縮小推定のはなし

@utaka233

Tokyo.R #76, 03/02/2019

Table of Contents

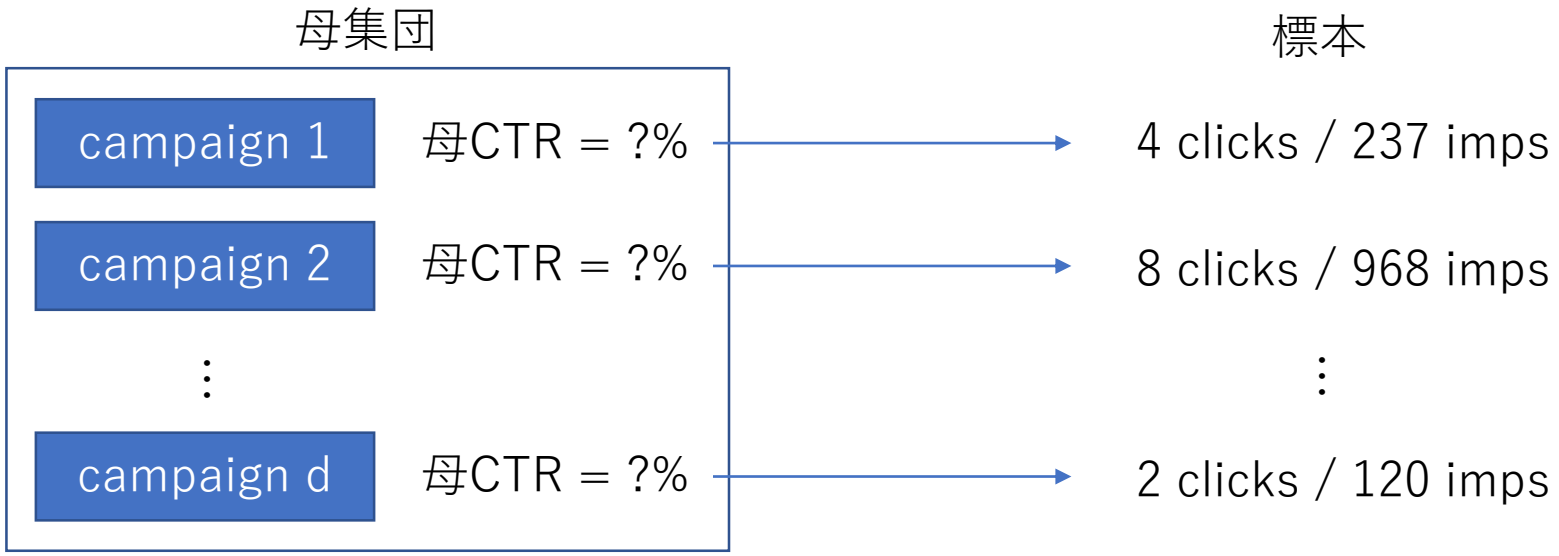
- 1. motivation
- 2. 縮小推定とは
- 3. 縮小推定の可能性

