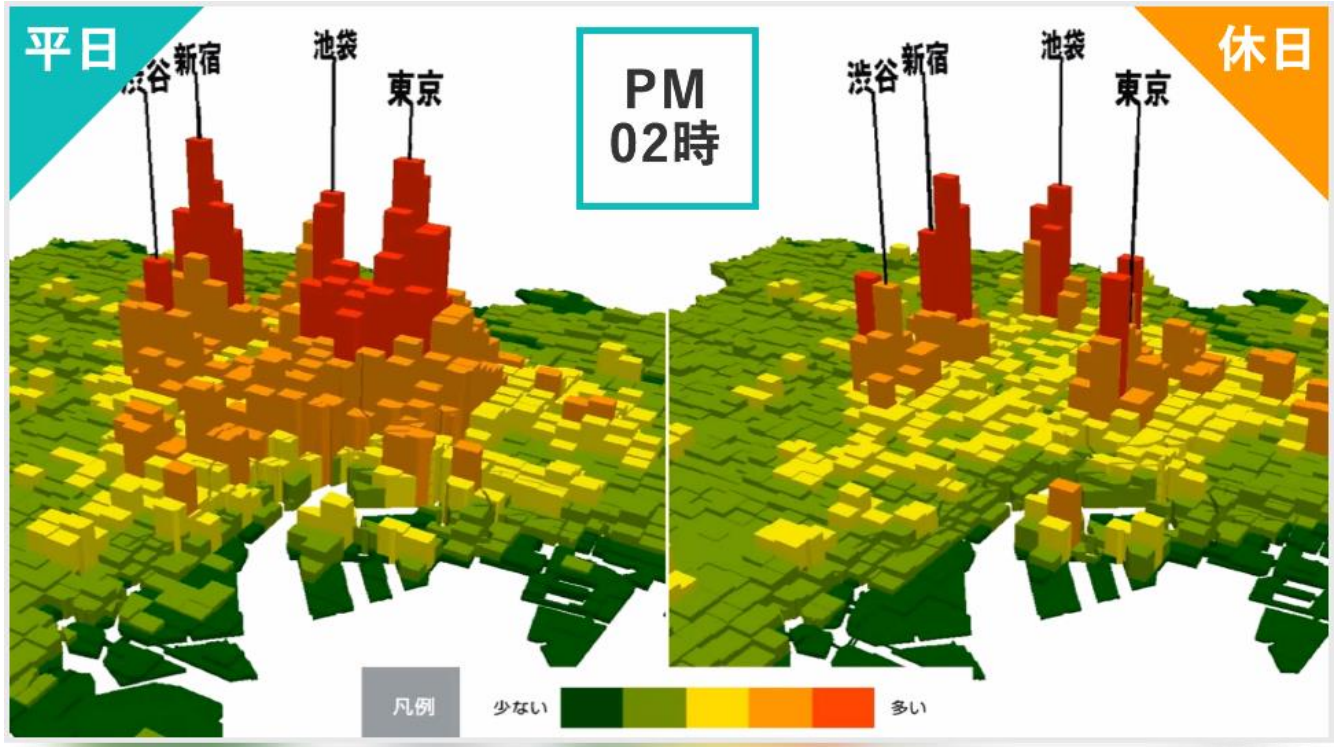
Local Differential Privacyによりプライバシーを考慮した位置情報分布推定

