

信州大学工学部

学士論文

レート・歪み問題における平均を用いた2つの規範の
関係について

指導教員 西新 幹彦 准教授

学科 電子情報システム工学科
学籍番号 18T2086K
氏名 高尾 優介

2023 年 10 月 2 日

目次

1	はじめに	1
2	レート・歪み理論	1
2.1	平均規範	2
2.2	CK 平均規範	3
3	CK 順定理	4
4	CK 逆定理	7
5	系	8
6	まとめ	9
	謝辞	9
	参考文献	9
	付録 A 補題 1	10
	付録 B 定理 1 の証明	12
	B.1 順定理の証明	13
	B.2 逆定理の証明	17
	付録 C 定理 1 の証明の補足	19
	C.1 Markov の不等式	19
	C.2 Chebychev の不等式	20
	C.3 大数の法則	20
	C.4 相互情報量の連続性の証明	21
	C.5 γ に関して連続であることの証明	22

1 はじめに

スマートフォンの通話では、話し手が声を出すと、その声をスマートフォンのマイクが拾い、それを聞き手側に送信する。聞き手側では、受け取った音から話し手の元の声が聞こえるような音を再現する。このとき聞き取りにあまり影響を与えない音や不要な雑音等を除去すると効率よくデータを送ることができるが、除去する量が多ければ元のデータとの差異は大きくなる。この差異を一般に歪みという。実数値などをとる連続情報源を有限のビット長で符号化する場合、復元された値を元の値と完全に一致させることはできない。よって元のデータと復号されたデータとの間に歪みを許容しなければならない。本研究では情報源を確率変数列として表すので、歪みも確率変数列となる。歪みが満たすべき条件を数学的に表したものを規範と呼ぶ。歪みが規範を満たす範囲内で符号化レートを小さくすることを考えていくのがレート・歪み問題である。

今日の情報理論において一般に用いられている規範と Csiszár and Körner の 1981 年の著書 [1] で定義される規範の 2 つを本研究で扱う。これらの規範は、歪みの期待値を上から抑えるという点で同じであるが、上極限を用いるか否かという数学的な定義方法が異なる。

本論文では 2 章でレート・歪み問題を考える上で前提となる情報源や歪み測度、2 つの規範などの概念を定義する。3 章、4 章で 2 つの規範のレート・歪み関数の関係を明らかにし、5 章で 3 章、4 章の結果から得られる系を述べ、最後に 6 章で本論文をまとめる。

2 レート・歪み理論

本章ではレート・歪み問題を定式化する。特に、歪みに対する規範を 2 種類導入する。

$\mathbf{X} = \{X^n = (X_1, X_2, \dots, X_n)\}_{n=1}^{\infty}$ を一般情報源とし、情報源アルファベットを $\mathcal{X}$ と表す。 $\mathcal{X}$ は有限集合であると仮定する。また復号アルファベットを $\mathcal{Y}$ とする。長さ n のデータ $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathcal{X}^n$ と $\mathbf{y} = (y_1, y_2, \dots, y_n) \in \mathcal{Y}^n$ の間の歪みを関数 $d_n : \mathcal{X}^n \times \mathcal{Y}^n \rightarrow [0, +\infty)$ で測る。これを歪み測度という。

情報源 $\mathbf{X}$ の符号化は次の様に行う。 M_n 個の符号語の集合を $\mathcal{M}_n = \{1, 2, \dots, M_n\}$ とする。そして、符号器 $\varphi_n : \mathcal{X}^n \rightarrow \mathcal{M}_n$ と、復号器 $\psi_n : \mathcal{M}_n \rightarrow \mathcal{Y}^n$ を考える。情報源の出力 $\mathbf{x} \in \mathcal{X}^n$ は符号器 φ_n によって符号化されたあと復号器 ψ_n によって歪みを含んだ状態で復元される。また、符号器 φ_n と復号器 ψ_n の組 (φ_n, ψ_n) を符号と呼ぶ。

レート・歪み問題を定式化するには上記のもとで歪みに対する規範を定めなければならない。以下の 2.1, 2.2 節ではそれぞれ異なる 2 つの歪み規範に基づいて問題の定式化を行う。

2.1 平均規範

この節で述べるのは、レート・歪み問題の標準的な定式化である。

定義 1 レート・歪み領域の点 (R, D) が平均規範達成可能とは

$$\limsup_{n \rightarrow \infty} \mathbb{E} \left[\frac{1}{n} d_n(X^n, \psi_n(\varphi_n(X^n))) \right] \leq D \quad (1)$$

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log M_n \leq R \quad (2)$$

を満たす符号の列 $\{(\varphi_n, \psi_n)\}_{n=1}^{\infty}$ が存在することである。式 (1) を平均規範と呼ぶ。

定義 1 は即ち、任意の小さな数 $\delta_1 > 0, \delta > 0$ に対して、十分大きな全ての n で

$$\mathbb{E} \left[\frac{1}{n} d_n(X^n, \psi_n(\varphi_n(X^n))) \right] < D + \delta_1 \quad (3)$$

$$\frac{1}{n} \log M_n < R + \delta \quad (4)$$

を満たす符号の列 $\{(\varphi_n, \psi_n)\}_{n=1}^{\infty}$ が存在することである。

定義 2 平均規範に基づくレート・歪み関数を

$$R(D) \triangleq \inf \{ R \mid (R, D) \text{ が平均規範達成可能} \} \quad (5)$$

と定義する。

以上の定義のもとで、情報源が定常無記憶である場合に知られている結果を紹介する。まず一般に、2 つの確率変数 X, Y に対して

$$I(X; Y) \triangleq \sum_{(x, y) \in \mathcal{X} \times \mathcal{Y}} P_{XY}(x, y) \log \frac{P_{Y|X}(y | x)}{P_Y(y)} \quad (6)$$

を定義する。これは X と Y の間の相互情報量と呼ばれる。また、歪み測度 d_n は加法的であると仮定する。すなわち、系列の間の歪みは文字の歪み $d(x, y) = d_1(x, y)$ を用いて

$$d_n(\mathbf{x}, \mathbf{y}) = \sum_i d(x_i, y_i) \quad (7)$$

と表されると仮定する。以上の定義のもとで次の定理が成り立つことが知られている [3]。

定理 1 定常無記憶情報源に対する平均規範のレート・歪み関数は

$$R(D) = \min_{Y: \mathbb{E}[d(X, Y)] \leq D} I(X; Y) \quad (8)$$

で与えられる．ただし，右辺の最小値は X と相関のある確率変数 Y のうち， $\mathbb{E}[d(X; Y)] \leq D$ を満たす範囲でとる．

定理 1 の証明は付録 B に記載する．

2.2 CK 平均規範

この節では Csiszár and Körner の 1981 年の著書 [1] で定義された規範に基づいて問題の定式化を行う．

定義 3 レート・歪み領域の点 (R, D) が CK 平均規範達成可能とは，任意の小さな数 $\delta > 0$ に対して，十分大きな全ての n で

$$\mathbb{E} \left[\frac{1}{n} d_n(X^n, \psi_n(\varphi_n(X^n))) \right] \leq D \quad (9)$$

$$\frac{1}{n} \log M_n < R + \delta \quad (10)$$

を満たす符号の列 $\{(\varphi_n, \psi_n)\}_{n=1}^{\infty}$ が存在することである．式 (9) を CK 平均規範と呼ぶ．

定義 1 と定義 3 の違いは，歪みを評価するために上極限を用いるかどうかである．

定義 4 CK 平均規範に基づくレート・歪み関数を

$$R_{\text{CK}}(D) \triangleq \inf \{ R \mid (R, D) \text{ が CK 平均規範達成可能} \} \quad (11)$$

と定義する．

次の定理が知られている [1]．

定理 2 定常無記憶情報源に対する CK 平均規範のレート・歪み関数は

$$R_{\text{CK}}(D) = \min_{Y: \mathbb{E}[d(X, Y)] \leq D} I(X; Y) \quad (12)$$

を満たす．

この定理は順定理と逆定理によって示される．本研究では，順定理を文献 [4] に基づいて文献 [1] より手短かに証明した．これを 3 章に記す．

さらに本研究では，一般情報源に対する CK 平均規範の結果として次を示した．

定理 3 一般情報源に対する CK 平均規範のレート・歪み関数は

$$R_{\text{CK}}(D) \geq R(D) \quad (13)$$

を満たす．

この定理は 4 章に記す.