1. motivation

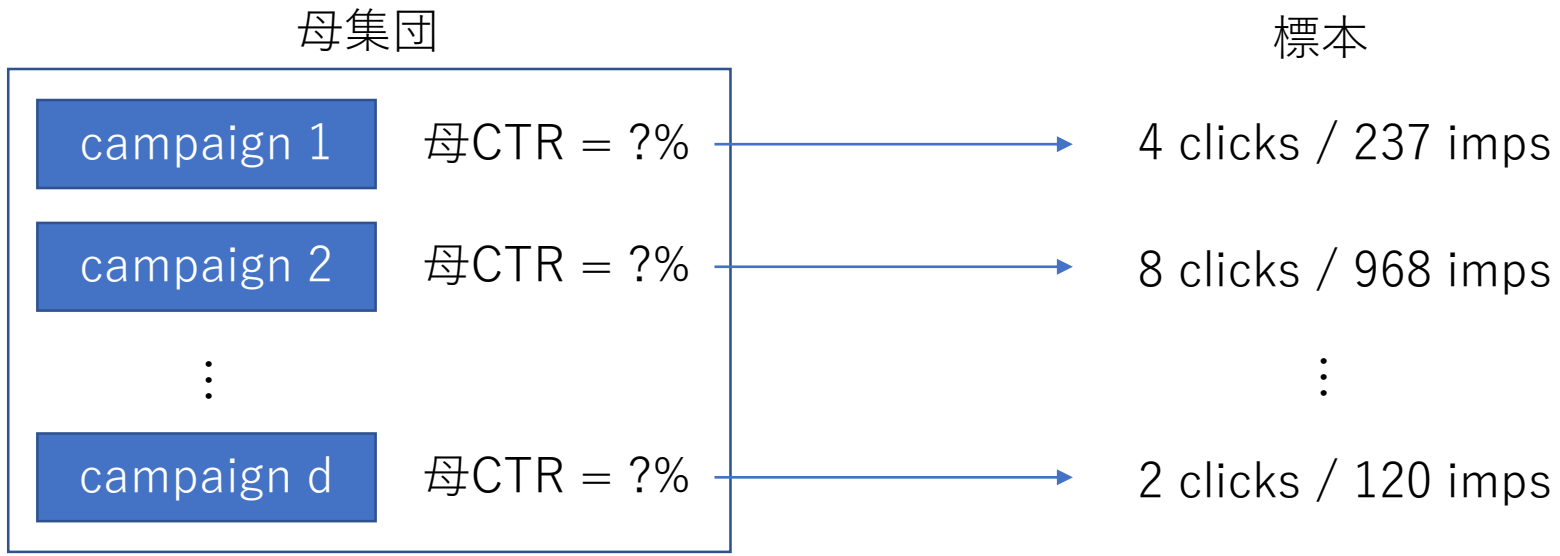
打者の生涯打率推定を例に

今回考える問題

- こんな問題を考えたい。
  - web広告：各キャンペーンのCTRの推定
  - セイバーメトリクス：各打者の打率の推定
  - 社会科学：各県の1世帯あたりの平均教育費の推定



定式化：多母集団の推定



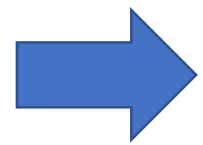
- campaignの母CTRを推定するにはどうすればよいか？
  - 直感：標本CTR = click数 / imp数で推定
  - 理屈：標本CTRは有効性と一致性を持つ。
    - 二項分布の母比率に対する最尤推定量
    - 有効推定量（∵不偏かつ最尤 ⇒ 有効）

例：打者の生涯打率推定

- 打者の生涯打率推定
  - 対象：通算で500打席以上に立った打者
  - デビューした年度の打率を用いて生涯打率を推定する。
- library(Lahman)のBattingデータセットを用いる。

```
> as_tibble(Batting)
# A tibble: 102,816 x 22
   playerID yearID stint teamID lgID      G    AB     R     H   X2B
   <chr>     <int> <int> <fct>  <fct> <int> <int> <int> <int> <int>
 1 abercda~   1871     1 TRO    NA        1     4     0     0     0
 2 addybo01   1871     1 RC1    NA       25   118    30    32     6
 3 allisar~   1871     1 CL1    NA       29   137    28    40     4
 4 allisdo~   1871     1 WS3    NA       27   133    28    44    10
 5 ansonca~   1871     1 RC1    NA       25   120    29    39    11
 6 armstbo~   1871     1 FW1    NA       12    49     9    11     2
 7 barkeal~   1871     1 RC1    NA        1     4     0     1     0
 8 barnero~   1871     1 BS1    NA       31   157    66    63    10
 9 barrebi~   1871     1 FW1    NA        1     5     1     1     1
10 barrofr~   1871     1 BS1    NA       18    86    13    13     2
# ... with 102,806 more rows, and 12 more variables: X3B <int>,
#   HR <int>, RBI <int>, SB <int>, CS <int>, BB <int>, SO <int>,
#   IBB <int>, HBP <int>, SH <int>, SF <int>, GIDP <int>
```