明治大学　総合数理学部
堀込光　菊池浩明

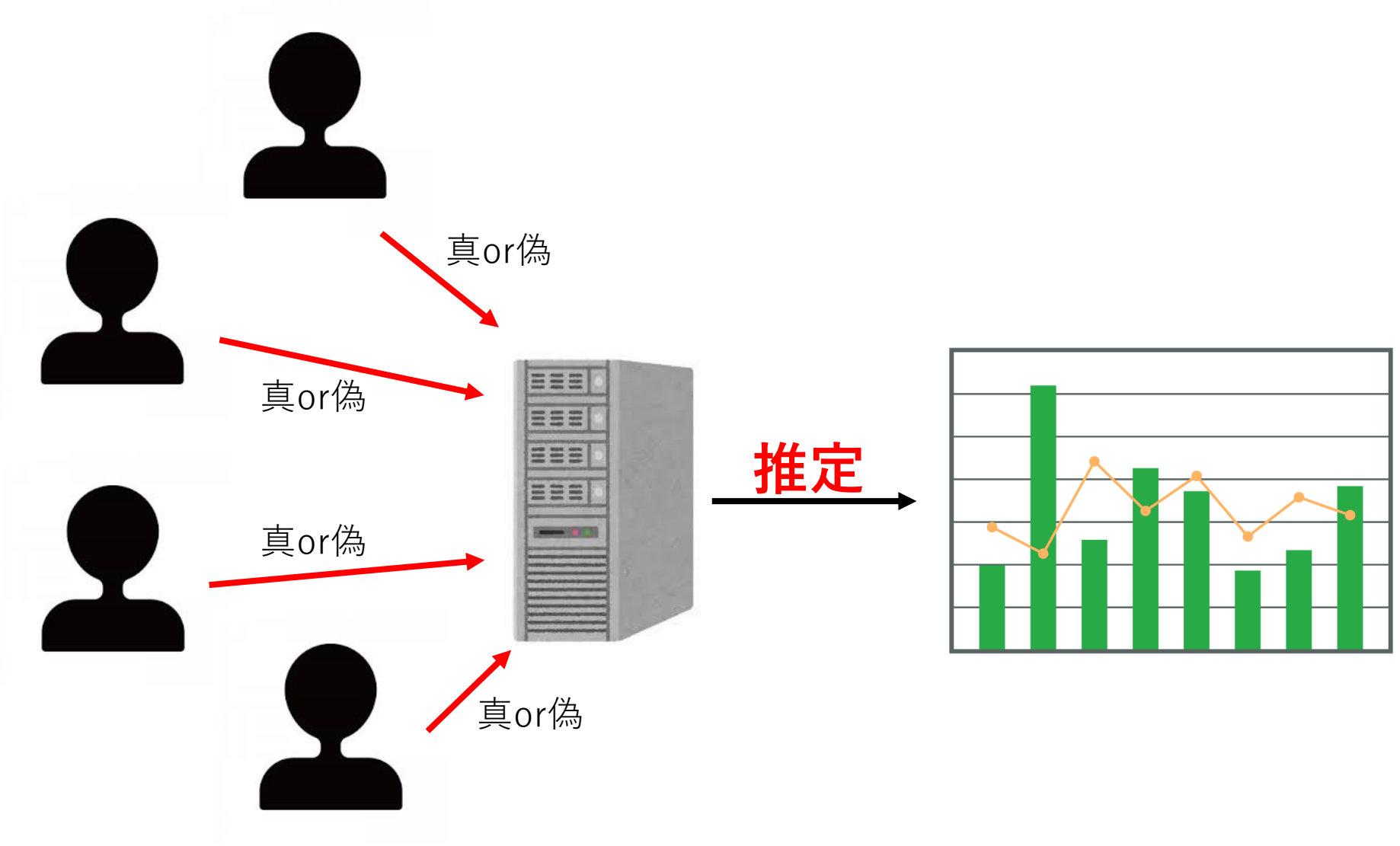
研究背景

- モバイル空間統計(NTTドコモ)

