# 離散や連続を仮定しないユニバーサル符号化と一般的な Shannon-MacMillan-Breiman 定理

鈴木 譲

大阪大学大学院理学研究科

2010 年 5 月 21 日
情報理論研究会

## あらまし

# 研究のねらい: ユニバーサル圧縮

$\{X_i\}_{i=1}^{\infty}$: 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: Boris Ryabko

**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: 安易に離散や連続を仮定するな

## 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、

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

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 $P^\infty$ 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 $P^\infty$ 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 $P^\infty$ 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 $\sigma$-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

例: $\Omega := [0, 1)$

- $A_0 := \{[0, 1/2), [1/2, 1)\}$
  $s_0(1/4) = [0, 1/2)$, $s_0(3/8) = [1/2, 1)$

- $A_1 := \{[0, 1/4), [1/4, 1/2), [1/2, 3/4), [3/4, 1)\}$
  $s_1(0) = [0, 1/4]$, $s_1(0.8) = [3/4, 1)$
- $\cdots$

$i = 0, 1, \cdots$ とともに、有限集合の列 $\{A_i\}$ が細分化されていく

# 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 $f^\infty$ 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 $f_i^\infty$ 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

$\Omega$: 全体集合
$\mathcal{G}$: $\Omega$ の $\sigma$-集合体

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

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

$\nu$: 測度

**$A\in\mathcal{G}$ 上の $\nu$ 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

$\mu, \nu$: $\sigma$-有限の測度

**$\mu$ が $\nu$ 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

# Radon-Nykodim の定理の適用例

$X$: 確率変数
$\mu_X(-\infty, x] = F_X(x)$
$\lambda_X(D) := |D|$ (Lebesgue 測度)

**確率密度関数 $f_X$ が存在: $F_X(x) = \int_0^x f_X(t)dt$**

$$\mu_X << \lambda_X$$

$$f_X(x) = \frac{d\mu_X}{d\nu_X}(x)$$

# Kullback-Leibler 情報量

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

## 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 $\nu^n$

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

$\eta^n$: 測度 (事前知識を反映、Lebesgue 測度 $\lambda^n$ でもよい)　$\mu^n << \eta^n$

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

**提案する測度 $\nu^n$**

$(D_1, \cdots, D_n) \in \mathcal{B}^n$ 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 $\lambda^n$ で Radon-Nicodym 微分すると、Ryabko に一致

# 提案する測度 $\nu^n$ (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 $\mu^\infty$ に対して、確率 1 で

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

**系**

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

# A. Barron の Shannon-MacMillan-Breiman (1985)

$\mu << \eta$ (B. Ryabko の場合、Lesbeg$\lambda$)
$\mu$: 任意の定常エルゴード
確率 1 で

$$-\frac{1}{n}\log\frac{d\nu^n}{d\eta^n} \to D(\nu||\eta)$$

**Shannon-MacMillan-Breiman (A. Barron の一般的なバージョン)**

確率 1 で

$$-\frac{1}{n}\log\frac{d\mu^n}{d\eta^n} \to D(\mu||\eta)$$

# 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 $\mu^\infty$ 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

# 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