信州大学大学院工学系研究科

修士論文

一般情報源に対する最大歪み測度を用いた忠実度規範付き情報源符号化定理

指導教員　西新 幹彦 准教授

専攻　電気電子工学専攻
学籍番号　11TA241A
氏名　原　寛貴

2013 年 2 月 21 日

目次

1 序章

1.1 研究の背景と目的

現代の情報社会において扱うデータ量は膨大なものとなってきている. これをそのまま伝送することは非常に効率が悪い. そこで, 情報を効率よく符号化し通信時間, 通信容量を抑える情報圧縮技術は必要不可欠なものである. また, 音声や画像などの本来連続的な情報は, デジタル回線において歪みなく伝送することは不可能である. これらの情報を伝送するために, 元の情報源とこれを符号化・復号化したものとの間に一定の歪みを許容し, その上でどれだけ圧縮できるかを考えることが現実的である. 本研究では, 歪み測度が平均歪みであるか最大歪みであるかの 2 種類を用意する. さらに予め与えられた歪み測度に対して, 受信者が受け取った情報源が元の情報源にどれだけ忠実であるかという忠実度規範を 4 種類用意する. そして, 許容する歪みの範囲内で忠実度規範を達成する最小符号化レートを求めることを目的とする. 本論文の主な内容は [2] の定理 5.1 に対してより詳細な証明を与えたこと（定理 1）, [1] の Problem 2.2.12 に対して証明を与えたこと（定理 3), 最大歪み測度に対して CK 最大忠実度規範で達成可能な最小符号化レートを導出したこと（定理 4）である.

1.2 構成

以降, 本論文では, 2 章で本研究の概要を述べ, 3 章では歪み測度に対してレート R が達成可能な最小符号化レートを与え, 4 章でまとめを述べる.

2 準備

2.1 情報源

情報源とは, 情報源アルファベットと呼ばれる有限集合 $\mathcal{X}$ に値をとる確率変数列 $\{X_i\}_{i=1}^{\infty}$ である. 本研究では特に断らない限り, 定常無記憶情報源

$$X^n = (X_1, X_2, \cdots, X_n) \tag{1}$$

のレート・歪み理論について考える.

2.2 符号

情報源 X^n の符号化を以下のように行う. まず, X^n に対する符号を

$$\mathcal{M}_n = \{1, 2, \cdots, M_n\}$$

として符号化を $\varphi_n : \mathcal{X}^n \to \mathcal{M}_n$ によって定める. $\varphi_n(x)$ を $\boldsymbol{x} \in \mathcal{X}^n$ の符号語と呼ぶ. 復元アルファベットとして有限集合 $\mathcal{Y}$ を用意し, 復号化を $\psi_n : \mathcal{M}_n \to \mathcal{Y}^n$ によって定める. ここで, 符号化レートを $r_n \triangleq \frac{1}{n} \log |\mathcal{M}_n|$ と定める. 情報源からの出力 $\boldsymbol{x} \in \mathcal{X}^n$ は符号器 φ_n により符号化された後, 復号器 ψ_n により $\boldsymbol{y} \in \mathcal{Y}^n$ と変換される. 本研究では x と y の間の情報源文字 1 個当りの歪み $d(x, y)$ を一定の大きさ以下に抑えた上で符号化レート r_n をできるだけ小さくしたい.

2.3 歪み測度

歪み測度と呼ばれる関数 $d : \mathcal{X} \times \mathcal{Y} \to [0, +\infty)$ を定め, $d(x, y)(x \in \mathcal{X}, y \in \mathcal{Y})$ を x と y の間の歪みと呼ぶ. 長さ n のデータ $\boldsymbol{x} = (x_1, x_2, \cdots, x_n) \in \mathcal{X}^n$ と $\boldsymbol{y} = (y_1, y_2, \cdots, y_n) \in \mathcal{Y}^n$ の間の歪み $d_n(\boldsymbol{x}, \boldsymbol{y})$ を平均歪み測度と最大歪み測度の 2 種類を定義する. ここで, d_n は系列の歪みであり, d は 1 シンボルの歪みである.

$$d_n(\boldsymbol{x}, \boldsymbol{y}) \triangleq \begin{cases} \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} d(x_i, y_i) & \text{(平均歪み測度)} \\ \displaystyle\max_{1 \le i \le n} d(x_i, y_i) & \text{(最大歪み測度)} \end{cases} \tag{2}$$

と定義する.

2.4 忠実度規範

符号器 φ_n によって, 出力系列 $X^n \in \mathcal{X}^n$ を符号化し, 復号器 ψ_n によって復号化して得られる系列を $Y^n \in \mathcal{Y}^n$ とする. すなわち $Y^n \triangleq \psi_n(\varphi_n(X^n))$ とする. このとき, 系列の歪み $d_n(X^n, Y^n)$ が小さい程, 符号器と復号器の列 (φ_n, ψ_n) は情報源に忠実であると考える. そこで, 任意の小さな数 $\epsilon > 0$ に対して, 情報源からの出力系列 X^n が D より大きな歪みを伴って復元される確率が ϵ 以下であることを意味する規範を最大忠実度規範と呼ぶ. また, 系列の歪みの平均が D 以下であることを意味する規範を平均忠実度規範と呼ぶ. 次に, この忠実度規範を 4 種類定義する.

定義 1 最大忠実度規範 D で達成可能とは

$$\text{p-}\limsup_{n \to \infty} d_n(X^n, Y^n) \le D \tag{3}$$

$$\limsup_{n \to \infty} \frac{1}{n} \log |\mathcal{M}_n| \le R \tag{4}$$

が成り立つことである. ただし, 確率変数列 $\{Z_n\}_n$ に対して

$$\text{p-}\limsup_{n \to \infty} Z_n \triangleq \inf\{\alpha | \lim_{n \to \infty} \Pr\{Z_n > \alpha\} = 0\} \tag{5}$$

と定める. これにより, (3), (4) は次のように言い換えられる.
任意の小さな数 $\epsilon > 0$, $\delta > 0$, δ_1 に対して, 十分大きなすべての n で

$$\Pr\{d_n(X^n, Y^n) > D + \delta_1\} < \epsilon$$
$$\frac{1}{n}\log|\mathcal{M}_n| < R + \delta$$

を満たす符号器と復号器の列 (φ_n, ψ_n) が存在することである.

定義２ 平均忠実度規範 D で達成可能とは

$$\limsup_{n\to\infty} \boldsymbol{E}[d_n(X^n, Y^n)] \le D$$
$$\limsup_{n\to\infty} \frac{1}{n}\log|\mathcal{M}_n| \le R$$

が成り立つことである. すなわち, 任意の小さな数 $\delta > 0$, $\delta_1 > 0$ に対して, 十分大きなすべての n で

$$\boldsymbol{E}[d_n(X^n, Y^n)] < D + \delta_1$$
$$\frac{1}{n}\log|\mathcal{M}_n| < R + \delta$$

を満たす符号器と復号器の列 (φ_n, ψ_n) が存在することである.

定義３ CK 最大忠実度規範 D で達成可能とは

$$\lim_{n\to\infty} \Pr\{d_n(X^n, Y^n) > D\} = 0$$
$$\limsup_{n\to\infty} \frac{1}{n}\log\mathcal{M}_n \le R$$

が成り立つことである. すなわち, 任意の小さな数 $\epsilon > 0$, $\delta > 0$ に対して, 十分大きなすべての n で

$$\Pr\{d_n(X^n, Y^n) > D\} < \epsilon$$
$$\frac{1}{n}\log|\mathcal{M}_n| < R + \delta$$

を満たす符号器と復号器の列 (φ_n, ψ_n) が存在することである.

定義４ CK 平均忠実度規範 D で達成可能とは, 任意の小さな数 $\delta > 0$ に対して, 十分大きなす

べての n で

$$\boldsymbol{E}[d_n(X^n, Y^n)] \le D$$
$$\frac{1}{n}\log|\mathcal{M}_n| < R + \delta$$

を満たす符号器と復号器の列 (φ_n, ψ_n) が存在することである.

3　レート・歪み理論

3.1　平均歪み測度に対してレート R が平均忠実度規範 D で達成可能

定義 5

$$R_1(D|X) \triangleq \inf\{R|\ \text{平均歪み測度に対して}\ R\ \text{は平均忠実度規範}\ D\ \text{で達成可能}\ \}$$

定理 1　定常無記憶情報源 X のレート・歪み関数 $R_1(D|X)$ は

$$R_1(D|X) = \min_{Y:\boldsymbol{E}d(X,Y)\le D} I(X;Y) \tag{6}$$

で得られる. また, 右辺の min は条件 $\boldsymbol{E}d(X,Y) \le D$ を満たす集合 $\mathcal{Y}$ の中に値をとるすべての確率変数 Y にわたる最小値を示す. このような確率変数全体は有界閉集合になるので最小値が存在することが保証される.

補題 1　(6) の右辺は D の関数として下に凸である.
（証明）D の関数を

$$f(D) \triangleq \min_{Y:\boldsymbol{E}d(X,Y)\le D} I(X;Y)$$

と定義する. $Y_1,\ Y_2$ をそれぞれ $\boldsymbol{E}d(X,Y_1) \le D_1,\ \boldsymbol{E}d(X,Y_2) \le D_2$ を満たす任意の確率変数とし, $\alpha_1 \ge 0,\ \alpha_2 \ge 0$ を $\alpha_1 + \alpha_2 = 1$ を満たす任意の定数とする. まず, X とは独立な確率変数 Q を $P_Q(1) = \alpha_1,\ P_Q(2) = \alpha_2$ と定め, 確率変数 Y を $Q = i(i = 1, 2)$ のとき $Y = Y_i$ と定めれば

$$\boldsymbol{E}d(X,Y) = \alpha_1\boldsymbol{E}d(X,Y_1) + \alpha_2\boldsymbol{E}d(X,Y_2) \le \alpha_1 D_1 + \alpha_2 D_2 \tag{7}$$

が成り立つ. また, Q と X は独立ということに注意すると, $I(X;Q) = 0$ であるため,

$$\begin{aligned}\alpha_1 I(X;Y_1) + \alpha_2 I(X;Y_2) &= I(X;Y|Q) \\ &= I(X;QY) \\ &\ge I(X;Y)\end{aligned} \tag{8}$$

を得る. (7) と (8) を合わせれば

$$\begin{aligned}\alpha_1 f(D_1)+\alpha_2 f(D_2) &= \alpha_1 \min_{Y_1:\boldsymbol{E}d(X,Y_1)\le D_1} I(X;Y_1)+\alpha_2 \min_{Y_2:\boldsymbol{E}d(X,Y_2)\le D_2} I(X;Y_2)\\ &= \min_{Y_1:\boldsymbol{E}d(X,Y_1)\le D_1} \alpha_1 I(X;Y_1)+\min_{Y_2:\boldsymbol{E}d(X,Y_2)\le D_2} \alpha_2 I(X;Y_2)\\ &= \min_{\substack{Y_1:\boldsymbol{E}d(X,Y_1)\le D_1\\ Y_2:\boldsymbol{E}d(X,Y_2)\le D_2}} (\alpha_1 I(X;Y_1)+\alpha_2 I(X;Y_2))\\ &\ge \min_{Y:\boldsymbol{E}d(X,Y)\le \alpha_1 D_1+\alpha_2 D_2} I(X,Y)\\ &= f(\alpha_1 D_1+\alpha_2 D_2)\end{aligned}$$

となり, $f(D)$ が D の関数として下に凸であることがわかる. ゆえに, $f(D)$ は D に関して連続である. □

（定理 1 の証明） 証明は [2] に負うが, より詳細に記述する.

(1)Direct part: 初めに, 条件 $\boldsymbol{E}d(X,Y)\le D-\gamma$ を満たす任意の Y を与えたときに, $R\triangleq I(X;Y)+2\gamma$ が平均忠実度規範 D で達成可能であることを示す. ここで, $\gamma>0$ は任意の小さい定数である. $X^nY^n \triangleq (X_1Y_1, X_2Y_2,\cdots,X_nY_n)$ を XY の定常独立系列として $X^n\triangleq(X_1,X_2,\cdots,X_n), Y_n\triangleq(Y_1,Y_2,\cdots,Y_n)$ とおく. そこで

$$T_n^{(1)} \triangleq \left\{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{X}^n\times\mathcal{Y}^n \middle| \left|\frac{1}{n}\log\frac{P_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})}{P_{Y^n}(\boldsymbol{y})}-I(X;Y)\right|<\gamma\right\} \tag{9}$$

$$T_n^{(2)} \triangleq \left\{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{X}^n\times\mathcal{Y}^n \middle| \left|d_n(\boldsymbol{x},\boldsymbol{y})-\boldsymbol{E}d(X,Y)\right|<\gamma\right\} \tag{10}$$

として $T_n\triangleq T_n^{(1)}\cap T_n^{(2)}$ とおく. そして, $\boldsymbol{E}d(X,Y)\le D-\gamma$ と (10) より

$$(\boldsymbol{x},\boldsymbol{y})\in T_n \text{ ならば } d_n(\boldsymbol{x},\boldsymbol{y})\le D \tag{11}$$

であることに気をつける. また, $X^nY^n=(X_1Y_1,X_2Y_2,\cdots,X_nY_n)$ が定常独立系列なので

$$\frac{1}{n}\log\frac{P_{Y^n|X^n}(Y^n|X^n)}{P_{Y^n}(Y^n)}=\frac{1}{n}\sum_{i=1}^{n}\log\frac{P_{Y|X}(Y_i|X_i)}{P_Y(Y_i)} \tag{12}$$

$$d_n(X^n,Y^n)=\frac{1}{n}\sum_{i=1}^{n}d(X_i,Y_i) \tag{13}$$

が成立する. そして, (12), (13) 右辺の和の中の各項は独立でその期待値は

$$\begin{aligned}&\boldsymbol{E}\left[\log\frac{P_{Y|X}(Y_i|X_i)}{P_Y(Y_i)}\right]=I(X;Y) \qquad (i=1,2,\cdots)\\ &\boldsymbol{E}d(X_i,Y_i)=\boldsymbol{E}d(X,Y) \qquad (i=1,2,\cdots)\end{aligned}$$

であり, その分散は有界である. 以下に分散が有界である事を示す. $Z \triangleq \frac{1}{n}\sum_{i=1}^{n} Z_i,\ Z_i \triangleq \log\frac{P_{Y|X}(Y_i|X_i)}{P_Y(Y_i)}$ とおく. Z の分散 $V(Z)$ は

$$\begin{aligned}
V(Z) &= \boldsymbol{E}(Z^2) - (\boldsymbol{E}(Z))^2 \\
&= \boldsymbol{E}\Bigg\{\bigg(\frac{1}{n}\sum_{i=1}^{n} Z_i\bigg)^2\Bigg\} - \Bigg\{\boldsymbol{E}\bigg(\sum_{i=1}^{n} Z_i\bigg)\Bigg\}^2 \\
&= \boldsymbol{E}\Bigg\{\bigg(\frac{1}{n}\sum_{i=1}^{n} Z_i\bigg)^2\Bigg\} - \Bigg\{\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{E}(Z_i))\Bigg\}^2 \\
&= \boldsymbol{E}\Bigg\{\bigg(\frac{1}{n}\sum_{i=1}^{n} Z_i\bigg)^2\Bigg\} - (\boldsymbol{E}[Z_1])^2
\end{aligned} \tag{14}$$

となる. ここで, (14) 右辺第 1 項は

$$\begin{aligned}
\boldsymbol{E}\Bigg\{\bigg(\frac{1}{n}\sum_{i=1}^{n} Z_i\bigg)^2\Bigg\} &= \frac{1}{n^2}\boldsymbol{E}\Bigg\{\bigg(\sum_{i=1}^{n} Z_i\bigg)^2\Bigg\} \\
&= \frac{1}{n^2}\boldsymbol{E}\{(Z_1 + Z_2 + \cdots + Z_n)^2\} \\
&= \frac{1}{n^2}\{n\boldsymbol{E}(Z^2) + (n^2 - n)\boldsymbol{E}(Z_1)^2\} \\
&= \frac{1}{n}\{\boldsymbol{E}(Z_1^2) + (n-1)\boldsymbol{E}(Z_1)^2\}
\end{aligned}$$

と変形できるので, 分散 $V(Z)$ は

$$\begin{aligned}
V(Z) &= \frac{1}{n}\{\boldsymbol{E}(Z_1^2) + (n-1)\boldsymbol{E}(Z_1)^2\} - \{\boldsymbol{E}(Z_1)\}^2 \\
&= \frac{1}{n}[\boldsymbol{E}(Z_1^2) + (n-1)\boldsymbol{E}(Z_1)^2 - n\{\boldsymbol{E}(Z_1)\}^2] \\
&= \frac{1}{n}[\boldsymbol{E}(Z_1) - \{\boldsymbol{E}(Z_1)\}^2] \\
&= \frac{1}{n}V(Z_1)
\end{aligned}$$

となり, 分散は有界である. よって, $T_n^{(1)},\ T_n^{(2)}$ の定義 (9), (10) に注意して Chebyshev の不等式を適用すれば, $n \to \infty$ によって

$$\begin{aligned}
\Pr\{X^nY^n \in T_n^{(1)}\} &= \Pr\Bigg\{\bigg|\frac{1}{n}\log\frac{P_{Y^n|X^n}(Y^n|X^n)}{P_{Y^n}(Y^n)} - I(X;Y)\bigg| < \gamma\Bigg\} \\
&= \Pr\Bigg\{\bigg|\frac{1}{n}\sum_{i=1}^{n}\log\frac{P_{Y|X}(Y_i|X_i)}{P_Y(Y_i)} - I(X;Y)\bigg| < \gamma\Bigg\} \\
&\geq 1 - \frac{\frac{1}{n}\sigma^2}{\gamma^2} \\
&\to 1
\end{aligned}$$

$$
\begin{aligned}
\Pr\{X^nY^n \in T_n^{(2)}\} &= \Pr\left\{\left| d_n(X^n, Y^n) - \boldsymbol{E}d(X,Y)\right| < \gamma\right\} \\
&= \Pr\left\{\left| \frac{1}{n}\sum_{i=1}^{n} d(X_i, Y_i) - \boldsymbol{E}d(Y,Y)\right| < \gamma\right\} \\
&\geq 1 - \frac{\frac{1}{n}\sigma^2}{\gamma^2} \\
&\to 1
\end{aligned}
$$

を得る. よって

$$
\begin{aligned}
\Pr\{X^nY^n \notin T_n\} &= \Pr\{X^nY^n \notin T_n^{(1)} \cap T_n^{(2)}\} \\
&\leq \Pr\{X^nY^n \notin T_n^{(1)}\} + \Pr\{X^nY^n \notin T_n^{(2)}\} \\
&\to 0
\end{aligned}
$$

であることから $n \to \infty$ で

$$
\Pr\{X^nY^n \in T_n\} \to 1 \tag{15}
$$

が成立する.

さて, ここで

$$
M_n \triangleq e^{nR} = e^{n(I(X;Y)+2\gamma)}
$$

として, $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_{M_n} \in \mathcal{Y}^n$ を互いに独立に分布 P_{Y_n} に従って, 発生させる. さらに, これらを符号語として $\mathcal{C} \triangleq \{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_{M_n}\}$ とおく (ランダム符号化). 次に, 各 $\boldsymbol{x} \in \mathcal{X}^n$ に対して

$$
d_n(\boldsymbol{x}, \boldsymbol{v}_{i_0}) = \min_{1 \leq i \leq M_n} d_n(\boldsymbol{x}, \boldsymbol{v}_i) \tag{16}
$$

となる番号 i_0 を用いて符号器 φ_n を $\varphi_n(\boldsymbol{x}) = i_0$ と定義する. すなわち

$$
\varphi(\boldsymbol{x}) \triangleq \arg \min_{1 \leq i \leq M_n} d_n(\boldsymbol{x}_i, \boldsymbol{v}_i)
$$

である. また, 復号器 ψ_n は符号器から番号 i を受け取ったとき $\psi_n(i) = \boldsymbol{v}_i$ と定める. このように定義した符号器と復号器の組 (φ_n, ψ_n) に対して

$$
\overline{P_e}^{(n)} \triangleq \Pr\left\{ d_n(X^n, \psi_n(\varphi_n(X^n))) > D \right\} \tag{17}
$$

とおく. ここで, $A \triangleq \{d_n(X^n, \psi_n(\varphi_n(X^n))) > D\}$, $\overline{A} \triangleq \{d_n(X^n, \psi_n(\varphi_n(X^n))) \leq D\}$ とすれば

$$
\begin{aligned}
\overline{\rho_n} &\triangleq \boldsymbol{E}_{\mathcal{C}}\boldsymbol{E} d_n(X_n, \psi_n(\varphi_n(X^n))) \\
&= \Pr\{A\}\boldsymbol{E}_{\mathcal{C}}\boldsymbol{E}\left[d_n(X_n, \psi_n(\varphi_n(X^n))) \middle| A \right] + \Pr\{\overline{A}\}\boldsymbol{E}_{\mathcal{C}}\boldsymbol{E}\left[d_n(X_n, \psi_n(\varphi_n(X^n))) \middle| \overline{A} \right] \\
&\leq \overline{P_e}^{(n)} d_{\max} + D
\end{aligned} \tag{18}
$$

を得る. ただし, $d_{\max} \triangleq \max_{x\in\mathcal{X}, y\in\mathcal{Y}} d(x,y)$ であり, $\boldsymbol{E}$ は X^n に関する期待値を示し, $\boldsymbol{E}_{\mathcal{C}}$ はランダム符号 $\mathcal{C}$ に関する期待値を表す.

