

Kanazawa University,  
Faculty of Economics and Management

# Discussion Paper Series

No. 068

多項式型Histogramの平滑化と  
その漸近的性質

齊藤実祥  
寒河江雅彦

[saito\\_misaki@stu.kanazawa-u.ac.jp](mailto:saito_misaki@stu.kanazawa-u.ac.jp)  
[sagae.masahiko@gmail.com](mailto:sagae.masahiko@gmail.com)

15 September 2022



金沢大学経済学経営学系  
〒920-1192 金沢市角間町

Faculty of Economics and Management,  
Kanazawa University

Kakumamachi, Kanazawa-shi, Ishikawa, 920-1192, Japan

[https://keikei.w3.kanazawa-u.ac.jp/research\\_dp.html](https://keikei.w3.kanazawa-u.ac.jp/research_dp.html)

# 多項式型 Histogram の平滑化とその漸近的性質

齊藤実祥 (金沢大学 人間社会環境研究科)

寒河江雅彦 (金沢大学 人間社会研究域)

## The Smoothing of Polynomial Histogram and Its Asymptotic Properties

Misaki Saito (Kanazawa University)

Masahiko Sagae (Kanazawa University)

Histogram の拡張の一つに Polynomial Histogram 推定法 (PH) がある。PH は各ビン内の局所モーメント情報を用いた多項式型の密度推定法であり、ビンごとに推定するため各ビンの境界で連続性を持たない。本稿では、局所モーメント条件と各ビンの境界での連続性の条件を同時に満たす多項式型の Smoothed-PH 推定法 (S-PH) を提案し、その漸近的性質と有限標本における特性を明らかにする。大標本と小標本とに関わらず、S-PH は従来の主なビン型推定量よりも推定精度を改良することを示した。

Polynomial Histogram (PH) is one of estimations expanding Histogram. PH is used the local moment information of each divided interval. However, it's discontinuous at the boundaries of each bins. In this paper, we propose a polynomial density estimation that satisfies both of continuity conditions and local moment conditions. We show its large sample properties and investigate its characteristics when our model analyzes finite samples. Our proposed model performs better an accuracy of estimate than some major bin type estimation methods.

### 1 研究背景と目的

Polynomial Histogram(PH) 推定法は各ビンにおける平均や分散からなる局所モーメント情報を用いた多項式型の密度推定法である。Sagae and Scott(1997a, b) は高次の局所モーメントを用いた PH を提案し、平均積分二乗誤差 (MISE) の意味で高次オーダーの収束性を持つことを示している。しかしながら、PH はビンごとに推定するため、各ビンの境界で連続性を持たない。これまでに Histogram の不連続性を改良する方法についてはいくつか提案されている。Scott(1985) は各ビンの中点で線形補間する Frequency Polygon(FP) を提案し、その MISE 収束レートが  $O(n^{-4/5})$  で、Histogram の  $O(n^{-2/3})$  を改良することを示した。FP の関連研究では、Minnotte(1996, 1998) の Bias-Optimized Frequency Polygon や、Jones et al.(1998) の Edge Frequency Polygon がある。その他に、滑らかな曲線で隣接ビン間を接続する方法として、Boneva, Kendall and Stefanov(BKS)(1971) と Schoenberg(1973) によって Histospline が提案され、Lii and Rosenblatt(LR)(1974) はその MISE 収束レートが  $O(n^{-6/7})$  であることを示した。齊藤・寒河江 (2021) は BKS 型と LR 型モデルの理論的同等性を示し、MISE の陽な漸近表現を導

いた。局所モーメント情報の利用と Histogram の平滑化を同時に行う推定量のモデルについては未整備の状態であるため、本稿ではここに焦点を当てる。

したがって、前述した先行研究の拡張として、局所モーメント条件と各ビンの境界での連続性の条件を同時に満たす多項式型の平滑化 Polynomial Histogram 推定法 (S-PH) を提案する。その理論的特性として漸近的性質を導出し、有限標本における有効性を数値実験から示す。

## 2 平滑化 Polynomial Histogram

### 2.1 S-PH 推定量の構築

$p$  を累積分布関数の連続性の次数、 $q$  を局所モーメントの次数とする。このとき、提案モデルの S-PH を  $S\text{-PH}_{(p,q)}$  で表記する。これは、従来のビン型推定量を包括した表現であり、Histogram は  $S\text{-PH}_{(0,0)}$ 、BFP は  $S\text{-PH}_{(1,0)}$ 、1 次 PH は  $S\text{-PH}_{(0,1)}$ 、Histospline は  $S\text{-PH}_{(2,0)}$  に対応する。以降、本稿では累積分布関数の 2 次までの連続性条件と、局所 1 次までのモーメント条件を同時に満たす  $S\text{-PH}_{(2,1)}$  の場合を扱う。

