Байесовское A/Б тестирование

Максим Кочуров

МГУ им. М.В. Ломоносова

Лекция 3

Содержание

Как это делается: классический подход

“Если ваше p-значение равно 0.05, это значит, что если нулевая гипотеза верна, вы получите такое же или более экстремальное значение тестовой статистики в 5% случаев.”

1. p-значения используются в тысячах статей
2. p-значения очень популярны из-за простоты интерпретации
3. легко вычислить доверительные интервалы

Как это делается: классический подход

"Если ваше p-значение равно 0.05, это значит, что если нулевая гипотеза верна, вы получите такое же или более экстремальное значение тестовой статистики в 5% случаев."

1. p-значения используются в тысячах статей
2. p-значения очень популярны из-за простоты интерпретации
3. легко вычислить доверительные интервалы

Вы уверены?

Действительно ли вы понимаете смысл p-значений?

Понимаете ли вы p-значения?

Что из нижеперечисленного верно?

1. p – вероятность того, что верна нулевая гипотеза
2. $(1 - p)$ – вероятность того, что верна альтернативная гипотеза
3. $P \leq 0.05$ значит, что нулевая гипотеза неверна и дожна быть отвергнута
4. Значение $P > 0.05$ значит, что эффект отсутствует

Понимаете ли вы p-значения?

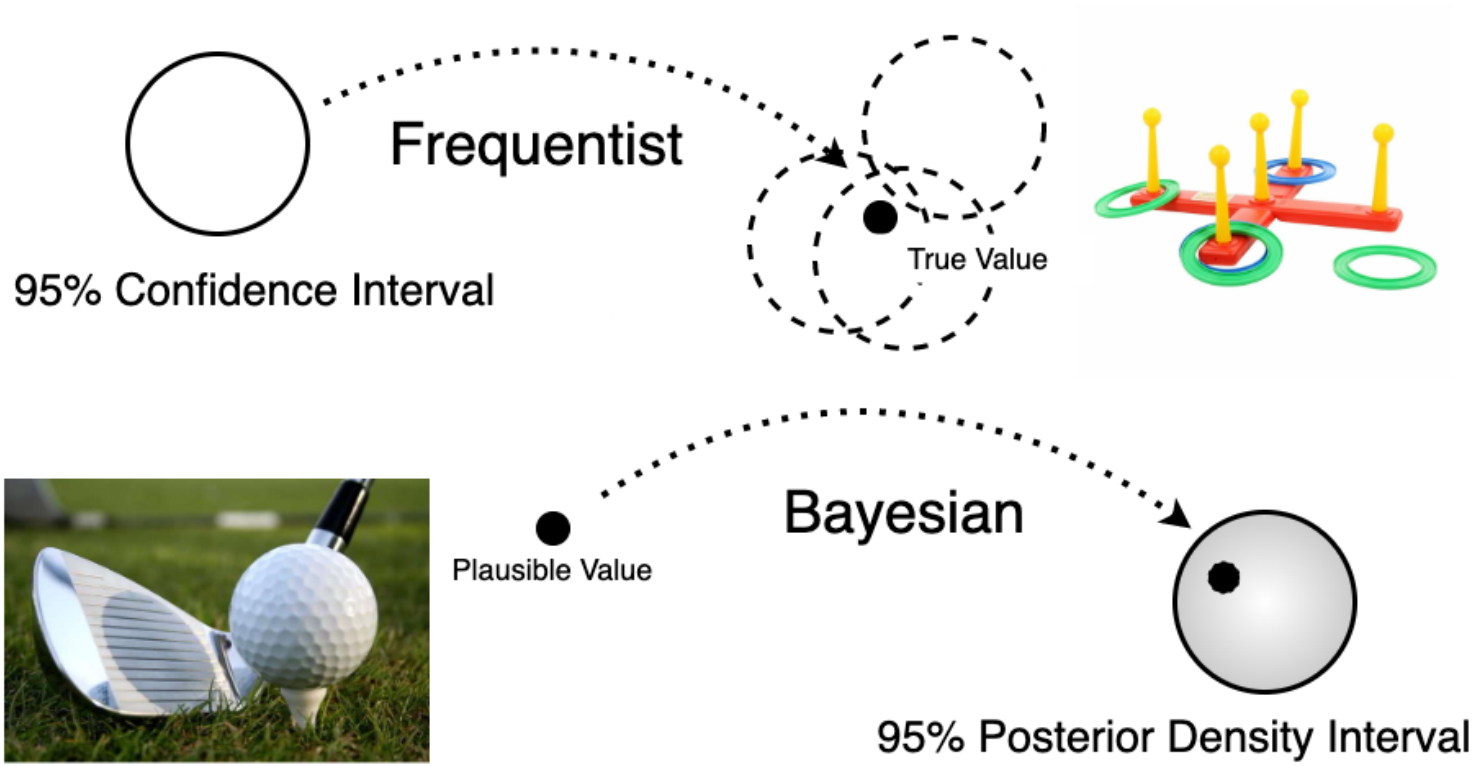
Что из нижеперечисленного верно?

1. p – вероятность того, что верна нулевая гипотеза
2. $(1 - p)$ – вероятность того, что верна альтернативная гипотеза
3. $P \leq 0.05$ значит, что нулевая гипотеза неверна и дожна быть отвергнута
4. Значение $P > 0.05$ значит, что эффект отсутствует

Использовать p-значение плохо?

Я не предлагаю отказаться от p-значений, я предлагаю лучше понять их

Интерпретация p-значений



Тестирование гипотез во фреймворке "H0 против H1"

Вы уже знаете, что такое тестирование гипотез, t-тест, p-значение.

- гипотеза о матожидании в 1 выборке: $t = \frac{Z}{s} = \frac{\bar{X} - \mu}{\widehat{\sigma}/\sqrt{n}}$
- гипотеза о равенстве матожиданий в 2-х выборках

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p\sqrt{\frac{2}{n}}}, \quad s_p = \sqrt{\frac{s^2_{X_1} + s^2_{X_2}}{2}}, \ldots$$

- гипотеза о равенстве дисперсий

$$\ldots, s_p = \sqrt{\frac{(n_1 - 1)\, s^2_{X_1} + (n_2 - 1)\, s^2_{X_2}}{n_1 + n_2 - 2}}$$

Слишком сложно

Чем меньше предположений, тем сложнее вычисления и реализация

Байесовский подход

Осторожно с интерпретацией p-значений!

- Частотнические доверительные интервалы – это не наиболее вероятные значения
- p-значение – не вероятность того, что “нет эффекта”

Байесовский подход – про интерпретацию:

Хорошо

- Проще объяснить
- Проще превратить в действия

Плохо

- Нужно разбираться в предметной области

Байесовские инструменты для визуализации

1. Интервал наибольшей плотности (Highest Density Interval, HDI)
2. Region of Practical Equivalence (RoPE)
3. Коэффициент Байеса (Bayes Factor)
4. Возможность кастомизации

Интервал наибольшей плотности

Наиболее популярный способ интерпретации апостериорного распределения

1. Область наиболее вероятных значений
2. Просто вычислить, интерпретировать и визуализировать

