論　文

符号シンボルのコストを考慮したユニバーサル情報源符号化

植松　友彦[†a)]　　川上　秀彦[†]

Universal Data Compression Considering Cost for Codeword

Tomohiko UYEMATSU[†a)] and Hidehiko KAWAKAMI[†]

あらまし　情報源符号化に用いられる多くの符号は，各符号シンボルを伝送あるいは記録するためのコストが均一であるという条件を暗黙のうちに仮定し，符号語長を短くすることを目的としている．しかしながら，情報の伝送や記録においては，各符号シンボルに不均一のコストを仮定することが自然な場合があり，このような場合，従来の符号語長を最小にする符号はもはや最適ではなく，符号語のコストを小さくする符号が要求される．本論文では，加法的コストを含む，より一般的なコストである正則コストあるいは有限状態コストが与えられたとき，指定された範囲内のひずみを許した場合に符号系列のコストを最小にするユニバーサル符号，及び符号語コストレートを指定する範囲内で制限した場合にひずみを最小にするユニバーサル符号の構成法を提案し，これらの符号が漸近的に最良であることを示す．更に，ひずみを許さない場合についても同様の考察を行っている．

キーワード　符号語コスト，有ひずみユニバーサル符号，無ひずみユニバーサル符号，レートひずみ理論，概収束符号化定理

1. まえがき

情報源符号化に用いられる多くの符号は，与えられた情報源系列に対して符号語長を短くすることを目的として設計されている．このことは，各符号シンボルを伝送あるいは記録するためのコストは均一であるという条件を暗黙のうちに仮定している．しかしながら，On-Off Keying で変調を行う場合，1 と 0 の伝送に対して必要な電力が異なるし，モールス符号では，長点と短点とで伝送に必要な時間が異なっている．また，CD や磁気記録等では，1 の続く長さ（ランレングス）に制約があり，この場合の符号化の問題も文脈に依存するコストとして定式化される．これらの例のように，実際の伝送や記録においては，各符号シンボルに不均一のコストを仮定することが自然な場合があり，このような場合，符号語長はもはや最適な尺度とは言えず，符号語のコストを小さくする符号が要求される．この要求に対し，シャノン以来，様々な検討が行われている [1], [2], [4], [10], [13]．中でも，岩田らは，定常エルゴード情報源について符号語コストレートが漸近的に最良となる実用的な無ひずみユニバーサル符号の構成法を提案している [10]．一方，内田と植松は，岩田らによる結果を有ひずみ符号に対して拡張した [13]．しかしながら，彼らの提案した符号では，符号語コストが加法的であるという強い制限を有していた．このため，ランレングス制約などをコストとして取り扱うことができず，応用が限られていた．他方，シャノン [1] や Csiszár [2] は，ランレングス制約などを取り扱うことのできる有限状態コストについて無ひずみ符号化定理を示しているが，有ひずみ符号化あるいはユニバーサル符号化についてはいまだ検討されていない．

本論文では，加法的コストを含む，より一般的なコストである正則コストあるいは有限状態コストに対するユニバーサル情報源符号化について考察している．すなわち，定常エルゴード情報源に対し，指定された範囲内のひずみを許した場合に，これらのコストで測られる符号語コストを最小にするユニバーサル符号，及び符号語コストレートを指定された範囲に制限したときにひずみを最小にするユニバーサル符号が，符号語長に基づいて設計されたユニバーサル符号から構成できることを明らかにし，これらの符号が漸近的に最良であることを示す．更に，無ひずみ符号化についても同様の考察を行っている．

† 東京工業大学 大学院 理工学研究科 集積システム専攻，東京都
Dept. of Communications and Integrated Systems, Tokyo Institute of Technology, Tokyo, 152–8552 Japan
a) E-mail: uematsu@ss.titech.ac.jp

2. コスト関数

本章では，本論文で用いている符号語コストについて説明する．コストのクラスや性質等の詳細は文献 [9] を参照されたい．本論文を通じて，$\mathcal{Y}$ を有限の符号アルファベットとし，アルファベット $\mathcal{Y}$ の元による長さ n の系列の集合を $\mathcal{Y}^n$，系列 $y_1y_2\cdots y_n \in \mathcal{Y}^n$ を y^n と表記する．更に，$\mathcal{Y}$ の元による有限系列全体の集合を $\mathcal{Y}^*$，その元を y^* と表記する．log 並びに exp の底は $|\mathcal{Y}|$ とする．ただし，$|\cdot|$ は集合の要素数を表す記号である．

2.1 条件付コストと正則コスト

条件付コストは次のように定義される [9].

［定義 1］ 系列 $y^i = y_1y_2\cdots y_i \in \mathcal{Y}^i$ における y_i の条件付コストを

$$0 \leqq \mathrm{cst}\left(y_i|y^{i-1}\right) \leqq +\infty$$

によって定義し，系列 $y^n \in \mathcal{Y}^n$ のコストを

$$\mathrm{cst}\left(y^n\right) \stackrel{\triangle}{=} \sum_{i=1}^{n} \mathrm{cst}\left(y_i|y^{i-1}\right)$$

によって定義する． □

条件付コスト容量及び（条件付）正則コストは次のように定義される [9].