https://mobaku.jp/

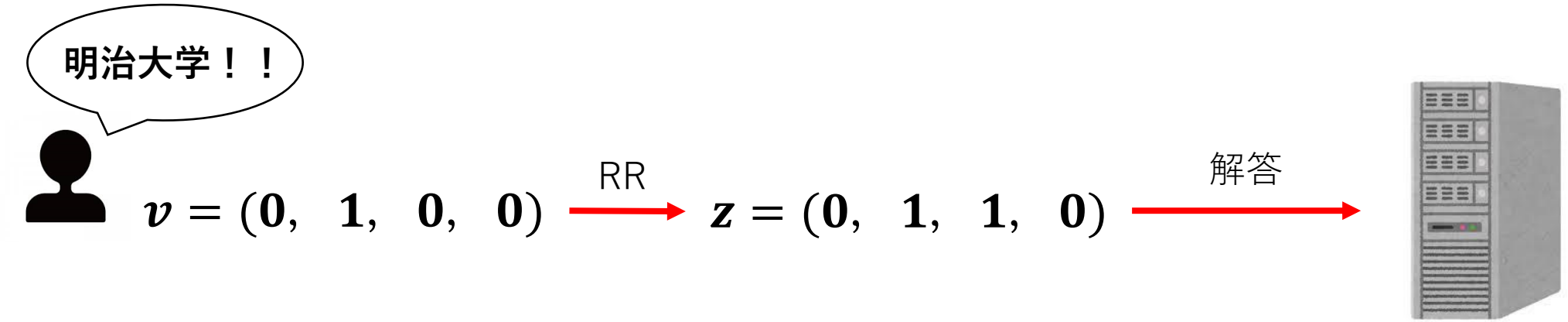


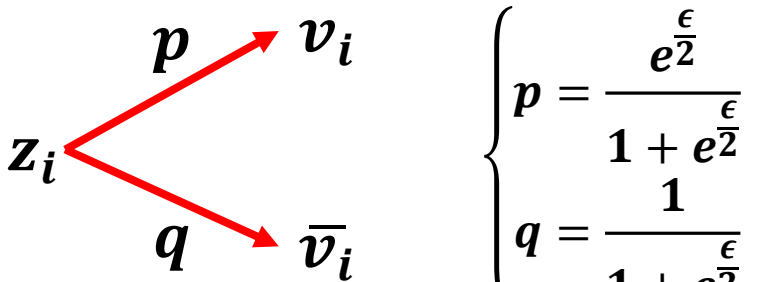
Local Differential Privacy



RAPPOR

Q.どこの大学の学生ですか？(東京大学，明治大学，立教大学，東洋大学)