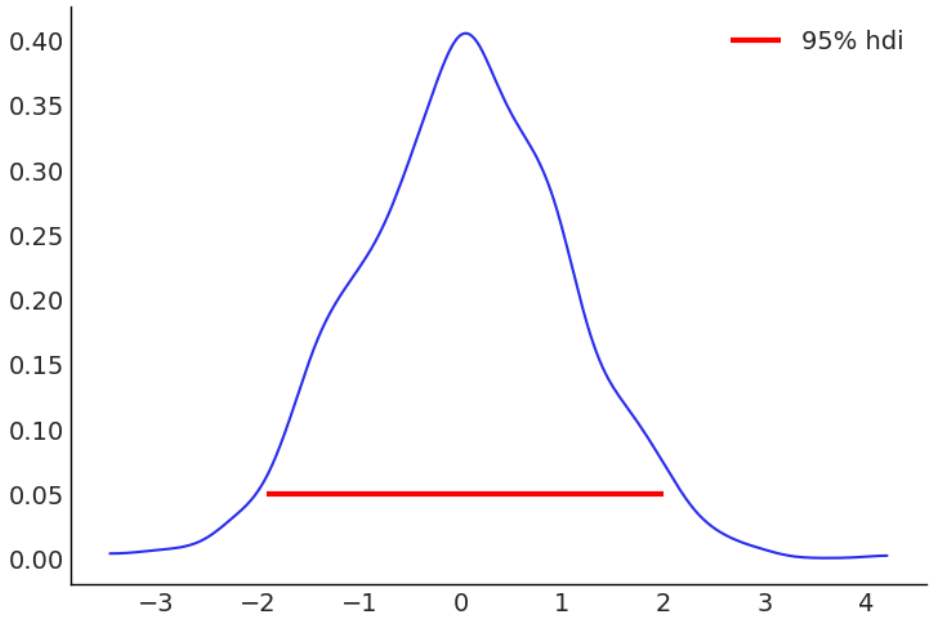


Figure: Интервал наибольшей плотности

Пример

- Размер эффекта – в интервале $[A, B]$ с вероятностью 0.95
- Интервал $[A, B]$ содержит 95% наиболее вероятных размеров эффекта

Region of Practical Equivalence (RoPE)

RoPE – способ оценить “значимость” оценки параметра.
Алгоритм использования:

1. Пусть эффект меньше 0.1 “незначим”
2. Нарисовать “регион незначимости” на одном графике с апостериорными распределениями
3. Определить значимость

Пример

Размер эффекта эксперимента “E” находится вне RoPE, поэтому эффект значим.

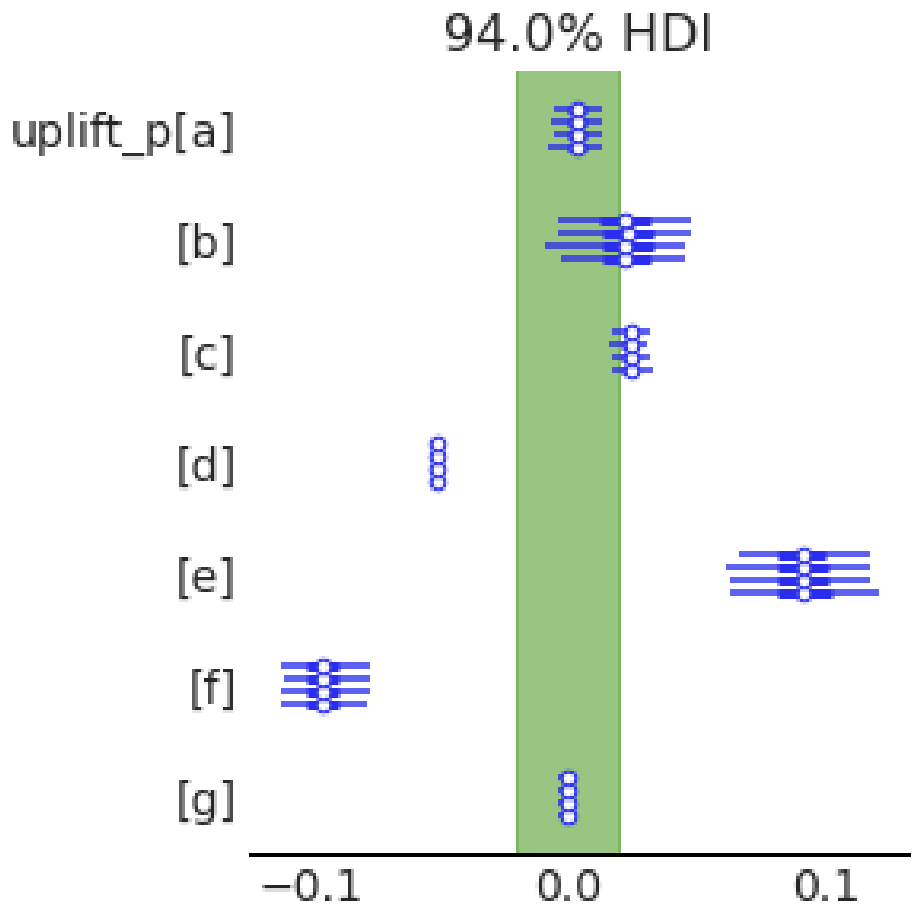


Figure: График RoPE

Коэффициент Байеса

По-моему, эту статистику сложнее всего объянить.

1. Похоже на частотническое p-значение
2. Сложнее интерпретировать и объяснять
3. Проверяет H0 против H1 для x_0

Определение

Коэффициент Байеса – отношение правдоподобия одной гипотезы к правдоподобию второй гипотезы

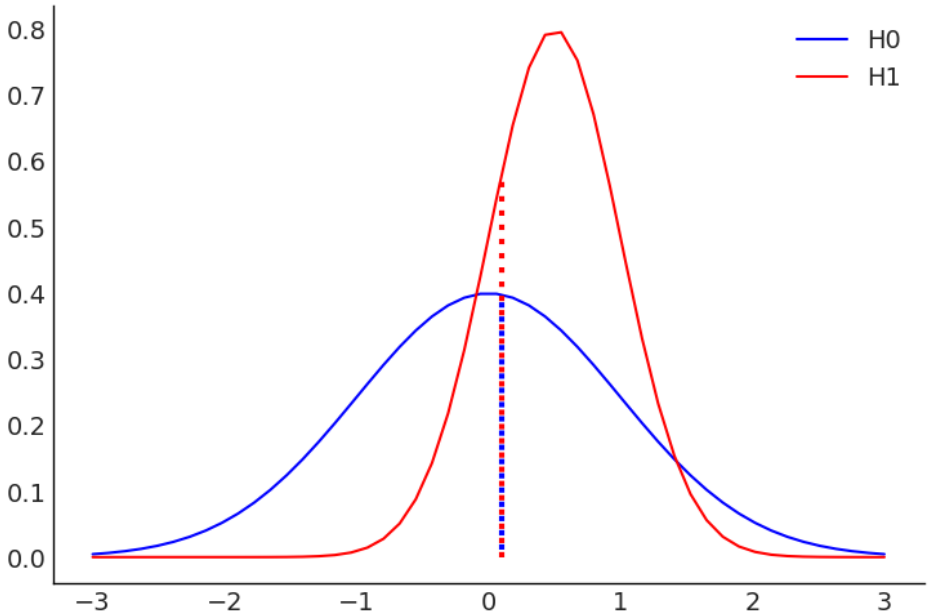


Figure: $\mathrm{BF} = \frac{\mathrm{pdf}_{H1}(x_0)}{\mathrm{pdf}_{H0}(x_0)}$

Другие запросы к апостериорному распределению

Можно пойти дальше:

1. $P(A < 0)$
2. $P(A > B)$
3. $P(\max(A) > \max(B))$
4. $P(A = \arg\max(A, B, C, D))$
5. $P(\text{выручка}(X, \Theta) > \$100)$
6. Квантили - $Q_{0.05}(\text{выручка}(X, \Theta))$

