統計数理（1998）
第 46 巻 第 2 号 461-476

「原著論文」

近似主領域での修正情報量に基づく確率分布間の一様近似と多変量一般指数型分布族の摇動の定量的評価への応用

総合研究大学院大学* 山 田 智 哉
統計数理研究所 松 縄 規

（受付 1997 年 10 月 3 日；改訂 1998 年 7 月 23 日）

要 旨

互いに絶対連続な二つの確率分布間のいくつかの代表的隔たりの尺度を，近似主領域と呼ばれる殆どの確率がそこに分布する領域で定義する．またそれらを利用して分布間の近似を定量的に評価するための基礎理論を展開する．特に，修正 Kullback-Leibler 情報量を隔たりの尺度とする時の近似の評価理論を，いくつかの有用な両側不等式を与えることにより論じる．応用として，多変量一般指数型分布族の正準パラメータまたは分布の平均値が揺らぐ時に，それに伴う密度関数の揺動の大きさを，近似主領域で定義された修正情報量を用いて定量的に評価する．その場合，近似される分布と近似する分布，あるいは統計基礎モデルと発展モデルと言った近似の方向性が意味を持っている．一方，通常の多くの統計理論における様に，近似の方向性を考慮しない場合には，本稿で主として扱っている近似の強さは全変動距離に基づく一様近似に相当する．このことを明示するために，この距離に関連するいくつかの有用な不等式も与える．

キーワード：近似主領域，修正情報量，全変動距離，一様近似，一般指数型分布族．

1. はじめに

本稿の研究目的は，後述の近似主領域での確率測度の近似の概念を積極的に利用した，ある強い意味での一様近似理論の展開を行うことと，その応用として有界区間上での Barron and Sheu (1991) の一変量指数型分布族による近似理論，Matsunawa (1995) の多変量指数型分布族に関連する近似理論を，系統的に改善・拡張することである．

指数型分布族は，統計学の理論において重要な立場を占めており，様々な観点からの研究がなされてきた (cf. Brown (1986))．また，関連諸領域においても，その名称は別として，早くから登場していた．例えば，L. Boltzmann, W. Gibbs, H. Poincaré, J. M. Keynes 等の仕事に密接な関連事項を見ることができる．特に，Gibbs の正準分布，大正準分布は統計力学の分野で重要な役割を演じてきたことは良く知られている．この分布を統計的観点から多変量分布へ一般化したものを考え，そこでのパラメータが変化する時の近似について考察することは理論および応用両面で興味深い．特殊な場合として，Barron and Sheu (1991) は，有界区間上で定

* 数物科学研究科 統計科学専攻：〒106-8569 東京都港区南麻布 4-6-7.

義される一次元指数型分布族について，経験的制約条件の下でK-L情報量最小化法に基づき，有界で線形独立な関数系を用いて，その密度関数に関する近似を考察した．このために，彼らは他の多くの事柄と共に，近似に必要となる情報量に関するいくつかの両側不等式を与えた．そこでのBarron and Sheuの与えた結果を見ると，彼らは述べてはいないが，分布の変化を背後の考察対象の系の物理量の変化としてかなり良く捉えている．これはある点で物理系における変動の大きさの定量的評価が統計的観点からより解釈し易くなったと言える．しかしながら，彼らの問題設定が数学的簡便さから有限区間に限定されていることと，K-L情報量の適用の仕方に物理的には必ずしも自然ではないものがある．そこで本稿では，そのような制限と適用を排除し，近似主領域の概念（cf. Matsunawa (1982, 1986)）に基づいて，多変量一般指数型分布族内で正準パラメータ等が微小変動する時に，密度関数の揺動がどうなるかを，定量的な近似誤差の評価問題として修正・拡張することを試みる．また，情報量の解釈と定義も，Barron and Sheu (1991) と違った，Kullback (1959) の線に沿って行う．これらの改善によって統計理論と関連諸科学の基礎の部分が有機的に整合し，今後のこの方面の近似理論の研究に見通しの良い方向を与えることが期待される．

さて，上記の事柄を議論するにあたり，二つの確率分布間の隔たりを評価することが主要な課題になる．しかしこのことに直接取り掛かる前に，まず広く様々な近似問題の解析にも役立ついくつかの量について，基礎的かつ一般的な設定で考え，近似のために必要な道具立てを整備することから始める．

PとQを測度空間$(R, \boldsymbol{B})$上で定義される確率測度とする．Rは空でない抽象空間，$\boldsymbol{B}$はRの部分集合のσ-集合体とする．本稿の以下の議論ではPとQはある集合$A(\in \boldsymbol{B})$上で互いに絶対連続とする．我々はそこでの二つの分布の隔たりを基に，それらの間のある程度大域的な近似に興味があるものとする．本稿ではこれに取り組む際に，PとQは可測集合A上にその確率測度の殆どを保持している状況：

$$\min\{P(A), Q(A)\} \approx 1, \qquad A(\in \boldsymbol{B}) \tag{1.1}$$

を想定する．そのような状況は，各種の中心極限定理や二項分布のポアソン近似など様々な近似問題の本質的部分に内在している．なお，上の条件(1.1)は次章において数学的により扱いやすい定量的な表現で与えられる．

次に，(1.1)に加えて，集合A上でのPとQの隔たりを評価するために次の諸量を考える(cf. Matsunawa (1982))：

$$\begin{aligned} &I^*(Q, P; A) := \int_A \left(\ln \frac{dQ}{dP}\right) dQ, \quad \rho^*(P, Q; A) := \int_A (dP \cdot dQ)^{1/2}, \\ &W^*(Q, P; A) := \int_A \frac{(dQ - dP)^2}{dQ}. \end{aligned} \tag{1.2}$$

これらの量は我々の目的に非常に有用で，例えば，$I^*(Q, P; A) \approx 0$でかつ条件(1.1)が成立するなら，PとQは集合A上のみならずR上においても，絶対誤差の意味において，互いに十分近い分布であることが言える．即ち，上記の観点からの近似を考える長所は，二分布がたとえ余集合$R-A$で相対的にかなり食い違っていても，(1.2)の諸量を評価することにより，PとQの間のかなり強い意味での近似を理論的に扱え，定量的な近似誤差評価も相当精密に行えることである．このことを明確に扱うために次の概念を導入する．「与えられた微小な数$\varepsilon > 0$に対して可測集合A_εが存在して$\min\{P(A_\varepsilon), Q(A_\varepsilon)\} > 1 - \varepsilon$となるとき，$A_\varepsilon$を確率分布$P, Q$の**近似主領域**（approximate main domain）と呼ぶ．」ここでεをどう与えるか，A_εをどう構成するかは，考察する近似問題毎に工夫が必要になる．上記の近似主領域での修正情報量に基づく近似は次の一様距離

$$D(Q, P; \boldsymbol{B}) := \sup_{E \in \boldsymbol{B}} |Q(E) - P(E)| \tag{1.3}$$

の強さの意味での近似に関して，上下からの有意義な誤差限界を与える上で有用であることも言える（cf. Matsunawa (1982), Barron and Sheu (1991)，本稿定理 4.3）．ただし修正情報量に基づく議論は，P と Q の非対称性，すなわち基礎分布 Q とその発展分布 P の区別という方向性があるが，(1.3) に基づくものは方向性を持たない．この点で，近似問題に物理的な背景を考える時，修正情報量に基づく近似の方が自然な概念であると言える．なお，特に $A=R$ と取れる場合，(1.2) の諸量はそれぞれ Kullback-Leibler (K-L) 情報量，Matusita の Affinity, Kagan の W-divergence に帰着する．

本稿の構成は以下の通りである．第 2 章で本稿の議論の基本となる若干の補題を与える．第 3 章で多変量一般指数型分布族を定義し，三つの修正 K-L 情報量間に成立する重要な不等式を与える．第 4 章で一般指数型分布族の近似について，修正された Affinity と W-divergence を援用して，修正 K-L 情報量を両側から評価して得た結果を与える．またこの分布族のパラメータの変動に対する K-L 情報量の揺動についても考察する．本稿の標題の一様近似の意味を明示する意味で，上記で触れた (1.3) の方向性を持たない一様距離を評価するために，いくつかの新しい両側不等式を与える．付録として，第 2 章で準備した基本的不等式 (補題 2.3) の関連事項の証明と数値例等を示す．

2. 補　　題

本章では，修正 K-L 情報量に基づく分布間の理論近似を定量的に評価するためのいくつかの補題を与える．特に，この情報量と前述の他の隔たりの諸量との関係に関する両側不等式を与える．

補題 2.1. 任意の可測集合 $A(\in \boldsymbol{B})$ に対し

$$\rho^*(P, Q; A) \geq \max\left\{P(A)\cdot\exp\left[\frac{-1}{2P(A)}I^*(P, Q; A)\right],\ Q(A)\cdot\exp\left[\frac{-1}{2Q(A)}I^*(Q, P; A)\right]\right\} \tag{2.1}$$

$$\geq \max\left\{P(A) - \frac{1}{2}I^*(P, Q; A),\ Q(A) - \frac{1}{2}I^*(Q, P; A)\right\} \tag{2.2}$$

が成立する．特に A として

$$B_\varepsilon = \{x : |\ln(dP(x)/dQ(x))| \leq \varepsilon,\quad x \in R\} \tag{2.3}$$

を考える時，次の結果が成立する：

$$\rho^*(P, Q; B_\varepsilon) \geq e^{-\varepsilon/2}\max\{P(B_\varepsilon), Q(B_\varepsilon)\}, \tag{2.4}$$

$$\min\{P(B_\varepsilon), Q(B_\varepsilon)\} \geq e^{-\varepsilon}\max\{P(B_\varepsilon), Q(B_\varepsilon)\}, \tag{2.5}$$

$$|P(B_\varepsilon) - Q(B_\varepsilon)| \leq \max\{P(B_\varepsilon), Q(B_\varepsilon)\}\cdot(1 - e^{-\varepsilon}) \leq 1 - e^{-\varepsilon}, \tag{2.6}$$

$$|P(B_\varepsilon) - Q(B_\varepsilon)| \geq \max\{P(B_\varepsilon), Q(B_\varepsilon)\}\cdot \ell(e^{-\varepsilon})\cdot\varepsilon, \tag{2.7}$$

