目次

第 1 章

情報の数量的認識

情報とは，「確率分布の変化」として形式化する．そこで，確率分布の変化の大きさを KL-変分などで測り，この値を「情報量」として再定義する．確率的な現象が確定的な現象となるとき，確定に必要な分だけの情報の流入が起こった，とする．また，「複製可能なものが情報である」という議論もあり得る (Feller,[4]).

1.1 エントロピー

測度空間 $(\Omega, \mathcal{F})$ 上の不変量を考える．

(1) 各 $P \in P(\Omega)$ に対して**情報量**なる測度 $I_P : \mathcal{F} \to [0,\infty]$ が定まる：$I_- : P(\Omega) \to M(\Omega; [0,\infty])$.
(2) P の ζ が誘導する情報量 $I_P(\omega|\zeta)$ の平均を**分割 ζ のエントロピー**という：$H_P(\zeta) : \Delta \times P(\Omega) \to [0,\infty]$.
(3) Δ は有向集合であるから，その極限を取ると**確率分布のエントロピー** $H_v : P(\Omega) \to [0,\infty]$ が定まる．
(4) 情報量 $I_Q(\omega|\zeta)$ の別の確率分布 P についての平均を**交差エントロピー** $H_{P,Q}$ という．
(5) 情報量の差 $I_Q(\omega|\zeta) - I_P(\omega|\zeta)$ の P についての平均を**相対エントロピー** $D(P|Q) = H_{P,Q} - H_P$ という．

こうしてエントロピーの概念を得たら，情報量から離陸する．

(1) エントロピー $H_P(\xi \cap \pi_\zeta(\omega))$ の P-平均 $H_P(\xi|\zeta) : \Delta^2 \to [0,\infty]$ を**条件付きエントロピー**という．
(2) ある確率変数 Y で X を条件付けた際のエントロピーの変化分 $I_{(X,Y)} = H(X) - H(X|Y)$ を 2 つの観測 X, Y の**相互情報量**という．

記法 1.1.1. 双対的な対象を意識して，エントロピー H，情報量 I については引数として分割・確率変数を取り，右下添字として分布を取る．逆に変分は引数として分布を取る．このように，エントロピーは分割と分布の 2 つの双対的な対象の組に対して定まるペアリングのようなものであることを意識しないと大局を見失う．

1.1.1 分割のエントロピー

事象の積の情報量は和になってほしいならば，情報量の定義には対数が必要である．底は，情報理論では 2 が，統計力学では特定の物理定数である．基本的にエントロピーは，Lebesgue 測度からみた KL-変分 $H_\mu = H_m - D(\mu|m)$ であるが，台集合が離散でない場合 H_m が発散するので，この項を取り除いたものを連続エントロピーという．ただの平行移動なので，性質そのものは本質的には同じである．

定義 1.1.2 ((self) information, entropy). $(\Omega, \mathcal{F}, P)$ を確率空間とする．

(1) $I(p) := -\log p$ で定まる関数 $I : [0,1] \to [0,\infty]$ について，合成によって定まる事象上の関数

$$I_P := I \circ P : \mathcal{F} \to [0,1] \to [0,\infty]$$

を**事象の (自己) 情報量**という．

(2) **分割 ζ の情報量** $I_P(-;\zeta):\Omega\to[0,\infty]$ とは，合成 $I_P\circ\pi_\zeta:\Omega\to\Omega/\zeta\to[0,\infty]$ をいう：

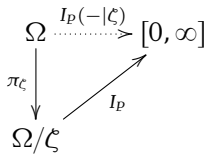
$$I_P(\omega;\zeta)=E_P[I_P|\zeta]=-\sum_{A\in\zeta}1_A\log P[A].$$

(3) **分割 ζ のエントロピー**または**平均情報量**とは，ζ の情報量 $I_P(-|\zeta):\Omega\to[0,\infty]$ の平均

$$H_P(\zeta):=E[I_P\circ\pi_\zeta]=\int_\Omega I_P(\omega;\zeta)P(d\omega)=-\int_X\log_2 P(\pi_\zeta(\omega))P(d\omega).$$

をいう．

要諦 1.1.3. 「情報量」なる概念が分割 $\zeta\in\Delta$ を通じてしか X 上に定まらないことが見通しが悪い原因である．すなわち，$I_P(-|\zeta)=E[I_P|\zeta](-)$ というように，可測分割 ζ を通じてしか Ω 上の関数として得られない：



ただし，次の命題の (5) から，可測分割としては有限なもののみを考えれば十分であることが分かる．積分で書いたが，次が成り立っている：$(A_i)_{i\in\mathbb{N}}$ を ξ の元で正の測度をもつもの，$A:=\cup_{i\in\mathbb{N}}A_i$ をその合併とする．

$$H(\xi):=\begin{cases}-\sum_{i\in\mathbb{N}}P[A_i]\log P[A_i] & P[A]=1,\\ +\infty & P[A]<1.\end{cases}$$

すなわち，P が連続分布であるときは，エントロピーは発散する．これは $H(\mu)=H(U([n]))-D(\mu|U([n]))$ であるが，KL-変分が絶対連続な分布にしか定まらないことにつながる 1.1.26．そこで，このような場合は連続エントロピーで測るが，これは Lebesgue 測度 m について $h(\mu)=-D(\mu|m)$ としたものである．

命題 1.1.4 (対数関数の特徴付け). 次の条件を満たす関数 $I:[0,1]\to\mathbb{R}_+$ は定数倍を除いて $I=-\log$ に限る：

$$\forall_{p,q\in[0,1]}\ f(pq)=f(p)+f(q).$$

要諦 1.1.5. 独立な事象 $A,B\in\mathcal{F}$ に対して，$I(P[A\cap B])=I(P[A])+I(P[B])$ が成り立つ関数は対数に限り，あとは底の問題になる．

定義 1.1.6 (continuous entropy). $\mu,\nu\in P(\mathbb{R})$ は絶対連続で，それぞれ密度 p,q を持つとする．

(1) 次で定まる $h:P(\mathbb{R})\to\overline{\mathbb{R}}$ を**絶対連続分布のエントロピー**という．[†1]

$$h_\mu:=-\int_{\mathbb{R}}p(x)\log p(x)dx.$$

(2) 次で定まる $h:P(\mathbb{R})\times P(\mathbb{R})\to\overline{\mathbb{R}}$ を**絶対連続分布の交差エントロピー**という．

$$h_{\mu,\nu}:=-\int_{\mathbb{R}}p(x)\log q(x)dx.$$

命題 1.1.7 . $X,Y\in P(\mathbb{R})$ を絶対連続とする．

(1) X_n,Y_n を X,Y に収束する単関数とする．このとき，

$$\lim_{n\to\infty}(H(X_n)-H(Y_n))=h(X)-h(Y).$$

(2) $h:P(\mathbb{R})\to\overline{\mathbb{R}}$ は極限も含めれば全射である．

(3) 交差エントロピーはエントロピーを最小値とする：$h_{\mu,\nu}\geqslant h_\mu$

$$h_\mu=-\int_{\mathbb{R}}p(x)\log p(x)dx\leqslant-\int_{\mathbb{R}}p(x)\log q(x)dx=h_{\mu,\nu}$$

等号成立条件は $p=q$ a.e.

[†1] これを Shannon は differential entropy と呼んだ．H の記法は Boltzmann による．

[証明].

(2)　$U([0,n])$ の連続エントロピーは $\log n$ となり，これは $\overline{\mathbb{R}}$ への全射である．

■

例 1.1.8. X は $S \subset \mathbb{R}^d$ 上の d 次元，分布族としては 0 次の指数型分布

$$p(x) = Ae^{-\varphi(x)}1_S$$

に従うとする．このエントロピーは $h(X) = -\log A + E[\varphi(X)]$ となるが，これはある $\varphi \in L(\mathbb{R})$ を通じて得られる条件

$$\{X \in L(\Omega) \mid E[\varphi(X)] = m\}$$

のうち，エントロピーが最大のものである．

1.1.2　エントロピーの性質

$H : \Delta \to \mathbb{R}_+$ は正な線型汎函数に非常に近い．Δ の束の構造を保つ「連続」写像であること，ネットの極限として特徴づけられること，これらを含めてみても，明らかに積分の変種としての性質を持つ．

記法 1.1.9. Δ を Ω の可測分割の全体とする，

$$\Delta_{<\infty} := \bigcup_{n\in\mathbb{N}} \mathcal{L}(\Omega; n).$$

を有限なものの全体のなす有向集合とする．

命題 1.1.10 (エントロピーの性質). $H(\Delta)$ は再び束の良い構造を持ち，「連続」である：

(1) 正：$H(\xi) \geqslant 0$．等号成立は ξ が自明な分割であるとき．
(2) $\xi \leqslant \eta \Rightarrow H(\xi) \leqslant H(\eta)$．なお，$\xi \leqslant \eta \Rightarrow H(\xi) = H(\eta) < \infty$ ならば $\xi = \eta$.
(3) $\xi_n \nearrow \xi \Rightarrow H(\xi_n) \nearrow H(\xi)$.
(4) $\xi_n \searrow \xi \wedge H(\xi_1) < \infty \Rightarrow H(\xi_n) \searrow H(\xi)$.
(5) ネットの極限としての特徴付け：$H(\xi) = \sup\{H(\eta) \in \mathbb{R}_+ \mid \xi \geqslant \eta \in \Delta_{<\infty}\}$.
(6) $\xi = \{A_i\}_{i\in[n]}$ のとき，$H(\xi) \leqslant \log n$ で，等号成立は離散一様分布に限る．

[証明].

