

離散や連続を区別しない相互情報量の推定と、その強一致性について

鈴木 譲

一般に、確率変数 X, Y の相互情報量が 0 であることと、両者が独立であることは同値である。しかしながら、既存の相互情報量の推定量は、常に正の値をとるため、 X, Y が独立であることと、0 の値をとることが一致しない。本研究では、強一致性を満足すること以外に、 X, Y が独立であることと、非正の値をとる (負の値をとることもある) ことが確率 1 で一致するような、相互情報量の推定量を提案する。推定量は、独立である場合とそうでない場合で、BIC (Bayesian Information Criterion) の値の差をとったものである。この BIC は、近年講演者によって、 X, Y が有限個の値をとるとか、正規分布にしたがうとか、特定の仮定を置かないで導き出されたものである。また、提案方式に基づいて、独立性検定を行う際には、帰無仮説をシミュレートして推定量のしきい値を設定するような手間は生じない。また、サンプル数 n に対して、 $O(n \log n)$ の計算量で、推定が完結するので、HSIC (Hilbert Schmidt Independence Criterion、 $O(n^3)$ の計算量) などのカーネルに基づく方法とくらべて、効率がよく、ビッグデータにも適用できる。最後に、R による数値実験を示して、提案方式の実際面での有効性を示す。

キーワード: 相互情報量の推定、ベイズ推定、(強) 一致性、HSIC、BIC、サンプル数に対する効率