ここで，$\ell(t)$ $(t>0)$ は後述の補題 2.3 の (2.19) 式で与えられる非負単調増加関数である．

注 2.1. 不等式 (2.4)～(2.7) に量 $\max\{P(B_\varepsilon), Q(B_\varepsilon)\}$ が共通して現われている．即ち B_ε を基

盤にして分布間の近似を考える時，$\max\{P(B_\varepsilon), Q(B_\varepsilon)\}\approx 1$ であるか否かを吟味すれば十分であることが分かる．B_ε が例えば分布 P の support である場合，このことは自動的に満たされることになり，近似問題は分布間の隔たり，例えば修正情報量 $I^*(Q, P; B_\varepsilon)$，の精密な評価に集中できることになる．なお，一般に条件

$$\max\{P(B_\varepsilon), Q(B_\varepsilon)\}\geq e^{-\varepsilon^\#}, \qquad (\varepsilon^\#\geq 0) \tag{2.8}$$

を考えると問題が扱い易くなることがある．この場合 ε が微小量のとき $\varepsilon^\#=\varepsilon^\#(\varepsilon)$ も微小であれば考察対象とする近似にとって好都合になることが普通である．

補題 2.1 の証明． 不等式(2.1)は Jensen の不等式を利用して

$$\begin{aligned}\rho^*(P, Q; A)&=\int_A (dP\cdot dQ)^{1/2}=Q(A)\cdot\exp\left[\ln\left\{\int_A\left(\frac{dP}{dQ}\right)^{1/2}d\left\{\frac{Q}{Q(A)}\right\}\right\}\right]\\ &\geq Q(A)\cdot\exp\left[\frac{1}{2Q(A)}\int_A \ln\frac{dP}{dQ}dQ\right]=Q(A)\cdot\exp\left[\frac{-1}{2Q(A)}I^*(Q, P; A)\right]\end{aligned}$$

さらに，不等式(2.2)は一般に成り立つ関係 $e^{-t}>1-t\,(t:\mathrm{real})$ を用いて，

$$\geq Q(A)-\frac{1}{2}\int_A \ln\frac{dQ}{dP}dQ$$

と求まる．同様にして

$$\rho^*(P, Q; A)\geq P(A)\cdot\exp\left[\frac{-1}{2P(A)}I^*(P, Q; A)\right]\geq P(A)-\frac{1}{2}\int_A \ln\frac{dP}{dQ}dP.$$

よって(2.1)，(2.2)が従う．

不等式(2.4)は，$A=B_\varepsilon$ として

$$I^*(Q, P; B_\varepsilon)\leq\int_{B_\varepsilon}|\ln(dQ(x)/dP(x))|dQ(x)\leq\varepsilon Q(B_\varepsilon),$$
$$I^*(P, Q; B_\varepsilon)\leq\int_{B_\varepsilon}|\ln(dP(x)/dQ(x))|dP(x)\leq\varepsilon P(B_\varepsilon)$$

であるから，これらと不等式(2.1)を利用して，次を得る：

$$\rho^*(P, Q; B_\varepsilon)\geq\max\{P(B_\varepsilon)e^{-\varepsilon/2}, Q(B_\varepsilon)e^{-\varepsilon/2}\}=e^{-\varepsilon/2}\max\{P(B_\varepsilon), Q(B_\varepsilon)\}.$$

不等式(2.5)は Cauchy-Schwarz の不等式と(2.4)を用いて次のように求まる．

$$\begin{aligned}\min\{P(B_\varepsilon), Q(B_\varepsilon)\}&=\frac{P(B_\varepsilon)\cdot Q(B_\varepsilon)}{\max\{P(B_\varepsilon), Q(B_\varepsilon)\}}\geq\frac{\{\rho^*(P, Q; B_\varepsilon)\}^2}{\max\{P(B_\varepsilon), Q(B_\varepsilon)\}}\\ &\geq\frac{\{e^{-\varepsilon/2}\max\{P(B_\varepsilon), Q(B_\varepsilon)\}\}^2}{\max\{P(B_\varepsilon), Q(B_\varepsilon)\}}=e^{-\varepsilon}\max\{P(B_\varepsilon), Q(B_\varepsilon)\}.\end{aligned}$$

不等式(2.6)は

$$|P(B_\varepsilon)-Q(B_\varepsilon)|=\max\{P(B_\varepsilon), Q(B_\varepsilon)\}-\min\{P(B_\varepsilon), Q(B_\varepsilon)\}$$

不等式(2.5)により

$$\leq\max\{P(B_\varepsilon), Q(B_\varepsilon)\}\cdot(1-e^{-\varepsilon})\leq 1-e^{-\varepsilon}.$$

不等式(2.7)は後述の補題 2.3 を用いて

$$|P(B_\varepsilon)-Q(B_\varepsilon)|=P(B_\varepsilon)\left|\frac{Q(B_\varepsilon)}{P(B_\varepsilon)}-1\right|\geq P(B_\varepsilon)\cdot \ell\left(\inf_{x\in B_\varepsilon}\frac{dQ(x)}{dP(x)}\right)$$
$$\cdot\left|\ln\left(\inf_{x\in B_\varepsilon}\frac{dQ(x)}{dP(x)}\right)\right|\geq P(B_\varepsilon)\cdot \ell(e^{-\varepsilon})\cdot\varepsilon,$$

同様に

$$|P(B_\varepsilon)-Q(B_\varepsilon)|=Q(B_\varepsilon)\left|\frac{P(B_\varepsilon)}{Q(B_\varepsilon)}-1\right|\geq Q(B_\varepsilon)\cdot \ell(e^{-\varepsilon})\cdot\varepsilon.$$

よって，この二つの不等式から所要の結果を得る．□

系 2.1. 上記の補題の(2.1)式から次の不等式が得られる：

$$I^*(P,Q;A)\geq -2P(A)\cdot\ln\frac{\rho^*(P,Q;A)}{P(A)},\ I^*(Q,P;A)\geq -2Q(A)\cdot\ln\frac{\rho^*(P,Q;A)}{Q(A)}, \tag{2.9}$$

$$I^*(P,Q;A)\geq -2P(A)\cdot\ln\frac{\sqrt{P(A)\cdot Q(A)}}{P(A)}=P(A)\cdot\ln\frac{P(A)}{Q(A)}, \tag{2.10}$$

$$I^*(P,Q;A)\geq P(A)\left(1-\frac{Q(A)}{P(A)}\right)=P(A)-Q(A), \tag{2.11}$$

$$I^*(Q,P;A)\geq -2Q(A)\cdot\ln\frac{\sqrt{P(A)\cdot Q(A)}}{Q(A)}=Q(A)\cdot\ln\frac{Q(A)}{P(A)}. \tag{2.12}$$

補題 2.2. 修正 K-L 情報量 $I^*(Q,P;A)$ と修正 Kagan's W-divergence $W^*(Q,P;A)$ の間に次の不等式関係が成立する（cf. Matsunawa (1982, 1986)）：

$$I^*(Q,P;A)\leq Q(A)-P(A)+\frac{1}{2}\left(1+\frac{1}{3}\sup_{x\in R}\left|\frac{dQ}{dP}(x)-1\right|\right)W^*(Q,P;A), \tag{2.13}$$

および

$$I^*(Q,P;A)\geq Q(A)-P(A)+\frac{1}{2}\left(1-\frac{1}{3}\sup_{x\in R}\left|\frac{dQ}{dP}(x)-1\right|\right)W^*(Q,P;A). \tag{2.14}$$

次の補題は，修正 K-L 情報量，W-divergence 等の定量的評価に有効である：

補題 2.3. $t>0$, α: real $(\neq 0)$ に対し

$$\ln t=(t-1)\left(\frac{2}{t^{1/\alpha}+1}\right)^{\alpha}+O\left(\frac{\alpha^2(3-\alpha)}{3}\left(\frac{t^{1/\alpha}-1}{t^{1/\alpha}+1}\right)^3\right),\qquad (\alpha\neq 3), \tag{2.15}$$

$$\ln t=(t-1)\left(\frac{2}{t^{1/3}+1}\right)^{3}+O\left(\frac{6}{5}\left(\frac{t^{1/3}-1}{t^{1/3}+1}\right)^5\right). \tag{2.16}$$

$$\ell(t)|\ln t|\leq|t-1|\leq u(t)|\ln t|,\qquad (\text{equality signs}\iff t=1), \tag{2.17}$$

ここに $u(t)$, $\ell(t)$ は以下に与えられる非負単調増加関数である：

$$u(t)=\frac{35(1+t^{1/3})^5(1+t^{1/3}+t^{2/3})(1+16t^{1/3}+t^{2/3})}{457+5014t^{1/3}+14471t^{2/3}+20596t+14471t^{4/3}+5014t^{5/3}+457t^2}, \tag{2.18}$$

(2.19)

$$\ell(t)=\frac{210t^{1/3}(1+t^{1/3})^7(1+t^{1/3}+t^{2/3})}{35+1832t^{1/3}+7796t^{2/3}+18968t+23378t^{4/3}+18968t^{5/3}+7796t^2+1832t^{7/3}+35t^{8/3}}.$$