［定義 2］ $y^{i-1} \in \mathcal{Y}^{i-1}$ に対する条件付コスト容量 $\alpha_0\left(y^{i-1}\right)$ は，

$$\sum_{y\in\mathcal{Y}} \exp\left\{-\alpha \mathrm{cst}\left(y|y^{i-1}\right)\right\} = 1 \qquad (1)$$

を満足する α の正の根である． □

［定義 3］ ある正数 d_1, d_2 $(0 < d_1 \leqq d_2 < \infty)$ が存在し，任意の y $(\in \mathcal{Y})$ 及び y^{i-1} $\left(\in \mathcal{Y}^{i-1}\right)$ について

$$d_1 \leqq \mathrm{cst}\left(y|y^{i-1}\right) \leqq d_2 \qquad (2)$$

を満足し，かつ条件付コスト容量 $\alpha_0\left(y^{i-1}\right)$ が i 及び y^{i-1} によらずに一定の値 α_0（定コスト容量と呼ぶ）を取るならば，このコストを（条件付）正則コストと呼ぶ． □

（注意 1）正則コストの定義において，式 (2) を $0 \leqq \mathrm{cst}\left(y|y^{i-1}\right) \leqq d_2$ に緩めることはできない．なぜなら，ある y が $\mathrm{cst}\left(y|y^{i-1}\right) = 0$ を満足するならば，式 (1) を満足する有限な α は存在しないからである．

2.2 有限状態コスト

$S = \{1, 2, \cdots, |S|\}$ を状態集合とし，s_1 $(\in S)$ を初期状態とする．また，関数 $F(s, y)$ によって，状態 $s \in S$ で $y \in \mathcal{Y}$ が生じたときの次の状態を表す．このとき，条件付コストの特別な場合である有限状態コストとコスト容量を次のように定義する [1], [2].

［定義 4］ 任意の $s_1, s' \in S$ に対し，ある有限系列 $y_1y_2\cdots y_m \in \mathcal{Y}^*$ が存在して，

$$s_2 = F(s_1, y_1), s_3 = F(s_2, y_2), \cdots, s' = F(s_m, y_m)$$

を満足するとき，状態 $s \in S$ における $y \in \mathcal{Y}$ の条件付コストを

$$0 \leqq \mathrm{cst}(y|s) \leqq +\infty$$

によって定義し，系列 $y^n \in \mathcal{Y}^n$ のコストを

$$\mathrm{cst}\left(y^n\right) \stackrel{\triangle}{=} \sum_{i=1}^{n} \mathrm{cst}(y_i|s_i)$$

によって定義する．ただし，$s_{i+1} = F(s_i, y_i)$ $(i = 1, 2, \cdots)$ である． □

［定義 5］ 有限状態コストのコスト容量 α_0 は，$|S| \times |S|$ 行列 $M(\alpha) = (m_{ik}(\alpha))$ を

$$m_{ik}(\alpha) = \sum_{y\in\mathcal{Y}:F(i,y)=k} \exp\{-\alpha \mathrm{cst}(y|i)\}$$

によって定めるとき，$\det[M(\alpha) - I_{|S|}] = 0$ を満足する最大の α である．ただし，$I_{|S|}$ は $|S|$ 次の単位行列であり，$\mathrm{cst}(y|i) = \infty$ ならば $\exp\{-\alpha \mathrm{cst}(y|i)\} = 0$ と約束する． □

（例 1）符号アルファベットを $\mathcal{Y} = \{0, 1\}$ とするとき，0 の数が 2 個以上続かないようなランレングス制約は，状態集合 $S = \{1, 2\}$ を用いた有限状態コスト

$$F(1,1) = F(2,1) = 1,\ F(1,0) = 2,$$
$$\mathrm{cst}(0|1) = \mathrm{cst}(1|1) = \mathrm{cst}(1|2) = 1,$$
$$\mathrm{cst}(0|2) = \infty$$

によって記述される．ただし，初期状態 $s_1 = 1$ である．このとき，

$$M(\alpha) = \begin{bmatrix} \exp(-\alpha) & \exp(-\alpha) \\ \exp(-\alpha) & 0 \end{bmatrix}$$

であり，コスト容量 α_0 は，

$$\det \begin{bmatrix} \exp(-\alpha)-1 & \exp(-\alpha) \\ \exp(-\alpha) & -1 \end{bmatrix} = 0$$

を満足する最大の α であるから，

$$\alpha_0 = \log_2 \frac{1+\sqrt{5}}{2}$$

となる．なお，このコスト容量 α_0 は，このランレングス制約を有する無雑音通信路の通信路容量 [11] と一致している． □

正則コストと有限状態コストの間には包含関係がないことに注意する．例えば，正則コストは，それまでの系列に依存して次の記号の条件付コストを定めることができるため，可算無限の状態を取り得る．正則コストであり有限状態コストではない具体的なコストの例を付録に示す．他方，有限状態コストでは，条件付コストの値として零と無限大を許しており，ランレングス制約をコストとして自然に取り扱うことができる．以下では，特に断らない限り，コストとして正則コストあるいは有限状態コストのみを取り扱うことにし，いずれのコストも $\mathrm{cst}(\cdot)$ によって表すことにする．

3. コスト付符号化における基本定理

本章では，コストを考慮した情報源符号化において重要な基本定理を述べる．

まず，正則コストについての基本定理を述べる．

［基本定理 1］ $\mathcal{A}$ を有限集合とするとき，任意の語頭符号 $\psi : \mathcal{A} \to \mathcal{Y}^*$ と正則コストに対し，次の条件を満足する語頭符号 $\tilde{\psi} : \mathcal{A} \to \mathcal{Y}^*$ が存在する．

$$\alpha_0 \mathrm{cst}\left(\tilde{\psi}(a)\right) \leqq \ell(\psi(a)) + \alpha_0 d_2 \quad \forall a \in \mathcal{A} \qquad (3)$$

ただし，$\ell(\cdot)$ は系列長を表す関数である．

また逆に，与えられた語頭符号 $\psi : \mathcal{A} \to \mathcal{Y}^*$ に対し，

$$\ell\left(\hat{\psi}(a)\right) < \alpha_0 \mathrm{cst}(\psi(a)) + 1 \quad \forall a \in \mathcal{A} \qquad (4)$$

