ネットワーク基礎論2

小川朋宏　　　　2008-10-1

日程	:	水曜日3限 (13:00〜14:30) 10/8 休講, 11/19 調布祭準備 12/24, 12/31 冬休み
評価	:	レポート 2回〜3回, 出席状況

イントロダクション

情報理論とは

1948 Shannon, A mathematical theory of communications

○「情報」を確率論に基いて体系化

○ 情報源符号化定理
データ圧縮率の最適値 = 情報源のエントロピー

○ 通信路符号化定理
通信速度の最適値 = 通信路の入力と出力の相互情報量

本講義の最初の目標

○ このため, エントロピー, 相対エントロピー
相互情報量について学ぶ

○ 1-1, 1-2, 1-3 は確率論についての
予備知識の確認

1. 情報量とその性質

1-1. 確率変数

例: $\mathcal{X}=\{1,2,3,4,5,6\}$

○ $\mathcal{X}$, $\mathcal{Y}$, ⋯ 有限集合

○ 関数 $P: \mathcal{X} \to \mathbb{R}$ (実数の集合) は
以下の条件をみたすとき
$\mathcal{X}$ 上の確率関数 (probability function) という

(a) $P(x) \geq 0 \quad (x \in \mathcal{X})$

(b) $\sum_{x \in \mathcal{X}} P(x) = 1$

[(a) (b) より $0 \leq P(x) \leq 1 \quad (x \in \mathcal{X})$ が示される]

○ $\mathcal{X}$ の要素に値を取る変数 X に
確率関数 $P(x)$ が付与されているとき
X は $\mathcal{X}$ に値をとる (の上の) 確率変数 (random variable) という

$P(x)$: $X = x$ となる確率

$\Pr\{X = x\}$

○ 記法: X の確率関数を $P_X(x)$ と書く

X, Y, Z ⋯ の確率関数を
P, Q, R ⋯ とすると文字が足りなくなる
対応が付けにくい
P_X, P_Y, P_Z ⋯ と書く

○ 確率変数は大文字 X, Y, Z, ⋯ で
集合の要素は小文字 x, y, z, ⋯ で書く
(実現値)

○ 部分集合 $A \subset \mathcal{X}$ を <u>事象</u> (event) という

A が起こる確率

$$P(A) := \Pr\{X \in A\} = \sum_{x \in A} P_X(x)$$

例: サイコロ

$$\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$$

$$P_X(x) = \frac{1}{6} \quad (x \in \mathcal{X})$$

□ $A = \{x \in \mathcal{X} \mid x \text{ は奇数}\} = \{1, 3, 5\}$

$$P(A) = P_X(1) + P_X(3) + P_X(5) = \frac{3}{6} = \frac{1}{2}$$

□ $A = \mathcal{X}$, $P(\mathcal{X}) = 1$

□ $A = \phi$ (空集合) , $P_X(\phi) = 0$

○ $\mathcal{X}$, $\mathcal{Y}$ 有限集合

X : $\mathcal{X}$ 上の確率変数

関数 $f : \mathcal{X} \longrightarrow \mathcal{Y}$ が与えられたとき、

$Y = f(X)$ は確率変数となり、

確率関数 $P_Y(y)$ $(y \in \mathcal{Y})$ は以下で与えられる

$$P_Y(y) = \sum_{x : y = f(x)} P_X(x)$$

例: サイコロ続き

$$\mathcal{X} = \{1, 2, \cdots, 6\}, \quad \mathcal{Y} = \{a, b\}$$

$$f(x) = \begin{cases} a & \text{if } x \text{ is odd} \\ b & \text{if } x \text{ is even} \end{cases}$$

$$P_Y(a) = P_X(1) + P_X(2) + P_X(3)$$

$$P_Y(b) = P_X(2) + P_X(4) + P_X(6)$$

1-2、多次元確率変数

○ $\mathcal{X}$, $\mathcal{Y}$ の直積

$$\mathcal{X} \times \mathcal{Y} = \{(x, y) \mid x \in \mathcal{X}, y \in \mathcal{Y}\}$$

例：$\mathcal{X} = \{1, 2\}$　　$\mathcal{Y} = \{1, 2, 3\}$

$\mathcal{X}$ \ $\mathcal{Y}$	1	2	3
1	(1, 1)	(1, 2)	(1, 3)
2	(2, 1)	(2, 2)	(2, 3)

○ $\mathcal{X} \times \mathcal{Y}$ 上の $\left\{\begin{array}{l} \text{確率関数　} P(x, y) \\ \text{確率変数　} (X, Y) \end{array}\right.$

が同様に定義される

○ 確率変数 (X, Y) の確率関数を $P_{XY}(x, y)$ と書く

○ 周辺確率関数

$$P_X(x) = \sum_{y \in \mathcal{Y}} P_{XY}(x, y)$$

$$P_Y(y) = \sum_{x \in \mathcal{X}} P_{XY}(x, y)$$

例：続き

$P_{XY}(x, y)$

X \ Y	1	2	3	
1	1/12	2/12	3/12	6/12
2	1/12	1/12	4/12	6/12
	2/12	3/12	7/12	1

$P_X(x)$

$P_Y(y)$

○ 条件付き確率

$\forall x \in \mathcal{X},\ \forall y \in \mathcal{Y}$

$$P_{XY}(x,y) = P_X(x)\, P_{Y|X}(y|x)$$

をみたす関数 $P_{Y|X}(y|x)$ を<u>条件付き確率</u>という

□ $P_X(x) > 0$ のとき

$$P_{Y|X}(y|x) = \frac{P_{XY}(x,y)}{P_X(x)}$$

$X=x$ が起ったもとでの Y の確率

各 $x \in \mathcal{X}$ に対して $P_{Y|X}(y|x)$ は $\mathcal{Y}$ 上の確率関数

□ $P_X(x) = 0$ のとき $P_{XY}(x,y) = 0$ だから

$P_{Y|X}(y|x)$ は勝手な $\mathcal{Y}$ 上の確率関数でよい

例：続き

	1	2	3	
1	1/6	2/6	3/6	← $P_{Y|X}(y|1)$
2	1/6	1/6	4/6	← $P_{Y|X}(y|2)$

○ $\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}$ 上の確率関数 $P_{XYZ}(x,y,z)$

確率変数 (X,Y,Z)

周辺確率 $P_{XY}(x,y)$ $P_{YZ}(y,z)$ $P_{XZ}(x,z)$

$P_X(x)$ $P_Y(y)$ $P_Z(z)$

条件付き確率 $P_{Z|XY}(z|x,y)$ $P_{XY|Z}(x,y|z)$

$\cdots$

が同様に定義される

Def 確率変数 $X_1, X_2, \cdots, X_n$ が独立

$$\overset{\text{def}}{\iff} \quad P_{X_1 X_2 \cdots X_n}(x_1, x_2, \cdots, x_n) = P_{X_1}(x_1) \cdot P_{X_2}(x_2) \cdot \cdots \cdot P_{X_n}(x_n)$$

for all $x_1, \cdots, x_n$

特に X, Y が独立

$$\iff \quad P_{XY}(x, y) = P_X(x) P_Y(y) \quad \text{for } \forall x \in \mathcal{X}, \forall y \in \mathcal{Y}$$

1-3. 期待値と分散

○ $\mathcal{X} \subset \mathbb{R}$ に値をとる確率変数の期待値 (expectation)

$$E[X] := \sum_{x \in \mathcal{X}} P_X(x)\, x$$

Lem $Y = f(X)$ のとき Y の期待値について

$$E[Y] = \sum_{x \in \mathcal{X}} P_X(x) f(x)$$

∵ $f : \mathcal{X} \to \mathcal{Y}$ での $y \in \mathcal{Y}$ の逆像

$$f^{-1}(y) = \{ x \in \mathcal{X} \mid f(x) = y \}$$

は $\mathcal{X}$ の分割を与える, i.e.,

$$\begin{cases} \mathcal{X} = \bigcup_{y \in \mathcal{Y}} f^{-1}(y) \\ f^{-1}(y_1) \cap f^{-1}(y_2) = \phi \quad (y_1 \neq y_2) \end{cases}$$

よって

$$\text{右辺} = \sum_{y \in \mathcal{Y}} \sum_{x \in f^{-1}(y)} P_X(x) \underbrace{f(x)}_{= y}$$

$$= \sum_{y \in \mathcal{Y}} y \underbrace{\sum_{x \in f^{-1}(y)} P_X(x)}_{= P_Y(y)}$$

$$= E[Y]$$

□

○ 期待値の性質

(a) (X, Y) 確率変数, $a, b \in \mathbb{R}$ について

$$E[aX + bY] = aE[X] + bE[Y]$$

∵ $f((x, y)) = ax + by$ として Lem を用いると

$$E[aX + bY] = \sum_x \sum_y P_{XY}(x, y)(ax + by)$$

$$= a \sum_x \underbrace{\sum_y P_{XY}(x, y)}_{= P_X(x)} x + b \sum_y \underbrace{\sum_x P_{XY}(x, y)}_{= P_Y(y)} y$$

$$= aE[X] + bE[Y] \qquad \square$$

(b) X, Y が独立のとき

$$E[XY] = E[X] \cdot E[Y]$$

(証明略)

○ 分散 (variance)

$$V[X] := E[(X - \mu)^2]$$

ただし $\mu = E[X]$

○ 分散の性質

(a) $V[X] \geq 0$

(b) $V[X] = E[X^2] - \mu^2$

(c) $V[aX + b] = a^2 V[X]$ $(a, b \in \mathbb{R})$

後の大数の法則でより詳しく学ぶ

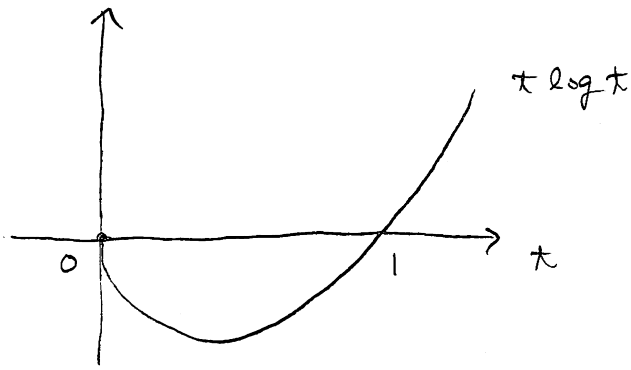
1-4. エントロピー

Def 確率変数 X の エントロピー (entropy)

$$H(X) := -\sum_{x \in \mathcal{X}} P_X(x) \log P_X(x) = E[-\log P_X(X)]$$

Remark

□ $0 \log 0 = \lim_{t \to 0} t \log t = 0$ とする



□ 対数の底を明示するとき

$$H_a(X) = E[-\log_a P_X(X)]$$

とかく

□ 底の変換

$$\begin{aligned} H_a(X) &= E[-\log_a P_X(X)] \\ &= E\left[-\frac{\log_b P_X(X)}{\log_b a}\right] \\ &= \frac{1}{\log_b a} H_b(X) \end{aligned}$$

定数倍の差

□ 実用的には底として 2 を用いる (bit)

理論計算には e を用いる (nat)

ことが多い

Lem (エントロピーの正値性)

$$H(X) \geq 0$$

等号成立 $\iff$ $\exists\, x_0 \in \mathcal{X},\ P_X(x_0) = 1$

∵

$$H(X) = \sum_{x \in \mathcal{X}} P_X(x) \underbrace{\{-\log P_X(x)\}}_{\geq 0} \geq 0$$

等号成立 $\iff$ $\forall x \in \mathcal{X},\ P_X(x) \log P_X(x) = 0$

$\iff$ $\forall x \in \mathcal{X},\ P_X(x) = 0 \text{ or } 1$

$\iff$ $\exists\, x_0 \in \mathcal{X}$ s.t.

$$\begin{cases} P_X(x_0) = 1 \\ P_X(x) = 0 \quad (x \neq x_0) \end{cases}$$

$\square$

例: binary entropy

$$h(p) := -p \log p - (1-p) \log(1-p)$$

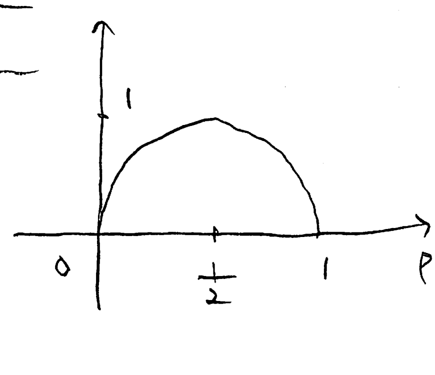
$$h'(p) = -\log p - 1 + \log(1-p) + 1 \qquad (\text{底 } e)$$

$$= -\log p + \log(1-p)$$

$$h''(p) = -\frac{1}{p} - \frac{1}{1-p} = -\frac{1}{p(1-p)} \leq 0$$

p	0		$1/2$		1
$h'(p)$	$+\infty$	$+$	0	$-$	$-\infty$
$h(p)$	0		1		0

↑ (底 2)

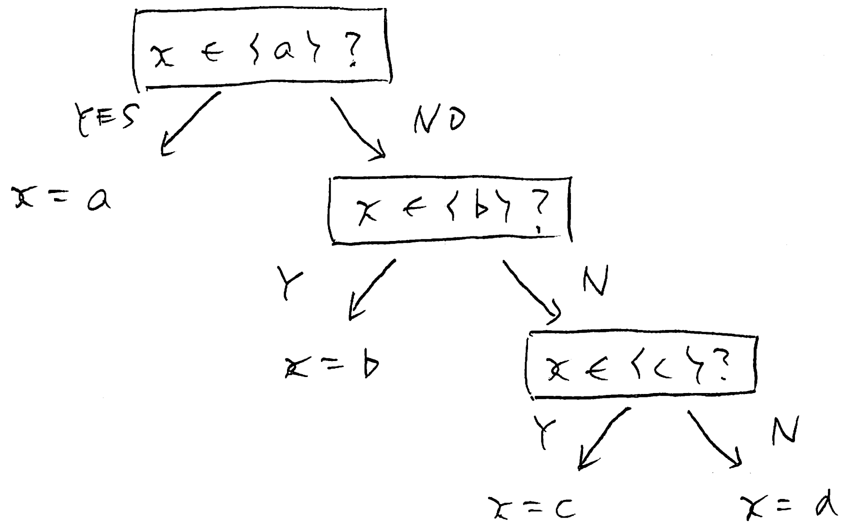


例：二分探索

- $\mathcal{X}=\{a, b, c, d\}$ 上の確率変数 X

x	a	b	c	d
$P_X(x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$

- X の実現値 x について
  部分集合 $A\subset\mathcal{X}$ により
  「$x\in A$ or $x\in A^c$」の形で質問
  ↑補集合
- 平均質問回数の最適値は？
- 探索木



□ 平均回数

$$\frac{1}{2}\cdot 1+\frac{1}{4}\cdot 2+\underbrace{\frac{1}{8}\cdot 3+\frac{1}{8}\cdot 3}_{\frac{3}{8}}=\frac{11}{8}$$

□ X のエントロピー

$$H_2(X)\quad -\frac{1}{2}\log_2\frac{1}{2}-\frac{1}{4}\log_2\frac{1}{4}-\frac{1}{8}\log_2\frac{1}{8}-\frac{1}{8}\log_2\frac{1}{8}$$

$$=\frac{1}{2}\cdot 1+\frac{1}{4}\cdot 2+\frac{1}{8}\cdot 3+\frac{1}{8}\cdot 3=\frac{11}{8}$$

- 一般に以下が示される

$H(X) \leq$ 平均質問回数の最適値 $\leq H(X)+1$

⇝ 可変長データ圧縮

- エントロピーの意味
  - □ X のあいまいさの尺度
  - □ X の実現値を得た時の「知識」の増分
  - □ 情報源符号化定理を通してより意味が明確になる

1-5 通信路

○ 集合 $\mathcal{X}, \mathcal{Y}$ 上の条件付き確率

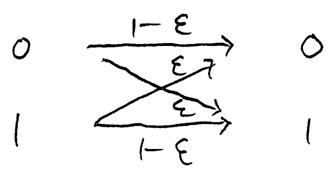
$$W(y|x) \quad \left[\sum_{y\in\mathcal{Y}} W(y|x) = 1,\ W(y|x) \ge 0 \right]$$

を <u>通信路</u> (channel) とよぶ

○ x を入力したときに確率 $W(y|x)$ で
y が出力される

$$x \longrightarrow \boxed{W} \longrightarrow y$$

○ 例: binary symmetric channel



○ $\mathcal{X}$ 上の確率 $P(x)$ と 通信路 $W(y|x)$
により 同時分布

$$P(x,y) = P(x)W(y|x)$$

が定まる

○ 関数 $f: \mathcal{X} \longrightarrow \mathcal{Y}$ は

$$W_f(y|x) = \begin{cases} 1 & \text{if} \quad y = f(x) \\ 0 & \text{if} \quad y \neq f(x) \end{cases}$$