標本数  $n$ 、節点数  $N$ 、ビン節点  $x_j$  ( $j = 0, 1, \dots, N$ )、定義域  $[x_0, x_N]$  とする。 $j$  番目のビン  $B_j = [x_{j-1}, x_j]$  ( $j = 1, 2, \dots, N$ )、ビン幅  $h$ 、ビンの中点  $t_j$ 、各ビンの面積  $\mu_j^{(0)}$ 、局所 1 次モーメント  $\mu_j^{(1)}$ 、 $B_j$  における  $S\text{-PH}_{(2,1)}$  推定量  $\hat{F}_j(x)$ 、 $x_j$  における分布関数の高さ  $G_j = \hat{F}_j(x_j)$ 、 $\hat{F}_j(x)$  の係数  $a_0^{(j)}, a_1^{(j)}, \dots, a_4^{(j)}$  とし、簡単のために  $a_0, a_1, \dots, a_4$  と略す。

この時、 $S\text{-PH}_{(2,1)}$  推定量  $\hat{F}_j(x)$  は次の通りである。

$$\hat{F}_j(x) = a_0 + a_1(x - t_j) + a_2(x - t_j)^2 + a_3(x - t_j)^3 + a_4(x - t_j)^4, \quad x \in B_j. \quad (1)$$

ここで、局所モーメント条件と各ビンの境界での連続性の条件を満たすために、以下の制約条件を設ける。

- (i) 面積相等性 :  $\int_{B_j} (x - t_j)^0 \hat{F}'_j(x) dx = \mu_j^{(0)}$
- (ii) 局所 1 次モーメント情報保持 :  $\int_{B_j} (x - t_j)^1 \hat{F}'_j(x) dx = \mu_j^{(1)}$
- (iii) 累積分布関数の 1 次連続性 :  $\hat{F}'(x_j-) = \hat{F}'(x_j+)$
- (iv) 累積分布関数の 2 次連続性 :  $\hat{F}''(x_j-) = \hat{F}''(x_j+)$

ただし、 $\hat{F}^{(i)}(x_j-)$  ( $i = 0, 1, \dots$ ) は節点  $x_j$  における  $\hat{F}^{(i)}(x)$  の左方  $i$  次微分係数、 $\hat{F}^{(i)}(x_j+)$  は右方  $i$  次微分係数である。上記の条件のみでは未知数に対して方程式の数が 2 つ不足する。そのため、節点  $x_j$  における  $\hat{F}_j(x)$  の 1 次微係数  $M_j$  ( $j = 0, 1, \dots, N$ ) に対し、付加条件として端条件  $M_0 = M_N = 0$  を仮定する。



## 2.2 漸近的性質

本節ではビンごとに推定された  $\hat{f}_j(x)$  で構成された S-PH<sub>(2,1)</sub> 密度推定量の漸近的 MISE を導き、漸近正規性が成り立つことを示す。

S-PH<sub>(2,1)</sub> 密度推定量に関して次の正則条件を満たすものとする。

- (i) ビン幅  $h$  について、 $n \rightarrow \infty$  のとき、 $h \rightarrow 0$  かつ  $nh \rightarrow \infty$ ,
- (ii) 関数  $f^{(4)}(x)$  は絶対連続関数で、 $R(f^{(5)}) = \int f^{(5)}(x)^2 dx < \infty$ .

この条件のもとで、以下の定理及び系が成り立つ。

*Theorem 1* : S-PH<sub>(2,1)</sub> の AMISE

全体での S-PH<sub>(2,1)</sub> 密度推定量  $\hat{f}(x)$  の漸近的な MISE(AMISE) は、

$$\begin{aligned} \text{AMISE} \left\{ \hat{f}(x) \right\} &= \left( \frac{992 + 695\sqrt{2}}{840} \right) \frac{1}{nh} + \frac{h^8}{1209600} R(f^{(4)}) \\ &\doteq \frac{2.351}{nh} + \frac{h^8}{1209600} R(f^{(4)}). \end{aligned} \quad (10)$$

このときの最適ビン幅  $h^*$  及び最小 AMISE は、

$$h^* = \left( \frac{125100\sqrt{2} + 178560}{R(f^{(4)})} \right)^{\frac{1}{9}} n^{-\frac{1}{9}}, \quad (11)$$

$$\text{最小 AMISE} = \left( \frac{2976 + 2085\sqrt{2}}{2240} \right) \left( \frac{R(f^{(4)})}{178560 + 125100\sqrt{2}} \right)^{\frac{1}{9}} n^{-\frac{8}{9}}. \quad (12)$$

*Theorem 2* : S-PH<sub>(2,1)</sub> の漸近正規性

S-PH<sub>(2,1)</sub> 密度推定量について、

$$\frac{\sum_{s=1}^n E |Z_s - E[Z_s]|^3}{\sigma \left[ \hat{f}_j(x) \right]^3} = \frac{O\left(\frac{1}{n^2 h^2}\right)}{O\left(\frac{1}{n^{3/2} h^{3/2}}\right)} = O\left(\frac{1}{n^{1/2} h^{1/2}}\right) = o(1), \quad (13)$$

ただし、 $\sigma \left[ \hat{f}_j(x) \right]^2 = \text{Var} \left\{ \hat{f}_j(x) \right\}$  で、 $Z_s$  ( $s = 1, 2, \dots, n$ ) は  $\hat{f}_j(x)$  に従う独立な確率変数である。したがって、リアプノフの条件を満たすことから、S-PH<sub>(2,1)</sub> 密度推定量は漸近正規性が成り立つ。

