

Байесовское А/Б тестирование

Максим Кочуров

МГУ им. М.В. Ломоносова

Лекция 3



Содержание

① Классический подход

- Предположения

② Байесовское тестирование гипотез

- Интервал наибольшей плотности
- Region of Practical Equivalence
- Произвольные гипотезы

③ А/Б тестирование

- Априорные распределения

④ Пример

- Априорное распределение
- Подготовка эксперимента
- Parameter Recovery
- Симуляции из апостериорного распределения



Как это делается: классический подход

“Если ваше р-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 р-значения используются в тысячах статей
- ② р-значения очень популярны из-за простоты интерпретации
- ③ легко вычислить доверительные интервалы



Как это делается: классический подход

“Если ваше р-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 р-значения используются в тысячах статей
- ② р-значения очень популярны из-за простоты интерпретации
- ③ легко вычислить доверительные интервалы

Вы уверены?

Действительно ли вы понимаете смысл р-значений?



Понимаете ли вы р-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 должна быть отвергнута
- ④ Значение $P > 0.05$ значит, что эффект отсутствует



Понимаете ли вы р-значения?

Что из нижеперечисленного верно?

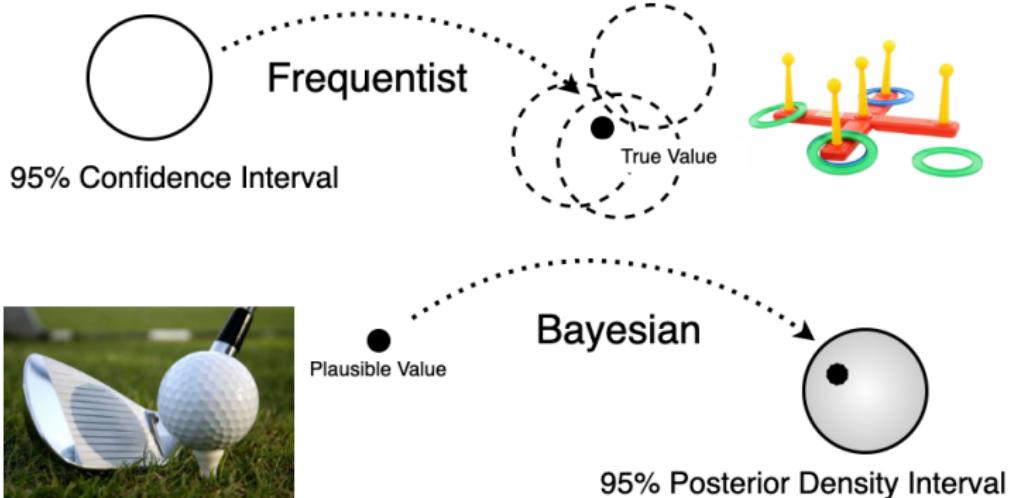
- ① p – вероятность того, что верна нулевая гипотеза
- ② $(1 - p)$ – вероятность того, что верна альтернативная гипотеза
- ③ $P \leq 0.05$ значит, что нулевая гипотеза неверна и должна быть отвергнута
- ④ Значение $P > 0.05$ значит, что эффект отсутствует

Использовать р-значение плохо?

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



Интерпретация р-значений





Тестирование гипотез во фреймворке “Н0 против Н1”

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

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

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

- гипотеза о равенстве дисперсий

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

Слишком сложно

Чем меньше предположений, тем сложнее вычисления и реализация



Байесовский подход

Осторожно с интерпретацией р-значений!

- Частотнические доверительные интервалы – это не наиболее вероятные значения
- р-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

Пример

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

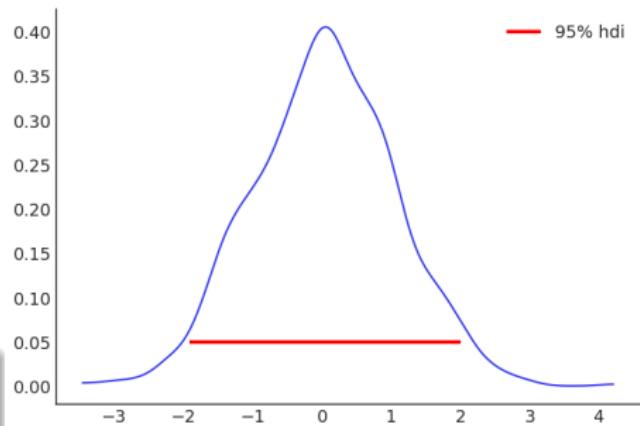


Figure: Интервал наибольшей плотности



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.