により 通信路とみなせる
(deterministic channel)

1-6 条件付きエントロピー

- 同時確率変数 (X, Y) のエントロピーは定義より

$$H(X,Y) = -\sum_{(x,y)\in \mathcal{X}\times\mathcal{Y}} P_{XY}(x,y) \log P_{XY}(x,y)$$

$$= -\sum_{x}\sum_{y} P(x,y) \log P(x,y)$$

- 条件付き確率 $P_{Y|X}(y|x)$ は $x \in \mathcal{X}$ を固定して yの関数とみると確率分布

↑ $P(\cdot|x)$ とかく

$$H(Y|x) := H(P(\cdot|x))$$

Def (条件付きエントロピー)

$$H(Y|X) := \sum_{x} P(x) H(Y|x) \quad \text{— ①}$$

$$= -\sum_{x} P(x) \sum_{y} P(y|x) \log P(y|x)$$

$f(x,y) = -\log P(y|x)$ として $E[f(X,Y)]$

$$= -\sum_{x}\sum_{y} P(x,y) \log P(y|x)$$

$$= E[-\log P(Y|X)]$$

Lem (条件付きエントロピーの正値性)

$$H(Y|X) \geq 0$$

等号成立 $\Longleftrightarrow$ $\forall x \in \mathcal{X}$,

$P(x) > 0 \Rightarrow \exists y_x \in \mathcal{Y}$ $P(y_x|x) = 1$

[確率1で出力が確定]

$H(Y|x) \geq 0$ だから ① より $H(Y|X) \geq 0$

等号成立 $\Longleftrightarrow$ $\forall x \in \mathcal{X}$, $P(x)H(Y|x) = 0$

$\Longleftrightarrow$ $\forall x \in \mathcal{X}$, $P(x) = 0$ or $H(Y|x) = 0$

$\neg P \vee Q \equiv P \Rightarrow Q$ $\longrightarrow$ $\Longleftrightarrow$ $\forall x \in \mathcal{X}$, $P(x) > 0 \Rightarrow \underbrace{H(Y|x) = 0}_{\parallel}$

$\exists y_x \in \mathcal{Y}$,
$W(y_x|x) = 1$
$(\because$ 1-4$)$ □

1-7. エントロピーの chain rule

Lem $H(X,Y) = H(X) + H(Y|X)$

$$\log P(x,y) = \log P(x)P(y|x) = \log P(x) + \log P(y|x)$$

両辺マイナスして平均をとる □

○ より一般に

□ $P(x_1, x_2, \cdots, x_n)$

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

$= P(x_1, \cdots, x_{n-2}) P(x_{n-1} | x_1, \cdots, x_{n-2}) P(x_n | x_1, \cdots, x_{n-1})$

$\vdots$

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

□ $H(X_1, X_2, \cdots, X_n)$

$= H(X_1) + H(X_2|X_1) + \cdots + H(X_n|X_1, \cdots, X_{n-1})$

が成りたつ

1-8 ダイバージェンス

Def 集合 $\mathcal{X}$ 上の確率分布 P, Q に対して

$$D(P \| Q) := \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}$$

ただし

- $0 \log \frac{0}{b} := \lim_{a \searrow 0} a \log \frac{a}{b} = 0 \quad (b > 0)$
- $a \log \frac{a}{0} := \lim_{b \searrow 0} a \log \frac{a}{b} = +\infty \quad (a > 0)$
- $0 \log \frac{0}{0} := 0$ とする

[$a \to 0$, $b \to 0$ のとき
$a \log \frac{a}{b}$ の極限は定まらない
$P(x) = Q(x) = 0$ となる x は集合 $\mathcal{X}$ から除外
したと考える]

確率変数 X, Y について

$$D(X \| Y) := D(P_X \| P_Y)$$

Remark (名前について)

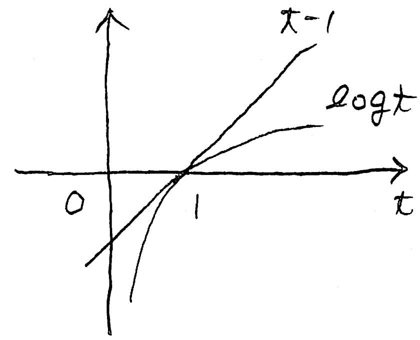
- ダイバージェンス (divergence)　情報理論
- 相対エントロピー (relative entropy)　統計物理
- Kullback-Leibler 情報量 (KL-information)　統計学

様々な分野で用いられる重要な量

Lem （ダイバージェンスの正値性）

$$D(P\|Q) \geq 0$$

等号成立 $\iff P = Q$



(∵)

右図より

$\log t \leq t-1 \quad (t>0)$

$t = \frac{1}{s} \; (s>0)$ とおくと

$\underbrace{\log\frac{1}{s}}_{= -\log s} \leq \frac{1}{s} - 1$

等号成立 $\iff s=1$

よって $\quad \log s \geq 1 - \frac{1}{s} \quad (s>0)$ —— ①

$$D(P\|Q) = \sum_{x\in\mathcal{X}} P(x)\log\frac{P(x)}{Q(x)}$$

$$= \sum_{\substack{x\in\mathcal{X}\\ P(x)>0}} P(x)\log\frac{P(x)}{Q(x)} + \underbrace{\sum_{\substack{x\in\mathcal{X}\\ P(x)=0}} P(x)\log\frac{P(x)}{Q(x)}}_{=0}$$

[∵ $P(x)=0$ のとき
$Q(x)=0$ と $Q(x)>0$ のどちらにせよ
$P(x)\log\frac{P(x)}{Q(x)} = 0$]

$$\overset{ⓐ}{\geq} \sum_{\substack{x\in\mathcal{X}\\ P(x)>0}} P(x)\left(1 - \frac{Q(x)}{P(x)}\right)$$

① ⟶ ⓐ

$$= \sum_{\substack{x\in\mathcal{X}\\ P(x)>0}} P(x) - \sum_{\substack{x\in\mathcal{X}\\ P(x)>0}} Q(x)$$

$$\overset{ⓑ}{\geq} 1 - 1 = 0$$

ⓑ ← $\left[\sum_{\substack{x\in\mathcal{X}\\ P(x)>0}} Q(x) \leq 1\right]$

等号成立

$$\Leftrightarrow \begin{cases} ⓐ \quad \dfrac{P(x)}{Q(x)} = 1 \quad \text{if} \quad P(x) > 0 \\ \text{かつ} \\ ⓑ \quad \displaystyle\sum_{x:P(x)>0} Q(x) = 1 \Leftrightarrow \sum_{x:P(x)=0} Q(x) = 0 \\ \qquad \Leftrightarrow Q(x) = 0 \quad \text{if} \quad P(x) = 0 \end{cases}$$

$$\Leftrightarrow \quad P = Q \qquad \square$$

10/22 <u>Remark</u> $D(P \| Q) \geq 0$

$$\Leftrightarrow \sum_x P(x) \log P(x) \geq \sum_x P(x) \log Q(x)$$

これは Gibbs の不等式ともよばれる

<u>例</u> (レポート) 2値ダイバージェンス

$$d(p, q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}$$

$$(0 \leq p \leq 1, \quad 0 \leq q \leq 1)$$

これは確率分布 $(p, 1-p)$ と $(q, 1-q)$ のダイバージェンス

(1) $y = d(p, q)$ を3次元プロットせよ

(2) $d(p, q) \neq d(q, p)$ となる例を示せ

ダイバージェンスの意味

(1) X の確率分布 P のとき（$X \sim P$ と書く）

□ $X = x$ のとき符号語長 $-\log P(x)$
となるような符号がほぼ最適

$$\boxed{\text{平均符号語長} = -\sum_x P(x) \log P(x) = H(X)} \quad \text{—— ①}$$

□ 一方間違えて、X の分布を Q だと思うと
符号語長 $-\log Q(x)$

$$\text{平均符号語長} = -\sum_x P(x) \log Q(x) \quad \text{—— ②}$$

$$\boxed{\text{最適値からのずれ} = ② - ① = D(P \| Q)}$$

(2) 仮説検定（単純仮説検定）

$D(P \| Q)$ ： P と Q の識別のし易さ

□ Def $X^n = X_1 X_2 \cdots X_n \underset{\text{i.i.d.}}{\sim} P$
(independently and identically distributed, 独立同一分布)

$\iff$ $P_{X^n}(x^n) = P(x_1) P(x_2) \cdots P(x_n) =: P^n(x^n)$ とかく
$(x^n = x_1 x_2 \cdots x_n \in \mathcal{X}^n)$

□ 仮説 $X^n \underset{\text{i.i.d.}}{\sim} P$ または $X^n \underset{\text{i.i.d.}}{\sim} Q$

□ P の受容域 $\mathcal{A}_n \subset \mathcal{X}^n$

X^n の実現値 $x^n \in \mathcal{A}_n$
$\longrightarrow$ P が真と判定

$x^n \notin \mathcal{A}_n$
$\longrightarrow$ Q が真と判定

□ 誤り確率

$$\alpha_n(\mathcal{A}_n) := \sum_{x^n \in \mathcal{A}_n^c} P^n(x^n) \quad \left(\begin{array}{l} P\text{が真のとき} \\ Q\text{だと判定} \end{array}\right)$$

$$\beta_n(\mathcal{A}_n) := \sum_{x^n \in \mathcal{A}_n} Q^n(x^n) \quad \left(\begin{array}{l} Q\text{が真のとき} \\ P\text{だと判定} \end{array}\right)$$

□ $\alpha_n(\mathcal{A}_n)$ と $\beta_n(\mathcal{A}_n)$ はトレードオフの関係

$$\beta_n^*(\varepsilon) := \min_{\substack{\mathcal{A}_n \subset \mathcal{X}^n \\ \alpha_n(\mathcal{A}_n) \leq \varepsilon}} \beta_n(\mathcal{A}_n)$$ とおくと

Thm（Stein の補題）

$$\lim_{n\to\infty} \frac{1}{n} \log \beta_n^*(\varepsilon) = -D(P\|Q)$$
$$\text{for } 0 < \forall \varepsilon < 1$$

すなわち
$\alpha_n(\mathcal{A}_n) \leq \varepsilon$ のとき $\beta_n(\mathcal{A}_n)$ の最適値は

$$\beta_n(\mathcal{A}_n) \simeq e^{-nD(P\|Q)}$$

□ $D(P\|Q)$ が大きい $\Rightarrow$ 識別し易い

1-9 通信路と Markov map

○ $W(y|x)$ を通信路とする $(x \in \mathcal{X},\ y \in \mathcal{Y})$
（条件付き確率）

○ $\mathcal{P}(\mathcal{X})$ ： $\mathcal{X}$ 上の確率分布全体

$\mathcal{P}(\mathcal{Y})$ ： $\mathcal{Y}$ 上の確率分布全体

○ 通信路 W は次の写像 (Markov map) とみなせる

$$
\begin{array}{cccc}
W: & \mathcal{P}(\mathcal{X}) & \longrightarrow & \mathcal{P}(\mathcal{Y}) \\
 & \Psi & & \Psi \\
 & P & \longmapsto & PW
\end{array}
$$

ただし $PW(y) = \sum_{x \in \mathcal{X}} P(x) W(y|x)$

解説

$$x \longrightarrow \boxed{W} \longrightarrow y$$

入力確率変数 　　通信路 　　出力確率変数

$X \sim P_X$ 　　$W(y|x)$ 　　Y 　　とすると

□ 同時分布 $P_{XY}(x,y) = P_X(x) W(y|x)$

$\downarrow$

□ 周辺分布 $P_Y(y) = \sum_{x \in \mathcal{X}} P_X(x) W(y|x)$

$\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad \shortparallel$

$\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad P_X W(y)$

Markov map の例 (周辺分布をとる操作)

□ $P_{XY}(x,y)$ が与えられているとする

□ $f: (x,y) \longmapsto x$ に対する通信路は

$$W_f(x'|(x,y)) = \begin{cases} 1 & \text{if } f(x,y) = x' \\ 0 & f(x,y) \neq x' \end{cases}$$

$\mathcal{X} \times \mathcal{Y}$ から $\mathcal{X}$ への通信路

□ $$P_{XY} W(x') = \sum_{x,y} P_{XY}(x,y) \underbrace{W(x'|x,y)}_{\shortparallel \; \delta_{x,x'}}$$

$$= \sum_{y} P_{XY}(x',y)$$

$$= P_X(x)$$

1-10 条件付ダイバージェンス

Def □ $P_{Y_1|X_1}(y|x)$, $P_{Y_2|X_2}(y|x)$

$(x \in \mathcal{X}, y \in \mathcal{Y})$ 条件付き確率

□ $Q(x)$ $(x \in \mathcal{X})$ $\mathcal{X}$ 上の確率分布

$$D(P_{Y_1|X_1} \| P_{Y_2|X_2} | Q)$$

$$:= \sum_{x \in \mathcal{X}} Q(x) \underbrace{D\left(P_{Y_1|X_1}(\cdot|x) \| P_{Y_2|X_1}(\cdot|x)\right)}_{\text{$x \in \mathcal{X}$ を固定したときの確率分布のダイバージェンス}}$$

Thm (ダイバージェンスの chain rule)

同時分布 $P_{X_1Y_1}$, $P_{X_2Y_2}$ に対して

$$D(P_{X_1Y_1} \| P_{X_2Y_2}) = D(P_{X_1} \| P_{X_2}) + D(P_{Y_1|X_1} \| P_{Y_2|X_2} | P_{X_1})$$

proof

$$P_{X_1Y_1}(x,y) = P_{X_1}(x) P_{Y_1|X_1}(y|x)$$
$$P_{X_2Y_2}(x,y) = P_{X_2}(x) P_{Y_2|X_2}(y|x)$$ を用いると

$$D(P_{X_1Y_1} \| P_{X_2Y_2})$$

$$= \sum_{x,y} P_{X_1Y_1}(x,y) \underbrace{\log \frac{P_{X_1Y_1}(x,y)}{P_{X_2Y_2}(x,y)}}_{= \log \frac{P_{X_1}(x)}{P_{X_2}(x)} + \log \frac{P_{Y_1|X_1}(y|x)}{P_{Y_2|X_2}(y|x)}}$$

$$= \sum_x \underbrace{\sum_y P_{X_1Y_1}(x,y)}_{P_{X_1}(x)} \log \frac{P_{X_1}(x)}{P_{X_2}(x)} + \sum_x \sum_y \underbrace{P_{X_1Y_1}(x,y)}_{= P_{X_1}(x) P_{Y_1|X_1}(y|x)} \log \frac{P_{Y_1|X_1}(y|x)}{P_{Y_2|X_2}(y|x)}$$

$$= D(P_{X_1} \| P_{X_2}) + \sum_x P_{X_1}(x) \sum_y P_{Y_1|X_1}(y|x) \log \frac{P_{Y_1|X_1}(y|x)}{P_{Y_2|X_2}(y|x)}$$

$$= (\text{右辺})$$ □

1−11 ダイバージェンスの単調性

Thm 任意の $P, Q \in \mathcal{P}(\mathcal{X})$
（$\mathcal{X}$ 上の確率分布）
$W(y|x)$ 通信路（$x \in \mathcal{X}, y \in \mathcal{Y}$）に対して

$$D(P \| Q) \geq D(PW \| QW) \quad \text{（単調性）}$$

$\boxed{\text{等号成立}} \iff$ $\exists\, V(x|y)$（$y \in \mathcal{Y}, x \in \mathcal{X}$）逆向きの通信路
s.t. $PWV = P$, $QWV = Q$ ㊍

proof 同時分布を

$$P_{X_1Y_1}(x, y) = P_{X_1}(x) W(y|x)$$
$$P_{X_2Y_2}(x, y) = P_{X_2}(x) W(y|x)$$

で定義すると

chain rule より

$$D(P_{X_1Y_1} \| P_{X_2Y_2}) = D(P_{X_1} \| P_{X_2}) + \underbrace{D(\underset{W}{P_{Y_1|X_1}} \| \underset{W}{P_{Y_2|X_1}} | P_{X_1})}_{=0}$$

$$= D(P_{Y_1} \| P_{Y_2}) + \underbrace{D(P_{X_1|Y_1} \| P_{X_2|Y_2} | P_{Y_1})}_{\geq 0}$$

$$\geq D(P_{Y_1} \| P_{Y_2})$$

等号成立

$$\iff D(P_{X_1|Y_1} \| P_{X_2|Y_2} | P_{Y_1}) = \sum_y P_{Y_1}(y) D(P_{X_1|Y_1}(\cdot|y) \| P_{X_2|Y_2}(\cdot|y)) = 0$$

$$\iff \forall y \in \mathcal{Y},\ P_{Y_1}(y) = 0 \text{ または } \underbrace{D(P_{X_1|Y_1}(\cdot|y) \| P_{X_2|Y_2}(\cdot|y)) = 0}_{\equiv\ P_{X_1|Y_1}(\cdot|y) = P_{X_2|Y_2}(\cdot|y)}$$

$$\iff \forall y \in \mathcal{Y},\ P_{X_1|Y_1}(\cdot|y) = P_{X_2|Y_2}(\cdot|y) \text{ if } P_{Y_1}(y) > 0 \quad \text{—— ①}$$