*Corollary 1* : 各ビンにおける S-PH<sub>(2,1)</sub> の漸近正規性

$h \propto O(n^{-\alpha})$ ,  $x \in B_j$  に対して、

$\alpha = \frac{1}{9}$  のとき、

$$\sqrt{nh} \left\{ \hat{f}_j(x) - f(x) \right\} \xrightarrow{d} N \left( \text{Bias} \left[ \hat{f}_j(x) \right], \left( \frac{992 + 695\sqrt{2}}{840} \right) f(\xi_j) \right), \quad (14)$$

$\alpha > \frac{1}{9}$  のとき、

$$\sqrt{nh} \left\{ \hat{f}_j(x) - f(x) \right\} \xrightarrow{d} N \left( o(1), \left( \frac{992 + 695\sqrt{2}}{840} \right) f(\xi_j) \right), \quad (15)$$

が漸近的に成り立つ。ただし、 $f(\xi_j)$  は  $p_j = \int_{B_j} f(t)dt = hf(\xi_j)$ ,  $\xi_j \in B_j$  を満たす  $B_j$  内のある点とする。Theorem 1 の S-PH<sub>(2,1)</sub> の AMISE 及び Theorem 2 の漸近正規性の証明については Appendix 1, 2 で後述する。

### 3 数値実験

S-PH<sub>(2,1)</sub> の有限標本における推定精度を調べるため、積分二乗誤差 (ISE) について数値実験を行う。定義域  $[-8, 8]$  の混合正規分布  $f(x)$  は  $\frac{3}{5}N(-3, 2^2) + \frac{2}{5}N(4, 1^2)$  に従う確率密度関数とし、標本数を  $n = 10^2, 10^3, 10^4, 10^5$  とする。ビン幅は理論上の最適ビン幅を用いて、各標本数における ISE の数値実験 1000 回の平均 (MISE) を算出する。同様の設定の下での Histogram、1 次 PH、Histospline と比較する。

数値実験に用いる各推定量の最適ビン幅  $h^*$  は次の通りである。

表 1 各推定量の最適ビン幅

(a)Histogram	$h_{HIST}^* = \left( \frac{6}{R(f^{(1)})} \right)^{\frac{1}{3}} n^{-\frac{1}{3}}$
(b)1 次 PH	$h_{PH(1)}^* = \left( \frac{360}{R(f^{(2)})} \right)^{\frac{1}{5}} n^{-\frac{1}{5}}$
(c)Histospline	$h_{HSP}^* = \left( \frac{2520\sqrt{3}+1512}{R(f^{(3)})} \right)^{\frac{1}{7}} n^{-\frac{1}{7}}$
(d)S-PH <sub>(2,1)</sub>	$h_{S-PH(2,1)}^* = \left( \frac{125100\sqrt{2}+178560}{R(f^{(4)})} \right)^{\frac{1}{9}} n^{-\frac{1}{9}}$

図 1 は  $n = 10^3$  のときの Histogram、1 次 PH、Histospline と本稿で提案した S-PH<sub>(2,1)</sub> の数値実験結果の例で、実線が推定結果、破線が真の密度関数である。

