

区間定常無記憶情報源 (PSMS) の学習アルゴリズム

金澤 宏紀*

Hiroki Kanazawa

山西 健司*

Kenji Yamanishi

Abstract: 非定常情報源の一種として区間定常無記憶情報源 (PSMS) が存在する。これは、定常区間の変化を潜在変数とする Latent Dynamics の 1 モデルと考えられる。情報理論の文脈で Merhav は PSMS に対する符号長期待値の下限を示した。この研究を基に Willems や Shamir and Merhav 等により PSMS に対する符号化アルゴリズムが提案されてきた。一方 Kleinberg はテキストマイニングの文脈でテキストストリームから特定の単語が頻出するバースト区間を推定するアルゴリズムを構成した。Kanazawa and Yamanishi は Kleinberg のアルゴリズムを MDL 原理に基づいて PSMS 学習向けに拡張し、Merhav の限界を達成するアルゴリズムを提案した。本発表では Kanazawa and Yamanishi のアルゴリズムと他手法との比較を中心に、PSMS 学習の最新動向を紹介する。

Keywords: 区間定常無記憶情報源, 記述長最小原理 (MDL 原理), 動的モデル選択, piecewise stationary memoryless source (PSMS), minimum description length principle (MDL principle), dynamic model selection (DMS)

1 まえがき

時間変化する情報源の変化検知は応用上極めて重要なトピックである。本研究では、記述長最小原理 (MDL 原理) 特に動的モデル選択 (dynamic model selection, DMS) の観点からこの問題を考える。すなわち情報源が時間変化するモデル列を考え、データ列を「モデル列が与えられたときのデータ記述長」と「モデル列自身の記述長」の和で表したときこの総記述長を最小にするモデルを選ぶことにより、情報源がいつどの程度変化したかの推定を行う。

MDL 原理は情報量規準・機械学習のヒューリスティクスの 1 つとして Rissanen[7] によって提案されたモデル選択規準である。Yamanishi and Maruyama[13], [14] は MDL 原理を非定常情報源に拡張し、単一のモデルを選択するのではなく、モデル間に遷移確率構造を入れてモデル列選択する DMS を提案した。DMS は Latent Dynamics を MDL 原理の観点から捉える上で重要な概念である。

Merhav[6] は非定常情報源の一種として、パラメータが区分的に変化する区間定常無記憶情報源 (piecewise stationary memoryless source, PSMS) を提案し、さら

に有限離散分布に対する PSMS の記述長下限を示した。これは Rissanen[7] の定常無記憶情報源に対する記述長下限の拡張になっている。Merhav[6] は逐次 Bayes 推定とリセット確率を用いた計算複雑性の高いアルゴリズムにより、この下限を漸近的に達成できることを示した。Willems[12] はリセット確率の推定量を適切に定義し時間計算量 $O(n^2)$ の PSMS 符号化アルゴリズムを提案した、ここで n はサンプル数である。しかしこれは Merhav の下限を達成するものではなかった。Shamir and Merhav[11] は Willems[12] の推定量を拡張し、時間計算量 $O(n^2)$ で理論上 Merhav の下限を漸近的に達成する。PSMS 符号化アルゴリズムを構成した。また、逐次推定に逐次正規化最尤推定 (SNML 推定) を用い、これら手法をモデル選択に適用した研究も Sakurai and Yamanishi[10] によってなされている。本稿では、このモデル選択アルゴリズムの枠組みを手法 1 として紹介する。

一方データマイニングの分野で、Kleinberg[4] は時系列テキストデータから特定の単語が頻出する時期—バースト区間—を推定するアルゴリズムを提案した。これは単語の出現確率を 1 次元パラメトリック確率分布で表したとき、パラメータ空間を離散化してそれぞれを 1 つの状態とみなし、与えられたコスト関数を最小にする状態列を動的計画法 (DP) で求めることにより、値の大きいパラメータを持つ区間を推定するというアルゴリズム

*東京大学大学院 情報理工学系研究科, 〒 113-0033 東京都文京区本郷 7-3-1, {hiroki_kanazawa, yamanishi}@mist.i.u-tokyo.ac.jp, Graduate School of Information Science and Technology, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo, 113-0033 Japan.

