Байесовские многорукие бандиты

Интеллектуальный анализ данных, 20!7

По материалам open data science

Малютин Евгений Алексеевич

Введение

В прошлых сериях

- доверительные и предсказательные интервалы
- гипотезы и их проверка
- A/B тестирование

Постановка задачи:

Фирма "Усы и Хвосты"

- у фирмы есть сайт, на сайте есть кнопка
- при нажатии на кнопку где-то мурлыкает котик
- цель фирмы – максимизировать мурлыкание. Дизайнеры нарисовали много новых кнопок
- вопрос – какая кнопка лучше?

Постановка задачи

Формально:

- клик – случайная переменная $k = \{0, 1\}$

Постановка задачи

Формально:

- клик – случайная переменная $k = \{0, 1\}$
- при этом:

$$\begin{aligned} k &\sim \mathrm{Bernoulli}\left(\theta\right) \\ p\left(k\right) &= \theta^k \left(1 - \theta\right)^{1-k} \end{aligned}$$

Формально:

- клик – случайная переменная $k = \{0, 1\}$
- при этом:

$$\begin{aligned} k &\sim \text{Bernoulli}\left(\theta\right) \\ p\left(k\right) &= \theta^k \left(1 - \theta\right)^{1-k} \end{aligned}$$

- гипотезы:

Постановка задачи

Формально:

- клик – случайная переменная $k = \{0, 1\}$
- при этом:

$$\begin{aligned} k &\sim \text{Bernoulli}(\theta) \\ p(k) &= \theta^k (1-\theta)^{1-k} \end{aligned}$$

- гипотезы:
  - нулевая гипотеза $H_0 : \sigma_c = \sigma_t$ нет никакой разницы;

Постановка задачи

Формально:

- клик – случайная переменная $k = \{0, 1\}$
- при этом:

$$\begin{aligned} k &\sim \text{Bernoulli}\,(\theta) \\ p\,(k) &= \theta^k\,(1-\theta)^{1-k} \end{aligned}$$

- гипотезы:
    - нулевая гипотеза $H_0 : \sigma_c = \sigma_t$ нет никакой разницы;
    - альтернативная гипотеза: $H_1 : \sigma_c < \sigma_t$

Постановка задачи

Формально:

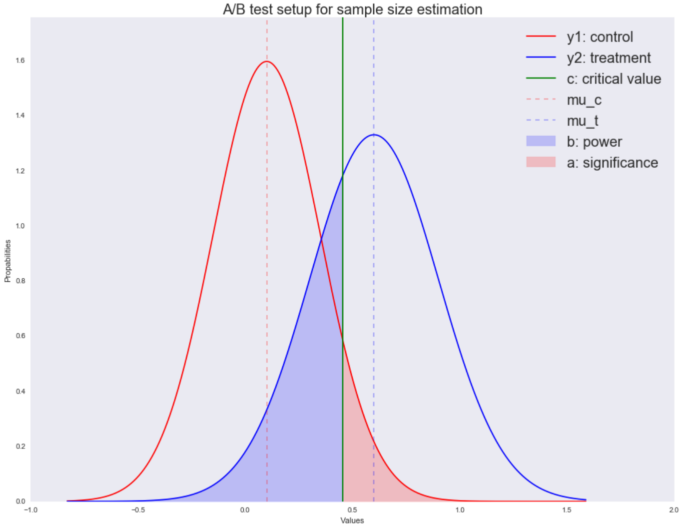
- клик – случайная переменная $k = \{0, 1\}$
- при этом:

$$\begin{aligned} k &\sim \text{Bernoulli}\left(\theta\right) \\ p\left(k\right) &= \theta^{k}\left(1-\theta\right)^{1-k} \end{aligned}$$

- гипотезы:
  - нулевая гипотеза $H_0 : \sigma_c = \sigma_t$ нет никакой разницы;
  - альтернативная гипотеза: $H_1 : \sigma_c < \sigma_t$
  - По ЗБЧ и ЦПТ:

$$\bar{\theta} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Классический A/B/n тест



