統計数理 第35巻 第1号(1987)

特別寄稿

情報量と統計

統計数理研究所 河田敬義*

(1987年2月 受付)

目次

1. はじめに
2. いろいろな情報量
2.1 KL情報量, Pearson情報量, 角谷情報量, I^{λ}
2.2 相対性と一意性
3. 情報量の公理系
3.1 公理系
3.2 基本情報量
3.3 基準情報量, 双曲的情報量, 楕円的情報量の特徴づけ
3.4 可微分基本情報量
4. L集合と情報量
4.1 情報量のL集合による特徴づけ
4.2 弧長型・面積型・幅型情報量
4.3 情報量の族の弱完備性と強完備性
5. 情報量と統計
5.1 正則情報量とその評価
5.2 分布族のパラメータの推定
5.3 AIC
謝辞

1. はじめに

今日, 情報量(Information)といえば, ふつうKullback-Leiblerの情報量を指している. (岩波数学辞典第3版, 178 情報理論, p. 464; 285 統計的決定関数, p. 834). すなわち, 二つの有限分布 $\boldsymbol{p}=(p_1,\cdots,p_m)$, $\boldsymbol{q}=(q_1,\cdots,q_m)$ に対して, KL情報量は

$$I_{KL}(\boldsymbol{p};\boldsymbol{q})=\sum_{k=1}^{m} p_k \log\frac{p_k}{q_k} \tag{1}$$

と定義される. この量は1951年にKullback-Leibler (1951)によって与えられたが, それは古くは19世紀のBoltzmannの統計力学におけるエントロピー理論に始まり, Shannon (1948)の情報理論における情報量(エントロピー)の定義をへて, 次第に発展して来たものである.

情報量の概念は, 情報理論にとどまらず, また統計理論においても有効に用いられている. そ

---

* 東京大学名誉教授, 統計数理研究所名誉所員(元所長)

式番号は各章ごとに(1), (2), …とつけてある. また引用は章内に限った.

の一例として赤池弘次氏の AIC（Akaike Information Criterion, 赤池情報量規準）の理論 Akaike（1973, 1974）を挙げることができる．すなわち，実験データとして与えられた分布 $\boldsymbol{p}^0=\left(\frac{n_1}{n},\cdots,\frac{n_m}{n}\right)$ と統計モデル $\boldsymbol{q}=(q_1(\theta_1,\cdots,\theta_r),\cdots,q_m(\theta_1,\cdots,\theta_r))$ とに対して

$$\text{(a)} \qquad I_{KL}(\boldsymbol{p}^0;\ \boldsymbol{q}(\theta))=\min$$

となる．$\theta=\hat{\theta}$（最尤推定値）を求め，近似の良さと自由度の両者を含む量：

$$\text{(b)} \qquad \mathrm{AIC}=-2(\text{最大対数尤度})+2r=2nI_{KL}(\boldsymbol{p}^0;\ \boldsymbol{q}(\hat{\theta}))+2r+c_0(\boldsymbol{p}^0)$$

を定義する．与えられたデータ $\boldsymbol{p}^0$ と，それに対するいくつかの統計モデルがあるとき，AIC の値が最小となるようなモデルを選択すればよいというのが，モデル選択の最小 AIC 法である．もちろん統計への応用はこれだけではない．

赤池氏の AIC の定義は予測の立場に立つ情報量の利用であるが(例えば，坂元 他(1983)，第四章参照)，情報量の考えが，明らさまにではないが，統計に用いられるようになったのは 1920 年代の Fisher にまでさかのぼることができる．例えば，Fisher の与えた最尤推定量が実は情報量より (a) によって与えられること，また K. Pearson のカイ自乗検定法に用いられる χ^2 は近似的に $2nI_{KL}(\boldsymbol{p}^0;\ \boldsymbol{q})$ に等しく（5.1, 注意参照），Fisher（1925）に見られるように，カイ自乗検定法は単に与えられた有意水準（例えば 5%）による検定と見るよりは，χ^2 の値自身に着目していることが指摘される．

このように情報量の概念がだんだんと広い分野で応用されて行くにつれて，"何故に情報量を (1) のように定義しなければならないか" ということが自ら問題となるであろう．(1) の定義にいたるには歴史的発展があり，確かにその意味での必然性があった．しかし Kolmogorov が確率の公理系を与えて，それまで多岐にわたる確率概念に明確な見とおしを与え，以来確率論の飛躍的発展を見たように，情報量に対しても，その公理系をしかるべく定義して，より広い統一的な立場に立って理論と応用を発展させることができないかと考えるのは，極めて自然なことであると思われる．

(1) 式で与えられた $I_{KL}(\boldsymbol{p};\ \boldsymbol{q})$ は，$\boldsymbol{p}$, $\boldsymbol{q}$ に関して対称的ではなく，分布 $\boldsymbol{p}$ の分布 $\boldsymbol{q}$ への近さ，或は分布 $\boldsymbol{p}$ の分布 $\boldsymbol{q}$ からのへだたり（Divergence）を表わすと見られる．(1) で定義された $I_{KL}(\boldsymbol{p};\ \boldsymbol{q})$ が満足する性質として，Kullback（1959）は，非負性，不変性，凸性，加法性，相対性などを挙げている．これに対して，従来情報量とは呼ばれてはいないが，類似の性質を持つ量がいくつか知られている．その一つは古くから知られている K. Pearson (1900) のカイ自乗検定法に用いられた量で，実験データの分布 $\boldsymbol{p}=\left(\frac{n_1}{n},\cdots,\frac{n_m}{n}\right)$ と理論分布 $\boldsymbol{q}(q_1,\cdots,q_m)$ とに対して

$$\text{(2)} \qquad I_P(\boldsymbol{p};\ \boldsymbol{q})=\frac{1}{n}\sum_{k=1}^{m}\frac{(n_k-nq_k)^2}{nq_k}=\left(\sum_{k=1}^{m}\frac{p_k^2}{q_k}\right)-1$$

で与えられたものである．また別に角谷静夫氏が 1948 年に無限直積測度の収束を論じたときに用いられた量（Kakutani (1948)）：

$$\text{(3)} \qquad I_K(\boldsymbol{p};\ \boldsymbol{q})=2\left(1-\sum_{k=1}^{m}\sqrt{p_kq_k}\right)$$

も一つの情報量と考えられる．さらに (2), (3) を特殊な場合として含む量：

$$\text{(4)} \qquad I^\lambda(\boldsymbol{p};\ \boldsymbol{q})=\frac{1}{\lambda}\left\{\left(\sum_{k=1}^{m}p_k^{1+\lambda}q_k^{-\lambda}\right)-1\right\}, \qquad -\frac{1}{2}\leqq\lambda<\infty,\quad \lambda\neq0$$

(すなわち，$I_P=I^1$, $I_K=I^{-1/2}$) を情報量の一種として定義することができる (2.1 参照).

これらの例を勘案して，ここで "情報量" $I(\boldsymbol{p}\,;\,\boldsymbol{q})$ を五つの性質：(I) 簡約性，(II) 対称性，(III) 非負性，(IV) 不変性，(V) 凸性を公理系として抽象的に定義しよう (3.1 参照).

特に，上記諸例のように

$$I(\boldsymbol{p}\,;\,\boldsymbol{q})=\sum_{k=1}^{m} L(p_k,\ q_k) \tag{5}$$

の形に表わされる場合に，$I(\boldsymbol{p}\,;\,\boldsymbol{q})$ を "基本情報量" と呼んで，その場合についてはやや立ち入って性質をしらべることができる (3.2 参照).

さらに KL 情報量 I_{KL} (および I^λ) が加法性 (擬加法性) を持つ "基本情報量" として特徴づけられることが示される (3.3 参照). これは何故に I_{KL} がもっぱら用いられるかということに対する一つの答えと見なされる. 以上によって，われわれの "情報量" の定義は一応妥当なものと思われる.

これまでの結果を1985年夏に筆者が工藤弘吉氏に知らせたところ，工藤氏は直ちに1952-1953年の結果を知らせて下さった (Kudō (1952), 工藤 (1953)). そこでは二つの分布 $\boldsymbol{p}, \boldsymbol{q}$ に対して Lyapunov 集合 (L 集合) を

$$L(\boldsymbol{p},\ \boldsymbol{q})=\left\{(x,\ y)\ \middle|\ x=\sum_{k=1}^{m} \alpha_k p_k,\ y=\sum_{k=1}^{m} \alpha_k q_k,\ 0\leqq\alpha_k\leqq 1,\ k=1,\cdots,m\right\} \tag{6}$$

と定義するとき，$L(\boldsymbol{p},\ \boldsymbol{q})$ は $\left(\frac{1}{2}, \frac{1}{2}\right)$ を中心とする凸集合で，統計に関する多くの結果が L 集合を用いることによって簡明に導かれることが示されている. そして，工藤氏は "情報量" がこの L 集合を用いて特徴づけられるであろうことを示唆された. それによって，実際に "情報量" が L 集合全体の族の上の非負・単調な汎関数として特徴づけられることが容易にわかった (4.1 参照). また，その特徴づけによって "基本情報量" の幾何学的解釈を与えることができ，さらに，"情報量" のいくつかの新しい例を与えることができた (4.2 参照). また，工藤氏の論文 (工藤 (1953)) にならって，"情報量" の或る種の族の (弱・強) 完備性 (すなわち，逆に L 集合の相等・包含を決定するという性質) を示すことができた (4.3 参照).

このような L 集合と "情報量" との深い関係を見るとき，われわれの "情報量" の公理系による定義の妥当性が深められたように思われた.

最後に，統計への一つの応用を試みた (第5章参照). 赤池氏の AIC の理論では I_{KL} を用いていろいろな計算を行っているが，I_{KL} の代りに他の情報量を用いて同様な計算ができないかということを考えた. そして，良く知られている計算の組み合わせによって，上記 I^λ について成り立つような或る種の十分条件を満足すればよいということを示すことができた.

以上，"情報量" $I(\boldsymbol{p}\,;\,\boldsymbol{q})$ は $\boldsymbol{p}, \boldsymbol{q}$ が有限分布である場合についてのみ考察した. これを一般の分布 $\boldsymbol{p}, \boldsymbol{q}$ の場合に拡張できないかという問題が残る. 事実，Kullback (1959), Kudō (1952), 工藤 (1953), Ali-Silvey (1966), Csiszár (1978) たちは一般の分布の場合を扱っている. 恐らく大部分の結果は適当な修正の下でほとんどそのまま成り立つであろうと思われるが，ここでは立ち入らないことにする.

2. いろいろな情報量

2.1 KL情報量，Pearson情報量，角谷情報量，I^{λ}

全事象 I の分割

$$\boldsymbol{E}=\{E_1,\cdots,E_m\},\qquad I=E_1\cup\cdots\cup E_m,\ E_k\neq\emptyset,\qquad E_i\cap E_j=\emptyset,\ (i\neq j)$$

を定め，その上の二つの確率分布

$$\boldsymbol{p}=(p_1,\cdots,p_m),\ p_k\geqq 0,\ p_1+\cdots+p_m=1,\quad \boldsymbol{q}=(q_1,\cdots,q_m),\ q_k\geqq 0,\ q_1+\cdots+q_m=1$$

をとる．

定義 1. (i) 分布 $\boldsymbol{p}$ の分布 $\boldsymbol{q}$ に対する KL 情報量を

$$I_{KL}(\boldsymbol{p}\,;\,\boldsymbol{q})=\sum_{k=1}^{m}p_k\log\frac{p_k}{q_k} \tag{1}$$

と定める．但し， $0\log\dfrac{0}{q_k}=0$, $p_k\log\dfrac{p_k}{0}=\infty$, $(p_k\neq 0)$, $0\log\dfrac{0}{0}=0$ と定める（Kullback-Leibler (1951)）．

(ii) 分布 $\boldsymbol{p}$ の分布 $\boldsymbol{q}$ に対する Pearson 情報量を

$$I_P(\boldsymbol{p}\,;\,\boldsymbol{q})=\sum_{k=1}^{m}\frac{p_k^2}{q_k}-1 \tag{2}$$

と定める．但し， $\dfrac{0}{q_k}=0$, $\dfrac{p_k^2}{0}=\infty$, $(p_k\neq 0,\ q_k\neq 0)$, $\dfrac{0^2}{0}=0$ と定める．

$$\boldsymbol{p}=\left(\frac{n_1}{n},\frac{n_2}{n},\cdots,\frac{n_m}{n}\right),\qquad n=n_1+\cdots+n_m$$

の場合には，Pearson の χ^2 と比べると

$$\begin{aligned}\chi^2&=\sum_{k=1}^{m}\frac{(n_k-nq_k)^2}{nq_k}=n\sum_{k=1}^{m}\frac{1}{q_k}\left(\frac{n_k}{n}-q_k\right)^2=n\sum_{k=1}^{m}\frac{(p_k-q_k)^2}{q_k}\\&=n\left\{\sum_{k=1}^{m}\frac{p_k^2}{q_k}-1\right\}=nI_P(\boldsymbol{p}\,;\,\boldsymbol{q})\end{aligned}$$

である（Pearson (1900)）．

(iii) 分布 $\boldsymbol{p}$ の分布 $\boldsymbol{q}$ に対する角谷情報量を

$$I_K(\boldsymbol{p}\,;\,\boldsymbol{q})=2\left(1-\sum_{k=1}^{m}\sqrt{p_kq_k}\right) \tag{3}$$

と定める．ベクトル $\sqrt{\boldsymbol{p}}=(\sqrt{p_1},\cdots,\sqrt{p_m})$, $\sqrt{\boldsymbol{q}}=(\sqrt{q_1},\cdots,\sqrt{q_m})$ とおくと，ユークリッドのノルム $\|\ \|$ に関して $\|\sqrt{\boldsymbol{p}}\|=\|\sqrt{\boldsymbol{q}}\|=1$ であるから

$$I_K(\boldsymbol{p}\,;\,\boldsymbol{q})=\|\sqrt{\boldsymbol{p}}-\sqrt{\boldsymbol{q}}\|^2 \tag{4}$$

と表わされる（Kakutani (1948), Matusita (1951), 工藤 (1953)）．

これらの情報量 I_{KL}, I_P, I_K は定理2に見るように，いずれも非負性，不変性，凸性などを持っているが，I_{KL}, I_P, I_K をすべて特殊な場合として含むような連続パラメータ λ を持つ情報量をも簡単に定義することができる．

定義 2.　分布 $\boldsymbol{p}$ の分布 $\boldsymbol{q}$ に対するパラメータ λ を持つ情報量を

$$I^{\lambda}(\boldsymbol{p}\,;\,\boldsymbol{q})=\frac{1}{\lambda}\left(\sum_{k=1}^{m} p_k^{1+\lambda}q_k^{-\lambda}-1\right),\qquad -\frac{1}{2}\leqq\lambda<\infty,\quad \lambda\neq 0 \tag{5}$$

と定める ($\lambda>0$ のとき，$p_k=0$ または $q_k=0$ に対しては I_P と同様に定める．また I^{λ}, $-1/2\leqq\lambda<0$ は工藤 (1953), Ali and Silvey (1966), Csiszār (1978) にも挙げてある).

これは

$$I^{\lambda}(\boldsymbol{p}\,;\,\boldsymbol{q})=\frac{1}{\lambda}\left(\sum_{k=1}^{m}\frac{p_k^{1+\lambda}}{q_k^{\lambda}}-1\right),\qquad \lambda=0 \tag{6}$$

$$I^{-\mu}(\boldsymbol{p}\,;\,\boldsymbol{q})=\frac{1}{\mu}\left(1-\sum_{k=1}^{m} p_k^{1-\mu}q_k^{\mu}\right),\qquad 0<\mu\leqq\frac{1}{2} \tag{7}$$

と書く方が見易い．特殊な場合として

$$\begin{aligned} I^{1}(\boldsymbol{p}\,;\,\boldsymbol{q})&=I_P(\boldsymbol{p}\,;\,\boldsymbol{q}), &&\text{Pearson 情報量}\\ I^{-1/2}(\boldsymbol{p}\,;\,\boldsymbol{q})&=I_K(\boldsymbol{p}\,;\,\boldsymbol{q}), &&\text{角谷情報量}\end{aligned} \tag{8}$$

である．また (5) で定義されなかった $\lambda=0$ に対しては

$$I^{0}(\boldsymbol{p}\,;\,\boldsymbol{q})=\sum_{k=1}^{m} p_k\log\frac{p_k}{q_k}=I_{KL}(\boldsymbol{p}\,;\,\boldsymbol{q}),\qquad \text{KL 情報量} \tag{9}$$

と定める．

連続パラメータ λ を持つ情報量については，まず λ について次の単調性と連続性が成り立つ．

定理 1.　情報量 $I^{\lambda}(\boldsymbol{p}\,;\,\boldsymbol{q})$, $-1/2\leqq\lambda<\infty$ は

(I)　単調性：$\lambda<\lambda'$ であれば，任意の $\boldsymbol{p}$, $\boldsymbol{q}$ に対して

$$I^{\lambda}(\boldsymbol{p}\,;\,\boldsymbol{q})\leqq I^{\lambda'}(\boldsymbol{p}\,;\,\boldsymbol{q}) \tag{10}$$

等号は $\boldsymbol{p}=\boldsymbol{q}$ または $I^{\lambda}(\boldsymbol{p}\,;\,\boldsymbol{q})=\infty$ の場合に限る．

(II)　連続性：$\lim_{n\to\infty}\lambda_n=\lambda_0$ であれば，任意の $\boldsymbol{p}$, $\boldsymbol{q}$ に対して

$$\lim_{n\to\infty} I^{\lambda_n}(\boldsymbol{p}\,;\,\boldsymbol{q})=I^{\lambda_0}(\boldsymbol{p}\,;\,\boldsymbol{q}) \tag{11}$$

である．

(I) の特殊な場合として $\lambda=-1/2,\,0,\,1$ の値を比べれば

$$I_K(\boldsymbol{p}\,;\,\boldsymbol{q})\leqq I_{KL}(\boldsymbol{p}\,;\,\boldsymbol{q})\leqq I_P(\boldsymbol{p}\,;\,\boldsymbol{q}) \tag{12}$$

であって，等号が成り立つのは，$\boldsymbol{p}=\boldsymbol{q}$ の場合（または ∞ の値をとる場合）に限る．

証明　$x>0$ に対して

$$K(x,\lambda)=\frac{1}{\lambda}(x^{-\lambda}-1),\qquad -\frac{1}{2}\leqq\lambda<\infty,\quad \lambda\neq 0 \tag{13}$$

$$K(x,0)=-\log x \tag{13*}$$

とおく．(5) と比べて $\left(\sum_{k=1}^{m} p_k=1\text{ を用いて}\right)$

(14) $$I^{\lambda}(\boldsymbol{p}\,;\,\boldsymbol{q})=\sum_{k=1}^{m}p_k K\left(\frac{q_k}{p_k},\,\lambda\right)$$

が成り立つ．一方 $x^{\lambda}=\exp(\lambda\log x)$ であるから，$K(x,\lambda)$ は $(-1/2\leqq\lambda<\infty$ の範囲で$)$ λ について連続かつ微分可能で，かつ

$$\frac{\partial K}{\partial\lambda}(x,\,\lambda)\geqq 0,\qquad x>0,\ -\frac{1}{2}\leqq\lambda<\infty$$

である．よって(14)より単調性および連続性が成り立つ．

$I^{\lambda}(\boldsymbol{p}\,;\,\boldsymbol{q})$, $\lambda\geqq 0$ の取り得る値は

(15) $$0\leqq I^{\lambda}(\boldsymbol{p}\,;\,\boldsymbol{q})\leqq\infty$$

で，∞ となるのは，或る k について $p_k>0$, $q_k=0$ の場合に限る．また $I^{-\mu}(\boldsymbol{p}\,;\,\boldsymbol{q})$, $0<\mu\leqq 1/2$ については

(16) $$0\leqq I^{-\mu}(\boldsymbol{p}\,;\,\boldsymbol{q})\leqq\frac{1}{\mu}$$

である．特に $I^{-\mu}(\boldsymbol{p}\,;\,\boldsymbol{q})=1/\mu$ となるのは $\boldsymbol{p}\perp\boldsymbol{q}$ (すなわち，各 $k=1,\cdots,m$ について $p_k=0$ または $q_k=0$) の場合に限る．

$I^{-1/2}$ の場合には，任意の $\boldsymbol{p}$, $\boldsymbol{q}$ に対して

(17) $$I^{-1/2}(\boldsymbol{p}\,;\,\boldsymbol{q})=I^{-1/2}(\boldsymbol{q}\,;\,\boldsymbol{p})$$

であるが，$\lambda\neq-1/2$ であれば，$\boldsymbol{p}\neq\boldsymbol{q}$ に対して

(18) $$I^{\lambda}(\boldsymbol{p}\,;\,\boldsymbol{q})\neq I^{\lambda}(\boldsymbol{q}\,;\,\boldsymbol{p})$$

である．

注意 (13), (13)*, (14)より，形式的に

(19) $$I^{\lambda}=\frac{1}{\lambda}\{\exp(\lambda I^{0})-1\},\qquad \lambda\neq 0$$

と表わされる．

定義 3. $I^{0}=I_{KL}$ を放物的情報量または基準情報量 (Canonical information), I^{λ}, $\lambda>0$ を双曲的情報量，$I^{-\mu}$, $0<\mu\leqq 1/2$ を楕円的情報量と呼ぶことにする．

注意 $I^{\lambda}(\boldsymbol{p}\,;\,\boldsymbol{q})$ の定義(5)において $\lambda<-1/2$ の範囲にまで形式的に定義を拡張すれば

(20) $$I^{-\lambda}(\boldsymbol{p}\,;\,\boldsymbol{q})=-\frac{\lambda-1}{\lambda}I^{\lambda-1}(\boldsymbol{q}\,;\,\boldsymbol{p}),\qquad \lambda>1$$

(20)* $$I^{-1}(\boldsymbol{p}\,;\,\boldsymbol{q})=0$$

(20)** $$I^{-\mu}(\boldsymbol{p}\,;\,\boldsymbol{q})=\frac{1-\mu}{\mu}I^{\mu-1}(\boldsymbol{q}\,;\,\boldsymbol{p}),\qquad \frac{1}{2}\leqq\mu<1$$

となる．

定理 2. 情報量 $I^{\lambda}(\boldsymbol{p}\,;\,\boldsymbol{q})$, $-1/2\leqq\lambda<\infty$ は，各 λ に対して，次の諸性質を持つ：

$$I^\lambda(\boldsymbol{p}\,;\ \boldsymbol{q})=I^\lambda(p_1,\cdots,p_m\,;\ q_1,\cdots,q_m)$$

は

(I)　簡約性：

(21)　$$I^\lambda(p_1,\cdots,p_{m-1},\ 0\,;\ q_1,\cdots,q_{m-1},\ 0)=I^\lambda(p_1,\cdots,p_{m-1}\,;\ q_1,\cdots,q_{m-1})$$

(II)　対称性：$(1,\ 2,\cdots,m)$ の任意の置換 $(i_1,\cdots,i_m)$ に対して

(22)　$$I^\lambda(p_{i_1},\ p_{i_2},\cdots,p_{i_m}\,;\ q_{i_1},\ q_{i_2},\cdots,q_{i_m})=I^\lambda(p_1,\cdots,p_m\,;\ q_1,\cdots,q_m)$$

(III)　非負性：任意の $\boldsymbol{p},\ \boldsymbol{q}$ に対して

(23)　$$I^\lambda(\boldsymbol{p}\,;\ \boldsymbol{q})\geqq 0$$

特に等号が成り立つのは $\boldsymbol{p}=\boldsymbol{q}$ の場合に限る．

次に分割 $\boldsymbol{E}=\{E_1,\ E_2,\cdots,E_m\}$ に対して，或る二つの事象（例えば E_1 と E_2）を合併して $\boldsymbol{E}'=\{E_1\cup E_2,\ E_3,\cdots,E_m\}$ を作る．$\boldsymbol{E}$ 上の分布

$$\boldsymbol{p}=(p_1,\cdots,p_m),\qquad \boldsymbol{q}=(q_1,\cdots,q_m)$$

から $\boldsymbol{E}'$ 上の分布

$$\boldsymbol{p}'=(p_1+p_2,\ p_3,\cdots,p_m),\qquad \boldsymbol{q}'=(q_1+q_2,\ q_3,\cdots,q_m)$$

が定まる．このとき

(IV)　不変性：もしも

(24)　$$\frac{q_1}{p_1}=\frac{q_2}{p_2}=\frac{q_1+q_2}{p_1+p_2}$$

であれば

(25)　$$I^\lambda(\boldsymbol{p}\,;\ \boldsymbol{q})=I^\lambda(\boldsymbol{p}'\,;\ \boldsymbol{q}')$$

すなわち

(25)*　$$I^\lambda(p_1,\ p_2,\cdots,p_m\,;\ q_1,\ q_2,\cdots,q_m)=I^\lambda(p_1+p_2,\ p_3,\cdots,p_m\,;\ q_1+q_2,\ q_3,\cdots,q_m)$$

である．

(V)　凸性（Strictly convex）：上において，一般に

(26)　$$I^\lambda(\boldsymbol{p}\,;\ \boldsymbol{q})\geqq I^\lambda(\boldsymbol{p}'\,;\ \boldsymbol{q}')$$

すなわち

(26)*　$$I^\lambda(p_1,\ p_2,\cdots,p_m\,;\ q_1,\ q_2,\cdots,q_m)\geqq I^\lambda(p_1+p_2,\ p_3,\cdots,p_m\,;\ q_1+q_2,\ q_3,\cdots,q_m)$$

で，等号が成り立つのは (24) の場合に限る．

(VI)　（擬）加法性：分割 $\boldsymbol{E}=\{E_1,\cdots,E_m\}$ と $\boldsymbol{E}'=\{E_1',\cdots,E_r'\}$ とから，直積の分割 $\boldsymbol{E}\otimes\boldsymbol{E}'=\{E_k\cap E_j'\mid k=1,\cdots,m,\ j=1,\cdots,r\}$（但し，$E_k\cap E_j'\neq\emptyset$）を作る．$\boldsymbol{E}$ 上の分布 $\boldsymbol{p}=(p_1,\cdots,p_m)$, $\boldsymbol{q}=(q_1,\cdots,q_m)$ および $\boldsymbol{E}'$ 上の分布 $\boldsymbol{p}'=(p_1',\cdots,p_r')$, $\boldsymbol{q}'=(q_1',\cdots,q_r')$ とから $\boldsymbol{E}\otimes\boldsymbol{E}'$ 上の分布

$$\boldsymbol{p}\otimes\boldsymbol{p}'=\{p_k\cdot p_j'\mid k=1,\cdots,m,\ j=1,\cdots,r\},\quad \boldsymbol{q}\otimes\boldsymbol{q}'=\{q_k\cdot q_j'\mid k=1,\cdots,m,\ j=1,\cdots,r\}$$

が定まる．そのとき，$\lambda=0$ であれば加法性：

(27)　$$I^0(\boldsymbol{p}\otimes\boldsymbol{p}'\,;\ \boldsymbol{q}\otimes\boldsymbol{q}')=I^0(\boldsymbol{p}\,;\ \boldsymbol{q})+I^0(\boldsymbol{p}'\,;\ \boldsymbol{q}')$$

また $\lambda\neq 0$ であればこれを一般化した擬加法性：

(28) $$I^\lambda(\boldsymbol{p}\otimes\boldsymbol{p}';\ \boldsymbol{q}\otimes\boldsymbol{q}')=I^\lambda(\boldsymbol{p};\ \boldsymbol{q})+I^\lambda(\boldsymbol{p}';\ \boldsymbol{q}')+\lambda I^\lambda(\boldsymbol{p};\ \boldsymbol{q})\times I^\lambda(\boldsymbol{p}';\ \boldsymbol{q}')$$

が成り立つ.

証明 KL 情報量 I^0 についての上記性質は Kullback (1951) に述べてある. 一般に $I^\lambda(\boldsymbol{p};\ \boldsymbol{q})$ の定義 (5) を用いれば, いずれも直接に容易に確かめられる.

注意 以上の他に

(VII) 連続性：$\boldsymbol{p}$, $\boldsymbol{q}$ のベクトルとしてのユークリッド位相に関して, $\lim_{n\to\infty}\boldsymbol{p}_n=\boldsymbol{p}_0$, $\lim_{n\to\infty}\boldsymbol{q}_n=\boldsymbol{q}_0$ であれば

(29) $$\lim_{n\to\infty} I^\lambda(\boldsymbol{p}_n;\ \boldsymbol{q}_n)=I^\lambda(\boldsymbol{p}_0;\ \boldsymbol{q}_0)$$

も成り立つ.