3 CK 順定理

定理 2 の順定理の証明を行う.

定理 2 の順定理

$$R_{CK}(D) \leq \min_{Y: \mathbb{E}[d(X,Y)] \leq D} I(X;Y) \quad (14)$$

〔証明〕 D を固定し, 任意の $\gamma > 0$ を考える. ここで

$$\mathbb{E}[d(X,Y)] + 2\gamma \leq D \quad (15)$$

を満たす任意の Y をとり, その上で

$$R \triangleq I(X;Y) + 2\gamma \quad (16)$$

とおく. $X^n Y^n \triangleq (X_1 Y_1, X_2 Y_2, \dots, X_n Y_n)$ を XY の独立同一分布な系列として $X^n \triangleq (X_1, X_2, \dots, X_n)$, $Y^n \triangleq (Y_1, Y_2, \dots, Y_n)$ とおく. そこで

$$\mathcal{T}_n^{(1)} \triangleq \left\{ (\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n \left| \left| \frac{1}{n} \log \frac{P_{Y^n|X^n}(\mathbf{y} | \mathbf{x})}{P_{Y^n}(\mathbf{y})} - I(X;Y) \right| < \gamma \right. \right\} \quad (17)$$

$$\mathcal{T}_n^{(2)} \triangleq \left\{ (\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n \left| \left| \frac{1}{n} d_n(\mathbf{x}, \mathbf{y}) - \mathbb{E}[d(X,Y)] \right| < \gamma \right. \right\} \quad (18)$$

として $\mathcal{T}_n \triangleq \mathcal{T}_n^{(1)} \cap \mathcal{T}_n^{(2)}$ とおく. $X^n Y^n$ が独立同一分布なので,

$$\frac{1}{n} \log \frac{P_{Y^n|X^n}(Y^n | X^n)}{P_{Y^n}(Y^n)} = \frac{1}{n} \log \frac{\prod_{i=1}^n P_{Y|X}(Y_i | X_i)}{\prod_{i=1}^n P_Y(Y_i)} \quad (19)$$

$$= \frac{1}{n} \sum_{i=1}^n \log \frac{P_{Y|X}(Y_i | X_i)}{P_Y(Y_i)} \quad (20)$$

となる. 式 (20) の和の中の各項は独立でその期待値は

$$\mathbb{E} \left[\log \frac{P_{Y|X}(Y_i | X_i)}{P_Y(Y_i)} \right] = \sum_{x,y} P_{XY}(x,y) \log \frac{P_{Y|X}(y | x)}{P_Y(y)} \quad (21)$$

$$= I(X;Y) \quad (22)$$

であり, 分散は有界である [2]. $\mathcal{T}_n^{(1)}$ の定義に注意して大数の法則を適用すれば

$$\Pr\{X^n Y^n \in \mathcal{T}_n^{(1)}\} \rightarrow 1 \quad (n \rightarrow \infty) \quad (23)$$

を得る. 大数の法則の証明は付録 C に記載する. 同様にして $\mathbb{E}[d(X,Y)]$ が有界ならば

$$\Pr\{X^n Y^n \in \mathcal{T}_n^{(2)}\} \rightarrow 1 \quad (n \rightarrow \infty) \quad (24)$$

を得る。したがって、

$$\Pr\{X^n Y^n \in \mathcal{T}_n\} = 1 - \Pr\{X^n Y^n \notin \mathcal{T}_n\} \quad (25)$$

$$\geq 1 - \Pr\{X^n Y^n \notin \mathcal{T}_n^{(1)}\} - \Pr\{X^n Y^n \notin \mathcal{T}_n^{(2)}\} \quad (26)$$

$$= 1 - (1 - \Pr\{X^n Y^n \in \mathcal{T}_n^{(1)}\}) - (1 - \Pr\{X^n Y^n \in \mathcal{T}_n^{(2)}\}) \quad (27)$$

$$\rightarrow 1 \quad (n \rightarrow \infty) \quad (28)$$

が成立する。ここで

$$\mathcal{B}_n(\mathbf{y}) \triangleq \left\{ \mathbf{x} \in \mathcal{X}^n \mid \frac{1}{n} \log \frac{P_{X^n|Y^n}(\mathbf{x} | \mathbf{y})}{P_{X^n}(\mathbf{x})} \leq R - \gamma \right\} \quad (29)$$

$$\mathcal{S}_n(\mathbf{y}) \triangleq \{ \mathbf{x} \in \mathcal{X}^n \mid (\mathbf{x}, \mathbf{y}) \in \mathcal{T}_n^{(2)} \} \quad (30)$$

とおく。すると付録 A に示す補題 1 より

$$\frac{1}{n} \log M_n \leq R \quad (31)$$

$$\Pr\{X^n \notin \mathcal{S}_n(\psi_n(\varphi_n(X^n)))\} \leq \Pr\{X^n \notin \mathcal{B}_n(Y^n)\} + \Pr\{X^n \notin \mathcal{S}_n(Y^n)\} + e^{-n\gamma} \quad (32)$$

を満たす符号 (φ_n, ψ_n) が存在する。式 (32) の右辺の第 1 項に関して、式 (29), (17) より、

$$\Pr\{X^n \in \mathcal{B}_n(Y^n)\} = \Pr\left\{ \frac{1}{n} \log \frac{P_{X^n|Y^n}(X^n | Y^n)}{P_{X^n}(X^n)} \leq I(X; Y) + \gamma \right\} \quad (33)$$

$$= \Pr\left\{ \frac{1}{n} \log \frac{P_{Y^n|X^n}(Y^n | X^n)}{P_{Y^n}(Y^n)} \leq I(X; Y) + \gamma \right\} \quad (34)$$

$$= \Pr\left\{ \frac{1}{n} \log \frac{P_{Y^n|X^n}(Y^n | X^n)}{P_{Y^n}(Y^n)} - I(X; Y) \leq \gamma \right\} \quad (35)$$

$$\geq \Pr\left\{ \left| \frac{1}{n} \log \frac{P_{Y^n|X^n}(Y^n | X^n)}{P_{Y^n}(Y^n)} - I(X; Y) \right| < \gamma \right\} \quad (36)$$

$$= \Pr\{X^n Y^n \in \mathcal{T}_n^{(1)}\} \quad (37)$$

$$\stackrel{(23)}{\rightarrow} 1 \quad (n \rightarrow \infty) \quad (38)$$

となる。よって

$$\Pr\{X^n \notin \mathcal{B}_n(Y^n)\} \rightarrow 0 \quad (n \rightarrow \infty) \quad (39)$$

