Asymptotic Properties of Histogram Smoothing Using a Cubic Spline Function～ Theoretical Equivalence between Boneva, Kendall, and Stefanov Model and Lii and Rosenblatt Model～

3次スプライン関数によるヒストグラム平滑化とその漸近的性質
～ Boneva, Kendall and Stefanov型とLii and Rosenblatt型モデルの理論的同等性～

金沢大学大学院　人間社会環境研究科　人間社会環境学専攻
齊　藤　実　祥
金沢大学　人間社会研究域　経済学経営学系
寒河江　雅　彦

要旨

ヒストグラムはデータの構造を把握するための最も簡単な統計量としてよく知られている。他方で，欠点として不連続であることが指摘される。この問題の解消のため，スプライン平滑化を考える。ヒストグラムのスプライン平滑化に関して，Boneva, Kendall and Stefanov（1971）（以下，BKS）がヒストスプラインを提案し，Schoenberg（1972）が定式化した。しかしながら，BKSとSchoenbergはモデルの提案に留まり，理論的性質については言及していない。他方で，Lii and Rosenblatt（1974）（以下，L＆R）はBKSとSchoenbergと異なる3次スプライン平滑化によるヒストグラムを提案し，その漸近的性質を導出した。その中で，漸近積分分散が$O(\frac{1}{nh})$，漸近積分二乗バイアスが$O(h^6)$となることと，漸近正規性が成り立つことを示している。しかしながら，L&RはBKSとSchoenbergとの差異については言及していない。また，推定量の明示的な表現までは導いていない。本研究では，BKS＋SchoenbergとL&Rの２つの問題について同等性を示し，推定量について正確な漸近表現を導く。有限標本時の特性に関しては，ISEの標本平均と標準偏差について数値実験を行い，ヒストグラムとヒストスプラインの推定精度について比較する。

以上２つの未解決な問題に関して議論する。最初にBKS＋SchoenbergとL&Rの推定量が同等であることを示した。次に，推定量のAMISEは分散項が$\frac{5\sqrt{3}+3}{10}\frac{1}{nh}$，二乗バイアス項が$\frac{R(f''')}{30420}h^6$と表されることを示した。ヒストグラムのAMISEと比較すると，分散は大きい一方で，二乗バイアスが小さいことが明らかになった。更に，ヒストスプライン推定量の平均積分誤差の上限と，漸近正規性を証明した。

数値実験の結果，ヒストグラムと比較してヒストスプラインの方が標本サイズに関わらずISE値が小さかった。一方で，ISE標準偏差については，どの標本サイズでもヒストグラムの方が小さく，大標本特性を裏付ける結果となった。この結果から，ヒストスプラインの分散は大きくなるが，バイアスを減少させる効果の方が大きく，全体の推定精度としては改良されることが理論と数値実験で明らかになった。

キーワード

ヒストグラム，平滑化，スプライン

Asymptotic Properties of Histogram Smoothing Using a Cubic Spline Function

～ Theoretical Equivalence between Boneva, Kendall, and Stefanov Model and Lii and Rosenblatt Model ～

Division of Human and Socio-Environmental Studies
Graduate School of Human and Socio-Environmental Studies, Kanazawa University
SAITO Misaki

Faculty of Economics and Management
Institute of Human and Social Sciences, Kanazawa University
SAGAE Masahiko

Abstract

Histograms are discontinuous between adjacent bins. We consider histogram smoothing using a cubic spline function. Boneva, Kendall, and Stefanov (BKS) (1971) proposed the histospline and Schoenberg (1972) formulated it. However, they did not show asymptotic properties of histogram smoothing estimate using spline functions. In related research, Lii and Rosenblatt (L&R) (1974) set different conditions from BKS to apply a cubic function for histogram smoothing and derived asymptotic properties. They showed that the asymptotic integrated variance(AIV) and the asymptotic integrated squared bias(AISB) for estimate are $O(\frac{1}{nh})$ and $O(h^6)$, respectively. However, they did not mention the theoretical equivalence to the proposal by BKS. In addition, they did not show explicit AIV and AISB. Therefore, we reveal whether there is a theoretical equivalence between the BKS and L&R models. We also derive an explicit AIV and AISB of the estimate in the BKS and L&R models.

As a result, the BKS and L&R models were found to have the same equation and the AMISE of the estimate has $\frac{5\sqrt{3}+3}{10}\frac{1}{nh}$ for AIV and $\frac{R(f''')}{30420}h^6$ for AISB. This shows that the histospline has a larger AIV and smaller AISB than the histogram. We also showed the explicit mean and variance related to the asymptotic normality of the estimate.

To examine histograms and histogram smoothing by cubic spline functions in finite samples, we compare numerical experiment of sample means and standard deviations of ISEs. The numerical experiment indicated that the ISEs of the histospline were smaller, but the standard deviations of the ISEs were larger than those of the histogram. In other words, we can enjoy a significant decrease in the bias of histogram smoothing, while its variance increases. The entire ISE of a histogram smoothing estimate overcomes that of the histogram.

Keyword

Histograms, Smoothing, Spline functions

1．研究背景と目的

ノンパラメトリック密度関数の代表的な推定法に，ヒストグラムが挙げられる。ここで「ヒストグラム」とは，ヒストグラム型密度関数を指す。ヒストグラム型密度関数とは，各分割区間（以降,

ビンと呼ぶ）に入る度数データに比例した高さを持つ連続分布のことである。

ヒストグラムは，ビンごとに区分的定数関数である。そのため，隣接ビンの間では不連続となる。この不連続性の問題に対して，Scott（1985）が各ビンの中点を直線で結んだものを推定量とするFrequency Polygon（以降，FPと呼ぶ）を提案した。その中で，FPがヒストグラムの推定精度を改良できることを示している。FPの関連研究として，Minnotte（1996）が各ビンの中点を節点とし，その高さを各ビンの面積相等性[1)]を満たすように決定するBias-Optimized Frequency Polygon（以降，BFPと呼ぶ）を提案している。また，Jones, Samiuddin, Al-Harbey and Maatouk（1998）がビンの端点を節点とし，隣接ビンの高さの中点を節点の高さとするEdge Frequency Polygon（以降, EFPと呼ぶ）がある。FP，BFP，EFPともに隣接ビン同士を一次関数で接続することによってヒストグラムの不連続性を解消する手法である。

ここで，2次以上の滑らかな曲線で隣接ビン間を接続するために，ヒストグラムを3次スプライン関数によって平滑化することが考えられる。スプライン関数とは，多項式を何らかの連続条件を満たすように接続する区分的多項式であり，点同士を滑らかな曲線で繋ぐことができる。スプライン関数はSchoenberg（1946）の提案以降，盛んに研究が行われており，その数学的な性質が明らかとなっている。ヒストグラムのスプライン平滑化に関しては，Boneva, Kendall and Stefanov（1971）（以降，BKSと略す）がヒストスプラインを提唱し，Schoenberg（1972）が定式化した。しかしながら，BKSとSchoenbergはモデルの提案に留まり，理論的性質については導出していない。一方，Lii and Rosenblatt（1974）（以降，L&Rと略す）がBKSとSchoenbergとは異なるアプローチでヒストグラムをスプライン平滑化し，その漸近的性質について導出した。その中で，スプライン平滑化したヒストグラムは漸近的にバイアス$O(h^3)$，分散$O(1/nh)$であることが示された。しかしながら，L&RはBKS＋Schoenbergとの差異もしくは同等性については明示していない。また，平均積分二乗誤差（以降，MISEと呼ぶ）について，定数項を含む明示的な表現までは導いていない。

以上より，本稿では上記のBKS＋SchoenbergとL&Rのヒストグラムのスプライン平滑化は同等であることを示し，ヒストスプライン推定量の漸近表現を陽な形で導く。加えて，その漸近正規性を示す。また，有限標本におけるヒストグラムとヒストスプラインの推定精度を比較するため，数値実験を行う。

2．BKS＋SchoenbergとL&Rの設定の違いについて

2.1. Boneva, Kendall and Stefanov＋Schoenbergの設定

まず，BKS＋Schoenbergによるヒストグラムの3次スプライン平滑化の定式化について説明する。スプライン関数はヒストグラムの累積分布関数に対応する。サンプル数n，区間[0,1]で等間隔の節点$x_j\,(j=0,1,...,N)$を決め，度数$v_j \in [x_{j-1}, x_j)\,(j=1,2,...,N)$のヒストグラムの面積$s_j = \frac{v_j}{n}$ $(j = 1, 2, \dots, N)$を得る。ヒストグラムについて累積経験分布関数$G_j\,(j=0,1,...,N)$は以下の通りに与えられる。

$$\begin{cases} G_0 = 0, \\ G_i = \sum_{j=1}^{i} s_j \quad (i = 1,2, \dots, N). \end{cases} \tag{2.1}$$