ここで, $\overline{P}_e^{(n)}$ の値を評価する. そのために, $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_{M_n}$ は互いに独立で同一の分布 P_{Y^n} に従うことを考慮すると, 符号 (φ_n, ψ_n) の定義と (17) から

$$
\begin{aligned}
\overline{P}_e^{(n)} &= \sum_{\boldsymbol{x}\in\mathcal{X}^n} \mathrm{Pr}\{X^n = \boldsymbol{x} \text{ かつ } d_n(\boldsymbol{x}, \psi_n(\varphi_n(\boldsymbol{x}))) > D\} \\
&= \sum_{\boldsymbol{x}\in\mathcal{X}^n} P_{X^n}(\boldsymbol{x}) \mathrm{Pr}\Big\{ d_n(\boldsymbol{x}, \psi_n(\varphi_n(\boldsymbol{x}))) > D \Big\} \\
&= \sum_{\boldsymbol{x}\in\mathcal{X}^n} P_{X^n}(\boldsymbol{x}) \mathrm{Pr}\Big\{ \text{どの } \boldsymbol{v}_i (i = 1, 2, \cdots, M_n) \text{ に対しても } d_n(\boldsymbol{x}, \boldsymbol{v}_i) > D \Big\} \\
&= \sum_{\boldsymbol{x}\in\mathcal{X}^n} P_{X^n}(\boldsymbol{x}) \prod_{i=1}^{M_n} \mathrm{Pr}\Big\{ d_n(\boldsymbol{x}, \boldsymbol{v}_i) > D \Big\} \\
&= \sum_{\boldsymbol{x}\in\mathcal{X}^n} P_{X^n}(\boldsymbol{x}) \left(\mathrm{Pr}\Big\{ d_n(\boldsymbol{x}, Y^n) > D \Big\} \right)^{M_n} \\
&= \sum_{\boldsymbol{x}\in\mathcal{X}^n} P_{X^n}(\boldsymbol{x}) \left(1 - \mathrm{Pr}\Big\{ d_n(\boldsymbol{x}, Y^n) \le D \Big\} \right)^{M_n} \\
&= \sum_{\boldsymbol{x}\in\mathcal{X}^n} P_{X^n}(\boldsymbol{x}) \left(1 - \sum_{\boldsymbol{y}\in\mathcal{Y}^n} P_{Y^n}(\boldsymbol{y}) \mathbf{1}\Big\{ d_n(\boldsymbol{x}, \boldsymbol{y}) \le D \Big\} \right)^{M_n} \qquad (19)
\end{aligned}
$$

が導かれる. ここで

$$
J(\boldsymbol{x}, \boldsymbol{y}) \triangleq \begin{cases} 1 & \text{for } (\boldsymbol{x}, \boldsymbol{y}) \in T_n \\ 0 & \text{otherwise} \end{cases} \qquad (20)
$$

とおくと, (11) より

$$
J(\boldsymbol{x}, \boldsymbol{y}) \le \mathbf{1}\{d_n(\boldsymbol{x}, \boldsymbol{y}) \le D\}
$$

が成立する. ゆえに, (19) の $\overline{P}_e^{(n)}$ は

$$
\overline{P}_e^{(n)} \le \sum_{\boldsymbol{x}\in\mathcal{X}^n} \mathrm{P}_{X^n}(\boldsymbol{x}) \left(1 - \sum_{\boldsymbol{y}\in\mathcal{Y}^n} P_{Y^n}(\boldsymbol{y}) J(\boldsymbol{x}, \boldsymbol{y}) \right)^{M_n} \qquad (21)
$$

と上から抑えることができる. しかし, (9) より $(\boldsymbol{x},\boldsymbol{y}) \in T_n$ ならば

$$\begin{aligned}
\frac{1}{n}\log\frac{P_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})}{P_{Y^n}(\boldsymbol{y})} - I(X;Y) < \gamma \\
\log\frac{P_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})}{P_{Y^n}(\boldsymbol{y})} &< n\{I(X;Y)+\gamma\} \\
\frac{P_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})}{P_{Y^n}(\boldsymbol{y})} &< e^{n(I(X;Y)+\gamma)} \\
P_{Y^n}(\boldsymbol{y}) &> e^{-n(I(X;Y)+\gamma)}P_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})
\end{aligned}$$

なので, 任意の $(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ に対して

$$P_{Y^n}(\boldsymbol{y})J(\boldsymbol{x},\boldsymbol{y}) > e^{-n(I(X;Y)+\gamma)}P_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})J(\boldsymbol{x},\boldsymbol{y})$$

が成り立つ. これを (21) の右辺に代入すれば

$$\overline{P}_e^{(n)} \leq \sum_{\boldsymbol{x}\in\mathcal{X}^n} P_{X^n}(\boldsymbol{x})\Biggl(1 - e^{-n(I(X;Y)+\gamma)}\sum_{\boldsymbol{y}\in\mathcal{Y}^n} P_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})J(\boldsymbol{x},\boldsymbol{y})\Biggr)^{M_n} \tag{22}$$

を得る. ここで, 不等式

$$(1-x)^y \leq e^{-xy} \qquad (0 \leq x \leq 1, y \geq 0) \tag{23}$$

を用いて $y = M_n,\ x = e^{-n(I(X;Y)+\gamma)}\sum_{\boldsymbol{y}\in\mathcal{Y}^n} P_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})J(\boldsymbol{x},\boldsymbol{y})$ と代入すれば

$$\begin{aligned}
\overline{P}_e^{(n)} &\leq \sum_{\boldsymbol{x}\in\mathcal{X}^n} P_{X^n}(\boldsymbol{x})\exp\Biggl\{-M_n e^{-n(I(X;Y)+\gamma)}\sum_{\boldsymbol{y}\in\mathcal{Y}^n} P_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})J(\boldsymbol{x},\boldsymbol{y})\Biggr\} \\
&= \sum_{\boldsymbol{x}\in\mathcal{X}^n} P_{X^n}(\boldsymbol{x})\exp\Biggl\{-e^{n\gamma}\sum_{\boldsymbol{y}\in\mathcal{Y}^n} P_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})J(\boldsymbol{x},\boldsymbol{y})\Biggr\}
\end{aligned}$$

となる. さらに, 不等式

$$e^{-xy} \leq 1 - y + e^{-x} \qquad (x \geq 0, 0 \leq y \leq 1)$$

に対して $y=\sum_{\boldsymbol{y}\in\mathcal{Y}^n} P_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})J(\boldsymbol{x},\boldsymbol{y}),\ x=e^{n\gamma}$ を用いると

$$\begin{aligned}
\overline{P}_e^{(n)} &\le \sum_{\boldsymbol{x}\in\mathcal{X}^n} P_{X^n}(\boldsymbol{x})\Big\{1-\sum_{\boldsymbol{y}\in\mathcal{Y}^n} P_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})J(\boldsymbol{x},\boldsymbol{y})+e^{-e^{n\gamma}}\Big\} \\
&= 1-\sum_{\boldsymbol{x}\in\mathcal{X}^n} P_{X^n}(\boldsymbol{x})\sum_{\boldsymbol{y}\in\mathcal{Y}^n} P_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})J(\boldsymbol{x},\boldsymbol{y})+e^{-e^{n\gamma}} \\
&= 1-\sum_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{X}^n\mathcal{Y}^n} P_{X^nY^n}(\boldsymbol{x},\boldsymbol{y})J(\boldsymbol{x},\boldsymbol{y})+e^{-e^{n\gamma}} \\
&= 1-\sum_{(\boldsymbol{x},\boldsymbol{y})\in T_n} P_{X^nY^n}(\boldsymbol{x}\boldsymbol{y})+e^{-e^{n\gamma}} \\
&= 1-P(X^nY^n\in T_n)+e^{-e^{n\gamma}} \\
&= P(X^nY^n\notin T_n)+e^{-e^{n\gamma}} \\
&\to 0 \quad (n\to\infty)
\end{aligned} \tag{24}$$

となる. 一方, (18) により少なくとも 1 つのランダムでない符号器と復号器の組 (φ_n,ψ_n) が存在し

$$\rho_n \triangleq \boldsymbol{E}d_n(X_n,\psi_n(\varphi_n(X^n))) \le D+\overline{P_e}^{(n)}d_{\max}$$

を満たさなければならない. よって (φ_n,ψ_n) の列を適切に選ぶことによって

$$\begin{aligned}
\limsup_{n\to\infty}\rho_n &= \limsup_{n\to\infty}\boldsymbol{E}d_n(X_n,\psi_n(\varphi_n(X^n))) \\
&\le D
\end{aligned}$$

とできる. ところが, $M_n=e^{n(I(X;Y)+2\gamma)}$ なので

$$\limsup_{n\to\infty}\frac{1}{n}\log M_n \le I(X;Y)+2\gamma$$

が成立している. よって, $R=I(X;Y)+2\gamma$ は平均忠実度規範 D で達成可能である. さらに, Y は条件 $\boldsymbol{E}d(X,Y)\le D-\gamma$ を満たす範囲で任意にとったので

$$R_1(D|X) \le \min_{Y:\boldsymbol{E}d(X,Y)\le D-\gamma} I(X;Y)+2\gamma$$

である. ここで, 補題 1 によりこの不等式の右辺は γ に関して連続である. $\gamma>0$ は任意だったため $\gamma\to 0$ とすれば

$$R_1(D|X) \le \min_{Y:\boldsymbol{E}d(X,Y)\le D} I(X;Y)$$

を得る.

(2) Converse part: R が平均忠実度規範 D で達成可能であるとすれば

$$\limsup_{n\to\infty} \boldsymbol{E} d_n(X^n, \psi_n(\varphi_n(X^n))) \le D \tag{25}$$

$$\limsup_{n\to\infty} \frac{1}{n}\log M_n \le R \tag{26}$$

なる符号器と復号器の組 (φ_n, ψ_n) が存在する. そこで,

$$Y^n \triangleq (Y_1^{(n)}, Y_2^{(n)}, \cdots, Y_n^{(n)}) = \psi_n(\varphi_n(X^n))$$

とおけば, 符号化関数 φ_n のとり得る値の数は高々 M_n, よって Y^n のとり得る値の数も高々 M_n なので

$$\begin{aligned}\log M_n &\ge H(Y^n)\\ &\ge I(X^n;Y^n)\\ &= H(X^n) - H(X^n \mid Y^n)\\ &= \sum_{i=1}^{n} H(X_i) - \sum_{i=1}^{n} H(X_i \mid Y^n X^{i-1})\end{aligned}$$

となる. ただし, 最後の等式で情報源の無記憶性とエントロピーのチェイン則を用いて, $X^{i-1} = (X_1, X_2, \cdots, X_{i-1})$ とおいた. そこで

$$H(X_i|Y^n X^{i-1}) \le H(X_i|Y_i^{(n)})$$

に注意すると

$$\begin{aligned}\log M_n &\ge \sum_{i=1}^{n} H(X_i) - \sum_{i=1}^{n} H(X_i|Y_i^{(n)})\\ &= \sum_{i=1}^{n} I(X_i;Y_i^{(n)})\end{aligned}$$

を得る. ゆえに

$$\frac{1}{n}\log M_n \ge \frac{1}{n}\sum_{i=1}^{n} I(X_i;Y_i^{(n)}) \tag{27}$$

となる. そして, 確率変数 $Q_n \in \{1, 2, \cdots, n\}$ の分布を $P_{Q_n}(i) = 1/n (i = 1, 2, \cdots, n)$ とし, 確率変数 $XY^{(n)}$ を $Q_n = i$ のとき $XY^{(n)} \triangleq X_i Y_i^{(n)}$ と定義すると (26) は

$$\frac{1}{n}\log M_n \ge I(X;Y^{(n)}|Q_n) \tag{28}$$

とかける. しかし, 情報源の定常性より

$$
\begin{aligned}
P_{Q_n X}(i,x) &= P_{Q_n}(i) \cdot P_{X|Q_n}(x|i) \\
&= P_{Q_n}(i) \cdot P_{X_i}(x) \\
&= P_{Q_n}(i) \cdot P_X(x)
\end{aligned}
$$

となり, Q_n と X は独立となる. よって

$$
I(X;Y^{(n)}|Q_n) = I(X;Q_n Y^{(n)}) \geq I(X;Y^{(n)})
$$

となる. したがって, (28) は

$$
\frac{1}{n}\log M_n \geq I(X;Y^{(n)}) \tag{29}
$$

となる. 一方

$$
\begin{aligned}
\boldsymbol{E} d_n(X^n, \psi_n(\varphi_n(X^n))) &= \boldsymbol{E} d_n(X^n, Y^n) \\
&= \boldsymbol{E}\left\{\frac{1}{n}\sum_{i=1}^{n} d(X_i, Y_i^{(n)})\right\} \\
&= \sum_{i=1}^{n}\frac{1}{n}\boldsymbol{E}\{d(X_i, Y_i)\} \\
&= \sum_{i=1}^{n} P_{Q_n}\boldsymbol{E}\{d(X, Y^{(n)})|Q_n = i\} \\
&= \boldsymbol{E} d(X, Y^{(n)})
\end{aligned}
$$

なので, (25) から

$$
\limsup_{n\to\infty} \boldsymbol{E} d(X, Y^{(n)}) \leq D
$$

を得る. ゆえに, 任意に小さい定数 $\gamma > 0$ に対して十分大きなすべての n で

$$
\boldsymbol{E} d(X, Y^{(n)}) \leq D + \gamma
$$

となる. そうすると, (29) より

$$
\begin{aligned}
\frac{1}{n}\log M_n &\geq I(X;Y\ (n)) \\
&\geq \min_{Y:\boldsymbol{E}d(X,Y)\leq D+\gamma} I(X;Y)
\end{aligned}
$$

となる. よって, (26) から n が十分大きければ

$$
R \geq \min_{Y:\boldsymbol{E}d(X,Y)\leq D+\gamma} I(X;Y) - \gamma
$$

が得られる. しかし, 補題 1 により, この不等式の右辺は γ に関して連続である. そこで, $\gamma > 0$ は任意であったことに注意して $\gamma \to 0$ とすれば

$$R \geq \min_{Y:\boldsymbol{E}d(X,Y)\leq D} I(X;Y)$$

が結論される. R は平均忠実度規範 D で達成可能な任意のレートと仮定されていたので

$$R_1(D|X) \geq \min_{Y:\boldsymbol{E}d(X,Y)\leq D} I(X;Y)$$

が成り立つ. □

3.2 平均歪み測度に対してレート R が CK 最大忠実度規範 D で達成可能

定義 6

$$R_2(D|X) \triangleq \inf\{R|\ \text{平均歪み測度に対して}\ R\ \text{は CK 最大忠実度規範}\ D\ \text{で達成可能}\ \}$$

定理 2 ([1]) 定常無記憶情報源 X のレート・歪み関数 $R_2(D|X)$ は

$$R_2(D|X) = \min_{\boldsymbol{Y}:\boldsymbol{E}[d(X,Y)]\leq D} I(X;Y) \tag{30}$$

で与えられる. この定理は定理 29 として付録で証明する.

3.3 最大歪み測度に対してレート R が CK 最大忠実度規範 D で達成可能

定義 7

$$R_3(D|X) \triangleq \inf\{R|\ \text{最大歪み測度に対して}\ R\ \text{は CK 最大忠実度規範}\ D\ \text{で達成可能}\ \} \tag{31}$$

定理 3 ([1]) 定常無記憶情報源 X のレート・歪み関数 $R_3(D|X)$ は

$$R_3(D|X) = \min_{Y:\Pr\{d(X,Y)\leq D\}=1} I(X;Y) \tag{32}$$

で与えられる.

定理 3 の証明は文献 [1] では与えられてない. これは次のように証明することができる.
(証明) まず

$$d'(x,y) \triangleq \begin{cases} 0 & \text{if } d(x,y) \leq D \\ 1 & \text{otherwise} \end{cases} \tag{33}$$

と定める. このとき

$$\boldsymbol{E}[d'(x,y)] = 0 \ \Leftrightarrow\ \Pr\{d(x,y) \leq D\} = 1$$

が成り立つ. よって

$$\min_{Y:\boldsymbol{E}[d'(x,y)]=0} I(X;Y) = \min_{Y:\Pr\{d(x,y)\leq D\}=1} I(X;Y)$$

となる. 左辺は「歪み測度 d' による平均歪みに対して符号化レート R は CK 最大忠実度規範 0 で達成可能」の意味でレートの下限である. つまり, 定理 2 より $R_2(0|X)$ に等しい. またこれは, 任意の $\delta > 0$ に対して十分大きなすべての n で次の条件を満たす符号 (φ_n, ψ_n) が存在することを意味する.

$$\begin{aligned} &\frac{1}{n}\log\|\varphi_n\| < R + \delta \\ &\Pr\left\{\frac{1}{n}\sum_{i=1}^{n} d'(X_i, \hat{Y}_i^{(n)}) = 0\right\} > 1 - \delta \end{aligned} \tag{34}$$

ただし $\hat{Y}^{(n)} \triangleq \hat{Y}_1^{(n)} \cdots \hat{Y}_n^{(n)} = \psi_n(\varphi_n(X^n))$ とおいた. ここで

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} d'(X_i, \hat{Y}_i^{(n)}) = 0 &\Leftrightarrow \sum_{i=1}^{n} d'(X_i, \hat{Y}_i^{(n)}) = 0 \\ &\Leftrightarrow i = 1, \cdots, n \text{ に対して } d'(X_i, \hat{Y}_i^{(n)}) = 0 \\ &\Leftrightarrow i = 1, \cdots, n \text{ に対して } d(X_i, \hat{Y}_i^{(n)}) \leq D \\ &\Leftrightarrow \max_{1\leq i\leq n} d(X_i, \hat{Y}_i^{(n)}) \leq D \end{aligned}$$

に注意すると

$$\Pr\left\{\frac{1}{n}\sum_{i=1}^{n} d'(X_i, \hat{Y}_i^{(n)}) = 0\right\} = \Pr\left\{\max_{1\leq i\leq n} d(X_i, \hat{Y}_i^{(n)}) \leq D\right\}$$

が成り立つ. よって (34) は

$$\Pr\left\{\max_{1\leq i\leq n} d(X_i, \hat{Y}_i^{(n)}) \leq D\right\} > 1 - \delta \Leftrightarrow \Pr\left\{\max_{1\leq i\leq n} d(X_i, \hat{Y}_i^{(n)}) > D\right\} < \delta$$

と表すことができる. 以上より, 「歪み測度 d' による平均歪みに対して符号化レート R は CK 最大忠実度規範 0 で達成可能」と「歪み測度 d による最大歪みに対して符号化レート R は CK 最大忠実度規範 D で達成可能」は同値である. よって, 達成可能なレートの下限は

$$R_3(D|X) = \min_{Y:\Pr\{d(X,Y)\leq D\}=1} I(X;Y) \tag{35}$$

である. □

3.4 最大歪み測度に対してレート R が最大忠実度規範 D で達成可能

本節では一般情報源に対する定理を扱う.

定義８

$$R_4(D|X) \triangleq \inf\{R| \text{ 最大歪み測度に対して } R \text{ は最大忠実度規範 } D \text{ で達成可能 }\} \tag{36}$$

次の定理は本研究の主要な成果である.

定理４　一般情報源 X のレート・歪み関数 $R_4(D|X)$ は

$$R_4(D|X) = R_3(D|X) \tag{37}$$

で与えられる.

(証明)　(31), (36) より, 最大忠実度規範は CK 最大忠実度規範と比較して, より緩い定義なので

$$R_4(D|X) \le R_3(D|X) \tag{38}$$

である.

次に, 逆を示すために最大歪み測度に対して R が最大忠実度規範 D で達成可能とする. すなわち

$$\limsup_{n\to\infty} \frac{1}{n} \log \|\varphi_n\| \le R \tag{39}$$

$$\text{p-}\limsup_{n\to\infty} d_n(X^n, \psi_n(\varphi_n(X^n)) \le D \tag{40}$$

となるような (φ_n, ψ_n) が存在する. なお $\|\varphi_n\|$ は符号語のサイズを表す. ここで

$$d_n \triangleq d_n(X^n, \psi_n(\varphi_n(X^n))$$

とおくと, (40) より任意の $\delta > 0$ に対して

$$\lim_{n\to\infty} \Pr\{d_n > D + \delta\} = 0 \tag{41}$$

となる. 一方, $\mathcal{X}$, $\mathcal{Y}$ が有限集合であることから

$$\mathcal{D} \triangleq \{d(x, y)|x \in \mathcal{X}, y \in \mathcal{Y}\}$$

は有限集合である. よって, 任意の n に対して

$$d_n(x^n, y^n) = \max_{1\le i\le n} d(x_i, y_i) \in \mathcal{D}$$

である. $\delta \triangleq \frac{1}{2}\min\{\Delta - D | \Delta \in \mathcal{D},\ \Delta > D\} > 0$ とおくと, $\mathcal{D}$ が有限集合なので, 任意の $\Delta \in \mathcal{D}$ に対して

$$\Delta > D + \delta \Leftrightarrow \Delta > D$$

となる. ゆえに

$$\Pr\{d_n > D + \delta\} = \Pr\{d_n > D\}$$

となる. ここで (41) より

$$\lim_{n \to \infty} \Pr\{d_n > D\} = 0$$

が得られ, これと (39) より R が CK 最大忠実度規範 D で達成可能であることがわかる. これは

$$R_4(D|X) \geq R_3(D|X) \tag{42}$$

を意味する.

(38), (42) より, 最大歪み測度に対して最大忠実度規範 D で達成可能なレートの下限と最大歪み測度に対して CK 最大忠実度規範 D で達成可能なレートの下限は等しい. □

系 1 定常無記憶情報源 X のレート・歪み関数 $R_4(D|X)$ は

$$R_4(D|X) = \min_{Y:\Pr\{d(X,Y) \leq D\} = 1} I(X;Y) \tag{43}$$

で与えられる.

(証明) 定理 3 と定理 4 より明らかである. □

4 まとめ

連続的な情報の伝送は無歪みでは不可能であるが, 有歪み圧縮を用いることで伝送が可能になる. 本研究では, 一定の歪みを許容した上で情報をどれだけ圧縮できるかを多角的に明らかにすることを目的とした. そのために, 歪み測度を 2 種類, 忠実度規範を 4 種類考え, それぞれの歪み測度に対してそれぞれの忠実度規範を適用した. 全部で 8 種類の組み合わせがあり, 本研究ではそのうち 4 種類の組み合わせにおける最小符号化レートを考察した. 具体的には, 次の 4 つである. 平均歪み測度に対して平均忠実度規範 D で達成可能な最小符号化レート $R_1(D|X)$ はより詳細に証明を与えた. 平均歪み測度に対して CK 最大忠実度規範 D で達成可能な最小符号化レート $R_2(D|X)$ は付録にて詳細に証明を与えた. 最大歪み測度に対して CK 最大忠実度規範 D で達成可能な最小符号化レート $R_3(D|X)$ を導出した. 最大歪み測度に対して最大忠実度規範 D で達成可能な最小符号化レート $R_4(D|X)$ を一般情報源に対して与えた.