なお, 工藤 (1953, pp. 106-108) において, 情報量の不変性, 凸性に対して "層化は情報量を増加する" という適切な表現が与えられている. このことは, 次の相対性 (定理 3) でさらに具体的に表わされる.

2.2 相対性と一意性

定理 2 の諸性質 (IV)-(VI) を含む強い内容を持つ相対性を考えよう.

分割 $\boldsymbol{E}=\{E_1,\cdots,E_m\}$ に対して, 各 E_k, $k=1,\cdots,m$ をさらに分割して

$$\boldsymbol{E}_k=\{E_{k1},\cdots,E_{kr_k}\},\quad E_k=E_{k1}\cup\cdots\cup E_{kr_k},\quad E_{ki}\cap E_{kj}=\emptyset,\qquad i\neq j$$

をとる ($1\leqq r_k<\infty$). これらすべての $\boldsymbol{E}_k$ を合わせて, 分割 $\boldsymbol{E}$ の細分

$$\boldsymbol{E}^*=\{E_{kj}\mid k=1,\cdots,m,\ j=1,\cdots,r_k\}$$

を定義する.

$\boldsymbol{E}^*$ 上の分布 $\boldsymbol{p}^*=(p_{kj})$, $\boldsymbol{q}^*=(q_{kj})$ が与えられれば, これから $\boldsymbol{E}$ 上の分布

$$\boldsymbol{p}=(p_1,\cdots,p_m),\quad p_k=p_{k1}+\cdots+p_{kr_k},\qquad k=1,\cdots,m$$
$$\boldsymbol{q}=(q_1,\cdots,q_m),\quad q_k=q_{k1}+\cdots+q_{kr_k},\qquad k=1,\cdots,m$$

および $\boldsymbol{E}_k$, $k=1,\cdots,m$ 上の相対分布

$$\boldsymbol{p}^{(k)}=\left(\frac{p_{k1}}{p_k},\cdots,\frac{p_{kr_k}}{p_k}\right)\qquad (\text{但し}, p_k>0 \text{ とする})$$
$$\boldsymbol{q}^{(k)}=\left(\frac{q_{k1}}{q_k},\cdots,\frac{q_{kr_k}}{q_k}\right)\qquad (\text{但し}, q_k>0 \text{ とする})$$

が定まる.

定理 3. 情報量 $I^\lambda(\boldsymbol{p};\ \boldsymbol{q})$, $-1/2\leqq\lambda<\infty$ は各 λ に対して

(VIII) 相対性：上の諸記号を用いて, $\lambda=0$ の場合 KL 情報量について良く知られている性質：

(30) $$I^0(\boldsymbol{p}^*;\ \boldsymbol{q}^*)=I^0(\boldsymbol{p};\ \boldsymbol{q})+\sum_{k=1}^{m} p_k I^0(\boldsymbol{p}^{(k)};\ \boldsymbol{q}^{(k)})$$

$\lambda\neq0$ のときはこれを一般化した性質：

(31)
$$I^{\lambda}(\boldsymbol{p}^*;\ \boldsymbol{q}^*)=I^{\lambda}(\boldsymbol{p};\ \boldsymbol{q})+\sum_{k=1}^{m} p_k^{1+\lambda}q_k^{-\lambda}I^{\lambda}(\boldsymbol{p}^{(k)};\ \boldsymbol{q}^{(k)})$$

が成り立つ.

証明 $I^{\lambda}(\boldsymbol{p};\ \boldsymbol{q})$ の定義(5)により容易に直接に計算される. 例えば(31)の右辺を計算すれば

$$\begin{aligned}
&I^{\lambda}(\boldsymbol{p};\ \boldsymbol{q})+\sum_{k=1}^{m} p_k^{1+\lambda}q_k^{-\lambda}I^{\lambda}(\boldsymbol{p}^{(k)};\ \boldsymbol{q}^{(k)})\\
&=\frac{1}{\lambda}\left[\left(\sum_{k=1}^{m} p_k^{1+\lambda}q_k^{-\lambda}-1\right)+\sum_{k=1}^{m} p_k^{1+\lambda}q_k^{-\lambda}\left(\sum_{j=1}^{r_k}\left(\frac{p_{kj}}{p_k}\right)^{1+\lambda}\left(\frac{q_{kj}}{q_k}\right)^{-\lambda}-1\right)\right]\\
&=\frac{1}{\lambda}\left[\sum_{k=1}^{m}\sum_{j=1}^{r_k} p_{kj}^{1+\lambda}q_{kj}^{-\lambda}-1\right]=I^{\lambda}(\boldsymbol{p}^*;\ \boldsymbol{q}^*)
\end{aligned}$$

である (Kullback (1951)では $\sum p_k I^0(\boldsymbol{p}^{(k)};\ \boldsymbol{q}^{(k)})$ を条件付情報量と呼んでいる).

注意 (VIII) (相対性) ⇒ (VI) ((擬)加法性):

証明 $\boldsymbol{E}^*=\boldsymbol{E}\otimes\boldsymbol{E}'$ 上の $\boldsymbol{p}^*=\boldsymbol{p}\otimes\boldsymbol{p}'$, $\boldsymbol{q}^*=\boldsymbol{q}\otimes\boldsymbol{q}'$ に対しては $\boldsymbol{p}^{(k)}=\boldsymbol{p}'$, $\boldsymbol{q}^{(k)}=\boldsymbol{q}'$, $k=1,\cdots,m$ となるから $\sum_{k=1}^{m} p_k=1$ を用いれば(30)から(27)となる. また $\sum_{k=1}^{m} p_k^{1+\lambda}q_k^{-1}=1+\lambda I^{\lambda}(\boldsymbol{p};\ \boldsymbol{q})$ を用いれば, (31)から(28)となる.

(III) (非負性), (VIII) (相対性) ⇒ (IV) (不変性), (V) (凸性):

証明 記号をかえて $\boldsymbol{E}=\{E_1\cup E_2,\ E_3,\cdots,E_m\}$, $\boldsymbol{E}^*=\{E_1,\ E_2,\cdots,E_m\}$ とすれば

$$\boldsymbol{E}^{(1)}=\{E_1,\ E_2\},\qquad \boldsymbol{E}^{(2)}=\{E_3\},\cdots,\boldsymbol{E}^{(m-1)}=\{\boldsymbol{E}_m\}$$

である. したがって, $\boldsymbol{E}^*$ 上の分布 $\boldsymbol{p}^*=\{p_1,\ p_2,\cdots,p_m\}$, $\boldsymbol{q}^*=\{q_1,\ q_2,\cdots,q_m\}$ に対して $\boldsymbol{p}=\{p_1+p_2,\ p_3,\cdots,p_m\}$, $\boldsymbol{q}=\{q_1+q_2,\ q_3,\cdots,q_m\}$, $\boldsymbol{p}^{(1)}=\left\{\frac{p_1}{p_1+p_2},\ \frac{p_2}{p_1+p_2}\right\}$, $\boldsymbol{q}^{(1)}=\left\{\frac{q_1}{q_1+q_2},\ \frac{q_2}{q_1+q_2}\right\}$, $\boldsymbol{p}^{(2)}=\cdots=\boldsymbol{p}^{(m-1)}=\boldsymbol{q}^{(2)}=\cdots=\boldsymbol{q}^{(m-1)}=\{1\}$ となる. 故に(30)と(31)とから

$$\begin{aligned}
&I^0(\boldsymbol{p}^*;\ \boldsymbol{q}^*)=I^0(\boldsymbol{p};\ \boldsymbol{q})+(p_1+p_2)I^0(\boldsymbol{p}^{(1)};\ \boldsymbol{q}^{(1)})\\
&I^{\lambda}(\boldsymbol{p}^*;\ \boldsymbol{q}^*)=I^{\lambda}(\boldsymbol{p};\ \boldsymbol{q})+(p_1+p_2)^{1+\lambda}(q_1+q_2)^{-\lambda}I^{\lambda}(\boldsymbol{p}^{(1)};\ \boldsymbol{q}^{(1)})
\end{aligned}$$

となる. これより(V)が導かれる. 不変性(IV)が成り立つのは $I^{\lambda}(\boldsymbol{p}^{(1)};\ \boldsymbol{q}^{(1)})=0$, すなわち(III) より $\boldsymbol{p}^{(1)}=\boldsymbol{q}^{(1)}$, すなわち(24)が成り立つ場合に限る.

定理 4. λ, $-1/2\leqq\lambda<\infty$ を定めておく. もしも $\boldsymbol{p}, \boldsymbol{q}$ の関数 $I^*(p_1,\cdots,p_m;\ q_1,\cdots,q_m)$ に対して (I) 簡約性, (II) 対称性, (III) 非負性, (VII) 連続性, (VIII) 相対性が成り立てば, 或る正の定数 c によって

$$I^*(\boldsymbol{p};\ \boldsymbol{q})=cI^{\lambda}(\boldsymbol{p};\ \boldsymbol{q})$$

が成り立つ.

証明 (i) $\lambda\neq0$ の場合: $\boldsymbol{E}^*=\boldsymbol{E}\otimes\boldsymbol{E}'$, $m=r$ とし $\boldsymbol{p}^*=\boldsymbol{p}\otimes\boldsymbol{p}'$, $\boldsymbol{q}^*=\boldsymbol{q}\otimes\boldsymbol{q}'$ をとれば(VIII) ⇒ (VI) と同様に

$$I^*(\boldsymbol{p}\otimes\boldsymbol{p}'\,;\ \boldsymbol{q}\otimes\boldsymbol{q}')=I^*(\boldsymbol{p}\,;\ \boldsymbol{q})+\left(\sum_{k=1}^{m} p_k^{1+\lambda}q_k^{-\lambda}\right)I^*(\boldsymbol{p}'\,;\ \boldsymbol{q}')$$

また

$$I^*(\boldsymbol{p}'\otimes\boldsymbol{p}\,;\ \boldsymbol{q}'\otimes\boldsymbol{q})=I^*(\boldsymbol{p}'\,;\ \boldsymbol{q}')+\left(\sum_{k=1}^{m} p_k'^{1+\lambda}q_k'^{-\lambda}\right)I^*(\boldsymbol{p}\,;\ \boldsymbol{q})$$

が成り立つ．(II) 対称性によって上の二つの式の左辺は等しいから，右辺を等しいとおけば

$$\frac{I^*(\boldsymbol{p}\,;\ \boldsymbol{q})}{\sum_k p_k^{1+\lambda}q_k^{-\lambda}-1}=\frac{I^*(\boldsymbol{p}'\,;\ \boldsymbol{q}')}{\sum_k p_k'^{1+\lambda}q_k'^{-\lambda}-1}$$

となる．これから（非負性を用いて）

$$I^*(\boldsymbol{p}\,;\ \boldsymbol{q})=c\left(\sum_{k=1}^{m} p_k^{1+\lambda}q_k^{-\lambda}-1\right)=cI^\lambda(\boldsymbol{p}\,;\ \boldsymbol{q}),\qquad c>0$$

となる．

(ii) $\lambda=0$ の場合：分割 $\boldsymbol{E}^*=\boldsymbol{E}\otimes\boldsymbol{E}'$，$m=r$ とし，$\boldsymbol{E}$，$\boldsymbol{E}'$ 上の分布 $\boldsymbol{p}$，$\boldsymbol{q}$ に対して $\boldsymbol{E}^*$ 上の分布 $\varDelta\boldsymbol{p}$ および $\boldsymbol{p}\otimes\boldsymbol{q}$ をとる．但し，$p_1>0,\cdots,p_m>0$ とし

$$\varDelta\boldsymbol{p}=\begin{bmatrix} p_1 & 0 & \cdots & 0\\ 0 & p_2 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & p_m\end{bmatrix},\quad \boldsymbol{p}\otimes\boldsymbol{q}=\begin{bmatrix} p_1q_1 & p_2q_1 & \cdots & p_mq_1\\ p_1q_2 & p_2q_2 & \cdots & p_mq_2\\ \vdots & \vdots & & \vdots\\ p_1q_m & p_2q_m & \cdots & p_mq_m\end{bmatrix}$$

とおく．(30) によって

$$I^*(\varDelta\boldsymbol{p}\,;\ \boldsymbol{p}\otimes\boldsymbol{q})=I^*(\boldsymbol{p}\,;\ \boldsymbol{q})+\sum_{k=1}^{m} p_kI^*((\varDelta\boldsymbol{p})^{(k)}\,;\ (\boldsymbol{p}\otimes\boldsymbol{q})^{(k)})$$

である．ここで

$$(\varDelta\boldsymbol{p})^{(k)}=(0,\cdots,0,\ 1,\ 0,\cdots,0),\qquad (\boldsymbol{p}\otimes\boldsymbol{q})^{(k)}=\boldsymbol{p},\qquad k=1,\cdots,m$$

であり，かつ (III)，(VIII) ⇒ (IV)（不変性）によって

$$I^*((\varDelta\boldsymbol{p})^{(k)}\,;\ (\boldsymbol{p}\otimes\boldsymbol{q})^{(k)})=I^*(1,\ 0\,;\ p_k,\ 1-p_k)$$

が成り立つ．よって

$$I^*(\varDelta\boldsymbol{p}\,;\ \boldsymbol{p}\otimes\boldsymbol{q})=I^*(\boldsymbol{p}\,;\ \boldsymbol{q})+\sum_{k=1}^{m} p_kI^*(1,\ 0\,;\ p_k,\ 1-p_k)$$

となる．次に $\boldsymbol{p}$ と $\boldsymbol{q}$ を入れかえて

$$I^*(\varDelta\boldsymbol{p}\,;\ \boldsymbol{q}\otimes\boldsymbol{p})=I^*(\boldsymbol{p}\,;\ \boldsymbol{p})+\sum_{k=1}^{m} p_kI^*(1,\ 0\,;\ q_k,\ 1-q_k)$$

となる．(II) 対称性によって $I^*(\varDelta\boldsymbol{p}\,;\ \boldsymbol{p}\otimes\boldsymbol{q})=I^*(\varDelta\boldsymbol{p}\,;\ \boldsymbol{q}\otimes\boldsymbol{p})$ であり，また $I^*(\boldsymbol{p}\,;\ \boldsymbol{p})=0$ であるから，上の二つの式より

$$I^*(\boldsymbol{p}\,;\ \boldsymbol{q})=\sum_{k=1}^{m} p_k\{I^*(1,\ 0\,;\ q_k,\ 1-q_k)-I^*(1,\ 0\,;\ p_k,\ 1-p_k)\}$$

と表わされる．

次に分割 $\boldsymbol{E}=\{E_1,\ E_2\}$，$\boldsymbol{E}'=\{E_1',\ E_2'\}$ 上の分布 $\mathbf{1}=(1,\ 0)$ および $\boldsymbol{p}=(p,\ 1-p)$，$0<p<1$，$\boldsymbol{q}=(q,\ 1-q)$，$0<q<1$ に対して (V)（加法性）をあてはめれば

$$I^*(\mathbf{1}\otimes\mathbf{1};\ \boldsymbol{p}\otimes\boldsymbol{q})=I^*(\mathbf{1};\ \boldsymbol{p})+I^*(\mathbf{1};\ \boldsymbol{q})$$

である．ここで

$$\mathbf{1}\otimes\mathbf{1}=\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix},\qquad \boldsymbol{p}\otimes\boldsymbol{q}=\begin{pmatrix}pq & p(1-q)\\ (1-p)q & (1-p)(1-q)\end{pmatrix}$$

であるから（IV）不変性によって

$$I^*(\mathbf{1}\otimes\mathbf{1};\ \boldsymbol{p}\otimes\boldsymbol{q})=I^*(\mathbf{1};\ pq,\ 1-pq)$$

となる．故に $f(p)=I^*(\mathbf{1};\ p,\ 1-p)$ とおくとき

$$f(pq)=f(p)+f(q)$$

である．ここで $f(1)=0$, $f(p)>0$, $0<p<1$ および f の（VII）連続性によって，或る定数 $c>0$ によって

$$f(p)=-c\log p,\qquad 0<p\leqq 1$$

と表わされる．よって

$$I^*(\boldsymbol{p};\ \boldsymbol{q})=c\sum_{k=1}^{m}p_k(\log p_k-\log q_k)=cI^0(\boldsymbol{p};\ \boldsymbol{q})$$

が導かれた．

注意　Shannon のエントロピーについて，対応する定理は，A.I. Khinchin によって証明されている（*Uspehi. Mat. Nauk*, 1953）．

3．情報量の公理系

3.1　公理系

2.1 において KL 情報量，Pearson 情報量，角谷情報量，$I^\lambda(\boldsymbol{p};\ \boldsymbol{q})$, $-1/2\leqq\lambda<\infty$ などの例について見た．これらが満足するいくつかの性質をとり出して，一般の情報量の定義を与えよう．

定義 4.　任意の分割：$\boldsymbol{E}=\{E_1,\cdots,E_m\}$ 上の任意の二つの分布

$$\boldsymbol{p}=(p_1,\cdots,p_m),\quad p_k\geqq 0,\quad \sum_k p_k=1$$

$$\boldsymbol{q}=(q_1,\cdots,q_m),\quad q_k\geqq 0,\quad \sum_k q_k=1$$

に対して定義される（実数値）関数

(1)　$$I(\boldsymbol{p};\ \boldsymbol{q})=I(p_1,\cdots,p_m;\ q_1,\cdots,q_m),\qquad m=1,\ 2,\cdots$$

が，定理 2 に述べた（I）簡約性，（II）対称性，（III）非負性，（IV）不変性，（V）凸性を満足するとき，I を情報量（Information）と呼ぶ．

但し，I^λ, $\lambda\geqq 0$ をも含めるためには，$q_k=0$, $p_k>0$ の場合には $I(\boldsymbol{p};\ \boldsymbol{q})=\infty$ となることも許しておく．

例 1.　2.1 で定義した I_{KL}, I_P, I_K, I^λ はすべて情報量である．

例 2.

$$d(\boldsymbol{p}\,;\ \boldsymbol{q})=\sum_{k=1}^{m}|p_k-q_k|$$

は情報量としてのほとんどすべての性質を持つが，(V) の凸性に関しては (26)* の等号が成り立っても必ずしも (24) とならないから，この $d(\boldsymbol{p}\,;\ \boldsymbol{q})$ は情報量ではない．また

$$D(\boldsymbol{p},\ \boldsymbol{q})=\left(\sum_{k=1}^{m}(p_k-q_k)^2\right)^{1/2}$$

に対しては不変性が成り立たないから $D(\boldsymbol{p},\ \boldsymbol{q})$ も情報量ではない．

補題 1. $f(x_1,\cdots,x_r)$ が $x_1\geqq 0,\cdots,x_r\geqq 0$ で定義された実関数で

(i) $f(x_1,\cdots,x_r)\geqq 0$ である．かつ等号が成り立つのは $x_1=\cdots=x_r=0$ の場合に限る．

(ii) $x_1\leqq x_1',\cdots,x_r\leqq x_r'$ であれば

$$f(x_1,\cdots,x_r)\leqq f(x_1',\cdots,x_r')$$

である．かつここで等号が成り立つのは $x_1=x_1',\cdots,x_r=x_r'$ の場合に限るという二つの性質を持つものとする．

そのとき，任意の r 個の情報量 $I_1,\cdots,I_r$ に対して

$$I(\boldsymbol{p}\,;\ \boldsymbol{q})=f(I_1(\boldsymbol{p}\,;\ \boldsymbol{q}),\cdots,I_r(\boldsymbol{p}\,;\ \boldsymbol{q})) \tag{2}$$

もまた情報量である．

証明 $I(\boldsymbol{p}\,;\ \boldsymbol{q})$ が情報量としての性質：簡約性，対称性，非負性，不変性，凸性を満足することを個々に見ればよい．

例 3. $f(x_1,\ x_2)=ax_1+bx_2,\ a>0,\ b>0,\ f(x_1,\ x_2)=ax_1^2+bx_2^2,\ a>0,\ b>0$ など．

例 4. $\lambda>0$ に対して

$$f(x)=\frac{1}{\lambda}\log(1+\lambda x)$$

とおく．また，$0<\mu\leqq 1/2$ に対して

$$f(x)=\frac{-1}{\mu}\log(1-\mu x),\qquad 0\leqq x\leqq\frac{1}{\mu}$$

とおく．これらは補題の条件 (i)，(ii) を満足する．したがって

$$\tilde{I}^{\lambda}(\boldsymbol{p}\,;\ \boldsymbol{q})=\frac{1}{\lambda}\log(1+\lambda I^{\lambda}(\boldsymbol{p}\,;\ \boldsymbol{q})),\qquad \lambda>0 \tag{3}$$

$$\tilde{I}^{-\mu}(\boldsymbol{p}\,;\ \boldsymbol{q})=\frac{-1}{\mu}\log(1-\mu I^{-\mu}(\boldsymbol{p}\,;\ \boldsymbol{q})),\qquad 0<\mu\leqq\frac{1}{2} \tag{4}$$

も情報量である．さらに $\tilde{I}^{\lambda}$, $\tilde{I}^{-\mu}$ は加法性を満足する：

$$\tilde{I}^{\lambda}(\boldsymbol{p}\otimes\boldsymbol{p}'\,;\ \boldsymbol{q}\otimes\boldsymbol{q}')=\tilde{I}^{\lambda}(\boldsymbol{p}\,;\ \boldsymbol{q})+\tilde{I}^{\lambda}(\boldsymbol{p}'\,;\ \boldsymbol{q}') \tag{5}$$

$$\tilde{I}^{-\mu}(\boldsymbol{p}\otimes\boldsymbol{p}'\,;\ \boldsymbol{q}\otimes\boldsymbol{q}')=\tilde{I}^{-\mu}(\boldsymbol{p}\,;\ \boldsymbol{q})+\tilde{I}^{-\mu}(\boldsymbol{p}'\,;\ \boldsymbol{q}'). \tag{6}$$

証明　定理2より

$$
\begin{aligned}
\lambda\tilde{I}^{\lambda}(\boldsymbol{p}\otimes\boldsymbol{p}';\ \boldsymbol{q}\otimes\boldsymbol{q}')&=\log(1+\lambda I^{\lambda}(\boldsymbol{p}\otimes\boldsymbol{p}';\ \boldsymbol{q}\otimes\boldsymbol{q}'))\\
&=\log(1+\lambda(I^{\lambda}(\boldsymbol{p};\ \boldsymbol{q})+I^{\lambda}(\boldsymbol{p}';\ \boldsymbol{q}')+\lambda(I^{\lambda}(\boldsymbol{p};\ \boldsymbol{q})\cdot I^{\lambda}(\boldsymbol{p}';\ \boldsymbol{q}'))))\\
&=\log(1+\lambda I^{\lambda}(\boldsymbol{p};\ \boldsymbol{q}))(1+\lambda I^{\lambda}(\boldsymbol{p}';\ \boldsymbol{q}'))\\
&=\lambda\tilde{I}^{\lambda}(\boldsymbol{p};\ \boldsymbol{q})+\lambda\tilde{I}^{\lambda}(\boldsymbol{p}';\ \boldsymbol{q}')
\end{aligned}
$$

また $\tilde{I}^{-\mu}$ についても同様である（$\tilde{I}^{-\mu}$ の定義と加法性は工藤（1953）による）.

注意　(3), (4) より

(7)
$$\lim_{\lambda\to 0}\tilde{I}^{\lambda}(\boldsymbol{p};\ \boldsymbol{q})=\lim_{\mu\to 0}\tilde{I}^{-\mu}(\boldsymbol{p};\ \boldsymbol{q})=I^{0}(\boldsymbol{p};\ \boldsymbol{q})$$

である.

3.2　基本情報量

いろいろな情報量のなかで，取り扱い易いものとして，次の基本情報量が考えられる.

定義 5.　情報量 $I(\boldsymbol{p};\ \boldsymbol{q})$ が基本情報量（Fundamental information）であるとは，$0\leqq x\leqq 1$, $0\leqq y\leqq 1$ に対して定義される或る実関数 $L(x,\ y)$ によって

(8)
$$I(\boldsymbol{p};\ \boldsymbol{q})=L(p_1,\ q_1)+\cdots+L(p_m,\ q_m)$$

（但し，$\boldsymbol{p}=(p_1,\cdots,p_m)$, $\boldsymbol{q}=(q_1,\cdots,q_m)$ とする）と表わされることをいう.

例えば第2章で述べた $I_{KL}, I_P, I_K, I^{\lambda}$ はすべて基本情報量である．しかし第3章，例4の (3), (4) 式の $\tilde{I}^{\lambda}$, $\tilde{I}^{-\mu}$ は基本情報量ではない.

定理 5.　(8) で表わされる $I(\boldsymbol{p};\ \boldsymbol{q})$ が情報量であるために $L(p,\ q)$ が満足すべき必要十分条件は

(I)*
$$L(0,\ 0)=0,\qquad L(1,\ 1)=0$$

(II)*
$$\frac{p_1}{q_1}=\frac{p_2}{q_2}=\frac{p_1+p_2}{q_1+q_2}$$

（但し，$0\leqq p_1+p_2\leqq 1$,　$0\leqq q_1+q_2\leqq 1$）であれば

(9)
$$L(p_1,\ q_1)+L(p_2,\ q_2)=L(p_1+p_2,\ q_1+q_2).$$

(III)*　$0\leqq p_1+p_2\leqq 1$,　$0\leqq q_1+q_2\leqq 1$ であれば一般に

(9)*
$$L(p_1,\ q_1)+L(p_2,\ q_2)\geqq L(p_1+p_2,\ q_1+q_2)$$

で，等号が成り立つのは (II)* の場合に限る.

証明　(i)　或る情報量 $I(\boldsymbol{p};\ \boldsymbol{q})$ が (8) のように表わされたとする．(I)* $L(0,\ 0)=0$ は簡約性より，$L(1,\ 1)=0$ は非負性 $I(\{1\};\ \{1\})=0$ よりわかる．(II)* は不変性より，(III)* は凸性より導かれる.

(ii)　逆に $L(p,\ q)$ が (I)*, (II)*, (III)* を満足すれば (8) の $I(\boldsymbol{p};\ \boldsymbol{q})$ が情報量となることは直ちにわかる．例えば非負性は (III)* をくりかえし用いて

$$L(p_1,\ q_1)+\cdots+L(p_m,\ q_m)\geqq L(p_1+\cdots+p_m,\ q_1+\cdots+q_m)=L(1,\ 1)=0.$$

注意 $I(\boldsymbol{p}\,;\,\boldsymbol{q})$ の連続性を仮定すれば，$L(p,q)$ も連続であり，かつ任意の $0<p<1$ に対しても

$$L(p,\ p)=0$$

となる．何となれば $p=r/s$（$r,\ s$ は正整数）に対しては

$$L\left(\frac{1}{s}\right)+\cdots+L\left(\frac{1}{s}\right)=L(1,\ 1)=0$$

より

$$L\left(\frac{1}{s}\right)=0,\qquad L\left(\frac{r}{s},\ \frac{r}{s}\right)=L\left(\frac{1}{s}\right)+\cdots+L\left(\frac{1}{s}\right)=0$$

である．よって，一般の p に対しても $p=\lim_{n\to\infty}\frac{r_n}{s_n}$ と表わせば $L(p,\ p)=0$ となる．

補題 2. 定理5の (II)* が成り立つための必要十分条件は ($L(p,\ q)$ の連続性を仮定して)，$x>0$ で定義される或る連続関数 $K(x)$ によって

$$L(p,\ q)=pK\left(\frac{q}{p}\right),\qquad p>0,\ q>0 \tag{10}$$

と表わされることである．

証明 $p>0,\ q>0$ に対して $q=pu$ とおいて

$$L(p,\ q)=L(p,\ pu)=F(p,\ u)$$

とおく．(II)* で

$$u=\frac{q_1}{p_1}=\frac{q_2}{p_2}=\frac{q_1+q_2}{p_1+p_2}$$

であるから，(9) は

$$F(p_1,\ u)+F(p_2,\ u)=F(p_1+p_2,\ u)$$

と表わされる．$L(p,q)$ の連続性から，$F(p,\ u)$ も $p,\ u$ について連続である．したがって，u を固定すれば，$F(p,\ u)$ は p の一次関数となり

$$F(p,\ u)=K(u)p$$

の形に表わされる．すなわち

$$L(p,\ q)=F\left(p,\ \frac{q}{p}\right)=K\left(\frac{q}{p}\right)\cdot p$$

と (10) の形になる．逆に (10) の形の $L(p,\ q)$ に対して (II)* が成り立つ．但し，$p=0$ または $q=0$ のときは，連続性を用いる．

補題 3. 定理5で $L(p,q)$ は連続とし，かつ (I)*, (II)* を仮定する．そのとき，(III)* が成り立つための必要十分条件は，$K(x)$ が $x>0$ で定義された凸関数であることである．すなわち $x_1>0,\quad x_2>0,\quad 1>\alpha>0$ に対して

$$\alpha K(x_1)+(1-\alpha)K(x_2)\geqq K(\alpha x_1+(1-\alpha)x_2) \tag{11}$$

但し，等号が成り立つのは $x_1=x_2$ の場合に限る．

証明　(10) を (III)* に代入すれば

(12)
$$p_1K\left(\frac{q_1}{p_1}\right)+p_2K\left(\frac{q_2}{p_2}\right)\geqq(p_1+p_2)K\left(\frac{q_1+q_2}{p_1+p_2}\right)$$

となる．よって

(13)
$$x_1=\frac{q_1}{p_1},\qquad x_2=\frac{q_2}{p_2},\qquad \alpha=\frac{p_1}{p_1+p_2},\qquad 1-\alpha=\frac{p_2}{p_1+p_2}$$

とおくと，$x_1>0$，$x_2>0$，$1>\alpha>0$ である．よって (12) は (11) となる．逆に，任意に $x_1>0$，$x_2>0$，$1>\alpha>0$ を与えるとき，$0<p_1<1$，$0<p_2<0$，$0<q_1<1$，$0<q_2<1$ をとって (13) が成り立つようにとることができる．よって (11) を書き直せば (12) となる．

$L(1,1)=0$ を (10) に代入すれば

(14)
$$K(1)=0$$

となり，逆に (14) ならば $L(1,1)=0$ である．

補題 4.

