都市形成史研究の基礎と応用

-第二回-

2019年度スタートアップゼミ

2019/06/20

小林 里瑳 (M2)

図：1866年（慶應二年）頃の道後温泉本館周辺　道後温泉本館調査委員会編『道後温泉本館の歴史』, 松山市, (1994)

今日の内容

- **都市形成史研究の基礎②**
  - 史料の種類（復習）
  - 史料批判の方法と概要

今日のメイン

- **EMアルゴリズムの導入**
  - 隠れ変数を含むロジットモデルのパラメータ推定方法
  - EMアルゴリズムの定式化
  - 実際の計算結果

モデルのパラメータ推定の一例

今日の目標

① 史料批判の重要性を知る

② EMアルゴリズムの導入方法を知る


THE FULL VIEW OF YUNOMACHI OF
DOGO FROM THE AIR, DOGO.
景全町の湯後道ろたし瞰俯りよ上機（勝名後道）

前回の復習

復習

具体的にどんな史料があるのか

都市形成史研究で
主に使う史料

沈黙史料

建築、土木建造物そのもの（もしくは遺構）

発言史料

無形

伝承史料

風俗、習慣、行事、地名、字名、音声など

有形

グラフィック史料

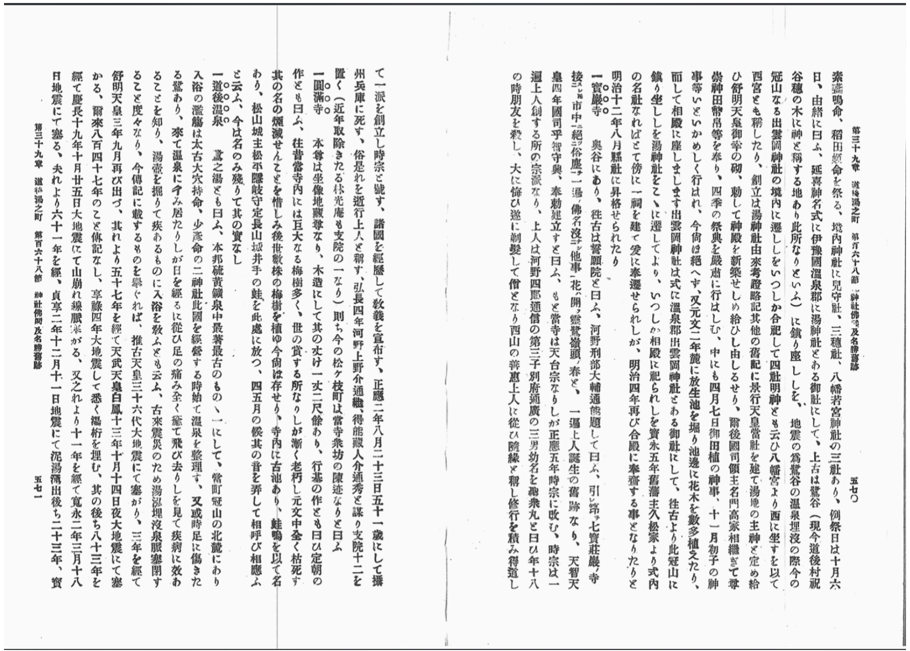
地図、絵図、絵画、図面、写真、映像など

キャラクタ史料

文書、記録、文学作品、紀行、日記、帳簿、議事録、新聞など

建築史家である玉井哲雄の史料分類を参考に作成

復習



第三十九章　道後湯之町　第百六十八節　神社佛閣及名勝舊跡　五七〇

素盞嗚命、稻田姫命を祭る、境内神社に兒守社、三穗社、八幡若宮神社の三社あり、例祭日は十月六日、由緒に曰ふ、延喜神名式に伊豫國溫泉郡に湯神社とある御社にして、上古は鷺谷（現今道後村祝谷）穗の木に神と稱する地あり此所なりといふ）に鎮り座ししを、地震の爲鷺谷の溫泉埋沒の際今の冠山なる出雲岡神社の境内に遷ししをいつしか合祀して四社明神とも云ひ八幡宮より西に坐すを以て西宮とも稱したり、創立は湯神社由來考證略記其他の舊記に景行天皇當社を建て湯地の主神と定め給ひ舒明天皇御幸の砌、勅して神殿を新築せしめ給ひし由しるせり、爾後國司領主名門高家相繼ぎて尊崇神田幣帛等を奉り、四季の祭典を嚴肅に行はしむ、中にも四月七日御田植の神事、十一月初子の神事等いといかめしく行はれ、今尙は絶へず、又元文二年麓に放生池を掘り池邊に花木を數多植えたり、而して相殿に座しまします出雲岡神社は式に溫泉郡出雲岡神社とある御社にして、往古より此冠山に鎮り坐ししを湯神社をこゝに遷してより、いつしか相殿に祀られしを寶永五年舊藩主久松家より式内の名社なればとて傍に一祠を建て爰に奉遷せられしが、明治四年再び合殿に奉齋する事となりたりと