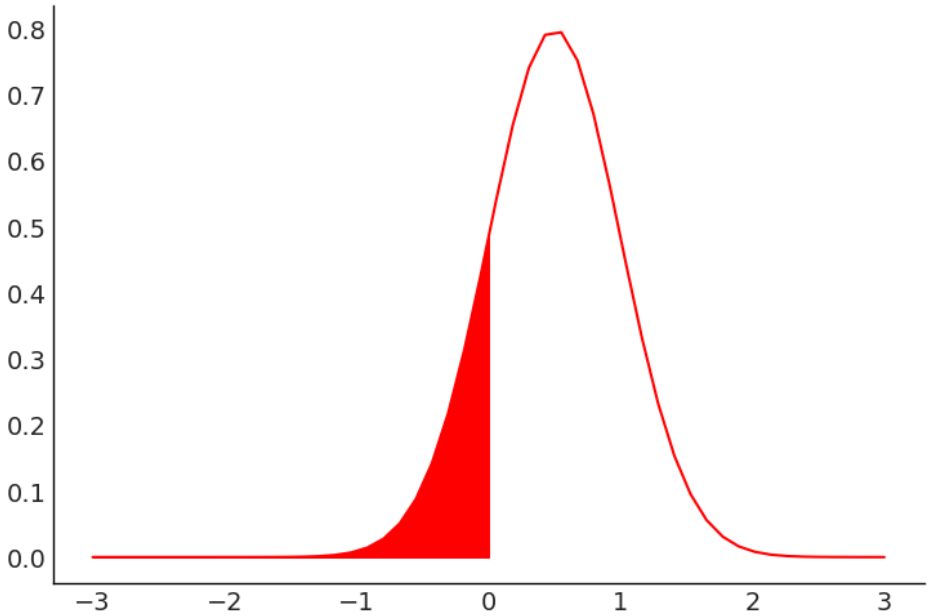


Figure: $P(A < 0)$

Выводы

Байесовский нож для проверки гипотез

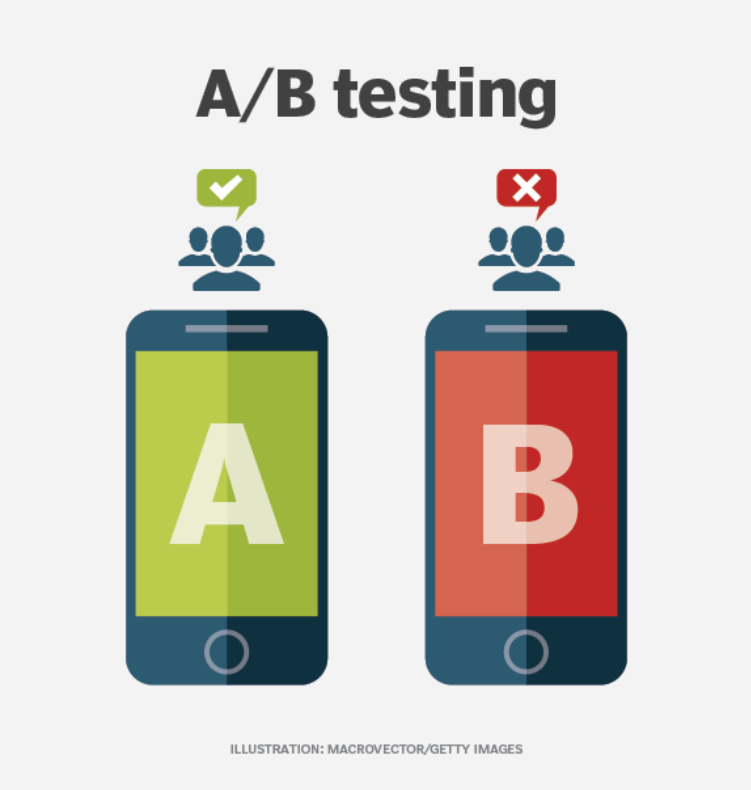
1. Много способов интерпретации результатов
2. Ответ не в формате “да/нет”
3. Отражает неопределённость
4. Гибкость анализа
5. Просто реализовать
6. Просто интерпретировать

Figure: байесовская проверка гипотез

Типы задач

Байесовское А/Б тестирование широко применимо

1. Дискретные наблюдения (просмотры и клики)
2. Непрерывные наблюдения (время чтения, потраченные деньги)
3. С предикторами контекста (Context Predictors; CUPED[1])
4. С иерархией (регионы)



Типы задач

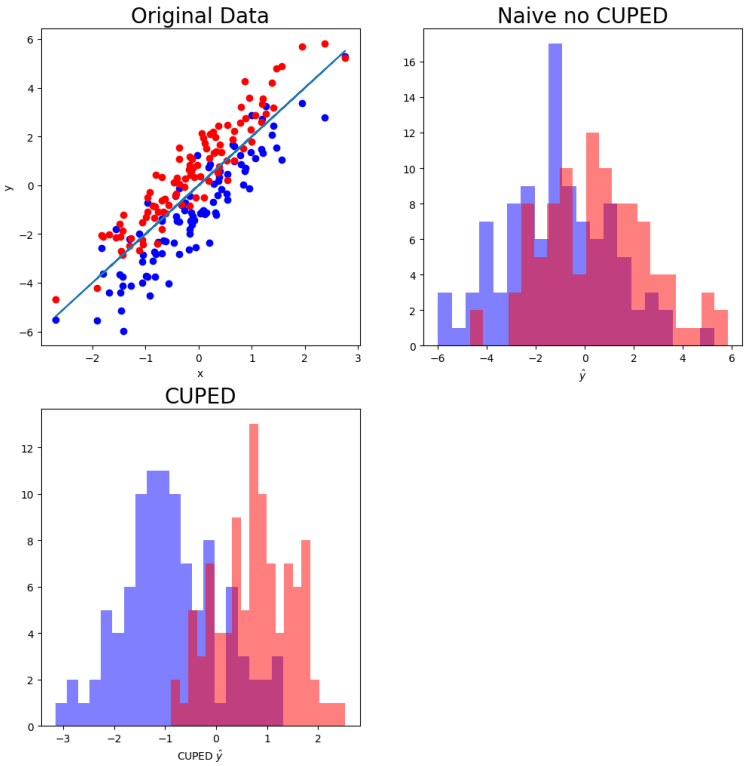
Байесовское А/Б тестирование широко применимо

1. Дискретные наблюдения (просмотры и клики)
2. Непрерывные наблюдения (время чтения, потраченные деньги)
3. С предикторами контекста (Context Predictors; CUPED[1])
4. С иерархией (регионы)

Типы задач

Байесовское А/Б тестирование широко применимо

1. Дискретные наблюдения (просмотры и клики)
2. Непрерывные наблюдения (время чтения, потраченные деньги)
3. С предикторами контекста (Context Predictors; CUPED[1])
4. С иерархией (регионы)



Типы задач

Байесовское А/Б тестирование широко применимо

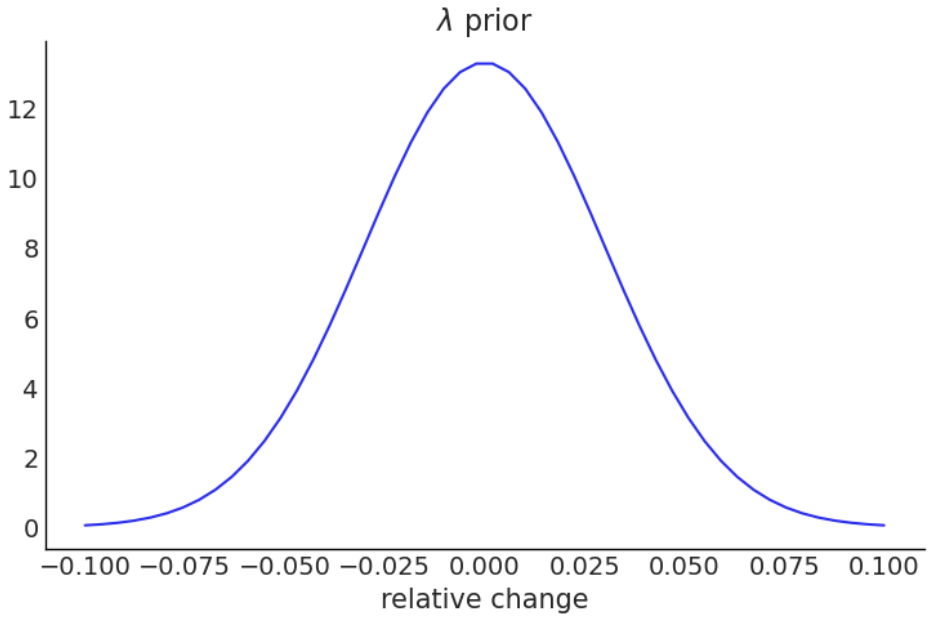
1. Дискретные наблюдения (просмотры и клики)
2. Непрерывные наблюдения (время чтения, потраченные деньги)
3. С предикторами контекста (Context Predictors; CUPED[1])
4. С иерархией (регионы)

Подходы к априорным распределениям

Прирост (uplift) λ

Изменение относительно базового уровня

Когда вы начинаете эксперимент, не знаете ли вы что-то о множестве возможных результатов воздействия?



