

情報理論研究会

区間ごとに文脈木モデルが
変化する情報源における
効率的ベイズ符号化アルゴリズム

島田 航志・齋藤 翔太・松嶋 敏泰

早稲田大学 基幹理工学研究科 数学応用数理専攻

2021年3月4日

発表の内容

1. 従来研究
2. 提案する情報源モデル
3. 提案モデルに対するベイズ符号化
4. 効率的なベイズ符号化確率の計算
5. 実験結果
6. まとめ

発表の内容

1. 従来研究

2. 提案する情報源モデル

3. 提案モデルに対するベイズ符号化

4. 効率的なベイズ符号化確率の計算

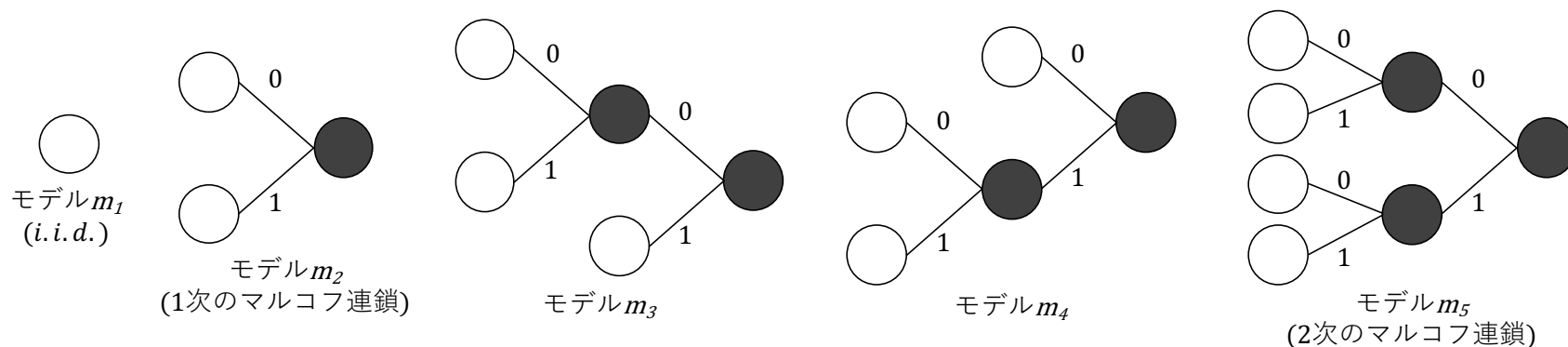
5. 実験結果

6. まとめ

従来研究①：文脈木情報源 (1/3)

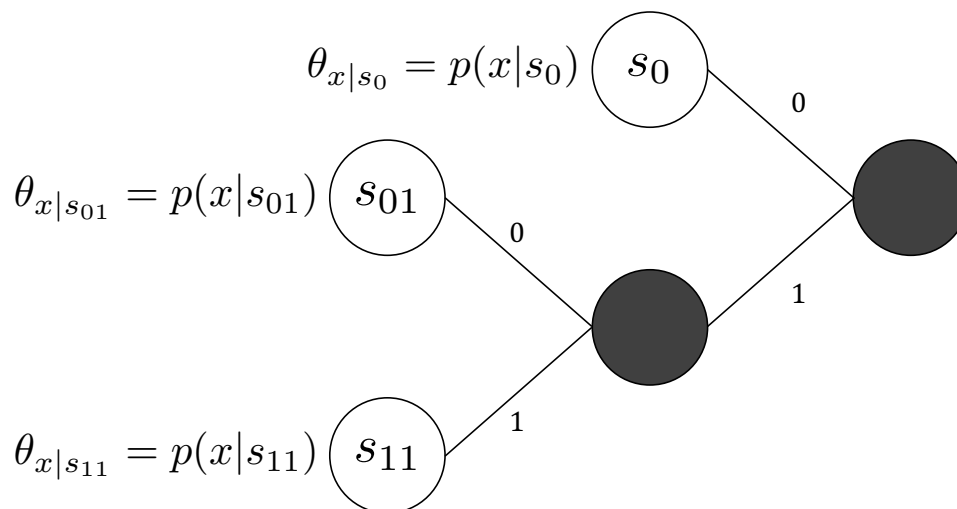
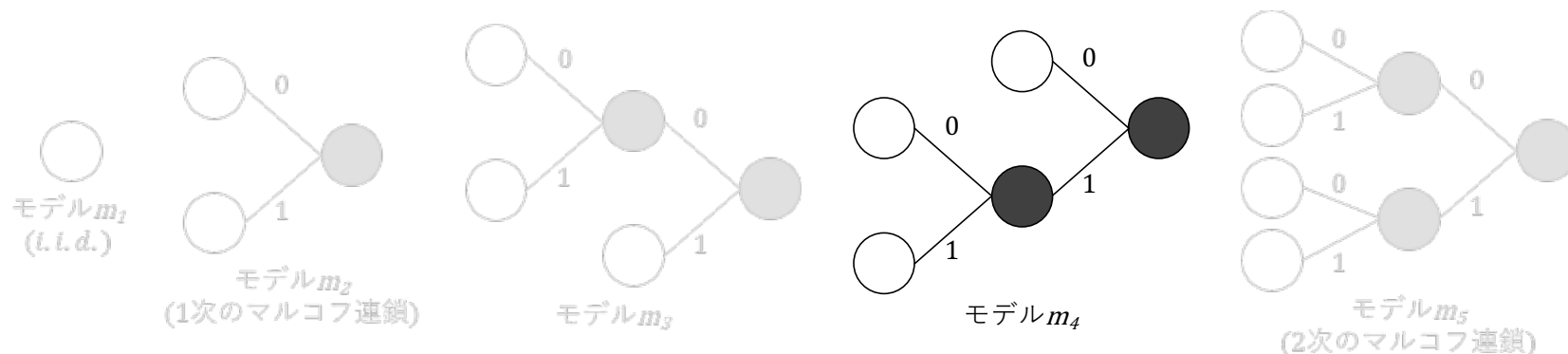
一文脈木情報源とは一