である。Kleinberg アルゴリズムを基に Kanazawa and Yamanishi[3] は MDL 原理に基づく 1 次元パラメータでの変化検出アルゴリズムを提案した。離散化手法と離散化点数および動的計画法の遷移確率を適切に定めることにより、アルゴリズムの期待冗長度の上限が Merhav[6] の下限に漸近的に一致することを示した。また、金澤・山西 [15] ではパラメータ空間の基底が直交する条件のもとで、多次元パラメータ空間に対する離散化手法の拡張がなされている。本稿では、この手法を手法 2 として紹介する。

本稿を通して対数の底は e とする。

2 問題設定

データ列を $x^n \stackrel{\text{def}}{=} x_1^n \stackrel{\text{def}}{=} x_1 \dots x_n$ とする。各データ x_t ($t = 1, \dots, n$) は \mathcal{X} に属し、 \mathcal{X} は離散値あるいは連続値を取るものとする。またデータ部分列を $x_{t_1}^{t_2} \stackrel{\text{def}}{=} x_{t_1} \dots x_{t_2}$ と表記する。 \mathcal{X} 上で定義される確率分布 $f(x; \theta)$ のパラメトリッククラスを

$$\mathcal{F} = \{f(x; \theta) : \theta \in \Theta\} \text{ for } x \in \mathcal{X},$$

と定める。ここで Θ を k 次元のコンパクトなパラメータ空間とする。また $f(x; \theta)$ は任意の $\theta \in \Theta$ に関して十分解析的であるとする。

2.1 区間定常無記憶情報源 (PSMS)

定義 1. パラメトリッククラス \mathcal{F} において区間定常無記憶情報源 (piecewise stationary memoryless source; PSMS) を次のように定義する：

データ列 x^n は、各 x_t が独立に確率分布

$$\begin{cases} x_t \sim f(x; \theta(0)) & (1 \leq t \leq m(1)), \\ x_t \sim f(x; \theta(1)) & (m(1) + 1 \leq t \leq m(2)), \\ \vdots \\ x_t \sim f(x; \theta(c)) & (m(c) + 1 \leq t \leq n), \end{cases} \quad (1)$$

から発生しているとする。ここで $0 < m(1) < m(2) < \dots < m(c) < n$ を変化点とする。また $m(0) = 0$, $m(c+1) = n$ と定義する。この c , $\theta(p)$ ($p = 0, \dots, c$), $m(p)$ ($p = 1, \dots, c$) で特徴付けられるデータ列 x^n の発生確率分布 (1) を PSMS とよぶ。事前に変化回数 c についての情報は無いものとする。

なお、 c は n によらない定数とし、漸近解析を行う際 n を大きくしていく一方で c は固定とする。すなわち各区間の長さは線形に大きくなる。

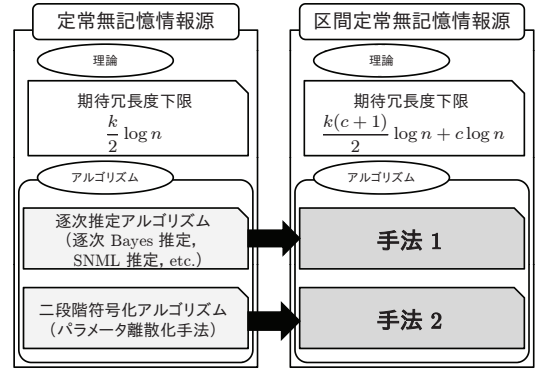


図 1: PSMS 変化検出手法周辺の関係図。

2.2 PSMS の期待冗長度下限

定義 2 (期待冗長度). PSMS に対しての無歪符号化アルゴリズム \mathcal{A} に対し、 \mathcal{A} によるデータ列 x^n の記述長を $\mathcal{L}_{\mathcal{A}}(x^n)$ とする。このとき \mathcal{A} の期待冗長度 $\mathcal{R}_{\mathcal{A}}^{(n)}$ を

$$\mathcal{R}_{\mathcal{A}}^{(n)} \stackrel{\text{def}}{=} \mathbb{E} \left[\mathcal{L}_{\mathcal{A}}(x^n) - \sum_{p=0}^c \sum_{t=m(p)+1}^{m(p+1)} (-\log f(x_t; \theta(p))) \right],$$