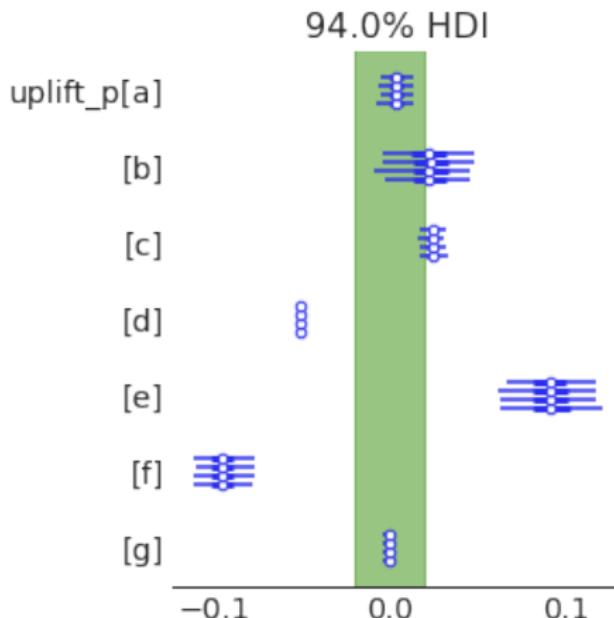


Figure: График RoPE



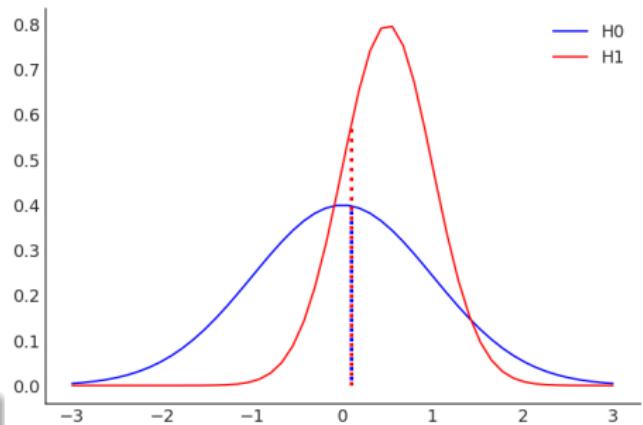
Коэффициент Байеса

По-моему, эту статистику сложнее всего объяснить.

- ① Похоже на частотническое р-значение
- ② Сложнее интерпретировать и объяснить
- ③ Проверяет H_0 против H_1 для x_0

Определение

Коэффициент Байеса – отношение правдоподобия одной гипотезы к правдоподобию второй гипотезы



$$\text{Figure: } BF = \frac{\text{pdf}_{H_1}(x_0)}{\text{pdf}_{H_0}(x_0)}$$

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 -
 $Q_{0.05}(\text{выручка}(X, \Theta))$

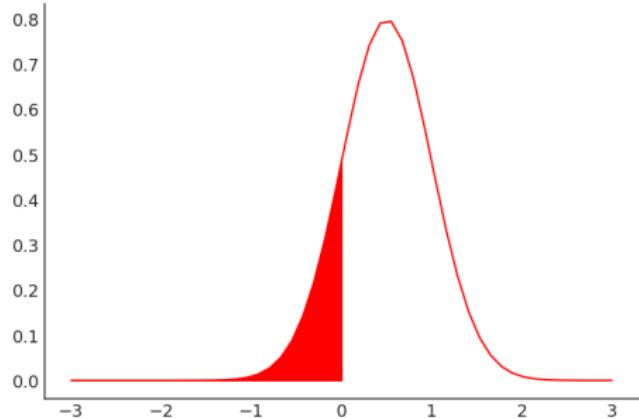


Figure: $P(A < 0)$



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[?])
- 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[?])
- ④ С иерархией (регионы)