不等式(2.17)は全変動距離を修正 K-L 情報量で評価するために，Matsunawa (1986) で与えられた．そこでの $u(t)$, $\ell(t)$ はそれぞれ上記の(2.18)，(2.19)より少し精度の低い形で与えられている．証明は全く同じ手法でできるので(2.17)の詳細な証明は省略する．この二つの関数が単調増加関数であることは，t に関する微分を実行することにより確かめ得る．ところで，上記の論文では何故 $u(t)$, $\ell(t)$ に t の立方根に関連する量を使うことが適切なのかの理論的説明はしなかった．その説明を(2.15)，(2.16)の証明と併せて付録1に与える．

3. 多変量一般指数型分布族

Rを実空間，$\boldsymbol{B}$ をRの Borel 集合族とし，$\boldsymbol{A}$ を $R^{n\times n}$ 上で定義される $n\times n$ ランダム行列，$\boldsymbol{u}_i(\boldsymbol{A})(i=1,\dots,m)$ を可測空間 $(R^{n\times n}, \boldsymbol{B}^{n\times n})$ 上で定義される実数行列値関数とする．Pをこの空間上で定義される基準確率測度（統計基礎モデル）とする．Pをどう取るかは重要であるが本稿では与えられたものとして議論を行う（cf. Matsunawa (1986, 1995)）．さて，Pにより誘導される大きさ（dimension）mの多変量一般指数型分布族の密度を次式で定義する：

$$(3.1)\quad dQ_m^*(\boldsymbol{A};\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_m)=\exp\left[-\mathrm{tr}\left\{\sum_{i=1}^m \boldsymbol{\beta}_i^t\boldsymbol{u}_i(\boldsymbol{A})\right\}-\Psi(\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_m)\right]\cdot dP(\boldsymbol{A}),\quad \boldsymbol{A}\in R^{n\times n},$$

但し，$\boldsymbol{\beta}_i$ $(i=1,\dots,m)(\in R^{n\times n})$ は一般正準パラメーター行列，$\boldsymbol{\beta}_i^t$ は $\boldsymbol{\beta}_i$ の転置とし，

$$(3.2)\quad \Psi(\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_m)=\ln Z(\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_m)=\ln\left[\int_{R^{n\times n}}\exp\left\{-\mathrm{tr}\left(\sum_{i=1}^m \boldsymbol{\beta}_i^t\boldsymbol{u}_i(\boldsymbol{A})\right)\right\}\cdot dP(\boldsymbol{A})\right]<\infty$$

はPに関するキュムラント母関数を表す．

注 3.1. Barron and Sheu (1991) は(3.1)においてm母数指数型分布族を扱った．彼等はPを reference distribution と呼び，その確率測度が閉区間 $[0,1]$ にのみ分布するものに限定した．これは関連する数学的な展開を非常に容易なものとした半面，問題の適用範囲をかなり狭くしてしまっている．本稿ではこの制限を弛めると共に，広い分野への適用可能性を考慮して，(3.1)の多変量一般指数型分布族を扱う．なお，Matsunawa (1995) はある規則に従う Legendre 変換を逐次用いてキュムラント母関数が

$$\Psi(\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_m)=\sum_{i=1}^m \Psi_i(\boldsymbol{\beta}_i),\ \text{with}\ \ \Psi_i(\boldsymbol{\beta}_i)=\ln\int_{R^{n\times n}}\exp[-\mathrm{tr}\{\boldsymbol{\beta}_i^t\boldsymbol{u}_i(\boldsymbol{A})\}]dP(\boldsymbol{A})$$

で与えられる多変量指数型分布族を扱った．本稿の議論はこの分布に対しても有効である．

さて，Barron and Sheu (1991) の類推として，$\boldsymbol{u}_i(\boldsymbol{A})$ $(i=1,\dots,m)$ の平均値

$$(3.3)\quad \boldsymbol{U}_i(\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_m)=\int_{R^{n\times n}}\boldsymbol{u}_i(\boldsymbol{A})dQ_m^*=-\frac{\partial}{\partial\boldsymbol{\beta}_i}\Psi(\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_m),\qquad (i=1,\dots,m)$$

が存在するという条件の下で，Q_m^* が同じ統計基礎モデルPを持つ多変量一般指数型分布族の中で次の分布 Q_T へ微小変化した状況を考える：

$$(3.4)\quad dQ_T(\boldsymbol{A};\boldsymbol{\beta}_{1T},\dots,\boldsymbol{\beta}_{mT})=\exp\left[-\mathrm{tr}\left\{\sum_{i=1}^m \boldsymbol{\beta}_{iT}^t\boldsymbol{u}_i(\boldsymbol{A})\right\}-\Psi(\boldsymbol{\beta}_{1T},\dots,\boldsymbol{\beta}_{mT})\right]\cdot dP(\boldsymbol{A}),\quad \boldsymbol{A}\in R^{n\times n}.$$

この様な変動によりひき起こされる揺らぎのためにPと Q_T あるいは Q_m^* の間の修正 K-L 情報

量の大きさがどのように影響されるかを知ることは興味深い．このことに関して次の命題が成立する．

補題 3.1. 近似問題を次の領域

$$C_\varepsilon := \left\{ \boldsymbol{A} ; \left| \ln \frac{dQ_T(\boldsymbol{A})}{dQ_m^*(\boldsymbol{A})} \right| \leq \varepsilon, \quad \boldsymbol{A} \in R^{n\times n} \right\} \tag{3.5}$$

で考える．有界な数 $U_{\varepsilon,m}^*$ が存在して，

$$\left| \mathrm{tr}\left\{ \sum_{i=1}^m \boldsymbol{\beta}_i^t \int_{C_\varepsilon} \boldsymbol{u}_i(\boldsymbol{A}) d(Q_T - Q_m^*) \right\} \right| \leq |U_{\varepsilon,m}^*| \cdot |Q_T(C_\varepsilon) - Q_m^*(C_\varepsilon)|, \tag{3.6}$$

が成立するものとする．このとき三つの修正情報量の間に次の不等式関係が成立する：

$$\begin{aligned} &|I^*(Q_T, P ; C_\varepsilon) - I^*(Q_T, Q_m^* ; C_\varepsilon) - I^*(Q_m^*, P ; C_\varepsilon)| \\ &\qquad \leq (|U_{\varepsilon,m}^*| + |\Psi(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m)|) \cdot e^\varepsilon (e^\varepsilon - 1) . \end{aligned} \tag{3.7}$$

証　明.

$$\begin{aligned} I^*(Q_T, P ; C_\varepsilon) - I^*(Q_T, Q_m^* ; C_\varepsilon) &= \int_{C_\varepsilon} \ln \frac{dQ_T}{dP} dQ_T - \int_{C_\varepsilon} \ln \frac{dQ_T}{dQ_m^*} dQ_T \\ &= \int_{C_\varepsilon} \ln \frac{dQ_m^*}{dP} dQ_T = -\mathrm{tr}\left\{ \sum_{i=1}^m \boldsymbol{\beta}_i^t \int_{C_\varepsilon} \boldsymbol{u}_i(\boldsymbol{A}) dQ_T \right\} - \Psi(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m) Q_T(C_\varepsilon) . \end{aligned}$$

一方，

$$I^*(Q_m^*, P ; C_\varepsilon) = \int_{C_\varepsilon} \ln \frac{dQ_m^*}{dP} dQ_m^* = -\mathrm{tr}\left\{ \sum_{i=1}^m \boldsymbol{\beta}_i^t \int_{C_\varepsilon} \boldsymbol{u}_i(\boldsymbol{A}) dQ_m^* \right\} - \Psi(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m) Q_m^*(C_\varepsilon) .$$

従って

$$\begin{aligned} &I^*(Q_T, P ; C_\varepsilon) - I^*(Q_T, Q_m^* ; C_\varepsilon) - I^*(Q_m^*, P ; C_\varepsilon) \\ &\quad = -\mathrm{tr}\left[\sum_{i=1}^m \boldsymbol{\beta}_i^t \int_{C_\varepsilon} \boldsymbol{u}_i(\boldsymbol{A}) d(Q_T - Q_m^*) \right] - \Psi(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m)(Q_T(C_\varepsilon) - Q_m^*(C_\varepsilon)) . \end{aligned}$$

(3.6) を考慮して

$$\begin{aligned} &|I^*(Q_T, P ; C_\varepsilon) - I^*(Q_T, Q_m^* ; C_\varepsilon) - I^*(Q_m^*, P ; C_\varepsilon)| \\ &\qquad \leq (|U_{\varepsilon,m}^*| + |\Psi(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m)|) \cdot |Q_T(C_\varepsilon) - Q_m^*(C_\varepsilon)| . \end{aligned}$$

(3.5) および (2.5) を考慮すると

$$|Q_T(C_\varepsilon) - Q_m^*(C_\varepsilon)| \leq (e^\varepsilon - 1) \max\{Q_T(C_\varepsilon), Q_m^*(C_\varepsilon)\} \leq e^\varepsilon (e^\varepsilon - 1) \min\{Q_T(C_\varepsilon), Q_m^*(C_\varepsilon)\}$$

が言える．よって所要の (3.7) を得る．□

注 3.2. (3.5) における C_ε の定義で ε は任意の正数を表わしている．大局的に良い近似を得るには $\varepsilon \to 0$ の時 $\min\{Q_T(C_\varepsilon), Q_m^*(C_\varepsilon)\} \to 1$ となることが望まれる．それには二分布 Q_T と Q_m^* のパラメータ間に適切な条件が必要になる．一方，(3.7) から $\varepsilon \to 0$ で $I^*(Q_T, P ; C_\varepsilon) = I^*(Q_T, Q_m^* ; C_\varepsilon) + I^*(Q_m^*, P ; C_\varepsilon)$ となり，修正情報量の間に収支関係が成り立つ．本補題の特別な場合として，$C_\varepsilon = R^{n\times n}$ の時，対応する三つの K-L 情報量の間に収支関係が成り立つ (cf. Matsunawa (1995))．なお，Barron and Sheu (1991) ではここでの関係と双対な $I(P, Q_T) = I(P, Q_m^*) + I(Q_m^*,$