- 過去の系列に依存してシンボルを生成する文脈木モデル
- i.i.d.情報源やマルコフ情報源を含む，広い情報源クラス



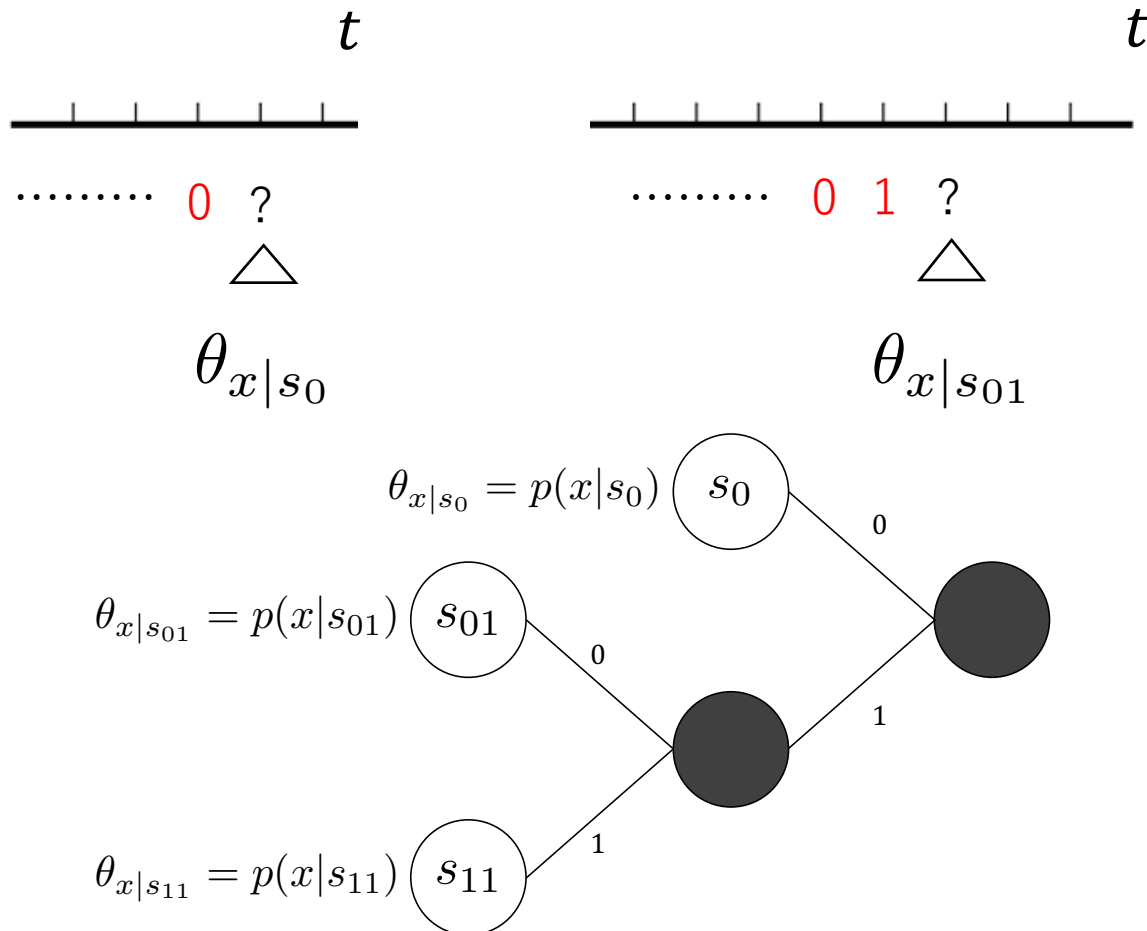
従来研究①：文脈木情報源 (2/3)

一文脈木モデルに基づく生起確率



従来研究①：文脈木情報源 (2/3)

— 文脈木モデルに基づく生起確率 —



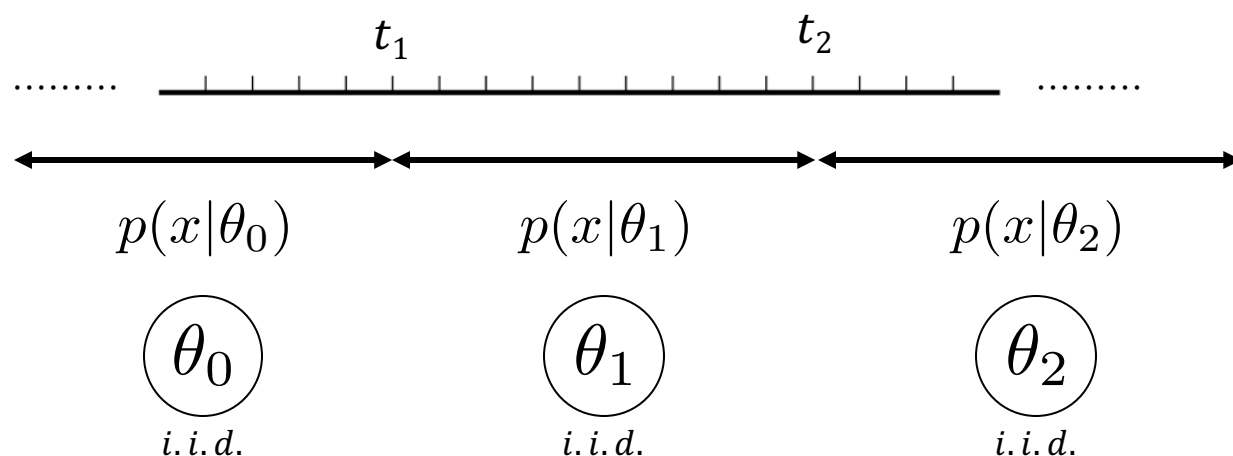
従来研究①：文脈木情報源 (3/3)

—既存の符号化法—

- **文脈木重み付け法** (CTW法) [Willemssら, 1995]
 - 最も知られている文脈木情報源のユニバーサル符号化法
- **各文脈木モデルの事前確率分布のクラスを定め, 効率的なベイズ符号化** [松嶋ら, 2007]
 - CTW法はこのベイズ符号化アルゴリズムの特殊な場合

従来研究②：i.p.i.d.情報源

- 区分的に異なるi.i.d.系列 [Merhav, 1993]
 - i.p.i.d. = independent and piecewise identically distributed
- 代表的な非定常情報源で，変化点検出問題などにも応用