$(\because \neg A \vee B \equiv A \Rightarrow B)$

ここで W と逆向きの通信路を

$$V = P_{X_2|Y_2} \quad — ②$$

とおく

明らかに

$$P_{X_2Y_2}(x,y) = P_{Y_2}(y)\,P_{X_2|Y_2}(x|y) = P_{Y_2}(y)\,V(x|y)$$

だから周辺分布をとると

$$P_{X_2}(x) = \sum_y P_{Y_2}(y)\,V(x|y)$$

すなわち

$$P_{X_2} = P_{Y_2}V = P_{X_2}WV \quad — ③$$

一方，$P_{Y_1}(y) > 0$ のとき ①② より

$$P_{X_1Y_1}(x,y) = P_{Y_1}(y)\,P_{X_1|Y_1}(x|y) = P_{Y_1}(y)\,V(x|y) \quad — ④$$

$P_{Y_1}(y) = 0$ のとき $P_{X_1Y_1}(x,y) = 0$ $(\forall x \in \mathcal{X})$ だから

このときも ④ は「$0=0$」で成立

よって ④ の周辺分布をとると

$$P_{X_1}(x) = \sum_y P_{Y_1}(y)\,V(x|y)$$

すなわち

$$P_{X_1} = P_{Y_1}V = P_{X_1}WV \quad — ⑤$$

よって，$P_{X_1} = P$, $P_{X_2} = Q$ だから，③⑤より，等号成立 $\Rightarrow$ ⊛ が示せた

逆に ⊛ が成立していると仮定すると，単調性を2回使うと

$$\begin{aligned} D(P\|Q) &\geq D(PW\|QW) \\ &\geq D(\underbrace{PWV}_{=P}\|\underbrace{QWV}_{=Q}) = D(P\|Q) \end{aligned}$$

よって $D(P\|Q) = D(PW\|QW)$ □

1-12 ダイバージェンスの結合凸性

○ 確率分布の凸結合

$\mathcal{X}$ 上の確率分布 $P_1, P_2, \cdots, P_m \in \mathcal{P}(\mathcal{X})$
$\{1, 2, \cdots, m\}$ 上の確率分布 $\pi = (\pi_1, \pi_2, \cdots, \pi_m)$
$\left(\begin{array}{l} \pi_i \geq 0 \ (i=1,2,\cdots,m) \\ \sum \pi_i = 1 \end{array}\right)$

$$\Downarrow$$

$$P_\pi(x) := \sum_{i=1}^{m} \pi_i P_i(x) \qquad (x \in \mathcal{X})$$

とおくと P_π は $\mathcal{X}$ 上の確率分布になる

このとき $P_\pi = \sum_{i=1}^{m} \pi_i P_i$ と書いて

確率分布 $\{P_i\}_{i=1}^{m}$ の π による

凸結合とよぶ

例：$\mathcal{X} = \{1, 2\}$

$P = (p, 1-p)$, $Q = (q, 1-q)$

$(0 \leq p \leq 1)$ $(0 \leq q \leq 1)$

$\pi = (t, 1-t)$ $(0 \leq t \leq 1)$

凸結合 $tP + (1-t)Q = (r, 1-r)$

ただし $r = tp + (1-t)q$

Thm（ダイバージェンスの結合凸性）

P_i, Q_i $(i = 1, 2, \cdots, m)$ を確率分布

$(\pi_1, \pi_2, \cdots, \pi_m)$ を $\{1, \cdots, m\}$ 上の確率分布

とすると

$$\sum_{i=1}^{m} \pi_i D(P_i \| Q_i) \geq D\left(\sum_{i=1}^{m} \pi_i P_i \,\Big\|\, \sum_{i=1}^{m} \pi_i Q_i\right) \quad \text{①}$$

特に
$$\sum_{i=1}^{m} \pi_i D(P_i \| Q) \geq D\left(\sum_{i=1}^{m} \pi_i P_i \,\Big\|\, Q\right)$$
$$\sum_{i=1}^{m} \pi_i D(P \| Q_i) \geq D\left(P \,\Big\|\, \sum_{i=1}^{m} \pi_i Q_i\right)$$

(証明)

$$\tilde{P}(i,x) = \pi_i P(x) \quad , \quad \tilde{Q}(i,x) = \pi_i Q(x)$$

$$(i = 1,2,\cdots,m \quad , \quad x \in \mathcal{X})$$

とおくと，これらは $\{1,2,\cdots,m\} \times \mathcal{X}$ 上の確率分布で，

周辺分布

$$\sum_{i=1}^{m} \tilde{P}(i,x) = \sum_{i=1}^{m} \pi_i P(x) \quad , \quad \sum_{i=1}^{m} \tilde{Q}(i,x) = \sum_{i=1}^{m} \pi_i Q_i(x)$$

は凸結合である ―― ②

$$D(\tilde{P} \| \tilde{Q}) = \sum_{i=1}^{m} \sum_{x \in \mathcal{X}} \pi_i P_i(x) \log \frac{\pi_i P_i(x)}{\pi_i Q_i(x)}$$

$$= \sum_{i=1}^{m} \pi_i \sum_{x \in \mathcal{X}} P_i(x) \log \frac{P_i(x)}{Q_i(x)}$$

$$= \sum_{i=1}^{m} \pi_i D(P_i \| Q_i) = (\text{① 左辺})$$

周辺分布をとる操作は通信路であるから

(Markov map 20ページ)

ダイバージェンスの単調性 (22ページ)

と ② より ① が示された □

1-13 相互情報量 (mutual information)

Def 同時分布 $P_{XY}(x,y)$ に対して

$$I(X;Y) := \sum_{x} \sum_{y} P_{XY}(x,y) \log \frac{P_{XY}(x,y)}{P_X(x) P_Y(y)}$$

同時分布 $P_{XY}(x,y)$ と周辺分布の積 $P_X(x) P_Y(y)$

のダイバージェンス

○ 相互情報量の様々な表式

□ $I(X;Y) = E_{P_{XY}}\left[\log \frac{P_{XY}(X,Y)}{P_X(X)P_Y(Y)}\right]$

□ $I(X;Y) = H(X) + H(Y) - H(X,Y)$ —— ①

$= H(X) - H(X|Y)$ —— ②

$= H(Y) - H(Y|X)$ —— ③

$I(X;Y) = E[-\log P_X(X)] + E[-\log P_Y(Y)] - E[-\log P_{XY}(X,Y)]$

$= H(X) + H(Y) - H(X,Y) = ①$

エントロピーの chain rule (14ページ)

$H(X,Y) = H(X) + H(Y|X)$

$H(X,Y) = H(Y) + H(X|Y)$

より

$H(Y|X) = H(X,Y) - H(X)$

$H(X|Y) = H(X,Y) - H(Y)$

これと ① より ② ③ が示される

□ $I(X;Y) = \sum_{x\in\mathcal{X}} P_X(x)\, D(P_{Y|X}(\cdot|x) \| P_Y)$

$P_{XY}(x,y) = P_X(x)P_{Y|X}(y|x)$ を用いると

$$I(X;Y) = \sum_x\sum_y P_{XY}(x,y)\log\frac{P_{XY}(x,y)}{P_X(x)P_Y(y)}$$

$$= \sum_x\sum_y P_X(x)P_{Y|X}(y|x)\log\frac{P_X(x)P_{Y|X}(y|x)}{P_X(x)P_Y(y)}$$

$$= \sum_x P_X(x)\underbrace{\sum_y P_{Y|X}(y|x)\log\frac{P_{Y|X}(y|x)}{P_Y(y)}}_{\parallel\ D(P_{Y|X}(\cdot|x)\|P_Y)}$$

Lem（相互情報量の正値性）

$$I(X:Y) \geq 0$$

等号成立 $\Leftrightarrow$ X, Y 独立

∵ ダイバージェンスの正値性（16ページ）より明らか

等号成立 $\Leftrightarrow$ $P_{XY}(x,y) = P_X(x) P_Y(y)$

$\Leftrightarrow$ X, Y 独立 □

Def（条件付き相互情報量）

conditional mutual information

$P_{XYZ}(x,y,z)$ に対して

$$I(X;Z|Y) := E_{P_{XYZ}}\left[\log \frac{P_{XZ|Y}(X,Z|Y)}{P_{X|Y}(X|Y)\,P_{Z|Y}(Z|Y)}\right]$$

$$= \sum_x \sum_y \sum_z \underbrace{P_{XYZ}(x,y,z)}_{= P_Y(y) P_{XZ|Y}(x,z|y)} \log \frac{P_{XZ|Y}(x,z|y)}{P_{X|Y}(x|y)\,P_{Z|Y}(z|y)}$$

$$= \sum_y P(y) \underbrace{\sum_x \sum_z P(x,z|y) \log \frac{P(x,z|y)}{P(x|y)\,P(z|y)}}_{= I(X:Z|y)\text{ とかく}}$$

④

［ y を固定したときの x と z の条件付き分布 $P_{XZ|Y}(x,z|y)$ についての相互情報量 ］

Lem（条件付き相互情報量の正値性）

$$I(X;Z|Y) \geq 0$$

等号成立 $\Longleftrightarrow$ $P(x,z|y) = P(x|y)P(z|y)$ if $P(y)>0$

∵ ④と相互情報量の正値性より明らか

等号成立 $\Longleftrightarrow$ $I(X;Z|y)=0$ if $P(y)>0$

$\Longleftrightarrow$ $P(x,z|y) = P(x|y)P(z|y)$ if $P(y)>0$

上式の等号成立条件が成立するとき

<u>X, Y, Z は Markov（または short Markov）</u>

と言い <u>$X \to Y \to Z$ と書く</u>

yの条件のもとで
xとzが独立

○ $X \to Y \to Z \overset{def}{\Longleftrightarrow} P(x,z|y) = P(x|y)P(z|y)$ if $P(y)>0$

$$\Longleftrightarrow \underbrace{P(y)P(x,z|y)}_{P(x,y,z)} = \underbrace{P(y)P(x|y)}_{P(x,y)}P(z|y)$$

$$\Longleftrightarrow P(x,y,z) = P(x)P(y|x)P(z|y)$$

すなわち $X \to Y \to Z$ は下図の状況

$$X \to \boxed{P_{Y|X}} \to Y \to \boxed{P_{Z|Y}} \to Z$$

○ xとzの対称性から

$$X \to Y \to Z \Longleftrightarrow Z \to Y \to X$$

1-14 相互情報量の単調性 (データ処理不等式)

Lem

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

∵
$$
\begin{aligned}
I(X:Z|Y) &= E\left[\log \frac{P_{XZ|Y}(X,Z|Y)}{P_{X|Y}(X|Y)\,P_{Z|Y}(Z|Y)}\right] \\
&= E[-\log P_{X|Y}(X|Y)] + E[-\log P_{Z|Y}(Z|Y)] \\
&\qquad - E[-\log P_{XZ|Y}(X,Z|Y)] \\
&= H(X|Y) + H(Z|Y) - H(X,Z|Y) \qquad \square
\end{aligned}
$$

Thm (相互情報量の chain rule)

$$I(X;YZ) = I(X;Y) + I(X;Z|Y)$$

(証明)

$$I(X;Y) = H(X) + H(Y) - H(XY)$$
$$I(X:Z|Y) = H(X|Y) + H(Z|Y) - H(X,Z|Y)$$

より

$$
\begin{aligned}
I(X:Y) + I(X;Z|Y) &= H(X) + \underbrace{\cancel{H(Y) + H(X|Y)}}_{H(XY)} + H(Z|Y) \\
&\qquad \cancel{-H(XY)} - H(X,Z|Y) \\
&= H(X) + H(Z|Y) - H(X,Z|Y)
\end{aligned}
$$

chain rule
$$
\begin{aligned}
&= H(X) + \{H(YZ) - \cancel{H(Y)}\} - \{H(XYZ) - \cancel{H(Y)}\} \\
&= H(X) + H(YZ) - H(XYZ) \\
&= I(X;YZ) \qquad \square
\end{aligned}
$$

Thm（相互情報量の単調性）

$X \to Y \to Z$ のとき

$$I(X;Y) \geq I(X;Z) \qquad \text{—— ①}$$

等号成立 $\Longleftrightarrow$ $Y \to Z \to X$

（証明）

chain rule より

$$I(X;YZ) = I(X;Y) + \underbrace{I(X;Z|Y)}_{=0} \qquad \text{—— ②}$$

仮定 $X \to Y \to Z$ と 28ページ

一方で，再び chain rule より

$$I(X;YZ) = I(X;Z) + I(X;Y|Z)$$

$$\geq I(X;Z) \qquad \text{—— ③}$$

$\uparrow$ 条件付き相互情報量の正値性（28ページ）

$I(X;Y|Z) \geq 0$

② ③ 合わせて

$$I(X;Y) \geq I(X;Z)$$

等号成立 $\Longleftrightarrow$ $I(X;Y|Z) = 0$

$\Longleftrightarrow$ $Y \to Z \to X$ $\square$

Remark

① の等号成立条件は

$$X \to \boxed{P_{Y|X}} \to Y \to \boxed{P_{Z|Y}} \to Z \to \boxed{P_{X|Z}} \to X$$

のように Z のみから X を復元可能なことである

1-15 エントロピーとダイバージェンス

○ エントロピーとダイバージェンスの関係

$$P_U(u) = \frac{1}{|\mathcal{X}|} \quad (x \in \mathcal{X}) \quad \text{一様分布}$$

(uniform distribution)

とすると

$$H(X) = \log|\mathcal{X}| - D(P_X \| P_U) \quad \text{—— ①}$$

(∵) $D(P_X \| P_U) = \sum_x P_X(x) \log \dfrac{P_X(x)}{1/|\mathcal{X}|}$

$= \log|\mathcal{X}| - H(X)$ □

—— ②

Lem (エントロピーの性質)

(a) $H(X) \leq \log|\mathcal{X}|$

等号成立 ⇔ X は一様分布

(b) エントロピーは上に凸 (concave)

$$H\left(\sum_{i=1}^{m} \pi_i P_i\right) \geq \sum \pi_i H(P_i)$$

(c) $H(X) \geq H(X|Y)$

等号成立 ⇔ X と Y が独立

(d) $H(X,Y) \leq H(X) + H(Y)$ (劣加法性)

等号成立 ⇔ X と Y が独立

(∵) (a) ② より $D(P_X \| P_U) = \log|\mathcal{X}| - H(X) \geq 0$

等号成立 ⇔ $P_X = P_U$

(b) ① とダイバージェンスの凸性 (24ページ) より

$$H\left(\sum_i \pi_i P_i\right) = \log|\mathcal{X}| - D\left(\sum_i \pi_i P_i \| P_U\right)$$
$$\geq \log|\mathcal{X}| - \sum_i \pi_i D(P_i \| P_U)$$
$$= \sum_i \pi_i \{\underbrace{\log|\mathcal{X}| - D(P_i \| P_U)}_{= H(P_i)}\}$$

(c) (d) $I(X;Y) = H(X) - H(X|Y)$

$= H(X) + H(Y) - H(X,Y)$

と相互情報量の正値性 (27ページ) より示される

2. データ圧縮

2-1 大数の法則 (Law of large numbers)

Thm $X^n = X_1 X_2 \cdots X_n \underset{\text{i.i.d.}}{\sim} P$ のとき

$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ (算術平均) とおくと

$\forall \varepsilon > 0$ に対して

$$\lim_{n\to\infty} \Pr\{ |S_n - \mu| > \varepsilon \} = 0$$

ただし $\mu = E[X_1]$

(大数の弱法則)

以下で証明を行う

Lem (Markov の不等式)

Z を非負の値をとる確率変数とすると

$\forall a > 0$ に対して

$$\Pr\{ Z \geq a \} \leq \frac{E[Z]}{a}$$

(∵)
$$E[Z] = \sum_{z:z\geq 0} z P(z)$$
$$= \sum_{z:z\geq a} z P(z) + \underbrace{\sum_{z:0\leq z<a} z P(z)}_{\geq 0}$$
$$\geq \sum_{z:z\geq a} a P(z)$$
$$= a \underbrace{\sum_{z:z\geq a} P(z)}_{=\Pr\{Z\geq a\}}$$

両辺を a でわればよい □

Lem (Chebyshevの不等式)

$\forall a > 0$ に対して

$$\Pr\{ |X-\mu| \geq a \} \leq \frac{V[X]}{a^2}$$

ただし $\mu = E[X]$

∵ $x \in \mathcal{X}$ について

$$|x-\mu| \geq a \iff (x-\mu)^2 \geq a^2$$

だから

$$\Pr\{ |X-\mu| \geq a \} = \Pr\{ (X-\mu)^2 \geq a^2 \}$$

$$\underset{\text{Markovの不等式}}{\leq} \frac{E[(X-\mu)^2]}{a^2} = \frac{V[X]}{a^2}$$

