

エントロピー・モデルにおける数理計画

神 保 雅 一

エントロピー・モデルは、情報が少なく、不確定要素が多い場合に、情報理論を応用してその内部構造の一端を探る1つの便利な手法であろうと思われる。ここでは、2, 3のエントロピー・モデルとそれに関する数理計画問題について論じる。

1. 1因子情報路(国沢[8])

ある商品に m 銘柄があり、それぞれの販売価格(コスト)が、 $l_1, \dots, l_m (> 0)$ である。このとき、大衆のランダムな選択による各銘柄の選択比率 $\mathbf{p} = (p_1, \dots, p_m)$ を、

$$\frac{H(\mathbf{p})}{\bar{l}(\mathbf{p})} \xrightarrow{\mathbf{p}} \max$$

となる \mathbf{p} を用いて推定するのが1因子情報路による方法である。ただし、

$$H(\mathbf{p}) = -\sum_{i=1}^m p_i \log p_i$$

$$\bar{l}(\mathbf{p}) = \sum_{i=1}^m l_i p_i$$

である。すなわち、 $H(\mathbf{p})$ は、大衆が銘柄を選択する際の選択のあいまいさの度合いを表わす measure であり、 \bar{l} は平均コストである。1因子情報路においては、単位コストあたりのあいまいさを最大にするような比率 \mathbf{p} で推定しようというわけである。この問題は、

$$\left. \begin{array}{l} \text{制約条件} \\ \sum_{i=1}^m p_i = 1 \quad p_i \geq 0 \quad (i=1, \dots, m) \\ \text{のもとで、} \\ H(\mathbf{p})/\bar{l}(\mathbf{p}) \\ \text{を最大にする。} \end{array} \right\} (1)$$

という数理計画問題と考えることができるが、その最適解は、 $p_i = w^{-l_i}$ (ただし w は $\sum w^{-l_i} = 1$ の最大正根) として得られ、その最大値は $\log w$ である。この問題に関連して、平均コスト \bar{l} を一定に保ち、エントロピーを最大にする問題

$$\left. \begin{array}{l} \text{制約条件} \\ \sum_{i=1}^m l_i p_i = \bar{l} \\ \sum_{i=1}^m p_i = 1, \quad p_i \geq 0 \\ \text{のもとで、} H(\mathbf{p}) \text{ を最大にする} \end{array} \right\} (2)$$

を考える。(2)の最適解を \bar{l} の関数とみなし $H(\bar{l})$ と書くと、(1)の解は、原点を通り $H(\bar{l})$ の曲線に接する直線の傾きである。

2. 判別関数によるモデル

2種類の確率分布 $\mathbf{p} = (p_1, \dots, p_m)$ と $\mathbf{r} = (r_1, \dots, r_m)$ に対して、

$$D(\mathbf{p}, \mathbf{r}) = -\sum_{i=1}^m p_i \log \frac{p_i}{r_i}$$

を、 \mathbf{p} と \mathbf{r} の判別関数とよぶ。(2)の形のモデルを拡張して、

$$\left. \begin{array}{l} \text{制約条件} \\ \sum_{i=1}^m b_{si} p_i = k_s \quad s=1, \dots, d \\ p_i \geq 0 \quad i=1, \dots, m \\ \text{のもとで、} D(\mathbf{p}, \mathbf{r}) \text{ を } \mathbf{p} \text{ に関して、} \\ \text{最小にする。 (ただし、} b_{si}, k_s \text{ は定数} \\ \text{であり、} \mathbf{r} \text{ は与えられた確率分布)} \end{array} \right\} (3)$$

なる凸計画問題を考える。この問題の最適解を求

めるには、反復尺度法 (Iterative Scaling Method) によるのが便利である (Darroch & Ratchif [6], 国沢 [8]). (3) の問題は、一般性を失うことなく、つぎのように修正できる。

$$\left. \begin{array}{l} \text{制約条件} \\ \sum_{i=1}^m a_{si} p_i = h_s \quad s=1, \dots, c \\ p_i \geq 0 \quad i=1, \dots, m \end{array} \right\} (3)'$$

