ТПОЭ - 23/24

Лекция 6

Нерсес Багиян

В предыдущих сериях...

“Статистическая” классификация

Средняя

- User Average Metrics (ARPU/ ARPPU/etc)

Ratio

- User-level Conversion Metrics (Retention / etc)
- Page-level Conversion Metrics (Global CTR / etc)

Квантиль

- Ну тут просто квантиль (.99 latency / перцентиль чека)

Абсолюты

- Метрики (GMV / Выручка / Просмотры)

Вот у нас есть дерево метрик

Аксессуары для ванной для повышения среднего чека

1. Предпосылка

   Аналитика показывает, что средний чек покупки увеличивается на 20%, когда клиенты добавляют в корзину товары из категории "Аксессуары для ванны". Интервью с пользователями выявили, что многие не замечают эту категорию при обычном посещении магазина.

2. Возможность

   Ввести функцию персонализированных рекомендаций аксессуаров для ванны на странице оформления заказа мыла.

3. Кого коснется

   Покупатели, оформляющие заказ мыла и заинтересованные в улучшении своего опыта ванны.

4. Мотивация

   Предложение аксессуаров для ванны на этапе оформления заказа мыла удобно для пользователей, так как позволяет дополнить их покупку, не отвлекаясь на поиск этих товаров отдельно.

5. Эффект, который мы ожидаем

   Увеличение среднего чека на 15% за счет добавления аксессуаров для ванны к заказу мыла, что приведет к росту ARPU.

Растим ARPU с помощью рекомендаций аксессуаров

1. Формулировка гипотезы с сформулированным ожидаемым размером эффекта

   Предложение мыла с подборкой популярных аксессуаров на основе анализа предпочтений покупателей и данных о самых продаваемых товарах увеличит средний доход на пользователя (ARPU) на 20%.

2. Описание аудитории

   Покупатели онлайн-магазина мыла, включая как новых, так и возвращающихся пользователей.

3. Описание вариантов с размером каждой группы

   Контрольная группа (A): Покупателям предлагается стандартный ассортимент без акцентов на комплекты.
   Экспериментальная группа (B): Покупателям активно предлагаются комплекты мыла с популярными ароматами на главной странице и в разделе рекомендаций.
   Размер каждой группы составляет 50% от общего числа посетителей в период эксперимента.

4. Ожидаемые исходы и метрики

   Основная метрика: Увеличение ARPU в экспериментальной группе по сравнению с контрольной.
   Второстепенные метрики: Увеличение среднего размера заказа, конверсии в покупку комплектов.

5. Продолжительность

   Эксперимент продлится 4 недели, чтобы собрать достаточно данных для статистически значимых результатов, учитывая недельные колебания трафика и поведения покупателей.

6. Результаты

   TBD

Пример дизайн дока: https://docs.google.com/document/d/1y1uawhMsQ3tyZ8dl8W9jDmOWMICPnZhJnsnayt9UX5c/edit?usp=sharing

В чем может быть проблема тестирования ARPU?

Вопрос аудитории

ARPU состоит из разных частей

ARPU

ARPPU X Conv

Avg. order amount X Purchase frequency

C1 X CN

... ...

Как лучше подобрать метрики, чтобы избавиться от шума?

Вопрос аудитории

Переходим на более низкоуровневые метрики в дизайне

1. Формулировка гипотезы с сформулированным ожидаемым размером эффекта

   Предложение комплектов мыла с подборкой популярных ароматов на основе анализа предпочтений покупателей и данных о самых продаваемых товарах увеличит средний доход на пользователя (ARPU) на 20%.

2. Описание аудитории

   Покупатели онлайн-магазина мыла, включая как новых, так и возвращающихся пользователей.