提案の概要

- 本発表では，区間ごとに文脈木モデルが変化する情報源の数理モデルを示す．
- 文脈木モデルおよびその変化パターンに事前確率分布のクラスを仮定することで，上記の情報源モデルに対する効率的なベイズ符号化アルゴリズムを提案する．
- 提案アルゴリズムがモデルの変化を構造的に捉えて符号化していることを，人工的に発生させたデータによる実験で示す．

発表の内容

1. 従来研究

2. 提案する情報源モデル

3. 提案モデルに対するベイズ符号化

4. 効率的なベイズ符号化確率の計算

5. 実験結果

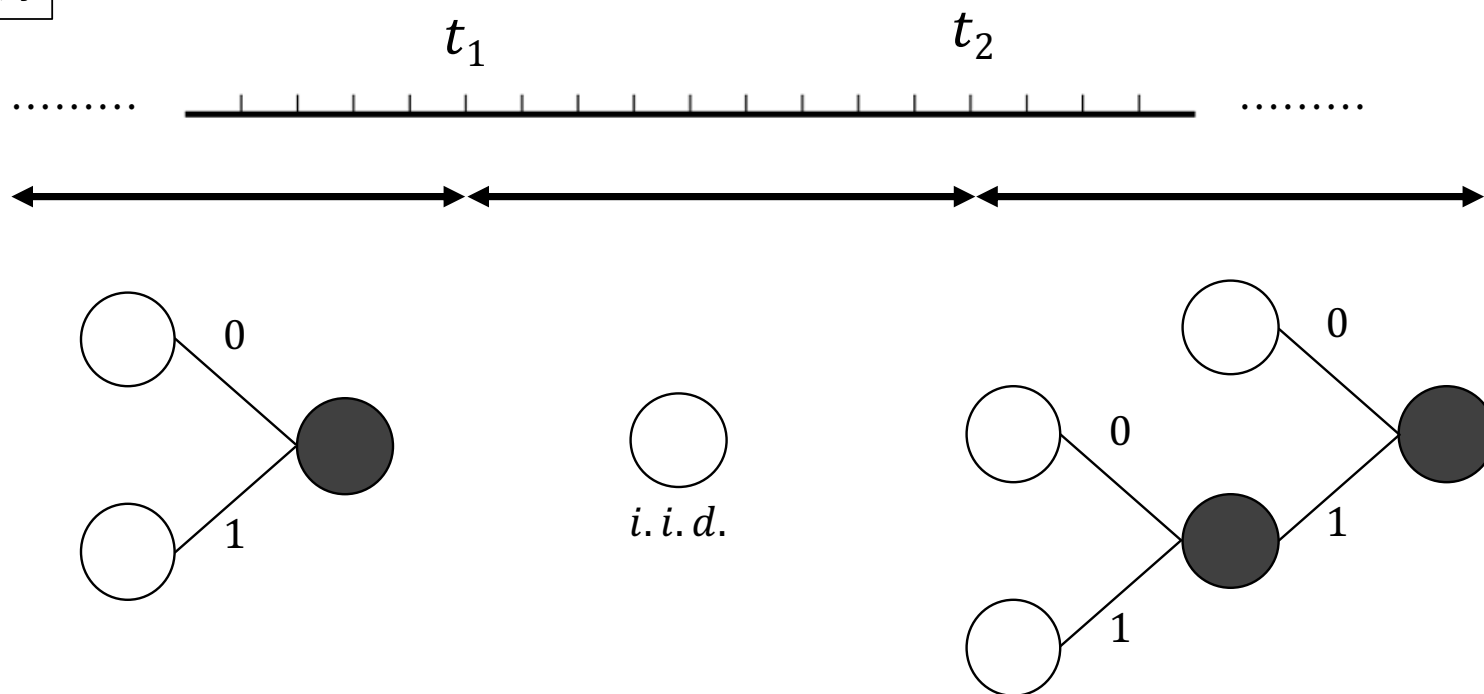
6. まとめ

提案する情報源モデル (1/5)

—提案モデルのイメージ—

区分的に異なる文脈木モデルから生成された情報源系列

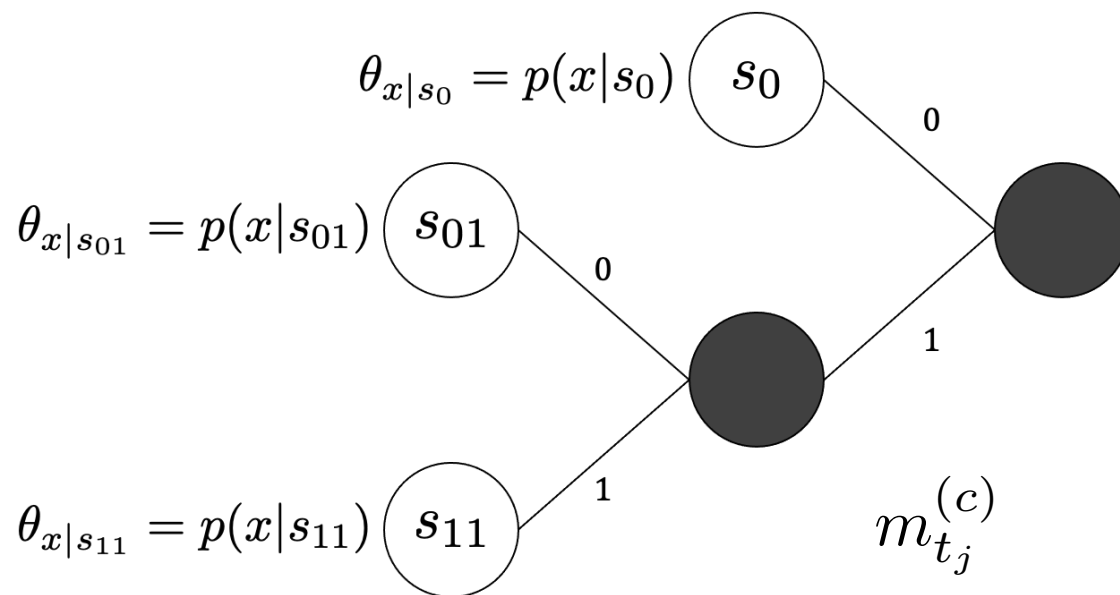
例



提案する情報源モデル (3/5)

— 文脈木モデルに基づく生起確率 —

j 回目の変化点 t_j から $j+1$ 回目の変化点 t_{j+1} までは, ある単一の文脈木モデル $m_{t_j}^{(c)}$ に従ってシンボルが生起する.

