U-divergence による密度推定

内藤貫太（島根大・総合理工）　江口真透（統計数理研究所）

1　はじめに

いわゆる Kullback-Leibler divergence が，統計科学のみならず，様々な分野において重要な役割を演じているのは周知の事実である．近年，Kullback-Leibler divergence を特殊な場合として含むような，より一般化された divergence が議論されており，その 1 つに，凸関数 U から構成される U-divergence がある．本講演では，U-divergence の最適化を通して得られる，stagewise な密度推定について議論した．提案した密度推定量の精度評価としての非漸近的誤差限界について解説した．

2　問題設定と方法

d 変量の確率密度関数 $f(x)$ の推定を考える．その推定は $\mathbb{R}^d$ のデータ $X_1, \cdots, X_n$ に基づいて実行される．機械学習の分野で議論が盛んなブースティングと類似した手法である，経験リスクの逐次最小化に基づいて密度推定量を構成する．そこで divergence が必要となる．

3　U-divergence と U-loss 関数

単調増加な凸関数 U を考え，その導関数を u とする．また，u の逆関数を ξ とし，三つ組 (U, u, ξ) を常に連想するものとする．密度関数 f と g の U-divergence は

$$D_U(f,g) = \int_{\mathbb{R}^d} \left[U\left(\xi(g(x))\right) - U\left(\xi(f(x))\right) - f(x)\left\{\xi(g(x)) - \xi(f(x))\right\}\right] dx$$

で定義され，f にのみ依存する項を落としたその経験版である U-loss 関数は

$$\widehat{L}_U(g) = -\frac{1}{n}\sum_{i=1}^{n} \xi(g(X_i)) + \int_{\mathbb{R}^d} U\left(\xi(g(x))\right) dx$$

で定義される．実際には，この $\widehat{L}_U(g)$ の最小化を通して密度推定量が構成される．

　ここで，$U(t) = \exp(t)$ を用いた場合には U-divergence は Kullback-Leibler divergence に他ならない．また，

$$U_\beta(t) = \frac{1}{\beta+1}(1+\beta t)^{\frac{\beta+1}{\beta}}, \quad 0 < \beta < 1$$

という選択は β-power divergence を導き，ロバスト推定などで議論されているものとなる．

4　アルゴリズムと推定量

$\mathcal{D}$ を密度関数の集合，$\epsilon > 0$ を近似限界，$0 < \pi_k < 1 (k = 1, ..., M-1)$ を Mixing Coefficients とする．密度推定量をは以下のアルゴリズムを経て得られる：

Step 1　Choose $\tilde{f}_0 \in \mathcal{D}$ so that

$$\widehat{L}_U(\tilde{f}_0) \leq \inf_{\phi \in \mathcal{D}} \widehat{L}_U(\phi) + \epsilon$$

Step 2　For $k = 1, ..., M-1$, let

$$\tilde{f}_k = u\left((1-\pi_k)\xi(\tilde{f}_{k-1}) + \pi_k \xi(\tilde{\phi})\right),$$

where $\tilde{\phi} \in \mathcal{D}$ is chosen so that

$$\widehat{L}_U(\tilde{f}_k) \leq \inf_{\phi \in \mathcal{D}} \widehat{L}_U\left(u\left((1-\pi_k)\xi(\tilde{f}_{k-1}) + \pi_k \xi(\phi)\right)\right) + \pi_k \epsilon.$$

Step 3　Let $\widehat{f} = \tilde{f}_{M-1}$.

5　理論的精度評価

精度評価として，$\widehat{f}$ に関する非漸近的誤差限界

$$E_f D_U\left(f, \widehat{f}\right) \leq \inf_{g \in \mathrm{CO}(\xi(\mathcal{D}))} D_U(f, u(g)) + 2E_f \sup_{\phi \in \mathcal{D}} |\nu_n\left(\xi(\phi)\right)| + \frac{\theta^2 B_U^2}{M + (\theta - 1)} + \epsilon$$

が得られる．ここで，

$$\nu_n(\xi(\phi)) = \frac{1}{n}\sum_{i=1}^{n} \xi(\phi(X_i)) - \int_{\mathbb{R}^d} \xi(\phi(x)) f(x) dx$$

であり，$\mathrm{co}(\xi(\mathcal{D}))$ は $\xi(\mathcal{D})$ の凸包である．

6　数値的精度評価

$\widehat{f}$ の挙動を核型推定量との比較を交えながらシミュレーションにより検証した結果を報告した．提案する推定量は，多くの場合で核型推定量よりも精度が良いことが確認できた．