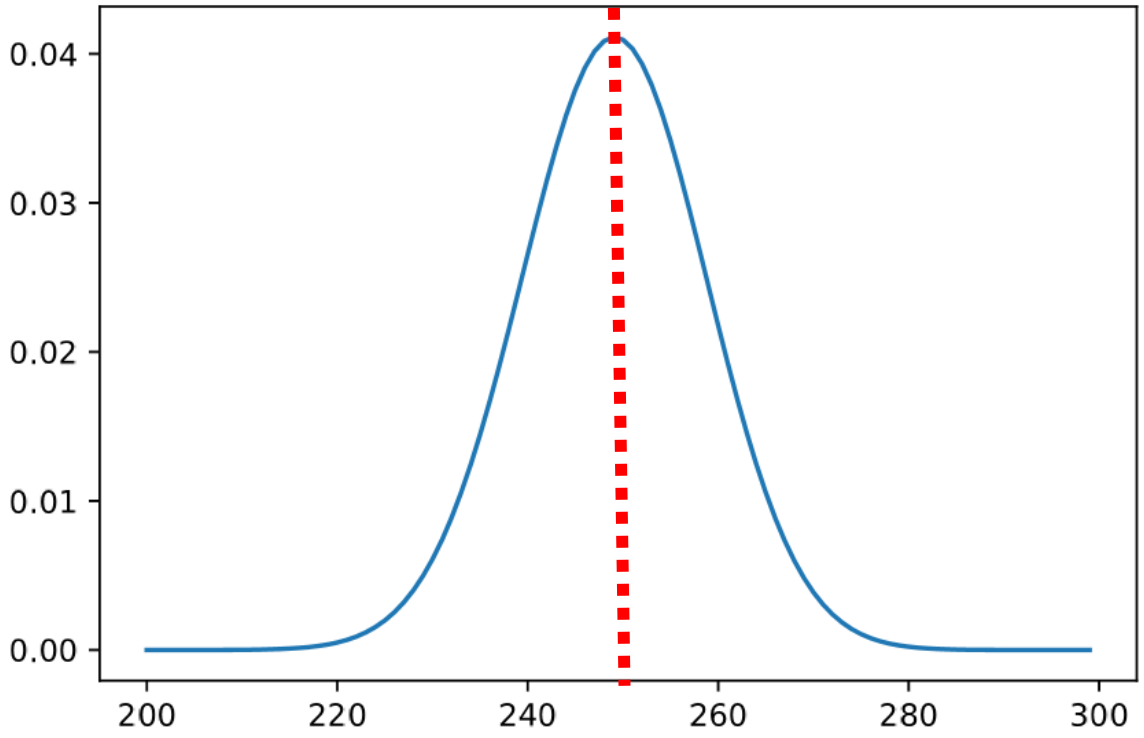


$$\begin{cases} p = \dfrac{e^{\frac{\epsilon}{2}}}{1 + e^{\frac{\epsilon}{2}}} \\ q = \dfrac{1}{1 + e^{\frac{\epsilon}{2}}} \end{cases}$$

[4]Ulfar Erlingsson, Vasyl Pihur,Aleksandra Korolova,“RAPPOR: Randomized Aggregatable PrivacyPreserving Ordinal Response”,ACM Conference on Computer and Communications Security, pp. 1054-1067, 2014.

