

講義を始めるにあたって

- ## テキスト等の入手方法

- スライドpdfファイルのダウンロード先

<http://www.cs.tsukuba.ac.jp/~kudo/japanese.html>

1.1 情報理論とは

1948年のシャノンの論文

“A mathematical theory of communication”

- ## (1) 情報の定量化(第2章)

情報を数学的にどのように数えるか？

- 例 ① 今年水星が爆発する どちらが情報量が大きい？
 ② 今年水星が爆発しない

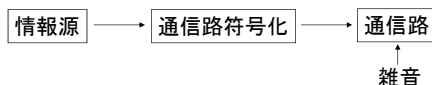
- (2) 情報源符号化の概念とその限界(第3章)



“情報を伝送する際にどこまで圧縮できるか”

① A → 00, B → 01, C → 10, D → 11
AAABCABAADを符号化するのに20個の0,1
② A → 0, B → 10, C → 110, D → 111
AAABCABAADを符号化するのに14個の0,1

- (3) 通信路符号化の概念とその限界(第4章)



“通信路の雑音や誤りに対処し高信頼化を図る”

例 A,B,C,Dの4文字を符号化

ABACD -> AABBAACDDとして伝送すると誤りが分かる

- (1) シヤノン理論 -> 情報の定量化, 符号化の限界
- (2) 符号理論 -> 符号化の具体的手法
- (3) 信号理論 -> 雑音中に埋もれた信号の推定, 信号の設計

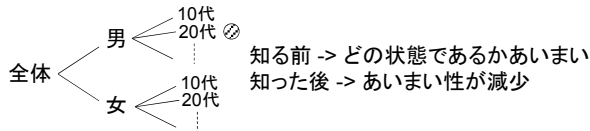
情報理論の応用分野

- (1) 情報源符号化
音声、画像のデータ圧縮
- (2) 通信路符号化
通信、計算機の記憶装置、オーディオ・ビデオ(CD・DVD)
- (3) 信号理論
音声・画像処理、生体信号や地震波の解析

1.2 情報とは何か

情報＝“知ることによりあいまい性(不明確さ)が減少するもの”
 情報量＝“あいまい性が減少する度合い”

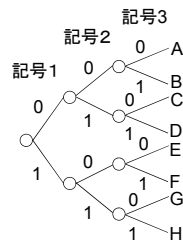
例1 “Aさんは男で20代である”



情報量 -> 情報によりどれぐらい少人数に特定できるか

例2 A,B,C,D,E,F,G,Hからなる文字の伝送

“A -> 000, B -> 001, C -> 010, D -> 011, E -> 100, F -> 101, G -> 110, H -> 111”として3つの1,0で伝送



記号が1つ送られる度に、あいまい性が 8 -> 4 -> 2 -> 1と減少

例3 天気予報の情報

仮定 ① 晴と雨が両方とも0.5の確率で起こる
 ② 天気予報の当たる確率0.8

天気予報を聞く前 晴 -> 0.5, 雨 -> 0.5
 晴の天気予報を聞いた後 晴 -> 0.8, 雨 -> 0.2
 完全に特定できないがあいまい性は減少

情報の表現方法

情報を記号の系列で抽象的に表現 -> 通報
 情報の持つ内容や意味は考えない(数学的な扱い困難)

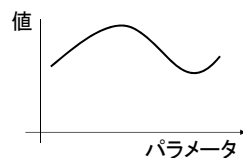
例 “本日は晴天なり”と“今日は良いお天気です”

意味を考えると同じ情報
 記号の列(通報)としては違う情報

意味を扱う情報理論 -> 現在のホットな話題

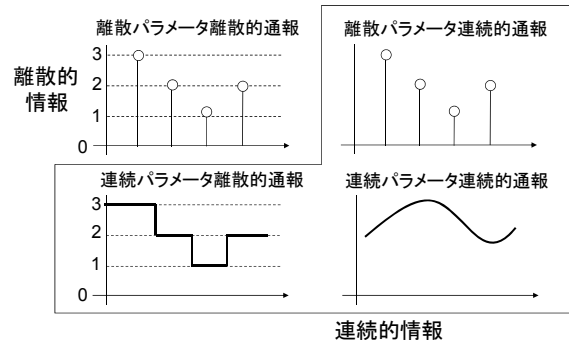
1.3 情報の種類

通報はあるパラメータの関数



例 音声 -> パラメータ＝時間, 値＝空気の圧力
 文章 -> パラメータ＝文字番号, 値＝文字

パラメータと値が離散(デジタル)か連続(アナログ)かで分類

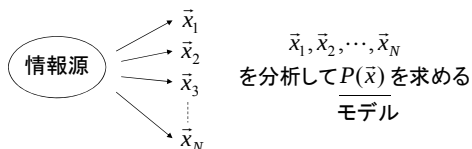


1.4 情報源の統計的モデル

情報理論の出発点＝情報源が確率モデルで表現できる

確率モデルとは？

ある情報源 X から出た多数のデータ $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N$



$\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N$ に完全な規則性があるわけではないが、統計的に共通の性質がある時に用いる

例 音声, 自然言語, 実験データの解析

(a) 情報源のモデル化

情報源 $\xrightarrow{a_3 a_1 a_3 \dots a_M a_2}$ 記号系列
 単位時間に1つの記号を発生する

a_i -> 情報源記号

$A = \{a_1, a_2, \dots, a_M\}$ -> 情報源アルファベット

例 英語

$A = \{A, B, \dots, Z, \text{space}, \text{punctuation}\}$

MY NAME IS HIROYUKI KUDO.
 (MY NAME IS) (HIROYUKI) (KUDO) (.)
 単語 文法
 文
 情報源記号 -> 単語 -> 文

(b) 自然言語の統計的性質

(1) 第0次近似(各記号独立, 等確率)

教科書 p.14 単語が長い, X,Z,Wが多い

(2) 第1次近似(各記号独立, 英語における文字の出現頻度)

教科書 p.14 文字と文字の関係が無視されている

(3) 第2次近似(2字組の出現頻度)

..... O ?
 $\left\{ \begin{array}{l} \text{N,R} \rightarrow \text{高確率} \\ \text{その他} \rightarrow \text{低確率} \end{array} \right.$

各文字の頻度は直前の文字の影響を受ける $\rightarrow$ マルコフ性
 教科書 p.15

(4) 第3次近似(3字組の出現頻度)

統計モデルで作った英語の文章(教科書p.14~15)

(1) 第0次近似(各記号独立, 等確率)

XFOML RXKHRJFFJUJ ZLPWCFWKCYJ
 FFJEYVKCQSGHYD QPAAMKBZAACIBZLHJQD

(2) 第1次近似(各記号独立, 英語における文字の出現頻度)

OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA
 TH EEI ALHENHTTPA OOBTTVA NAH BRL

(3) 第2次近似(2字組の出現頻度)

ON IE ANTSOUTINYS ARE T INCTORE ST S
 DEAMY ACHIN D ILONASIVE TUCOOWE AT
 TEASONARE FUSO TIZIN ANDY TOBE SEACE
 CTISBE

(4) 第3次近似(3字組の出現頻度)

IN NO IST WHEY CRATICT FROURE BIRS
 GROCID PONDENOME OF DEMONSTURES OF THE
 REPTAGIN IS REGOACIONA OF CRE

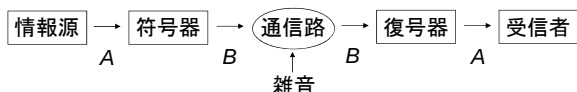
..... AR ?
 $\left\{ \begin{array}{l} \text{E, L} \rightarrow \text{高確率} \\ \text{その他} \rightarrow \text{低確率} \end{array} \right.$

各文字の頻度は直前の2文字の影響を受ける
 $\rightarrow$ 2次のマルコフ性

近似の精度を上げる $\rightarrow$ 精度は向上 but モデルは複雑

情報源のモデル $\rightarrow$ m 重マルコフモデルが適切
 ある時点の記号の発生確率は直前の m 個の記号に依存

1.5 情報伝送系のモデル



各パーツの役割

(1) 情報源 $\rightarrow$ 通報 $a_1 a_2 \dots a_M a_3$ を発生する

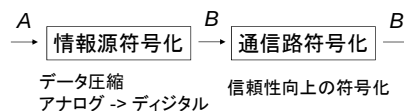
(2) 受信者 $\rightarrow$ 通報 $a_1 a_2 \dots a_M a_3$ を受け取る

(3) 通信路(channel) $\rightarrow$ 情報伝送を行う物理的手段

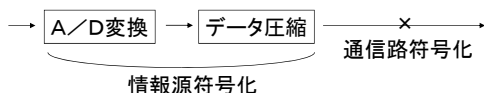
通信路記号 $B = \{b_1, b_2, \dots, b_N\}$
 $\rightarrow$ 通信路に送り込まれ取り出される記号

例 電気通信 $B = \{0, 1\}$

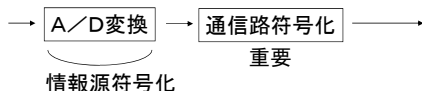
(4) 符号器(coder) $\rightarrow$ 情報源記号Aを通信路記号Bに変換



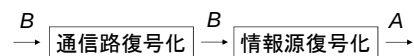
例 デジタルテレビ放送



CD(コンパクトディスク)



(5) 復号器(decoder) $\rightarrow$ 符号化の逆



情報源符号化 $\rightarrow$ 情報源の統計的性質を利用してデータ圧縮
 通信路符号化 $\rightarrow$ 通信路の誤りや雑音に対処し信頼性向上
 (目的の違うものは別々に考える)

第2章:情報量

概要

(1) 情報の定量化

(2) 2つの事象の間の相互情報量

事象AとBに関係があるときBを知ってAに関する情報は
 どれだけ得られるか?

例 天気予報を知って天気に関する情報

(3) 連続的情報を離散的情報に変換する操作

2.1 情報量

(a) 情報量の単位

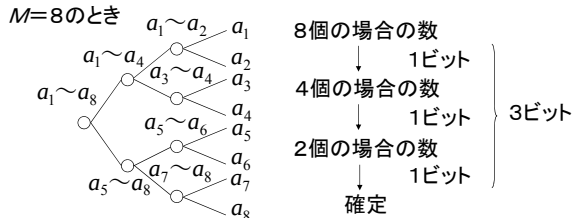
事象 a_1 と a_2 が各0.5で起こる確率試行Xの結果
 $= 1$ (ビット)の情報

- 例 ① 0か1か, 右か左か, 黒か白かなどに対する解答
 ② サイコロの目が偶数であることを知った
 ③ 昨日生まれた子供が女であることを知った

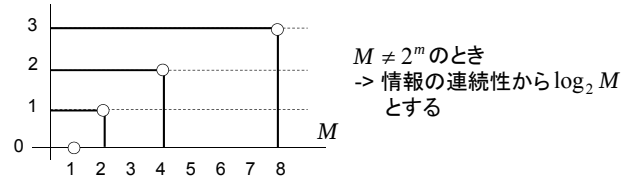
注意 $P(a_1) = P(a_2) = 0.5$ でなくてはならない

(b) 事象の数が多い場合

事象 $a_1, a_2, \dots, a_M$ が各 $\frac{1}{M}$ の確率で起こる試行 X の結果



$M = 2^m$ のとき $\rightarrow m$ ビット



(c) 事象の生起確率が等しくないとき

確率 $P(a)$ の事象が生じたことを知った時の情報量 $I(a)$

$$I(a) = \log_2 \frac{1}{P(a)} \quad \text{式(2.5)}$$

Why ?