のもとで、 $D(\mathbf{p}, \mathbf{r})$ を最小にする。
(ただし、 $a_{si} \geq 0$, $\sum_s a_{si} = 1$, $\sum_s h_s = 1$, $h_s \geq 0$)

(3)' の形の最適解は、つぎの反復尺度法によって求められる。

反復尺度法

- (i) $p_i^{(0)} = q_i$ を初期確率分布とする。
- (ii) $p_i^{(k+1)} = p_i^{(k)} \prod_{s=1}^c \left(\frac{h_s}{h_s^{(k)}} \right)^{a_{si}}$ $i=1, \dots, m$ とする。ここに、 $h_s^{(k)} = \sum_i a_{si} p_i^{(k)}$, $s=1, \dots, c$.

この手順によって、 $p_i^{(k)}$ は、(3)' の最適解を与える確率分布に収束する。また、(3) の最適解を与える確率分布は、

$$p_i = r_i \mu \prod_{s=1}^d \mu_s^{b_{si}} \quad i=1, 2, \dots, m$$

の形で与えられることも知られている。

また、(3)' の双対問題

$$\left. \begin{array}{l} \text{制約条件} \\ t_s > 0 \quad s=1, \dots, c \end{array} \right\} (3)''$$

のもとで、 $\theta(\mathbf{t}) = e \sum_i q_i \prod_s t_s^{a_{si}} - \sum_s h_s \log t_s$ を最小にする。

と (3)' との duality に関する議論が Charne^s and Cooper [3] にある。

3. Discrete Memoryless Channel における Capacity と Constrained Capacity

n 個の入力シンボルと m 個の出力シンボルをもつ Discrete Memoryless Channel を、

$$P = (p_{j|i}) \quad (i=1, \dots, m; j=1, \dots, n)$$

$$p_{j|i} \geq 0, \quad \sum_{j=1}^n p_{j|i} = 1$$

なる推移行列で表わす。そして、その相互情報量

$$I(\mathbf{p}, P) = \sum_i p_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_j}$$

の最大値を Capacity とよんでいる。ただし、 $\mathbf{p} = (p_1, \dots, p_m)$ および $\mathbf{q} = (q_1, \dots, q_n)$ はそれぞれ入力確率分布、出力確率分布である。Capacity を求める直接的な手法が、Muroga [9], Cheng [5], Takano [11] 等によって研究されている。一方で、Arimoto [1] と Blahut [2] は、Capacity を求める簡単な逐次近似法を提案した。さらに、Blahut は、

$$\left. \begin{array}{l} \text{制約条件} \\ \sum a_{si} p_i \leq h_s \quad s=1, \dots, c \\ p_i \geq 0 \quad i=1, \dots, m \end{array} \right\} (4)$$

なるもとで相互情報量 $I(\mathbf{p}, P)$ を最大にする

なる形の凸計画問題の最適解を求める近似法も導いている。この最適解を constrained capacity とよんでいる。

4. Relative Capacity

入力シンボル B_i を送るのに、コスト $l_i (> 0)$ を要する場合には、相互情報量を大きくする \mathbf{p} を求めるより、単位コストあたりの相互情報量を大きくするほうがより妥当であると思われる。ここでは、つぎのような数理計画問題について論じる。

$$\left. \begin{array}{l} \text{制約条件} \\ \sum_{i=1}^m p_i = 1 \quad p_i \geq 0 \end{array} \right\} (5)$$

のもとで、 $R(\mathbf{p}) = \frac{I(\mathbf{p}, P)}{\bar{l}(\mathbf{p})}$ を最大にする。ただし、 $\bar{l}(\mathbf{p}) = \sum_i l_i p_i$

この問題の最適解を relative capacity とよんでいる (Reza [10]). $R(\mathbf{p})$ は \mathbf{p} に関して quasi-concave であり、この問題は、quasi-convex programming であるが、Jimbo & Kunisawa [7] は、(5) の最適解を求める Iteration 法を提

案した. この方法は, 簡単であり, また systematic である. また, capacity を求める Arimoto, Blahut の結果を $l_i=1 (i=1, \dots, m)$ なる特殊ケースとして含んでいる. その Iteration 法はつぎのとおりである.