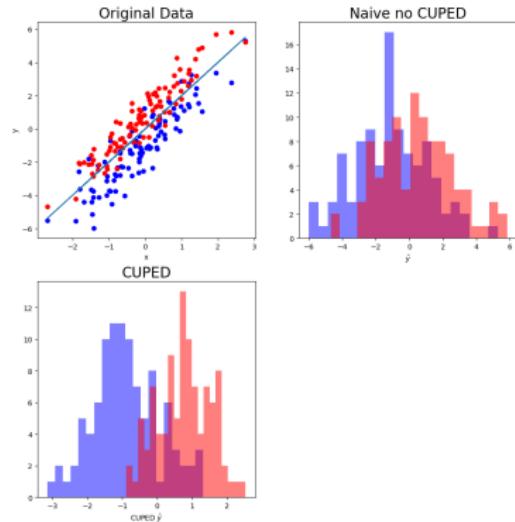




Типы задач

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[?])
- ④ С иерархией (регионы)

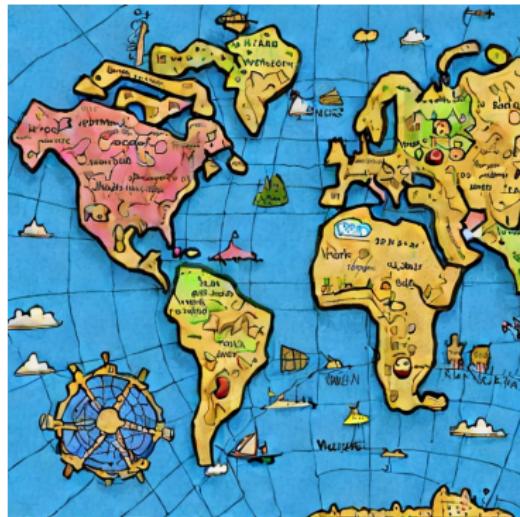




Типы задач

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[?])
- ④ С иерархией (регионы)



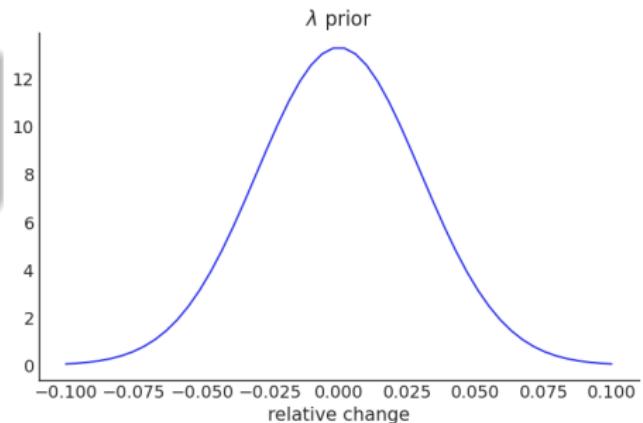


Подходы к априорным распределениям

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 В с контрольной группой А. Вы можете быть заинтересованы в увеличении какой-либо статистики (например, среднего чека).



Подготовка

Вы готовитесь к проведению эксперимента над группой В с контрольной группой А. 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 В с контрольной группой А. 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 В с контрольной группой А. 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 В с контрольной группой А. 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 В с контрольной группой А. 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 В с контрольной группой А. 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?



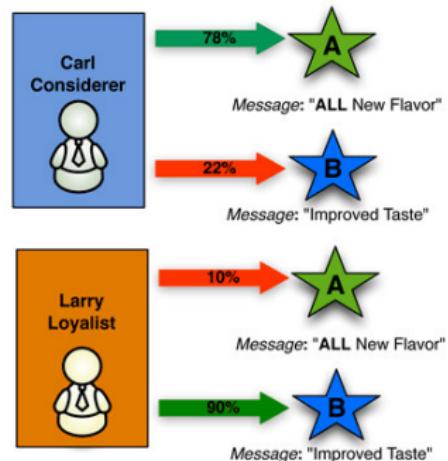
Пример с биномиальной моделью

- 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

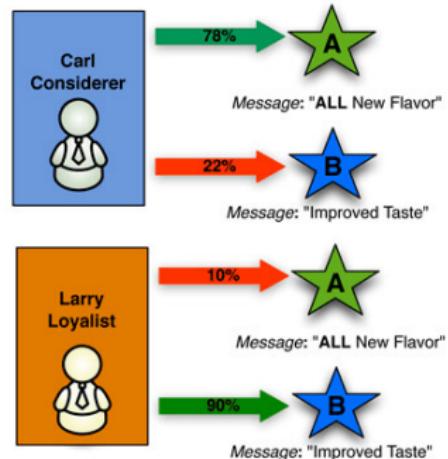
- Бинарный ответ: да/нет
- Наблюдения имеют распределение Бернулли

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

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

Есть ли у нас дополнительная информация?