□

○ 分散の性質

□ $V[X] = E[(X-\mu)^2]$ （定義）

□ $V[aX+b] = a^2 V[X] \quad a, b \in \mathbb{R}$

∵ $E[aX+b] = aE[X] + b = a\mu + b$ だから

$$V[aX+b] = E[\{(aX+b)-(a\mu+b)\}^2]$$
$$= E[\{a(X-\mu)\}^2]$$
$$= a^2 E[(X-\mu)^2] = a^2 V[X]$$

□ X と Y が独立なとき

$$V[X+Y] = V[X] + V[Y]$$

∵ $\mu_X = E[X]$，$\mu_Y = E[Y]$ とおくと
$E[X+Y] = \mu_X + \mu_Y$ だから

$$V[X+Y] = E[\{(X+Y)-(\mu_X+\mu_Y)\}^2]$$
$$= E[\{(X-\mu_X)+(Y-\mu_Y)\}^2]$$
$$= \underset{= V[X]}{E[(X-\mu_X)^2]} + \underbrace{E[(X-\mu_X)(Y-\mu_Y)]}_{\underset{\text{独立性}}{=} E[(X-\mu_X)]E[(Y-\mu_Y)] = 0} + \underset{= V[Y]}{E[(Y-\mu_Y)^2]}$$

□

大数の法則の証明

Chebyshev の不等式より

$$\Pr\{ |S_n - \mu| \geq \varepsilon \} \leq \frac{V[S_n]}{\varepsilon^2}$$

ここで
$$V[S_n] = V\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]$$
$$= \frac{1}{n^2} V\left[\sum_{i=1}^{n} X_i\right]$$
$$= \frac{1}{n^2} \cdot \sum_{i=1}^{n} \underbrace{V[X_i]}_{= V[X_1]}$$
$$= \frac{1}{n} V[X_1]$$

よって
$$\Pr\{ |S_n - \mu| \geq \varepsilon \} \leq \frac{V[X_1]}{n\varepsilon^2} \xrightarrow[(n\to\infty)]{} 0 \qquad \square$$

2-2 典型系列 (typical sequence)

Def $X^n = X_1 X_2 \cdots X_n \underset{\text{i.i.d.}}{\sim} P$ のとき

$\varepsilon > 0$ に対して

$$A_{n,\varepsilon} := \left\{ x^n \in \mathcal{X}^n \;\middle|\; \left| -\frac{1}{n} \log P^n(x^n) - H(P) \right| \leq \varepsilon \right\}$$

とおき $A_{n,\varepsilon}$ の要素 $x^n = x_1 x_2 \cdots x_n$ を典型系列とよぶ

Thm (典型系列の性質)

(1) $\forall \varepsilon > 0$ に対して $\lim_{n\to\infty} \Pr\{A_{n,\varepsilon}\} = 1$

(2) $x^n \in A_{n,\varepsilon}$
$$\Leftrightarrow \quad e^{-n\{H(P)+\varepsilon\}} \leq P^n(x^n) \leq e^{-n\{H(P)-\varepsilon\}}$$

(3) $|A_{n,\varepsilon}| \leq e^{n\{H(P)+\varepsilon\}}$

(4) $\forall \delta > 0$ に対して n が十分大きいとき
$$|A_{n,\varepsilon}| \geq (1-\delta)\, e^{n\{H(P)-\varepsilon\}}$$

（証明）

(1) $P^n(x^n) = P(x_1)P(x_2)\cdots P(x_n)$ だから

$$-\frac{1}{n}\log P^n(x^n) = \frac{1}{n}\sum_{i=1}^{n} -\log P(x_i)$$

一方 $E[-\log P(X_i)] = H(P)$ なので，大数の法則より

$$\Pr\{A_{n,\varepsilon}^c\} = \Pr\{|-\frac{1}{n}\log P^n(X^n) - H(P)| > \varepsilon\}$$

補集合

$$= \Pr\{|\frac{1}{n}\sum_{i=1}^{n} -\log P(X_i) - H(P)| > \varepsilon\}$$

$$\longrightarrow 0 \quad (n\to\infty)$$

よって $\lim_{n\to\infty}\Pr\{A_{n,\varepsilon}\} = 1 - \lim_{n\to\infty}\Pr\{A_{n,\varepsilon}^c\}$

$$= 1$$

(2) $x^n \in A_{n,\varepsilon}$

$\iff |-\frac{1}{n}\log P^n(x^n) - H(P)| \le \varepsilon$

$\iff -\varepsilon \le -\frac{1}{n}\log P^n(x^n) - H(P) \le \varepsilon$

$\iff H(P) - \varepsilon \le -\frac{1}{n}\log P^n(x^n) \le H(P)+\varepsilon$

$\iff -n\{H(P)+\varepsilon\} \le \log P^n(x^n) \le -n\{H(P)-\varepsilon\}$

$\iff e^{-n\{H(P)+\varepsilon\}} \le P^n(x^n) \le e^{-n\{H(P)-\varepsilon\}}$

(3) $$1 = \sum_{x^n\in\mathcal{X}^n} P^n(x^n) \ge \sum_{x^n\in A_{n,\varepsilon}} P^n(x^n)$$

$$\overset{(2)}{\ge} \sum_{x^n\in A_{n,\varepsilon}} e^{-n\{H(P)+\varepsilon\}}$$

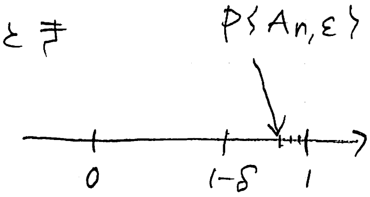
$$= |A_{n,\varepsilon}|\, e^{-n\{H(P)+\varepsilon\}}$$

両辺 $e^{n\{H(P)+\varepsilon\}}$ をかければよい

(4) (1) より $\lim_{n\to\infty} \Pr\{A_{n,\varepsilon}\} = 1$ だから

$\forall \delta > 0$ に対して，n が十分大きいとき

$$\Pr\{A_{n,\varepsilon}\} \geq 1-\delta$$



これより

$$1-\delta \leq \Pr\{A_{n,\varepsilon}\} = \sum_{x^n \in A_{n,\varepsilon}} P^n(x^n)$$

$$\xrightarrow{(2)} \leq \sum_{x^n \in A_{n,\varepsilon}} e^{-n\{H(P)-\varepsilon\}}$$

$$= |A_{n,\varepsilon}|\, e^{-n\{H(P)-\varepsilon\}}$$

両辺に $e^{n\{H(P)-\varepsilon\}}$ をかければよい $\square$

2-3 情報源符号化定理

○ データ圧縮の種類

□ FF 符号化 ： データのブロック長 n (固定 fixed) $\longrightarrow$ 符号語ブロック長 m (固定 fixed)

小さな誤りを許す．

□ FV 符号化 ： データのブロック長 n (固定 fixed) $\longrightarrow$ 符号語 (可変長 variable)

□ VF，VV もある

○ FF 符号化 (binary の場合)

$$x^n = x_1 x_2 \cdots x_n \xrightarrow{f_n} y_1 y_2 \cdots y_m \xrightarrow{g_n} x_1 x_2 \cdots x_n$$

$\boxed{\text{情報源}}$ $\quad x^n \in \{0,1\}^n$ $\quad \boxed{\text{符号器}}$ encoder $\quad \boxed{\text{符号語}}$ codeword $y_1\cdots y_m \in \{0,1\}^m$ $\quad \boxed{\text{復号器}}$ decoder $\quad x_1\cdots x_n \in \{0,1\}^n$

$$\text{圧縮率（レート）} = \frac{m}{n}$$

$\boxed{\text{目的}}$

圧縮率 $\longrightarrow$ 小さく

誤り $\longrightarrow$ 0 $(n\to\infty)$

} 圧縮率の限界は？

Remark

誤りを全く許さない場合は圧縮できない！

この場合 符号器は単射でなければならないので

集合の要素数は

$$|\{0,1\}^n| \leq |\{0,1\}^m|$$

を満たす必要がある。良くても $m=n$

圧縮にならない

○ 簡単のため符号語（codeword）を単なるインデックスにする

$$\boxed{\begin{array}{ccccc} x^n = x_1 x_2 \cdots x_n & \xrightarrow{f_n} & f(x^n) = k & \xrightarrow{g_n} & g_n(f_n(x^n)) \\ \cap & & \cap & & \\ \mathcal{X}^n & & \{1,2,\cdots,M_n\} & & \end{array}}$$

□ $$|\{1,2,\cdots,M_n\}| \leq |\{0,1\}^m|$$

$$\iff M_n \leq 2^m$$

$$\iff \log_2 M_n \leq m \quad \text{だから}$$

$m = \lceil \log_2 M_n \rceil$ のとき $\{1,2,\cdots,M_n\}$ は m ケタの二進列で表せる

よって圧縮率は $\dfrac{\lceil \log_2 M_n \rceil}{n}$

（注： $\lceil a \rceil$ は a の小数を切り上げた整数
$b \geq a$ となる最小の整数）

□ $n \to \infty$ の状況では

$$0 \leq \frac{\lceil \log_2 M_n \rceil - \log_2 M_n}{n} \leq \frac{1}{n} \xrightarrow[n\to\infty]{} 0$$

なので

$\boxed{\text{圧縮率は } \dfrac{\log M_n}{n}}$ としてよい

□ log の底の違いは定数倍の違いなので
当面気にしなくてよい

- 情報源は

$$X^n = X_1 X_2 \cdots, X_n \underset{\text{i.i.d.}}{\sim} P$$

とする

- 誤り確率

$$\varepsilon_n = \Pr\{ g_n(f_n(X^n)) \neq X^n \}$$

- 達成可能レート (achievable rate)

R は achievable

$$\overset{\text{def}}{\Longleftrightarrow} \quad \exists f_n, g_n \quad (n=1,2,\cdots)$$

$$\Pr\{ g_n(f_n(X^n)) \neq X^n \} \longrightarrow 0 \quad (n\to\infty)$$

かつ $\displaystyle \lim_{n\to\infty} \frac{\log M_n}{n} \leq R$

- 達成可能限界

$$R^*(P) = \inf\{ R \mid R \text{ は achievable} \}$$

<u>Thm</u> (固定長情報源符号化定理)

$$R^*(P) = H(P)$$

(証明) (a) $R > H(P) \Rightarrow R$ は achievable (direct part)

(b) $R < H(P) \Rightarrow R$ は achievable でない (converse part)

を示す

<u>(a) direct part</u>

$R > H(P)$ を仮定する。

このとき $R \geq H(P) + \varepsilon$ をみたす $\varepsilon > 0$ を固定

$$M_n = e^{n\{H(P)+\varepsilon\}} \quad \text{——} \ ①$$

P_X の典型系列の集合 $A_{n,\varepsilon}$ について

$$|A_{n,\varepsilon}| \leq M_n = e^{n\{H(P)+\varepsilon\}} \quad \text{——} \ ②$$

(2-2, 34ページ)

□ そこで 符号器 f_n を以下のように構成する

(1) $A_{n,\varepsilon}$ の元を順番に並べて

$1, 2, \cdots$ の番号を付ける．この番号は $\overbrace{\text{高々} M_n \text{まで}}^{②\text{より}}$

この対応を $\alpha_n : A_{n,\varepsilon} \longrightarrow \{1, 2, \cdots, M_n\}$

$$x^n \longmapsto \alpha_n(x^n)$$

とする

(2)

$$f_n(x^n) = \begin{cases} \alpha_n(x_n) & \text{if } x^n \in A_{n,\varepsilon} \\ 1 & \text{otherwise} \end{cases}$$

□ 復号器 g_n を

$$g_n(k) = \alpha_n^{-1}(k)$$

とすると、作り方から $A_{n,\varepsilon}$ の元は正しく復号される

$$g_n(f_n(x^n)) = x^n \quad (x^n \in A_{n,\varepsilon})$$

このとき誤り確率は

$$\varepsilon_n = \Pr\{g_n(f_n(x_n)) \neq x_n\}$$

$$\leq \Pr\{A_{n,\varepsilon}^c\} \longrightarrow 0 \quad (n \to \infty) \quad \text{— ③}$$

（34ページ Thm (1)）

圧縮レートは ① より

$$\lim_{n\to\infty} \underbrace{\frac{1}{n} \log M_n}_{= H(P)+\varepsilon} = H(P) + \varepsilon \leq R \quad \text{— ④}$$

をみたす，よって ③④ より R は achievable □

(b) converse part

(b) $\overset{\text{対偶}}{\Longleftrightarrow}$ 「R が achievable $\Rightarrow R \geq H(P)$」

を示す

R が achievable と仮定すると

$f_n, g_n \ (n=1,2,\cdots)$ が存在して

$$\begin{cases} \varepsilon_n = \Pr\{g_n(f_n(X_n)) \neq X_n\} \underset{(n\to\infty)}{\longrightarrow} 0 & \text{— ⑤} \\ \lim_{n\to\infty} \frac{1}{n}\log M_n \leq R & \text{— ⑥} \end{cases}$$

$S_n = \{x^n \in \mathcal{X}^n \mid g_n(f_n(x^n)) = x^n\}$ とおくと

$$S_n = S_n \cap (\underbrace{A_{n,\varepsilon} \cup A_{n,\varepsilon}^c}_{\mathcal{X}^n \text{ 全体}}) = (S_n \cap A_{n,\varepsilon}) \cup (S_n \cap A_{n,\varepsilon}^c)$$

ド・モルガン

だから

$$1-\varepsilon_n = \Pr\{S_n\} = \Pr\{(S_n \cap A_{n,\varepsilon}) \cup (S_n \cap A_{n,\varepsilon}^c)\}$$

$$\leq \Pr\{S_n \cap A_{n,\varepsilon}\} + \Pr\{S_n \cap A_{n,\varepsilon}^c\}$$

$$\leq \sum_{x^n \in S_n \cap A_{n,\varepsilon}} P^n(x^n) + \Pr\{A_{n,\varepsilon}^c\}$$

$$\sum_{x^n \in S_n \cap A_{n,\varepsilon}} P^n(x^n) \leq \sum_{x^n \in S_n \cap A_{n,\varepsilon}} e^{-n\{H(P)-\varepsilon\}} = |S_n \cap A_{n,\varepsilon}| e^{-n\{H(P)-\varepsilon\}}$$

(34ページ Thm (2))

$$\leq |S_n| e^{-n\{H(P)-\varepsilon\}}$$

よって

$$|M_n| \geq |S_n| \geq e^{n\{H(P)-\varepsilon\}} [1-\varepsilon_n - \Pr\{A_{n,\varepsilon}^c\}]$$

$$\therefore \quad \frac{1}{n}\log M_n \geq H(P) - \varepsilon - \frac{1}{n}\log[1-\varepsilon_n - \Pr\{A_{n,\varepsilon}^c\}]$$

$n\to\infty$ で $\to 0$

よって

$$R \underset{⑥}{\geq} \lim \frac{1}{n}\log M_n \geq H(P) - \varepsilon$$

$\varepsilon > 0$ は任意にとれるので

$$R \geq H(P)$$

□

3. 情報理論と大偏差原理

3-1. タイプ(経験分布)

Def 系列 $x^n = x_1 x_2 \cdots x_n \in \mathcal{X}^n$ に対して

$$P_{x^n}(a) = \frac{N(a|x^n)}{n} \quad (a \in \mathcal{X})$$

ただし $N(a|x^n)$ = 「x^n における a の生起回数」

$$= \sum_{i=1}^{n} \delta_{x_i, a}$$

とおくと P_{x^n} は $\mathcal{X}$ 上の確率分布になる

□ P_{x^n} を x^n の 経験分布 (empirical distribution) または タイプ (type) とよぶ

Def

□ $T^n(P)$ = 「タイプが P となる系列の集合」

$$= \{ x^n \in \mathcal{X}^n \mid P_{x^n} = P \}$$

□ $\mathcal{P}_n$ = 「長さ n の系列から作られるタイプ全体」

$$= \{ P_{x^n} \mid x^n \in \mathcal{X}^n \}$$

例 $\mathcal{X} = \{0, 1\}$ の場合

□ $x^n = 01001 \quad (n=5)$

$$P_{x^n}(0) = \frac{3}{5}, \quad P_{x^n}(1) = \frac{2}{5}$$

□
$$\mathcal{P}_n = \left\{ \left(\frac{0}{n}, \frac{n}{n}\right), \left(\frac{1}{n}, \frac{n-1}{n}\right), \left(\frac{2}{n}, \frac{n-2}{n}\right), \cdots, \left(\frac{n-1}{n}, \frac{1}{n}\right), \left(\frac{n}{n}, \frac{0}{n}\right) \right\}$$