従来方式

・最尤推定法

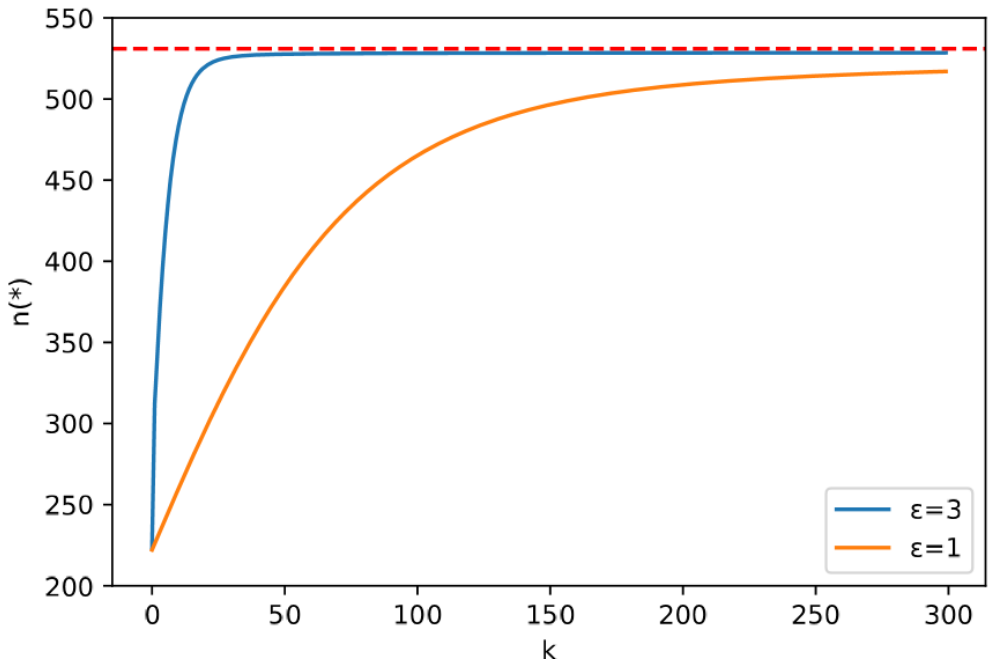


提案方式

・EMアルゴリズム

1，E-step　：　事象に対する条件付き確率の期待値を計算

2，M-step　：　その期待値を最大にするパラメータを事後確率として更新



実験概要

1. 目的

・RAPPORにおける推定精度の向上

2. 方法

・東京23区内の疑似人流データを使用

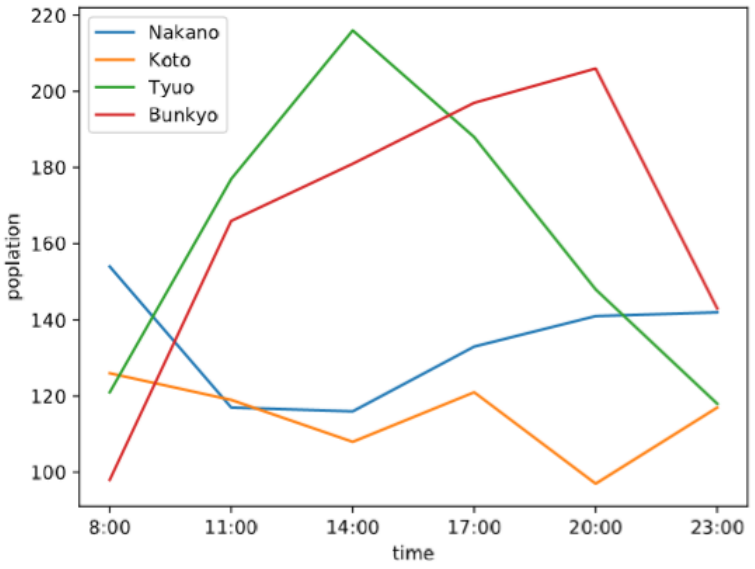
・従来の推定方式である最尤推定と提案方式であるEMアルゴリズムの各々で推定を行い誤差を測定

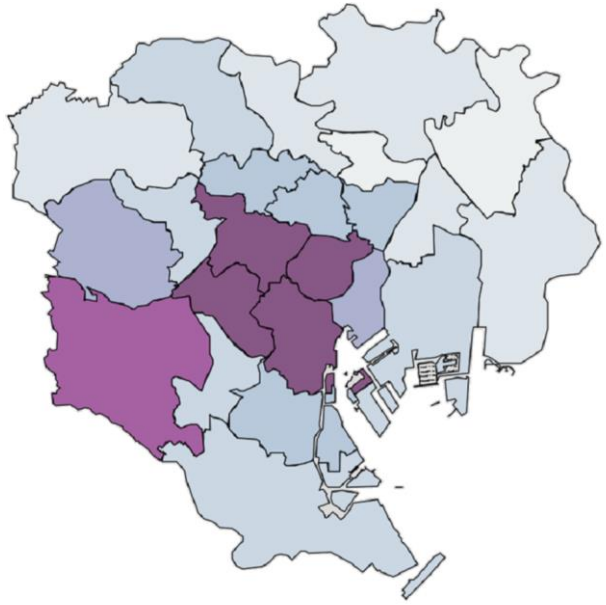
収集データ

(35.67159, 139.7078) Google Map API → “渋谷区”

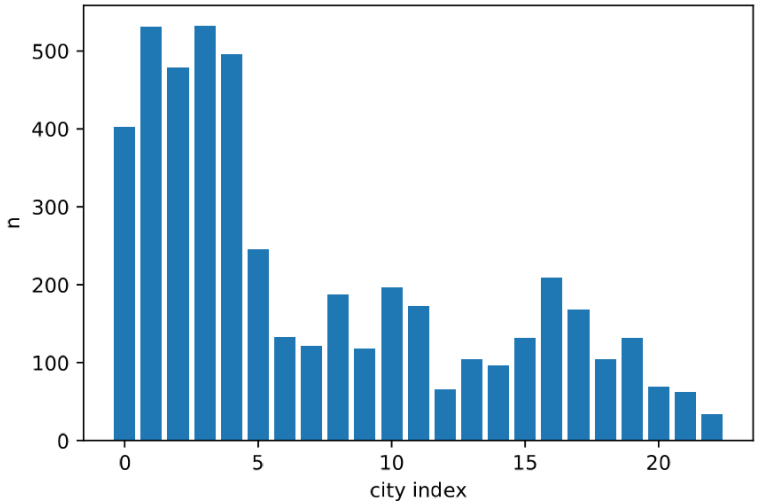
$$\boldsymbol{v} = \left(\boldsymbol{n}_{\text{世田谷区}}, \boldsymbol{n}_{\text{中野区}}, \boldsymbol{n}_{\text{渋谷区}}, \cdots, \boldsymbol{n}_{\text{葛飾区}}, \boldsymbol{n}_{\text{江戸川区}}\right)$$