(5)　任意の可測分割 ξ に対して，これに下から収束する有限分割の列が取れるためである．

■

記法 1.1.11. $\Gamma^n := \partial B_+ \subset \mathbb{R}^n$ を l^1-ノルムに関する単位円周の非負部分とすると，これは $[n]$ 上の確率分布の全体に等しい．

定理 1.1.12 (離散エントロピー関数の特徴付け [6]). 関数 $I : \Gamma^n \to \mathbb{R}$ が次の 6 条件を満たすならば，$I_n = \sum_{i=1}^{n} -p_i \log_2 p_i$ である：

(1) 正規性：$I_2(1/2, 1/2) = 1$.
(2) 展開：$\forall_{n\geqslant 2}\ \forall_{p\in\Gamma^n}\ I_n(p) = I_{n+1}(0,p) = \cdots = I_{n+1}(p_1, \cdots, p_k, 0, p_{k+1}, \cdots, p_n) = \cdots = I_{n+1}(p, 0)$.
(3) 決定性：$I_2(1,0) = I_2(0,1) = 0$.
(4) 強加法性：$\forall_{n,m\geqslant 2}\ \forall_{p\in\Gamma^n, p_i\in\Gamma^m}$ 行列を $P := (p_1\ \cdots\ p_n) \in M_{mn}(\mathbb{R})$ とする．$I_{nm}(p^\top P) = I_n(p) + p^\top (I_m(p_1), \cdots, I_m(p_n))^\top$.
(5) 最大性：$\forall_{n\geqslant 2}\ \forall_{p\in\Gamma^n}\ I_n(p) \leqslant I_n(1/n, \cdots, 1/n)$.
(6) 連続性：$I_2(p, 1-p)$ は $p \in [0,1]$ の連続関数.

1.1.3　分布の交差エントロピー

エントロピー $H_P = H_{P,P}$ は，交差エントロピーを最小化する．Boltzman 機械の学習算譜は交差エントロピーの最小化として特徴付けられている (1985[a])．

[a] David H. Ackley, Geoffrey E. Hilton, Terrence J. Sejnowski. A learning algorithm for Boltzmann machines, Cognitive Science, 9 (1985), 147 – 169.

定義 1.1.13 (cross entropy). $\mu, \nu \in P(\Omega)$ とする．

$$H_{\mu,\nu} := \sup_{\zeta \in \Delta_{<\infty}} E_\mu[I_\nu(-;\zeta)]$$

実は sup は Δ 全域で取っている．よって，

$$H_\mu := H_{\mu,\mu} = \sup_{\xi \in \Delta} H_\mu(\xi)$$

となっており，これを確率空間のエントロピーという．これは $H_\mu(\mathcal{F})$ に一致するためである．

命題 1.1.14 .

(1) $H_{\mu,\nu} \geqslant H_\mu$ で，等号成立は $\nu = \mu$ のときに限る．
(2) $H_{\mu,\nu} \leqslant 2H_\mu$?

1.1.4　条件付きエントロピー

エントロピーでは情報量 $I_P : \Omega/\zeta \to [0,\infty]$ を平均したが，今回は ξ-エントロピーの平均 $H(\xi \cap -) : \Omega/\zeta \to [0,\infty]$ の平均を取る．すると，エントロピーを取る手続きを，条件付き期待値のように一段階挟む「計算途中」のような量である．このときの「縮退具合」によって相互情報量という概念が抽出出来る．

定義 1.1.15 (conditional entropy). ξ, ζ を可測分割とする．

(1) 殆ど至る所の $C \in \zeta$ に対して，ξ は C 上の可測分割 $\xi_C := \xi \cap C$ を定める．それぞれの上でのエントロピーを対応させる写像 $H(\xi_-) : \Omega/\zeta \to [0,\infty]$ は可測である．
(2) この写像の平均

$$H(\xi|\zeta) := \int_{\Omega/\zeta} H(\xi_C) P^{\pi_\zeta}(dC) = \int_\Omega H(\xi_{\pi(\omega)}) P(d\omega).$$

を**条件付きエントロピー**という．

要諦 1.1.16. 同様の記法，$\omega \in A(\omega) \in \xi, \omega \in C(\omega) \in \zeta$ について，$P[\omega;\xi|\zeta] := P_{C(\omega)}[A(\omega)]$ と定めると，

$$H(\xi|\zeta) = -\int_\Omega \log P[\omega;\xi|\zeta] dP$$

と表せる．

命題 1.1.17 . ν を自明な分割とする．

(1) $H(\xi|\nu) = H(\xi)$.
(2) $\eta \leqslant \xi \Rightarrow H(\xi \vee \eta|\zeta) = H(\xi|\zeta)$.
(3) $H(\xi|\zeta) \geqslant 0$. 等号成立条件は $\xi \leqslant \zeta$.
(4) $\xi \leqslant \eta \Rightarrow H(\xi|\zeta) \leqslant H(\eta|\zeta)$. なお，$\xi \leqslant \eta \wedge H(\xi|\zeta) = H(\eta|\zeta) < \infty$ ならば，$\xi \vee \zeta = \eta \vee \zeta$.
(5) 単調な収束列について連続である．
(6) $H(\xi|\zeta) = \sup\{H(\eta|\zeta) \in \mathbb{R}_+ \mid \xi \geqslant \eta \in \Delta_{<\infty}\}$.
(7) $\eta \leqslant \zeta \Rightarrow H(\xi|\eta) \geqslant H(\xi|\zeta)$.

1.1.5　条件付きエントロピーの性質

命題 1.1.18 (連鎖律と三角不等式).

$$H(\xi \vee \eta|\zeta) = H(\xi|\zeta) + H(\eta|\zeta \vee \xi) \leqslant H(\xi|\zeta) + H(\eta|\zeta).$$

系 1.1.19 (確率変数の場合). $X, Y, Z \in L(\Omega)$ について,

(1) $H((X,Y)) = H(X) + H(Y|X)$.
(2) $H((X,Y)|Z) = H(X|Z) + H(Y|(X,Z))$.

要諦 1.1.20. $H((X,Y))$ は**結合エントロピー**という.

命題 1.1.21 . $\zeta_n \nearrow \zeta$ かつ $H(\xi|\zeta_1) < \infty$，または，$\zeta_n \searrow \zeta$ ならば,

$$\lim_{n\to\infty} H(\xi|\zeta_n) = H(\xi|\zeta).$$

[証明]. 条件付き期待値に関する Doob の定理より,

$$\forall_{A\in\mathcal{F}} \lim_{n\to\infty} P[A|\zeta_n;\omega] = P[A|\zeta;\omega] \text{ a.e.}$$

■

命題 1.1.22 (独立性の特徴付け). $\xi, \eta \in \Delta$ について,

(1) $H(\xi) < \infty$ ならば，$\xi \perp\!\!\!\perp \eta \Leftrightarrow H(\xi|\eta) = H(\xi)$.
(2) $H(\xi), H(\eta) < \infty$ ならば，$\xi \perp\!\!\!\perp \eta \Leftrightarrow H(\xi \vee \eta) = H(\xi) + H(\eta)$.

1.1.6　相対エントロピー

条件付きエントロピーの「有限分割のネットの極限」としての特徴付けから,「確率測度の間の相対エントロピー」定義する．これは，統計的多様体上 $P(\Omega)$ の計量と思える.

定義 1.1.23 (relative entropy / KL-divergence). $\mu, \nu \in P(\Omega)$ について，ν に関する μ の**相対エントロピー**または **KL-変分** $D : P(\Omega) \times P(\Omega) \to [0,\infty]$ とは,

$$D(\mu|\nu) := \sup_{\xi\in\Delta_{<\infty}} E_\mu[I_\nu(-|\xi) - I_\mu(-|\xi)] = H_{(\mu,\nu)} - H_\mu.$$

要諦 1.1.24.

(1) パラメトリックなモデルの中で，観測の範囲での KL-変分の最小化問題は，尤度の最大化問題に等しい.
(2) $D(P|Q)$ は，事前分布 Q から事後分布 P に変化した際に，流入した情報量の平均値と解釈出来る．この観点から，KL-変分は**情報利得**ともいう.

命題 1.1.25 (非退化性 (Gibbs)).

(1) $\forall_{\mu,\nu\in M(\Omega;\mathbb{R}_+)} D(\mu|\nu) \geqslant 0$，かつ，等号成立条件は $\mu = \nu$.
(2) 対称性，三角不等式はいずれもそのままの意味では満足しない．これはまず，変分が，距離の自乗に対応するためである.