標本抽出

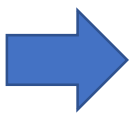


```
> data
# A tibble: 25 x 3
   playerID    hit    AB
   <chr>     <int> <int>
 1 allisbo01   149   570
 2 blancjo01   115   404
 3 brownto02    27   138
 4 canoro01    179   624
 5 cervefr01    86   326
 6 curtrgu01   142   488
 7 donaljo01    80   363
 8 frymatr01   321  1183
 9 hairsje01    61   225
10 harrebu01   137   564
# ... with 15 more rows
```

例：推定量の比較

- ２つの推定量を比較してみよう。
  - MLE：手元のデータから計算できる打率（標本比率）
  - mystery：何者？？

```
> data
# A tibble: 25 x 3
   playerID    hit    AB
   <chr>     <int> <int>
 1 allisbo01   149   570
 2 blancjo01   115   404
 3 brownto02    27   138
 4 canoro01    179   624
 5 cervefr01    86   326
 6 curtrgu01   142   488
 7 donaljo01    80   363
 8 frymatr01   321  1183
 9 hairsje01    61   225
10 harrebu01   137   564
# ... with 15 more rows
```



```
# A tibble: 25 x 4
   playerID    MLE mystery truth
   <chr>     <dbl>   <dbl> <dbl>
 1 allisbo01 0.261   0.263 0.255
 2 blancjo01 0.285   0.275 0.239
 3 brownto02 0.196   0.229 0.241
 4 canoro01  0.287   0.277 0.307
 5 cervefr01 0.264   0.265 0.280
 6 curtrgu01 0.291   0.279 0.276
 7 donaljo01 0.220   0.242 0.238
 8 frymatr01 0.271   0.268 0.274
 9 hairsje01 0.271   0.268 0.258
10 harrebu01 0.243   0.254 0.236
# ... with 15 more rows
```

例：平均2乗誤差による評価

- MSE（平均2乗誤差）の比較

```
# A tibble: 1 x 3
   mse_MLE mse_mystery efficiency
     <dbl>       <dbl>      <dbl>
1 0.000550    0.000326      0.593
```

  - MSEとは：

$$\mathrm{MSE}(\theta, \hat{\theta}) = \mathbb{E}[(\theta - \hat{\theta})^2]$$

  - どうやらmysteryはMLE（標本比率）より**良い推定量**らしい。
    - efficiency = mysteryのMSE / MLEのMSE
    - MLEよりmysteryのほうが、全体的にはground truthに近い値をとっている。

例：たまたまでは？

- もう一度やってみる。たまたまでは？

```
# A tibble: 1 x 3
   mse_MLE mse_mystery efficiency
     <dbl>       <dbl>      <dbl>
1 0.000561    0.000320      0.571
```

```
# A tibble: 1 x 3
   mse_MLE mse_mystery efficiency
     <dbl>       <dbl>      <dbl>
1 0.000195    0.000151      0.775
```

```
# A tibble: 1 x 3
   mse_MLE mse_mystery efficiency
     <dbl>       <dbl>      <dbl>
1 0.000321    0.000237      0.738
```

```
# A tibble: 1 x 3
   mse_MLE mse_mystery efficiency
     <dbl>       <dbl>      <dbl>
1 0.000694    0.000460      0.662
```

単なる偶然ではなさそう…？

mysteryは何者？

- mysteryの正体

$$\hat{p}^{JS} = \bar{p} + \left[1 - \frac{(n-3)\hat{\sigma^2}}{\sum(p_i - \bar{p})^2}\right](p_i - \bar{p})$$

  - 平均方向に縮小する推定量
    - 他の打者の情報をつかって推定効率を良くする。そんなことができるのか？
    - James-Stein型推定量という。

試してみてください。

- GitHubにスクリプトを貼っておいたので、試してみてください。
  - URL : https://github.com/utaka233/tokyor76/tree/master
  - stein.R : 例に掲げた計算を行うためのスクリプト

2. 縮小推定とは

原点や平均方向への縮小がもたらす平均2乗誤差の効率性

良く用いられる推定量の良さとは

- 不偏性と標準誤差
  - MSEのバイアス・バリアンス分解

$$\begin{aligned}\mathrm{MSE}(\theta, \hat{\theta}) &= \mathbb{E}[(\theta - \hat{\theta})^2] \\ &= (\mathbb{E}[\hat{\theta}] - \theta)^2 + \mathbb{V}[\hat{\theta}]\end{aligned}$$

