情報理論

第2章 情報理論の問題

2.2 問題の設定

中心課題：(1)通信路使用の効率の向上〈ディジタル通信路：記号数↓ / アナログ通信路：時間と周波数帯域↓

(2)信頼性(reliability)の向上〈ディジタル通信路：誤り率↓ / アナログ：雑音↓

情報源 → 情報源符号化 → 通信路符号化 → 通信路 → 通信路復号 → 情報源復号 → あて先

情報源系列　符号系列　受信系列　符号系列　情報源系列

情報源の符号化の目標：(1) 1情報源記号当たりの平均符号長ができるだけ小さいこと

平均符号長をどこまで小さくできるか。 (2) 装置化の簡単さ (3) 復号による遅延が小さいこと

(4) (optional, 非可逆符号化) ひずみ(元の通報との違い)を平均として，許容される値以下に抑える

通信路符号化の目標：(1) 付加した冗長性が信頼性向上にできるだけ有効に用いられる

限界を知ること： (2) 復号後の記号誤り率をある値以下に抑える

付加すべき冗長度をどこまで小さくできるか　(3) 装置化の簡単さ　(4) 復号による遅延が小さいこと

第3章 情報源と通信路のモデル

3.1 情報源のモデル　3.1.1 情報源の統計的表現

情報源アルファベット：$A=\{a_1,a_2,\dots,a_M\}$ 時点iの情報源の出力をX_i $(i=0,1,2\dots)$で表す。

$P(x_0,x_1,\dots,x_{n-1})=[X_0=x_0, X_1=x_1,\dots,X_{n-1}=x_{n-1}$となる確率$]$

3.1.2 記憶のない定常情報源 (定常：同一の確率分布に従う)　さいころ(投げる)

出力$X_0,X_1,\dots,X_{n-1}$の結合確率分布は，任意のnについて $P(x_0,\dots x_{n-1})=\prod_{i=0}^{n-1}P_X(x_i)$

3.1.3 定常情報源とエルゴード情報源 (エルゴード，遍歴)

定常情報源：$P_{X_0X_1\cdots X_{n-1}}(x_0,x_1,\dots,x_{n-1})=P_{X_iX_{i+1}\cdots X_{i+n-1}}(x_0,x_1,\dots,x_{n-1})$ が成立する

エルゴード情報源 (ergodic information sourse)：それが発生する十分長い任意の系列に，その情報源の統計的性質が完全に現れている定常情報源である。

無数の同じ情報源の上で考えた割合が確率である。

$\overline{f(x)}=\sum_{x\in A}f(x)P_X(x)$　A：情報源アルファベット，無数の同じ情報源についての平均であり f(x)の集合平均

一つの情報源からの出力系列$x_0x_1x_2\cdots$について $\langle f(x)\rangle=\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f(x_i)$ 時間平均

エルゴード情報源：$\overline{f(x)}=\langle f(x)\rangle$

記憶のない定常情報源はエルゴード情報源となる

定常情報源であってもエルゴード情報源でないものもいくらでも存在する。

例：$\frac{1}{2}$の確率で0だけからなる系列，$\frac{1}{2}$の確率で1だけからなる系列。 $P_X(1)=\frac{1}{2}$　$\overline{f(x)}=\frac{1}{2}$　$\langle f(x)\rangle=1$ or 0

3.2 マルコフ情報源

あるnがn>mを満すとき，任意の時点iについて，その直前のn個の出力$X_{i-1},\dots X_{i-n}$で条件を付けたX_iの条件付確率分布が直前のm個の出力$X_{i-1},X_{i-2},\dots X_{i-m}$だけで条件を付けた$X_i$の条件付確率と一致するとき，すなわち $P_{X_i|X_{i-1}\cdots X_{i-n}}(x_i|x_{i-1},\dots x_{i-n})=P_{X_i|X_{i-1}\cdots X_{i-m}}(x_i|x_{i-1},\dots x_{i-m})$

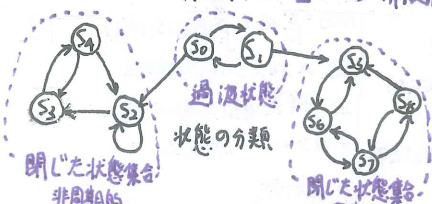
m重マルコフ情報源 (その直前のm個の出力だけから)

1重マルコフ情報源＝単純マルコフ情報源. m重マルコフ2元情報源が2^m個の状態を持ち.