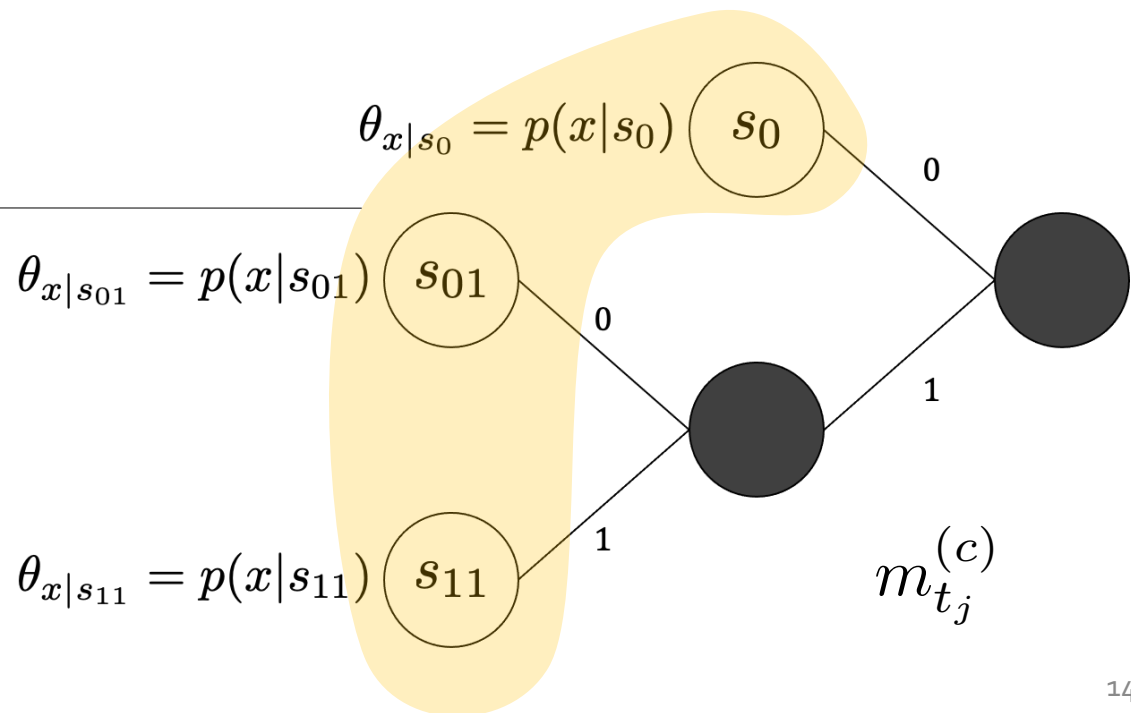


提案する情報源モデル (3/5)

一文脈木モデルに基づく生起確率

j 回目の変化点 t_j から $j+1$ 回目の変化点 t_{j+1} までは, ある単一の文脈木モデル $m_{t_j}^{(c)}$ に従ってシンボルが生起する.

$m_{t_j}^{(c)}$ に対するパラメータ
 $\theta^{m_{t_j}^{(c)}} = \{\theta_{s_0}, \theta_{s_{01}}, \theta_{s_{11}}\}$.
ただし, $\theta_s = {}^T(\theta_{0|s}, \theta_{1|s})$



提案する情報源モデル (4/5)

一文脈木モデルに基づく生起確率

定義：文脈木モデルのパラメータ

変化パターン c において，区間 $[t_j, t_{j+1})$ における文脈木モデル

$m_{t_j}^{(c)}$ に対し，その葉ノードの集合を $L_{m_{t_j}^{(c)}}^{(c)}$ と表すことにする．

このとき， $m_{t_j}^{(c)}$ に対するパラメータ $\theta^{m_{t_j}^{(c)}}$ を次のように定める．

$$\theta^{m_{t_j}^{(c)}} \stackrel{\text{def}}{=} \left\{ \theta_s \in (0, 1)^{|\mathcal{X}|} \mid s \in L_{m_{t_j}^{(c)}}^{(c)} \right\}$$

$$\theta_s \stackrel{\text{def}}{=} {}^T(\theta_{0|s}, \theta_{1|s}, \dots, \theta_{|\mathcal{X}|-1|s}),$$

$\theta_{a|s}$: ノード s に対応する状態における $a \in \mathcal{X}$ の生起確率

$\mathcal{X}$: 情報源アルファベット
の集合

提案する情報源モデル (4/5)

一文脈木モデルに基づく生起確率

定義：文脈木モデルのパラメータ

変化パターン c において，区間 $[t_j, t_{j+1})$ における文脈木モデル

$m_{t_j}^{(c)}$ に対し，その葉ノードの集合を $L_{m_{t_j}^{(c)}}^{(c)}$ と表すことにする．

このとき， $m_{t_j}^{(c)}$ に対するパラメータ $\theta^{m_{t_j}^{(c)}}$ を次のように定める．

$$\theta^{m_{t_j}^{(c)}} \stackrel{\text{def}}{=} \left\{ \theta_s \in (0, 1)^{|\mathcal{X}|} \mid s \in L_{m_{t_j}^{(c)}}^{(c)} \right\}$$

葉ノードに載っている
パラメータの集合

$$\theta_s \stackrel{\text{def}}{=} \top(\theta_{0|s}, \theta_{1|s}, \dots, \theta_{|\mathcal{X}|-1|s}),$$

$\theta_{a|s}$: ノード s に対応する状態における $a \in \mathcal{X}$ の生起確率

提案する情報源モデル (5/5)

—情報源系列の確率分布—

従って、変化パターン $c \in \{0, 1\}^N$ における $X^N = X_1 \cdots X_N$ の確率分布は次式で表される。

変化点の集合 $\mathcal{T}_c \stackrel{\text{def}}{=} \left\{ t \mid w_t^{(c)} = 1 \right\}$

$$\begin{aligned} p(x^N \mid \Theta^c, c) &= \prod_{j=0}^{|\mathcal{T}_c|-1} p\left(x_{t_j}^{t_{j+1}-1} \mid \theta^{m_{t_j}}, m_{t_j}\right) \\ &= \prod_{j=0}^{|\mathcal{T}_c|-1} \prod_{t=t_j}^{t_{j+1}-1} p\left(x_t \mid x_{t_j}^{t-1}, \theta^{m_{t_j}}, m_{t_j}\right). \end{aligned}$$