と定義する。ここで期待値はデータ列の確率分布 (1) に対してとる。

PSMS の期待冗長度に対し Merhav[6] は下限を示した。

定理 1 (Merhav[6]). 定義域 \mathcal{X} を有限集合とする。その他いくつかの仮定をおく (詳細は Merhav[6] 参照のこと)。このとき、ほとんどすべての PSMS において、任意の $\varepsilon > 0$ と十分大きな n に対し期待冗長度の下限は

$$\inf_{\mathcal{A}} \mathcal{R}_{\mathcal{A}}^{(n)} \geq (1 - \varepsilon) \left(\frac{k(c+1)}{2} \log n + c \log n \right), \quad (2)$$

となる。ここで $\inf_{\mathcal{A}}$ はすべての無歪符号化アルゴリズムを考え、その下限を取る。

式 (2) の下限は 2 つの項の和で表されている。右辺第 1 項は、 $c+1$ 個のパラメータに関して各パラメータ $(k/2) \log n$ の記述長が必要であることを示している (Rissanen[7] 参照)。右辺第 2 項は c 個の変化点に関して各変化点 $\log n$ の記述長が必要であることを示している。

3 PSMS の変化検出アルゴリズム

PSMS の変化検出として 2 種類のアプローチを紹介する。1 つ目は逐次推定とリセット確率を用いるアプローチであり、2 つ目はパラメータ空間離散化と動的計画法を用いるアプローチである。本稿ではそれぞれを手法 1, 手法 2 とよぶ。手法 1 と手法 2 を定常情報源の学習手法と対比させた図が図 1 となる。

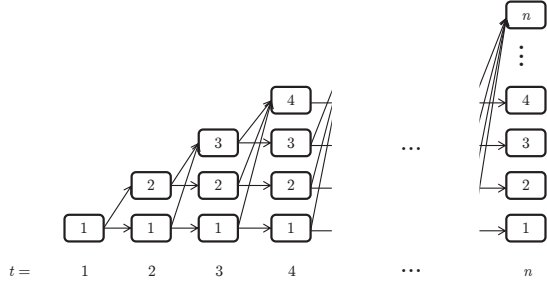


図 2: 手法 1 の状態遷移図.

3.1 手法 1

手法 1 の基本的なアイデアは逐次推定とリセット確率である. また最適なモデルを選択する際に動的計画法を用いる.

状態遷移図として図 2 を考える. 時刻 t における状態 s_t は $1, \dots, t$ までの値を取る. $s_t = 1, \dots, t-1$ はそれまでの逐次推定を継続し, $s_t = t$ は推定をリセットする. すなわち時刻 $t-1$ の状態 s_{t-1} からの遷移は自身か t への 2 通りとなる.

遷移の確率および逐次推定を以下のように表記する:

$$P_{\mathcal{A}}(s_t | s_{t-1}) = \begin{cases} P_{\text{reset}}(t | s_{t-1}) & s_t = t, \\ 1 - P_{\text{reset}}(t | s_{t-1}) & s_t = s_{t-1}, \end{cases} \quad (3)$$

$$P_{\mathcal{A}}(x_t | s_t) = \begin{cases} P_{\text{update}}(x_t | \emptyset) & s_t = t, \\ P_{\text{update}}(x_t | x_{s_t}^{t-1}) & s_t = s_{t-1}. \end{cases} \quad (4)$$

$P_{\text{reset}}(t | s_{t-1})$ は Willems 推定量 [12]

$$P_{\text{reset}}(t | s_{t-1}) = \frac{1/2}{t - s_{t-1}}, \quad (5)$$

や, Shamir and Merhav 推定量 [11]: ある $\varepsilon > 0$ に対し

$$\pi(j) = 1/j^{1+\varepsilon}, Z_{\infty} = \sum_{j=1}^{\infty} \pi(j), Z_t = \sum_{j=1}^t \pi(j), \quad (6)$$

$$P_{\text{reset}}(t | s_{t-1}) = \frac{\pi(t - s_t)}{Z_{\infty} - Z_{t-s_{t-1}-1}},$$

などがある.