3. Описание вариантов с размером каждой группы

   Контрольная группа (A): Покупателям предлагается стандартный ассортимент без акцентов на комплекты.
   Экспериментальная группа (B): Покупателям активно предлагаются комплекты мыла с популярными ароматами на главной странице и в разделе рекомендаций.
   Размер каждой группы составляет 50% от общего числа посетителей в период эксперимента.

4. Ожидаемые исходы и метрики

   Основная метрика: Увеличение **среднего чека** в экспериментальной группе по сравнению с контрольной.
   Контрметрика: Не падение **конверсии в покупку**

5. Продолжительность

   Эксперимент продлится 4 недели, чтобы собрать достаточно данных для статистически значимых результатов, учитывая недельные колебания трафика и поведения покупателей.

6. Результаты

   TBD

Пример дизайн дока: https://docs.google.com/document/d/1y1uawhMsQ3tyZ8dl8W9jDmOWMICPnZhJnsnayt9UX5c/edit?usp=sharing

Чем будем тестировать средний чек? Как размер выборки считать?

Вопрос аудитории

Почему t-test не работает?

Вопрос аудитории

Зависимые данные ломают t-test

При наличии зависимых данных оценка дисперсии в тесте Стьюдента ломается. Это приводит к повышению вероятности ошибки первого рода

Пример:

Мы в нашем магазине мыла проводим эксперимент, чтобы понять увеличивают ли наши рекомендации средней чек. У нас есть разные юзеры, некоторые из них могут купить несколько раз за эксперимент. **Данные в выборке в таком случае являются зависимыми**

Rookie mistake подход

Усредним средние чеки по пользователям и потом проведем тест

Группа А

Пользователь 1

2 покупки по 1000 рублей

Пользователь 2

1 покупка по 400 рублей

Группа Б

Пользователь 1

2 покупки по 400 рублей

Пользователь 2

1 покупка по 1300 рублей

Как изменится средний чек?

Вопрос аудитории

Группа А

Пользователь 1

2 покупки по 1000 рублей

Пользователь 2

1 покупка по 400 рублей

Группа Б

Пользователь 1

2 покупки по 400 рублей

Пользователь 2

1 покупка по 1300 рублей

Как изменится средний чек?

Вопрос аудитории

Группа А

Пользователь 1
2 покупки по 1000 рублей

Пользователь 2
1 покупка по 400 рублей

Средний чек: 800 рублей

Группа Б

Пользователь 1
2 покупки по 400 рублей

Пользователь 2
1 покупка по 1300 рублей

Средний чек: 700 рублей

А среднее средних пользователей?

Вопрос аудитории

Группа А

Пользователь 1
2 покупки по 1000 рублей

Пользователь 2
1 покупка по 400 рублей

Среднее среднего: 700 рублей

Группа Б

Пользователь 1
2 покупки по 400 рублей

Пользователь 2
1 покупка по 1300 рублей

Среднее среднего: 700 рублей

Среднее средних это не путь

Сонаправленность не соблюдается

При использовании среднего средних мы не обеспечиваем сонаправленность изменений. По сути, это означает, что в любом тесте мы можем сказать, что изменение есть, хотя его нет или наоборот

Reminder: свойства прокси метрик

Прокси метрика - косвенная метрика целевой метрики, с которой она сильно коррелирует и с которой есть причинно-следственная связь.

Корреляция

Высоко скоррелирована с основной метрикой

Казуальная связь

Имеет причинно-следственную связь с целевой метрикой

Интерпретируемая

Метрику может объяснить каждый сотрудник

Можно заменить на “Сонаправленность”

Чувствительная

Насколько долго нужно ждать, чтобы увидеть изменения в метрике

Достоверная

Можно ли получить точное подтверждение из данных?

Как нам оценить дисперсию метрики отношения в таком случае?