を満足する語頭符号 $\hat{\psi} : \mathcal{A} \to \mathcal{Y}^*$ が存在する． □

（証明） 文献 [2] の Proposition 2.2 の証明と同様の手法を用いる．$\mathcal{A} = \{1, 2, \cdots |\mathcal{A}|\}$ とし，一般性を失うことなく

$$\ell(\psi(1)) \leqq \ell(\psi(2)) \leqq \cdots \leqq \ell(\psi(|\mathcal{A}|))$$

が成り立つとする．このとき，$i \in \mathcal{A}$ に対して実数 $\beta_i \in [0,1]$ を

$$\beta_i = \sum_{k=1}^{i-1} \exp\{-\ell(\psi(i))\}$$

によって対応させる，ただし，$\beta_1 = 0$ とする．

次に，区間 $[0,1]$ を $|\mathcal{Y}|$ 個の右半開区間 $J(y_1)$ $(y_1 \in \mathcal{Y})$ に分割する．ただし，区間 $J(y_1)$ の幅は $\exp\{-\alpha_0 \mathrm{cst}(y_1)\}$ に等しい．このような分割が可能なことは，α_0 の定義からわかる．次に，各々の区間 $J(y_1)$ $(y_1 \in \mathcal{Y})$ を $|\mathcal{Y}|$ 個の右半開区間 $J(y_1 y_2)$ $(y_2 \in \mathcal{Y})$ に分割する．このとき，区間 $J(y_1 y_2)$ の幅は $\exp\{-\alpha_0 \mathrm{cst}(y_1 y_2)\}$ に等しい．以下，同様の分割を繰り返し，n 回目の分割では，$J(y_1 y_2 \cdots y_{n-1})$ $(y_1 y_2 \cdots y_{n-1} \in \mathcal{Y}^{n-1})$ を $|\mathcal{Y}|$ 個の右半開区間 $J(y_1 y_2 \cdots y_n)$ $(y_n \in \mathcal{Y})$ に分割する．ただし，区間 $J(y_1 y_2 \cdots y_n)$ の幅は $\exp\{-\alpha_0 \mathrm{cst}(y_1 y_2 \cdots y_n)\}$ に等しい．このとき，正則コストの定義（式 (2)）から，1 回の分割によって得られる各々の区間の幅は分割前の区間幅のたかだか $\exp\{-\alpha_0 d_1\}(< 1)$ 倍であり，分割を繰り返すことで得られる各々の区間の幅はすべてに 0 に収束する．したがって，任意の $i \in \mathcal{A}$ に対し，$\beta_i \in J(y^*)$ でありかつ i 以外のすべての $j \in \mathcal{A}$ について $\beta_j \notin J(y^*)$ を満足する有限系列 $y^* \in \mathcal{Y}^*$ が必ず存在する．区間 $J(y^*)$ の作り方から，このような有限系列 y^* の中で $\mathrm{cst}(y^*)$ を最小にする系列 $y^*_{\min}$ は一意的に定まる．そこで写像 $\tilde{\psi} : \mathcal{A} \to \mathcal{Y}^*$ を，$i \in \mathcal{A}$ に対してそのような系列 $y^*_{\min}$ を対応させることによって定義すれば，$\tilde{\psi}$ は語頭符号である．他方，$\tilde{\psi}(i) = y_1 y_2 \cdots y_m$ のとき，明らかに

$$\exp\{-\alpha_0 \mathrm{cst}(y_1 y_2 \cdots y_{m-1})\} \geqq \exp\{-\ell(\psi(i))\}$$

が成り立ち，上式と式 (2) から導かれる不等式

$$\begin{aligned} &\mathrm{cst}(y_1 y_2 \cdots y_m) \\ &\quad = \mathrm{cst}(y_1 y_2 \cdots y_{m-1}) + \mathrm{cst}(y_m | y_1 y_2 \cdots y_{m-1}) \\ &\quad \leqq \mathrm{cst}(y_1 y_2 \cdots y_{m-1}) + d_2 \end{aligned}$$

から式 (3) が得られる．

次に，式 (4) の成立を示す．まず，$\mathcal{Y}^*$ の部分集合 C を $C \triangleq \{\psi(a) : a \in \mathcal{A}\}$ によって定める．そして，C を表す符号木（$|\mathcal{Y}|$ 分木）に最小本数の枝を付け加え，完全木としたものを $\tilde{C}$ とすれば，$C \subset \tilde{C}$ と式 (1) か

ら直ちに，

$$\sum_{y^*\in C}\exp\{-\alpha_0\mathrm{cst}(y^*)\}\leqq\sum_{y^*\in\tilde{C}}\exp\{-\alpha_0\mathrm{cst}(y^*)\}$$
$$=1$$

が成り立つ．すなわち，

$$\sum_{a\in\mathcal{A}}\exp\{-\alpha_0\mathrm{cst}(\psi(a))\}\leqq 1$$

が成り立つので，$L(a)\equiv\lceil\alpha_0\mathrm{cst}(\psi(a))\rceil$ とすると，

$$\sum_{a\in\mathcal{A}}\exp\{-L(a)\}\leqq 1$$

が成り立つ．このことと文献 [14] の定理 3.10 から，任意の $a\in\mathcal{A}$ に対して $L(a)$ を符号語長とする語頭符号 $\hat{\psi}:\mathcal{A}\to\mathcal{Y}^*$ が存在し，符号 $\hat{\psi}$ は式 (4) を満足することがわかる．これで基本定理 1 は証明された． □

（注意 2）基本定理 1 は語頭符号に対して，符号長と正則コストの尺度としての等価性を述べたものである．なお，語頭符号のかわりに Kolmogorov 複雑量を考えた場合が加藤 [9] によって示されている．