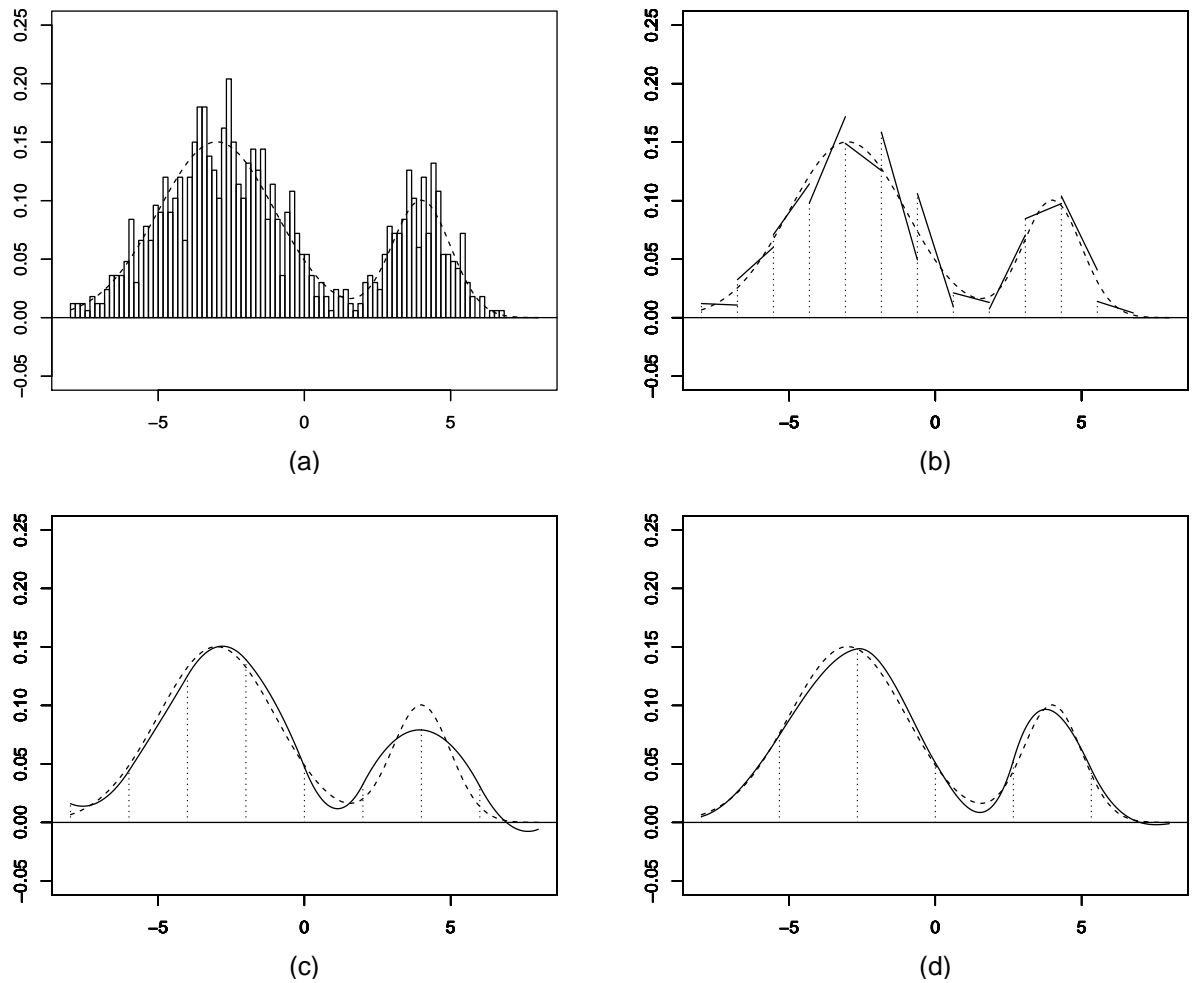


図1  $n = 10^3$  での推定例

(実線が推定結果、破線が真の分布。(a)Histogram、(b)1次 PH、(c)Histospline、(d)S-PH<sub>(2,1)</sub>)

表2 主な3つの推定量と S-PH<sub>(2,1)</sub> の MISE 比較 (繰り返し数 1000 回)

	MISE*	$n = 10^2$	$n = 10^3$	$n = 10^4$	$n = 10^5$
Histogram	$O\left(n^{-\frac{2}{3}}\right)$	0.026454	0.005948	0.001301	0.000280
PH <sub>(1)</sub>	$O\left(n^{-\frac{4}{5}}\right)$	0.009111	0.001575	0.000265	0.000043
Histospline	$O\left(n^{-\frac{6}{7}}\right)$	<u>0.003927</u>	0.001752	0.000187	0.000020
S-PH <sub>(2,1)</sub>	$O\left(n^{-\frac{10}{9}}\right)$	0.005953	<u>0.000849</u>	<u>0.000112</u>	<u>0.000016</u>

表2はデータ数  $n$  を変化した時の4つの推定法を比較するための MISE の実験結果である。推定精度が良いほど MISE は 0 に近いため、各  $n$  ごとに比較して最小の値に下線を引いてある。ど

の推定量も標本数が大きくなるにつれて ISE は小さくなり、ISE の各推定量の差は小さくなる。標本数が  $n = 10^3$  以上の時に、S-PH<sub>(2,1)</sub> の ISE が最も小さくなる。

表 3 ビン幅比

	$h^*$	$n = 10^2$	$n = 10^3$	$n = 10^4$	$n = 10^5$
Histogram	$O\left(n^{-\frac{1}{3}}\right)$	1.0	1.0	1.0	1.0
PH <sub>(1)</sub>	$O\left(n^{-\frac{1}{5}}\right)$	5.5	7.4	9.9	13.1
Histospline	$O\left(n^{-\frac{1}{7}}\right)$	7.3	12.0	18.8	27.9
S-PH <sub>(2,1)</sub>	$O\left(n^{-\frac{1}{9}}\right)$	<u>11.0</u>	<u>16.0</u>	<u>25.9</u>	<u>44.6</u>

表 3 は標本数  $n = 10^2 \sim 10^5$  に固定し、Histogram の最適ビン幅でそれぞれの最適なビン幅を割った値である。つまり、Histogram のビン幅を 1.0 としたときの 1 次 PH、Histospline、S-PH<sub>(2,1)</sub> の最適なビン幅である。 $n$  を変化させた時の最大の値に下線を引いてある。標本数に関わらず S-PH<sub>(2,1)</sub> のビン幅が他と比較して最も大きい。例えば、 $n = 10^5$  では Histogram の約 45 倍のビン幅となる。推定したビン幅で定義域を割ることでビン数が得られるが、S-PH<sub>(2,1)</sub> は他の推定法と比較して最も少ないビン数で推定できることが分かる。

## 4 結論と考察

本研究では、Polynomial Histogram の各ビンの境界で連続性を持たないという問題を解消するため、局所モーメント情報の保持と各ビンの境界での連続性の条件を同時に満たす平滑化 Polynomial Histogram(S-PH) 推定量を提案した。また、その理論的特性について漸近的性質を導出し、実証面については数値実験を行い、理論と実証の両面において提案モデルが有効であることを示した。