定理 1.1.26 (積分表示). 有限測度 $\mu, \nu \in M(\Omega;\mathbb{R}_+)$ について,

(1) μ が ν に対して絶対連続でないならば，$D(\mu|\nu) = \infty$.
(2) $\mu \ll \nu$ ならば,

$$D(\mu|\nu) = \int_\Omega \log\left(\frac{d\mu}{d\nu}(\omega)\right)\mu(d\omega) = E_\mu\left[\log\frac{d\mu}{d\nu}\right].$$

1.1.7　相対エントロピーが生成する位相

相対エントロピーは $\mathcal{P} \simeq \Theta$ 上に Riemann 計量を，これについて統計多様体になる．

命題 1.1.27 . 相対エントロピーに関する収束が生成する位相は，全変動ノルムが生成するノルムよりも強い．

定理 1.1.28 (Fisher 計量). C^3-級のパラメトリック模型 $\mathcal{P} = (P_\theta)_{\theta\in\Theta}$ $(\Theta \subset \mathbb{R}^p)$ について，

(1) 1 次の微小項が消える：$\forall_{j\in[p]}\ \left.\dfrac{\partial}{\partial\theta_j}\right|_{\theta=\theta_0} D(P_\theta|P_{\theta_0}) = 0.$

(2) KL-変分の合成 $\theta \mapsto D(P_\theta|P_{\theta_0})$ の Hesse 行列を $G = \left(g_{jk}(\theta_0) := \left.\dfrac{\partial^2}{\partial\theta_j\partial\theta_k}\right|_{\theta=\theta_0} D(P_\theta|P_{\theta_0})\right)$ とする：

$$D(P_\theta|P_{\theta_0}) = \frac{1}{2}\Delta\theta_j\Delta\theta_k g_{jk}(\theta_0) + o(|\theta-\theta_0|^2) \qquad \Delta\theta_j := (\theta-\theta_0)_j.$$

このとき，G は半正定値で，Θ 上に Riemann 計量を定める．

1.1.8　条件付き相対エントロピー

KL-変分が最大になるのは 2 つの実験が独立なときで，$H_{(P_2|P_1)} \leqslant H[P_2]$ で，等号成立は 2 つの実験が独立なときに限る．

定義 1.1.29 (条件付き相対エントロピー).

命題 1.1.30 .

$$D(P|Q) \leqslant H_P$$

1.1.9　相互情報量

2 つの確率変数が，独立であった場合と比べて条件付きエントロピーがどれほど小さくなっているかによって，情報量の概念が一般化出来る．そのアイデアは $X \perp\!\!\!\perp Y \Rightarrow H(X) = H(X|Y)$ より 1.1.22，$I(X,Y) := H(X) - H_{(X|Y)}$ と取ることが出来る．

定義 1.1.31 (mutual information). 2 つの分割 $\xi, \zeta \in \Delta$ の**相互情報量** $I_{(-,-)} : L(\Omega) \times L(\Omega) \to [0,\infty]$ を

$$I_{(X,Y)} := D(P^{(X,Y)}|P^X \times P^Y).$$

要諦 1.1.32. 通信路については，**伝送容量**とも呼ぶ．Gauss 分布 $X \sim N_k(a,A), Y \sim N_l(b,B), (X,Y) \sim N_{k+l}(c,C)$ について，

$$I(X,Y) = \frac{1}{2}\log\frac{\det(AB)}{\det(C)}$$

特に $k = l = 1$ のとき，

$$I(X,Y) = -\frac{1}{2}\log(1 - \mathrm{Corr}[X,Y]^2).$$

命題 1.1.33 .

(1) $I_{(X,Y)} \geqslant 0$．等号成立条件は $X \perp\!\!\!\perp Y$.

(2) $I_{(X,Y)} = I_{(Y,X)}$.

(3) $\forall_{f\in L(\Omega,\Omega)}\ I(X,Y) \geqslant I(X,f(Y))$．等号成立条件は f が可測な逆を持つこと．

定理 1.1.34 (特徴付け).

(1) 条件付きエントロピーによる特徴付け：

$$I_{(X,Y)} = H(X) - H(X|Y) = H(X) + H(Y) - H((X,Y)).$$

(2) 積分表示：X, Y は密度 p, q，(X, Y) は密度 r を持つとする．このとき，

$$I(X, Y) = E_{(X,Y)}\left[\log \frac{r}{pq}\right].$$

系 1.1.35．相互情報量はエントロピー／平均情報量の一般化である：

$$I(X, X) = H(X) - H(X|X) = H(X).$$

1.1.10　条件付き相互情報量

定義 1.1.36 (conditional mutual information). 条件付き確率の間の KL-変分の平均として，**条件付き相互情報量**を

$$I(X, Y|Z) := E_Z[D(P_{(X,Y)|Z}|P_{X|Z} \otimes P_{Y|Z})].$$

と定める．

定理 1.1.37 (条件付きエントロピーによる特徴付け).

$$I(X, Y|Z) := H(X|Z) - H(X|Y, Z).$$

命題 1.1.38 (相互情報量による特徴付け).

(1) $I(X, Y|Z) \geqslant 0$ で，等号成立条件は，$X \perp\!\!\!\perp Y|Z$．すなわち，X, Z, Y は Markov 連鎖である．
(2) $(X, Y) \perp\!\!\!\perp Z \Rightarrow I(X, Y|Z) = I(X, Y)$.
(3) 特徴付け：$I(X, Y|Z) = I(X, (Y, Z)) - I(X, Z)$.

系 1.1.39 (連鎖律).

$$I((X_1, \cdots, X_n), Y) = \sum_{i \in [n]} I(X_i, Y|X_{i-1}, \cdots, X_1).$$

1.1.11　情報変分

KL-変分に加えて，相互情報量から (こちらは真の) 計量が作り出せる．感覚的には，2 つの分割の間の「共有された情報」の量を測っている．

命題 1.1.40 (variation of information).

$$d(X, Y) = H((X, Y)) - I(X, Y) = H(X) + H(Y) - 2I(X, Y) = H(X|Y) + H(Y|X) = 2H((X, Y)) - H(X) - H(Y)$$

は距離の公理を満たす．

要諦 1.1.41. 正規化

$$d_0(X, Y) = \frac{d(X, Y)}{H((X, Y))} = 1 - \frac{I(X, Y)}{H((X, Y))} \leqslant 1$$

も距離を定め，Rajski 距離と呼ばれる．分割についての Jaccard 距離の考え方に等しい．なお，

$$d_1(X, Y) = 1 - \frac{I(X, Y)}{\max(H(X), H(Y))}$$

も距離になる．

1.2　Shannon の定理

記法 1.2.1. $\Omega := [r]$ 上の確率分布を $(p_k)_{k \in [r]}$ とし，ω_k の起こる回数を表す Ω^n 上の確率変数 $\nu_k := \sum_{l \in [r]} 1_{\{\omega_k\}}(\omega_l)$ を考える．

$$\mathcal{A}_\epsilon^{(n)} := \bigcap_{k \in [r]} \left(\left|\frac{\nu_k}{n} - p_k\right| < \epsilon\right).$$

Bernoulli の定理から，この集合は充満集合に収束する．

定理 1.2.2 (Shannon (1948)). 任意の $\epsilon > 0$ に対して，ϵ_1, n_0 が存在して，任意の $n > n_0$ に対して，次の 3 条件が成り立つ：

(1) $P[\mathcal{A}_{\epsilon_1} - (n)] > 1 - \epsilon$.
(2) $e^{n(H(P)-\epsilon)} \leqslant |\mathcal{A}_{\epsilon_1}^{(n)}| \leqslant e^{n(H(P)+\epsilon)}$.
(3) $\forall_{\omega \in \mathcal{A}_{\epsilon_1}^{(n)}}\ e^{-n(H(P)+\epsilon)} \leqslant P[\omega] \leqslant e^{-n(H(P)-\epsilon)}$.

第 2 章

最大エントロピー原理

この原理は Jaynes, E. T. (57)[†1] において，統計物理の分野で導入された．

(1) 何の事前知識もない場合の先験的等確率の原理もこの原理によって正当化される．

(2) 平衡状態における微視的状態の確率分布であるカノニカル分布は最大エントロピーの原理を用いても導かれる．

(3) 一般化線型モデルでは，残差に平均と分散以外の仮定を置かないから，観測 Y は $\varphi : \mathbb{R}^p \to \mathbb{R}^n$ を通じた条件 $E[\varphi(X)] = 0$ のみ与えられている．このとき，エントロピーを最大にするような絶対連続分布は $Ae^{-\varphi(x)}$ という形の密度を持つものであり，A の部分にパラメータが入る余地がある．

(4) 定常過程の推定においても，最大エントロピー原理が極めて有効であることが認められつつある．

2.1 定常過程とエントロピー

定義 2.1.1 (entropy rate). $X = (X_n)$ を離散時間確率過程とする．

$$\overline{H}(X) = \limsup_{n\to\infty} \frac{1}{n} H(X_1, \cdots, X_n), \quad \overline{h}(X) = \limsup_{n\to\infty} \frac{1}{n} h(X_1, \cdots, X_n).$$

をそれぞれ**単位時間あたりのエントロピー**という．

定理 2.1.2 . $X = (X_n)_{n\in\mathbb{Z}}$ を強定常過程とし，$X_0^- := (X_0, X_{-1}, X_{-2}, \cdots)$ とする．

(1) X_1 は離散的であるとする．

$$\overline{H}(X) = \lim_{n\to\infty} \frac{1}{n} H(X_1, \cdots, X_n) = H(X_1 | X_0^-).$$

(2) X_1 は連続的であるとする．

$$\overline{h}(X) = \lim_{n\to\infty} \frac{1}{n} h(X_1, \cdots, X_n) = h(X_1 | X_0^-).$$

2.2 平衡分布

一般の統計的推定において最大エントロピー原理は「仮定」であり，決定理論の枠組みで考えられるべきであるが，統計力学においては，エントロピー増大の法則を認める限り，最大エントロピー原理は「原理」ではない．

2.2.1 ミクロカノニカル分布

模型 2.2.1. N 個の同種の粒子からなる孤立系で，相互作用を無視できる系を扱う．この粒子のエネルギー準位の取り合える値を $E = \{e_i\}_{i\in[L]}$ とし，列 $n : [N] \to [L]$ を用いて状態点 $(e_{n_i})_{i\in[N]} \in E^N$ であって総エネルギー $e \in \mathbb{R}$ を持つ状態空間を

$$W(e) := \left\{ (e_{n_1}, \cdots, e_{n_N}) \in E^N \mid e = \sum_{k\in[N]} e_{n_k} \right\}$$

[†1] Information Theory and Statistical Mechanics, Phys. Rev.