Q_T) が触れられている．この違いは，K-L 情報量の解釈の違いから起きてくる．(3.1)の Q_m^* が正準分布の一般化であり，それに対応する系がどう変化するかを追及する時は，Kullback (1959)に沿った我々の K-L 情報量の与え方の方が，関連する大偏差確率の解釈や物理的観点からも，適切であると思われる(cf. Matsunawa (1995))．しかしそのことについてはより詳細な考察を要するので，別の機会に論じることにする．

4. 多変量一般指数型分布族に関する近似

本章では多変量一般指数型分布に関連する近似を，第2，第3章で準備した補題を利用して行う．なお，前述の K-L 情報量の与え方の差異から，我々の結果は Barron and Sheu (1991)の一変量の場合の並行する結果の直接的な拡張ではないことに注意する．我々の問題への接近法は近似主領域の概念に基づいている点においても彼らのものと異なっている．

定理 4.1. Q_m^* および Q_T をそれぞれ(3.1)および(3.4)で与えられる，基礎モデル P から誘導される多変量一般指数型分布とする．次の領域

$$D_\varepsilon := \left\{ \boldsymbol{A}\,;\max\left(\left| \ln \frac{dQ_T(\boldsymbol{A})}{dQ_m^*(\boldsymbol{A})} \right|,\ \left| \ln \frac{dQ_m^*(\boldsymbol{A})}{dP(\boldsymbol{A})} \right| \right) \le \varepsilon,\quad \boldsymbol{A} \in R^{n\times n} \right\}$$

を考える．この時，補題 3.1 の仮定の下で，次の一連の不等式が成立する：

$$I^*(Q_T, Q_m^*\,; D_\varepsilon) \le \varepsilon, \tag{4.1}$$

$$I^*(Q_T, Q_m^*\,; D_\varepsilon) \ge Q_T(D_\varepsilon)\cdot \ln \frac{Q_T(D_\varepsilon)}{Q_m^*(D_\varepsilon)} \ge Q_T(D_\varepsilon) - Q_m^*(D_\varepsilon) \ge e^{-\varepsilon} - 1, \tag{4.2}$$

$$I^*(Q_m^*, P\,; D_\varepsilon) \le 1 - e^{-\varepsilon} + \frac{1}{2}\{\varepsilon\cdot u(e^\varepsilon)\}^2 + \frac{1}{6}\{\varepsilon\cdot u(e^\varepsilon)\}^3, \tag{4.3}$$

但し $u(\cdot)$ は(2.18)で与えられた関数で，$u(1)=1$.

$$I^*(Q_m^*, P\,; D_\varepsilon) \ge (e^{-\varepsilon} - 1) + \frac{1}{2}\left[\{\ell(e^{-\varepsilon})\}^2 e^{-\varepsilon} P(D_\varepsilon) - \frac{1}{3}\{u(e^\varepsilon)\}^3 \varepsilon \right]\cdot \varepsilon^2, \tag{4.4}$$

ここで $\ell(\cdot)$ は(2.19)式で与えられた関数で，$\ell(1)=1$ である．さらに

$$I^*(Q_T, P\,; D_\varepsilon) \le (1 - e^{-\varepsilon})\{1 + |U_{\varepsilon,m}^*| + |\Psi(\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_m)|\} + \varepsilon + \frac{1}{2}\{\varepsilon\cdot u(e^\varepsilon)\}^2 + \frac{1}{6}\{\varepsilon\cdot u(e^\varepsilon)\}^3, \tag{4.5}$$

$$\begin{aligned} I^*(Q_T, P\,; D_\varepsilon) \ge{}& (e^{-\varepsilon} - 1)\{2 + |U_{\varepsilon,m}^*| + |\Psi(\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_m)|\} \\ &+ \frac{1}{2}\left[\{\ell(e^{-\varepsilon})\}^2 e^{-\varepsilon} P(D_\varepsilon) - \frac{1}{3}\{u(e^\varepsilon)\}^3 \varepsilon \right]\cdot \varepsilon^2 \end{aligned} \tag{4.6}$$

が成立する．

注 4.1. 上記定理で $\varepsilon \to 0$ とすれば，どの修正 K-L 情報量も 0 に収束することが分かる．よって，考慮の対象としている分布が集合 D_ε 上に確率測度の殆どを保持していることが分かれば，考察している分布間の近似は意味のあるものであることが言える．

証 明. 不等式(4.1)は，D_ε の定義から

$$I^*(Q_T, Q_m^*\,; D_\varepsilon) = \int_{D_\varepsilon} \ln \frac{dQ_T}{dQ_m^*} dQ_T \le \int_{D_\varepsilon} \left| \ln \frac{dQ_T}{dQ_m^*} \right| dQ_T \le \varepsilon Q_T(D_\varepsilon) \le \varepsilon.$$

不等式(4.2)は，系 2.1 と補題 2.1 から

$$\begin{aligned}I^*(Q_T, Q_m^*; D_\varepsilon) &\geq Q_T(D_\varepsilon)\cdot\ln\frac{Q_T(D_\varepsilon)}{Q_m^*(D_\varepsilon)} \geq Q_T(D_\varepsilon)-Q_m^*(D_\varepsilon)\\ &\geq \min\{Q_T(D_\varepsilon), Q_m^*(D_\varepsilon)\}-\max\{Q_T(D_\varepsilon), Q_m^*(D_\varepsilon)\}\\ &\geq \max\{Q_T(D_\varepsilon), Q_m^*(D_\varepsilon)\}\cdot(e^{-\varepsilon}-1)\geq e^{-\varepsilon}-1\,.\end{aligned}$$

不等式(4.3)の証明は次の通り．最初に補題 2.3 の不等式(2.17)より

$$\text{(4.7)}\quad \{\ell(e^{-\varepsilon})\cdot\varepsilon\}^2\cdot Q_m^*(D_\varepsilon)\leq W^*(Q_m^*, P; D_\varepsilon)=\int_{D_\varepsilon}\left(\frac{dQ_m^*}{dP}-1\right)^2 dQ_m^*\leq\{u(e^\varepsilon)\cdot\varepsilon\}^2\cdot Q_m^*(D_\varepsilon)\,,$$

および

$$\text{(4.8)}\qquad \sup_{A\in R^{n\times n}}\left|\frac{dQ_m^*(\boldsymbol{A})}{dP(\boldsymbol{A})}-1\right|\leq\sup_{A\in R^{n\times n}}\left[u\left(\left|\frac{dQ_m^*(\boldsymbol{A})}{dP(\boldsymbol{A})}\right|\right)\left|\ln\frac{dQ_m^*(\boldsymbol{A})}{dP(\boldsymbol{A})}\right|\right]\leq u(e^\varepsilon)\cdot\varepsilon$$

に注意する．補題 2.3 の(2.13)より

$$\begin{aligned}I^*(Q_m^*, P; D_\varepsilon) &\leq Q_m^*(D_\varepsilon)-P(D_\varepsilon)+\frac{1}{2}\left(1+\frac{1}{3}\sup_{A\in R^{n\times n}}\left|\frac{dQ_m^*(\boldsymbol{A})}{dP(\boldsymbol{A})}-1\right|\right)W^*(Q_m^*, P; D_\varepsilon)\\ &\leq|Q_m^*(D_\varepsilon)-P(D_\varepsilon)|+\frac{1}{2}\left(1+\frac{1}{3}u(e^\varepsilon)\cdot\varepsilon\right)\{u(e^\varepsilon)\cdot\varepsilon\}^2\cdot Q_m^*(D_\varepsilon)\,.\end{aligned}$$

補題 2.1 の(2.6)から

$$\begin{aligned}&\leq\max\{Q_m^*(D_\varepsilon), P(D_\varepsilon)\}\cdot(1-e^{-\varepsilon})+\frac{1}{2}\left(1+\frac{1}{3}u(e^\varepsilon)\cdot\varepsilon\right)\{u(e^\varepsilon)\cdot\varepsilon\}^2\cdot Q_m^*(D_\varepsilon)\\ &\leq 1-e^{-\varepsilon}+\frac{1}{2}\left(1+\frac{1}{3}u(e^\varepsilon)\cdot\varepsilon\right)\{u(e^\varepsilon)\cdot\varepsilon\}^2\,.\end{aligned}$$

不等式(4.4)は，補題 2.3 の不等式(2.14)および(4.7)，(4.8)を用いる．

$$\begin{aligned}I^*(Q_m^*, P; D_\varepsilon) &\geq Q_m^*(D_\varepsilon)-P(D_\varepsilon)+\frac{1}{2}\left(1-\frac{1}{3}\sup_{A\in R^{n\times n}}\left|\frac{dQ_m^*(\boldsymbol{A})}{dP(\boldsymbol{A})}-1\right|\right)W^*(Q_m^*, P; D_\varepsilon)\\ &\geq-|Q_m^*(D_\varepsilon)-P(D_\varepsilon)|+\frac{1}{2}\left(\{\ell(e^{-\varepsilon})\cdot\varepsilon\}^2-\frac{1}{3}u(e^\varepsilon)\cdot\varepsilon\{u(e^\varepsilon)\cdot\varepsilon\}^2\right)\cdot Q_m^*(D_\varepsilon)\end{aligned}$$

補題 2.1 の不等式(2.6)により

$$\geq\max\{Q_m^*(D_\varepsilon), P(D_\varepsilon)\}\cdot(e^{-\varepsilon}-1)+\frac{1}{2}\left(\{\ell(e^{-\varepsilon})\cdot\varepsilon\}^2-\frac{1}{3}\{u(e^\varepsilon)\cdot\varepsilon\}^3\right)\cdot Q_m^*(D_\varepsilon)\,.$$