定常無記憶情報源に対して $R_1(D|X)$ と $R_2(D|X)$ は同じ値になった. これは n を十分大きくすると CK 最大忠実度規範の δ が 0 に収束するため最大忠実度規範と同義とみなされ, 同じ値となったと考えられる. また, CK 最大忠実度規範を平均歪み測度と最大歪み測度それぞれに対して適用したときの最小符号化レート $R_2(D|X)$, $R_3(D|X)$ は確率変数 Y の条件が異なる値が得られた. 最大歪み測度のほうが平均歪み測度に比べ歪みが大きくなるため, 確率変数 Y の条件がより厳しくなり, その結果 $R_3(D|X) \geq R_2(D|X)$ となっている. $R_3(D|X)$ と $R_4(D|X)$ は最大歪み測度に対して, それぞれ異なる忠実度規範を用いたが, 達成可能な最小符号化レートは同値となった. これは一般情報源に対して成り立つ.

謝辞

本研究を終えるにあたり, 終始一貫して丁寧なご指導を賜わりました指導教員の西新幹彦准教授に心より感謝と敬意の意を申し上げます.

参考文献

[1] I. Csiszár and J. Körner, *Information Theory: Cording Theorems for Discrete Memoryless Systems*, Academic Press, New York, 1981.

[2] 韓 太舜, 情報理論における情報スペクトル的方法, 培風館, 1998.

[3] 植松 友彦, 現代シャノン理論, 培風館, 1998.

付録 A　レート・歪み理論の基礎

A.1　タイプ理論

情報源アルファベット $\mathcal{X}$ とその上の分布 $P_{\boldsymbol{X}}$ を持つ定常独立な情報源から系列 $\boldsymbol{x} \in \mathcal{X}^n$ を得る確率は, $\mathcal{X}$ の中の各シンボルが $\boldsymbol{x}$ 中に何回出現しているかによって完全に定められる. シンボル $a \in \mathcal{X}$ が系列 $\boldsymbol{x}$ 中に現れている回数を $N(a|\boldsymbol{x})$ で表す. 長さ n の系列 $\boldsymbol{x} \in \mathcal{X}^n$ と $a \in \mathcal{X}$ に対して

$$P_{\boldsymbol{x}}(a) \triangleq \frac{1}{n} N(a|\boldsymbol{x})$$

と定める. $P_{\boldsymbol{x}}$ を $\boldsymbol{x}$ のタイプという. タイプは $\mathcal{X}$ 上の分布である. また, タイプ P に対して $\mathcal{X}^n$ に属する系列の中でタイプが P であるような系列全体を

$$\mathcal{T}_P^n \triangleq \left\{ \boldsymbol{x} \in \mathcal{X}^n \,\middle|\, \text{任意の } a \in \mathcal{X} \text{ に対して } P(a) = \frac{1}{n} N(a|\boldsymbol{x}) \right\}$$

と表す. また, $\mathcal{X}$ 上の任意の分布 P, 任意の定数 $\delta \geq 0$ を用いて

$$\mathcal{T}_{[P]\delta}^n \triangleq \left\{ \boldsymbol{x} \in \mathcal{X}^n \,\middle|\, \begin{array}{ll} \text{任意の } a \in \mathcal{X} \text{ に対して} & \\ |\frac{1}{n} N(a|\boldsymbol{x}) - P(a)| \leq \delta & (P(a) > 0 \text{ のとき }) \\ \qquad N(a|\boldsymbol{x}) = 0 & (P(a) = 0 \text{ のとき }) \end{array} \right\}$$

と定義する. また, $n \to \infty$ のとき $\delta_n \to 0$, $\sqrt{n} \cdot \delta_n \to \infty$ となる δ_n に対して

$$\mathcal{T}_{[P]}^n \triangleq \mathcal{T}_{[P]\delta_n}^n \tag{44}$$

を定義する. これに属する系列 $\boldsymbol{x} \in \mathcal{X}^n$ を P-ティピカルであるという.

次に, 復元アルファベットを $\mathcal{Y}$ とする. 系列 $\boldsymbol{x} \in \mathcal{X}^n$ と $\boldsymbol{y} \in \mathcal{Y}^n$ の組 $(\boldsymbol{x}, \boldsymbol{y})$ に対する同時タイプを系列 $\{(x_i, y_i)\}_{i=1}^n \in (\mathcal{X} \times \mathcal{Y})^n$ のタイプと定義する. すなわち, 同時タイプは

$$P_{\boldsymbol{x},\boldsymbol{y}}(a, b) \triangleq \frac{1}{n} N(a, b|\boldsymbol{x}, \boldsymbol{y})$$

によって定まる $\mathcal{X} \times \mathcal{Y}$ 上の分布である. ただし, $N(a, b|\boldsymbol{x}, \boldsymbol{y})$ は $x_i = a$ かつ $y_i = b$ となっている場所 i の総数である. そして, 与えられた $\boldsymbol{x} \in \mathcal{X}^n$ に対する $\boldsymbol{y} \in \mathcal{Y}^n$ の条件付きタイプ V はすべての $a \in \mathcal{X}$ および $b \in \mathcal{Y}$ について

$$N(a, b|\boldsymbol{x}, \boldsymbol{y}) = N(a|\boldsymbol{x}) V(b|a)$$

を満足する確率遷移行列 $V : \mathcal{X} \to \mathcal{Y}$ として定義される. 特に, すべての $a \in \mathcal{X}$ について $N(a|\boldsymbol{x}) > 0$ ならば条件付きタイプは一意に定まり

$$V(b|a) = \frac{N(a, b|\boldsymbol{x}, \boldsymbol{y})}{N(a|\boldsymbol{x})}$$

によって与えられる. さらに, $\boldsymbol{x} \in \mathcal{X}^n$ に対し, 条件付きタイプが V となるような $\mathcal{Y}^n$ の部分集合を

$$\mathcal{T}_V^n(\boldsymbol{x}) \triangleq \{\boldsymbol{y} \in \mathcal{Y}^n | \text{ すべての } a \in \mathcal{X},\ b \in \mathcal{Y} \text{ に対して } N(a, b|\boldsymbol{x}, \boldsymbol{y}) = N(a|\boldsymbol{x})V(b|a)\}$$

と表す.

遷移確率行列 $W : \mathcal{X} \to \mathcal{Y}$ と, 任意の定数 $\delta \geq 0$, 系列 $\boldsymbol{x} \in \mathcal{X}^n$ に対して

$$\mathcal{T}_{[W]\delta}^n(\boldsymbol{x}) \triangleq \left\{ \boldsymbol{y} \in \mathcal{Y}^n \middle| \begin{array}{cc} \multicolumn{2}{c}{\text{すべての } a \in \mathcal{X},\ b \in \mathcal{Y} \text{ に対して}} \\ |\frac{1}{n}N(a, b|\boldsymbol{x}, \boldsymbol{y}) - \frac{1}{n}N(a|\boldsymbol{x})W(b|a)| \leq \delta & (W(b|a) > 0 \text{ のとき })\\ N(a, b|\boldsymbol{x}, \boldsymbol{y}) = 0 & (W(b|a) = 0 \text{ のとき }) \end{array} \right\}$$

を定義する. そして, (44) を満たす δ_n に対して

$$\mathcal{T}_{[W]}^n(\boldsymbol{x}) \triangleq \mathcal{T}_{[W]\delta_n}^n(\boldsymbol{x})$$

を定義する. これに属する系列 $\boldsymbol{y} \in \mathcal{Y}^n$ を条件 $\boldsymbol{x} \in \mathcal{X}^n$ の下で W-ティピカルであるという.

ここで, ΣPW を

$$\Sigma PW(b) \triangleq \sum_a P(a)W(b|a)$$

によって定まる $\mathcal{Y}$ 上の分布とする.

補題 1 ([1, Lemma 2.10]) 任意の分布 P と任意の遷移確率行列 W に対して, $\boldsymbol{x} \in \mathcal{T}_{[P]\delta}^n$ かつ $\boldsymbol{y} \in \mathcal{T}_{[W]\delta}^n(\boldsymbol{x})$ のとき, $(\boldsymbol{x},\ \boldsymbol{y}) \in \mathcal{T}_{[PW]2\delta}^n$ かつ $\delta' \triangleq 2|\mathcal{X}|\delta$ を用いて $\boldsymbol{y} \in \mathcal{T}_{[\Sigma PW]\delta'}^n$ となる.

(証明) $\boldsymbol{x} \in \mathcal{T}_{[P]\delta}^n$ かつ $\boldsymbol{y} \in \mathcal{T}_{[W]\delta}^n(\boldsymbol{x})$ とする. 初めに, $(\boldsymbol{x},\ \boldsymbol{y}) \in \mathcal{T}_{[PW]2\delta}^n$ を示す. $P(a)W(b|a) > 0$ のとき

$$\begin{aligned}
&\left|\frac{1}{n}N(a, b|\boldsymbol{x}, \boldsymbol{y}) - P(a)W(b|a)\right| \\
=&\left|\frac{1}{n}N(a, b|\boldsymbol{x}, \boldsymbol{y}) - \frac{1}{n}N(a|\boldsymbol{x})W(b|a) + \frac{1}{n}N(a|\boldsymbol{x})W(b|a) - P(a)W(b|a)\right| \\
\leq&\left|\frac{1}{n}N(a, b|\boldsymbol{x}, \boldsymbol{y}) - \frac{1}{n}N(a|\boldsymbol{x})W(b|a)\right| + \left|\frac{1}{n}N(a|\boldsymbol{x})W(b|a) - P(a)W(b|a)\right| \\
\leq&\delta + \delta W(b|a) \\
\leq&2\delta
\end{aligned}$$

となる. $P(a)W(b|a) = 0$ ならば, $P(a) = 0$ または $W(b|a) = 0$ である. $P(a) = 0$ のとき $\boldsymbol{x} \in \mathcal{T}_{[P]\delta}^n$ より $N(a|\boldsymbol{x}) = 0$ である. よって, $N(a, b|\boldsymbol{x}, \boldsymbol{y}) = 0$ となる. $W(b|a) = 0$ のとき $\boldsymbol{y} \in \mathcal{T}_{[W]\delta}^n(\boldsymbol{x})$ より $N(a, b|\boldsymbol{x}, \boldsymbol{y}) = 0$ である. ゆえに, いずれにせよ $N(a, b|\boldsymbol{x}, \boldsymbol{y}) = 0$ となる. よって $(\boldsymbol{x},\ \boldsymbol{y}) \in \mathcal{T}_{[PW]2\delta}^n$ が成り立つ.

次に，$\boldsymbol{y} \in \mathcal{T}^n_{[\Sigma PW]\delta'}$ を示す．

$$
\begin{aligned}
\left|\frac{1}{n}N(b|\boldsymbol{y}) - \Sigma PW(b)\right| &= \left|\frac{1}{n}\sum_a N(a,b|\boldsymbol{x},\boldsymbol{y}) - \sum_a P(a)W(b|a)\right| \\
&\leq \sum_a \left|\frac{1}{n}N(a,b|\boldsymbol{x},\boldsymbol{y}) - P(a)W(b|a)\right| \\
&\leq \sum_a \left\{\left|\frac{1}{n}N(a,b|\boldsymbol{x},\boldsymbol{y}) - \frac{1}{n}N(a|\boldsymbol{x})W(b|a)\right| + \delta W(b|a)\right\} \\
&= \sum_a \left|\frac{1}{n}N(a,b|\boldsymbol{x},\boldsymbol{y}) - \frac{1}{n}N(a|\boldsymbol{x})W(b|a)\right| + \sum_a \delta W(b|a) \\
&\leq \sum_a \delta + \delta \sum_a W(b|a) \\
&\leq |\mathcal{X}|\delta + |\mathcal{X}|\delta \\
&= 2|\mathcal{X}|\delta
\end{aligned}
$$

となる．特に $\Sigma PW(b) = 0$ のとき

$$
0 = \sum_a P(a)W(b|a)
$$

である．このとき $P(a) > 0$ なる a に対して，$W(b|a) = 0$ であるため，$\boldsymbol{y} \in \mathcal{T}^n_{[W]}(\boldsymbol{x})$ より $N(a,b|\boldsymbol{x},\boldsymbol{y}) = 0$ である．一方，$P(a) = 0$ なる a に対しては $\boldsymbol{x} \in \mathcal{T}^n_{[P]}$ より $N(a|\boldsymbol{x}) = 0$ である．ゆえに，$N(a,b|\boldsymbol{x},\boldsymbol{y}) = 0$ である．いずれにせよどの $a \in \mathcal{X}$ に対しても $N(a,b|\boldsymbol{x},\boldsymbol{y}) = 0$ となる．以上より，$\sum PW(b) = 0$ ならば

$$
\frac{1}{n}N(b|\boldsymbol{y}) = \frac{1}{n}\sum_a N(a,b|\boldsymbol{x},\boldsymbol{y}) = 0
$$

となる．したがって，$\delta' \triangleq 2|\mathcal{X}|\delta$ に対して

$$
\boldsymbol{y} \in \mathcal{T}^n_{[\sum PW]\delta'}
$$

が成り立つ．□

補題 2 ([1, Lemma 2.12]) 分布 P に対して，$n \to \infty$ としたとき

$$
P^n(\mathcal{T}^n_{[P]}) \to 1
$$

となる．

（証明） 分布 P に従う確率変数 X を考える．$N(a|X)$ の期待値は

$$
\begin{aligned}
\boldsymbol{E}[N(a|X)] &= \sum_{\boldsymbol{x} \in \mathcal{X}} P(\boldsymbol{x})N(a|\boldsymbol{x}) \\
&= P(a)
\end{aligned}
$$

となる．分布 P^n に従う確率変数 X^n を考える．$a \in \mathcal{X}$ に対して $\frac{1}{n}N(a|X^n)$ の期待値は

$$\begin{aligned}
\boldsymbol{E}\left[\frac{1}{n}N(a|X^n)\right] &= \frac{1}{n}\boldsymbol{E}\left[\sum_{i=1}^{n} N(a|X_i)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{E}[N(a|X_i)] \\
&= \frac{1}{n}\sum_{i=1}^{n} P(a) \\
&= P(a)
\end{aligned}$$

となる．次に，$N(a|X)$ の分散は

$$\begin{aligned}
Var[N(a|X)] &= \boldsymbol{E}[\{N(a|X)\}^2] - \{\boldsymbol{E}[N(a|X)]\}^2 \\
&= \sum_{\boldsymbol{x}\in\mathcal{X}} P(\boldsymbol{x})\{N(a|\boldsymbol{x})\}^2 - \{P(a)\}^2 \\
&= P(a) - \{P(a)\}^2 \\
&= P(a)\{1-P(a)\}
\end{aligned}$$

となる．さらに，$\frac{1}{n}N(a|X^n)$ の分散は

$$\begin{aligned}
Var\left[\frac{1}{n}N(a|X^n)\right] &= \frac{1}{n^2}Var\left[\sum_{i=1}^{n} N(a|X_i)\right] \\
&= \frac{1}{n^2}\sum_{i=1}^{n} Var[N(a|X_i)] \quad \text{（情報源の独立性より）} \\
&= \frac{1}{n^2}\sum_{i=1}^{n} P(a)\{1-P(a)\} \\
&= \frac{1}{n}P(a)\{1-P(a)\}
\end{aligned}$$

となる. 以上のことから

$$
\begin{aligned}
1-P^n(\mathcal{T}^n_{[P]}) &= 1-\Pr\{X^n \in \mathcal{T}^n_{[P]}\} \\
&= \Pr\{X^n \notin \mathcal{T}^n_{[P]}\} \\
&= \Pr\left\{\begin{array}{ll} \text{ある } a \in \mathcal{X} \text{ に対して} & \\ |\frac{1}{n}N(a|X^n)-P(a)|>\delta & (P(a)>0 \text{ のとき }) \\ N(a|X^n)>0 & (P(a)=0 \text{ のとき }) \end{array}\right\} \\
&= \Pr\left\{\begin{array}{l} P(a)>0 \text{ なるある } a \in \mathcal{X} \text{ に対して} \\ |\frac{1}{n}N(a|X^n)-P(a)|>\delta \end{array}\right\} \\
&\leq \sum_{\substack{a\in\mathcal{X} \\ P(a)>0}} \Pr\left\{\left|\frac{1}{n}N(a|X^n)-P(a)\right|>\delta\right\} \\
&\leq \sum_{\substack{a\in\mathcal{X} \\ P(a)>0}} \frac{P(a)\{1-P(a)\}}{n\delta_n^2} \quad \text{（チェビシェフの不等式より）} \\
&\leq \sum_{a\in\mathcal{X}} \frac{1}{4n\delta_n^2} \\
&= \frac{|\mathcal{X}|}{4n\delta_n^2} \\
&\to 0
\end{aligned}
$$

を得る. よって, $n\to\infty$ のとき

$$
P^n(\mathcal{T}^n_{[P]}) \to 1
$$

となる. □

補題 3 ([1, Lemma 2.12 条件付き]) 任意の遷移確率行列 $W:\mathcal{X}\to\mathcal{Y}$ を考える. 任意の $\epsilon>0$ に対して, 十分大きな n をとれば, どの $\boldsymbol{x}\in\mathcal{X}^n$ に対しても

$$
W^n(\mathcal{T}^n_{[W]}(\boldsymbol{x})|\boldsymbol{x}) \geq 1-\epsilon
$$

となる.

（証明） 任意の文字 $x\in\mathcal{X}$ を考える. $W(\cdot|x)$ に従う確率変数 Y を考える. $N(a,b|x,Y)$ の期待値は

$$
\begin{aligned}
\boldsymbol{E}[N(a,b|x,Y)] &= \sum_{y\in\mathcal{Y}} W(y|x)N(a,b|x,y) \\
&= N(a|x)W(b|a)
\end{aligned}
$$

となる．任意の系列 $\boldsymbol{x} \in \mathcal{X}^n$ に対して $W^n(\cdot|\boldsymbol{x})$ に従う確率変数 Y^n を考える．そして，$N(a,b|\boldsymbol{x},Y^n)$ の期待値は

$$\begin{aligned}\boldsymbol{E}[N(a,b|\boldsymbol{x},Y^n)] &= \boldsymbol{E}\left[\sum_{i=1}^{n} N(a,b|x_i,Y_i)\right] \\ &= \sum_{i=1}^{n} \boldsymbol{E}[N(a,b|x_i,Y_i)] \\ &= \sum_{i=1}^{n} N(a|x_i)W(b|a) \\ &= W(b|a)\sum_{i=1}^{n} N(a|x_i) \\ &= W(b|a)N(a|\boldsymbol{x})\end{aligned}$$

となる．次に，$N(a,b|x,Y)$ の分散は

$$\begin{aligned}Var[N(a,b|x,Y)] &= \boldsymbol{E}[\{N(a,b|x,Y)\}^2] - \{\boldsymbol{E}[N(a,b|x,Y)]\}^2 \\ &= \sum_{y\in\mathcal{Y}} W(y|x)\{N(a,b|x,y)\}^2 - \{N(a|x)W(b|a)\}^2 \\ &= \{N(a|x)\}^2 W(b|a) - \{N(a|x)W(b|a)\}^2 \\ &= \{(N(a|x))^2 W(b|a)\}(1-W(b|a))\end{aligned}$$

となる．さらに，$N(a,b|\boldsymbol{x},Y^n)$ の分散は

$$\begin{aligned}Var[N(a,b|\boldsymbol{x},Y)] &= Var\left[\sum_{i=1}^{n} N(a,b|x_i,Y_i)\right] \\ &= \sum_{i=1}^{n} Var[N(a,b|x_i,Y_i)] \quad \text{（情報源の独立性より）} \\ &= \sum_{i=1}^{n}\{(N(a|x_i))^2 W(a|b)\}(1-W(b|a)) \\ &= (1-W(b|a))W(b|a)\sum_{i=1}^{n}(N(a|x_i))^2 \\ &= (1-W(b|a))W(b|a)\sum_{i=1}^{n} N(a|x_i) \\ &= (1-W(b|a))W(b|a)N(a|\boldsymbol{x})\end{aligned}$$

となる．そして

$$
\begin{aligned}
1 - W^n(\mathcal{T}^n_{[W]}(\boldsymbol{x})|\boldsymbol{x}) &= 1 - \Pr\{Y^n \in \mathcal{T}^n_{[W]}(\boldsymbol{x})\} \\
&= \Pr\{Y^n \notin \mathcal{T}^n_{[W]}(\boldsymbol{x})\} \\
&= \Pr\left\{ \begin{array}{ll} \text{ある } a \in \mathcal{X}, b \in \mathcal{Y} \text{ に対して} & \\ |N(a,b|\boldsymbol{x},Y^n) - N(a|\boldsymbol{x})W(b|a)| > n\delta_n & (W(b|a) > 0 \text{ のとき })\\ N(a,b|\boldsymbol{x},Y^n) > 0 & (W(b|a) = 0 \text{ のとき }) \end{array} \right\} \\
&= \Pr\left\{ \begin{array}{l} W(b|a) > 0 \text{ なるある } a \in \mathcal{X}, b \in \mathcal{Y} \text{ に対して} \\ |N(a,b|\boldsymbol{x},Y^n) - N(a|\boldsymbol{x})W(b|a)| > n\delta_n \end{array} \right\} \\
&\leq \sum_{\substack{a\in\mathcal{X},b\in\mathcal{Y}\\ W(b|a)>0}} \Pr\{|N(a,b|\boldsymbol{x},Y^n) - N(a|\boldsymbol{x})W(b|a)| > n\delta_n\} \\
&\leq \sum_{\substack{a\in\mathcal{X},b\in\mathcal{Y}\\ W(b|a)>0}} \frac{Var[N(a,b|\boldsymbol{x},Y^n)]}{n^2{\delta_n}^2} \quad \text{（チェビシェフの不等式より）} \\
&\leq \sum_{\substack{a\in\mathcal{X},b\in\mathcal{Y}\\ W(b|a)>0}} \frac{(1-W(b|a))W(b|a)N(a|\boldsymbol{x})}{n^2\delta_n^2} \\
&\leq \sum_{a\in\mathcal{X},b\in\mathcal{Y}} \frac{1\cdot n}{4n^2\delta_n^2} \\
&= \sum_{a\in\mathcal{X},b\in\mathcal{Y}} \frac{1}{4n\delta_n^2} \\
&\leq \frac{|\mathcal{X}||\mathcal{Y}|}{4n\delta_n^2}
\end{aligned}
$$