変化パターン c に対するパラメータ $\Theta^c \stackrel{\text{def}}{=} \left\{ \theta^{m_t^{(c)}} \mid t \in \mathcal{T}_c \right\}$

提案する情報源モデル (5/5)

—情報源系列の確率分布—

従って、変化パターン $c \in \{0, 1\}^N$ における $X^N = X_1 \cdots X_N$ の確率分布は次式で表される。

変化点の集合 $\mathcal{T}_c \stackrel{\text{def}}{=} \{t \mid w_t^{(c)} = 1\}$

$$\begin{aligned} p(x^N \mid \Theta^c, c) &= \prod_{j=0}^{|\mathcal{T}_c|-1} p(x_{t_j}^{t_{j+1}-1} \mid \theta^{m_{t_j}}, m_{t_j}) \\ &= \prod_{j=0}^{|\mathcal{T}_c|-1} \prod_{t=t_j}^{t_{j+1}-1} p(x_t \mid x_{t_j}^{t-1}, \theta^{m_{t_j}}, m_{t_j}). \end{aligned}$$

各区間の分布の積

変化パターン c に対するパラメータ $\Theta^c \stackrel{\text{def}}{=} \{\theta^{m_t^{(c)}} \mid t \in \mathcal{T}_c\}$

発表の内容

1. 従来研究
2. 提案する情報源モデル
- 3. 提案モデルに対するベイズ符号化**
4. 効率的なベイズ符号化確率の計算
5. 実験結果
6. まとめ

ベイズ符号化する上で

- 前節で導入した変化パターン・文脈木モデル・パラメータは**未知**とする.
- これらを確率変数ととらえ, 事前確率分布を仮定し, 事後確率を計算する.
- ベイズ符号の符号化確率は, 事後確率分布による重み付け計算があり, 計算量が大きくなる傾向がある.

提案モデルに対するベイズ符号化

定理：ベイズ符号化確率

提案モデルに対する逐次型ベイズ符号の符号化確率は,

$$\text{AP}^*(x_t|x^{t-1}) = \sum_{c \in \{0,1\}^N} \pi(c|x^{t-1}) \left[\sum_{m^{(c)} \in \mathcal{M}} P(m^{(c)}|x^{t-1}, c) \int p(x_t|x^{t-1}, \boldsymbol{\theta}^{m^{(c)}}, m^{(c)}, c) w(\boldsymbol{\theta}^{m^{(c)}}|x^{t-1}, m^{(c)}, c) d\boldsymbol{\theta}^{m^{(c)}} \right].$$



ベイズ符号は, この符号化確率 AP^* を用いて
算術符号化を行う.

提案モデルに対するベイズ符号化

定理：ベイズ符号化確率

提案モデルに対する逐次型ベイズ符号の符号化確率は、

$$\text{AP}^*(x_t|x^{t-1}) = \sum_{c \in \{0,1\}^N} \pi(c|x^{t-1}) \left[\sum_{m^{(c)} \in \mathcal{M}} P(m^{(c)}|x^{t-1}, c) \int p\left(x_t|x^{t-1}, \boldsymbol{\theta}^{m^{(c)}}, m^{(c)}, c\right) w\left(\boldsymbol{\theta}^{m^{(c)}}|x^{t-1}, m^{(c)}, c\right) d\boldsymbol{\theta}^{m^{(c)}} \right].$$

$\mathcal{M}$: とり得る文脈木モデル全体の集合.
文脈木モデルの最大深さに対して指数的に存在する.

提案モデルに対するベイズ符号化

定理：ベイズ符号化確率

提案モデルに対する逐次型ベイズ符号の符号化確率は、

$$\text{AP}^*(x_t|x^{t-1}) = \sum_{c \in \{0,1\}^N} \pi(c|x^{t-1}) \left[\sum_{m^{(c)} \in \mathcal{M}} P(m^{(c)}|x^{t-1}, c) \int p(x_t|x^{t-1}, \boldsymbol{\theta}^{m^{(c)}}, m^{(c)}, c) w(\boldsymbol{\theta}^{m^{(c)}}|x^{t-1}, m^{(c)}, c) d\boldsymbol{\theta}^{m^{(c)}} \right].$$

- m
- 効率的なベイズ符号化アルゴリズム [松嶋ら, 2007]

提案モデルに対するベイズ符号化

定理：ベイズ符号化確率

提案モデルに対する逐次型ベイズ符号の符号化確率は、

$$\text{AP}^*(x_t|x^{t-1}) = \sum_{c \in \{0,1\}^N} \pi(c|x^{t-1}) \left[\sum_{m^{(c)} \in \mathcal{M}} P(m^{(c)}|x^{t-1}, c) \int p(x_t|x^{t-1}, \theta^{m^{(c)}}, m^{(c)}, c) w(\theta^{m^{(c)}}|x^{t-1}, m^{(c)}, c) d\theta^{m^{(c)}} \right].$$



依然として、変化パターンの事後分布による
重み付け計算の計算量が大きい

発表の内容

1. 従来研究
2. 提案する情報源モデル
3. 提案モデルに対するベイズ符号化
- 4. 効率的なベイズ符号化確率の計算**
5. 実験結果
6. まとめ

効率的なベイズ符号化確率の計算 (1/3)

—変化パターンの事前分布—

定義：変化パターンの事前分布 [須子ら, 2003]

変化パターン $c \stackrel{\text{def}}{=} (w_1^{(c)}, \dots, w_t^{(c)}, \dots, w_N^{(c)})$ に関して, 各 $w_t^{(c)}$ が独立にベルヌーイ分布 $\text{Ber}(\alpha)$ に従うと仮定する. すなわち

$$\pi(c) \stackrel{\text{def}}{=} \alpha^{|\mathcal{T}_c|-1} (1 - \alpha)^{N-|\mathcal{T}_c|}.$$

変化点の集合 $\mathcal{T}_c \stackrel{\text{def}}{=} \left\{ t \mid w_t^{(c)} = 1 \right\}$

効率的なバイズ符号化確率の計算 (2/3)

—最終変化点の導入—

- 現在時点 t から見た最終変化点 τ_t を導入
- τ_t 以前の文脈木モデルの変化は, x_t の生起確率に影響を与えない.