Продолжаем про A/B тесты:

Оценка размера выборки:

- Порог c:

$$c = \mu + t\frac{\sigma}{\sqrt{n}}$$

Продолжаем про A/B тесты:

Оценка размера выборки:

- Порог c:

$$c = \mu + t\frac{\sigma}{\sqrt{n}}$$

- Квантильные оценки:

$$\begin{cases} c = \theta_c + t_\alpha\sqrt{\frac{\theta_c(1-\theta_c)}{n}} \\ c = \theta_t + t_\beta\sqrt{\frac{\theta_t(1-\theta_t)}{n}} \end{cases}$$

Продолжаем про A/B тесты:

Оценка размера выборки:

- Порог c:

$$c = \mu + t\frac{\sigma}{\sqrt{n}}$$

- Квантильные оценки:

$$\begin{cases} c = \theta_c + t_\alpha\sqrt{\frac{\theta_c(1-\theta_c)}{n}} \\ c = \theta_t + t_\beta\sqrt{\frac{\theta_t(1-\theta_t)}{n}} \end{cases}$$

- Оценка sample size размера:

$$\begin{aligned} \theta_c + t_\alpha\sqrt{\frac{\theta_c(1-\theta_c)}{n}} &= \theta_t + t_\beta\sqrt{\frac{\theta_t(1-\theta_t)}{n}} \\ \theta_c\sqrt{n} + t_\alpha\sqrt{\theta_c\left(1-\theta_c\right)} &= \theta_t\sqrt{n} + t_\beta\sqrt{\theta_t\left(1-\theta_t\right)} \\ n &= \left(\frac{t_\beta\sqrt{\theta_t(1-\theta_t)} - t_\alpha\sqrt{\theta_c(1-\theta_c)}}{\theta_c-\theta_t}\right)^2 \end{aligned}$$

Продолжаем про A/B тесты:

Оценка размера выборки:

- Порог c:

$$c = \mu + t\frac{\sigma}{\sqrt{n}}$$

- Квантильные оценки:

$$\begin{cases} c = \theta_c + t_\alpha\sqrt{\frac{\theta_c(1-\theta_c)}{n}} \\ c = \theta_t + t_\beta\sqrt{\frac{\theta_t(1-\theta_t)}{n}} \end{cases}$$

- Оценка sample size размера:

$$\begin{aligned} \theta_c + t_\alpha\sqrt{\frac{\theta_c(1-\theta_c)}{n}} &= \theta_t + t_\beta\sqrt{\frac{\theta_t(1-\theta_t)}{n}} \\ \theta_c\sqrt{n} + t_\alpha\sqrt{\theta_c\left(1-\theta_c\right)} &= \theta_t\sqrt{n} + t_\beta\sqrt{\theta_t\left(1-\theta_t\right)} \\ n &= \left(\frac{t_\beta\sqrt{\theta_t(1-\theta_t)} - t_\alpha\sqrt{\theta_c(1-\theta_c)}}{\theta_c-\theta_t}\right)^2 \end{aligned}$$

- Пусть мы меряем 0.001 vs 0.0011, при $\alpha = \beta = 0.01$ нужно прогнать 2269319 (и A, и B). Новая вариация должна показать = 0.00104.

Поговорим про n

Ну а теперь про n:

- Пусть $\alpha = 0.05$, а тестов в эксперименте 5, посмотрим на ошибку:

$$\begin{aligned} P\left(\text{хотя бы один результат значимый}\right) &= 1 - P\left(\text{все результаты незначимы}\right) \\ &= 1 - (1 - 0.05)^5 \\ &= 1 - 0.95^5 \\ &\approx 0.2262 \end{aligned}$$

Поговорим про n

Ну а теперь про n:

- Пусть $\alpha = 0.05$, а тестов в эксперименте 5, посмотрим на ошибку:

$$\begin{aligned} P\,(\text{хотя бы один результат значимый}) &= 1 - P\,(\text{все результаты незначимы}) \\ &= 1 - (1 - 0.05)^5 \\ &= 1 - 0.95^5 \\ &\approx 0.2262 \end{aligned}$$

- а теперь посмотрим с $\alpha = 0.01$:

$$\begin{aligned} P\,(\text{хотя бы один результат значимый}) &= 1 - P\,(\text{все результаты незначимы}) \\ &= 1 - (1 - 0.01)^5 \\ &= 1 - 0.99^5 \\ &\approx 0.0491 \end{aligned}$$

Поговорим про n

Ну а теперь про n:

- Пусть $\alpha = 0.05$, а тестов в эксперименте 5, посмотрим на ошибку:

$$\begin{aligned} P\,(\text{хотя бы один результат значимый}) &= 1 - P\,(\text{все результаты незначимы}) \\ &= 1 - (1 - 0.05)^5 \\ &= 1 - 0.95^5 \\ &\approx 0.2262 \end{aligned}$$

- а теперь посмотрим с $\alpha = 0.01$:

$$\begin{aligned} P\,(\text{хотя бы один результат значимый}) &= 1 - P\,(\text{все результаты незначимы}) \\ &= 1 - (1 - 0.01)^5 \\ &= 1 - 0.99^5 \\ &\approx 0.0491 \end{aligned}$$

- а значит нам надо 2853873 vs 2269319 (на каждую!), а это на 26% больше траффика

Поговорим про n

Ну а теперь про n:

- Пусть $\alpha = 0.05$, а тестов в эксперименте 5, посмотрим на ошибку:

$$\begin{aligned} P\,(\text{хотя бы один результат значимый}) &= 1 - P\,(\text{все результаты незначимы}) \\ &= 1 - (1 - 0.05)^5 \\ &= 1 - 0.95^5 \\ &\approx 0.2262 \end{aligned}$$

- а теперь посмотрим с $\alpha = 0.01$:

$$\begin{aligned} P\,(\text{хотя бы один результат значимый}) &= 1 - P\,(\text{все результаты незначимы}) \\ &= 1 - (1 - 0.01)^5 \\ &= 1 - 0.99^5 \\ &\approx 0.0491 \end{aligned}$$

- а значит нам надо 2853873 vs 2269319 (на каждую!), а это на 26% больше траффика
- зомби-лосось

К слову в защиту A/B:

Особенности

- пришли из медицины, сельского хозяйства, экономика

К слову в защиту A/B:

Особенности

- пришли из медицины, сельского хозяйства, экономика
- стоимость одного эксперимента существенна

К слову в защиту A/B:

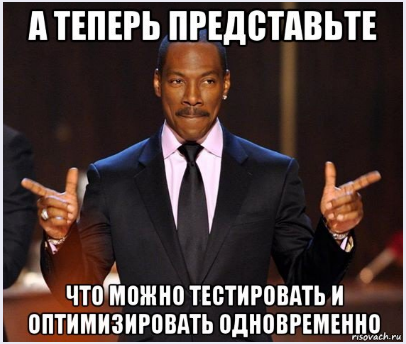
Особенности

- пришли из медицины, сельского хозяйства, экономика
- стоимость одного эксперимента существенна
- в итоге, зная **практическую** значимость мы можем заранее оценить sample size и бюджет

К слову в защиту A/B:

Особенности

- пришли из медицины, сельского хозяйства, экономика
- стоимость одного эксперимента существенна
- в итоге, зная **практическую** значимость мы можем заранее оценить sample size и бюджет
- в онлайне всё немного по-другому...