(15)
$$I(\boldsymbol{p}\,;\,\boldsymbol{q})=L_1(p_1,q_1)+\cdots+L_1(p_m,q_m)=L_2(p_1,q_1)+\cdots+L_2(p_m,q_m)$$

と表わされ，かつ

(16)
$$L_1(p,q)=pK_1\left(\frac{q}{p}\right),\qquad L_2(p,q)=pK_2\left(\frac{q}{p}\right)$$

であるとき，或る定数 a によって

(17)
$$K_1(x)-K_2(x)=a(x-1)$$

と表わされる．逆に (17) であれば (15) が成り立つ．

証明　(5)，(6) とすれば

$$\sum_k L_1(p_k,q_k)=\sum_k p_kK_1\left(\frac{q_k}{p_k}\right)=\sum_k p_kK_2\left(\frac{q_k}{p_k}\right)+a\sum_k p_k\left(\frac{q_k}{p_k}-1\right)=\sum_k L_2(p_k,q_k)$$

が成り立つ．また (15)，(16) のとき

$$K_1(x)-K_2(x)=G(x)$$

とおく．$m=2$，$p_1+p_2=1$，$q_1+q_2=1$ に対して (15) は

$$p_1K_1\left(\frac{q_1}{p_1}\right)+p_2K_1\left(\frac{q_2}{p_2}\right)=p_1K_2\left(\frac{q_1}{p_1}\right)+p_2K_2\left(\frac{q_2}{p_2}\right)$$

したがって，$x_1=q_1/p_1>0$，$x_2=q_2/p_2>0$，$p_1x_1+p_2x_2=1$ に対して

(18)
$$p_1G(x_1)+p_2G(x_2)=0$$

である．よって補題 4 は，次の補題 5 より導かれる．

補題 5.　任意の $x_1>0$，$x_2>0$，$p_1+p_2=1$，$p_1x_1+p_2x_2=1$ (但し，$0<p_1$，$0<p_2$) に対して

(18) が成り立つならば，或る定数 a によって

$$G(x)=a(x-1) \tag{19}$$

と表わされる．

証明　$p_1+p_2=1$, $p_1x_1+p_2x_2=1$ を解けば $p_1=\frac{x_2-1}{x_2-x_1}$, $p_2=\frac{1-x_1}{x_2-x_1}$ となる．よって (20) に代入すれば

$$\frac{G(x_1)}{x_1-1}=\frac{G(x_2)}{x_2-1}$$

となり，(19) が成り立つ．

最後に凸関数 $K(x)$, $K(1)=0$ の $x=1$ における（一つの）支持関数を $y=a(x-1)$ とすれば，$K(x)$ の代りに，次の $K_1(x)$ をとれば

$$K_1(x)=K(x)-a(x-1)\geqq 0$$

となる．よって $L(p,\ q)$ を (10) の形に表わすとき

$$L(p,\ q)=pK_1\left(\frac{q}{p}\right), \qquad K_1(x)\geqq 0 \tag{10)*}$$

とすることができる．以上をまとめて

定理 6.　連続基本情報量 $I(\boldsymbol{p}\ ;\ \boldsymbol{q})$ は $x>0$ で定義された或る凸関数 $K(x)$ (但し，$K(1)=0$) によって

$$I(\boldsymbol{p}\ ;\ \boldsymbol{q})=\sum_{k=1}^{m} p_k K\left(\frac{q_k}{p_k}\right) \tag{20}$$

と表わされる (凸関数は必ず連続である)．また $p_k K\left(\frac{0}{p_k}\right)=p_k\cdot\lim_{x\to 0}K(x)$; $0\times K\left(\frac{q_k}{0}\right)=q_k\cdot\lim_{x\to\infty}\frac{1}{x}K(x)$ $(q_k\neq 0)$; $0\times K\left(\frac{0}{0}\right)=0$ とおく．ここに $K(x)$ は I に対して一意に定まらないで，$K(x)$ の代りに $K(x)-a(x-1)$ (a: 定数) をとることができる．したがって，$y=a(x-1)$ を $K(x)$ の点 $x=1$ における（一つの）支持関数にすれば

$$K(x)\geqq 0 \tag{21}$$

に選ぶことができる (等号は $x=1$ に限る)．逆にこのような $K(x)$ に対して (20) は連続基本情報量となる．

例 5.　(i)　$I^0(\boldsymbol{p}\ ;\ \boldsymbol{q})=\sum_{k=1}^{m} p_k \log\frac{p_k}{q_k}$ に対して，凸関数

$$K^0(x)=-\log x+(x-1)\geqq 0 \tag{22}$$

をとれば

$$I^0(\boldsymbol{p}\ ;\ \boldsymbol{q})=\sum_{k=1}^{m} p_k K^0\left(\frac{q_k}{p_k}\right)$$

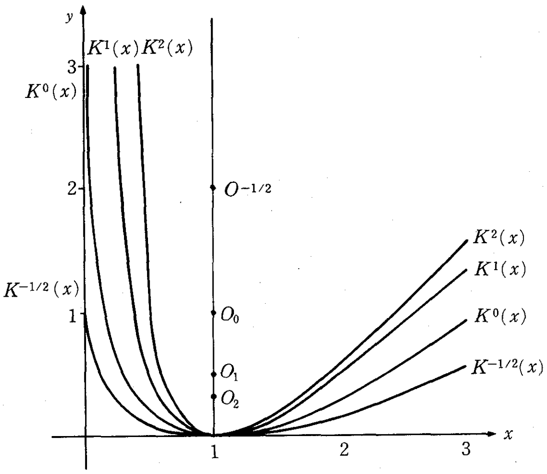

図 1.　$K^\lambda(x)$ のグラフ

$K^\lambda(x)=\frac{1}{\lambda}(\frac{1}{x^\lambda}-1)+(x-1),\ \lambda\neq 0$

$K^0(x)=-\log x+(x-1)$

O_λ : $y=K^\lambda(x)$ の点 (1, 0) における曲率の中心，$O_\lambda=\left(1, \frac{1}{K^{\lambda\prime\prime}(1)}\right)$

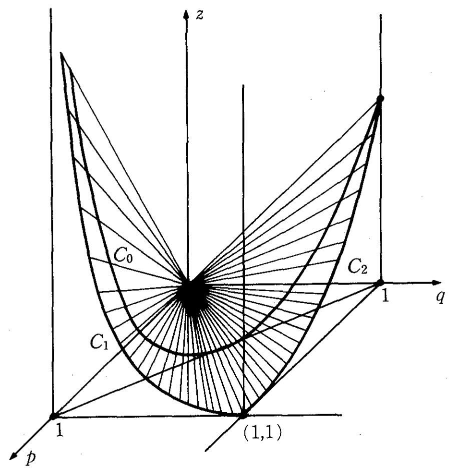

図 2.　$z=L^0(p,q)$ のグラフ

$z=L^0(p,q)=pK^0\left(\frac{q}{p}\right)$

$K^0(u)=-\log u+(u-1)$

$0\leqq p\leqq 1,\ 0\leqq q\leqq 1,\ 0<u<\infty$

C_0 : $\begin{cases} p+q=1 \\ z=\frac{1}{u+1}K^0(u), \\ u=q/p \end{cases}$　C_1 : $\begin{cases} p=1 \\ z=K^0(q) \end{cases}$

C_2 : $\begin{cases} q=1 \\ z=pK^0\left(\frac{1}{p}\right) \end{cases}$

と表わされる．

(ii)　$I^\lambda(\boldsymbol{p}\,;\,\boldsymbol{q})=\frac{1}{\lambda}\left(\sum_{k=1}^{m} p_k^{1+\lambda}q_k^{-\lambda}-1\right)$,　　$\lambda\neq 0$ に対して，凸関数

(23)
$$K^\lambda(x)=\frac{1}{\lambda}(x^{-\lambda}-1)+(x-1)\geqq 0$$

をとれば

$$I^\lambda(\boldsymbol{p}\,;\,\boldsymbol{q})=\sum_{k=1}^{m} p_k K^\lambda\left(\frac{q_k}{p_k}\right)$$

と表わされる（図 1，2 参照）．

定義 6.　情報量 $I(\boldsymbol{p}\,;\,\boldsymbol{q})$ に対して

(24)
$$I^*(\boldsymbol{p}\,;\,\boldsymbol{q})=I(\boldsymbol{q}\,;\,\boldsymbol{p})$$

もまた情報量となる．I^* を I の双対情報量（Dual）と呼ぶ．

特に I が基本情報量で

(25)
$$I(\boldsymbol{p}\,;\,\boldsymbol{q})=\sum_{k=1}^{m}L(p_k,\,q_k),\qquad L(p,\,q)=pK\left(\frac{q}{p}\right)$$
$$K(1)=0,\qquad K(x)>0,\qquad x\neq 1$$

と表わせば，その双対情報量 I^* も基本情報量で

(26)
$$I^*(\boldsymbol{p}\,;\,\boldsymbol{q})=\sum_{k=1}^{m}L^*(p_k,\,q_k)$$
$$L^*(p,\,q)=L(q,\,p)=qK\left(\frac{p}{q}\right)=pK^*\left(\frac{q}{p}\right)$$

但し，凸関数 $K^*(x)$ は

(27)
$$K^*(x)=xK\left(\frac{1}{x}\right),\qquad\qquad x>0$$

と表わされ，$K^*(1)=0$ および $K^*(x)>0,\ \ x\neq 1$ である．

例 6. (i) $I^0(\boldsymbol{p}\,;\,\boldsymbol{q})$ の双対情報量 I^{0*} は

(28)
$$I^{0*}(\boldsymbol{p}\,;\,\boldsymbol{q})=\sum_{k=1}^{m}q_k\log\frac{p_k}{q_k}=\sum_{k=1}^{m}p_kK^{0*}\left(\frac{q_k}{p_k}\right)$$
$$K^{0*}(x)=x\log x-x+1\geqq 0$$

となる．

(ii) $I^\lambda(\boldsymbol{p}\,;\,\boldsymbol{q})$ の双対情報量 $I^{\lambda*}$ は

(29)
$$I^{\lambda*}(\boldsymbol{p}\,;\,\boldsymbol{q})=\frac{1}{\lambda}\left(\sum_{k=1}^{m}q_k^{1+\lambda}p_k^{-\lambda}-1\right)=\sum_{k=1}^{m}p_kK^{\lambda*}\left(\frac{q_k}{p_k}\right),\qquad -\frac{1}{2}<\lambda<\infty,\ \lambda\neq 0$$
$$K^{\lambda*}(x)=\frac{1}{\lambda}\{(x^{\lambda+1}-1)+(1+\lambda)(1-x)\}\geqq 0$$

となる．

注意 情報量 $I(\boldsymbol{p}\,;\,\boldsymbol{q})$ に対して

(30)
$$J(\boldsymbol{p}\,;\,\boldsymbol{q})=I(\boldsymbol{p}\,;\,\boldsymbol{q})+I(\boldsymbol{q}\,;\,\boldsymbol{p})$$

とおけば，J は“対称性” $J(\boldsymbol{p}\,;\,\boldsymbol{q})=J(\boldsymbol{q}\,;\,\boldsymbol{p})$ を持つ情報量である．

3.3 基準情報量，双曲的情報量，楕円的情報量の特徴づけ

基本情報量が 2.2 で与えた基準・双曲的・楕円的情報量であるための条件を考えよう．

定理 7. (i) 可微分基本情報量 $I(\boldsymbol{p}\,;\,\boldsymbol{q})$ に対して，加法性

(31)
$$I(\boldsymbol{p}\otimes\boldsymbol{p}'\,;\,\boldsymbol{q}\otimes\boldsymbol{q}')=I(\boldsymbol{p}\,;\,\boldsymbol{q})+I(\boldsymbol{p}'\,;\,\boldsymbol{q}')$$

が成り立つのは，基準情報量 $I^0(\boldsymbol{p}\,;\,\boldsymbol{q})$, $I^0=I_{KL}$ によって

(32)
$$I(\boldsymbol{p}\,;\,\boldsymbol{q})=c_1I^0(\boldsymbol{p}\,;\,\boldsymbol{q})+c_2I^0(\boldsymbol{q}\,;\,\boldsymbol{p}),\qquad c_1\geqq 0,\ c_2\geqq 0$$

と表わされる場合に限る．特に $I(\boldsymbol{p}\,;\,\boldsymbol{q})=I(\boldsymbol{q}\,;\,\boldsymbol{p})$ を条件とすれば，$I(\boldsymbol{p}\,;\,\boldsymbol{q})=cJ^0(\boldsymbol{p}\,;\,\boldsymbol{q})$ $(c>0)$ に限る．

(ii) 可微分基本情報量 $I(\boldsymbol{p}\,;\,\boldsymbol{q})$ に対して

(33) $$I(\boldsymbol{p}\otimes\boldsymbol{p}';\ \boldsymbol{q}\otimes\boldsymbol{q}')=I(\boldsymbol{p};\ \boldsymbol{q})+I(\boldsymbol{p}';\ \boldsymbol{q}')+I(\boldsymbol{p};\ \boldsymbol{q})\times I(\boldsymbol{p}';\ \boldsymbol{q}')$$

が成り立つのは，双曲的情報量 $I^{\lambda}(\boldsymbol{p};\ \boldsymbol{q})$, $\lambda>0$ に対して

(34) $$I(\boldsymbol{p};\ \boldsymbol{q})=\lambda I^{\lambda}(\boldsymbol{p};\ \boldsymbol{q})\quad \text{または}\quad \lambda I^{\lambda}(\boldsymbol{q};\ \boldsymbol{p})$$

となる場合に限る．

(iii) 可微分基本情報量 $I(\boldsymbol{p};\ \boldsymbol{q})$ に対して

(35) $$I(\boldsymbol{p}\otimes\boldsymbol{p}';\ \boldsymbol{q}\otimes\boldsymbol{q}')=I(\boldsymbol{p};\ \boldsymbol{q})+I(\boldsymbol{p}';\ \boldsymbol{q}')-I(\boldsymbol{p};\ \boldsymbol{q})\times I(\boldsymbol{p}';\ \boldsymbol{q}')$$

が成り立つのは，楕円的情報量 $I^{-\mu}(\boldsymbol{p};\ \boldsymbol{q})$, $1/2\geqq\mu>0$ に対して

(36) $$I(\boldsymbol{p};\ \boldsymbol{q})=\mu I^{-\mu}(\boldsymbol{p};\ \boldsymbol{q})\quad \text{または}\quad \mu I^{-\mu}(\boldsymbol{q};\ \boldsymbol{p})$$

となる場合に限る．

基本情報量の可微分性については，3.4 で改めて考察する．

証明　(i) $I(\boldsymbol{p};\ \boldsymbol{q})=\sum_k p_k K\left(\frac{q_k}{p_k}\right)$ と表わすとき，(31) は $\boldsymbol{p}=(p_1,\cdots,p_m),\boldsymbol{q}=(q_1,\cdots,q_m)$, $\boldsymbol{p}'=(p'_1,\cdots,p'_r)$, $\boldsymbol{q}'=(q'_1,\cdots,q'_r)$ に対して

$$\sum_{k=1}^{m}\sum_{j=1}^{r}p_k p'_j K\left(\frac{q_k}{p_k}\frac{q'_j}{p'_j}\right)=\sum_{k=1}^{m}p_k K\left(\frac{q_k}{p_k}\right)+\sum_{j=1}^{r}p'_j K\left(\frac{q'_j}{p'_j}\right)$$

と表わされる．特に $m=r=2$ とすれば

(37) $$\begin{aligned}&p_1p'_1K(x_1y_1)+p_1p'_2K(x_1y_2)+p_2p'_1K(x_2y_1)+p_2p'_2K(x_2y_2)\\&\quad=p_1K(x_1)+p_2K(x_2)+p'_1K(y_1)+p'_2K(y_2)\end{aligned}$$

となる．但し，$x_1=q_1/p_1$, $x_2=q_2/p_2$, $y_1=q'_1/p'_1$, $y_2=q'_2/p'_2$ とし，$p_1x_1+p_2x_2=1$, $p'_1y_1+p'_2y_2=1$ である．

(38) $$F(x,\ y)=K(xy)-K(x)-K(y)$$

とおくと (37) は

(39) $$p_1p'_1F(x_1,\ y_1)+p_1p'_2F(x_1,\ y_2)+p_2p'_1F(x_2,\ y_1)+p_2p'_2F(x_2,\ y_2)=0$$

となる．

補題 6.　$x_1>0$, $x_2>0$, $y_1>0$, $y_2>0$, $p_1x_1+p_2x_2=1$, $p_1+p_2=1$, $p'_1y_1+p'_2y_2=1$, $p'_1+p'_2=1$ に対して (39) が成り立てば，或る定数 c および関数 $A(y)$, $B(x)$ によって

(40) $$F(x,\ y)=c(x-1)(y-1)+A(y)(x-1)+B(x)(y-1)$$

と表わされる．

また $F(1,\ y)=0$, $F(x,\ 1)=0$ とすれば $F(x,\ y)$ が可微分な場合には (40) において $A(1)=B(1)=0$, かつ $A'(1)=B'(1)=0$ にとることができる．その条件をつけると，(40) の分解は一意である．

証明　$p_1+p_2=1$, $p_1x_1+p_2x_2=1$, かつ $p'_1+p'_2=1$, $p'_1y_1+p'_2y_2=1$ より

$$p_1=\frac{x_2-1}{x_2-x_1},\quad p_2=\frac{x_1-1}{x_1-x_2},\quad p'_1=\frac{y_2-1}{y_2-y_1},\quad p'_2=\frac{y_1-1}{y_1-y_2}$$

となる．故に (39) より

$$\frac{F(x_1, y_1)}{(x_1-1)(y_1-1)}-\frac{F(x_1, y_2)}{(x_1-1)(y_2-1)}-\frac{F(x_2, y_1)}{(x_2-1)(y_1-1)}+\frac{F(x_2, y_2)}{(x_2-1)(y_2-1)}=0$$

したがって，$x_1=x$, $y_1=y$, $x_2=a$, $y_2=b$ とおけば

$$\frac{F(x, y)}{(x-1)(y-1)}=-\frac{F(a, b)}{(a-1)(b-1)}+\frac{F(x, b)}{(x-1)(b-1)}+\frac{F(a, y)}{(a-1)(y-1)}$$

となる．故に

$$c=-\frac{F(a, b)}{(a-1)(b-1)},\quad B(x)=\frac{F(x, b)}{b-1},\quad A(y)=\frac{F(a, y)}{a-1} \tag{41}$$

とおけば (40) が成り立つ．$F(1, y)=0$, $F(x, 1)=0$ であれば $A(1)=B(1)=0$ となる．また $a\to1$, $b\to1$ とすれば $F(x, y)$ の可微分性より

$$c=-\frac{\partial^2 F}{\partial x\partial y}(1, 1),\quad B(x)=\frac{\partial F}{\partial y}(x, 1),\quad A(y)=\frac{\partial F}{\partial x}(1, y) \tag{42}$$

となる．さらに $c_0=-c$, $A_0(y)=A(y)-c(y-1)$, $B_0(x)=B(x)-c(x-1)$ とおくと，$A_0(1)=B_0(1)=A_0'(1)=B_0'(1)=0$ となり，この条件の下に $A_0(x)$, $B_0(y)$, c_0 は (40) より一意に定められる．

(i) の証明にもどる．補題 6 によって

$$K(xy)-K(x)-K(y)=c(x-1)(y-1)+A(y)(x-1)+B(x)(y-1)$$

となる．ここで

$$K_1(x)=K(x)-c(x-1) \tag{43}$$

とおくと

$$K_1(xy)-K_1(x)-K_1(y)=A(y)(x-1)+B(x)(y-1) \tag{44}$$

となる．この両辺に $\frac{\partial}{\partial x}$, $\frac{\partial}{\partial y}$ を施すと

$$yK_1'(xy)-K_1'(x)=A(y)+B'(x)(y-1)$$
$$K_1'(xy)+xyK_1''(xy)=A'(y)+B'(x)$$

となる．ここで $z=xy$, $y=z/x$ とし，z, x を独立変数にとると

$$K_1'(z)+zK_1''(z)=A'\left(\frac{z}{x}\right)+B'(x)$$

となる．$\frac{\partial}{\partial x}$ を施せば

$$0=-\frac{z}{x^2}A''\left(\frac{z}{x}\right)+B''(x)$$

すなわち

$$0=-\frac{y}{x}A''(y)+B''(x)$$

すなわち

$$yA''(y)=xB''(x)=c_0$$

故に

$$A''(y)=\frac{c_0}{y}, \qquad B''(x)=\frac{c_0}{x}$$

より

$$A'(y)=c_0 \log x+d, \qquad B'(x)=c_0 \log y+e$$

となるが，A′(1)=$B'(1)=0$ より $d=e=0$ である．よって

$$A(y)=c_0(y \log y-y)+f, \qquad B(x)=c_0(x \log x-x)+g$$

となるが，$A(1)=B(1)=0$ より $f=g=c_0$ となり

$$A(y)=c_0\{y \log y-(y-1)\}, \qquad B(x)=c_0\{x \log x-(x-1)\}$$

となる．これらを (44) に代入して

$$K_1(xy)-K_1(x)-K_1(y)=c_0(x-1)\{y \log y-(y-1)\}+c_0(y-1)\{x \log x-(x-1)\}$$

となる．故に

$$K_2(x)=K_1(x)-c_0 x \log x+2c_0(x-1)$$

とおくと　$K_2(xy)-K_2(x)-K_2(y)=0$ となる．故に或る定数 d_0 により（$K_2(1)=0$ を用いて）$K_2(x)=d_0 \log x$，故に

(45)
$$K(x)=c_0 x \log x+d_0 \log x+e_0(x-1)$$

と表わされる．これから $I(\boldsymbol{p}\,;\,\boldsymbol{q})=\sum_k p_k K\left(\frac{q_k}{p_k}\right)$ に代入して

$$I(\boldsymbol{p}\,;\,\boldsymbol{q})=c_0 I^0(\boldsymbol{q}\,;\,\boldsymbol{p})+d_0 I^0(\boldsymbol{p}\,;\,\boldsymbol{q})$$

と表わされることがわかった．

(ii) (31) の代りに (33) を仮定する．

(46)
$$F(x,\, y)=K(xy)-K(x)-K(y)-K(x)K(y)$$

とおくと，$F(x,\, 1)=F(1,\, y)=0$，かつ (40) が成り立つことがわかる．

$$K(x)=L(x)-1$$

とおくと，(46) は

(47)
$$F(x,\, y)=L(xy)-L(x)L(y)$$

となる．故に補題 6 によって

(48)
$$L(xy)-L(x)L(y)=c(x-1)(y-1)+A(y)(x-1)+B(x)(y-1)$$

と表わされる．両辺に $\frac{\partial^2}{\partial x \partial y}$ を施すと

$$L'(xy)+xy\,L''(xy)+L'(x)L'(y)=A'(y)+B'(x)+c$$

と表わされる．両辺に $\frac{\partial^2}{\partial x^2} \cdot \frac{\partial^2}{\partial y^2}$ を施すと

$$4xyL'''(xy)+x^2y^2L''''(xy)+L''(x)L''(y)+2L''(xy)=0$$

となる．ここで $xy=z$ とし，z と x とを独立変数と見ると

$$4zL'''(z)+z^2L''''(z)+L''\left(\frac{z}{x}\right)L''(x)+2L''(z)=0$$

さらに $\dfrac{\partial}{\partial x}$ を施して

$$L''\left(\frac{z}{x}\right)L'''(x)-\frac{z}{x^2}L'''\left(\frac{z}{x}\right)L''(x)=0$$

となる．すなわち

$$\frac{L'''(x)}{L''(x)}\cdot x=\frac{L'''(y)}{L''(y)}y=c_0$$

となる．よって積分して

$$\log L''(x)=c_0\log x+c_1$$

もう一度積分して

$$L(x)=ax^\lambda+c(x-1)+d, \qquad L''(x)=a\lambda(\lambda-1)x^{\lambda-2}>0$$

の形になる．但し，$L(1)=a+d=1$ である．これを(48)に代入して

$$\begin{aligned}L(xy)-L(x)L(y)=a(1-a)(xy)^\lambda&+ac(x^\lambda-1)(y-1)+ac(y^\lambda-1)(x-1)\\&+c(xy-1)-c^2(x-1)(y-1)+ac(x-1)+ac(y-1)\\&+d(1-ax^\lambda-ay^\lambda-c(x-1)-c(y-1)-d)\end{aligned}$$

$\lambda\neq0$, $\lambda\neq1$ より $a=1$, $d=0$．したがって

$$L(xy)-L(x)L(y)=c(x^\lambda-1)(y-1)+c(y^\lambda-1)(x-1)+c(1-c)(x-1)(y-1)$$

と表わされる．すなわち，任意の c に対して

$$L(x)=x^\lambda+c(x-1), \qquad L''(x)=\lambda(\lambda-1)x^{\lambda-2}>0$$

は求める解となる．$\lambda(\lambda-1)>0$ より $\lambda<0$ または $\lambda>1$ となる．これから $K(x)=(x^\lambda-1)+c(x-1)$ で（λ の代りに $-\lambda$ とおいて）(23)と比べて $I(\boldsymbol{p}\,;\,\boldsymbol{q})=\lambda I^\lambda(\boldsymbol{p}\,;\,\boldsymbol{q})$ または $\lambda I^\lambda(\boldsymbol{q}\,;\,\boldsymbol{p})$, $\lambda>0$ となる．

(iii) についても同様である．

3.4 可微分基本情報量

定義 7. 基本情報量 $I(\boldsymbol{p}\,;\,\boldsymbol{q})=\sum_{k=1}^{m}L(p_k, q_k)$ において，$0\leqq x\leqq1$, $0\leqq y\leqq1$ で定義された関数 $L(x,y)$ が2回(または3回)連続的微分可能であるとき，可微分基本情報量(Differentiable)という．

注意 (i) 補題2によって $L(x,y)=xK\left(\dfrac{y}{x}\right)$ と表わすとき，$L(x,y)$ が可微分であることは $K(x)$ が $x>0$ で可微分であることと同値である．$K(x)$ は凸関数であるから

$$\frac{d^2K}{dx^2}(x)\geqq0, \qquad x>0 \tag{49}$$

である．

(ii)　$x>0$ で定義された凸関数 $K(x)$ に対して $K(1)=0$, かつ $K(x)\geqq 0$ であるという条件は

$$\frac{dK}{dx}(1)=0 \tag{50}$$

という条件と同値である．

定義 8.　可微分基本情報量 $I(\boldsymbol{p}\,;\,\boldsymbol{q})$ に対して

$$\alpha=\frac{d^2K}{dx^2}(1)\geqq 0 \tag{51}$$

を I の不変数（Invariant）という．

$K(x)$ の代りに $K(x)-a(x-1)$ をとっても，不変数は同じである．

例 7.　(i)　基準情報量 $I^0(\boldsymbol{p}\,;\,\boldsymbol{q})$ に対して

$$\begin{aligned}&K^0(x)=-\log x+x-1, \qquad \frac{dK^0}{dx}(1)=0,\\ &\frac{d^2}{dx^2}K^0(x)=\frac{1}{x^2}>0, \qquad x>0, \qquad \alpha=\frac{d^2}{dx^2}K^0(1)=1\end{aligned} \tag{52}$$

すなわち

$$\alpha(I^0)=1 \tag{52)*}$$

である．

(ii)　双曲的情報量 I^λ, $\lambda>0$ および楕円的情報量 $I^{-\mu}$, $0<\mu\leqq 1/2$ に対して

$$\begin{aligned}&K^\lambda(x)=\frac{1}{\lambda}(x^{-\lambda}-1)+(x-1), \qquad \frac{dK^\lambda}{dx}(1)=0,\\ &\frac{d^2}{dx^2}K^\lambda(x)=(1+\lambda)x^{-\lambda-2}>0, \qquad x>0\end{aligned} \tag{53}$$

したがって

$$\alpha(I^\lambda)=1+\lambda>0, \qquad \lambda>0, \qquad \alpha(I^{-\mu})=1-\mu>0, \qquad 0<\mu\leqq\frac{1}{2} \tag{53)*}$$

である．特に $I_{KL}=I^0$, $I_P=I^1$, $I_K=I^{-1/2}$ に対して

$$\alpha(I_{KL})=1, \qquad \alpha(I_P)=2, \qquad \alpha(I_K)=\frac{1}{2} \tag{54}$$

となる．

注意　可微分基本情報量 $I(\boldsymbol{p}\,;\,\boldsymbol{q})$ の双対を $I^*(\boldsymbol{p}\,;\,\boldsymbol{q})$ とするとき

$$\alpha(I)=\alpha(I^*) \tag{55}$$

である．

証明　$K^*(x)=\dfrac{1}{x}K\!\left(\dfrac{1}{x}\right)$, $K(1)=\dfrac{dK}{dx}(1)=0$ であるから

$$\frac{dK^*}{dx}=-\frac{1}{x^2}K\left(\frac{1}{x}\right)-\frac{1}{x^3}K'\left(\frac{1}{x}\right), \qquad K'=\frac{dK}{dx}$$
$$\frac{d^2K^*}{dx^2}=\frac{2}{x^3}K\left(\frac{1}{x}\right)+\frac{4}{x^4}K'\left(\frac{1}{x}\right)+\frac{1}{x^5}K''\left(\frac{1}{x}\right)$$

したがって

$$K^*(1)=0, \qquad \frac{dK^*}{dx}(1)=0, \qquad \frac{d^2K^*}{dx^2}(1)=K''(1)=\alpha$$

が成り立つ.