とおく．このとき，同様の系を複数用意したときの空間的出現割合の分布を知りたい．

(1) これが時間的割合に等しいという仮定を**エルゴード仮説**という．
(2) いずれの状態も他と区別する理由がないため，$W(e)$ 上に離散一様分布 $U_{|W(e)|}$ を仮定するとき，これを**小正準集団** (microcanonical ensemble) という．

定義 2.2.2 . k を Boltzmann 定数とする．ミクロカノニカル集団のエントロピーを

$$S(e) := k \log W$$

で定める．

2.2.2　カノニカル分布

模型 2.2.3. 2 つの力学系 A, B の粒子交換は伴わず，エネルギー交換のみを伴う結合 $A + B$ からなる孤立系で，結合によるエネルギーの消失は無視できるものとする．$A + B$ の平衡状態において，系 A のエネルギーが e_A のとき，A は小正準分布に従うとみなせるから，A の取り得るエネルギー準位は，空間 E 上の確率分布として

$$p(e) := \frac{1}{\Phi} e^{-\lambda e}, \quad \Phi := \sum_{e \in E} |W(e)| e^{-\lambda e}$$

なる形の密度を持つと仮定できる．これを**正準分布**といい，Φ を**分配関数**という．

要諦 2.2.4. この分布は指数型で，エントロピー最大の原理から正当化出来る．

命題 2.2.5 . 理想気体からなる正準集団の分子の速度は次の正規分布 $N(0, \Sigma)$ $(\Sigma = \mathrm{diag}((\lambda m)^{-1}))$ に従う：

$$f(v) = \left(\frac{\lambda m}{2\pi}\right)^{3/2} \exp\left(-\frac{\lambda m}{2}(v_x^2 + v_y^2 + v_z^2)\right)$$

これを **Maxwell 分布**という．

要諦 2.2.6. これは，系のエネルギーの平均値が与えられたときの，連続エントロピーを最大にする分布に他ならない．

2.2.3　グランドカノニカル分布

模型 2.2.7. 2 つの力学系 A, B の粒子交換を伴う結合 $A + B$ からなる孤立系で，結合によるエネルギーの消失は無視できるものとする．このとき，2 つの変数，系 A のエネルギー $e_A \in E$ と粒子の個数 $N \in [|W(e)|]$ とからなる確率空間 $E \times |W(e)|$ 上の分布

$$\frac{1}{\Phi} \exp(-\lambda(E - \mu N)), \quad \Phi := \sum_{e_A \in E, N \in [|W(e)|]} \exp(-\lambda(E - \mu N)).$$

これを**大正準集団**という．

注 2.2.8. より一般に，l 個の物理量 $\alpha_1, \cdots, \alpha_l$ を交換する結合系 $A + B$ について，大正準分布が考えられる．

第 3 章

情報源のモデル

前章では確率変数の情報量を定義した．（情報源アルファベット上の）確率変数の列 (確率過程) とその上の制約の組 $(X_i)_{i\in\mathbb{N}}$ を**情報源**という．したがって，情報源のエントロピーとは，定常離散確率過程のエントロピーになる．

続いて，情報源（の出力）系列を効率よく符号化する，データ圧縮の問題を扱う．48 の Shannon の論文は，Weiner の確率論の方法を用いて，データ圧縮の方法を議論し，符号化の問題から情報理論を創始した．ここではまず，容易で，曖昧さのない復号が存在する符号の構成を考える．

3.1 情報源の定義

定義 3.1.1 (memoryless information source, Markov source).

(1) 情報源 $(X_i)_{i\in\mathbb{N}}$ が Markov 連鎖であるとき，**記憶のない情報源**または **i.i.d. 情報源**という．
(2) Markov 性が直前の有限個の記号に限定されていて，それ以前の記号列からは影響を受けない情報源を**（m-重）Markov 情報源**という．特に $m = 1$ の場合は**単純**であるという．

3.2 Markov 情報源

記憶を持つ情報源のうち，ある種の有限性を持つのが Markov 情報源である．主な問題は定常分布を知ることであるが，これは遷移確率行列の定める力学系の極限を求める問題である．特に重要な atom として，エルゴード情報源を考える．

定義 3.2.1 (state transition diagram). m 重 Markov 情報源について，状態空間 A^m 上の遷移として捉えたものを，**状態遷移図**または **Shannon 線図**という．

要諦 3.2.2. このような捉え方によって，$m = 1$ として一般性は失われない．

3.3 情報源のエントロピー

定義 3.3.1 . 情報源の出力 1 記号当たりの情報量の期待値を**情報源のエントロピー**といい，$H_r(\mathcal{S})$ で表す．

例 3.3.2.

(1) 記憶のない情報源のエントロピーは任意の i について $H(X_i)$ である．
(2) Markov 情報源のエントロピーは，各状態の出力分布のエントロピーを，定常分布に関して積分したものを言う．

3.4 エントロピーと平均符号長

定理 3.4.1 . 情報源 $\mathcal{S}$ の一意復号可能な r-元符号 C について，$L(C) \geqslant H_r(\mathcal{S})$ が成り立つ．

要諦 3.4.2. $H_r(S)$ は，S から発信されるシンボルの持つ自己情報量の期待値である．C が一意復号可能ならば，この情報量が保存されているはずである．

系 3.4.3 . S を生起確率 (p_i) の情報源とする．

(1) 一意復号可能な r 元符号 $\mathbb{C}$ で，$L(C) = H_r(S)$ を満たすものが存在する．
(2) $\exists_{e_i \in \mathbb{Z}_{\leqslant 0}}\ p_i = r^{e_i}$.

定義 3.4.4 . 情報源 S の r 元符号 C について，$\eta := \dfrac{H_r(S)}{L(C)}$ を**効率**といい，$0 \leqslant \eta \leqslant 1$ を満たす．$\overline{\eta} := 1 - \eta$ を**冗長性**という．

3.5　Shannon-Fano 符号

Huffman 符号は平均符号長の計算が複雑である．一方で，Shannon-Fano 符号は最適に近いことが保証されている上に，平均符号長の評価も容易である．

定義 3.5.1 . $l_i := \lfloor \log_r(1/p_i) \rfloor$ は Kraft の不等式を満たすので，これが定める瞬時符号が存在する．これを **Shannon-Fano 符号**という．

命題 3.5.2 . $H_r(S) \leqslant L(C) \leqslant 1 + H_r(S)$.

3.6　拡大情報源と積のエントロピー

定理 3.6.1 . S を任意の情報源とする．$H_r(S^n) = nH_r(S)$ が成り立つ．

定理 3.6.2 . 独立な情報源 $S, \mathbb{T}$ について，$H_r(S \times \mathbb{T}) = H_r(S) + H_r(\mathbb{T})$.

3.7　Shannon の第一定理

定理 3.7.1 (Noiseless Coding Theorem (Shannon 48)). 任意の $\epsilon > 0$ に対して，十分大きな $n \in \mathbb{N}$ が存在して，S^n を符号化することによって，情報源 S の一意復号可能な r 元符号 C であって，平均符号長がエントロピー $H_r(S)$ に対して，$L(C) - H_r(S) < \epsilon$ を満たす．

3.8　情報源符号化

記法 3.8.1.

(1) $T^* = \cup_{n \geqslant 0}[T]^n, T^+ := \cup_{n>0}[T]^n$.
(2) $C :=$ Im C と混用する．

定義 3.8.2 (coding, non-singular, decoding, uniquely decodable).

(1) アルファベット A, B について，写像 $f : A \to B^+$ を**符号**，値 $f(a) \in B^+$ を**符号語**という．始域アルファベットを**情報源アルファベット**，終域を生成するアルファベットを**符号アルファベット**という．
(2) f が単射であるとき，**正則符号**という．以降正則符号のみを考える．
(3) f の延長 $f^* : A^* \to B^*$ の逆写像を求めることを**復号**という．逆写像が B^* の全域で定まるとき，すなわち，f^* が単射であるとき**一意復号可能**であるという．

注 3.8.3. 符号アルファベットは主に通信路の技術に依るので，A, B は応用上も別物である．$|B|$ を基数 (radix) といい，多くの例では $r = 2$ である．モールス信号は空白を含めて $r = 3$ の例である．

例 3.8.4.