このとき

$$S(x_j) = G_j \quad (j = 0,1,...,N),$$

を満たす3次スプライン関数$S(x)$が存在する。$S_{BKS}(x)$は以下の制約条件のもとで決定される。

1. 面積相等性：$\int_{x_{j-1}}^{x_j} \hat{f}(x)dx = S_{BKS}(x_j) - S_{BKS}(x_{j-1}) = G_j - G_{j-1} = s_j$
2. 1次導関数の連続性：$S'_{BKS}(x_j -) = S'_{BKS}(x_j +)$
3. 2次導関数の連続性：$S''_{BKS}(x_j -) = S''_{BKS}(x_j +)$
4. 端条件：$S'_{BKS}(x_0) = S'_{BKS}(x_N) = 0$

ただし，$\hat{f}(x)$ はスプライン平滑化したヒストグラムの密度推定量，$S'(x_j)$ は節点 x_j における $S(x)$ の１次微係数，$S''(x_j)$ は節点 x_j における $S(x)$ の２次微係数，$S'(x_j-)$ は，節点 x_j における $S(x)$ の左方微分係数，$S'(x_j+)$ は右方微分係数である。

ここで，まず$S_{BKS}(x)$について，以下のように表わされる。

$$S_{BKS}(x)=m_{j-1}\frac{(x_j-x)^2(x-x_{j-1})}{h^2}-m_j\frac{(x-x_{j-1})^2(x_j-x)}{h^2}+$$
$$G_{j-1}\frac{(x_j-x)^2[2(x-x_{j-1})+h]}{h^3}+G_j\frac{(x-x_{j-1})^2[2(x_j-x)+h]}{h^3}, \quad (2.2)$$

ただし，m_j（j=0,1,...,N）は節点x_jにおける$S(x)$の１次微係数（=$S'(x_j)$），h=x_j-x_{j-1}（=ヒストグラムのビン幅）である。

（2.2）式の微分によってスプライン平滑化したヒストグラムの密度推定量$\hat{f}(x)$は得られ，ビン B_j, $x\in[x_{j-1}, x_j)$ において次の表現を得る：

$$\hat{f}_j(x)=m_{j-1}\frac{(x_j-x)(2x_{j-1}+x_j-3x)}{h^2}-$$
$$m_j\frac{(x-x_{j-1})(2x_j+x_{j-1}-3x)}{h^2}+$$
$$\frac{6s_j}{h^3}(x_j-x)(x-x_{j-1}). \quad (2.3)$$

ただし，$\hat{f}_j(x)$は$x\in B_j$を意味する。

（2.3）式でm_jは未知であるため，$\hat{f}_j(x)$の導関数の連続性$S''_{BKS}(x_j-)=S''_{BKS}(x_j+)$ から，

$$\frac{1}{6}m_{j-1}+\frac{2}{3}m_j+\frac{1}{6}m_{j+1}=\frac{s_j+s_{j+1}}{2h}$$
$$=\frac{v_j+v_{j+1}}{2nh}\quad (j=1,2,\dots,N-1), \quad (2.4)$$

をm_jについて解く。

m_jについて解いた$\hat{f}_j(x)$ の表現は

$$\hat{f}_j(x)=\frac{1}{h^3}\left\{-2h(x_j-x)+3(x_j-x)^2\right\}\sum_{k=1}^{N-1}w_{j-1,k}\left(\frac{s_k+s_{k+1}}{2}\right)$$
$$+\frac{1}{h^3}\left\{h^2-4(x_j-x)+3(x_j-x)^2\right\}\sum_{l=1}^{N-1}w_{j,l}\left(\frac{s_l+s_{l+1}}{2}\right)$$
$$-\frac{6}{h^3}\left\{-h(x_j-x)+(x_j-x)^2\right\}s_j, \quad (2.5)$$

ただし，$w_{j,l}$は重み$\sum_{l=1}^{N-1}w_{j,l}=1$で，十分大きな$N$において，

$$w_{j,l}=\frac{3}{\sqrt{3}}\left(\sqrt{3}-2\right)^{|j-l|},$$

である。$w_{j,l}$の導出については文献(4)を参照のこと。

2.2. Lii and Rosenblattの設定

L＆Rによるヒストグラムの３次スプライン平滑化の定式化について説明する。L&Rは，ヒストグラムの累積分布関数の推定量として３次スプライン関数を使用する仮定で制約条件を決定している。2.1節と同じく，サンプル数n, 区間[0,1]で，等間隔の節点x_j (j=0,1,...,N) はビンの端点とする。３次スプライン関数$S(x)$の２次導関数$S''(x)$は線形となることから，以下の$S''(x)$の連続性をまず

$$S''_{LR}(x)=M_{j-1}\frac{x_j-x}{h}+M_j\frac{x-x_{j-1}}{h}, \quad (2.6)$$

制約条件として設定する。

ただし，M_jは節点x_jにおける$S(x)$の２次微係数である。x_jにおけるヒストグラムの累積分布関数の高さをG_jとし，（2.6）式を２回積分して，条件$S(x_{j-1})=G_{j-1}$，$S(x_j)=G_j$より積分定数を求めることで，

$$S_{LR}(x)=M_{j-1}\left\{\frac{(x_j-x)^3}{6h}-\frac{x_j-x}{6}h\right\}$$
$$+M_j\left\{\frac{(x-x_{j-1})^3}{6h}-\frac{x-x_{j-1}}{6}h\right\}$$
$$+G_{j-1}\frac{x_j-x}{h}+G_j\frac{x-x_{j-1}}{h}, \quad (2.7)$$

を得る。また，条件$S_{LR}(x_{j-1})=G_{j-1}$，$S_{LR}(x_j)=G_j$により面積相等性の条件が満たされる。

$S_{LR}(x)$の微分は，

$$S'_{LR}(x)=-M_{j-1}\frac{(x-x_j)^2}{2h}+M_j\frac{(x-x_{j-1})^2}{2h}$$
$$-\frac{h}{6}(M_j-M_{j-1})+\frac{s_j}{h}, \quad (2.8)$$

となり，これはスプライン平滑化したヒストグラムの密度推定量である。節点x_jにおける$S'_{LR}(x)$の左方微分係数及び右方微分係数はそれぞれ次のようになる。

$$\begin{cases} S'_{LR}(x_j-) = \frac{h}{6}M_{j-1} + \frac{h}{3}M_j + \frac{G_j - G_{j-1}}{h}, \\ S'_{LR}(x_j+) = -\frac{h}{3}M_j - \frac{h}{6}M_{j+1} + \frac{G_{j+1} - G_j}{h}. \end{cases} \quad (2.9)$$

点x_jでの一次連続性を満たすには，（2.9）式が等しくなる必要があるため，度数$\nu_j \in [x_{j-1}, x_j)$，（$j=1,2,...,N$）とすると，$G_j - G_{j-1} = S_j = \frac{\nu_j}{n}$より，

$$\frac{1}{6}M_{j-1} + \frac{2}{3}M_j + \frac{1}{6}M_{j+1} = \frac{\nu_{j+1} - \nu_j}{nh^2}, \quad (2.10)$$

となる。この制約条件により得られる方程式には未知のM_jが含まれており，このM_jについて解く問題となる。しかしながら，$N-1$個の方程式よりも$N+1$個の未知数の方が多く，更に2つの制約条件が必要であるため，端条件として$M_0 = M_N = 0$を設定する。これにより，$M_0, ..., M_N$について解くことが可能となり，目的の推定量を得る。

L&Rは上記の設定で，$S'_{LR}(x)$のバイアスの主要項を導出しており，

$$Bias\{S'_{LR}(x)\} = \frac{f'''(x)}{4!}h^3\{(1-r)^4 - r^4 - (1-r)^2 + r^2\}, \quad (2.11)$$

ただし，$r = (x - x_{j-1})/h$である。また，$S'_{LR}(x)$の分散の主要項について以下のように導出している。

$$Var\{S'_{LR}(x)\} = \frac{f(x)}{nh}A(r), \quad (2.12)$$

ただし，$A(r)$は$\sigma = \sqrt{3} - 2$とし，以下の通りである。

$$A(r) = 1 - \frac{3(1-\sigma)}{2+\sigma}\left(2r^2 - 2r + \frac{1}{3}\right) + \frac{9}{4}\left(\frac{1-\sigma}{2+\sigma}\right)^2\left[\left(2r^2 - 2r + \frac{1}{3}\right)^2 + \left[\left(r^2 - \frac{1}{3}\right) + \sigma\left(\frac{1}{3} - (1-r)^2\right)\right]^2 \frac{1}{1-\sigma^2} + \left[\left(r^2 - \frac{1}{3}\right) + \frac{1}{\sigma}\left(\frac{1}{3} - (1-r)^2\right)\right]^2 \frac{\sigma^2}{1-\sigma^2}\right]. \quad (2.13)$$