Подготовка

Вы готовитесь к проведению эксперимента над группой B с контрольной группой A. Вы можете быть заинтересованы в увеличении какой-либо статистики (например, среднего чека).

Подготовка

Вы готовитесь к проведению эксперимента над группой B с контрольной группой A. Вы можете быть заинтересованы в увеличении какой-либо статистики (например, среднего чека).

- Ожидаете ли вы роста на 1000%? Точно нет

Относительное или абсолютное изменение?

Должно быть понятно, относительное это изменение или абсолютное!

Подготовка

Вы готовитесь к проведению эксперимента над группой B с контрольной группой A. Вы можете быть заинтересованы в увеличении какой-либо статистики (например, среднего чека).

- Ожидаете ли вы роста на 1000%? Точно нет
- Ожидаете ли вы роста на 100%? Точно нет

Относительное или абсолютное изменение?

Должно быть понятно, относительное это изменение или абсолютное!

Подготовка

Вы готовитесь к проведению эксперимента над группой B с контрольной группой A. Вы можете быть заинтересованы в увеличении какой-либо статистики (например, среднего чека).

- Ожидаете ли вы роста на 1000%? Точно нет
- Ожидаете ли вы роста на 100%? Точно нет
- Ожидаете ли вы роста на 10%? Скорее нет

Относительное или абсолютное изменение?

Должно быть понятно, относительное это изменение или абсолютное!

Подготовка

Вы готовитесь к проведению эксперимента над группой B с контрольной группой A. Вы можете быть заинтересованы в увеличении какой-либо статистики (например, среднего чека).

- Ожидаете ли вы роста на 1000%? Точно нет
- Ожидаете ли вы роста на 100%? Точно нет
- Ожидаете ли вы роста на 10%? Скорее нет
- Ожидаете ли вы роста на 3%? Возможно

Относительное или абсолютное изменение?

Должно быть понятно, относительное это изменение или абсолютное!

Подготовка

Вы готовитесь к проведению эксперимента над группой B с контрольной группой A. Вы можете быть заинтересованы в увеличении какой-либо статистики (например, среднего чека).

- Ожидаете ли вы роста на 1000%? Точно нет
- Ожидаете ли вы роста на 100%? Точно нет
- Ожидаете ли вы роста на 10%? Скорее нет
- Ожидаете ли вы роста на 3%? Возможно
- Ожидаете ли вы падения на 3%? Возможно

Относительное или абсолютное изменение?

Должно быть понятно, относительное это изменение или абсолютное!

Подготовка

Вы готовитесь к проведению эксперимента над группой B с контрольной группой A. Вы можете быть заинтересованы в увеличении какой-либо статистики (например, среднего чека).

- Ожидаете ли вы роста на 1000%? Точно нет
- Ожидаете ли вы роста на 100%? Точно нет
- Ожидаете ли вы роста на 10%? Скорее нет
- Ожидаете ли вы роста на 3%? Возможно
- Ожидаете ли вы падения на 3%? Возможно
- Ожидаете ли вы падения на X%? Ваш ответ

Относительное или абсолютное изменение?

Должно быть понятно, относительное это изменение или абсолютное!

Примерный алгоритм работы

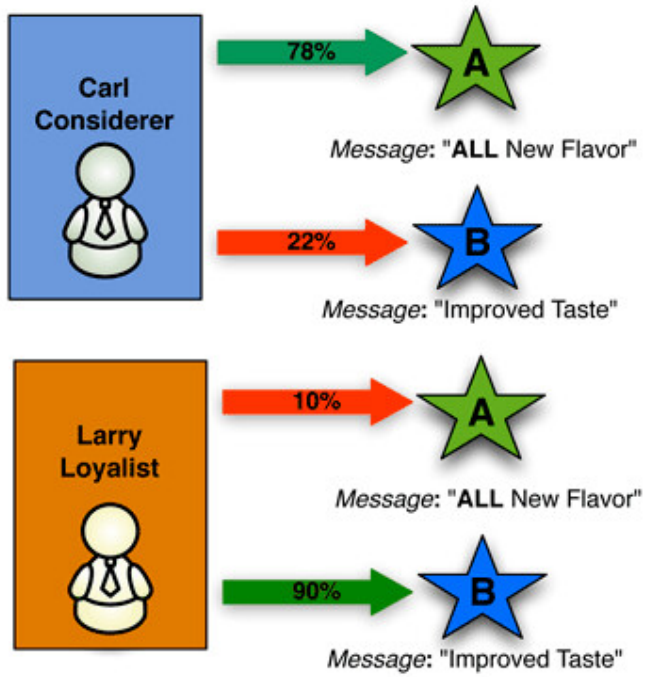
- Как составить эксперимент?
- Как спланировать его исполнение?
- Как интерпретировать результаты?

Пример с биномиальной моделью

- Бинарный ответ: да/нет
- Наблюдения имеют распределение Бернулли

$$x_i^A \sim \text{Bernoulli}(p_A)$$
$$x_i^B \sim \text{Bernoulli}(p_B)$$

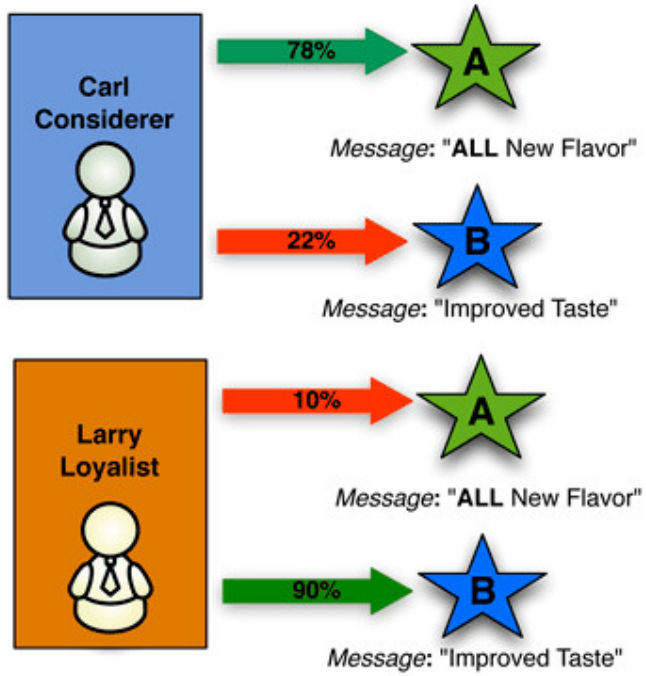
Есть ли у нас дополнительная информация?



Пример с биномиальной моделью

- Бинарный ответ: да/нет
- Наблюдения имеют распределение Бернулли

$$x_i^A \sim \text{Bernoulli}(p_A)$$
$$x_i^B \sim \text{Bernoulli}(p_B)$$

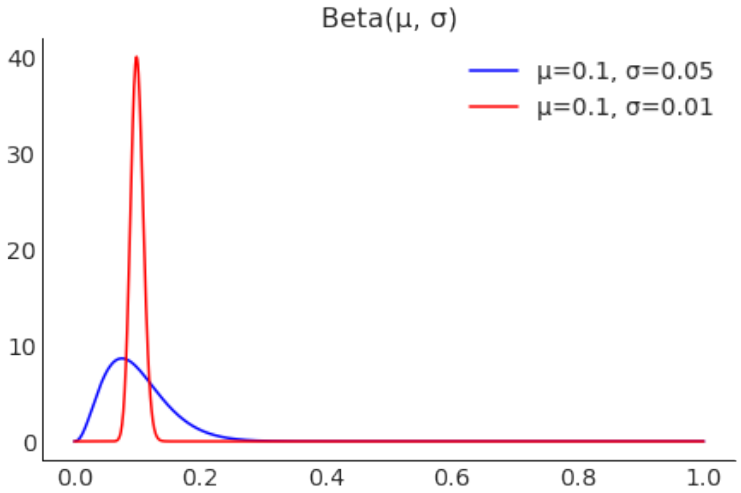


Есть ли у нас дополнительная информация?