定理 8. $0\leqq x\leqq 1$, $0\leqq y\leqq 1$ で定義された実関数 $L(x, y)$ が, $L(0, 0)=L(1, 1)=0$ で, かつ可微分であるとき, $I(\boldsymbol{p}\,;\,\boldsymbol{q})=\sum_{k=1}^{m}L(p_k, q_k)$ が可微分基本情報量であるための条件は

(IV) 不変性: に対しては

(56) $$x\frac{\partial L}{\partial x}+y\frac{\partial L}{\partial y}=L$$

が必要十分である.

(V) 凸性: に対しては

(57) $$\frac{\partial^2 L}{\partial x^2}\geqq 0, \qquad \frac{\partial^2 L}{\partial y^2}\geqq 0, \qquad \frac{\partial^2 L}{\partial x\partial y}\leqq 0$$

が必要であり,

(58) $$\frac{\partial^2 L}{\partial x^2}>0, \qquad \text{または} \qquad \frac{\partial^2 L}{\partial y^2}>0, \qquad \text{または} \qquad \frac{\partial^2 L}{\partial x\partial y}<0$$

が十分である.

証明 (IV) 不変性: $L(x, y)=xK\left(\frac{y}{x}\right)$ と表わされるならば, $K'=\frac{dK}{dx}$, $K''=\frac{d^2K}{dx^2}$ とおくとき

(59) $$\frac{\partial L}{\partial x}=K\left(\frac{y}{x}\right)-\frac{y}{x}K'\left(\frac{y}{x}\right), \qquad \frac{\partial L}{\partial y}=K'\left(\frac{y}{x}\right)$$

したがって

$$x\frac{\partial L}{\partial x}+y\frac{\partial L}{\partial y}=xK\left(\frac{y}{x}\right)=L$$

が成り立つ. 逆に (56) が成り立つならば

$$L(x, y)=xG(x, y)=xG(x, ux), \qquad u=\frac{y}{x}$$

とおいて, $G(x, ux)=F(x,u)$ と見るとき

$$\frac{\partial F}{\partial x}=\frac{-1}{x^2}L+\frac{1}{x}\frac{\partial L}{\partial x}+\frac{u}{x}\frac{\partial L}{\partial y}=\frac{1}{x^2}\left(x\frac{\partial L}{\partial x}+y\frac{\partial L}{\partial y}-L\right)=0$$

すなわち, $F(x, y)=K(u)$ と表わされ, $L(x, y)=xK\left(\frac{y}{x}\right)$ と表わされる.

(V)　凸性：(56)の両辺の $\frac{\partial}{\partial x}$ および $\frac{\partial}{\partial y}$ をとれば

$$x^2\frac{\partial^2 L}{\partial x^2}=y^2\frac{\partial^2 L}{\partial y^2}=-xy\frac{\partial^2 L}{\partial x\partial y} \tag{60}$$

である．一方，$L(x, y)=xK\left(\frac{y}{x}\right)$ とおくとき

$$\frac{\partial^2 L}{\partial x^2}=\frac{y^2}{x^3}K''\left(\frac{y}{x}\right),\qquad \frac{\partial^2 L}{\partial y^2}=\frac{1}{x}K''\left(\frac{y}{x}\right),\qquad \frac{\partial^2 L}{\partial x\partial y}=\frac{-y}{x^2}K''\left(\frac{y}{x}\right) \tag{61}$$

したがって

$$x^2\frac{\partial^2 L}{\partial x^2}=y^2\frac{\partial^2 L}{\partial y^2}=-xy\frac{\partial^2 L}{\partial x\partial y}=\frac{y^2}{x}K''\left(\frac{y}{x}\right) \tag{62}$$

である．故に $x>0$ で $K''(x)\geqq 0$ と $\frac{\partial^2 L}{\partial x^2}\geqq 0$ とは同値であり，$K''(x)>0$ と $\frac{\partial^2 L}{\partial x^2}>0$ とは同値である．

注意　(i) (59)より $K(1)=K'(1)=0$ を用いれば

$$\frac{\partial L}{\partial x}(p, p)=\frac{\partial L}{\partial y}(p, p)=0,\qquad 0<p<1 \tag{63}$$

が成り立つ．また(62)より I の不変数 α に対して

$$\alpha=K''(1)=p\frac{\partial^2 L}{\partial x^2}(p, p)=p\frac{\partial^2 L}{\partial y^2}(p, p)=-p\frac{\partial^2 L}{\partial x\partial y}(p, p),\qquad 0<p<1 \tag{64}$$

が成り立つ．

定理 9.　$I(\boldsymbol{p}\,;\,\boldsymbol{q})$ を（3回連続的微分可能な）可微分基本情報量とする．

$$\boldsymbol{p}=(p_1,\cdots,p_m),\qquad \boldsymbol{q}=(q_1,\cdots,q_m),\qquad \sum_k p_k=1,\qquad \sum_k q_k=1$$
$$\boldsymbol{p}^0=(p_1^0,\cdots,p_m^0),\qquad \boldsymbol{q}^0=(q_1^0,\cdots,q_m^0),\qquad \sum_k p_k^0=1,\qquad \sum_k q_k^0=1$$

に対して

$$\begin{aligned} &p_k=p_k^0+u_k, && q_k=q_k^0+v_k, && k=1,\cdots,m \\ &\sum_{k=1}^m u_k=0, && \sum_{k=1}^m v_k=0 \end{aligned} \tag{65}$$

とおく，いま

$$|u_k|<\varepsilon,\qquad |v_k|<\varepsilon,\qquad k=1,\cdots,m \tag{66}$$

であれば

$$\begin{aligned} I(\boldsymbol{p}\,;\,\boldsymbol{q})=I(\boldsymbol{p}^0\,;\,\boldsymbol{q}^0)&+\sum_{k=1}^m \frac{\partial L}{\partial x}(p_k^0, q_k^0)u_k+\sum_{k=1}^m \frac{\partial L}{\partial y}(p_k^0, q_k^0)v_k \\ &+\frac{1}{2}\sum_{k=1}^m p_k^0\frac{\partial^2 L}{\partial x^2}(p_k^0, q_k^0)\left(\frac{u_k}{p_k^0}-\frac{v_k}{q_k^0}\right)^2+R,\qquad |R|=O(\varepsilon^3) \end{aligned} \tag{67}$$

が成り立つ．特に $\boldsymbol{p}^0=\boldsymbol{q}^0$ であれば

(68) $$I(\boldsymbol{p}\,;\ \boldsymbol{q})=\frac{\alpha}{2}\sum_{k=1}^{m}\frac{1}{q_k^0}(u_k-v_k)^2+R, \qquad R=O(\varepsilon^3)$$

となる．さらに，$\boldsymbol{p}^0=\boldsymbol{q}^0=\boldsymbol{q}$ であれば

(69) $$I(\boldsymbol{p}\,;\ \boldsymbol{q}^0)=\frac{\alpha}{2}\sum_{k=1}^{m}\frac{1}{q_k^0}u_k^0+R, \qquad R=O(\varepsilon^3)$$

となる．

証明 $L(x,y)$ を $x=p_k^0$, $y=q_k^0$ のまわりの Taylor 展開を用いれば

$$\begin{aligned}
I(\boldsymbol{p}\,;\ \boldsymbol{q}) &= \sum_{k=1}^{m} L(p_k, q_k)\\
&= \sum_{k=1}^{m}\Big\{L(p_k^0, q_k^0)+\frac{\partial L}{\partial x}(p_k^0, q_k^0)u_k+\frac{\partial L}{\partial y}(p_k^0, q_k^0)v_k\\
&\quad +\frac{1}{2}\Big(\frac{\partial^2 L}{\partial x^2}(p_k^0, q_k^0)u_k^2+2\frac{\partial^2 L}{\partial x\partial y}(p_k^0, q_k^0)u_kv_k+\frac{\partial^2 L}{\partial y^2}(p_k^0, q_k^0)v_k^2\Big)\Big\}+O(\varepsilon^3)\\
&= I(\boldsymbol{p}^0\,;\ \boldsymbol{q}^0)+\sum_{k=1}^{m}\Big(\frac{\partial L}{\partial x}(p_k^0, q_k^0)u_k+\frac{\partial L}{\partial y}(p_k^0, q_k^0)v_k\Big)\\
&\quad +\frac{1}{2}\sum_{k=1}^{m} p_k^{02}\frac{\partial^2 L}{\partial x^2}(p_k^0, q_k^0)\Big(\frac{u_k}{p_k^0}-\frac{v_k}{q_k^0}\Big)^2+O(\varepsilon^3)
\end{aligned}$$

となる（但し (60) を用いた）．よって (67) が成り立つ．(67) において $\boldsymbol{p}^0=\boldsymbol{q}^0$ とし，(64) を用いれば (68) となる．(68) において $\boldsymbol{q}=\boldsymbol{q}^0$ とおけば (69) となる．

注意 （追記）査読者より御注意のあった論文 Rathie and Kannappan (1972), Rényi (1961) と比べて気づいた点を挙げる．

(i) P.N. Ratie and P.L. Kannappan (1972) において，われわれが定義した情報量 $I^\lambda(\boldsymbol{p}\,;\ \boldsymbol{q})$, $\lambda\neq 0$ (2.1, 定義 2) の定義が述べられ，定理 4（相対性による一意性）の証明が与えられている．

(ii) A. Rényi (1961) では工藤弘吉 (1953) において与えられた加法性を持つ情報量 $\tilde{I}^\lambda(\boldsymbol{p}\,;\ \boldsymbol{q})$, $\lambda\neq 0$ (3.1, (3) 式) を（独立に）定義している．そして $I^0(\boldsymbol{p}\,;\ \boldsymbol{q})$ および $\tilde{I}^\lambda(\boldsymbol{p}\,;\ \boldsymbol{q})$ の公理系による決定がなされている．但し，Rényi の場合には $I^0(\boldsymbol{p}\,;\ \boldsymbol{q})$ および $\tilde{I}^\lambda(\boldsymbol{p}\,;\ \boldsymbol{q})$ が $(\boldsymbol{p}\,;\ \boldsymbol{q})=(p_1,\cdots,p_m\,;\ q_1,\cdots,q_m)$, $(p_k\geqq 0,\ q_k\geqq 0,\ p_1+\cdots+p_m\leqq 1,\ q_1+\cdots+q_m\leqq 1)$ に対しても定義されており，公理系の中に，それらに対する拡張された不変性の公理が含まれているので，Rényi の結果をそのままわれわれの場合に適用することは出来ない（すなわち，Rényi の結果と定理 7 とは互いに独立である）．

4. L 集合と情報量

4.1 情報量の L 集合による特徴づけ

次の定義 9 と定理 10 は Kudō (1952) による．

定義 9. 有限分布 $\boldsymbol{p}=(p_1,\cdots,p_m)$, $0\leqq p_k\leqq 1$, $\sum_{k=1}^{m} p_k=1$ および $\boldsymbol{q}=(q_1,\cdots,q_m)$, $0\leqq q_k\leqq 1$, $\sum_{k=1}^{m} q_k=1$ に対して L 集合：

(1) $$L(\boldsymbol{p},\boldsymbol{q})=\left\{(x,y)\,\middle|\,x=\sum_{k=1}^{m}\alpha_k p_k,\ y=\sum_{k=1}^{m}\alpha_k q_k,\ 0\leqq\alpha_k\leqq 1,\ k=1,\cdots,m\right\}$$

と定義する.

定理 10. $L(\boldsymbol{p},\boldsymbol{q})$ は次の諸性質を持つ:

(I)　$(0,0)\in L(\boldsymbol{p},\boldsymbol{q})$,　$(1,1)\in L(\boldsymbol{p},\boldsymbol{q})$.

(II)　$L(\boldsymbol{p},\boldsymbol{q})\subset[0,1]\times[0,1]$.

(III)　$L(\boldsymbol{p},\boldsymbol{q})$ は点 $\left(\frac{1}{2},\frac{1}{2}\right)$ を中心とする凸集合である.

(VI)　p_k/q_k,　$k=1,\cdots,m$ を大きさの順に並べて

(2) $$0\leqq\frac{q_1}{p_1}\leqq\cdots\leqq\frac{q_m}{p_m}$$

とするとき $\left(\text{但し, }\frac{0}{0}=0\text{ とおく}\right)$, $L(\boldsymbol{p},\boldsymbol{q})$ は

(3) $$(0,0),\ (p_1,q_1),\ (p_1+p_2,q_1+q_2),\cdots,(p_1+\cdots+p_{m-1},\ q_1+\cdots+q_{m-1}),\ (1,1)$$

を $(0,0)$ と $(1,1)$ を結ぶ線分 $\varDelta$ の下方の折線とし,

(4) $$(0,0),\ (p_m,q_m),\ (p_m+p_{m-1},q_m+q_{m-1}),\cdots,(p_m+\cdots+p_2,\ q_m+\cdots+q_2),\ (1,1)$$

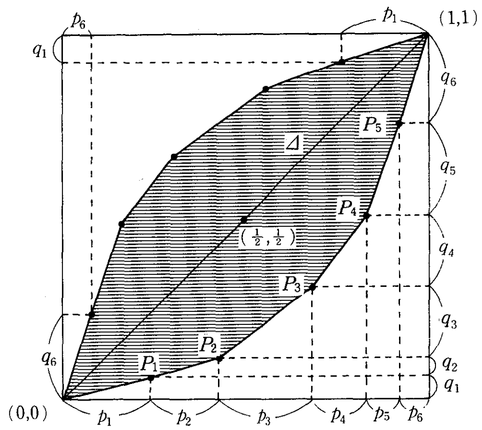

図 3.　L集合: $L(\boldsymbol{p},\boldsymbol{q})$ のグラフ

を $\varDelta$ の上方の折線とし, 両者にはさまれる部分に等しい.

証明　はじめに $p_1=\cdots=p_k=0$, $q_1=\cdots=q_k=0$ であれば, $(p_1,q_1)=\cdots=(p_1+\cdots+p_k,q_1+\cdots+q_k)=(0,0)$ であるから, 定義により

$$L((p_1,\cdots,p_m),(q_1,\cdots,q_m))=L((p_{k+1},\cdots,p_m),(q_{k+1},\cdots,q_m))$$

を注意しておく.

(I), (II) は自明.

(III)　$\lambda(\sum \alpha_k p_k, \sum \alpha_k q_k)+\mu(\sum \beta_k p_k, \sum\beta_k q_k)=(\sum \gamma_k p_k, \sum\gamma_k q_k)$, 但し, $\gamma_k=\lambda\alpha_k+\mu\beta_k$, $0<\lambda<1$, $0<\mu<1$, $\lambda+\mu=1$ より L が凸集合であることがわかる．また

$$\frac{1}{2}\{(\sum\alpha_k p_k, \sum\alpha_k q_k)+(\sum(1-\alpha_k)p_k, \sum(1-\alpha_k)q_k)\}=\left(\frac{1}{2}, \frac{1}{2}\right)$$

より，L は $\left(\frac{1}{2}, \frac{1}{2}\right)$ を中心とすることがわかる．

(IV)　まず $L(\boldsymbol{p}, \boldsymbol{q})$ は 2^m 個の点 $(p_{i_1}+\cdots+p_{i_l}, q_{i_1}+\cdots+q_{i_l})$, $i_1<\cdots<i_l$; $l=0,1,\cdots,m$ (但し，$l=0$ のときは $(0,0)$ とする)の張る凸集合であることを見よう．いま仮に $1\geqq\alpha_1\geqq\cdots\geqq\alpha_n\geqq0$ とすれば

$$(\sum\alpha_k p_k, \sum\alpha_k q_k)=(1-\alpha_1)(0,0)+(\alpha_1-\alpha_2)(p_1, q_1)+\cdots +(\alpha_{m-1}-\alpha_m)(p_1+\cdots+p_{m-1}, q_1+\cdots+q_{m-1})+\alpha_m(1,1)$$

と表わされることよりわかる．次に(2)のようにとれば，上記 2^m 個の点は(3)と(4)とに囲まれる図形に含まれることがわかる(図3参照)．よって(IV)が成り立つ．

注意　(i)　(I), (II), (III) であるような任意の凸多角形は或る $\boldsymbol{p}, \boldsymbol{q}$ によって $L(\boldsymbol{p}, \boldsymbol{q})$ と表わされる．

(ii)　$L(\boldsymbol{p}, \boldsymbol{q})$ が Δ ((0,0) と (1,1) を結ぶ直線)であるのは $p_1=q_1,\cdots,p_m=q_m$,　すなわち $\boldsymbol{p}=\boldsymbol{q}$ の場合に限る．

(iii)　$L(\boldsymbol{p}, \boldsymbol{q})=[0,1]\times[0,1]$ となるのは $p_k q_k=0$,　$k=1,\cdots,m$,　すなわち $\boldsymbol{p}\perp\boldsymbol{q}$ の場合に限る．

(iv)　$L_1=L(\boldsymbol{p}_1, \boldsymbol{q}_1)$,　$L_2=L(\boldsymbol{p}_2, \boldsymbol{q}_2)$ に対して

(5)　$$L_1 * L_2=L(\boldsymbol{p}_1\otimes\boldsymbol{p}_2, \boldsymbol{q}_1\otimes\boldsymbol{q}_2)$$

と定義する($L_1 * L_2$ は，$\boldsymbol{p}_i, \boldsymbol{q}_i$ を用いることなく，L_1 と L_2 とから直接に幾何学的に定義することができる)．このとき

(6)　$$L_1 * L_2=L_2 * L_1$$

(7)　$$L_1 * L_2\supset L_1, \qquad L_1 * L_2\supset L_2$$

(8)　$$L_1 * L_2=L_1 \iff L_2=\Delta$$
$$L_1 * L_2=L_2 \iff L_1=\Delta$$

である(図4参照)．

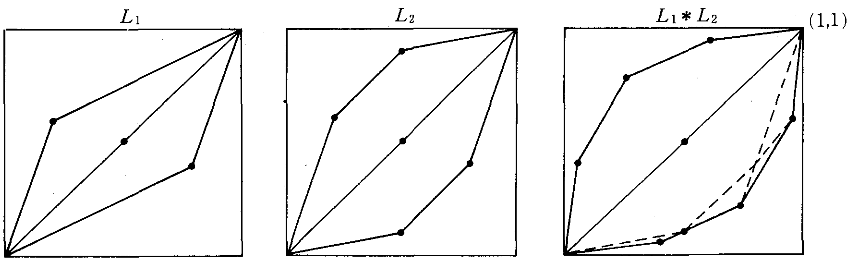

図4.　$L_1 * L_2$ のグラフ

定理 11. $I(\boldsymbol{p}\,;\,\boldsymbol{q})$を有限分布$(\boldsymbol{p},\ \boldsymbol{q})$に対する情報量とする．そのとき

(I) $L(\boldsymbol{p}_1, \boldsymbol{q}_1)=L(\boldsymbol{p}_2, \boldsymbol{q}_2) \ \Rightarrow\ I(\boldsymbol{p}_1\,;\,\boldsymbol{q}_1)=I(\boldsymbol{p}_2\,;\,\boldsymbol{q}_2)$

(II) $L(\boldsymbol{p}_1, \boldsymbol{q}_1)\supsetneqq L(\boldsymbol{p}_2, \boldsymbol{q}_2) \ \Rightarrow\ I(\boldsymbol{p}_1\,;\,\boldsymbol{q}_1)>I(\boldsymbol{p}_2\,;\,\boldsymbol{q}_2)$, または$I(\boldsymbol{p}_2\,;\,\boldsymbol{q}_2)=\infty$

である．

証明 (I) $L(\boldsymbol{p}_1, \boldsymbol{q}_1)=L(\boldsymbol{p}_2, \boldsymbol{q}_2)$となるのは，(i) $\boldsymbol{p}_1=(p_1,\cdots,p_m)$, $\boldsymbol{q}_1=(q_1,\cdots,q_m)$, $p_1=\cdots=p_k=0$, $q_1=\cdots=q_k=0$ であるとき $\boldsymbol{p}_2=(p_{k+1},\cdots,p_m)$, $\boldsymbol{q}_2=(q_{k+1},\cdots,q_m)$とする場合，(ii) $q_k/p_k=q_{k+1}/p_{k+1}$のとき$P_k=(p_k, q_k)$を取り除いて，折れ線P_{k-1}, P_k, P_{k+1}を線分P_{k-1}, P_{k+1}でおきかえる場合，およびこれらを何回かくりかえすことによって得られる．(i) に対して (I) が成り立つことはIの簡約性による．(ii) に対して (I) が成り立つことはIの不変性による．

(II) $I(\boldsymbol{p}_2\,;\,\boldsymbol{q}_2)<\infty$とする．(i) $\dfrac{q_k}{p_k}<\dfrac{q_{k+1}}{p_{k+1}}$の場合に $\boldsymbol{p}_1=(p_1,\cdots,p_m)$, $\boldsymbol{p}_2=(p_1,\cdots,p_{k-1}, p_k+p_{k+1}, p_{k+2},\cdots,p_m)$, $\boldsymbol{q}_1=(q_1,\cdots,q_m)$, $\boldsymbol{q}_2=(q_1,\cdots,q_{k-1}, q_k+q_{k+1}, q_{k+2},\cdots,q_m)$ とすれば $L(\boldsymbol{p}_1, \boldsymbol{q}_1)-L(\boldsymbol{p}_2, \boldsymbol{q}_2)=\varDelta\, P_{k-1}\, P_k\, P_{k+1}+$(その対称図形)である．一方，情報量$I$の凸性によって$I(\boldsymbol{p}_1\,;\,\boldsymbol{q}_1)>I(\boldsymbol{p}_2\,;\,\boldsymbol{q}_2)$である（図5参照）．(ii) 一般に分布$(\boldsymbol{p}_1, \boldsymbol{q}_1)$, $(\boldsymbol{p}_2, \boldsymbol{q}_2)$に対して$L(\boldsymbol{p}_1, \boldsymbol{q}_1)\supsetneqq L(\boldsymbol{p}_2, \boldsymbol{q}_2)$であれば，上の操作を有限回くりかえして$L(\boldsymbol{p}_1, \boldsymbol{q}_1)$より$L(\boldsymbol{p}_2, \boldsymbol{q}_2)$に達することを見ればよい．これを一般に記述することは複雑になるが図6のように示すことができる．

Kudō (1952)，工藤 (1953) では，一般の分布$(\boldsymbol{p}, \boldsymbol{q})$に対して，$I_{KL}(\boldsymbol{p}\,;\,\boldsymbol{q})$に対して (I), (II) を証明している．

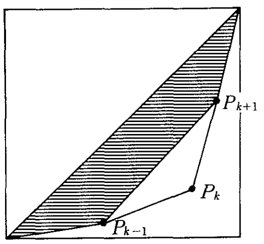

図5.

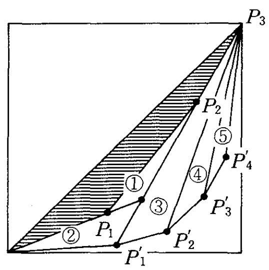

図6.

定理 12. （定理11の逆命題）任意のL集合$L(\boldsymbol{p}, \boldsymbol{q})$に対して実数$I(\boldsymbol{p}\,;\,\boldsymbol{q})$が対応して

(0) $L(\boldsymbol{p}, \boldsymbol{q})=\varDelta \iff I(\boldsymbol{p}\,;\,\boldsymbol{q})=0$

(I) $L(\boldsymbol{p}_1, \boldsymbol{q}_1)=L(\boldsymbol{p}_2, \boldsymbol{q}_2) \ \Rightarrow\ I(\boldsymbol{p}_1\,;\,\boldsymbol{q}_1)=I(\boldsymbol{p}_2\,;\,\boldsymbol{q}_2)$

(II) $L(\boldsymbol{p}_1, \boldsymbol{q}_1)\supsetneqq L(\boldsymbol{p}_2, \boldsymbol{q}_2) \ \Rightarrow\ I(\boldsymbol{p}_1\,;\,\boldsymbol{q}_1)>I(\boldsymbol{p}_2\,;\,\boldsymbol{q}_2)$, または $I(\boldsymbol{p}_2\,;\,\boldsymbol{q}_2)=\infty$

であれば，$I(\boldsymbol{p}\,;\,\boldsymbol{q})$は一つの情報量である．

証明 (i) $I(\boldsymbol{p}\,;\,\boldsymbol{q})$の対称性は (I) よりわかる．

(ii) $I(\boldsymbol{p}\,;\,\boldsymbol{q})$の簡約性は

$$L((0, p_2,\cdots,p_m), (0, q_2,\cdots,q_m)) = L((p_2,\cdots,p_m), (q_2,\cdots,q_m))$$

に対して (I) よりわかる.

(iii) $I(\boldsymbol{p};\boldsymbol{q})$の非負性は (0), (I), (II) よりわかる.

(iv) $I(\boldsymbol{p};\boldsymbol{q})$の不変性は (I) よりわかる.

(v) $I(\boldsymbol{p};\boldsymbol{q})$の凸性は，いま (2) が成り立つとすると

(イ) $\boldsymbol{p}_1=(p_1,\cdots,p_m)$, $\boldsymbol{q}_1=(q_1,\cdots,q_m)$, $\boldsymbol{p}_2=(p_1,\cdots,p_k+p_{k+1},\cdots,p_m)$, $\boldsymbol{q}_2=(q_1,\cdots,q_k+q_{k+1},\cdots,q_m)$に対しては (II) よりわかる.