Iteration Procedure(*)

$a = \min l_i$ とし,

- (i) 初期確率 $p_i^{(0)} (>0)$ を任意に選ぶ.
- (ii) k ステップ目の確率分布 $\mathbf{p}^{(k)}$ を用いて, $(k+1)$ ステップ目の確率分布 $\mathbf{p}^{(k+1)}$ をつぎのようにする.

$$\tilde{p}_i^{(k+1)} = p_i^{(k)} \exp [a D_i^{(k)} / l_i]$$

ただし,

$$D_i^{(k)} = \sum p_{j|i} \log \frac{p_{j|i}}{q_j^{(k)}}$$

$$q_j^{(k)} = \sum p_i^{(k)} p_{j|i}$$

とし,

$$w^{(k)} = \sum_i \tilde{p}_i^{(k+1)}$$

$$p_i^{(k+1)} = \tilde{p}_i^{(k+1)} / w^{(k)}$$

によって $p_i^{(k+1)}$ を定める.

この Iteration 法に関して, つぎの定理が成り立つ.

定理 この Iteration Procedure(*) によって作られた $\mathbf{p}^{(k)} (k=0, 1, 2, \dots)$ に関して, $R(\mathbf{p}^{(k)})$ は単調に増加し, relative capacity C_R に収束する. また C_R を達成する出力確率分布 $\mathbf{q} = (q_1, \dots, q_n)$ は, 一意に決まる.

5. Iteration Procedure(*) の収束の証明

ここで, 前節の Iteration Procedure(*) の収束性を証明しておこう.

$$D_i(\mathbf{p}) = \sum_{j=1}^n p_{j|i} \log \frac{p_{j|i}}{q_j} \quad (i=1, \dots, m)$$

とおくと, 相互情報量 $I(\mathbf{p}, P)$ は,

$$I(\mathbf{p}, P) = \sum p_i D_i(\mathbf{p})$$

と書ける. k ステップ目と $(k+1)$ ステップ目の確率分布 $\mathbf{p}^{(k)}, \mathbf{p}^{(k+1)}$ を簡単のために, \mathbf{p}, \mathbf{r} で表わす. そして,

$$f_i = \exp [a D_i(\mathbf{p}) / l_i]$$

$$w = \sum_i p_i f_i$$

とおくと, $r_i = p_i f_i / w$ であり, つぎの補題が成り立つ.

補題 上で与えた確率分布 \mathbf{p}, \mathbf{r} に対して,

$$R(\mathbf{p}) \leq \frac{\log \lambda}{a} \leq R(\mathbf{r})$$

が成り立つ. ただし, $R(\mathbf{p}) = I(\mathbf{p}, P) / \bar{l}(\mathbf{p})$ であり, $\lambda = \sum_{i=1}^n u_i f_i$.

証明 Jensen の不等式より,

$$R(\mathbf{p}) \leq \frac{\log \lambda}{a}$$

であることが容易にわかる. $\mathbf{q} = (q_1, \dots, q_n)$, $\mathbf{s} = (s_1, \dots, s_n)$ を入力確率分布 \mathbf{p}, \mathbf{r} に対応する出力確率分布とすると,

$$D(\mathbf{r}, \mathbf{p}) \geq D(\mathbf{s}, \mathbf{q})$$

が成り立つ. ただし, $D(\cdot, \cdot)$ は判別関数である. このことに注意すると,

$$\begin{aligned} I(\mathbf{r}, P) &= \sum_i r_i D_i(\mathbf{r}) \\ &= \sum_i r_i D_i(\mathbf{p}) - D(\mathbf{s}, \mathbf{q}) \\ &\geq \frac{1}{a} \sum_i r_i l_i \log f_i - D(\mathbf{r}, \mathbf{p}) \end{aligned}$$