漸近的性質に関して、S-PH<sub>(2,1)</sub> 推定量の推定精度が MISE の意味で  $O(n^{-8/9})$  であることを示した。従来の主なビン型推定法と推定精度を比較すると、Histogram の  $O(n^{-2/3})$ 、局所モーメントの条件のみを満たす 1 次 PH の  $O(n^{-4/5})$ 、連続性の条件のみを満たす Histospline の  $O(n^{-6/7})$  を改良している。カーネル密度推定との関係では、S-PH<sub>(2,1)</sub> の MISE 収束オーダーは 4 次の高次カーネル推定に相当する。また、S-PH<sub>(2,1)</sub> は漸近正規性が成り立つことを示し、その陽な表現を導出した。

有限標本における特性を調べるため、二山の分布を例として ISE の標本平均 (MISE) について数値実験を行った。その結果、データ数が  $10^3$  以上の場合で、S-PH<sub>(2,1)</sub> の MISE が Histogram、1 次 PH、Histospline より小さく、良い推定結果を示した。各推定量の最適ビン幅に注目すると、データ数が  $10^5$  の時、今回の実験における S-PH<sub>(2,1)</sub> のビン幅は Histogram の約 45 倍である。このことは、S-PH<sub>(2,1)</sub> が広いビン幅を取ることで各ビンにおけるデータ数が多くなり、標本モー



ント情報が推定に効果的に利用できることを示している。視点を変え、ビン数の観点から見ると、S-PH<sub>(2,1)</sub> は最も少ないビン数で、より優れた推定精度を得た。本稿で提案した S-PH<sub>(2,1)</sub> による密度関数の推定法は大規模データの縮約に対して有効的な手法であると考えられる。

## Appendix 1 : AMISE $\{\hat{f}(x)\}$ の証明

MISE の定義は次の通りである。

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

ただし、IV と ISB は次のように定義される。

$$\begin{aligned} \text{IV}[\hat{f}(t)] &= \int \text{Var}[\hat{f}(t)] dt, \\ \text{ISB}[\hat{f}(t)] &= \int \text{Bias}[\hat{f}(t)]^2 dt. \end{aligned}$$

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

まず漸近積分二乗バイアスについて示す。S-PH<sub>(2,1)</sub> の密度推定量は (9) 式より、

$$\begin{aligned} \hat{f}_j(x) &= \left\{-\frac{6}{h^3}(x-t_j)^2 + \frac{3}{2h}\right\} \mu_j^{(0)} + \left\{-\frac{120}{h^5}(x-t_j)^3 + \frac{30}{h^3}(x-t_j)\right\} \mu_j^{(1)} \\ &\quad + \left\{\frac{10}{h^3}(x-t_j)^3 + \frac{3}{h^2}(x-t_j)^2 - \frac{3}{2h}(x-t_j) - \frac{1}{4}\right\} \\ &\quad \times \frac{1}{2h} \sum_{k=1}^{N-1} w_{j,k} \left\{(\mu_{k+1}^{(0)} + \mu_k^{(0)}) - \frac{10}{h}(\mu_{k+1}^{(1)} - \mu_k^{(1)})\right\} \\ &\quad + \left\{-\frac{10}{h^3}(x-t_j)^3 + \frac{3}{h^2}(x-t_j)^2 + \frac{3}{2h}(x-t_j) - \frac{1}{4}\right\} \\ &\quad \times \frac{1}{2h} \sum_{k=1}^{N-1} w_{j-1,k} \left\{(\mu_{k+1}^{(0)} + \mu_k^{(0)}) - \frac{10}{h}(\mu_{k+1}^{(1)} - \mu_k^{(1)})\right\}. \end{aligned} \quad (16)$$

ここで (9) 式の期待値を取ると、

$$\begin{aligned} E[\hat{f}_j(x)] &= c_1 E[\mu_j^{(0)}] + c_2 E[\mu_j^{(1)}] \\ &\quad + c_3 \sum_{k=1}^{N-1} w_{j,k} \left\{E[\mu_{k+1}^{(0)}] + E[\mu_k^{(0)}] - \frac{10}{h}(E[\mu_{k+1}^{(1)}] - E[\mu_k^{(1)}])\right\} \\ &\quad + c_4 \sum_{k=1}^{N-1} w_{j-1,k} \left\{E[\mu_{k+1}^{(0)}] + E[\mu_k^{(0)}] - \frac{10}{h}(E[\mu_{k+1}^{(1)}] - E[\mu_k^{(1)}])\right\}, \end{aligned} \quad (17)$$

ただし、

$$c_1 = -\frac{6}{h^3}(x-t_j)^2 + \frac{3}{2h}, \quad (18)$$

$$c_2 = -\frac{120}{h^5}(x-t_j)^3 + \frac{30}{h^3}(x-t_j), \quad (19)$$

$$c_3 = \frac{10}{h^3}(x-t_j)^3 + \frac{3}{h^2}(x-t_j)^2 - \frac{3}{2h}(x-t_j) - \frac{1}{4}, \quad (20)$$

$$c_4 = -\frac{10}{h^3}(x-t_j)^3 + \frac{3}{h^2}(x-t_j)^2 + \frac{3}{2h}(x-t_j) - \frac{1}{4}. \quad (21)$$