Вопрос аудитории

T-test для ratio метрик

Определяем ratio метрику

Введем обозначения для ratio метрик:

$$R_C = \frac{X_1^C + \ldots + X_N^C}{Y_1^C + \ldots + Y_N^C}$$

$$R_T = \frac{X_1^T + \ldots + X_N^T}{Y_1^C + \ldots + Y_N^T}$$

T-test для ratio metric

Определяем ratio метрику

В более общем виде:

$$R_C = \frac{X_1^C + \ldots + X_N^C}{Y_1^C + \ldots + Y_N^C} = \frac{\sum_{u \in C} X(u)}{\sum_{u \in C} Y(u)}$$

$$R_T = \frac{X_1^T + \ldots + X_N^T}{Y_1^C + \ldots + Y_N^T} = \frac{\sum_{u \in T} X(u)}{\sum_{u \in T} Y(u)}$$

T-test для ratio metric
Формулируем нулевую гипотезу

В классическом случае:

$$H_0 : \theta = \mu_C - \mu_T = 0$$
$$H_1 : \theta = \mu_C - \mu_T \neq 0 = \Delta$$

В нашем случае:

$$H_0 : \theta = R_C - R_T = 0$$
$$H_1 : \theta = R_C - R_T \neq 0$$

T-test для ratio metric

Записываем тест статистику

Запишем тест статистику T нашего теста в случае ratio метрик:

$$T = \frac{R_C - R_T}{\sqrt{\sigma^2(R_C) + \sigma^2(R_T)}}$$

Какие проблемы со статистикой у нас есть?

Вопрос аудитории

T-test для ratio metric
Записываем тест статистику

Запишем тест статистика T нашего теста в случае ratio метрик:

$$T = \frac{R_C - R_T}{\sqrt{\sigma^2(R_C) + \sigma^2(R_T)}}$$

Главные вопросы, на которые нужно найти ответ:

1. Как выглядит дисперсия ratio метрики?
2. Как выглядит распределение T?

Расчет дисперсии ratio метрики

Записываем тест статистику

Запишем тест статистика T нашего теста в случае ratio метрик:

$$T = \frac{R_C - R_T}{\sqrt{\sigma^2(R_C) + \sigma^2(R_T)}}$$

Главные вопросы, на которые нужно найти ответ:

1. Как выглядит дисперсия ratio метрики?
2. Как выглядит распределение T?

Расчет дисперсии ratio метрики
Начинаем с математического ожидания

Наша метрика является более сложным случаем вот такого соотношения:

$$R = \frac{X}{Y}$$

В функциональной форме легко записать:

$$f(X, Y) = \frac{X}{Y}$$

Расчет дисперсии ratio метрики

Начинаем с математического ожидания

Одним из способов для нахождения некоторого приближения математического ожидания является работа с аппроксимацией функции f в точке $\theta = (EX, EY) = (\mu_x, \mu_y)$

$$f(X, Y) = f(\theta) + f'_x(\theta)(X - \mu_x) + f'_y(\theta)(Y - \mu_y) + R$$

Посчитаем математическое ожидание штуки выше:

$$E[f(X, Y)] = E[f(\theta) + f'_x(\theta)(X - \mu_x) + f'_y(\theta)(Y - \mu_y) + R]$$

Расчет дисперсии ratio метрики

Начинаем с математического ожидания

Посчитаем математическое ожидание штуки выше:

$$E[f(X,Y)] = E[f(\theta) + f'_x(\theta)(X-\mu_x) + f'_y(\theta)(Y-\mu_y) + R]$$

Эту штуку можно разложить на слагаемые и вынести $f(\theta)$ за скобки:

$$E[f(X,Y)] = E[f(\theta)] + f'_x(\theta)E[(X-\mu_x)] + f'_y(\theta)E[(Y-\mu_y)]$$

