2021年6月18日　統計数理研究所　オープンハウス

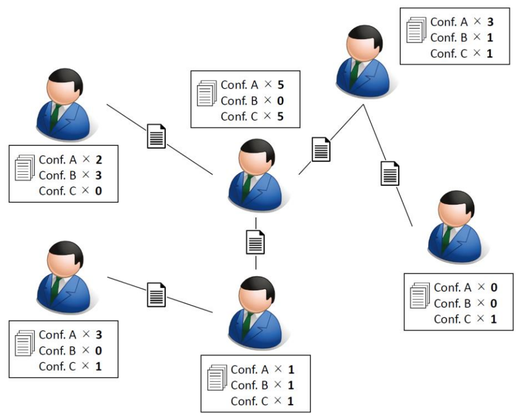
ノンパラメトリックなリンク予測に関する理論

奥野　彰文　　統計思考院　助教

※本研究は矢野恵佑准教授（統計数理研究所・数理推論系）との共同研究であり https://arxiv.org/abs/2012.13106 にてプレプリントを公開済です.

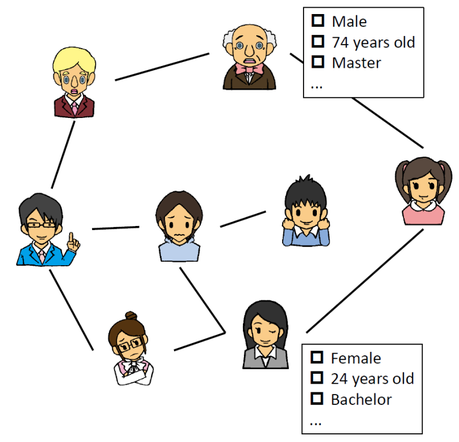
1. 背景

グラフのノードの特徴量（共変量）を $\{\boldsymbol{x}_i\} \subset \chi \subset \mathbb{R}^p$，ノード間のリンクの重みを $\{y_{ij}\}_{1\le i\neq j\le n} \subset \mathcal{Y}\ (\subset \mathbb{R})$ と書きます：



例）共著ネットワーク：

$y_{ij} \in \mathbb{N}$：研究者i,jが2人で書いた論文の本数
$\boldsymbol{x}_i \in \mathbb{N}^p$：研究者$i$が各雑誌に乗せた論文の本数を並べたもの



例）友達ネットワーク：

$y_{ij} \in \{0,1\}$：個人i,jが友達であるかどうか
$\boldsymbol{x}_i \in \mathbb{N}^p$：個人$i$の性質（年齢など）を並べたもの

リンク回帰問題	組 $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ から 重み y_{ij} を予測したい：$y_{ij} \approx f_\theta(\boldsymbol{x}_i, \boldsymbol{x}_j)$

※類似度学習などはリンク回帰に含まれる.

2. 提案法

ノンパラメトリックリンク回帰

$$\hat{f}_{n,h}(\boldsymbol{x}, \boldsymbol{x}') := \frac{\sum_{1\le i<j\le n} y_{ij} K_h(\boldsymbol{x}-\boldsymbol{x}_i) K_h(\boldsymbol{x}-\boldsymbol{x}_j) + \lambda_{n,h}}{\sum_{1\le i<j\le n} K_h(\boldsymbol{x}-\boldsymbol{x}_i) K_h(\boldsymbol{x}-\boldsymbol{x}_j) + \lambda_{n,h}}$$

（ただし$h = h_n > 0$はバンド幅，K_hはカーネル関数，$\lambda_{n,h} > 0$は正則化パラメータ．）

$h = h_n \to 0$ のとき推定量は一致性を持つ：$\hat{f}_{n,h}(\boldsymbol{x}, \boldsymbol{x}') \to f_*(\boldsymbol{x}, \boldsymbol{x}') = \mathbb{E}[Y \mid X = \boldsymbol{x}, X' = \boldsymbol{x}']$.