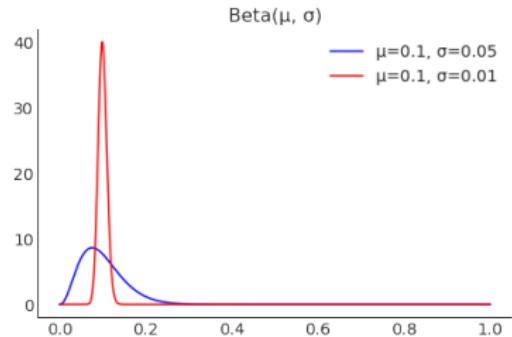
- Историческое значение \bar{p}
- Ожидаемое улучшение $\pm\bar{\sigma}\%$
(например, $\pm 0.01\%$)





Добавление дополнительной информации

Бета распределение можно параметризовать специальным образом



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

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

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

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

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



Добавление дополнительной информации

Бета распределение можно параметризовать специальным образом

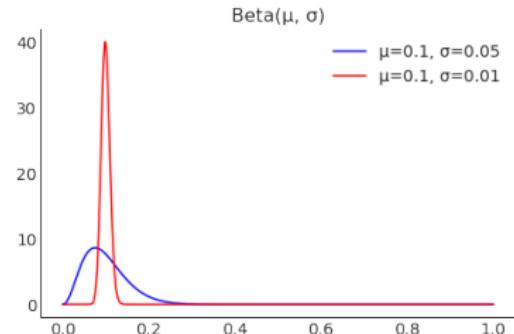
$$X \sim \text{Beta}(\alpha, \beta)$$

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

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

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

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



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

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

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

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

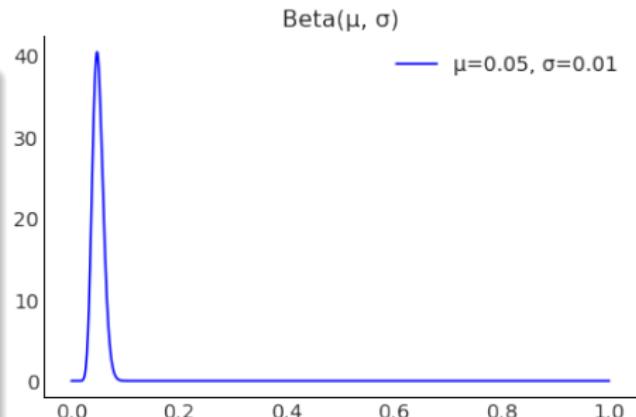
$$\text{Var } p_G = \bar{\sigma}^2$$



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 ($\bar{\delta}$).



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

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

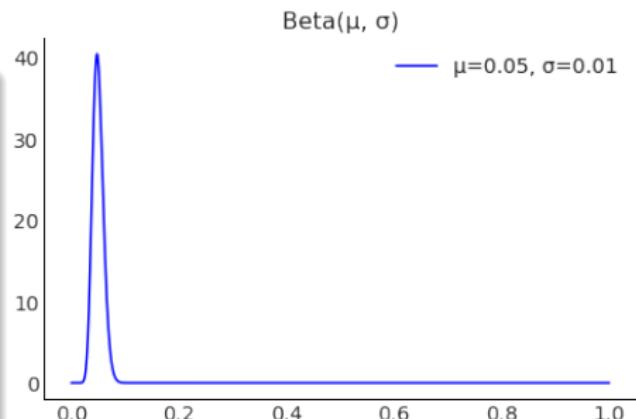
$$p_G \sim \text{Beta}(\mu = \bar{p}, \sigma = \bar{\sigma})$$



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 ($\bar{\delta}$).