$I(a) = f(p)$ とおき $f(p)$ の満たすべき条件

条件1 $f(p)$ は p の単調減少関数

確率 p が小さいほど情報量は大きい

条件2 $f(p_1 p_2) = f(p_1) + f(p_2)$ 加法性

互いに独立な事象 A と B の結果を知ったときの情報量
 $= (A \text{ の結果に関する情報量}) + (B \text{ の結果に関する情報量})$
 $f(p_1) \quad f(p_2)$

条件3 $f(p)$ は p の連続関数

条件4 $f(\frac{1}{2}) = 1$ 単位情報量の定義

条件1~4を同時に満足する関数形 $\rightarrow I(a) = \log_2 \frac{1}{P(a)}$
 事象 a の自己情報量

(d) ビットとナット

$I(a) = \log \frac{1}{P(a)}$ における対数の底の取り方は任意

$\log_2 \rightarrow$ 情報量の単位ビット (通常)
 $\log_e \rightarrow$ 情報量の単位ナット

2.2 エントロピー

事象 $a_1, a_2, \dots, a_M$ が確率 $P(a_1), P(a_2), \dots, P(a_M)$ で生起する試行 X

各事象 a_i の情報量 $I(a_i) = \log_2 \frac{1}{P(a_i)}$
 $(a_i \text{ に依存する})$

一回の試行で得られる情報量の期待値 $H(X)$

$$H(X) = E[I(a_i)] = \sum_{i=1}^M P(a_i) \log_2 \frac{1}{P(a_i)} \quad (\text{ビット}) \quad \text{式(2.7)}$$

期待値
エントロピー (平均情報量)

注意 $\left\{ \begin{array}{l} \text{情報量 } I(a_i) \rightarrow a_i \text{ に依存する} \\ \text{エントロピー } H(X) \rightarrow \text{試行に固有の定数} \end{array} \right.$

例 情報源から $A = \{a_1, a_2, a_3\}$ のいずれかの記号が発生

$P(a_1) = 0.3, P(a_2) = 0.6, P(a_3) = 0.1$

① 情報量

$$I(a_1) = -\log_2 0.3 = 1.74 \text{ (ビット)}$$

$$I(a_2) = -\log_2 0.6 = 0.74 \text{ (ビット)}$$

$$I(a_3) = -\log_2 0.1 = 3.32 \text{ (ビット)}$$

② エントロピー

$$H(X) = -(0.3 \log_2 0.3 + 0.6 \log_2 0.6 + 0.1 \log_2 0.1) = 1.298 \text{ (ビット)}$$

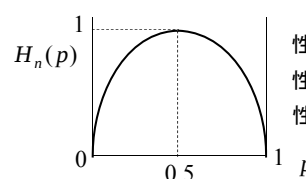
エントロピーの性質

(a) 記号数2で $A = \{0, 1\}$ の場合

$$P(0) = 1 - p, P(1) = p$$

エントロピー

$$H(X) = -p \log_2 p - (1 - p) \log_2 (1 - p) = H_n(p)$$



性質(1) $0 \leq H_n(p) \leq 1$

性質(2) $H_n(p)$ は $p = 0.5$ で最大値1

性質(3) $H_n(p)$ は凸関数

エントロピー → 不確定さの度合い

$p = 0$ or $p = 1$ のときは不確定さが無い → $H(X) = 0$

$p = 0.5$ のとき不確定さは最大 → $H(X) = 1$

(b) 記号数 $M > 2$ で $A = \{a_1, a_2, \dots, a_M\}$ の場合

$$\text{エントロピー } H(X) = - \sum_{i=1}^M P(a_i) \log_2 P(a_i)$$

性質(1) $0 \leq H(X) \leq \log_2 M$

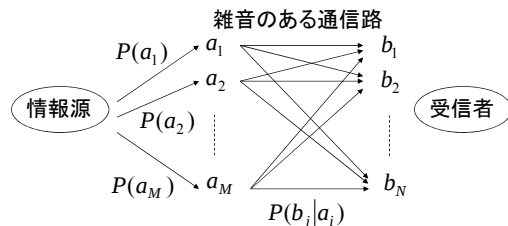
性質(2) $H(X)$ は $P(a_1) = P(a_2) = \dots = P(a_M) = \frac{1}{M}$ のとき最大値 $\log_2 M$ をとる

証明 → 教科書 p.28~29

2.3 相互情報量

最重要

問題の設定



受信者が b_j を受け取ったとき、 a_i についてどれだけの情報量が得られるか？

一般的な定式化

仮定1 事象 $a_1, a_2, \dots, a_M$ の生起確率は $P(a_1), P(a_2), \dots, P(a_M)$

仮定2 事象 $a_1, a_2, \dots, a_M$ と事象 $b_1, b_2, \dots, b_N$ の間に確率的な依存関係 → $P(b_j|a_i)$ で表現

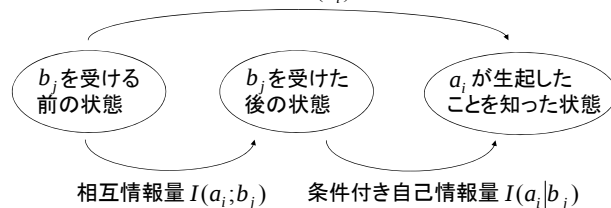
事象 b_j を知った時事象 a_i に関して得られる情報量
→ 相互情報量 $I(a_i; b_j)$

例 ① 天気予報を知って実際の天気について得られる情報量
② 誤りが起こるCDの再生音を聞いて得られる情報量

(a) 相互情報量 $I(a_i; b_j)$ の導出

考え方

自己情報量 $I(a_i)$



$$I(a_i; b_j) = I(a_i) - I(a_i|b_j)$$

(1) 必要な確率 $P(a_i, b_j)$, $P(b_j)$, $P(a_i|b_j)$ の計算

$P(a_i)$ と $P(b_j|a_i)$ は与えられている

情報源 通信路

$$P(a_i, b_j) = P(a_i)P(b_j|a_i)$$

$$P(b_j) = \sum_{i=1}^M P(a_i, b_j) = \sum_{i=1}^M P(a_i)P(b_j|a_i)$$

$$P(a_i|b_j) = \frac{P(a_i, b_j)}{P(b_j)} = \frac{P(a_i)P(b_j|a_i)}{\sum_{i=1}^M P(a_i)P(b_j|a_i)} \quad \text{ベイズの公式}$$

$P(a_i|b_j)$ → 事後確率 $P(a_i)$ → 事前確率

(2) 条件付き自己情報量 $I(a_i|b_j)$

b_j を知った後で a_i を知ったときの情報量
→ 情報量の計算で $P(a_i)$ を事後確率 $P(a_i|b_j)$ に置き換える

$$I(a_i|b_j) = -\log_2 P(a_i|b_j)$$

(3) 相互情報量 $I(a_i; b_j)$

$$I(a_i; b_j) = \overbrace{I(a_i)}^{\text{自己情報量}} - \overbrace{I(a_i|b_j)}^{\text{条件付き自己情報量}} = -\log_2 P(a_i) + \log_2 P(a_i|b_j)$$

自己情報量 条件付き自己情報量

$$= \log_2 \frac{P(a_i|b_j)}{P(a_i)} = \log_2 \frac{P(a_i, b_j)}{P(a_i)P(b_j)} \quad \text{式(2.21)}$$

(b) 相互情報量 $I(a_i; b_j)$ の性質

性質1 a_i と b_j に関する対称性

$$I(a_i; b_j) = I(b_j; a_i)$$

b_j による a_i に関する情報量 = a_i による b_j に関する情報量

例 天気予報が実際の天気について与える情報量
= 実際の天気が天気予報について与える情報量

性質2 $P(a_i|b_j) < P(a_i)$ のとき $I(a_i; b_j) < 0$ 嘘はマイナスの情報量

$P(a_i|b_j) < P(a_i) \rightarrow b_j$ が a_i に関する嘘の情報のとき

性質3 a_i と b_j が全く独立 ($P(a_i, b_j) = P(a_i)P(b_j)$) のとき $I(a_i; b_j) = 0$

b_j を聞いて a_i に関する情報は全く得られない

性質4 全く誤りがない ($P(b_j|a_i) = \delta_{ij}$) のとき $I(a_i; b_i) = I(a_i)$
クロネッカーのデルタ

性質5 $I(a_i; b_j) \leq I(a_i)$, $I(a_i; b_j) \leq I(b_j)$

相互情報量は自己情報量より大きくならない

例 天気予報の情報量 $\leq$ 実際の天気を知ったときの情報量
(百聞は一見にしかず)

(c) 相互情報量の計算例

例 天気予報の相互情報量

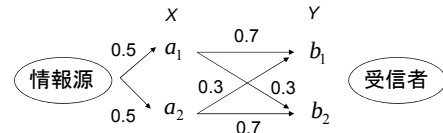
仮定 実際の天気 X ($a_1 \rightarrow$ 晴, $a_2 \rightarrow$ 雨)

雨と晴は50%の確率 $\rightarrow P(a_1) = P(a_2) = 0.5$

天気予報 Y ($b_1 \rightarrow$ 晴, $b_2 \rightarrow$ 雨)

70%の確率で当たる $\rightarrow P(b_1|a_1) = P(b_2|a_2) = 0.7$

$P(b_2|a_1) = P(b_1|a_2) = 0.3$



(1) 事後確率 $P(a_i|b_j)$ の計算

$$P(b_1) = P(b_2) = 0.5 \times 0.7 + 0.5 \times 0.3 = 0.5$$

$$P(a_1|b_1) = P(a_2|b_2) = \frac{P(a_i, b_i)}{P(b_i)} = \frac{0.5 \times 0.7}{0.5} = 0.7$$

($i = 1, 2$)

$$P(a_1|b_2) = P(a_2|b_1) = \frac{P(a_i, b_j)}{P(b_j)} = \frac{0.5 \times 0.3}{0.5} = 0.3$$

($i \neq j$)

(2) 自己情報量 $I(a_i)$

$$I(a_1) = I(a_2) = -\log_2 P(a_i) = -\log_2 0.5 = 1 \text{ (ビット)}$$

($i = 1, 2$)

(3) 条件付き自己情報量 $I(a_i|b_j)$

$$I(a_1|b_1) = I(a_2|b_2) = -\log_2 P(a_i|b_i) = -\log_2 0.7 = 0.515 \text{ (ビット)}$$

($i = 1, 2$)

$$I(a_2|b_1) = I(a_1|b_2) = -\log_2 P(a_i|b_j) = -\log_2 0.3 = 1.737 \text{ (ビット)}$$

($i \neq j$)

(4) 相互情報量 $I(a_i; b_j)$

$$I(a_1; b_1) = I(a_2; b_2) = I(a_i) - I(a_i|b_i) = 1 - 0.515 = 0.485 \text{ (ビット)}$$

($i = 1, 2$)

$$I(a_1; b_2) = I(a_2; b_1) = I(a_i) - I(a_i|b_j) = 1 - 1.737 = -0.737 \text{ (ビット)}$$

($i \neq j$)

$I(a_1; b_1) = I(a_2; b_2) > 0 \rightarrow$ 正しい天気予報は正の情報量
 $I(a_1; b_2) = I(a_2; b_1) < 0 \rightarrow$ 嘘の天気予報は負の情報量

2.4 平均相互情報量

最重要

期待値 (平均)

	個々の事象の情報量	全事象に関する平均情報量
単一確率変数 X	自己情報量 $I(a_i)$	エントロピー $H(X)$
2つの確率変数 X, Y	相互情報量 $I(a_i; b_j)$	平均相互情報量 $\bar{I}(X; Y)$

平均相互情報量 $\bar{I}(X; Y)$

相互情報量 $I(a_i; b_j)$ を結合確率 $P(a_i, b_j)$ で平均

$$\bar{I}(X; Y) = E[I(a_i; b_j)] = E\left[\log_2 \frac{P(a_i|b_j)}{P(a_i)}\right]$$

次ページへ続く

$$= \sum_i \sum_j P(a_i, b_j) \log_2 \frac{1}{P(a_i)} - \sum_i \sum_j P(a_i, b_j) \log_2 \frac{1}{P(a_i|b_j)}$$

エントロピー $H(X)$ 条件付きエントロピー $H(X|Y)$

$$= H(X) - H(X|Y) \quad \text{式(2.36)}$$

(a) 平均相互情報量 $\bar{I}(X; Y)$ の性質

性質1 X と Y に関する対称性

$$\bar{I}(X; Y) = \bar{I}(Y; X) = H(Y) - H(Y|X)$$

性質2 結合エントロピーによる表現

$$\bar{I}(X; Y) = H(X) + H(Y) - H(X, Y) \quad \text{式(2.39)}$$

$$H(X, Y) = - \sum_i \sum_j P(a_i, b_j) \log_2 P(a_i, b_j) \quad \text{結合エントロピー} \quad \text{式(2.40)}$$

まとめ

$$\begin{aligned}\bar{I}(X;Y) &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) \\ &= H(X) + H(Y) - H(X,Y)\end{aligned}$$

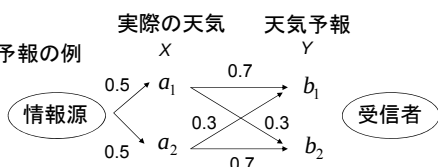
性質3 $\bar{I}(X;Y) \geq 0$ 平均相互情報量は負にならない