であり, $\lambda = \sum_i u_i f_i = w \bar{l}(\mathbf{r}) / \bar{l}(\mathbf{p})$ を用いて,

$$\begin{aligned} a I(\mathbf{r}, P) - \bar{l}(\mathbf{r}) \log \lambda &\geq \sum_i r_i l_i \log \frac{f_i}{w} - a D(\mathbf{r}, \mathbf{p}) - \bar{l}(\mathbf{r}) \log \frac{\bar{l}(\mathbf{r})}{\bar{l}(\mathbf{p})} \\ &= \sum r_i (l_i - a) \log \frac{r_i}{p_i} - \bar{l}(\mathbf{r}) \log \frac{\bar{l}(\mathbf{r})}{\bar{l}(\mathbf{p})} \\ &\geq (\bar{l}(\mathbf{r}) - a) \log \frac{\bar{l}(\mathbf{r}) - a}{\bar{l}(\mathbf{p}) - a} - \bar{l}(\mathbf{r}) \log \frac{\bar{l}(\mathbf{r})}{\bar{l}(\mathbf{p})} \end{aligned}$$

が成り立つ (上の不等式の最後の不等号は判別関数の凸性より導かれる). この不等式の最右辺が非負であることは, つぎの関数

$$f(z) = (z-a) \log \frac{z-a}{b-a} - z \log \frac{z}{b} \quad (b \geq a > 0)$$

が, $z \geq a$ で非負であることより直ちに導かれる.

これで収束の単調性が言えた. つぎに, Iteration Procedure(*) の収束性を示す.

定理の証明 $f_i^{(k)} = \exp [a D_i(\mathbf{p}^{(k)}) / l_i]$,

$u_i^{(k)} = p_i^{(k)} l_i / \bar{l}(\mathbf{p}^{(k)}) \quad (i=1, \dots, m; k=0, 1, 2, \dots)$ とおき,

$$w^{(k)} = \sum_{i=1}^n p_i^{(k)} f_i^{(k)},$$

$$\begin{aligned} \lambda^{(k)} &= \sum_{i=1}^n p_i^{(k)} l_i f_i^{(k)} / \bar{l}(\mathbf{p}^{(k)}) \\ &= \sum_{i=1}^n u_i^{(k)} f_i^{(k)} \end{aligned}$$

とおくと、定義より、 $p_i^{(k+1)} = p_i^{(k)} f_i^{(k)} / w^{(k)}$ である。補題により、任意の $k=0, 1, 2, \dots$ に対して、

$$R(\mathbf{p}^{(k)}) \leq \frac{\log \lambda^{(k)}}{a} \leq R(\mathbf{p}^{(k+1)}) \leq C_R$$

である。したがって、 $\{R(\mathbf{p}^{(k+1)})\}$ および $\{\log \lambda^{(k)}/a\}$ はともに単調に増加し、同じ値に収束する。さて、 \mathbf{p}^* を relative capacity を達成する確率分布とし、 $u_i^* = p_i^* l_i / \bar{l}(\mathbf{p}^*)$ とおく。さらに、 $\mathbf{q}^{(k)} (k=0, 1, 2, \dots)$ および \mathbf{q}^* を、それぞれ、入力確率 $\mathbf{p}^{(k)}$ 、 \mathbf{p}^* に対応する出力確率とする。 $u_i^{(k+1)} / u_i^{(k)} = f_i^{(k)} / \lambda^{(k)}$ に注意すれば、

$$\begin{aligned} &D(\mathbf{u}^*, \mathbf{u}^{(k)}) - D(\mathbf{u}^*, \mathbf{u}^{(k+1)}) \\ &= \sum_{i=1}^m u_i^* \log \frac{u_i^{(k+1)}}{u_i^{(k)}} \\ &= a \sum_{i=1}^m u_i^* D_i(\mathbf{p}^{(k)}) - \log \lambda^{(k)} \\ &= \frac{a}{\bar{l}(\mathbf{p}^*)} \left\{ \sum_{i=1}^m p_i^* D_i(\mathbf{p}^*) + D(\mathbf{q}^*, \mathbf{q}^{(k)}) - \log \lambda^{(k)} \right\} \\ &= a C_R - \log \lambda^{(k)} + \frac{a}{\bar{l}(\mathbf{p}^*)} D(\mathbf{q}^*, \mathbf{q}^{(k)}) \quad (6) \end{aligned}$$