を得る．よって，任意の $\epsilon > 0$ に対して，十分大きな n をとれば，どの $\boldsymbol{x} \in \mathcal{X}^n$ に対しても

$$W^n(\mathcal{T}^n_{[W]}(\boldsymbol{x})|\boldsymbol{x}) \geq 1 - \epsilon$$

となる．□

補題 4 ([1, Lemma 2.6]) $\mathcal{X}$ 上の分布 Q とタイプ P をもつ系列 $\boldsymbol{x} \in \mathcal{X}^n$ に対して

$$Q^n(\boldsymbol{x}) = \exp\{-n(H(P) + D(P \parallel Q))\} \tag{45}$$

となる．

（証明）

$$
\begin{aligned}
Q^n(\boldsymbol{x}) &= \prod_{a\in\mathcal{X}} Q(a)^{N(a|\boldsymbol{x})} \\
&= \exp\left\{\log \prod_{a\in\mathcal{X}} Q(a)^{N(a|\boldsymbol{x})}\right\} \\
&= \exp\left\{\sum_{a\in\mathcal{X}} \log Q(a)^{N(a|\boldsymbol{x})}\right\} \\
&= \exp\left\{n \sum_{a\in\mathcal{X}} \frac{N(a|\boldsymbol{x})}{n} \log Q(a)\right\} \\
&= \exp\left\{n \sum_{a\in\mathcal{X}} P(a) \log Q(a)\right\} \\
&= \exp\left\{n \sum_{a\in\mathcal{X}} P(a)\left(\log P(a) - \log \frac{P(a)}{Q(a)}\right)\right\} \\
&= \exp\left\{n \sum_{a\in\mathcal{X}} P(a) \log P(a) - n \sum_{a\in\mathcal{X}} P(a) \log \frac{P(a)}{Q(a)}\right\} \\
&= \exp\{-nH(P) - nD(P \parallel Q)\} \\
&= \exp\{-n(H(P) + D(P \parallel Q))\}
\end{aligned}
$$

□

補題 5 ([3, 補題 2.6])（対数和不等式）任意の正の数 $\{a_i\}_{i=1}^n, \{b_i\}_{i=1}^n$ において

$$
\sum_{i=1}^n a_i \log \frac{a_i}{b_i} \geq a \log \frac{a}{b} \tag{46}
$$

が成り立つ．ただし

$$
\left\{ \begin{array}{ll} a & \triangleq \sum_{i=1}^n a_i \\ b & \triangleq \sum_{i=1}^n b_i \end{array} \right. \tag{47}
$$

であり，等号成立の必要十分条件は $a_i/b_i = a/b$ がすべての i で成り立つことである．
（証明）

$$
x - 1 \geq \log x \Leftrightarrow 1 - x \leq \log \frac{1}{x}
$$

に注意する. $x=1$ のとき等号成立となる. これを用いると

$$\begin{aligned}\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} - a \log \frac{a}{b} &= \sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} - \sum_{i=1}^{n} a_i \log \frac{a}{b} \\ &= \sum_{i=1}^{n} a_i \left(\log \frac{a_i}{b_i} - \log \frac{a}{b} \right) \\ &= \sum_{i=1}^{n} \log \frac{a_i/a}{b_i/b} \\ &\geq \sum_{i=1}^{n} a_i \left(1 - \frac{b_i/b}{a_i/a} \right) \\ &= \sum_{i=1}^{n} \left(a_i - \frac{ab_i}{b} \right) \\ &= a - \frac{a}{b} b \\ &= 0\end{aligned}$$

となる. 等号成立条件は $a_i/a = b_i/b$ がすべての i について成り立つことである. □

補題 6 ([1, Lemma 2.7]) 有限アルファベット $\mathcal{X}$ 上の分布 $P,\ Q$ について $\|P-Q\| \leq \Theta \leq \frac{1}{2}$ ならば

$$|H(P) - H(Q)| \leq -\Theta \log \frac{\Theta}{|\mathcal{X}|} \tag{48}$$

が成り立つ. ここで, $\|P-Q\| \triangleq \sum_{a \in \mathcal{X}} |P(a) - Q(a)|$ であり, P と Q の間の変動距離と呼ばれる.

(証明) $f(t) \triangleq -t \log t$ とおく. ここで, $0 \leq \tau \leq \frac{1}{2},\ 0 \leq t \leq 1-\tau$ なる $\tau,\ t$ を考える. 次に, $g(t) \triangleq f(t) - f(t+\tau) = -t\log t + (t+\tau)\log(t+\tau)$ を考える. このとき

$$\begin{aligned}\frac{d}{dt} g(t) &= -\log t - 1 + \log(t+\tau) + (t+\tau)\frac{1}{t+\tau} \\ &= \log \frac{t+\tau}{t} \\ &\geq 0\end{aligned}$$

となるので $g(t)$ は単調増加であり, $t = 1-\tau$ で最大, $t=0$ で最小となる. よって

$$-f(\tau) = g(0) \leq g(t) \leq g(1-\tau) = f(1-\tau)$$

となり，したがって

$$
\begin{aligned}
|f(t) - f(t+\tau)| &= |g(t)| \\
&= \max\{-g(t),\ g(t)\} \\
&\leq \max\{f(\tau),\ f(1-\tau)\}
\end{aligned}
$$

を得る．さらに，ここで

$$
\begin{aligned}
h(\tau) &\triangleq f(\tau) - f(1-\tau) \\
&= -\tau \log \tau + (1-\tau)\log(1-\tau)
\end{aligned}
$$

とおくと，$h(0) = h(\frac{1}{2}) = 0$ である．また

$$
\begin{aligned}
\frac{d}{d\tau} h(\tau) &= -\log \tau - 1 - \log(1-\tau) - (1-\tau)\frac{1}{1-\tau} \\
&= -\log\{\tau(1-\tau)\} - 2
\end{aligned}
$$

は $0 \leq \tau \leq \frac{1}{2}$ で単調減少かつ

$$
\frac{dh}{d\tau}(0) = +\infty, \qquad \frac{dh}{d\tau}\left(\frac{1}{2}\right) < 0
$$

である．よって，$0 \leq \tau \leq \frac{1}{2}$ で $h(\tau) \geq 0$ すなわち

$$
f(\tau) \geq f(1-\tau)
$$

となる．ゆえに，(14) より

$$
|f(t) - f(t+\tau)| \leq f(\tau) \tag{49}
$$

となる．さらに

$$
\begin{aligned}
|H(P)-H(Q)| &= \left|-\sum_{a\in\mathcal{X}} P(a)\log P(a) + \sum_{a\in\mathcal{X}} Q(a)\log Q(a)\right| \\
&= \left|\sum_{a\in\mathcal{X}} \bigl\{f(P(a)) - f(Q(a))\bigr\}\right| \\
&\le \sum_{a\in\mathcal{X}} \Bigl|f(P(a)) - f(Q(a))\Bigr| \\
&= \sum_{\substack{a\in\mathcal{X}\\ P(a)>Q(a)}} \Bigl|f(Q(a)) - f(P(a))\Bigr| + \sum_{\substack{a\in\mathcal{X}\\ P(a)\le Q(a)}} \Bigl|f(P(a)) - f(Q(a))\Bigr| \\
&= \sum_{\substack{a\in\mathcal{X}\\ P(a)>Q(a)}} \Bigl|f(Q(a)) - f(|P(a)-Q(a)|+Q(a))\Bigr| \\
&\qquad + \sum_{\substack{a\in\mathcal{X}\\ P(a)\le Q(a)}} \Bigl|f(P(a)) - f(|Q(a)-P(a)|+P(a))\Bigr|
\end{aligned}
$$

であるから，ここで $|P(a)-Q(a)| \le \frac{1}{2}$ を用いると

$$
\begin{aligned}
|H(P)-H(Q)| &\le \sum_{a\in\mathcal{X}} f(|P(a)-Q(a)|) \quad \text{((49) より)} \\
&= -\sum_{a\in\mathcal{X}} |P(a)-Q(a)|\log|P(a)-Q(a)| \\
&= -\|P-Q\| \sum_{a\in\mathcal{X}} \frac{|P(a)-Q(a)|}{\|P-Q\|}\left(\log\frac{|P(a)-Q(a)|}{\|P-Q\|} + \log\|P-Q\|\right) \\
&= -\|P-Q\| \sum_{a\in\mathcal{X}} r(a)\bigl(\log r(a) + \log\|P-Q\|\bigr) \\
&= -\|P-Q\|\left(-H(r) + \sum_{a\in\mathcal{X}} r(a)\log\|P-Q\|\right) \\
&\le \|P-Q\|(\log|\mathcal{X}| - \log\|P-Q\|) \\
&= -\|P-Q\|\log\frac{\|P-Q\|}{|\mathcal{X}|} \\
&\le -\Theta\log\frac{\Theta}{|\mathcal{X}|}
\end{aligned}
$$

となる．ただし，$r(a) \triangleq \frac{|P(a)-Q(a)|}{\|P-Q\|}$ とおいた．r は $\mathcal{X}$ 上の分布である．□

補題 7 ([3, 定理 2.7]) P を任意の分布とする．任意の $\epsilon > 0$ に対して十分大きなすべての n で

$\boldsymbol{x} \in \mathcal{T}^n_{[P]}$ は

$$\left| -\frac{1}{n} \log P^n(\boldsymbol{x}) - H(P) \right| < \epsilon$$

を満たす.

（証明） $\boldsymbol{x} \in \mathcal{T}^n_{[P]}$ のタイプを $P_{\boldsymbol{x}}$ とあらわす. $\delta_n \le \frac{1}{2|\mathcal{X}|}$ となるほど n を大きくすれば

$$\begin{aligned} |P_{\boldsymbol{x}} - P\| &= \sum_{a \in \mathcal{X}} |P_{\boldsymbol{x}}(a) - P(a)| \\ &\le |\mathcal{X}|\delta_n \\ &\le \frac{1}{2} \end{aligned}$$

したがって補題 6 より

$$|H(P_{\boldsymbol{x}}) - H(P)| \le -|\mathcal{X}|\delta_n \log \delta_n \tag{50}$$

となる. 次に $P(a) > 0$ なる a を考える. このとき $|P_{\boldsymbol{x}}(a) - P(a)| \le \delta_n$ なので

$$\frac{P_{\boldsymbol{x}}(a)}{P(a)} \le 1 + \frac{\delta_n}{P(a)}$$

となるので

$$\begin{aligned} D(P_{\boldsymbol{x}} \parallel P) &= \sum_{a \in \mathcal{X}} P_{\boldsymbol{x}}(a) \log \frac{P_{\boldsymbol{x}}(a)}{P(a)} \\ &= \sum_{\substack{a \in \mathcal{X} \\ P(a) > 0}} P_{\boldsymbol{x}}(a) \log \frac{P_{\boldsymbol{x}}(a)}{P(a)} \quad \text{（}P(a) = 0 \text{ ならば } P_{\boldsymbol{x}}(a) = 0 \text{ より）} \\ &\le \sum_{\substack{a \in \mathcal{X} \\ P(a) > 0}} P_{\boldsymbol{x}}(a) \log\left(1 + \frac{\delta_n}{P(a)}\right) \\ &\le \log\left(1 + \frac{\delta_n}{P_{\min}}\right) \end{aligned}$$

を得る. ただし, $P_{\min} \triangleq \min_{\substack{a \in \mathcal{X} \\ P(a) > 0}} P(a)$ とおいた. 以上より補題 4 を用いて

$$\begin{aligned} \left| -\frac{1}{n} \log P^n(\boldsymbol{x}) - H(P) \right| &= \left| D(P_{\boldsymbol{x}} \parallel P) + H(P_{\boldsymbol{x}}) - H(P) \right| \\ &\le D(P_{\boldsymbol{x}} \parallel P) + \left| H(P_{\boldsymbol{x}}) - H(P) \right| \\ &\le \log\left(1 + \frac{\delta_n}{P_{\min}}\right) - |\mathcal{X}|\delta_n \log \delta_n \\ &< \epsilon \end{aligned}$$

となる. ただし, 最後の不等式は十分大きな n で成り立つ. □

補題 8 ([1, Lemma 2.2]) $\mathcal{X}^n$ の系列の相異なるタイプは $(n+1)^{|\mathcal{X}|}$ 個を超えない.
（証明） 任意の $a \in \mathcal{X}$ に対して, $N(a|\boldsymbol{x})$ は高々 $(n+1)$ 種類の値をとる. 有限集合 $\mathcal{X}$ に含まれるシンボルの数は $|\mathcal{X}|$ なので, 相異なるタイプの数は高々 $(n+1)^{|\mathcal{X}|}$ となる. □

補題 9 ([3, 補題 2.4]) 系列 $\boldsymbol{x} \in \mathcal{X}^n$ に対し, $\mathcal{Y}^n$ の系列の条件付きタイプのうち相異なるものは $(n+1)^{|\mathcal{X}||\mathcal{Y}|}$ 個を超えない.
（証明） 任意の $a \in \mathcal{X}, b \in \mathcal{Y}$ に対して, $V(b|a)$ は高々 $(n+1)$ 種類の値をとる. 有限集合 $\mathcal{X}, \mathcal{Y}$ に含まれるシンボルの数はそれぞれ $|\mathcal{X}|, |\mathcal{Y}|$ なので, 相異なるタイプの数は高々 $(n+1)^{|\mathcal{X}||\mathcal{Y}|}$ となる. □

補題 10 ([1, Lemma 2.3]) $\boldsymbol{x} \in \mathcal{X}^n$ のタイプ P に対して

$$(n+1)^{-|\mathcal{X}|} \exp\{nH(P)\} \le |\mathcal{T}_P^n| \le \exp\{nH(P)\}$$

が成り立つ.
（証明）

$$\begin{aligned}
P^n(\boldsymbol{x}) &= \prod_{a \in \mathcal{X}} P(a)^{N(a|\boldsymbol{x})} \\
&= \prod_{a \in \mathcal{X}} P(a)^{nP(a)} \\
&= \exp\left\{\log \prod_{a \in \mathcal{X}} P(a)^{nP(a)}\right\} \\
&= \exp\left\{\sum_{a \in \mathcal{X}} \log P(a)^{nP(a)}\right\} \\
&= \exp\left\{n \sum_{a \in \mathcal{X}} P(a) \log P(a)\right\} \\
&= \exp\{-nH(P)\}
\end{aligned}$$

となる. これを用いると

$$\begin{aligned}
P^n(\mathcal{T}_P^n) &= \sum_{\boldsymbol{x} \in \mathcal{T}_P^n} P^n(\boldsymbol{x}) \\
&= \sum_{\boldsymbol{x} \in \mathcal{T}_P^n} \exp\{-nH(P)\} \\
&= |\mathcal{T}_P^n| \cdot \exp\{-nH(P)\} \qquad (51)
\end{aligned}$$

となる. よって

$$\begin{aligned}
|\mathcal{T}_P^n| &= P^n(\mathcal{T}_P^n) \cdot \exp\{nH(P)\} \\
&\le \exp\{nH(P)\} \quad （P^n(\mathcal{T}_P^n) \le 1 \text{ より}）
\end{aligned}$$

を得る．

次に，P とは異なるタイプ $\hat{P}$ を考える．$\hat{\boldsymbol{x}} \in \mathcal{T}^n_{\hat{P}}$ ならば $\hat{P}(a) = \frac{1}{n}N(a|\hat{\boldsymbol{x}})$ であるので

$$\begin{aligned}
P^n(\mathcal{T}^n_{\hat{P}}) &= \sum_{\hat{\boldsymbol{x}} \in \mathcal{T}^n_{\hat{P}}} P^n(\hat{\boldsymbol{x}}) \\
&= \sum_{\hat{\boldsymbol{x}} \in \mathcal{T}^n_{\hat{P}}} \prod_{a \in \mathcal{X}} P(a)^{N(a|\hat{\boldsymbol{x}})} \\
&= \sum_{\hat{\boldsymbol{x}} \in \mathcal{T}^n_{\hat{P}}} \prod_{a \in \mathcal{X}} P(a)^{n\hat{P}(a)} \\
&= |\mathcal{T}^n_{\hat{P}}| \prod_{a \in \mathcal{X}} P(a)^{n\hat{P}(a)} \\
&= \frac{n!}{\displaystyle\prod_{a \in \mathcal{X}} (n\hat{P}(a))!} \cdot \prod_{a \in \mathcal{X}} P(a)^{n\hat{P}(a)}
\end{aligned}$$

となる．$P^n(\mathcal{T}^n_P)$ についても同様となるので

$$\frac{P^n(\mathcal{T}^n_{\hat{P}})}{P^n(\mathcal{T}^n_P)} = \prod_{a \in \mathcal{X}} \frac{(nP(a))!}{(n\hat{P}(a))!} \prod_{a \in \mathcal{X}} P(a)^{n(\hat{P}(a) - P(a))}$$

が得られる．ここで，不等式

$$\frac{n!}{m!} \leq n^{n-m}$$

に注目する．この不等式が成立するのは

$$\frac{n!}{m!} = \begin{cases} \underbrace{n(n-1)\cdots(m+1)}_{n-m\ \text{項}} \leq n^{n-m} & (n \geq m) \\ \dfrac{1}{\underbrace{m(m-1)\cdots(n+1)}_{m-n\ \text{項}}} \leq \left(\frac{1}{n+1}\right)^{m-n} & (n < m) \end{cases}$$

からわかる．これを用いて

$$\begin{aligned}
\frac{P^n(\mathcal{T}^n_{\hat{P}})}{P^n(\mathcal{T}^n_P)} &\leq \prod_{a \in \mathcal{X}} (nP(a))^{n(P(a) - \hat{P}(a))} \prod_{a \in \mathcal{X}} P(a)^{n(\hat{P}(a) - P(a))} \\
&= \prod_{a \in \mathcal{X}} n^{n(P(a) - \hat{P}(a))} \\
&= n^{\sum_{a \in \mathcal{X}} n(P(a) - P(\hat{a}))} \\
&= n^{\sum_{a \in \mathcal{X}} nP(a) - \sum_{a \in \mathcal{X}} n\hat{P}(a)} \\
&= n^{n-n} \\
&= 1
\end{aligned}$$

となるので

$$P^n(\mathcal{T}^n_{\hat{P}}) \leq P^n(\mathcal{T}^n_P)$$

を得る. また, 異なるタイプ全体で総和をとれば

$$\begin{aligned}
1 &= P^n(\mathcal{X}^n) \\
&= \sum_{\hat{P}} P^n(\mathcal{T}^n_{\hat{P}}) \\
&\leq \sum_{\hat{P}} P^n(\mathcal{T}^n_P) \\
&\leq (n+1)^{|\mathcal{X}|} P^n(\mathcal{T}^n_P) \quad \text{（補題 8 より）} \\
&= (n+1)^{|\mathcal{X}|} |\mathcal{T}^n_P| \exp\{-nH(P)\} \quad \text{((51) より)}
\end{aligned}$$

を得る. ただし, 最後の等式でタイプ P における系列の出現確率は全て等しいということを使った. よって

$$|\mathcal{T}^n_P| \geq (n+1)^{-|\mathcal{X}|} \exp\{nH(P)\}$$

が得られる. □

補題 11 ([1, Lemma 2.13]) 任意の $\epsilon > 0$ に対して十分大きなすべての n で

$$\left| \frac{1}{n} \log |\mathcal{T}^n_{[P]}| - H(P) \right| < \epsilon$$

が成り立つ.

（証明） まず, $\boldsymbol{x} \in \mathcal{T}^n_{[P]}$ を考える. このとき, $\mathcal{T}^n_{P_{\boldsymbol{x}}} \subset \mathcal{T}^n_{[P]}$ であるから, 補題 10 より

$$(n+1)^{-|\mathcal{X}|} \exp\{nH(P_{\boldsymbol{x}})\} \leq |\mathcal{T}^n_{P_{\boldsymbol{x}}}| \leq |\mathcal{T}^n_{[P]}|$$

であるから,

$$-\frac{|\mathcal{X}|}{n} \log(n+1) + H(P_{\boldsymbol{x}}) - H(P) \leq \frac{1}{n} \log |\mathcal{T}^n_{[P]}| - H(P) \tag{52}$$

が成り立つ. ここで n を十分大きくとれば

$$\begin{aligned}
\|P_{\boldsymbol{x}} - P\| &= \sum_{a \in \mathcal{X}} |P_{\boldsymbol{x}}(a) - P(a)| \\
&\leq |\mathcal{X}| \delta_n \\
&\leq \frac{1}{2}
\end{aligned}$$

となるので, 補題 6 より, このとき

$$\begin{aligned}|H(P_{\boldsymbol{x}}) - H(P)| &\leq -|\mathcal{X}|\delta_n \log \delta_n \\ &\leq \frac{\epsilon}{2}\end{aligned}$$

となる. 一方, n を十分大きくとれば $-\frac{\epsilon}{2} < -\frac{|\mathcal{X}|}{n} \log(n+1)$ であるので (52) は

$$-\epsilon < \frac{1}{n} \log |\mathcal{T}_{[P]}^n| - H(P) \tag{53}$$

となる.

