２０１９年5月6日（川端）

データ圧縮基礎資料

5　情報源

5.1　情報源と定常性

同一分布空間 $\mathcal{X}$ 上の確率変数列 $\{X_l\}_{l\in\mathcal{N}}$ を情報源あるいは確率過程という.

$$\begin{aligned} & p_{X_1X_2\ldots X_k}(x_1,\ldots,x_k) \\ = \;& p_{X_1}(x_1)p_{X_2|X_1}(x_2|x_1)\ldots \\ & p_{X_k|X_1X_2\ldots X_{k-1}}(x_k|x_1\ldots x_{k-1}) \end{aligned}$$

が成り立つ. 以前の例からも類推できるがこれも

$$p(x_1,\ldots,x_k)=p(x_1)p(x_2|x_1)\ldots p(x_k|x_1\ldots x_{k-1})$$

と簡略して記述できる.

あらゆる $k\geq 1$ について, 組 $\{X_l\}_{1\leq l\leq k}$ と組 $\{X_l\}_{2\leq l\leq k+1}$ が同じ同時確率分布を持つとき, 情報源は定常であるという．定常情報源ではあらゆる $m\geq 1, k\geq 1$ について組 $\{X_l\}_{1\leq l\leq k}$ と組 $\{X_l\}_{m\leq l\leq k+m-1}$ が同じ同時確率分布を持つことが帰結される. またこのことから定常情報源からとった任意のサンプルがなす確率変数の同時確率分布は任意時間シフトしても同じ同時確率分布を持つことが分かる.

通信では情報源の確率分布は外部から与えられることは滅多になく情報源からのサンプルにより推定しなければ実用的なシステムにはならない. 定常情報源についてその確率分布が出力列から推定可能になっていることが望まれるが，この条件を満足する定常情報源はエルゴート的であるという. 任意の定常情報源は定常エルゴート情報源の混合（確率的重ね合わせ）として表現できることが知られている. この知識に基づけば定常情報源からの出力はある定常エルゴート情報源の出力であると考えて構わないことになる，つまり情報源が定常であれば，符号化装置としては任意の定常エルゴート情報源が処理できれば十分であるといえるかもしれない.（現実に現れる情報源は都合よく定常情報源になっている保証はない．定常でない情報源は非定常であるという．非定常な情報源であっても処理は区分的に定常エルゴート的な情報源を仮定して設計されることが多い. ）

定常エルゴート情報源を推定するには無限の長さのサンプル列が必要であるがこれは現実には期待できない．そこでまず情報源のアルファベット $\mathcal{X}$ は有限であると仮定し，情報源系列の長さ n も有限であるとする．情報源は $(\sharp\mathcal{X})^n-1$ 個のパラメータで指定される．定常情報源の場合にはこれが $(\sharp\mathcal{X}-1)(\sharp\mathcal{X})^{n-1}$ 個に少なくなる（定常エルゴート情報源でも同じである）.

このパラメータ数が n に対して指数的であることからも分かるように，定常エルゴート情報源はサンプルから推定するには表現能力が高すぎる．実際に用いるには確率分布のクラスを限定しパラメータ数を小さく抑えることを念頭に置く必要があるであろう．例えば次節以降で述べる定常マルコフ連鎖は定常情報源の近似を行うときに役に立つ工学的に重要な情報源である.

5.2　独立情報源とマルコフ連鎖

情報源 $\{X_l\}_{l\in\mathcal{N}}$ の同時確率分布があらゆる $k\geq 1$ について

$$p_{X_1X_2\ldots X_k}(x_1,\ldots,x_k)=p_{X_1}(x_1)p_{X_2}(x_2)\ldots p_{X_k}(x_k)$$

を満たすとき, 独立であるという. この関係も

$$p(x_1,\ldots,x_k)=p(x_1)p(x_2)\ldots p(x_k)$$

と略記できる．長さ n の独立情報源のパラメータ数は $n(\sharp\mathcal{X}-1)$ である．定常独立情報源は特に重要な情報源であり，パラメータ数は $\sharp\mathcal{X}-1$ である．

情報源 $\{X_l\}_{1\le l\le n}$ の同時確率分布が

$$\begin{aligned} & p_{X_1X_2\ldots X_k}(x_1,\ldots,x_n) \\ & \quad = \quad p_{X_1}(x_1)p_{X_2|X_1}(x_2|x_1)\ldots p_{X_n|X_{n-1}}(x_n|x_{n-1}) \end{aligned}$$