(1) ASCII (American Standard Code for Information Interchange) は二元符号で，$f(A) \subset B^7$ を満たす 7 ビットの符号化

である．

(2) A に線形順序があり，隣接する符号語の Hamming 距離が 1 になるような二元符号を **Gray 符号**という．

(3) バーコードや受験番号，ISBN の最後の桁は誤り訂正符号となっている．ISBN はハイフンを除いて長さ 10 の $\mathbb{Z}_{11}$ 上の符号語で，$a_1 0$ は $a_1 + 2a_2 + \cdots + 10a_{10} \equiv 0 \mod 11$ を満たすように定まっている．これは，単一誤りと，2 つの記号の置換を検出できる，人為エラーに特化された符号である．

定義 3.8.5 .

(1) $L : \mathrm{Map}(S, T^+) \to \mathbb{R}$ を，関数 $|C^*(X_n)| : S \to \mathbb{N}$ の期待値として定める．これを**平均符号長**という．

3.9　一意復号可能性

定理 3.9.1 . C は単射とする．$\mathrm{Im}\, C$ に含まれる符号語の長さが全て同じならば，C は一意復号可能である．このとき，C を長さ l の**ブロック符号**であるという．

定義 3.9.2 . $C_0 := \mathrm{Im}\, C, C_n := \left\{ w \in T^+ \mid \exists_{u \in \mathrm{Im}\, C, v \in C_{n-1}}\ uw = v \vee vw = u \right\}$ と帰納的に定める．$C_\infty := \cup_{n=1}^{\infty} C_n$ とする．

注 3.9.3. $C_1 = \left\{ w \in T^+ \mid \exists_{u,v \in C}\ uw = v \right\}$ となる．

定理 3.9.4 (Sardinas, Patterson 53). 次の 2 条件は同値．

(1) C は一意復号可能．

(2) $C \cap C_\infty = \varnothing$.

定理 3.9.5 (McMillan 56). 次の 2 条件は同値．

(1) 符号長が $l_1, \cdots, l_q$ である一意復号可能な r 元符号 C が存在する．

(2) $\displaystyle\sum_{i=1}^{q} \frac{1}{r^{l_i}} \leqslant 1$.

3.10　瞬時符号

瞬時符号とは，語頭符号である．これが存在するための条件は，Kraft の不等式によって特徴づけることができる．組み合わせ論的な本質は，木構造である．

3.10.1　語頭符号

定義 3.10.1 (instantaneously decodable codee).

(1) 符号 C が**瞬時復号可能な符号**であるとは，任意の符号語列 $t \in T^+$ に対して，t で始まる全ての符号列が，その後の符号に依らず，一意に復号されることをいう．

(2) 符号 C が**語頭符号**であるとは，どの符号語も，他の符号語の語頭には来ないことをいう：$\forall_{w_i \in \mathrm{Im}\, C}\ \forall_{w \in T^*}\ \forall_{w_j \in \mathrm{Im}\, C}\ i \neq j \Rightarrow w_j \neq w_i w$．すなわち，$C_1 = \varnothing$.

定理 3.10.2 . 次の 2 条件は同値．

(1) 誤頭符号である．

(2) 瞬時符号である．

3.10.2　木と構成法

議論 3.10.3. T^* は自然な包含関係に関して r 元根付き木の構造を持ち，ϵ を根とする．この木の頂点集合 $(\epsilon \notin)C$ が，性質 $\forall_{x,y\in C}\ x \neq y \Rightarrow x \wedge y = \epsilon$ を満たすとき，C は瞬時符号である．

3.10.3　Kraft の不等式

定理 3.10.4 (Kraft 49). 次の 2 条件は同値.

(1) 符号長が $l_1, \cdots, l_q$ であるような r 元瞬時符号 C が存在する．
(2) $\displaystyle\sum_{i=1}^{q} \frac{1}{r^{l_i}} \leqslant 1.$

要諦 3.10.5. 符号長が l であるとは，木構造の中の頂点としては高さ l に存在することを表す．これは，高さ $h(>l)$ の頂点 r^h 個のうち，r^{h-l} 個を使用不可とする．これを根に引き戻して考えると，$\dfrac{1}{r_i^l}$ の和が 1 を超えると，どのようにうまく選ぼうと瞬時符号は構成できないことがわかる．

系 3.10.6 . 次の 2 条件は同値.

(1) 符号長が $l_1, \cdots, l_q$ の r 元瞬時符号が存在する．
(2) 符号長が $l_1, \cdots, l_q$ の r 元一意復号可能な符号が存在する．

3.11　網羅性

一番効率よく符号を作るには，$\displaystyle\sum_{i=1}^{q} \frac{1}{r^{l_i}} = 1$ を狙いたい．これを捉える概念が存在する．

定義 3.11.1 (exhaustive). $\exists_{n\in\mathbb{N}}\ \forall_{w\in T^+}\ |w| > n \Rightarrow [\exists_{w_0\in \mathrm{Im}\ C}\ \exists_{w_1\in T*}\ w_0w_1 = w]$ を満たすとき，C を**網羅的**であるという．

定理 3.11.2 . 次の 2 条件は同値.

(1) 符号長が $l_1, \cdots, l_q$ であるような r 元網羅的符号 C が存在する．
(2) $\displaystyle\sum_{i=1}^{q} \frac{1}{r^{l_i}} \geqslant 1.$

第 4 章

情報通信路

雑音の多い／信頼性の低い通信路を介してメッセージを届ける情報源.

エントロピーは情報量の平均であったが，これが「字数」に対応づくことは，ある種「情報量」の概念の well-definedness をあらわす．エントロピーはこちらを定義として採用することも出来る.

4.1 記法と定義

定義 4.1.1 (channel). 情報源 $\mathcal{A}, \mathcal{B}$ について，それぞれのアルファベットを $\{a_1, \cdots, a_r\}, \{b_1, \cdots, b_s\}$ とする.

(1) それぞれのアルファベットの間の写像 Γ を**情報通信路**という.
(2) 各成分を $P_{ij} := P(b = b_j | a = a_i)$ とする行列を，**通信路行列**という.

例 4.1.2 (binary symmetric channel, binary erasure channel)**.**

(1) $A = B = 2$ で，シンボルに依らず成功・失敗確率が一様であるとき，**二元対称通信路**という.
(2) $A = 2, B = 2 \cup \{?\}$ であるとき，**二元消失通信路**という.

定義 4.1.3 . Γ, Γ' を通信路とする.

(1) 和 $\Gamma + \Gamma'$ とは，入力アルファベット $A \sqcup A'$ と出力アルファベット $B \sqcup B'$ について，通信路行列を直和 $M \oplus M'$ とする通信路である.
(2) 積 $\Gamma \times \Gamma$ とは，入力アルファベット $A \times A'$ と出力アルファベット $B \times B'$ について，通信路行列を Kronecker 積 $M \otimes M'$ とする通信路である.
(3) 積 Γ^n を，n **次拡大**という.
(4) 合成 $\Gamma \circ \Gamma'$ とは，入力アルファベット A と出力アルファベット B' について，通信路行列を積 MM' とする通信路である．これを通信路の**カスケード**という.

4.2 システムエントロピー

$H(\mathcal{A}), H(\mathcal{B}), H(\mathcal{A}|\mathcal{B}), H(\mathcal{B}|\mathcal{A}), H(\mathcal{A}, \mathcal{B})$ をシステムエントロピーという.

議論 4.2.1 (equivocation)**.** 条件付きエントロピー $H(\mathcal{A}|b_i)$ の積分 $H(\mathcal{A}|\mathcal{B})$ を**あいまい度**という.

4.3 通信路に関するシャノンの第一基本定理

定理 4.3.1 . 通信路の出力 $\mathcal{B}$ が既知ならば，任意の $\epsilon > 0$ に対して，十分大きな $n \in \mathbb{N}$ が存在して，$\mathcal{A}^n$ を符号化することによって，入力 $\mathcal{A}$ の一意復号可能で，$L(C) - H(\mathcal{A}|\mathcal{B}) < \epsilon$ を満たす符号化が存在する.

4.4　相互情報量

議論 4.4.1. 相互情報量 $I(\mathcal{A},\mathcal{B}) := H(\mathcal{A}) - H(\mathcal{A}|\mathcal{B})$ は次の 3 通りの解釈がある．

(1) $\mathcal{B}$ を知ることで解消する $\mathcal{A}$ についての不確かさの総量．
(2) $\mathcal{B}$ によって伝送される $\mathcal{A}$ についての情報量の総量．
(3) $\mathcal{A}$ に対する符号語に含まれるシンボルで，$\mathcal{B}$ の符号語に出てくるものの平均個数．

要諦 4.4.2 (数え上げ測度の例)**.** $|A \cap B| = |A| - |A\backslash B|$ における左辺が相互情報量と見れる．

4.5　通信路容量

定義 4.5.1 (capacity). 通信路 $\Gamma : A \to B$ の**容量**を，$I(\mathcal{A},\mathcal{B})$ の最大値と定める．

定理 4.5.2 .