(ロ) $k+1<l$ の場合に，$\boldsymbol{p}_1=(p_1,\cdots,p_m)$, $\boldsymbol{q}_1=(q_1,\cdots,q_m)$, $\boldsymbol{p}_2=(p_1,\cdots,\check{p}_k,\cdots,p_k+p_l,\cdots,\check{p}_l,\cdots,p_m)$, $\boldsymbol{q}_2=(q_1,\cdots,\check{q}_k,\cdots,q_k+q_l,\cdots,\check{q}_l,\cdots,q_m)$ ($\check{}$ 印はその値が欠けていることを示す) とする．但し

$$\frac{q_1}{p_1}\leqq\cdots\leqq\frac{q_{k-1}}{p_{k-1}}\leqq\frac{q_{k+1}}{p_{k+1}}\leqq\cdots\leqq\frac{q_h}{p_h}\leqq\frac{q_k+q_l}{p_k+p_l}\leqq\frac{q_{h+1}}{p_{h+1}}\leqq\cdots\leqq\frac{q_{l-1}}{p_{l-1}}\leqq\frac{q_{l+1}}{p_{l+1}}\leqq\cdots\leqq\frac{q}{p}$$

とする．$L(\boldsymbol{p}_1,\boldsymbol{q}_1)$の$\Delta$の下方にある頂点$P_0,\cdots,P_m(P_k=(p_1+\cdots+p_k,\ q_1+\cdots+q_k))$の張る折線$P_0,P_1,\cdots,P_m$に対して$L(\boldsymbol{p}_2,\boldsymbol{q}_2)$の$\Delta$の下方にある頂点$(0,0),(p_1,q_1),\cdots,(p_1+\cdots+p_{k-1},\ q_1+\cdots+q_{k-1}),(p_1+\cdots+p_{k-1}+p_{k+1},\ q_1+\cdots+q_{k-1}+q_{k+1}),\cdots,(1,1)$はL集合の定義によってすべて$L(\boldsymbol{p}_1,\boldsymbol{q}_1)$に含まれる．よって，$L(\boldsymbol{p}_1,\boldsymbol{q}_1)\supset L(\boldsymbol{p}_2,\boldsymbol{q}_2)$である．したがって (I) または (II) によって$I(\boldsymbol{p}_1;\boldsymbol{q}_1)\geqq I(\boldsymbol{p}_2,;\boldsymbol{q}_2)$が成り立つ (図7参照).

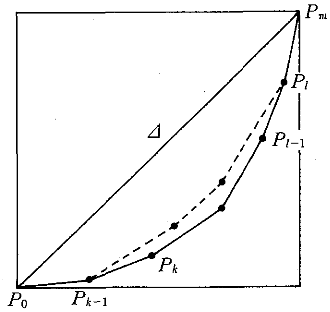

図7.

以上の定理11, 12によって，情報量$I(\boldsymbol{p};\boldsymbol{q})$は，L集合全体の族$\mathscr{L}$の上の単調な非負汎関数として特徴づけられた.

注意 情報量$I(\boldsymbol{p};\boldsymbol{q})$の加法性は，幾何学的には

"$L(\boldsymbol{p}_1,\boldsymbol{q}_1)*L(\boldsymbol{p}_2,\boldsymbol{q}_2)$に対応する情報量は$I(\boldsymbol{p}_1;\boldsymbol{q}_1)+I(\boldsymbol{p}_2;\boldsymbol{q}_2)$に等しい"

という命題で表わすことができる.

4.2 弧長型・面積型・幅型情報量

A. 弧長型情報量

定理12によってL集合，$L(\boldsymbol{p},\boldsymbol{q})$を用いていろいろな情報量を定義することができるわけである.

3.2で特にくわしく扱った基本情報量を凸図形$L(\boldsymbol{p},\boldsymbol{q})$を用いて特徴づけることをしらべよう．この場合に基本情報量$I(\boldsymbol{p};\boldsymbol{q})$は，$x>0$で定義されたある凸関数$K(x)$によって

(9) $$I(\boldsymbol{p};\boldsymbol{q})=\sum_{k=1}^{m}p_kK\left(\frac{q_k}{p_k}\right)$$

と表わされる．但し，$K(x)\geqq0$, $K(1)=K'(1)=0$, $K''(x)>0$とする．いま凸集合$L(\boldsymbol{p},\boldsymbol{q})$の$\Delta$の下側の折れ線を

(10) $$C(\boldsymbol{p};\boldsymbol{q})=\{(x,y)\mid y=\varphi(x)\}$$

とおく．(9) はC上の線積分

(11) $$I(\boldsymbol{p}\,;\,\boldsymbol{q})=\int_c K\left(\frac{d\varphi}{dx}\right)dx$$

と表わされる．

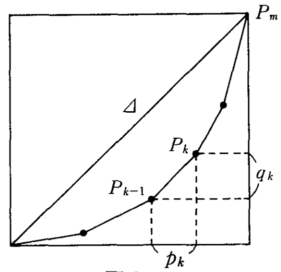

図 8．

証明　線分 $\overline{P_{k-1}P_k}$ 上では $\dfrac{d\varphi}{dx}=\dfrac{q_k}{p_k}$ であり，また $\overline{P_{k-1}P_k}$ の x 軸上への正射影の長さは p_k であるから，(11) の右辺の積分は (9) の右辺の和に等しい（図 8 参照）．特に $I^\lambda(\boldsymbol{p}\,;\,\boldsymbol{q})$ に対しては

$$K^\lambda(x)=\frac{1}{\lambda}(x^{-\lambda}-1)+(x-1),\qquad \lambda\neq 0$$
$$K^0(x)=-\log x+(x-1)$$

であった．

例 8．　定理 12 の性質 (0), (I), (II) を持つ量として

(12) $$I(\boldsymbol{p}\,;\,\boldsymbol{q})=(L(\boldsymbol{p},\boldsymbol{q})\text{の周の長さ})\times\frac{1}{2}-\sqrt{2}$$

をとることができる．これを式で表わすと，やはり (9) の形に表わすことができる．すなわち

(12)* $$K(x)=\sqrt{1+x^2}-\frac{1}{\sqrt{2}}(x+1)$$

とおく．$K(x)\geqq 0$, $K(1)=K'(1)=0$, $K''(x)>0$ であって

$$I_K(\boldsymbol{p}\,;\,\boldsymbol{q})=\int_c K\left(\frac{d\varphi}{dx}\right)dx=\sum_{k=1}^m p_k\left(\sqrt{1+\left(\frac{q_k}{p_k}\right)^2}-\frac{1}{\sqrt{2}}\left(\frac{q_k}{p_k}+1\right)\right)$$
$$=\left(\sum_{k=1}^m\sqrt{p_k^2+q_k^2}\right)-\sqrt{2}$$

すなわち (12) と一致することがわかる．よって基本情報量のことを弧長型情報量と呼ぶことにする．

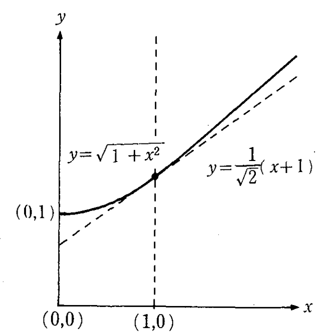

図 9．$y=\sqrt{1+x^2}$ のグラフ

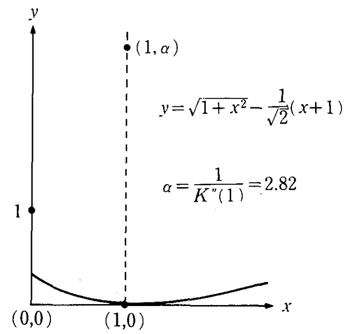

図 10．$y=\sqrt{1+x^2}-\dfrac{1}{\sqrt{2}}(x+1)$ のグラフ

B.　面積型情報量

L 集合の全体 $\mathscr{L}$ 上の非負・単調な汎関数として，まず思いつくのは $L(\boldsymbol{p}, \boldsymbol{q})$ の面積である．すなわち

$$I_A(\boldsymbol{p}\,;\ \boldsymbol{q})=\mathrm{A}(L(\boldsymbol{p}, \boldsymbol{q}))=L(\boldsymbol{p}, \boldsymbol{q}) \text{ の面積} \tag{13}$$

とおくと，I_A は定理 12 の条件 (0),(I),(II) を満足する．したがって，一つの情報量である．$\boldsymbol{p}=(p_1,\cdots,p_m)$, $\boldsymbol{q}=(q_1,\cdots,q_m)$ とすれば

$$\mathrm{A}(L(\boldsymbol{p}, \boldsymbol{q}))=\frac{1}{2}\sum_{k=1}^{m}\sum_{j=1}^{m}|p_kq_j-p_jq_k| \tag{14}$$

と表わされる．或は

$$0\leqq\frac{q_1}{p_1}\leqq\cdots\leqq\frac{q_m}{p_m}$$

とすれば，(14) は

$$\mathrm{A}(L(\boldsymbol{p}, \boldsymbol{q}))=\sum_{k<j}(p_kq_j-p_jq_k) \tag{14*}$$

となる．

証明　図 11 において ($m=4$ の場合)

$$\frac{1}{2}\mathrm{A}=\triangle\ P_0P_1P_2+\triangle\ P_0P_2P_3+\cdots+\triangle P_0P_{m-1}P_m$$

$$\triangle\ (\ \triangle\ P_0P_kP_{k+1})=\frac{1}{2}\begin{bmatrix}1 & 0 & 0\\ 1 & p_1+\cdots+p_k & q_1+\cdots+q_k\\ 1 & p_1+\cdots+p_{k+1} & q_1+\cdots+q_{k+1}\end{bmatrix}$$

$$=\frac{1}{2}\sum_{i=1}^{k}(p_iq_{k+1}-p_{k+1}q_i),\qquad k=1,\cdots,m-1$$

であるから，合わせて (14)* となる(各三角形 $\triangle\ P_0P_kP_{k+1}$ は図 11 のように P_{k+1} を頂点とする k 個の小さい三角形に分割され，その各々の面積が $(p_iq_{k+1}-p_{k+1}q_i)/2$ となる)．

注意　面積型情報量は基本情報量ではなくて

$$I(\boldsymbol{p}\,;\ \boldsymbol{q})=\sum_{k=1}^{m}\sum_{j=1}^{m}L\begin{pmatrix}p_k & p_j\\ q_k & q_j\end{pmatrix} \tag{15}$$

$$L\begin{pmatrix}p & p\\ q & q\end{pmatrix}=0,\qquad L\begin{pmatrix}p_1 & p_2\\ q_1 & q_2\end{pmatrix}=L\begin{pmatrix}p_2 & p_1\\ q_2 & q_1\end{pmatrix}$$

$$L\begin{pmatrix}\alpha p_1 & p_2\\ \alpha q_1 & q_2\end{pmatrix}=\alpha L\begin{pmatrix}p_1 & p_2\\ q_1 & q_2\end{pmatrix},\qquad 0\leqq\alpha\leqq1 \tag{15*}$$

によって特徴づけられる．一般に (15) の形に表わされるが，必ずしも (15)* が成り立たない情報量とはどのようなものであるかは明らかでない．

次に I_A を一般にして，$0\leqq x\leqq1$, $0\leqq y\leqq1$ で定義される連続関数 $f(x, y)$ で，$x>0$, $y>0$ において $f(x, y)>0$ となるものをとり

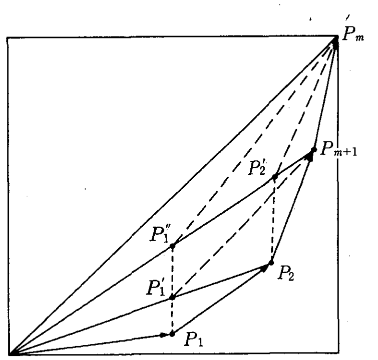

図 11.

(16)
$$I_{A,f}(\boldsymbol{p}\,;\,\boldsymbol{q})=\iint_{L(\boldsymbol{p},\,\boldsymbol{q})} f(x,y)dxdy$$

とおく．$I_{A,f}$ は $L(\boldsymbol{p},\boldsymbol{q})\in\mathscr{L}$ に対する非負・単調な汎関数であるから，定理 12 によって一つの情報量である．特に

(17)
$$f(x,y)=x^iy^j, \qquad i,j=0,1,2,\cdots$$

とすれば

(18)
$$I_{A,f}(\boldsymbol{p}\,;\,\boldsymbol{q})=\iint_{L(\boldsymbol{p},\,\boldsymbol{q})} x^iy^jdxdy=\int_0^1\left\{\int_{\varphi_1(x)}^{\varphi_0(x)} y^jdy\right\}x^idx$$
$$=\frac{1}{j+1}\int_0^1(\varphi_0(x)^{j+1}-\varphi_1(x)^{j+1})x^idx$$

で与えられる．但し，$L(\boldsymbol{p},\boldsymbol{q})$ を囲む上側の折線を $y=\varphi_0(x)$，下側の折線を $y=\varphi_1(x)$ とおく（図 12 参照）．一般に (16) の形に表わされる情報量を面積型情報量と呼ぶことにする．

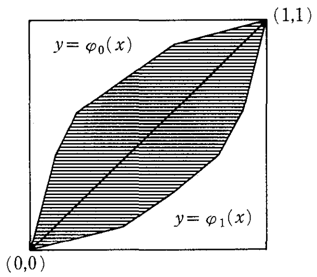

図 12．

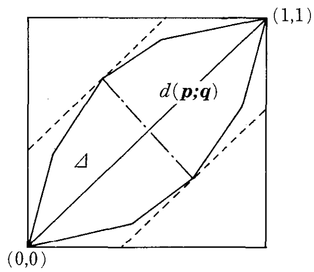

図 13．$d(\boldsymbol{p}\,;\,\boldsymbol{q})$

C．幅型情報量

3.1，　例 2 に挙げた $d(\boldsymbol{p}\,;\,\boldsymbol{q})=\frac{1}{\sqrt{2}}\sum_{k=1}^{m}|p_k-q_k|$ は，$L(\boldsymbol{p},\boldsymbol{q})$を用いて表わせば，凸集合 L の Δ の方向に垂直な幅である（工藤 (1953)，pp. 118–119）（図 13 参照）．一般に $L(\boldsymbol{p},\boldsymbol{q})$ の x 軸の正方向と θ，$0\leqq\theta<\pi$ なる角をなす方向への幅 $B_\theta(L)$は

$$B_\theta(L)=\sum_{k=1}^{m}|p_k\cos\theta+q_k\cos\theta|$$

と表わされ，上記 $d(\boldsymbol{p}\,;\,\boldsymbol{q})$ は $\theta=\frac{3}{4}\pi$ の場合である．いま $L(\boldsymbol{p},\boldsymbol{q})$ の幅の平均

$$B(L(\boldsymbol{p},\boldsymbol{q}))=\frac{1}{\pi}\int_0^{\pi}B_\theta(L)\,d\theta$$

をとれば，定理 12 の条件 (I),(II) を満足する．但し，(0) を満足しないので，$B(\Delta)=\frac{1}{\pi}2\sqrt{2}$ を引いて

$$I_B(\boldsymbol{p}\,;\,\boldsymbol{q})=B(L(\boldsymbol{p},\,\boldsymbol{q}))-B(\Delta)$$

とおけば，I_B は一つの情報量となる．I_B を $(p_1,\cdots,p_m,q_1,\cdots,q_m)$ で表わせば（その表示はむずかしくない），弧長型でも面積型でもないことがわかる．また一般に，連続関数 $f(\theta)>0$，$0\leqq\theta<$

π を用いて

$$I_{B,f}(\boldsymbol{p}\,;\,\boldsymbol{q})=\frac{1}{\pi}\int_0^{\pi}(B_\theta(L)-B_\theta(\varDelta))f(\theta)d\theta$$

も情報量となる．これを幅型情報量と呼ぶ．

4.3 情報量の族の弱完備性と強完備性

4.1で情報量がL集合の全体$\mathscr{L}$上の非負・単調な汎関数として特徴づけられることを見たが，ここでは逆に或る情報量の集合から，L集合の相等・包含関係が特徴づけられないかという問題を考えよう．この問題は工藤（1953），pp. 115-119で取り上げられたものである．

定義10． 情報量の族 $\{I_\omega(\boldsymbol{p}\,;\,\boldsymbol{q})\,|\,\omega\in\Omega\}$ が弱完備（Weakly complete）とは

(19) $$\forall\omega\in\Omega:I_\omega(\boldsymbol{p}_1\,;\,\boldsymbol{q}_1)=I_\omega(\boldsymbol{p}_2\,;\,\boldsymbol{q}_2)\quad\Rightarrow\quad L(\boldsymbol{p}_1,\boldsymbol{q}_1)=L(\boldsymbol{p}_2,\boldsymbol{q}_2)$$

が成り立つことをいう．

定義11． 情報量の族 $\{I_\omega(\boldsymbol{p}\,;\,\boldsymbol{q})\,|\,\omega\in\Omega\}$ が強完備（Strongly complete）とは

(20) $$\forall\omega\in\Omega:I_\omega(\boldsymbol{p}_1\,;\,\boldsymbol{q}_1)\leqq I_\omega(\boldsymbol{p}_2\,;\,\boldsymbol{q}_2)\quad\Rightarrow\quad L(\boldsymbol{p}_1,\boldsymbol{q}_1)\subset L(\boldsymbol{p}_2,\boldsymbol{q}_2)$$

が成り立つことをいう．

注意 (19), (20)で $\Leftarrow$ が成り立つことは定理11である．また $\{I_\omega\,|\,\omega\in\Omega\}$ が強完備であれば必ず弱完備であるが，逆は必ずしも真でない（後に反例を挙げる）．また強完備の条件(20)は

(21) $$L(\boldsymbol{p}_1,\boldsymbol{q}_1)\not\subset L(\boldsymbol{p}_2,\boldsymbol{q}_2)\quad\Rightarrow\quad\exists\omega\in\Omega:I_\omega(\boldsymbol{p}_1\,;\,\boldsymbol{q}_1)>I_\omega(\boldsymbol{p}_2\,;\,\boldsymbol{q}_2)$$

と同値である．以下弱完備または強完備となる情報量の族の例をいくつか挙げる．

定理13． (i) 面積型情報量の族

(22) $$\left\{I_A^{(i,j)}(\boldsymbol{p}\,;\,\boldsymbol{q})=\iint_{L(\boldsymbol{p},\boldsymbol{q})}x^iy^jdxdy\;\middle|\;i,j=0,1,2,\cdots\right\}$$

は弱完備である．

(ii) 面積型情報量の族

(23) $$\left\{I_{A,f}(\boldsymbol{p}\,;\,\boldsymbol{q})=\iint_{L(\boldsymbol{p},\boldsymbol{q})}f(x,y)dxdy\;\middle|\;f(x,y)\geqq 0,\text{連続}\right\}$$

は強完備である．

証明 (i) (18)によって

$$I_A^{(i,j)}=\frac{1}{j+1}\int_0^1(\varphi_0(x)^{j+1}-\varphi_1(x)^{j+1})x^idx$$

であった．ここで良く知られている事実：

"[0, 1]上の連続関数 $f(x)$ の i 次のMomentを

$$M_i(f)=\int_0^1 x^i f(x)dx, \qquad i=0,1,2,\cdots$$

とすれば，二つの連続関数 $f(x)$, $g(x)$ に対して

$$M_i(f)=M_i(g),\quad i=0,1,\cdots \iff f(x)=g(x)''$$

を用いる．まず，$j=0$ とすれば，$\varphi_0(x)-\varphi_1(x)$ は $I_A^{(i,0)}$, $i=0,1,2,\cdots$ により決定される．次に，$j=1$ とすれば $\varphi_0^2(x)-\varphi_1^2(x)$ は $I_A^{(i,1)}$, $i=0,1,2,\cdots$ により決定される．故に，$\varphi_0(x)+\varphi_1(x)=(\varphi_0^2(x)-\varphi_1^2(x))/(\varphi_0(x)+\varphi_1(x))$ も $I_A^{(i,j)}$, $j=0,1$; $i=0,1,2,\cdots$ により決定される．したがって，$\varphi_0(x)$, $\varphi_1(x)$ も $I_A^{(i,j)}$ より決定される．すなわち (22) の族の弱完備性が示された．

(ii)　(23) の族に対して (21) を示そう．いま $L(\boldsymbol{p}_1,\boldsymbol{q}_1)\not\subset L_1(\boldsymbol{p}_2,\boldsymbol{q}_2)$ とすれば

$$D=(L(\boldsymbol{p}_1,\boldsymbol{q}_1)\cup L(\boldsymbol{p}_2,\boldsymbol{q}_2))-L(\boldsymbol{p}_2,\boldsymbol{q}_2)\neq\emptyset$$

である（図 14 参照）．したがって，連続関数 $f_\varepsilon(x,y)$, $0\leqq x\leqq 1$, $0\leqq y\leqq 1$を

(i)　$(x,y)\in D \Rightarrow f_\varepsilon(x,y)=1$

(ii)　$d((x,y),D)>\varepsilon \Rightarrow f_\varepsilon(x,y)=\varepsilon$

(iii)　$\forall(x,y) \Rightarrow 1\geqq f_\varepsilon(x,y)\geqq\varepsilon$

にとる．ここで ε を十分に小さくすれば

$$\iint_{L(\boldsymbol{p}_1,\boldsymbol{q}_1)} f_\varepsilon(x,y)dxdy-\iint_{L(\boldsymbol{p}_2,\boldsymbol{q}_2)} f_\varepsilon(x,y)dxdy\geqq(D\text{ の面積})-\varepsilon>0$$

となる．よって，(23) の族は強完備である．

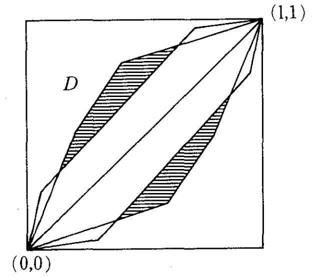

図 14.

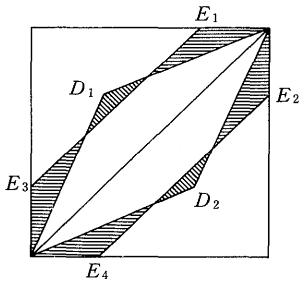

図 15.

注意　(22) の族が強完備でないことは，次の反例によって示される．$L_1=L(\boldsymbol{p}_1,\boldsymbol{q}_1)$, $L_2=L(\boldsymbol{p}_2,\boldsymbol{q}_2)$, $L_1\cup L_2-L_2=D_1\cup D_2\neq\emptyset$, $L_1\cup L_2-L_1=E_1\cup E_2\cup E_3\cup E_4\neq\emptyset$, かつ（$E_1$ の面積）>（D_1 の面積），（E_2 の面積）>（D_2 の面積）とする（図 15 参照）．このとき

$$(x_1,y_1)\in E_1,(x_2,y_2)\in D_1 \Rightarrow x_2<x_1,\ y_2<y_1 \Rightarrow x_1^iy_1^j>x_2^iy_2^j$$
$$(x_1,y_1)\in E_2,(x_2,y_2)\in D_2 \Rightarrow x_2<x_1,\ y_2<y_1 \Rightarrow x_1^iy_1^j>x_2^iy_2^j$$

である．よって

$$\iint_{L_2} x^iy^jdxdy>\iint_{L_1} x^iy^jdxdy, \qquad i,j=0,1,\cdots$$

が成り立つ．よって，(22) の族は強完備でない．

注意 凸集合 $L(\boldsymbol{p}, \boldsymbol{q})$ は幅 $B_\theta(L(\boldsymbol{p}, \boldsymbol{q}))$, $0\leqq\theta<\pi$ を与えれば決定されることに注意すれば，定理 13 と類似の定理が幅型情報量についても成り立つことがわかる．

定理 14． 弧長型情報量（基本情報量）について

(i) $\forall\alpha>0$, $\forall\varepsilon>0$ に対して，族

(24) $$\Omega^{(\alpha,\alpha+\varepsilon)}=\left\{I^\lambda(\boldsymbol{p}\,;\,\boldsymbol{q})=\frac{1}{\lambda}\left(\sum_{k=1}^m\frac{p_k^{1+\lambda}}{q_k^\lambda}-1\right)\middle|\,\alpha<\lambda<\alpha+\varepsilon\right\}$$

は弱完備である

(ii) $\forall\,\alpha>0$, $\forall\,\varepsilon>0$ （但し，$1/2\geqq\alpha>\varepsilon$）に対して，族

(25) $$\Omega^{(-\alpha,-\alpha+\varepsilon)}=\left\{I^{-\mu}(\boldsymbol{p}\,;\,\boldsymbol{q})=\frac{1}{\mu}\left(1-\sum_{k=1}^m p_k^{1-\mu}q_k^\mu\right)\middle|\,\alpha>\mu>\alpha-\varepsilon\right\}$$

は弱完備である．

(iii) 族

(26) $$\Omega=\left\{I_K(\boldsymbol{p}\,;\,\boldsymbol{q})=\sum_{k=1}^m p_kK\left(\frac{q_k}{p_k}\right)\middle|\,K(1)=K'(1)=0,\ K''(x)>0,(x>0)\right\}$$

は強完備である．

証明 (i) (24)において $\boldsymbol{p}=(p_1,\cdots,p_m)$, $\boldsymbol{q}=(q_1,\cdots,q_m)$, $0<\frac{q_1}{p_1}<\frac{q_2}{p_2}<\cdots<\frac{q_m}{p_m}<\infty$ とする．そのとき

$$\lambda I^\lambda(\boldsymbol{p}\,;\,\boldsymbol{q})+1=f(\lambda)=\sum_{k=1}^m p_k\exp(-\lambda\xi_k)$$

$$\xi_k=\log\frac{q_k}{p_k},\qquad k=1,\cdots,m,\quad \xi_1<\xi_2<\cdots<\xi_m$$

である．すなわち

(27) $$\alpha(t)=\begin{cases}0, & t<\xi_1\\ p_1+\cdots+p_k, & \xi_k\leqq t<\xi_{k+1}\\ 1, & t\geqq\xi_m\end{cases}$$

とおけば

$$f(\lambda)=\int_{-\infty}^\infty\exp(-\lambda t)\,d\alpha(t)$$

と Laplace-Stieltjes 積分で表わされる．反転公式（岩波数学辞典第2版，Bromwich 積分，p. 463)によって

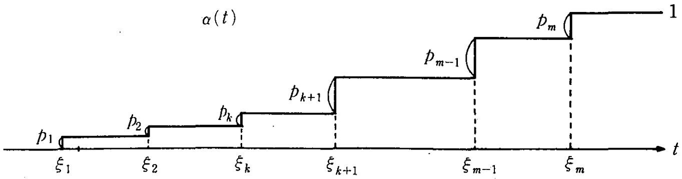

図 16. $\alpha(t)$ のグラフ

(28) $$\alpha(t)=\lim_{T\to\infty}\frac{1}{2\pi i}\int_{c-iT}^{c+iT}\frac{f(\tilde{\lambda})}{\tilde{\lambda}}e^{\tilde{\lambda}t}\,d\tilde{\lambda}$$

と表わされる（但し，$\tilde{\lambda}$: 複素数）．$\tilde{\lambda}=\lambda+i\mu$, λ, $\mu\in\boldsymbol{R}$ に対して

$$\begin{aligned} f(\lambda+i\mu)&=\sum_{k=1}^{m}p_k e^{-(\lambda+i\mu)\xi_k}=\sum_{k=1}^{m}p_k\,e^{-i\mu\xi_k}e^{-\lambda\xi_k}\\ &=\sum_{k=1}^{n}p_k\left\{\sum_{n=0}^{\infty}\frac{1}{n!}(-i\mu\xi_n)^n\,e^{-\lambda\xi_k}\right\}\\ &=\sum_{n=0}^{\infty}\frac{1}{n!}(i\mu)^n\left\{\sum_{k=1}^{m}p_k(-\xi_k)^n\,e^{-\lambda\xi_k}\right\}\end{aligned}$$

である．しかるに $f(\lambda)=\sum_{k=1}^{m}p_k e^{-\lambda\xi_k}$ に対して

$$\frac{d^nf(\lambda)}{d\lambda^n}=\sum_{k=1}^{m}p_k(-\xi_k)^n\,e^{-\lambda\xi_k}$$

であるから，

(29) $$f(\lambda+i\mu)=\sum_{n=0}^{\infty}\frac{1}{n!}(i\mu)^n\,\frac{d^nf(\lambda)}{d\lambda^n}$$

と表わされる．よって，$c\in(\alpha,\alpha+\varepsilon)$ にとれば，$\alpha(t)$，したがって，$\xi_1,\cdots,\xi_m$, $p_1,\cdots,p_m$ は $f(\lambda)$, $\alpha<\lambda<\alpha+\varepsilon$ によって決定される．すなわち (24) の弱完備性が示された．

(ii)　(25) の族の弱完備性についても同様である．

(iii)　(26) の族の強完備性について：二組の分布の対（$\boldsymbol{p}$, $\boldsymbol{q}$），($\boldsymbol{p}'$, $\boldsymbol{q}'$）に対して $L(\boldsymbol{p},\boldsymbol{q})$ と $L(\boldsymbol{p}',\boldsymbol{q}')$ との包含関係をしらべるとき，必要に応じて事象を細分して，$\boldsymbol{p}=\boldsymbol{p}'$ の場合に直すことができる．そこで

$$\boldsymbol{p}=\boldsymbol{p}'=(p_1,\cdots,p_m),\quad \boldsymbol{q}=(q_1,\cdots,q_m),\quad \boldsymbol{q}'=(q_1',\cdots,q_m')$$
$$\frac{q_1}{p_1}\leqq\cdots\leqq\frac{q_m}{p_m},\quad \frac{q_1'}{p_1}\leqq\cdots\leqq\frac{q_m'}{p_m}$$

とし，$L(\boldsymbol{p},\boldsymbol{q})$ と $L(\boldsymbol{p},\boldsymbol{q}')$ の $\varDelta$ の下側の折線をそれぞれ

$$y=\varphi(x),\qquad y=\psi(x)$$

とする．いま

(30) $$L(\boldsymbol{p},\boldsymbol{q})\not\supset L(\boldsymbol{p},\boldsymbol{q}')$$

と仮定するとき，都合のよい凸関数 $K(x)$ をとって

(31) $$I_K(\boldsymbol{p}\,;\,\boldsymbol{q})=\sum_{k=1}^{m}p_kK\left(\frac{q_k}{p_k}\right)<I_K(\boldsymbol{p}\,;\,\boldsymbol{q}')=\sum_{k=1}^{m}p_kK\left(\frac{q_k'}{p_k}\right)$$

となることを証明しよう．(30) より

$$\varphi(x)\leqq\psi(x),\qquad 0\leqq x\leqq 1$$

となることはない．よって，$\exists x_1$, $\exists x_2$, $0\leqq x_1<x_2\leqq 1$ に対して

(32) $$\varphi(x_1)=\psi(x_1),\quad \varphi(x_2)=\psi(x_2),\quad \varphi(x)>\psi(x),\qquad x_1<x<x_2$$

が成り立つ（図 17 参照）．いま

(33) $$x_1=p_1+\cdots+p_h,\quad \varphi(x_1)=q_1+\cdots+q_h=\psi(x_1)=q_1'+\cdots+q_h',\qquad h\geqq 0$$

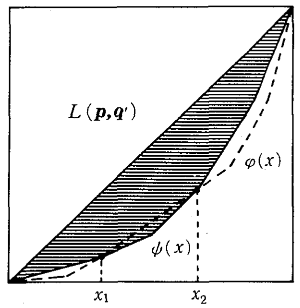

図 17.