逐次推定は代表的には逐次 Bayes 推定

$$P_{\text{update}}(x | \emptyset) = f(x; \theta_0),$$

$$P_{\text{update}}(x | x_{t_1}^{t_2}) = \frac{\int_{\theta \in \Theta} p(\theta) f(x; \theta) \prod_{t=t_1}^{t_2} f(x_t; \theta) d\theta}{\int_{\theta \in \Theta} p(\theta) \prod_{t=t_1}^{t_2} f(x_t; \theta) d\theta}, \quad (7)$$

が存在する. 期待冗長度を小さくするには事前分布 $p(\theta)$ に Jeffreys の事前分布

$$I(\theta) \stackrel{\text{def}}{=} \mathbb{E}_{\theta}[\nabla_{\theta}^2(-\log f(x; \theta))], \quad (8)$$

$$p_J(\theta) \stackrel{\text{def}}{=} \frac{\sqrt{|I(\theta)|}}{\int_{\theta \in \Theta} \sqrt{|I(\theta)|} d\theta},$$

が用いられる, ここで ∇_{θ}^2 は θ に関する Hesse 行列であ

る. \mathcal{X} が有限離散集合のとき, Jeffreys 事前分布を用いた逐次 Bayes 推定は, Krichevsky and Trofimov 推定 [5]

$$P_{\text{update}}(x | \emptyset) = 1/|\mathcal{X}|,$$

$$P_{\text{update}}(x = a | x_{t_1}^{t_2}) = \frac{N_a(x_{t_1}^{t_2}) + 1/2}{t_2 - t_1 + |\mathcal{X}|/2},$$

として知られている, ここで $N_a(x_{t_1}^{t_2})$ は $x_{t_1}^{t_2}$ 中に出現する a の回数を示す.

式 (3), (4) を用いると, 手法 1 の記述長は

$$\mathcal{L}_{\mathcal{A}}(x^n) = \min_{s^n} \mathcal{L}_{\mathcal{A}}(x^n, s^n),$$

$$\mathcal{L}_{\mathcal{A}}(x^n, s^n) = \sum_{t=1}^n (-\log P_{\mathcal{A}}(x_t | s_t)) + \sum_{t=1}^n (-\log P_{\mathcal{A}}(s_t | s_{t-1})),$$

となる, ここで $P_{\mathcal{A}}(s_1 | s_0) \stackrel{\text{def}}{=} 1$ とする. $\mathcal{L}(x^n, s^n)$ を最小とする状態列 $s^n = s_1 \dots s_n$ を得ることにより, データ列の変化検出を行うことができる. このとき, 動的計画法により, 各 P_{update} にかかる計算時間が $O(k)$ であるならば, 総時間計算量 $O(kn^2)$ で最適状態列 s^n 求めることができる, ここで k はパラメータの次元である.

逐次推定の方法とリセット確率の推定量を定めると, 手法 1 はアルゴリズムとして一意に定まる. すなわちアルゴリズムの性能は逐次推定とリセット確率に依存する.

手法 1 の一般的な期待冗長度上限評価の概略を以下に示す.

期待冗長度評価 図の状態遷移を動的計画法で解くので, 手法 1 の記述長はある状態列を考えたときの記述長以下となる. \tilde{s}^n を $\tilde{s}_{m(p)+1} = \dots = \tilde{s}_{m(p+1)} = m(p) + 1$ ($p = 0, \dots, c$) と定めると $\mathcal{L}_{\mathcal{A}}(x^n) \leq \mathcal{L}_{\mathcal{A}}(x^n, \tilde{s}^n)$ となるので

$$\mathcal{R}_{\mathcal{A}}^{(n)} \leq \mathbb{E} \left[\sum_{t=1}^n (-\log P_{\mathcal{A}}(x_t | \tilde{s}_t)) + \sum_{t=1}^n (-\log P_{\mathcal{A}}(\tilde{s}_t | \tilde{s}_{t-1})) - \sum_{p=0}^c \sum_{t=m(p)+1}^{m(p+1)} (-\log f(x_t; \theta(p))) \right]$$

$$= \sum_{p=0}^c \mathbb{E}_{\theta(p)} \left[\sum_{t=m(p)+1}^{m(p+1)} \left((-\log P_{\mathcal{A}}(x_t | m(p) + 1)) - (-\log f(x_t; \theta(p))) \right) + \sum_{t=1}^n (-\log P_{\mathcal{A}}(\tilde{s}_t | \tilde{s}_{t-1})) \right]$$