(1) $\mathcal{P} := \left\{ p \in \mathbb{R}^r \mid p_i \geqslant 0, \sum_i p_i = 1 \right\}$ はコンパクト．
(2) $\mathcal{I}$ は，入力確率分布 $(p_i) \in \mathcal{P}$ の連続関数である．
(3) 任意の通信路について，容量が存在する．

4.6　連続 Gauss 型通信路

模型 4.6.1. Gauss 過程 (Z_t) について，次の 4 条件を満たすモデル $Y_t = X_t + Z_t$ を (遅延も雑音もない) **フィードバック ξ を持つ連続 Gauss 型通信路**という：

(1) $Z \perp\!\!\!\perp \xi$.
(2) X_t は $(\xi_s)_{s\in[0,t-]}$ と $(Y_s)_{s\in[0,t-]}$ の関数である：

$$X_t(\omega) = X(t, (\xi_s(\omega))_{s\in[0,t-]}, (Y_s(\omega))_{s\in[0,t-]}).$$

(3) $Y_t = X_t + Z_t$ は Y について一意な解を持つ．

4.7　白色 Gauss 型通信路

定義 4.7.1 (white noise). $(\nu_t)_{t\in\mathbb{R}}$ は Gauss 過程で，$s \neq t \Rightarrow \nu_s \perp\!\!\!\perp \nu_t$ が成り立つとする．これを**白色雑音**という．

模型 4.7.2. 平均過程 $\varphi(t)$ と白色雑音 ν_t が定める連続 Gauss 型通信路

$$Y_t = \int_0^t \varphi(u)du + B_t$$

を**白色 Gauss 型通信路**という．

4.8　信頼できない通信路の利用

Bravo と Victor のように，互いに十分異なる符号語を使えば，符号語に含まれるシンボルのいくつかが不正確な場合でも，受信側が混乱する可能性は低いというアイデアが Shannon の定理に含まれている．このように，情報理論は存在性を保証するが，実際の構成とアルゴリズムに対する考察は符号理論の範疇である．

4.8.1　決定則

定義 4.8.1 .

(1) 関数 $\Delta : B \to A$ を**決定則**という．$\Delta(b_j) = a_{j*}$ とあらわす．
(2) 正しい決定則を用いて，送信元のアルファベットを求めることを**復号**という．
(3) 正しく復号される確率 $P_C = \sum_j q_j P(a = a_{j*}|b = b_j)$ を最大にする決定則を**理想的観測者規則**という．
(4) 一方で，確率 $P(a = a_{j*}|b = b_j)$ が不可知であることも多い．その場合は，前向き確率 P_{ij}（通信路についての知識）のみが判断基準となり，$\forall_i\ P_{j*j} \geqslant P_{ij}$ を満たす決定則を**最尤法**という．

4.8.2　信頼性を高める例

例 4.8.2 (binary repetition code, majority decoding)**.** 同じ入力を n 回繰り返すとする．そこで，届いた符号語のうち，一番多いシンボルを採用して復号する．これを，r 元**反復符号** R_n という．この代償は，伝送速度が遅くなることである（n 倍の時間がかかるはず）．これを伝送レートという概念で測る．$|R_n| = r$ なので，伝送レートは $R = \dfrac{\log_r(r)}{n} = 1/n$ である．

定義 4.8.3 (transmission rate). 符号 $C \subset [A]^n$ の**伝送レート**とは，

$$R := \frac{\log_r |C|}{n}$$

である．

補題 4.8.4 . $0 \leqslant R \leqslant 1$.

4.9　Hamming 距離

最尤法は，反復符号については Hamming 距離の言葉によって特徴づけられ，これを**最近傍復号**という．

4.10　Shannon の基本定理

定理 4.10.1 . Γ を通信路容量 $C > 0$ の通信路とする．任意の $\delta, \epsilon > 0$ について，十分大きな $N \in \mathbb{N}$ が存在して，任意の $n \geqslant N$ について，$C - \epsilon \leqslant R < C$ を満たし，誤り確率が $P_E < \delta$ となる決定則を持つような，長さ n で伝送レート R の符号 C が存在する．

系 4.10.2 .

(1) Γ を $P > 1/2$ となる二元対称通信路とすると，Γ の容量は $C = 1 - H(P) > 0$ となる．
(2) 任意の $\delta, \epsilon > 0$ について，十分大きな $n \in \mathbb{N}$ が存在して，伝送レート R が $C - \epsilon \leqslant R < C$ を満たし，最近傍復号が誤り確率 $P_E < \delta$ を与えるような符号 $C \subset 2^n$ が存在する．

4.11　Shannon の基本定理の逆

定理 4.11.1 (Fano bound). Γ を通信路，r 元入力を $\mathcal{A}$，出力を $\mathcal{B}$ とする．Γ に対する任意の決定則 Δ の誤り確率 P_E は，次を満たす：

$$H(\mathcal{A}|\mathcal{B}) \leqslant H(P_E) + P_E[\log(r-1)].$$

系 4.11.2 . 通信路容量 C について，任意の $C' > C$ について，Shannon の定理は成り立たない．すなわち，C は任意の精度での伝送を可能にするためのレートの上限である．

第 5 章

符号理論

良い符号とは，ある程度は符号語が長くなければならないことを組み合わせ論的に観察し，２つの不等式の形で結果を得た．そこで，その制約の中でも，平均符号長 $L(C)$ がなるべき短い符号の構成方法を考える．

最適符号

誤り訂正符号　なるべく高い伝送レート R と，低い誤り率 P_E を兼ね備えた符号 C の構成法を考える．以降は符号理論であり，構成論であるが，その時の主要な道具が代数であり，特に線形代数である．

線型符号　線型符号は最小距離の計算が簡単であることをみた．また，最尤復号が最近傍復号となり，誤り訂正の計算の枠組みも最小距離の言葉で統一的である．ここで，線型符号を統一的に扱う枠組みを鳥瞰する．

5.1　最適符号の定義と存在

定義 5.1.1 (optimal code / compact code). 平均符号長 $L : \mathrm{Map}(S, T^+) \to \mathbb{R}$ を最小にする r 元瞬時符号を**最適符号**または**コンパクト符号**という．

補題 5.1.2 (定義の well-defined 性). 情報源 S と整数 r について，S の一意復号可能な r 元符号 C の平均符号長 $L(C)$ の値域は，S の r 元瞬時符号 C の平均符号長 $L(C)$ の値域に一致する．

定理 5.1.3 . 任意の情報源 S と整数 $r \geqslant 2$ について，最適な r 元符号が存在する．

5.2　２元 Huffman 符号

符号アルファベットを $\{0,1\}$ とする．Huffman 52 の理論．生起確率に基づいて，符号を定める．

定義 5.2.1 (reduced source).

(1) 情報源 S の値域を $\{s_1, \cdots, s_q\}$ とし，それぞれの生起確率を $p_1 \geqslant \cdots \geqslant p_q$ とする．新たなシンボル $s' := s_{q-1} \vee s_q$ を定め，**縮退情報源** S' を構成する．

(2) ２つの情報源の間に，写像 $\Phi : \{S' \text{の符号 }\} \to \{S \text{ の符号 }\}$ を構成する．S' の符号 C に対して，符号 C' は，$\{s_1, \cdots, s_{q-2}\}$ 上では同じだが，s_{q-1} を $w'0$，s_q を $w'1$ に対応させることとする．

(3) 縮退情報源を取る操作を $q-1$ 回繰り返すと，値域が一点集合の情報源となる．この自明な符号 $C^{(q-1)}$ の値域を $\{\epsilon\}$ とし（これは明らかに瞬時符号），Φ の値を $q-1$ 回とることで，**Huffman 符号** C を得る．

補題 5.2.2 . Φ は瞬時符号を保つ．すなわち，C' が瞬時符号ならば，C も瞬時符号である．

[証明]. 瞬時符号は語頭符号であることから従う． ■

要諦 5.2.3. 一意復元可能性は失われかねないが，瞬時符号ならうまくいく．絶妙．

5.3　2 元 Huffman 符号の最適性

定義 5.3.1 (sibling). 2 つの 2 元符号語 w_1, w_2 が，ある符号語 $x \in T^*$ が存在して $x0, x1$ とあらわせるとき，w_1, w_2 は**兄弟**であるという．

補題 5.3.2 . 任意の情報源 S は，符号長が最も長い 2 つの符号語が兄弟であるような 2 元最適符号 D を持つ．

定理 5.3.3 . C が情報源 S の 2 元 Huffman 符号ならば，最適符号である．

5.4　拡大情報源

よりマクロな構造に注目し，情報源アルファベットをうまく取り直すことで，生起確率が高い特定の列を符号化すると，さらに平均符号長を小さくすることが出来る．

5.5　誤り訂正の枠組みと線型符号

記法 5.5.1.

(1) 通信路はアルファベットを $\mathbb{F} := A = B$ として一般性を失わない．
(2) すべての符号語は等しい長さ n を持つとする（ブロック符号）．これは Shannon の定理による帰結．
(3) 任意の有限体は，p 群であり，$\mathbb{Z}/p^n\mathbb{Z}$ と表せる．

定義 5.5.2 .