次に, $\mathcal{T}_{[P]}^n$ に属する系列をタイプごとに仕分けると

$$\mathcal{T}_{[P]}^n = \bigcup_Q \mathcal{T}_Q^n$$

とかける. ただし, 右辺の集合和はある $\boldsymbol{x} \in \mathcal{T}_{[P]}^n$ に対して $Q = P_{\boldsymbol{x}}$ となるようなタイプ Q 全体に渡っている. したがって, 再び補題 10 より

$$\begin{aligned}|\mathcal{T}_{[P]}^n| &= \sum_Q |\mathcal{T}_Q^n| \\ &\leq \sum_Q \exp(nH(Q))\end{aligned} \tag{54}$$

を得る. ここで, Q はある $\boldsymbol{x} \in \mathcal{T}_{[P]}^n$ を用いて $Q = P_{\boldsymbol{x}}$ となることから, 十分大きい n に対して

$$\begin{aligned}\|Q - P\| &= \sum_{a \in \mathcal{X}} |Q(a) - P(a)| \\ &\leq |\mathcal{X}|\delta_n \\ &< \frac{1}{2}\end{aligned}$$

が成り立つ. よって (54) は補題 6, 補題 8 より

$$\begin{aligned}|\mathcal{T}_{[P]}^n| &\leq \sum_Q \exp\left(n\left(H(P) + \frac{\epsilon}{2}\right)\right) \\ &\leq (n+1)^{|\mathcal{X}|} \exp\left(n\left(H(P) + \frac{\epsilon}{2}\right)\right)\end{aligned}$$

となる. これを変形すると

$$\frac{1}{n} \log |\mathcal{T}_{[P]}^n| \leq \frac{|\mathcal{X}|}{n} \log(n+1) + H(P) + \frac{\epsilon}{2}$$

を得る. 十分大きい n に対しては $\frac{|\mathcal{X}|}{n} \log(n+1) < \frac{\epsilon}{2}$ なので

$$\frac{1}{n} \log |\mathcal{T}_{[P]}^n| - H(P) < \epsilon \tag{55}$$

を得る. (53), (55) をまとめると

$$\left|\frac{1}{n}\log|\mathcal{T}^n_{[P]}| - H(P)\right| < \epsilon$$

となる. □

補題 12 ($|\mathcal{T}^n_V(\boldsymbol{x})|$ の値) 遷移確率行列 $V:\mathcal{X}\to\mathcal{Y}$ を考える. $|\mathcal{T}^n_V(\boldsymbol{x})|$ の値は $\boldsymbol{x}\in\mathcal{X}^n$ のタイプで決まる.

(証明) タイプの等しい *2* つの相異なる系列 $\boldsymbol{x},\ \boldsymbol{x}'\in\mathcal{X}^n$ を考える. すなわち, 任意の $a\in\mathcal{X}$ に対して

$$N(a|\boldsymbol{x}) = N(a|\boldsymbol{x}') \tag{56}$$

とする. このとき $\boldsymbol{x}$ の文字の位置を適切に置換すれば $\boldsymbol{x}'$ を得る. 任意の $\boldsymbol{y}\in\mathcal{T}^n_V(\boldsymbol{x})$ を考える. すると $\mathcal{T}^n_V(\boldsymbol{x})$ の定義より, 任意の $a\in\mathcal{X},\ b\in\mathcal{Y}$ に対して

$$N(a,b|\boldsymbol{x},\boldsymbol{y}) = N(a|\boldsymbol{x})V(b|a) \tag{57}$$

が成り立っている. ここで $\boldsymbol{x}$ から $\boldsymbol{x}'$ を得るのと同じ文字の位置の置換を $\boldsymbol{y}$ に対して施したものを $\boldsymbol{y}'$ とおく. すると

$$\begin{aligned} N(a,b|\boldsymbol{x}',\boldsymbol{y}') &= N(a,b|\boldsymbol{x},\boldsymbol{y}) \\ &= N(a|\boldsymbol{x})V(b|a) \quad ((57)\text{ より }) \\ &= N(a|\boldsymbol{x}')V(b|a) \quad ((56)\text{ より }) \end{aligned}$$

となって, $\boldsymbol{y}'\in\mathcal{T}^n_V(\boldsymbol{x})$ が成り立つ. 文字の位置の置換が単射であることに注意すれば

$$|\mathcal{T}^n_V(\boldsymbol{x})| \le |\mathcal{T}^n_V(\boldsymbol{x}')|$$

がわかる. 逆も真なので結局

$$|\mathcal{T}^n_V(\boldsymbol{x})| = |\mathcal{T}^n_V(\boldsymbol{x}')|$$

を得る. □

補題 13 ([1, Lemma 2.5]) 全ての $\boldsymbol{x}\in\mathcal{X}^n$ と遷移確率行列 $V:\mathcal{X}\to\mathcal{Y}$ について $\mathcal{T}^n_V(\boldsymbol{x})$ が空でないならば

$$(n+1)^{-|\mathcal{X}||\mathcal{Y}|}\exp\{nH(V|P_{\boldsymbol{x}})\} \le |\mathcal{T}^n_V(\boldsymbol{x})| \le \exp\{nH(V|P_{\boldsymbol{x}})\}$$

が成り立つ.

(証明) $\mathcal{T}^n_V(\boldsymbol{x})\neq\emptyset$ とし

$$n_a \triangleq N(a|\boldsymbol{x})$$

とおく. 各 $a \in \mathcal{X}$ に対して定まる文字列 a^{n_a} を連接したものを $\boldsymbol{x}'$ とおくと, $\boldsymbol{x}'$ と $\boldsymbol{x}$ のタイプは等しい. ゆえに補題 12 より

$$|\mathcal{T}_V^n(\boldsymbol{x}')| = |\mathcal{T}_V^n(\boldsymbol{x})|$$

となって, $|\mathcal{T}_V^n(\boldsymbol{x}')|$ を評価すれば十分である. $\boldsymbol{y}' \in \mathcal{T}_V^n(\boldsymbol{x}')$ を考える. $\boldsymbol{y}'$ を先頭から順にそれぞれ長さ n_a となるように分節し, 各文字列を $\boldsymbol{y}_a$ とおく. $\boldsymbol{y}_a$ をすべて連接すると $\boldsymbol{y}'$ となる. このとき任意の $b \in \mathcal{Y}$ に対して

$$N(b|\boldsymbol{y}_a) = N(a, b|\boldsymbol{x}', \boldsymbol{y}')$$

であることに注意すれば, $n_a \neq 0$ のとき

$$\frac{1}{n_a} N(b|\boldsymbol{y}_a) = \frac{N(a, b|\boldsymbol{x}', \boldsymbol{y}')}{N(a|\boldsymbol{x}')} = V(b|a)$$

であるので, $n_a \neq 0$ なる $\boldsymbol{y}_a$ のタイプは $V(\cdot|a)$ となる.

また逆に, $n_a \neq 0$ なる各 $a \in \mathcal{X}$ に対して $\boldsymbol{y}_a \in \mathcal{T}_{V(\cdot|a)}^{n_a}$ を選んで連接したものを $\boldsymbol{y}$ とおくと, $\boldsymbol{y} \in \mathcal{T}_V^n(\boldsymbol{x}')$ となる.

したがって, $\boldsymbol{y} \in \mathcal{T}_V^n(\boldsymbol{x}')$ であることと, $n_a \neq 0$ なる任意の a に対して $\boldsymbol{y}_a \in \mathcal{T}_{V(\cdot|a)}^{n_a}$ となることは同値である. このことから

$$|\mathcal{T}_V^n(\boldsymbol{x}')| = \prod_{\substack{a \in \mathcal{X} \\ n_a \neq 0}} |\mathcal{T}_{V(\cdot|a)}^{n_a}|$$

が成り立つ. ここで補題 10 より

$$\prod_{\substack{a \in \mathcal{X} \\ n_a \neq 0}} (n_a + 1)^{-|\mathcal{Y}|} \exp\{n_a H(V(\cdot|a))\} \leq \prod_{\substack{a \in \mathcal{X} \\ n_a \neq 0}} |\mathcal{T}_{V(\cdot|a)}^{n_a}| \leq \prod_{\substack{a \in \mathcal{X} \\ n_a \neq 0}} \exp\{n_a H(V(\cdot|a))\}$$

であるので

$$\begin{aligned}
\prod_{\substack{a \in \mathcal{X} \\ n_a \neq 0}} \exp\{n_a H(V(\cdot|a))\} &= \exp\left\{\sum_{a \in \mathcal{X}} n_a H(V(\cdot|a))\right\} \\
&= \exp\left\{n \sum_{a \in \mathcal{X}} \left\{\frac{N(a|\boldsymbol{x})}{n} H(V(\cdot|a))\right\}\right\} \\
&= \exp\left\{n \sum_{a \in \mathcal{X}} \left\{P_{\boldsymbol{x}}(a) H(V(\cdot|a))\right\}\right\} \\
&= \exp\{n H(V|P_{\boldsymbol{x}})\}
\end{aligned}$$

と

$$\prod_{\substack{a\in\mathcal{X}\\ n_a\neq 0}}(n_a+1)^{-|\mathcal{Y}|}\geq\prod_{a\in\mathcal{X}}(n+1)^{-|\mathcal{Y}|}$$
$$=(n+1)^{-|\mathcal{X}||\mathcal{Y}|}$$

より

$$(n+1)^{-|\mathcal{X}||\mathcal{Y}|}\exp\{nH(V|P_{\boldsymbol{x}})\}\leq|\mathcal{T}_V^n(\boldsymbol{x})|\leq\exp\{nH(V|P_{\boldsymbol{x}})\}$$

が成り立つ．□

補題 14 ([1, Lemma 2.6 条件付き]) 全ての $\boldsymbol{x}\in\mathcal{X}^n$ と遷移確率行列 $V:\mathcal{X}\to\mathcal{Y}$, $W:\mathcal{X}\to\mathcal{Y}$ について $\mathcal{T}_V^n(\boldsymbol{x})$ が空でないならば, $\boldsymbol{y}\in\mathcal{T}_V^n(\boldsymbol{x})$ に対して,

$$W^n(\boldsymbol{y}|\boldsymbol{x})=\exp\{-n(D(V\|W|P_{\boldsymbol{x}})+H(V|P_{\boldsymbol{x}}))\}$$

が成り立つ.

（証明） 任意の $\boldsymbol{x}\in\mathcal{X}^n$, $\boldsymbol{y}\in\mathcal{T}_V^n(\boldsymbol{x})$ について $W^n(\boldsymbol{y}|\boldsymbol{x})$ を求める.

$$\begin{aligned}
W^n(\boldsymbol{y}|\boldsymbol{x})&=\prod_{i=1}^n W(y_i|x_i)\\
&=\prod_{a\in\mathcal{X},b\in\mathcal{Y}}W(b|a)^{N(a,b|\boldsymbol{x},\boldsymbol{y})}\\
&=\exp\left\{\log\prod_{a\in\mathcal{X},b\in\mathcal{Y}}W(b|a)^{N(a,b|\boldsymbol{x},\boldsymbol{y})}\right\}\\
&=\exp\left\{\sum_{a\in\mathcal{X},b\in\mathcal{Y}}\log W(b|a)^{N(a,b|\boldsymbol{x},\boldsymbol{y})}\right\}\\
&=\exp\left\{\sum_{a\in\mathcal{X},b\in\mathcal{Y}}N(a,b|\boldsymbol{x},\boldsymbol{y})\log W(b|a)\right\}\\
&=\exp\left\{n\sum_{a\in\mathcal{X},b\in\mathcal{Y}}\frac{N(a|\boldsymbol{x})}{n}V(b|a)\log W(b|a)\right\}\\
&=\exp\left\{n\sum_{a\in\mathcal{X}}P_{\boldsymbol{x}}(a)\sum_{b\in\mathcal{Y}}V(b|a)\left(\log\frac{W(b|a)}{V(b|a)}+\log V(b|a)\right)\right\}\\
&=\exp\left\{n\sum_{a\in\mathcal{X}}P_{\boldsymbol{x}}(a)\sum_{b\in\mathcal{Y}}V(b|a)\log\frac{W(b|a)}{V(b|a)}+n\sum_{a\in\mathcal{X}}P_{\boldsymbol{x}}(a)\sum_{b\in\mathcal{Y}}V(b|a)\log V(b|a)\right\}\\
&=\exp\{-nD(V\|W|P_{\boldsymbol{x}})-nH(V|P_{\boldsymbol{x}})\}\\
&=\exp\{-n(D(V\|W|P_{\boldsymbol{x}})+H(V|P_{\boldsymbol{x}})\}
\end{aligned}$$

□

補題 15 (条件付きエントロピーの差 1) 任意の $\epsilon > 0$ に対して十分大きな n をとれば, 任意の $\boldsymbol{x} \in \mathcal{X}^n$ に対し, $\boldsymbol{y} \in \mathcal{T}^n_{[W]}(\boldsymbol{x})$ の条件付きタイプ V は

$$|H(V|P_{\boldsymbol{x}}) - H(W|P_{\boldsymbol{x}})| \leq \epsilon$$

を満たす. ただし $P_{\boldsymbol{x}}$ は $\boldsymbol{x}$ のタイプである.

（証明） $P_{\boldsymbol{x}}(a) > 0$ なる $a \in \mathcal{X}$ を考える. すると $\boldsymbol{y} \in \mathcal{T}^n_{[W]}(\boldsymbol{x})$ より任意の $b \in \mathcal{Y}$ に対して

$$|N(a,b|\boldsymbol{x},\boldsymbol{y}) - N(a|\boldsymbol{x})W(b|a)| \leq \delta_n$$

であるが, $\boldsymbol{y} \in \mathcal{T}^n_V(\boldsymbol{x})$ であるから

$$N(a,b|\boldsymbol{x},\boldsymbol{y}) = N(a|\boldsymbol{x})V(b|a)$$

なので

$$|V(b|a) - W(b|a)| \leq \frac{\delta_n}{N(a|\boldsymbol{x})} \leq \delta_n$$

となる. したがって, n が十分大きければ

$$\begin{aligned}
\|V(\cdot|a) - W(\cdot|a)\| &= \sum_{b\in\mathcal{Y}} |V(b|a) - W(b|a)| \\
&\leq \delta_n|\mathcal{Y}| \\
&\leq \frac{1}{2}
\end{aligned}$$

となる. よって補題 6 より

$$|H(V(\cdot|a)) - H(W(\cdot|a))| \leq -\delta_n|\mathcal{Y}| \log \frac{\delta_n|\mathcal{Y}|}{|\mathcal{Y}|}$$

となり, 十分大きなすべての n で

$$\begin{aligned}
|H(V|P_{\boldsymbol{x}}) - H(W|P_{\boldsymbol{x}})| &= \left| \sum_{a\in\mathcal{X}} P_{\boldsymbol{x}}(a)H(V(\cdot|a)) - \sum_{a\in\mathcal{X}} P_{\boldsymbol{x}}(a)H(W(\cdot|a)) \right| \\
&\leq \sum_{a\in\mathcal{X}} P_{\boldsymbol{x}}(a)|H(V(\cdot|a)) - H(W(\cdot|a))| \\
&\leq -\delta_n|\mathcal{Y}| \log \delta_n \\
&\leq \epsilon
\end{aligned}$$

を得る. □

補題 16 (条件付きエントロピーの差 2) 任意の $\epsilon > 0$, $\mathcal{X}$ 上の分布 P に対して十分大きな n をとれば, 任意の $\boldsymbol{x} \in \mathcal{T}^n_{[P]}$ に対して

$$|H(W|P_{\boldsymbol{x}}) - H(W|P)| \leq \epsilon$$

となる．
（証明）

$$\begin{aligned}|H(W|P_{\boldsymbol{x}})-H(W|P)| &= \left|\sum_{a\in\mathcal{X}}(P_{\boldsymbol{x}}(a)H(W(\cdot|a))-P(a)H(W(\cdot|a))\right| \\ &= \left|\sum_{a\in\mathcal{X}}(P_{\boldsymbol{x}}(a)-P(a))H(W(\cdot|a))\right| \\ &\le \sum_{a\in\mathcal{X}}\left|P_{\boldsymbol{x}}(a)-P(a)\right|H(W(\cdot|a)) \\ &\le \delta_n|\mathcal{X}|\log|\mathcal{Y}| \\ &\le \epsilon\end{aligned}$$

□

補題 17 ([3, 定理 2.7 条件付き]) $\mathcal{X}$ 上の任意の分布 P, 遷移確率行列 $W:\mathcal{X}\to\mathcal{Y}$ を考える．任意の $\boldsymbol{x}\in\mathcal{T}^n_{[P]}$ に対して $\boldsymbol{y}\in\mathcal{T}^n_{[W]}(\boldsymbol{x})$ ならば, 任意の $\epsilon>0$ に対して十分大きな n をとれば

$$\left|-\frac{1}{n}\log W^n(\boldsymbol{y}|\boldsymbol{x})-H(W|P)\right|\le\epsilon$$

が成り立つ．
（証明） $\boldsymbol{x}$ のタイプを $P_{\boldsymbol{x}}$ とし, $\boldsymbol{y}\in\mathcal{T}^n_{[W]}(\boldsymbol{x})$ に対して, $N(a,b|\boldsymbol{x},\boldsymbol{y})=V(b|a)N(a|\boldsymbol{x})$ を満たす V を考える．すると V は $\boldsymbol{x}$ に対する $\boldsymbol{y}$ の条件付タイプである．また, 補題 14 より

$$\begin{aligned}&\left|-\frac{1}{n}\log W^n(\boldsymbol{y}|\boldsymbol{x})-H(W|P)\right| \\ =&|D(V\|W|P_{\boldsymbol{x}})+H(V|P_{\boldsymbol{x}})-H(W|P)| \\ \le&D(V\|W|P_{\boldsymbol{x}})+|H(V|P_{\boldsymbol{x}})-H(W|P_{\boldsymbol{x}})|+|H(W|P_{\boldsymbol{x}})-H(W|P)|\end{aligned}$$

を得る．ここで

$$D(V\|W|P_{\boldsymbol{x}})\triangleq\sum_{a\in\mathcal{X}}P_{\boldsymbol{x}}(a)\sum_{b\in\mathcal{Y}}V(b|a)\log\frac{V(b|a)}{W(b|a)}$$

と定義した．$P_{\boldsymbol{x}}(a)>0$ なる a, $V(b|a)>0$ なる b を考える．すると，$\frac{1}{n}N(a,b|\boldsymbol{x},\boldsymbol{y})>0$ より $W(b|a)>0$ であるので

$$|P_{\boldsymbol{x}}(a)V(b|a)-P_{\boldsymbol{x}}(a)W(b|a)|\le\delta_n \tag{58}$$

より

$$\frac{V(b|a)}{W(b|a)}\le\frac{\delta_n}{P_{\boldsymbol{x}}(a)W(b|a)}+1$$

を得る. ゆえに, 任意の $\epsilon > 0$ に対して十分大きなすべての n で

$$
\begin{aligned}
D(V\|W|P_{\boldsymbol{x}}) &\le \sum_{\substack{a\in\mathcal{X}\\ P_{\boldsymbol{x}}(a)>0}} P_{\boldsymbol{x}}(a) \sum_{\substack{b\in\mathcal{Y}\\ V(b|a)>0}} V(b|a)\log\left(\frac{\delta_n}{P_{\boldsymbol{x}}(a)W(b|a)}+1\right)\\
&\le \log\left(\frac{\delta_n}{(P_{\boldsymbol{x}}W)_{\min}}+1\right)\\
&\le \frac{1}{3}\epsilon
\end{aligned}
\tag{59}
$$

となる. ただし

$$
(P_{\boldsymbol{x}}W)_{\min} \triangleq \min_{\substack{a\in\mathcal{X},b\in\mathcal{Y}\\ P_{\boldsymbol{x}}(a)>0,W(b|a)>0}} P_{\boldsymbol{x}}(a)W(b|a)
$$

とおいた. 補題 15, 補題 16, (59) より, 十分大きな n をとれば

$$
\begin{aligned}
\left|-\frac{1}{n}\log W^n(\boldsymbol{y}|\boldsymbol{x}) - H(W|P)\right| &\le \frac{1}{3}\epsilon + \frac{1}{3}\epsilon + \frac{1}{3}\epsilon\\
&= \epsilon
\end{aligned}
$$

となる. □

補題 18 ([1, Lemma 2.13 条件付き]) $\mathcal{X}$ 上の任意の分布 P を考える. 任意の $\epsilon > 0$ に対し, 十分大きなすべての n で, すべての $\boldsymbol{x} \in \mathcal{T}^n_{[P]}$ に対して

$$
\left|\frac{1}{n}\log|\mathcal{T}^n_{[W]}(\boldsymbol{x})| - H(W|P)\right| \le \epsilon
$$

が成り立つ.

(証明) 上界と下界を別々に証明する.