を得る。また、式 (32) の右辺の第 2 項に関して式 (30), (24) より

$$\Pr\{X^n \in \mathcal{S}_n(Y^n)\} = \Pr\{X^n Y^n \in \mathcal{T}_n^{(2)}\} \quad (40)$$

$$\rightarrow 1 \quad (n \rightarrow \infty) \quad (41)$$

となる。よって

$$\Pr\{X^n \notin \mathcal{S}_n(Y^n)\} \rightarrow 0 \quad (n \rightarrow \infty) \quad (42)$$

を得る．以上より，

$$\Pr\{X^n \notin \mathcal{S}_n(\psi_n(\varphi_n(X^n)))\} \rightarrow 0 \quad (n \rightarrow \infty) \quad (43)$$

となる．一方で式 (32) の左辺は式 (30), (18) より

$$\Pr\{X^n \notin \mathcal{S}_n(\psi_n(\varphi_n(X^n)))\} = \Pr\{(X^n, \psi_n(\varphi_n(X^n))) \notin \mathcal{T}_n^{(2)}\} \quad (44)$$

$$= \Pr\left\{\left|\frac{1}{n}d_n(X^n, \psi_n(\varphi_n(X^n))) - \mathbb{E}[d(X, Y)]\right| \geq \gamma\right\} \quad (45)$$

$$\geq \Pr\left\{\frac{1}{n}d_n(X^n, \psi_n(\varphi_n(X^n))) \geq \mathbb{E}[d(X, Y)] + \gamma\right\} \quad (46)$$

である．したがって式 (24) より

$$\Pr\left\{\frac{1}{n}d_n(X^n, \psi_n(\varphi_n(X^n))) \geq \mathbb{E}[d(X, Y)] + \gamma\right\} \rightarrow 0 \quad (47)$$

となる．ここで

$$\mathcal{U}'_n \triangleq \left\{\mathbf{x} \in \mathcal{X}^n \left| \frac{1}{n}d_n(\mathbf{x}, \psi_n(\varphi_n(\mathbf{x}))) \leq D - \gamma \right.\right\} \quad (48)$$

と定義すると d_n の期待値は

$$\mathbb{E}\left[\frac{1}{n}d_n(X^n, \psi_n(\varphi_n(X^n)))\right] \quad (49)$$

$$= \frac{1}{n} \sum_{\mathbf{x}} P_{X^n}(\mathbf{x}) d_n(\mathbf{x}, \psi_n(\varphi_n(\mathbf{x}))) \quad (50)$$

$$= \frac{1}{n} \sum_{\mathbf{x} \in \mathcal{U}'_n} P_{X^n}(\mathbf{x}) d_n(\mathbf{x}, \psi_n(\varphi_n(\mathbf{x}))) + \frac{1}{n} \sum_{\mathbf{x} \notin \mathcal{U}'_n} P_{X^n}(\mathbf{x}) d_n(\mathbf{x}, \psi_n(\varphi_n(\mathbf{x}))) \quad (51)$$

となる．

式 (51) の第 1 項は

$$\frac{1}{n} \sum_{\mathbf{x} \in \mathcal{U}'_n} P_{X^n}(\mathbf{x}) d_n(\mathbf{x}, \psi_n(\varphi_n(\mathbf{x}))) \quad (52)$$

$$\leq \sum_{\mathbf{x} \in \mathcal{U}'_n} P_{X^n}(\mathbf{x}) (D - \gamma) \quad (53)$$

$$\leq D - \gamma \quad (54)$$

となる．次に第 2 項を評価する．情報源アルファベットが有限なので $d(x, y)$ の最大値 $d_{\max}$

が存在する．すると $d(x, y) \leq d_{\max}$ より， $\frac{1}{n}d_n(\mathbf{x}, \mathbf{y}) \leq d_{\max}$ となるので第 2 項は

$$\frac{1}{n} \sum_{\mathbf{x} \notin \mathcal{U}'_n} P_{X^n}(\mathbf{x}) d_n(\mathbf{x}, \psi_n(\varphi_n(\mathbf{x}))) \quad (55)$$

$$\leq \sum_{\mathbf{x} \notin \mathcal{U}'_n} P_{X^n}(\mathbf{x}) d_{\max} \quad (56)$$

$$= d_{\max} \Pr \left\{ \frac{1}{n} d_n(X^n, \psi_n(\varphi_n(X^n))) > D - \gamma \right\} \quad (57)$$

$$\leq d_{\max} \Pr \left\{ \frac{1}{n} d_n(X^n, \psi_n(\varphi_n(X^n))) \geq \mathbb{E}[d(X, Y)] + \gamma \right\} \quad (58)$$

$$\stackrel{(47)}{\rightarrow} 0 \quad (n \rightarrow \infty) \quad (59)$$

となる．以上より n が十分大きいとき

$$\mathbb{E} \left[\frac{1}{n} d_n(X^n, \psi_n(\varphi_n(X^n))) \right] \leq D \quad (60)$$

を得る．よって式 (60), 式 (31) からレート・歪み領域の点 (R, D) は CK 平均規範を達成可能である．よってレート・歪み関数は

$$R_{\text{CK}}(D) \leq I(X; Y) + 2\gamma \quad (61)$$

を満たす．式 (15) より式 (61) は $\mathbb{E}[d(X, Y)] \leq D - 2\gamma$ を満たす任意の Y に対して成り立つので

$$R_{\text{CK}}(D) \leq \inf_{Y: \mathbb{E}[d(X, Y)] \leq D - 2\gamma} I(X; Y) + 2\gamma \quad (62)$$

である． $\gamma > 0$ は任意で右辺は γ に関して連続なので， $\gamma \rightarrow 0$ とすると付録 C.5 より

$$R_{\text{CK}}(D) \leq \inf_{Y: \mathbb{E}[d(X, Y)] \leq D} I(X; Y) \quad (63)$$

である． $I(X; Y)$ は Y に関して連続で Y の範囲は閉じているので

$$R_{\text{CK}}(D) \leq \min_{Y: \mathbb{E}[d(X, Y)] \leq D} I(X; Y) \quad (64)$$

を得る．□

4 CK 逆定理

定理 3 の証明を行う．

〔証明〕 一般情報源 $\mathbf{X}$ に対して (R, D) が CK 平均規範達成可能である仮定すると, 任意の $\delta > 0$ に対して十分大きな全ての n で

$$\mathbb{E} \left[\frac{1}{n} d_n(X^n, Y^n) \right] \leq D \quad (65)$$

$$\frac{1}{n} \log M_n < R + \delta \quad (66)$$

を満たす符号の列 $\{(\varphi_n, \psi_n)\}_{n=1}^{\infty}$ が存在する. このとき式 (65) から, 任意の $\delta_1 > 0$ に対して

$$\mathbb{E} \left[\frac{1}{n} d_n(X^n, Y^n) \right] < D + \delta_1 \quad (67)$$

が自明に成り立つ. よって, 式 (67), 式 (66) より

$$\limsup_{n \rightarrow \infty} \mathbb{E} \left[\frac{1}{n} d_n(X^n, Y^n) \right] \leq D \quad (68)$$

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log M_n \leq R \quad (69)$$

がいえる. 即ち (R, D) が CK 平均規範達成可能ならば平均規範達成可能でもある. 従って, レート歪み関数の定義から

$$R_{\text{CK}}(D) \geq R(D) \quad (70)$$

を得る. $\square$

5 系

定理 3, 定理 1 より特に情報源が定常無記憶のとき, 直ちに

系 1

$$R_{\text{CK}}(D) \geq R(D) = \min_{Y: \mathbb{E}[d(X, Y)] \leq D} I(X; Y) \quad (71)$$

を得る.