有限状態コストについての同様な基本定理を次に示す．

［基本定理 2］ $\mathcal{A}$ を有限集合とするとき，任意の語頭符号 $\psi:\mathcal{A}\to\mathcal{Y}^*$ と有限状態コストに対し，次の条件を満足する語頭符号 $\tilde{\psi}:\mathcal{A}\to\mathcal{Y}^*$ が存在する．

$$\alpha_0\mathrm{cst}\left(\tilde{\psi}(a)\right)\leqq\ell(\psi(a))+B\quad\forall a\in\mathcal{A}\qquad(5)$$

ただし，B は与えられた有限状態コストによって定まる定数である．

また逆に，与えられた語頭符号 $\psi:\mathcal{A}\to\mathcal{Y}^*$ に対し，

$$\ell\left(\hat{\psi}(a)\right)<\alpha_0\mathrm{cst}(\psi(a))+B'\quad\forall a\in\mathcal{A}\qquad(6)$$

を満足する語頭符号 $\hat{\psi}:\mathcal{A}\to\mathcal{Y}^*$ が存在する．ただし，B' は与えられた有限状態コストによって定まる定数である． □

（証明） 文献 [2] の Proposition 2.2 と本質的に同一である．式 (5) は文献 [2] の式 (2.24) にほかならない．そこで，式 (6) を示す．文献 [2] の式 (2.12) から，与えられた有限状態コストに対し，

$$\sum_{k=1}^{|S|}m_{ik}(\alpha_0)\cdot a_k=a_i\quad i=1,2,\cdots,|S|$$

を満足する $a_i>0$ $(i=1,2,\cdots,|S|)$ が存在する．このとき，文献 [2] の p.296 で述べられた単位区間の分割を考えれば，$C\triangleq\{\psi(a):a\in\mathcal{A}\}$ に対し，直ちに，

$$\sum_{y^*\in C}\exp\{-\alpha_0\mathrm{cst}(y^*)+t(y^*)\}\leqq 1$$

が成り立つ．ただし，$y^n=y_1y_2\cdots y_n$ とするとき，$s_{i+1}=F(s_i,y_i)$ $(i=1,2,\cdots,n)$ であり，

$$t(y^n)\triangleq\log\frac{a_{s_{n+1}}}{a_{s_1}}$$

である．後は，$t(y^*)$ が $y^*\in\mathcal{Y}^*$ に依存しない正の定数で上から抑えられることに注目して，式 (4) の導出と同様に行える． □

以下では，これらの基本定理を用いることによって，符号長に関する従来の符号化定理（あるいは符号化逆定理）から直ちに，コストに関する符号化定理（あるいは符号化逆定理）が導けることを明らかにする．

4. ひずみを許容したコスト付符号化定理

本章では，ひずみ D (>0) を許したときに符号語のコストを最小にする符号 $f_D:\mathcal{X}^n\to\mathcal{Y}^*$，及び符号語コストレートを r (>0) で制限したときにひずみを最小にする符号 $f_r:\mathcal{X}^n\to\mathcal{Y}^*$ の存在を述べる．

まず，レートひずみ理論に関する基本的な定義と表記法について述べる．レートひずみ理論に関する詳細は，例えば文献 [6] を参考にされたい．以下では，$\mathcal{X}$ を（有限とは限らない）情報源アルファベット，$\hat{\mathcal{X}}$ を有限の再生アルファベットとし，情報源 $\{X_i\}_{i=1}^{\infty}$ は $\mathbf{X}$ と表記する．本論文で取り扱う情報源 $\mathbf{X}$ は定常エルゴード情報源に限定し，μ_X によって $\mathbf{X}$ に対応する確率測度を表す．

シンボル $x\in\mathcal{X}$ とシンボル $\hat{x}\in\hat{\mathcal{X}}$ との間のひずみ測度 d は，$\mathcal{X}\times\hat{\mathcal{X}}\to[0,\infty)$ の可測関数であるとする．このとき，ブロックひずみを次のように定義する．

［定義 6］ 系列 $x^n\in\mathcal{X}^n$ と系列 $\hat{x}^n\in\hat{\mathcal{X}}^n$ との間のブロックひずみは，

$$d_n(x^n,\hat{x}^n)\equiv\frac{1}{n}\sum_{i=1}^{n}d(x_i,\hat{x}_i)\qquad(7)$$

によって定義される． □

レートひずみ関数を次のように定義する．

［定義 7］ 情報源 $\mathbf{X}$ のレートひずみ関数 $R_X(D)$，及びひずみレート関数 $D_X(R)$ は

$$R_X(D) \equiv \lim_{n\to\infty} \inf_{p^n \in \mathcal{P}_D^n} \frac{1}{n} I\left(X^n; \hat{X}^n\right)$$

$$D_X(R) \equiv \lim_{n\to\infty} \inf_{p^n \in \mathcal{P}_R^n} E_{p^n}\left\{d_n\left(X^n, \hat{X}^n\right)\right\}$$

によって定義される．ただし，$\mathcal{P}_D^n$ は $\mathcal{X}^n \times \hat{\mathcal{X}}^n$ 上の確率測度 p^n で，

(i) $\displaystyle\sum_{\hat{x}^n \in \hat{\mathcal{X}}^n} p^n(x^n, \hat{x}^n) = \mu_X^n(x^n)$

(ii) $E_{p^n}\left\{d_n\left(X^n, \hat{X}^n\right)\right\} \leqq D$

を満足するものの集合，$\mathcal{P}_R^n$ は

(i) $\displaystyle\sum_{\hat{x}^n \in \hat{\mathcal{X}}^n} p^n(x^n, \hat{x}^n) = \mu_X^n(x^n)$

(ii) $\dfrac{1}{n} I\left(X^n; \hat{X}^n\right) \leqq R$

を満足するものの集合であり，$E_{p^n}[\cdot]$ は測度 p^n に対する平均を表す．また，$I\left(X^n; \hat{X}^n\right)$ は X^n と $\hat{X}^n$ との間の相互情報量

$$I\left(X^n; \hat{X}^n\right) \equiv \sup_{\mathcal{F}} \sum_{F_i^n \times \hat{F}_i^n \in \mathcal{F}} p^n\left(F_i^n \times \hat{F}_i^n\right) \cdot \log \frac{p^n\left(F_i^n \times \hat{F}_i^n\right)}{\mu_X^n\left(F_i^n\right) \mu_{\hat{X}}^n\left(\hat{F}_i^n\right)}$$