Bayesian multi-armed bandits

Неформально

- вы в казино, у вас конечное количество денег и времени

Bayesian multi-armed bandits

Неформально

- вы в казино, у вас конечное количество денег и времени
- вы хотите быстро найти лучший автомат и получить денег

Bayesian multi-armed bandits

Неформально

- вы в казино, у вас конечное количество денег и времени
- вы хотите быстро найти лучший автомат и получить денег
- существует много подходов, мы рассмотрим один из них, основанный на семплировании по Томпсону

Bayesian multi-armed bandits

Неформально

- вы в казино, у вас конечное количество денег и времени
- вы хотите быстро найти лучший автомат и получить денег
- существует много подходов, мы рассмотрим один из них, основанный на семплировании по Томпсону
- Yahoo так, например, оптимизируют выдачу. И Microsoft баннеры крутят

Bayesian multi-armed bandits

Формально:

- пусть к моменту времени t мы наблюдаем последовательность наград $y_t = (y_1, y_2, ..., y_t)$.

Формально:

- пусть к моменту времени t мы наблюдаем последовательность наград $y_t = (y_1, y_2, ..., y_t)$.
- Обозначим действие, принятое в момент времени t как a_t (индекс бандита, цвет кнопки).

Формально:

- пусть к моменту времени t мы наблюдаем последовательность наград $y_t = (y_1, y_2, ..., y_t)$.
- Обозначим действие, принятое в момент времени t как a_t (индекс бандита, цвет кнопки).
- Также считаем, что каждый y_t сгенерирован независимо из некоторого распределения наград своего бандита $f_{a_t}(y|\overrightarrow{\theta})$, где $\overrightarrow{\theta}$ — это некоторый вектор параметров.

Формально:

- пусть к моменту времени t мы наблюдаем последовательность наград $y_t = (y_1, y_2, ..., y_t)$.
- Обозначим действие, принятое в момент времени t как a_t (индекс бандита, цвет кнопки).
- Также считаем, что каждый y_t сгенерирован независимо из некоторого распределения наград своего бандита $f_{a_t}(y|\overrightarrow{\theta})$, где $\overrightarrow{\theta}$ — это некоторый вектор параметров.
- Награда – это мат. ожидание бандита. Знай мы распределение, легко бы выбрали и радовались, но мы не знаем.

Bayesian multi-armed bandits

Формально:

- пусть к моменту времени t мы наблюдаем последовательность наград $y_t = (y_1, y_2, ..., y_t)$.
- Обозначим действие, принятое в момент времени t как a_t (индекс бандита, цвет кнопки).
- Также считаем, что каждый y_t сгенерирован независимо из некоторого распределения наград своего бандита $f_{a_t}(y|\overrightarrow{\theta})$, где $\overrightarrow{\theta}$ — это некоторый вектор параметров.
- Награда – это мат. ожидание бандита. Знай мы распределение, легко бы выбрали и радовались, но мы не знаем.
- Для оценки параметров воспользуемся Бета-распределением.

Бета-распределение

- $f_X(x) = \frac{1}{B(\alpha,\beta)} x^{(\alpha-1)}(1-x)^{(\beta-1)}$ – плотность вероятности, $B(\alpha,\beta)$ – Бета-функция

Bayesian multi-armed bandits

Бета-распределение

- $f_X(x) = \frac{1}{B(\alpha,\beta)} x^{(\alpha-1)}(1-x)^{(\beta-1)}$ – плотность вероятности, $B(\alpha,\beta)$ – Бета-функция
- B-распределение является априорно-сопряженным к распределению Бернулли:

$$
\begin{aligned}
p(\theta \mid y) &\propto p(\theta) \cdot p(y \mid \theta) \\
&\propto \frac{1}{\mathrm{B}(\alpha,\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \cdot \theta^{y} (1-\theta)^{1-y} \\
&\propto \theta^{\alpha-1+y} (1-\theta)^{\beta-1+1-y}
\end{aligned}
$$