さらに系 1, 定理 2 の順定理より情報源が定常無記憶のとき, 直ちに

系 2

$$R_{\text{CK}}(D) = \min_{Y: \mathbb{E}[d(X, Y)] \leq D} I(X; Y) \quad (72)$$

を得る.

6 まとめ

本研究ではレート・歪み問題の平均規範と CK 平均規範について、その関係とレート歪み関数について考察した。結果として CK 平均規範に着目することで CK 順定理, CK 逆定理を示し, そこから系 1, 系 2 を得ることができた。特に CK 逆定理に関しては一般情報源に拡張して証明することができた。本研究では情報源が定常無記憶情報源の場合の CK 平均規範のレート・歪み関数の数学的表現を求めることができたが, 一般情報源におけるレート歪み関数の数学的表現を求めるには至らなかった。また, 今回考察した 2 つ以外の規範についても検討の余地があると考ええる。

謝辞

本研究を行うにあたり, 丁寧なご指導を賜りました指導教員の西新幹彦准教授に感謝の意を表する。

参考文献

- [1] I. Csiszár and J. Körner, Information Theory: Coding Theorems For Discrete Memoryless Systems, Academic Press, New York, 1981.
- [2] 澤田真子, 「観測値に雑音のあるシステムのレート・歪み関数の数値計算と基礎的考察」, 信州大学工学部学士論文 (指導教員: 西新幹彦), 2020 年 2 月.
- [3] 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.
- [4] M.Nishiara, Y.Ito, “Proof of achievability part of rate-distortion theorem without random coding,” IEICE Trans. Fundamentals, accepted.

付録 A 補題 1

ここでは、定理 1 を証明するのに利用する補題を証明する。

補題 1([4]) 定常無記憶情報源 X^n と、それと相関のある Y^n を考える。 R, γ を任意に与えられた正の定数とする。 さらに

$$\mathcal{B}_n(\mathbf{y}) \triangleq \left\{ \mathbf{x} \in \mathcal{X}^n \left| \frac{1}{n} \log \frac{P_{X^n|Y^n}(\mathbf{x} | \mathbf{y})}{P_{X^n}(\mathbf{x})} \leq R - \gamma \right. \right\} \quad (73)$$

とおき、 $\mathcal{S}_n(\mathbf{y})$ を任意の $\mathbf{y} \in \mathcal{Y}^n$ を引数とする $\mathcal{X}^n$ の部分集合とすると

$$\frac{1}{n} \log M_n \leq R \quad (74)$$

$$\Pr\{X^n \notin \mathcal{S}_n(\psi_n(\varphi_n(X^n)))\} \leq \Pr\{X^n \notin \mathcal{B}_n(Y^n)\} + \Pr\{X^n \notin \mathcal{S}_n(Y^n)\} + e^{-n\gamma} \quad (75)$$

を満たす符号 (φ_n, ψ_n) が存在する。ただし M_n は符号語数である。

〔証明〕 $\mathcal{T}_n(\mathbf{y}) \triangleq \mathcal{B}_n(\mathbf{y}) \cap \mathcal{S}_n(\mathbf{y})$ とおく。復号語を以下のように順次選んでいく。最初に条件

$$P_{X^n|Y^n}(\mathcal{T}_n(\mathbf{y}) | \mathbf{y}) \geq e^{-n\gamma} \quad (76)$$

を満たす任意の $\mathbf{y} \in \mathcal{Y}$ を選びこれを最初の復号語 $\psi_n(1)$ とする。さらに $\varphi_n^{-1}(1) \triangleq \mathcal{T}_n(\psi_n(1))$ とおく。以降 $m \geq 2$ に対し m 番目の復号語として

$$P_{X^n|Y^n} \left(\mathcal{T}_n(\mathbf{y}) \setminus \bigcup_{m' < m} \varphi_n^{-1}(m') \middle| \mathbf{y} \right) \geq e^{-n\gamma} \quad (77)$$

を満たす $\mathbf{y} \in \mathcal{Y}^n$ を任意に選び、 $\psi_n(m)$ とする。さらに

$$\varphi_n^{-1}(m) \triangleq \mathcal{T}_n(\psi_n(m)) \setminus \bigcup_{m' < m} \varphi_n^{-1}(m') \quad (78)$$

とする。可能な限りこの操作を続け、復号語を M_n 個取れたとする。これで符号器 φ_n と復号器 ψ_n が完成した。ここで $\varphi_n^{-1}(m)$, $m = 1, \dots, M_n$ が互いに素であることに注意し

$$\mathcal{D} \triangleq \bigcup_m \varphi_n^{-1}(m) \quad (79)$$

とおく。すると M_n の定義より任意の $\mathbf{y} \in \mathcal{Y}^n$ に対して

$$P_{X^n|Y^n}(\mathcal{T}_n(\mathbf{y}) \setminus \mathcal{D} | \mathbf{y}) < e^{-n\gamma} \quad (80)$$

が成り立つことに注意しておく。

以上のように定義された (φ_n, ψ_n) 符号の性能を評価する。

まず符号化レートを調べる．そのために $\mathbf{x} \in \varphi_n^{-1}(m)$ となる $\mathbf{x}$ を考える．すると $\varphi_n^{-1}(m)$ の定義より $\mathbf{x} \in \mathcal{T}_n(\psi_n(m))$ となる．さらに $\mathcal{T}_n(y)$ の定義より $\mathbf{x} \in \mathcal{B}_n(\psi_n(m))$ となるため $\mathcal{B}_n(y)$ の定義より

$$\frac{1}{n} \log \frac{P_{X^n|Y^n}(\mathbf{x} \mid \psi_n(m))}{P_{X^n}(\mathbf{x})} \leq R - \gamma \quad (81)$$

が得られる．即ち $\mathbf{x} \in \varphi_n^{-1}(m)$ ならば

$$P_{X^n}(\mathbf{x}) \geq P_{X^n|Y^n}(\mathbf{x} \mid \psi_n(m))e^{-n(R-\gamma)} \quad (82)$$

が成り立つ．したがって

$$1 \geq P_{X^n}(\mathcal{D}) \quad (83)$$

$$\stackrel{(79)}{=} P_{X^n} \left(\bigcup_m \varphi_n^{-1}(m) \right) \quad (84)$$

$$\stackrel{(78)}{=} \sum_m P_{X^n}(\varphi_n^{-1}(m)) \quad (85)$$

$$= \sum_m \sum_{\mathbf{x} \in \varphi_n^{-1}(m)} P_{X^n}(\mathbf{x}) \quad (86)$$

$$\stackrel{(82)}{\geq} \sum_m \sum_{\mathbf{x} \in \varphi_n^{-1}(m)} P_{X^n|Y^n}(\mathbf{x} \mid \psi_n(m))e^{-n(R-\gamma)} \quad (87)$$

$$= \sum_m P_{X^n|Y^n}(\varphi_n^{-1}(m) \mid \psi_n(m))e^{-n(R-\gamma)} \quad (88)$$

$$\stackrel{(77)}{\geq} \sum_m e^{-n\gamma} \cdot e^{-n(R-\gamma)} \quad (89)$$

$$= \sum_m e^{-nR} \quad (90)$$

$$= M_n e^{-nR} \quad (91)$$

となり

$$\frac{1}{n} \log M_n \leq R \quad (92)$$

が得られる．

次に式 (75) を導く． $\mathbf{x} \in \mathcal{D}$ とすると，ある m に対して $\mathbf{x} \in \varphi_n^{-1}(m)$ となる．したがって $\mathbf{x} \in \mathcal{T}_n(\psi_n(m))$ となるので， $\mathcal{T}_n(\mathbf{y})$ の定義より $\mathbf{x} \in \mathcal{S}_n(\psi_n(m))$ となる．一方 $\varphi_n(\mathbf{x}) = m$ なので $\mathbf{x} \in \mathcal{S}_n(\psi_n(\varphi_n(\mathbf{x})))$ となる．したがって

$$\Pr\{X^n \in \mathcal{D}\} \leq \Pr\{X^n \in \mathcal{S}_n(\psi_n(\varphi_n(X^n)))\} \quad (93)$$

