

日本統計学会誌
第 39 卷, 第 2 号, 2010 年 3 月
243 頁 ~ 263 頁

カーネル・データスカッシング —カーネル密度推定法のデータマイニングへの応用—

小暮 厚之*, 寒河江 雅彦†

Kernel Data Squashing: An Application of Kernel Density Estimation
Techniques to Mining Massive Data Sets

Atsuyuki Kogure* and Masahiko Sagae†

DuMouchel *et al.* (1999) は、大量データの統計分析の計算負荷を緩和するために、データスカッシング (data squashing) というマイニング手法を提案した。この手法の考え方は、大量データを分析するための計算アルゴリズムを「拡大させる」のではなく、大量データセットを代表的なサンプル (squashed data) へと「縮小させる」というものである。しかし、大量データの統計的情報を保持するようにスカッシングを行うためには、大規模な（非線形）連立方程式を解くことが必要となり、データスカッシングの既存の手法は依然として計算負荷が大きい。この困難を解決するために、本稿ではスカッシング実装への新たなアプローチとしてカーネル・データスカッシング (KDS) を提案する。KDS は、その名前が示唆するようにカーネル密度推定法を利用したものである。KDS は連立方程式を解くステップを必要とせず、簡単に実装可能である。KDS が大量データの統計的情報を十分に保存することを示すために、KDS に基づく最尤法の漸近的なバイアスと漸近分散を導出する。さらに、我々の結果を例示するためのいくつかのシミュレーションを行う。

The data squashing is proposed by DuMouchel *et al.* (1999) to deal with massive data sets. The idea is to scale data sets down to smaller representative samples, *squashed data*, instead of scaling up algorithms to large data sets. However, the original scheme for the construction of the squashed data may still be computationally burdensome because it usually requires solving a large system of equations. To overcome this difficulty we propose a new method, *the Kernel Data Squashing (KDS)*, which does not involve solving such equations and thus is easy to implement. To examine if our approach preserves much of the statistical information of the massive data set, we study the accuracy of the maximum likelihood estimator based on the KDS. In particular, we derive the asymptotic bias and variance of the KDS-based maximum likelihood estimator. The results show that the KDS preserves the statistical information of the original massive data set as required. Some simulation study is provided to explain our results.

キーワード: 大量データ, データ・スカッシング, ビニング, モーメント・マッチング, カーネル密度推定, 多項式カーネル, 最尤法, 漸近理論

* 慶應義塾大学 総合政策学部, 〒 252-8520 神奈川県藤沢市遠藤 5322 (E-mail: kogure@sfc.keio.ac.jp).

† 金沢大学 人間社会研究域経済学経営学系, 〒 920-1192 金沢市角間町 (E-mail: sagae.masahiko@gmail.com).

1. はじめに

DuMouchel (2002) は、大量データを「コンピュータ・メモリーの制約のために、自分の望むように組み合わせたハードウエアとソフトウエアによっては分析できないデータ」と定義し、このような大量データに対して以下の 2 つのアプローチを提示している：

- 他のハードウエアとソフトウエアを組み合わせること
- 小規模ではあるが、元の大規模データと近似的に同一な分析結果が導かれるような代替的なデータを構成すること。

後者のアプローチは、DuMouchel *et al.* (1999) によってデータスカッシング (data squashing) の名前のもとに導入された。その考え方は、大量データに対する計算アルゴリズムの規模を「拡大させる」のではなく、大量データを代表的なサンプルへと「縮小させる」というものである。DuMouchel (2002) では、「データマイニングを「統計的情報の保存を目的とする不可逆圧縮 (lossy compression) である」と述べている。Madigan *et al.* (2002) は、あてはめるべき統計モデルの尤度を用いてスカッシングを改善する提案を行っている。また、鈴木 (2005) はデータスカッシングの興味深い解説を行っている。

データスカッシングは、統計的情報を保存するために、データ空間をいくつかのビンに分割し、オリジナルな大量データとスカッシュ化データの各ビンごとのモーメントを一致させようとする。これは、多くの計算を必要とする大規模な（非線形）連立方程式を解くことを要求するため、スカッシュ化データを構成するための計算負荷が非常に大きくなってしまう。

本稿では、そのような連立方程式を解くことを必要とはせずにモーメント・マッチングを実現する新たな手法を提案する。我々の手法はカーネル密度推定法に基づいており、それをカーネル・データスカッシング (KDS) と呼ぶことにする。我々は、KDS によるスカッシュ化データに基づく最尤推定量の理論的性質を求めることにより、KDS のパフォーマンスを評価する。特に、Hall and Wand (1996), Sagae and Scott (1997), Kogure (1998) で用いられたビニングに基づく密度推定法の手法を用いて、KDS 最尤推定量のバイアスと分散の漸近的な表現を求め、大量データに基づく通常の最尤推定量に対する近似の正確さを評価する。

我々が知る限りにおいて、この論文はデータスカッシングの解析的性質を導出する最初の試みである。我々の主な結果は、データのサイズが大きくなる場合の漸近理論に基づいているが、本稿で考察する大量データの分析にはそれは適切なアプローチであると考える。最後に、導いた結果を例示するために、シミュレーションによる考察も行う。

2. データスカッシングとは何か

データスカッシングの基本的な考え方を説明するために、本稿では单一変量の大規模データセット $\{X_1, X_2, \dots, X_N\}$ を考える。ここで、 N は巨大な数（例えば 100,000）とする。大量データ $\{X_1, X_2, \dots, X_N\}$ を「押し潰して」、小規模のデータ $\{s_1, s_2, \dots, s_M\}$ を作ろう。ここで、小規模データの大きさ M は、オリジナルな大量データに比べてはるかに小さい（例えば 10）ものとする。各 s_j には、「加重」 w_j が付随していると仮定する。 w_j は小規模データにおける s_j の比率を表す定数であり

$$\sum_{j=1}^M w_j = 1 \quad (2.1)$$

を満たすものとする。 $\{(s_j, w_j), j = 1, 2, \dots, M\}$ をスカッシュ化データ (squashed data) と呼ぶ。

大量データ $\{X_1, X_2, \dots, X_N\}$ はある確率分布からの無作為標本であると仮定する。また、この分布の密度関数 $f(\cdot|\theta)$ は、パラメータ θ の値を除いて完全に既知であると仮定する。このとき、 $\{X_1, X_2, \dots, X_N\}$ の対数尤度は

$$l(\theta) \equiv \sum_{i=1}^N \log f(X_i|\theta) \quad (2.2)$$

によって与えられ、対応するスカッシュ化データ $\{(s_j, w_j)\}$ の対数尤度は

$$\tilde{l}(\theta) \equiv \sum_{j=1}^M w_j \log f(s_j|\theta) \quad (2.3)$$

と定義される。もしも M が小さければ、特別な計算アルゴリズムなしに、(2.3)に基づいて θ に関する推測を容易に実行できるであろう。

データスカッシングの背後にある基本的な考え方は、既存の統計手法にも見られる。簡単な例は、サブサンプルである。 $\{X_1, X_2, \dots, X_N\}$ からの大きさ M の無作為なサブサンプルを s_1, s_2, \dots, s_M とすれば、それはデータスカッシングの特別な場合となる。この場合、 $w_j = 1/M$ ($j = 1, \dots, M$) となる。無作為なサブサンプルは、オリジナルな大量データをよく代表するものであるが、 M が小さい場合には推測の効率性は低い。

別の古典的な例は、データのグルーピングである。それは、データ範囲をいくつかのビンに区分し、各ビンにおけるデータの度数分布によって元のデータを代表させる方法である。各ビンの中点を s_j 、相対度数を w_j とすれば、データのグルーピングはデータスカッシングの特殊なケースとなる。この場合の (2.3) は、コンピュータが未発達であった時代にしばしば用いられた正確な尤度 (2.2) に対する近似式である（例えば、Tallis (1967) を見よ）。