を満たすとき，確率変数列 $X_1,X_2,\ldots,X_n$ はこの順にマルコフ連鎖をなすといい，$X_1\to X_2\to\ldots\to X_n$ と略記する．$\mathcal{X}$ はマルコフ連鎖の状態空間ともいいその要素は状態と呼ぶ．マルコフ連鎖は遷移する状態の列であるといえる．状態空間 $\mathcal{X}$ 上の長さが n のマルコフ連鎖のパラメータ数は初期状態の確率分布に対して $\sharp\mathcal{X}-1$ その後の $n-1$ 回の遷移に対して各々$\sharp\mathcal{X}(\sharp\mathcal{X}-1)$ ある．

$\mathcal{Y}$ 上の情報源 $\{Y_l\}$ と状態空間 $\mathcal{X}$ 上のマルコフ連鎖 $\{X_l\}$ があり，関数 $f:\mathcal{X}\to\mathcal{Y}$ を用いて $Y_l=f(X_l)$ と書けるとき，$\{Y_l\}$ は（$\mathcal{Y}$ 上の）マルコフ情報源という．マルコフ情報源は初期状態が与えられたとき情報源系列から状態の系列が復元できるとき，ユニフィラ（糸手繰り可能）と呼ばれる．ユニフィラでない情報源は隠れマルコフ情報源であると呼ばれる．

ユニフィラマルコフ情報源の例として，k 次マルコフ情報源 $\{Y_l\}_{l=1}^n$ がある．これは状態空間 $\mathcal{X}=\mathcal{Y}^k$ をもつ $\mathcal{Y}$ 上のマルコフ情報源であり，$Y_1^k:=(Y_1,Y_2,\ldots,Y_k)$ を初期状態とし，$Y_1^k\to Y_2^{(k+1)}\to Y_3^{(k+2)}\to Y_4^{(k+3)}\to\ldots\to Y_{(n-k+1)}^n$ を遷移とする．関数 $f:\mathcal{X}\to\mathcal{Y}$ は，$f(a_1,a_2,\ldots,a_k)=a_k$ で与えられる．パラメータ数は，初期状態の分布に対する $\sharp\mathcal{Y}^k-1$ 個と，$n-1$ 回の遷移における各 $\sharp\mathcal{Y}^k(\sharp\mathcal{Y}-1)$ 個の総計である．$k=1$ 次マルコフ情報源は $\mathcal{Y}$ 上のマルコフ連鎖に他ならない．また，0 次マルコフ情報源は，独立情報源に他ならない．

マルコフ連鎖についても定常性を考えることができるが，次々節で述べる．

5.3 マルコフ連鎖とデータ処理不等式

補題 1 $X_0\to X_1\to\cdots\to X_n \Leftrightarrow X_n\to X_{n-1}\to X_{n-2}\to\cdots\to X_0$ が成り立つ．従って $\to$ を単に $-$ で置き換えて，左辺などを $X_0-X_1-\cdots-X_n$ とかいてもよい．

証明： 3 つの確率変数 $X\to Y\to Z$ の場合についてのみ証明する．

$$\begin{aligned} p(x,y,z) &= p(x)p(y|x)p(z|y) \\ &= p(x)\frac{p(y)p(x|y)}{p(x)}\cdot\frac{p(z)p(y|z)}{p(y)} \\ &\qquad \text{(Bayes の定理を 2 回利用)} \\ &= p(z)p(y|z)p(x|y) \end{aligned}$$

例： 確率変数 $X\to Y\to Z$ について，

i) $Z=g(Y)$ なら，$X-Y-Z$

ii) $X\to Y\to Z$ の定義で X の確率分布には何の制約もないことに注意する．実際 X は確率変数でなくてもよく，一般のパラメータ x であってもよい．

iii) $X-Y-Z\Leftrightarrow p(x,y,z)=p(y)p(x|y)p(z|y)$．つまりこの条件は $Y=y$ が与えられたとき，X と Z が条件付に独立でることを意味しており，$I(X;Z|Y)=0$ が成立することと同値である．

