

【講義の要点】

まえがき (introduction, p. 567-568)

情報理論 (information theory): 通信, 情報科学, 確率論などと深く関係

情報源エントロピー (entropy of source): 情報源の確率的性質に由来

通信路容量 (capacity of channel): 通信路の雑音と関係

不確かさ, 情報, エントロピー (uncertainty, information and entropy, p. 568-573)

アルファベット (alphabet) と出現確率 (probability) (9.1)-(9.3)

離散無記憶情報源 (discrete memoryless source)

驚きの度合い, 不確かさの度合い ~ 情報量

各アルファベットがもつ情報量の定義 (9.4)

情報量の性質

- ・ 確率が1なら情報量は0 (9.5)
- ・ 各アルファベットが持つ情報量は非負 (9.6)
- ・ 確率が低いほど情報量は大きい (9.7)
- ・ 統計的に無相関なら2つのアルファベットの情報量は各アルファベットの情報量の和

対数の底は任意: 2 とすれば単位は bit (binary digit) (9.8)

二元等確率事象 (two possible and equiprobable event): 1 bit ~ 0/1 が等確率で出現

エントロピー (entropy): 情報量の期待値 (9.9)

~統計力学 (statistical mechanics) におけるエントロピー (熱力学的乱雑さ) と式の表現が同じ
エントロピーの性質

- ・ 不等式 (9.10)
- ・ どれかの確率が1ならエントロピーは0
- ・ 全アルファベットが等確率のときにエントロピーが最大
- ・ 証明に必要な不等式 (9.11)

例 9.1 二元無記憶情報源のエントロピー: 等確率時に最大値 1 bit

ブロック: n 個のアルファベットの並び ~ 拡張情報源 (extended source) から生成

⇒ エントロピーは n 倍

例 9.2 拡張情報源のエントロピー

情報源符号化定理 (source-coding theorem, p. 574-575)

情報源符号化: 効率のよい (efficient) データ表現

情報源の統計的性質を活用: 頻度の高い文字は短い符号 / 頻度の低い文字は長い符号

例 モールス符号: 頻度の高い E “.”, 頻度の低い Q “--.-“

効率のよい情報源符号化器

- ・ 符号語 (code word) は二値 (binary) 形式
- ・ 一意に復号可能 (uniquely decodable)

平均符号長 (average code-word length) (9.18)

符号化効率 (coding efficiency) (9.19), (9.21)

シャノンの情報源符号化定理 (Shannon's source coding theorem)

離散無記憶情報源のエントロピー $\leq$ 平均符号長 (9.20)

本日の課題 (p. 618-620)

9.2, 9.3, 9.8

- ・ 提出締切 10/22 (月) : 南 6 号館 1F メールボックス S6-4

演習 9.7

Consider a discrete memoryless source with source alphabet $S = \{s_0, s_1, s_2\}$ and source statistics $\{0.7, 0.15, 0.15\}$.

- (a) Calculate the entropy of the source.
- (b) Calculate the entropy of the second-order extension of the source.

参考 : $\log_2 0.7 \simeq -0.5, \log_2 0.15 \simeq -2.7$