離散や連続を仮定しない情報源のオンライン学習とノンパラメトリック推定の評価

鈴木 譲

大阪大学大学院理学研究科

2010 年 3 月 30 日
統計数理とデーターマイニング研究会

あらまし

研究のねらい

$\{X_i\}_{i=1}^{\infty}$: 定常エルゴード

オンライン予測

X_i が有限の値をとる $\Longrightarrow$ ユニバーサルデータ圧縮

$X^n \sim P^n$ に対して、$\exists Q^n$ s. t.

$$\sum_{x^n} Q^n(x^n) \leq 1$$

$$\frac{1}{n} \sum_{x^n} P^n(x^n) \log \frac{P^n(x^n)}{Q^n(x^n)} \to 0$$

$Q(x_{n+1}|x^n) := \dfrac{Q^{n+1}(x^{n+1})}{Q^n(x^n)}$ が、x^n のもとでの x_{n+1} の予測になる

研究のねらい (続)

オンライン予測

X^n に確率密度関数 f_{X^n} が存在 $\Longrightarrow$ Boris Ryabko 2009

B. Ryabko. “Compression-Based Methods for Nonparametric Prediction and Estimation of Some Characteristics of Time Series”. *IEEE Trans. on Information Theory*, VOL. 55, NO. 9, 2009, pp. 4309-4315.

研究のねらい (続)

オンライン予測

一般の場合 $\Longrightarrow$ 今回の対象

例: $\int_0^\infty g(x)dx = 1$ として、

$$F_X(x) = \begin{cases} 0 & x < -1, \\ \frac{1}{2}, & -1 \leq x < 0 \\ \int_0^x \frac{1}{2} g(t)dt, & 0 \leq x \end{cases}$$

であれば、$F_X(x) = \int_{-\infty}^x f_X(t)dt$ なる f_X (確率密度関数) は存在しない。

一般の定常エルゴードな確率変数 $\{X_n\}_{n=1}^\infty$ の列の場合

- オンライン予測
- パラメータ推定をしないという意味で、ノンパラメトリック推定

有限の値をとるとき

$X^n \sim P^n,\ n = 1, 2, \cdots$ (定常エルゴード)
$A := X_i(\Omega),\ i = 1, \cdots, n$ (有限集合)

$\{X_i\}_{i=1}^{\infty}$ のエントロピー

$$H(P^{\infty}) := \lim_{n\to\infty} -\frac{1}{n} \sum_{x^n \in A^n} P^n(x^n) \log P^n(x^n)$$

$\varphi^n : A^n \to \{0,1\}^*$ が符号化

$x^n, y^n \in A^n$ について、

$$\varphi^n(x^n) = \varphi^n(y^n) \Longrightarrow x^n = y^n$$

有限の値をとるとき (続)

ユニバーサル符号化 $\varphi^n : A^n \to \{0,1\}^*$ が存在

任意の定常エルゴードな P^∞ について、確率 1 で

$$\frac{|\varphi^n(X_1, \cdots, X_n)|}{n} \to H(P^\infty)$$

Shannon-MacMillan-Breiman (有限の場合)

任意の定常エルゴードな P^∞ について、確率 1 で

$$\frac{-\log P^n(x_1, \cdots, x_n)}{n} \to H(P^\infty)$$

$Q^n(x_1, \cdots, x_n) := 2^{-|\varphi^n(x_1, \cdots, x_n)|}$

任意の定常エルゴードな P^∞ について、確率 1 で

$$\frac{1}{n} \log \frac{P^n(x_1, \cdots, x_n)}{Q^n(x_1, \cdots, x_n)} \to 0$$

確率密度関数が存在するとき

$X^n \sim f^n_{X^n},\ n = 1, 2, \cdots$ (定常エルゴード)

$\{X_i\}_{i=1}^{\infty}$ の微分エントロピー

$$h(f^{\infty}) := \lim_{n\to\infty} -\frac{1}{n} \int f^n(x^n) \log f^n(x^n)$$

$\mathcal{B}$: $\mathbb{R}$ の Borel 集合
$\{A_i\}_{i=0}^{\infty}$: 有限集合の列

- A_{i+1} は A_i をさらに分割したもの
- A_i で生成される σ-集合体は、$i \to \infty$ で $\mathcal{B}$ に近づく

$s_i : \mathbb{R} \to A_i$: A_i への射影
$\{s_i(X_j)\}_{j=1}^{\infty}$ も定常エルゴード

確率密度関数が存在するとき (続)

$s_i(X^n) \sim P_i^n$

$$Q_i^n(a_1, \cdots, a_n) := 2^{-|\varphi_i^n(a_1, \cdots, a_n)|}$$

$a_1, \cdots, a_n \in A_i$

$$f_i^n(x_1, \cdots, x_n) := \frac{P_i^n(s_i(x_1), \cdots, s_i(x_n))}{\lambda_i^n(s_i(x_1), \cdots, s_i(x_n))}$$

$$g_i^n(x_1, \cdots, x_n) := \frac{Q_i^n(s_i(x_1), \cdots, s_i(x_n)}{\lambda_i^n(s_i(x_1), \cdots, s_i(x_n))}$$