である．ここで上限は，$F_i^n \subset \mathcal{X}^n$，$\hat{F}_i^n \subset \hat{\mathcal{X}}^n$，$F_i^n \times \hat{F}_i^n \subset \mathcal{X}^n \times \hat{\mathcal{X}}^n$ がそれぞれ μ_X^n，$\mu_{\hat{X}}^n$，p^n に関して可測集合であるような $\mathcal{X}^n \times \hat{\mathcal{X}}^n$ 上のすべての分割 $\mathcal{F}$ に関してとる．ただし，$\mathcal{P}_D^n$ が空集合ならば，

$$\inf_{p^n \in \mathcal{P}_D^n} \frac{1}{n} I\left(X^n; \hat{X}^n\right) = \infty$$

とし，$\mathcal{P}_R^n$ が空集合ならば，

$$\inf_{p^n \in \mathcal{P}_R^n} E_{p^n}\left\{d_n\left(X^n, \hat{X}^n\right)\right\} = \infty$$

とする． □

4.1 ひずみを制限した場合

次の定理は，符号長に関する有ひずみ概収束符号化定理と概収束符号化逆定理から，コストに関する概収束符号化定理と概収束符号化逆定理が導けることを示している．

［定理 1］ 与えられた定常エルゴード情報源に対し，ひずみ D を許す n 次の語頭符号を $C_n \equiv (\phi_n, \psi_n, \mathcal{A}_n)$ で定義する．ただし，$\mathcal{A}_n$ は $\hat{\mathcal{X}}^n$ の部分集合，ϕ_n は任意の $x^n \in \mathcal{X}^n$ に対し，

$$\sup_{x^n \in \mathcal{X}^n} d_n(x^n, \phi_n(x^n)) \leqq D$$

を満足する $\mathcal{X}^n$ から $\mathcal{A}_n$ への可測関数，ψ_n は $\mathcal{A}_n$ から語頭条件を満足する $\mathcal{Y}^*$ の部分集合への 1 対 1 写像である [7]．このとき，もし

$$\limsup_{n\to\infty} \frac{1}{n} \ell(\psi_n(\phi_n(x^n))) \leqq R_X(D) \quad \text{a.s.} \qquad (8)$$

を満足する符号列 $\{C_n\}$ が存在すれば，任意の正則コストあるいは有限状態コストについて，

$$\limsup_{n\to\infty} \frac{1}{n} \mathrm{cst}\left(\tilde{\psi}_n(\phi_n(x^n))\right) \leqq \frac{R_X(D)}{\alpha_0} \quad \text{a.s.} \ (9)$$

を満足する符号列 $\left\{\left(\phi_n, \tilde{\psi}_n, \mathcal{A}_n\right)\right\}$ が存在する．

また逆に，ひずみ D を許す任意の符号列 $\{C_n\}$ は

$$\liminf_{n\to\infty} \frac{1}{n} \mathrm{cst}(\psi_n(\phi_n(x^n))) \geqq \frac{R_X(D)}{\alpha_0} \quad \text{a.s.} \ (10)$$

を満足する． □

（証明） ここでは，正則コストに限って証明を行う．有限状態コストについては，基本定理 1 のかわりに基本定理 2 を用いればよい．

(a) Achievability Part の証明

与えられた符号 $C_n = (\phi_n, \psi_n, \mathcal{A}_n)$ に対し，写像 ψ_n を基本定理 1 を満たす写像 $\tilde{\psi}_n$ に取り換えて得られる符号 $\tilde{C}_n = \left(\phi_n, \tilde{\psi}_n, \mathcal{A}_n\right)$ を考えよう．このとき，式 (3) から

$$\frac{1}{n} \mathrm{cst}\left(\tilde{\psi}_n(\phi_n(x^n))\right) \leqq \frac{\ell(\psi_n(\phi_n(x^n)))}{n\alpha_0} + \frac{d_2}{n}$$

が成り立つ．両辺の上極限を取り，式 (8) の仮定を用いると直ちに式 (9) が導かれる．

(b) Converse Part の証明

与えられた符号 $C_n = (\phi_n, \psi_n, \mathcal{A}_n)$ に対し，写像 ψ_n を基本定理 1 を満たす写像 $\hat{\psi}_n$ に取り換えて得られる符号 $\hat{C}_n = \left(\phi_n, \hat{\psi}_n, \mathcal{A}_n\right)$ は，ひずみ D を許す符号となるので，文献 [7] の Proposition 4 a) より

$$\liminf_{n\to\infty} \frac{1}{n} \ell\left(\hat{\psi}_n(\phi_n(x^n))\right) \geqq R_X(D) \quad \text{a.s.} \quad (11)$$

が成り立つ．ここで，式 (4) より，

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n}\mathrm{cst}(\psi_n(\phi_n(x^n)))&\geqq\liminf_{n\to\infty}\left\{\frac{1}{n\alpha_0}\ell\left(\hat{\psi}_n(\phi_n(x^n))\right)-\frac{1}{n\alpha_0}\right\}\\&\geqq\frac{R_X(D)}{\alpha_0}\quad\text{a.s.}\end{aligned}$$

が得られ，式 (10) が導かれた． □

（注意 3）定常エルゴード情報源やブロックひずみの仮定（式 (7)）は，定理 1 において本質的ではない．より一般的には，ある情報源とあるブロックひずみに対し，概収束符号化定理（式 (8)）及び概収束符号化逆定理（式 (11)）が成り立てば，定理 1 は成立する．そのような ブロックひずみの例は文献 [7] を参照されたい．