が成立する. ここから, 各項を評価すればよいが, 詳細は省略する. 逐次推定に Jeffreys 事前分布 (8) の逐次 Bayes 推定 (7) を用い, リセット確率に Shamir and

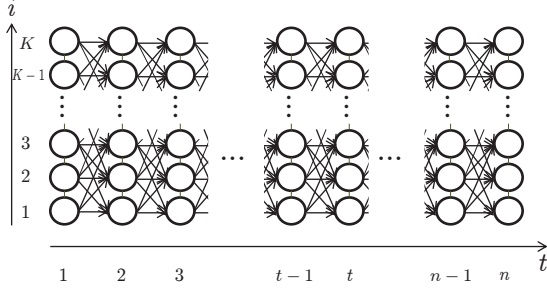


図 3: 手法 2 の状態遷移図.

Merhav 推定量 (6) を用いた場合

$$\begin{aligned} \mathcal{R}_A^{(n)} \leq & \frac{k(c+1)}{2} \log \frac{n}{2\pi ec} + c \log \frac{n}{c} \\ & + (c+1)\varepsilon \log \frac{n}{c} + c \log \int_{\theta \in \Theta} \sqrt{|I(\theta)|} d\theta \quad (9) \\ & + (c+1) \log(1+\varepsilon) - c \log \varepsilon + o(1), \end{aligned}$$

となる (Clarke and Barron[2], Shamir and Merhav[11] 参照), ここで $o(1)$ は $n \rightarrow \infty$ で零となる項である. この $\mathcal{R}_A^{(n)}$ 上限は $\varepsilon \rightarrow 0, n \rightarrow \infty$ の極限で Merhav の下限 (2) に一致する.

一方, 逐次推定に Jeffreys 事前分布の逐次 Bayes 推定を用い, リセット確率に Willems 推定量 (5) を用いると, 式 (9) で $\varepsilon = 0.5$ とおいた場合に相当することが知られており (Shamir and Merhav[11] 参照), Merhav の下限 (2) を達成しない.

3.2 手法 2

手法 2 の基本的なアイデアはパラメータ離散化による二段階符号化と動的計画法である. 本稿では 1 次元パラメータに対する手法を扱う.

パラメータ空間 Θ を K 点に離散化し各離散化点を 1 つの状態と考える. 離散化点を $\bar{\theta}_1, \dots, \bar{\theta}_K \in \Theta$ と表記する. 状態間に遷移確率を導入して状態の変化を表し, DMS を適用し最も総記述長が小さくなる状態列を選ぶ. 手法 2 は手法 1 に比べて, パラメータの変化を直接探索しているという点において, より Latent Dynamics に迫る手法といえる.

手法 2 の状態遷移図として図 3 を考える. 状態列の初期確率を $P_A(i_1)$, 遷移確率を $P_A(i_t | i_{t-1})$ と記す. このとき手法 2 の記述長は次のようになる:

$$\begin{aligned} \mathcal{L}_A(x^n) &= \min_{i_1^n} \mathcal{L}_A(x^n, i_1^n), \\ \mathcal{L}_A(x^n, i_1^n) &= \sum_{t=1}^n (-\log f(x_t; \bar{\theta}_{i_t})) \\ &\quad + \sum_{t=1}^n (-\log P_A(i_t | i_{t-1})), \end{aligned}$$

ここで $P_A(i_1 | i_0) \stackrel{\text{def}}{=} P_A(i_1)$ である.

手法 2 の性能は離散化手法と離散化点数および遷移確率の設定による. この 3 点を適切に定義することにより, 手法 2 の期待冗長度は Merhav の下限を達成する. 以下, 離散化手法と離散化点数および遷移確率それぞれについて, Kanazawa and Yamanishi[3] の内容を紹介する.