定義：最終変化点の事前分布 [須子ら, 2003]

τ_t の事前確率分布 $v(\tau_t)$ を

$$v(\tau_t) \stackrel{\text{def}}{=} \sum_{c: \tau_t \in \mathcal{T}_c} \pi(c), \text{ where } \tau_t = 1, 2, \dots, t$$

と定める.

効率的なベイズ符号化確率の計算 (3/3)

—主結果—

効率的なベイズ符号化確率の計算

符号化確率を次式で計算する.

$$\tilde{p}(x_t|x^{t-1}) = \sum_{\tau_t=1}^t v(\tau_t|x^{t-1}) \tilde{q}(x_t|x_{\tau_t} \cdots x_{t-1}).$$

効率的なベイズ符号化確率の計算 (3/3)

—主結果—

効率的なベイズ符号化確率の計算

符号化確率を次式で計算する.

$$\tilde{p}(x_t|x^{t-1}) = \sum_{\tau_t=1}^t v(\tau_t|x^{t-1}) \tilde{q}(x_t|x_{\tau_t} \cdots x_{t-1}).$$

- 文脈木情報源に対する効率的なベイズ符号化アルゴリズム
[松嶋ら, 2007]
- 計算量のオーダーが $\mathcal{O}(d)$ まで軽減
※ d : 文脈木モデルの最大深さ

効率的なベイズ符号化確率の計算 (3/3)

—主結果—

効率的なベイズ符号化確率の計算

符号化確率を次式で計算する.

$$\tilde{p}(x_t|x^{t-1}) = \sum_{\tau_t=1}^t v(\tau_t|x^{t-1}) \tilde{q}(x_t|x_{\tau_t} \cdots x_{t-1}).$$

計算量のオーダーが $\mathcal{O}(2^N)$ から $\mathcal{O}(N^2)$ まで軽減

N : 系列全体の長さ

効率的なベイズ符号化確率の計算 (3/3)

—主結果—

効率的なベイズ符号化確率の計算

符号化確率を次式で計算する.

$$\tilde{p}(x_t|x^{t-1}) = \sum_{\tau_t=1}^t v(\tau_t|x^{t-1}) \tilde{q}(x_t|x_{\tau_t} \cdots x_{t-1}).$$