定理 1 と有ひずみユニバーサル符号化定理（例えば [8, Theorem 2]）を組み合せることで，定理 2 が得られる．なお，定理 2 は文献 [13] の定理 1 を正則コスト及び有限状態コストに拡張したものになっている．

［定理 2］ 正則コストあるいは有限状態コストとひずみ D が与えられたとき，定常エルゴード情報源 $\mathbf{X}$ からの出力系列 $x^n\in\mathcal{X}^n$ に対し，

$$\lim_{n\to\infty}\frac{1}{n}\mathrm{cst}(\psi_n(\phi_n(x^n)))=\frac{R_X(D)}{\alpha_0}\quad\text{a.s.}$$

を満足するようなユニバーサル符号列 $\{(\phi_n,\psi_n,\mathcal{A}_n)\}$ が存在する． □

4.2 符号語コストレートを制限した場合

次の定理は，符号語コストレートを r で制限した場合にも，同様な概収束符号化定理と概収束符号化逆定理が成り立つことを示している．

［定理 3］ 与えられた定常エルゴード情報源に対して，レート R を有する n 次の語頭符号を $C_n=(\phi_n,\psi_n,\mathcal{A}_n)$ によって表す．ただし，$|\mathcal{A}_n|\leqq\left|\hat{\mathcal{X}}\right|^{nR}$ である．このとき，もし

$$\limsup_{n\to\infty}d_n(x^n,\phi_n(x^n))\leqq D_X(R)\quad\text{a.s.}\qquad(12)$$

を満足する符号列 $\{C_n\}$ が存在すれば，式 (12) を満足し，かつ任意の正則コストあるいは有限状態コストと任意の $\varepsilon>0$ に対し，1 記号当りのコストとして定義される符号語コストレートが

$$\frac{1}{n}\mathrm{cst}\left(\tilde{\psi}_n(\phi_n(x^n))\right)\leqq\frac{R}{\alpha_0}+\varepsilon\quad\forall n>N(\varepsilon)\ (13)$$

を満足する符号列 $\left\{\left(\phi_n,\tilde{\psi}_n,\mathcal{A}_n\right)\right\}$ が存在する．

また逆に，符号語コストレートを r に制限した任意の語頭符号列 $\{C_n\}$ は

$$\liminf_{n\to\infty}d_n(x^n,\phi_n(x^n))\geqq D_X(\alpha_0 r)\quad\text{a.s.}\qquad(14)$$

を満足する． □

（証明）定理 1 と同様に，コストが正則コストであると仮定して証明を行う．

(a) Achievability Part の証明

仮定を満足する符号 $C_n=(\phi_n,\psi_n,\mathcal{A}_n)$ として，ψ_n が符号長 $\lceil nR\rceil$ の固定長符号であるとしても一般性を失わない．このとき，基本定理 1 より，ある語頭符号 $\tilde{\psi}_n:\mathcal{A}_n\to\mathcal{Y}^*$ が存在して，任意の $a\in\mathcal{A}_n$ について

$$\alpha_0\mathrm{cst}\left(\tilde{\psi}_n(a)\right)\leqq\lceil nR\rceil+\alpha_0 d_2<nR+\alpha_0 d_2+1$$

を満足する．これから直ちに，十分大きな n に対し，符号 $C'_n=\left(\phi_n,\tilde{\psi}_n,\mathcal{A}_n\right)$ が式 (12) 及び式 (13) を満足することがわかる．

(b) Converse Part の証明

符号語コストレートを r で制限した語頭符号 $C_n=(\phi_n,\psi_n,\mathcal{A}_n)$ において，定理 1 の Converse Part の証明と同様にして，語頭符号 $\hat{\psi}_n:\mathcal{A}_n\to\mathcal{Y}^*$ を構成すると，すべての $\hat{x}^n\in\mathcal{A}_n$ について

$$\begin{aligned}\frac{1}{n}\ell\left(\hat{\psi}_n(\hat{x}^n)\right)&\leqq\frac{\alpha_0}{n}\mathrm{cst}(\psi_n(\hat{x}^n))+\frac{1}{n}\\&\leqq\alpha_0 r+\frac{1}{n}\end{aligned}$$

が成り立つ．したがって，符号 $\hat{C}_n=\left(\phi_n,\hat{\psi}_n,\mathcal{A}_n\right)$ はたかだかレート $\alpha_0 r+1/n$ を有するので，文献 [7] の Proposition 2 a) と $D_X(R)$ の連続性から

$$\begin{aligned}\liminf_{n\to\infty}d_n(x^n,\phi_n(x^n))&\geqq\lim_{n\to\infty}D_X(\alpha_0 r+1/n)\\&=D_X(\alpha_0 r)\quad\text{a.s.}\end{aligned}$$

が得られ，式 (14) が導かれた． □

（注意 4）定理 1 と同様に，定常エルゴード情報源とブロックひずみの仮定は，定理 3 でも本質的ではない．一般に，ある情報源とあるブロックひずみに対し，概収束符号化定理とその逆定理が成り立てば，定理 3 は成立する．

定理 3 と有ひずみユニバーサル符号化定理（例え

ば [8, Theorem 1]）を組み合せることで，定理 4 が得られる．なお，定理 4 は，文献 [13] の定理 2 を正則コスト及び有限状態コストに拡張したものになっている．

［定理 4］ 正則コストあるいは有限状態コストが与えられたとき，符号語コストレートが r に制限された符号で，定常エルゴード情報源 $\mathbf{X}$ からの出力系列 $x^n \in \mathcal{X}^n$ に対し，

$$\lim_{n\to\infty} d_n(x^n, \phi_n(x^n)) = D_X(\alpha_0 r) \quad \text{a.s.}$$

を満足するようなユニバーサル符号列 $\{(\phi_n, \psi_n, \mathcal{A}_n)\}$ が存在する． □

5. 無ひずみコスト付符号化定理

本章では，ひずみを許さないときに符号語のコストを最小にする符号 $f : \mathcal{X}^n \to \mathcal{Y}^*$ の存在と平均コスト冗長度について述べる．なお，本章を通じて，$\mathcal{X}$ は有限集合とする．