を得る。この等式の右辺の最後の項は非負であるから、

$$\frac{1}{a} \left\{ D(\mathbf{u}^*, \mathbf{u}^{(k)}) - D(\mathbf{u}^*, \mathbf{u}^{(k+1)}) \right\} \geq C_R - \frac{\log \lambda^{(k)}}{a}$$

が $k=0, 1, 2, \dots$ に対して成り立つ。これらを $k=0$ から $k=N-1$ まで加えると、任意の N に対して、

$$\sum_{k=0}^{N-1} \left(C_R - \frac{\log \lambda^{(k)}}{a} \right) \leq \frac{1}{a} D(\mathbf{u}^*, \mathbf{u}^0) (< \infty)$$

が成り立つ。したがって、 $\{\log \lambda^{(k)}/a\}$ は relative capacity に収束し、ゆえに、 $\{R(\mathbf{p}^{(k)})\}$ も C_R に収束する。また、不等式(6)から、

$$0 \leq \frac{a}{\bar{l}(\mathbf{p}^*)} D(\mathbf{q}^*, \mathbf{q}^{(k)}) \leq D(\mathbf{u}^*, \mathbf{u}^{(k)}) - D(\mathbf{u}^*, \mathbf{u}^{(k+1)})$$

が、任意の $k=0, 1, 2, \dots$ に対して成り立つが、これらを前と同様に $k=0$ から $N-1$ まで加えると、

$$D(\mathbf{q}^*, \mathbf{q}^N) \longrightarrow 0 \quad (N \rightarrow \infty)$$

であることがわかる。したがって \mathbf{q}^* は一意である。

これで定理の証明が終わったわけであるが、この Iteration Procedure の収束の速さについては、relative capacity を達成する入力確率 \mathbf{p}^* が一意で任意の i に対して、 $p_i^* > 0$ であるときには、この収束の誤差は、指数オーダーで減少することがわかっている。また、 \mathbf{p}^* が relative capacity を達成するならつぎの Kuhn-Tucker 条件

$$\begin{aligned} \frac{D_i(\mathbf{p}^*)}{l_i} &= C_R & p_i^* > 0 \text{ の時} \\ &\leq C_R & p_i^* = 0 \text{ の時} \end{aligned}$$

を満たさなければならないことも容易にわかる。

ところで relative capacity をつぎのように解釈することもできる。すなわち、計画問題(4)において、制約条件を2つにした場合、

$$\left. \begin{aligned} &\text{制約条件} \\ &\sum_{i=1}^m l_i p_i \leq \bar{l} \\ &\sum_{i=1}^m p_i = 1, \quad p_i \geq 0 \end{aligned} \right\} (4)'$$

のもとで、相互情報量 $I(\mathbf{p}, P)$ を最大にする

なる問題の最適解を、とくに capacity at cost とよび、 $C(\bar{l})$ と書くと、 $C(\bar{l})$ は \bar{l} に関して単調増加であり凹である。この曲線 $C(\bar{l})$ に接して、原点を通る直線の傾きが relative capacity である。

6. Relative Capacity のエントロピー・モデル的解釈

市場に n 銘柄 B_1, \dots, B_n があり、その販売価格は市場の状態 $S_i (i=1, \dots, m)$ によって異なるとする。状態が S_i であるときに銘柄 B_j の販売価格は $l_{ij} (> 0)$ であるとする。また、状態 S_i のもとの各銘柄の選択比率 $p_{j|i} (j=1, \dots, n)$ はわかっているものとする。もしこの選択比率がわからないならば、1 因子情報路を用いて $p_{j|i} (j=1, \dots,$

n)を推定すればよい。われわれは、どの状態が生起するかということに関してなら情報を得られないとして、各銘柄 B_j が選択される比率を推定したい。ここで、 $\mathbf{p}=(p_1, \dots, p_m)$ を状態 $\mathbf{S}=(S_1, \dots, S_m)$ の生起確率とし、つぎの3つの仮説を設定してみる。

