正定値カーネルによる条件付確率の推定とその応用

新機軸創発センター 関数解析的統計推論グループ
教授　　福水　健次

1　はじめに

近年，正定値カーネルを用いた方法論いわゆる「カーネル法」の新しい展開として，確率変数を再生核ヒルベルト空間に写像した特徴ベクトルの平均によって確率分布を表現することにより，分布の均一性や独立性などの統計的問題を議論する方法が開発されてきた．本報告では，ヒルベルト空間上で条件付確率を表現する方法と，その時系列データへの応用に関して述べる．

2　正定値カーネルによる条件付確率の表現

$k_{\mathcal{X}}, k_{\mathcal{Y}}$ をそれぞれ $\mathcal{X}, \mathcal{Y}$ 上の有界な正定値カーネルとする．(X, Y) を $\mathcal{X} \times \mathcal{Y}$ 上に値を持つ確率変数で，その分布を P_{XY}，周辺分布を P_X, P_Y とする．特徴写像 $\Phi_{\mathcal{X}} : \mathcal{X} \to \mathcal{H}_{\mathcal{X}}, x \mapsto k_{\mathcal{X}}(\cdot, x)$ によって，確率変数 X を $\mathcal{H}_{\mathcal{X}}$ 上の確率変数 $\Phi_{\mathcal{X}}(X)$ に写像し，その $\mathcal{H}_{\mathcal{X}}$ での平均

$$m_X \equiv E[\Phi_{\mathcal{X}}(X)] = E[k_{\mathcal{X}}(x, X)]$$

を考える．このとき

$$\langle m_X, f \rangle_{\mathcal{H}_{\mathcal{X}}} = E[f(X)] \qquad (\forall f \in \mathcal{H}_{\mathcal{X}})$$

が成り立つ．m_X は X の高次の統計量の情報を有し，ガウス RBF カーネル $k(x_1, x_2) = \exp(-\|x_1 - x_2\|^2/\sigma^2)$ など様々なカーネルに対して，m_X によって確率分布 P_X が一意的に定まることが知られている (Fukumizu et al., 2009)．そこで m_X を分布 P_X の表現として用いることができる．

X, Y をそれぞれ $\Phi_{\mathcal{X}}(X), \Phi_{\mathcal{Y}}(Y)$ と埋め込み，$\mathcal{H}_{\mathcal{X}} \otimes \mathcal{H}_{\mathcal{Y}}$ 上で共分散 Σ_{YX} を考えると，

$$\langle g, \Sigma_{YX} f \rangle_{\mathcal{H}_{\mathcal{Y}}} = \mathrm{Cov}[f(X), g(Y)] \qquad (\forall f \in \mathcal{H}_{\mathcal{X}}, \forall g \in Hy)$$

が成り立つ．

いま $g \in \mathcal{H}_{\mathcal{Y}}$ に関して，条件付平均 $E[g(Y)|X = x]$ を $\mathcal{X}$ 上の関数とみたとき，$\mathcal{H}_{\mathcal{X}}$ に属すると仮定する．このとき

$$\Sigma_{XX} E[g(Y)|X] = \Sigma_{XY} g$$

が成り立つ (Fukumizu et al., 2004)．そこで，$\Sigma_{YX}\Sigma_{XX}^{-1}$ が考えられれば，

$$\langle g, \Sigma_{YX}\Sigma_{XX}^{-1}\Phi_{\mathcal{X}}(x)\rangle_{\mathcal{H}_{\mathcal{Y}}} = \langle \Sigma_{XX}^{-1}\Sigma_{XY} g, \Phi_{\mathcal{X}}(x)\rangle_{\mathcal{H}_{\mathcal{X}}} = E[g(Y)|X = x] \tag{2.1}$$

となり，$\Sigma_{YX}\Sigma_{XX}^{-1}\Phi_{\mathcal{X}}(x)$ は条件付確率を表現すると考えることができる[1]．

そこで $\varepsilon_n > 0$ を正則化定数として，$E[g(Y)|X]$ の推定量を

$$\widehat{C}_{Y|X}^{(n)} g = (\widehat{\Sigma}_{XX}^{(n)} + \varepsilon_n I)^{-1} \widehat{\Sigma}_{XY}^{(n)} g$$

によって定める．このとき次のように，$\widehat{C}_{Y|X}^{(n)} g$ は $E[g(Y)|X]$ の一致推定量である．

[1] $E[g(Y)|X = x] \in \mathcal{H}_{\mathcal{X}}$ は成り立たない場合もあり，$\Sigma_{YX}\Sigma_{XX}^{-1}$ も well-defined とは限らないため，必ずしも厳密な議論とは言えない．