領域 D_ε の条件から $e^{-\varepsilon}P(D_\varepsilon)\leq Q_m^*(D_\varepsilon)\leq\min\{e^\varepsilon P(D_\varepsilon), 1\}$ となることに注意して

$$\geq(e^{-\varepsilon}-1)+\frac{1}{2}\left(\{\ell(e^{-\varepsilon})\}^2e^{-\varepsilon}P(D_\varepsilon)-\frac{1}{3}\{u(e^\varepsilon)\}^3\varepsilon\right)\cdot\varepsilon^2\,.$$

不等式(4.5)の証明には，補題 3.1 の不等式(3.7)と，(4.1)，(4.3)，(2.6)を利用する：

$$\begin{aligned}I^*(Q_T, P; D_\varepsilon) &\leq I^*(Q_T, Q_m^*; D_\varepsilon)+I^*(Q_m^*, P; D_\varepsilon)\\ &\qquad+(|U_{\varepsilon,m}^*|+|\Psi(\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_m)|)\cdot|Q_T(D_\varepsilon)-Q_m^*(D_\varepsilon)|\\ &\leq\varepsilon+1-e^{-\varepsilon}+\frac{1}{2}\{\varepsilon\cdot u(e^\varepsilon)\}^2+\frac{1}{6}\{\varepsilon\cdot u(e^\varepsilon)\}^3\\ &\qquad+(|U_{\varepsilon,m}^*|+|\Psi(\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_m)|)\cdot\max\{Q_T(D_\varepsilon), Q_m^*(D_\varepsilon)\}\cdot(1-e^{-\varepsilon})\\ &\leq(1-e^{-\varepsilon})\{1+|U_{\varepsilon,m}^*|+|\Psi(\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_m)|\}+\varepsilon+\frac{1}{2}\{\varepsilon\cdot u(e^\varepsilon)\}^2+\frac{1}{6}\{\varepsilon\cdot u(e^\varepsilon)\}^3\,.\end{aligned}$$

不等式(4.6)の証明は，(3.7)のもう一方の不等式と，(4.2)，(4.4)，(2.6)を利用する：

$$
\begin{aligned}
I^*(Q_T, P; D_\varepsilon) \ge & I^*(Q_T, Q_m^*; D_\varepsilon) + I^*(Q_m^*, P; D_\varepsilon) \\
& -(|U_{\varepsilon,m}^*| + |\Psi(\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_m)|)\cdot|Q_T(D_\varepsilon) - Q_m^*(D_\varepsilon)| \\
\ge & e^{-\varepsilon} - 1 + (e^{-\varepsilon}-1) + \frac{1}{2}\left[\{\ell(e^{-\varepsilon})\}^2 e^{-\varepsilon} P(D_\varepsilon) - \frac{1}{3}\{u(e^{\varepsilon})\}^3\varepsilon\right]\cdot\varepsilon^2 \\
& -(|U_{\varepsilon,m}^*| + |\Psi(\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_m)|)\cdot\max\{Q_T(D_\varepsilon), Q_m^*(D_\varepsilon)\}\cdot(1-e^{-\varepsilon}) \\
\ge & (e^{-\varepsilon}-1)\{2 + |U_{\varepsilon,m}^*| + |\Psi(\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_m)|\} + \frac{1}{2}\left[\{\ell(e^{-\varepsilon})\}^2 e^{-\varepsilon} P(D_\varepsilon) - \frac{1}{3}\{u(e^{\varepsilon})\}^3\varepsilon\right]\cdot\varepsilon^2
\end{aligned}
$$

□

さて，多変量一般指数型分布族に属する分布 dQ_m^* のパラメータが $\{\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_m\}\to\{\boldsymbol{\beta}_{1T},\ldots,\boldsymbol{\beta}_{mT}\}$ と変化し，同じ分布族内の dQ_T に移行する場合を考える．このような設定は統計的にも，物理的にも極く自然である．我々の関心は，この場合の二つの分布の変動を，パラメータの変化を用いてどう評価できるか，さらには，$\boldsymbol{u}_i(\boldsymbol{A})$ $(i=1,\ldots,m)$ の平均値を使ってどう評価できるかである．このことを以下で考察する．このために，dQ_m^*, dQ_T を少し扱い易い形で考える：dQ_m^*, dQ_T およびそれらを誘導する基礎分布 dP が可測空間 $(R^{n\times n}, \boldsymbol{\beta}^{n\times n})$ 上で定義される σ-有限測度 $\mu^{n\times n}$ に関して絶対連続とし，

$$dP = p(\boldsymbol{A})d\mu^{n\times n}, \quad dQ_m^* = q(\boldsymbol{A}; \boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_m)p(\boldsymbol{A})d\mu^{n\times n},$$
$$dQ_T = q(\boldsymbol{A}; \boldsymbol{\beta}_{1T},\ldots,\boldsymbol{\beta}_{mT})p(\boldsymbol{A})d\mu^{n\times n},$$
$$q(\boldsymbol{A}; \boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_m) = \exp\left[-\mathrm{tr}\left\{\sum_{i=1}^m \boldsymbol{\theta}_i^t \boldsymbol{u}_i(\boldsymbol{A})\right\} - \Psi(\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_m)\right], \tag{4.9}$$

と表わす．ここに，$q(\boldsymbol{A}; \cdot)$ は考察対象の指数型分布族の $\mu^{n\times n}$ に関する Radon-Nikodym 導関数を表わし，$\boldsymbol{\theta}_i$ $(i=1,\ldots,m)\in R^{n\times n}$ で一般の正準パラメータ行列を代表させる．

ところで上の設定で $q(\boldsymbol{A}; \cdot)$ は存在は仮定されているがその形は明示されていない．そこで，以下では $\boldsymbol{\theta}_i$ $(i=1,\ldots,m)$ の関数として $q(\boldsymbol{A}; \boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_m) =: q$ と記し，q がある可測集合 K $(\in \boldsymbol{B}^{n\times n})$ 上で次の Kullback (1959) に準じた正則条件を満たすものと仮定する：

1. 任意の $\boldsymbol{A}\in\boldsymbol{B}^{n\times n}$ と，その要素が非退化区間 $\min(\beta_{i\mu\nu}, \beta_{iT\mu\nu}) < \beta_{i\mu\nu}^* < \max(\beta_{i\mu\nu}, \beta_{iT\mu\nu})$, $i=1, \ldots, m$; $\mu, \nu=1,\ldots,n$ に属する任意の正準パラメータ行列 $\boldsymbol{\beta}_i^* = (\beta_{i\mu\nu}^*)$ $[\boldsymbol{\beta}_i^* \in ((\boldsymbol{\beta}_i, \boldsymbol{\beta}_{iT}))$ と以下で略記する$]$ に対して，$(\partial \ln q/\partial\boldsymbol{\theta}_i), (\partial^2 \ln q/\partial\boldsymbol{\theta}_i^t\partial\boldsymbol{\theta}_i)$ $(i=1,\ldots,m)$ が存在する．
2. 任意の $\boldsymbol{\beta}_i^* \in ((\boldsymbol{\beta}_i, \boldsymbol{\beta}_{iT}))$ に対して，$\|(\partial \ln q/\partial\boldsymbol{\theta}_i)\| < L(\boldsymbol{A}), \|(\partial^2 \ln q/\partial\boldsymbol{\theta}_i^t\partial\boldsymbol{\theta}_i)\| < M(\boldsymbol{A})$, が全ての $i=1,\ldots,m$ に対して成立するものとする．ここで，記号 $\|\cdot\|$ は次の意味での行列の Frobenius ノルムを表わす (cf. Golub and van Loan (1989), Chatelin (1993))：

$$\|\boldsymbol{A}\| = \left(\sum_{i=1}^n\sum_{j=1}^n |a_{ij}|^2\right)^{1/2} = \sqrt{\mathrm{tr}(\boldsymbol{A}^t\boldsymbol{A})} \quad \text{for} \quad \boldsymbol{A} = (a_{ij}) \in R^{n\times n}.$$

また，$L(\boldsymbol{A}), M(\boldsymbol{A}), N(\boldsymbol{A})$ は K 上で可積分 $[\mu^{n\times n}]$ で，$\sup_{A\in K}|L(\boldsymbol{A}), M(\boldsymbol{A})/2| < H < \infty$ を満たし，H は $\boldsymbol{\theta}_i, i=1,\ldots,m$ に依存しないものとする．

3. パラメータ行列による微分と，$\boldsymbol{A}$ に関連する積分の順序が交換が可能とし，この時分布 dQ_T と dQ_m^* に関連して

$$\frac{\partial}{\partial\boldsymbol{\beta}_{iT}}(-\Psi(\boldsymbol{\beta}_{1T},\ldots,\boldsymbol{\beta}_{mT})) = \int_{R^{n\times n}} \boldsymbol{u}_i(\boldsymbol{A})q(\boldsymbol{A}; \boldsymbol{\beta}_{1T},\ldots,\boldsymbol{\beta}_{mT})d\mu^{n\times n}(\boldsymbol{A}^t) =: \boldsymbol{U}_{iT}(\boldsymbol{\beta}_{1T},\ldots,\boldsymbol{\beta}_{mT}),$$

$$\frac{\partial}{\partial\boldsymbol{\beta}_{iT}^t}\frac{\partial}{\partial\boldsymbol{\beta}_{iT}}\Psi(\boldsymbol{\beta}_{1T},\ldots,\boldsymbol{\beta}_{mT}) = \int_{R^{n\times n}} (\boldsymbol{u}_i(\boldsymbol{A}) - \boldsymbol{U}_{iT})^t(\boldsymbol{u}_i(\boldsymbol{A}) - \boldsymbol{U}_{iT})q\,d\mu^{n\times n}(\boldsymbol{A}^t) =: \boldsymbol{G}_i,$$

$$\frac{\partial}{\partial\boldsymbol{\beta}_i}(-\Psi(\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_m)) = \int_{R^{n\times n}} \boldsymbol{u}_i(\boldsymbol{A})q(\boldsymbol{A}; \boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_m)d\mu^{n\times n}(\boldsymbol{A}^t) =: \boldsymbol{U}_i(\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_m),$$

$(i=1,\dots,m)$ の諸量の存在と関係式が成り立つものとする.