更に，$S'_{LR}(x)$の漸近正規性について，リアプノフの条件を満たすことから中心極限定理を証明している。しかしながら，その平均と分散の明示的な表現については示していない。

表1は上記で述べたBKS＋SchoenbergとL&Rの制約条件等の違いを示す。表中の記号について，節点x_j, ($j=0,1,...,N$)，x_jでのヒストグラムの累積分布関数の値G_j，スプライン関数による累積分布関数の推定量$S(x)$，スプライン関数の1次導関数による密度関数の推定量$S'(x)$，スプライン関数の2次導関数$S''(x)$，x_jにおける$S'(x)$の左方微分係数$S'(x_j-)$, 右方微分係数$S'(x_j+)$である。

スプライン関数の設定条件の内，面積相等性はBKS＋SchoenbergとL＆Rともに設けている。他の設定条件として，BKS＋Schoenbergは1次導関数の連続性，L＆Rは2次導関数の連続性を設けた。スプライン関数を一意に定めるための付加条件として，BKS＋Schoenbergは2次導関数の連続性，L&Rは1次導関数の連続性を設けた。これら条件の設定順による推定量の違いについては言及されていないため，本研究で示す。

3．定理

スプライン平滑化したヒストグラム密度推定量の大標本特性は，次の2つの条件

表1　BKS+SchoenbergとLii and Rosenblattの制約条件

	スプライン関数の次数	スプライン表現	節点箇所	設定条件			付加条件
				面積相等性 $S(x_j)=G_j$	1次導関数の連続性	2次導関数の連続性	
BKS, Schoernberg	3（累積分布関数）	3次スプライン	ビンの端点	○	○	−	S''の連続性 $S''(x_j-)=S''(x_j+)$, $S'(x_0)=S'(x_N)=0$
Lii and Rosenblatt	3（累積分布関数）	3次スプライン	ビンの端点	○	−	○	S'の連続性 $S'(x_j-)=S'(x_j+)$, $S''(x_0)=S''(x_N)=0$

1．ビン幅hについて，$n\to\infty$のとき，
$h\to 0$かつ$nh\to\infty$
2．関数$f(x)$は絶対連続関数で，導関数の二階微分可能

を満たすとき，以下の通りである。

BKS＋Schoenbergが提案したヒストスプラインと，L&Rが提案した3次スプライン平滑化によるヒストグラムについて次の定理が成り立つ。

定理1. BKS＋SchoenbergとL&Rの同等性

BKS+Schoenbergの設定における推定量：

$$S'_{BKS}(x)=m_{j-1}\frac{(x_j-x)(2x_{j-1}+x_j-3x)}{h^2}-m_j\frac{(x-x_{j-1})(2x_j+x_{j-1}-3x)}{h^2}+\frac{6s_j}{h^3}(x_j-x)(x-x_{j-1}),$$

および，L&Rの設定における推定量：

$$S'_{LR}(x)=-M_{j-1}\frac{(x-x_j)^2}{2h}+M_j\frac{(x-x_{j-1})^2}{2h}-\frac{h}{6}(M_j-M_{j-1})+\frac{s_j}{h},$$

が方程式として同等である。説明は4章で示す。

定理1で同等性が示されたため，BKS+SchoenbergとL&Rの推定量では同じAMISEを得る。ヒストスプライン推定量の明示的なAMISEは次の通りである。

定理2. ヒストスプラインのAMISE

ヒストスプライン推定量$\hat{f}(x)$の漸近的なMISE（AMISE）は，

$$\mathrm{AMISE}(\hat{f}(x))=\mathrm{AIV}(\hat{f}(x))+\mathrm{AISB}(\hat{f}(x))=\left(\frac{5\sqrt{3}+3}{10}\right)\frac{1}{nh}+\frac{R(f''')}{30240}h^6,$$

ただし，AIVは漸近積分分散，AISBは漸近積分二乗バイアスを表し，$R(f''')=\int f'''(x)^2dx$である。

最小AMISE*は

$$\mathrm{AMISE}^*=\frac{35\sqrt{3}+21}{60}\left(\frac{R(f''')}{2520\sqrt{3}+1512}\right)^{\frac{1}{7}}n^{-\frac{6}{7}},$$

であり，このときの最適ビン幅h^*は

$$h^*=\left(\frac{2520\sqrt{3}+1512}{R(f''')}\right)^{\frac{1}{7}}n^{-\frac{1}{7}},$$

である。

AMISE$(\hat{f}(x))$をヒストグラムのAMISE$_{hist}(\hat{f}(x))=\frac{1}{nh}+\frac{R(f')}{12}h^2$と比較すると，分散項は大きくなる一方で二乗バイアス項が小さくなる。これは，ビン幅についての条件：$n\to\infty$のとき，$h\to 0$かつ$nh\to\infty$からも明らかな通り，分散とバイアスがトレードオフの関係にあるからである。説明は4章で示す。

ヒストスプライン推定量の平均積分誤差（以下，MSEと呼ぶ）の上限について次の通りである。

系1．ヒストスプライン推定量のMSEの上限

$x\in[x_{j-1},x_j)$，$0\le|x_j-x|\le h$とすると，ヒストスプライン推定量のMSEの上限は，

$$\mathrm{MSE}\{S'_{BKS}(x)\}\le\sqrt{nh}\frac{f'''(x)}{2}h^3+\left(\frac{16\sqrt{3}}{9}+\frac{9}{4}\right)\frac{1}{nh}f(x).$$

BKS＋Schoenbergによるヒストスプラインと，L&Rによる3次スプライン平滑化によるヒストグラムについて次の補助定理が成り立つ。

補助定理．ヒストスプラインとL&Rによる推定量の同等性

ヒストスプライン推定量：

$$S'_{BKS}(x)=m_{j-1}\frac{(x_j-x)(2x_{j-1}+x_j-3x)}{h^2}-m_j\frac{(x-x_{j-1})(2x_j+x_{j-1}-3x)}{h^2}+\frac{6s_j}{h^3}(x_j-x)(x-x_{j-1}),$$

および，L&Rによる漸近正規性の証明における推定量（文献(4)，p.229）：

$$S'_{LR}(x)=\frac{1}{h}(G_j-G_{j-1})+\sum_{i=0}^{N}\frac{3a_{j,i}}{h^2}[G_{i+1}-2G_i+G_{i-1}],$$

について，$S'_{BKS}(x)=S'_{LR}(x)$が成り立つ。ただし，

$$a_{j,i}=\left\{\frac{(x-x_{j-1})^2}{2h}-\frac{h}{6}\right\}A^{-1}_{j,i}+\left\{\frac{h}{6}-\frac{(x_j-x)^2}{2h}\right\}A^{-1}_{j-1,i},\tag{3.1}$$

であり，$A^{-1}_{j,i}$, (i=0,...,N) は（2.4），（2.10）式のm_jおよびM_jの係数についての逆行列A^{-1}の(j,i)要素である。$A^{-1}_{j,i}$の導出については文献(4)を参照のこと。説明は4章で示す。

以上の補助定理より，スプライン平滑化によるヒストグラムの漸近正規性の成立については，L&R（1974）で示されたTheorem 4. に帰着する。従って，ヒストスプラインの漸近正規性について定数項を含む明示的な表現は次の通りである。

系2. 各ビンにおけるヒストスプラインの漸近正規性

$h\infty 0\,(n^{-\alpha})$, $x\in B_j$に対して，

$\alpha=\frac{1}{7}$のとき

$$\sqrt{nh}\{f_j(x)-f(x)\}\xrightarrow{d}N\left(\sqrt{nh}\frac{f'''(x)}{6}h^3\left(r^3-\frac{3}{2}r^2+\frac{1}{2}r\right),\ \frac{5\sqrt{3}+3}{10}f(\xi_j)\right)$$

$\alpha>\frac{1}{7}$のとき

$$\sqrt{nh}\{f_j(x)-f(x)\}\xrightarrow{d}N\left(0,\frac{5\sqrt{3}+3}{10}f(\xi_j)\right),$$

が漸近的に成り立つ。ただし，$r=\frac{1}{h}(x_j-x)$，$f(\xi_j)$，$\xi_j\in B_j$は$p_j=\int_{B_j}f(t)\,dt=hf(\xi_j)$ を満たす点である。説明は4章で示す。

4．定理と系の証明

4.1．定理1．BKS＋SchoenbergとL&Rの同等性の証明

BKS＋SchoenbergとL&Rの同等性について以下に示す。BKS＋Schoenbergの設定による密度推定量は，

$$S'_{BKS}(x)=m_{j-1}\frac{(x_j-x)(2x_{j-1}+x_j-3x)}{h^2}-m_j\frac{(x-x_{j-1})(2x_j+x_{j-1}-3x)}{h^2}+\frac{6s_j}{h^3}(x_j-x)(x-x_{j-1}).\quad(4.1)$$