$\nu_k \sim B(n, p_k)$  ( $k = 1, 2, \dots, j-1, j, \dots, n$ ) で、 $p_k = \int_{B_k} f(t)dt$  の時、

$$E[\mu_k^{(0)}] = \frac{1}{n}E[\nu_k] = \frac{1}{n}np_k = \int_{B_k} f(t)dt, \quad (22)$$

$$E[\mu_k^{(1)}] = \int_{B_k} (t-t_k)f(t)dt, \quad (23)$$

ここで、未知の  $f(t)$  についてテイラー級数による近似から以下の通り表せられる。

$$\begin{aligned} \int_{B_k} f(t)dt \sim \int_{B_k} \left\{ f(x) + (t-x)f^{(1)}(x) + \frac{1}{2!}(t-x)^2f^{(2)}(x) + \frac{1}{3!}(t-x)^3f^{(3)}(x) \right. \\ \left. + \frac{1}{4!}(t-x)^4f^{(4)}(x) + \frac{1}{5!}(t-x)^5f^{(5)}(x) + \dots \right\} dt. \end{aligned} \quad (24)$$

(17) 式で (??)~(24) を用いて、 $\sum_{k=1}^{N-1} w_{j,k}k = j$ ,  $\sum_{k=1}^{N-1} w_{j,k}k^2 = j^2 + \frac{1}{2}$ ,  $\sum_{k=1}^{N-1} w_{j,k}k^3 = j^3 + \frac{3}{2}j$ ,  $\sum_{k=1}^{N-1} w_{j,k}k^4 = j^4 + 3j^2 + 2$  であることを利用して整理すると、 $\hat{f}_j(x)$  のバイアスは

$$\begin{aligned} \text{Bias}[\hat{f}_j(x)] &= E[\hat{f}_j(x)] - f(x) \\ &= -\frac{1}{24}f^{(4)}(x) \left\{ (x-t_j)^4 - \frac{h^2}{2}(x-t_j)^2 + \frac{7}{240}h^4 \right\}. \end{aligned} \quad (25)$$

したがって、ビン  $B_j$  における AISB $\{\hat{f}_j(x)\}$  は、

$$\begin{aligned} \text{AISB}\{\hat{f}_j(x)\} &= \int_{B_j} \frac{1}{576}f^{(4)}(\xi_j)^2 \left\{ (x-t_j)^8 - h^2(x-t_j)^6 + \frac{127}{120}(x-t_j)^4 \right. \\ &\quad \left. - \frac{7}{240}h^6(x-t_j)^2 + \frac{49}{57600}h^8 \right\} dx \\ &= \frac{h^9}{1209600}f^{(4)}(\xi_j)^2. \end{aligned} \quad (26)$$

以上より、全体での AISB $\{\hat{f}(x)\}$  はリーマン積分近似  $\sum_j f^{(4)}(\xi_j)^2 h = [\int f^{(4)}(x)^2 dx + o(1)]$  を用いて、