これらの条件の下，まず，正準パラメータが $\{\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_m\}\to\{\boldsymbol{\beta}_{1T},\dots,\boldsymbol{\beta}_{mT}\}$ と変化したとする時，$I(Q_T, Q_m^*\,;K)$ の変動がどうなるかを知ることは多変量一般指数型分布族の揺動の定量的評価という意味で，統計理論面および物理的観点からも大いに興味のあることである．次の命題が成立する：

定理 4.2. 上記の正則条件の下で次の不等式が成立する：

$$\left|\ln\frac{dQ_T}{dQ_m^*}\right|\le H\cdot\sum_{i=1}^{m}\max\{\|\boldsymbol{\beta}_i-\boldsymbol{\beta}_{iT}\|,\ \|\boldsymbol{\beta}_i-\boldsymbol{\beta}_{iT}\|^2\}\,, \tag{4.10}$$

$$I(Q_T, Q_m^*\,;K)\le\sum_{i=1}^{m}\left[\|\boldsymbol{\beta}_i-\boldsymbol{\beta}_{iT}\|\cdot\|\boldsymbol{U}_{iT}^*-\boldsymbol{U}_{iT}\cdot Q_T^*\|+\frac{1}{2}\operatorname{tr}\{\boldsymbol{G}_i^*(\boldsymbol{\beta}_i-\boldsymbol{\beta}_{iT})^t(\boldsymbol{\beta}_i-\boldsymbol{\beta}_{iT})\}\cdot Q_T^*\right], \tag{4.11}$$

$$I(Q_T, Q_m^*\,;K)\ge\sum_{i=1}^{m}\left[-\|\boldsymbol{\beta}_i-\boldsymbol{\beta}_{iT}\|\cdot\|\boldsymbol{U}_{iT}^*-\boldsymbol{U}_{iT}\cdot Q_T^*\|+\frac{1}{2}\operatorname{tr}\{\boldsymbol{G}_i^*(\boldsymbol{\beta}_i-\boldsymbol{\beta}_{iT})^t(\boldsymbol{\beta}_i-\boldsymbol{\beta}_{iT})\}\cdot Q_T^*\right], \tag{4.12}$$

ここで，H は正則条件 2 で定義した量，$\boldsymbol{G}_i^*, Q_T^*, \boldsymbol{U}_{iT}^*$ はそれぞれ次の諸量を表わす：

$$\boldsymbol{G}_i^*:=\left(\frac{\partial}{\partial\boldsymbol{\beta}_{iT}^t}\frac{\partial}{\partial\boldsymbol{\beta}_{iT}}\Psi(\boldsymbol{\beta}_{1T},\dots,\boldsymbol{\beta}_{mT})\right)_{\boldsymbol{\beta}_{iT}=\boldsymbol{\beta}_i^*},\qquad (i=1,\dots,m)\,,$$

$$Q_T^*:=Q_T(K\,;\boldsymbol{\beta}_{1T},\dots,\boldsymbol{\beta}_{mT})=\int_K q(\boldsymbol{A}\,;\boldsymbol{\beta}_{1T},\dots,\boldsymbol{\beta}_{mT})p(\boldsymbol{A})d\mu^{n\times n}(\boldsymbol{A})\,,$$

$$\boldsymbol{U}_{iT}^*:=\boldsymbol{U}_{iT}^*(\boldsymbol{\beta}_{1T},\dots,\boldsymbol{\beta}_{mT})=\int_K \boldsymbol{u}_i(\boldsymbol{A})q(\boldsymbol{A}\,;\boldsymbol{\beta}_{1T},\dots,\boldsymbol{\beta}_{mT})p(\boldsymbol{A})d\mu^{n\times n}(\boldsymbol{A})\,.$$

注 4.2. Barron and Sheu (1991) は $\ln dQ_m^*, \ln dQ_T$ が有界かつ線形独立な直交関数系で張られる線形空間に属するものとして，上の命題に対応する一変量の場合の結果を得ている．そこでの有界性が彼らの証明の本質的な役割をしているが，統計理論の観点からはその意味が理解しにくい．本稿では，彼らの接近法と違って，情報量密度比 $\ln(dQ_T/dQ_m^*)$ に Taylor 展開を用いて K-L 情報量の近似に繋げるという Kullback (1959) の接近法を多変量の場合に適用する．この方法では，$\{\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_m\}\to\{\boldsymbol{\beta}_{1T},\dots,\boldsymbol{\beta}_{mT}\}$ の変化量だけでなく，その係数としていわゆる Fisher の情報量が現われる利点がある．しかし Barron and Sheu (1991) の接近法にはこの利点が現われない．なお，本定理の条件(4.10)を補題 3.1 の C_ε と組み合わせることにより近似主領域を構成して議論を展開できるが本稿では省略する．

定理 4.2 の証明. 正則条件の下，$\boldsymbol{\beta}_i=\boldsymbol{\beta}_{iT}+(\boldsymbol{\beta}_i-\boldsymbol{\beta}_{iT})=:\boldsymbol{\beta}_{iT}+\Delta\boldsymbol{\beta}_{iT}$ $(i=1,\dots,m)$ として $dQ_m^*=q(\boldsymbol{A}\,;\boldsymbol{\beta}_{1T}+\Delta\boldsymbol{\beta}_{1T},\dots,\boldsymbol{\beta}_{mT}+\Delta\boldsymbol{\beta}_{mT})p(\boldsymbol{A})d\mu^{n\times n}(\boldsymbol{A})$ に Taylor 近似を適用して

$$\ln\frac{dQ_T}{dQ_m^*}=-\operatorname{tr}\left[\sum_{i=1}^{m}(\boldsymbol{\beta}_i-\boldsymbol{\beta}_{iT})^t\frac{\partial\ln q}{\partial\boldsymbol{\beta}_{iT}}+\frac{1}{2}\sum_{i,j=1}^{m}(\boldsymbol{\beta}_i-\boldsymbol{\beta}_{iT})^t(\boldsymbol{\beta}_i-\boldsymbol{\beta}_{iT})\left(\frac{\partial^2\ln q}{\partial\boldsymbol{\beta}_{iT}^t\partial\boldsymbol{\beta}_{jT}}\right)_{\boldsymbol{\beta}_{iT}=\boldsymbol{\beta}_i^*}\right], \tag{4.13}$$

ここで $\boldsymbol{\beta}_i^*\in((\boldsymbol{\beta}_i, \boldsymbol{\beta}_{iT}))$ は $\boldsymbol{\beta}_i^*$ $(i=1,\dots,m)$ の各要素が，$\boldsymbol{\beta}_i$ と $\boldsymbol{\beta}_{iT}$ の対応する要素の間の値を取るパラメータ行列であることを表わす．即ち，$\boldsymbol{\beta}_i^*=\boldsymbol{\beta}_{iT}+\boldsymbol{\eta}_i(\boldsymbol{\beta}_i-\boldsymbol{\beta}_{iT})$, $\boldsymbol{\eta}_i=(\eta_{ij})\in R^{n\times n}$, $0<\eta_{ij}<1$, $i, j=1,\dots,m$ を意味する．正則条件 2 によって

$$\begin{aligned}\left|\ln\frac{dQ_T}{dQ_m^*}\right|&\le\sum_{i=1}^{m}\left\{\|\boldsymbol{\beta}_i-\boldsymbol{\beta}_{iT}\|\cdot L(\boldsymbol{A})+\frac{1}{2}\|\boldsymbol{\beta}_i-\boldsymbol{\beta}_{iT}\|^2\cdot M(\boldsymbol{A})\right\}\\&\le\sup_{A\subset R^{n\times n}}\{L(\boldsymbol{A}), M(\boldsymbol{A})/2\}\cdot\sum_{i=1}^{m}\max\{\|\boldsymbol{\beta}_i-\boldsymbol{\beta}_{iT}\|,\ \|\boldsymbol{\beta}_i-\boldsymbol{\beta}_{iT}\|^2\}\,.\end{aligned} \tag{4.14}$$

正則条件2の最後の条件により，(4.10)が得られる．

次に，(4.11)を証明する．式(4.13)を，K上で，dQ_Tで積分する．正則条件3から

$$\mathrm{tr}\left[\int_K\left(\frac{\partial}{\partial\boldsymbol{\beta}_{iT}}\ln q\right)dQ_T\right]=-\mathrm{tr}(\boldsymbol{U}_{iT}^*-\boldsymbol{U}_{iT}\cdot Q_T^*),$$
$$\mathrm{tr}\left[\int_K\left(\frac{\partial}{\partial\boldsymbol{\beta}_{iT}^t}\frac{\partial}{\partial\boldsymbol{\beta}_{jT}}\ln q\right)_{\beta_{iT}=\beta_i^*}\cdot dQ_T\right]=-\mathrm{tr}\ \boldsymbol{G}_i^*\cdot Q_T^*$$

