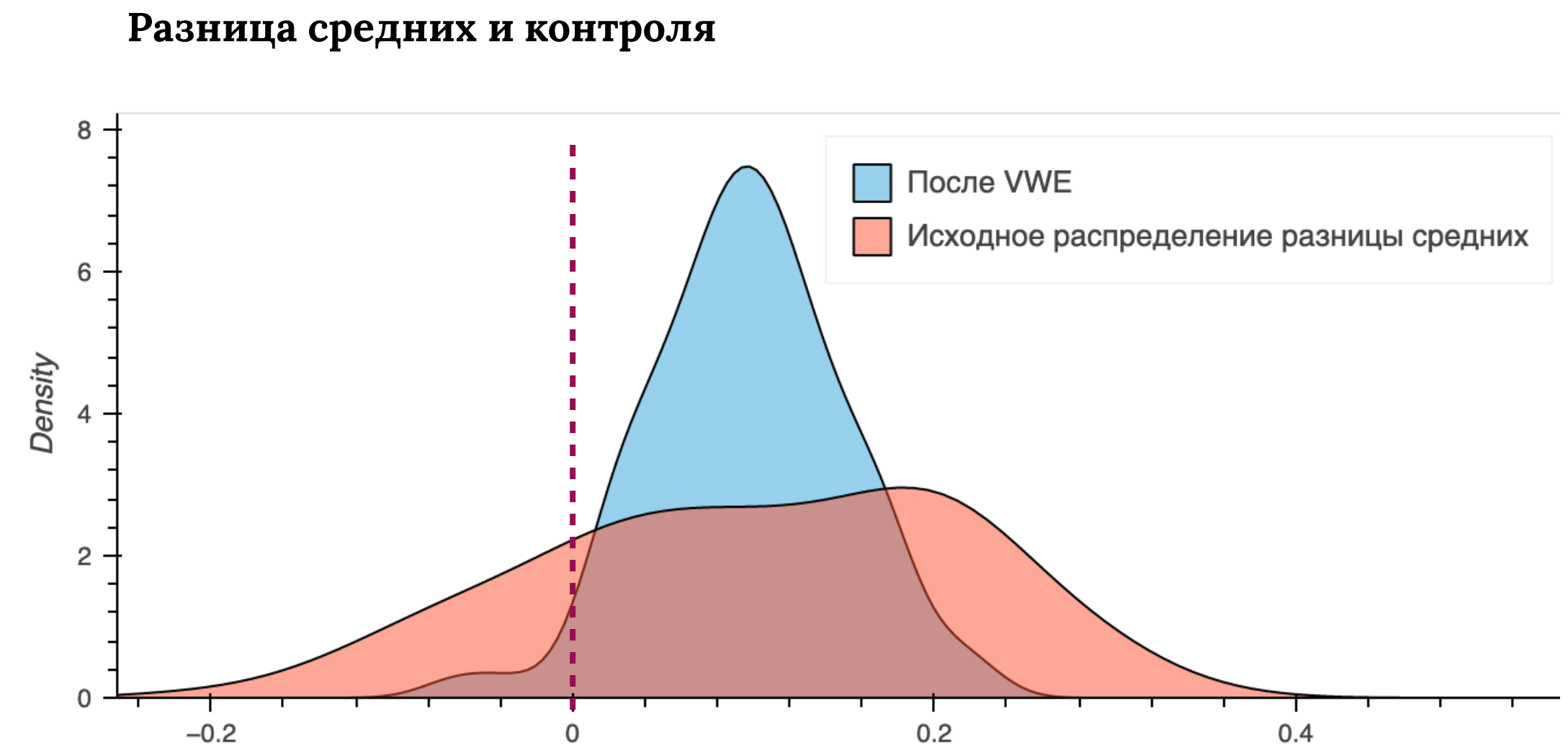
$$Var(\hat{\delta}_{opt}) = \left(\sum_i \sigma_i^{-2}\right)^{-1}$$



Variance-Weighted Estimator

Наблюдения

- Мы проводили эксперимент с новым форматом страницы маркетплейса
- Исследуемая метрика: количество переходов в карточку товара на сеанс (pdp_per_sessions)
- После применения VWE на основе предрасчитанной дисперсии пользователей получилось уменьшить дисперсию на **79%** и показать стат. значимые изменения.



Variance-Weighted Estimator

В среднем же на наших экспериментах использование vwe даёт **5-8%** сокращение дисперсии.

Проблема остаётся в том, что мы можем применять данный подход только на метриках с данными по их дисперсии (т.е. должна быть богатая история по кликам пользователя)

Стратификация

1 %

Стратификация +
VWE

6,4 %

Сокращение
дисперсии

Применение VWE к ratio-метрикам

Что делать с ratio?

- Метрики отношений наиболее чувствительны к проблеме сокращения дисперсии, потому что мы имеем уже заложенную корреляцию числителя и знаменателя.
- Как раз эту проблему решает, например, **дельта-метод**:

$$E\left[\frac{X}{Y}\right] \approx \frac{E[X]}{E[Y]} - \frac{\text{cov}[X, Y]}{E[Y]^2} + \frac{E[X]}{E[Y]^3} \text{var}[Y]$$

$$\text{var}\left[\frac{X}{Y}\right] \approx \frac{\text{var}[X]}{E[Y]^2} - \frac{2E[X]}{E[Y]^3} \text{cov}[X, Y] + \frac{E[X]^2}{E[Y]^4} \text{var}[Y]$$