Bayesian multi-armed bandits

Бета-распределение

- $f_X(x) = \frac{1}{B(\alpha,\beta)} x^{(\alpha-1)}(1-x)^{(\beta-1)}$ – плотность вероятности, $B(\alpha,\beta)$ – Бета-функция
- B-распределение является априорно-сопряженным к распределению Бернулли:

$$
\begin{aligned}
p(\theta \mid y) &\propto p(\theta) \cdot p(y \mid \theta) \\
&\propto \frac{1}{\mathrm{B}(\alpha,\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} \cdot \theta^{y}(1-\theta)^{1-y} \\
&\propto \theta^{\alpha-1+y}(1-\theta)^{\beta-1+1-y}
\end{aligned}
$$

- при $\alpha = \beta = 1$ бета-распределение принимает форму равномерного распределения

Bayesian multi-armed bandits

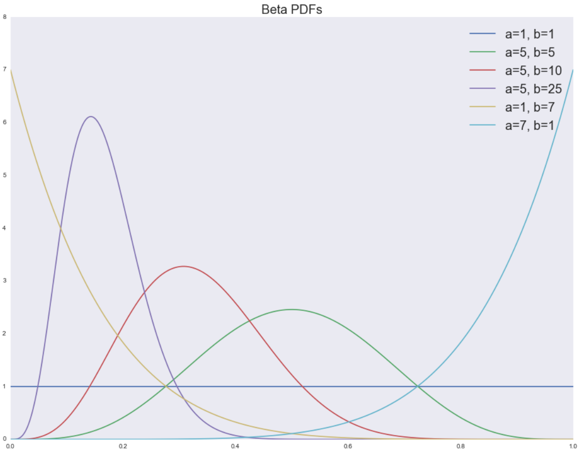
Бета-распределение

- $f_X(x) = \frac{1}{B(\alpha,\beta)} x^{(\alpha-1)}(1-x)^{(\beta-1)}$ – плотность вероятности, $B(\alpha,\beta)$ – Бета-функция
- B-распределение является априорно-сопряженным к распределению Бернулли:

$$\begin{aligned} p(\theta \mid y) &\propto p(\theta) \cdot p(y \mid \theta) \\ &\propto \frac{1}{\mathrm{B}(\alpha,\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \cdot \theta^{y} (1-\theta)^{1-y} \\ &\propto \theta^{\alpha-1+y} (1-\theta)^{\beta-1+1-y} \end{aligned}$$

- при $\alpha = \beta = 1$ бета-распределение принимает форму равномерного распределения
- модель легко интерпретируема

Bayesian multi-armed bandits



Наша модель

$$\begin{aligned} \theta_i &\sim \text{Beta}(\alpha_i, \beta_i) \\ y_i &\sim \text{Bernoulli}(\theta_i) \end{aligned}$$

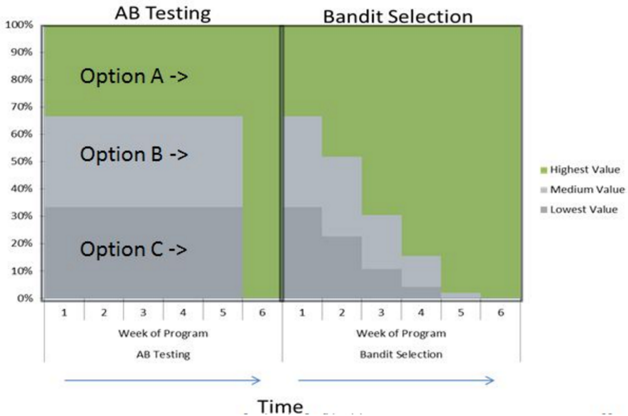
Pray for our bayesian lord!