    - 第1項：バイアス, 第2項：推定量の標準誤差
    - 不偏推定量 = バイアスのない推定量
      - 平均2乗誤差が最小の推定量を見つけるのは困難。不偏推定量はそこまででもない。
      - 標準誤差が最小の不偏推定量を求めればよい。→ 一様最小分散不偏推定量
        - Cramer-Rao下限（達成できる場合、有効性を持つという。）
      - 例：母平均に対する標本平均, 母分散に対する不偏分散, …

平均2乗誤差最小推定量

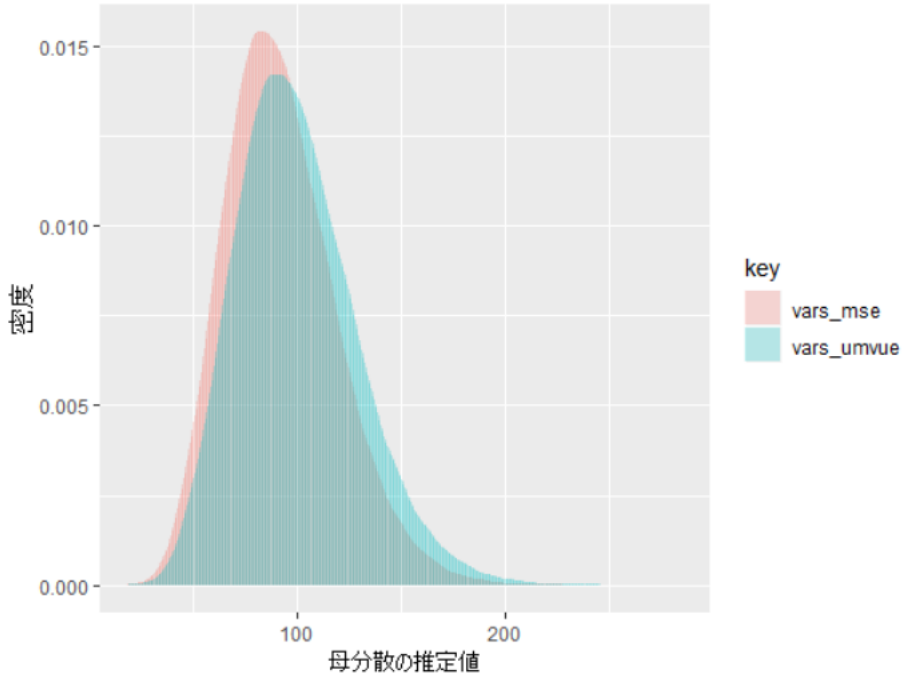
- 平均2乗誤差最小推定量 ≠ 一様最小分散不偏推定量
  - 代表例：正規分布の母分散の推定

$$\hat{\sigma^2}^{MSE} = \frac{1}{n+1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

  - バイアスを許してしまう。
  - その代わりに標準誤差を小さくする。

```
# A tibble: 6 x 2
  key                value
  <chr>              <dbl>
1 bias_umvue       -0.0795
2 bias_mse          7.62
3 std_error_umvue  28.9
4 std_error_mse    26.7
5 mse_umvue       834.
6 mse_mse         769.
```

N(0, 100)から25個の標本をとる。



2つの推定量の比較

- 一様最小分散不偏推定量
  - 各推定時に期待される値は真のパラメータの値そのもの。
  - 推定ごとに得られる値はやや不安定。

- 平均2乗誤差最小推定量
  - 各推定時に期待される値は真のパラメータより少しズレている。
  - 推定ごとに得られる値は安定。
    - 要するに、真のパラメータより少しズレた値ではあろうけれど、言うて近い値を安定して得ることが出来る。

Stein現象

- 問題設定
  - 3群以上の正規母集団を考えてください。
    - 母平均は未知とします。
    - 母分散は既知、すべての群で等しいとしてよいことにします。
  - 各群からサイズ1の標本をひとつずつ抽出しましょう。
  - 各群の母平均を推定してください。