データスカッシングを実装するために最も重要かつ困難な課題は、オリジナルな大量データの統計的情報を保存するために、いかに $\{(s_j, w_j)\}$ を選択するかである。膨大な大

量データを扱う今日の統計分析に直面して, DuMouchel *et al.* (1999) は, それらを選択する巧妙なアプローチを編み出した. このアプローチでは, スカッシュ化データとオリジナルデータのビンごとのモーメントが一致するように $\{s_j, w_j\}$ を決める. より具体的に述べるために, $\{B_r\}$ を観測値の範囲 \mathbb{R}_1 の分割

$$\cup_r B_r = \mathbb{R}_1, \quad B_r \cap B_s = \emptyset \quad \text{for } r \neq s$$

とする. B_r を r 番目のビンと呼ぶ. このとき, B_r におけるオリジナルな大量データの k 次ビン・モーメントは

$$m_k(B_r) \equiv \frac{1}{N} \sum_{i=1}^N X_i^k \mathbb{I}\{X_i \in B_r\}, \quad k = 0, 1, \dots, p$$

であり, 対応するスカッシュ化データの k 次ビン・モーメントは

$$\tilde{m}_k(B_r) \equiv \frac{1}{N} \sum_{j=1}^M w_j s_j^k \mathbb{I}\{s_j \in B_r\}, \quad k = 0, 1, \dots, p$$

と定義される. ここで, \mathbb{I} は指示関数であり

$$\mathbb{I}\{A\} \equiv \begin{cases} 1 & A \text{ が成立する} \\ 0 & A \text{ が成立しない} \end{cases}$$

と定義される. DuMouchel *et al.* (1999) は, 次数 Q までのビンモーメントがマッチするように, 各 $r \in \mathbb{Z}$ に対して

$$\sum_{k=0}^Q (\tilde{m}_k(B_r) - m_k(B_r))^2 \tag{2.4}$$

を最小にするように, $\{(s_j, w_j)\}$ を選択することを提案した. 彼らは, このようなモーメントマッチングの理論的根拠として, 各ビンにおいて (2.2) が (2.3) に「十分よく」近似されるという議論を展開している.

3. カーネル・データスカッシング

3.1 構成

DuMouchel *et al.* (1999) によって提案されたモーメント・マッチングというスカッシュ化データの実装は, (2.1) という制約の下で (2.4) の最小化によって行われる. 通常は, この最小化は s_j と w_j に関する導関数をゼロと置いて得られる大規模な連立方程式を解くことによって実現される. 例えば, データ範囲を 10 個のビンに分割し, 各ビンに 5 個のスカッシュ化データが含まれるとすると, 50 組の $\{(s_j, w_j)\}$ に対する 99 ($= 50 \times 2 - 1$) 本の

方程式からなる連立方程式を持つことになる。そのような大規模な連立方程式を解くことは計算負荷が大きく、データスキャッシングのそもそもの精神と相容れない。

本稿では、そのような計算負荷を伴わない $\{(s_j, w_j)\}$ の新たな選択法を提案する。我々のアプローチは以下の通りである：

- s_j の選択：

データの範囲をビン幅 δ が同一のビン $\{B_j\}$ に分割し、 s_j を B_j の中点とする。一般性を失うことなく $s_j = j\delta$ とし、

$$B_j = [j\delta - \delta/2, \quad j\delta + \delta/2)$$

と設定する。

- w_j の選択：

加重 w_j をカーネル関数 K を用いて

$$w_j \equiv \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{j\delta - X_i}{h}\right) \times \delta \quad (3.1)$$

と選択する。ここで、 h はカーネル関数のバンド幅であり、 $h = \delta/2$ と設定する。

(3.1) におけるカーネル関数 K として、ここでは $[-1, 1]$ 上で定義され

$$\int_{-1}^1 K(x)dx = 1, \quad \int_{-1}^1 K^2(x)dx < \infty.$$

を満たす任意の左右対称な関数を採用する。したがって、 K が負の値を取ることも許容する。カーネル関数とバンド幅 h が与えられるとき、オリジナルな大量データの共通の確率密度 $f(x)$ に対するカーネル推定量¹⁾は

$$\hat{f}(x) \equiv \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{x - X_i}{h}\right), \quad -\infty < x < \infty.$$

と与えられる。したがって、(3.1) と比較すると、 w_j/δ は、 $f(j\delta)$ のカーネル推定値であると理解できる。このように $\{(s_j, w_j)\}$ を選択したデータスキャッシングをカーネル・データスキャッシング (KDS) と呼ぶことにする。KDS を構成する上で必要となるのは、 δ 及び K の選択だけである。

KDS の簡単な例は、 K を

$$K(u) = \frac{1}{2}I\{|u| \leq 1\} = [-1, 1] \text{ 上の一様分布}$$

¹⁾ カーネル推定量に関する一般的な議論については、Wand and Jones (1995) 2章を参照されたい。

と選択することによって得られる。このとき w_j は

$$\begin{aligned} w_j &= \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{j\delta - X_i}{h}\right) \delta \\ &= \frac{1}{N} \sum_{i=1}^N \mathbb{I}\left(\frac{j\delta - X_i}{\delta/2}\right) \\ &= m_0(B_j)\delta, \end{aligned}$$

となり、KDS は上で述べたデータのグルーピングに帰着する。

より戦略的な K の選択は多項式カーネルである。各 $p = 0, 1, 2, \dots$ に対して、 $2p$ 次の多項式カーネルは

$$W_{2p}(u) = \sum_{r=0}^p \left(2r + \frac{1}{2}\right) Q_{2r}(0) Q_{2r}(u) \quad (3.2)$$

によって定義される。ここで、 $Q_{2r}(u)$ は $[-1, 1]$ 上で定義される次数 $2r$ のルジャンドル多項式である。最初のいくつかの p の値に対して、多項式カーネルは

$$\begin{aligned} W_0(u) &= 1/2, \quad W_2(u) = \frac{3}{8}(3 - 5u^2), \\ W_4(u) &= \frac{15}{128}(15 - 70u^2 + 63u^4) \end{aligned}$$

となり、対応する w_j は

$$w_j = \begin{cases} m_0(B_j)\delta & \text{for } p = 0 \\ \frac{3}{4}[3m_0(B_j) - 5m_2(B_j)]\delta & \text{for } p = 1 \\ \frac{15}{64}[15m_0(B_j) - 70m_2(B_j) \\ \quad + 63m_4(B_j)]\delta & \text{for } p = 2 \end{cases}$$

となる。 $p = 0$ のとき、多項式カーネルは一様カーネルとなる。一様カーネルに基づく加重 w_j は bin の度数（0 次 bin モーメント）のみに依存するのに対して、次数の高い多項式カーネルは、偶数次の高次 bin モーメントを必要とする。次の定理は、 W_{2p} は $2(p+1)$ のオーダーのカーネル関数であることを示す：

定理 3.1 W_{2p} は (3.2) で定義されるものとする。その（カーネル関数としての）オーダーは $2(p+1)$ となる。すなわち：

$$\int_{-1}^1 u^k W_{2p}(u) du = \begin{cases} 1 & \text{for } k = 0 \\ 0 & \text{for } 1 \leq k \leq 2p+1 \end{cases} \quad (3.3)$$

及び

$$\int_{-1}^1 u^{2(p+1)} W_{2p}(u) du \neq 0$$

が成立する。

証明は補遺で与える。3.3節で示すように、 W_{2p} を用いることは、次数 $2p$ までのモーメントをマッチさせることと実質的に同値である。

3.2 理論

スカッシュ化データに基づく最尤推定量の漸近的性質を導く。簡単化のために、パラメータ θ は一次元空間 Θ に属するものとする。 θ_0 を θ の真の値とし

$$G(\theta) \equiv \int f(x|\theta_0) \log \frac{f(x|\theta)}{f(x|\theta_0)} dx, \quad \theta \in \Theta. \quad (3.4)$$