状態図： S_0からS_1への矢印に付けられている$1/p$ (S_0=直前の出力0とし)

(1) 直前の出力が0 (2) 確率pで出力1を発生し (3) 状態S_1に遷移する

一般化されたマルコフ情報源 ≠ m重マルコフ情報源 (例：0だけからなる出力系列が続く場合)



過渡状態：十分時間が経過すればこれらの状態から抜け出してしまう

状態の分類

非周期的：ある時間が経過した後の任意の時点において，どの状態にある確率も0でない。

周期的：ある周期的な時点においてのみ現れ得るという場合

既約マルコフ情報源：一つの閉じた状態集合だけからなるマルコフ情報源である。

非周期的マルコフ情報源を正規マルコフ情報源という

$\Pi=\begin{bmatrix}P_{00}&\cdots&P_{0,N-1}\\ \vdots&&\vdots\\ P_{N-1,0}&\cdots&P_{N-1,N-1}\end{bmatrix}$ 遷移確率行列　各行が現在の状態に対応し 各列が次の状態に対応する

各列の和は1になるとは限らない (列和不一定)

状態S_iから出発し，t時点後にS_jに到達する確率を$P_{ij}^{(t)}$で表そう：明らかに

(1) $P_{ij}^{(1)}=P_{ij}$　(2) $P_{ij}^{(t)}=\sum_{k=0}^{N-1}P_{ik}^{(t-1)}P_{kj}$　(3) $P_{ij}^{(t)}=\Pi^t$の(i,j)要素となることが直ちに導ける

正規マルコフ情報源は，ある正整数に対し，$P_{ij}^{(t_0)}>0$. 正規マルコフ情報源では，$t\to\infty$と移るとき $P_{ij}^{(t)}$ は i に無関係な値に収束する

$\lim_{t\to\infty}\Pi^t=U=\begin{bmatrix}u_0&\cdots&u_{N-1}\\ \vdots&&\vdots\\ u_0&\cdots&u_{N-1}\end{bmatrix}$ n×n

$\lim_{t\to\infty}P_{ij}^{(t)}=u_j\ (j=0,1,\dots N-1)$

時点tにおいて状態S_jにいる確率を$w_j^{(t)}$で表わし，$w_t=(w_0^{(t)},w_1^{(t)},\dots w_{N-1}^{(t)})$ というN次元ベクトルを定義する。状態確率分布ベクトル　$w_j^{(t)}=\sum_{i=0}^{N-1}w_i^{(t-1)}P_{ij}$　w_0は時期0における状態分布である 初期分布

$w_t=w_{t-1}\Pi \Rightarrow w_t=w_0\Pi^t$

極限分布：正規マルコフ情報源では，$\lim_{t\to\infty}w_t=w_0\lim_{t\to\infty}\Pi^t=w_0U=u$

定常分布：$w=(w_0,w_1,\cdots w_{N-1})$　$w_0+w_1+\cdots+w_{N-1}=1$　$w\Pi=w$ ⇐ 唯一つ存在し

正規マルコフ情報源は，十分な時間の後には，エルゴード情報源とみることができる。極限分布と一致すること

(しかし，周期的既約情報源では，定常分布以外の初期分布を与えると，十分時間が経過した後も一般には定常情報源とはならない)

3.3 通信路のモデル

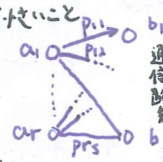
入力系列 $X_0X_1\cdots X_{n-1}$ $\xrightarrow{対応}$ 出力系列 $Y_0Y_1\cdots Y_{n-1}$　$P_{Y_0\cdots Y_{n-1}|X_0\cdots X_{n-1}}(y_0\cdots y_{n-1}|x_0\cdots x_{n-1})$

記憶のない定常通信路

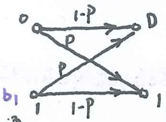
入力 $A=\{a_1,a_2\cdots a_r\}$，出力 $B=\{b_1,b_2\cdots b_s\}$　$P_{ij}=P_{Y|X}(b_j|a_i)$

$$\begin{array}{cc} & \begin{array}{cccc}b_1 & b_2 & & b_s\end{array}\\ T= & \begin{bmatrix}P_{11}&P_{12}&\cdots&P_{1s}\\ P_{21}&P_{22}&\cdots&P_{2s}\\ \vdots&&&\vdots\\ P_{r1}&P_{r2}&\cdots&P_{rs}\end{bmatrix}\begin{array}{c}a_1\\a_2\\ \\a_r\end{array}\end{array}$$