- Историческон значение $\bar{p}$
- Ожидаемое улучшение $\pm\bar{\sigma}\%$ (например, $\pm 0.01\%$)

Добавление дополнительной информации

Бета распределение можно параметризовать специальным образом



$$G \in \{A, B\}$$
$$x_i^G \sim \text{Bernoulli}(p_G)$$
$$p_G \sim \text{Beta}(\alpha_G, \beta_G) \; s.t.$$
$$\mathbb{E} p_G = \bar{p},$$
$$\text{Var}\, p_G = \bar{\sigma}^2$$

Добавление дополнительной информации

Бета распределение можно параметризовать специальным образом

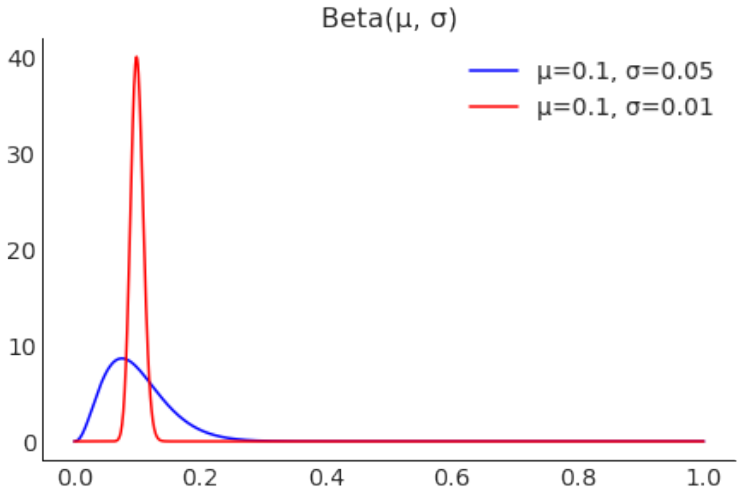
$$X \sim \text{Beta}(\alpha, \beta)$$

$$\mu = \frac{\alpha}{\alpha + \beta}$$

$$\sigma = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

$$X \sim \text{Beta}(\mu, \sigma) \Rightarrow$$

$$\Rightarrow \begin{cases} \alpha & = \mu\kappa \\ \beta & = (1-\mu)\kappa \\ \text{where} & \kappa = \frac{\mu(1-\mu)}{\sigma^2} - 1 \end{cases}$$



$$G \in \{A, B\}$$

$$x_i^G \sim \text{Bernoulli}(p_G)$$

$$p_G \sim \text{Beta}(\alpha_G, \beta_G) \; s.t.$$

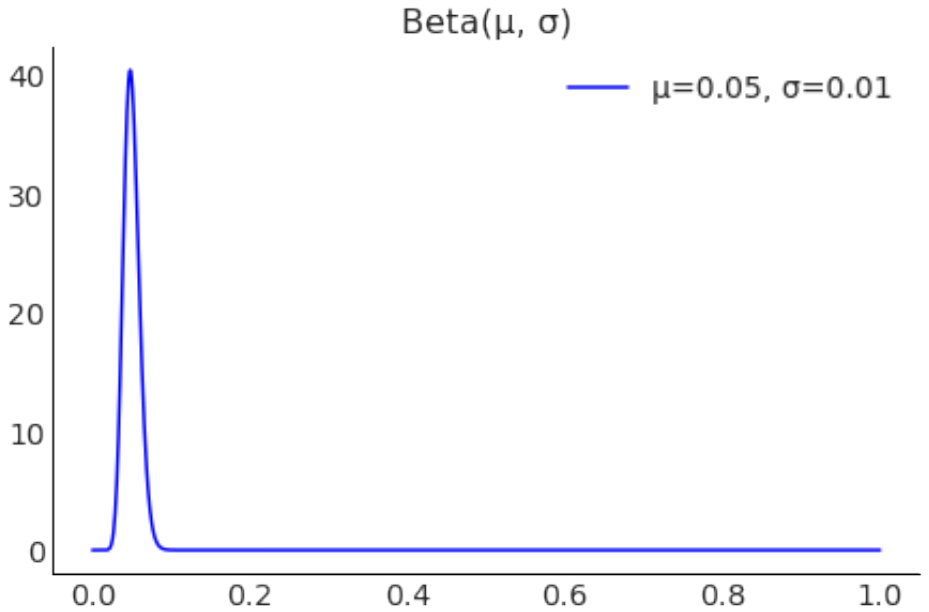
$$\mathbb{E}p_G = \bar{p},$$

$$\text{Var}\, p_G = \bar{\sigma}^2$$

Спецификация априорного распределения

Кейс

Исторические уровни конверсии – около 5% (и фиксированы). Мы ожидаем, что после внедрения решения произойдет изменение примерно на 1% в абсолютном выражении ($\bar{\sigma}$), или 20% в **относительном** ($\bar{\delta}$).



$$\bar{p} = 0.05,\ \bar{\sigma} = 0.01 = \bar{\delta} \cdot 0.05$$

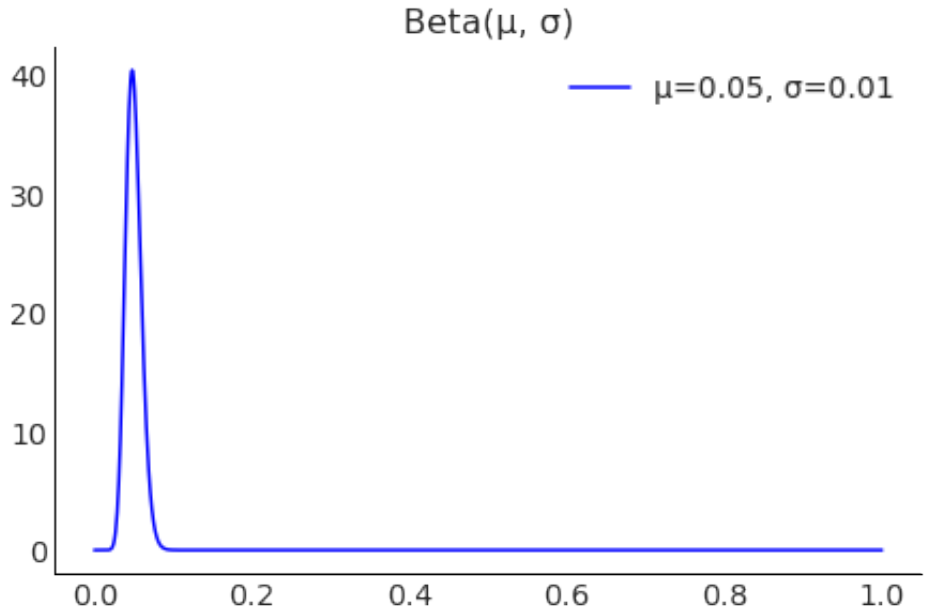
$$G \in \{A, B\}$$

$$p_G \sim \text{Beta}(\mu = \bar{p}, \sigma = \bar{\sigma})$$

Спецификация априорного распределения

Кейс

Исторические уровни конверсии – около 5% (и фиксированы). Мы ожидаем, что после внедрения решения произойдет изменение примерно на 1% в абсолютном выражении ($\bar{\sigma}$), или 20% в **относительном** ($\bar{\delta}$).



$$\bar{p} = 0.05,\ \bar{\sigma} = 0.01 = \bar{\delta} \cdot 0.05$$

$$G \in \{A, B\}$$

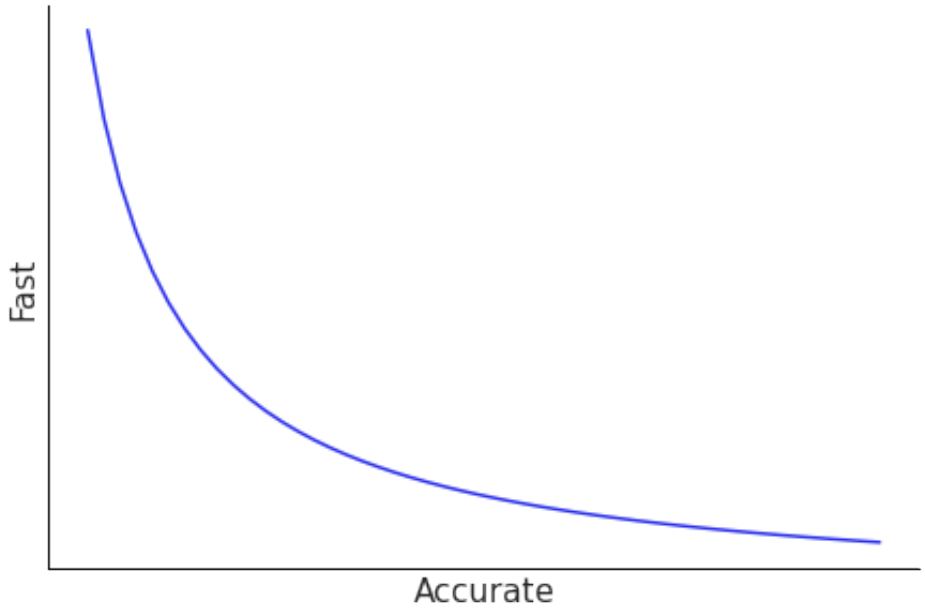
$$p_G \sim \text{Beta}(\mu = \bar{p}, \sigma = \bar{\sigma})$$