$$= (0, 0, 1, \cdots, 0, 0)$$

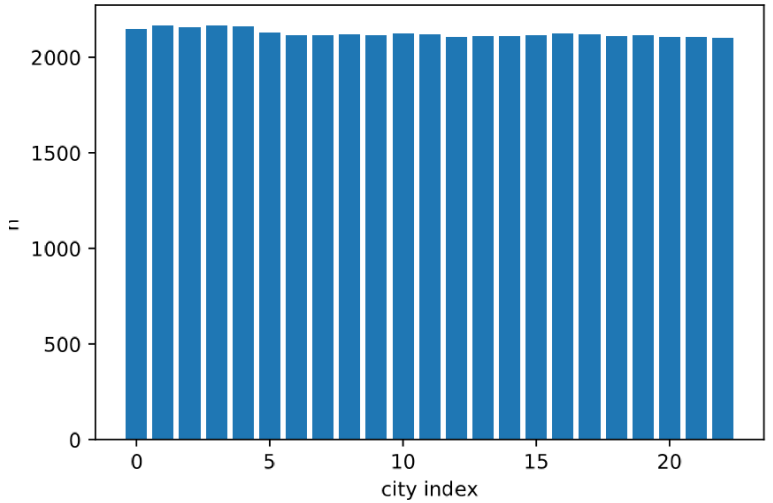




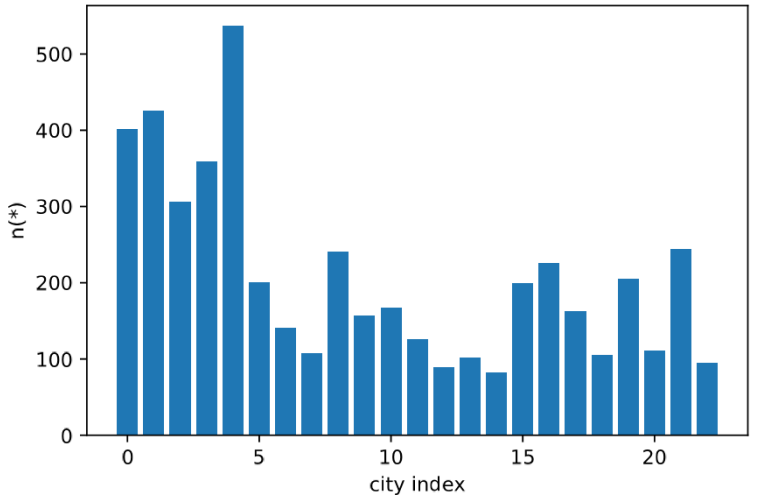
実験結果(1)