5.1 コスト付符号化定理

次の定理が，無ひずみ符号化における主要成果である．

［定理 5］ 与えられた定常エルゴード情報源 $\mathbf{X}$ に対して，n 次の無ひずみ符号を $f_n : \mathcal{X}^n \to \mathcal{Y}^*$ で定義する．このとき，もし

$$\limsup_{n\to\infty} \frac{1}{n}\ell(f_n(x^n)) \leqq H(\mathbf{X}) \quad \text{a.s.} \tag{15}$$

を満足する符号列 $\{f_n\}$ が存在すれば，任意の正則コストあるいは有限状態コストについて，

$$\limsup_{n\to\infty} \frac{1}{n}\mathrm{cst}\left(\tilde{f}_n(x^n)\right) \leqq \frac{H(\mathbf{X})}{\alpha_0} \quad \text{a.s.} \tag{16}$$

を満足する無ひずみ符号列 $\left\{\tilde{f}_n\right\}$ $\left(\tilde{f}_n : \mathcal{X}^n \to \mathcal{Y}^*\right)$ が存在する．ただし，$H(\mathbf{X})$ は定常エルゴード情報源 $\mathbf{X}$ のエントロピーレートであり，

$$H(\mathbf{X}) = \lim_{k\to\infty} -\frac{1}{k} \sum_{x^k \in \mathcal{X}^k} \mu_X^k\left(x^k\right) \log \mu_X^k\left(x^k\right)$$

によって定義される．

また逆に，任意の無ひずみ符号列 $\{f_n\}$ は

$$\liminf_{n\to\infty} \frac{1}{n}\mathrm{cst}(f_n(x^n)) \geqq \frac{H(\mathbf{X})}{\alpha_0} \quad \text{a.s.} \tag{17}$$

を満足する． □

定理 5 は定理 1 の特別な場合（$\mathcal{X} = \hat{\mathcal{X}}$，$d$：ハミング距離，$D = 0$）なので証明を省略する．

定理 5 と無ひずみユニバーサル符号化定理（例えば文献 [3, Theorem 4]）を組み合せることで，定理 6 が得られる．なお，定理 6 は，文献 [10] の Theorem 2 を正則コスト及び有限状態コストに拡張したものになっている．

［定理 6］ 正則コストあるいは有限状態コストが与えられたとき，定常エルゴード情報源 $\mathbf{X}$ からの出力系列 $x^n \in \mathcal{X}^n$ に対し，

$$\lim_{n\to\infty} \frac{1}{n}\mathrm{cst}(f_n(x^n)) = \frac{H(\mathbf{X})}{\alpha_0} \quad \text{a.s.}$$

を満足するような無ひずみユニバーサル符号列 $\{f_n\}$ が存在する． □

5.2 平均コスト冗長度

本節では，符号語のコストを最小にする符号の平均コスト冗長度について考察する．なお，本節に限り，情報源を $\mathcal{X}$ 上のユニフィラー情報源あるいは FSMX 情報源 [5] に制限し，語頭符号 $f_n : \mathcal{X}^n \to \mathcal{Y}^*$ は正規（regular），すなわち，不等式

$$\ell(f_n(x^n)) \geqq -\log \mu_X^n(x^n)$$

をすべての $x^n \in \mathcal{X}^n$ について満足すると仮定する [5].

平均コスト冗長度は次のように定義される．

［定義 8］ 語頭符号 $f_n : \mathcal{X}^n \to \mathcal{Y}^*$ の平均コスト冗長度 $r_n^{\mathrm{cst}}(f_n)$ を

$$r_n^{\mathrm{cst}}(f_n) = \frac{1}{n} E_{\mu_X^n}[\mathrm{cst}(f_n(X^n))] - \frac{H(\mathbf{X})}{\alpha_0}$$

によって定義する． □

このとき，無ひずみユニバーサル符号の平均コスト冗長度に関して次の定理が成り立つ．

［定理 7］ 正則コストあるいは有限状態コストが与えられたとき，S 個の状態を有するユニフィラー情報源に対し，無ひずみユニバーサル符号 $f_n : \mathcal{X}^n \to \mathcal{Y}^*$ は，任意の $\varepsilon > 0$ と十分大きな n について

$$r_n^{\mathrm{cst}}(f_n) \geqq \frac{S - \varepsilon}{2\alpha_0 n} \log n \tag{18}$$

を満足する．

また逆に，

$$r_n^{\mathrm{cst}}\left(\hat{f}_n\right) \leqq \frac{S}{2\alpha_0 n}\log n + O\left(\frac{1}{n}\right) \qquad (19)$$

を満足するような無ひずみユニバーサル符号 $\hat{f}_n : \mathcal{X}^n \to \mathcal{Y}^*$ が存在する．□

（証明）式 (18) の証明は，文献 [5] の Theorem 1 と基本定理 1 または 2 を組み合せることで容易に行える．他方，式 (19) の証明は，文献 [5] の Theorem 2 と基本定理 1 または 2 を組み合せればよい．□

本節では，コストレート及び平均コスト冗長度についてのみ考察を行ったが，ミニマックス冗長度 [14] やコスト付ユニバーサル符号の存在条件 [12] などについても基本定理を用いることで同様の定理が容易に導けることに注意しておく．