といえる．一方, 式 (80) により任意の $\mathbf{y}$ について

$$P_{X^n|Y^n}(\mathcal{T}_n(\mathbf{y}) \mid \mathbf{y}) = P_{X^n|Y^n}(\mathcal{T}_n(\mathbf{y}) \cap \mathcal{D} \mid \mathbf{y}) + P_{X^n|Y^n}(\mathcal{T}_n(\mathbf{y}) \setminus \mathcal{D} \mid \mathbf{y}) \quad (94)$$

$$\leq P_{X^n|Y^n}(\mathcal{D} \mid \mathbf{y}) + e^{-n\gamma} \quad (95)$$

となるので式 (93) の左辺は

$$\Pr\{X^n \in \mathcal{D}\} = P_{X^n}(\mathcal{D}) \quad (96)$$

$$= \sum_{\mathbf{y}} P_{X^n Y^n}(\mathcal{D}, \mathbf{y}) \quad (97)$$

$$= \sum_{\mathbf{y}} P_{Y^n}(\mathbf{y}) P_{X^n|Y^n}(\mathcal{D} \mid \mathbf{y}) \quad (98)$$

$$\stackrel{(95)}{\geq} \sum_{\mathbf{y}} P_{Y^n}(\mathbf{y}) (P_{X^n|Y^n}(\mathcal{T}_n(\mathbf{y}) \mid \mathbf{y}) - e^{-n\gamma}) \quad (99)$$

$$= \sum_{\mathbf{y}} P_{Y^n}(\mathbf{y}) P_{X^n|Y^n}(\mathcal{T}_n(\mathbf{y}) \mid \mathbf{y}) - \sum_{\mathbf{y}} P_{Y^n}(\mathbf{y}) e^{-n\gamma} \quad (100)$$

$$= \sum_{\mathbf{y}} \Pr\{X^n \in \mathcal{T}_n(Y^n), Y^n = \mathbf{y}\} - e^{-n\gamma} \quad (101)$$

$$= \Pr\{X^n \in \mathcal{T}_n(Y^n)\} - e^{-n\gamma} \quad (102)$$

$$= \Pr\{X^n \in \mathcal{B}_n(Y^n) \text{ かつ } X^n \in \mathcal{S}_n(Y^n)\} - e^{-n\gamma} \quad (103)$$

$$= 1 - \Pr\{X^n \notin \mathcal{B}_n(Y^n) \text{ または } X^n \notin \mathcal{S}_n(Y^n)\} - e^{-n\gamma} \quad (104)$$

$$\geq 1 - \Pr\{X^n \notin \mathcal{B}_n(Y^n)\} - \Pr\{X^n \notin \mathcal{S}_n(Y^n)\} - e^{-n\gamma} \quad (105)$$

となる．よって式 (93), (105) より

$$\Pr\{X^n \in \mathcal{S}_n(\psi_n(\varphi_n(X^n)))\} \geq 1 - \Pr\{X^n \notin \mathcal{B}_n(Y^n)\} - \Pr\{X^n \notin \mathcal{S}_n(Y^n)\} - e^{-n\gamma} \quad (106)$$

となるので

$$\Pr\{X^n \notin \mathcal{S}_n(\psi_n(\varphi_n(X^n)))\} \leq \Pr\{X^n \notin \mathcal{B}_n(Y^n)\} + \Pr\{X^n \notin \mathcal{S}_n(Y^n)\} + e^{-n\gamma} \quad (107)$$

が導かれる．□

付録 B 定理 1 の証明

ここでは定理 1 を証明する．

定理 1 (再掲) 平均規範のレート・歪み関数は

$$R(D) = \min_{Y: \mathbb{E}[d(X, Y)] \leq D} I(X; Y) \quad (108)$$

を満たす。

定理 1 は順定理と逆定理を示すことで得られる。以下の節では順定理と逆定理をそれぞれ証明する。

B.1 順定理の証明

平均規範のレート・歪み関数は

$$R(D) \leq \min_{Y: \mathbb{E}[d(X, Y)] \leq D} I(X; Y) \quad (109)$$

を満たす。

〔証明〕 D を固定し、任意の $\gamma > 0$ を考える。ここで

$$\mathbb{E}[d(X, Y)] + \gamma \leq D \quad (110)$$

を満たす任意の Y をとり、その上で

$$I(X; Y) + 2\gamma = R \quad (111)$$

とおく。 $X^n Y^n \triangleq (X_1 Y_1, X_2 Y_2, \dots, X_n Y_n)$ を XY の独立同一分布な系列として $X^n \triangleq (X_1, X_2, \dots, X_n)$, $Y^n \triangleq (Y_1, Y_2, \dots, Y_n)$ とおく。そこで、

$$\mathcal{T}_n^{(1)} \triangleq \left\{ (\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n \left| \left| \frac{1}{n} \log \frac{P_{Y^n|X^n}(\mathbf{y} | \mathbf{x})}{P_{Y^n}(\mathbf{y})} - I(X; Y) \right| < \gamma \right. \right\} \quad (112)$$

$$\mathcal{T}_n^{(2)} \triangleq \left\{ (\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n \left| \left| \frac{1}{n} d_n(\mathbf{x}, \mathbf{y}) - \mathbb{E}[d(X, Y)] \right| < \gamma \right. \right\} \quad (113)$$

として $\mathcal{T}_n \triangleq \mathcal{T}_n^{(1)} \cap \mathcal{T}_n^{(2)}$ とおく。 $X^n Y^n$ が独立同一分布なので、

$$\frac{1}{n} \log \frac{P_{Y^n|X^n}(Y^n | X^n)}{P_{Y^n}(Y^n)} = \frac{1}{n} \log \frac{\prod_{i=1}^n P_{Y|X}(Y_i | X_i)}{\prod_{i=1}^n P_Y(Y_i)} \quad (114)$$

$$= \frac{1}{n} \sum_{i=1}^n \log \frac{P_{Y|X}(Y_i | X_i)}{P_Y(Y_i)} \quad (115)$$

となる。式 (115) の右辺の和の中の各項は独立でその期待値は

$$\mathbb{E} \left[\log \frac{P_{Y|X}(Y_i | X_i)}{P_Y(Y_i)} \right] = \sum_{x, y} P_{XY}(x, y) \log \frac{P_{Y|X}(y | x)}{P_Y(y)} \quad (116)$$

$$= I(X; Y) \quad (117)$$

であり、分散は有界である^{*1}。 $\mathcal{T}_n^{(1)}$ の定義に注意して大数の法則^{*2}を適用すれば

$$\Pr\{X^n Y^n \in \mathcal{T}_n^{(1)}\} \rightarrow 1 \quad (n \rightarrow \infty) \quad (118)$$

^{*1} 証明は参考文献 [2]

^{*2} 証明は付録 C に記載

を得る．同様にして $\mathbb{E}[d(X, Y)]$ が有界ならば

$$\Pr\{X^n Y^n \in T_n^{(2)}\} \rightarrow 1 \quad (n \rightarrow \infty) \quad (119)$$