(1) $C \subset \mathbb{F}^n$ が線型部分空間であるとき，C を**線型符号**または**群符号**という．
(2) $k := \dim C$ のとき，これを線型 $[n, k]$ 符号という．

補題 5.5.3 . 線型 $[n, k]$ 符号の伝送レートは $R = k/n$ と表せる．

要諦 5.5.4. 情報は k 成分に乗せて，残りは誤り訂正に使われる，という比率である．

5.6　符号の例

例 5.6.1 (repetition code)**.** $\mathbb{F}$ 上の反復符号 R_n は，$\mathbb{F}^n$ 内の一次元の線型符号で，符号語 $11\cdots1$ によって張られる空間となる．誤り率が十分小さい時，最近傍復号によって誤りが訂正される．しかし，伝送レートは悪く，$R = 1/n$.

例 5.6.2 (parity-check code)**. パリティ検査符号** P_n とは，

$$P_n := \left\{ (u_i) \in \mathbb{F}_q^n \;\middle|\; \sum_{i=1}^{n} u_i = 1 \right\}$$

によって定まる $n-1$ 次元の線型符号で，u_n を検査桁とする．伝送レートは優秀だが，誤り訂正は不可能で，検出も穴がある．

例 5.6.3 (binary Hamming code (47))**.** 2 元ハミング符号 H_7 は，長さ $n = 7$ の $\mathbb{F}_2$ 上の 4 次元の線型符号である．Bell 研究所の Hamming によって開発された．長さ 4 の 2 元記号列 $a_1a_2a_3a_4$ を，7 桁で符号化する．u_3, u_5, u_6, u_7 に写し，残りの u_1, u_2, u_4 は誤り訂正桁である．単一の誤りが修正可能．

例 5.6.4 (extended code)**.** 長さ n の体 $\mathbb{F}$ 上の符号 C に対して，**拡大符号** $\overline{C}$ とは，追加の桁を $\sum_{i=1}^{n+1} u_i = 0$ となるように選ぶことで，濃度の変わらない長さ $n+1$ の符号のことである．C が線型ならば，$\overline{C}$ も線型．

例 5.6.5 (punctured code)**.** 長さ n の符号 C に対して，**パンクチャド符号** C° とは，定めた桁数 $i \in [n]$ に対して，シンボル u_i を各符号語 $u_1 \cdots u_n \in C$ から取り除くことで定義される．

5.7　最小距離

誤り訂正の精度を上げる技法

最近傍復号を行うにあたって，符号の最小距離は大きければ大きいほど精度が高い．長さ n，符号語濃度 M，最小距離 d の符号を (n, M, d)-符号という．d に対して，$t := \left\lfloor \dfrac{d-1}{2} \right\rfloor$ ビット以内の誤りを正確に訂正できる．

補題 5.7.1 . C が線型符号のとき，最小距離は $d = \min\{d(v,0) \geqslant 0 \mid v \in C, v \neq 0\}$.

要諦 5.7.2. $w(v) := d(v,0)$ を **(Hamming) 重み**という．Hamming 距離が定めるノルムである．

定義 5.7.3 . 符号 C が t **重誤り訂正**であるとは，最大 t 桁の誤りまでは（誤りが距離 t 以下であるときは），常に正しく訂正されることをいう．

注 5.7.4. 一方で，最小距離 d の符号は，$d-1$ 個の誤りを検出する．

定義 5.7.5 . ベクトル $e := v - u$ を**誤りパターン**という．

定理 5.7.6 . 最小距離 d の符号 C について，

(1) t 個の誤りを訂正する．
(2) $d \geqslant 2t+1$ を満たす．

5.8　ハミングの球充填限界式

符号語 u に復号される範囲は，球 $S_t(u)$ である．誤り訂正の精度は，この球の大きさに単調増加し，伝送レートは，球の総数に単調増加する．$M := |C|$ の上界を与える．

定理 5.8.1 (Hamming's sphere-packing bound). C を q 元 t 重誤り訂正符号で，長さが n の M 個の符号語からなるとする．このとき，

$$M\left(1 + \binom{n}{1}(q-1) + \binom{n}{2}(q-1)^2 + \cdots + \binom{n}{t}(q-1)^t\right) \leqslant q^n.$$

要諦 5.8.2. この条件を満たす符号を**完全**符号という．これは，交わらない球 $S_t(u)$ が C を埋め尽くす条件と同値である．

5.9　Gilbert-Varshamov 限界

長さ n，最小距離 d に対して，符合語の数 M の最大値を $A_q(n,d)$ とすると，この下限を与える．

定理 5.9.1 (Gilbert-Varshamov bound). $q \geqslant 2$ かつ $n \geqslant d \geqslant 1$ ならば，

$$A_q(n,d)\left(1 + \binom{n}{1}(q-1) + \binom{n}{2}(q-1)^2 + \cdots + \binom{n}{d-1}(q-1)^{d-1}\right) \geqslant q^n$$

命題 5.9.2 (singleton の限界式, MDS : maximum distance separable code). $\mathbb{F}_q$ 上の符号が長さ n，最小距離 d，濃度 M であるとする．

$$\log_q M \leqslant n - d + 1.$$

等号を成立させるときの符号を，**最大距離分離符号**という．

5.10　Hadamard 符号

符号の構成の中でも，Hadamard 行列から構成できるクラスがある．

定理 5.10.1 (Hadamard).

(1) $|h_{ij}| \leqslant 1$ を満たす行列 $H = (h_{ij})$ の行列式について，$|\det H| \leqslant n^{n/2}$ が成り立つ．
(2) 等号成立条件は，$h_{ij} = \pm 1$ かつ H の任意の 2 つの異なる行は直交するとき．これを満たす行列を **Hadamard 行列**という．

注 5.10.2. $h_{ij} = \pm 1$ は，各行が長さ $\sqrt{n}$ であることを含意する $r_i \cdot r_i = n$．$HH^T = nI_n$ である．これより，$(\det H)^2 = n^n$ がすぐに従う．

補題 5.10.3 . Hadamard 行列 H に対し，$H' := \begin{pmatrix} H & H \\ H & -H \end{pmatrix}$ は $2n$ 次元の Hadamard 行列である．

系 5.10.4 (Sylvester matrix). 任意の $m \in \mathbb{N}$ について，2^m 次元の Hadamard 行列が存在する．

[証明]. $H := (1)$ は 1 次の Hadamard 行列である．これに対して補題を繰り返し適用すれば，帰納法より従う．この算譜で構成される Hadamard 行列を Sylvester 行列という． ■

命題 5.10.5 .

(1) Hadamard 行列の次数 $n > 1$ は偶数である．
(2) Hadamard 行列の次数 $n > 2$ は 4 の倍数である．

注 5.10.6. すべての 4 の倍数について，その次数の Hadamard 行列が存在するかは未解決．

定理 5.10.7 (Hadamard code). H を n 次元 Hadamard 行列とする．最小距離 $d = n/2$ で，符号語濃度が $M = 2n$ であるような長さ n の 2 元符号を，各行ベクトルの $\pm$ 計 $2n$ 個を，-1 を 0 とみなして定める．

歴史 5.10.8. $n = 32$ のものが，1969 年の火星探査機マリナーからの写真伝送に使われた．

命題 5.10.9 . 長さ n の Hadamard 符号の伝送レートは，$R = \dfrac{\log_2(2n)}{n} = \dfrac{1 + \log_2 n}{n} \xrightarrow{n\to\infty} 0.$

5.11　線型符号の行列表現

5.11.1　生成行列

線型符号化は，生成行列 G を用いて $x = uG$ と表せる．

定義 5.11.1 . k 行 n 列の行列 G の各行が C の基底からなるとき，k 次元線型符号 C の**生成行列**という．

要諦 5.11.2. この行列 G が定める線型同型 $\mathbb{F}^k \to \mathbb{F}^n$ が，情報源からの符号化を定めるとみなせる．

例 5.11.3. 反復符号 R_n は $G = (1\ 1\ \cdots\ 1)$ によって生成される．

5.11.2　パリティ検査行列

$HG^T = O$ を満たす行列をパリティ検査行列という．行ベクトルが線型符号を生成する行列を生成行列としたが，行ベクトルが C の直交補空間を生成する行列をパリティ検査行列という．

定義 5.11.4 . 線型符号 C が，c 個の一次方程式 $vH^T = 0$ で規定されるとき，これらを**パリティ検査方程式**といい，係数行列 H を**パリティ検査行列**という．

定義 5.11.5 . H を生成行列とみなして得る符号を，元の符号 C の**双対符号** D という．

補題 5.11.6 . $D = C^{\perp}$.

5.12　線型符号の同値性

線型符号の統一論

線型符号の分類を行う．代表元は組織符号が選ばれる．組織符号 $G = (I_k|P)$ の I_k を情報ビット，P をパリティ検査ビットという．$H = (G^T|I_{n-k})$ となる．

定義 5.12.1 (equivalent, systematic code form).