と定義する。各 $\varepsilon > 0$ に対して

$$\sup_{\theta: |\theta - \theta_0| \geq \varepsilon} G(\theta) < G(\theta_0), \quad (3.5)$$

と仮定する。このような仮定は、最尤推定法の一致性を証明するための標準的な仮定である²⁾。このとき、次の定理が成立する：

定理 3.2 $\tilde{\theta}$ を (2.3) を最大化する KDS 最尤推定量とする。 K は左右対称な $\int u^2 K(u) du < \infty$ を満たすカーネル関数とする。このとき、補遺で与える f に関する一定の条件の下で、 $\tilde{\theta}$ は一致性を持つ。ただし、 δ は

$$\delta \rightarrow 0, \quad \delta \sqrt{\frac{N}{\log N}} \rightarrow \infty \quad \text{as } N \rightarrow \infty.$$

を満たすように選択するものとする。

証明は補遺で与える。

次に $\tilde{\theta}$ のバイアスと分散の漸近的な表現を求める：

定理 3.3 カーネル関数 K のオーダーは $2m$ であり、オリジナルなデータの密度関数 $f(x|\theta_0)$ が $2m+1$ 回連続微分可能であると仮定する。このとき、以下が成立する：

$$E[\sqrt{N}(\tilde{\theta} - \theta_0)] = O\left(\sqrt{N}\delta^{2m}\right) \quad \text{as } N \rightarrow \infty \text{ and } \delta \rightarrow 0.$$

証明は補遺で与える。

この定理は、KDS 最尤推定量の漸近的バイアスは δ に強く依存することを示す。また、この漸近的バイアスはオーダーの大きいカーネルを用いることによって減少可能であるこ

²⁾ 例えば、Wasserman (2004) を参照されたい。

とが分かる。カーネル関数として多項式カーネルを用いる場合には、この定理は高次のビンモーメントを用いてスカッシュ化データを構成することに対する理論的根拠を与える。

よく知られているように、一定の条件の下で、大量データ $\{X_i\}$ に基づく通常の最尤推定量 $\hat{\theta}$ は漸近的に正規分布に従う：

$$\sqrt{N}(\hat{\theta} - \theta_0) \rightsquigarrow N\left(0, \frac{1}{I(\theta_0)}\right)$$

ここで、 $I(\theta)$ はフィッシャーの情報量

$$I(\theta) \equiv - \int \left(\frac{\partial^2 \log f(x|\theta)}{\partial \theta^2} \right) f(x|\theta) dx.$$

である。したがって、もしも $\delta = o(N^{-1/4m})$ と設定すれば、定理 3.3 から

$$E[\sqrt{N}(\tilde{\theta} - \theta_0)] = o(1)$$

となり、漸近的バイアスは無視できることになる。

定理 3.4 定理 3.3 と同じ仮定の下で以下が成立する：

$$\text{Var}\left(\sqrt{N}(\tilde{\theta} - \theta_0)\right) \sim \frac{2R(K)}{I(\theta_0)} \text{ as } N \rightarrow \infty \text{ and } \delta \rightarrow 0.$$

ここで、 $R(K) \equiv \int K^2(z) dz$ である。

証明は補遺で与える。

この定理は、ビン幅 δ は漸近的には分散に対してほとんど影響を与えないことを意味する。さらに、コーチー・シュワルツの不等式を用いると、 $[-1, 1]$ 上で定義される任意のカーネル関数に対して

$$1 = \left(\int_{-1}^1 K(t) dt \right)^2 \leq \int_{-1}^1 K(t)^2 dt \times \int_{-1}^1 dt = 2R(K) \quad (3.6)$$

が示される。したがって、当然のことではあるが、定理 3.4 で与えられる KDS 最尤推定量 $\tilde{\theta}$ の漸近分散は通常の最尤推定量の漸近分散より大きい。 K が一様カーネルであるとき $2R(K) = 1$ となり、漸近分散は最小化される。もしも K として多項式カーネル W_{2p} を用いると

$$R(K) = \sum_{r=0}^p \frac{2r+1}{2} Q_{2r}^2(0),$$

となる。これは p とともに増加する。したがって、高次ビンモーメントを用いると、KDS 最尤推定量の漸近分散は増加する。

3.3 多項式カーネルのモーメント・マッチング性

以上の定理は多くの一般的なカーネル関数に対して成立する。しかし、我々は多項式カーネルの使用を推奨する。多項式カーネルは bin モーメントのみに依存しており計算上の効率性が高いからである。これに加えて、多項式カーネルは、DuMouchel *et al.* (1999) によって提唱されたモーメントマッチングを実現する。実際、多項式カーネル W_{2p} を用いて KDS を行うと、補遺で示すように、 $\hat{f}(x)$ に基づく bin モーメントはオリジナルな大量データの bin モーメントと完全に一致する。すなわち以下が成立する：

$$m_k(B_j) = \int_{B_j} x^k \hat{f}(x) dx, \quad k = 0, 1, \dots, 2p. \quad (3.7)$$

ここで、 B_j における KDS の k 次 bin モーメント

$$\tilde{m}(B_j) = s_j^k w_j$$

は (3.7) の右辺の離散近似であることに注意されたい。

4. シミュレーション

KDS を例示するために 2 種類の簡単なシミュレーションを行った。すべてのシミュレーションにおいて多項式カーネルを用いた。

4.1 正規分布

正規分布 $N(0, \sigma^2)$ から大きさ 100,000 の大量データが与えられているとき、KDS によって分散 σ^2 を推定する問題を考える。 σ^2 の真の値は 1 とする。以下のようにシミュレーションを進めていく。

1. 正規分布 $N(0, 1)$ に従う乱数を 100,000 個発生する。
2. $[-5, 5]$ を M 個の同一幅の bin に分割し、各 bin の中点を s_j とする。
3. $K = W_{2p}, p = 0, 1, 2$ を用いて、KDS 最尤推定量を計算する。

1-3 を 100 回繰り返してバイアスと標準偏差を算出した。 M を 4 から 10 まで変化させて得られた結果が表 1 に表示されている。

図 1 は、この表のグラフ表示である。左のグラフは、各 p に対して M とともにバイアスがいかに変化するかを示す。各 p について、バイアス（の絶対値）が M とともに急速に減少することが分かる。また、 M をそれほど大きくしなくとも、高次モーメントを用いた $p = 1$ あるいは $p = 2$ の場合には、バイアスを小さくできることが分かる。図 1 の右のグラフは、各 p に対して M とともに標準偏差がいかに変化するかを示す。各 p について標準偏差は広い範囲の M の値に対して安定的であることが分かる。一方、各 M に対して、 p

表 1 正規標本に対する KDS 最尤推定量のバイアス (Bias) と標準偏差 (SD)

M	$p = 0$		$p = 1$		$p = 2$	
	Bias	Sd	Bias	Sd	Bias	Sd
4	0.7177	0.0049	0.4568	0.0077	0.4549	0.0093
5	0.3008	0.0062	-0.1184	0.0098	-0.2294	0.0120
6	0.2355	0.0051	0.0182	0.0076	0.0207	0.0086
7	0.1697	0.0052	-0.0011	0.0080	-0.0007	0.0091
8	0.1299	0.0052	-0.0005	0.0079	-0.0005	0.0093
9	0.1032	0.0051	-0.0003	0.0076	-0.0002	0.0099
10	0.0832	0.0046	0.0004	0.0072	0.0006	0.0098