個々の事象に対する $I(a_i; b_j)$ は負になる

証明 → 教科書 p.40~41

(b) 計算例

前回の天気予報の例



$\bar{I}(X;Y) = H(X) - H(X|Y)$ から平均相互情報量を求める

(1) 情報源のエントロピー $H(X)$

$$H(X) = 1 \text{ (ビット)}$$

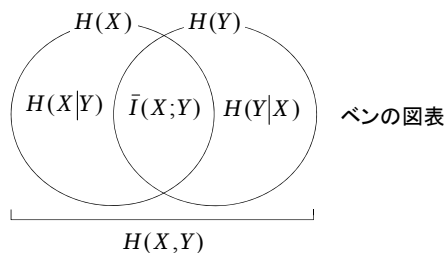
(2) 条件付きエントロピー $H(X|Y)$

$$\begin{aligned}H(X|Y) &= -\sum_i \sum_j P(a_i)P(b_j|a_i) \log_2 P(a_i|b_j) \quad \text{対称性} \\ &= \{-P(a_1)P(b_1|a_1) \log_2 P(a_1|b_1) - P(a_1)P(b_2|a_1) \log_2 P(a_1|b_2)\} \times 2 \\ &= \{-0.5 \times 0.7 \times \log_2 0.7 - 0.5 \times 0.3 \times \log_2 0.3\} \times 2 = 0.88 \text{ (ビット)}\end{aligned}$$

(3) 平均相互情報量 $\bar{I}(X;Y)$

$$\bar{I}(X;Y) = H(X) - H(X|Y) = 1 - 0.88 = 0.12 \text{ (ビット)}$$

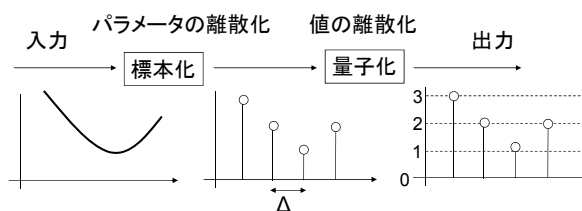
(c) 簡単な公式の覚え方(教科書に書いていない)



$H(X)$ → Xに含まれる平均情報量
 $H(Y)$ → Yに含まれる平均情報量
 $\bar{I}(X;Y)$ → XとYの両方に含まれる平均情報量
 $H(X|Y)$ → Xには含まれるがYには含まれない平均情報量
 $H(Y|X)$ → Yには含まれるがXには含まれない平均情報量
 $H(X,Y)$ → XとYを同時に知った時の平均情報量

2.5 標本化と量子化

音声や画像などの連続的パラメータ連続的情報を離散的パラメータ離散的情報に変換する手法

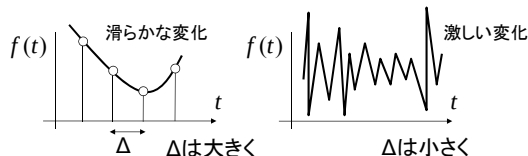


(a) 標本化(サンプリング)

関数 $f(t)$ の間隔 Δ おきの値 $\dots, f(-2\Delta), f(-\Delta), f(0), f(\Delta), f(2\Delta), \dots$ を取り出す

Δ → サンプリング間隔

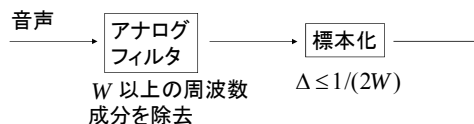
サンプリング間隔 Δ の選び方



根拠 → 標本化定理

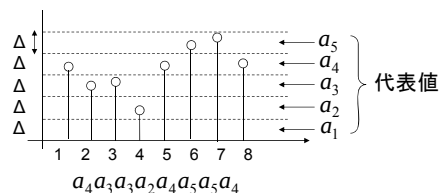
ある関数 $f(t)$ が W ヘルツ以上の周波数成分を含まないとき、 $1/(2W)$ 以下の間隔 Δ で取った標本値 $\dots, f(-2\Delta), f(-\Delta), f(0), f(\Delta), f(2\Delta), \dots$ から一意的に $f(t)$ が復元できる

実際の音声や画像 → W ヘルツ以上の周波数成分を含む
 前置フィルタを使う



(b) 量子化

各標本値 $\dots, f(-2\Delta), f(-\Delta), f(0), f(\Delta), f(2\Delta), \dots$ に対し、実数をいくつかの区間に分割し代表値で置き換える



$$a_i - \frac{\Delta}{2} \leq f(k\Delta) \leq a_i + \frac{\Delta}{2} \text{ ならば } f(k\Delta) \text{ を } a_i \text{ で置き換える}$$

画像 -> 8ビット(0~255の整数), 音楽 -> 16ビット(0~65535)

量子化による誤差 -> 量子化雑音

量子化の方法

(1) 一様量子化と非線形量子化

一様量子化 -> 量子化幅 Δ を一定にとる

非線形量子化 -> 量子化幅 Δ を不均一にとる
(頻度が高い値を細かく量子化)

(2) スカラー量子化とベクトル量子化

スカラー量子化 -> 各標本値を独立に量子化

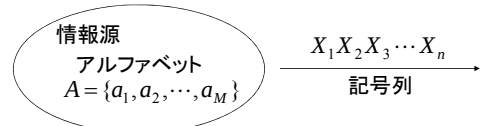
ベクトル量子化 -> 複数標本値を組にして量子化
(データ圧縮の効果がある)

第3章:情報源と符号化

概要

- (1) マルコフ過程による情報源モデルとそのエントロピー
- (2) 情報圧縮のための符号化とその限界
- (3) 具体的な符号化の方法

3.1 マルコフ情報源



ある時点の記号 X_i の生起確率

-> 直前の m 個 $X_{i-m}, X_{i-m+1}, \dots, X_{i-1}$ に依存する

m 重マルコフ過程

($m=1$ のとき単純マルコフ過程)

ある一つの系列 $x_1 x_2 \dots x_n$ の生起確率の一般式

$$P(x_1, x_2, \dots, x_n) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2) \dots P(x_n|x_1, x_2, \dots, x_{n-1})$$

式(3.1)

(1) 無記憶情報源

各時点の記号の生起は独立

$$P(x_1, x_2, \dots, x_n) = P(x_1)P(x_2) \dots P(x_n)$$

(2) m 重マルコフ情報源

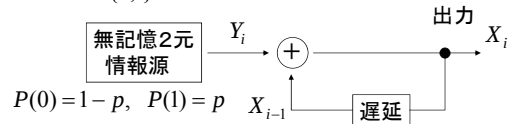
$$P(x_i|x_1, x_2, \dots, x_{i-1}) = P(x_i|x_{i-m}, x_{i-m+1}, \dots, x_{i-1}) \quad i > m$$

例 $m=1$ のとき

$$P(x_1, x_2, \dots, x_n) = P(x_1)P(x_2|x_1)P(x_3|x_2) \dots P(x_n|x_{n-1})$$

単純マルコフ情報源の例

$$A = \{0, 1\}$$



$$P(0) = 1 - p, \quad P(1) = p$$

$$X_i = Y_i \oplus X_{i-1} = (Y_i + X_{i-1}) \bmod 2 \quad \text{排他的論理和}$$

2で割った余り

$$P(0|0) = P(1|1) = 1 - p, \quad P(0|1) = P(1|0) = p$$

(a) 状態遷移図

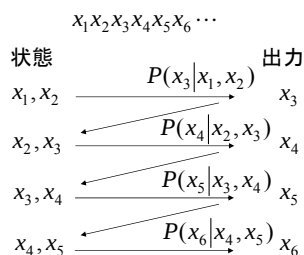
m 重マルコフ過程の動作を記述

状態 -> 直前の m 個の記号列

マルコフ過程

-> 各時点で記号を出力し別の状態に遷移する動作を繰り返す

例 2重マルコフ過程

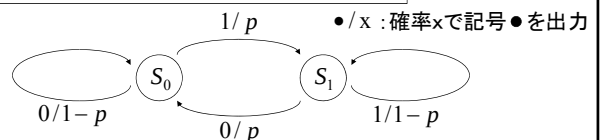


例1 先の単純マルコフ情報源

状態 $S_0 = (0), S_1 = (1)$

$$S_0 = (0), \quad S_1 = (1)$$

$$P(0|0) = P(1|1) = 1 - p, \quad P(0|1) = P(1|0) = p$$



例2 2重マルコフ情報源 ($A = \{0, 1\}$)

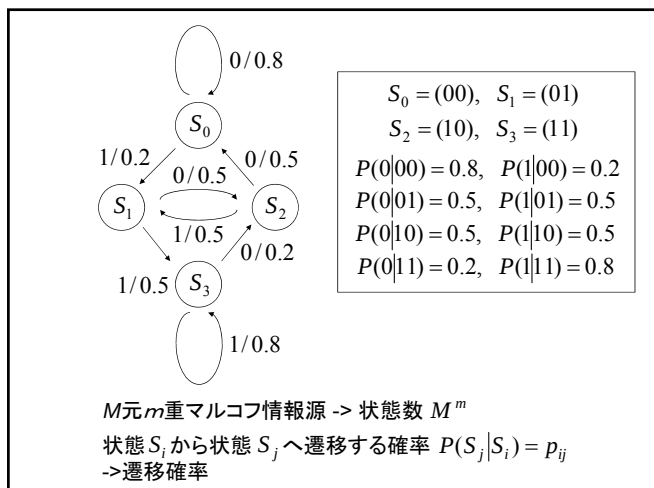
$$P(0|00) = 0.8, \quad P(1|00) = 0.2$$

$$P(0|01) = 0.5, \quad P(1|01) = 0.5$$

$$P(0|10) = 0.5, \quad P(1|10) = 0.5$$

$$P(0|11) = 0.2, \quad P(1|11) = 0.8$$

状態 $S_0 = (00), S_1 = (01), S_2 = (10), S_3 = (11)$



$$P_t = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{bmatrix}$$
 遷移確率行列

$$\sum_{j=1}^N p_{ij} = 1$$

(b) マルコフ過程の数学的基礎

(1) 過渡状態

十分時間が経過した後には到達できなくなる状態

(2) 既約

どの状態 S_i から出発してどの状態 S_j にも行ける

(3) 高次の遷移確率

状態 S_i から出発して n 回の遷移で状態 S_j に到達する確率 $\rightarrow n$ 次の遷移確率 $p_{ij}^{(n)}$

$$P_t^{(n)} = \begin{bmatrix} p_{11}^{(n)} & p_{12}^{(n)} & \cdots & p_{1N}^{(n)} \\ p_{21}^{(n)} & p_{22}^{(n)} & \cdots & p_{2N}^{(n)} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1}^{(n)} & p_{N2}^{(n)} & \cdots & p_{NN}^{(n)} \end{bmatrix}$$
 n 次の遷移確率行列

$P_t^{(n)}$ を P_t を用いて表現する

① $n=1$ のとき ($S_i \rightarrow S_j$)

$p_{ij}^{(1)} = p_{ij}, P_t^{(1)} = P_t$

② $n=2$ のとき ($S_i \rightarrow S_v \rightarrow S_j$)

何でも良い

$$p_{ij}^{(2)} = \sum_{v=1}^N p_{iv} p_{vj}, P_t^{(2)} = P_t P_t$$

③ 一般の n のとき ($S_i \xrightarrow{n-1回} S_v \xrightarrow{何でも良い} S_j$)

最重要

$$p_{ij}^{(n)} = \sum_{v=1}^N p_{iv}^{(n-1)} p_{vj}, P_t^{(n)} = P_t^{(n-1)} P_t = \underbrace{P_t P_t \cdots P_t}_{n回}$$

(4) 定常確率

状態 S_i から出発するマルコフ過程

時刻 $t=0 \quad t=1 \quad t=2 \quad t=3 \quad t=\infty$

状態 $S_i \rightarrow S_\alpha \rightarrow S_\beta \rightarrow S_\gamma \cdots \rightarrow S_j$

無限に時間が経過した後で状態 S_j にいる確率分布

$\lim_{n \rightarrow \infty} p_{ij}^{(n)} \rightarrow$ 定常確率

① エルゴート定理

既約で非周期的なマルコフ過程

初期状態 S_i に依存しない唯一の定常確率 $P(S_j)$ に収束

$$P(S_j) = \lim_{n \rightarrow \infty} p_{ij}^{(n)}$$

② 定常確率 $P(S_j)$ の計算法

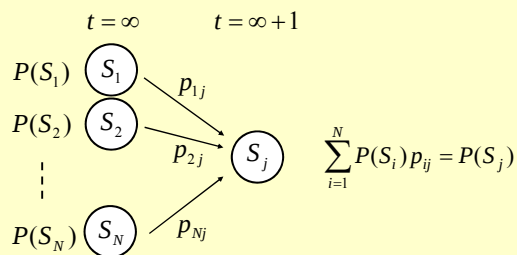
公式

$$\begin{cases} \sum_{j=1}^N P(S_j) = 1 & (\text{確率の総和は1}) \\ \sum_{i=1}^N P(S_i) p_{ij} = P(S_j) \end{cases}$$