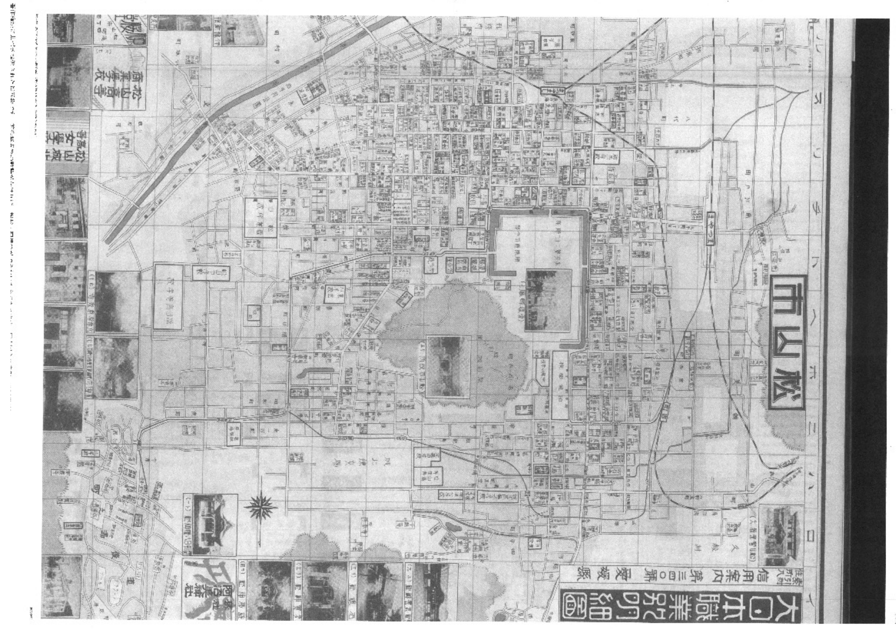
明治十二年八月縣社に昇格せられたり

一寶嚴寺　奥谷にあり、往古は誓願院と曰ふ、河野刑部大輔通能題して曰ふ、引籌七寶莊嚴寺接市中絶俗塵一遍佛名沒他事花開靈鷲嶺頭春と、一遍上人誕生の舊跡なり、天智天皇四年國司乎智守興、奉勅建立すと曰ふ、もと當寺は天台宗なりしが正應五年時宗に改む、時宗は一遍上人創する所の宗派なり、上人は河野四郎通信の第三子別府通廣の三男幼名を[illegible]丸と曰ひ年十八の時朋友を殺し、大に悔ひ遂に剃髮して僧となり西山の善惠上人に從ひ隨縁と稱し修行を積み得道して一派を創立し時宗と號す、諸國を經歷して教義を宣布す、正應二年八月二十三日五十一歲にして攝州兵庫に死す、俗是れを遊行上人と稱す、弘長四年河野上野介通繼、得能藏人介通秀と謀り支院十二を置く（近年取除きたる林光庵も支院の一なり）則ち今の松ヶ枝町は當寺衆坊の陳迹なりと曰ふ

一圓滿寺　本尊は坐像地藏尊なり、木造にして其の丈け一丈二尺餘あり、行基の作とも曰ひ定朝の作とも曰ふ、往昔當寺内には巨大なる梅樹多く、世の賞する所なりしが漸く老朽し元文中全く枯死す其の名の煙滅せんことを惜しみ後世數株の梅樹を植ゆ今尙は存せり、寺内に古池あり、蛙鳴を以て名あり、松山城主松平隱岐守定長山城井手の蛙を此處に放つ、四五月の候其の音を弄して相呼び相應ふと云ふ、今は名のみ殘りて其の實なし