例（続き）

□ $P = \left(\frac{3}{5}, \frac{2}{5}\right)$, $n=5$ のとき

$T^n(P)$ = 「0が3回、1が2回現れる長さ5の系列の集合」

[$= \{00011, 00101, 00110, 01001, 01010, 01100, 10010, 10001, 10100, 11000\}$]

$$|T^n(P)| = \frac{5!}{3!\,2!}$$

<u>Lem</u>

$$|\mathcal{P}_n| \le (n+1)^{|\mathcal{X}|-1} \le (n+1)^{|\mathcal{X}|}$$

すなわち<u>タイプの数は高々nの多項式オーダー</u>

☺ $\mathcal{X} = \{1, 2, \cdots, m\}$ のとき示せばよい

$$P_{x^n} = (P_{x^n}(1), P_{x^n}(2), \cdots, P_{x^n}(m-1), P(m))$$

↑

$\frac{0}{n}, \frac{1}{n} \sim \frac{n}{n}$ の$(n+1)$通り

P_{x^n}は確率分布（和が1）だから

$P_{x^n}(1) \sim P_{x^n}(m-1)$

が定まると $P_{x^n}(m)$ も定まる。

よって $|\mathcal{P}_n| \le (n+1)^{m-1} = (n+1)^{|\mathcal{X}|-1}$ □

Remark $|\mathcal{P}_n| = \binom{n+|\mathcal{X}|-1}{|\mathcal{X}|-1}$

Thm　$Q^n(x^n) = Q(x_1)Q(x_2)\cdots Q(x_n)$ のとき

$$Q^n(x^n) = e^{-n\{H(P_{x^n}) + D(P_{x^n}\|Q)\}} \quad \text{―― ①}$$

（x^n の確率はタイプ P_{x^n} にしかよらない）

（証明）

$$Q^n(x^n) = \prod_{a\in\mathcal{X}} Q(a)^{N(a|x^n)}$$

$$= e^{\sum_a N(a|x^n)\log Q(a)}$$

$$= e^{n\sum_a P_{x^n}(a)\log Q(a)} \quad \text{―― ②}$$

$\leftarrow$ （$N(a|x^n) = nP_{x^n}(a)$）

一方で，

$$H(P_{x^n}) + D(P_{x^n}\|Q)$$

$$= -\sum_{a\in\mathcal{X}} P_{x^n}(a)\cancel{\log P_{x^n}(a)} + \sum_{a\in\mathcal{X}} P_{x^n}(a)\{\cancel{\log P_{x^n}(a)} - \log Q(a)\}$$

$$= -\sum_{a\in\mathcal{X}} P_{x^n}(a)\log Q(a) \quad \text{―― ③}$$

よって ② ③ より ① が示された □

系　$P_{x^n} = P$ のとき（x^n のタイプが P のとき）

$$P^n(x^n) = e^{-nH(P)}$$

∵ ① において $Q = P$ とおけばよい □

Lem　整数 m, n について

$$\frac{m!}{n!} \geq n^{m-n} \quad \text{——④}$$

∵　$m \geq n$ のとき

$$\frac{m!}{n!} = \underbrace{m \times (m-1) \times \cdots \times (n+1)}_{(m-n)}$$

$$\geq n \times n \times \cdots \times n$$

$$= n^{m-n}$$

$m < n$ のとき

$$\frac{m!}{n!} = \frac{1}{\underbrace{n \times (n-1) \times \cdots \times (m+1)}_{(n-m)}}$$

$$\geq \frac{1}{n \times n \times \cdots \times n}$$

$$= \frac{1}{n^{n-m}} = n^{m-n}$$

□

Lem　すべてのタイプ $P, P' \in \mathcal{P}_n$ について

$$P^n(T^n(P)) \geq P^n(T^n(P')) \quad \text{——⑤}$$

が成りたつ

(証明)　任意のタイプ $P' \in \mathcal{P}_n$ に対して

$$P^n(T^n(P')) = \sum_{x^n \in T^n(P')} P^n(x^n)$$

$$= \sum_{x^n \in T^n(P')} \underbrace{\prod_{a \in \mathcal{X}} P(a)^{nP'(a)}}_{\text{タイプにしかよらない}}$$

$$= |T^n(P')| \prod_{a \in \mathcal{X}} P(a)^{nP'(a)}$$

［系列の文字がすべて異るときの順列数］

ここで

$$|T^n(P')| = \frac{n!}{\prod_{a \in \mathcal{X}} (nP'(a))!}$$

［各 a について，a の生起回数 $(nP'(a))$ の順列 $(nP'(a))!$ だけ重複］

だから

$$P^n(T^n(P')) = \frac{n!}{\prod_a (nP'(a))!} \prod_a P(a)^{nP'(a)} \quad \text{— ⑥}$$

よって

$$\frac{P^n(T^n(P))}{P^n(T^n(P'))} = \frac{\prod_a (nP'(a))!}{\prod_a (nP(a))!} \frac{\prod_a P(a)^{nP(a)}}{\prod_a P(a)^{nP'(a)}}$$

［⑥と⑥で $P'=P$ とした式より］

$$④ \longrightarrow \geq \prod_a (n\cancel{P(a)})^{n\{P'(a)-P(a)\}} \cancel{P(a)}^{n\{P(a)-P'(a)\}}$$

$$= \prod_a n^{n\{P'(a)-P(a)\}}$$

$$= n^{n\{\sum_a P'(a)-P(a)\}}$$

$$= n^{n\{1-1\}} = n^0 = 1$$

□

Thm 任意のタイプ $P \in \mathcal{P}_n$ に対して

$$\frac{1}{(n+1)^{|\mathcal{X}|-1}} e^{nH(P)} \leq |T^n(P)| \leq e^{nH(P)} \quad \text{— ⑦}$$

(証明)

$$1 \geq P^n(T^n(P))$$

$$= \sum_{x^n \in T^n(P)} \underbrace{P^n(x^n)}_{= e^{-nH(P)}}$$

(43ページ系)

$$= |T^n(P)|\, e^{-nH(P)} \quad \text{—— ⑧}$$

よって $\quad |T^n(P)| \leq e^{nH(P)}$

一方で,

$$1 = \sum_{x^n \in \mathcal{X}^n} P^n(x^n)$$

$$= \sum_{P' \in \mathcal{P}_n} \underbrace{\sum_{x^n \in T^n(P')} P^n(x^n)}_{= P^n(T^n(P'))}$$

$$\leq \sum_{P' \in \mathcal{P}_n} \underbrace{\max_{P' \in \mathcal{P}_n} P^n(T^n(P'))}_{= P^n(T^n(P))} \quad (\because ⑤)$$

$$= |\mathcal{P}_n| \cdot P^n(T^n(P))$$

$$\leq (n+1)^{|\mathcal{X}|-1} P^n(T^n(P))$$

↑ 42ページ Lem

$$= (n+1)^{|\mathcal{X}|-1} |T^n(P)|\, e^{-nH(P)}$$

(⑧より)

よって

$$\frac{1}{(n+1)^{|\mathcal{X}|-1}} e^{nH(P)} \leq |T^n(P)|$$

□

Thm 任意のタイプ $P \in \mathcal{P}_n$ と
確率分布 $Q \in \mathcal{P}(\mathcal{X})$ に対して

$$\frac{1}{(n+1)^{|\mathcal{X}|-1}} e^{-nD(P\|Q)} \leq Q^n(T^n(P)) \leq e^{-nD(P\|Q)} \quad ——⑨$$

が成りたつ，これより

$$\lim_{n\to\infty} \frac{1}{n} \log Q^n(T^n(P)) = -D(P\|Q) \quad ——⑩$$

（証明）

$$Q^n(T^n(P)) = \sum_{x^n \in T^n(P)} Q^n(x^n)$$

43ページ ① →
$$= \sum_{x^n \in T^n(P)} e^{-n\{H(P)+D(P\|Q)\}}$$

$$= |T^n(P)|\, e^{-nH(P)} \cdot e^{-nD(P\|Q)}$$

これと45ページ ⑦ を用いると ⑨ が示される

⑨ より

$$-D(P\|Q) - \frac{1}{n}\log(n+1)^{|\mathcal{X}|-1} \leq \log Q^n(T^n(P)) \leq -D(P\|Q)$$

だから $n\to\infty$ とすると ⑩ が導かれる □

3-2, タイプと大数の法則

$$\begin{array}{ccccl} x^n & \longmapsto & P_{x^n} & (タイプ) & \\ \cap & & \cap & & \\ \mathcal{X}^n & \longrightarrow & \mathcal{P}_n & \subset & \mathcal{P}(\mathcal{X}) \end{array}$$

は関数であることに注意

よって

$$X^n = X_1 X_2 \cdots X_n \underset{i.i.d.}{\sim} Q$$

のとき　<u>確率変数 X^n のタイプ P_{X^n} は確率変数</u>

（確率変数の関数は確率変数）

<u>Thm</u>　$X^n \underset{i.i.d.}{\sim} Q$ のとき

$\forall \varepsilon > 0$ に対して

$$\lim_{n\to\infty} \Pr\{ D(P_{X^n} \| Q) > \varepsilon \} = 0$$

↑ X^n のタイプ

（証明）

$$\Pr\{ D(P_{X^n} \| Q) > \varepsilon \}$$

$$= \sum_{\substack{P \in \mathcal{P}_n \\ D(P\|Q) > \varepsilon}} Q^n(T^n(P))$$

$$\leq \sum_{\substack{P \in \mathcal{P}_n \\ D(P\|Q) > \varepsilon}} e^{-nD(P\|Q)} \quad (\because ⑨式)$$

$$\leq \sum_{P \in \mathcal{P}_n} e^{-n\varepsilon} = |\mathcal{P}_n| e^{-n\varepsilon}$$

$$\leq (n+1)^{|\mathcal{X}|-1} e^{-n\varepsilon} \xrightarrow[(n\to\infty)]{} 0$$

（42ページ Lem）

□

タイプは真の確率分布に近づく

3-3. $\mathbb{R}^m$ 上の位相

□ 距離 (distance)

$x^m = (x_1, x_2, \cdots, x_m) \in \mathbb{R}^m$

$y^m = (y_1, y_2, \cdots, y_m) \in \mathbb{R}^m$ の距離

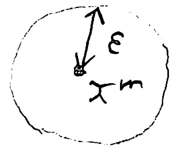
$$d(x^m, y^m) = \left\{ \sum_{i=1}^{m} (x_i - y_i)^2 \right\}^{\frac{1}{2}}$$

□ ε-近傍 (ε-neighborhood)

$x^m \in \mathbb{R}^m$ に対して

$$U(x^m, \varepsilon) = \{ y^m \in \mathbb{R}^m \mid d(x^m, y^m) < \varepsilon \}$$

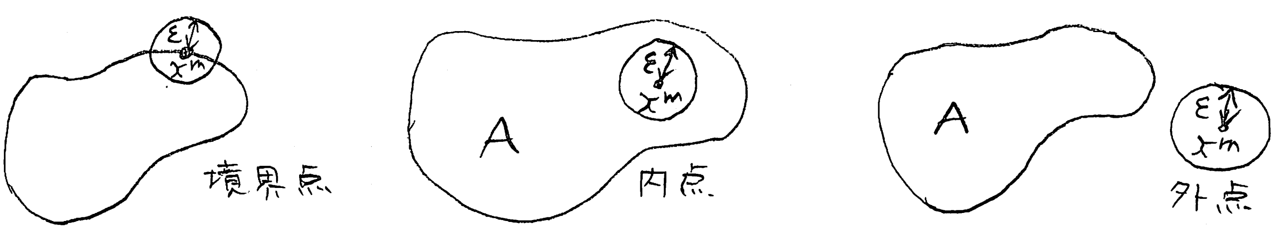
を x^m の ε-近傍という



<u>Def</u>　$A \subset \mathbb{R}^m$, $x^m \in \mathbb{R}^m$ について

□ x^m は A の <u>内点 (interior point)</u>

$\overset{\text{def}}{\iff}$ $\exists \varepsilon > 0$, $U(x^m, \varepsilon) \subset A$



□ x^m は A の <u>外点 (exterior point)</u>

$\overset{\text{def}}{\iff}$ $\exists \varepsilon > 0$, $A \cap U(x^m, \varepsilon) = \phi$

□ x^m は A の <u>境界点 (boundary point)</u>

$\overset{\text{def}}{\iff}$ $\forall \varepsilon > 0$, $A \cap U(x^m, \varepsilon) \neq \phi$

$A^c \cap U(x^m, \varepsilon) \neq \phi$

□ A の <u>内部（内点の集合, interior）</u>

$$A^{\circ} := \{ x^m \in \mathbb{R}^m \mid x^m \text{は} A \text{の内点} \}$$

□ A の <u>閉包（closure）</u>

$$\overline{A} := \{ x^m \in \mathbb{R}^m \mid x^m \text{は} A \text{の内点または境界点} \}$$

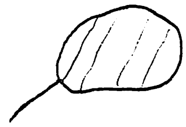
<u>例</u> $U(x^m, \varepsilon)^{\circ} = U(x^m, \varepsilon)$

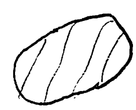
$$\overline{U(x^m, \varepsilon)} = \{ y^m \in \mathbb{R}^m \mid d(x^m, y^m) \leq \varepsilon \}$$

<u>例</u> $\mathbb{R}^2$ において

$A =$  のとき

（点線は含まない）

$\overline{A} =$ 

$A^{\circ} =$  $\overline{A^{\circ}} =$ 

<u>Remark</u> 体積（2次元のときは面積）を持たない「ヒゲ」は<u>境界点</u>

$\Downarrow$

内点をとる操作 A° で取り除かれる

<u>Remark</u>

A が開集合 $\overset{\text{def}}{\Longleftrightarrow}$ $A^{\circ} = A$
（open set）

A が閉集合 $\overset{\text{def}}{\Longleftrightarrow}$ $\overline{A} = A$
（closed set）

Remark 性質 $\overline{A^\circ} = A$ —— ⊛

- □ Aは「ヒゲ」を持たない閉集合
- □ Aのすべての点は内点の閉包

3-4 Sanov の定理

Thm $X^n = X_1 X_2 \cdots X_n \underset{\text{i.i.d.}}{\sim} Q$

$E \subset \mathcal{P}(\mathcal{X})$ のとき

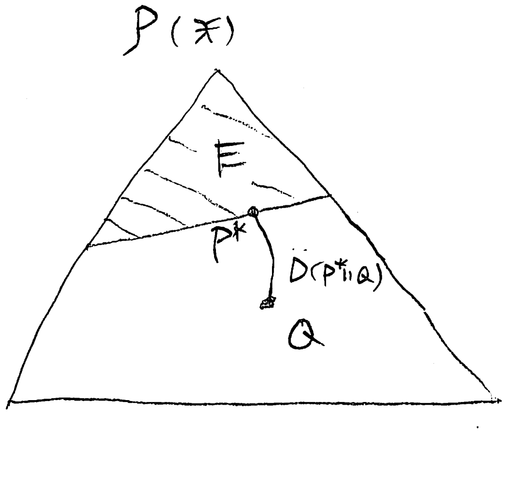
$$\Pr\{ P_{X^n} \in E \} \le (n+1)^{|\mathcal{X}|-1} e^{-nD(P^* \| Q)} \quad \text{—— ①}$$

P_{X^n} ← X^n のタイプ

ただし $P^* = \underset{P \in E}{\operatorname{argmin}}\ D(P\|Q)$

さらに $\overline{E^\circ} = E$ をみたすとき

$$\lim_{n\to\infty} \frac{1}{n} \log \Pr\{ P_{X^n} \in E \} = -D(P^* \| Q) \quad \text{—— ②}$$



（証明）

$$\Pr\{P_{x^n} \in E\} = \sum_{\substack{P \in \mathcal{P}_n \\ P \in E}} Q^n(T^n(P))$$

$$\le \sum_{P \in E \cap \mathcal{P}_n} e^{-nD(P\|Q)}$$

$$\le \sum_{P \in E \cap \mathcal{P}_n} e^{-nD(P^*\|Q)}$$

$$\left[\because \ D(P^*\|Q) \le D(P\|Q) \right]$$

$$\le (n+1)^{|\mathcal{X}|-1} e^{-nD(P^*\|Q)}$$

① が示された

$\overline{E^\circ} = E$ をみたすとき（性質 ＊）

□ P^* は E の内点 E° の境界だから

$\forall \varepsilon > 0$ に対して $U(P^*, \varepsilon) \cap E^\circ \neq \phi$ —— ③

□ $\overline{\bigcup_{n=1}^{\infty} \mathcal{P}_n} = \mathcal{P}(\mathcal{X})$

[$\bigcup_n \mathcal{P}_n$ は $\mathcal{P}(\mathcal{X})$ において稠密 (dense)]

