統計数理研究所 75 周年記念事業（2019）

ノンパラメトリックエントロピー推定とその応用
Non-parametric Entropy Estimation

モデリング研究系 日野 英逸 （Hideitsu Hino）

情報理論において最も基本的な量の一つである情報量は $I_f(x) = -\log f(x)$ で定義される．ここで $f(x)$ はデータ $x \in \mathbb{R}^d$ が従う分布の確率密度関数である．情報量の期待値はエントロピーと呼ばれる: $H(f) = E_f[I_f(X)] = -\int f(x) \log f(x) \mathrm{d}x$. 情報量，エントロピーあるいはこれらを用いて導出できる KL ダイバージェンス及び相互情報量は統計学や機械学習など非常に多くの分野で重要な役割を果たしている．

微分エントロピーの推定方法として最も良く利用されている手法の一つが，k-近傍法と確率密度関数の 1 次の展開に基づく推定法である．最近傍法に基づくエントロピー推定量は, Kozachenko and Leonenko（1987）により提案され，任意の次元の確率変数に対して mean square consistency を持つことが示されている．この結果は一般の k-近傍に基づく推定量に拡張され（Goria et al., 2005），その後も各種の拡張と理論的解析がなされている（Beirlant et al., 1997; Paninski, 2003）．確率密度関数 $f(z)$ の検査点 $z \in \mathbb{R}^p$ における値を観測データ集合 $\mathcal{D} = \{x_i\}_{i=1}^n$ を用いて推定する問題を考える．検査点 z を中心とする半径 ε の p 次元超球を $b(z;\varepsilon) = \{x \in \mathbb{R}^p | \|z - x\| < \varepsilon\}$ で表す．この超球の体積は $|b(z;\varepsilon)| = c_p \varepsilon^p$ である．ただし，$c_p = \pi^{p/2}/\Gamma(p/2+1)$ である．中心 z の ε 球に含まれる確率質量を $q_z(\varepsilon) = \int_{x \in b(z;\varepsilon)} f(x) \mathrm{d}x$ で定義する．この定義式の被積分関数を Taylor 展開すると $q_z(\varepsilon) = |b(z;\varepsilon)| \left(f(z) + O(\varepsilon^2)\right) = c_p \varepsilon^p f(z) + O(\varepsilon^{p+2})$ を得る．超球の半径 ε を十分小さいと仮定してその 2 次以上の項を無視し，確率質量を全観測データに占める超球内の点の割合で近似することで，確率密度関数の推定量 $\hat{f}(z;\varepsilon) = \frac{k_\varepsilon}{n c_p \varepsilon^p}$ を得る．ここで k_ε は観測データ集合 $\mathcal{D}$ の中で半径 ε の超球の中に含まれるものの個数である．一方，ε の代わりに超球に含まれるサンプル数 k を固定した場合，$\hat{f}(z;\varepsilon)$ は $\hat{f}^{nn}(z;k) = k/(n c_p \varepsilon_k^p)$ のようにかける．ここで超球の半径 ε_k は検査点 z からその k 番目に近い点までの距離で決定されることになる．$\hat{f}_i^{nn}(x_i;k)$ を，データ集合 $\mathcal{D} \backslash \{x_i\}$ を用いて k-近傍法により推定した推定量として，$-\ln \hat{f}_i^{nn}(x_i;k)$ の経験期待値を計算することで，k-近傍エントロピー推定量 $\hat{H}^{nn}(\mathcal{D};k) = -\sum_{i=1}^n \ln \hat{f}_i^{nn}(x_i;k)$ を得る（Goria et al., 2005）．この方法は確率質量関数の一次展開に基づく方法であるが，筆者らはより高次の展開に基づく手法を提案した（Hino et al., 2015）．検査点 z を中心とした半径 ε の超球内の確率質量 $q_z(\varepsilon)$ は ε に関する二次の Taylor 展開をすると，$q_z(\varepsilon) = c_p f(z) \varepsilon^p + \frac{n}{4(p/2+1)} c_p \mathrm{Tr} \nabla^2 f(z) \varepsilon^{p+2} + O(\varepsilon^{p+4})$ の形で表わされる．上式左辺の $q_z(\varepsilon)$ を比 k_ε/n で近似し，両辺を $c_p \varepsilon^p$ で割ることで，$\frac{k_\varepsilon}{n c_p \varepsilon^p} = f(z) + C\varepsilon^2 + O(\varepsilon^4)$ を得る．ここで $C = n \mathrm{Tr} \nabla^2 f(z)/4(p/2+1)$ である．さらに，$Y_\varepsilon = \frac{k_\varepsilon}{n c_p \varepsilon^p}$ と $X_\varepsilon = \varepsilon^2$ を導入し，ε に関する 4 次以上の項を無視することで，応答変数 Y の説明変数 X に関する一次式 $Y_\varepsilon \simeq f(z) + C X_\varepsilon$ が得られる．この式は説明変数 X による応答変数 Y の線形回帰式とみなせる．複数の半径 $\mathcal{E} = \{\varepsilon_i\}_{i=1}^m$ と置き，$\mathcal{E}$ に含まれる各 ε で定まる $\{(X_\varepsilon, Y_\varepsilon)\}_{\varepsilon \in \mathcal{E}}$ の組を観測データとみなして，二乗誤差 $R = \frac{1}{m} \sum_{\varepsilon \in \mathcal{E}} (Y_\varepsilon - f(z) - C X_\varepsilon)^2$ を最小化するように $f(z)$ と C を求める．これは単回帰に他ならず，この回帰によって得られた切片が，$f(z)$ の推定量 $\hat{f}(z)$ である．以上より検査点 z における密度の推定量 $\hat{f}^s(z)$ が得られたので，leave-one-out 推定量としてエントロピーの