定理 1 (データ処理不等式) $X-Y-Z\Rightarrow I(X;Y)\ge I(X;Z)$

証明： 相互情報量 $I(X;(Y,Z))$ を省略して $I(X;Y,Z)$ とかこう．この情報量は次の 2 通りに変形できる．

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

ここで，最初の等号の右辺の第二項は零であり，第二の等号の右辺の第二項は非負である．従って，定理が成り立つ．この変形の副産物として次の系もなり立つ．

定義 1 $I(X;Y;Z) := I(X;Y) - I(X;Y|Z)$ と定義する．この式は X,Y,Z について対称である．

系 1 $X-Y-Z$ ならば $I(X;Y|Z) \leq I(X;Y)$ が成り立つ．すなわち，$I(X;Y;Z) \geq 0$ が成り立つ．

5.4 定常独立情報源と定常マルコフ連鎖

定常独立情報源は別名ベルヌーイ情報源, あるいは i.i.d（独立同一分布）情報源とよばれる. パラメータ数が $\mathcal{X}$ のサイズマイナス 1 であり, 取り扱いやすく基本的である. マルコフ連鎖は状態遷移確率 $p_{X_l|X_{l-1}}(b|a)$ が時刻 l によらないならば, 定常であるという. マルコフ連鎖の確率分布は初期状態の分布 $p_{X_1}(a)$ とともに決まる. $\mathcal{X} = \mathcal{Y}$ とみなすとマルコフ連鎖はマルコフ情報源となるが, この情報源が定常であるためには, 初期状態の分布に制約が生じる. この必要な制約条件は例えば $p_{X_1}(a) = p_{X_2}(a), a \in \mathcal{X}$ である. 実は, この方程式は必要かつ十分な条件を与えることが簡単に見て取れる. その意味で, この方程式を状態方程式という. またそれを満たす初期状態の確率分布を定常状態確率分布という. 少し具体的に, 方程式を記述し解いてみよう. 状態集合 $\{0,1,2\}$ 上に定常マルコフ連鎖が状態遷移確率

$$
\begin{pmatrix}
p(0|0) & p(1|0) & p(2|0) \\
p(0|1) & p(1|1) & p(2|1) \\
p(0|2) & p(1|2) & p(2|2)
\end{pmatrix}
=
\begin{pmatrix}
\bar{\alpha} & \alpha & 0 \\
\beta & 0 & \bar{\beta} \\
1 & 0 & 0
\end{pmatrix}
$$

で与えられている. ただし $0 < \alpha, \beta < 1$ とする. この行列を P と表す. 定常状態の確率分布を行ベクトル $q = (q_0, q_1, q_2)$ で表すと. 状態方程式は

$$
qP = q
$$

と書ける. 最後の 2 要素と, 全確率性に対応する方程式は具体的に

$$
\begin{aligned}
q_1 &= \alpha q_0 \\
q_2 &= \bar{\beta} q_1 \\
q_0 + q_1 + q_2 &= 1
\end{aligned}
$$

となる. 従って,

$$
(q_0, q_1, q_2) = \frac{1}{1 + \alpha + \alpha\bar{\beta}} \times (1, \alpha, \alpha\bar{\beta})
$$

となる.

5.5 定常情報源のエントロピーレート

定義 2 極限

$$\lim_{n\to\infty}\frac{1}{n}H(X_1,\ldots,X_n)$$

が存在するときこれを情報源 $\{X_l\}_{l\in\mathcal{N}}$ のエントロピーあるいはエントロピーレートという．この値を H_∞ とかく．

次の定理が成り立つ.

定理 2 $\{X_l\}_{l\in\mathcal{Z}}$ が定常であり, $H(X_1)<\infty$ のとき, H_∞ が存在しその値は

$$H_\infty=\lim_{n\to\infty}H(X_n|X_{n-1},X_{n-2},\ldots,X_1)$$

で与えられる.

証明:

補題 2 有限な極限が存在する数列に対しては，その算術平均のなす数列も同じ極限を有する.