まず, 下界を示す. $\boldsymbol{x} \in \mathcal{T}^n_{[P]}$ のタイプを $P_{\boldsymbol{x}}$, $\boldsymbol{y} \in \mathcal{T}^n_{[W]}(\boldsymbol{x})$ の $\boldsymbol{x}$ に対する条件付きタイプを V とする. すなわち, $N(a,b|\boldsymbol{x},\boldsymbol{y}) = N(a|\boldsymbol{x})V(b|a)$ である. このとき, $\boldsymbol{y}' \in \mathcal{T}^n_V(\boldsymbol{x})$ に対して

$$
N(a,b|\boldsymbol{x},\boldsymbol{y}') = N(a|\boldsymbol{x})V(b|a)
$$

なので $N(a,b|\boldsymbol{x},\boldsymbol{y}') = N(a,b|\boldsymbol{x},\boldsymbol{y})$ である. よって $\boldsymbol{y} \in \mathcal{T}^n_{[W]}(\boldsymbol{x})$ より $\boldsymbol{y}' \in \mathcal{T}^n_{[W]}(\boldsymbol{x})$ となる. これは $\mathcal{T}^n_V(\boldsymbol{x}) \subset \mathcal{T}^n_{[W]}(\boldsymbol{x})$ を意味する. ここで, $|\mathcal{T}^n_V(P_{\boldsymbol{x}})| \le |\mathcal{T}^n_{[W]}(\boldsymbol{x})|$ をさらに補題 13 で下から抑えると

$$
\begin{aligned}
(n+1)^{-|\mathcal{X}||\mathcal{Y}|}\exp\{nH(V|P_{\boldsymbol{x}})\} &\le |\mathcal{T}^n_{[W]}(\boldsymbol{x})|\\
-\frac{|\mathcal{X}||\mathcal{Y}|}{n}\log(n+1) + H(V|P_{\boldsymbol{x}}) &\le \frac{1}{n}\log|\mathcal{T}^n_{[W]}(\boldsymbol{x})|
\end{aligned}
\tag{60}
$$

を得る. また, n を十分大きくすると, $\frac{\epsilon}{3} > \frac{|\mathcal{X}||\mathcal{Y}|}{n}\log(n+1)$ が成り立ち, さらに補題 15, 補題 16 より $|H(V|P_{\boldsymbol{x}}) - H(W|P_{\boldsymbol{x}})| \le \frac{\epsilon}{3}$ および $|H(W|P_{\boldsymbol{x}}) - H(W|P)| \le \frac{\epsilon}{3}$ がいえるので

$$
\begin{aligned}
\frac{1}{n}\log|\mathcal{T}^n_{[W]}(\boldsymbol{x})| - H(W|P) &\ge -\frac{|\mathcal{X}||\mathcal{Y}|}{n}\log(n+1) + H(V|P_{\boldsymbol{x}}) - H(W|P) \quad ((60)\text{ より })\\
&> -\frac{\epsilon}{3} + H(V|P_{\boldsymbol{x}}) - H(W|P_{\boldsymbol{x}}) + H(W|P_{\boldsymbol{x}}) - H(W|P)\\
&> -\frac{\epsilon}{3} - \frac{\epsilon}{3} - \frac{\epsilon}{3}\\
&= -\epsilon
\end{aligned}
$$

となる.

次に, 上界を示す. $\mathcal{T}^n_{[W]}(\boldsymbol{x})$ に属する系列を条件付きタイプごとに仕分けると

$$
\mathcal{T}^n_{[W]}(\boldsymbol{x}) = \bigcup_{V} \mathcal{T}^n_V(\boldsymbol{x})
$$

となる. ただし, 右辺の集合和はある $\boldsymbol{y} \in \mathcal{T}^n_{[W]}(\boldsymbol{x})$ に対して $N(a,b|\boldsymbol{x},\boldsymbol{y}) = N(a|\boldsymbol{x})V(b|a)$ となるような条件付きタイプ V 全体に渡っている. ここで, 補題 13, 補題 15 より

$$
\begin{aligned}
|\mathcal{T}^n_V(\boldsymbol{x})| &\le \exp(nH(V|P_{\boldsymbol{x}}))\\
&\le \exp\left(n\left(H(W|P_{\boldsymbol{x}}) + \frac{\epsilon}{3}\right)\right)
\end{aligned}
$$

となるので, 補題 9 より

$$
\begin{aligned}
|\mathcal{T}^n_{[W]}(\boldsymbol{x})| &= \sum_V |\mathcal{T}^n_V(\boldsymbol{x})|\\
&\le \sum_V \exp\left(n\left(H(W|P_{\boldsymbol{x}}) + \frac{\epsilon}{3}\right)\right)\\
&\le (n+1)^{|\mathcal{X}||\mathcal{Y}|}\exp\left(n\left(H(W|P_{\boldsymbol{x}}) + \frac{\epsilon}{3}\right)\right)
\end{aligned}
$$

となり, n が十分大きければ

$$
\begin{aligned}
\frac{1}{n}\log|\mathcal{T}^n_{[W]}(\boldsymbol{x})| &\le \frac{|\mathcal{X}||\mathcal{Y}|}{n}\log(n+1) + H(W|P_{\boldsymbol{x}}) + \frac{\epsilon}{3}\\
&< H(W|P_{\boldsymbol{x}}) + \frac{2}{3}\epsilon
\end{aligned}
$$

を得る. ここで両辺から $H(W|P)$ を引いて補題 16 を用いると

$$
\begin{aligned}
\frac{1}{n}\log|\mathcal{T}^n_{[W]}(\boldsymbol{x})| - H(W|P) &< H(W|P_{\boldsymbol{x}}) - H(W|P) + \frac{2}{3}\epsilon\\
&< \epsilon
\end{aligned}
$$

となる.

以上より, 十分大きなすべての n で

$$\left|\frac{1}{n}\log|\mathcal{T}_{[W]}^n(\boldsymbol{x})| - H(W|P)\right| < \epsilon$$

が成り立つ. □

補題 19 ([1, Corollary 2.14]) $0 < \eta < 1$ に対して, $(\eta, |\mathcal{X}|)$ に依存する系列 $\epsilon_n \to 0$ が存在し, 十分大きなすべての n で $P^n(\mathcal{A}) \geq \eta$ なる $\mathcal{A} \subset \mathcal{X}^n$ に対して

$$\frac{1}{n}\log|\mathcal{A}| \geq H(P) - \epsilon_n$$

が成り立つ.

(証明) 補題 2 より, 十分大きい n をとれば

$$P^n(\mathcal{T}_{[P]}^n) > 1 - \frac{\eta}{2}$$

が成り立つ. また, $P^n(\mathcal{A}) \geq \eta$ であるから

$$\begin{aligned} 1 &\geq P^n(\mathcal{A} \cup \mathcal{T}_{[P]}^n) \\ &= P^n(\mathcal{A}) + P^n(\mathcal{T}_{[P]}^n) - P^n(\mathcal{A} \cap \mathcal{T}_{[P]}^n) \end{aligned}$$

に注意すれば

$$\begin{aligned} P^n(\mathcal{A} \cap \mathcal{T}_{[P]}^n) &\geq P^n(\mathcal{A}) + P^n(\mathcal{T}_{[P]}^n) - 1 \\ &> \eta - \frac{\eta}{2} \\ &= \frac{\eta}{2} \end{aligned}$$

を得る. $\mathcal{T}_{[P]}^n$ をタイプごとに分類して

$$\mathcal{T}_{[P]}^n = \bigcup_{P'} \mathcal{T}_{P'}^n$$

と表せば

$$\begin{aligned} \frac{\eta}{2} &< P^n\left(\mathcal{A} \cap \bigcup_{P'} \mathcal{T}_{P'}^n\right) \\ &= \sum_{P'} P^n(\mathcal{A} \cap \mathcal{T}_{P'}^n) \\ &= \sum_{P'} \sum_{\boldsymbol{x} \in \mathcal{A} \cap \mathcal{T}_{P'}^n} P^n(\boldsymbol{x}) \\ &= \sum_{P'} |\mathcal{A} \cap \mathcal{T}_{P'}^n| C_{P'} \end{aligned} \tag{61}$$

と書ける. ただし, $\boldsymbol{x} \in \mathcal{T}_{P'}^n$ に対して $P^n(\boldsymbol{x})$ の値は同じであることからこれを $C_{P'}$ とおいた. 一方

$$
\begin{aligned}
1 &\geq P^n(\mathcal{T}_{[P]}^n) \\
&= \sum_{P'} P^n(\mathcal{T}_{P'}^n) \\
&= \sum_{P'} |\mathcal{T}_{P'}^n| C_{P'}
\end{aligned} \tag{62}
$$

に注意する. ここで $\mathcal{T}_{[P]}^n$ に含まれる任意の系列のタイプ P' に対して

$$
|\mathcal{A} \cap \mathcal{T}_{P'}^n| \leq \frac{\eta}{2} |\mathcal{T}_{P'}^n|
$$

と仮定すると, (61), (62) より

$$
\frac{\eta}{2} < \sum_{P'} |\mathcal{A} \cap \mathcal{T}_{P'}^n| C_{P'} \leq \frac{\eta}{2} \sum_{P'} |\mathcal{T}_{P'}^n| C_{P'} \leq \frac{\eta}{2}
$$

となり矛盾が生じる. よって $\mathcal{T}_{[P]}^n$ に含まれる系列のタイプ P' で

$$
|\mathcal{A} \cap \mathcal{T}_{P'}^n| > \frac{\eta}{2} |\mathcal{T}_{P'}^n|
$$

なるものが存在する. するとこの P' に補題 10 を用いると

$$
\begin{aligned}
|\mathcal{A}| &\geq |\mathcal{A} \cap \mathcal{T}_{P'}^n| \\
&> \frac{\eta}{2} |\mathcal{T}_{P'}^n| \\
&\geq \frac{\eta}{2} (n+1)^{-|\mathcal{X}|} \exp(nH(P'))
\end{aligned}
$$

$$
\frac{1}{n} \log |\mathcal{A}| > H(P') + \frac{1}{n} \log \frac{\eta}{2} - \frac{|\mathcal{X}|}{n} \log(n+1) \tag{63}
$$

を得る. ここで, n が十分大きければ

$$
\begin{aligned}
\|P' - P\| &= \sum_{a \in \mathcal{X}} |P'(a) - P(a)| \\
&\leq |\mathcal{X}| \delta_n \\
&\leq \frac{1}{2}
\end{aligned}
$$

に注意すると, 補題 6 より

$$
|H(P') - H(P)| \leq -|\mathcal{X}| \delta_n \log \delta_n
$$

となる. よって (63) は

$$
\frac{1}{n} \log |\mathcal{A}| > H(P) - \epsilon_n
$$

が成り立つ．ただし

$$\epsilon_n \triangleq -\frac{1}{n}\log\frac{\eta}{2}+\frac{|\mathcal{X}|}{n}\log(n+1)-|\mathcal{X}|\delta_n\log\delta_n>0$$

とおいた．$n\to\infty$ のとき $\epsilon_n\to 0$ である．□

補題 20 ([1, Corollary 2.14 条件付き]) 任意の遷移確率行列 $W:\mathcal{X}\to\mathcal{Y}$ を考える．$0<\eta<1$ に対して，$(\eta,|\mathcal{X}|,|\mathcal{Y}|)$ に依存する系列 $\epsilon'_n\to 0$ が存在し，十分大きなすべての n で $\boldsymbol{x}\in\mathcal{T}^n_{[P]}$，$W^n(\mathcal{B}|\boldsymbol{x})\geq\eta$ となるような $\mathcal{B}\subset\mathcal{Y}^n$ に対して

$$\frac{1}{n}\log|\mathcal{B}|\geq H(W|P)-\epsilon'_n$$

が成り立つ．

（証明） 補題 3 より，十分大きな n をとれば

$$W^n(\mathcal{T}^n_{[W]}(\boldsymbol{x})|\boldsymbol{x})>1-\frac{\eta}{2}$$

が成り立つ．また $W^n(\mathcal{B}|\boldsymbol{x})\geq\eta$ であるから

$$\begin{aligned}1&\geq W^n(\mathcal{B}\cup\mathcal{T}^n_{[W]}(\boldsymbol{x})|\boldsymbol{x})\\&=W^n(\mathcal{B}|\boldsymbol{x})+W^n(\mathcal{T}^n_{[W]}(\boldsymbol{x})|\boldsymbol{x})-W^n(\mathcal{B}\cap\mathcal{T}^n_{[W]}(\boldsymbol{x})|\boldsymbol{x})\end{aligned}$$

に注意すれば

$$\begin{aligned}W^n(\mathcal{B}\cap\mathcal{T}^n_{[W]}(\boldsymbol{x})|\boldsymbol{x})&\geq W^n(\mathcal{B}|\boldsymbol{x})+W^n(\mathcal{T}^n_{[W]}(\boldsymbol{x})|\boldsymbol{x})-1\\&>\eta-\frac{\eta}{2}\\&=\frac{\eta}{2}\end{aligned}$$

を得る．$\mathcal{T}^n_{[W]}(\boldsymbol{x})$ を条件付きタイプごとに分類して

$$\mathcal{T}^n_{[W]}(\boldsymbol{x})=\bigcup_V\mathcal{T}^n_V(\boldsymbol{x})$$

と表せば

$$\begin{aligned}\frac{\eta}{2}&<W^n(\mathcal{B}\cap\mathcal{T}^n_{[W]}(\boldsymbol{x})|\boldsymbol{x})\\&=\sum_V W^n(\mathcal{B}\cap\mathcal{T}^n_V(\boldsymbol{x})|\boldsymbol{x})\\&=\sum_V\sum_{\boldsymbol{y}\in\mathcal{B}\cap\mathcal{T}^n_V(\boldsymbol{x})}W^n(\boldsymbol{y}|\boldsymbol{x})\\&=\sum_V|\mathcal{B}\cap\mathcal{T}^n_V(\boldsymbol{x})|C_V(\boldsymbol{x})\end{aligned}\tag{64}$$

と書ける. ただし, $\boldsymbol{y} \in \mathcal{T}_V^n(\boldsymbol{x})$ に対して $W^n(\boldsymbol{y}|\boldsymbol{x})$ の値は同じであることから, これを $C_V(\boldsymbol{x})$ とおいた. 一方

$$\begin{aligned} 1 &\geq W^n(\mathcal{T}_{[W]}^n(\boldsymbol{x})|\boldsymbol{x}) \\ &= \sum_V W^n(\mathcal{T}_V^n(\boldsymbol{x})|\boldsymbol{x}) \\ &= \sum_V \sum_{\boldsymbol{y} \in \mathcal{T}_V^n(\boldsymbol{x})} W^n(\boldsymbol{y}|\boldsymbol{x}) \\ &= \sum_V |\mathcal{T}_V^n(\boldsymbol{x})| C_V(\boldsymbol{x}) \end{aligned} \tag{65}$$

に注意する. ここで $\mathcal{T}_{[W]}^n$ に含まれる任意の条件付きタイプ V が

$$|\mathcal{B} \cap \mathcal{T}_V^n(\boldsymbol{x})| \leq \frac{\eta}{2} |\mathcal{T}_V^n(\boldsymbol{x})|$$

であると仮定すると, (64), (65) より

$$\frac{\eta}{2} < \sum_V |\mathcal{B} \cap \mathcal{T}_V^n(\boldsymbol{x})| C_V(\boldsymbol{x}) \leq \frac{\eta}{2} \sum_V |\mathcal{T}_V^n(\boldsymbol{x})| C_V(\boldsymbol{x}) \leq \frac{\eta}{2}$$

となって矛盾が生じる. よって $\mathcal{T}_{[W]}^n(\boldsymbol{x})$ に含まれる条件付きタイプ V で

$$|\mathcal{B} \cap \mathcal{T}_V^n(\boldsymbol{x})| > \frac{\eta}{2} |\mathcal{T}_V^n(\boldsymbol{x})|$$

となるものが存在する. すると, この V に補題 13 を用いると

$$\begin{aligned} |\mathcal{B}| &\geq |\mathcal{B} \cap \mathcal{T}_V^n(\boldsymbol{x})| \\ &> \frac{\eta}{2} |\mathcal{T}_V^n(\boldsymbol{x})| \\ &\geq \frac{\eta}{2} (n+1)^{-|\mathcal{X}||\mathcal{Y}|} \exp(nH(V|P_{\boldsymbol{x}})) \\ \frac{1}{n} \log |\mathcal{B}| &\geq H(V|P_{\boldsymbol{x}}) + \frac{1}{n} \log \frac{\eta}{2} - \frac{|\mathcal{X}||\mathcal{Y}|}{n} \log(n+1) \end{aligned} \tag{66}$$

を得る. ここで任意の $a \in \mathcal{X}$ に対して, n が十分大きければ

$$\begin{aligned} \|V(\cdot|a) - W(\cdot|a)\| &= \sum_{b \in \mathcal{Y}} |V(b|a) - W(b|a)| \\ &\leq |\mathcal{Y}| \delta_n \\ &\leq \frac{1}{2} \end{aligned}$$

に注意すると, 補題 6 より

$$|H(V(\cdot|a)) - H(W(\cdot|a))| \leq -|\mathcal{Y}| \delta_n \log \delta_n$$

となる. これより

$$
\begin{aligned}
|H(V|P_{\boldsymbol{x}}) - H(W|P_{\boldsymbol{x}})| &= \left| \sum_{a \in \mathcal{X}} P_{\boldsymbol{x}}(a) H(V(\cdot|a)) - \sum_{a \in \mathcal{X}} P_{\boldsymbol{x}}(a) H(W(\cdot|a)) \right| \\
&= \left| \sum_{a \in \mathcal{X}} P_{\boldsymbol{x}}(a) \big(H(V(\cdot|a)) - H(W(\cdot|a)) \big) \right| \\
&\leq \sum_{a \in \mathcal{X}} P_{\boldsymbol{x}}(a) \big| H(V(\cdot|a)) - H(W(\cdot|a)) \big| \\
&\leq -|\mathcal{Y}| \delta_n \log \delta_n
\end{aligned}
$$

を得るので, (66) は

$$
\frac{1}{n} \log |\mathcal{B}| \geq H(W|P_{\boldsymbol{x}}) - \epsilon'_n
$$

が成り立つ. ただし

$$
\epsilon'_n \triangleq -\frac{1}{n} \log \frac{\eta}{2} + \frac{|\mathcal{X}||\mathcal{Y}|}{n} \log(n+1) - |\mathcal{Y}| \delta_n \log \delta_n > 0
$$

とおいた. $n \to \infty$ のとき $\epsilon'_n \to 0$ である. □

A.2　通信路符号化

入力集合 $\mathcal{X}$, 出力集合 $\mathcal{Y}$ の通信路に対する符号を (f, φ) とする. メッセージの集合を $\mathcal{M}$ とする. f は $\mathcal{M}$ から $\mathcal{X}$ への写像である. φ は $\mathcal{Y}$ から $\mathcal{M}'$ への写像である. φ の値域は $\mathcal{M}$ と異なることに注意する. $\mathcal{M}' \supset \mathcal{M}$ と仮定する. 任意の遷移確率行列 $W : \mathcal{X} \to \mathcal{Y}$ が与えられ W に対する符号が (f, φ) のとき, 符号 (f, φ) と W は新しい通信路 $T : \mathcal{M} \to \mathcal{M}'$ を次のように定義する.

$$
T(m'|m) \triangleq W(\varphi^{-1}(m')|f(m)) \quad (m \in \mathcal{M}, m' \in \mathcal{M}')
$$

メッセージ m の誤り確率は

$$
e_m = e_m(W, f, \varphi) \triangleq 1 - T(m|m)
$$

と定義される. 符号 (f, φ) の最大誤り確率は

$$
e = e(W, f, \varphi) \triangleq \max_{m \in \mathcal{M}} e_m
$$

と定義される. 集合 $\mathcal{M}$ 上の確率分布が与えられれば, メッセージの誤り確率の期待値が求められる. 特に, メッセージが等確率のとき対応するメッセージの誤り確率の期待値

$$
\bar{e} \triangleq \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} e_m
$$

は (f,φ) の平均誤り確率と呼ばれる.

通信路符号化問題は,「メッセージ集合 $\mathcal{M}$ をできるだけ大きくしつつ, 最大誤り確率 e をできるだけ小さくする.」ということに帰着される. 以降, 通信路が無記憶の場合を考える. 漸近的な意味でこの問題を解くために n 回の通信路の使用を $W^n : \mathcal{X}^n \to \mathcal{Y}^n$ と表す. つまり

$$W^n(\boldsymbol{y}|\boldsymbol{x}) \triangleq \prod_{i=1}^{n} W(y_i|x_i)$$

である. 無記憶通信路は $\{W : \mathcal{X} \to \mathcal{Y}\}$, または単純に $\{W\}$ とかく. 通信路 $W^n(\boldsymbol{y}|\boldsymbol{x})$ のブロック符号 (f,φ) に対し, 符号化レートは $\frac{1}{n}\log|\mathcal{M}_f|$ である. ただし, $\mathcal{M}_f$ は写像 f の定義域（メッセージの集合）である. さらに, 最大誤り確率 $e(W^n,f,\varphi) \le \epsilon$ のブロック符号を (n,ϵ)-符号と呼ぶ.

定義 21 ([1, Definition 2.1.1]) 任意の通信路 W を考える. $0 \le \epsilon < 1$ が与えられたとき, 任意の $\delta > 0$ に対して十分大きなすべての n で符号化レートが $R - \delta$ を超える (n,ϵ)-符号が存在するならば, 通信路のレート $R > 0$ は ϵ-達成可能であるという. 全ての $0 < \epsilon < 1$ で R が ϵ-達成可能ならば $R > 0$ は達成可能であるという. そして, ϵ-達成可能なレートの上限を ϵ-容量 C_ϵ と表し, 達成可能なレートの上限を容量 C と表す.

定義 22 ([1, Definition 2.1.2]) 集合 $\mathcal{A} \subset \mathcal{X}$, $\mathcal{B} \subset \mathcal{Y}$ を考える. 全ての $x \in \mathcal{A}$ に対して $W(\mathcal{B}|x) \ge \eta$ ならば, 通信路 $W : \mathcal{X} \to \mathcal{Y}$ を通った集合 $\mathcal{B} \subset \mathcal{Y}$ は集合 $\mathcal{A} \subset \mathcal{X}$ の η-*image* $(0 < \eta \le 1)$ であるという. $\mathcal{A}$ の η-*image* の最小濃度を $g_W(\mathcal{A},\eta)$ と表す.

定義 23 (f,φ) と $(\hat{f},\hat{\varphi})$ が共に通信路 $W : \mathcal{X} \to \mathcal{Y}$ に対する符号であり, $\mathcal{M}_f \subset \mathcal{M}_{\hat{f}}$ かつ $m \in \mathcal{M}_f$ に対して $f(m) = \hat{f}(m)$ ならば $(\hat{f},\hat{\varphi})$ は (f,φ) の *extension* である, 又は (f,φ) は $(\hat{f},\hat{\varphi})$ の *subcode* であるという.