一道後溫泉　鷺之湯とも曰ふ、本邦硫黄鑛泉中最著最古のもの、一にして、常町冠山の北麓にあり入浴の濫觴は太古大穴持命、少彥命の二神社此國を經營する時始て溫泉を整理す、又或時足に傷きたる鷺あり、來て溫泉に浴み居たりしが日を經るに從ひ足の痛み全く癒て飛び去りしを見て疾病に效あることを知り、湯壺を掘りて疾あるものに入浴を教ふとも云ふ、古來震災のため湯泓埋沒泉脈塞閉すること度々なり、今傳記に載するものを擧ぐれば、推古天皇三十六代大地震にて塞がり、三年を經て舒明天皇三年九月再び出づ、其れより五十七年を經て天武天皇白鳳十三年十月十四日夜大地震にて塞がる、爾來八百四十七年のこと傳記なし、享祿四年大地震して悉く湯桁を埋む、其の後ち八十三年を經て慶長十九年十月廿五日大地震にて山崩れ線脈塞がる、又之れより十一年を經て寬永二年三月十八日地震にて塞る、夫れより六十一年を經、貞享二年十二月十一日地震にて泥湯湧出後ち二十三年、寶

第三十九章　道後湯之町　第百六十八節　神社佛閣及名勝舊跡　五七一



図：史料の例

史料批判について

史料批判の概要

- 史料批判とは何か
  - ある史料の価値（史料の有効性と信頼性）を見極める作業のこと
  - 歴史研究におけるアプローチの基礎
- 一次史料、二次史料
  - 「その時」「その場で」「その人が」叙述＝「一次史料」（国立国会図書館の定義）
  - 作成された年、作成された場所，作成した人が特定可能かどうか
    - 例：伊佐庭如矢日誌＝日誌そのものが編纂され、関連文書の書誌が掲載されているので「二次史料」、掲載された書誌の原本が「一次史料」になる

形式批判と内容批判：キャラクタ史料の場合

- **形式批判＝史料の真贋の検証**
  - 検証事項の例
    - オリジナルな史料か（一次史料か否か）
    - 史料そのものの偽造や改変の可能性はないか
    - 史料の来歴（作成経緯や時代背景，作成者の社会的位置付けなど）
- **内容批判＝史料内容の真贋の検証**
  - 検証事項の例
    - 故意か過失かに関わらず、史料の内容に虚偽／錯誤の可能性はないか
    - 他の史料と比較して内容に矛盾がないか
      - 「記録」が目的の地図や台帳⇔日記や書簡「なぜ書かれたのか」
    - 例：旧土地台帳、旧土地台帳付属図面
      - 税務署→法務局へ移管され、公証において「正当な文書」とされていることから、公的な課税証明という範囲で有効かつ信頼性がある，と判断

形式批判と内容批判：キャラクタ史料の場合

- **形式批判＝史料の真贋の検証**
  - 検証事項の例
    - オリジナルな史料か（一次史料か否か）
    - 史料そのものの偽造や改変の可能性はないか
    - 史料の来歴（作成経緯や時代背景，作成者の社会的位置付けなど）
- **内容批判＝史料内容の真贋の検証**
  - 検証事項の例
    - 故意か過失かに関わらず、史料の内容に虚偽／錯誤の可能性はないか
    - 他の史料と比較して内容に矛盾がないか
      - 「記録」が目的の地図や台帳⇔日記や書簡「なぜ書かれたのか」
    - 例：旧土地台帳、旧土地台帳付属図面

史料に限らず研究で使用する実データの有効性と信頼性には常に細心の注意を

EMアルゴリズムの導入

はじめに

- 対象のモデル構造に応じて，推定手法と推定アルゴリズムを検討するのが良い　最尤推定でいい？推定アルゴリズムは準-Newton法でいい？
- 最小化（最大化）問題に対して最適化関数optimを用いた

MNLモデルのパラメータ推定

対数尤度関数の最大化

```
76  ##### 対数尤度関数 fr の最大化#####
77
78  ##パラメータ値の最適化
79  res <- optim(b0,fr, method = "Nelder-Mead", hessian = TRUE, control=list(fnscale=-1))
80
```

最適化関数optim　(Rがあらかじめ用意してくれている関数)

optim(par, fn, gr =NULL, method = “~~~~”, lower, upper, control = list(), hessian)

b0というパラメータを初期値として，frを最大化するようにパラメータを動かしながら反復して探索.
その時の探索方法は”~~~~”.
ヘッセ行列を返すように指示して，最小化ではなく最大化.

という命令をしている.

対数尤度を最大化したときの結果のもろもろを変数resに入れている.