$S'_{BKS}(x)$を微分して，

$$S''_{BKS}(x)=m_{j-1}\frac{(6x-4x_j-2x_{j-1})}{h^2}-m_j\frac{(-6x+2x_j+4x_{j-1})}{h^2}+\frac{6s_j}{h^3}(-2x+x_j+x_{j-1}).\quad(4.2)$$

M_jを節点x_jにおける$S_{BKS}(x)$の2次微係数とし，$S''_{BKS}(x)$にx_jとx_{j-1}をそれぞれ代入して，

$$\begin{cases}S''_{BKS}(x_j)=m_{j-1}\dfrac{(2x_j-2x_{j-1})}{h^2}-m_j\dfrac{(-4x_j+4x_{j-1})}{h^2}+\dfrac{6s_j}{h^3}(-x_j+x_{j-1})=M_j, & (4.3)\\ S''_{BKS}(x_{j-1})=m_{j-1}\dfrac{(-4x_j+4x_{j-1})}{h^2}-m_j\dfrac{(2x_j-2x_{j-1})}{h^2}+\dfrac{6s_j}{h^3}(x_j-x_{j-1})=M_{j-1}, & (4.4)\end{cases}$$

（4.3）−（4.4）より，

$$m_{j-1}\frac{(6x_j-6x_{j-1})}{h^2}-m_j\frac{(-6x_j+6x_{j-1})}{h^2}+\frac{6s_j}{h^3}(-2x_j+2x_{j-1})=M_j-M_{j-1}.\quad(4.5)$$

（4.3）＋（4.4）より，

$$m_{j-1}\frac{(-2x_j+2x_{j-1})}{h^2}-m_j\frac{(-2x_j+2x_{j-1})}{h^2}=M_j+M_{j-1}.\quad(4.6)$$

（4.6）式で項を入れ替えて，

$$m_{j-1}=(M_j+M_{j-1})\frac{h^2}{-2(x_j-x_{j-1})}+m_j.\quad(4.7)$$

（4.7）式を（4.5）式に代入してm_jについて解くと，

$$m_j=\frac{s_j}{h}+(2M_j+M_{j-1})\frac{h^2}{6(x_j-x_{j-1})}.\quad(4.8)$$

（4.8）式を（4.7）式に代入して，

$$m_{j-1}=\frac{s_j}{h}+(M_j+2M_{j-1})\frac{h^2}{-6(x_j-x_{j-1})}.\quad(4.9)$$

（4.8）式と（4.9）式を密度推定量（4.1）式に代入して，

$$\begin{aligned}S'_{BKS}(x)=&(M_j+2M_{j-1})\frac{h^2}{-6(x_j-x_{j-1})}\frac{(x_j-x)(2x_{j-1}+x_j-3x)}{h^2}\\&+\frac{s_j}{h}\frac{(x_j-x)(2x_{j-1}+x_j-3x)}{h^2}\\&-(2M_j+M_{j-1})\frac{h^2}{6(x_j-x_{j-1})}\frac{(x-x_{j-1})(2x_j+x_{j-1}-3x)}{h^2}\\&-\frac{s_j}{h}\frac{(x-x_{j-1})(2x_j+x_{j-1}-3x)}{h^2}+\frac{6s_j}{h^3}(x_j-x)(x-x_{j-1}),\end{aligned}\quad(4.10)$$

整理すると，

$$S'_{BKS}(x)=-M_{j-1}\frac{(x-x_j)^2}{2h}+M_j\frac{(x-x_{j-1})^2}{2h}-\frac{h}{6}(M_j-M_{j-1})+\frac{s_j}{h},$$

これは，L&Rの設定における密度推定量（2.8）式と一致するため，$S'_{BKS}(x)=S'_{LR}(x)$である。以上で定理1は証明された。

4.2. 定理2. AMISE（$\hat{f}(x)$）の証明

定理２の証明について，MISEの定義は以下の通りである。

$$\begin{aligned}\mathrm{MISE}&:=\mathrm{E}[\mathrm{ISE}]\\&=\mathrm{E}\left\{\int\left[\hat{f}(t)-f(t)\right]^2dx\right\}=\int E\left[\hat{f}(t)-f(t)\right]^2dx\\&=\mathrm{IV}[\hat{f}(t)]+\mathrm{ISB}[\hat{f}(t)],\end{aligned}$$

ただし，IVとISBは次のように定義される。

$$\mathrm{IV}[\hat{f}(t)]=\int Var\left[\hat{f}(t)\right]dt,$$

$$\mathrm{ISB}[\hat{f}(t)]=\int Bias\left[\hat{f}(t)\right]^2dt.$$

MISEは分散項IVと二乗バイアス項ISBに分解でき，MISEの値が０に近いほど推定量と真の密度との誤差が小さいことを意味する。

AMISE($\hat{f}(x)$)は漸近積分分散AIV($\hat{f}(x)$)と漸近積分二乗バイアスAISB($\hat{f}(x)$)のそれぞれについて導出する。

まず漸近積分二乗バイアスについて示す。ヒストスプライン推定量は（2.5）式より，

$$\begin{aligned}\widehat{f}_j(x)=&\frac{1}{h^3}\left\{-2h(x_j-x)+3(x_j-x)^2\right\}\sum_{k=1}^{N-1}w_{j-1,k}\left(\frac{s_k+s_{k+1}}{2}\right)\\&+\frac{1}{h^3}\left\{h^2-4(x_j-x)+3(x_j-x)^2\right\}\sum_{l=1}^{N-1}w_{j,l}\left(\frac{s_l+s_{l+1}}{2}\right)\\&-\frac{6}{h^3}\left\{-h(x_j-x)+(x_j-x)^2\right\}s_j,\end{aligned}$$

ただし，$w_{j,l}$は重み$\sum_{l=1}^{N-1}w_{j,l}=1$で，

$$w_{j,l}=\frac{3}{\sqrt{3}}\left(\sqrt{3}-2\right)^{|j-l|},$$

である。(2.5)式について期待値を取ると，$S_j=\frac{\nu_j}{n}$より，

$$\begin{aligned}E\left[\widehat{f}_j(x)\right]=&\frac{1}{2n}C_{1j}(x)\sum_{k=1}^{N-1}w_{j-1,k}(E[\nu_k]+E[\nu_{k+1}])\\&-\frac{1}{2n}C_{2j}(x)\sum_{l=1}^{N-1}w_{j,l}(E[\nu_l]+E[\nu_{l+1}])+\frac{1}{n}C_{3j}(x)E\left[\nu_j\right],\end{aligned}\tag{4.11}$$

ただし，

$$C_{1j}(x)=\frac{1}{h^3}\left\{-2h(x_j-x)+3(x_j-x)^2\right\},$$

$$C_{2j}(x)=\frac{1}{h^3}\left\{-2h(x-x_{j-1})+3(x-x_{j-1})^2\right\},$$

$$C_{3j}(x)=\frac{6}{h^3}(x_j-x)(x-x_{j-1}).$$

$\nu_j\sim B(n,p_k)$で，$p_k=\int_{B_j}f(t)dt$とすると，(4.11)式は

$$\begin{aligned}E[\widehat{f}_j(x)]=&\frac{1}{2n}C_{1j}(x)\sum_{k=1}^{N-1}w_{j-1,k}(np_k+np_{k+1})\\&-\frac{1}{2n}C_{2j}(x)\sum_{l=1}^{N-1}w_{j,l}(np_l+np_{l+1})+\frac{1}{n}C_{3j}(x)np_j\\=&C_{1j}(x)\sum_{k=1}^{N-1}w_{j-1,k}\frac{1}{2}\left(\int_{B_k}f(t)dt+\int_{B_{k+1}}f(t)dt\right)\\&-C_{2j}(x)\sum_{l=1}^{N-1}w_{j,l}\frac{1}{2}\left(\int_{B_l}f(t)dt+\int_{B_{l+1}}f(t)dt\right)+C_{3j}(x)\int_{B_j}f(t)dt.\end{aligned}\tag{4.12}$$