補題 24 ([1, Lemma 2.1.3]) 無記憶通信路 $\{W : \mathcal{X} \to \mathcal{Y}\}$ と $\mathcal{X}$ 上の分布 P が与えられているとする. 任意の $\tau,\epsilon \in (0,1)$ を考える. このとき, n が十分大きければ, ある (n,ϵ)-符号 (f,φ) が存在し, 任意の $m \in \mathcal{M}_f$ に対して

$$f(m) \in \mathcal{T}^n_{[P]} \tag{67}$$

$$\varphi^{-1}(m) \subset \mathcal{T}^n_{[W]}(f(m)) \tag{68}$$

でありかつ, 符号化レートは

$$\frac{1}{n}\log|\mathcal{M}_f| \ge I(P,W) - \tau$$

を満たす．

（証明） 任意の n を考える．このときある (n,ϵ)-符号 (f,φ) が存在し，任意の $m\in\mathcal{M}_f$ に対して，(67), (68) が成り立ちかつ任意の $\boldsymbol{x}\in\mathcal{T}^n_{[P]}$ に対して

$$W^n(\mathcal{T}^n_{[W]}(\boldsymbol{x})-\mathcal{B}|\boldsymbol{x})<1-\epsilon \tag{69}$$

を満たす．ただし $\mathcal{B}\triangleq\bigcup_{m\in\mathcal{M}_f}\varphi^{-1}(m)$ とおいた．まず，このような符号が存在することを示す．もしある (n,ϵ)-符号 (f,φ) が (67), (68) を満たして (69) を満たさないとすると

$$W^n(\mathcal{T}^n_{[W]}(\boldsymbol{x})-\mathcal{B}|\boldsymbol{x})\ge 1-\epsilon$$

なる $\boldsymbol{x}\in\mathcal{T}^n_{[P]}$ が存在する．もしこの $\boldsymbol{x}$ が f の符号語ならば，つまり $f(m)=\boldsymbol{x}$ なる $m\in\mathcal{M}_f$ が存在するならば，復号器 φ を $\boldsymbol{y}\in\mathcal{T}^n_{[W]}(\boldsymbol{x})-\mathcal{B}$ に対しても m を復号するように復号領域を拡大し，$\hat{\varphi}$ とおく．すると $\hat{\mathcal{B}}\triangleq\bigcup_{m\in\mathcal{M}_f}\hat{\varphi}^{-1}(m)$ に対して

$$W^n(\mathcal{T}^n_{[W]}(\boldsymbol{x})-\hat{\mathcal{B}}|\boldsymbol{x})=0$$

となり，(69) を満たす．また，$\boldsymbol{x}$ が f の符号語でなければ，新しいメッセージ m' を追加し，$\boldsymbol{x}$ をその符号語とする．新しい符号器を $\hat{f}$ とおく．そして，復号器を $\boldsymbol{y}\in\mathcal{T}^n_{[W]}(\boldsymbol{x})-\mathcal{B}$ に対しては m' を復号するように変更し $\hat{\varphi}$ とおく．すると m' の誤り確率は

$$\begin{aligned}1-W^n(\hat{\varphi}^{-1}(m')|f(m'))&=1-W^n(\mathcal{T}^n_{[W]}(\boldsymbol{x})-\mathcal{B}|\boldsymbol{x})\\&\le\epsilon\end{aligned}$$

となる．$(\hat{f},\hat{\varphi})$ は (n,ϵ) 符号となる．$\mathcal{T}^n_{[P]}$ が有限集合であることから，以上の変更を高々有限回繰り返せば (67), (68), (69) を満たす (n,ϵ)-符号を得ることができる．

次に，得られた符号 (f,φ) の性質を調べる．まず，$0<\tau'<\min\{\tau,\epsilon\}$ なる τ' をとる．補題 3 より n が十分大きければ任意の $\boldsymbol{x}\in\mathcal{X}^n$ で

$$W^n(\mathcal{T}^n_{[W]}(\boldsymbol{x})|\boldsymbol{x})\ge 1-\tau'$$

となるので，任意の $\boldsymbol{x}\in\mathcal{T}^n_{[P]}$ で

$$\begin{aligned}W^n(\mathcal{B}|\boldsymbol{x})&\ge W^n(\mathcal{B}\cap\mathcal{T}^n_{[W]}(\boldsymbol{x})|\boldsymbol{x})\\&=W^n(\mathcal{T}^n_{[W]}(\boldsymbol{x})|\boldsymbol{x})-W^n(\mathcal{T}^n_{[W]}(\boldsymbol{x})-\mathcal{B}|\boldsymbol{x})\\&>(1-\tau')-(1-\epsilon)\\&=\epsilon-\tau'\end{aligned} \tag{70}$$

となる．よって $\epsilon' \in (0,1)$ に対して n が十分大きければ

$$\begin{aligned}(\Sigma PW)^n(\mathcal{B}) &= \sum_{\boldsymbol{x}\in\mathcal{X}^n} W^n(\mathcal{B}|\boldsymbol{x})P^n(\boldsymbol{x}) \\ &\geq \sum_{\boldsymbol{x}\in\mathcal{T}^n_{[P]}} W^n(\mathcal{B}|\boldsymbol{x})P^n(\boldsymbol{x}) \\ &\geq (\epsilon-\tau')P^n(\mathcal{T}^n_{[P]}) \\ &\geq (\epsilon-\tau')(1-\epsilon')\end{aligned}$$

となる．$0 < (\epsilon-\tau')(1-\epsilon') < 1$ であるから補題 19 より，n が十分大きければ

$$\frac{1}{n}\log|\mathcal{B}| \geq H(\Sigma PW) - \frac{\tau}{2} \tag{71}$$

となる．

一方，(67) より補題 18 を用いれば，十分大きな n に対して

$$\begin{aligned}|\mathcal{B}| &= \sum_{m\in\mathcal{M}_f} |\varphi^{-1}(m)| \\ &\leq \sum_{m\in\mathcal{M}_f} |\mathcal{T}^n_{[W]}(f(m))| \\ &\leq |\mathcal{M}_f|\exp\left(n\left(H(W|P)+\frac{\tau}{2}\right)\right) \\ \frac{1}{n}\log|\mathcal{M}_f| &\geq \frac{1}{n}\log|\mathcal{B}| - H(W|P) - \frac{\tau}{2}\end{aligned}$$

を得る．ここで (71) を用いれば

$$\begin{aligned}\frac{1}{n}\log|\mathcal{M}_f| &\geq H(\Sigma PW) - H(W|P) - \tau \\ &= I(P,W) - \tau\end{aligned}$$

が成り立つ．□

補題 25 ([1, Corollary 2.1.4]) 無記憶通信路 $\{W : \mathcal{X} \to \mathcal{Y}\}$ に対し，すべての符号語が $\mathcal{T}^n_{[P]}$ に入る符号を考える．このとき任意の $\epsilon, \tau \in (0,1)$ に対して十分大きな n をとれば，どんな (n,ϵ)-符号 (f,φ) でも

$$\frac{1}{n}\log|\mathcal{M}_f| < I(P,W) + \tau$$

を満たす．または，$\mathcal{M}_f$ は符号語が $\mathcal{T}^n_{[P]}$ に入るメッセージからなる集合である．

（証明） 任意の $m \in \mathcal{M}_f$ を考え，$\boldsymbol{x} \triangleq f(m)$ とおく．このとき $\boldsymbol{x} \in \mathcal{T}^n_{[P]}$ であるので補題 1 より $\boldsymbol{y} \in \mathcal{T}^n_{[W]}(\boldsymbol{x})$ ならば $\boldsymbol{y} \in \mathcal{T}^n_{[\Sigma PW]\delta'_n}$ である．ただし $\delta'_n \triangleq 2\delta_n|\mathcal{X}|$ とおいた．ここで

$$\mathcal{B} \triangleq \mathcal{T}^n_{[\Sigma PW]\delta'_n} \tag{72}$$

とおくと

$$\mathcal{T}^n_{[W]}(\boldsymbol{x}) \subset \mathcal{B}$$

なので

$$W^n(\mathcal{T}^n_{[W]}(\boldsymbol{x})|\boldsymbol{x}) \leq W^n(\mathcal{B}|\boldsymbol{x}) \tag{73}$$

が成り立つ．一方，補題 3 より $\epsilon < \tau' < 1$ に対して，n が十分大きければ

$$W^n(\mathcal{T}^n_{[W]}(\boldsymbol{x})|\boldsymbol{x}) > \tau' \tag{74}$$

となるので (73), (74) より

$$W^n(\mathcal{B}|\boldsymbol{x}) > \tau'$$

を得る．ここで誤り確率が ϵ 以下であることから

$$\begin{aligned} W^n(\mathcal{B} \cap \varphi^{-1}(m)|\boldsymbol{x}) &= W^n(\mathcal{B}|\boldsymbol{x}) - W^n(\mathcal{B} - \varphi^{-1}(m)|\boldsymbol{x}) \\ &\geq W^n(\mathcal{B}|\boldsymbol{x}) - (1 - W^n(\varphi^{-1}(m)|\boldsymbol{x})) \\ &\geq W^n(\mathcal{B}|\boldsymbol{x}) - \epsilon \\ &> \tau' - \epsilon \end{aligned}$$

となるが，$0 < \tau' - \epsilon < 1$ であることから補題 20 より，n が十分大きければ

$$\begin{aligned} \frac{1}{n} \log |\mathcal{B} \cap \varphi^{-1}(m)| &\geq H(W|P_{\boldsymbol{x}}) - \frac{\tau}{3} \\ &\geq H(W|P) - \frac{2}{3}\tau \end{aligned}$$

が成り立つ．ここで，$m \in \mathcal{M}_f$ に関して $\varphi^{-1}(m)$ は互いに素なので

$$\begin{aligned} |\mathcal{B}| &\geq \left| \mathcal{B} \cap \bigcup_{m \in \mathcal{M}_f} \varphi^{-1}(m) \right| \\ &= \sum_{m \in \mathcal{M}_f} |\mathcal{B} \cap \varphi^{-1}(m)| \\ &\geq |\mathcal{M}_f| \exp\left(n \left(H(W|P) - \frac{2}{3}\tau \right) \right) \end{aligned}$$

$$\frac{1}{n} \log |\mathcal{M}_f| \leq \frac{1}{n} \log |\mathcal{B}| - H(W|P) + \frac{2}{3}\tau \tag{75}$$

を得る．一方，(72) に注意すれば補題 11 より，n が十分大きければ

$$\frac{1}{n} \log |\mathcal{B}| < H(\Sigma PW) + \frac{\tau}{3} \tag{76}$$

となる. したがって (75), (76) より

$$\begin{aligned}\frac{1}{n}\log|\mathcal{M}_f| &< H(\Sigma PW) - H(W|P) + \tau \\ &= I(P,W) + \tau\end{aligned}$$

が成り立つ. □

定理 26 ([1, Theorem 2.1.5]) $0 < \epsilon < 1$ に対して, 無記憶通信路 $\{W\}$ の ϵ-容量は

$$C_\epsilon = C = \max_P I(P,W)$$

である.

（証明） まず, 下界を証明する. 補題 24 より $\mathcal{X}$ 上の任意の分布 P と任意の $\delta > 0$ に対して十分大きなすべての n で

$$\frac{1}{n}\log|\mathcal{M}_f| \geq I(P,W) - \delta$$

が成り立つ (n,ϵ)-符号が存在する. よって $I(P,W) - \delta$ は ϵ-達成可能なレートとなる. δ は任意なので

$$C_\epsilon \geq I(P,W)$$

となり, さらに P も任意なので

$$C_\epsilon \geq \sup_P I(P,W)$$

となる. I は連続関数であり, かつ sup をとる P の領域は有界閉集合なので

$$C_\epsilon \geq \max_P I(P,W)$$

である.

次に, 上界を証明するために無記憶通信路 $\{W\}$ に対する任意の (n,ϵ)-符号を考える. 系列 $\boldsymbol{x}^n$ のタイプ P に対して, タイプ P の系列に符号化されるメッセージ集合を $\mathcal{M}_f(P)$ とおく.

$$\mathcal{M}_f(P) \triangleq \{m \in \mathcal{M}_f | f(m) \in \mathcal{T}_P^n\}$$

全ての $\tau > 0$, タイプ P, 十分大きなすべての n に対して f において補題 25 を用いると

$$\begin{aligned}\frac{1}{n}\log|\mathcal{M}_f(P)| &< I(P,W) + \tau \\ |\mathcal{M}_f(P)| &< \exp\{n(I(P,W) + \tau)\}\end{aligned}$$

となる．さらに，補題 8 を用いると

$$\begin{aligned}
|\mathcal{M}| &= \sum_P |\mathcal{M}_f(P)| \\
&\le (n+1)^{|\mathcal{X}|} \max_P |\mathcal{M}_f(P)| \\
&< (n+1)^{|\mathcal{X}|} \max_P \{\exp\{n(I(P,W)+\tau)\}\} \\
&= (n+1)^{|\mathcal{X}|} \exp\{n(\max_P I(P,W)+\tau)\}
\end{aligned}$$

を得る．P は系列 $\boldsymbol{x}^n$ のタイプにわたるが，P が $\mathcal{X}$ 上の分布全体にわたるように広げると，

$$\begin{aligned}
|\mathcal{M}| &< (n+1)^{|\mathcal{X}|} \exp\{n(\sup_P I(P,W)+\tau)\} \\
&= (n+1)^{|\mathcal{X}|} \exp\{n(\max_P I(P,W)+\tau)\} \\
\frac{1}{n}\log|\mathcal{M}| &< \max_P I(P,W) + \tau + \frac{|\mathcal{X}|}{n}\log(n+1) \\
&< \max_P I(P,W) + 2\tau \quad \text{（n を十分大きくすれば）}
\end{aligned}$$

となる．したがって，任意の (n,ϵ)-符号に対して

$$C_\epsilon \le \max_P I(P,W) + 2\tau$$

となり，τ は任意なので

$$C_\epsilon \le \max_P I(P,W)$$

である．□

A.3 忠実度規範付き情報源符号化

補題 27 ([1, Lemma 2.2.2]) $\mathcal{X}$ 上の任意の分布 P，無記憶通信路 $\{W\}$ を考える．Δ は歪みの大きさを表し，文字単位の歪みを d で与える．そして

$$\tilde{R}(P,\Delta) \triangleq \min_{W:d(P,W)\le\Delta} I(P,W) \tag{77}$$

$$d(P,W) \triangleq \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x)W(y|x)d(x,y) \tag{78}$$

と定義する．固定された分布 P に対して，$\tilde{R}(P,\Delta)$ は有限値で $\Delta \ge 0$ の非増加凸関数である．さらに $\tilde{R}(P,\Delta)$ は (P,Δ) の連続関数である．

（証明）まず，P を固定する．歪み関数の定義よりどの $x \in \mathcal{X}$ に対して少なくとも一つの $y \in \mathcal{Y}$ で $d(x,y)=0$ となる．よって，最小化の範囲は空にならないので $\tilde{R}(P,\Delta)$ は有限値である．さ

らに, 最小化の範囲は有界閉集合であるので $I(P, W)$ が W の連続関数であることから, (77) の最小値は存在する. Δ に関する単調減少性は明らかである. では, $\tilde{R}(P, \Delta)$ の Δ に関する凸性を示す. $\Delta_1 \geq 0, \Delta_2 \geq 0, 0 < \alpha < 1$ を考える. すると

$$
\begin{aligned}
\alpha\tilde{R}(P, \Delta_1) + (1-\alpha)\tilde{R}(P, \Delta_2) &= \alpha \min_{W : d(P,W) \leq \Delta_1} I(P, W) + (1-\alpha) \min_{W : d(P,W) \leq \Delta_2} I(P, W) \\
&= \alpha I(P, W_1) + (1-\alpha) I(P, W_2) \\
&\geq I(P, \alpha W_1 + (1-\alpha) W_2) \quad \text{（相互情報量の凸性より）} \\
&= I(P, W_3) \\
&\geq \min_{W : d(P,W) \leq \alpha\Delta_1 + (1-\alpha)\Delta_2} I(P, W)
\end{aligned}
$$

となり, $\tilde{R}(P, \Delta)$ は Δ に関して凸関数である. ただし, $W : d(P, W) \leq \Delta_1$ の範囲で $I(P, W)$ を最小とする W を W_1, $W : d(P, W) \leq \Delta_2$ の範囲で $I(P, W)$ を最小とする W を W_2, $W_3 = \alpha W_1 + (1-\alpha) W_2$ とおいた. 最後の不等式では

$$
\begin{aligned}
d(P, W_3) &= \sum_{x,y} P(x) W_3(y|x) d(x, y) \\
&= \sum_{x,y} P(x) d(x, y) \{\alpha W_1(b|a) + (1-\alpha) W_2(b|a)\} \\
&= \sum_{x,y} P(x) d(x, y) \alpha W_1(b|a) + \sum_{x,y} P(x) d(x, y) (1-\alpha) W_2(b|a) \\
&= \alpha d(P, W_1) + (1-\alpha) d(P, W_2) \\
&\leq \alpha\Delta_1 + (1-\alpha)\Delta_2
\end{aligned}
$$

であることを用いた.

さらに, $\tilde{R}(P, \Delta)$ の (P, Δ) に関する連続性を証明するために $P_n \to P, \Delta_n \to \Delta$ なる列 $\{P_n\}$, $\{\Delta_n\}$ を考える.

$\Delta > 0$ の場合を考える. $\tilde{R}(P, \Delta)$ は Δ に関して連続である. 任意の $\epsilon > 0$ に対してある $\delta > 0$ が存在し

$$
\tilde{R}(P, \Delta) + \frac{\epsilon}{2} > \tilde{R}(P, \Delta - \delta)
$$

となる. よってある

$$
d(P, W) \leq \Delta - \delta \tag{79}
$$

なる W が存在して

$$
\tilde{R}(P, \Delta) + \frac{\epsilon}{2} > \tilde{R}(P, \Delta - \delta) = I(P, W) \tag{80}
$$

と書ける. また, $d(P,W)$ の P に関する連続性と $\Delta_n \to \Delta$ より, 十分大きなすべての n に対して

$$d(P_n, W) - d(P, W) < \frac{\delta}{2} \tag{81}$$

$$\Delta - \Delta_n < \frac{\delta}{2} \tag{82}$$

が成り立つ. よって (79), (81), (82) より

$$\begin{aligned} d(P,W) + d(P_n, W) - d(P,W) + \Delta - \Delta_n &< \Delta - \delta + \frac{\delta}{2} + \frac{\delta}{2} \\ d(P_n, W) &< \Delta_n \end{aligned} \tag{83}$$

となる. さらに $I(P,W)$ の P に関する連続性より十分大きなすべての n で

$$I(P,W) > I(P_n, W) - \frac{\epsilon}{2} \tag{84}$$

となる. すると十分大きなすべての n で

$$\begin{aligned} \tilde{R}(P_n,\ \Delta_n) &= \min_{W: d(P_n, W) \le \Delta_n} I(P_n, W) \\ &\le I(P_n, W) \qquad ((83) \text{ より }) \\ &< I(P,W) + \frac{\epsilon}{2} \qquad ((84) \text{ より }) \\ &< \tilde{R}(P, \Delta) + \epsilon \qquad ((80) \text{ より }) \end{aligned}$$

が成り立つ. これは

$$\limsup_{n \to \infty} \tilde{R}(P_n, \Delta_n) \le \tilde{R}(P, \Delta)$$

を意味する.

$\Delta = 0$ の場合を考える. 改めて

$$\tilde{R}(P, 0) = I(P, W) \tag{85}$$

なる W を考える. このとき $d(P,W) = 0$ であるので $P(x) > 0,\ d(x,y) > 0$ なる x, y に対して $W(y|x) = 0$ でなくてはならない. ところが $P(x) = 0$ なる x に対して $W(y|x)$ がどのような値をとろうとも $d(P,W)$ と $I(P,W)$ には影響しないので $d(x,y) > 0$ なる任意の x, y に対して $W(y|x) = 0$ であるとみなしてよい. このとき W は任意の P_n に対して

$$d(P_n, W) = 0 \tag{86}$$

を満たす. また, $I(P,W)$ の P に関する連続性より, 任意の $\epsilon > 0$ に対して十分大きなすべての n で

$$I(P_n, W) < I(P, W) + \epsilon \tag{87}$$

となる．すると

$$
\begin{aligned}
\tilde{R}(P_n, \Delta_n) &= \min_{W:d(P_n,W)\le\Delta_n} I(P_n, W) \\
&\le I(P_n, W) \qquad ((86)\text{ より }) \\
&< I(P, W) + \epsilon \qquad ((87)\text{ より }) \\
&= \tilde{R}(P, 0) + \epsilon \qquad ((85)\text{ より })
\end{aligned}
$$

が成り立つ．これは

$$\limsup_{n\to\infty} \tilde{R}(P_n, \Delta_n) \le \tilde{R}(P, 0)$$

を意味する．

一方

$$\tilde{R}(P_n, \Delta_n) = I(P_n, W_n) \tag{88}$$

となる W_n を選ぶ．このとき

$$d(P_n, W_n) \le \Delta_n$$

である．ここで

$$\tilde{R}(P_{n_k}, \Delta_{n_k}) \to \liminf_{n\to\infty} \tilde{R}(P_n, \Delta_n) \tag{89}$$

かつある W に対して $W_{n_k} \to W$ となるような整数の列 $\{n_k\}$ を考える[*1]．$d(P, W)$ の連続性より

$$d(P, W) = \lim_{k\to\infty} d(P_{n_k}, W_{n_k}) \le \Delta \tag{90}$$

であるから

$$
\begin{aligned}
\tilde{R}(P, \Delta) &= \min_{W:d(P,W)\le\Delta} I(P, W) \\
&\le I(P, W) \\
&= \lim_{k\to\infty} I(P_{n_k}, W_{n_k}) \\
&= \lim_{k\to\infty} \tilde{R}(P_{n_k}, \Delta_{n_k}) \qquad ((88)\text{ より }) \\
&= \liminf_{n\to\infty} \tilde{R}(P_n, \Delta_n)
\end{aligned}
$$

となる．□

[*1] 適当な距離を導入すれば遷移確率行列全体の直径は有限になることに注意する．

定義 28（定義 3 再掲）任意の分布 P を考える. $0 < \epsilon < 1$ が与えられた時, 任意の $\delta > 0$ に対して十分大きなすべての n で

$$\Pr\{d_n(X^n, Y^n) > \Delta\} < \epsilon$$
$$\frac{1}{n}\log|\mathcal{M}_n| < R + \delta$$

を満たす符号器と復号器の列 $\{(\varphi_n, \psi_n)\}_n$ が存在するならば, 符号化レート $R > 0$ は歪み Δ で CK 最大忠実度規範に対して ϵ-達成可能であるという. X^n は分布 P で決まる確率変数であり, Y^n は X^n を符号化し, 復号化したものである. ただし, $\mathcal{M}_n = \varphi_n(\mathcal{X}^n)$ である. 全ての $0 < \epsilon < 1$ に対して R が ϵ-達成可能ならば $R > 0$ は CK 最大忠実度規範に対して達成可能であるという. そして, 上記の意味で ϵ-達成可能なレートの下限を $R_\epsilon(P, \Delta)$ と表し, 達成可能なレートの下限を $R(P, \Delta)$ と表す.