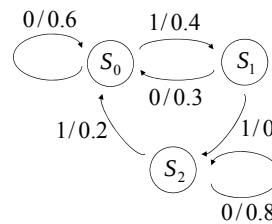
(定常状態から1回遷移を行っても確率分布は変化しない)

板書

2つ目の公式の導出



例



遷移確率行列

$$P_t = \begin{bmatrix} 0.6 & 0.4 & 0 \\ 0.3 & 0 & 0.7 \\ 0.2 & 0 & 0.8 \end{bmatrix}$$

定常確率は？

公式より

$$\begin{cases} P(S_0) + P(S_1) + P(S_2) = 1 \\ 0.6P(S_0) + 0.3P(S_1) + 0.2P(S_2) = P(S_0) \\ 0.4P(S_0) = P(S_1) \\ 0.7P(S_1) + 0.8P(S_2) = P(S_2) \end{cases}$$

$P(S_0), P(S_1), P(S_2)$ について解くと

$$P(S_0) = 0.3571, P(S_1) = 0.1429, P(S_2) = 0.5$$

※ 実際に $t=0$ で $(P(S_0), P(S_1), P(S_2)) = (1, 0, 0)$ とすると

$$t=1 \quad (P(S_0), P(S_1), P(S_2)) = (0.6, 0.4, 0)$$

$$t=2 \quad (P(S_0), P(S_1), P(S_2)) = (0.48, 0.24, 0.28)$$

$$t=3 \quad (P(S_0), P(S_1), P(S_2)) = (0.416, 0.192, 0.392)$$

$$t=4 \quad (P(S_0), P(S_1), P(S_2)) = (0.3856, 0.1664, 0.448)$$

3.2 マルコフ情報源のエントロピー

情報源 $\rightarrow X_1 X_2 X_3 \dots X_n$

1記号あたりのエントロピー $H(X)$

(1) 無記憶情報源

板書

別の計算方法

$$[1, 0, 0](P_t)^n$$

$X_1, X_2, X_3, \dots, X_n$ は無関係

$$H(X) = -\sum_x P(x) \log_2 P(x) \quad \text{1つの記号だけを考えて十分}$$

(2) マルコフ情報源

$X_1, X_2, X_3, \dots, X_n$ に確率的依存関係

$$H(X) = -\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{x_1} \sum_{x_2} \dots \sum_{x_n} P(x_1, x_2, \dots, x_n) \log_2 P(x_1, x_2, \dots, x_n)$$

全記号をまとめて考える必要 $\rightarrow$ 計算がたいへん

最重要

(a) 定常確率を用いたマルコフ情報源のエントロピー

状態 S_i にあるとき記号 a_k の発生確率 $\rightarrow P(a_k | S_i)$

S_i に条件付けたエントロピー

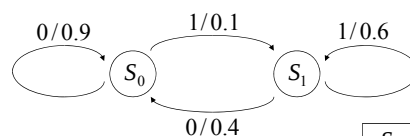
$$H_{S_i}(X) = -\sum_{k=1}^M P(a_k | S_i) \log_2 P(a_k | S_i)$$

状態 S_i の定常確率 $P(S_i)$ で平均

$$H(X) = -\sum_{i=1}^N \sum_{k=1}^M P(S_i) P(a_k | S_i) \log_2 P(a_k | S_i) \quad \text{式(3.24)}$$

定義式から計算したのと等価

例



エントロピーは？

$$\begin{matrix} S_0 = (0) \\ S_1 = (1) \end{matrix}$$

① 定常確率

$$P(S_0) = 0.8, P(S_1) = 0.2$$

② エントロピー

S_0 に条件付けたエントロピー

$$\begin{aligned} H_{S_0}(X) &= -P(0|S_0) \log_2 P(0|S_0) - P(1|S_0) \log_2 P(1|S_0) \\ &= -0.9 \log_2 0.9 - 0.1 \log_2 0.1 \\ &= 0.469 \text{ (ビット)} \end{aligned}$$

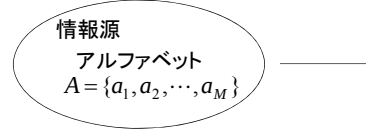
S_1 に条件付けたエントロピー

$$\begin{aligned} H_{S_1}(X) &= -P(0|S_1) \log_2 P(0|S_1) - P(1|S_1) \log_2 P(1|S_1) \\ &= -0.4 \log_2 0.4 - 0.6 \log_2 0.6 \\ &= 0.971 \text{ (ビット)} \end{aligned}$$

定常確率 $P(S_i)$ で平均をとると

$$\begin{aligned} H(X) &= P(S_0)H_{S_0}(X) + P(S_1)H_{S_1}(X) \\ &= 0.8 \times 0.469 + 0.2 \times 0.971 \\ &= 0.5694 \text{ (ビット)} \end{aligned}$$

(b) 相対エントロピーと冗長度



エントロピーの最大値 $H_{\max}$

マルコフ性がなく $P(a_1) = P(a_2) = \dots = P(a_M) = \frac{1}{M}$ のとき

$$H_{\max} = \log_2 M$$

各記号の発生頻度の偏りやマルコフ性があるときのエントロピー

$$H < H_{\max}$$

$$h = \frac{H}{H_{\max}} \quad \text{相対エントロピー}$$

(偏りやマルコフ性がない時を基準に)

$$r = 1 - \frac{H}{H_{\max}} \quad \text{冗長度}$$

(情報にどれだけ無駄があるか)

例 ① $P(0) = 0.8, P(1) = 0.2$ の無記憶情報源

$$H_{\max} = 1 \text{ (ビット)}$$

$$H = -0.8 \log_2 0.8 - 0.2 \log_2 0.2 = 0.72 \text{ (ビット)}$$

$$h = 0.72, r = 0.28$$

② 英語の冗長度は約50%

3.3 情報源の符号化

$$\begin{aligned} X_1 X_2 X_3 \dots X_n &\xrightarrow{\text{符号化}} Y_1 Y_2 Y_3 \dots Y_{n'} \\ A = \{a_1, a_2, \dots, a_M\} &\xrightarrow{\text{符号化}} B = \{b_1, b_2, \dots, b_N\} \end{aligned}$$

記号の変換

目的

- (1) 通信路で使用する記号列への変換
- (2) データの圧縮

例

- (1) 音声をデジタル伝送のため2元符号に変換
- (2) 数値を計算機に記録するため2元符号に変換

符号化の方法

情報源記号	符号語
a_1	b_1
a_2	$b_1 b_2$
$\vdots$	$\vdots$
a_M	$b_1 b_N b_2$

符号化テーブル
情報源の記号列を1つずつ符号語に変換

一対一 $a_1 a_2 a_M \longrightarrow b_1 \underbrace{b_1 b_2}_{b_1 b_2} \underbrace{b_1 b_N b_2}_{b_1 b_N b_2}$

符号化の基礎概念

$\{A, B, C, D\}$ を2元符号 $\{0, 1\}$ に

記号	確率	可変長					
		等長					
		C_1	C_2	C_3	C_4	C_5	C_6
A	0.6	00	0	0	0	0	0
B	0.25	01	10	10	01	10	10
C	0.1	10	110	110	011	11	11
D	0.05	11	1110	111	111	01	0
平均符号長		2.0	1.6	1.55	1.55	1.4	1.35

$$L = \sum_{A, B, C, D} \text{確率} \times \text{符号語長}$$

(1) C_6 は正しく復号できない → 特異符号

AとDに同じ符号語

(2) C_5 は特異でないが非可逆 → 一意復号不可能

$\begin{array}{ccc} 0 & 1 & 1 & 0 \\ \square & \square & \square & \square \\ A & C & A & \end{array} \quad \begin{array}{ccc} 0 & 1 & 1 & 0 \\ \square & \square & \square & \square \\ D & B & & \end{array} \quad \text{2つ以上の解釈}$

(3) $C_1 \sim C_4$ は一意復号可能

C_3 と C_4 の相違

0 1 1 1 1 1 1 0
 A D D A

$C_3 \rightarrow$ 符号のパターンが現れた時点で復号できる

$C_4 \rightarrow$ 符号のパターンが現れた時点で復号できない

0 1 1 1 1 1 1 0 0 1 1 1 1 1 1
 B D C D

符号のパターンが現れた時点で復号可能 $\rightarrow$ 瞬時復号可能

好ましい符号の条件

- (1) 一意復号可能, 瞬時復号可能
- (2) 平均符号長ができるだけ短い

平均符号長は情報源の統計的性質に依存

	符号1	符号2
A	00	0
B	01	10
C	10	110
D	11	111

(1) $P(A) = P(B) = P(C) = P(D) = \frac{1}{4}$ のとき

符号1 $L_1 = \frac{1}{4} \times 2 + \frac{1}{4} \times 2 + \frac{1}{4} \times 2 + \frac{1}{4} \times 2 = 2$ $L_1 < L_2$

符号2 $L_2 = \frac{1}{4} \times 1 + \frac{1}{4} \times 2 + \frac{1}{4} \times 3 + \frac{1}{4} \times 3 = 2.25$

(2) $P(A) = \frac{1}{2}, P(B) = \frac{1}{4}, P(C) = \frac{1}{8}, P(D) = \frac{1}{8}$ のとき

符号1 $L_1 = 2$ 符号2 $L_2 = 1.75$ $L_1 > L_2$

符号化の問題

- (1) いかなる情報源にはいかなる符号が適するか?
- (2) 平均符号長はどこまで短くできるか?

(a) 瞬時復号可能性と符号の木

A $\rightarrow$ 0, B $\rightarrow$ 01, C $\rightarrow$ 011, D $\rightarrow$ 111 が瞬時復号できない理由

0 が 01, 011 の語頭になっている
 01 が 011 の語頭になっている

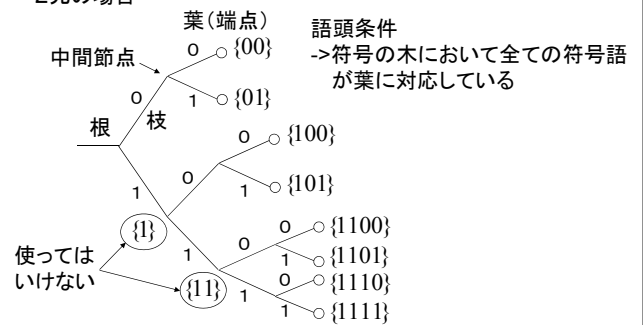
瞬時復号可能の必要十分条件 (語頭条件)

どの符号語 x も別の符号語 y の語頭になっていない

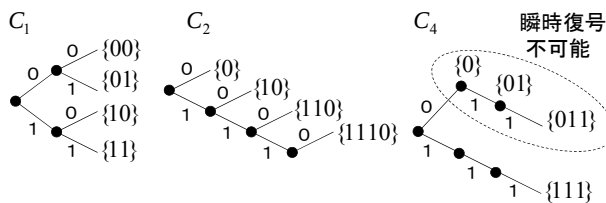
符号の木

語頭条件の図解

2元の場合



例 C_1, C_2, C_4 の符号の木



(b) クラフトの不等式

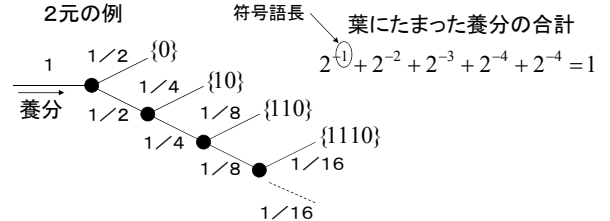
$a_1 \rightarrow \otimes \otimes \otimes, a_2 \rightarrow \otimes \otimes \otimes \otimes, \dots, a_M \rightarrow \otimes \otimes \otimes \otimes$ と符号化

$l_1, l_2, \dots, l_M \rightarrow$ 符号長

瞬時復号可能な符号のとき $l_1, l_2, \dots, l_M$ に関する制約

定性的説明

2元の場合



$l=1 \quad l=2 \quad l=3$

全ての符号語が葉に対応している $\Rightarrow \sum_{i=1}^M 2^{-l_i} \leq 1$

全ての葉に対応する符号語を使う $\rightarrow \sum_{i=1}^M 2^{-l_i} = 1$ 完全符号
 全ての葉に対応する符号語を使わない $\rightarrow \sum_{i=1}^M 2^{-l_i} < 1$

r 元符号の場合も一般化可能

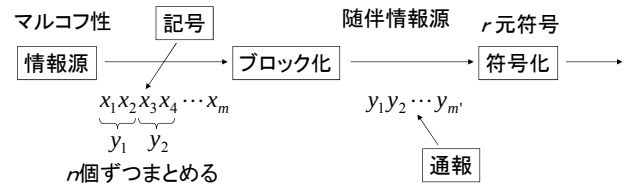
定理

記号の数 r 個で符号長 $l_1, l_2, \dots, l_M$ の瞬時復号可能な符号語が作り得る必要十分条件

$$\sum_{i=1}^M r^{-l_i} \leq 1 \quad \text{クラフトの不等式}$$

(符号語はいくらでも短くできるわけではない)

3.4 符号化の基本定理



ブロック符号化 $n \rightarrow$ ブロック長

1記号あたり(1通報あたり)の平均符号長をどこまで短くできるか?