ここで，$f(t)$は未知のため，テイラー級数により近似すると，

$$\begin{aligned}E[\widehat{f}_j(x)]=&C_{1j}(x)\sum_{k=1}^{N-1}w_{j-1,k}\left\{hf(x)+hf'(x)(x_k-x)+\frac{f''(x)}{2}h\left(x_k^2-2x_kx+\frac{h^2}{3}+x^2\right)\right.\\&\left.+\frac{f'''(x)}{6}h(x_k^3+x_kh^2-(3x_k^2+h^2)x+3x_kx^2-x^3)\right\}\\&-C_{2j}(x)\sum_{l=1}^{N-1}w_{j,l}\left\{hf(x)+hf'(x)(x_l-x)+\frac{f''(x)}{2}h\left(x_l^2-2x_lx+\frac{h^2}{3}+x^2\right)\right.\\&\left.+\frac{f'''(x)}{6}h(x_l^3+x_lh^2-(3x_l^2+h^2)x+3x_lx^2-x^3)\right\}\\&+C_{3j}(x)\left[hf(x)+hf'(x)\left(x_j-\frac{h}{2}-x\right)\right.\\&+\frac{f''(x)}{2}h\left(x_j^2-x_jh+\frac{h^2}{3}+(-2x_j+h)x+x^2\right)\\&\left.+\frac{f'''(x)}{6}h\left\{x_j^3-\frac{3}{2}x_j^2h+x_jh^2-\frac{h^3}{4}\right.\right.\\&\left.\left.+(-3x_j^2+3x_jh-h^2)x-\frac{3}{2}(-2x_j+h)x^2-x^3\right\}\right].\end{aligned}\tag{4.13}$$

(4.13)式について，$x_0=0,\ x_j=x_0+jh=jh$とし，$\sum_{l=1}^{N-1}w_{j,l}\,l=j,\ \sum_{l=1}^{N-1}w_{j,l}\,l^2=j^2-\frac{1}{3},\ \sum_{l=1}^{N-1}w_{j,l}\,l^3=j^3-j$ であることを利用して整理すると，

$$\begin{aligned}E\left[\widehat{f}_j(x)\right]=&f(x)+\frac{f'''(x)}{6}\left\{\left(j^3h^3-\frac{3}{2}j^2h^3+\frac{1}{2}jh^3\right)\right.\\&\left.+\left(-3j^2h^2+3jh^2-\frac{1}{2}h^2\right)x+\left(3jh-\frac{3}{2}h\right)x^2-x^3\right\}.\end{aligned}\tag{4.14}$$

したがって，$\hat{f}_j(x)$ のバイアスは，

$$\begin{aligned}Bias\left(\widehat{f}_j(x)\right)=&E\left[\widehat{f}_j(x)\right]-f(x)\\=&\frac{f'''(x)}{6}\left\{\left(j^3h^3-\frac{3}{2}j^2h^3+\frac{1}{2}jh^3\right)\right.\\&\left.+\left(-3j^2h^2+3jh^2-\frac{1}{2}h^2\right)x+\left(3jh-\frac{3}{2}h\right)x^2-x^3\right\},\end{aligned}\tag{4.15}$$

となる。このことから，ビンB_jにおける漸近積分二乗バイアス（AISB）は，

$$\begin{aligned}\mathrm{AISB}_j=&\int_{x_j-h}^{x_j}\frac{f'''(x)^2}{36}\left\{\left(j^3h^3-\frac{3}{2}j^2h^3\frac{1}{2}jh^3\right)\right.\\&\left.+\left(-3j^2h^2+3jh^2-\frac{1}{2}\mathrm{h}^2\right)x+\left(3jh-\frac{3}{2}h\right)x^2-x^3\right\}^2dx\\=&\frac{1}{30240}f'''(x)^2h^7.\end{aligned}\tag{4.16}$$

以上より，全体でのAISBは，リーマン積分近似 $\sum_k f'''(\xi_k)^2 h = [\int f'''(x)^2\,dx + o(1)]$ を用いて，

$$\mathrm{AISB} = \frac{R(f''')}{30240}h^6, \tag{4.17}$$

ただし，$R(f''')=\int f'''(x)^2\,dx$ である。

続いて，分散について，

$$\begin{aligned} Var\left(\hat{f}_j(x)\right) &= Var\left(\frac{1}{2n}\sum_{k=1}^{N-1} w_{j-1,k}(v_k + v_{k+1})\right) \\ &+ Var\left(\frac{1}{2n}\sum_{l=1}^{N-1} w_{j,l}(v_l + v_{l+1})\right) + Var\left(C_{3j}(x)\frac{v_j}{n}\right) \\ &+ 2Cov\left(\frac{1}{2n}\sum_{k=1}^{N-1} w_{j-1,k}(v_k + v_{k+1}),\ \frac{1}{2n}\sum_{l=1}^{N-1} w_{j,l}(v_l + v_{l+1})\right) \\ &+ 2Cov\left(\frac{1}{2n}\sum_{k=1}^{N-1} w_{j-1,k}(v_k + v_{k+1}),\ C_{3j}(x)\frac{v_j}{n}\right) \\ &+ 2Cov\left(\frac{1}{2n}\sum_{l=1}^{N-1} w_{j,l}(v_l + v_{l+1}),\ C_{3j}(x)\frac{v_j}{n}\right). \end{aligned} \tag{4.18}$$

第1項は，

$$\begin{aligned} &Var\left(\frac{1}{2n}\sum_{k=1}^{N-1} w_{j-1,k}(v_k + v_{k+1})\right) \\ &= \frac{1}{4n^2}C_{1j}(x)^2\left\{Var\left(\sum_{k=1}^{N-1} w_{j-1,k}\,v_k\right) + Var\left(\sum_{k=1}^{N-1} w_{j-1,k}\,v_{k+1}\right)\right. \\ &\left. + 2Cov\left(\sum_{k=1}^{N-1} w_{j-1,k}\,v_k, \sum_{k=1}^{N-1} w_{j-1,k}\,v_{k+1}\right)\right\} = 2\sqrt{3}\frac{h}{n}C_{1j}(x)^2 f(x), \end{aligned}$$

ここで，$Var(\cdot)$ を積分したものを$\mathrm{AI}Var(\cdot)$とすると，

$$\mathrm{AI}Var\left(\frac{1}{2n}\sum_{k=1}^{N-1} w_{j-1,k}(v_k + v_{k+1})\right) = \frac{\sqrt{3}}{15}\frac{1}{nh}f(x)h. \tag{4.19}$$

第2項も第1項と同様に，

$$\begin{aligned} &Var\left(\frac{1}{2n}\sum_{l=1}^{N-1} w_{j,l}(v_l + v_{l+1})\right) \\ &= \frac{1}{4n^2}C_{2j}(x)^2\left\{Var\left(\sum_{l=1}^{N-1} w_{j,l}\,v_l\right) + Var\left(\sum_{l=1}^{N-1} w_{j,l}\,v_{l+1}\right)\right. \\ &\left. + 2Cov\left(\sum_{l=1}^{N-1} w_{j,l}\,v_l, \sum_{l=1}^{N-1} w_{j,l}\,v_{l+1}\right)\right\} = 2\sqrt{3}\frac{h}{n}C_{2j}(x)^2 f(x), \end{aligned}$$

積分して，

$$\mathrm{AI}Var\left(\frac{1}{2n}\sum_{l=1}^{N-1} w_{j,l}(v_l + v_{l+1})\right) = \frac{\sqrt{3}}{15}\frac{1}{nh}f(x)h. \tag{4.20}$$

第3項は，

$$Var\left(C_{3j}(x)\frac{v_j}{n}\right) = \frac{1}{n^2}C_{3j}(x)^2 Var(v_j) = \frac{h}{n}C_{3j}(x)^2 f(x),$$

積分して，

$$\mathrm{AI}Var\left(C_{3j}(x)\frac{v_j}{n}\right) = \frac{6}{5}\frac{1}{nh}f(x)h. \tag{4.21}$$

第4項は，

$$\begin{aligned} &2Cov\left(\frac{1}{2n}\sum_{k=1}^{N-1} w_{j-1,k}(v_k + v_{k+1}),\ \frac{1}{2n}\sum_{l=1}^{N-1} w_{j,l}(v_l + v_{l+1})\right) \\ &= \left(5\sqrt{3}-9\right)\frac{h}{n}C_{1j}(x)C_{2j}(x)f(x) + O\left(\frac{h^2}{n}\right), \end{aligned}$$

ここで，$Cov(\cdot)$ を積分したものを$\mathrm{AI}Cov(\cdot)$とすると，

$$\begin{aligned} &2\mathrm{AI}Cov\left(\frac{1}{2n}\sum_{k=1}^{N-1} w_{j-1,k}(v_k + v_{k+1}),\ \frac{1}{2n}\sum_{l=1}^{N-1} w_{j,l}(v_l + v_{l+1})\right) \\ &= \left(\frac{5\sqrt{3}-9}{30}\right)\frac{1}{nh}f(x)h. \end{aligned} \tag{4.22}$$

第5項は，

$$\begin{aligned} &2Cov\left(\frac{1}{2n}\sum_{k=1}^{N-1} w_{j-1,k}(v_k + v_{k+1}),\ C_{3j}(x)\frac{v_j}{n}\right) \\ &= \left(3-\sqrt{3}\right)\frac{h}{n}C_{1j}(x)C_{3j}(x)f(x) + O\left(\frac{h^2}{n}\right), \end{aligned}$$