計算量のオーダーが $\mathcal{O}(2^N)$ から $\mathcal{O}(N^2)$ まで軽減

N : 系列全体の長さ

▶ 全体の計算量が $\mathcal{O}(d \cdot N^2)$ まで削減

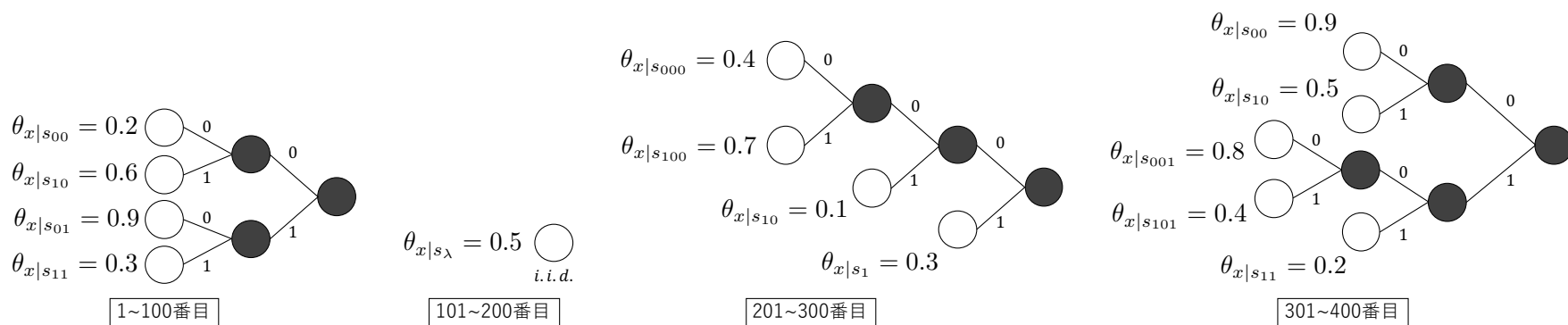
発表の内容

1. 従来研究
2. 提案する情報源モデル
3. 提案モデルに対するベイズ符号化
4. 効率的なベイズ符号化確率の計算
- 5. 実験結果**
6. まとめ

実験条件

—人工的に生成した系列の生成モデル—

- 長さ400で、100シンボル毎に異なる文脈木モデルに従って発生させた人工的な情報源系列



実験①：圧縮性能

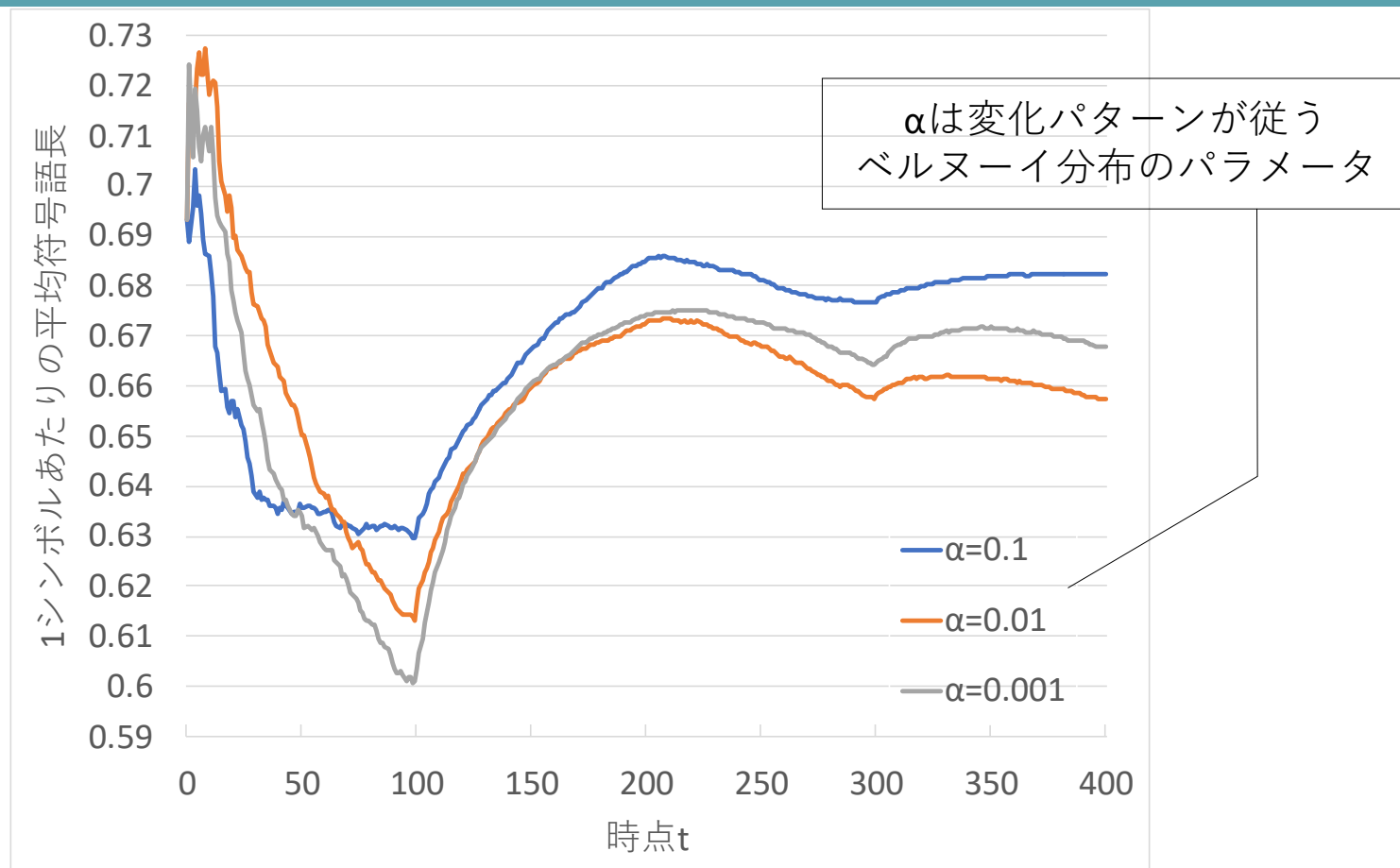


図1. 1シンボルあたりの平均符号語長の推移（100回平均）

実験②：最終変化点のMAP推定

※ MAP = 最大事後確率

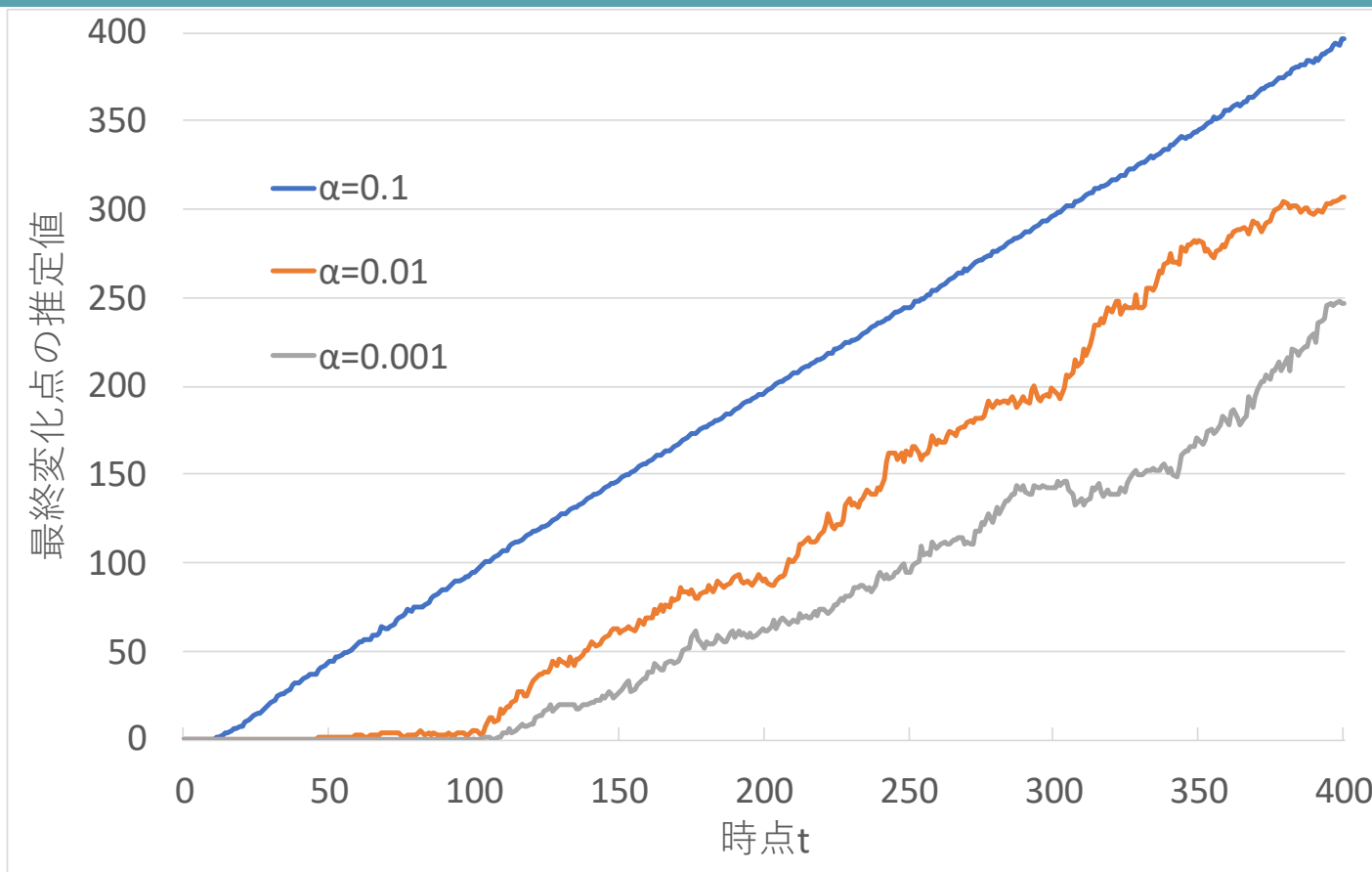


図2. 最終変化点の推定値の推移 (100回平均)