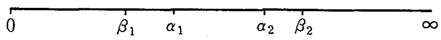

図 18.

および

(34) $\quad x_2=p_1+\cdots+p_l,\quad \varphi(x_2)=q_1+\cdots+q_l=\psi(x_2)=q_1'+\cdots+q_l',\qquad h<l\leqq m$

とする．そのとき(32)より

$$x_1-\varepsilon \quad \text{で} \quad \varphi(x)\leqq\psi(x) \qquad \therefore\ \alpha_1=\left(\frac{d\varphi}{dx}\right)^{-}(x_1)\geqq\beta_1=\left(\frac{d\psi}{dx}\right)^{-}(x_1)$$

$$x_2+\varepsilon \quad \text{で} \quad \varphi(x)\leqq\psi(x) \qquad \therefore\ \alpha_2=\left(\frac{d\varphi}{dx}\right)^{+}(x_2)\leqq\beta_2=\left(\frac{d\psi}{dx}\right)^{+}(x_2)$$

となる(但し，$\alpha_1\leqq\alpha_2$, 図 18 参照)．いま或る定数 a_1, a_2, b_1, b_2 によって

(35) $\quad K(x)=a_1x+b_1,\quad 0<x\leqq\alpha_1,\qquad K(x)=a_2x+b_2,\qquad \alpha_2<x<\infty$

とおく．

$$I_K(\boldsymbol{p}\,;\,\boldsymbol{q})=\sum_{k=1}^{m}p_kK\left(\frac{q_k}{p_k}\right)=\sum_{k=1}^{h}+\sum_{k=h+1}^{l}+\sum_{k=l+1}^{m}=I_1+I_2+I_3$$

$$I_K(\boldsymbol{p}\,;\,\boldsymbol{q}')=\sum_{k=1}^{m}p_kK\left(\frac{q_k'}{p_k}\right)=\sum_{k=1}^{h}+\sum_{k=h+1}^{l}+\sum_{k=l+1}^{m}=I_1'+I_2'+I_3'$$

とおけば

$$I_1=\sum_{k=1}^{h}p_k\left(a_1\frac{q_k}{p_k}+b_1\right)=a_1\varphi(x_1)+b_1x_1$$

$$I_1'=\sum_{k=1}^{h}p_k\left(a_1\frac{q_k'}{p_k}+b_1\right)=a_1\psi(x_1)+b_1x_1$$

であるから，$I_1=I_1'$ となる．全く同様に $I_3=I_3'$ となる．よって，$K(x)$が $x_1\leqq x\leqq x_2$ で凸になるように接続すれば

$$I_2<I_2'$$

となることが示されれば，$I_K(\boldsymbol{p}\,;\,\boldsymbol{q})<I_K(\boldsymbol{p}\,;\,\boldsymbol{q}')$ が導びかれたことになる．すなわち，α_1, α_2, 1 の相対的関係によって凸関数 $K(x)$, $K(x)\geqq 0$, $K(1)=K'(1)=0$ を図 19 のようにとる．

(i) (33), (34) で $l=h+2$, $\dfrac{q_{h+1}}{p_{h+1}}=\dfrac{q_{h+2}}{p_{h+2}}$ とする(図 20 参照).

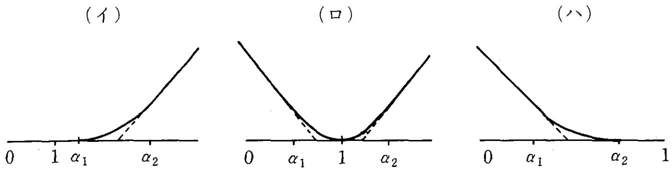

図 19.

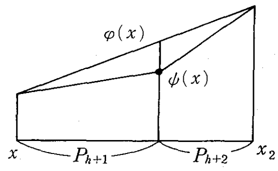

図 20.

$$
\begin{aligned}
I_2' &= p_{h+1}K\left(\frac{q_{h+1}'}{p_{h+1}}\right)+p_{h+2}K\left(\frac{q_{h+2}'}{p_{h+2}}\right)\\
&>(p_{h+1}+p_{h+2})K\left(\frac{q_{h+1}'+q_{h+2}'}{p_{h+1}+p_{h+2}}\right)=(p_{h+1}+p_{h+2})K\left(\frac{q_{h+1}+q_{h+2}}{p_{h+1}+p_{h+2}}\right)\\
&=p_{h+1}K\left(\frac{q_{h+1}}{p_{h+1}}\right)+p_{h+2}K\left(\frac{q_{h+2}}{p_{h+2}}\right)=I_2
\end{aligned}
$$

すなわち，$I_2<I_2'$ となる．

(ii)　一般の場合の $I_2<I_2'$ の証明は定理 11，(II) の証明の場合と同様である．

最後に上で構成した $K(x)$ は，凸関数 ($K''(x)>0$) の条件を満足していない．よって，$K(x)$ に十分近い凸関数 $K_1(x)$ (但し，$K_1(x)\geqq 0,\ K_1(1)=K_1'(1)=0$)

$$|K(x)-K_1(x)|<\varepsilon, \qquad 0<x<\infty$$

をとれば，$|I_K(\boldsymbol{p}\,;\ \boldsymbol{q})-I_{K_1}(\boldsymbol{p}\,;\ \boldsymbol{q})|\leqq\varepsilon$ であるから，$2\varepsilon<I_2'-I_2$ に ε をとれば

$$I_{K_1}(\boldsymbol{p}\,;\ \boldsymbol{q})<I_{K_1}(\boldsymbol{p}\,;\ \boldsymbol{q}')$$

が成り立つ．

注意　定理 14 において (25) の族が強完備でないことは，工藤(1953)，pp. 117-118 において（伊関兼四郎氏による反例を挙げて）示されている．すなわち

$$\boldsymbol{p}_1=\boldsymbol{p}_2=\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right),\quad \boldsymbol{q}_1=\left(\frac{44}{136},\frac{44}{136},\frac{48}{136}\right),\quad \boldsymbol{q}_2=\left(\frac{43}{136},\frac{46}{136},\frac{47}{136}\right)$$

に対して

$$I^{-\mu}(\boldsymbol{p}_1\,;\ \boldsymbol{q}_1)>I^{-\mu}(\boldsymbol{p}_2\,;\ \boldsymbol{q}_2), \qquad 0<\mu<\frac{1}{2}$$

であるが，$L(\boldsymbol{p}_1,\boldsymbol{q}_1)\ \not\supset\ L(\boldsymbol{p}_2,\boldsymbol{q}_2)$ となっている．同様に (24) の族も強完備でない．

5.　情報量と統計

5.1　正則情報量とその評価

これまで考察した情報量の統計への応用を見よう．歴史的に見れば，1900 年代の K. Pearson のカイ自乗適合度検定法（Test of goodness of fit）(1900) および 1920 年代の R. A. Fisher の最尤推定法（Maximum likelihood estimate）(1925) がある．前者は情報量 $I_P=I^1$ による近似の評価であり，後者は $I_{KL}=I^0$ の値を最小にするようなパラメータの推定法である．そして 1970 年代に情報量 $I_{KL}=I^0$ を用いて予測の立場に立つモデル選択の理論として AIC の理論（赤池（1973，1974））が生まれた．

これらは特定の情報量を用いているが，その計算法の多くは次に定義する正則情報量についてもそのままあてはめられる．以下にその計算を確めよう．

定義 12． 情報量 $I(\boldsymbol{p};\ \boldsymbol{q})$ が正則情報量(Regular Information)であるとは，次の性質(A)，(B)が成り立つことをいう．

(A) $\boldsymbol{E}$ 上の分布 $\boldsymbol{q}^0=(q_1^0,\cdots,q_m^0)$ を固定し，これに十分近い分布 $\boldsymbol{p}=(p_1,\cdots,p_m)$，$\boldsymbol{q}=(q_1,\cdots,q_m)$ を考える．すなわち

$$
\begin{aligned}
&(1) \qquad p_k=q_k^0+u_k, \qquad q_k=q_k^0+v_k, \qquad k=1,\cdots,m\\
&\qquad\quad u_1+\cdots+u_m=0, \quad v_1+\cdots+v_m=0,\\
&(2) \qquad |u_k|<\varepsilon, \qquad |v_k|<\varepsilon, \qquad k=1,\cdots,m
\end{aligned}
$$

であるとき，$I(\boldsymbol{p};\ \boldsymbol{q})$ は $(u_1,\cdots,u_{m-1},\ v_1,\cdots,v_{m-1})$ に関して3回連続的微分可能であって，或る定数 $\alpha>0$ に対して

$$
(3) \qquad I(\boldsymbol{p};\ \boldsymbol{q})=\frac{\alpha}{2}\sum_{k=1}^{m}\frac{1}{q_k^0}(u_k-v_k)^2+R, \qquad R=O(\varepsilon^3)
$$

が成り立つ．ここに α を I の不変数（Invaviant）という．

(B) 分布 $\boldsymbol{q}$ を定めるとき，分布 $\boldsymbol{p}$ をどのように変化させても，$\boldsymbol{q}$ によって定まる或る定数 $c(\boldsymbol{q})$ によって

$$
(4) \qquad I(\boldsymbol{p};\ \boldsymbol{q})\ <\ c(\boldsymbol{q})
$$

が成り立つ．

注意 3.4，定理9で見たように，可微分基本情報量 $I(\boldsymbol{p};\ \boldsymbol{q})$ で，不変数 $\alpha>0$ の場合にはすべて条件(A)が成り立つ（例えば，I_P, I_{KL}, I_K, I^λ, $-1/2\leqq\lambda<\infty$ の場合)．

また基本情報量 $I(\boldsymbol{p};\ \boldsymbol{q})=\sum_{k=1}^{m}L(p_k,\ q_k)$, $L(p,\ q)=pK\left(\frac{q}{p}\right)$ に対しては，q を定めるとき，p をどのように変えても

$$
(4)^* \qquad pK\left(\frac{q}{p}\right)\ <\ c_0(q)
$$

であることと，(4)とは同値である．特に

$$
K^\lambda(x)=\frac{1}{\lambda}(x^{-\lambda}-1)+(x-1), \qquad -\frac{1}{2}\leqq\lambda<\infty,\ \lambda\neq0
$$

に対しては

$$
pK^\lambda\left(\frac{q}{p}\right)=\frac{p}{\lambda}\left(\left(\frac{p}{q}\right)^\lambda-1\right)+q-p\leqq\begin{cases}\dfrac{1}{\lambda q^\lambda}+1, & \lambda>0\\[2ex] \dfrac{1+\lambda}{-\lambda}, & 0>\lambda\geqq-\dfrac{1}{2}\end{cases}
$$

であり，また

$$
K^0(x)=-\log x+(x-1)
$$

に対しては

$$pK^0\left(\frac{q}{p}\right)=-p\log\left(\frac{q}{p}\right)+p\left(\frac{q}{p}-1\right)\leqq\log q+1$$

が成り立つ．

よって，基準・双曲的・楕円的情報量 I^λ, $-1/2\leqq\lambda<\infty$ はすべて正則情報量である．同様にこれらの関数，例えば $\tilde{I}^\lambda$（例 4）も正則情報量である．

注意　面積型情報量 $I_A(\boldsymbol{p};\ \boldsymbol{q})$ や $I_A(\boldsymbol{p};\ \boldsymbol{q})^2$ に対しては，条件 (A) が成り立たない．

条件 (A) に対して $q_k=q_k^0$（すなわち，$v_1=\cdots=v_m=0$）とおけば

(5)
$$x_k=\frac{p_k-q_k^0}{\sqrt{q_k^0}}\quad \text{すなわち}\quad p_k=q_k^0\left(1+\frac{x_k}{\sqrt{q_k^0}}\right),\qquad k=1,\cdots,m$$

とおくとき，$|x_k|<\varepsilon$,　$k=1,\cdots,m$ であれば

(6)
$$I(\boldsymbol{p};\ \boldsymbol{q}^0)=\frac{\alpha}{2}\left(\sum_{k=1}^m x_k^2\right)+O(\varepsilon^3)$$

が成り立つ．

(A) の評価の例を一つ挙げよう．KL 情報量 $I^0=I_{KL}$ に対しては $\alpha(I^0)=1$, Pearson 情報量 $I^1=I_P$ に対しては $\alpha(I^1)=2$ であるから

$$I^0(\boldsymbol{p};\ \boldsymbol{q}^0)=\frac{1}{2}I^1(\boldsymbol{p};\ \boldsymbol{q}^0)+O(\varepsilon^3)$$

である．また角谷情報量に対しては $\alpha=1/2$ であるから，ε が小さいとき

$$I_K:I_{KL}:I_P=1:2:4$$

である．例えば $m=2$ の場合にも表 1 のように計算される．

表 1.

	I_K : I_{KL} : I_P
$\boldsymbol{p}=(0.6,\ 0.4)$　$\boldsymbol{q}^0=(0.5,\ 0.5)$	1 : 1.98 : 3.94
$\boldsymbol{p}=(0.6,\ 0.4)$　$\boldsymbol{q}^0=(0.4,\ 0.6)$	1 : 2.00 : 4.12
$\boldsymbol{p}=(0.7,\ 0.3)$　$\boldsymbol{q}^0=(0.4,\ 0.6)$	1 : 1.97 : 4.02

さて分割 $\boldsymbol{E}=\{E_1,\cdots,E_m\}$ 上の分布 $\boldsymbol{q}^0=(q_1^0,\cdots,q_m^0)$ に従う n 回の独立試行によって，$E_1,\cdots,E_m$ がそれぞれ

(7)
$$(N_1,\cdots,N_m)\text{ 回},\qquad n=N_1+\cdots+N_m$$

起ったとする．このとき，N_k は 2 項分布 $B(n,\ q_k^0)$

$$\text{prob}\,(N_k=n_k)=\frac{n!}{n_k!(n-n_k)!}q_k^{0\,n_k}(1-q_k^0)^{n-n_k},\qquad 0\leqq n_k\leqq n$$

に従う．よって，確率変数

(8)
$$X_k=\frac{1}{\sqrt{n}}\frac{N_k-nq_k^0}{\sqrt{q_k^0}},\qquad k=1,\cdots,m$$

とおけば，期待値 $\mathrm{E}(X_k)=0$, 分散 $\sigma^2(X_k)=1-q_k^0$ で，$n\to\infty$ のとき

$$X_k \Longrightarrow N(0,\ 1-q_k^0), \qquad \text{（法則収束）} \tag{9}$$

となる．但し，$N(a,\sigma^2)$ は平均 a, 分散 σ^2 の正規分布を示す．また，$(N_1,\cdots,N_m)$ の同時分布は多項分布

$$\mathrm{prob}\,(N=n_1,\cdots,N_m=n_m)=\frac{n!}{n_1!\cdots n_m!}\,q_1^{0n_1}\cdots q_m^{0n_m}, \qquad n=n_1+\cdots+n_m$$

に従う．$(X_1,\cdots,X_m)$ の同時分布を考えると，$\mathrm{E}(X_kX_j)=-\sqrt{q_k^0q_j^0}$, $k\neq j$, かつ $\sum_{k=1}^{m}\sqrt{q_k^0}\,X_k=0$ で，$n\to\infty$ のとき

$$(X_1,\cdots,X_m) \Longrightarrow N((0,\cdots,0),\ \boldsymbol{A}), \qquad \text{（法則収束）} \tag{10}$$

である．但し，$N((a_1,\cdots,a_m),\ \boldsymbol{A})$ は平均ベクトル $(a_1,\cdots,a_m)$, 共分散行列 $\boldsymbol{A}$ の m 次元正規分布とする．(10) の場合には

$$\begin{aligned}\boldsymbol{A}&=(a_{kj}), \quad a_{kk}=1-q_k^0, \quad a_{kj}=-\sqrt{q_k^0\,q_j^0}, \qquad k\neq j\\ &=\boldsymbol{E}_m-{}^t\sqrt{\boldsymbol{q}^0}\cdot\sqrt{\boldsymbol{q}^0}, \qquad \sqrt{\boldsymbol{q}^0}=(\sqrt{q_1^0},\cdots,\sqrt{q_m^0})\end{aligned}$$

である（但し，$\boldsymbol{E}_m$ は m 次元単位行列とする）．さらに良く知られているように，$n\to\infty$ のときに

$$X_1^2+\cdots+X_m^2 \Longrightarrow \chi_{m-1}^2, \qquad \text{（法則収束）} \tag{11}$$

である．但し，χ_{m-1}^2 は自由度 $m-1$ の χ^2 分布とする．

定理 15． $\boldsymbol{E}=\{E_1,\cdots,E_m\}$ 上の分布 $\boldsymbol{q}^0=(q_1^0,\cdots,q_m^0)$ に従う n 回の独立な試行によって，$E_1,\cdots,E_m$ がそれぞれ $N_1,\cdots,N_m$ 回，$(n=N_1+\cdots+N_m)$ 起ったとする．そのとき $\boldsymbol{E}$ 上の分布（確率変数）

$$\boldsymbol{P}=\left(\frac{N_1}{n},\cdots,\frac{N_m}{n}\right) \tag{12}$$

とおくとき，不変数 α の正則情報量 I に対して，$n\to\infty$ のとき確率変数

$$\frac{2n}{\alpha}I(\boldsymbol{P}\,;\,\boldsymbol{q}^0) \Longrightarrow \chi_{m-1}^2, \qquad \text{（法則収束）} \tag{13}$$

である．

$I=I_P=I^1$, $\alpha=2$ の場合が Pearson の定理である．

証明 $$X_k=\frac{1}{\sqrt{n}}\frac{N_k-nq_k^0}{\sqrt{q_k^0}}, \qquad Z_k=\frac{1}{\sqrt{n}}X_k=\frac{1}{\sqrt{q_k^0}}\left(\frac{N_k}{n}-q_k^0\right), \qquad k=1,\cdots,m$$

とおく．いま $\varepsilon=n^{-1/2+1/10}$ とおくと

$$|Z_k|<\varepsilon \iff |X_k|<n^{1/10}$$

である．ここで，$|X_k|<n^{1/10}$, $k=1,\cdots,m$ が同時に起るという事象を E とおくと，

$$E^c=\{|X_1|\geqq n^{1/10}\}\cup\cdots\cup\{|X_m|\geqq n^{1/10}\}$$

である．中心極限定理によって，任意の $\varepsilon'>0$ に対して $n\geqq n_0(\varepsilon')$ に対して，

$$\operatorname{prob}\{|X_k|>a\}\leqq\sqrt{\frac{2}{\pi}}\int_a^\infty e^{-x^2/2}dx+R_n\leqq\sqrt{\frac{2}{\pi}}\,\frac{1}{a}e^{-a^2/2}+R_n,\qquad |R_n|<\frac{\varepsilon'}{2m}$$

となる．ここで，$a=n^{1/10}$ にとれば，$n\geqq n_1(\varepsilon')$ に対して，

$$\operatorname{prob}\{|X_k|>n^{1/10}\}\leqq\frac{1}{m}\varepsilon'$$

となる．よって

$$\operatorname{prob}(E)\geqq 1-\varepsilon' \tag{14}$$

となる．一方，事象 E においては $|Z_k|<\varepsilon$, $\varepsilon=n^{-1/2+1/10}$, $k=1,\cdots,m$ が成り立つから，(5), (6) によって

$$I(\boldsymbol{P};\ \boldsymbol{q}^0)=\frac{\alpha}{2}\sum_{k=1}^m Z_k^2+O(\varepsilon^3)$$

である．故に

$$nI(\boldsymbol{P};\ \boldsymbol{q}^0)=\frac{\alpha}{2}\sum_{k=1}^m X_k^2+nO(\varepsilon^3),\qquad nO(\varepsilon^3)=O(n^{-1/2+3/10}) \tag{15}$$

である．したがって，(11) と (14) とから (13) が導びかれる．

定理 16． $I(\boldsymbol{p};\ \boldsymbol{q})$ を不変数 α の正則情報量とするとき，確率変数 $\boldsymbol{P}$ に関する期待値 E に関して

$$\lim_{n\to\infty}\frac{2n}{\alpha}\mathrm{E}(I(\boldsymbol{P};\ \boldsymbol{q}^0))=m-1 \tag{16}$$

が成り立つ．

証明　$\mathrm{E}(X_1^2+\cdots+X_m^2)=\sum_{k=1}^m(1-q_k^0)=m-1$ である．よって

$$\mathrm{E}\left(nI(\boldsymbol{P};\ \boldsymbol{q}^0)-\frac{\alpha}{2}\sum_{k=1}^m X_k^2\right)=R_n$$

とおくとき，$\lim_{n\to\infty}R_n=0$ を見ればよい．事象 E を定理 15 の証明にあるように，$E=\bigcap_{k=1}^m\{|X_k|<n^{1/10}\}$ とおく．

$$\mathrm{E}\left(nI(\boldsymbol{P};\ \boldsymbol{q}^0)-\frac{\alpha}{2}\sum_{k=1}^m X_k^2\right)=\int_E+\int_{E^c}$$

とおくとき，第 1 項については (15) より

$$\left|\int_E\right|=O(n^{-1/2+3/10})$$

となる．また第 2 項については，正則情報量の条件 (B) より

$$\left|\int_{E^c}nI(\boldsymbol{P};\ \boldsymbol{q}^0)dp\right|\leqq n\cdot c(\boldsymbol{q}^0)\cdot m\ \operatorname{prob}\{|X_1|>n^{1/10}\}$$

および

$$\left|\int_{E^c}\frac{\alpha}{2}\sum_{k=1}^{m}X_k^2dp\right|\leqq\frac{m\alpha}{2}\int_{|X_1|>n^{1/10}}X_1^2dp+\frac{1}{2}m(m-1)\alpha\cdot n^{2/10}\ \mathrm{prob}\ \{|X_1|>n^{1/10}\}$$

と評価される． ここに，$E^c=\{|X_1|>n^{1/10}\}\cup\bigcup_{k=2}^{m}\{|X_1|\leqq n^{1/10}\}\cap\{|X_k|>n^{1/10}\}$ を用いた．したがって，$\lim R_n=0$ を証明するためには

$$\mathrm{prob}\ (|X_1|>n^{1/10})=O(n^{-3/2}) \tag{17}$$

および

$$\int_{|X_1|>n^{1/10}}X_1^2dp=O(n^{-1/2}) \tag{18}$$

が成り立つことを見ればよい．すなわち

補題 7. 分布 $\boldsymbol{q}^0$ に従う n 回の独立試行によって事象 E_1 $(p(E_1)=q^0)$ が N_1 回起ったとすれば，確率変数

$$X_1=\frac{1}{\sqrt{n}}\ \frac{N_1-nq^0}{\sqrt{q^0}},\qquad \mathrm{E}(X_1)=0,\qquad \sigma^2(X_1)=1-q^0 \tag{19}$$

に対して，n が十分大きいとき，(17), (18) が成り立つ．

証明 次の Cramér の定理（Cramér (1937), p. 81, Theorem 25）を用いる．

〝$Y_1, Y_2,\cdots$ が同一の分布関数 $F(x)$ をもつ互いに独立な確率変数で，それらの平均値は 0, 分散が σ^2, かつ s 次の絶対モーメント $(s\geqq3)$ が有限であるとする． さらに $F(x)$ の絶対連続部分が 0 でないとき，$\mathscr{F}_n(x)$ を $(Y_1+\cdots+Y_n)/\sigma\sqrt{n}$ の分布関数とすれば

$$\mathscr{F}_n(x)=\Phi(x)+\sum_{r=1}^{s-3}\frac{p_{3r-1}(x)}{n^{r/2}}e^{-x^2/2}+R_{sn}(x).$$

ここに， $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-x^2/2}dx$, $p_t(x)$ は或る t 次多項式，かつ

$$|R_{sn}(x)|<\frac{M}{n^{(s-2)/2}}$$

となる．M は s と $F(x)$ にのみ依存し，n と x とには無関係である．〟

われわれの場合に Y_k は E_1 が起れば $1-q^0$, 起らねば $-q^0$ ととるから，$\sigma=\sqrt{q^0(1-q^0)}$ で，$F(x)$ は $x=1-q^0$ で q^0, $x=-q^0$ で $1-q^0$ の飛びをもつ階段関数となり，上記の Cramér の定理の仮定を満足しない．そこで次のような工夫をする．

$Z_1,\cdots,Z_n$ を $Y_1,\cdots,Y_n$ と独立，かつ相互にも独立な確率変数で，$N(0,\sigma_0^2)$ に従うものとする．そのとき，$Y_1+Z_1,\cdots,Y_n+Z_n$ は平均値が 0, 分散 $\sigma^2=\sigma_0^2+q^0(1-q^0)$ を持つ相互に独立な確率変数である．また (19) の X_1 は

$$X_1=\frac{1}{\sqrt{nq_1^0}}(Y_1+\cdots+Y_n)$$

と表わされる．ここで Cramér の定理において，$Y_1,\cdots,Y_n$ の代りに $Y_1+Z_1,\cdots,Y_n+Z_n$ をとれば，その仮定が成り立ち

(20) $$\frac{1}{\sqrt{n}\cdot\sigma}\{(Y_1+Z_1)+\cdots+(Y_n+Z_n)\}=\frac{\sqrt{q^0}}{\sigma}X_1+\frac{1}{\sqrt{n}\cdot\sigma}(Z_1+\cdots+Z_n)$$

の分布関数 $\mathscr{F}_n(x)$ は，$s=5$ にとるとき

$$\mathscr{F}_n(x)=\Phi(x)+\left(\frac{p_2(x)}{\sqrt{n}}+\frac{p_5(x)}{n}\right)e^{-x^2/2}+O(n^{-3/2})$$