積分して，

$$\begin{aligned} &2\mathrm{AI}Cov\left(\frac{1}{2n}\sum_{k=1}^{N-1} w_{j-1,k}(v_k + v_{k+1}),\ C_{3j}(x)\frac{v_j}{n}\right) \\ &= \left(\frac{\sqrt{3}-3}{10}\right)\frac{1}{nh}f(x)h. \end{aligned} \tag{4.23}$$

第6項も第5項と同様に，

$$\begin{aligned} &2Cov\left(\frac{1}{2n}\sum_{l=1}^{N-1} w_{j,l}(v_l + v_{l+1}),\ C_{3j}(x)\frac{v_j}{n}\right) \\ &= \left(3-\sqrt{3}\right)\frac{h}{n}C_{2j}(x)C_{3j}(x)f(x), \end{aligned}$$

積分して，

$$\begin{aligned} &2\mathrm{AI}Cov\left(\frac{1}{2n}\sum_{l=1}^{N-1} w_{j,l}(v_l + v_{l+1}),\ C_{3j}(x)\frac{v_j}{n}\right) \\ &= \left(\frac{\sqrt{3}-3}{10}\right)\frac{1}{nh}f(x)h. \end{aligned} \tag{4.24}$$

したがって，(4.19)～(4.24) 式より，ビンB_jにおける漸近積分分散は，

$$\begin{aligned} \mathrm{AIV}_j &= \left(\frac{\sqrt{3}}{15} + \frac{\sqrt{3}}{15} + \frac{6}{5} + \frac{5\sqrt{3}-9}{30} + \frac{\sqrt{3}-3}{10} + \frac{\sqrt{3}-3}{10}\right)\frac{1}{nh}f(x)h \\ &= \left(\frac{5\sqrt{3}+3}{10}\right)\frac{1}{nh}f(x)h. \end{aligned} \tag{4.25}$$

以上より，全体でのAIVはリーマン積分近似 $\sum_k f(\xi_k)h = [\int f(x)dx + o(1)]$ より，

$$\mathrm{AIV}=\left(\frac{5\sqrt{3}+3}{10}\right)\frac{1}{nh}. \tag{4.26}$$

まとめると，$\hat{f}(x)$ のAMISEは

$$\begin{aligned}\mathrm{AMISE}(\hat{f}(x))&=\mathrm{AIV}(\hat{f}(x))+\mathrm{AISB}(\hat{f}(x))\\&=\left(\frac{5\sqrt{3}+3}{10}\right)\frac{1}{nh}+\frac{R(f''')}{30240}h^6,\end{aligned} \tag{4.27}$$

となる。以上より，定理２は証明された。

4.3. 系1. ヒストスプライン推定量のMSEの上限の証明

ヒストスプライン推定量のMSEの上限の導出について示す。まず，ヒストスプライン推定量のバイアスは（4.15）式から，

$$\begin{aligned}Bias(S'_{BKS}(x))&=\sqrt{nh}\,\frac{f'''(x)}{6}\left\{\left({x_j}^3-\frac{3}{2}{x_j}^2h+\frac{1}{2}x_jh^2\right)\right.\\&\left.+\left(-3{x_j}^2+3x_jh-\frac{1}{2}h^2\right)x+\left(3x_j-\frac{3}{2}h\right)x^2-x^3\right\}\\&=\sqrt{nh}\frac{f'''(x)}{6}\left\{(x_j-x)^3-\frac{3}{2}\mathrm{h}(x_j-x)^2+\frac{1}{2}h^2(x_j-x)\right\},\end{aligned}$$

ここで，$0\le|x_j-x|\le h$より，

$$\begin{aligned}Bias(S'_{BKS}(x))&\le\sqrt{nh}\frac{f'''(x)}{6}\left(h^3+\frac{3}{2}h^3+\frac{1}{2}h^3\right)\\&=\sqrt{nh}\frac{f'''(x)}{2}h^3.\end{aligned} \tag{4.28}$$

続いて，分散について（4.19）～（4.24）式の導出において，$0\le|x_j-x|\le h$より，$C_{1j}(x)$，$C_{2j}(x)$，$C_{3j}(x)$ の絶対値に関して上限を求めると，

$$\begin{cases}|C_{1j}(x)|\le\dfrac{1}{h}\\|C_{2j}(x)|\le\dfrac{1}{3h}\\|C_{3j}(x)|\le\dfrac{3}{2h}\end{cases} \tag{4.29}$$

であることから，

$$\begin{aligned}Var\left(S'_{BKS}(x)\right)&\le\left(\sqrt{3}+\frac{1}{3\sqrt{3}}+\frac{9}{4}+\frac{5\sqrt{3}-9}{3}+\frac{3(3-\sqrt{3})}{2}\right.\\&\left.-\frac{3-\sqrt{3}}{2}\right)\frac{1}{nh}f(x)=\left(\frac{16\sqrt{3}}{9}+\frac{9}{4}\right)\frac{1}{nh}f(x).\end{aligned} \tag{4.30}$$

以上より系１が示された。

4.4. 補助定理. ヒストスプラインとL&Rによる推定量の同等性の証明

ヒストスプラインの漸近正規性の証明に関する補助定理として，ヒストスプライン推定量とL&Rによる推定量が同等であることを示す。

（2.5）式からヒストスプライン推定量：

$$\begin{aligned}S'_{BKS}(x)&=m_{j-1}\frac{(x_j-x)(2x_{j-1}+x_j-3x)}{h^2}\\&-m_j\frac{(x-x_{j-1})(2x_j+x_{j-1}-3x)}{h^2}\\&+\frac{6s_j}{h^3}(x_j-x)(x-x_{j-1}).\end{aligned}$$

（2.4）式を書き換えると，

$$m_j=\sum_{i=0}^{N}A_{j,i}^{-1}d_i, \tag{4.31}$$

ただし，d_i は（2.4）式の右辺を変形したものに対応し，

$$\begin{cases}d_i=\dfrac{3}{2h}(G_{i+1}-G_{i-1}),\ (G_1-G_0),\\d_0=\dfrac{3}{2h}\ (j=1,\dots,N-1),\\d_N=\dfrac{3}{2h}(G_N-G_{N-1}),\end{cases} \tag{4.32}$$

である。そのため，（4.31）式は以下のように表される。

$$m_j=\frac{3}{2h}\sum_{i=0}^{N}A_{j,i}^{-1}(G_{i+1}-G_{i-1}). \tag{4.33}$$

ここで，表記の簡便化のため，

$$\begin{aligned}b_{j,i}&=\left\{\frac{3(x-x_{j-1})^2}{h^2}-\frac{2(x-x_{j-1})}{h}\right\}A_{j,i}^{-1}\\&+\left\{\frac{3(x_j-x)^2}{h^2}-\frac{2(x_j-x)}{h}\right\}A_{j-1,i}^{-1},\end{aligned} \tag{4.34}$$

とする。（4.33），（4.34）式を用いて（2.5）式を書き換えると，

$$\begin{aligned}S'_{BKS}(x)&=\frac{6}{h^3}(G_j-G_{j-1})(x_j-x)(x-x_{j-1})\\&+\frac{3}{2h}\sum_{i=0}^{N}b_{j,i}[(G_{i+1}-G_i)+(G_i-G_{i-1})]\\&=\frac{1}{h}(G_j-G_{j-1})\\&+\frac{1}{h}(G_j-G_{j-1})\left[-\frac{6(x-x_{j-1})^2}{h^2}+\frac{6(x-x_{j-1})-h}{h}\right]\\&+\frac{3}{h^2}\sum_{i=0}^{N}\left[\left\{\frac{3(x-x_{j-1})^2}{2h}-(x-x_{j-1})\right\}A_{j,i}^{-1}\right.\\&\left.+\left\{\frac{3(x_j-x)^2}{2h}-(x_j-x)\right\}A_{j-1,i}^{-1}\right][G_{i+1}-2G_i+G_{i-1}]\end{aligned}$$

$$+\left\{\frac{9(x-x_{j-1})^2}{h^3}-\frac{6(x-x_{j-1})}{h^2}\right\}\sum_{i=0}^{N}A_{j,i}^{-1}(G_i-G_{i-1})$$
$$+\left\{\frac{9(x_j-x)^2}{h^3}-\frac{6(x_j-x)}{h^2}\right\}\sum_{i=0}^{N}A_{j-1,i}^{-1}(G_i-G_{i-1}).$$

(4.35)