であるから

$$(4.15)\quad I(Q_T,Q_m^*;K)=\sum_{i=1}^m\mathrm{tr}\left[(\boldsymbol{\beta}_i-\boldsymbol{\beta}_{iT})^t(\boldsymbol{U}_{iT}^*-\boldsymbol{U}_{iT}\cdot Q_T^*)+\frac{1}{2}\{(\boldsymbol{\beta}_i-\boldsymbol{\beta}_{iT})^t(\boldsymbol{\beta}_i-\boldsymbol{\beta}_{iT})\boldsymbol{G}_i^*\cdot Q_T^*\}\right]$$
$$\leq\sum_{i=1}^m\left[\|\boldsymbol{\beta}_i-\boldsymbol{\beta}_{iT}\|\cdot\|\boldsymbol{U}_{iT}^*-\boldsymbol{U}_{iT}\cdot Q_T^*\|+\frac{1}{2}\mathrm{tr}\{\boldsymbol{G}_i^*(\boldsymbol{\beta}_i-\boldsymbol{\beta}_{iT})^t(\boldsymbol{\beta}_i-\boldsymbol{\beta}_{iT})\}\cdot Q_T^*\right].$$

これによって不等式(4.11)が証明された．同様にして不等式(4.12)も証明される．□

最後に，標題の一様近似の意味を明示的に表現するための結果を記す．次の一様距離に基づいて，近似主領域からの確率測度の逃げも考慮した分布間の近似誤差の評価が可能である．証明はMatsunawa (1986) に基づいて行えるので省略する．

定理 4.3. $D(Q_T,P;\boldsymbol{B}^{n\times n})=\sup_{E\in\boldsymbol{B}^{n\times n}}|Q_T(E)-P(E)|=2^{-1}\int_R|dQ_T-dP|=:D$とし，前と同様に

$$(4.16)\qquad E_\varepsilon:=\left\{\boldsymbol{A};\left|\ln\frac{dQ_T(\boldsymbol{A})}{dP(\boldsymbol{A})}\right|\leq\varepsilon,\quad \boldsymbol{A}\in R^{n\times n}\right\}$$

を考える．これが

$$(4.17)\qquad \max\{P(E_\varepsilon),Q_T(E_\varepsilon)\}\geq e^{-\varepsilon^*},\qquad (0\leq\varepsilon^*)$$

を満たすものとする．この時，一様距離Dに対して次の不等式が成り立つ：

$$(4.18)\qquad D\geq\frac{1}{2}\left[\ell\left(\inf_{E_\varepsilon}\frac{dP}{dQ_T}\right)\cdot|I^*(Q_T,P;E_\varepsilon)|+|Q_T(E_\varepsilon)-P(E_\varepsilon)|\right]$$

$$(4.19)\qquad \geq\frac{1}{2}\ell(e^{-\varepsilon})\cdot|I^*(Q_T,P;E_\varepsilon)|+\varepsilon\cdot\ell(e^{-\varepsilon})\cdot e^{-\varepsilon^*}\geq\frac{1}{2}\varepsilon\cdot\ell(e^{-\varepsilon})\cdot e^{-\varepsilon^*}$$

$$(4.20)\qquad D\leq\frac{1}{2}u\left(\sup_{E_\varepsilon^-}\frac{dP}{dQ_T}\right)\cdot|I^*(Q_T,P;E_\varepsilon)|+\left\{1-\frac{Q_T(E_\varepsilon)+P(E_\varepsilon)}{2}\right\}$$

$$(4.21)\qquad \leq\frac{1}{2}u\left(\sup_{E_\varepsilon}\frac{dP}{dQ_T}\right)\cdot I_a^*(Q_T,P;E_\varepsilon)+\{1-\min(Q_T(E_\varepsilon),P(E_\varepsilon))\}$$
$$\leq\frac{1}{2}u(e^\varepsilon)\cdot\varepsilon+(1-e^{-(\varepsilon+\varepsilon^*)})$$

ただし，$\ell(\cdot),u(\cdot)$は補題2.3で定義された関数である．また，(4.20)において$E_\varepsilon^-:=E_\varepsilon\cap\{\boldsymbol{A};dP(\boldsymbol{A})/dQ_T(\boldsymbol{A})<1,\boldsymbol{A}\in R^{n\times n}\}$を表わし，(4.21)において

$$I_a^*(Q_T,P;E_\varepsilon):=\int_{E_\varepsilon}\left|\ln\frac{dQ_T}{dP}\right|dQ_T\leq\varepsilon Q_T(E_\varepsilon)\leq\varepsilon$$

である．□

注 4.3. 上記定理で $\max\{\varepsilon, \varepsilon^*\}\to 0$ ならば $D\to 0$ であることが分かる．即ち，D に基づく二つの確率分布間の一様漸近理論あるいは極限理論は，この定理によってより精密な誤差評価を含む議論が可能になる．なお，D は距離としての対称性を有する．即ち D は，K-L 情報量や W-divergence と違って，近似に方向性を持たない．このことから，D の評価として $I^*(P, Q_T ; E_\varepsilon)$ や $I_\alpha^*(P, Q_T ; E_\varepsilon)$ も使えることが分かる．

5. あとがき

確率分布の間の近似を評価することは統計学の理論研究の根幹と言える．本稿での研究もこの分野への貢献を意図してなされた．近似主領域を意識しての一様近似理論の組織的な研究は Matsunawa (1982, 1986) において論じられたが，本稿ではそこで得たいくつかの基本的不等式を改善した．我々の近似は，従来の確率分布列の近似理論における吸引域の問題と思想的に近いとも言えるが，一般には両者では興味の対象となる近似の強さが違っている．我々の場合，分布間に絶対連続性を仮定している分だけ近似の結果は強くなっている．近似の度合いを，主として K-L 情報量に基づいて議論することを試みたが，この量に密接に係わる Affinity, W-divergence, Half variation との関係も従来より精密になった．それらの結果を利用して，小標本の近似問題における定量的な誤差評価を可能にする一連の両側不等式を与え得た．我々の近似理論の長所は問題の設定と解決の見通しが理解しやすいことにある．修正 K-L 情報量の精密な評価と，近似主領域の構成を追及することを念頭に問題に取り組めばよいからである．従来のよく知られているかなり多くの近似問題は，実のところ，本稿の問題設定の中で見通しよく，しかも場合によってはより強い結果で与え得ると思われる．また，本稿では単なる分布間の近似だけではなく，分布の変化の方向にも興味があった．この点からは，近似主領域における修正 K-L 情報量による近似は，物理的エントロピーをその背景に持ち計算も比較的しやすい，非常に有用な独立した近似概念と捉え得る．応用として，多変量一般指数型分布族の近似に限ったが，今後さらに具体的な問題へ適用し，よい事例をえることが課題である．

謝　　辞

本稿を精読し貴重なコメントをして頂いた査読者と担当編集者に深く感謝します．

付録 1

(2.15)，(2.16)の証明. $t>0, \alpha:\mathrm{real}(\neq 0)$ に対し，$t\to 1$ の時式(2.15)

$$\ln t=(t-1)\left(\frac{2}{t^{1/\alpha}+1}\right)^{\alpha}+O\left(\frac{\alpha^2(3-\alpha)}{3}\left(\frac{t^{1/\alpha}-1}{t^{1/\alpha}+1}\right)^3\right), \qquad (\alpha\neq 3)$$

が成立することを証明する．左辺を変形して

$$\begin{aligned}\ln t=\alpha\ln t^{1/\alpha}&=-\alpha\ln\{1+(1-t^{1/\alpha})/t^{1/\alpha}\}\\&=-\alpha\ln\left[\left\{1+\frac{(1-t^{1/\alpha})/t^{1/\alpha}}{2+(1-t^{1/\alpha})/t^{1/\alpha}}\right\}\Big/\left\{1-\frac{(1-t^{1/\alpha})/t^{1/\alpha}}{2+(1-t^{1/\alpha})/t^{1/\alpha}}\right\}\right]\\&=-\alpha\ln\left\{\left(1+\frac{1-t^{1/\alpha}}{1+t^{1/\alpha}}\right)\Big/\left(1-\frac{1-t^{1/\alpha}}{1+t^{1/\alpha}}\right)\right\}=:-\alpha\ln\left(\frac{1+u}{1-u}\right)\end{aligned}$$

ここで

$$u := \frac{1-t^{1/\alpha}}{1+t^{1/\alpha}}, \qquad (-1<u<1)$$

と置いた．これは展開により次の様に表わされる．

$$\text{(A.1)} \qquad \ln t = -2\alpha\left(u+\frac{1}{3}u^3+\frac{1}{5}u^5+\frac{1}{7}u^7+\cdots\right)$$

一方，(2.15)の右辺の主要項は

$$\text{(A.2)} \qquad 2^\alpha(t-1)(1+t^{1/\alpha})^{-\alpha} = \left(1-\frac{1-t^{1/\alpha}}{1+t^{1/\alpha}}\right)^\alpha - \left(1+\frac{1-t^{1/\alpha}}{1+t^{1/\alpha}}\right)^\alpha = (1-u)^\alpha-(1+u)^\alpha$$
$$= -2\left\{\binom{\alpha}{1}u+\binom{\alpha}{3}u^3+\binom{\alpha}{5}u^5+\cdots\right\}$$

と表現できる．そこで，誤差を評価するために，(A.1)－(A.2)を計算すると

$$\text{(A.3)} \qquad \ln t - 2^\alpha(t-1)(1+t^{1/\alpha})^{-\alpha}$$
$$= -2\alpha\left[\frac{1}{3}\left\{1-\frac{(\alpha-1)(\alpha-2)}{2!}\right\}u^3+\frac{1}{5}\left\{1-\frac{(\alpha-1)(\alpha-2)(\alpha-3)(\alpha-4)}{4!}\right\}u^5+\cdots\right].$$

これより(2.15)が得られる．次に $\ln t$ の精密な近似を得るためには右辺の u^3 の係数をゼロと置く：

$$1-\frac{(\alpha-1)(\alpha-2)}{2!}=0 \qquad \therefore \quad \alpha(\alpha-3)=0.$$