と評価される．また

$$\Phi(-x)=\frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-x^2/2}dx\leqq\frac{1}{\sqrt{2\pi}}\frac{1}{x}e^{-x^2/2},\qquad x>0$$

である．(20) より $\mathscr{F}_n(x)$ はその第1項および第2項の分布関数 $F_1(x)$ と $F_2(x)$ との convolution として表わされる．

$$\mathscr{F}_n(x)=F_1(x)*F_2(x)$$

一般に，$F_2(0)=1/2$ であれば

$$(F_1*F_2)(x)\geqq\frac{1}{2}F_1(x),\qquad 1-(F_1*F_2)(x)\geqq\frac{1}{2}(1-F_1(x))$$

である．よって，$c=\sqrt{q^0/(\sigma_0^2+q^0(1-q^0))}$ とおくとき

$$\begin{aligned}\text{prob }\{|X_1|>n^{1/10}\}&=\text{ prob }\{X_1<-n^{1/10}\}+\text{prob }\{X_1>n^{1/10}\}\\&=F_1(-c\,n^{1/10})+(1-F(c\,n^{1/10}))\leqq2\mathscr{F}_n(-c\,n^{1/10})+2(1-\mathscr{F}_n(c_1n^{1/10}))\\&=2\Phi(-c\,n^{1/10})+2\left(\frac{p_2(-c\,n^{1/10})}{\sqrt{n}}+\frac{p_2(c\,n^{1/10})}{\sqrt{n}}+\frac{p_5(-c\,n^{1/10})}{n}+\frac{p_5(c\,n^{1/10})}{n}\right)\\&\quad\times e^{-n^{2/10}c_1^2/2}+O(n^{-3/2})\end{aligned}$$

が成り立つ．したがって，n が十分大きいとき

$$\begin{aligned}\text{prob }\{|X_1|>n^{1/10}\}&\leqq\sqrt{\frac{2}{\pi}}\frac{1}{cn^{1/10}}e^{-c^2n^{2/10}/2}+O(n^{-1/2+2/10}+n^{-1+5/10})e^{-c^2n^{2/10}/2}\\&\quad+O(n^{-3/2})=O(n^{-3/2})\end{aligned}$$

と評価され (17) が成り立つ．次に

$$\begin{aligned}\int_{|X_1|>n^{1/10}}X_1^2dp&\leqq\sum_{i=1}^\infty n^{2(1+i)/10}\text{ prob }(n^{i/10}<|X|\leqq n^{(1+i)/10})\\&\leqq\sum_{i=1}^\infty n^{(1+i)/5}\text{ prob }\{|X_1|>n^{i/10}\}\leqq M\sum_{i=1}^\infty n^{(1+i)/5}\cdot n^{-3i/2}\\&=Mn^{-3/2+2/5}\frac{1}{1-n^{1/5-3/2}}=O(n^{-1/2})\end{aligned}$$

となって (18) が成り立つ．

5.2 分布族のパラメータの推定

分割 $\boldsymbol{E}=\{E_1,\cdots,E_m\}$ 上の分布 $\boldsymbol{q}=(q_1,\cdots,q_m)$ に関して，$\boldsymbol{q}$ 自身が定められておらず，代りに r 個の連続パラメータを含む分布族

(21) $$\boldsymbol{q}(\theta_1,\cdots,\theta_r)=(q_1(\theta_1,\cdots,\theta_r),\cdots,q_m(\theta_1,\cdots,\theta_r)),\qquad\boldsymbol{\theta}=(\theta_1,\cdots,\theta_r)\in\Omega^{(r)}$$

が定められているとする．一方，この分布族に含まれている $\boldsymbol{E}$ 上の未知の真の分布

$$\boldsymbol{q}^0=(q_1^0,\cdots,q_m^0),\quad \boldsymbol{q}^0=\boldsymbol{q}(\theta_1^0,\cdots,\theta_r^0),\quad \boldsymbol{\theta}^0=(\theta_1^0,\cdots,\theta_r^0)\in\Omega^{(r)}$$

に関して，n 回独立に試行して $E_1,\cdots,E_m$ がそれぞれ $n_1,\cdots,n_m$ 回，$(n=n_1+\cdots+n_m)$ 起ったとする．そのとき

(22) $$\boldsymbol{p}=(p_1,\cdots,p_m),\qquad p_k=\frac{n_k}{n},\qquad k=1,\cdots,m$$

とおいて，$\boldsymbol{p}$ から $\boldsymbol{q}^0=\boldsymbol{q}(\boldsymbol{\theta}^0)$ のパラメータ値 $\boldsymbol{\theta}^0$ を推定(Estimate)する問題を考える．これまでよく用いられる方法として

例 9.　χ^2 最小法

(23) $$\chi^2=nI^1(\boldsymbol{p}\,;\,\boldsymbol{q}(\boldsymbol{\theta}))=\sum_{k=1}^{m}\frac{(p_k-q_k(\boldsymbol{\theta}))^2}{q_k(\boldsymbol{\theta})}=\text{最小}$$

となる $\boldsymbol{\theta}$ の値を求める方法（Cramér (1937), p. 506)．

例 10.　最尤法（Maximum likelihood method）

$$L(\theta_1,\cdots,\theta_r)=q_1(\boldsymbol{\theta})^{n_1}\cdots q_m(\boldsymbol{\theta})^{n_m}=\text{最大}$$

となる θ の値を求める方法．これは，$\boldsymbol{p}$ を固定して

(24) $$\begin{aligned}I^0(\boldsymbol{p}\,;\,\boldsymbol{q}(\boldsymbol{\theta}))&=\sum_{k=1}^{m}p_k\log\frac{p_k}{q_k}=\sum_{k=1}^{m}p_k\log p_k-\sum_{k=1}^{m}p_k\log q_k\\&=\left(\sum_{k=1}^{m}p_k\log p_k\right)-\frac{1}{n}\log L(\theta_1,\cdots,\theta_r)=\text{最小}\end{aligned}$$

ということと同値である．

よって，一般に与えられた分布 $\boldsymbol{p}$ に対して，或る正則情報量 $I(\boldsymbol{p}\,;\,\boldsymbol{q})$ に関して

(25) $$I(\boldsymbol{p}\,;\,\boldsymbol{q}(\boldsymbol{\theta}))=\text{最小}$$

となる $\boldsymbol{\theta}$ の値 $\hat{\boldsymbol{\theta}}$ を求め，この $\hat{\boldsymbol{\theta}}=(\hat{\theta}_1,\cdots,\hat{\theta}_r)$ を $\boldsymbol{p}$ より推定される $\boldsymbol{q}$ の値 $\hat{\boldsymbol{q}}=\boldsymbol{q}(\hat{\boldsymbol{\theta}})$ とする．この方法を I 最小法と呼び，$\hat{\boldsymbol{\theta}}$ を I 最小推定値（I Minimum estimate）という．今後は $q_1(\boldsymbol{\theta}),\cdots,q_m(\boldsymbol{\theta})$ は $(\theta_1,\cdots,\theta_r)$ に関して 3 回連続的可微分とする．したがって，$I(\boldsymbol{p}\,;\,\boldsymbol{q}(\boldsymbol{\theta}))$ も $(\theta_1,\cdots,\theta_r)$ に関して 3 回連続的可微分である．さて

(26) $$p_k=q_k^0+u_k=q_k^0\left(1+\frac{x_k}{\sqrt{q_k^0}}\right),\qquad k=1,\cdots,m$$

とおき

(27) $$|\theta_j-\theta_j^0|<\varepsilon,\quad |x_k|<\varepsilon,\qquad j=1,\cdots,r\,;\ k=1,\cdots,m$$

の範囲で考える．

$$q_k=q_k^0+\sum_{j=1}^{r}q_k^{(j)0}\,(\theta_j-\theta_j^0)+O(\varepsilon^2)$$

と展開する．但し

$$q_k^{(j)0}=\left(\frac{\partial q_k}{\partial\theta_j}\right)_{\theta=\theta^0}$$

とする．また

$$q_k = q_k^0 + v_k, \qquad k=1,\cdots,m$$

とおく．

$$I(\boldsymbol{p}\,;\ \boldsymbol{q}(\boldsymbol{\theta})) = F(x_1,\cdots,x_m,\ \theta_1,\cdots,\theta_r)$$

を，正則情報量の性質（A）と（3）の展開とによって考察しよう．与えられた $\boldsymbol{p}$ に対して，$I(\boldsymbol{p}\,;\ \boldsymbol{q}(\boldsymbol{\theta}))$＝最小 を求めるに当たって（簡単のため，$q_1,\cdots,q_m$ の順序を並べかえて）$\boldsymbol{q}$ が $\boldsymbol{q}^0$ の近傍で

$$\frac{\partial(q_1,\cdots,q_r)}{\partial(\theta_1,\cdots,\theta_r)} \neq 0$$

であると仮定する．まず I＝最小 の必要条件として

$$\frac{\partial I}{\partial\theta_j} = 0, \qquad j=1,\cdots,r \tag{28}$$

を解く．その解として $(\hat\theta_1,\cdots,\hat\theta_r)$ が

$$\hat\theta_j = \hat\theta_j(x_1,\cdots,x_m), \qquad j=1,\cdots,r \tag{29}$$

のように得られたとする．

補題 8.　r 次行列

$$\boldsymbol{I} = \left(\left(\frac{\partial^2 I}{\partial\theta_i\partial\theta_j}\right)^0\right)_{i,\,j=1,\cdots,r} \tag{30}$$

は正の定値行列である．

$\alpha^{-1}\boldsymbol{I}$ は Fisher の情報量行列に当たる．(31) に見るように正則情報量 I に対して，$\alpha^{-1}\boldsymbol{I}$ は分布族 $\Omega^{(r)}$ に対して定まり，I に無関係となる．

証明　(27) の範囲で $|u_k| = O(\varepsilon)$, $|v_k| = O(\varepsilon)$, $k=1,\cdots,m$ となるから，I は(3)の形に展開される．したがって

$$\frac{\partial I}{\partial\theta_2} = \frac{\alpha}{2}\sum_{k=1}^{m}\frac{(-1)}{q_k^0}\cdot 2(u_k - v_k)\frac{\partial v_k}{\partial\theta_i} + \frac{\partial R}{\partial\theta_i}, \qquad \frac{\partial v_k}{\partial\theta_i} = \frac{\partial q_k}{\partial\theta_i}$$

$$\frac{\partial^2 I}{\partial\theta_i\partial\theta_j} = \alpha\sum_{k=1}^{m}\frac{(-1)}{q_k^0}\left\{(u_k - v_k)\frac{\partial^2 q_k}{\partial\theta_i\partial\theta_j} - \frac{\partial q_k}{\partial\theta_i}\frac{\partial q_k}{\partial\theta_j}\right\} + \frac{\partial^2 R}{\partial\theta_i\partial\theta_j}$$

となる．かつ

$$\left(\frac{\partial R}{\partial\theta_i}\right)^0 = 0, \qquad \left(\frac{\partial^2 R}{\partial\theta_i\partial\theta_j}\right)^0 = 0, \qquad i,\ j=1,\cdots,r$$

である．ここで，$\boldsymbol{p}=\boldsymbol{q}=\boldsymbol{q}^0$ とすれば $u_1=\cdots=u_m=v_1=\cdots=v_m=0$ となるから

$$\left(\frac{\partial^2 I}{\partial\theta_i\partial\theta_j}\right)^0 = \alpha\sum_{k=1}^{m}\frac{1}{q_k^0}\ q_k^{(i)0}q_k^{(j)0}, \qquad i,\ j=1,\cdots,r \tag{31}$$

となる．したがって

$$\sum_{i=1}^{r}\sum_{j=1}^{r}\left(\frac{\partial^2 I}{\partial\theta_i\partial\theta_j}\right)z_iz_j=\alpha\sum_{k=1}^{m}\sum_{i=1}^{r}\sum_{j=1}^{r}\left(\frac{1}{q_k^0}q_k^{(i)0}q_k^{(j)0}z_iz_j\right)$$
$$=\alpha\sum_{k=1}^{m}\left(\sum_{i=1}^{r}\frac{1}{\sqrt{q_k^0}}q_k^{(i)0}z_i\right)^2\geqq 0$$

となる．かつここで $=0$ となるのは

(32) $$\sum_{i=1}^{r}\frac{1}{\sqrt{q_k^0}}q_k^{(i)0}z_i=0,\qquad k=1,\cdots,m$$

となる場合であるが，これを $(z_1,\cdots,z_r)$ の連立一次方程式と見て，その最初の $k=1,\cdots,r$ の係数の行列式は

$$\det\left(\frac{1}{\sqrt{q_i^0}}q_i^{(j)0}\right)_{i,j=1,\cdots,r}=\frac{1}{\sqrt{q_1^0\cdots q_r^0}}\left(\frac{\partial(q_1,\cdots,q_r)}{\partial(\theta_1,\cdots,\theta_r)}\right)^0\neq 0$$

であるから，(32) となるのは $z_1=\cdots=z_r=0$ の場合に限る．したがって，(30) の行列 $\boldsymbol{I}$ は正の定値行列である．

補題 9． $(x_1,\cdots,x_m,\ \theta_1,\cdots,\theta_r)$ が $(0,\cdots,0,\ \theta_1^0,\cdots,\theta_r^0)$ の近傍にあるとき，(26) の $(p_1,\cdots,p_m)$ に対して $I(\boldsymbol{p}\,;\,\boldsymbol{q}(\boldsymbol{\theta}))=$ 最小 となる $\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}=(\hat{\theta}_1,\cdots,\hat{\theta}_r)$ は $\boldsymbol{\theta}^0=(\theta_1^0,\cdots,\theta_r^0)$, $\boldsymbol{x}=(x_1,\cdots,x_m)$ に対して

(33) $$\boldsymbol{Q}=\begin{pmatrix}\dfrac{q_1^{(1)0}}{\sqrt{q_1^0}}&\cdots&\dfrac{q_m^{(1)0}}{\sqrt{q_m^0}}\\ \vdots&&\vdots\\ \dfrac{q_1^{(r)0}}{\sqrt{q_1^0}}&\cdots&\dfrac{q_m^{(r)0}}{\sqrt{q_m^0}}\end{pmatrix},\qquad \boldsymbol{I}=\alpha\boldsymbol{Q}\cdot{}^t\boldsymbol{Q}$$

によって

(34) $${}^t\hat{\boldsymbol{\theta}}={}^t\boldsymbol{\theta}^0+\alpha\cdot\boldsymbol{I}^{-1}\cdot\boldsymbol{Q}\cdot{}^t\boldsymbol{x}+O(\varepsilon^2)$$

と表わされる．

証明 $\dfrac{\partial I}{\partial\theta_i}=0$, $i=1,\cdots,r$ の解 $\hat{\boldsymbol{\theta}}=(\hat{\theta}_1,\cdots,\hat{\theta}_r)$ は $(x_1,\cdots,x_m)$ が $(0,\cdots,0)$ の近傍にあるとき

(35) $$\hat{\theta}_j=\theta_j^0+\sum_{k=1}^{m}b_{jk}\,x_k+\hat{R}_j,\qquad j=1,\cdots,r$$

と表わされる．これを $\dfrac{\partial I}{\partial\theta_i}=0$ に代入すれば

$$u_k=\sqrt{q_k^0}\,x_k,\quad q_k(\hat{\boldsymbol{\theta}})=q_k^0+\sum_{j=1}^{r}q_k^{(j)0}(\hat{\theta}_j-\theta_j)+\hat{S}_k,\qquad k=1,\cdots,m$$

より，$\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}$ において

$$\left(\frac{\partial I}{\partial\theta_i}\right)_{\theta=\hat{\theta}}=\alpha\sum_{k=1}^{m}\frac{1}{q_k^0}\left\{\left(\sqrt{q_k^0}\,x_k-\sum_{j=1}^{r}q_k^{(j)0}(\hat{\theta}_j-\theta_j^0)+\hat{S}_k\right)\left(\frac{\partial q_k}{\partial\theta_i}\right)_{\theta=\hat{\theta}}\right\}+\left(\frac{\partial\hat{R}_j}{\partial\theta_i}\right)=0$$

である．ここで $|x_1|<\varepsilon_1,\cdots,|x_m|<\varepsilon_1$ に対して $\hat{S}_k=O(\varepsilon_1^2)$, $\left(\dfrac{\partial q_k}{\partial\theta_i}\right)_{\theta=\hat{\theta}}=q_k^{(i)0}+O(\varepsilon_1)$, $\left(\dfrac{\partial\hat{R}}{\partial\theta_i}\right)=O(\varepsilon_1^2)$ となるから

$$\sum_{k=1}^{m} \frac{q_k^{(i)0}}{\sqrt{q_k^0}} x_k - \sum_{j=1}^{r} \left(\sum_{k=1}^{m} \frac{q_k^{(j)0} q_k^{(i)0}}{q_k^0} \right) (\hat{\theta}_j - \theta_j^0) + O(\varepsilon_1^2) = 0$$

となる．故に(31)と(35)から

$$\boldsymbol{Q} \cdot {}^t\boldsymbol{x} = \frac{1}{\alpha} \boldsymbol{I} \cdot \boldsymbol{B} \cdot {}^t\boldsymbol{x}, \qquad \boldsymbol{B} = \begin{pmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & & \vdots \\ b_{m1} & \cdots & b_{mm} \end{pmatrix}$$

となる．これを(35)に代入すれば(34)となる．

次に $\boldsymbol{E} = \{E_1, \cdots, E_m\}$ 上の分布 $\boldsymbol{q}^0 = (q_1^0, \cdots, q_m^0)$, $q_k^0 = q_k(\boldsymbol{\theta}^0)$, $k = 1, \cdots, m$ に従う n 回の独立試行によって $E_1, \cdots, E_m$ がそれぞれ $N_1, \cdots, N_m$ 回，$(n = N_1 + \cdots + N_m)$ 起ったとして

(36) $$\boldsymbol{P} = \left(\frac{N_1}{n}, \cdots, \frac{N_m}{n} \right)$$

にとる．ここで確率変数

(37) $$X_k = \frac{1}{\sqrt{n}} \frac{1}{\sqrt{q_k^0}} (N_k - n q_k^0) = \sqrt{n}\, x_k, \qquad k = 1, \cdots, m$$

にとると(8), (10)によって，正方行列 $A = \boldsymbol{E}_m - {}^t\sqrt{\boldsymbol{q}^0} \cdot \sqrt{\boldsymbol{q}^0}$ に対して

$$(X_1, \cdots, X_m) \implies N((0, \cdots, 0), A), \qquad \text{(法則収束)}$$

であった．

補題 10． (37)の確率変数 $(X_1, \cdots, X_n)$ に対して，確率変数 $\tilde{\boldsymbol{\Theta}} = (\tilde{\Theta}_1, \cdots, \tilde{\Theta}_r)$ を

(38) $${}^t\tilde{\boldsymbol{\Theta}} = {}^t\boldsymbol{\theta}^0 + \frac{\alpha}{\sqrt{n}} \boldsymbol{I}^{-1} \cdot \boldsymbol{Q} \cdot {}^t\boldsymbol{X}$$

と定めれば(但し，$\boldsymbol{\theta}^0 = (\theta_1^0, \cdots, \theta_r^0)$, $\boldsymbol{X} = (X_1, \cdots, X_m)$ とおく)，$n \to \infty$ のとき $\sqrt{n}\,(\tilde{\boldsymbol{\Theta}} - \boldsymbol{\theta}^0)$ は r 次元正規分布 $N((0, \cdots, 0), \alpha \boldsymbol{I}^{-1})$ に法則収束する．

証明 $\sqrt{n}\tilde{\boldsymbol{\Theta}}$ の共分散行列を計算すれば

$$n\mathrm{E}\,({}^t(\tilde{\boldsymbol{\Theta}} - \boldsymbol{\theta}^0) \cdot (\tilde{\boldsymbol{\Theta}} - \boldsymbol{\theta}^0)) = \boldsymbol{I}^{-1} \cdot \boldsymbol{Q} \cdot \mathrm{E}\,({}^t\boldsymbol{X} \cdot \boldsymbol{X}) \cdot {}^t\boldsymbol{Q} \cdot \boldsymbol{I}^{-1}$$

ここで，$\mathrm{E}\,({}^t\boldsymbol{X} \cdot \boldsymbol{X}) = \boldsymbol{E}_m - {}^t\sqrt{\boldsymbol{q}^0} \cdot \sqrt{\boldsymbol{q}^0}$ および $\boldsymbol{Q} \cdot {}^t\sqrt{\boldsymbol{q}^0} = 0$ $\left(\because \sum_{k=1}^{m} q_k^{(i)0} = 0 \right)$ および $\boldsymbol{I} = \alpha \boldsymbol{Q} \cdot {}^t\boldsymbol{Q}$ を代入すれば

$$n\,\mathrm{E}\,({}^t(\tilde{\boldsymbol{\Theta}} - \boldsymbol{\theta}^0) \cdot (\tilde{\boldsymbol{\Theta}} - \boldsymbol{\theta}^0)) = \alpha \boldsymbol{I}^{-1}$$

を得る．よって，$\sqrt{n}\,(\tilde{\boldsymbol{\Theta}} - \boldsymbol{\theta}^0)$ は r 次元正規分布 $N((0, \cdots, 0), \alpha \boldsymbol{I}^{-1})$ に法則収束する．

以上の補題を用いれば，定理 15, 16 の証明と同様に次の定理が証明される．

定理 17． 正則情報量 I と，分布族 $\Omega^{(r)} = \{\boldsymbol{q}(\boldsymbol{\theta})\}$ および(36)の $\boldsymbol{P}$ に対して $I(\boldsymbol{P}\,;\,\boldsymbol{q}(\boldsymbol{\theta})) =$ 最小 となる $\hat{\boldsymbol{\Theta}} = (\hat{\Theta}_1, \cdots, \hat{\Theta}_r)$ をとる．$\hat{\boldsymbol{\Theta}} = (\hat{\Theta}_1, \cdots, \hat{\Theta}_r)$ は r 次元確率変数であって，$n \to \infty$ のと

き

$$(39) \qquad \sqrt{n}\,(\hat{\Theta}_1-\theta_1^0,\cdots,\hat{\Theta}_r-\theta_r^0) \implies N((0,\cdots,0),\ \alpha \boldsymbol{I}^{-1}), \qquad \text{(正規分布)}$$

に法則収束する．

補題 11. $\boldsymbol{p}=(p_1,\cdots,p_m)$, $p_k=q_k^0\left(1+\dfrac{x_k}{\sqrt{q_k^0}}\right)$, $k=1,\cdots,m$ とし，$|x_k|<\varepsilon$, $k=1,\cdots,m$ とすれば，

$$(40) \qquad {}^t(\tilde{\theta}_1,\cdots,\tilde{\theta}_r)={}^t(\theta_1^0,\cdots,\theta_r^0)+\boldsymbol{I}^{-1}\cdot\boldsymbol{Q}\cdot{}^t(x_1,\cdots,x_m)$$

に対して，

$$(41) \qquad I(\boldsymbol{p}\,;\ \boldsymbol{q}(\tilde{\boldsymbol{\theta}}))=I(\boldsymbol{p}\,;\ \boldsymbol{q}^0)-\frac{1}{2}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}^0)\cdot\boldsymbol{I}\cdot{}^t(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta})+O(\varepsilon^3)$$

となる．

証明 $I(\boldsymbol{p}\,;\ \boldsymbol{q}(\tilde{\boldsymbol{\theta}}))$

$$=I(\boldsymbol{p}\,;\ \boldsymbol{q}(\boldsymbol{\theta}^0))+\sum_{i=1}^{r}I^{(i)0}(\tilde{\theta}_i-\theta_i^0)+\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}I^{(i,j)0}(\tilde{\theta}_i-\theta_i^0)(\tilde{\theta}_j-\theta_j^0)+O(\varepsilon^3)$$

である．一方

$$\begin{aligned}0=I^{(i)}(\hat{\boldsymbol{\theta}})&=I^{(i)0}+\sum_{j=1}^{r}I^{(i,j)0}(\hat{\theta}_j-\theta_j^0)+O(\varepsilon^2)\\&=I^{(i)0}+\sum_{j=1}^{r}I^{(i,j)0}(\tilde{\theta}_j-\theta_j^0)+O(\varepsilon^2)\end{aligned}$$

である．これを上の式に代入して $I^{(i)0}$ を消去すれば (41) となる．

定理 18. 定理 17 と同じ記号 $\boldsymbol{P}$, $\hat{\boldsymbol{\Theta}}$ を用いるとき，$\boldsymbol{P}$, $\hat{\boldsymbol{\Theta}}$ は確率変数であって，$n\to\infty$ に対して

$$(42) \qquad \frac{2n}{\alpha}(I(\boldsymbol{P}\,;\ \boldsymbol{q}^0)-I(\boldsymbol{P}\,;\ \boldsymbol{q}(\hat{\boldsymbol{\Theta}}))) \implies \chi_r^2,$$

に法則収束する．但し，χ_r^2 は自由度 r の χ^2 分布を示す．また，$\boldsymbol{P}$, $\hat{\boldsymbol{\Theta}}$ を確率変数とみて，期待値をとれば

$$(43) \qquad \lim_{n\to\infty}\frac{2n}{\alpha}\mathrm{E}\,(I(\boldsymbol{P}\,;\ \boldsymbol{q}^0)-I(\boldsymbol{P}\,;\ \boldsymbol{q}(\hat{\boldsymbol{\Theta}})))=r$$

が成り立つ．

証明 いま，$\boldsymbol{X}=(X_1,\cdots,X_m)$ に対して，(33) の行列 $\boldsymbol{Q}$ および $\alpha\boldsymbol{I}^{-1}={}^t\boldsymbol{A}\cdot\boldsymbol{A}$ となる r 次正方行列 $\boldsymbol{A}$ を用いて

$$(44) \qquad {}^t\boldsymbol{I}=\boldsymbol{A}\cdot\boldsymbol{Q}\cdot{}^t\boldsymbol{X}$$

とおく．そのとき (38) の $\tilde{\boldsymbol{\Theta}}$ に対して，$\boldsymbol{I}=\alpha\cdot\boldsymbol{Q}\cdot{}^t\boldsymbol{Q}$ を用いれば

$$\begin{aligned}\frac{n}{\alpha}(\tilde{\boldsymbol{\Theta}}-\boldsymbol{\theta}^0)\cdot\boldsymbol{I}\cdot{}^t(\tilde{\boldsymbol{\Theta}}-\boldsymbol{\theta}^0)&=\alpha\,{}^t(\boldsymbol{I}^{-1}\cdot\boldsymbol{Q}\cdot{}^t\boldsymbol{X})\cdot\boldsymbol{I}\cdot(\boldsymbol{I}^{-1}\cdot\boldsymbol{Q}\cdot{}^t\boldsymbol{X})\\&={}^t(\boldsymbol{A}\cdot\boldsymbol{Q}\cdot{}^t\boldsymbol{X})\cdot(\boldsymbol{A}\cdot\boldsymbol{Q}\cdot{}^t\boldsymbol{X})={}^t\boldsymbol{Z}\cdot\boldsymbol{Z}\end{aligned}$$

となる．一方，$\boldsymbol{Q}\cdot{}^t\sqrt{\boldsymbol{q}^0}=0$ であるから

$$\begin{aligned}\mathrm{E}\,({}^t\boldsymbol{Z}\cdot\boldsymbol{Z}) &= \mathrm{E}\,((\boldsymbol{A}\cdot\boldsymbol{Q}\cdot{}^t\boldsymbol{X})(\boldsymbol{X}\cdot{}^t\boldsymbol{Q}\cdot{}^t\boldsymbol{A}))\\ &= \boldsymbol{A}\cdot\boldsymbol{Q}\cdot(\boldsymbol{E}_m-{}^t\sqrt{\boldsymbol{q}^0}\cdot\sqrt{\boldsymbol{q}^0})\cdot{}^t\boldsymbol{Q}\cdot{}^t\boldsymbol{A}\\ &= \boldsymbol{A}\cdot\boldsymbol{Q}\cdot{}^t\boldsymbol{Q}\cdot{}^t\boldsymbol{A}=\frac{1}{\alpha}\boldsymbol{A}\cdot\boldsymbol{I}\cdot{}^t\boldsymbol{A}=\boldsymbol{A}\cdot({}^t\boldsymbol{A}\cdot\boldsymbol{A})^{-1}\cdot\boldsymbol{A}=\boldsymbol{E}_r\end{aligned}$$