定理 29 ([1, Theorem 2.2.3]) 分布 P をもつ離散無記憶情報源 $\{X_i\}_{i=1}^{\infty}$ に対して, 任意の $0 < \epsilon < 1$ と $\Delta \geq 0$ に対して

$$R_\epsilon(P, \Delta) = R(P, \Delta) = \min_{\substack{P_X = P \\ Y : E[d(X,Y)] \leq \Delta}} I(X;Y)$$

となる.

（証明）はじめに $R(P, \Delta) \leq \min_{\substack{P_X = P \\ Y : E[d(X,Y)] \leq \Delta}} I(X;Y)$ を示す. すなわち, $\tilde{R}(P, \Delta)$ が歪み Δ で ϵ-達成可能であることを示す. そのために以下のようにして, 逆向き無記憶通信路 $\{\hat{W} : \mathcal{Y} \to \mathcal{X}\}$ を構成し, $\{\hat{W}\}$ に対する長さ n のブロック符号を構成する. $\Delta > 0$ のとき, X と Y を $P_X = P, E[d(X,Y)] < \Delta$ を満たす相関のある確率変数とおく. $0 < 2\tau < \hat{\epsilon} < 1$ とする. $\mathcal{Y}_0 \triangleq \{y \in \mathcal{Y} : P_Y(y) > 0\}$ とし, $\hat{W} \triangleq P_{X|Y}$ で定義される無記憶通信路 $\{\hat{W} : \mathcal{Y}_0 \to \mathcal{X}\}$ を考える. $(\hat{f}, \hat{\varphi})$ をこの無記憶通信路に対する $(n, \hat{\epsilon})$-符号とおき, n が十分大きいとき, すべての $m \in \mathcal{M}_{\hat{f}}$ に対して

$$\hat{f}(m) \in \mathcal{T}^n_{[Y]} \subset \mathcal{Y}^n_0 \tag{91}$$
$$\hat{\varphi}^{-1}(m) \subset \mathcal{T}^n_{[X|Y]}(\hat{f}(m)) \tag{92}$$

が成り立つとする. さらに, $(\hat{f}, \hat{\varphi})$ は上の特性を持つような *extension* をもたないとする. 集合

$$\mathcal{B} \triangleq \bigcup_{m \in \mathcal{M}_{\hat{f}}} \hat{\varphi}^{-1}(m) \subset \mathcal{X}^n$$

は (70) より $\boldsymbol{y} \in \mathcal{T}^n_{[Y]}$ に対して十分大きい n で

$$\hat{W}^n(\mathcal{B}|\boldsymbol{y}) > \hat{\epsilon} - \tau$$

を満たす．補題 2 より $P_Y^n(\mathcal{T}_{[Y]}^n) \to 1$ なので

$$\begin{aligned}
P^n(\mathcal{B}) &= P_X^n(\mathcal{B}) \\
&= \sum_{\boldsymbol{y}} P_{XY}^n(\mathcal{B}, \boldsymbol{y}) \\
&= \sum_{\boldsymbol{y}} P_Y^n(\boldsymbol{y}) \cdot \hat{W}^n(\mathcal{B}|\boldsymbol{y}) \\
&\geq \sum_{\boldsymbol{y} \in \mathcal{T}_{[Y]}^n} P_Y^n(\boldsymbol{y}) \cdot \hat{W}^n(\mathcal{B}|\boldsymbol{y}) \\
&\geq \sum_{\boldsymbol{y} \in \mathcal{T}_{[Y]}^n} P_Y^n(\boldsymbol{y})(\hat{\epsilon} - \tau) \\
&= (\hat{\epsilon} - \tau) \cdot P_Y^n(\mathcal{T}_{[Y]}^n) \\
&> \hat{\epsilon} - 2\tau \quad (n \text{ を十分大きくすれば})
\end{aligned}$$

となる．補題 25 より

$$\frac{1}{n} \log |\mathcal{M}_{\hat{f}}| < I(P_Y, \hat{W}) + 2\tau \tag{93}$$

である．

復号器 $\hat{\varphi}$ は $\mathcal{X}^n$ を $\mathcal{M}' \supset \mathcal{M}_{\hat{f}}$ に写す．元のメッセージ集合の中には復号できないものを受信したとき $m^* \notin \mathcal{M}_{\hat{f}}$ を復号することにして，$\mathcal{M}' = \mathcal{M}_{\hat{f}} \cup \{m^*\}$ とする．よって $|\mathcal{M}'| = |\mathcal{M}_{\hat{f}}| + 1$ である．ここで

$$f(\boldsymbol{x}) \triangleq \hat{\varphi}(\boldsymbol{x}) \tag{94}$$

$$\varphi(m) \triangleq \begin{cases} \hat{f}(m) & m \in \mathcal{M}_{\hat{f}} \\ arbitary & m \notin \mathcal{M}_{\hat{f}} \end{cases} \tag{95}$$

$$d_{\max} \triangleq \max_{a \in \mathcal{X}, b \in \mathcal{Y}} d(a, b)$$

とおく．(91), (92) より $\boldsymbol{x} \in \mathcal{B}$, $\boldsymbol{y} \triangleq \varphi(f(\boldsymbol{x}))$ に対して $\boldsymbol{y} \in \mathcal{T}_{[Y]}^n$, $\boldsymbol{x} \in \mathcal{T}_{[X|Y]}(\boldsymbol{y})$ となり，補題

1 の効果より

$$
\begin{aligned}
d_n(\boldsymbol{x}, \varphi(f(\boldsymbol{x}))) &= d_n(\boldsymbol{x}, \boldsymbol{y}) \\
&= \frac{1}{n} \sum_{a \in \mathcal{X}} \sum_{b \in \mathcal{Y}} N(a, b|\boldsymbol{x}, \boldsymbol{y}) d(a, b) \\
&= \sum_{a \in \mathcal{X}} \sum_{b \in \mathcal{Y}} \frac{1}{n} N(a, b|\boldsymbol{x}, \boldsymbol{y}) d(a, b) \\
&\leq \sum_{a \in \mathcal{X}} \sum_{b \in \mathcal{Y}} (P_{XY}(a, b) + 2\delta_n) d(a, b) \\
&= \sum_{a \in \mathcal{X}} \sum_{b \in \mathcal{Y}} P_{XY}(a, b) d(a, b) + \sum_{a \in \mathcal{X}} \sum_{b \in \mathcal{Y}} 2\delta_n d(a, b) \\
&\leq \sum_{a \in \mathcal{X}} \sum_{b \in \mathcal{Y}} P_{XY}(a, b) d(a, b) + |\mathcal{X}||\mathcal{Y}| 2\delta_n d_{\max} \\
&= E[d(X, Y)] + 2|\mathcal{X}||\mathcal{Y}|\delta_n d_{\max} \\
&< \Delta + 2|\mathcal{X}||\mathcal{Y}|\delta_n d_{\max}
\end{aligned}
$$

である．よって，任意の $\zeta > 0$ に対して十分大きなすべての n で

$$
d_n(\boldsymbol{x}, \varphi(f(\boldsymbol{x}))) \leq \Delta + \zeta \tag{96}
$$

となる．ここで，$\epsilon = 1 - \hat{\epsilon} + 2\tau$ とおく．すると十分大きいすべての n で $P^n(\mathcal{B}) > 1 - \epsilon$ となるので (96) より歪みについては

$$
\begin{aligned}
&\Pr\{d_n(X^n, \varphi(f(X^n))) \leq \Delta + \zeta\} > 1 - \epsilon \\
&\Pr\{d_n(X^n, \varphi(f(X^n))) > \Delta + \zeta\} < \epsilon
\end{aligned}
$$

と表すことができる．一方，レートについては十分小さい $\tau > 0$ を選び，(93) より

$$
\frac{1}{n} \log \|f\| < I(X; Y) + 2\tau
$$

となる．よって，$I(X; Y)$ は歪み $\Delta + \zeta$ で ϵ-達成可能なレートである．ここで，

$$
R > \tilde{R}(P, \Delta) = \min_{\substack{P_X = P \\ Y: E[d(X;Y)] \leq \Delta}} I(X; Y)
$$

なる任意の R をとる．$\tilde{R}$ は Δ に関して連続なので $\zeta > 0$ を十分小さくとれば

$$
R > \tilde{R}(P, \Delta - \zeta) \geq \tilde{R}(P, \Delta)
$$

が成り立つ．$\tilde{R}(P, \Delta - \zeta)$ の定義からある X, Y が存在して $P_X = P,\ E[d(X; Y)] \leq \Delta - \zeta < \Delta,\ \tilde{R}(P, \Delta - \zeta) = I(X; Y)$ が成り立つ．すると，上の議論より $\tilde{R}(P, \Delta - \zeta) = I(X; Y)$ は歪

み Δ で ϵ-達成可能である．よって R も歪み Δ で ϵ-達成可能である．$R > \tilde{R}(P,\Delta)$ は任意だったので

$$\tilde{R}(P,\Delta) \geq R_\epsilon(P,\Delta)$$

となる．特に ϵ が任意であったことから R は歪み Δ で達成可能である．したがって $\tilde{R}(P,\Delta) \geq R(P,\Delta)$ となる．

$\Delta = 0$ のとき $P_X = P, E[d(X,Y)] = 0$ を満たす X と Y について前のことを繰り返す．(92) に注意して，復号領域はティピカルなものの集合なので $\hat{\varphi}(\boldsymbol{x}) = m \in \mathcal{M}_{\hat{f}}$ は $\hat{W}(\boldsymbol{x}|\hat{f}(m)) > 0$ を意味する．このとき

$$\begin{aligned}
E[d_n(X^n,Y^n)] &= \frac{1}{n}\sum_{i=1}^{n} E[d(X_i,Y_i)] \\
&= d_n(\boldsymbol{x},\hat{f}(m)) \\
&= 0
\end{aligned}$$

となる．一方，(94) より $\boldsymbol{x} \in \mathcal{B}$ に対して $m \in \mathcal{M}_{\hat{f}}$ が存在して $f(\boldsymbol{x}) = \hat{\varphi}(\boldsymbol{x}) = m$ となる．このとき φ は $\varphi(m) = \hat{f}(m)$ を復号するので

$$d_n(\boldsymbol{x},\varphi(f(\boldsymbol{x}))) = d_n(\boldsymbol{x},\hat{f}(m)) = 0$$

となる．よって，十分大きいすべての n で $P^n(\mathcal{B}) > 1-\epsilon$ となるので

$$\begin{gathered}
\Pr\{d_n(\boldsymbol{X},\varphi(f(\boldsymbol{X}))) = 0\} > 1-\epsilon \\
\Pr\{d_n(\boldsymbol{X},\varphi(f(\boldsymbol{X}))) > 0\} < \epsilon
\end{gathered}$$

となる．レートは前と同様に十分小さい $\tau > 0$ を選び，(93) より

$$\frac{1}{n}\log\|f\| < I(X;Y) + 2\tau$$

となる．以上より，$I(X;Y)$ は歪み 0 で ϵ-達成可能なレートである．つまり，$I(X;Y) \geq R_\epsilon(P,0)$ となる．今，X,Y は条件 $P_X = P, E[d(X,Y)] = 0$ を満たす任意の確率変数だったので

$$\tilde{R}(P,0) = \min_{\substack{P_X = P \\ Y: E[d(X,Y)]=0}} I(X;Y) \geq R_\epsilon(P,0) \tag{97}$$

である．特に ϵ が任意であったことから $\tilde{R}(P,0)$ は歪み 0 で達成可能である．したがって $\tilde{R}(P,0) \geq R(P,0)$ となる．

次に, $\Delta \geq 0$ に対して $R(P,\Delta) \geq \min_{\substack{P_X=P \\ Y:E[d(X,Y)]\leq\Delta}} I(X;Y)$ を示す. これを示すために, 任意の $\epsilon \in (0,1)$ と $\delta > 0$ に対して, n が十分大きいとき $\Pr\{d_n(X^n, \varphi(f(X^n)) > \Delta\} < \epsilon$ で

$$\frac{1}{n}\log\|f\| \geq \tilde{R}(P,\Delta) - \delta \tag{98}$$

となることをを示す. (f,φ) をそのような符号とする. すなわち, $g(\boldsymbol{x}) \triangleq \varphi(f(\boldsymbol{x}))$ とおき, $\Pr\{d_n(X^n, g(X^n)) \leq \Delta\} \geq 1-\epsilon$ を満たすとする. ここで,

$$\mathcal{A} \triangleq \{\boldsymbol{x} : \boldsymbol{x} \in \mathcal{T}^n_{[P]}\ d_n(\boldsymbol{x}, g(\boldsymbol{x})) \leq \Delta\}$$

と定める. すると, 十分大きな n に対して

$$\begin{aligned} P^n(\mathcal{A}^c) &\leq \Pr\{X^n \notin \mathcal{T}^n_{[n]}\} + \Pr\{d_n(X^n, g(X^n)) > \Delta\} \\ &\leq 2\epsilon \end{aligned}$$

すなわち

$$P^n(\mathcal{A}) > 1 - 2\epsilon$$

となる. $\epsilon > 0$ が任意であることから補題 19 より, 任意の $\tau > 0$ に対して

$$\begin{aligned} \frac{1}{n}\log|\mathcal{A}| &\geq H(P) - \tau \\ |\mathcal{A}| &\geq \exp\{n(H(P)-\tau)\} \end{aligned} \tag{99}$$

となる. 任意の $\boldsymbol{y} \in \mathcal{Y}^n$ と, $\mathcal{X}^n \times \mathcal{Y}^n$ 上の同時タイプ $\tilde{P}$ に対して $\boldsymbol{x} \in \mathcal{X}^n$ の数は, 補題 13 より上限 $\exp\{n(H(\tilde{X}|\tilde{Y}))\}$ を持つ. ただし, $\tilde{X}$, $\tilde{Y}$ は同時分布 $\tilde{P}$ に従う確率変数である. $d_n(\boldsymbol{x},\boldsymbol{y}) \leq \Delta$ と $\boldsymbol{x} \in \mathcal{T}^n_{[P]}$ ならば

$$\begin{aligned} E[d(\tilde{X},\tilde{Y})] &= \sum_{a,b} d(a,b)\tilde{P}(a,b) \\ &= \sum_{a,b} d(a,b)P_{\boldsymbol{xy}}(a,b) \\ &= \sum_{a,b} d(a,b)\frac{1}{n}N(a,b|\boldsymbol{x},\boldsymbol{y}) \\ &= \frac{1}{n}\sum_{a,b} d(a,b)N(a,b|\boldsymbol{x},\boldsymbol{y}) \\ &= \frac{1}{n}\sum_{i=1}^{n} d(x_i,y_i) \\ &= d_n(\boldsymbol{x},\boldsymbol{y}) \end{aligned}$$

となる．また

$$\begin{aligned}
P_{\tilde{X}}(a) &= \sum_b \tilde{P}(a,b) \\
&= \sum_b P_{\boldsymbol{x}\boldsymbol{y}}(a,b) \\
&= \sum_b \frac{1}{n} N(a,b|\boldsymbol{x},\boldsymbol{y}) \\
&= \frac{1}{n} N(a|\boldsymbol{x}) \\
&= P_{\boldsymbol{x}}(a)
\end{aligned}$$

を得るので

$$E[d(\tilde{X},\tilde{Y})] \le \Delta \tag{100}$$

$$|P_{\tilde{X}}(a) - P(a)| \le \delta_n \tag{101}$$

であることに注意する．(101) のもとで $H(\tilde{X}|\tilde{Y})$ を最大にする $(\tilde{X},\tilde{Y})$ の組を考える．ある $\boldsymbol{x} \in \mathcal{A}$ に対して，$\boldsymbol{y} = g(\boldsymbol{x})$ となる $\boldsymbol{y}$ の集合を $\mathcal{C}$ とかく．

$$\mathcal{C} \triangleq \{\boldsymbol{y} \in \mathcal{Y}^n | \boldsymbol{y} = g(\boldsymbol{x})\} = g(\mathcal{A})$$

さらに, 集合 $\mathcal{A}$ の大きさは

$$
\begin{aligned}
|\mathcal{A}| &= |\{\boldsymbol{x} \in \mathcal{T}_{[P]}^n | d_n(\boldsymbol{x}, g(\boldsymbol{x})) \le \Delta\}| \\
&= \left| \bigcup_{\boldsymbol{y} \in \mathcal{C}} \{\boldsymbol{x} \in \mathcal{T}_{[P]}^n | d_n(\boldsymbol{x}, \boldsymbol{y}) \le \Delta, \boldsymbol{y} = g(\boldsymbol{x})\} \right| \\
&\le \left| \bigcup_{\boldsymbol{y} \in \mathcal{C}} \{\boldsymbol{x} \in \mathcal{T}_{[P]}^n | d_n(\boldsymbol{x}, \boldsymbol{y}) \le \Delta\} \right| \\
&\le \sum_{\boldsymbol{y} \in \mathcal{C}} |\{\boldsymbol{x} \in \mathcal{T}_{[P]}^n | d_n(\boldsymbol{x}, \boldsymbol{y}) \le \Delta\}| \\
&= \sum_{\boldsymbol{y} \in \mathcal{C}} \sum_{\tilde{P}: \Sigma_{\boldsymbol{x}} \tilde{P} = P_{\boldsymbol{y}}} |\{\boldsymbol{x} \in \mathcal{T}_{[P]}^n | d_n(\boldsymbol{x}, \boldsymbol{y}) \le \Delta, P_{\boldsymbol{x}\boldsymbol{y}} = \tilde{P}\}| \\
&\le \sum_{\boldsymbol{y} \in \mathcal{C}} \sum_{\tilde{P}} \big| \{\boldsymbol{x} \big| E[d(\tilde{X}, \tilde{Y})] \le \Delta, |P_{\tilde{X}}(a) - P(a)| \le \delta_n, P_{\boldsymbol{x}\boldsymbol{y}} = \tilde{P}\} \big| \\
&= \sum_{\boldsymbol{y} \in \mathcal{C}} \sum_{\substack{\tilde{P} \\ E[d(\tilde{X}, \tilde{Y})] \le \Delta, |P_{\tilde{X}}(a) - P(a)| \le \delta_n}} \big| \{\boldsymbol{x} \big| P_{\boldsymbol{x}\boldsymbol{y}} = \tilde{P}\} \big| \\
&\le \sum_{\boldsymbol{y} \in \mathcal{C}} \sum_{\substack{\tilde{P} \\ E[d(\tilde{X}, \tilde{Y})] \le \Delta, |P_{\tilde{X}}(a) - P(a)| \le \delta_n}} \exp\{nH(\tilde{X}|\tilde{Y})\} \\
&\le \max_{\substack{\tilde{P} \\ E[d(\tilde{X}, \tilde{Y})] \le \Delta, |P_{\tilde{X}}(a) - P(a)| \le \delta_n}} \{\exp(nH(\tilde{X}|\tilde{Y}))\} \cdot \sum_{\boldsymbol{y} \in \mathcal{C}} \sum_{\tilde{P}} 1 \\
&= \exp(nH(\tilde{X}|\tilde{Y}))\} \|g\| (n+1)^{|\mathcal{X}||\mathcal{Y}|}
\end{aligned} \tag{102}
$$

となる. ただし, 最後の行で (100), (101) のもとで $H(\tilde{X}|\tilde{Y})$ を最大化する $\tilde{X}, \tilde{Y}$ をとった. (101) より補題 6 を用いて十分大きなすべての n で

$$
|H(P) - H(\tilde{X})| < \tau
$$

となり, (102) より十分大きなすべての n で

$$
\exp\{n(H(\tilde{X}) - 2\tau)\} \le |\mathcal{A}| \le \|g\| \exp\{n(H(\tilde{X}|\tilde{Y}) + \tau)\}
$$

を得る. ゆえに

$$
\begin{aligned}
\|g\| &\ge \exp\{n(H(\tilde{X}) - 2\tau)\} \frac{1}{\exp\{n(H(\tilde{X}|\tilde{Y}) + \tau)\}} \\
&= \exp\{n(H(\tilde{X}) - 2\tau) - n(H(\tilde{X}|\tilde{Y}) + \tau)\} \\
&= \exp\{n(H(\tilde{X}) - H(\tilde{X}|\tilde{Y}) - 3\tau)\} \\
&= \exp\{n(I(\tilde{X}; \tilde{Y}) - 3\tau)\}
\end{aligned}
$$

となる．そして

$$
\begin{aligned}
\frac{1}{n}\log\|f\| &\geq \frac{1}{n}\log\|g\| \\
&\geq \frac{1}{n}\log\exp\{n(I(\tilde{X};\tilde{Y})-3\tau)\} \\
&= I(\tilde{X};\tilde{Y})-3\tau \\
&\geq \min_{\substack{W\cdot P_Y=P_{\tilde{X}}\cdot W \\ E[d(\tilde{X},Y)]\leq\Delta}} I(\tilde{X};Y)-3\tau \\
&= \tilde{R}(P_{\tilde{X}},\Delta)-3\tau \\
&\geq \tilde{R}(P,\Delta)-4\tau
\end{aligned}
$$

を得る．ただし，最後の不等式では (101) より $|\tilde{R}(P_{\tilde{X}},\Delta)-\tilde{R}(P,\Delta)|<\tau$ となることを用いた．$\tau=\frac{\delta}{4}$ とすれば，(98) を得る．□