２０１９年5月２０日（川端）

データ圧縮基礎資料

8　レート歪関数の計算（離散分布の場合）

本節では，離散定常無記憶情報源 X^n,$p_{X^n}(x^n) = \prod_{l=1} p_X(x_l)$ のレート歪み関数

$$R(D) = \inf_{p_{Y|X}:E[\rho(X,Y)]\le D} I(p_X, p_{Y|X}) \tag{1}$$

の計算について述べる.（ただし，$I(p_X, p_{Y|X})$ は $I(X;Y)$ のことである.）

8.1　ハミング歪に対する二値情報源の R(D)

定理 1 情報源アルファベットを $\mathcal{X} = \{0,1\}$，再生アルファベットを $\mathcal{Y} = \{0,1\}$ とする．歪はハミング歪すなわち $\rho(x,y) = x \oplus y$ とする．$\mathcal{X}$ 上の二値定常無記憶情報源を確率分布が $p = p_X(1) = 1 - p_X(0)$ となるように与える．このとき，

$$R(D) = \begin{cases} h(p) - h(D) & if \quad D \le p \\ 0 & if \quad D > p \end{cases}$$

証明: まず $p \le 1/2$ の場合を証明する.

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= h(p) - H(X \oplus Y|Y) \\ &\ge h(p) - H(X \oplus Y) \quad (\text{等号条件は } X \oplus Y \perp Y) \qquad (2) \\ &= h(p) - h(E(X \oplus Y)) \qquad (3) \end{aligned}$$

まず，$D \le p$ の場合を考える．このとき

$$(3) \ge h(p) - h(D) \quad (\text{等号には } E(X \oplus Y) = D \text{ が必要}) \tag{4}$$

が成立する．以上の等号条件を同時に満たす Y が存在することが等号のための必要十分条件である．$X = Y \oplus (Y \oplus X)$ に注意すると (2) の等号条件は，入力 Y に独立な加法的雑音 $Y \oplus X$ が加わる二元対称通信路の出力が X であることと同じである．$E(X \oplus Y) = D$ のとき，交差確率は D である．$r = P(Y = 1)$ と定義すると，以上の条件は

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

と書ける．これは

$$D(1-r) + r(1-D) = p$$

と同じであるから，r について解くと

$$r = \frac{p-D}{(1-D)-D},$$

となる．仮定した $D \le p$ の範囲では r が確率として存在していることがわかり，この範囲で定理が示された.

次に，$D > p$ の場合を考えると,

$$(3) \ge 0 \quad ((2) \text{ の等号条件と合わせて，等号は } Y \text{ が定数 } 0 \text{ であるときに成立})$$

がいえるので，この範囲で定理が示された.

$p \ge 1/2$ のときには $\mathcal{X}$,$\mathcal{Y}$ の 2 つの記号を同時に交換すれば上の議論がそのまま成り立つ．従って結果は変わらない．**証明終**

8.2 Lagrange 乗数を用いた R(D) に対する表現と Arimoto-Brahut 法

$R(D)$ に対する (1) の表現は，$E\rho(X,Y) \le D$ に対する Lagrange 未定乗数 $s \le 0$ を固定すると，以下の最小化問題として表現できる．

$$R(D) \quad = \quad sD + \min_{q_Y} \min_{p_{Y|X}} \left[\sum_x \sum_y p_X(x) p_{Y|X}(y|x) \log \frac{p_{Y|X}(y|x)}{q_Y(y) \exp_2(s\rho(x,y))} \right] \tag{5}$$

s はKuhn-Tucker条件すなわち，最小値を達成する $p^*_{Y|X}$ に対しては $D \ge \sum_x \sum_y p_X(x) p^*_{Y|X}(y|x)\rho(x,y)$ を満足する必要があり，特に不等号が厳密である場合には $s=0$ である必要がある．（以上を満たせば最適性の十分条件であることも知られている．）

そこで負の s に対して，(5) における D を $D_s = \sum_x \sum_y p_X(x) p^*_{Y|X}(y|x)\rho(x,y)$ と書いておくことにする．(5) の最適化は 2 重最小化問題になっている．

まず $p_{Y|X}$ 固定の下で $\min_{q_Y}$ を実行すると $q^*_Y(\cdot) = \sum_x p_X(x) p_{Y|X}(\cdot|x)$ において最小値

$$R(D_s) = I(p_X, p_{Y|X}) \tag{6}$$

がえられる．また，q_Y 固定の下で $\min_{p_{Y|X}}$ を実行すると $p^*_{Y|X}(y|x) = \frac{q_Y(y)\exp_2\{s\rho(x,y)\}}{\sum_y q_Y(y)\exp_2\{s\rho(x,y)\}}$ において最小値

$$R(D_s) \quad = \quad sD_s + \min_{q_Y} \left[\sum_x p_X(x) \log \left(\sum_y q_Y(y) \exp_2(s\rho(x,y)) \right)^{-1} \right] \tag{7}$$

が得られる．このときも，やはり，最小値を達成する $p^*_{Y|X}$ を用いて $D_s = \sum_x \sum_y p_X(x) p^*_{Y|X}(y|x)\rho(x,y)$. が成り立つものとしてよい．

最小化 (6) および最小化 (7) を繰り返すことで $R(D_s)$ に対する単調減少する上界が得られるが，この上界は $R(D_s)$ に収束することも知られている．これを Arimoto-Brahut アルゴリズムという．

8.3 R(D) に対する双対最適化

次に，$R(D_s)$ には，以下の双対表現がある．

$$R(D_s) = \max_{\lambda(\cdot)\in\Lambda_s} \left(sD + \sum_x p_X(x) \log \lambda(x) \right)$$

ただし，

$$\Lambda_s := \{\lambda(\cdot) : (\forall x)(\lambda(x) \ge 0), \quad (\forall y)(\sum_x p_X(x)\lambda(x) \exp_2 s\rho(x,y) \le 1)\}$$

証明: まず $\lambda(\cdot) \in \Lambda_s$ について

$$R(D_s) \ge sD_s + \sum_x p_X(x) \log \lambda(x)$$

を示そう．左辺引く右辺は

$$\begin{aligned} I(X;Y) - sD - \sum_x p_X(x) \log \lambda(x) \quad &= \quad \sum_x \sum_y p_X(x) p_{Y|X}(y|x) \log \frac{p_{YX}(y|x)}{q_Y(y)\lambda(x) \exp_2(s\rho(x,y))} \\ &= \quad \sum_x \sum_y p_X(x) p_{Y|X}(y|x) \log \frac{p_X(x) p_{YX}(y|x)}{p_X(x) q_Y(y)\lambda(x) \exp_2(s\rho(x,y))} \\ &\ge \quad \log \frac{1}{\sum_y q_Y(y)(\sum_x p_X(x)\lambda(x)\exp_2(s\rho(x,y)))} \\ &\ge \quad 0 \end{aligned} \tag{8}$$

最初の不等式は Log 和不等式を用いた．次に，等号の達成を示そう．(7) の最小化を与える q_Y を q_Y^* とし

$$\lambda(x) := \left(\sum_y q_Y^*(y) \exp_2 (s\rho(x,y)) \right)^{-1}$$

と定義すると $p_{Y|X}^*(y|x)p_X(x)/q_Y^*(y)$ が $\mathcal{X}$ 上の確率分布であることから

$$\sum_x p_X(x)\lambda(x) \exp_2 s\rho(x,y) = 1$$

が成り立つが，この $\lambda(x)$ を用いると，

$$R(D_s) := sD_s + \sum_x p_X(x) \log \lambda(x)$$

と書けている．証明終