Вывод

Такая параметризация бета распределения даёт более интерпретируемые априорные распределения

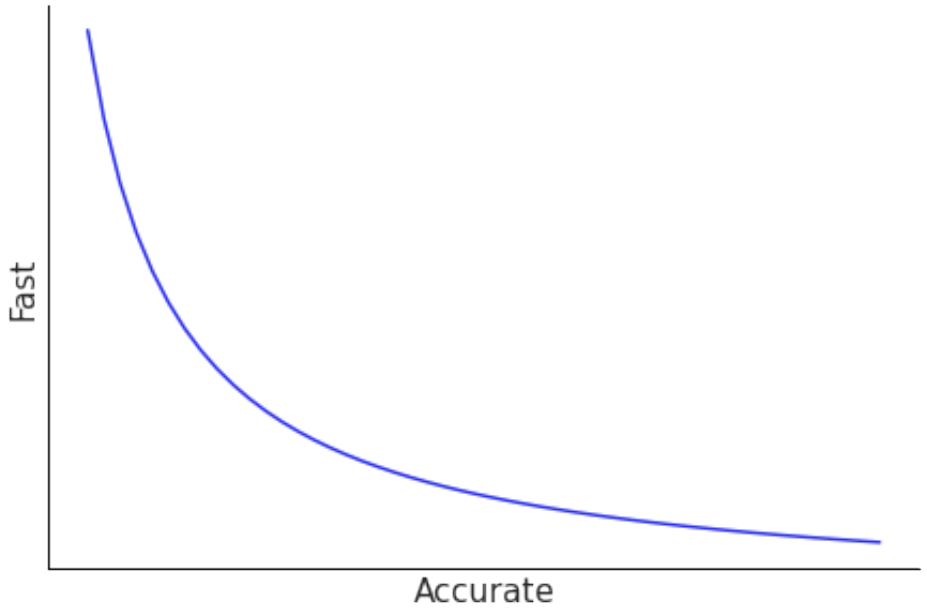
Важные вопросы перед началом

- Сколько времени можно выделить для эксперимента?
  - Насколько точным тогда будет решение?
- Насколько точным должно быть решение?
  - Сколько времени долно быть выделено, чтобы достичь заданной точности?



Важные вопросы перед началом

- Сколько времени можно выделить для эксперимента?
  - Насколько точным тогда будет решение?
- Насколько точным должно быть решение?
  - Сколько времени долно быть выделено, чтобы достичь заданной точности?



Невозможность

Нельзя одновременно быстро собирать данные и быть точным

Parameter Recovery Study

Использование симуляций, чтобы лучше понять свойства модели.

1. Сгенерировать данные из модели с известными параметрами
2. Сделать вид, что параметры неизвестны
3. Оценить параметры из симулированных данных
4. Сравнить оценки с истинными значениями

Используя результаты, ответить на вопросы:

- Насколько хорошо оцениваются параметры?
- Как результаты зависят от размера выборки?
- Есть ли неидентифицируемые параметры?

Рекомендуемое чтение

Глава 4 в Bayesian Workflow

Parameter Recovery Study

Использование симуляций, чтобы лучше понять свойства модели.

1. Сгенерировать данные из модели с известными параметрами
2. Сделать вид, что параметры неизвестны
3. Оценить параметры из симулированных данных
4. Сравнить оценки с истинными значениями

Используя результаты, ответить на вопросы:

- Насколько хорошо оцениваются параметры?
- Как результаты зависят от размера выборки?
- Есть ли неидентифицируемые параметры?

Рекомендуемое чтение

Глава 4 в Bayesian Workflow

Parameter Recovery в A/Б тестировании

Имеем:

- Эффект значим, если $|p - \bar{p}| > \bar{\sigma}$

Модель

$$i \in 1 \dots N$$
$$x_i \sim \text{Bernoulli}(p)$$
$$p \sim \text{Beta}(\mu = \bar{\mu}, \sigma = \bar{\sigma})$$

Parameter Recovery в A/Б тестировании

Имеем:

- Эффект значим, если $|p - \bar{p}| > \bar{\sigma}$
- Игнорируем эффект, если $|p - \bar{p}| < \bar{\sigma}$

Модель

$$i \in 1 \dots N$$
$$x_i \sim \text{Bernoulli}(p)$$
$$p \sim \text{Beta}(\mu = \bar{\mu}, \sigma = \bar{\sigma})$$

Parameter Recovery в A/Б тестировании

Имеем:

- Эффект значим, если $|p - \bar{p}| > \bar{\sigma}$
- Игнорируем эффект, если $|p - \bar{p}| < \bar{\sigma}$
- Насколько большое нужно N, чтобы определить значимость?

Модель

$$i \in 1 \dots N$$
$$x_i \sim \text{Bernoulli}(p)$$
$$p \sim \text{Beta}(\mu = \bar{\mu}, \sigma = \bar{\sigma})$$

Parameter Recovery в А/Б тестировании

Имеем:

- Эффект значим, если $|p - \bar{p}| > \bar{\sigma}$
- Игнорируем эффект, если $|p - \bar{p}| < \bar{\sigma}$
- Насколько большое нужно N, чтобы определить значимость?
- $N = 0$, $N = 1000$, $N = 100000$?

Модель

$$i \in 1 \dots N$$

$$x_i \sim \text{Bernoulli}(p)$$

$$p \sim \text{Beta}(\mu = \bar{\mu}, \sigma = \bar{\sigma})$$

Parameter Recovery в A/Б тестировании

Имеем:

- Эффект значим, если $|p - \bar{p}| > \bar{\sigma}$
- Игнорируем эффект, если $|p - \bar{p}| < \bar{\sigma}$
- Насколько большое нужно N, чтобы определить значимость?
- $N = 0$, $N = 1000$, $N = 100000$?
- Какую метрику использовать для оценки эффективности?

Модель

$$i \in 1 \dots N$$

$$x_i \sim \text{Bernoulli}(p)$$

$$p \sim \text{Beta}(\mu = \bar{\mu}, \sigma = \bar{\sigma})$$

Parameter Recovery в А/Б тестировании

Имеем:

- Эффект значим, если $|p - \bar{p}| > \bar{\sigma}$
- Игнорируем эффект, если $|p - \bar{p}| < \bar{\sigma}$
- Насколько большое нужно N, чтобы определить значимость?
- $N = 0$, $N = 1000$, $N = 100000$?
- Какую метрику использовать для оценки эффективности?

Ключевое наблюдение

Обнаружение эффекта – задача классификации: отрицательные, нейтральные, положительные эффекты. Можно использовать ROC-AUC.