$$\text{AISB}\{\hat{f}(x)\} = \frac{R(f^{(4)})}{1209600}h^8. \quad (27)$$

次に漸近積分分散について示す。 $\hat{f}_j(x)$  の分散は、

$$\begin{aligned}
\text{Var} \left[ \hat{f}_j(x) \right] &= \text{Var} \left( c_1 \mu_j^{(0)} \right) + \text{Var} \left( c_2 \mu_j^{(1)} \right) \\
&+ \text{Var} \left[ c_3 \left\{ \sum_{k=1}^{N-1} w_{j,k} \left( \mu_{k+1}^{(0)} + \mu_k^{(0)} \right) - \frac{10}{h} \sum_{k=1}^{N-1} w_{j,k} \left( \mu_{k+1}^{(1)} - \mu_k^{(1)} \right) \right\} \right] \\
&+ \text{Var} \left[ c_4 \left\{ \sum_{k=1}^{N-1} w_{j-1,k} \left( \mu_{k+1}^{(0)} + \mu_k^{(0)} \right) - \frac{10}{h} \sum_{k=1}^{N-1} w_{j-1,k} \left( \mu_{k+1}^{(1)} - \mu_k^{(1)} \right) \right\} \right] \\
&+ 2\text{Cov} \left( c_1 \mu_j^{(0)}, c_2 \mu_j^{(1)} \right) \\
&+ 2\text{Cov} \left( c_1 \mu_j^{(0)}, c_3 \left\{ \sum_{k=1}^{N-1} w_{j,k} \left( \mu_{k+1}^{(0)} + \mu_k^{(0)} \right) - \frac{10}{h} \sum_{k=1}^{N-1} w_{j,k} \left( \mu_{k+1}^{(1)} - \mu_k^{(1)} \right) \right\} \right) \\
&+ 2\text{Cov} \left( c_1 \mu_j^{(0)}, c_4 \left\{ \sum_{k=1}^{N-1} w_{j-1,k} \left( \mu_{k+1}^{(0)} + \mu_k^{(0)} \right) - \frac{10}{h} \sum_{k=1}^{N-1} w_{j-1,k} \left( \mu_{k+1}^{(1)} - \mu_k^{(1)} \right) \right\} \right) \\
&+ 2\text{Cov} \left( c_2 \mu_j^{(1)}, c_3 \left\{ \sum_{k=1}^{N-1} w_{j,k} \left( \mu_{k+1}^{(0)} + \mu_k^{(0)} \right) - \frac{10}{h} \sum_{k=1}^{N-1} w_{j,k} \left( \mu_{k+1}^{(1)} - \mu_k^{(1)} \right) \right\} \right) \\
&+ 2\text{Cov} \left( c_2 \mu_j^{(1)}, c_4 \left\{ \sum_{k=1}^{N-1} w_{j-1,k} \left( \mu_{k+1}^{(0)} + \mu_k^{(0)} \right) - \frac{10}{h} \sum_{k=1}^{N-1} w_{j-1,k} \left( \mu_{k+1}^{(1)} - \mu_k^{(1)} \right) \right\} \right) \\
&+ 2\text{Cov} \left( c_3 \left\{ \sum_{k=1}^{N-1} w_{j,k} \left( \mu_{k+1}^{(0)} + \mu_k^{(0)} \right) - \frac{10}{h} \sum_{k=1}^{N-1} w_{j,k} \left( \mu_{k+1}^{(1)} - \mu_k^{(1)} \right) \right\}, \right. \\
&\quad \left. c_4 \left\{ \sum_{k=1}^{N-1} w_{j-1,k} \left( \mu_{k+1}^{(0)} + \mu_k^{(0)} \right) - \frac{10}{h} \sum_{k=1}^{N-1} w_{j-1,k} \left( \mu_{k+1}^{(1)} - \mu_k^{(1)} \right) \right\} \right). \tag{28}
\end{aligned}$$

以降、分散については主要項のみ取り出して記述する。

(28) の第 1 項は、

$$\begin{aligned}
\text{Var} \left( c_1 \mu_j^{(0)} \right) &= \frac{c_1^2}{n^2} \text{Var} \left( \nu_j \right) = \frac{c_1^2}{n^2} n p_j (1 - p_j) \\
&\sim \frac{c_1^2}{n} h f(x). \tag{29}
\end{aligned}$$

(28) の第 2 項は、

$$\text{Var} \left( c_2 \mu_j^{(1)} \right) = c_2^2 \text{Var} \left( \mu_j^{(1)} \right) \sim c_2^2 \frac{h^3}{12n} f(x). \tag{30}$$

(28) の第 3 項は、

$$\begin{aligned}
& \text{Var} \left[ c_3 \left\{ \sum_{k=1}^{N-1} w_{j,k} \left( \mu_{k+1}^{(0)} + \mu_k^{(0)} \right) - \frac{10}{h} \sum_{k=1}^{N-1} w_{j,k} \left( \mu_{k+1}^{(1)} - \mu_k^{(1)} \right) \right\} \right] \\
&= c_3^2 \left( 1 + \frac{100}{h^2} \right) \sum_{k=1}^{N-1} w_{j,k}^2 \left\{ \text{Var} \left( \mu_{k+1}^{(0)} + \mu_k^{(0)} \right) + \text{Var} \left( \mu_{k+1}^{(1)} - \mu_k^{(1)} \right) \right\} \\
&+ 2\text{Cov} \left( c_3 \sum_{k=1}^{N-1} w_{j,k} \left( \mu_{k+1}^{(0)} + \mu_k^{(0)} \right), -\frac{10}{h} c_3 \sum_{k=1}^{N-1} w_{j,k} \left( \mu_{k+1}^{(1)} - \mu_k^{(1)} \right) \right) \\
&\sim \frac{31\sqrt{2}}{6} c_3^2 \left( \frac{h}{n} \right) f(x). \tag{31}
\end{aligned}$$

(28) の第 4 項は、(31) と同様に、

$$\begin{aligned}
& \text{Var} \left[ c_4 \left\{ \sum_{k=1}^{N-1} w_{j-1,k} \left( \mu_{k+1}^{(0)} + \mu_k^{(0)} \right) - \frac{10}{h} \sum_{k=1}^{N-1} w_{j-1,k} \left( \mu_{k+1}^{(1)} - \mu_k^{(1)} \right) \right\} \right] \\
&\sim \frac{31\sqrt{2}}{6} c_4^2 \left( \frac{h}{n} \right) f(x). \tag{32}
\end{aligned}$$

(28) の第 5 項は、

$$\begin{aligned}
2\text{Cov} \left( c_1 \mu_j^{(0)}, c_2 \mu_j^{(1)} \right) &= 2c_1 c_2 \text{Cov} \left( \mu_j^{(0)}, \mu_j^{(1)} \right) \\
&\sim c_1 c_2 \frac{h^3}{6n} f'(x). \tag{33}
\end{aligned}$$