推定量 $\hat{H}^s(\mathcal{D}) = -\frac{1}{n}\sum_{i=1}^{n} \ln \hat{f}_i^s(x_i)$, を得る．ここで $\hat{f}_i^s(x_i)$ は x_i を用いずに求めた密度の推定値である．筆者らは，この推定量を始めとして，より直接的に一度の回帰問題を解くことでエントロピーを推定する手法，誤差構造として Poisson 分布を仮定した手法（Hino et al., 2016），及び確率質量関数の高次展開に基づく局所フラクタル次元推定量（Hino et al., 2017）を提案し，ノンパラメトリックなアプローチによるデータ分布の特徴付けの方法論の開発を進めている．

また，関連して筆者らは一般にデータに重みが与えられている場合に適用可能な情報量推定手法を提案している（Hino and Murata, 2013）．例えば観測されたデータそれぞれに対して，その信頼度が与えられている場合や，あるいは同一の事象が重複して観測されるとして個々の事象を観測した頻度を重みとして表現する場合のように，データ $\mathcal{D}_x$ が与えられた上で，各データ点 $x_i \in \mathcal{D}_x$ の重要さとして重みが付与されることが考えられる．過去のデータの重みが小さくなっていくような忘却係数付きのオンライン観測データもこうした重み付きデータの一例である．この重み付き情報量推定量の応用として情報論的クラスタリング（Hino and Murata, 2014），変化点検知（Koshijima et al., 2015）などの方法を開発した．情報論的クラスタリング手法は同種の手法と比較してより正確なクラスタリングを実現し，変化点検知の手法は従来の手法では捉えられなかったデータの背後にある構造的な変化を抽出することに成功している．

参 考 文 献

Beirlant, J., Dudewicz, E. J., Györfi, L. and Meulen, E. C. (1997). Nonparametric Entropy Estimation: An Overview, *International Journal of the Mathematical Statistics Sciences*, **6**, 17–39.

Goria, M. N., Leonenko, N. N., Mergel, V. V. and Novi Inverardi, P. L. (2005). A new class of random vector entropy estimators and its applications in testing statistical hypotheses, *Journal of Nonparametric Statistics*, **17** (3), 277–297.

Hino, H., Fujiki, J., Akaho, S. and Murata, N. (2017). Local intrinsic dimension estimation by generalized linear Modeling, *Neural Computation*, **29** (7).

Hino, H. and Murata, N. (2013). Information estimators for weighted observations, *Neural Networks*, **46** (0), 260 – 275.

Hino, H. and Murata, N. (2014). A Non-parametric information theoretic clustering algorithm based on Quantile-based entropy estimator, *Neural Computation*, **26** (9), 2074–2101.

Hino, H., Koshijima, K. and Murata, N. (2015). Non-parametric entropy estimators based on simple linear regression, *Computational Statistics & Data Analysis*, **89** (0), 72 – 84.

Hino, H., Akaho, S. and Murata, N. (2016). An Entropy Estimator Based on Polynomial Regression with Poisson Error Structure, *Neural Information Processing - 23rd International Conference, ICONIP 2016, Kyoto, Japan, October 16-21, 2016, Proceedings, Part II*, 11–19.

Koshijima, K., Hino, H. and Murata, N. (2015). Change-Point Detection in a Sequence of Bags-of-Data, *Knowledge and Data Engineering, IEEE Transactions on*, **27** (10), 2632–2644, Oct.

Kozachenko, L. F. and Leonenko, N. N. (1987). Sample estimate of entropy of a random vector, *Problems of Information Transmission*, **23**, 95–101.

Paninski, L. (2003). Estimation of entropy and mutual information, *Neural Comput.*, **15**, 1191–1253, June.