$x_1, \cdots, x_n \in \mathbb{R}$

$\lambda_i^n(a_1, \cdots, a_n)$: $(a_1, \cdots, a_n) \in A_i^n$ の Lebesgue 測度 (体積)

$\{\omega_i\}_{i=0}^{\infty}$: $\sum_{i=0}^{\infty} \omega_i = 1$, $\omega_i > 0$

$$g^n(x_1, \cdots, x_n) := \sum_{i=0}^{\infty} \omega_i g_i^n(x_1, \cdots, x_n)$$

確率密度関数が存在するとき (続)

Shannon-MacMillan-Breiman (確率密度関数が存在する場合)

任意の定常エルゴードな f^∞ について、確率 1 で

$$\frac{-\log f^n(x_1, \cdots, x_n)}{n} \to h(f^\infty)$$

$i = 0, 1, \cdots$ についても、任意の定常エルゴードな f_i^∞ について、確率 1 で

$$\frac{-\log f_i^n(x_1, \cdots, x_n)}{n} \to h(f_i^\infty)$$

Ryabko, 2009

$h(f_i^\infty) = h(f^\infty)$ $(i \to \infty)$ のとき、
任意の定常エルゴードな f^n について確率 1 で、

$$\frac{1}{n} \log \frac{f^n(x_1, \cdots, x_n)}{g^n(x_1, \cdots, x_n)} \to 0$$

Lebesgue 積分

Ω: 全体集合
$\mathcal{G}$: Ω の σ-集合体

$g : \Omega \to \mathbb{R}$ が $\mathcal{G}$-可測

$$D \in \mathcal{B} \Longrightarrow \{\omega \in \mathcal{G} | X(\omega) \in D\} \in \mathcal{G}$$

ν: 測度

$A \in \mathcal{G}$ 上の ν に関する Lebesgue 積分

$$\int_A g(\omega) d\nu(\omega) := \sup_{\{A_i\}} \sum_i \inf_{\omega \in A_i} g(\omega)\nu(A_i)$$

(sup の $\{A_i\}$ は、A の分割を動く)

Radon-Nykodim の定理

μ, ν: σ-有限の測度

μ が ν について絶対連続 $\mu << \nu$

$$\nu(A) = 0 \Longrightarrow \mu(A) = 0$$

Radon-Nykodim

$\mu << \nu$ のとき、

$$\mu(A) = \int_A g(\omega) d\nu(\omega) \ , A \in \mathcal{G}$$

となる $\mathcal{G}$-可測な $\dfrac{d\mu}{d\nu} := g : \Omega \to \mathbb{R}$ が存在

Kullback-Leibler 情報量

$(\Omega, \mathcal{F}, \mu)$: 確率空間
ν: $\nu(\Omega) \leq 1$

Kullback-Leibler 情報量

$\mu << \nu$ のとき、

$$D(\mu||\nu) := \int_\Omega d\mu(\omega) \log \frac{d\mu}{d\nu}(\omega)$$

Pinsker の不等式

$$\sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)| \leq \sqrt{\frac{2}{\log e} D(\mu||\nu)}$$

$X : \Omega \to \mathbb{R}$ が確率変数

X が $\mathcal{F}$-可測

提案する測度 ν^n

$\{X_n\}_{n=1}^{\infty} \sim \mu^{\infty}$

η^n: 測度 (事前知識を反映、Lebesgue 測度 λ^n でもよい)

$\sum_{i=0}^{\infty} \omega_i = 1,\ \omega_i > 0$

提案する測度 ν^n

$(D_1, \cdots, D_n) \in \mathcal{B}^n$ に対して、

$$\nu_i^n(D_1, \cdots, D_n) := \sum_{a_1, \cdots, a_n \in A_i} \frac{\eta^n(a_1 \cap D_1, \cdots, a_n \cap D_n)}{\eta^n(a_1, \cdots, a_n)} Q_i^n(a_1, \cdots, a_n)$$

$\eta^n = \lambda^n$ として、両辺を λ^n で Radon-Nicodym 微分すると、Ryabko に一致

提案する測度 ν^n (続)

$\{A_i\}_{i=0}^{\infty}$ に関する仮定

:

$$\mu_i^n(D_1, \cdots, D_n) := \sum_{a_1, \cdots, a_n \in A_i} \frac{\eta^n(a_1 \cap D_1, \cdots, a_n \cap D_n)}{\eta^n(a_1, \cdots, a_n)} P_i^n(a_1, \cdots, a_n)$$

$$\lim_{i \to \infty} \lim_{n \to \infty} \frac{1}{n} \log \frac{d\mu^n}{d\mu_i^n}(x_1, \cdots, x_n) = 0$$

証明したこと (その 1)