Модель

$$i \in 1 \dots N$$

$$x_i \sim \text{Bernoulli}(p)$$

$$p \sim \text{Beta}(\mu = \bar{\mu}, \sigma = \bar{\sigma})$$

A/Б тестирование как классификация

Определения для задачи классификации

Модель

$$i \in 1 \dots N$$
$$x_i \sim \text{Bernoulli}(p)$$
$$p \sim \text{Beta}(\mu = \bar{\mu}, \sigma = \bar{\sigma})$$

Апостериорное распределение $p(p \mid X_{1:N})$

A/Б тестирование как классификация

Определения для задачи классификации

1. Таргет $\hat{p}$ – для генерации данных

Модель

$$i \in 1 \dots N$$
$$x_i \sim \text{Bernoulli}(p)$$
$$p \sim \text{Beta}(\mu = \bar{\mu}, \sigma = \bar{\sigma})$$

Апостериорное распределение $p(p \mid X_{1:N})$

A/Б тестирование как классификация

Определения для задачи классификации

1. Таргет $\hat{p}$ – для генерации данных
2. Метки
   - "0" при $\hat{p} < \bar{p} - \bar{\sigma}$ (отриц.)
   - "1" при $\bar{p} - \bar{\sigma} < \hat{p} < \bar{p} + \bar{\sigma}$ (нейтр.)
   - "2" при $\hat{p} > \bar{p} + \bar{\sigma}$ (положит.)

Модель

$$i \in 1 \dots N$$

$$x_i \sim \text{Bernoulli}(p)$$

$$p \sim \text{Beta}(\mu = \bar{\mu}, \sigma = \bar{\sigma})$$

Апостериорное распределение $p(p \mid X_{1:N})$

A/Б тестирование как классификация

Определения для задачи классификации

1. Таргет $\hat{p}$ – для генерации данных
2. Метки
   - "0" при $\hat{p} < \bar{p} - \bar{\sigma}$ (отриц.)
   - "1" при $\bar{p} - \bar{\sigma} < \hat{p} < \bar{p} + \bar{\sigma}$ (нейтр.)
   - "2" при $\hat{p} > \bar{p} + \bar{\sigma}$ (положит.)
3. Предсказания (вероятности на основе апостериорного распределения): $P(p < 0 \mid X_{1:N})$, $P(p \approx 0 \mid X_{1:N})$, $P(p > 0 \mid X_{1:N})$

Модель

$$i \in 1 \dots N$$

$$x_i \sim \text{Bernoulli}(p)$$

$$p \sim \text{Beta}(\mu = \bar{\mu}, \sigma = \bar{\sigma})$$

Апостериорное распределение $p(p \mid X_{1:N})$

A/Б тестирование как классификация

Определения для задачи классификации

1. Таргет $\hat{p}$ – для генерации данных
2. Метки
   - "0" при $\hat{p} < \bar{p} - \bar{\sigma}$ (отриц.)
   - "1" при $\bar{p} - \bar{\sigma} < \hat{p} < \bar{p} + \bar{\sigma}$ (нейтр.)
   - "2" при $\hat{p} > \bar{p} + \bar{\sigma}$ (положит.)
3. Предсказания (вероятности на основе апостериорного распределения): $P(p < 0 \mid X_{1:N})$, $P(p \approx 0 \mid X_{1:N})$, $P(p > 0 \mid X_{1:N})$

Симуляционный эксперимент

1. для $\hat{p} \in \ldots$ и $N \in \ldots$ получить $p(p \mid X_{1:N})$
2. для $N \in \ldots$ вычислить ROC-AUC

Модель

$$i \in 1 \ldots N$$

$$x_i \sim \text{Bernoulli}(p)$$

$$p \sim \text{Beta}(\mu = \bar{\mu}, \sigma = \bar{\sigma})$$

Апостериорное распределение $p(p \mid X_{1:N})$

ROC-AUC в действии

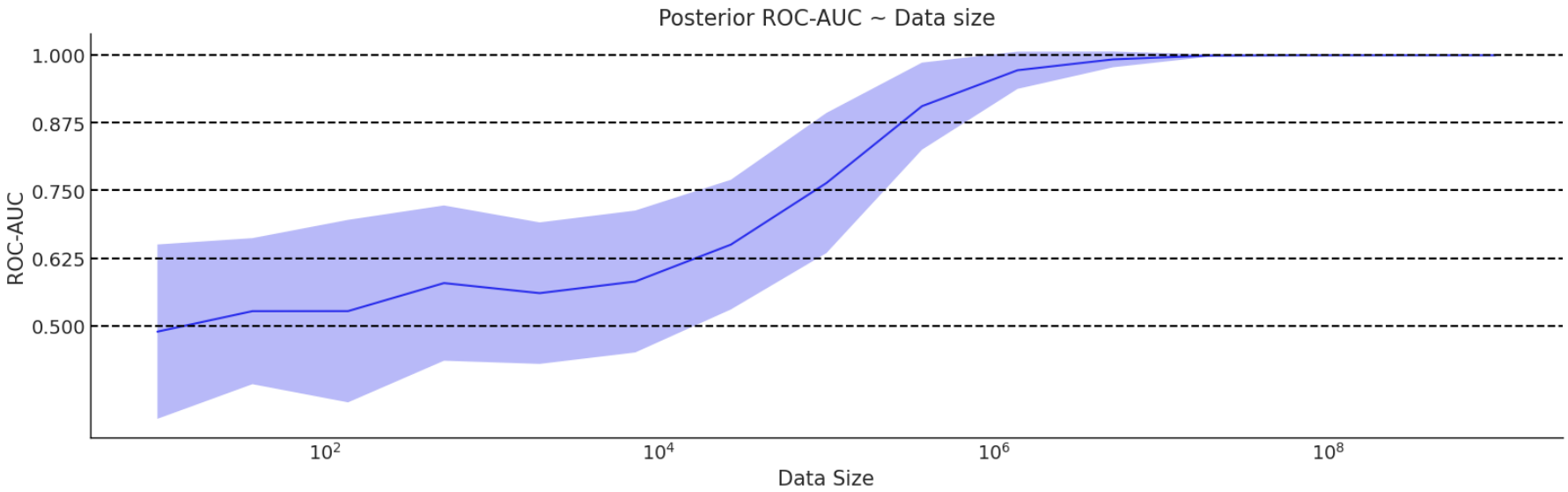


Figure: ROC-AUC возрастает с ростом числа наблюдений

Ограничение: время

1. Обсудите максимальный запас времени
2. Опирайтесь на график ROC-AUC

Ограничение: ROC-AUC

1. Обсудите минимальное требуемое значение ROC-AUC
2. График подскажет ожидаемый объём выборки

После оценки параметров

Ситуация: вы провели эксперимент в течении заданного времени. Главные вопросы:

1. Какую альтернативу выбрать?
2. Какой критерий сравнения?
3. Сопоставим ли этот критерий с реальной жизнью?

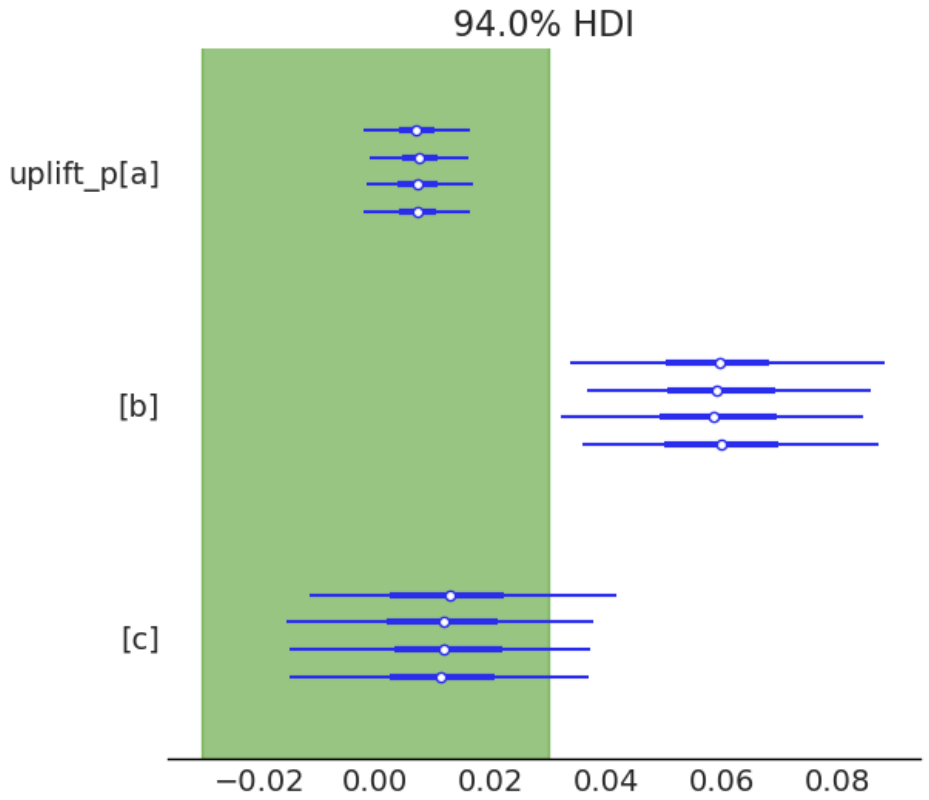


Figure: Пример графика ROPE

После оценки параметров

Ситуация: вы провели эксперимент в течении заданного времени. Главные вопросы:

1. Какую альтернативу выбрать?
2. Какой критерий сравнения?
3. Сопоставим ли этот критерий с реальной жизнью?