(3.1)式の$a_{j,i}$を用いて(4.35)式を整理すると，

$$S'_{BKS}(x)=\frac{1}{h}(G_j-G_{j-1})+\frac{3}{h^2}\sum_{i=0}^{N}a_{j,i}[G_{i+1}-2G_i+G_{i-1}]$$
$$+\frac{1}{h}(G_j-G_{j-1})\left[-\frac{6(x-x_{j-1})^2}{h^2}+\frac{6(x-x_{j-1})-h}{h}\right]$$
$$+\frac{3}{h^2}\left\{\frac{(x-x_{j-1})^2}{h}-(x-x_{j-1})+\frac{h}{6}\right\}\sum_{i=0}^{N}A_{j,i}^{-1}[G_{i+1}-2G_i+G_{i-1}]$$
$$+\frac{3}{h^2}\left\{\frac{2(x_j-x)^2}{h}-(x_j-x)-\frac{h}{6}\right\}\sum_{i=0}^{N}A_{j-1,i}^{-1}[G_{i+1}-2G_i+G_{i-1}]$$
$$+\left\{\frac{9(x-x_{j-1})^2}{h^3}-\frac{6(x-x_{j-1})}{h^2}\right\}\sum_{i=0}^{N}A_{j,i}^{-1}(G_i-G_{i-1})$$
$$+\left\{\frac{9(x_j-x)^2}{h^3}-\frac{6(x_j-x)}{h^2}\right\}\sum_{i=0}^{N}A_{j-1,i}^{-1}(G_i-G_{i-1}).$$

(4.36)

(4.36) 式の第3～7項について，

$$\frac{1}{h}(G_j-G_{j-1})\left[-\frac{6(x-x_{j-1})^2}{h^2}+\frac{6(x-x_{j-1})-h}{h}\right]$$
$$+\frac{3}{h^2}\left\{\frac{(x-x_{j-1})^2}{h}-(x-x_{j-1})+\frac{h}{6}\right\}\sum_{i=0}^{N}A_{j,i}^{-1}[G_{i+1}-2G_i+G_{i-1}]$$
$$+\frac{3}{h^2}\left\{\frac{2(x_j-x)^2}{h}-(x_j-x)-\frac{h}{6}\right\}\sum_{i=0}^{N}A_{j-1,i}^{-1}[G_{i+1}-2G_i+G_{i-1}]$$
$$+\left\{\frac{9(x-x_{j-1})^2}{h^3}-\frac{6(x-x_{j-1})}{h^2}\right\}\sum_{i=0}^{N}A_{j,i}^{-1}(G_i-G_{i-1})$$
$$+\left\{\frac{9(x_j-x)^2}{h^3}-\frac{6(x_j-x)}{h^2}\right\}\sum_{i=0}^{N}A_{j-1,i}^{-1}(G_i-G_{i-1})$$
$$=\frac{1}{h}(G_j-G_{j-1})\left[-\frac{6(x-x_{j-1})^2}{h^2}+\frac{6(x-x_{j-1})-h}{h}\right]$$
$$+\left\{\frac{3(x-x_{j-1})^2}{h^3}-\frac{3(x-x_{j-1})}{h^2}+\frac{1}{2h}\right\}\sum_{i=0}^{N}A_{j,i}^{-1}(G_{i+1}-G_i)$$
$$+\left\{\frac{6(x-x_{j-1})^2}{h^3}-\frac{3(x-x_{j-1})}{h^2}-\frac{1}{2h}\right\}\sum_{i=0}^{N}A_{j,i}^{-1}(G_i-G_{i-1})$$
$$+\left\{\frac{6(x_j-x)^2}{h^3}-\frac{3(x_j-x)}{h^2}-\frac{1}{2h}\right\}\sum_{i=0}^{N}A_{j-1,i}^{-1}(G_{i+1}-G_i)$$
$$+\left\{\frac{3(x_j-x)^2}{h^3}-\frac{3(x_j-x)}{h^2}+\frac{1}{2h}\right\}\sum_{i=0}^{N}A_{j-1,i}^{-1}(G_i-G_{i-1}),$$

(4.37)

ここで，文献(4)の(12)式，p.228の結果を用いて，$y_i=G_i$であるため，

$$G_i-G_{i-1}=f(x)h+O(h^2),$$

となり，また，$A_{j,i}^{-1}$について文献(4)の（21）～（24）式から，$\sum_{i=0}^{N}A_{j,i}^{-1}=1/3$であるため，これらを用いて（4.37）式を整理すると，

$$\left[-\frac{6(x-x_{j-1})^2}{h^3}+\frac{6(x-x_{j-1})-h}{h^2}\right]f(x)h$$
$$+\left\{\frac{3(x-x_{j-1})^2}{h^3}-\frac{3(x-x_{j-1})}{h^2}+\frac{1}{2h}\right\}\sum_{i=0}^{N}A_{j,i}^{-1}f(x)h$$
$$+\left\{\frac{6(x-x_{j-1})^2}{h^3}-\frac{3(x-x_{j-1})}{h^2}-\frac{1}{2h}\right\}\sum_{i=0}^{N}A_{j,i}^{-1}f(x)h$$
$$+\left\{\frac{6(x_j-x)^2}{h^3}-\frac{3(x_j-x)}{h^2}-\frac{1}{2h}\right\}\sum_{i=0}^{N}A_{j-1,i}^{-1}f(x)h$$
$$+\left\{\frac{3(x_j-x)^2}{h^3}-\frac{3(x_j-x)}{h^2}+\frac{1}{2h}\right\}\sum_{i=0}^{N}A_{j-1,i}^{-1}f(x)h.$$
$$=\left\{\frac{3(x-x_{j-1})^2}{h^3}-\frac{2(x-x_{j-1})}{h^2}+\frac{3(x-x_{j-1})^2}{h^3}-\frac{6(x-x_{j-1})}{h^2}+\frac{3}{h}\right.$$
$$\left.+\frac{2(x-x_{j-1})}{h^2}-\frac{2}{h}-\frac{6(x-x_{j-1})^2}{h^3}+\frac{6(x-x_{j-1})}{h^2}-\frac{1}{h}\right\}f(x)h=0.$$

(4.38)

以上より，（4.36）式は

$$S'_{BKS}(x)=\frac{1}{h}(G_j-G_{j-1})+\frac{3}{h^2}\sum_{i=0}^{N}a_{j,i}[G_{i+1}-2G_i+G_{i-1}],$$

となり，これはL&Rによる推定量（文献(4), p.229）と同等である。以上より，補助定理が示された。

4.5. 系2. ヒストスプラインの漸近正規性の証明

上記の補助定理から，ヒストスプラインの漸近正規性の成立は，L&R（1974）のTheorem 4. において示される。これを踏まえて，平均と分散の明示的な表現を示す。

4.2節のAMISE($\hat{f}(x)$)の導出から，各ビンにおけるスプライン推定量について$Bias\{\hat{f}_j(x)\}=E[\hat{f}_j(x)]-f_j(x)$であり，$h\propto O(n^{-\alpha})$, $x\in B_j$に対して，$\alpha=\frac{1}{7}$のとき，$\sqrt{nh}\{f_j(x)-f(x)\}$の平均は$\sqrt{nh}Bias\{f_j(x)\}$となることが示されるため（4.15）式より，

$$\sqrt{nh}\,\frac{f'''(x)}{6}\left\{\left(x_j{}^3-\frac{3}{2}x_j{}^2h+\frac{1}{2}x_jh^2\right)\right.$$
$$\left.+\left(-3x_j{}^2+3x_jh-\frac{1}{2}h^2\right)x+\left(3x_j-\frac{3}{2}h\right)x^2-x^3\right\},$$

(4.39)

ここで，$r=\frac{1}{h}(x_j-x)$とおくと，

$$\sqrt{nh}\,\frac{f'''(x)}{6}\left\{\left({x_j}^3-\frac{3}{2}{x_j}^2h+\frac{1}{2}x_jh^2\right)\right.$$
$$\left.+\left(-3{x_j}^2+3x_jh-\frac{1}{2}h^2\right)x+\left(3x_j-\frac{3}{2}h\right)x^2-x^3\right\}$$
$$=\sqrt{nh}\frac{f'''(x)}{6}\left\{\left(x_j-x\right)^3-\frac{3}{2}\mathrm{h}\left(x_j-x\right)^2+\frac{1}{2}h^2\left(x_j-x\right)\right\}$$
$$=\sqrt{nh}\frac{f'''(x)}{6}h^3\left(r^3-\frac{3}{2}r^2+\frac{1}{2}r\right). \quad (4.40)$$

また，$\sqrt{nh}\{\hat{f}_j(x)-f(x)\}$の分散についてはAMISE$(\hat{f}(x))$の分散項（4.27）式より，

$$\frac{5\sqrt{3}+3}{10}f(\xi_j), \quad (4.41)$$

である。

$a>\frac{1}{7}$のとき，$Bias\{\hat{f}_j(x)\}$よりもビン幅hの収束スピードの方が速いことから，平均は0となる。以上より，ヒストスプラインの漸近正規性について平均と分散の明示的な表現が証明された。