Тогда можем записать:

$$E[f(X,Y)] = E[f(\theta)] + 0 + 0 = \frac{\mu_x}{\mu_y} = f(\theta)$$

Расчет дисперсии ratio метрики

Имеем математическое ожидание, переходим к дисперсии

Мы получили математическое ожидание:

$$E[f(X, Y)] = \frac{\mu_x}{\mu_y}$$

Перейдем теперь к дисперсии:

$$V[f(X, Y)] = E\{(f(X, Y) - E[f(X, Y)])^2\}$$

Можем заменить $E[f(X, Y)]$ **на** $f(\theta)$ **(считали на прошлом слайде):**

$$V[f(X, Y)] = E\{(f(X, Y) - f(\theta))^2\}$$

Расчет дисперсии ratio метрики

Имеем математическое ожидание, переходим к дисперсии

Имея такой вид дисперсии:

$$V[f(X,Y)] = E\{(f(X,Y) - f(\theta))^2\}$$

Пришло самое время ее расписать:

$$V[f(X,Y)] = E\{(f(\theta) + f'_x(\theta)(X - \theta_x) + f'_y(\theta)(Y - \theta_y) - f(\theta))^2\}$$

Немного можно упростить:

$$V[f(X,Y)] = E\{(f'_x(\theta)(X - \theta_x) + f'_y(\theta)(Y - \theta_y))^2\}$$

Расчет дисперсии ratio метрики

Имеем математическое ожидание, переходим к дисперсии

Иии, никакого трюка нет, надо это расписывать:

$$V[f(X,Y)] = E\{(f'(\theta)(X-\theta_x) + f'(\theta)(Y-\theta_y))^2\}$$

УПРОЩАЕМ:

$$V[f(X,Y)] = E\left[f_x'(\theta)^2(X-\theta_x)^2 + 2f_x'(\theta)f_y'(\theta)(X-\theta_x)(Y-\theta_y) + f_y'(\theta)^2(Y-\theta_y)^2\right]$$

Посмотрим внимательно:

$$E[(X-\theta_x)^2] = V(X)$$
$$E[(X-\theta_x)(Y-\theta_y)] = Cov(X,Y)$$
$$E[(Y-\theta_y)^2] = V(Y)$$

Расчет дисперсии ratio метрики

Имеем математическое ожидание, переходим к дисперсии

Учитывая этот факт:

$$E[(X-\theta_x)^2] = V(X)$$
$$E[(X-\theta_x)(Y-\theta_y)] = Cov(X,Y)$$
$$E[(Y-\theta_y)^2] = V(Y)$$

Запишем снова дисперсию:

$$V[f(X,Y)] = f_x'^2(\theta)V(X) + 2f_x'(\theta)f_y'(\theta)cov(X,Y) + f_y'^2(\theta)V(Y)$$

Если посчитать все производные, то получим:

$$V(\frac{X}{Y}) \approx \frac{1}{\mu_y^2}V(X) - 2\frac{\mu_x}{\mu_y^3}cov(X,Y) + \frac{\mu_x^2}{\mu_y^4}V(Y)$$

Расчет дисперсии ratio метрики

Давайте обобщим на чуть более общий случай

В более общем случае, возвращаясь к нашей ratio метрике:

$$V(R) = V\left(\frac{X_1 + \ldots + X_N}{Y_1 + \ldots + Y_N}\right) = V\left(\frac{\bar{X}}{\bar{Y}}\right)$$

Возвращаясь к формуле с предыдущего слайда:

$$\frac{1}{\mu_y^2}V(\bar{X}) - 2\frac{\mu_x}{\mu_y^3}cov(\bar{X}, \bar{Y}) + \frac{\mu_x^2}{\mu_y^4}V(\bar{Y})$$

И тут просто упрощается:

$$V(R) = \frac{1}{N\mu_y^2}V(X) - 2\frac{\mu_x}{\mu_y^3}cov(X, Y) + \frac{\mu_x^2}{N\mu_y^4}V(Y)$$

T-test для ratio metric

Записываем тест статистику

Запишем тест статистика T нашего теста в случае ratio метрик:

$$T = \frac{R_C - R_T}{\sqrt{\sigma^2(R_C) + \sigma^2(R_T)}}$$

Главные вопросы, на которые нужно найти ответ:

1. Как выглядит дисперсия ratio метрики?

$$E[R] = \frac{\mu_x}{\mu_y}$$

$$V(R) = \frac{1}{N\mu_y^2}V(X) - 2\frac{\mu_x}{\mu_y^3}cov(X, Y) + \frac{\mu_x^2}{N\mu_y^4}V(Y)$$

2. Как выглядит распределение T?