実験②：最終変化点のMAP推定

※ MAP = 最大事後確率

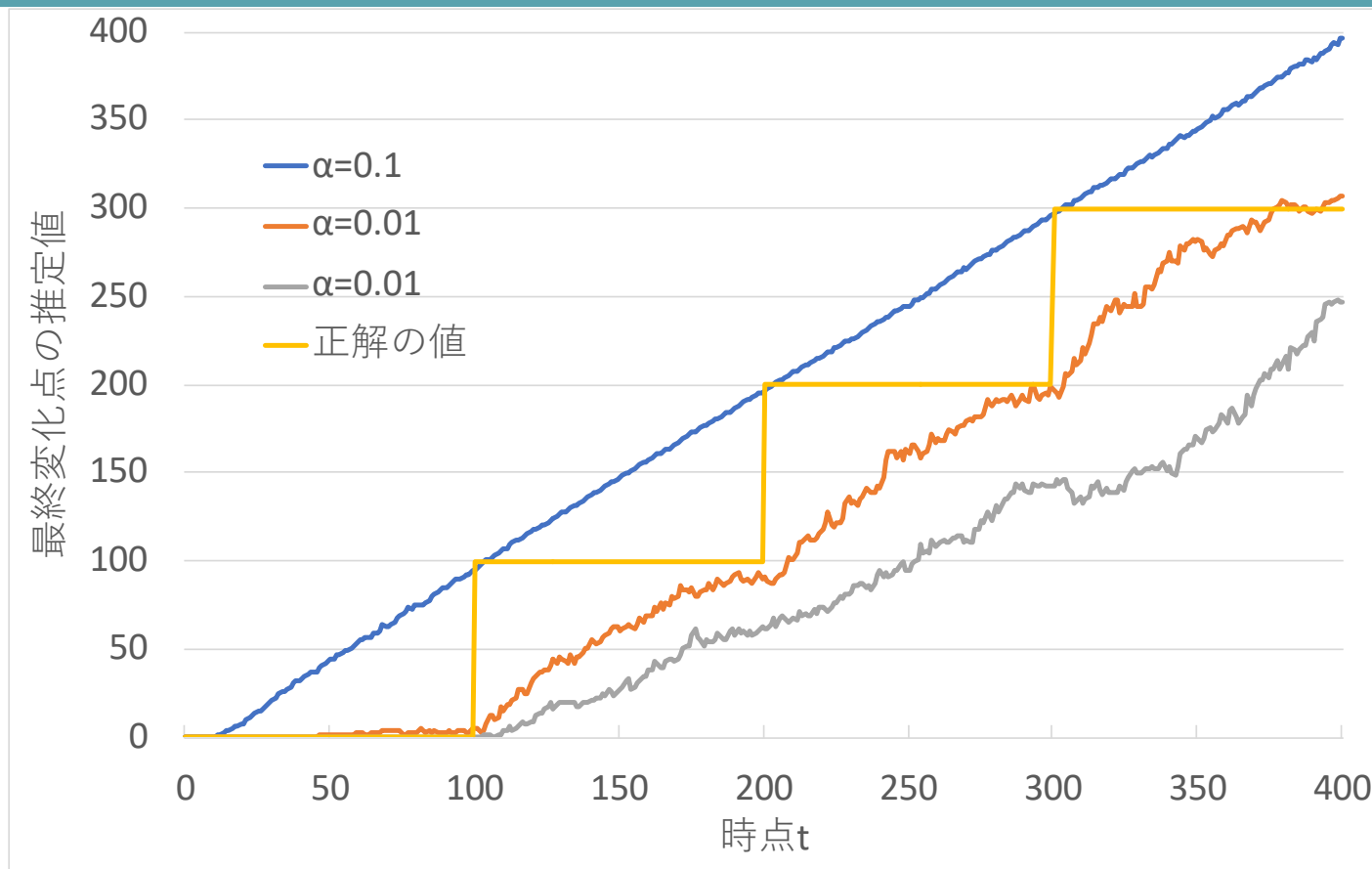


図2. 最終変化点の推定値の推移 (100回平均)

実験結果

- 実際のモデル変化の頻度に即した $\alpha = 0.01$ の場合が最も圧縮率が良かったが、概ね α の値によらず圧縮できる.
- 変化点の推定という目的で本アルゴリズムを用いる場合、 α は実際のモデル変化の頻度に即した値である必要がある.
- 他にも文脈木モデルのMAP推定などが可能であるが、今回は省略する.

発表の内容

1. 従来研究
2. 提案する情報源モデル
3. 提案モデルに対するベイズ符号化
4. 効率的なベイズ符号化確率の計算
5. 実験結果
- 6. まとめ**

まとめ

- 区間ごとに文脈木モデルが変化する情報源の数理モデルを示した.
- 上記の情報源モデルに対する効率的なベイズ符号化アルゴリズムを示した.
- 人工的に発生させた情報源系列を提案アルゴリズムで符号化する実験の結果を示した.

参考文献

[Willemssら, 1995] F. M. J. Willems, Y. M. Shtarkov and T. J. Tjalkens, “The Context-Tree Weighting Method: Basic Properties,” in *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 653-664, May 1995.

[松嶋ら, 2007] T. Matsushima and S. Hirasawa, “A Class of Prior Distributions on Context Tree Models and an Efficient Algorithm of the Bayes Codes Assuming It,” *2007 IEEE International Symposium on Signal Processing and Information Technology*, Giza, 2007, pp. 938-941.

[Merhav, 1993] N. Merhav, “On the Minimum Description Length Principle for Sources with Piecewise Constant Parameters,” in *IEEE Transactions on Information Theory*, vol. 39, no. 6, pp. 1962-1967, Nov. 1993.

[須子ら, 2003] 須子統太, 松嶋敏泰, 平澤茂一, “区間で一定なパラメータを持つ情報源におけるベイズ符号化法について,” 第 26 回情報理論とその応用シンポジウム予稿集, pp. 165-168, 2003.