1) 離散化手法: コンパクトな 1 次元パラメータ空間 $\Theta = [\theta_{\min}, \theta_{\max}]$ を K 個の有限状態に離散化する. \mathcal{F} に関する Fisher 情報量 $I(\theta)$ と L_I をそれぞれ

$$I(\theta) \stackrel{\text{def}}{=} \mathbb{E}_\theta \left[-\frac{\partial^2 \log f(x; \theta)}{\partial \theta^2} \right], \quad L_I \stackrel{\text{def}}{=} \int_{\theta_{\min}}^{\theta_{\max}} \sqrt{I(\theta)} d\theta,$$

と定義する. $d\sqrt{I(\theta)}/d\theta \leq 0$ または $d\sqrt{I(\theta)}/d\theta \geq 0$ が離散化幅を $\delta_I \stackrel{\text{def}}{=} L_I/(K-1)$ と定義し, 離散化点 $\bar{\theta}_{\kappa+1}, \dots, \bar{\theta}_K$ を

$$\int_{\theta_{\min}}^{\bar{\theta}_i} \sqrt{I(\theta)} d\theta = (i-1)\delta_I \quad (i=1, \dots, K), \quad (10)$$

と定める. ただしこのとき, $\bar{\theta}_i \leq \theta \leq \bar{\theta}_{i+1}$ ($i=1, \dots, K-1$) の区間において $d\sqrt{I(\theta)}/d\theta \leq 0$ または $d\sqrt{I(\theta)}/d\theta \geq 0$ が成立するよう離散化をする. 各離散化点を 1 つの状態 i と考える.

2) 離散化点数: 離散化総点数 K を

$$K = \lfloor \sqrt{n} \rfloor \quad (11)$$

と定める.

3) 遷移確率: $P_A(i_1)$ と $P_A(i_t | i_{t-1})$ を以下の形で与える:

$$\begin{aligned} P_A(i_1) &= 1/K, \\ P_A(i_t | i_{t-1}) &= \begin{cases} \alpha/(K-1) & (i_t \neq i_{t-1}), \\ 1-\alpha & (i_t = i_{t-1}). \end{cases} \quad (12) \end{aligned}$$

これは, 状態が変化する記述長は状態が変化しない記述長より $\log(K-1) + \log((1-\alpha)/\alpha)$ 大きく, 状態が変化するには空間的に一様な確率で変化することを意味する. いま遷移パラメータ α を $\alpha = 1/n$ と定める.

手法 2 の期待冗長度上限評価の概略を以下に示す.

期待冗長度評価 手法 1 と同様に, ある状態列を考えその記述長を上限とする. 状態列 \tilde{i}^n を $\bar{\theta}_{i_{m(p)+1}} = \dots = \bar{\theta}_{i_{m(p+1)}} = \bar{\theta}(p)$ ($p=0, \dots, c$) ここで $\bar{\theta}(p)$ は

$$\bar{\theta}(p) = \underset{\bar{\theta}_i}{\operatorname{argmin}} |\bar{\theta}_i - \theta(p)| \quad (13)$$

を満たす離散化点である. このとき, $\mathcal{L}(x^n) \leq \mathcal{L}(x^n, \tilde{i}^n)$

となるので、手法2の期待冗長度は

$$\begin{aligned} \mathcal{R}_A^{(n)} &\leq \mathbb{E} \left[\sum_{t=1}^n (-\log f(x_t; \bar{\theta}_{i_t})) \right. \\ &\quad + \sum_{t=1}^n (-\log P_A(\tilde{i}_t | \tilde{i}_{t-1})) \\ &\quad \left. - \sum_{p=0}^c \sum_{t=m(p)+1}^{m(p+1)} (-\log f(x; \theta(p))) \right] \\ &= \sum_{p=0}^c \sum_{t=m(p)+1}^{m(p+1)} D(\theta(p) \| \bar{\theta}(p)) \\ &\quad + \sum_{t=1}^n (-\log P_A(\tilde{i}_t | \tilde{i}_{t-1})), \quad (14) \end{aligned}$$

と上から抑えられる、ここで $D(\cdot \| \cdot)$ は Kullback-Leibler ダイバージェンス

$$D(\theta \| \bar{\theta}) \stackrel{\text{def}}{=} \mathbb{E}_\theta [\log(f(x; \theta) / f(x; \bar{\theta}))]$$