真の人口n_i



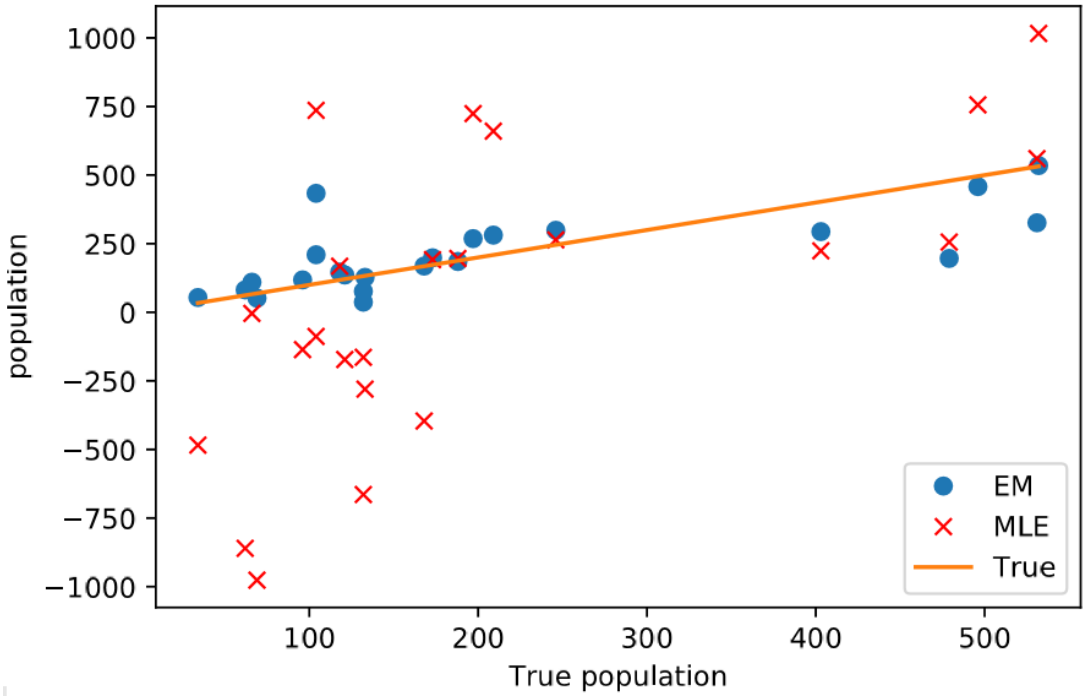
RAPPORにおける人口n'_i



EMあるごりズムによる推定人口$n^{(*)}_i$

実験結果：推定人口

推定人口($\epsilon = 0.1$)



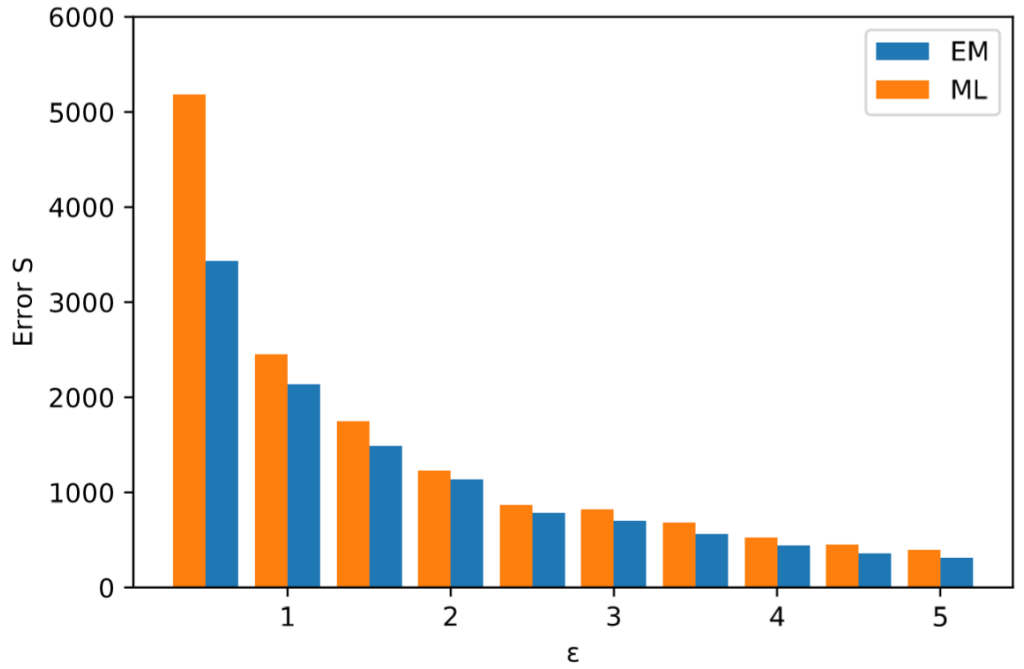
真の人口に対する推定人口(ϵ =0.1)

平均絶対誤差(MAE)

$\hat{n}_i$：最尤推定における推定人口

$n_i^{(*)}$：EMアルゴリズムにおける推定人口

$$MAE = \sum_{i=1}^{23} |n_i - \hat{n}_i|$$



まとめ

- 8:00から23:00までの3時間おきにϵを0.5づつ変化させてMAEを求めた.

$$6(\text{時間}) \times 10(\epsilon) \times 10(\text{回数}) = 600(\text{回数})$$

600回中555回=92.5%でEMアルゴリズムの方がMAEが小さかった.

また，$\epsilon = 0.5$の際，37.7%から98.1%の精度の向上がみられた.