を得る．したがって,

$$\Pr\{X^n Y^n \in T_n\} = 1 - \Pr\{X^n Y^n \notin T_n\} \quad (120)$$

$$\geq 1 - \Pr\{X^n Y^n \notin T_n^{(1)}\} - \Pr\{X^n Y^n \notin T_n^{(2)}\} \quad (121)$$

$$= 1 - (1 - \Pr\{X^n Y^n \in T_n^{(1)}\}) - (1 - \Pr\{X^n Y^n \in T_n^{(2)}\}) \quad (122)$$

$$\rightarrow 1 \quad (n \rightarrow \infty) \quad (123)$$

が成立する．ここで

$$\mathcal{B}_n(\mathbf{y}) \triangleq \left\{ \mathbf{x} \in \mathcal{X}^n \mid \frac{1}{n} \log \frac{P_{X^n|Y^n}(\mathbf{x} | \mathbf{y})}{P_{X^n}(\mathbf{x})} \leq R - \gamma \right\} \quad (124)$$

$$\mathcal{S}_n(\mathbf{y}) \triangleq \{\mathbf{x} \in \mathcal{X}^n \mid (\mathbf{x}, \mathbf{y}) \in \mathcal{T}_n^{(2)}\} \quad (125)$$

とおく．ただし式 (124) と式 (73) は同じである．すると補題 1 より

$$\frac{1}{n} \log M_n \leq R \quad (126)$$

$$\Pr\{X^n \notin \mathcal{S}_n(\psi_n(\varphi_n(X^n)))\} \leq \Pr\{X^n \notin \mathcal{B}_n(Y^n)\} + \Pr\{X^n \notin \mathcal{S}_n(Y^n)\} + e^{-n\gamma} \quad (127)$$

を満たす符号 (φ_n, ψ_n) が存在する．式 (127) の右辺の第 1 項に関して, 式 (124), (112) より,

$$\Pr\{X^n \in \mathcal{B}_n(Y^n)\} = \Pr\left\{ \frac{1}{n} \log \frac{P_{X^n|Y^n}(X^n | Y^n)}{P_{X^n}(X^n)} \leq I(X; Y) + \gamma \right\} \quad (128)$$

$$= \Pr\left\{ \frac{1}{n} \log \frac{P_{Y^n|X^n}(Y^n | X^n)}{P_{Y^n}(Y^n)} \leq I(X; Y) + \gamma \right\} \quad (129)$$

$$= \Pr\left\{ \frac{1}{n} \log \frac{P_{Y^n|X^n}(Y^n | X^n)}{P_{Y^n}(Y^n)} - I(X; Y) \leq \gamma \right\} \quad (130)$$

$$\geq \Pr\left\{ \left| \frac{1}{n} \log \frac{P_{Y^n|X^n}(Y^n | X^n)}{P_{Y^n}(Y^n)} - I(X; Y) \right| < \gamma \right\} \quad (131)$$

$$= \Pr\{X^n Y^n \in \mathcal{T}_n^{(1)}\} \quad (132)$$

$$\stackrel{(118)}{\rightarrow} 1 \quad (n \rightarrow \infty) \quad (133)$$

となる．よって

$$\Pr\{X^n \notin \mathcal{B}_n(Y^n)\} \rightarrow 0 \quad (n \rightarrow \infty) \quad (134)$$

を得る．また, 式 (127) の右辺の第 2 項に関して式 (125), (119) より

$$\Pr\{X^n \in \mathcal{S}_n(Y^n)\} = \Pr\{X^n Y^n \in \mathcal{T}_n^{(2)}\} \quad (135)$$

$$\rightarrow 1 \quad (n \rightarrow \infty) \quad (136)$$

となる． よって

$$\Pr\{X^n \notin \mathcal{S}_n(Y^n)\} \rightarrow 0 \quad (n \rightarrow \infty) \quad (137)$$

を得る． 以上より,

$$\Pr\{X^n \notin \mathcal{S}_n(\psi_n(\varphi_n(X^n)))\} \rightarrow 0 \quad (n \rightarrow \infty) \quad (138)$$

となる． 一方で式 (127) の左辺は式 (125), (113) より

$$\Pr\{X^n \notin \mathcal{S}_n(\psi_n(\varphi_n(X^n)))\} = \Pr\{(X^n, \psi_n(\varphi_n(X^n))) \notin \mathcal{T}_n^{(2)}\} \quad (139)$$

$$= \Pr\left\{\left|\frac{1}{n}d_n(X^n, \psi_n(\varphi_n(X^n))) - \mathbb{E}[d(X, Y)]\right| \geq \gamma\right\} \quad (140)$$

$$\geq \Pr\left\{\frac{1}{n}d_n(X^n, \psi_n(\varphi_n(X^n))) \geq \mathbb{E}[d(X, Y)] + \gamma\right\} \quad (141)$$

したがって式 (138) より

$$\Pr\left\{\frac{1}{n}d_n(X^n, \psi_n(\varphi_n(X^n))) \geq \mathbb{E}[d(X, Y)] + \gamma\right\} \rightarrow 0 \quad (142)$$

となる． ここで

$$\mathcal{U}_n \triangleq \left\{\mathbf{x} \in \mathcal{X}^n \mid \frac{1}{n}d_n(\mathbf{x}, \psi_n(\varphi_n(\mathbf{x}))) \leq D\right\} \quad (143)$$

と定義すると d_n の期待値は

$$\mathbb{E}\left[\frac{1}{n}d_n(X^n, \psi_n(\varphi_n(X^n)))\right] \quad (144)$$

$$= \frac{1}{n} \sum_{\mathbf{x}} P_{X^n}(\mathbf{x}) d_n(\mathbf{x}, \psi_n(\varphi_n(\mathbf{x}))) \quad (145)$$

$$= \frac{1}{n} \sum_{\mathbf{x} \in \mathcal{U}_n} P_{X^n}(\mathbf{x}) d_n(\mathbf{x}, \psi_n(\varphi_n(\mathbf{x}))) + \frac{1}{n} \sum_{\mathbf{x} \notin \mathcal{U}_n} P_{X^n}(\mathbf{x}) d_n(\mathbf{x}, \psi_n(\varphi_n(\mathbf{x}))) \quad (146)$$

となる． 式 (146) の第 1 項は

$$\frac{1}{n} \sum_{\mathbf{x} \in \mathcal{U}_n} P_{X^n}(\mathbf{x}) d_n(\mathbf{x}, \psi_n(\varphi_n(\mathbf{x}))) \quad (147)$$

$$\leq \sum_{\mathbf{x} \in \mathcal{U}_n} P_{X^n}(\mathbf{x}) D \quad (148)$$

$$\leq D \quad (149)$$

となる．次に第 2 項を評価する．情報源アルファベットが有限なので $d(x, y)$ の最大値 $d_{\max}$ が存在する．すると $d(x, y) \leq d_{\max}$ より， $\frac{1}{n}d_n(\mathbf{x}, \mathbf{y}) \leq d_{\max}$ となるので，第 2 項は

$$\frac{1}{n} \sum_{\mathbf{x} \notin \mathcal{U}_n} P_{X^n}(\mathbf{x}) d_n(\mathbf{x}, \psi_n(\varphi_n(\mathbf{x}))) \quad (150)$$

$$\leq \sum_{\mathbf{x} \notin \mathcal{U}_n} P_{X^n}(\mathbf{x}) d_{\max} \quad (151)$$

$$= d_{\max} \Pr \left\{ \frac{1}{n} d_n(X^n, \psi_n(\varphi_n(X^n))) > D \right\} \quad (152)$$

$$\leq d_{\max} \Pr \left\{ \frac{1}{n} d_n(X^n, \psi_n(\varphi_n(X^n))) \geq \mathbb{E}[d(X, Y)] + \gamma \right\} \quad (153)$$

$$\stackrel{(142)}{\rightarrow} 0 \quad (n \rightarrow \infty) \quad (154)$$

となる．以上より

$$\limsup_{n \rightarrow \infty} \mathbb{E} \left[\frac{1}{n} d_n(X^n, \psi_n(\varphi_n(X^n))) \right] \leq D \quad (155)$$