定理 1

任意の定常エルゴードな μ^∞ に対して、確率 1 で

$$\frac{1}{n}\log\frac{d\mu^n}{d\nu^n}(x_1,\cdots,x_n)\to 0$$

を満足する $\{i(n)\}_{n=1}^\infty$ が存在

系:

$$\frac{1}{n}D(\mu^n||\nu^n)=\frac{1}{n}\int d\mu^n\log\frac{d\mu^n}{d\nu^n}=E[\log\frac{d\mu^n}{d\nu^n}(x_1,\cdots,x_n)]\to 0$$

定理１の適用例

例題 1: $\Omega := [0,1)$, $\eta = \lambda$　(仮定を満足)

- $A_0 := \{[0,1/2), [1/2,1)\}$
- $A_1 := \{[0,1/4), [1/4,1/2), [1/2,3/4), [3/4,1)\}$
- $\cdots$

例題 2: $\Omega := \mathbb{N} = \{1,2,\cdots\}$, $\eta(j) = \dfrac{1}{j} - \dfrac{1}{j+1}$, $j \in \mathbb{N}$　(仮定を満足)

- $A_0 := \{\{1\}, \mathbb{N} - \{1\}\}$
- $A_1 := \{\{1\}, \{2\}, \mathbb{N} - \{1,2\}\}$
- $\cdots$

例題 3: $\Omega := \mathbb{R}$, $\eta = \lambda$　(仮定を満足しない)

- $A_0 := \{[0,1/2), [1/2,1)\}$
- $A_1 := \{[0,1/4), [1/4,1/2), [1/2,3/4), [3/4,1)\}$
- $\cdots$

証明したこと (その 2)

$r : \mathbb{R} \to \mathbb{R}$: $\mathcal{F}$-可測な有界関数

定理 2

任意の定常エルゴードな μ^∞ について、確率 1 で

$$\frac{1}{n} E \sum_{j=0}^{n-1} \{ \int r(x) d\mu^{j|j-1}(x|x_1, \cdots, x_{j-1}) - \int r(x) d\nu^{j|j-1}(x|x_1, \cdots, x_{j-1}) \}^2 \to 0$$

$$\frac{1}{n} E \sum_{j=0}^{n-1} | \int r(x) d\mu^{j|j-1}(x|x_1, \cdots, x_{j-1}) - \int r(x) d\nu^{j|j-1}(x|x_1, \cdots, x_{j-1}) | \to 0$$

スケッチ: Pinsker の不等式を用いる

$$
\begin{aligned}
& \{\int r(x) d\mu^{j|j-1}(x|x_1,\cdots,x_{j-1}) - \int r(x) d\nu^{j|j-1}(x|x_1,\cdots,x_{j-1})\}^2 \\
\leq\ & b^2\{\int d\mu^{j|j-1}(x|x_1,\cdots,x_{j-1}) - \int d\nu^{j|j-1}(x|x_1,\cdots,x_{j-1})\}^2 \\
\leq\ & \frac{2b^2}{\log e}\{\sup_A |\int_A d\mu^{j|j-1}(x|x_1,\cdots,x_{j-1}) - \int_A d\nu^{j|j-1}(x|x_1,\cdots,x_{j-1})|\}^2 \\
\leq\ & \frac{2b^2}{\log e}\int d\mu^{j|j-1}(x|x_1,\cdots,x_{j-1}) \log \frac{d\mu^{j|j-1}}{d\nu^{j|j-1}}(x|x_1,\cdots,x_{j-1}) \ ,
\end{aligned}
$$

オンライン予測の実際の適用

Radon-Nykodim の定理と、$\nu^{n+1} << \nu^n$ より、

$$\nu^{n+1}(D^{n+1}) = \int_{D^n} \eta(D_{n+1}|x^n) d\nu^n(x^n)$$

なる条件付確率測度 η が存在。
x^n のもとでの、x_{n+1} の条件付測度の計算: 十分小さな $\epsilon > 0$ について、

$$\begin{aligned} \eta(D_{n+1}|x^n) &= \frac{\nu^{n+1}(dx_1, \cdots, dx_n, D_{n+1})}{\nu^n(dx_1, \cdots, dx_n)} \\ &\approx \frac{\nu^{n+1}(x_1 - \epsilon, x_1 + \epsilon, \cdots, x_n - \epsilon, x_n + \epsilon, D_{n+1})}{\nu^n(x_1 - \epsilon, x_1 + \epsilon, \cdots, x_n - \epsilon, x_n + \epsilon)} \end{aligned}$$

まとめ

定常エルゴード情報源におけるノンパラメトリック推定とオンライン予測

一般の測度の場合について、漸近的な最良性 (定理 1, 定理 2) を証明した。

- 仮定も結論も、Ryabko の方法の一般化
- 予測の際の条件付確率の計算で近似を伴う
- 事前知識を考慮しないと、有限の n での学習測度は保証できない

今回提案した測度に対応する SBM

Krystina Ziemian (Warszawa, 1989)