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

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

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

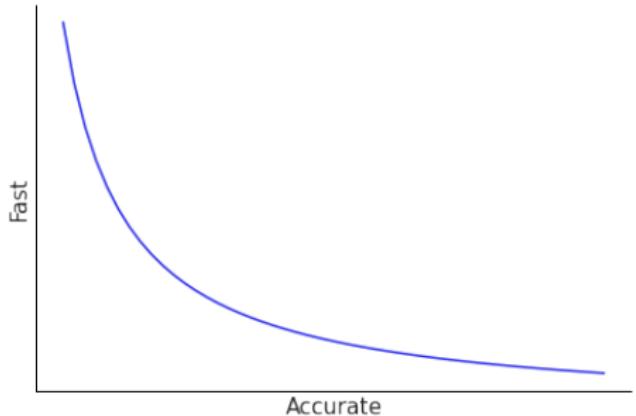
Вывод

Такая параметризация бета распределения даёт более интерпретируемые априорные распределения



Важные вопросы перед началом

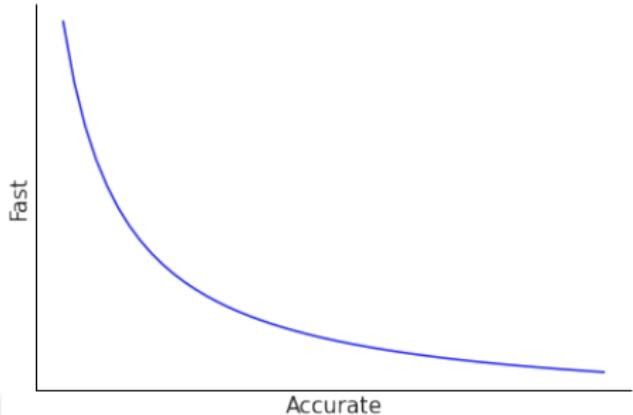
- Сколько времени можно выделить для эксперимента?
 - Насколько точным тогда будет решение?
- Насколько точным должно быть решение?
 - Сколько времени должно быть выделено, чтобы достичь заданной точности?





Важные вопросы перед началом

- Сколько времени можно выделить для эксперимента?
 - Насколько точным тогда будет решение?
- Насколько точным должно быть решение?
 - Сколько времени должно 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 А/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 А/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 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?

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 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?

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?

Модель

$$i \in 1 \dots N$$

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

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

Ключевое наблюдение

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



А/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 $p(p \mid X_{1:N})$



А/Б тестирование как классификация

Определения для задачи классификации

- ➊ 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 $p(p \mid X_{1:N})$



А/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 $p(p | X_{1:N})$



A/B 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):

$$P(p < 0 | X_{1:N}), P(p \approx 0 | X_{1:N}), \\ P(p > 0 | X_{1:N})$$

Модель

$$i \in 1 \dots N$$

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

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

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



A/B 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):

$$P(p < 0 | X_{1:N}), P(p \approx 0 | X_{1:N}),$$

$$P(p > 0 | X_{1:N})$$

Модель

$$i \in 1 \dots N$$

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

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

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

Симуляционный эксперимент

① для $\hat{p} \in \dots$ и $N \in \dots$ получить
 $p(p | X_{1:N})$

② для $N \in \dots$ вычислить ROC-AUC



ROC-AUC в действии

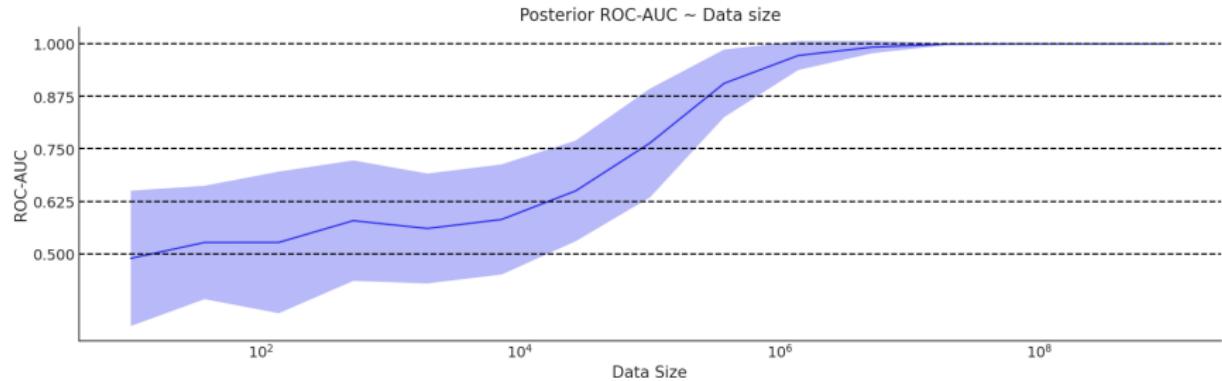


Figure: ROC-AUC возрастает с ростом числа наблюдений

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 течение заданного времени. Главные вопросы:

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

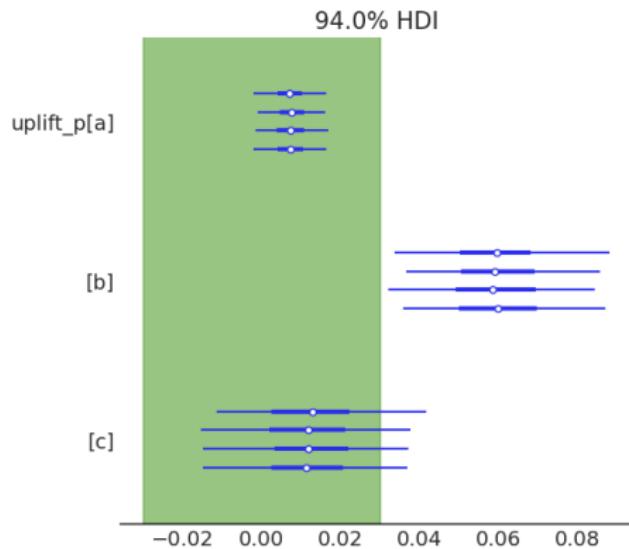


Figure: Пример графика ROPE



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.

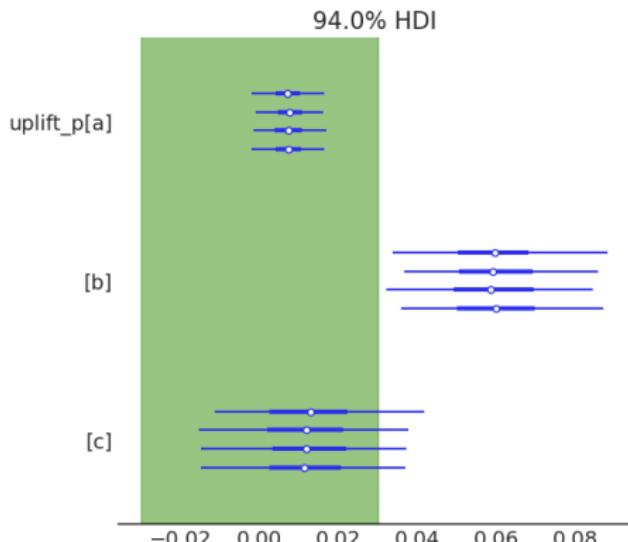


Figure: Пример графика ROPE



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 $p(\text{монетизация}_A | X_A)$ из $p(p_A | X_A)$



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 неопределенностью

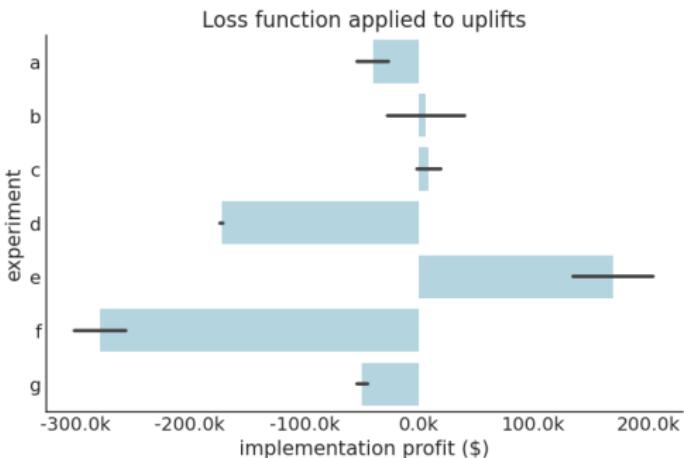


Figure: $p(\text{монетизация}_G \mid X_G)$



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