I 平均コスト $\bar{l}(\mathbf{p})=\sum_{ij} p_i p_{j|i} l_{ij}=\sum_i p_i l_i$ を小さくしたい(ただし、 $l_i=\sum_j p_{j|i} l_{ij}$)。

II 状態は、どの状態にもかたよらない生起をしようとする。

III 各状態における最適配分比率 $\mathbf{p}_{\cdot|i}=(p_{1|i}, \dots, p_{n|i})$ と $\mathbf{q}=(q_1, \dots, q_n)$ との「距離(判別関数)」

$$D_i=\sum_j p_{j|i} \log \frac{p_{j|i}}{q_j} \quad i=1, \dots, m$$

の平均 $\sum_i p_i D_i$ をできるだけ小さくしたい。すなわち、どの配分比率にもかたよらない配分をした。

これら3つの仮説を同時に満足するような配分比率 \mathbf{q} を推定することを考える。ところで、

$$\sum_{ij} p_i p_{j|i} \log \frac{p_{j|i}}{q_j} \geq \sum_{ij} p_i p_{j|i} \log \frac{p_{j|i}}{\sum_i p_i p_{j|i}}$$

であるから、仮説 III は、 $q_j=\sum_i p_i p_{j|i}$ とおくことによって満足される。したがって、 $\sum_i p_i D_i=I(\mathbf{p}, P)$ となり、

$$\frac{\sum_i p_i D_i}{\bar{l}(\mathbf{p})} = \frac{I(\mathbf{p}, P)}{\bar{l}(\mathbf{p})} = R(\mathbf{p})$$

を最大にする問題に帰着される。したがって、これは(5)の解、すなわちrelative capacityを求める問題に他ならない。

参 考 文 献

[1] Arimoto, S., An Algorithm for Computing the Capacity of Arbitrary Discrete Memory-

less Channels, *IEEE Trans. Inform. Theory*, IT-18, 14-20(1972).

[2] Blahut, R. E., Computation of Capacity and Rate Distortion Functions, *IEEE, Trans. Inform. Theory*, IT-18, 460-473(1972).

[3] Charnes and Cooper, Constrained Kullback-Leibler estimation; generalized Cobb-Douglas balance and unconstrained convex programming, *Atti. Accad. Naz. Lincei, Rc ser 8*, 58, 568-576.

[4] Charnes, Cooper, and Learner, Constrained Information Theoretic Characterization in Consumer Purchase Behaviour, *J. Opr. Res. Soc.* vol. 29.9 pp. 833-842.

[5] Cheng, M. C., On the Computation of Capacity of a Discrete Memoryless Channel, *Inform. Contr.*, 24, 292-298.

[6] Darroch and Ratchiff, Generalized iterative scaling for log-linear models, *Annals of Math. Stat.* vol. 43, 1972, 1470-1480.

[7] Jimbo, M. and Kunisawa, K., An Iteration Method for Calculating the Relative Capacity (to appear).

[8] 国沢清典, エントロピーモデル, 日科技連出版。

[9] Muroga, S., On the capacity of a Discrete Channel I, *J. Phys. Soc. Jap.*, vol. 8, 484-494 (1953).

[10] Reza, F. M., *An Introduction to Information Theory*, McGraw-Hill, New York.

[11] Takano, S., On a method of Calculating the Capacity of a Discrete Memoryless Channel, *Inform. Contr.* 29, 327-336(1975).

(じんぼ・まさかず 東京理科大学)

会員名簿ができます

先に皆様よりお送りいただいた名簿作成資料にもとづき、現在名簿発行の準備をすすめており、12月にはでき上る予定です。頒布は希望者のみとさせていただきますので、学会事務局 (03-815-3351) までお申込みください (頒布価格 1000円)。