Более хорошая метрика

Хорошая метрика та, которая связана с ожидаемым доходом.

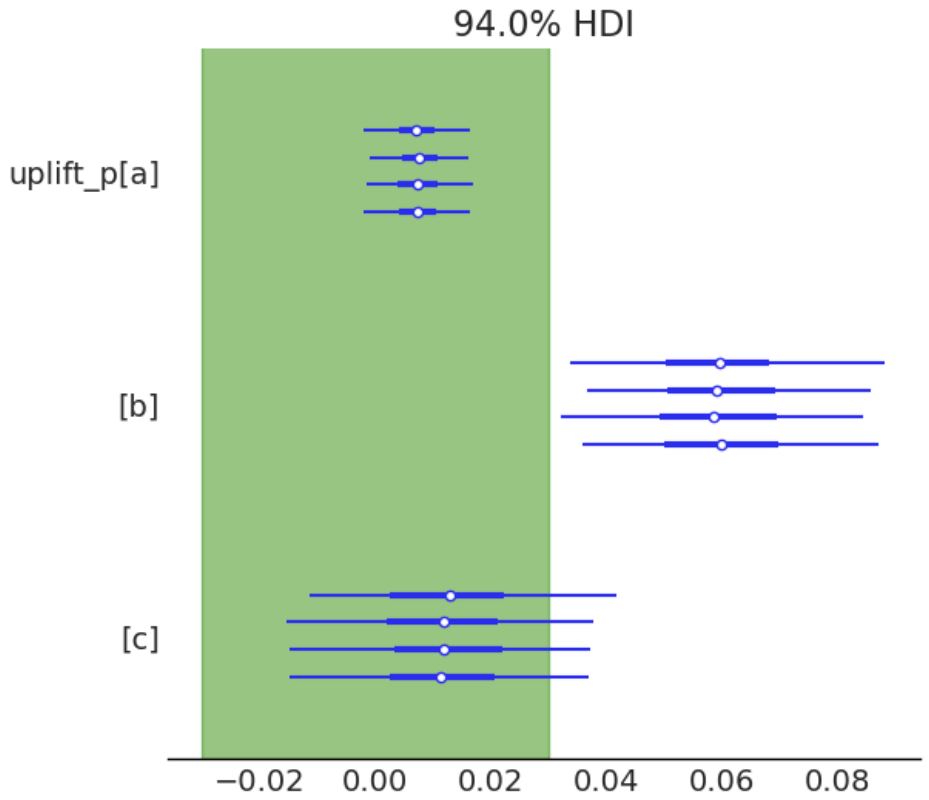


Figure: Пример графика ROPE

Интерпретация апостериорного распределения

Как можно вычислить метрику получше?

Это может выглядеть так:

Интерпретация апостериорного распределения

Как можно вычислить метрику получше?

- Привязать коэффициент конверсии p_A или p_B к количеству клиентов компании

Это может выглядеть так:

Интерпретация апостериорного распределения

Как можно вычислить метрику получше?

- Привязать коэффициент конверсии p_A или p_B к количеству клиентов компании
- Использовать “стоимость клиента” в качестве показателя денежного эффекта

Это может выглядеть так:

Интерпретация апостериорного распределения

Как можно вычислить метрику получше?

- Привязать коэффициент конверсии p_A или p_B к количеству клиентов компании
- Использовать “стоимость клиента” в качестве показателя денежного эффекта

Это может выглядеть так:

$$\text{монетизация}_A =$$

$$(\text{стоимость клиента}) \times (\text{число клиентов}) \times \Delta p_A - (\text{стоимость реализации})$$

Интерпретация апостериорного распределения

Как можно вычислить метрику получше?

- Привязать коэффициент конверсии p_A или p_B к количеству клиентов компании
- Использовать "стоимость клиента" в качестве показателя денежного эффекта

Это может выглядеть так:

$$\text{монетизация}_A =$$

$$(\text{стоимость клиента}) \times (\text{число клиентов}) \times \Delta p_A - (\text{стоимость реализации})$$

Используйте апостериорное распределение

Можно вычислить $p(\text{монетизация}_A \mid X_A)$ из $p(p_A \mid X_A)$

Апостериорное распределение монетизации

$$(\text{стоимость клиента}) \times (\text{число клиентов}) \times \Delta p_A - (\text{стоимость реализации})$$

- Стоимости реализации могут различаться
- Стоимости клиентов могут зависеть от сценариев
- Вы соединяете эксперимент с бизнесом
- Сравните результаты с неопределённостью

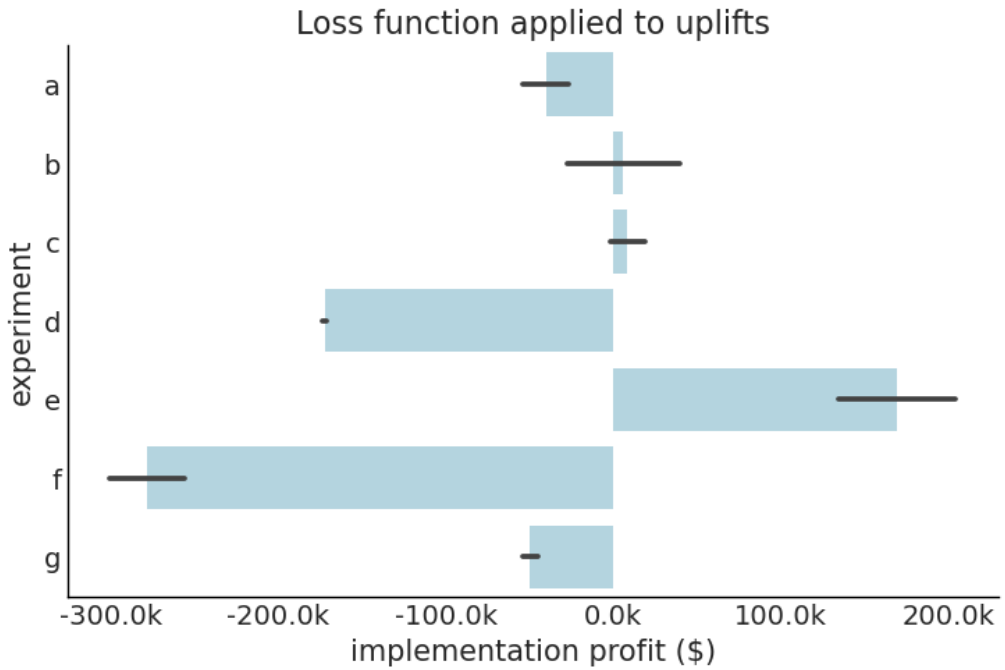


Figure: $p(\text{монетизация}_G \mid X_G)$

Выводы

Реальные А/Б тесты полны трудностей. Байесовские методы могут сделать гораздо больше для превращения данных в действие.

1. Формализация статистического теста
   - Определение априорных распределений
   - Определение правдоподобия
2. Планирование эксперимента
   - Parameter recovery study
3. Байесовское принятие решений для выбора действия
   - Функции потерь
   - Тестирование сценариев

Библиография I


R. Kohavi, A. Deng, Y. Xu, and T. Walker.
In *Improving the Sensitivity of Online Controlled Experiments by Utilizing Pre-Experiment Data*, 02 2013.