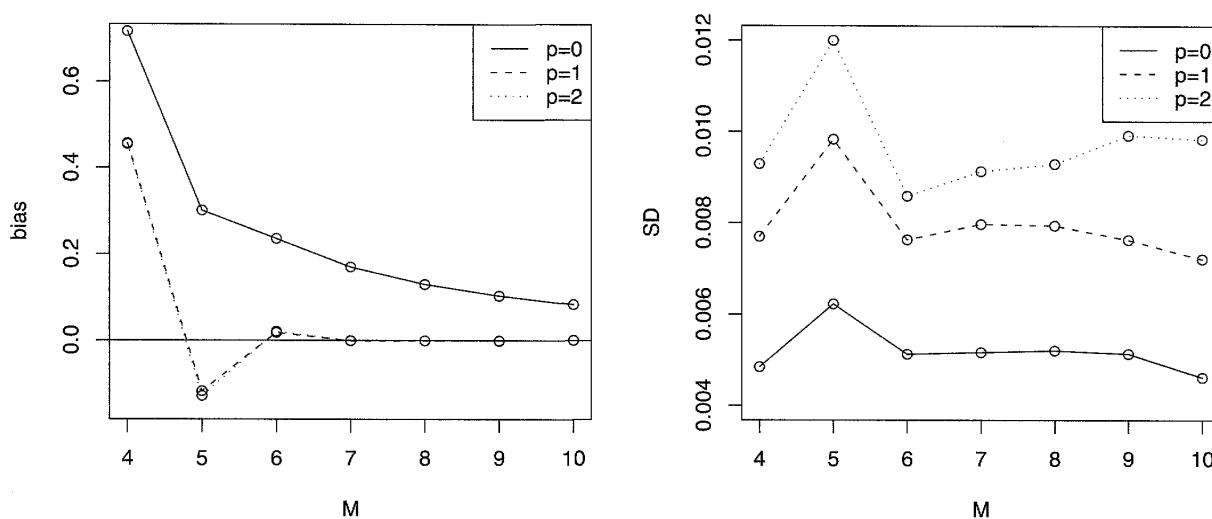


図 1 正規標本に対する KDS 最尤推定量のバイアス (bias) と標準偏差 (SD)

が大きいほど標準偏差は大きくなる。以上のシミュレーション結果は、前節で述べた漸近的な理論的結果と一致する。

図 2 は、バイアスと分散の両方を考慮した全体的な統計的効率性を示す平均平方誤差平方根のグラフ表示である。 $p = 0$ のケースに比べ、 $p = 1$ 及び $p = 2$ の平均平方誤差平方根の値は小さく、高次モーメントを利用するとの有利さを示している。しかし、4 次モーメントを追加する ($p = 2$ に対応) ことは効率性をほとんど改善しないことも示している。以上から、正規分布の分散を推定するには、0 次ビンモーメント及び 2 次ビンモーメントを用いて $M = 5$ あるいは $M = 6$ 程度のビンを用いれば十分であると結論づけられよう。

4.2 パレート分布

Dimakos (2000) は、裾の厚い分布に対するデータスカッシングの性質を探るために、一般化パレート分布に対するシミュレーション研究を行っている。ここでは、パレート分布

$$f(x|\theta) = \frac{\beta}{x^{\beta+1}}, \quad (x \geq 1).$$

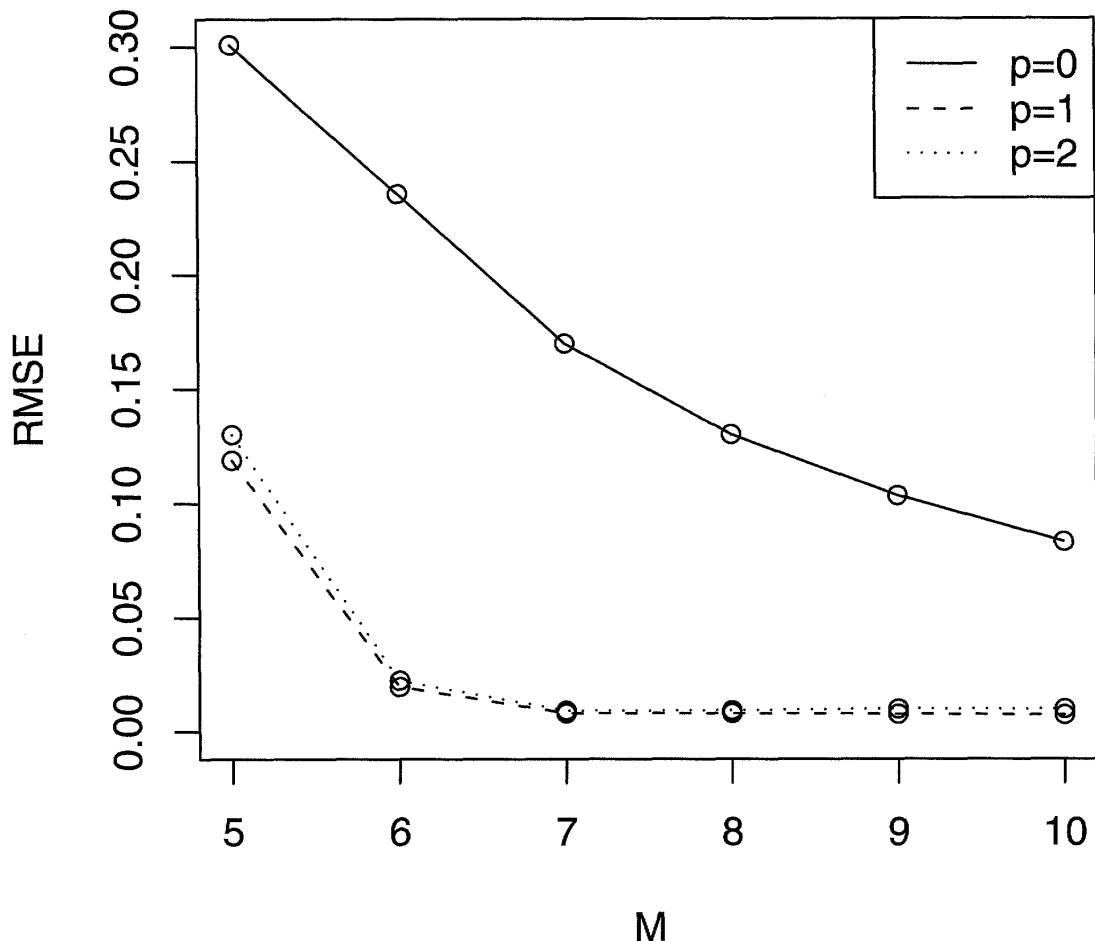


図 2 正規標本に対する KDS 最尤推定量の平均平方誤差平方根 (RMSE)

のパラメータ β の推定を考察する。以下のようにシミュレーションを進めていく：

1. パラメータ $\beta = 3$ であるパレート分布に従う乱数を 100,000 個発生する。
2. $[1, 5]$ を M 個の同一幅の bin に分割し、各 bin の中点を s_j とする。
3. $K = W_{2p}, p = 0, 1, 2$ を用いて、KDS 最尤推定量を計算する。

1-3 を 100 回繰り返してバイアスと標準偏差を算出した。 M を 4 から 10 まで変化させて得られた結果が表 2 に示されているこの表のグラフ表示が図 3 である。左のグラフは、各 p に対していかにバイアスが M とともに変化するかを示す。正規分布の場合と比較可能なように、ここでは、真のパラメータの大きさに対する比率として、推定バイアスを表示している。このグラフから、 M が小さい場合、KDS 最尤推定量は真のパラメータを過小推定していることが分かる。また、正規分布の場合に比べ、バイアスを小さくするためにより大きな M の値を必要とすること、高次の bin モーメントを用いてもバイアスはそれほど低下しないことが示唆される。

表 2 パレート標本に対する KDS 最尤推定量のバイアス (bias) と標準偏差 (SD)