なので、

$\forall \varepsilon > 0$ に対して ③より $\exists n_0$, $\forall n \ge n_0$

$$U(P^*, \varepsilon) \cap E^\circ \cap \mathcal{P}_n \neq \phi$$

これより

$$U(P^*, \varepsilon) \cap E \cap \mathcal{P}_n \neq \phi \quad \text{—— ④}$$

すなわち $\forall \varepsilon > 0$ に対して

$\exists n_0$, $\forall n \ge n_0$

$$d(P^*, P_n) < \varepsilon$$

となる タイプの列 $P_n \in E \cap \mathcal{P}_n$ が存在

なんとなく分かればOK

[4 と 5 により]

タイプの列 $P_n \in E \cap \mathcal{P}_n$ が存在して

$$\lim_{n\to\infty} D(P_n \| Q) = D(P^* \| Q) \quad — ⑤$$

このとき

$$\Pr\{P_{x^n} \in E\} = \sum_{P \in E \cap \mathcal{P}_n} Q^n(T^n(P))$$

$$\geq Q^n(T^n(P_n))$$

$$\geq \frac{1}{(n+1)^{|\mathcal{X}|-1}} e^{-nD(P_n \| Q)} \quad — ⑥$$

だから、① と ⑥ より

$$-D(P_n \| Q) - \frac{1}{n}\log(n+1)^{|\mathcal{X}|-1}$$

$$\leq \frac{1}{n}\log \Pr\{P_{x^n} \in E\} \leq -D(P^* \| Q) + \frac{1}{n}\log(n+1)^{|\mathcal{X}|-1}$$

$n \to \infty$ とすると ⑤ より ② が示される □

3-5 仮説検定と Stein の補題
(Hypothesis testing)

□ <u>仮説</u> $X^n = X_1 X_2 \cdots X_n \underset{i.i.d.}{\sim} P$

または

$X^n = X_1 X_2 \cdots X_n \underset{i.i.d.}{\sim} Q$

□ X^n の実現値 $x^n \in \mathcal{X}^n$ をもとに
どちらが真の分布であるかを判定

$$T_n \subset \mathcal{X}^n \quad (P \text{ の受容域, acceptance region})$$

を用意して,

$$\begin{cases} x^n \in T_n \longrightarrow P \text{ が真と判定} \\ x^n \in T_n^c \longrightarrow Q \text{ が真と判定} \end{cases}$$

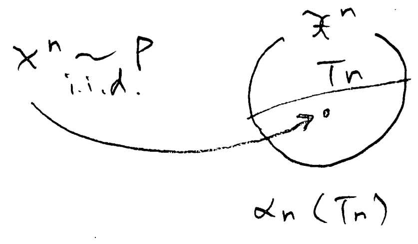
↑ 補集合

このとき <u>T_n を検定 (test)</u> とよぶ

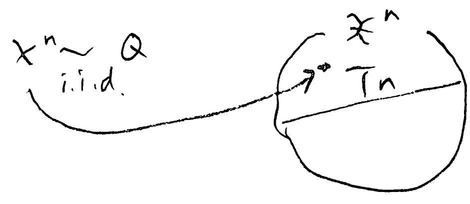
□ 誤り確率

$$\alpha_n(T_n) := P^n(T_n^c) \quad (\text{第一種誤り})$$

$$\beta_n(T_n) := Q^n(T_n) \quad (\text{第二種誤り})$$



$\alpha_n(T_n)$



$\beta_n(T_n)$

<u>Thm (Stein の補題)</u>

$$\beta_n^*(\varepsilon) = \min_{\substack{T_n \subset \mathcal{X}^n \\ \alpha_n(T_n) \le \varepsilon}} \beta_n(T_n) \quad \text{—— ①} \quad \text{とおくと}$$

$0 < \forall \varepsilon < 1$ に対して

$$\lim_{n \to \infty} \frac{1}{n} \log \beta_n^*(\varepsilon) = -D(P \| Q) \quad \text{—— ②}$$

以下で順次 Stein の補題を証明

Def (Neyman - Pearson test)

$$S_{n,a} := \left\{ x^n \in \mathcal{X}^n \;\middle|\; \frac{1}{n} \log \frac{P^n(x^n)}{Q^n(x^n)} > a \right\} \quad \text{— ③}$$

Remark

$$\frac{1}{n} \log \frac{P^n(x^n)}{Q^n(x^n)} > a \iff P^n(x^n) - e^{na} Q^n(x^n) > 0$$

だから

$$S_{n,a} = \left\{ x^n \in \mathcal{X}^n \;\middle|\; P^n(x^n) - e^{na} Q^n(x^n) > 0 \right\} \quad \text{— ④}$$

Lem

$\forall a \in \mathbb{R}$ と $\forall T_n \subset \mathcal{X}^n$ に対して

$$P^n(S_{n,a}) - e^{na} Q^n(S_{n,a}) \geq P^n(T_n) - e^{na} Q^n(T_n) \quad \text{— ⑤}$$

$\because$

$$\sum_{x^n \in S_{n,a}} \left\{ P^n(x^n) - e^{na} Q^n(x^n) \right\} \geq \sum_{x^n \in T_n} \left\{ P^n(x^n) - e^{na} Q^n(x^n) \right\} \quad \text{— ⑥}$$

を示せばよい.

$$\begin{aligned} T_n &= T_n \cap (S_{n,a} \cup S_{n,a}^c) \\ &= (T_n \cap S_{n,a}) \cup (T_n \cap S_{n,a}^c) \end{aligned}$$

より

$$\begin{aligned} (\text{⑥の右辺}) &= \sum_{x^n \in T_n \cap S_{n,a}} \left\{ P^n(x^n) - e^{na} Q^n(x^n) \right\} + \underbrace{\sum_{x^n \in T_n \cap S_{n,a}^c} \left\{ P^n(x^n) - e^{na} Q^n(x^n) \right\}}_{\leq 0 \ (\text{④より})} \\ &\leq \sum_{x^n \in T_n \cap S_{n,a}} \left\{ P^n(x^n) - e^{na} Q^n(x^n) \right\} \\ &\overset{④}{\leq} \sum_{x^n \in S_{n,a}} \left\{ P^n(x^n) - e^{na} Q^n(x^n) \right\} \\ &= (\text{⑥の左辺}) \end{aligned}$$

$\square$

↑ $P^n(x^n) - e^{na} Q^n(x^n)$ が正の x^n を過不足なく足したもの

Lem（Neyman-Pearson の補題）

$\forall a \in \mathbb{R}$ と $\forall T_n \subset \mathcal{X}^n$ について

$$\alpha_n(S_{n,a}) \geq \alpha_n(T_n) \quad — ⑦ \quad \Rightarrow \quad \beta_n(S_{n,a}) \leq \beta_n(T_n) \quad — ⑧$$

$\because$ ⑤より

$$1 - P^n(T_n) - e^{na}Q^n(S_{n,a}) \geq 1 - P^n(S_{n,a}) - e^{na}Q^n(T_n)$$

$$\alpha_n(T_n) - e^{na}\beta_n(S_{n,a}) \geq \alpha_n(S_{n,a}) - e^{na}\beta_n(T_n)$$

$$\alpha_n(T_n) - \alpha_n(S_{n,a}) \geq e^{na}\{\beta_n(S_{n,a}) - \beta_n(T_n)\} \quad — ⑨$$

ここで ⑦ が成立しているとすると

⑨ の左辺は負になるので、

$$0 \geq e^{na}\{\beta_n(S_{n,a}) - \beta_n(T_n)\}$$

だから ⑧ が成立する □

Lem　$\forall a \in \mathbb{R}$ について

$$P^n(S_{n,a}) - e^{na}Q^n(S_{n,a}) \geq 0 \quad — ⑩$$

特に　$Q^n(S_{n,a}) \leq e^{-na}$ — ⑪

$\because$

$$P^n(S_{n,a}) - e^{na}Q^n(S_{n,a})$$

$$= \sum_{x^n \in S_{n,a}} \{\underbrace{P^n(x^n) - e^{na}Q^n(x^n)}_{\geq 0 \ (④より)}\} \geq 0$$

⑪ は ⑩ より次のように示される

$$Q^n(S_{n,a}) \leq e^{-na} \underset{\substack{\wedge \\ 1}}{P^n(S_{n,a})} \leq e^{-na}$$

□

Thm $X^n = X_1 X_2 \cdots X_n \underset{\text{i.i.d.}}{\sim} P$ のとき

$$\lim_{n\to\infty} \underbrace{\Pr\left\{ \frac{1}{n} \log \frac{P^n(X^n)}{Q^n(X^n)} > a \right\}}_{= P^n(S_{n,a})} = \begin{cases} 1 & \text{if } a < D(P\|Q) \\ 0 & \text{if } a > D(P\|Q) \end{cases} \quad \text{— (12)}$$

(証明)

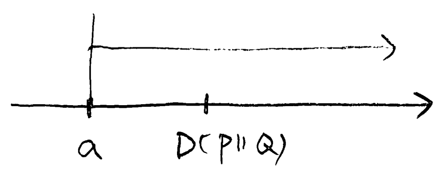
$$\frac{1}{n} \log \frac{P^n(X^n)}{Q^n(X^n)} = \frac{1}{n} \log \frac{P(X_1)P(X_2)\cdots P(X_n)}{Q(X_1)Q(X_2)\cdots Q(X_n)}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \log \frac{P(X_i)}{Q(X_i)} \quad \text{— (13)}$$

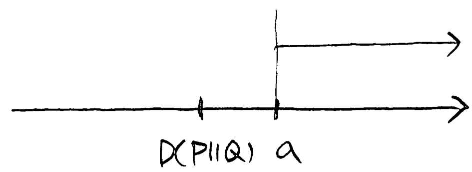
が独立な確率変数のサンプル平均であることと

$$E\left[\log \frac{P(X_i)}{Q(X_i)} \right] = D(P\|Q) \quad \text{— (14)}$$

に注意する



$a < D(P\|Q)$ のとき



$a > D(P\|Q)$ のとき

(13)(14)と大数の法則(32ページ)より

$\forall \varepsilon > 0$ に対して

$$\lim_{n\to\infty} \Pr\left\{ \left| \frac{1}{n} \log \frac{P^n(X^n)}{Q^n(X^n)} - D(P\|Q) \right| \leq \varepsilon \right\} = 1 \quad \text{— (15)}$$

<u>$a < D(P\|Q)$ のとき</u> $\varepsilon = D(P\|Q) - a > 0$ とおくと

$$\Pr\left\{ \frac{1}{n} \log \frac{P^n(X^n)}{Q^n(X^n)} > \underset{D(P\|Q)-\varepsilon}{\underset{\shortparallel}{a}} \right\} = \Pr\left\{ \frac{1}{n} \log \frac{P^n(X^n)}{Q^n(X^n)} - D(P\|Q) > -\varepsilon \right\}$$

$$\geq \Pr\left\{ -\varepsilon \leq \frac{1}{n} \log \frac{P^n(X^n)}{Q^n(X^n)} - D(P\|Q) \leq \varepsilon \right\}$$

$$= \Pr\left\{ \left| \frac{1}{n} \log \frac{P^n(X^n)}{Q^n(X^n)} - D(P\|Q) \right| \leq \varepsilon \right\}$$

$$\xrightarrow[\text{(15)}]{} 1 \quad (n \to \infty)$$

<u>$a > D(P\|Q)$ のとき</u>　　$\varepsilon = a - D(P\|Q) > 0$ とおくと

$$\Pr\left\{ \frac{1}{n}\log\frac{P^n(X^n)}{Q^n(X^n)} > \underset{D(P\|Q)+\varepsilon}{\overset{\|}{a}} \right\} = \Pr\left\{ \frac{1}{n}\log\frac{P^n(X^n)}{Q^n(X^n)} - D(P\|Q) > \varepsilon \right\}$$

$$\le \Pr\left\{ \frac{1}{n}\log\frac{P^n(X^n)}{Q^n(X^n)} - D(P\|Q) > \varepsilon \text{ または } \frac{1}{n}\log\frac{P^n(X^n)}{Q^n(X^n)} - D(P\|Q) < -\varepsilon \right\}$$

$$= \Pr\left\{ \left| \frac{1}{n}\log\frac{P^n(X^n)}{Q^n(X^n)} - D(P\|Q) \right| > \varepsilon \right\}$$

$$\xrightarrow[(15)]{} 0 \quad (n \to \infty)$$

□

本当は limsup

＜Stein の補題の証明＞

(i)　$\displaystyle \lim_{n\to\infty} \frac{1}{n}\log\beta_n^*(\varepsilon) \le -D(P\|Q)$

　　(direct part，性能の良い test の存在)

(ii)　$\displaystyle \lim_{n\to\infty} \frac{1}{n}\log\beta_n^*(\varepsilon) \ge -D(P\|Q)$

　　(converse part，性能の限界)

を示す．

本当は liminf

<u>(i) direct part</u>

$\varepsilon > 0$ を任意に固定して，$a = D(P\|Q) - \varepsilon$ とおく

このとき (12) より

$$\lim_{n\to\infty} \alpha_n(S_{n,a}) = \lim_{n\to\infty}\{1 - P^n(S_{n,a})\} = 0$$

また (11) より

$$\beta_n(S_{n,a}) \le e^{-na}$$

だから

$$\lim_{n\to\infty} \frac{1}{n}\log\beta_n(S_{n,a}) \le -a = -D(P\|Q) + \varepsilon$$

よって

$$\lim_{n\to\infty} \frac{1}{n} \log \beta_n^*(\varepsilon) \le \lim_{n\to\infty} \frac{1}{n} \log \beta_n(S_{n,a})$$

$$\le -D(P\|Q) + \varepsilon$$

$\varepsilon > 0$ は任意だから $\varepsilon \searrow 0$ とすれば (i) が示される

(ii) converse part

$1 > \varepsilon > 0$ を任意に固定して $a = D(P\|Q) + \varepsilon$ とおく

このとき ⑫ より

$$\lim_{n\to\infty} \alpha_n(S_{n,a}) = \lim_{n\to\infty} \{1 - P^n(S_{n,a})\} = 1 \quad ⑯$$

また、⑨ より

$$\beta_n(T_n) \ge \beta_n(T_n) - \beta_n(S_{n,a}) \overset{⑨}{\ge} e^{-na}\{\alpha_n(S_{n,a}) - \alpha_n(T_n)\}$$

よって $\alpha_n(T_n) \le \varepsilon$ であるすべての $T_n \subset \mathcal{X}^n$ について

$$\beta_n(T_n) \ge e^{-na}\{\alpha_n(S_{n,a}) - \alpha_n(T_n)\}$$

$$\ge e^{-na}\{\alpha_n(S_{n,a}) - \varepsilon\}$$

よって

$$\beta_n^*(\varepsilon) \ge e^{-na}\{\alpha_n(S_{n,a}) - \varepsilon\}$$

⑯ より 十分大きなすべての n に対して

$\alpha_n(S_{n,a}) - \varepsilon > 0$ であるから

$$\frac{1}{n} \log \beta_n^*(\varepsilon) \ge -a + \underbrace{\frac{1}{n} \log\{\alpha_n(S_{n,a}) - \varepsilon\}}_{⑯ \searrow 0}$$

これより

$$\lim_{n\to\infty} \frac{1}{n} \log \beta_n^*(\varepsilon) \ge -a = -D(P\|Q) - \varepsilon$$

$\varepsilon > 0$ は任意なので $\varepsilon \searrow 0$ として (ii) が示される

□

本当は limsup

Thm（強逆性, strong converse）

$\forall T_n \subset \mathcal{X}^n$ について

$$\lim_{n\to\infty} \frac{1}{n} \log \beta_n(T_n) < -D(P\|Q) \Rightarrow \lim_{n\to\infty} \alpha_n(T_n) = 1$$

— (17)

（証明）

(17) が成立しているとすると $\varepsilon > 0$ が存在して

$$\lim_{n\to\infty} \frac{1}{n} \log \beta_n(T_n) < -D(P\|Q) - \varepsilon$$

このとき十分大きなすべての n に対して

$$\frac{1}{n} \log \beta_n(T_n) < -D(P\|Q) - \varepsilon$$

$$\Longleftrightarrow \quad \beta_n(T_n) < e^{-n\{D(P\|Q)+\varepsilon\}} \quad \text{— (18)}$$

ここで (9) より

$$\begin{aligned}
\alpha_n(T_n) &\geq \alpha_n(S_{n,a}) + e^{na}\{\beta_n(S_{n,a}) - \beta_n(T_n)\} \\
&\geq \alpha_n(S_{n,a}) - e^{na}\beta_n(T_n) \\
&\geq \alpha_n(S_{n,a}) - e^{na} \cdot e^{-n\{D(P\|Q)+\varepsilon\}} \quad \text{— (19)}
\end{aligned}$$

ここで $a = D(P\|Q) + \frac{\varepsilon}{2}$ とおくと、(12) より

$\lim_{n\to\infty} \alpha_n(S_{n,a}) = 1$ だから

(19) より

$$\alpha_n(T_n) \geq \underbrace{\alpha_n(S_{n,a})}_{\to 1} - \underbrace{e^{n\{D(P\|Q)+\frac{\varepsilon}{2}\}} \cdot e^{-n\{D(P\|Q)+\varepsilon\}}}_{= e^{-n\cdot\frac{\varepsilon}{2}} \to 0}$$