(1) 生成行列が相似な 2 つの線型符号を，**同値**であるという．
(2) ある行列 P について，$G = (I_k|P)$ と表せるとき，G を**組織符号形式**であるという．

要諦 5.12.2. 定義上，行の置換は符号を変えない．列の置換は，部分空間を変えるかもしれないが，シンボルの順序が変わるだけで，基本的な特性量は変わらない．その同値類の代表元は，各 $a_1 \cdots a_k \in [\mathbb{F}]^k$ をそのまま情報桁に写し取り，検査桁が aP で定義される符号である．

補題 5.12.3 . 組織符号形式 $G = (I_k|P)$ のパリティ検査符号は $H = (-P^T|I_{n-k})$ である．

5.13　線型符号の最小距離

線型符号の最小距離は，

$$d = \min\{w(v) = d(v,0) \in \mathbb{R}_{\geqslant 0} \mid v \in C, v \neq 0\}$$

であったが，パリティ検査行列の言葉で特徴づけることが出来る．

定理 5.13.1 . C を最小距離 d の線型符号とし，H を C のパリティ検査符号とする．このとき，d は H の一次従属な列の数の最小数となる．

要諦 5.13.2. $Hx^T = 0$ は $\sum_i x_i h_i = 0$ だが，これは $x_i = 1$ なる i について，列ベクトル h_i を足したもの．すなわち，これらの h_i が一次従属であることを表す．すなわち，一次従属な列ベクトル h_i の最小個数は，1 の最小個数に一致する．

5.14　Hamming 符号

定義 5.14.1 . パリティ検査行列 H の列ベクトルとして，零ベクトル以外のすべての列ベクトルを取って得られる組織符号を，**Hamming 符号**という．

例 5.14.2. $n = 6, k = 3$ とする．

$$H = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}$$

がパリティ検査行列となり，主座な 3×4 行列が P^T に当たる．生成行列は

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

となる．最小距離は 3 で，$t = \lfloor (3-1)/2 \rfloor = 1$ ビット誤り訂正符号となる．

定義 5.14.3 . Hamming 符号に，さらにパリティビット $p := x_1 + \cdots + x_k \mod 2$ を加えて得る $y = (x\ p)$ からなる符号を，**拡**

大 Hamming 符号という．符号の Hamming 重みが偶数になるが，最小重みは小さくならない．よって，$n = 6, k = 3$ のとき，最小重みは 4 になる．最近傍復号により，単一誤りの訂正と二重誤りの検出が可能になる．

5.15　Golay 符号

5.16　標準配列

5.17　シンドローム復号

線型符号が定めるシンドローム $s^T = Hy^T$ を求め，これからコセットリーダー e が求まれば，$y - e = x$ である．

議論 5.17.1 (線型符号の最尤復号)**.** 符号語 $x \in C$ に対して，誤りパターン $e = y - x$ は $Hy^T = He^T$ を満たす．よって，入力と出力の組 (x, y) について次の 3 条件は同値．

(1) $d(x, y)$ が最小．
(2) $w(e)$ が $\{e = y - x \mid x \in C\}$ で最小．
(3) $w(e)$ が $\left\{e \in V \mid He^T = Hy^T\right\}$ で最小．

定義 5.17.2 .

(1) $s^T := Hy^T$ を**シンドローム**という．
(2) $\left\{e \in V \mid He^T = Hy^T\right\}$ を**コセット**という．
(3) $\arg\min\left\{w(e) \geqslant 0 \mid He^T = Hy^T\right\}$ を**コセットリーダー**という．

5.18　巡回符号

剰余環 $\mathbb{F}/(z^n - 1)$ のイデアルとみなせる線型符号を，巡回符号という．

5.18.1　定義

定義 5.18.1 .

(1) 線型符号 C の任意の符号語 $x = (x_1, \cdots, x_n)$ の巡回シフト $(x_n, x_1, \cdots, x_{n-1})$ も C の符号語であるとき，C を**巡回符号**という．
(2) C の符号語を多項式 $f(z) = \sum_{i=1}^{n} x_i z^{n-i} \in \mathbb{Z}_2[z]/(z^n - 1)$ と同一視すると，z 倍準同型が巡回シフトに対応する．

命題 5.18.2 .

(1) 巡回符号 C に対応する符号多項式の集合は，剰余環 $\mathbb{Z}_2[z]/(z^n - 1)$ のイデアルに対応する．
(2) 環 $Z_2[Z]/(z^n - 1)$ は単項イデアル整域である．

5.18.2　生成多項式

定義 5.18.3 . 巡回符号 C を生成する多項式 $g \in \mathbb{Z}_2[z]/(z^n - 1)$ を**生成多項式**という．

補題 5.18.4 . g は，$C \backslash \{0\}$ の中で次数が最小の多項式になる．

定理 5.18.5 . 生成多項式 g は $z^n - 1$ を割り切る．

定理 5.18.6 . 巡回符号 C のパリティ検査多項式 h は，以下を満たす：$f \in C \Leftrightarrow h(z)f(z) \equiv 0 \mod z^n - 1$.

第 6 章

暗号理論

信頼できないのは通信路の性質だけでなく，通信路の傍受である可能性もある．

6.1 RSA

Rivest-Shamir-Adleman 暗号 (78) は，「戻し方が簡単にはわからない Cieser 暗号」となる．このように暗号化と復号化の鍵が異なる暗号を，**公開鍵暗号**という．RSA は位数が秘密鍵を知らないと／ $n = pq$ の因数分解が出来ないと不明な群 $(\mathbb{Z}/\varphi(n)\mathbb{Z})^*$ によってこれを実現している．

定理 6.1.1 (Euler). 任意の正整数 m に対して，$\gcd(a,m) = 1 \Rightarrow a^{\varphi(m)} \equiv 1 \mod m$.

定義 6.1.2 .

(1) Alice は素数 p,q を選び，$n := pq$ と，$1 < e < \phi(n) = (p-1)(q-1)$ かつ $\gcd(e,\phi(n)) = 1$ を満たす整数 e とを公開鍵とし，$d := e^{-1} \mod \phi(n)$ を秘密鍵とする．
(2) Bob は平文 $m \in (\mathbb{Z}/n\mathbb{Z})^*$ を送るとき，暗号文 $c := m^e \mod n$ を送信する．
(3) Alice は $m = c^d \mod n$ を計算することで復号できる．

要諦 6.1.3. 平文・暗号文の空間 $(\mathbb{Z}/n\mathbb{Z})^*$ と，その乗数の空間 $(\mathbb{Z}/\varphi(n)\mathbb{Z})^*$ とがある．十分大きな $n := pq$ について，$e < |(\mathbb{Z}/n\mathbb{Z})^*| = \varphi(n)$ を用意すると，$e \in (\mathbb{Z}/\phi(n)\mathbb{Z})^*$ の逆元はなかなか判らない．n を p,q に因数分解できないと，Euler 関数 $d := e^{-1} \mod \phi(n)$ の計算の仕様もない（２つの素数を使うメカニズムはここにある）．すると，「戻し方が判らない Cieser 暗号」のようになる．戻る原理は Euler の定理 $m = (m^e)^d = m^{k\varphi(n)+1} = m \mod n$ による．

例 6.1.4 (discrete logarithm)**.** 一般に，巡回群 $G = \langle\alpha\rangle$ について，$\forall_{\beta\in G}\ \exists_{n\in\mathbb{N}}\ \alpha^n = \beta$ であるが，$n = \log_\alpha \beta$ を求める問題を離散対数問題という．これが計算困難であるから，離散群の指数をいじる方針が立つ．

6.2 Diffie-Hellman の鍵交換

RSA が（素因数分解をするか）離散対数問題を解かねばならないのと同様なメカニズムで，安全に暗号鍵を共有する方法 (76)．すべての公開鍵と秘密鍵の内１つを得ることで簡単に破られる．

定義 6.2.1 .

(1) Alice と Bob は素数 p と原子根 $g \in \mathbb{F}_p$ を定めて共有する．
(2) 整数 a,b を，Alice と Bob がそれぞれ定め，秘密鍵とする．
(3) Alice が $g^a \mod p$ を Bob に送信し，Bob は $g^b \mod p$ を Alice に送信する．すると，２人の間のみで $g^{ab} \mod p$ の値が秘密裏に共有される．

6.3　ElGamal 暗号

実際に Diffie-Hellman の鍵交換を用いた暗号化法 (85).

定義 6.3.1 .

(1) Alice が素数 p と原子根 $g \in \mathbb{F}_p$ を定め，整数 $0 < a < p-2$ を用意し，$h := g^a \mod p$ と併せて，3-組 (h, g, p) を Bob に共有する.

(2) Bob は同様に $0 < b < p-2$ を定め，共有された秘密鍵 $s := h^b \mod p$ を定める．平文 $m \in \mathbb{F}_p^\times$ の暗号化を $c := ms$ として行い，これを $g^b \mod p$ と共に送信する.

(3) Alice も $(g^b)^a = s$ を得るので，$s^{-1} \in \mathbb{F}_p$ を Euclid の互除法によって計算し，これを用いて復号する.

要諦 6.3.2. 今回は体の元なので，逆元の計算は簡単である代わりに，秘密鍵 a, b の取得が困難になる.

第 7 章

参考文献

参考文献

[1] 甘利俊一『情報理論』
[2] 井原俊輔 (1984) 『確率過程とエントロピー』(岩波書店).
[3] 横尾英俊『情報理論の基礎』
[4] 野口悠紀雄 (1974) 『情報の経済理論』(東洋経済新聞社).
[5] Claude E. Shannon "A mathematical Theory of Communication" (1948)
[6] 有本卓 (1980) 『確率・情報・エントロピー』(森北出版).