証明：数列 $\{a_n\}$ が $\lim_n a_n=a<\infty$ とするとき, $\forall\varepsilon>0,\exists n_0(\varepsilon),\forall n\geq n_0(\varepsilon)$ に対して，

$$|a_n-a|\leq\varepsilon$$

が成り立つ．従って,

$$\begin{aligned}|\frac{1}{n}\sum_{l=1}^{n}a_l-a| &\leq \frac{1}{n}\sum_{l=1}^{n}|a_l-a|\\ &\leq \frac{1}{n}\sum_{l=1}^{n_0(\varepsilon)}|a_l-a|+\frac{n-n_0(\varepsilon)}{n}\varepsilon.\end{aligned}$$

ここで，n を十分大きくすると，右辺は，2ε で上から上界される.

補題 3 定常情報源 $\{X_n\}$ について，$H(X_1)<\infty$ ならば，$H_\infty(X)$ が存在し，

$$H_\infty(X)=\lim_{n\to\infty}H(X_n|X_1,\ldots,X_{n-1})$$

がなりたつ．また，

$$H_\infty(X)\leq H(X_n|X_1,\ldots,X_{n-1})$$

が成り立つ.

証明：エントロピーに関するチェインルールより，

$$H(X^n)=\sum_{l=1}^{n}H(X_l|X_1,\ldots,X_{l-1})$$

が成り立つ.

$$\begin{aligned}a_l &:= H(X_l|X_1,\ldots,X_{l-1})\\ &= H(X_{l+1}|X_2,\ldots,X_l) \quad \text{（定常性）}\\ &\geq H(X_{l+1}|X_1,\ldots,X_l) \quad \text{（条件の追加）}\\ &= a_{l+1}\end{aligned}$$

より，数列 $\{a_l\}$ は，初項 $a_1=H(X_1)$ が有限であり，単調減少，かつ非負である．したがって，$\lim_{n\to\infty}a_n$ が有限の値として存在する．定理は補題をこの場合に解釈することにより得られる.

5.6 情報源符号化定理

情報源 $\{X_l\}_{l\in\mathcal{Z}}$ からの固定の長さのブロック $\{X_l\}_{1\le l\le n}$ の全てををメッセージの集合とみて, これに対する一意復号可能な符号の構成を考えてみよう. 期待符号長の下限は $H(X_1,\ldots,X_n)$ で与えられる. また, 上限は**??**節で示されたように $H(X_1,\ldots,X_n)+1$ である. 従って, 次の事実にいたる.

定理 3 情報源が定常ならば, 情報源 1 記号あたりの期待符号長の最適値は以下の不等式を満足する. 即ち, 任意の正数 ε に対して, 十分大きなすべての n について, 確率変数 $(X_1,\ldots,X_n)$ に対する一意復号可能な符号

$$c:\mathcal{X}^n\to\{0,1\}^*$$

が存在しその一記号当たりの平均符号長 $\frac{1}{n}El(X_1,\ldots,X_n)$ の最適値は

$$H_\infty\le\frac{1}{n}El(X_1,\ldots,X_n)\le H_\infty+\varepsilon$$

を満足する.

5.7 定常マルコフ情報源に対するエントロピーレート

前節で情報源のエントロピーレートの重要性が理解できたので定常マルコフ連鎖 $\{X_l\}_{l\in\mathcal{N}}$ についてそれを求めてみよう. X_1 は定常分布に従うとする.

$$\begin{aligned}H(X_1,\ldots,X_n)=&\\&H(X_1)+\sum_{l=2}^{n}H(X_l|X_1,\ldots,X_{l-1})\\=&\ H(X_1)+\sum_{l=2}^{n}H(X_l|X_{l-1})\\=&\ H(X_1)+(n-1)H(X_2|X_{1-1})\end{aligned}$$

が成り立つ. 第一の等式はチェイン規則と呼ばれるものである. 第二の等式はマルコフ性による. 第 3 の等式は定常性による. 即ち, $H_\infty=H(X_2|X_1)$ となる. 5.4 節の例では,

$$\begin{aligned}H_\infty &= H(X_2|X_1)\\&=\sum_{i=0}^{2}q(i)H(p(\cdot|i))\\&=q_0H(\alpha)+q_1H(\beta)\\&=q_0H(\alpha)+q_1H(\beta)\\&=\frac{H(\alpha)+\alpha H(\beta)}{1+\alpha+\alpha\bar{\beta}}\end{aligned}\tag{1}$$

ただし, 式では二進エントロピー関数

$$H(r)=-r\log r-\bar{r}\log\bar{r}$$

を用いて表している.