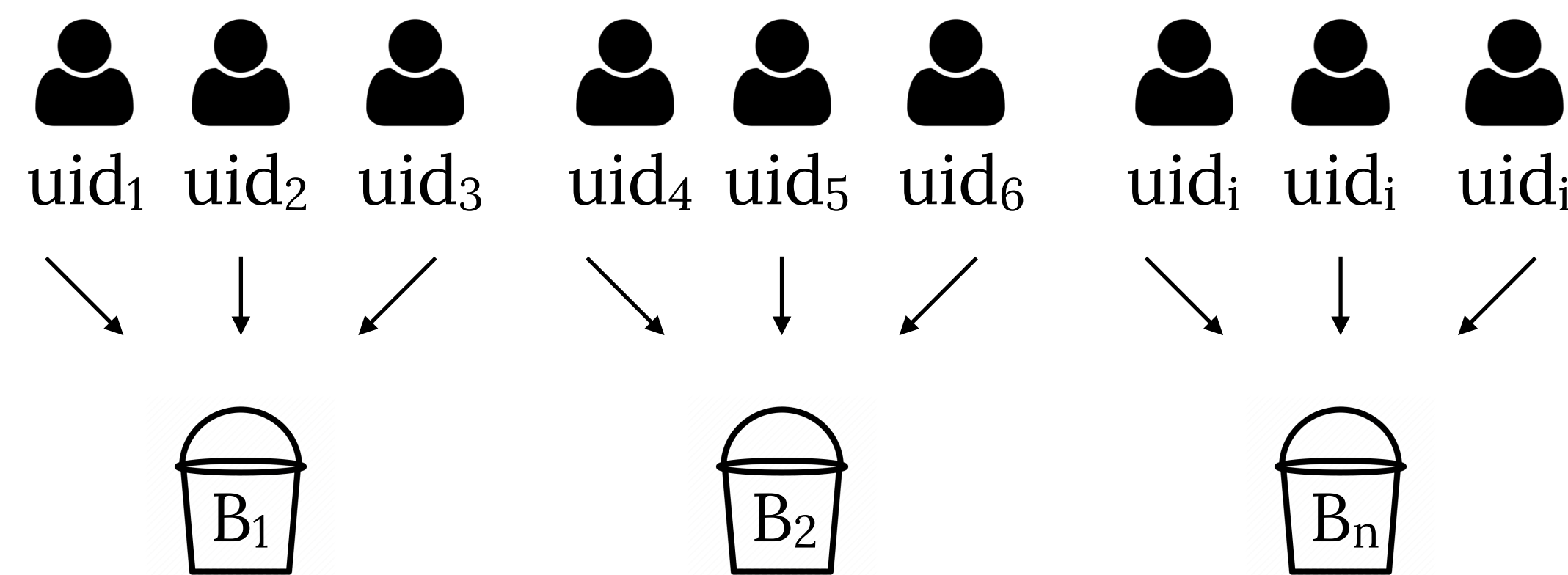
user	numerator	denominator
1	131 683	103
2	138 159	97
3	142 181	113
4	121 865	92
5	142 560	92
6	122 654	103
7	138 891	120
8	155 786	105
9	112 620	92
10	152 623	122

«сократили» дисперсию на 26%

$Var(value) = 16963 \rightarrow 12540$

При чём тут бакетизация

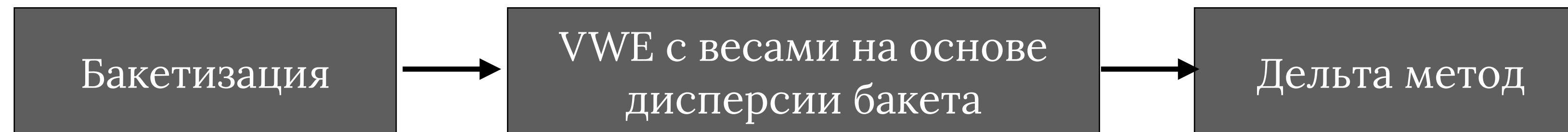
- После бакетизации для каждой метрики имеем два обязательных столбца: числитель и знаменатель метрики
- И плюс можно сохранить для каждого бакета его дисперсию по метрике



bucket_n	numerator	denominator	bucket_variance
1	148 618	113	1 127 430
2	147 381	118	915 738
3	166 340	120	1 279 555
4	150 589	113	1 206 999
5	159 382	122	1 398 143
6	157 678	116	2 001 283
7	147 707	112	2 333 948
8	154 680	114	825 101
9	154 780	113	1 710 700
10	168 127	114	1 542 361
11	177 280	109	2 117 537
12	148 197	115	822 935

Подход с VWE на бакетах

- Сохранённую дисперсию в бакетах можно использовать вместо весов в нашем методе.
- При этом тут уже берутся данные пользователей за эксперимент, а не исторические.
- Веса берём аналогично: обратно пропорциональные.
- По факту после бакетизации все метрики становятся метриками отношений



Пример реализации

В среднем с таким подходом на наших экспериментах нам удаётся сократить дисперсию на **7.1%** (при выборке 10 экспериментов)

Сокращение дисперсии

Стратификация	1 %
Стратификация + vwe	6,4 %
Бакетизация + стратификация + vwe	7,1 %

Особенности подхода

- **Комбинация VWE и других подходов.** Можно (и нужно) использовать в связке с другими подходами, в первую очередь со стратификацией и дельта-методом. Можно даже применять поверх того же VWE на предрасчитанной дисперсии
- **Направленность.** Алгоритм даёт смещение среднего и это необходимо учитывать (проверяйте направленность)

Заключение

- Мы рассмотрели как дисперсия влияет на время проведения эксперимента и для чего её сокращать
- Рассмотрели распространённые способы сокращения дисперсии
- Разобрались с VWE канонически и на бакектизации
- Увидели примеры реализации на кейсах

Спасибо

По любым вопросам
info@expf.ru