例 $n=2$ の3元情報源の2元ブロック符号化

入力 00120121
 $y_1 \ y_2 \ y_3 \ y_4$

通報	符号語	入力
$u_1 = 00$	00	00120121
$u_2 = 01$	01	$y_1 \ y_2 \ y_3 \ y_4$
$u_3 = 02$	100	
$u_4 = 10$	101	
$u_5 = 11$	1100	
$u_6 = 12$	1101	
$u_7 = 20$	1110	
$u_8 = 21$	11110	
$u_9 = 22$	11111	

9通り

出力 00 1101 01 11110

瞬時復号可能

定理1

M 種類の通報 $u_1, u_2, \dots, u_M$ を持つブロック符号化において、(通報あたりの)平均符号長 L を次式を満たすように符号化できる.

$$\frac{H(U)}{\log_2 r} \leq \bar{L} < \frac{H(U)}{\log_2 r} + 1$$

$$H(U) \rightarrow \text{通報のエントロピー} \quad H(U) = -\sum_{i=1}^M P(u_i) \log_2 P(u_i)$$

(証明は時間があれば板書で)

定理1から定理2の導出

定理1 $\rightarrow$ 1通報あたりの平均符号長 $\bar{L}$
 1記号あたりの平均符号長 $\bar{l}$ は?

$$\bar{l} = \frac{\bar{L}}{n}$$

$H(U)$ と $H(X)$ の関係

$$H(U) \rightarrow nH(X) \quad (n \rightarrow \infty) \quad (\text{式(3.26)})$$

$H(X) \rightarrow$ 1記号あたりのエントロピー, $n \rightarrow$ ブロック長

定理1に代入すると

$$\frac{H(X)}{\log_2 r} \leq \bar{l} < \frac{H(X)}{\log_2 r} + \left(\frac{1}{n}\right)$$

n を大きくすれば
いくらでも小さくなる

定理2

任意の $\varepsilon > 0$ に対して、ブロック長 n を十分大きくとれば、1記号あたりの平均符号長 $\bar{l}$ が次式を満たすように符号化できる.

$$\frac{H(X)}{\log_2 r} \leq \bar{l} < \frac{H(X)}{\log_2 r} + \varepsilon$$

定理1, 2 $\rightarrow$ シヤノンの第1符号化定理

意味

① 2元符号のとき

ブロック長 n を大きくとれば、平均符号長 $\bar{l}$ は情報源のエントロピー $H(X)$ にいくらでも近づけられる.

② r 元符号のとき

ブロック長 n を大きくとれば、平均符号長 $\bar{l}$ は情報源のエントロピー $H(X)/\log_2 r$ にいくらでも近づけられる.

実用上の問題点

n を大きくとる $\rightarrow$ 装置の複雑化, delay が長くなる

例 3記号 A, B, C を持つ無記憶情報源の符号化

$$P(A) = \frac{1}{2}, \quad P(B) = \frac{3}{8}, \quad P(C) = \frac{1}{8}$$

(1) 情報源エントロピー $H(X)$

$$H(X) = \frac{1}{2} \log_2 2 + \frac{3}{8} \log_2 \frac{8}{3} + \frac{1}{8} \log_2 8 = 1.4056 \text{ (ビット)}$$

(2) ブロック長 $n=1$ の符号化

	符号	平均符号長
A	1	$\bar{l} = \frac{1}{2} \times 1 + \frac{3}{8} \times 2 + \frac{1}{8} \times 2 = 1.5$
B	01	
C	00	

$$\frac{H(X)}{1.4056} \leq \frac{\bar{l}}{1.5} \leq \frac{H(X)}{1.4056} + \frac{1}{n} \text{ を満足している}$$

(3) ブロック長 $n=2$ の符号化

	確率	符号	確率が小さいブロックに長い符号
AA	1/4	00	
AB	3/16	10	$\frac{H(X)}{1.4056} \leq \frac{\bar{l}}{1.9056} \leq \frac{H(X)}{1.4056} + \frac{1}{n}$
AC	1/16	1101	
BA	3/16	010	を満足している
BB	9/64	011	
BC	3/64	1110	
CA	1/16	1100	
CB	3/64	11110	
CC	1/64	11111	

$$\bar{l} = \frac{\bar{L}}{n} = \frac{\frac{1}{4} \times 2 + \frac{3}{16} \times (2+3) + \frac{9}{64} \times 3 + \frac{1}{16} \times (4+4) + \frac{3}{64} \times (4+5) + \frac{1}{64} \times 5}{2} = \frac{183}{128} = 1.4297$$

ブロック長 n が長い方が平均符号長 $\bar{l}$ が短くなる

通信容量 (言葉の定義だけ)

一秒間に ν 個の r 元符号を送れる通信路

$$C = \nu \log_2 r \text{ (ビット/秒)}$$

r 元符号が1秒間に送ることができる最大の情報量
-> 符号の通信容量

3.5 ハフマンの符号化法

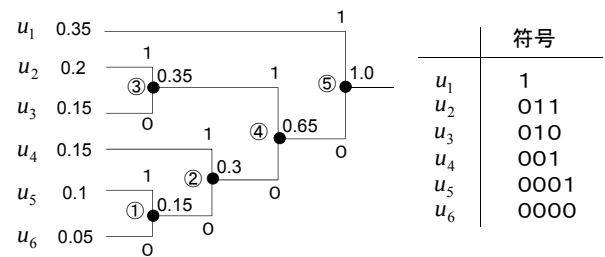
有限のブロック長 n に対して平均符号長 $\bar{l}$ を最小にする符号
-> 最短符号

ハフマンによる最短符号の構成法

文章による説明 -> 教科書 p.80

例 通報 $u_1, u_2, u_3, u_4, u_5, u_6$ を2元符号化

$$P(u_1) = 0.35, P(u_2) = 0.2, P(u_3) = 0.15, P(u_4) = 0.15, P(u_5) = 0.1, P(u_6) = 0.05$$



符号の木を作っている, 瞬時復号可能で完全

$$\bar{l} = 0.35 \times 1 + 0.2 \times 3 + 0.15 \times (3+3) + 0.1 \times 4 + 0.05 \times 4 = 2.45$$

注意

ハフマン符号は唯一に決まるとは限らない

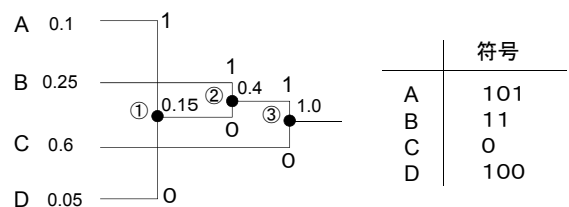
-> 確率が同じになるとき, 最も小さい2つの選び方に任意性

平均符号長 $\bar{l}$ はどのような選び方をしても等価

	符号1	符号2	符号3	平均符号長 $\bar{l}$ は いずれも2.45
u_1	1	11	0	
u_2	011	01	111	
u_3	010	00	101	
u_4	001	101	100	
u_5	0001	1001	1101	
u_6	0000	1000	1100	

例 A, B, C, D を発生する情報源のハフマン符号化

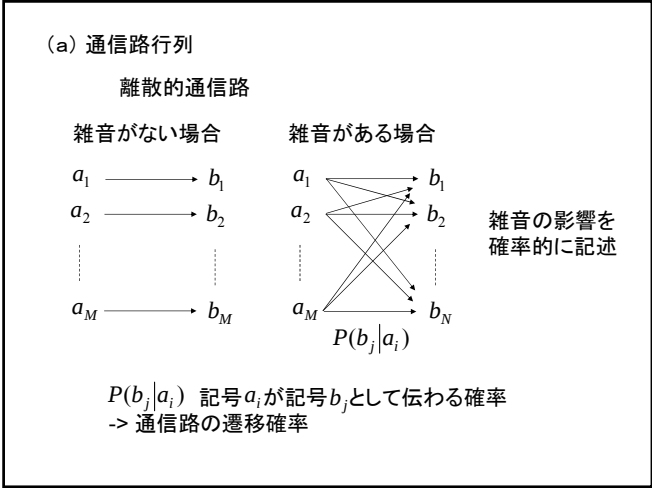
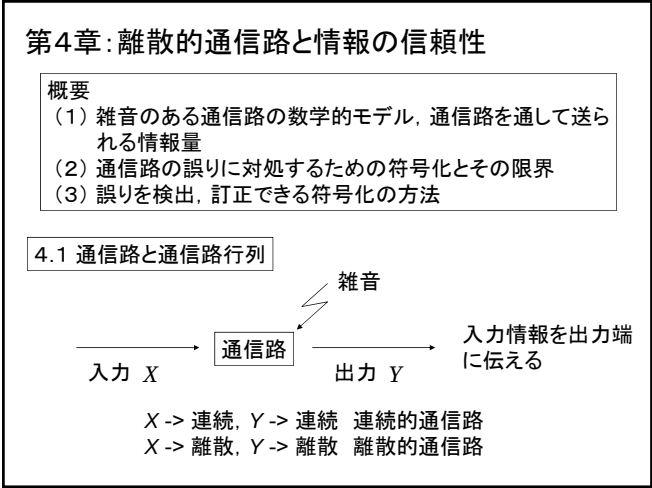
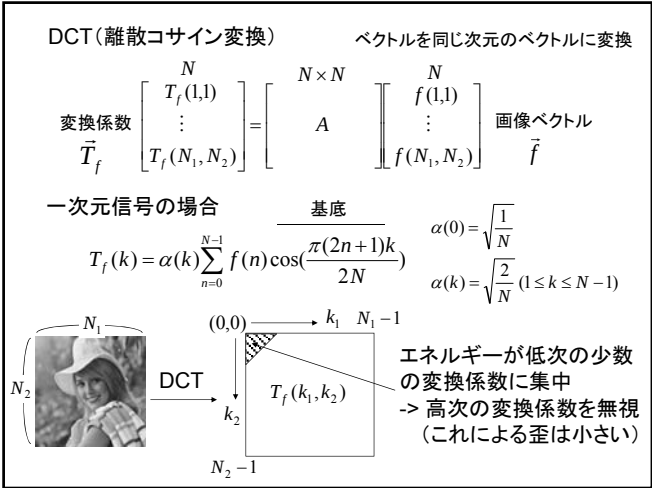
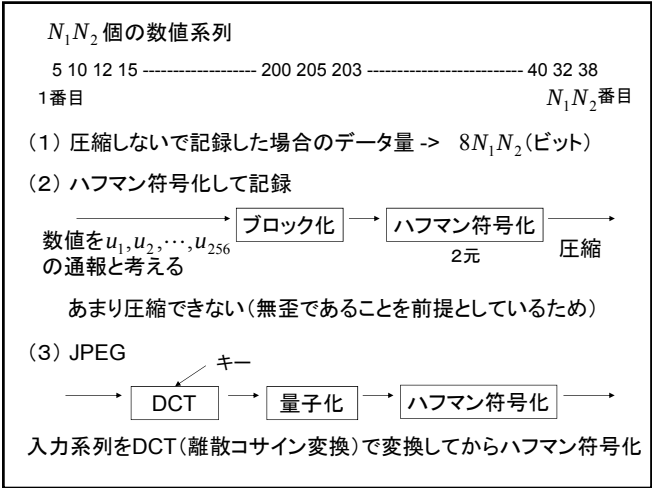
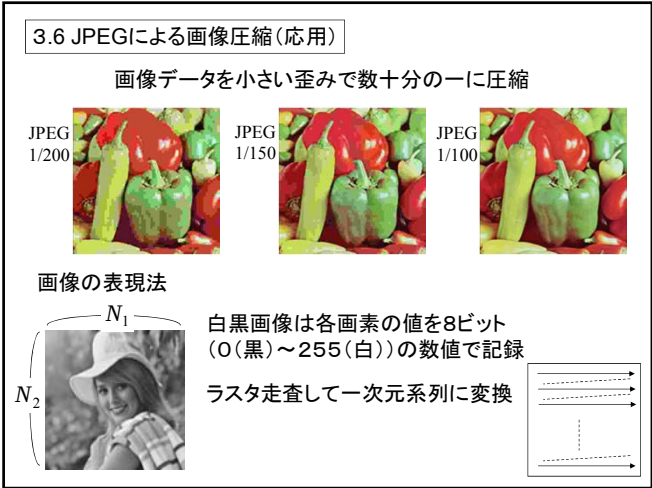
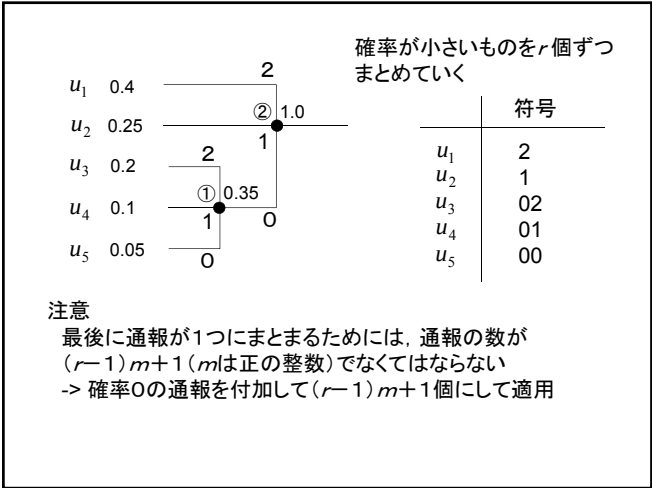
$$P(A) = 0.1, P(B) = 0.25, P(C) = 0.6, P(D) = 0.05$$