を得る．したがって式 (126), (155) より (R, D) は達成可能である．よってレート・歪み関数 $R(D)$ は式 (111) より

$$R(D) \leq I(X; Y) + 2\gamma \quad (156)$$

を満たす．さらに式 (110) より条件 $\mathbb{E}[d(X, Y)] \leq D - \gamma$ を満たすような任意の Y を選んだので

$$R(D) \leq \inf_{Y: \mathbb{E}[d(X, Y)] \leq D - \gamma} I(X; Y) + 2\gamma \quad (157)$$

となる．さらに $\gamma > 0$ は任意であり，式 (157) の右辺は γ に関して連続である*³ため $\gamma \rightarrow 0$ とすれば

$$R(D) \leq \inf_{Y: \mathbb{E}[d(X, Y)] \leq D} I(X; Y) \quad (158)$$

となる． $I(X; Y)$ は Y に関して連続であり*⁴，下限をとる Y の範囲は閉じているので

$$R(D) \leq \min_{Y: \mathbb{E}[d(X, Y)] \leq D} I(X; Y) \quad (159)$$

が導け，順定理が示された．□

*³ 証明は付録 C.5 に記載

*⁴ 証明は付録 C.4 に記載

B.2 逆定理の証明

平均規範のレート歪み関数は

$$R(D) \leq \min_{Y: \mathbb{E}[d(X,Y)] \geq D} I(X;Y) \quad (160)$$

を満たす.

〔証明〕 定常無記憶情報源 $\mathbf{X}$ に対して D を固定し, レート・歪み領域の点 (R, D) が達成可能であるとする

$$\limsup_{n \rightarrow \infty} \mathbb{E} \left[\frac{1}{n} d_n(X^n, \psi_n(\varphi_n(X^n))) \right] \leq D \quad (161)$$

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log M_n \leq R \quad (162)$$

となる符号の列 $\{\varphi_n, \psi_n\}_{n=1}^{\infty}$ が存在する. ここで $Y^n \triangleq \psi_n(\varphi_n(X^n)) = (Y_1^{(n)}, Y_2^{(n)}, \dots, Y_n^{(n)})$ とおくと $Y^{(n)}$ の取りうる値は M_n 個なので

$$\log M_n \geq H(Y^n) \quad (163)$$

$$\geq H(Y^n) - H(Y^n|X^n) \quad (164)$$

$$= I(X^n; Y^n) \quad (165)$$

$$= H(X^n) - H(X^n|Y^n) \quad (166)$$

$$= \sum_{i=1}^n H(X_i) - \sum_{i=1}^n H(X_i|X^{i-1}Y^n) \quad (167)$$

$$\geq \sum_{i=1}^n H(X_i) - \sum_{i=1}^n H(X_i|Y_i^{(n)}) \quad (168)$$

$$= \sum_{i=1}^n I(X_i; Y_i^{(n)}) \quad (169)$$

となる. さらに確率変数 Q_n を導入し, この分布を

$$P_{Q_n}(i) = \frac{1}{n}, i = 1, 2, \dots, n \quad (170)$$

とする． $XY^{(n)} \triangleq X_{Q_n}Y_{Q_n}^{(n)}$ とおくと，

$$\frac{1}{n} \sum_{i=1}^n (X_i; Y_i^{(n)}) = \sum_{i=1}^n P_{Q_n}(i) I(X_{Q_n}; Y_{Q_n}^{(n)} | Q = i) \quad (171)$$

$$= I(X_{Q_n}; Y_{Q_n}^{(n)} | Q_n) \quad (172)$$

$$= H(X | Q_n) - H(X | Y^{(n)} Q_n) \quad (173)$$

$$= H(X) - H(X | Y^{(n)} Q_n) \quad (174)$$

$$\geq H(X) - H(X | Y^{(n)}) \quad (175)$$

$$= I(X; Y^{(n)}) \quad (176)$$

よって

$$\frac{1}{n} \log M_n \geq I(X; Y^{(n)}) \quad (177)$$

を得る．一方で

$$\mathbb{E} \left[\frac{1}{n} d_n(X^n, \psi_n(\varphi_n(X^n))) \right] \quad (178)$$

$$= \mathbb{E} \left[\frac{1}{n} d_n(X^n, Y^n) \right] \quad (179)$$

$$= \frac{1}{n} \sum_{i=1}^n \mathbb{E}[d(X_i, Y_i^n)] \quad (180)$$

$$= \sum_{i=1}^n P_{Q_n}(i) \mathbb{E}[d(X_{Q_n}, Y_i^{(n)}) | Q_n = i] \quad (181)$$

$$= \mathbb{E}[d(X_{Q_n}, Y_{Q_n}^{(n)})] \quad (182)$$

$$= \mathbb{E}[d(X, Y^{(n)})] \quad (183)$$

$$(184)$$

したがって

$$\limsup_{n \rightarrow \infty} \mathbb{E}[d(X, Y^n)] \leq D \quad (185)$$

を得る． よって任意の $\gamma > 0$ に対して十分大きな全ての n で

$$\mathbb{E}[d(X, Y^{(n)})] \leq D + \gamma \quad (186)$$

$$(187)$$

となる． 確率変数 $Y^{(n)}$ はこれを満たす． n が十分大きい時

$$\frac{1}{n} \log M_n \geq I(X; Y^{(n)}) \quad (188)$$

$$\geq \min_{Y: \mathbb{E}[d(X, Y)] \leq D + \gamma} I(X; Y) \quad (189)$$

である． $\limsup_{n \rightarrow \infty} \frac{1}{n} \log M_n \leq R$ より

$$R \geq \min_{Y: \mathbb{E}[d(X, Y)] \leq D + \gamma} I(X; Y) \quad (190)$$

は γ に関して連続で， $\gamma > 0$ は任意だったので $\gamma \rightarrow 0$ とすると

$$R \geq \min_{Y: \mathbb{E}[d(X, Y)] \leq D} I(X; Y) \quad (191)$$

を得る． $R(D) = \inf\{R | (R, D) \text{ は達成可能} \}$ より

$$R(D) \geq \min_{Y: \mathbb{E}[d(X, Y)] \leq D} I(X; Y) \quad (192)$$

が導け，逆定理が示された． $\square$

付録 C 定理 1 の証明の補足

大数の法則の導出のため，Markov の不等式と Chebychev の不等式の証明を行う．

C.1 Markov の不等式

非負値をとる確率変数 Z を考える．このとき任意の $a > 0$ に対して

$$\Pr\{Z > a\} \leq \frac{\mathbb{E}[Z]}{a} \quad (193)$$

が成り立つ．

〔証明〕 Z の期待値は

$$\mathbb{E}[Z] = \sum_z z P_Z(z) \quad (194)$$

$$= \sum_{0 \leq z \leq a} z P_Z(z) + \sum_{z > a} z P_Z(z) \quad (195)$$

$$\geq \sum_{z > a} z P_Z(z) \quad (196)$$

$$\geq \sum_{z > a} a P_Z(z) \quad (197)$$

$$= a \sum_{z > a} P_Z(z) \quad (198)$$

$$= a \Pr\{Z > a\} \quad (199)$$

と評価される．よって

$$\Pr\{Z > a\} \leq \frac{\mathbb{E}[Z]}{a} \quad (200)$$

を得る．□

C.2 Chebychev の不等式

実数値をとる確率変数 Z を考える．その平均を μ , 分散を σ^2 とするとき, 任意の $a > 0$ に対して

$$\Pr\{|Z - \mu| > a\} \leq \frac{\sigma^2}{a^2} \quad (201)$$

が成り立つ．

〔証明〕 Z を $(Z - \mu)^2$ に, a を a^2 に置き換えると, Markov の不等式より

$$\Pr\{(Z - \mu)^2 > a^2\} \leq \frac{\mathbb{E}[(Z - \mu)^2]}{a^2} \quad (202)$$