M	$p = 0$		$p = 1$		$p = 2$	
	Bias	Sd	Bias	Sd	Bias	Sd
4	-0.9058	0.0028	-1.0053	0.0056	-0.9888	0.0064
6	-0.5164	0.0048	-0.6090	0.0077	-0.6013	0.0010
8	-0.2990	0.0061	-0.3744	0.0107	-0.3719	0.0127
10	-0.1716	0.0069	-0.2271	0.0123	-0.2231	0.0139
12	-0.0929	0.0073	-0.1397	0.0115	-0.1385	0.0146
14	-0.0410	0.0075	-0.0789	0.0107	-0.0783	0.0139
16	-0.0057	0.0077	-0.0346	0.0125	-0.0341	0.0170
18	0.0192	0.0081	-0.0061	0.0118	-0.0052	0.0159
20	0.0381	0.0083	0.0192	0.0138	0.0194	0.0167

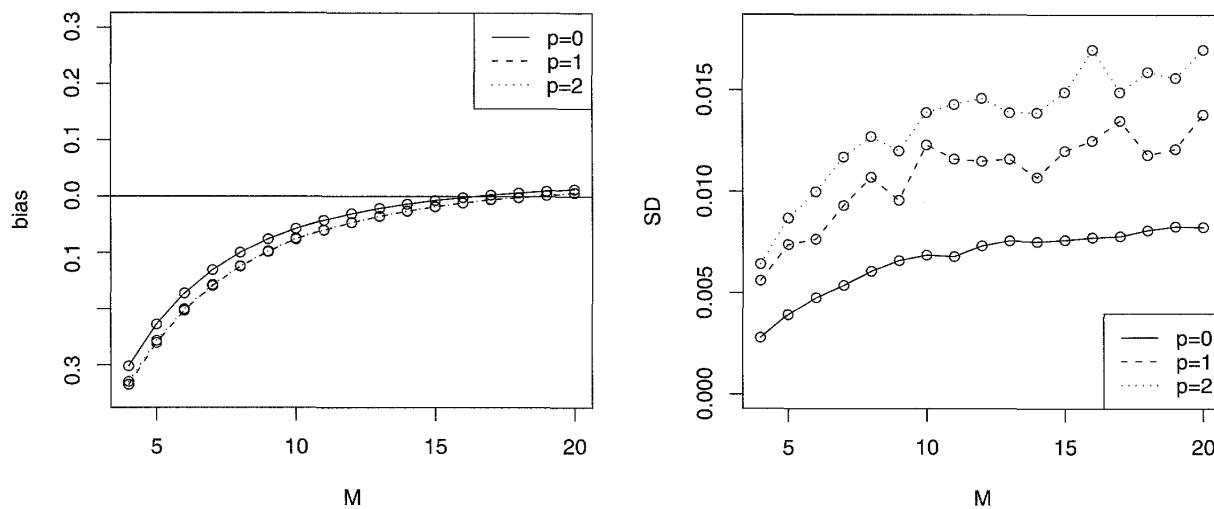


図 3 パレート標本に対する KDS 最尤推定量のバイアス (bias) と標準偏差 (SD)

図 3 の右のグラフは、各 p についていかに標準偏差が M に影響を受けるかを示している。このグラフから、正規分布の場合と同様に、各 M について p とともに標準偏差が増加することが分かる。一方、標準偏差は各 p について M とともに緩やかに増加しているよう見える。これは正規分布では見られなかった特徴である。図 4 に表示される平均平方誤差平方根のグラフから、パレート分布の場合でも、 M を 15 以上にすれば、効率性はかなり改善されることを示している。

5. 結語

本稿では、データスカッシングの新たな手法として KDS を提案した。既存の手法と異なり、我々の手法は実装が容易である。KDS 最尤推定量の漸近的性質を考察し、多項式カーネルを用いることによって統計的情報が十分に保存されることを示した。我々が知る限りにおいて、この論文はデータスカッシングの実際の効率性を理論的に解明した最初の論文

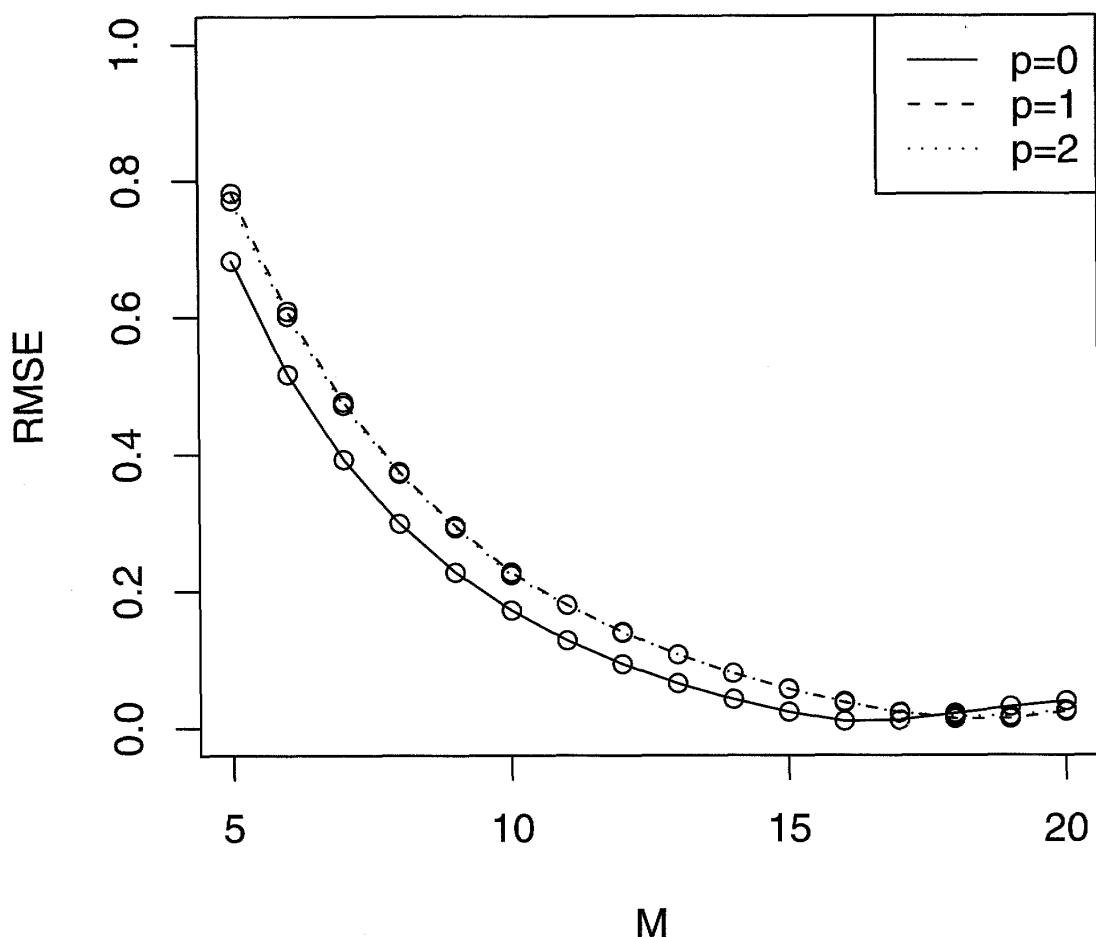


図4 パレート標本に対するKDS最尤推定量の平均平方誤差平方根(RMSE)

である。シミュレーションによる考察も行ったが、その結果は漸近的結果とほぼ一致するものであった。

本稿における議論は単一変量に限定された。実際には多変量データを扱うことが一般的である。ここでは、Hall and Wand (1996) の研究にならって、KDSの多変量版の一案を示唆する。いま d 次元の分布からの大量データ $(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N)$ を考える。このとき、 $\mathbf{j} = (j_1, j_2, \dots, j_d)$ を各要素が任意の整数である d -次元ベクター、 $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_d)$ を各要素が bin幅の d -次元ベクターとし、スカッシュ化データ点 $s_{\mathbf{j}}$ を

$$s_{\mathbf{j}} = (j_1 \delta_1, j_2 \delta_2, \dots, j_d \delta_d).$$

と定義する。対応する多変量 KDS の加重 $w_{\mathbf{j}}$ は単一変量 KDS の加重の積として以下のように定義する。各 $1 \leq m \leq d$ に対して、 w_{jm} を単一変量の KDS 加重

$$w_{jm} \equiv \frac{1}{Nh_m} \sum_{i=1}^N K_m \left(\frac{j_m \delta_m - X_i}{h_m} \right) \text{ with } h_m = \delta_m / 2$$