r 元符号への拡張

例 通報 u_1, u_2, u_3, u_4, u_5 を3元符号化

$$P(u_1) = 0.4, P(u_2) = 0.25, P(u_3) = 0.2, P(u_4) = 0.1, P(u_5) = 0.05$$



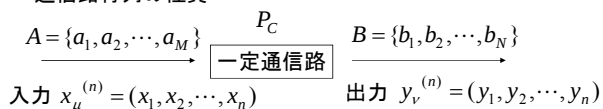
$$P_C = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{M1} & p_{M2} & \cdots & p_{MN} \end{bmatrix} \quad p_{ij} = P(b_j | a_i)$$

通信路行列

一定(無記憶)通信路 → P_C が時間で変化しない
 記憶がある通信路 → P_C がマルコフ性により変化する

例 バースト通信路(誤りが一定時間連続して起こる)
 → 記憶がある

通信路行列の性質



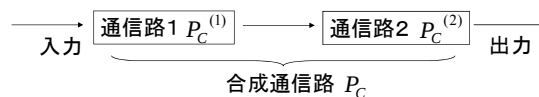
性質1 $(x_\mu^{(n)}, y_\nu^{(n)})$ に対する通信路行列

$$P_C = \underbrace{\begin{bmatrix} P(y_\nu^{(n)} | x_\mu^{(n)}) \end{bmatrix}}_{N^n} M^n \quad P(y_\nu^{(n)} | x_\mu^{(n)}) = \prod_{i=1}^n P(y_i | x_i)$$

性質2 出力記号 b_j の発生確率

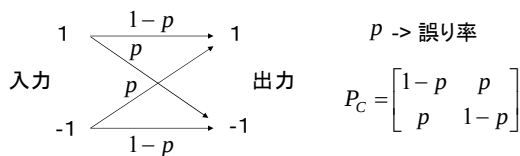
$$P(b_j) = \sum_{i=1}^M P(a_i) p_{ij} \quad P(a_i) \rightarrow \text{入力記号 } a_i \text{ の発生確率}$$

性質3 2つの通信路の縦続接続 $P_C = P_C^{(1)} P_C^{(2)}$

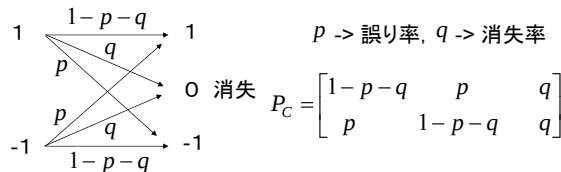


(b) 代表的な通信路

(1) 2元対称通信路(BSC)



(2) 2元対称消失通信路

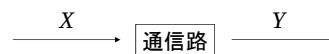


4.2 伝送情報量とあいまい度

通信路を通して送られる情報量
 = 送信記号の情報量 - 雑音により失われた情報量

入力と出力の相互情報量

(1) 独立な1記号だけを送る場合



平均相互情報量

$$\begin{aligned} \bar{I}(X; Y) &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) \\ &= H(X) + H(Y) - H(X, Y) \end{aligned}$$

次ページへ続く

$H(X) \rightarrow$ 送信記号のエントロピー

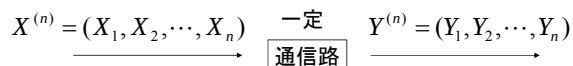
$H(Y) \rightarrow$ 受信記号のエントロピー

$H(X|Y) \rightarrow$ あいまい度(条件付きエントロピー)

(受信記号を受けた後どれだけあいまい性が残っているか)

$\bar{I}(X; Y) \rightarrow$ 情報伝送速度(ビット/秒), 情報伝送量(ビット/記号)
 (受信記号を受けて消滅したあいまい性)

(2) 情報源にマルコフ性がある場合



長さ n の系列を考える

情報伝送速度

$$\bar{I}(X; Y) = \lim_{n \rightarrow \infty} \frac{\bar{I}(X^{(n)}; Y^{(n)})}{n}$$

n 記号間の相互情報量

$$\bar{I}(X^{(n)}; Y^{(n)}) = H(Y^{(n)}) - H(Y^{(n)} | X^{(n)})$$

第1項

$$\begin{aligned} H(Y^{(n)}) &= H(Y_1) + H(Y_2 | Y_1) + H(Y_3 | Y_1, Y_2) + \cdots \\ &\quad + H(Y_n | Y_1, Y_2, \dots, Y_{n-1}) \\ &= \sum_{i=1}^n H(Y_i | Y_1, Y_2, \dots, Y_{i-1}) \end{aligned}$$

第2項

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

性質

$$\bar{I}(X^{(n)}; Y^{(n)}) \leq \sum_{i=1}^n \bar{I}(X_i; Y_i) \quad (\text{等号は情報源が無記憶のとき})$$

マルコフ性の分だけ損

4.3 離散的通信路の通信容量

最重要

情報伝送速度 $\bar{I}(X;Y)$ → 通信路と情報源の両者に依存
通信路を通して伝送できる最大の情報量は？

$$C = \max_{P(X)} \bar{I}(X;Y) = \max_{P(X)} \{H(Y) - H(Y|X)\}$$

無記憶情報源の記号発生確率

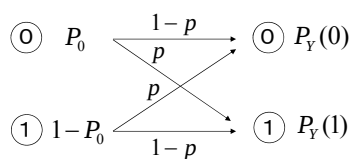
この表現の方が使いやすい

“情報源を変化させて伝送情報量が最大になるとき”

C → 通信容量(通信路容量) (ビット/秒) or (ビット/記号)

計算例

(1) 2元対称通信路



受信記号の確率

$$P_Y(0) = P_0(1-p) + (1-P_0)p, P_Y(1) = P_0p + (1-P_0)(1-p)$$

受信記号のエントロピー

$$H(Y) = -P_Y(0)\log_2 P_Y(0) - P_Y(1)\log_2 P_Y(1)$$

条件付きエントロピー

$$H(Y|X) = -\sum_x \sum_y P(x,y)\log_2 P(y|x)$$

次ページへ続く

$$= -[P_0(1-p)\log_2(1-p) + (1-P_0)p\log_2 p + P_0p\log_2 p + (1-P_0)(1-p)\log_2(1-p)]$$

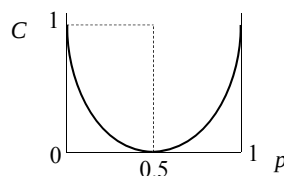
$$= -p\log_2 p - (1-p)\log_2(1-p)$$

情報伝送速度

$$\bar{I}(X;Y) = -P_Y(0)\log_2 P_Y(0) - P_Y(1)\log_2 P_Y(1) + p\log_2 p + (1-p)\log_2(1-p) \text{ (ビット/記号)}$$

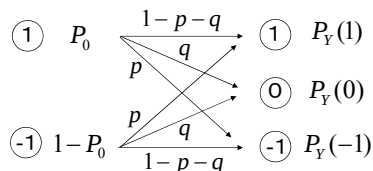
上式は $P_Y(0) = P_Y(1) = \frac{1}{2}$ のとき最大になるので

$$C = 1 + p\log_2 p + (1-p)\log_2(1-p) \text{ (ビット/記号)}$$



通信路で誤る確率が小さいほど
通信容量は大きい

(2) 2元対称消失通信路



受信記号の確率

$$P_Y(1) = P_0(1-p-q) + (1-P_0)p, P_Y(0) = q,$$

$$P_Y(-1) = P_0p + (1-P_0)(1-p-q)$$

受信記号のエントロピー

$$H(Y) = -P_Y(1)\log_2 P_Y(1) - P_Y(0)\log_2 P_Y(0) - P_Y(-1)\log_2 P_Y(-1)$$

条件付きエントロピー

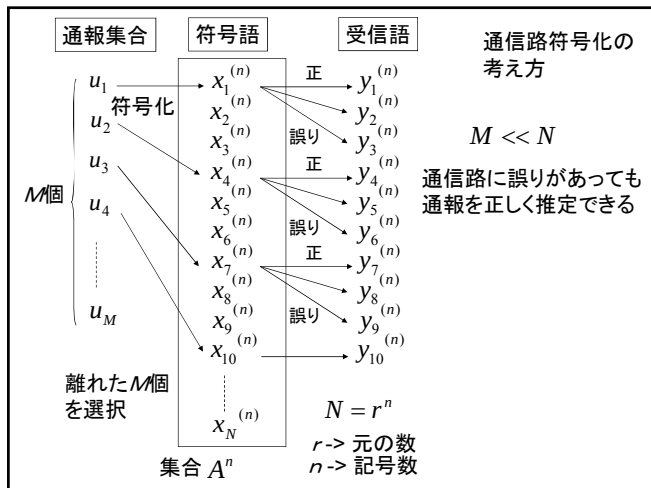
$$H(Y|X) = -(1-p-q)\log_2(1-p-q) - q\log_2 q - p\log_2 p$$

$H(Y)$ は $P_0 = \frac{1}{2}$ のとき最大になるから

$$C = (1-q)(1-\log_2(1-q)) + (1-p-q)\log_2(1-p-q) + p\log_2 p \text{ (ビット/記号)}$$

4.4 通信路の符号化および復号化

通信路で生じる誤りの影響を抑え信頼性向上
→ 通信路符号化



簡単な例

$$M=2 \quad u_1 = \{0\}, u_2 = \{1\}$$

$$N=8 \quad A^3 = \{000, 001, 010, 011, 100, 101, 110, 111\}$$

$$x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5 \quad x_6 \quad x_7 \quad x_8$$

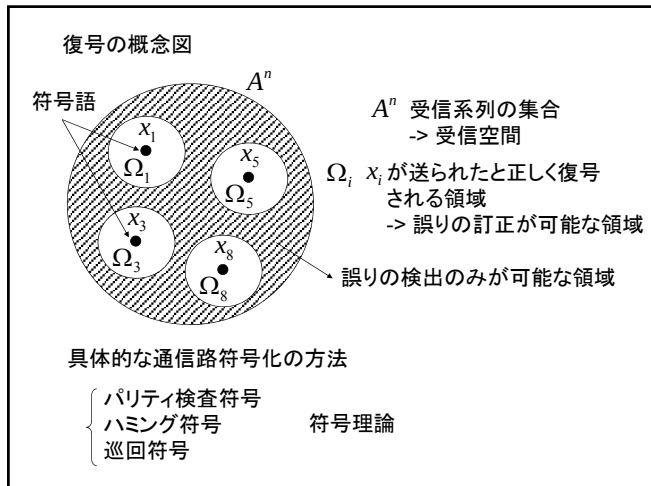
$$\text{符号化} \quad u_1 \rightarrow x_1 (0 \rightarrow 000), u_2 \rightarrow x_8 (1 \rightarrow 111)$$

$$\text{復号化} \quad x_1, x_2, x_3, x_5 \rightarrow u_1, x_4, x_6, x_7, x_8 \rightarrow u_2$$