よって

$$\lim_{n\to\infty} \alpha_n(T_n) = 1$$

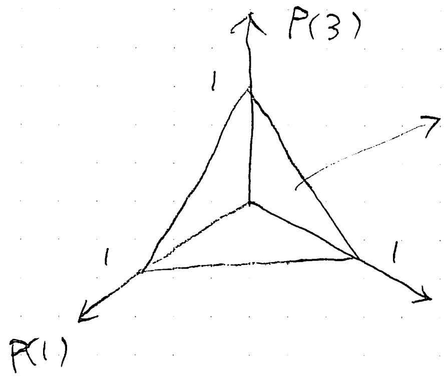
□

3－6　Sanov の定理の応用

○ 確率分布全体 $\mathcal{P}(\mathcal{X})$ の例

$\mathcal{X} = \{1, 2, 3\}$

$\mathcal{P}(\mathcal{X}) = \{ (P(1), P(2), P(3)) \in \mathbb{R}^3 \mid$
$P(1) + P(2) + P(3) = 1,$
$P(1) \geq 0,\ P(2) \geq 0,\ P(3) \geq 0 \}$



$P(1) + P(2) + P(3) = 1$
[法線ベクトル $(1, 1, 1)$ の平面]

○ $\mathcal{X} = \{1, 2, 3\}$

$Q = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) \in \mathcal{P}(\mathcal{X})$ とする

平均値　$\mu = \sum_{x \in \mathcal{X}} Q(x)\, x$

$= \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 2 + \frac{1}{3} \cdot 3 = 2$

○ $X^n = X_1 X_2 \cdots X_n \underset{\text{i.i.d.}}{\sim} Q$ のとき

大数の法則（32ページ）より

$$\lim_{n \to \infty} \Pr\left\{ \left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right| \leq \varepsilon \right\} = 1 \quad (\forall \varepsilon > 0)$$

[n が十分大きいとき，$\frac{1}{n} \sum_{i=1}^{n} X_i \approx \mu$]

$(\mu = 2)$

○ $c > \mu$ のとき $\frac{1}{n}\sum_{i=1}^{n} X_i > c$ の確率は？

（例 $c = 2.5$）　　（平均からずれる確率）

□ $x^n = x_1 x_2 \cdots x_n \in \mathcal{X}^n$ に対して

$$\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{a\in\mathcal{X}} N(a \mid x^n)\, a$$

$N(a \mid x^n)$ ← a の生起回数（41ページ）

$$= \sum_{x\in\mathcal{X}} P_{x^n}(a)\, a$$

$P_{x^n}(a)$ ← x^n のタイプ

例： $n = 6$, $x^n = 112132$

$$\frac{1}{n}\sum x_i = \frac{1}{6}(1+1+2+1+3+2)$$

$$= \frac{1}{6}(3\times 1 + 2\times 1 + 1\times 3)$$

1の数　2の数　3の数

$$= \frac{3}{6}\times 1 + \frac{2}{6}\times 1 + \frac{1}{6}\times 3$$

x^n のタイプ

サンプル平均 ＝ タイプに関する平均

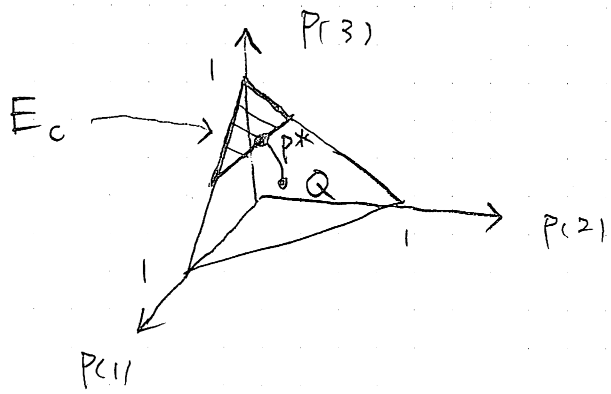
□ $\Pr\{x^n \in \mathcal{X}^n \mid \frac{1}{n}\sum_{i=1}^{n} x_i \geq c\}$

$$= \Pr\{x^n \in \mathcal{X}^n \mid \sum_{a\in\mathcal{X}} P_{x^n}(a)\, a \geq c\}$$

$$= \Pr\{x^n \in \mathcal{X}^n \mid P_{x^n} \in E_c\}$$

ただし $E_c = \{P \in \mathcal{P}(\mathcal{X}) \mid \sum_{a\in\mathcal{X}} P(a)\, a \geq c\}$

$$= \{(P(1), P(2), P(3)) \in \mathcal{P}(\mathcal{X}) \mid P(1)\cdot 1 + P(2)\cdot 2 + P(3)\cdot 3 \geq 2.5\}$$



よって Sanov の定理より （ $\overline{E^\circ} = E$ をみたすので）

$$\lim_{n\to\infty} \frac{1}{n} \log \Pr\left\{ \frac{1}{n}\sum_{i=1}^{n} x_i \geq c \right\}$$

$$= \lim_{n\to\infty} \frac{1}{n} \log \Pr\left\{ P_{x^n} \in E_c \right\}$$

$x^n = x_1 x_2 \cdots x_n$ のタイプ

$$= -D(P^* \| Q)$$

ただし $P^* = \underset{P \in E_c}{\arg\min}\, D(P \| Q)$

$$\Pr\left\{ \frac{1}{n}\sum_{i=1}^{n} x_i \geq c \right\} \approx e^{-n \cdot D(P^* \| Q)}$$

サンプル平均が期待値からずれる確率　　指数的に減少
deviation

○ P^* の求め方

□ 方法1： Lagrange 乗数法 (multiplier method)

□ 方法2： ダイバージェンスのピタゴラス定理

○ Lagrang 乗数法

関数 $f: \mathbb{R}^m \longrightarrow \mathbb{R}$ が 凸 関数 (convex) のとき
(下に凸)

拘束条件 $g_k(x^m) = 0 \quad (k = 1, 2, \cdots, d)$

(ただし、g_k は 凸 関数 $g_k : \mathbb{R}^m \longrightarrow \mathbb{R}$)

のもとで f を最小化

問題

$$
\begin{array}{lll}
\text{minimize} & f(x^m) & x^m \in \mathbb{R}^m \\
\text{s.t.} & g_k(x^m) = 0 & (k = 1, 2, \cdots, d)
\end{array}
\quad \text{--- ①}
$$

解法

(step 1) Lagrange 乗数 $\lambda_k \ (k = 1, 2, \cdots, d)$

を導入して

$$F(x^m) = f(x^m) + \sum_{k=1}^{d} \lambda_k g_k(x^m) \quad \text{--- ②}$$

を拘束条件なしで最小化する

そのためには ② を $x_i \ (i = 1, 2, \cdots, m)$

で偏微分してゼロとおけばよい

$$\frac{\partial F(x^m)}{\partial x_i} = \frac{\partial f(x^m)}{\partial x_i} + \sum_{k=1}^{d} \lambda_k \frac{\partial g_k(x^m)}{\partial x_i} = 0$$

$$(i = 1, 2, \cdots, m) \quad \text{--- ③}$$

(step 2) ③ より $x_i \ (i = 1, 2, \cdots, m)$ が

$\lambda_k \ (k = 1, 2, \cdots, d)$ の関数として定まる

ここから、$\lambda_k \ (k = 1, 2, \cdots, d)$ を ① をみたすように

定める

○ 63ページの場合にLagrange乗数法を適用

$m = |\mathcal{X}| = 3$

minimize $\quad D(P \| Q) \qquad P = (P(1), \cdots, P(m))$

s.t. $\quad \sum_{a \in \mathcal{X}} P(a)\, a = c \quad (P \in E_c)$ — ①

$\qquad \sum_{a \in \mathcal{X}} P(a) = 1 \quad (P \in \mathcal{P}(\mathcal{X}))$ — ②

$\left\{ \begin{array}{l} \text{①についての Lagrange 乗数を } \lambda \\ \text{② 〃 } \mu \end{array} \right.$ とおき

$$F(P) = D(P \| Q) + \lambda \left(\sum_a P(a)\, a - c \right) + \mu \left(\sum_a P(a) - 1 \right)$$ — ③ とおく

(step 1) $\dfrac{\partial F}{\partial P(b)} \quad (b \in \mathcal{X})$ を計算

$$\frac{\partial D(P \| Q)}{\partial P(b)} = \frac{\partial}{\partial P(b)} \sum_a P(a) \{ \log P(a) - \log Q(a) \}$$

$$= \{ \log P(b) - \log Q(b) \} + P(b) \cdot \frac{1}{P(b)}$$

$\left[\dfrac{\partial P(a)}{\partial P(b)} = \begin{cases} 1 & a = b \\ 0 & a \neq b \end{cases} \right.$ に注意 $\left. \right]$

より

$$\frac{\partial F}{\partial P(b)} = \log P(b) - \log Q(b) + 1 + \lambda \cdot b + \mu = 0$$

これを解くと

$$P(b) = e^{\log Q(b) - \lambda b - (\mu + 1)}$$

$$= \frac{Q(b)\, e^{-\lambda b}}{e^{\mu + 1}}$$ — ④

(step 2)　λは ① をみたすように
μは ② をみたすように　定める

□ μについて ② をみたすように定める
↑ 規格化条件

これは　$e^{\mu+1} = \sum_{b\in\mathcal{X}} Q(b)\, e^{-\lambda b}$　[④の分子の和]

とすればよい，すなわち

$$P(a) = \frac{Q(a)\, e^{-\lambda a}}{\sum_{b\in\mathcal{X}} Q(b)\, e^{-\lambda b}} \qquad (a\in\mathcal{X})$$

□ λについて ① をみたすように定める

解析的には求められないが
一意にλが求まることが知られている

これらより

$$\begin{cases} P^*(a) = \dfrac{Q(a)\, e^{-\lambda a}}{\sum_{b\in\mathcal{X}} Q(b)\, e^{-\lambda b}} \quad (a\in\mathcal{X}) & \text{—— ⑤} \\ \text{ただし}\quad \sum_{a\in\mathcal{X}} P^*(a)\, a = c & \text{—— ⑥} \end{cases}$$

$$D(P^*\|Q) = \sum_a P^*(a)\{\underbrace{\log P^*(a)}_{\log Q(a) - \lambda a - \log \sum_b Q(b) e^{-\lambda b}} - \log Q(a)\}$$

$$= \sum_a P^*(a)\{-\lambda a - \log \sum_b Q(b)\, e^{-\lambda b}\}$$

$$= -\lambda \underbrace{\sum_a P^*(a)\, a}_{c\ (⑥より)} - \log \sum_b Q(b)\, e^{-\lambda b}$$

$$= -\lambda c - \log \sum_b Q(b)\, e^{-\lambda b}$$

(λは ⑥ をみたす実数)

○ レポート

□ 仮説検定（54ページ）を考える

$$X^n = X_1 X_2 \cdots X_n \overset{iid}{\sim} P \text{ or } Q$$

観測値 $x^n = x_1 x_2 \cdots x_n \in \mathcal{X}^n$

について Neyman-Pearson 検定（55ページ）

$$S_{n,a} = \left\{ x^n \in \mathcal{X}^n \,\middle|\, \frac{1}{n} \log \frac{P^n(x^n)}{Q^n(x^n)} > a \right\}$$

を考える

□ $$x^n \in S_{n,a} \iff \frac{1}{n} \log \frac{P^n(x^n)}{Q^n(x^n)} > a$$

$$\iff \underbrace{\frac{1}{n} \sum_{i=1}^{n} \log \frac{P(x_i)}{Q(x_i)}}_{\shortparallel} > a$$

［62ページと同じ計算］

$$\frac{1}{n} \sum_{b \in \mathcal{X}} N(b|x^n) \log \frac{P(b)}{Q(b)}$$

$$= \sum_{b \in \mathcal{X}} P_{x^n}(b) \log \frac{P(b)}{Q(b)}$$

$$\iff P_{x^n} \in E_a$$

ただし $$E_a = \left\{ R \in \mathcal{P}(\mathcal{X}) \,\middle|\, \sum_{b \in \mathcal{X}} R(b) \log \frac{P(b)}{Q(b)} > a \right\}$$

□ Sanov の定理より

$$\alpha_n(S_{n,a}) = P^n(\overline{E_a}) \approx e^{-n \cdot \min_{R \in \overline{E_a}} D(R \| P)}$$

$$\beta_n(S_{n,a}) = Q^n(E_a) \approx e^{-n \cdot \min_{R \in E_a} D(R \| Q)}$$

□ 問題：$-D(Q \| P) < a < D(P \| Q)$ とするとき

(1) $\min_{R \in \overline{E_a}} D(R \| P)$ を達成する R を求めよ

(2) $\min_{R \in E_a} D(R \| Q)$ 〃

4. 通信路符号化

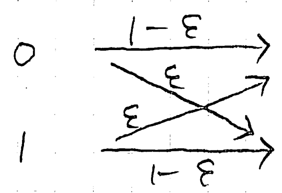
4-1. 通信路符号化定理 (Channel coding Thm)

○ 通信路 (channel) = 条件付確率

$W(y|x)$ $\left(x \in \mathcal{X} \text{ (入力アルファベット)}, \ y \in \mathcal{Y} \text{ (出力アルファベット)} \right)$

$$x \longrightarrow \boxed{W} \longrightarrow y$$

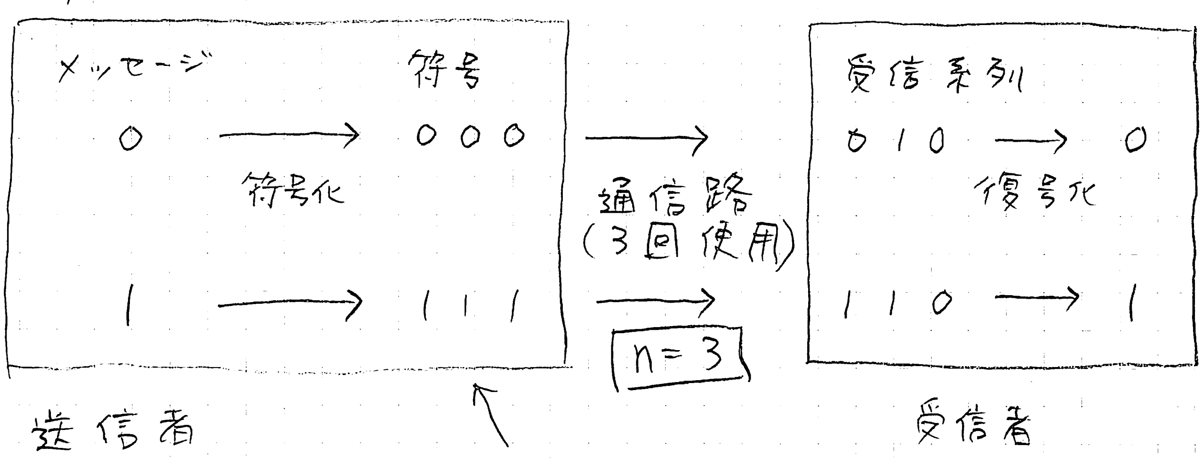
○ 例: binary symmetric channel

$\mathcal{X} = \{0, 1\}$  $\mathcal{Y} = \{0, 1\}$

$$W(0|0) = W(1|1) = 1 - \varepsilon$$
$$W(1|0) = W(0|1) = \varepsilon$$

○ 通信路を通したメッセージ伝送

例:



符号化方法は
あらかじめ
送信者と受信者で合意しておく

□ 誤りを0に近づけるには $n \to \infty$ とする
必要がある

○ 通信路符号化 (channel coding)

メッセージ (message) $k \in \{1, 2, \cdots, M_n\}$

↓ 符号化 (encoding)

符号語 (codeword) $x^n(k) = x_1(k)\, x_2(k) \cdots x_n(k) \in \mathcal{X}^n$

通信路 W n回使用

受信系列 $y^n(k) = y_1(k)\, y_2(k) \cdots y_n(k)$

$$W^n(y^n|x^n) = \prod_{i=1}^{n} W(y_i|x_i)$$

↓ 復号化 (decoding)

復号メッセージ ℓ

□ 符号化 $f_n : \{1, 2, \cdots, M_n\} \longrightarrow \mathcal{X}^n$

コードブック (codebook) $\{x^n(1), x^n(2), \cdots, x^n(M_n)\}$ を指定することに外ならない

□ 復号化 $g_n : \mathcal{Y}^n \longrightarrow \{1, 2, \cdots, M_n, 0\}$

(0:) 復号不可能を表すシンボル

□ 符号 = 符号化 + 復号化

□ 平均誤り確率 (avarage error prob.)

$$P_e(f_n, g_n) := \frac{1}{M_n} \sum_{k=1}^{M_n} \Pr\{g_n(f_n(k)) \neq k\}$$

□ 伝送レート (transmission rate)

$$\frac{1}{n} \log M_n \quad \text{(bit / 回)}$$

通信路1回使用あたりの伝送ビット数　　$\log_2$のとき

□ 通信速度との関係

1秒間にN回通信路を使用するとき

通信速度 = $N \cdot \frac{1}{n} \log M_n$ (bit/s)

○ 目的

$$\begin{cases} P_e(f_n, g_n) \longrightarrow 0 \quad (n \to \infty) \\ \frac{1}{n} \log M_n \longrightarrow 大きく \end{cases}$$

となる符号(encoder + decoder)を作る

符号化レートの限界は?