図：2019年度スタートアップゼミ4/24より抜粋

EMアルゴリズムとは

目的：隠れ(潜在)変数を持つモデルの最尤解を見つける

モデルが同時確率分布$P(x, s|\theta)$を持つ場合，対数尤度関数は

$$\ln P(x|\theta) = \ln \sum_{s} P(x, s|\theta)$$

x: 観測変数

s: 隠れ変数

θ: パラメータ

となり一般的に最大化問題を解くことが困難になる

そこで

基本方針

- 現在パラメータθ^{old}を用いてsの事後分布$P(s|x, \theta^{old})$を計算
- 同時確率分布の対数尤度$\log P(x, s|\theta)$の期待値を計算
  Expectation
- 期待値を最大にするθ^{new}を推定する
  Expectation　Maximization

EMアルゴリズムとは

1. E-step： θ^{old} を使って隠れ変数 sの事後分布$P(s|x,\theta^{old})$を計算
2. E-step：対数尤度 $\ln P(x,s|\theta)$の期待値(Q関数)を計算

$$Q(\theta,\theta^{old}) = \sum_{s} P(s|x,\theta^{old})\ \ln P(x,s|\theta)$$

3. M-step：Q関数をθについて最大化し新しいパラメータθ^{new}を決定

$$\theta^{new} = arg\max_{\theta} Q(\theta,\theta^{old})$$

4. 収束判定：対数尤度関数$\log P(x,s|\theta^{new})$を計算

   収束条件を満たしている：終了

   収束条件を満たしていない： $\theta^{old} \leftarrow \theta^{new}$ としてθを更新し2へ戻る

ロジットモデルへの適用例

- 隠れ変数の例
  - 選択主体（歩行者，避難する人etc)の行動規範には異質性(heterogeneity)がある→潜在クラスモデル(latent class model) を導入すればいいのでは
    - 潜在クラスモデルとは：ある説明変数βx_nで特徴付けられる主体nは，いずれかの潜在クラス$\mathbf{c}$に確率$\boldsymbol{s_c}$で帰属する，と考えるモデル
  - 「主体nがどのクラスにどれくらい帰属しているか」は観測データからはわからない＝**帰属確率 $\boldsymbol{s_c}$ が隠れ変数**
- 主体の潜在クラスへの帰属確率s_c(隠れ変数)とクラスごとのパラメータβ_{kc}(顕在変数)を同時に推定する． k: 説明変数xの番号
  - E, Train Kenneth (1st 2003, 2nd 2009) " Discrete choice methods with simulation" の"Chapter 14. EM Algorithms" を参考に実装.

ロジットモデルへの適用：定式化

- 考え方
  - 潜在クラスごとに効用(モデル)があるイメージ
  - ∴(潜在クラスの数 C)×(パラメータの数 K) 個ぶんのパラメータを推定することになる.
- 準備
  - 便宜上，パラメータ $\theta = \langle s_c, \beta_{kc}, c = 1, \dots, C, k = 1, \dots, K \rangle$ とおく
  - 主体nが選択肢jを選択した場合　効用関数は，$U_{ni} = \beta_{ni} x_{ni} + \varepsilon_{nj}$
    確定項　誤差項

    ε はガンベル分布を仮定.

ロジットモデルへの適用：定式化

- 潜在クラスcごとの選択確率

$$P'_{ni}(\beta_{kc}) = \frac{\exp(\beta_{kc} x_{kc})}{\sum_{i \in I} \exp(\beta_{kc} x_{kc})}$$

(主体の数)N行×(クラス数)C列

- 条件なしの選択確率

$$P_{ni}(s) = \sum_{c \in C} s_c P'_{ni}(\beta_c)$$

(主体の数)N行×1列

- モデルの尤度関数

$$LL(s) = \sum_{n} \log P_{ni}(s)$$

- 確率s_cでクラスcに所属する主体nが
  選択肢iを選ぶ場合の条件付き確率（sは固定）

$$h(\beta_{kc}; i_n, s_c) = \frac{s_c P'_{ni}(\beta_c)}{P_{ni}(s)}$$

(主体の数)N行×(クラス数)C列

ロジットモデルへの適用：定式化

- この問題におけるQ関数は

$$Q(\theta^0|\theta) = \sum_n \sum_c h(\beta_{kc}\,; i_n, s_c) \log\left(s_c P'_{ni}(\beta_c)\right)$$

- 潜在クラスごとの選択確率には隠れ変数s_c

  が含まれていないので以下の二式をそれぞれ扱ってやれば良い

  隠れ変数を含む／含まない式で分離

$$Q(s_c|\theta) = \sum_n \sum_c h(\beta_{kc}\,; i_n, s_c) \log s_c$$

$$Q(\beta_{kc}|\theta) = \sum_c h(\beta_{kc}\,; i_n, s_c) \log P'_{ni}(\beta_c)$$