直感的には、各群の標本の値そのもので推定するしかない。
しかし、もっと良い推定量がある。

$$\hat{\mu}^{JS} = \left(1 - \frac{(d-2)\sigma^2}{\sum_{i=1}^{d} x_i^2}\right) x$$

James-Stein推定量, Stein (1956)

James-Stein推定量

- James-Stein推定量
  - 原点への縮小
    - 標本の値をそのまま推定に使うより、少し0に近づけた値を使っている。
  - 不偏推定量ではない。要するにbiasを許している。
  - その代わり、平均2乗誤差は一様最小分散不偏推定量より小さい。
    - 要するに標準誤差が小さい。

$$\hat{\mu}^{JS} = \left(1 - \frac{(d-2)\sigma^2}{\sum_{i=1}^{d} x_i^2}\right) x$$

なぜ他の群の情報が役立つ？

- 経験ベイズ推定量による解釈
  - 実はJames-Stein推定量は、経験ベイズ推定量と一致している。
  - 以下、母分散を1として証明のoutlineを説明します。
    - 母平均パラメータの事前分布を正規分布とします。
      - 期待値を0, 分散をAとしましょう。

$$\mu \sim N(0, A)$$

      - 分散Aはmoment法で推定してしまう。（経験ベイズ）

$$\mathbb{E}\left[\frac{d-2}{\sum_{i=1}^{d} x_i^2}\right] = \frac{1}{A+1}$$

    - ベイズ更新により以下の事後分布を得る。あとはEAPを考えればよい。

$$\mu \mid x \sim N\left(\frac{A}{A+1}\mu, \frac{A}{A+1}E\right)$$

平均への縮小

- 平均への縮小
  - 群が4以上の場合には、全体平均へ縮小する推定量がある。

$$\hat{\mu}^{JS} = \bar{x} + \left[1 - \frac{(n-3)\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right](x - \bar{x})$$

$$\hat{p}^{JS} = \bar{p} + \left[1 - \frac{(n-3)\hat{\sigma^2}}{\sum(p_i - \bar{p})^2}\right](p_i - \bar{p})$$

二項分布の正規近似

注：母比率の場合の経験ベイズ推定量

- 母比率の(経験)ベイズ推定
  - beta-二項モデル：事前分布はbeta分布、母集団モデルは二項分布。
- library(ebbr)
  - beta-二項モデルの経験ベイズ推定を行うパッケージ

```
> data %>% add_ebb_estimate(x = hit, n = AB)
# A tibble: 23 x 9
   playerID    hit    AB .alpha1 .beta1 .fitted  .raw  .low .high
   <chr>     <int> <int>   <dbl>  <dbl>   <dbl> <dbl> <dbl> <dbl>
 1 amarial01    87   368    698.  1915.   0.267 0.236 0.250 0.284
 2 barfije01   156   539    767.  2017.   0.275 0.289 0.259 0.292
 3 beltrca01   194   663    805.  2103.   0.277 0.293 0.261 0.293
 4 brownje01   137   513    748.  2010.   0.271 0.267 0.255 0.288
 5 brubabi01   160   554    771.  2028.   0.275 0.289 0.259 0.292
 6 goodmiv01   139   489    750.  1984.   0.274 0.284 0.258 0.291
 7 harpeto01   166   646    777.  2114.   0.269 0.257 0.253 0.285
 8 herrejo03    68   281    679.  1847.   0.269 0.242 0.252 0.286
 9 johnsro02    72   260    683.  1822.   0.273 0.277 0.255 0.290
10 jordari02   149   523    760.  2008.   0.274 0.285 0.258 0.291
# ... with 13 more rows
```

```
# A tibble: 3 x 3
  estimator      MSE efficiency
  <chr>        <dbl>      <dbl>
1 mse_MLE   0.000368      1
2 mse_stein 0.000196      0.532
3 mse_ebbr  0.000287      0.780
```

