多次元空間における埋め込み 1 次元曲線の同時信頼領域

千葉大・融合理工学府　山添 滉弥
千葉大・理学研究院　内藤 貫太

はじめに:　道路交通網上の交通事故や，脳神経細胞網における樹状突起スピンの分布といった，枝分かれした 1 次元曲線上で観測されるデータをネットワークデータを呼ぶ．ネットワークデータに対して，与えられたネットワーク上における密度推定の手法についてはすでに研究されている (Liu and Ruppert, 2021; McSwiggan et al., 2017)．しかしながら，ネットワークを構成する 1 次元曲線の同時信頼領域をデータから構築する方法については，いまだ研究がないようである．本研究ではこの問題を多次元空間に埋め込まれた 1 次元曲線の同時信頼領域を構築する問題として捉え，ノンパラメトリック回帰のアプローチから取り組む．

問題:　$I=[a,b]\subset\mathbb{R}$ を閉区間とし，$\varphi_i : I\to\mathbb{R}$ $(i=1,\ldots,d)$ を滑らかな関数とする．このとき，$\mathbb{R}^d$ に埋め込まれた 1 次元曲線を

$$\boldsymbol{\varphi}(t)=\begin{bmatrix}\varphi_1(t)\\ \vdots\\ \varphi_d(t)\end{bmatrix}\in\mathbb{R}^d$$

で表す．確率ベクトル $\boldsymbol{Y}\in\mathbb{R}^d$ と共変量 $T\in I$ の組 $(T,\boldsymbol{Y})$ からの n 個の観測値 $(T_1,\boldsymbol{Y}_1),\ldots,(T_n,\boldsymbol{Y}_n)$ に基づき 1 次元曲線 $\boldsymbol{\varphi}$ の同時信頼領域を構築したい．つまり，$M_{\boldsymbol{\varphi}}=\left\{\boldsymbol{\varphi}(t)\in\mathbb{R}^d \mid t\in I\subset\mathbb{R}\right\}$ としたとき，十分小さい $\alpha>0$ に対して，

$$\mathbb{P}\left(\mathfrak{D}\supset M_{\boldsymbol{\varphi}}\right)\geq 1-\alpha$$

となるような有界閉領域 $\mathfrak{D}\subset\mathbb{R}^d$ の構築を目指す．

設定:　1 次元曲線の推定の関連研究である Hastie and Stuetzle (1989) を参照しつつ，モデルの仮定を行う．$\mathbb{R}^d$ に埋め込まれた 1 次元曲線 $\boldsymbol{\varphi}$ に対して，任意の点 t における勾配ベクトル $\boldsymbol{\varphi}'(t)$ を $\boldsymbol{\varphi}'(t)=[\varphi_1'(t)\ \cdots\ \varphi_d'(t)]^T$ で表す．また，$\boldsymbol{n}_1(t),\ldots,\boldsymbol{n}_{d-1}(t)$ を $\{\boldsymbol{\varphi}'(t)/\left\|\boldsymbol{\varphi}'(t)\right\|,\boldsymbol{n}_1(t),\ldots,\boldsymbol{n}_{d-1}(t)\}$ が $\mathbb{R}^d$ の正規直交基底となるようにとる．さらに，T を I 上の密度 f_T に従う確率変数とし，$\boldsymbol{V}=[V_1\ \cdots\ V_{d-1}]^T$ を原点中心半径 r の $(d-1)$ 次元球 B_r^{d-1} 上の密度 $f_{\boldsymbol{V}}$ に従う確率ベクトルとする．

$(T_1,\boldsymbol{V}_1),\ldots,(T_n,\boldsymbol{V}_n)\overset{\text{i.i.d.}}{\sim} f_T\times f_V$ と $E[\boldsymbol{V}_1]=\boldsymbol{0}$ を仮定する．確率ベクトル $\boldsymbol{Y}$ の d 次元標本 $\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_n$ として，

$$\boldsymbol{Y}_i=\boldsymbol{\varphi}(T_i)+N(T_i)\boldsymbol{V}_i$$

を考える．ただし，$\boldsymbol{Y}_i=[Y_{i1}\ \cdots\ Y_{id}]^T,\boldsymbol{V}_i=[V_{i1}\ \cdots\ V_{id-1}]^T,N(t)=[\boldsymbol{n}_1(t)\ \cdots\ \boldsymbol{n}_{d-1}(t)]$ である．