Tは通信路行列. 記憶のない定常通信路の統計的性質は完成に定まることになる

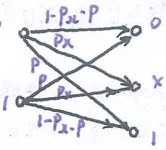


通信路線図



$$\begin{array}{cc} & \begin{array}{cc}0 & 1\end{array}\\ T= & \begin{bmatrix}1-p & p\\ p & 1-p\end{bmatrix}\begin{array}{c}0\\1\end{array}\end{array}$$

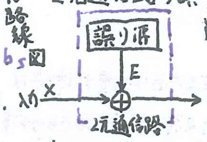
2元対称通信路 BSC　binary symmetric channel



$$\begin{array}{cc} & \begin{array}{ccc}0 & x & 1\end{array}\\ T= & \begin{bmatrix}1-P_x-P & P_x & P\\ P & P_x & 1-P_x-P\end{bmatrix}\begin{array}{c}0\\1\end{array}\end{array}$$

2元対称消失通信路

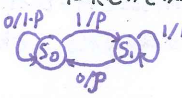
2元通信路の誤りによる表現



$Y=X\oplus E$　Eは，誤り源から各時点に一つずつ発生する。

誤り源の統計的性質は入力とは無関係　この2元通信路の統計的性質は誤り源の統計的性質だけで決定できる　加法的通信路

random error.　p=bit error rate



$\Pi=\begin{bmatrix}1-p & p\\ p & 1-p\end{bmatrix}$

3.3.4 burst error通信路.

誤りが一度生じると，その後しばらくの間誤りが続けて生じることがある.　$w_0=\frac{p}{P+p}$　$w_1=\frac{P}{P+p}$

$P_E(1)=w_0P+w_1(1-p)=w_1=\frac{P}{P+p}$　$P_E(1)\ll1$, $P<1-P$

$w_0P+w_1(1-p)=w_1$

$P-w_1P+w_1-w_1p=w_1$

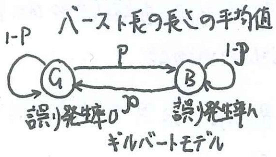
$P=w_1(P+p)$

$w_1=\frac{P}{P+p}$

誤り系列における1の連続(1のラン)を，一つ任意に取り出す

この長さがLとなる確率を$P_B(L)$とすれば $P_B(L)=(1-p)^{L-1}p$

バースト長の長さの平均値 $\overline{L}=\sum_{L=1}^{\infty}LP_B(L)=\frac{1}{p}$



第3章：

1. 情報源モデル

1.1 記憶のない定常情報源　無数の同じ情報源について平均

定常情報源：任何時刻概率不变　エルゴード情報源：$\overline{f(x)}=\langle f(x)\rangle$　集合　時間

定常情報源 ≠ エルゴード情報源

1.2 m重マルコフ情報源　只与之前m个出力有关　2^m個の状態を持ち　状態図

状態の分類 {過渡状態 / 閉じた状態集合(周期的/非周期的)　既約マルコフ情報源(一つの閉じた状態集合だけ)

遷移確率行列 $\Pi=\begin{bmatrix}P_{00}&\cdots&P_{0,N-1}\\ \vdots&&\vdots\\ P_{N-1,0}&\cdots&P_{N-1,N-1}\end{bmatrix}$　時点t 状態S_jの確率：$w_j^{(t)}$

極限分布：$\lim_{t\to\infty}w_t=w_0\lim_{t\to\infty}\Pi^t=w_0U=u$　ベクトル $w_t=(w_0^{(t)},w_1^{(t)}\cdots w_{N-1}^{(t)})$ 状態確率分布ベクトル

定常分布：$w=(w_0,w_1,\cdots w_{N-1})$　$w\Pi=w$ 唯一存在し

2 通信路モデル

2.1 記憶のない定常情報源　把这个当作情報源处理　誤り源 E　入力X ⊕ 出力Y

2元対称通信路 BSC, binary symmetric channel

2元対称消失通信路　2元通信路の誤りによる表現

2.2 バースト通信路

$P_E(1)=w_0P+w_1(1-p)=w_1=\frac{P}{P+p}$ $\xLeftarrow{描述}$ ギルバートモデル，フリッチマンモデル.

重要問題：演習問題3.2(c)