となる．よって $n\to\infty$ のとき，$\boldsymbol{Z}$ は r 次元正規分布 $N((0,\cdots,0),\ \boldsymbol{E}_r)$ に法則収束する．したがって，$Z_1^2+\cdots+Z_r^2$ は $n\to\infty$ のとき自由度 r の χ^2 分布に法則収束する．さて補題 11 によって

$$(45)\qquad \frac{2n}{\alpha}(I(\boldsymbol{P}\,;\ \boldsymbol{q}^0)-I(\boldsymbol{P}\,;\ \boldsymbol{q}(\tilde{\boldsymbol{\Theta}})))=\frac{n}{\alpha}(\tilde{\boldsymbol{\Theta}}-\boldsymbol{\theta}^0)\cdot\boldsymbol{I}\cdot{}^t(\tilde{\boldsymbol{\Theta}}-\boldsymbol{\theta}^0)+O(\varepsilon^3)$$
$$=(Z_1^2+\cdots+Z_r^2)+O(\varepsilon^3)$$

と表わされるから定理 17 と同様に $n\to\infty$ のとき，左辺の確率変数は自由度 r の χ^2 分布に法則収束する．また (43) も定理 16 の証明と同様に証明される．

定理 19． 確率変数

$$(46)\qquad \frac{2n}{\alpha}I(\boldsymbol{P}\,;\ \boldsymbol{q}(\hat{\boldsymbol{\Theta}}))$$

は，$n\to\infty$ のとき自由度 $m-1-r$ の χ^2 分布に法則収束する ($I=I_P$ の場合は，$\alpha=1$ で，この定理は Fisher によって与えられている．Cramér (1937), p. 424 参照)．

証明 (44) で与えた $Z_1,\cdots,Z_r$ に対しては，$\boldsymbol{Q}\cdot{}^t\sqrt{\boldsymbol{q}^0}=0$ より，$\mathrm{E}\left(Z_i\left(\sum_{k=1}^{m}\sqrt{q_k^0}\,X_k\right)\right)=0,\ i=1,\cdots,r$ であるから，同じく $\mathrm{E}\left(Z_l\left(\sum_{k=1}^{m}\sqrt{q_k^0}\ X_k\right)\right)=0,\quad l=r+1,\cdots,m-1$ となる $X_1,\cdots,X_m$ の一次式 $Z_{r+1},\cdots,Z_{m-1}$ を追加して

$$X_1^2+\cdots+X_m^2=Z_1^2+\cdots+Z_{m-1}^2$$

となり，$(Z_1,\cdots,Z_{m-1})$ の分布が $n\to\infty$ のとき，正規分布 $((0,\cdots,0),\ \boldsymbol{E}_{m-1})$ に収束するようにとることができる．定理 15 の証明中の (15) によって，

$$\frac{2n}{\alpha}I(\boldsymbol{P}\,;\ \boldsymbol{q}^0)=X_1^2+\cdots+X_m^2+R_1$$

と表わされるのであったから，定理 18 の証明中の (45) によって

$$\frac{2n}{\alpha}I(\boldsymbol{P}\,;\ \boldsymbol{q}(\hat{\boldsymbol{\Theta}}))=Z_{r+1}^2+\cdots+Z_{m-1}^2+R_2$$

と表わされ，$n\to\infty$ のとき χ^2_{m-r-1} に法則収束する．

次に，真の分布 $\boldsymbol{q}^0$ を含む r 次元パラメータを持つ分布族 $\Omega^{(r)}=\{\boldsymbol{q}(\boldsymbol{\theta})\}$ の他に，s 次元パラメータを持つ分布族 $\Omega_0^{(s)}=\{\boldsymbol{q}(\boldsymbol{\varphi})\}$ があって，

$$\boldsymbol{q}_1^0\in\Omega_0^{(s)}\subset\Omega^{(r)},\qquad s<r$$

としよう．いま $\Omega^{(r)}$ のパラメータ $(\theta_1,\cdots,\theta_r)$ の代りに，或る r 次直交行列 $\boldsymbol{O}$ (${}^t\boldsymbol{O}=\boldsymbol{O}^{-1}$) によって $(\theta_1'-\theta_1^0,\cdots,\theta_r'-\theta_r^0)=(\theta_1-\theta_1^0,\cdots,\theta_r-\theta_r^0)\boldsymbol{O}$ に $(\theta_1',\cdots,\theta_r')$ をとれば

$$\left(\frac{\partial^2 I}{\partial\theta_i\partial\theta_j}\right)^0_{i,j}={}^t\boldsymbol{O}\cdot\left(\frac{\partial^2 I}{\partial\theta'_i\partial\theta'_j}\right)^0\cdot\boldsymbol{O}$$

となる．そこで$I(\boldsymbol{P}\,;\,\boldsymbol{q}(\boldsymbol{\Theta}))=$最小 とする$\hat{\boldsymbol{\Theta}}'$をとれば

$$(\hat{\boldsymbol{\Theta}}'_1-\theta_1^0,\cdots,\hat{\boldsymbol{\Theta}}'_r-\theta_r^0)=(\hat{\boldsymbol{\Theta}}_1-\theta_1^0,\cdots,\hat{\boldsymbol{\Theta}}_r-\theta_r^0)\cdot\boldsymbol{O}$$

である．よって定理17によって，$n\to\infty$のとき

$$\sqrt{n}\,(\hat{\boldsymbol{\Theta}}'_1-\theta_1^0,\cdots,\hat{\boldsymbol{\Theta}}'_r-\theta_r^0)\implies N((0,\cdots,0),\ \alpha\,{}^t\boldsymbol{O}\cdot\boldsymbol{I}\cdot\boldsymbol{O})$$

に法則収束する．

さて，分布族$\Omega^{(r)}$に対して$I(\boldsymbol{P}\,;\,\boldsymbol{q}(\boldsymbol{\theta}))$, $\boldsymbol{q}(\boldsymbol{\theta})\in\Omega^{(r)}$を最小にするパラメータの値を$(\hat{\boldsymbol{\Theta}}_1,\cdots,\hat{\boldsymbol{\Theta}}_r)$とし，$\Omega_0^{(s)}$に対して，$I(\boldsymbol{P}\,;\,\boldsymbol{q}(\boldsymbol{\varphi}))$, $\boldsymbol{q}(\boldsymbol{\varphi})\in\Omega^{(s)}$を最小にするパラメータの値を$(\hat{\boldsymbol{\Phi}}_1,\cdots,\hat{\boldsymbol{\Phi}}_s)$とする．上記直交行列を適当にとることによって，$s$次行列$(I_\varphi^{(i,j)0})_{i,j=1,\cdots,s}$を$r$次行列$(I_\theta^{(i,j)0})_{i,j=1,\cdots,r}$の上左部分になるように変換することができる．このことを用いれば

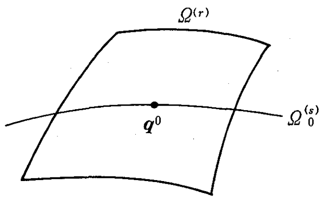

図21．　分布族の集合

定理20．　上記のような分布族$\Omega^{(r)}$, $\Omega_0^{(s)}$に対して

$$\frac{2n}{\alpha}\{I(\boldsymbol{P}\,;\,\boldsymbol{q}(\hat{\boldsymbol{\Theta}}))-I(\boldsymbol{P}\,;\,\boldsymbol{q}(\hat{\boldsymbol{\Phi}}))\}\tag{47}$$

は，$n\to\infty$のとき自由度$r-s$のχ^2分布に法則収束する．

$I=I_{KL}$の場合に，この定理はNeymann-E.S. Pearson-Wilksの結果として知られている（くわしい計算は省く）．Wilks (1962), 13.3.　The Likelihood Ratio Test, p. 402参照．

5.3　AIC

$\boldsymbol{E}=\{E_1,\cdots,E_m\}$上の未知の真の分布$\boldsymbol{q}^0$に関して$n$回の独立試行に対して，$E_1,\cdots,E_m$がそれぞれ$n_1,\cdots,n_m$回，$(n=n_1+\cdots+n_m)$起ったとする．それに対して$\boldsymbol{E}$上の分布を

$$\boldsymbol{p}^0=\left(\frac{n_1}{n},\cdots,\frac{n_m}{n}\right)$$

とおく．一方，$\boldsymbol{E}$上の分布モデルMODとして，r個のパラメータ$(\theta_1,\cdots,\theta_r)$を含む分布族

$$\Omega^{(r)}=\{\boldsymbol{q}(\boldsymbol{\theta})=(q_1(\boldsymbol{\theta}),\cdots,q_m(\boldsymbol{\theta}))\}$$

が与えられているとき，データ$\boldsymbol{p}^0$に対して，パラメータ$\boldsymbol{\theta}$の推定値として$\hat{\boldsymbol{\theta}}$を正則情報量$I$に関して$I$最小推定値とする．そのとき，モデルMODのAICが

$$\mathrm{AIC}\,(\mathrm{MOD})=\frac{2n}{\alpha}I(\boldsymbol{p}^0\,;\,\boldsymbol{q}(\hat{\boldsymbol{\theta}}))+2r\tag{48}$$

と定義される．

いま分布$\boldsymbol{q}^0$に対するモデル$\mathrm{MOD}_1,\cdots,\mathrm{MOD}_H$（それぞれのパラメータ数を$r_1,\cdots,r_H$とする）が与えられているとき，データ$\boldsymbol{p}^0$に関して，各モデルに関する$I$最小推定値をそれぞれ$\hat{\boldsymbol{\theta}}_1,\cdots,\hat{\boldsymbol{\theta}}_H$とする．それらに対して

$$\mathrm{AIC}\,(\mathrm{MOD}_h)=\frac{2n}{\alpha}I(\boldsymbol{p}^0\,;\ \boldsymbol{q}(\hat{\boldsymbol{\theta}}_h))+2r_h\,,\qquad h=1,\cdots,H$$

を計算して，その値が最も小さい h を選んで，MOD_h を採用しようというのが最小 AIC 法である．

もし，MOD_i に対する分布族 $\Omega_i^{(r_i)}$ が MOD_j に対する分布族 $\Omega_j^{(r_j)}$ に含まれるならば（したがって，$r_i<r_j$），当然 $I(\boldsymbol{p}^0\,;\ \boldsymbol{q}(\hat{\boldsymbol{\theta}}_i))\geqq I(\boldsymbol{p}^0\,;\ \boldsymbol{q}(\hat{\boldsymbol{\theta}}_j))$ であるが，一方，$r_i<r_j$ であるから，(48) の AIC に関して，$\mathrm{AIC}\,(\mathrm{MOD}_i)$ と $\mathrm{AIC}\,(\mathrm{MOD}_j)$ の大小については予め知ることができない．

ここで予測の立場で考察する．すなわち，上記の試行と独立に分布 $\boldsymbol{q}^0$ に関して N^* 回独立試行を行い，$E_1,\cdots,E_m$ がそれぞれ $N_1^*,\cdots,N_m^*$ 回，$(N^*=N_1^*+\cdots+N_m^*)$ 起ったとする．

(49) $$\boldsymbol{P}^*=\left(\frac{N_1^*}{n^*},\cdots,\frac{N_m^*}{n^*}\right)$$

を確率変数と見よう．そこで，最初のデータ $\boldsymbol{p}^0$ に対する I 最小推定値 $\hat{\boldsymbol{\theta}}$ を固定して，$\boldsymbol{P}^*$ に関する期待値

(50) $$\mathrm{E}^*\,(I(\boldsymbol{P}^*\,;\ \boldsymbol{q}(\hat{\boldsymbol{\theta}})))$$

を考えれば，これは分布 $\boldsymbol{q}(\hat{\boldsymbol{\theta}})$ に対する平均予測情報量を表わしている．このとき次の定理が成り立つ．

定理 21． 上記記号に関して，

(51) $$\begin{aligned}\mathrm{AIC}&=\frac{2n}{\alpha}I(\boldsymbol{p}^0\,;\ \boldsymbol{q}(\hat{\boldsymbol{\theta}}))+2r\\&=\frac{2n}{\alpha}\mathrm{E}^*\,(I(\boldsymbol{P}^*\,;\ \boldsymbol{q}(\hat{\boldsymbol{\theta}})))+\frac{2n}{\alpha}\{I(\boldsymbol{p}^0\,;\ \boldsymbol{q}^0)-\mathrm{E}^*\,(I(\boldsymbol{P}^*\,;\ \boldsymbol{q}^0))\}+R\end{aligned}$$

となる．ここに $\boldsymbol{R}$ は $\boldsymbol{p}^0$ を確率変数と見るとき

(52) $$\mathrm{E}(\boldsymbol{R})=0$$

となる項である（坂元 他（1983)，pp. 51-53)．

系 未知の分布 $\boldsymbol{q}^0$ に関する H 個のモデル $\mathrm{MOD}_1,\cdots,\mathrm{MOD}_H$ と，データ $\boldsymbol{p}^0$ が与えられているとき，$\mathrm{AIC}\,(\mathrm{MOD}_1),\cdots,\mathrm{AIC}\,(\mathrm{MOD}_H)$ の最小値を与えるモデルは（共通項 $(2n/\alpha)\{I(\boldsymbol{p}^0\,;\ \boldsymbol{q}^0)-\mathrm{E}^*(I(\boldsymbol{P}^*\,;\ \boldsymbol{q}^0))\}$ および $\boldsymbol{p}^0$ に関する期待値が 0 となる変動量を除けば)，平均予測情報量 $\mathrm{E}^*(I(\boldsymbol{P}^*\,;\ \boldsymbol{q}(\hat{\boldsymbol{\theta}}_h)))$, $h=1,\cdots,H$ が最小となるモデル MOD_h のことである．

定理 21 を証明するために次の補題を挙げる．

補題 12． 上記記号 $\boldsymbol{q}^0=\boldsymbol{q}(\boldsymbol{\theta}_0)$, $\boldsymbol{P}^*=\left(\frac{N_1^*}{n^*},\cdots,\frac{N_m^*}{n^*}\right)$, 分布族 $\Omega^{(r)}$ および正則情報量 I に関して，n^* を十分大きくとれば $|\theta_k-\theta_k^0|<\varepsilon$, $k=1,\cdots,m$, $\varepsilon=O(n^{*-1/2})$ に対して

(53) $$\mathrm{E}^*\,(I(\boldsymbol{P}^*\,;\ \boldsymbol{q}(\boldsymbol{\theta})))=\mathrm{E}^*\,(I(\boldsymbol{P}^*\,;\ \boldsymbol{q}^0))+\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\cdot\boldsymbol{I}\cdot{}^t(\boldsymbol{\theta}-\boldsymbol{\theta}_0)+O(n^{*\,-3/2+3/10})$$

が成り立つ.

証明

$$(54)\qquad I(\boldsymbol{P}^*;\ \boldsymbol{q}(\boldsymbol{\theta}))=I(\boldsymbol{P}^*;\ \boldsymbol{q}^0)+\sum_{i=1}^{r}\frac{\partial I}{\partial\theta_i}(\boldsymbol{P}^*;\ \boldsymbol{q}^0)(\theta_i-\theta_i^0)$$
$$+\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}\frac{\partial^2 I}{\partial\theta_i\partial\theta_j}(\boldsymbol{P}^*;\ \boldsymbol{q}^0)(\theta_i-\theta_i^0)(\theta_j-\theta_j^0)+R(\boldsymbol{P}^*;\ \boldsymbol{q}^0)$$

である.一方

$$X_k^*=\frac{\sqrt{n^*}}{\sqrt{q_k^0}}\left(\frac{N_k^*}{n^*}-q_k^0\right),\qquad k=1,\cdots,m$$

に対して

$$u_k^*=\frac{N_k^*}{n^*}-q_k^0=\frac{\sqrt{q_k^0}}{\sqrt{n^*}}X_k^*,\quad v_k^*=q_k-q_k^0,\qquad k=1,\cdots,n$$

とすれば,正則情報量の性質(A)によって,

$$|X_k^*|<n^{1/10},\quad u_k=O(n^{*\,-1/2+1/10}),\quad |\theta_k-\theta_k^0|<O(n^{*-1/2}),\quad v_k=O(n^{*-1/2})$$

のとき

$$I(\boldsymbol{P}^*;\ \boldsymbol{q})=\frac{\alpha}{2}\sum_{k=1}^{m}\frac{1}{q_k^0}\left(\frac{\sqrt{q_k^0}}{\sqrt{n^*}}X_k^*-v_k\right)^2+O(n^{*-3/2+3/10})$$

である.これより

$$\frac{\partial I}{\partial\theta_i}(\boldsymbol{P}^*;\ \boldsymbol{q})=\alpha\sum_{k=1}^{m}\frac{(-1)}{q_k^0}\left(\frac{\sqrt{q_k^0}}{\sqrt{n^*}}X_k^*-v_k\right)\frac{\partial q_k}{\partial\theta_i}+O(n^{*-1+1/5})$$

したがって

$$\frac{\partial I}{\partial\theta_i}(\boldsymbol{P}^*;\ \boldsymbol{q}^0)=-\frac{\alpha}{\sqrt{n^*}}\sum_{k=1}^{m}\frac{q_k^{(i)0}}{\sqrt{q_k^0}}X_k^*+O(n^{*-1+1/5})$$

また

$$\frac{\partial^2 I}{\partial\theta_i\partial\theta_j}(\boldsymbol{P}^*;\ \boldsymbol{q})=-\alpha\sum_{k=1}^{m}\left\{\left(\frac{\sqrt{q_k^0}}{\sqrt{n^*}}X_k^*-v_k\right)\frac{\partial^2 q_k}{\partial\theta_i\partial\theta_j}-\frac{\partial q_k}{\partial\theta_i}\frac{\partial q_k}{\partial\theta_j}\right\}+O(n^{-1/2+1/10})$$

故に

$$\frac{\partial^2 I}{\partial\theta_i\partial\theta_j}(\boldsymbol{P}^*;\ \boldsymbol{q}^0)=\alpha\sum_{k=1}^{m}\frac{1}{q_k^0}q_k^{(i)0}q_k^{(j)0}-\frac{\alpha}{\sqrt{n^*}}\sum_{k=1}^{m}\frac{1}{\sqrt{q_k^0}}q_k^{(i,j)0}X_k^*+O(n^{-1/2+1/10})$$

となる.以上を(54)に代入し,かつ $\mathrm{E}^*(X_k^*)=0$ であるから,定理16の証明と同様に

$$\mathrm{E}^*(I(\boldsymbol{P}^*;\ \boldsymbol{q}(\boldsymbol{\theta})))=\mathrm{E}^*(I(\boldsymbol{P}^*;\ \boldsymbol{q}^0))+\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}I^{(i,j)0}(\theta_i-\theta_i^0)(\theta_j-\theta_j^0)+O(n^{*-3/2+3/10})$$

すなわち,(53)を得る.

補題 13. $\boldsymbol{E}$ 上の分布 $\boldsymbol{q}^0=\boldsymbol{q}(\boldsymbol{\theta}^0)$ に関して n 回の独立試行により $E_1,\cdots,E_m$ がそれぞれ N_1,

$\cdots,N_m$ 回，$(n=N_1+\cdots+N_m)$ 起ったとして，$\boldsymbol{P}=\left(\frac{N_1}{n},\cdots,\frac{N_m}{n}\right)$ とし，これに対する I 最小推定値を $\hat{\boldsymbol{\Theta}}$ とする．但し，I は正則情報量とする．これらを確率変数と見て期待値をとるとき

$$\frac{2n}{\alpha}\{\mathrm{E}^*(I(\boldsymbol{P}^*;\ \boldsymbol{q}(\hat{\boldsymbol{\Theta}})))-\mathrm{E}^*(I(\boldsymbol{P}^*;\ \boldsymbol{q}^0))\} \tag{55}$$

は，$n\to\infty$ のとき自由度 r の χ^2 分布に法則収束する．また

$$\lim_{n\to\infty}\frac{2n}{\alpha}\mathrm{E}\{\mathrm{E}^*(I(\boldsymbol{P}^*;\ \boldsymbol{q}(\hat{\boldsymbol{\Theta}})))-\mathrm{E}^*(I(\boldsymbol{P}^*;\ \boldsymbol{q}^0))\}=r \tag{56}$$

も成り立つ．

証明　補題 12 の結果を用いれば定理 18 の証明と同様である．

定理 21 の証明

$$\begin{aligned}\mathrm{AIC}&=\frac{2n}{\alpha}I(\boldsymbol{p}^0;\ \boldsymbol{q}(\hat{\boldsymbol{\theta}}))+2r\\&=\frac{2n}{\alpha}\mathrm{E}^*(I(\boldsymbol{P}^*;\ \boldsymbol{q}(\hat{\boldsymbol{\theta}})))+\frac{2n}{\alpha}\{I(\boldsymbol{p}^0;\ \boldsymbol{q}^0)-\mathrm{E}^*(I(\boldsymbol{P}^*;\ \boldsymbol{q}^0))\}-R_1-R_2\end{aligned}$$

$$R_1=\frac{2n}{\alpha}(I(\boldsymbol{p}^0;\ \boldsymbol{q}^0)-I(\boldsymbol{p}^0;\ \boldsymbol{q}(\hat{\boldsymbol{\theta}})))-r$$

$$R_2=\frac{2n}{\alpha}(\mathrm{E}^*(I(\boldsymbol{P}^*;\ \boldsymbol{q}(\hat{\boldsymbol{\theta}})))-\mathrm{E}^*(I(\boldsymbol{P}^*;\ \boldsymbol{q}^0)))-r$$

と表わされる．ここで，データ $\boldsymbol{p}^0$ の代りに確率変数 $\boldsymbol{P}$ とおき，I 最小推定値 $\hat{\boldsymbol{\theta}}$ の代りに確率変数 $\hat{\boldsymbol{\Theta}}$ ととれば，定理 18，(43) および補題 13，(56) によって

$$\mathrm{E}(R_1)=\mathrm{E}(R_2)=0$$

を得る．したがって，$\boldsymbol{R}=-R_1-R_2$ ととれば，定理 21 の (51)，(52) を得る．

謝　辞

最後にいろいろとお教えをうけた赤池弘次，工藤弘吉両氏，原稿を閲読して種々の御注意を下さった田原秀敏氏および特に文献の不備を御注意いただいた査読者の方々に厚く御礼を申し上げます．

参 考 文 献

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *2nd Internat. Symp., Information Theory,* Akademiai Kiado, Budapest, 267-281.

Akaike, H. (1974). A new look at the statistical model identification, *IEEE. Trans. Automatic Control,* AC-**19**, 716-723.

Ali, S.M. and Silvey, S.D. (1966). A general class of divergence of one distribution from another, *J. R. Statist. Soc.,* B**28**, 131-142.

Cramér, R. (1937). *Random Variables and Probability Distributions,* Cambridge.

Cramér, R. (1946). *Mathematical Statistics,* Princeton.

Csiszár, I. (1978). Information measures: a critical survey, *Trans. 7th Prague Conference on Informa-*

tion Theory, Statistical Decision, Functions, Random Processes and of the 1974 European Meeting of Statisticians, Vol. B, 73-86.

Fisher, R.A. (1921). On the mathematical foundations of theoretical statistics, *PTRS*, A**222**, 309-368.

Fisher, R.A. (1925). *Statistical Methods for Research Workers,* Oliver and Boyd.

Kakutani, S. (1948). On equivalence of infinite product measures, *Ann. Math.,* **49**, 214-224.

Kudō, H. (1952). A theorem of Kakutani on infinite product measures, *Nat. Sci. Rep. Ochanomizu University,* **3**, 10-22.

工藤弘吉(1953). 時系列および情報の理論とその応用, 第2章 統計的実験とその情報, 日本科学技術連盟, 104-124.

Kullback, S. and Leibler, R.A. (1951). On information and sufficiency, *Ann. Math. Statist.,* **22**, 79-86.

Kullback, S. (1959). *Information Theory and Statistics*, Wiley.

Matusita, K. (1951). On the theory of statistical decision functions, *Ann. Inst. Statist. Math.,* **3**, 17-35.

Pearson, K. (1900). On a criterion that a system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen in random sampling, *Phil. Mag. 50,* ser. 5, **1**, 157-175.

Rathie, P.N. and Kannappan, P.L. (1972). A directed-divergence function of type β^*, *Inform. Control,* **20**, 38-45.

Rényi, A. (1961). On measures of entropy and information, *Proc. Fourth Berkeley Symp. Math. Statist. Prob.,* **1**, 547-561.

坂元慶行, 石黒真木夫, 北川源四郎 (1983). 情報量統計学, 共立出版.

Shannon, C.E. (1948). A mathematical theory of communications, *Bell. System. Tech. J.,* **27**, 379-423, 623-656.

Wilks, S. (1962). *Mathematical Statistics,* Wiley.

Information and Statistics

Yukiyosi Kawada
(The Institute of Statistical Mathematics)

We give firstly a system of axioms for an *information* $I(\boldsymbol{p}\,;\,\boldsymbol{q})$ where $\boldsymbol{p}=(p_1,\cdots,p_m)$, $\boldsymbol{q}=(q_1,\cdots,q_m)$ are any finite probability distributions $(m=1, 2,\cdots)$:

(I) Reducibility: $I(p_1,\cdots,p_{m-1}, 0\,;\, q_1,\cdots,q_{m-1}, 0)=I(p_1,\cdots,p_{m-1}\,;\, q_1,\cdots,q_{m-1})$,

(II) Symmetry: $I(p_1,\cdots,p_m\,;\, q_1,\cdots,q_m)=I(p_{i_1},\cdots,p_{i_m}\,;\, q_{i_1},\cdots,q_{i_m})$,

(III) Non-negativity: $I(\boldsymbol{p}\,;\,\boldsymbol{q})\geqq 0$,

(IV) Invariance and (V) Convexity: $I(p_1+p_2, p_3,\cdots,p_m\,;\, q_1+q_2, q_3,\cdots,q_m)\leqq I(p_1,\cdots, p_m\,;\, q_1,\cdots,q_m)$ and the equality holds if and only if $p_1/q_1=p_2/q_2$ (see 3.1, Definition 4). Several examples such as I_{KL}, I_P, I_K, I^λ are given in Chapter 2.

An information $I(\boldsymbol{p}\,;\,\boldsymbol{q})$ is called *fundamental* if we can express $I(\boldsymbol{p}\,;\,\boldsymbol{q})$ in the form $\sum_{k=1}^{m} L(p_k, q_k)$ by some function $L(x, y)$ (see 3.2, Definition 5). In this case we can express $I(\boldsymbol{p}\,;\,\boldsymbol{q})=\sum_{k=1}^{m} p_k K(q_k/p_k)$ by some non-negative convex function $K(x)$, (see 3.2, Theorem 6). If a fundamental information $I(\boldsymbol{p}\,;\,\boldsymbol{q})$ satisfies the Axiom of Additivity: $I(\boldsymbol{p}_1\otimes\boldsymbol{p}_2\,;\, \boldsymbol{q}_1\otimes\boldsymbol{q}_2)=I(\boldsymbol{p}_1\,;\,\boldsymbol{q}_1)+I(\boldsymbol{p}_2\,;\,\boldsymbol{q}_2)$, then we have $I(\boldsymbol{p}\,;\,\boldsymbol{q})=c_1 I_{KL}(\boldsymbol{p}\,;\,\boldsymbol{q})+c_2 I_{KL}(\boldsymbol{q}\,;\,\boldsymbol{p})$, $c_1, c_2\geqq 0$ by the Kullback-Leibler information I_{KL} (see 3.3, Theorem 7).

The *Liapunov set* $L(\boldsymbol{p}, \boldsymbol{q})$ is defined by $\left\{(x, y)\,\middle|\, x=\sum_{k=1}^{m} \alpha_k p_k,\ y=\sum_{k=1}^{m} \alpha_k q_k,\ 0\leqq\alpha_k\leqq 1, k=1,\cdots,m\right\}$. Then an information $I(\boldsymbol{p}\,;\,\boldsymbol{q})$ is characterized as a non-negative monotone functional for the system of all Liapunov sets (see 4.1, Theorems 11 and 12).

From this characterization we can define several new kinds of informations (see 4.2). Notion of weakly (strongly) complete set of informations are introduced and several examples are given (see 4.3, Theorems 13 and 14).

Finally we give some applications to statistics (see 5.1, 5.2, 5.3). Namely we define a class of *regular* informations (see 5.1, Definition 12), and prove several results, which are well-known for I_{KL}, for general regular informations (see Theorems 15-21).