通信路で1個の誤りが生じても訂正可能

$M < N$ のとき, 可能な通信路記号系列の一部を使用しない
 $\rightarrow$ 誤りの検知, 訂正が可能となる

A^n から選ばれた符号化に用いる系列集合 $\rightarrow$ 通信路符号



(a) 符号の効率

通信路符号の冗長性の尺度

可能なM個の通信路記号系列からM個を選び等確率で用いる

$$\rightarrow \text{情報伝送速度} \quad R = \bar{I}(X;Y) = \frac{\log_2 M}{n} \quad (\text{ビット/記号})$$

可能なM個を全て等確率で用いる

$$\rightarrow \text{情報伝送速度} \quad R_{\max} = \bar{I}(X;Y) = \frac{\log_2 N}{n} \quad (\text{ビット/記号})$$

$$\eta = \frac{R}{R_{\max}} \rightarrow \text{符号の効率(どれだけ冗長になるか)}$$

(b) 復号法

(1) 最大事後確率(MAP)復号法

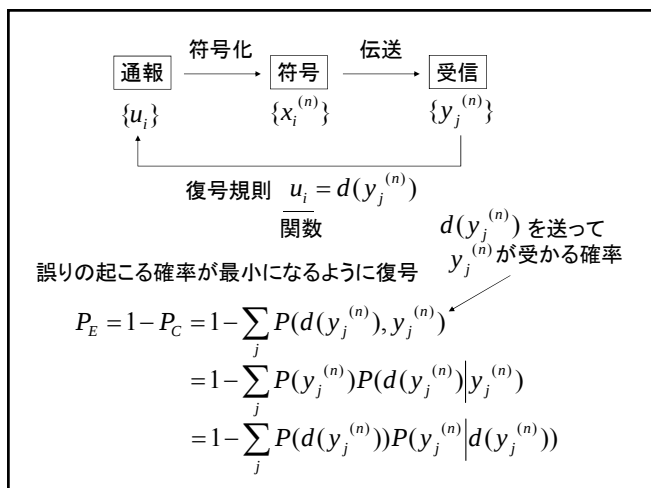
$$u_i = \arg \min_{u_i} P_E = \arg \max_{u_i} P(u_i | y_j^{(n)}) \quad P_E \text{ を最小化}$$

$$= \arg \max_{u_i} P(u_i) P(y_j^{(n)} | u_i)$$

(2) 最大尤度(ML)復号法

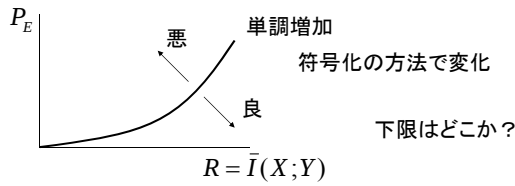
$P(u_i) = \text{一定と仮定して}$ P_E を最小化

$$u_i = \arg \max_{u_i} P(y_j^{(n)} | u_i)$$



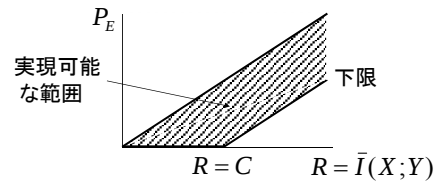
4.5 通信路符号化の基本定理

伝送速度 $R = \bar{I}(X; Y)$ と誤り率 P_E の関係



P_E を小さく保って R をいかに大きくできるか?
 R を一定に保って P_E をいかに小さくできるか?

答 -> シャノンの第2符号化定理



適切な符号化を行えば、いくらでも通信容量 C に近い伝送速度といくらでも小さい誤り率 P_E が達成できる

定理

通信容量が C (ビット/秒) の一定通信路
 情報源を R (ビット/秒) の伝送速度で符号化

ブロック長を $n \rightarrow \infty$ とすれば

- (1) $R \leq C \Rightarrow$ 誤り率 P_E をいくらでも小さくする符号化法が存在
- (2) $R > C \Rightarrow$ あいまい度 $H(X|Y)$ が次式となる符号化法が存在
 $R - C < H(X|Y) < R - C + \varepsilon$ P_E と同じ意味

存在定理 -> 具体的な符号化法を示したわけでない
 (証明は時間があれば板書で)

実用上の問題点

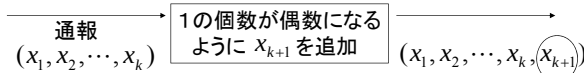
- (1) 具体的な符号化法が存在しない
- (2) $n \rightarrow \infty$ とすると装置の複雑化, delayが長くなる

4.6 パリティ検査符号

誤りの検出, 訂正が可能な具体的な符号化法

(a) パリティ検査

簡単な例(2元)



通信路で1個の誤りが起こる

-> 全体の1の個数が奇数となり誤り検出

誤り検出符号

$$\begin{cases} (x_1, x_2, \dots, x_k) \rightarrow \text{情報記号} \\ x_{k+1} \rightarrow \text{検査符号} \end{cases}$$

数学的表現

$$\text{符号化 } x_{k+1} = x_1 \oplus x_2 \oplus \dots \oplus x_k \text{ を追加}$$

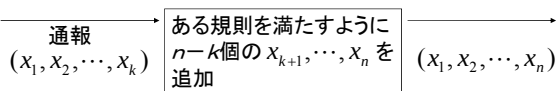
$$(a \oplus b = (a + b) \bmod 2)$$

$$\text{復号化 } S = x_1 \oplus x_2 \oplus \dots \oplus x_k \oplus x_{k+1} \text{ を計算}$$

$$S = \begin{cases} 0 \rightarrow \text{誤りなし} \\ 1 \rightarrow \text{誤りあり} \end{cases} \quad \text{パリティ検査}$$

誤りの検出は可能だが訂正はできない

(b) 一般化(ハミングの単一誤り訂正符号)



ある規則

$$\begin{cases} h_{11}x_1 \oplus h_{12}x_2 \oplus \dots \oplus h_{1n}x_n = 0 \\ h_{21}x_1 \oplus h_{22}x_2 \oplus \dots \oplus h_{2n}x_n = 0 \\ \vdots \\ h_{m1}x_1 \oplus h_{m2}x_2 \oplus \dots \oplus h_{mn}x_n = 0 \end{cases} \quad \begin{matrix} h_{ij} = 0 \text{ or } 1 \\ \text{パリティ検査方程式} \end{matrix}$$

$$(m = n - k)$$

$$H\vec{X} = \vec{0} \quad \text{パリティ検査符号(線形符号)} \quad (n, k)$$

復号の方法

$$\vec{S} = H\vec{Y} \text{ を計算} \quad (1) \vec{S} = \vec{0} \text{ のとき} \rightarrow \text{誤りなし}$$

$$\text{シンδροム 受信記号} \quad (2) \vec{S} \neq \vec{0} \text{ のとき} \rightarrow \text{誤りあり}$$

x_i に誤りがあるとき

$$\vec{S} = (h_{1i}, h_{2i}, \dots, h_{mi})^T \text{ 行列 } H \text{ の } i \text{ 列}$$

行列 H の各列を全て違うものにしておく
 -> 1つまでの誤りの箇所の特定, 訂正可能

例

$$H = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \quad \begin{matrix} \text{情報記号 } k=4 \\ \text{検査符号 } m=3 \end{matrix}$$

$$\vec{S} = (0, 0, 1)^T \rightarrow x_1 \text{ に誤り}, \vec{S} = (1, 0, 1)^T \rightarrow x_5 \text{ に誤り}$$

組織符号

$$H = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1k} & 1 & 0 & \cdots & 0 \\ h_{21} & h_{22} & \cdots & h_{2k} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{m1} & h_{m2} & \cdots & h_{mk} & 0 & 0 & \cdots & 1 \end{bmatrix} \quad m = [H' \ I] \quad \text{のとき}$$

$\underbrace{\hspace{10em}}_k$ 情報記号
 $\underbrace{\hspace{10em}}_m$ 検査記号

(c) 生成行列

符号化 $\rightarrow$ k 次元空間から n 次元空間への写像

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \underbrace{n \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1k} \\ g_{21} & g_{22} & \cdots & g_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ g_{n1} & g_{n2} & \cdots & g_{nk} \end{bmatrix}}_k \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_k \end{bmatrix} \quad \bar{X} = G \bar{U}$$

符号 通報
 $G \rightarrow$ 生成行列

例 組織符号

$$H = [H' \ I] \rightarrow G = \begin{bmatrix} I \\ H' \end{bmatrix}$$

4.7 CDにおけるデータ記録(応用)

CDは記録密度は高いが読み込みエラーが起きやすい
 $\rightarrow$ 誤り訂正符号不可欠

記録するものがデータか音楽かで2つの規格

① 音楽用CD (CD-DA, Compact Disc Digital Audio)

② データCD (CD-ROM, Compact Disc Read-Only-Memory)

音楽とデータで一番違う点
 $\rightarrow$ 誤り訂正符号で修正できないエラーの取り扱い(傷や劣化)

リードソロモン符号

組織符号の1つで、符号化・復号化の処理は複雑だがバースト誤り(誤りが2つ以上続く)を訂正できる

第5章: 情報伝送と信号

概要

- (1) 通信路内部での信号伝送の解析法
- (2) 標本化定理
- (3) 通信方式の基礎 (AM, FMの原理)