同時信頼領域の構築:　共変量 $\boldsymbol{T}=[T_1\ \cdots\ T_n]^T$ と $\boldsymbol{Y}$ の第 j-成分のデータ $\widetilde{\boldsymbol{Y}}_j=[Y_{1j}\ \cdots\ Y_{nj}]^T$ との組 $(\boldsymbol{T},\widetilde{\boldsymbol{Y}}_i)$ $(i=1,\ldots,d)$ に基づき，t における $\varphi_i(t)$ の局所線形推定量 $\widehat{\varphi}_i(t)(i=1,\ldots,d)$ が得られる．これらを並べることにより $\boldsymbol{\varphi}(t)$ の推定量 $\widehat{\boldsymbol{\varphi}}(t)$ が構築される．さらに $\widehat{\boldsymbol{\varphi}}(t)$ を用いてその勾配ベクトルを求める：

$$\widehat{\boldsymbol{\varphi}}(t)=\begin{bmatrix}\widehat{\varphi}_1(t)\\ \vdots\\ \widehat{\varphi}_d(t)\end{bmatrix},\quad \widehat{\boldsymbol{\varphi}}'(t)=\frac{d}{dt}\widehat{\boldsymbol{\varphi}}(t).$$

局所線形推定量については Wand and Jones (1995); Fan and Gijbels (1996) などを参照されたい．$\widehat{\boldsymbol{n}}_1(t),\ldots,\widehat{\boldsymbol{n}}_{d-1}(t)$ を $\{\widehat{\boldsymbol{\varphi}}'(t)/\|\widehat{\boldsymbol{\varphi}}'(t)\|, \widehat{\boldsymbol{n}}_1(t),\ldots,\widehat{\boldsymbol{n}}_{d-1}(t)\}$ が $\mathbb{R}^d$ の正規直交基底となるようにとり，

$$\begin{aligned}
\widehat{\mathfrak{D}}_a(r) &= \left\{ \widehat{\boldsymbol{\varphi}}(a) - R\frac{\widehat{\boldsymbol{\varphi}}'(a+)}{\|\widehat{\boldsymbol{\varphi}}'(a+)\|} + \widehat{N}(a)\boldsymbol{v} \,\middle|\, 0 \le R \le r, \boldsymbol{v} \in B^{d-1}_{\sqrt{r^2-R^2}} \right\}, \\
\widehat{\mathfrak{D}}_J(r) &= \left\{ \widehat{\boldsymbol{\varphi}}(t) + \widehat{N}(t)\boldsymbol{v} \,\middle|\, t \in J, \boldsymbol{v} \in B^{d-1}_r \right\}, \\
\widehat{\mathfrak{D}}_b(r) &= \left\{ \widehat{\boldsymbol{\varphi}}(b) + R\frac{\widehat{\boldsymbol{\varphi}}'(b-)}{\|\widehat{\boldsymbol{\varphi}}'(b-)\|} + \widehat{N}(b)\boldsymbol{v} \,\middle|\, 0 \le R \le r, \boldsymbol{v} \in B^{d-1}_{\sqrt{r^2-R^2}} \right\}, \\
\widehat{N}(t) &= \begin{bmatrix} \widehat{\boldsymbol{n}}_1(t) & \cdots & \widehat{\boldsymbol{n}}_{d-1}(t) \end{bmatrix}
\end{aligned}$$

とする．3 つの排反な領域を用いて同時信頼領域 $\widehat{\mathfrak{D}}(r) \subset \mathbb{R}^d$ を

$$\widehat{\mathfrak{D}}(r) = \widehat{\mathfrak{D}}_a(r) \cup \widehat{\mathfrak{D}}_J(r) \cup \widehat{\mathfrak{D}}_b(r)$$

として定義する (図 1 参照)．本講演では $\widehat{\boldsymbol{\varphi}}(t)$ と $\widehat{\mathfrak{D}}(r)$ の理論的性質および，信頼係数 α から $\widehat{\mathfrak{D}}(r)$ の r を決定する方法について解説を与える．また，共変量 $\boldsymbol{T}$ がデータとして得られていない場合において，$\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_n$ から $T_1,\ldots,T_n$ を作成する方法についても考察する．

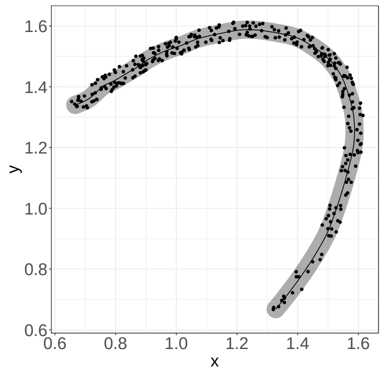

図 1　推定曲線 $\widehat{\boldsymbol{\varphi}}(t)$ から作られた同時信頼領域 $\widehat{\mathfrak{D}}(r)$ の例

適用例:　シミュレーション実験の結果および実データへの適用例については，シンポジウムにて報告する．

参考文献

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability 66*. Taylor & Francis.

Hastie, T. and Stuetzle, W. (1989). Principal curves. *Journal of the American Statistical Association*, 84, 502–516.

Liu, Y. and Ruppert, D. (2021). Density estimation on a network. *Computational Statistics & Data Analysis*, 156, 107128.

McSwiggan, G., Baddeley, A., and Nair, G. (2017). Kernel density estimation on a linear network. *Scandinavian Journal of Statistics*, 44, 324–345.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*. Chapman & Hall.