とし、

$$\mathbf{w}_j = \prod_{k=1}^d w_{jm}$$

と置く。このとき、多変量 KDS の対数尤度は

$$l(\theta) \equiv \sum_{j \in J} \mathbf{w}_j \log(f(s_j | \theta))$$

と定義される。ここで、 J はすべての j の集まりとする。

单一変量の場合と同様にして、多変量 KDS に対する漸近的性質を導くことができるかもしれない。しかし、 d の増大に連れて出現する「次元の呪い」³⁾によって、このような多変量 KDS のパフォーマンスは満足できるような水準とはならないかもしれない。

謝辞

この研究は科研費（No.16530216 及び No.19530285）による支援を受けた。

A. 補遺

A.1 定理 3.1 の証明

$Q_j^* = \sqrt{\frac{2j+1}{2}} Q_j$ とすると、 $\{Q_j^*\}$ は $[-1, 1]$ 上で直交基底となる：

$$\int_{-1}^1 Q_j^*(z) Q_{j'}^*(z) dz = \begin{cases} 1 & \text{for } j = j' \\ 0 & \text{for } j \neq j' \end{cases}.$$

このとき、 $t = 0, 1, \dots, p-1, p$ 及び $-1 \leq u \leq 1$ に対して、 u^t は

$$u^t = \sum_{j=0}^p \left(\int_{-1}^1 z^t Q_j^*(z) dz \right) Q_j^*(u)$$

と表現できる。上の方程式で $u = 0$ と置くと以下が成立する：

$$0^t = \sum_{j=0}^p \left(\int_{-1}^1 z^t Q_j^*(z) dz \right) Q_j^*(0) = \sum_{r=0}^m \left(\int_{-1}^1 z^t Q_{2r}^*(z) dz \right) Q_{2r}^*(0)$$

3) 例えば、Scott (1992) 7 章を見よ。

なぜならば, j が奇数ならば $Q_j^*(0) = 0$ となるからである. したがって,

$$\begin{aligned} \int_{-1}^1 u^t W_p(u) du &= \int_{-1}^1 u^t \sum_{r=0}^m \left(2r + \frac{1}{2} \right) Q_{2r}(0) Q_{2r}(u) du \\ &= \int_{-1}^1 u^t \sum_{r=0}^m Q_{2r}^*(0) Q_{2r}^*(u) du \\ &= \sum_{r=0}^m \left(\int_{-1}^1 u^t Q_{2r}^*(u) du \right) Q_{2r}^*(0) = 0^t = \begin{cases} 1 & \text{for } t = 0 \\ 0 & \text{for } 1 \leq t \leq p. \end{cases} \end{aligned}$$

となる. ここで, $p+1$ は奇数であり W_p が原点に関して左右対称であることから, $\int_{-1}^1 u^{p+1} W_p(u) du = 0$ となることに注意されたい. また, 直接的な計算から $\int_{-1}^1 u^{p+2} W_p(u) du \neq 0$. となる. \square

A.2 定理 3.2 の証明

(3.5) によって, 任意の $\varepsilon > 0$ に対してある $\eta > 0$ が存在し $|\theta - \theta_0| \geq \varepsilon$ ならば $G(\theta_0) - G(\theta) > \eta$ となる. したがって,

$$\Pr(|\tilde{\theta} - \theta_0| \geq \varepsilon) \leq \Pr(G(\theta_0) - G(\tilde{\theta}) > \eta).$$

が成立する. ここで, $\tilde{G}(\theta)$ を

$$\tilde{G}(\theta) \equiv \sum_{j=1}^M w_j \log \frac{f(s_j|\theta)}{f(s_j|\theta_0)}, \quad \theta \in \Theta.$$

と定義すると

$$\begin{aligned} G(\theta_0) - G(\tilde{\theta}) &= \tilde{G}(\theta_0) - G(\tilde{\theta}) + G(\theta_0) - \tilde{G}(\theta_0) \\ &\leq \tilde{G}(\tilde{\theta}) - G(\tilde{\theta}) + G(\theta_0) - \tilde{G}(\theta_0) \leq 2 \sup_{\theta} |\tilde{G}(\theta) - G(\theta)| \end{aligned}$$

を得る. したがって, 次を示せば十分である:

$$\sup_{\theta} |\tilde{G}(\theta) - G(\theta)| \rightarrow 0 \quad \text{in Probability.} \quad (\text{A.2.1})$$

(A.2.1) を示すために, 以下を仮定する:

1. $f(x|\theta_0)$ は x に関して 2 回連続微分可能であり一様有界である
2. ある $0 < \beta < 1/2$ に対して $\sum |f''(x|\theta_0)| / (f(x|\theta_0))^{\beta} < \infty$.
3. 各 θ に対して, $f(x|\theta)$ は x に関して微分可能であり, 以下を満たす:

$$\begin{aligned} \sup_{\theta} \int \left| (f(x|\theta_0))^{\beta} \log \left(\frac{f(x|\theta)}{f(x|\theta_0)} \right) \right| dx &< \infty, \\ \sup_{\theta} \int \left| \left((f(x|\theta_0))^{\beta} \log \left(\frac{f(x|\theta)}{f(x|\theta_0)} \right) \right)' \right| dx &< \infty \end{aligned} \quad (\text{A.2.2})$$

ここで'は x に関する微分を表す。

(A.2.1) の成立を示すために

$$\begin{aligned}
 |\tilde{G}(\theta) - G(\theta)| &\leq \left| \sum_j \delta f(s_j|\theta_0) \log \frac{f(s_j|\theta)}{f(s_j|\theta_0)} - \int f(x|\theta_0) \log \frac{f(x|\theta)}{f(x|\theta_0)} dx \right| \\
 &+ \left| \sum_j \left(\frac{\hat{f}(s_j) - f(s_j|\theta_0)}{(f(s_j|\theta_0))^\beta} \right) \times \delta(f(s_j|\theta_0))^\beta \log \frac{f(s_j|\theta)}{f(s_j|\theta_0)} \right| \\
 &\leq \left| \sum_j \int_{B_j} \left(f(s_j|\theta_0) \log \frac{f(s_j|\theta)}{f(s_j|\theta_0)} - f(x|\theta_0) \log \frac{f(x|\theta)}{f(x|\theta_0)} \right) dx \right| \\
 &+ \sup_x \left| \frac{\hat{f}(x) - f(x|\theta_0)}{(f(x|\theta_0))^\beta} \right| \times \sum_j \delta \left| (f(s_j|\theta_0))^\beta \log \frac{f(s_j|\theta)}{f(s_j|\theta_0)} \right|, \quad (\text{A.2.3})
 \end{aligned}$$

となることに注意する。各 j について Friedman and Diaconis (1981) の Lemma 2.2 を用いると、

$$\int_{B_j} \left(f(s_j|\theta_0) \log \frac{f(s_j|\theta)}{f(s_j|\theta_0)} - f(x|\theta_0) \log \frac{f(x|\theta)}{f(x|\theta_0)} \right) dx \leq \delta \int_{B_j} \left| \left(f(x|\theta_0) \log \frac{f(x|\theta)}{f(x|\theta_0)} \right)' \right|$$