である。 $f(x; \bar{\theta}(p))$ を $\theta(p)$ まわりで3次 Taylor 展開することにより

$$\begin{aligned} D(\theta(p) \| \bar{\theta}(p)) &= \frac{I(\theta(p))}{2} (\bar{\theta}(p) - \theta(p))^2 \\ &\quad + O(|\bar{\theta}(p) - \theta(p)|^3), \end{aligned}$$

を得る。

いま、式(10)の離散化手法を用いると次の補題が成立する。

補題 1 (Kanazawa and Yamanishi[3]). 式(10)の離散化法のもとで、式(13)のように離散化点を取ると

$$\left| \sqrt{I(\theta(p))} (\bar{\theta}(p) - \theta(p)) \right| < \delta_I = \frac{L_I}{K-1}, \quad (15)$$

が成立する。

最後に、式(12)の初期確率および遷移確率を用いると、 i^n の変化回数は c 以下であるので

$$\begin{aligned} &\sum_{t=1}^n (-\log P_A(i_t | i_{t-1})) \\ &\leq \log K + c \log(K-1) \\ &\quad + c \log \frac{1-\alpha}{\alpha} - (n-1) \log(1-\alpha) \\ &= \log K + c \log(K-1) \\ &\quad + c \log(n-1) - (n-1) \log(1-1/n) \\ &< (c+1) \log K + c \log n + \log e \quad (16) \end{aligned}$$

となる。

式(11), (15), (16)から、手法2の期待冗長度(14)は

$$\mathcal{R}_A^{(n)} < \frac{c+1}{2} \log n + c \log n + \frac{L_I^2}{2} + \log e + O(n^{-1/2}), \quad (17)$$

という上限を持つ。 $L_I^2/2 + \log e$ は定数であるため、式(17)の上限は、 $n \rightarrow \infty$ の極限で $k=1$ の場合の Merhav の

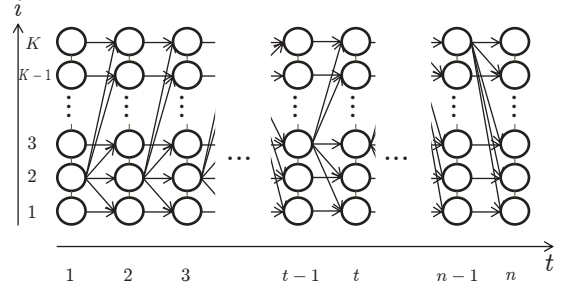


図 4: 一様遷移確率下での手法2の状態遷移図。

下限(2)と一致する。

最後に手法2の時間計算量について述べる。離散化点数を式(11)と設定し、図3の状態遷移を単純に実装した場合、時間計算量は $O(nK^2) = O(n^2)$ となる。ここで式(12)の遷移確率を用いると図4のように各時刻 t の遷移は、時刻 $t-1$ で記述長が最小となる状態から変化する遷移または変化しない遷移のみ考えればよいので、時間計算量は $O(nK) = O(n^{3/2})$ となる。

4 おわりに

本稿では、まずはじめに区間定常無記憶情報源 (PSMS) を定義し、その期待冗長度について Merhav の下限を紹介した。その後 PSMS に対する変化検出手法を2種類紹介した。手法1は逐次推定とリセット確率を用いる手法である。逐次推定に Jeffreys 事前分布を使った逐次 Bayes 推定を、リセット確率に Shamir and Merhav 推定量を用いた場合、期待冗長度が漸近的に Merhav の下限に一致し、時間計算量が $O(kn^2)$ となることを紹介した。手法2はパラメータ空間離散化と動的計画法を用いる手法である。パラメータ空間が1次元の場合、Fisher 情報量に基づく離散化をし、空間的に一様な状態遷移確率を用いることにより、期待冗長度が漸近的に Merhav の下限に一致し、時間計算量が $O(nK) = O(n^{3/2})$ となることを紹介した。

最後に、本稿より進んだ内容を簡単に紹介する。

手法1に関連して: Willems [12] や Sakurai and Yamanishi [10] ではリセット確率に Krichevsky and Trofimov 推定量を用いた手法を提案している。これは図2の範囲に入らず、手法1より計算複雑性が $O(n^3)$ に上がる手法となる。一方 Shamir and Merhav[11] では、Merhav の下限を達成しないが時間計算量を $O(n)$ に削減した手法を提案している。