Что говорит ваша интуиция? Какое распределение у T?

Вопрос аудитории

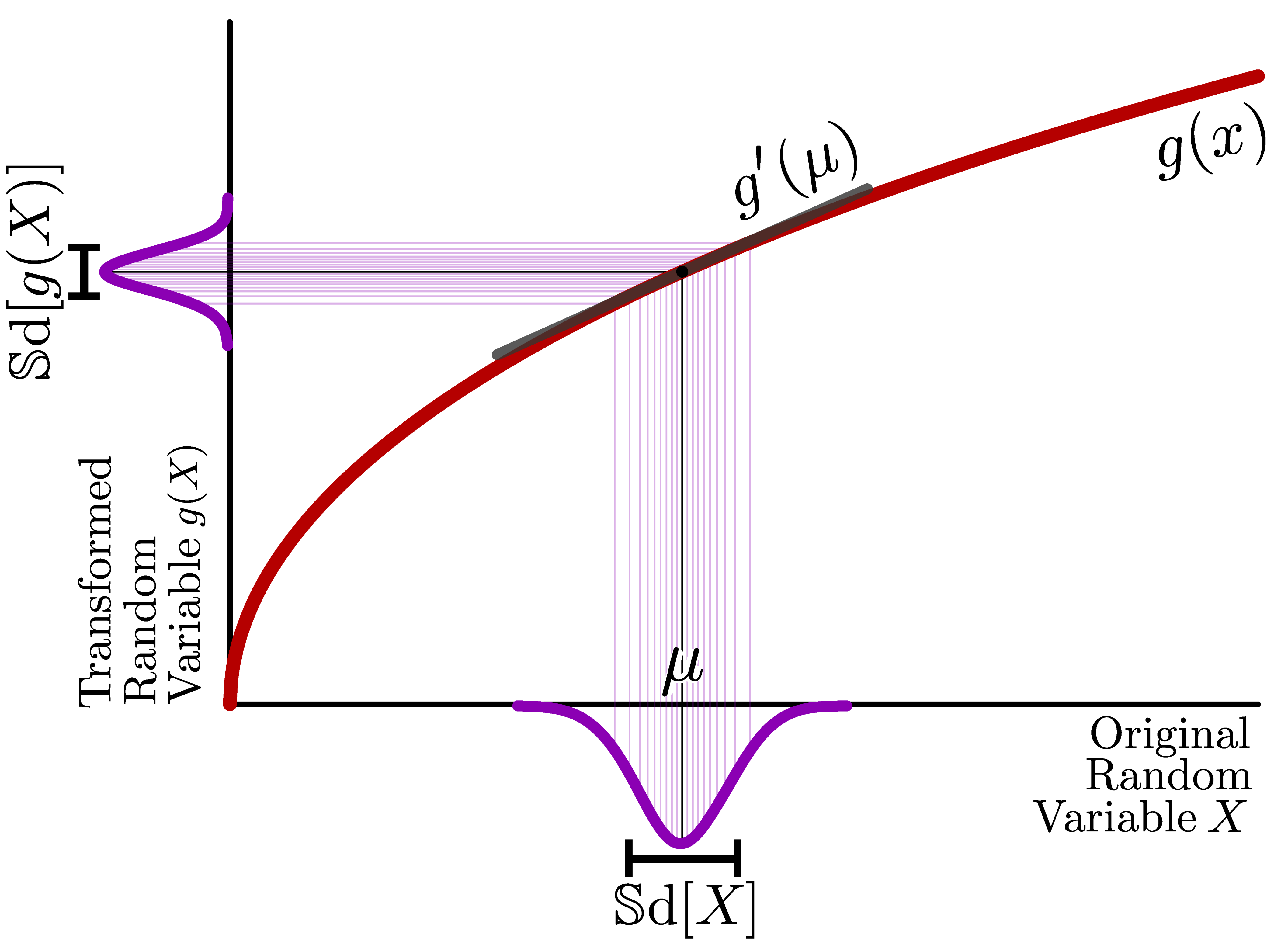
T-test для ratio metric
Вспоминаем дельта метод