となるが, $\mathbb{E}[(Z - \mu)^2] = \sigma^2$ であるから

$$\Pr\{|Z - \mu| > a\} \leq \frac{\sigma^2}{a^2} \quad (203)$$

が導かれる．□

C.3 大数の法則

独立同一分布な実数値確率変数の列 $\{Z_i\}_{i=1}^{\infty}$ を考える． $\mathbb{E}[Z] = \mathbb{E}[Z_i]$ とおく． Z_i の分散が有界ならば任意の $\gamma > 0$ に対して

$$\lim_{n \rightarrow \infty} \Pr \left\{ \left| \frac{1}{n} \sum_{i=1}^n Z_i - \mathbb{E}[Z] \right| < \gamma \right\} = 1 \quad (204)$$

が成り立つ．

〔証明〕 まず,

$$\mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n Z_i \right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[Z_i] = \mathbb{E}[Z] \quad (205)$$

に注意する．さらに $\mathbb{V}[Z] = \mathbb{V}[Z_i]$ とおくと,

$$\mathbb{V} \left[\frac{1}{n} \sum_{i=1}^n Z_i \right] = \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}[Z_i] = \frac{1}{n} \mathbb{V}[Z] \quad (206)$$

となる．ここで，Chebychev の不等式より

$$\Pr \left\{ \left| \frac{1}{n} \sum_{i=1}^n Z_i - \mathbb{E}[Z] \right| \geq \gamma \right\} < \frac{\mathbb{V}[Z]}{n\gamma^2} \quad (207)$$

となる．したがって

$$\lim_{n \rightarrow \infty} \Pr \left\{ \left| \frac{1}{n} \sum_{i=1}^n Z_i - \mathbb{E}[Z] \right| \geq \gamma \right\} = 0 \quad (208)$$

が成り立つ．□

C.4 相互情報量の連続性の証明

定義 5 一般に関数 $f(\mathbf{x})$ が x に関して下に凸とは， $0 \leq \forall \lambda \leq 1$ と $\forall \mathbf{x}_1, \forall \mathbf{x}_2$ に対して

$$\lambda f(\mathbf{x}_1) + (1 - \lambda)f(\mathbf{x}_2) \geq f(\lambda \mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2) \quad (209)$$

となることである．

定理 4 式 (6) で定義される $I(X; Y)$ は P_X を固定したとき $P_{Y|X}$ に関して下に凸な関数である．

〔証明〕それぞれ X と相関のある Y_1, Y_2 を考える．さらに X と独立な確率変数 Q を導入し，その分布を

$$P_Q(1) = \lambda, \quad P_Q(2) = 1 - \lambda \quad (210)$$

とする．その上で $Y = Y_Q$ とすれば

$$P_{Y|X}(y|x) = P_{Y_Q|X}(y|x) \quad (211)$$

$$= \sum_i P_{Y_Q Q|X}(y, i|x) \quad (212)$$

$$= \sum_i P_Q(i) P_{Y_Q|XQ}(y|x, i) \quad (213)$$

$$= \sum_i P_Q(i) P_{Y_i|X}(y|x) \quad (214)$$

即ち

$$P_{Y|X} = \lambda P_{Y_1|X} + (1 - \lambda) P_{Y_2|X} \quad (215)$$

となる．このとき

$$I(X; Y|Q) = P_Q(1)I(X; Y|Q = 1) + P_Q(2)I(X; Y|Q = 2) \quad (216)$$

$$= \lambda I(X; Y_1) + (1 - \lambda)I(X; Y_2) \quad (217)$$

となるが, 一方で

$$I(X; Y|Q) = H(X|Q) - H(X|YQ) \quad (218)$$

$$= H(X) - H(X|YQ) \quad (219)$$

$$\geq H(X) - H(X|Y) \quad (220)$$

$$= I(X; Y) \quad (221)$$

となるので

$$\lambda I(X; Y_1) + (1 - \lambda)I(X; Y_2) \geq I(X; Y) \quad (222)$$

を得る. ここで式 (215) に注意すれば $I(X; Y)$ は下に凸である. $\square$

C.5 γ に関して連続であることの証明

$\inf_{Y: \mathbb{E}[d(X, Y)] \leq D} I(X; Y)$ が D に関して連続であることを示す.

[証明] $\inf_{Y: \mathbb{E}[d(X, Y)] \leq D} I(X; Y)$ が D に関して下に凸であることを示せば十分である. すなわち

$$f(D) = \inf_{Y: \mathbb{E}[d(X, Y)] \leq D} I(X; Y) \quad (223)$$

として

$$\lambda f(D_1) + (1 - \lambda)f(D_2) \geq f(\lambda D_1 + (1 - \lambda)D_2) \quad (224)$$

となればよい. Y_1, Y_2 を

$$\mathbb{E}[d(X, Y_1)] \leq D_1 \quad (225)$$

$$\mathbb{E}[d(X, Y_2)] \leq D_2 \quad (226)$$

なる確率変数として $\lambda > 0$, Q を用いて

$$P_Q(1) = \lambda \quad (227)$$

$$P_Q(2) = 1 - \lambda \quad (228)$$

として Y を $Q = i$ のとき $Y = Y_i (i = 1, 2)$ と定めると

$$\mathbb{E}[d(X, Y)] = \lambda \mathbb{E}[d(X, Y_1)] + (1 - \lambda) \mathbb{E}[d(X, Y_2)] \leq \lambda D_1 + (1 - \lambda)D_2 \quad (229)$$

が成り立つ．ここで

$$\lambda f(D_1) + (1 - \lambda)f(D_2) = \inf_{Y_1: \mathbb{E}[d(X, Y_1)] \leq D_1, Y_2: \mathbb{E}[d(X, Y_2)] \leq D_2} (\lambda I(X; Y_1) + (1 - \lambda)I(X; Y_2)) \quad (230)$$

$$\geq \inf_{Y_1: \mathbb{E}[d(X, Y_1)] \leq D_1, Y_2: \mathbb{E}[d(X, Y_2)] \leq D_2, Y: \mathbb{E}[d(X, Y)] \leq \lambda D_1 + (1 - \lambda)D_2} I(X; Y) \quad (231)$$

$$\geq \inf_{Y: \mathbb{E}[d(X, Y)] \leq \lambda D_1 + (1 - \lambda)D_2} I(X; Y) \quad (232)$$

$$= f(\lambda D_1 + (1 - \lambda)D_2) \quad (233)$$

となる．よって $\inf_{Y: \mathbb{E}[d(X, Y)] \leq D} I(X; Y)$ は D に関して連続である．□