<u>Def</u>

レート R は achievable (達成可能)

$\overset{def}{\Longleftrightarrow}$ メッセージ数 M_n の符号列 f_n, g_n $(n = 1, 2, \cdots)$ が存在して

$$\begin{cases} \lim_{n\to\infty} P_e(f_n, g_n) = 0 \\ \lim_{n\to\infty} \frac{1}{n} \log M_n \geq R \end{cases}$$

本当は liminf

<u>Def</u> (通信路容量, channel capacity)

$C := \{R \mid R$ は achievable$\}$

(通信路 W に固有の量)

補足 同一の通信路 W を独立に n 回使用するとき、入力系列 $x^n = x_1 x_2 \cdots x_n \in \mathcal{X}^n$ に対して、出力系列 $y^n = y_1 y_2 \cdots y_n \in \mathcal{Y}^n$ を得る確率は

$$W(y^n | x^n) = W(y_1 | x_1) W(y_2 | x_2) \cdots W(y_n | x_n)$$

[定常無記憶通信路 とよぶ
stationary memoryless channel]

Thm（通信路符号化定理，Shannon の第2定理
channel coding theorem）

$$C = \max_{P_X} I(X;Y)$$

ただし　$I(X;Y)$ は　$P_{XY}(x,y) = P_X(x)W(y|x)$
の相互情報量　（25ページ）

4-2. 通信路符号化定理の証明

- Direct Part：性能の良い符号の存在を示す
  $$C \geq \max_{P_X} I(X;Y)$$
- Converse Part：性能の限界を示す
  $$C \leq \max_{P_X} I(X;Y)$$

○ ここでは Direct Part のみ示す

○ <u>証明の手順</u>

(1) 符号化の作成（f_n の作成）：$\boxed{\text{ランダムコーディング}}$（random coding）

$X^n(1), X^n(2), \cdots, X^n(M_n)$ を，
それぞれ独立に確率分布
$$P_{X^n}(x^n) = P_X(x_1)P_X(x_2)\cdots P_X(x_n)$$
に従って発生して，コードブックとする

すなわち
$$\left[\begin{array}{l} X^n(1) = X_1(1)X_2(1)\cdots X_n(1) \overset{iid}{\sim} P_X \\ X^n(2) = X_1(2)X_2(2)\cdots X_n(2) \overset{iid}{\sim} P_X \\ \vdots \\ X^n(M_n) = X_1(M_n)X_2(M_n)\cdots X_n(M_n) \overset{iid}{\sim} P_X \end{array}\right.$$

(2) 復号化の作成：後述，$X^n(1)\cdots X^n(M_n)$ に依存
(g_n の作成)

(3) $E[P_e(f_n, g_n)] \to 0 \quad (n \to \infty)$ を示す

↑ ランダムコーディングの平均

(4) (3)より少なくとも1つ

$$P_e(f_n, g_n) \longrightarrow 0 \quad (n \to \infty)$$

となる f_n, g_n の存在が示される

Lem (union bound)

$\boxed{\cup : \text{union}}$

P : $\mathcal{X}$ 上の確率分布

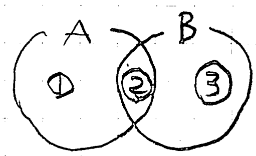
$A \subset \mathcal{X}$, $B \subset \mathcal{X}$ とすると

$$P(A \cup B) \le P(A) + P(B)$$

∵

$$A \setminus B := A \cap \overline{B} \quad ①$$

$$B \setminus A := B \cap \overline{A} \quad ③$$



とおくと

$$P(A \cup B) = \underset{①}{P(A \setminus B)} + \underset{③}{P(B \setminus A)} + \underset{②}{P(A \cap B)}$$

$$\le \underbrace{P(A \setminus B) + P(A \cap B)}_{} + \underbrace{P(B \setminus A) + P(A \cap B)}_{}$$

$$= P(A) + P(B)$$

□

(板書が良くなかったので記号を少し変えました)

通信路符号化定理の証明

○ 入力アルファベット上の確率分布 $P_X \in \mathcal{P}(\mathcal{X})$ を任意にとって，以後固定

(1) 手順(1)によりランダムにコードブック

$x^n(1), x^n(2), \cdots, x^n(M_n)$ を作成 —— ①

(2) 復号化：仮説検定の重ね合わせ

仮説検定 $\boxed{W(y^n|x^n)}$ v.s. $\boxed{P_{Y^n}(y^n)}$ を考える

$W(y^n|x^n)$ = x^nを入力したときのy^nの確率

$P_{Y^n}(y^n) = \sum_{x^n} P_{X^n}(x^n) W(y^n|x^n)$

□ Neyman-Pearson test

$$S_{n,a}(x^n) = \{ y^n \in \mathcal{Y}^n \mid W(y^n|x^n) - e^{na} P_{Y^n}(y^n) > 0 \} \quad \text{—— ②}$$

□ メッセージ k を送信するときは

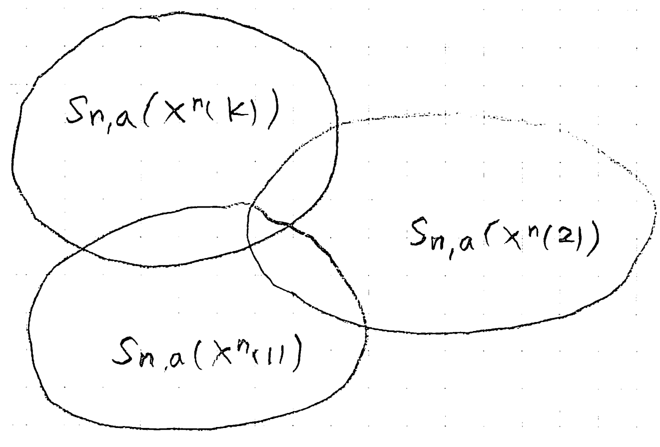
①より $x^n(k)$ が入力系列なので

y^n の確率分布は $W(y^n|x^n(k))$ のはず

y^n を受信したとき

$y^n \in S_{n,a}(x^n(k)) \Longrightarrow$ メッセージkを復号

としたい ⇝ これはダメ



交わったところでは，どちらのメッセージか決めれない

□ 交わりをとり除く（交わりはあきらめる）

$$T_{n,a}(x^n(k)) = S_{n,a}(x^n(k)) \setminus \bigcup_{\ell \neq k} S_{n,a}(x^n(\ell))$$

$$= S_{n,a}(x^n(k)) \cap \overline{\bigcup_{\ell \neq k} S_{n,a}(x^n(\ell))} \quad — ③$$

とおく

$$(\because \quad A \setminus B = A \cap \overline{B} \quad)$$

$T_{n,a}(x^n(k)) \quad (k = 1, 2, \cdots, M_n)$

には交わりがないので，y^nを受信したとき

$$y^n \in T_{n,a}(x^n(k)) \Rightarrow \text{メッセージ } k \text{ を復号} \quad (k = 1, 2, \cdots, M_n)$$

とする

(3) 誤り確率の評価

$$P_e = \frac{1}{M_n} \sum_{k=1}^{M_n} \Pr\{ Y^n \notin T_{n,a}(x^n(k)) \mid X^n = x^n(k) \}$$

$$= \frac{1}{M_n} \sum_{k=1}^{M_n} \underbrace{W(\overline{T_{n,a}(x^n(k))} \mid x^n(k))}_{\parallel} \quad — ④$$

$\sum_{y^n \in \overline{T_{n,a}(x^n(k))}} W(y^n \mid x^n(k))$ の意味

ここで ③ より

$$\overline{T_{n,a}(x^n(k))} = \overline{S_{n,a}(x^n(k))} \cup \left\{ \bigcup_{\ell \neq k} S_{n,a}(x^n(\ell)) \right\}$$

だから，union bound を用いると

$$④ \leq \frac{1}{M_n} \sum_{k=1}^{M_n} \left\{ W(\overline{S_{n,a}(x^n(k))} \mid x^n(k)) + \sum_{\ell \neq k} W(S_{n,a}(x^n(\ell)) \mid x^n(\ell)) \right\}$$

ここで $\boxed{\text{ランダムコーディングについての平均}}$ をとると

$$\underset{\substack{\uparrow \\ \boxed{\substack{X^n(1),\cdots,X^n(M_n) \\ \text{についての平均}}}}}{E[P_e]} \leq \frac{1}{M_n}\sum_{k=1}^{M_n}\Big\{ \underbrace{E\Big[W\big(\overline{S_{n,a}(X^n(k))} \mid X^n(k)\big)\Big]}_{⑤} + \underbrace{\sum_{l\neq k} E\Big[W\big(S_{n,a}(X^n(l)) \mid X^n(k)\big)\Big]}_{⑥} \Big\}$$

各項を評価していく

(3-1) ⑤ $= E_{X^n(1)\cdots X^n(M_n)}\Big[W\big(\overline{S_{n,a}(X^n(k))} \mid X^n(k)\big)\Big]$

$\uparrow$ $X^n(k)$ のみ

$= E_{X^n(k)}[\ \sim\ ]$

$= \displaystyle\sum_{x^n(k)\in\mathcal{X}^n} P_{X^n}(x^n(k))\, W\big(\overline{S_{n,a}(x^n(k))} \mid x^n(k)\big)$

Σの変数変換

$= \displaystyle\sum_{x^n\in\mathcal{X}^n} P_{X^n}(x^n)\, W\big(\overline{S_{n,a}(x^n)} \mid x^n\big)$ — ⑦

$= \displaystyle\sum_{x^n\in\mathcal{X}^n} P_{X^n}(x^n) \sum_{y^n\in\overline{S_{n,a}(x^n)}} W(y^n|x^n)$

$= \displaystyle\sum_{x^n\in\mathcal{X}^n}\ \sum_{y^n\in\overline{S_{n,a}(x^n)}} \underbrace{P_{X^n}(x^n)\, W(y^n|x^n)}_{=\,P_{X^nY^n}(x^n,y^n)}$

$= \displaystyle\sum_{x^n,y^n\in\overline{S_{n,a}}} P_{X^nY^n}(x^n,y^n)$ — ⑧

ただし

$S_{n,a} = \{ (x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n \mid y^n\in S_{n,a}(x^n) \}$ とおいた

ここで ② より

$S_{n,a} = \{ (x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n \mid W(y^n|x^n) - e^{na}P_{Y^n}(y^n) > 0 \}$

$= \{ (x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n \mid \underbrace{P_{X^nY^n}(x^n,y^n)}_{P_{X^n}(x^n)W(y^n|x^n)} - e^{na}P_{X^n}(x^n)P_{Y^n}(y^n) > 0 \}$

に注意すると

$$⑧ = P_{X^n Y^n}(\overline{S_{n,a}}) \longrightarrow 0 \quad (n \to \infty) \quad —⑨$$

$$\text{if} \quad a < I(X:Y) = D(P_{XY} \| P_X(\cdot)P_Y(\cdot)) \quad —⑩$$

[∵ $P(x,y) = P_X(x)W(y|x)$
$Q(x,y) = P_X(x)P_Y(y)$
として 仮説検定の定理(57ページ)
を用いる(大数の法則の帰結)]

(3-2)

$$⑥ = \sum_{\ell \neq k} E_{X^n(1)\cdots X^n(M_n)}\left[W(S_{n,a}(X^n(\ell)) \mid X^n(k)) \right]$$

↑ $X^n(k), X^n(\ell)$ のみ

$$= \sum_{\ell \neq k} E_{X^n(\ell)X^n(k)}\left[W(S_{n,a}(X^n(\ell)) \mid X^n(k)) \right]$$

$$= \sum_{\ell \neq k} \sum_{x^n(\ell) \in \mathcal{X}^n} \sum_{x^n(k) \in \mathcal{X}^n} P_{X^n}(x^n(\ell)) P_{X^n}(x^n(k)) W(S_{n,a}(x^n(\ell)) \mid x^n(k))$$

[∵ $X^n(\ell)$ と $X^n(k)$ の独立性]

$$= \sum_{\ell \neq k} \sum_{\hat{x}^n} \sum_{x^n} P_{X^n}(\hat{x}^n) P_{X^n}(x^n) \sum_{y^n \in S_{n,a}(\hat{x}^n)} W(y^n \mid x^n)$$

[∵ ⑦と同様の変数変換を行った, ℓによらない]

$$= (M_n - 1) \sum_{\hat{x}^n} P_{X^n}(\hat{x}^n) \sum_{x^n} \underbrace{\sum_{y^n \in S_{n,a}(\hat{x}^n)}}_{x^n\text{によらない}} \underbrace{P_{X^n}(x^n) W(y^n|x^n)}_{= P_{X^nY^n}(x^n, y^n)}$$

$$= (M_n - 1) \sum_{\hat{x}^n} P_{X^n}(\hat{x}^n) \sum_{y^n \in S_{n,a}(\hat{x}^n)} \underbrace{\sum_{x^n} P_{X^nY^n}(x^n, y^n)}_{= P_{Y^n}(y^n)} \quad —⑪$$

[$= (M_n - 1) E_{\hat{X}^n}\left[\Pr\{ Y^n \in S_{n,a}(\hat{X}^n) \} \right]$

$(\hat{X}^n, Y^n) \sim P_{X^n}(x^n) P_{Y^n}(y^n)$

□ $\hat{X}^n$ と Y^n が独立なのに
Y^n が $S_{n,a}(\hat{X}^n)$ に入る確率]

⑪ において,

$$y^n \in S_{n,a}(\hat{x}^n) \iff \underbrace{W(y^n|\hat{x}^n) - e^{na} P_{Y^n}(y^n) > 0}_{②\text{の定義}}$$

$$\iff e^{-na} W(y^n|\hat{x}^n) > P_{Y^n}(y^n)$$

を用いると

$$⑪ \leq (M_n - 1) \sum_{\hat{x}^n} P_{X^n}(\hat{x}^n) \sum_{y^n \in S_{n,a}(\hat{x}^n)} e^{-na} W(y^n|\hat{x}^n)$$

$$= (M_n - 1) e^{-na} \underbrace{\sum_{\hat{x}^n} \sum_{y^n \in S_{n,a}(\hat{x}^n)} P_{X^n}(\hat{x}^n) W(y^n|\hat{x}^n)}_{\text{確率なので }1\text{ 以下}}$$

$$\leq \frac{M_n}{e^{na}} \quad — ⑫$$

ここで $M_n = e^{nR}$　　$R < a$ — ⑬

とおくと 伝送レートは

$$\lim_{n\to\infty} \underbrace{\frac{1}{n} \log M_n}_{= R} = R$$

をみたして, ⑬ より ⑫ $\to 0$ $(n \to \infty)$ となる

以上より　$R < I(X;Y)$ ならば

$\underbrace{R < a}_{⑬\text{をみたす}} \underbrace{< I(X;Y)}_{⑩\text{をみたす}}$ となる a が存在して

よって

$$E[P_e] \to 0 \quad (n \to \infty)$$

これより, 少なくとも一つ符号(の列)が存在して

$$\begin{cases} \lim_{n\to\infty} P_e(f_n, g_n) = 0 & \text{かつ} \\ \lim_{n\to\infty} \frac{1}{n} \log M_n = R \end{cases}$$

をみたす

すなわち レート R は達成可能 (achievable)

通信路容量は達成可能なレートの上限(sup)
(達成可能: achievable)

$$C = \sup\{R \mid R \text{ is achievable}\}$$

であった。

$$R < I(X;Y) \Rightarrow R \text{ is achievable}$$

が示されたので，集合の包含関係

$$\{R \mid R < I(X;Y)\} \subset \{R \mid R \text{ is achievable}\}$$

が成立する。よって

$$C = \sup\{R \mid R \text{ is achievable}\}$$

$$\geq \sup\{R \mid R < I(X;Y)\} = I(X;Y)$$

が成立する。P_X は任意だったので，max をとると

$$C \geq \max_{P_X} I(X;Y)$$

$\square$

Remark（通信路符号化定理の意義）

- 通信速度の限界（＝通信路容量）を定め，限界までは良い符号が存在することを示している
  existence of good code
- 実用的な符号の構成には，さらに議論が必要 $\Longrightarrow$ 符号理論
  practical code $\rightsquigarrow$ theory of error-correcting code