Если существует последовательность случайных величин X_n, удовлетворяющая:

$$\sqrt{n}(X_n - \theta) \to^D N(0,\sigma^2)$$

где σ^2 и θ - конечные константы,
а D обозначает сходимость по распределению, то верно:

$$\sqrt{n}(g(X_n) - g(\theta)) \to^D N(0,\sigma^2[g'(\theta)]^2)$$



T-test для ratio metric
Попробуем применить дельта метод

Запишем тест статистика T нашего теста в случае ratio метрик:

$$T = \frac{R_C - R_T}{\sqrt{\sigma^2(R_C) + \sigma^2(R_T)}}$$

Для оценки дисперсии R_C и R_T мы применили многомерный дельта метод. Исходя из этого, можно предположить, что они асимптотически имеют нормальное распределение. Тогда при верности нулевой гипотезы:

$$R_C - R_T \sim N(0, \sigma(R_C)^2 + \sigma(R_T)^2)$$

https://stats.stackexchange.com/questions/291594/estimation-of-population-ratio-using-delta-method/291652#291652

А теперь? Какое распределение у T?

Вопрос аудитории

T-test для ratio metric

Попробуем применить дельта метод

Тест статистика в данном случае имеет следующее распределение при верности нулевой гипотезы:

$$T = \frac{R_C - R_T}{\sqrt{\sigma^2(R_C) + \sigma^2(R_T)}} \sim N(0,1)$$

https://stats.stackexchange.com/questions/291594/estimation-of-population-ratio-using-delta-method/291652#291652

T-test для ratio metric

Записываем тест статистику

Запишем тест статистика T нашего теста в случае ratio метрик:

$$T = \frac{R_C - R_T}{\sqrt{\sigma^2(R_C) + \sigma^2(R_T)}}$$

Главные вопросы, на которые нужно найти ответ:

1. Как выглядит дисперсия ratio метрики?

$$E[R] = \frac{\mu_x}{\mu_y}$$

$$V(R) = \frac{1}{N\mu_y^2}V(X) - 2\frac{\mu_x}{\mu_y^3}cov(X, Y) + \frac{\mu_x^2}{N\mu_y^4}V(Y)$$

2. Как выглядит распределение T?

$$T \sim N(0,1)$$

Смотрим на три типа метрик
Две из них уже свели к нормальному распределению

Средняя

- User Average Metrics (ARPU/ ARPPU/etc)

Ratio

- User-level Conversion Metrics (Retention / etc)
- Page-level Conversion Metrics (Global CTR / etc)

Квантиль

- Ну тут просто квантиль (.99 latency / перцентиль чека)

Абсолюты

- Метрики (GMV / Выручка / Просмотры)

Мы можем пользоваться той же формулой

Потому что свели все к t-test и нормальному распределению

$$n \geq \frac{2(F^{-1}(1 - \frac{\alpha}{2}) - F^{-1}(\beta))^2 s^2}{MDE^2}$$

Итого

1. Узнали, какие проблемы могут быть с ratio метриками
2. Вспомнили дельта метод
3. Посмотрели как свести все к нормальному распределению
4. Чуть не умерли от формул (дальше опять продолжение будет)