定理 1 $\varepsilon_n \to 0, \varepsilon_n n^{1/3} \to \infty$ のとき,

$$\left\| \left\{ \widehat{C}^{(n)}_{Y|X} g - \frac{1}{n}\sum_{i=1}^{n} \big(\widehat{C}^{(n)}_{Y|X} g\big)(X_i) \right\} - \big\{ E[g(Y)|X] - E[g(Y)] \big\} \right\|_{L^2(P_X)} = o_p(1) \quad (n \to \infty).$$

さらに $E[g(Y)|X] \in \mathcal{H}_{\mathcal{X}}$ が成り立つとき,

$$\big\| \widehat{C}^{(n)}_{Y|X} g - E[g(Y)|X] \big\|_{\mathcal{H}_{\mathcal{X}}} = o_p(1) \qquad (n \to \infty).$$

3　時系列データへの応用

(2.1) 式において X に関する期待値をとると,

$$\Sigma_{YX}\Sigma_{XX}^{-1} m_X = m_Y \tag{3.1}$$

が成り立つ．これは，確率分布に関する関係式 $\int p(y|x)p(x)dx = p(y)$ のカーネルによる表現である．この関係式を基礎として，有向グラフによるグラフィカルモデルなど，条件付確率によって定義されるモデルの推論に正定値カーネルの方法を用いることが可能である．本報告では，観測されない状態変数 Z_t と観測される変数 X_t を持つ定常な状態空間モデル

$$p(\{X_i\}_{i=1}^T, \{Z_i\}_{i=1}^T) = \pi(Z_0)\prod_{t=0}^{T} p(X_t|Z_t) \prod_{t=1}^{T-1} p(Z_{t+1}|Z_t)$$

におけるフィルタリングの問題に適用する．ここでは，$p(X_t|Z_t)$ や $p(Z_{t+1}|Z_t)$ にパラメトリックなモデルを仮定しない．そのかわりに，学習サンプルとして (X_t, Z_t) の例が与えられ，それに基づいて共分散作用素の推定を行い，その後あらたなデータに対して (3.1) 式の関係式によって条件付確率 $p(Z_t|x_0,\ldots,x_t)$ の逐次推定を行う (Song et al., 2009)．具体的には，まず学習サンプルによって，$\widehat{\Sigma}^{(n)}_{Z_tZ_t}$, $\widehat{\Sigma}^{(n)}_{Z_{t+1}Z_t}$, $\widehat{\Sigma}^{(n)}_{X_tX_t}$ を推定する．次に $x_0, x_1, \ldots$ を新たなデータとするとき，$\widehat{m}_{Z_0|x_0} = \widehat{C}_{Z|X}\Phi_{\mathcal{X}}(x_0)$ による初期化のあと，$p(Z_t|x_0,\ldots,x_t)$ の表現 $\widehat{m}_{Z_t|x_0,\ldots,x_t}$ を更新していく．このためには，$p(Z_{t+1}|X_0,\ldots,X_t,X_{t+1}) = p(Z_{t+1}|X_{t+1},Z_t)p(Z_t|X_0,\ldots,X_t,X_{t+1}) \approx p(Z_{t+1}|X_{t+1},Z_t)p(Z_t|X_0,\ldots,X_t)$ の近似に基づいて

$$\widehat{m}_{Z_{t+1}|x_0,\ldots,x_{t+1}} = \widehat{C}_{Z_{t+1}|Z_t}\widehat{m}_{Z_t|x_0,\ldots,x_t} + \widehat{C}^{(n)}_{Z_{t+1}|X_{t+1}}\Phi_{\mathcal{X}}(x_{t+1})$$

により逐次更新を行う，報告会においては，より詳しい方法と，ノイズを伴った観測から回転角を推定する数値シミュレーションの結果についても論じる．

参考文献

K. Fukumizu, F.R. Bach, and M.I. Jordan. Dimensionality reduction for supervised learning with reproducing kernel Hilbert spaces. *J. Machine Learning Research*, 5:73–99, 2004.

K. Fukumizu, F.R. Bach, and M.I. Jordan. Kernel dimension reduction in regression. *Ann. Stat.*, 37(4):1871–1905, 2009.

L. Song, J. Huang, A. Smola, and K. Fukumizu. Hilbert space embeddings of conditional distributions with applications to dynamical systems. In *Proc. ICML2009*, pages 961–968, 2009.