が成立し、(A.2.3) の右側の第1項は

$$\begin{aligned}
 &\left| \sum_j \int_{B_j} \left(f(s_j|\theta_0) \log \frac{f(s_j|\theta)}{f(s_j|\theta_0)} - f(x|\theta_0) \log \frac{f(x|\theta)}{f(x|\theta_0)} \right) dx \right| \\
 &\leq \delta \int \left| \left(f(x|\theta_0) \log \frac{f(x|\theta)}{f(x|\theta_0)} \right)' \right| dx
 \end{aligned}$$

となる。同様にして

$$\begin{aligned}
 \left| \sum_j \delta (f(s_j|\theta_0))^\beta \log \frac{f(s_j|\theta)}{f(s_j|\theta_0)} \right| &\leq \int \left| (f(s_j|\theta_0))^\beta \log \frac{f(x|\theta)}{f(x|\theta_0)} \right| dx \\
 &+ \delta \int \left| \left((f(s_j|\theta_0))^\beta \log \frac{f(x|\theta)}{f(x|\theta_0)} \right)' \right| dx.
 \end{aligned}$$

が成立する。また

$$\sup_x \left| \frac{\hat{f}(x) - f(x|\theta_0)}{(f(x|\theta_0))^\beta} \right| \leq \sup_x \left| \frac{\hat{f}(x) - E[\hat{f}(x)]}{(f(x|\theta_0))^\beta} \right| + \sup_x \left| \frac{E[\hat{f}(x)] - f(x|\theta_0)}{(f(x|\theta_0))^\beta} \right|$$

となる。Gine et al. (2004) の定理 2.1 を用いると、上方程式的第1項は $O\left(\delta + \delta^{-1}\sqrt{\frac{\log N}{N}}\right)$ であることが示される。 $\sup_x |f''(x|\theta)/f(x|\theta)^\beta| < \infty$ という仮定よ

り、2番目の項は

$$\begin{aligned} \sup_x \left| \frac{\mathbb{E}[\hat{f}(x) - f(x|\theta_0)]}{(f(x|\theta_0))^\beta} \right| &= \sup_x \left| \int_{-\infty}^{\infty} K(u) \frac{f(x-hu|\theta_0) - f(x|\theta_0)}{(f(x|\theta_0))^\beta} du \right| \\ &= \sup_x \left| \int_{-\infty}^{\infty} K(u) \frac{-huf'(x|\theta_0) + (1/2)h^2u^2f''(x_*)}{(f(x|\theta_0))^\beta} du \right| = O(\delta^2) \end{aligned}$$

となる。ここで、 x^* は $x - hu$ と x の間の値である。これらの結果をあわせると

$$\begin{aligned} |\tilde{G}(\theta) - G(\theta)| &\leq \delta \int \left| \left(f(x|\theta_0) \log \frac{f(x|\theta)}{f(x|\theta_0)} \right)' \right| dx \\ &\quad \times \left[\int \left| \log \frac{f(x|\theta)}{f(x|\theta_0)} \right| dx + \delta \int \left| \left(\log \frac{f(x|\theta)}{f(x|\theta_0)} \right)' \right| dx \right] \\ &\quad \times O\left(\delta^2 + \delta^{-1} \sqrt{\frac{\log N}{N}}\right), \end{aligned}$$

となり、(A.2.1) が示された。

A.3 定理 3.3 の証明

θ_0 を θ の真の値とし、

$$\tilde{l}'(\theta_0) = \frac{\partial}{\partial \theta} \sum_j w_j \log f(j\delta|\theta) \Big|_{\theta=\theta_0}.$$

とする。テーラー展開

$$0 = \tilde{l}'(\tilde{\theta}) \sim \tilde{l}'(\theta_0) + (\tilde{\theta} - \theta)\tilde{l}''(\theta_0)$$

より

$$\sqrt{N}(\tilde{\theta} - \theta_0) \sim \frac{-N^{-1/2}\tilde{l}'(\theta_0)}{N^{-1}\tilde{l}''(\theta_0)}.$$

となる。ここで

$$\begin{aligned} \mathbb{E}[w_j/\delta] &= \mathbb{E} \left[\frac{1}{Nh} \sum_{i=1}^N K \left(\frac{j\delta - X_i}{h} \right) \right] \\ &= \frac{1}{h} \int K \left(\frac{j\delta - x}{h} \right) f(x|\theta_0) dx \\ &= \int K(u) f(j\delta + uh|\theta_0) du \\ &= \int K(u) \left[f(j\delta|\theta_0) + uhf'(j\delta|\theta_0) + \frac{h^2}{2}u^2f''(j\delta|\theta_0) + \dots \right] du \\ &= f(j\delta|\theta_0) + \frac{h^{2m}}{(2m)!} \mu_{2m}(K) f^{(2m)}(j\delta|\theta_0) + O(h^{2m+1}) \end{aligned}$$

となるから

$$\begin{aligned}
 E[N^{-1}\tilde{l}'(\theta_0)] &= \sum_j \delta \frac{\partial}{\partial \theta} \log f(j\delta|\theta_0) E \left[\frac{1}{Nh} \sum_{i=1}^N K \left(\frac{j\delta - X_i}{h} \right) \right] \\
 &\sim \delta \sum_j \frac{\partial}{\partial \theta} \log f(j\delta|\theta_0) \\
 &\quad \times \left[f(j\delta|\theta_0) + \frac{h^{2m}}{(2m)!} \mu_{2m}(K) f^{(2m)}(j\delta|\theta_0) + o(h^{2m}) \right] \\
 &= J_0 + \frac{h^{2m}}{(2m)!} \mu_{2m}(K) J_{2m} + o(h^{2m})
 \end{aligned} \tag{A.3.4}$$

が成立する。ただし、 J_0 及び J_{2m} は

$$J_0 \equiv \delta \sum_j f(j\delta|\theta_0) \frac{\partial}{\partial \theta} \log f(j\delta|\theta_0), \quad J_{2m} \equiv \delta \sum_j f^{(2m)}(r\delta|\theta_0) \frac{\partial}{\partial \theta} \log f(j\delta|\theta_0)$$

と定義される。さらに、オイラー・マクローリンの公式⁴⁾を用いると

$$\begin{aligned}
 J_0 &= \int \frac{\partial}{\partial \theta} f(x|\theta_0) \log f(x|\theta_0) dx \\
 &\quad + \frac{\delta^{2t}}{(2t)!} B_{2m+1} \int \left(f(x|\theta_0) \frac{\partial}{\partial \theta} \log f(x|\theta_0) \right)^{(2m+1)} dx + o(\delta^{2m+1}) \\
 &= O(\delta^{2m+1})
 \end{aligned}$$

が成立する。ここで、 B_{2m+1} is the $2m+1$ 次のベルヌイ数である。同様にして

$$J_{2m} \sim \int f^{(2m)}(x|\theta_0) \frac{\partial}{\partial \theta} \log f(x|\theta_0) dx.$$

を得る。したがって、 $t > m$ に対して：

$$E[N^{-1/2}\tilde{l}'(\theta_0)] \sim N^{1/2} \frac{h^{2m}}{(2m)!} \mu_{2m}(K) J_{2m}$$

となる。また、

$$\begin{aligned}
 N^{-1}\tilde{l}''(\theta_0) &= \sum_j \delta \frac{1}{Nh} \sum_{i=1}^N K \left(\frac{j\delta - X_i}{h} \right) \frac{\partial^2}{\partial \theta^2} \log f(j\delta|\theta_0) \\
 &\rightarrow \sum_j \delta E \left[\frac{1}{h} K \left(\frac{j\delta - X_i}{h} \right) \right] \frac{\partial^2}{\partial \theta^2} \log f(j\delta|\theta_0) \quad (N \rightarrow \infty) \\
 &\approx \int \left(\frac{\partial^2}{\partial \theta^2} \log f(x|\theta_0) \right) f(x|\theta_0) dx = -I(\theta_0)
 \end{aligned} \tag{A.3.5}$$