Алгоритм

- Для всех бандитов введем два параметра бета-распределения и приравняем их к единице $\forall i, \alpha_i = \beta_i = 1$;
- повторяем в течении некоторого времени $t = 1, 2, \dots$
  - для каждого бандита семплируем $\theta_i \backsim Beta(\alpha_i, \beta_i)$;
  - выбираем бандита с максимальной наградой $k = argmax_i \theta_i$
  - используем k-ого бандита в текущем эксперименте и получаем награду $y \in \{0, 1\{$ (показываем ту кнопку текущему пользователю, которая по текущему семплу максимизирует награду);
  - обновляем параметры соответствующего априорного распределения (легко модифицируется для batch mode, если мы проводим не один, а серию экспериментов):
  - $\alpha_i = \alpha_i + y$
  - $\beta_i = \beta_i + 1 - y$

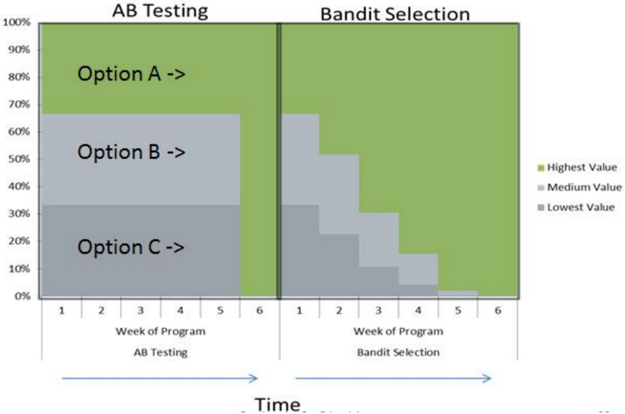
Особенности

- модель динамически настраивается, перегоняя траффик на модель



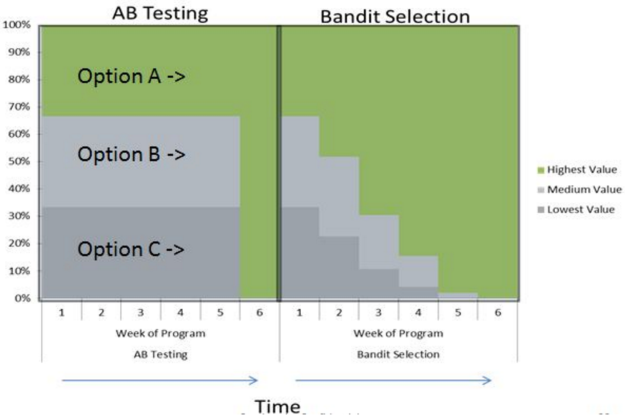
Особенности

- модель динамически настраивается, перегоняя траффик на модель
- в любой момент времени у нас есть параметры распределения вероятностей выигрыша



Особенности

- модель динамически настраивается, перегоняя траффик на модель
- в любой момент времени у нас есть параметры распределения вероятностей выигрыша
- мы, как бизнес, хотим максимизировать CTR

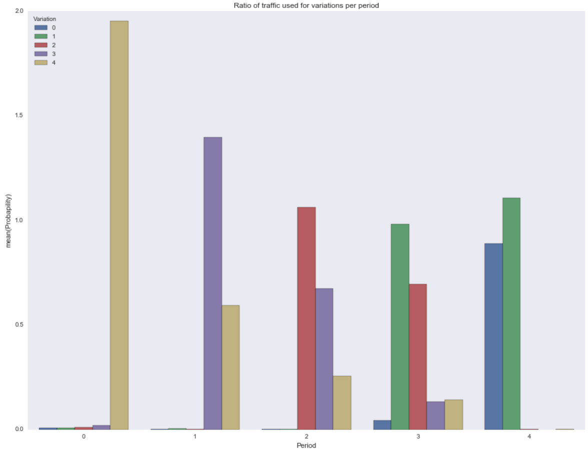


Эксперимент

- 5 поколений, на каждом меняются предпочтения
- чем "уже"распределение pdf – тем больше мы в нём уверены
- каждое новое поколение мы используем предыдущие оценки, поэтому частенько ошибаемся

Ещё особенности:

- (exploration vs exploitaion trade-off) (и осбенности бандитов)
- можно динамически добавлять элементы
- можно добавлять элементы с априорным знанием (завышать α)



Ещё эксперимент

Link

- По материалам open data science
  https://habrahabr.ru/company/ods/blog/325416/