5．数値実験

ヒストグラムとヒストスプラインの有限標本における密度推定の精度を比較するため，積分二乗誤差（以降，ISEと呼ぶ）について数値実験を行う。ここでは，MISEの変動をISEの標本平均と標準偏差で評価した。定義域[-3,3]の標準正規分布$N(0,1)$に従う標本について，標本サイズn=100,200,500,1000,5000と設定する。ビン幅はLeave-one-out CV[2)]により推定する。ヒストグラムとヒストスプラインそれぞれについてISEの計算シミュレーションを10000回行い，ISEの標本平均と標準偏差を算出する。

図1は，n=200のヒストグラムとヒストスプラインの数値実験結果を示す。実線が真の密度関数，破線がヒストスプラインである。

表2は，ISEの標本平均の数値実験結果を示す。推定精度が良いほどISEは0に近いため，比較して値が小さい方に下線を引いてある。ヒストグラムとヒストスプラインのどちらも，標本サイズが大きくなるにつれてISEは小さくなる。標本サイズに関わらず，ヒストスプラインの方がISEは小さい。しかしながら，標本サイズが大きくなるにつれて両者のISE差は小さくなる。

表3は，ISE標準偏差の数値実験結果を示す。表中で，ヒストグラムとヒストスプラインで比較して値が小さい方に下線を引いてある。ヒストグラムとヒストスプラインのどちらも，標本サイズが大きくなるにつれてISE標準偏差は小さくな

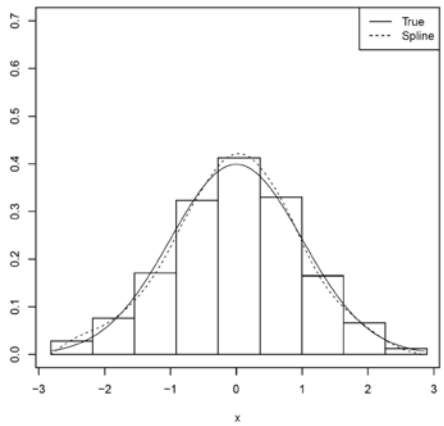

図1　数値実験結果（n=200）

表2　ISE標本平均の数値実験結果

	n=100	n=200	n=500	n=1000	n=5000
ヒストグラム	0.02783	0.01785	0.00924	0.00561	0.00180
ヒストスプライン	0.02387	0.01516	0.00782	0.00457	0.00143

表3　ISE標準偏差の数値実験結果

	n=100	n=200	n=500	n=1000	n=5000
ヒストグラム	0.01693	0.01051	0.00472	0.00241	0.00051
ヒストスプライン	0.02388	0.01482	0.00684	0.00345	0.00081

る。標本サイズに関わらず，ヒストグラムの方がISE標準偏差は小さい。しかしながら，標本サイズが大きくなるにつれて両者のISE標準偏差の差は小さくなる。

6．結論と考察

本研究では，ヒストグラムのスプライン平滑化に関するBKS＋SchoenbergとL&Rの２つの問題についての同等性を示し，また，ヒストスプライン推定量の漸近表現を陽な形で導出した。また，有限標本におけるヒストグラムとヒストスプラインの推定精度を比較する目的で，数値実験を行った。大標本特性として，一般的な正則条件の下で①BKS＋SchoenbergとL&Rの同等性，②ヒストスプラインの明示的なAMISE，⑤ヒストスプラインのMSEの上限，④補助定理及び明示的な漸近正規性を証明した。①BKS＋SchoenbergとL&Rの同等性について，$S'_{BKS}(x)$においてL&Rによる推定量と表現を揃えたときに$S'_{BKS}(x)$=$S'_{LR}(x)$であることを示した。②ヒストスプラインのAMISEについて，分散項が$\frac{5\sqrt{3}+3}{10}\frac{1}{nh}$，二乗バイアス項が$\frac{R(f''')}{30240}h^6$であることを示した。このAMISEはヒストグラムの$\mathrm{AMISE}_{hist}\left(f(x)\right)=\frac{1}{nh}+\frac{R(f')}{12}h^2$と比較すると，分散は増加する一方，二乗バイアス項は減少している。③ヒストスプラインのMSEの上限について，$0\leq|x_j-x|\leq h$とすると，分散項が$\left(\frac{16\sqrt{3}}{9}+\frac{9}{4}\right)\frac{1}{nh}f(x)$，二乗バイアス項が$\sqrt{nh}\frac{f'''(x)}{2}h^3$であることを示した。④明示的な漸近正規性について，各ビンにおけるヒストスプラインの正規性に関しては，$h\propto O(n^{-\alpha})$，$x\in B_j$に対して，$r=\frac{1}{h}(x_j-x)$とおくと，$\alpha=\frac{1}{7}$のときは$\sqrt{nh}\{f_j(x)-f(x)\}\xrightarrow{d}N\left(\sqrt{nh}\frac{f'''(x)}{6}h^3\left(r^3-\frac{3}{2}r^2+\frac{1}{2}r\right),\frac{5\sqrt{3}+3}{10}f(\xi_j)\right)$，$\alpha>\frac{1}{7}$のときは$\sqrt{nh}\{f_j(x)-f(x)\}\xrightarrow{d}N\left(0,\frac{5\sqrt{3}+3}{10}f(\xi_j)\right)$であることを示した。

有限標本時の特性について，ISEの標本平均と標準偏差についてのシミュレーション結果から，ヒストグラムとヒストスプラインのどちらの場合も，ISEの標本平均と標準偏差は標本サイズが大きくなるにつれてその値が小さくなる。ヒストスプラインの方が，どの標本サイズの時にもISE値は小さく，ヒストグラムよりも推定精度が改良される。両者のISE差は標本サイズが大きくなるほど縮まっていく。また，ISE標準偏差は，標本サイズに関わらずヒストグラムの方が値は小さいが，標本サイズが大きくなるほど両者の値は近づいていく。このことから，ヒストスプラインはヒストグラムよりも分散は増加するが，バイアスは減少する。バイアス減少の効果が推定精度に及ぼす影響の方が大きいため，全体のISEはヒストグラムよりも改良されることが分かった。

ここまで，ヒストグラムを３次のスプライン曲線で平滑化する問題について議論した。スプライン関数の次元を４次，５次，…と上げた際の一般化表現とその漸近的性質については明らかにされていないため，その導出が今後の課題である。

【注】

1) ヒストグラムの各ビンにおける面積と，スプライン平滑化後の推定量での各ビンにおける面積が等しくなるとき，面積相等性をもつという。

2) Leave-one-out CVとは，ビン幅推定法の一つである。具体的には，標本から１つデータ点を抜き出し，残りのデータ点でヒストグラムを構築し，抜き出したデータ点でそのヒストグラムを評価する。以上をデータ点ごとに繰り返し，それら評価について平均を算出する。この標本平均を最小化するようなビン幅を求め，それを推定ビン幅とする手法である。ヒストグラムの場合には，最終的な計算が陽に示され，標本サイズn，ビンB_kにおける度数をν_k，ビン幅hとすると，unbiased CV（UCV）は，

$$\mathrm{UCV}(h)=\frac{2}{(n-1)h}-\frac{n+1}{n^2(n-1)h}\sum_k \nu_k^2.$$

[引用・参考文献]

(1) D.W. Scott, “Frequency Polygons: Theory and Application”, *Journal of the American Statistical Association*, 80.390, 1985, pp.348－354.

⑵I. J. Schoenberg, "Splines and Histograms", *Spline Functions and Approximation Theory*, Birkhauser, Basel, 1973, pp.277－327.
⑶I. J. Schoenberg, "Contribution to The problem of Approximation of Equidistant Data by Analytic Functions", *Quartely of Applied Mathematics*, 4(2), 1946, pp.112－141.
⑷Keh-Shin Lii, and M. Rosenblatt, "Asymptotic Behavior of A Spline Estimate of A Density Function", *Computers & Mathematics with Applications*, 1 (2), 1975, pp.223－235.
⑸Liliana I. Boneva, David Kendall, and Ivan Stefanov, "Spline Transformations: Three New Diagnostic Aids for the Statistical Data-Analyst.", *Journal of the Royal Statistical Society. Series B (Methodological)*, 33.1, 1971, pp.1－71.
⑹M.C.Jones, M.Samiuddin, A. H.Al-Harbey, and T. A. H.Maatouk, "The Edge Frequency Polygon", *Biometrika*, 85 (1), 1998, pp.235－239.
⑺M.C.Minnotte, "The Bias-Optimized Frequency Polygon", *Computational Statistics*, 11, 1996, pp.35－48.