$\alpha\neq 0$ と仮定したから，結局 $\alpha=3$ と設定すればよい．即ち，$t\to t^{1/3}$ なる立方根変換を利用することにより，$\ln t$ のより優れた近似式，補題2.3の式(2.16)が求まる．実際，$t=1$ での自明な等号成立の場合を除き，後者の誤差の方が常に小さくなることが両者の比を計算して分かる：

$$\left(\frac{t^{1/3}-1}{t^{1/3}+1}\right)^5\Big/\left(\frac{t-1}{t+1}\right)^3 = \left(\frac{t^{1/3}-1}{t^{1/3}+1}\right)^2\left(\frac{t^{2/3}-t^{1/3}+1}{t^{2/3}+t^{1/3}+1}\right)^3 = \left(1-\frac{2}{t^{1/3}+1}\right)^2\left\{\frac{(t^{1/3}-1/2)^2+3/4}{(t^{1/3}+1/2)^2+3/4}\right\}^3 < 1.$$

付録 2

補題2.3で与えた不等式(2.17)は本稿の解析で有用であった．他の様々な理論や応用問題に役立つと思われるので，以下に幾つかの式と数値例を与える：

$$\ell_i(t)|\ln t| \le |t-1| \le u_i(t)|\ln t|, \qquad (t>0,\ i=2,4,6,\ldots)$$

$$u_2(t)=\frac{(1+t^{1/3})(1+t^{1/3}+t^{2/3})(1+8t^{1/3}+t^{2/3})}{11+384t^{1/3}+11t^{2/3}},\quad u_4(t)=\frac{5(1+t^{1/3})^3(1+t^{1/3}+t^{2/3})(1+12t^{1/3}+t^{2/3})}{61+436t^{1/3}+686t^{2/3}+436t+61t^{4/3}},$$

$$u_6(t)=\frac{35(1+t^{1/3})^5(1+t^{1/3}+t^{2/3})(1+16t^{1/3}+t^{2/3})}{457+5014t^{1/3}+14471t^{2/3}+20596t+14471t^{4/3}+5014t^{5/3}+457t^2},$$

$$\ell_4(t)=\frac{10t^{1/3}(1+t^{1/3})^3(1+t^{1/3}+t^{2/3})}{3+68t^{1/3}+98t^{2/3}+68t+3t^{4/3}},$$

$$\ell_6(t)=\frac{70t^{1/3}(1+t^{1/3})^5(1+t^{1/3}+t^{2/3})}{15+554t^{1/3}+1569t^{2/3}+2444t+1569t^{4/3}+554t^{5/3}+15t^2},$$

表 A.1. 不等式(2.17)による $|t-1|$ の近似の絶対誤差.

	$\lvert t-1\rvert - l_i[t]\lvert\ln t\rvert$			$u_i[t]\lvert\ln t\rvert - \lvert t-1\rvert$			
項数	4	6	8	2	4	6	8
t=0.1	1.47×10^{-4}	1.11×10^{-5}	9.52×10^{-7}	6.00×10^{-5}	3.32×10^{-6}	2.26×10^{-7}	1.74×10^{-8}
t=0.2	1.61×10^{-5}	6.18×10^{-7}	2.71×10^{-8}	6.51×10^{-6}	1.81×10^{-7}	6.22×10^{-9}	2.44×10^{-10}
t=0.3	2.53×10^{-6}	5.53×10^{-8}	1.38×10^{-9}	1.02×10^{-6}	1.60×10^{-8}	3.13×10^{-10}	6.99×10^{-12}
t=0.4	4.27×10^{-7}	5.46×10^{-9}	7.99×10^{-11}	1.71×10^{-7}	1.57×10^{-9}	1.79×10^{-11}	2.35×10^{-13}
t=0.5	6.72×10^{-8}	4.95×10^{-10}	4.17×10^{-12}	2.69×10^{-8}	1.42×10^{-10}	9.31×10^{-13}	6.99×10^{-15}
t=0.6	8.66×10^{-9}	3.47×10^{-11}	1.59×10^{-13}	3.47×10^{-9}	9.94×10^{-12}	3.55×10^{-14}	1.67×10^{-16}
t=0.7	7.54×10^{-10}	1.48×10^{-12}	3.44×10^{-15}	3.02×10^{-10}	4.23×10^{-13}	6.66×10^{-16}	*1.67×10^{-16}
t=0.8	3.02×10^{-11}	2.33×10^{-14}	2.78×10^{-17}	1.21×10^{-11}	6.63×10^{-15}	*2.78×10^{-17}	*1.11×10^{-16}
t=0.9	1.67×10^{-13}	0.00	*2.78×10^{-17}	6.70×10^{-14}	4.16×10^{-17}	4.16×10^{-17}	4.16×10^{-17}
t=1.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
t=1.1	9.18×10^{-14}	0.00	0.00	3.67×10^{-14}	4.16×10^{-17}	0.00	0.00
t=1.2	8.99×10^{-12}	4.55×10^{-15}	*5.55×10^{-17}	3.60×10^{-12}	1.36×10^{-15}	5.55×10^{-17}	1.11×10^{-16}
t=1.3	1.20×10^{-10}	1.27×10^{-13}	1.11×10^{-16}	4.79×10^{-11}	3.64×10^{-14}	5.55×10^{-17}	5.55×10^{-17}
t=1.4	7.09×10^{-10}	1.24×10^{-12}	2.61×10^{-15}	2.84×10^{-10}	3.54×10^{-13}	2.78×10^{-16}	*2.22×10^{-16}
t=1.5	2.71×10^{-9}	6.86×10^{-12}	2.00×10^{-14}	1.09×10^{-9}	1.96×10^{-12}	4.11×10^{-15}	*3.33×10^{-16}
t=1.6	7.88×10^{-9}	2.68×10^{-11}	1.05×10^{-13}	3.16×10^{-9}	7.67×10^{-12}	2.26×10^{-14}	*8.88×10^{-16}
t=1.7	1.90×10^{-8}	8.23×10^{-11}	4.08×10^{-13}	7.62×10^{-9}	2.36×10^{-11}	9.06×10^{-14}	0.00
t=1.8	4.01×10^{-8}	2.13×10^{-10}	1.29×10^{-12}	1.61×10^{-8}	6.10×10^{-11}	2.89×10^{-13}	2.66×10^{-15}
t=1.9	7.64×10^{-8}	4.82×10^{-10}	3.49×10^{-12}	3.06×10^{-8}	1.38×10^{-10}	7.80×10^{-13}	5.77×10^{-15}
t=2.0	1.34×10^{-7}	9.89×10^{-10}	8.33×10^{-12}	5.39×10^{-8}	2.84×10^{-10}	1.86×10^{-12}	1.33×10^{-14}

* 印は計算機の精度上の問題で負号のついたものを示す.

$$\ell_8(t)=\frac{210t^{1/3}(1+t^{1/3})^7(1+t^{1/3}+t^{2/3})}{35+1832t^{1/3}+7796t^{2/3}+18968t+23378t^{4/3}+18968t^{5/3}+7796t^2+1832t^{7/3}+35t^{8/3}}\cdot$$

$$\forall(t>0\ \&\ i)\ 0\le\ell_i(t)\le u_i(t)\,;\qquad \forall i\quad \ell_i(1)=u_i(1)=1\,.$$

$$\ell_i(t)\downarrow 0,\text{ as } t\to0\,;\ \ell_i(t)\uparrow\infty,\text{ as } t\to\infty\,;\ u_i(t)\downarrow,\text{ as } t\to0\,;\ u_i(t)\uparrow\infty,\text{ as } t\to\infty\,.$$

参 考 文 献

Barron, A. R. and Sheu, C.-H. (1991). Approximation of density functions by sequences of exponential families, *Ann. Statist.*, **19**, 1347-1369.

Brown, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*, IMS Lecture Notes-Monograph Series, Vol. 9, Hayward, California.

Chatelin, F. (1993). *Eigenvalues of Matrices*, Wiley, Chichester.

Golub, G. H. and van Loan, C. F. (1989). *Matrix Computations*, 2nd ed., The Johns Hopkins University Press, Baltimore.

Kullback, S. (1959). *Information Theory and Statistics*, Wiley, New York.

Matsunawa, T. (1982). Uniform ϕ-equivalence of probability distributions based on information and related measure of discrepancy, *Ann. Inst. Statist. Math.*, **34**, 1-17.

Matsunawa, T. (1986). Modified information criteria for a uniform approximate equivalence of probability distributions, *Ann. Inst. Statist. Math.*, **38**, 205-222.

Matsunawa, T. (1995). Development of distributions——The Legendre transformation and canonical information criteria——, *Proc. Inst. Statist. Math.*, **43**, 293-311 (in Japanese).

Uniform Approximations of Probability Distributions Based on K-L Information Defined on Approximate Main Domains —With Applications to Quantitative Evaluations of Fluctuations in Multivariate General Exponential Families—

Tomoya Yamada

(Department of Statistical Science, The Graduate University for Advanced Studies)

Tadashi Matsunawa

(The Institute of Statistical Mathematics)

Quantitative approximation theory in a certain strong sense for estimating the approximation error between two probability distributions is developed. The distributions are assumed that they are absolutely continuous with respect to each other on a measurable space. Some familiar measures of discrepancy between the two distributions, including K-L information number, Matusita's affinity and Kagan's W-divergence, are considered on certain approximate main domains on which almost all probability measures of the underlying distributions are lying. In case where the modified K-L information number is considered as the measure of discrepancy, detailed approximation theory is discussed by providing new double-sided inequalities for the information. As applications, some interesting inequalities are given to estimate the fluctuations of the multivariate general exponential family when its canonical parameters or its mean values have small changes. Apart from the direction of the approximation between two distributions, the mode of the approximation chiefly discussed in this article is equivalent to the uniform approximation in the sense of total variation distance. To clarify the fact some new useful lower and upper bounds to the half variation distance are also given.

Key words: Quantitative uniform approximation, modified K-L information, total variation distance, approximate main domain, multivariate general exponential family, fluctuation of canonical parameter.