が成立する。ここで、 $I(\theta)$ はフィッシャー情報量：

$$I(\theta) \equiv \int \left(\frac{\partial}{\partial \theta} \log f(x|\theta) \right)^2 f(x|\theta) dx.$$

⁴⁾ 例えば Lange (1999) を見よ。

である. (A.3.4) 及び (A.3.5) によって最終的に次式を得る:

$$\begin{aligned} \mathrm{E}[\sqrt{N}(\tilde{\theta} - \theta_0)] &\sim N^{1/2} \frac{h^{2m}}{(2m)!} \mu_{2m}(K) J_{2m} \frac{1}{I(\theta_0)} \\ &= O(N^{1/2} h^{2m}) \end{aligned}$$

A.4 定理 3.4 の証明

まず

$$\mathrm{Var}(w_j/\delta) = \mathrm{Var}\left(\frac{1}{Nh} \sum_{i=1}^N K\left(\frac{j\delta - X_i}{h}\right)\right) \sim \frac{1}{Nh} R(K) f(j\delta|\theta_0)$$

となる. また

$$\begin{aligned} \tilde{l}'(\theta) &= \sum_j w_j \frac{\partial}{\partial \theta} \log f(j\delta|\theta) \\ &= \delta \frac{1}{N} \sum_{i=1}^N \sum_j \frac{\partial}{\partial \theta} \log f(j\delta|\theta) \frac{1}{h} K\left(\frac{j\delta - X_i}{h}\right) \end{aligned}$$

より

$$\begin{aligned} \mathrm{Var}(\sqrt{N}\tilde{l}'(\theta_0)) &= \delta^2 \mathrm{Var}\left(\sum_j \frac{\partial}{\partial \theta} \log f(j\delta|\theta) \frac{1}{h} K\left(\frac{j\delta - X_1}{h}\right)\right) \\ &= \delta^2 \sum_j \left(\frac{\partial}{\partial \theta} \log f(r\delta|\theta)\right)^2 \mathrm{Var}\left(\frac{1}{h} K\left(\frac{j\delta - X_1}{h}\right)\right) \\ &= 2 \sum_r \left(\frac{\partial}{\partial \theta} \log f(j\delta|\theta_0)\right)^2 R(K) f(j\delta|\theta_0) \delta \\ &\sim 2R(K) \int \left(\frac{\partial}{\partial \theta} \log f(x|\theta_0)\right)^2 f(x|\theta_0) dx = 2R(K) I(\theta_0) \end{aligned}$$

が成立する. この最後の式と (A.2.2) から

$$\mathrm{Var}(\sqrt{n}(\tilde{\theta} - \theta_0)) \approx \frac{2R(K)}{I(\theta_0)}$$

が成立する.

A.5 (3.7) の証明

以下を証明すれば十分である:

$$\frac{1}{N} \sum_{i=1}^N \left(\frac{X_i - j\delta}{h}\right)^t \mathbb{I}\left\{\left|\frac{X_i - j\delta}{h}\right| \leq 1\right\} = \frac{1}{N} \sum_{i=1}^N \int_{B_j} \left(\frac{x - j\delta}{h}\right)^t \tilde{f}(x) dx. \quad (\text{A.5.6})$$

ここで, $\{Q_t^*\}_{t \in \mathbb{Z}}$ $[-1, 1]$ 上で正規直交基底となる:

$$\int_{-1}^1 Q_t^*(z) Q_r^*(z) dz = \begin{cases} 1 & t = r \\ 0 & t \neq r \end{cases}.$$

このとき, $t = 0, 1, \dots, p$ 及び $-1 \leq u \leq 1$ に対して, u^s は以下のように表現できる

$$\begin{aligned} u^t &= \sum_{t=0}^p \left(\int_{-1}^1 z^t Q_t^*(z) dz \right) Q_t^*(u) \\ &= \sum_{t=0}^p \left(\int_{B_j} \left(\frac{x - j\delta}{h} \right)^t Q_t^* \left(\frac{x - j\delta}{h} \right) dx \right) \frac{1}{h} Q_t^*(u) \\ &= \int_{B_j} \left(\frac{x - j\delta}{h} \right)^t \left[\sum_{r=0}^p Q_r^* \left(\frac{x - j\delta}{h} \right) \frac{1}{h} Q_r^*(u) \right] dx. \end{aligned}$$

ここで

$$u \equiv \frac{X_i - j\delta}{h} \mathbb{I} \left\{ \left| \frac{X_i - j\delta}{h} \right| \leq 1 \right\},$$

とおくと, 上の方程式は

$$\begin{aligned} &\left(\frac{X_i - j\delta}{h} \right)^t \mathbb{I} \left\{ \left| \frac{X_i - j\delta}{h} \right| \leq 1 \right\} \\ &= \int_{j\delta-h}^{j\delta+h} \left(\frac{x - j\delta}{h} \right)^t \left[\sum_{r=0}^p Q_r^* \left(\frac{x - j\delta}{h} \right) \frac{1}{h} Q_r^* \left(\frac{X_i - j\delta}{h} \right) \right] dx \end{aligned}$$

となる. この式を $X'_i s$ に関して和を取れば, 以下を得る:

$$\begin{aligned} &\frac{1}{N} \sum_{i=1}^N \left(\frac{X_i - j\delta}{h} \right)^t \mathbb{I} \left\{ \left| \frac{X_i - j\delta}{h} \right| \leq 1 \right\} \\ &= \frac{1}{N} \sum_{i=1}^N \int_{B_j} \left(\frac{x - j\delta}{h} \right)^t \sum_{r=0}^p Q_r^* \left(\frac{x - j\delta}{h} \right) \frac{1}{h} Q_r^* \left(\frac{X_i - j\delta}{h} \right) dx \end{aligned}$$

これは, (A.5.6) である. □

参考文献

- Dimakos, X. K. (2000). Data squashing for tail inference in the generalized Pareto distribution, Norwegian Computing Center Report No.957.
- DuMouchel, W. (2002). Data Squashing: Constructing summary data sets, in *Handbook of Massive Data Sets*, Kluwer Academic Publishers, 579–591.
- DuMouchel, W., Volinsky, C., Johnson, T., Cortes, C. and Pregibon, D. (1999). Squashing flat files flatter, in *Proceeding of the Fifth International Conference on Knowledge Discovery and Data Mining*, 6–15.
- Freedman, D. and Diaconis, D. (1981). On the histogram as a density estimator: L_2 theory, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **57**, 453–476.
- Gine, E., Koltchinskii, V. and Zinn, J. (2004). Weighted uniform consistency of kernel density estimators, *Ann. Probab.*, **32**, 2570–2605.
- Hall, P. and Wand, M. P. (1996). On the accuracy of binned kernel density estimators, *J. Multivar. Anal.*, **56**, 165–184.
- Kogure, A. (1998). Effective Interpolations for Kernel Density Estimators, *J. Nonparametr. Stat.*, **9**, 165–195.
- Lange, K. (1999). *Numerical Analysis for Statisticians*, Springer, New York.
- Madigan, D., Raghavan, N., Dumouchel, W., Nason, M., Posse, C. and Ridgeway, G. (2002). Likelihood-based data squashing: A modeling approach to instance construction, *Data Min. Knowl. Disc.*, **6**, 173–190.

- Sagae, M. and Scott, D. (1997). Bin Interval Method of Locally Adaptive Nonparametric Density Estimation, Statistics technical report series of RICE University, 1–21.
- Scott, D. W. (1992), *Multivariate Density Estimation: Theory, Practice, and Visualization*, Wiley, New York.
- 鈴木英之進 (2005). 「データスカッシング—逆転の発想によるスケールダウン戦略—」 『情報処理』 **46**, 12–19.
- Tallis, G. M. (1967). Approximate maximum likelihood estimates from grouped data, *Technometrics*, **9**, 599–606.
- Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*, Chapman & Hall.
- Wasserman, L. (2004) *All of Statistics: A Concise Course in Statistical Inference*, Springer.