5.1 信号とスペクトル

離散通信路の内部

信号 $\rightarrow$ 通報の通信路内部での物理的表現

(a) 信号の解析

フーリエ解析 $\rightarrow$ 任意の信号を正弦波の重ね合わせで表現

(1) 非周期波形

$e^{i\omega t} = \cos(\omega t) + i \sin(\omega t)$
 オイラーの公式

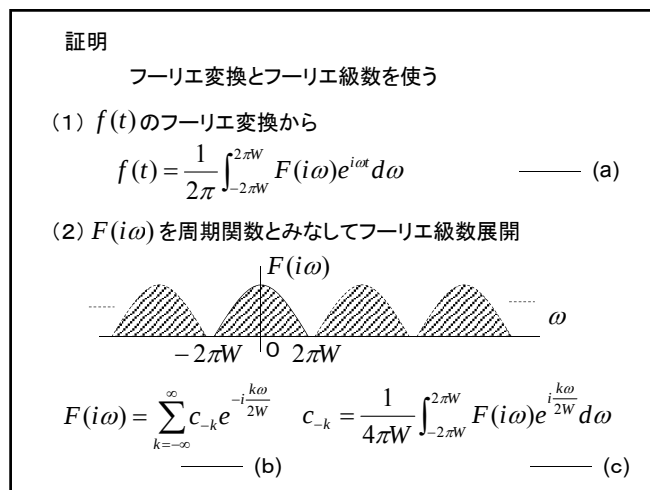
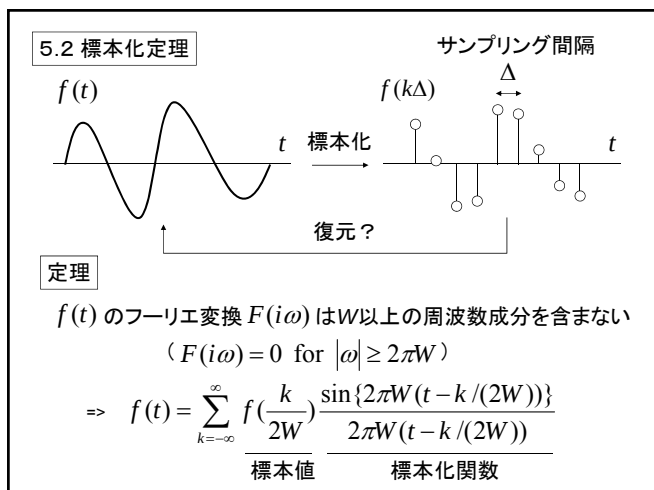
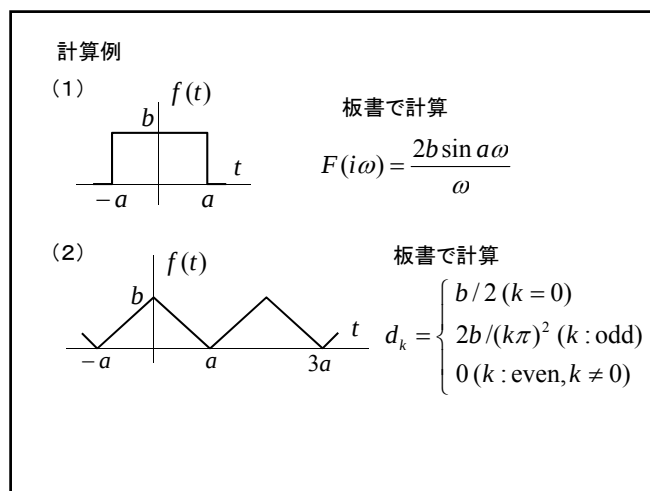
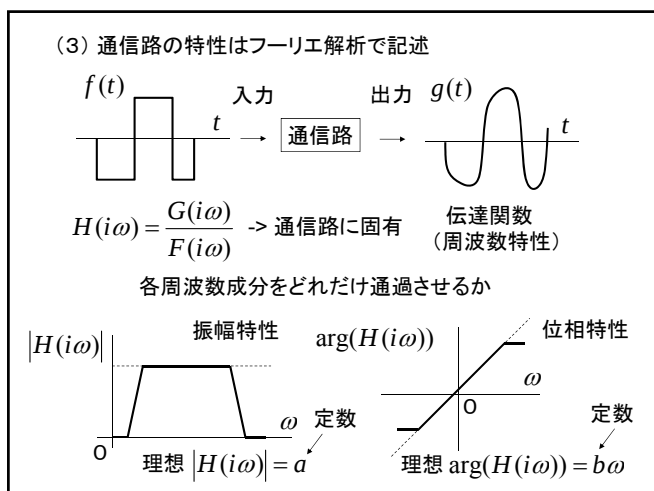
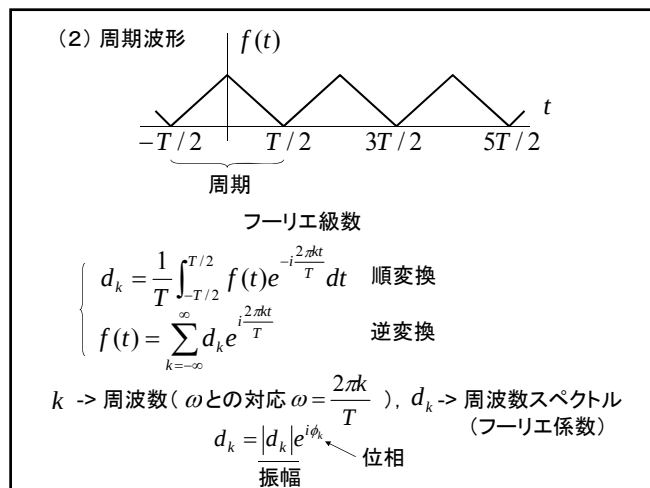
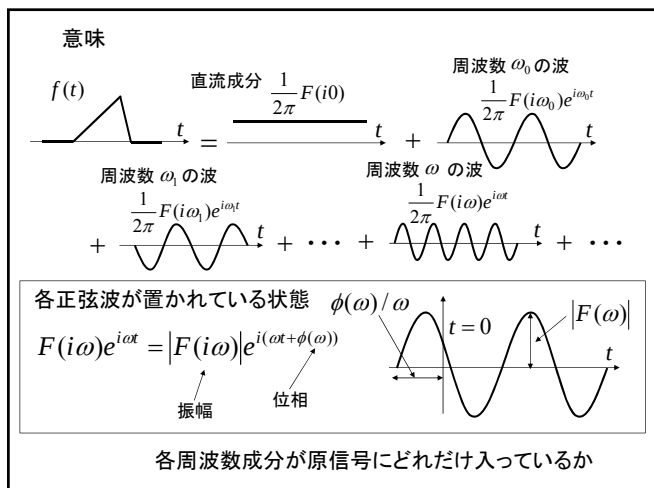
フーリエ変換

$$\begin{cases} F(i\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt & \text{順変換} \\ f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(i\omega) e^{i\omega t} d\omega & \text{逆変換} \end{cases}$$

重ね合わせ 正弦波

$\omega \rightarrow$ 角周波数, $F(i\omega) \rightarrow$ 周波数スペクトル

$$F(i\omega) = \underbrace{|F(i\omega)|}_{\text{振幅}} e^{i\phi(\omega)} \quad \leftarrow \text{位相}$$



式(a)と式(c)の比較により

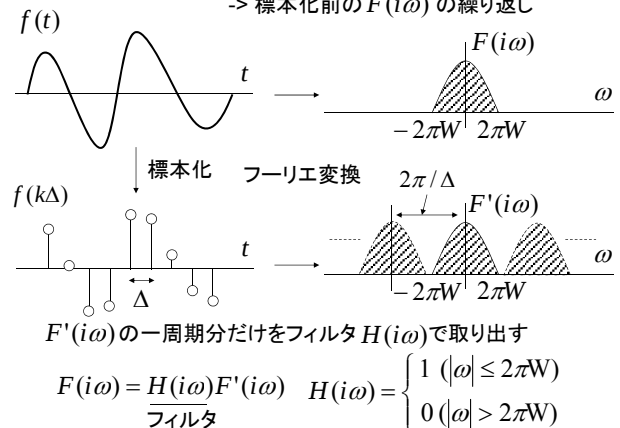
$$c_{-k} = \frac{1}{2W} f\left(\frac{k}{2W}\right) \quad \text{--- (d)}$$

式(a),(b),(d)から

$$\begin{aligned} f(t) &= \frac{1}{2\pi} \int_{-2\pi W}^{2\pi W} \left(\sum_{k=-\infty}^{\infty} \frac{1}{2W} f\left(\frac{k}{2W}\right) e^{-i\frac{k\omega}{2W}} \right) e^{i\omega t} d\omega \\ &= \sum_{k=-\infty}^{\infty} f\left(\frac{k}{2W}\right) \frac{1}{4\pi W} \int_{-2\pi W}^{2\pi W} e^{i\omega(t-k/(2W))} d\omega \\ &= \sum_{k=-\infty}^{\infty} f\left(\frac{k}{2W}\right) \frac{\sin\{2\pi W(t-k/(2W))\}}{2\pi W(t-k/(2W))} \quad // \end{aligned}$$

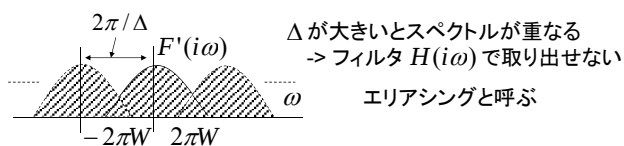
物理的意味

標本化後の $F'(i\omega)$
→ 標本化前の $F(i\omega)$ の繰り返し



両辺のフーリエ逆変換 → 標本化定理

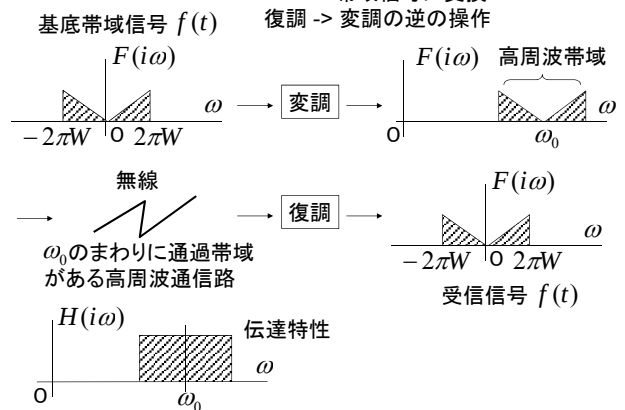
サンプリング間隔が大きいと何故復元できないか



5.3 通信方式の基礎

変調 → 基底帯域信号を送信に適した帯域信号に変換

復調 → 変調の逆の操作

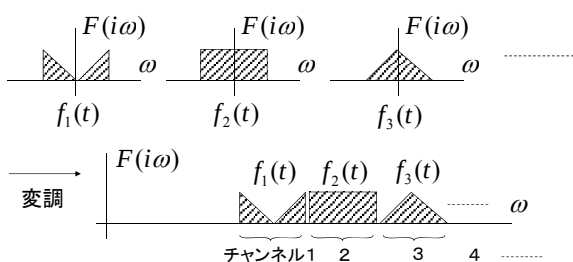


変調の目的

(1) 通信路の通過帯域が高周波の場合

無線, 有線の一部

(2) 周波数多重化(FDM)



受信側 → フィルタで希望チャンネルの帯域のみを取り出す

変調の方法

搬送波(高周波の正弦波)のパラメータを基底帯域信号で変化

$$A \cos(\omega_c t + \theta)$$

- { 振幅 A を変化 → 振幅変調(AM)
- { 周波数 ω を変化 → 周波数変調(FM)
- { 位相 θ を変化 → 位相変調(PM)

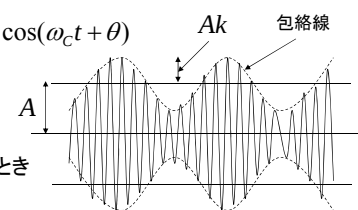
(a) 振幅変調(AM)

$$f(t) = A(1 + kv(t)) \cos(\omega_c t + \theta)$$

変調波 振幅

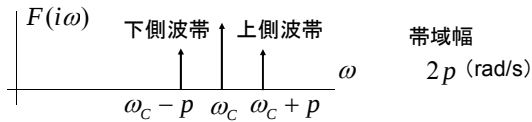
簡単のため

$$v(t) = \cos(pt + \phi) \text{ のとき}$$

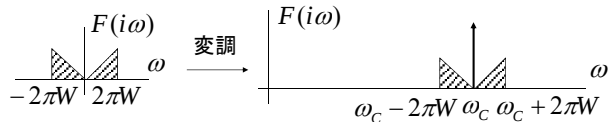


周波数スペクトル(単一正弦波 $v(t) = \cos pt$ のとき)

$$f(t) = A \cos(\omega_c t + \theta) + \frac{Ak}{2} \cos\{(\omega_c + p)t + \theta\} + \frac{Ak}{2} \cos\{(\omega_c - p)t + \theta\}$$



$v(t)$ が一般の信号の場合



3つの異なる伝送方式

- (1) 搬送波 + 両側波帯
- (2) 両側波帯
- (3) 単側波帯のみ → 帯域幅削減

(b) 位相変調と周波数変調 (PM, FM)

位相変調

$$f(t) = A \cos(\omega_c t + \underbrace{mv(t) + \theta_0}_{\text{位相}}) \quad m \rightarrow \text{変調指数}$$

周波数変調

$$f(t) = A \cos \psi(t) \text{ の瞬間周波数 } \rightarrow \omega(t) = \frac{d\psi(t)}{dt}$$

$$\omega(t) = \omega_c + \Omega_d v(t) \text{ として}$$

$$f(t) = A \cos\left(\int \omega(t) dt\right) = A \cos(\omega_c t + \Omega_d \int v(t) dt + \theta_0)$$

$\Omega_d / (2\pi) \rightarrow \text{周波数偏移}$

周波数変調と位相変調の関係

- $\left\{ \begin{array}{l} v(t) \text{ を微分した後周波数変調} \rightarrow \text{位相変調} \\ v(t) \text{ を積分した後位相変調} \rightarrow \text{周波数変調} \end{array} \right.$

周波数スペクトル(単一正弦波 $v(t) = \cos pt$ のとき)

位相変調

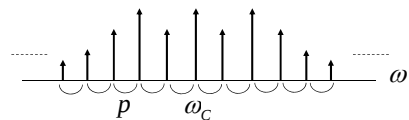
$$f(t) = A \cos(\omega_c t + \underbrace{m \cos pt + \theta_0}_{\text{等価}})$$

周波数変調

$$f(t) = A \cos(\omega_c t + \underbrace{\left(\frac{\Omega_d}{p}\right) \sin pt + \theta_0'}_{\text{等価}}) \quad m = \frac{\Omega_d}{p}$$

上式を正弦波の和に分解

$$f(t) = \sum_{\nu=-\infty}^{\infty} \frac{J_{\nu}(m)}{\nu \text{ 次ベッセル関数}} \cos\left\{(\omega_c + \nu p)t + \frac{\nu\pi}{2}\right\}$$



$$|\nu| > m + 1 \text{ のとき実質的に } J_{\nu}(m) \approx 0$$

$-(m+1) \leq \nu \leq m+1$ の範囲のスペクトルのみ考えればよい

$$m \ll 1 \Rightarrow \text{帯域幅 } 2p \text{ (rad/s) AMと同じ}$$

$$m \gg 1 \Rightarrow \text{帯域幅 } (2m+2)p \text{ (rad/s)}$$