ロジットモデルへの適用

1. 初期値パラメータ$\theta^0=\langle\beta_c^0, s^0\rangle$を設定

この時，パラメータβ_c^0は$C=1$の時の結果(理由は後述)，$s^0=\frac{1}{C}$とする．

2. E-step：負担率の計算

$$h^{old}\left(\beta_{kc}^{old}; i_n, s^{old}\right)=\frac{s_c^{old} P'_{ni}\left(\beta_c^{old}\right)}{P_{ni}(s^{old})} \qquad old：一つ前のstep$$

3. M-step: Q関数を最大化するパラメータを推定

$Q(s|\theta)$の場合

$$s_c^{new}=\frac{\sum_n h^{old}\left(\beta_{kc}^{old}; i_n, s_{c'}^{old}\right)}{\sum_n \sum_{c'} h^{old}\left(\beta_{kc}^{old}; i_n, s_{c'}^{old}\right)}$$

前のstepでQ関数を最大化するパラメータβを代入して更新する

$Q(\beta_{kc}|\theta)$の場合

$$\beta_{kc}^{new}=argmax\sum_c h^{old}\left(\beta_{kc}^{old}; i_n, s_c^{old}\right)\log P'_{ni}(\beta_c)$$

対数尤度最大化に対して最適化問題を解くやり方でβを推定する

4. 対数尤度　$LL=\sum_n \log P_{ni}(s^{new})$ を計算

5. 1-4のステップを繰り返し，収束するまでパラメータを更新

実装例：推定コード

```
# EMアルゴリズム
##結果を格納するマトリクスの設定 nrow=計算回数
###（省略）###

##初期尤度の算出
p0<-matrix(1/table(Data[,12]))
pc0<-sapply(1:hh,function(x){sum(sc*p0[x,])}) # 条件なしの選択確率
L0<-sum(log(pc0)) # 初期尤度

##アルゴリズムの本編
for(i in 1:num){
  # クラスごとの確率Pncを計算する（p12.段階1）
  pnc<-sapply(1:c,function(y){ ⟺ })
  # 帰属確率scを重みにして，条件なしの選択確率Pを計算する（p12.段階2）
  p<-sapply(1:hh,function(x){sum(sc*pnc[x,])})
  # クラスごとのパラメータβcで計算された選択確率にクラス帰属確率scを重みづけした事後確率hを計算する(E-step:期待値の計算)
  h<-t(sapply(1:hh,function(x){sc*pnc[x,]/p[x]}))
  # 次のステップのクラス帰属確率sc_newを計算しscを更新する（M-step:Q関数を最大化するパラメータsc_newを算出）
  sc<-apply(h,2,sum)/sum(apply(h,2,sum))
  # 次のステップのパラメータβを最尤推定でoptim関数を使って計算する（M-step:Q関数を最大化するパラメータb_newを算出）
  ## パラメータを更新する
  tval<-matrix(0,nrow=1,ncol=length(b0))
  for(j in 1:c){ ⟺ }
  # 更新したパラメータで対数尤度を計算する(p.14 step4)
  L_pnc<-sapply(1:c,function(y){ ⟺ })
  L_p<-sapply(1:hh,function(x){sum(sc*L_pnc[x,])})
  LL<-sum(log(L_p)) # 対数尤度の計算

  #結果を格納
  ###（省略）###
  if(abs(resLL[i+1]-resLL[i])<0.001){ #収束条件
    break
  }
```

メリット／デメリット

- 比較的扱いやすい／理解しやすいアルゴリズム
- 計算に時間がかかる
  - E-stepとM-stepを繰り返すのでそれはそう
  - 1回のstepごとに，クラス数分のモデルの顕在パラメータβを最適化関数で推定しているので，クラス数が増えた分計算時間がかかる

    今回は2クラスが最適なクラス数だったのでなんとかなった
- 初期値によっては局所最適解に陥る可能性がある
  - 潜在クラスを仮定しないC=1の場合のモデル構造で推定したパラメータを使用


さいごに

さいごに

- 重要な点として…
  - 観測データの特徴を吟味し，理解すること
  - 観測データの不確実性や潜在データを考慮して，モデリングすること
  - モデル構造にとって最適な推定方法や推定アルゴリズムを吟味すること
    - 今日紹介したEMアルゴリズムはその一例です
- 史料も観測された実データの一種です
  - 史料批判とは確からしい叙述への第一歩
  - どの範囲で価値があるのかを念頭に置きつつ，発見をやっていきましょう

楽しく研究をやっていきましょう！