3. 予測値の漸近分散に関する発見

ノンパラメトリック統計解析では共変量 $\boldsymbol{x}_i$ に以下のどちらかのデザインを仮定することが多い：

Random design：共変量 $\boldsymbol{x}_i$ が**実験ごとに**ある分布から**ランダム**に生成されると仮定する

Fixed design：共変量 $\boldsymbol{x}_i$ は**すべての実験で常に同じ値をとる**と仮定する

古典的なノンパラメトリック回帰では，上記のどちらのデザインを仮定しても漸近的に同等である (Brown and Low (AoS1996), Reiss (AoS2008) など).

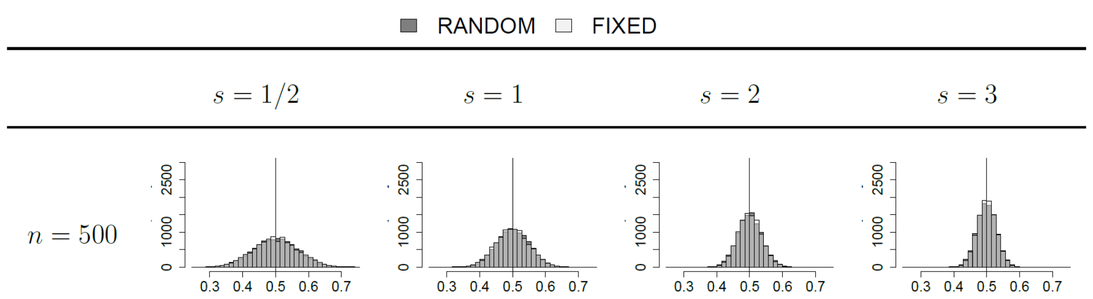


表1：古典的なノンパラメトリック回帰での予測値のヒストグラム．デザインによらずヒストグラムはほぼ同じ．バンド幅$h = n^{-\frac{1}{s+d}}$.

では，提案したノンパラメトリックリンク回帰ではどうなるか？

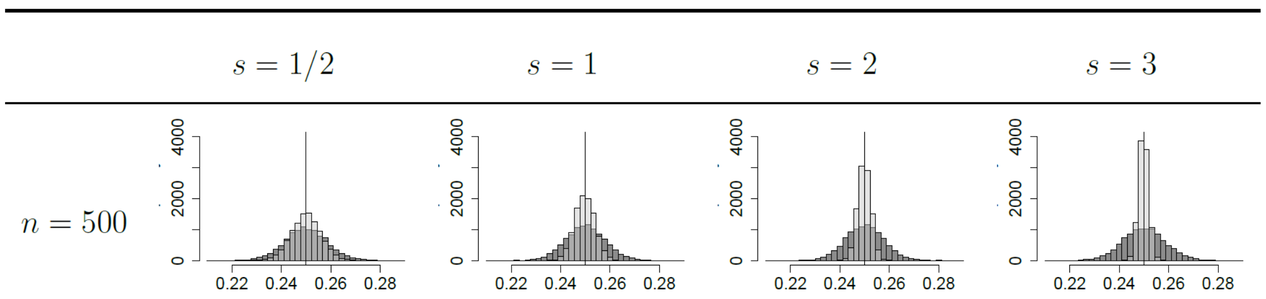


表2：ノンパラメトリックリンク回帰での予測値のヒストグラム．バンド幅によらず，ヒストグラムがデザインに依存する．バンド幅$h = n^{-\frac{1}{s+d}}$.

ノンパラメトリックリンク回帰の**予測値の漸近分散は共変量x_iのデザインに依存する**ことを発見し，理論的にも評価した．
これは既存のノンパラメトリック解析で**一般的に知られていない**特殊な現象であり，類似度学習などで予測値の（漸近的な）統計的性質を調べるには特別な注意を払う必要があることが分かった．

大学共同利用機関法人 情報・システム研究機構
統計数理研究所
The Institute of Statistical Mathematics