3. 縮小推定の可能性

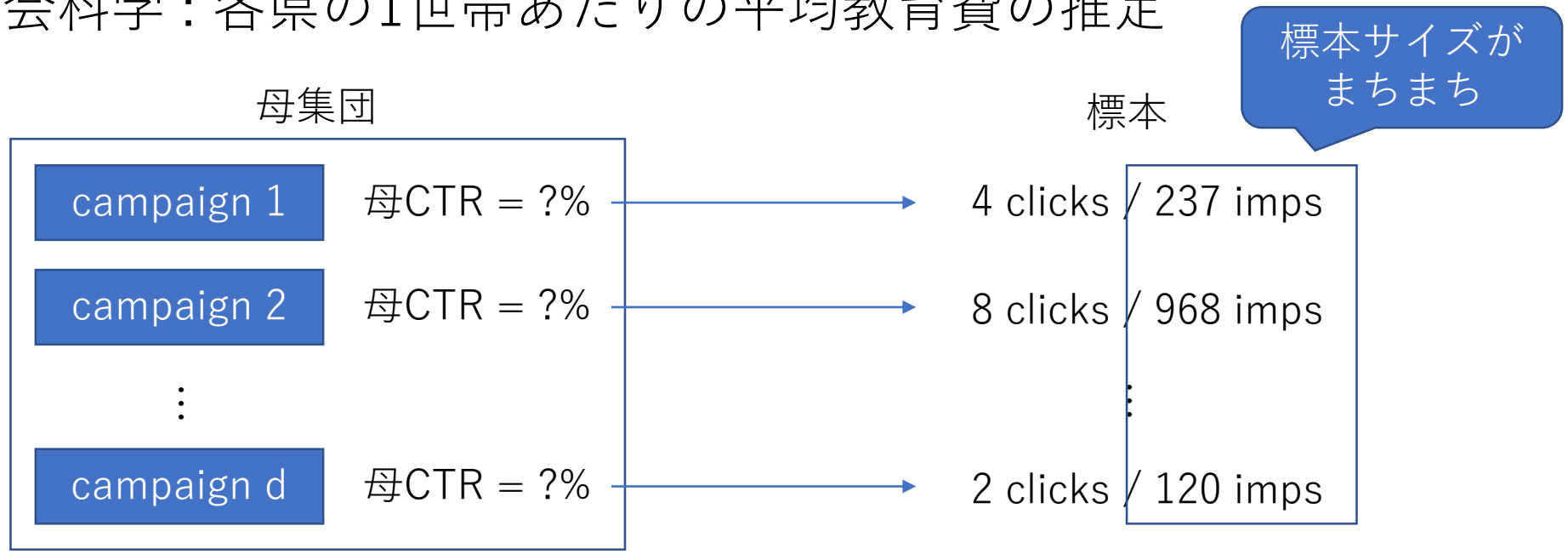
縮小推定が活躍する場面とは

縮小推定のプライオリティ

- 多母集団における標準誤差の改善
  - ドメイン知識が存在する場合
    - 広告のCTRは基本的に0に近い値を取るなど。
    - 原点や平均値など任意の値に対して推定量を縮小できる。
  - 小地域推定
    - 各母集団ごとに推定すると、各群で標本サイズが違う場合と標本サイズが小さい群のほうが大きい群より標準誤差が高くなってしまう。

最初に考えた問題

- 多母集団の推定問題（特に小地域推定）
  - web広告：各キャンペーンのCTRの推定
  - セイバーメトリクス：各打者の打率の推定
  - 社会科学：各県の1世帯あたりの平均教育費の推定



4. おわりに

自己紹介とか…。

自己紹介

- お仕事
  - 2014-現在：株式会社すうがくぶんか（現在：教務部 部長）
  - 2015-現在：株式会社オモロワークス データサイエンティスト
  - 2018-現在：株式会社スカイディスク 技術顧問
- 経歴
  - 2015年：修士（理学, 早稲田大学）代数幾何学専攻
  - 2015年：統計検定1級, 人文科学優秀者A

すうがく＋ぶんか
もっと社会に数学を

We Are Hiring !

マーベリックでは機械学習エンジニアを募集しています。

機械学習を活用し、広告配信システムの
機能開発を行いませんか？

実務経験のある方、実務未経験だけど意欲のある方、
ぜひお声がけください！

MAVERICK