6. むすび

本論文では，指定された範囲内のひずみを許した場合に，正則コストあるいは有限状態コストによる符号系列のコストを最小にするユニバーサル符号，及び符号語コストレートを指定する範囲内で制限したときにひずみを最小にするユニバーサル符号の構成法を提案し，これらの符号が漸近的に最良であることを示した．更に，ひずみを許さない場合についても考察し，符号系列のコストを最小にするユニバーサル符号の存在を示すとともに，そのような符号が符号長及び平均コスト冗長度の点から漸近的に最良であることを示した．

今後の課題としては，(1) 文献 [15] に示されたようなコスト表現の更なる一般化，(2) 計算量，記憶量などの点で実用に耐え得るアルゴリズムの開発，(3) 代表的なランレングス制約に対してコスト容量を算出し，ランレングス制約を有する無雑音通信路の容量と比較することなどがあげられる．

謝辞　筆者らに文献 [9] を御教示下さった，電気通信大学大学院情報システム学研究科 韓太舜教授に深謝する．

文　献

[1] C.E. Shannon, “A mathematical theory of communication,” Bell Syst. Tech. J., vol.27, pp.379–423, 623–656, July Oct. 1948.
[2] I. Csiszár, “Simple proofs of some theorems on noiseless channels,” Inf. Control, vol.14, pp.285–298, 1969.
[3] J. Ziv and A. Lempel, “Compression of individual sequence via variable-rate coding,” IEEE Trans. Inf. Theory, vol.IT-24, no.5, pp.530–536, Sept. 1978.
[4] I. Csiszár and J.Körner, Information Theory: Coding Theorems for Discrete Memoryless Systems, Academic Press, New York, 1981.
[5] J. Rissanen, “Complexity of strings in the class of Markov sources,” IEEE Trans. Inf. Theory, vol.IT-32, no.4, pp.526–532, July 1986.
[6] R.M. Gray, Entropy and Information Theory, Springer-Verlag, New York, 1990.
[7] J.C. Kieffer, “Sample converses in source coding theory,” IEEE Trans. Inf. Theory, vol.37, no.2, pp.263–268, March 1991.
[8] E.-H. Yang and J.C. Kieffer, “Simple univarsal lossy data compression schemes derived from the Lempel-Ziv algorithm,” IEEE Trans. Inf. Theory, vol.42, no.1, pp.239–245, Jan. 1996.
[9] A. Kato, Kolmogorov Complexity with Cost and the Randomness, Ph.D. Dissertation, The Univ. of Electro-Communications, Chofu, Tokyo, 1997.
[10] K. Iwata, M. Morii, and T. Uyematsu, “An efficient universal coding algorithm for noiseless channel with symbols of unequal cost,” IEICE Trans. Fundamentals, vol.E80-A, no.11, pp.2232–2237, Nov. 1997.
[11] 橋本　猛，情報理論，培風館，東京，1997.
[12] T. Yoshida, T. Matsushima, and S. Hirasawa, “A universal code considering the codeword cost,” Proc. of Int. Symp. on Inf. Theory and its Applications, pp.165–168, Mexico-city, Mexico, Oct. 1998.
[13] 内田　理，植松友彦，“符号語のコストを考慮した有ひずみユニバーサル情報圧縮，” 信学論（A），vol.J82-A, no.1, pp.104–111, Jan. 1999.
[14] 韓　太舜，小林欣吾，情報と符号化の数理，培風館，東京，1999.
[15] T. Kawabata, “Noiseless source coding theorems with stationary and mixing cost functions,” 第 22 回情報理論とその応用シンポジウム資料，pp.351–354, 1999.

付　録

有限状態コストではない正則コストの例

Q^i によって $\mathcal{Y}^i$ 上の確率分布を表す．また，$\left\{Q^i\right\}_{i=1}^{\infty}$ によって整合性条件

$$Q^{i-1}(y_1 y_2 \cdots y_{i-1}) = \sum_{y_i \in \mathcal{Y}} Q^i(y_1 y_2 \cdots y_i)$$

を満足する確率分布の列を表すことにする．ただし，$\left\{Q^i\right\}_{i=1}^{\infty}$ はすべての $y^i \in \mathcal{Y}^i$ 及び $i = 1, 2, \cdots$ に対し，ある q_1, q_2 が存在し

$$0 < q_1 \leqq Q^i\left(y^i\right)/Q^{i-1}\left(y^{i-1}\right) \leqq q_2 < 1 \quad (\mathrm{A}\cdot 1)$$

を満足するものとする．このとき，この確率分布列から導かれるコスト

$$\mathrm{cst}\left(y_i | y^{i-1}\right) = -\log \frac{Q^i\left(y^i\right)}{Q^{i-1}\left(y^{i-1}\right)}$$

は i 及び y^{i-1} によらず定コスト容量 1 を有する．また，式 (A·1) から，直ちに式 (2) を満足することがわかり，このコストは正則コストである．ここで，$\{Q^i\}_{i=1}^{\infty}$ は整合性条件と式 (A·1) を満たすように選べばよいので，必ずしも有限状態コストとは成り得ない．

（平成 12 年 2 月 9 日受付，5 月 23 日再受付）

植松　友彦　（正員）

昭 57 東工大・工・電気電子卒．昭 59 同大大学院修士課程了．同年同大・工・電気電子助手．同講師を経て平 3 同助教授．平 4 北陸先端大・情報科学研究科助教授．平 9 東工大・工・電気電子工学科助教授．工博．情報理論，特にシャノン理論の研究に従事．昭 63 年度本会篠原記念学術奨励賞受賞．平 4 年度並びに平 7 年度本会論文賞受賞．著書「文書データ圧縮アルゴリズム入門」,「よくわかる通信工学」,「現代シャノン理論」など．IEEE，情報理論とその応用学会各会員．

川上　秀彦

平 10 東工大・工・情工卒．平 12 同大大学院修士課程了．同年（株）KDD 勤務．在学中，情報理論，特に情報源符号化の研究に従事．