手法2に関連して: 山澤・山西 [15] では直交する k 次元パラメータ空間に対する離散化が提案されている。このアルゴリズムは、離散化総点数 $K = n^{k/2} + o(n^{k/2})$ となるため時間計算量は $O(nK) = O(n^{k/2+1})$ となる。

Fisher 情報行列を用いた離散化は, Balasubramanian[1] や Rissanen[8] など提案されている “distinguishability” という概念と関わりが深い. 一方で Rissanen[9] では最大容量分割 (maximum capacity partition) が提案されている. これは期待冗長度ではなく, データ列に対するミニマックス冗長度を対象とした離散化といえる.

謝辞 本研究の一部は, 科研費基盤研究 23240019(A), NTT によって助成されたものである. また, 本研究の一部は, 総合科学技術会議により制度設計された最先端研究開発支援プログラム (FIRST 合原最先端数理モデルプロジェクト) により, 日本学術振興会を通して助成されたものである.

参考文献

- [1] V. Balasubramanian, “Statistical inference, Occam’s razor and statistical mechanics on the space of probability distributions,” *Neural Comput.*, vol. 9, pp. 349–368, Feb. 1997.
- [2] B. S. Clarke and A. R. Barron, “Information-theoretic asymptotics of Bayes methods,” *IEEE Trans. Inform. Theory*, vol. 36, pp. 453–471, May 1990.
- [3] H. Kanazawa and K. Yamanishi, “An MDL-based change-detection algorithm with its applications to learning piecewise stationary memoryless sources,” In *Proc. of the 2012 IEEE Inform. Theory Workshop (ITW)*, Lausanne, Switzerland, Sept. 3–7 2012, pp. 562–566.
- [4] J. Kleinberg, “Bursty and hierarchical structure in streams,” *D. M. K. D.*, vol. 7, pp. 373–397, Nov. 2003.
- [5] R. E. Krichevsky and V. K. Trofimov, “The performance of universal encoding,” *IEEE Trans. Inform. Theory*, vol. 27, pp. 199–207, Mar. 1981.
- [6] N. Merhav, “On the minimum description length principle for sources with piecewise constant parameters,” *IEEE Trans. Inform. Theory*, vol. 39, pp. 1962–1967, Nov. 1993.
- [7] J. Rissanen, “Universal coding, information, prediction, and estimation,” *IEEE Trans. Inform. Theory*, vol. 30, pp. 629–636, July 1984.
- [8] J. Rissanen, *Information and complexity in statistical modeling*. Springer, New York, 2007.
- [9] J. Rissanen, *Optimal Estimation of Parameters*. Cambridge University Press, Cambridge, 2012.
- [10] E. Sakurai and K. Yamanishi, “Comparison of dynamic model selection with infinite HMM for statistical model change detection,” In *Proc. of the 2012 IEEE Inform. Theory Workshop (ITW)*, Lausanne, Switzerland, Sept. 3–7 2012, pp. 302–306.
- [11] G. I. Shamir and N. Merhav, “Low complexity sequential lossless coding for piecewise stationary memoryless sources,” *IEEE Trans. Inform. Theory*, vol. 45, pp. 1498–1519, July 1999.
- [12] F. M. J. Willems, “Coding for a binary independent piecewise-identically-distributed source,” *IEEE Trans. Inform. Theory*, vol. 42, pp. 2210–2217, Nov. 1996.
- [13] K. Yamanishi and Y. Maruyama, “Dynamic syslog mining for network failure monitoring,” In *Proc. of the 11th ACM SIGKDD Int’l. Conf. on Knowledge Discovery in Data Mining (KDD2005)*, Chicago, Illinois, USA, Aug. 21–24, 2005, pp. 499–508.
- [14] K. Yamanishi and Y. Maruyama, “Dynamic model selection with its applications to novelty detection,” *IEEE Trans. Inform. Theory*, vol. 53, pp. 2180–2189, June 2007.
- [15] 金澤 宏紀, 山西 健司 “多次元パラメータを有する区間定常無記憶情報源に対しての MDL 原理に基づく変化検出アルゴリズム,” 電子情報通信学会 第 9 回 IBISML 研究会, 京都, 2012 年 6 月 19–20 日.