(28) の第 6, 7 項は、

$$\begin{aligned}
& 2\text{Cov} \left( c_1 \mu_j^{(0)}, c_3 \left\{ \sum_{k=1}^{N-1} w_{j,k} \left( \mu_{k+1}^{(0)} + \mu_k^{(0)} \right) - \frac{10}{h} \sum_{k=1}^{N-1} w_{j,k} \left( \mu_{k+1}^{(1)} - \mu_k^{(1)} \right) \right\} \right) \\
&+ 2\text{Cov} \left( c_1 \mu_j^{(0)}, c_4 \left\{ \sum_{k=1}^{N-1} w_{j-1,k} \left( \mu_{k+1}^{(0)} + \mu_k^{(0)} \right) - \frac{10}{h} \sum_{k=1}^{N-1} w_{j-1,k} \left( \mu_{k+1}^{(1)} - \mu_k^{(1)} \right) \right\} \right) \\
&\sim 2c_1 c_3 \left\{ \text{Cov} \left( \mu_j^{(0)}, \sum_{k=1}^{N-1} w_{j,k} \mu_{k+1}^{(0)} \right) + \text{Cov} \left( \mu_j^{(0)}, \sum_{k=1}^{N-1} w_{j,k} \mu_k^{(0)} \right) \right\} \\
&+ 2c_1 c_4 \left\{ \text{Cov} \left( \mu_j^{(0)}, \sum_{k=1}^{N-1} w_{j-1,k} \mu_{k+1}^{(0)} \right) + \text{Cov} \left( \mu_j^{(0)}, \sum_{k=1}^{N-1} w_{j-1,k} \mu_k^{(0)} \right) \right\} \\
&\sim 2c_1 (c_3 + c_4) \frac{1}{n^2} (w_{j,j-1} + w_{j,j} + w_{j-1,j-1} + w_{j-1,j}) \text{Var}(\nu_j) \\
&= 2\sqrt{2} \left( 2 - \sqrt{2} \right) c_1 (c_3 + c_4) \left( \frac{h}{n} \right) f(x). \tag{34}
\end{aligned}$$

(28) の第 8, 9 項は、

$$\begin{aligned}
& 2\text{Cov} \left( c_2 \mu_j^{(1)}, c_3 \left\{ \sum_{k=1}^{N-1} w_{j,k} (\mu_{k+1}^{(0)} + \mu_k^{(0)}) - \frac{10}{h} \sum_{k=1}^{N-1} w_{j,k} (\mu_{k+1}^{(1)} - \mu_k^{(1)}) \right\} \right) \\
& + 2\text{Cov} \left( c_2 \mu_j^{(1)}, c_4 \left\{ \sum_{k=1}^{N-1} w_{j-1,k} (\mu_{k+1}^{(0)} + \mu_k^{(0)}) - \frac{10}{h} \sum_{k=1}^{N-1} w_{j-1,k} (\mu_{k+1}^{(1)} - \mu_k^{(1)}) \right\} \right) \\
& \sim -\frac{20}{h} c_2 c_3 \left\{ \text{Cov} \left( \mu_j^{(1)}, \sum_{k=1}^{N-1} w_{j,k} \mu_{k+1}^{(1)} \right) - \text{Cov} \left( \mu_j^{(1)}, \sum_{k=1}^{N-1} w_{j,k} \mu_k^{(1)} \right) \right\} \\
& - \frac{20}{h} c_2 c_4 \left\{ \text{Cov} \left( \mu_j^{(1)}, \sum_{k=1}^{N-1} w_{j-1,k} \mu_{k+1}^{(1)} \right) - \text{Cov} \left( \mu_j^{(1)}, \sum_{k=1}^{N-1} w_{j-1,k} \mu_k^{(1)} \right) \right\} \\
& \sim -\frac{20}{h} c_2 (c_3 + c_4) (w_{j,j-1} - w_{j,j} + w_{j-1,j-1} - w_{j-1,j}) \text{Var} \left( \mu_j^{(1)} \right) \\
& = \frac{5\sqrt{2}}{3} (1 - \sqrt{2}) c_2 (c_4 - c_3) \left( \frac{h^2}{n} \right) f(x). \tag{35}
\end{aligned}$$

(28) の第 10 項は、 $\sum_{k=1}^{N-1} w_{j,k} w_{j,k+1} = \frac{\sqrt{2}}{8}$ ,  $\sum_{k=1}^{N-2} w_{j,k} w_{j-1,k+1} = \frac{64-45\sqrt{2}}{8}$ ,  $\sum_{k=1}^{N-2} w_{j-1,k}^2 = \frac{3\sqrt{2}}{8}$  を用いて、

$$\begin{aligned}
& 2\text{Cov} \left( c_3 \left\{ \sum_{k=1}^{N-1} w_{j,k} (\mu_{k+1}^{(0)} + \mu_k^{(0)}) - \frac{10}{h} \sum_{k=1}^{N-1} w_{j,k} (\mu_{k+1}^{(1)} - \mu_k^{(1)}) \right\}, \right. \\
& \quad \left. c_4 \left\{ \sum_{k=1}^{N-1} w_{j-1,k} (\mu_{k+1}^{(0)} + \mu_k^{(0)}) - \frac{10}{h} \sum_{k=1}^{N-1} w_{j-1,k} (\mu_{k+1}^{(1)} - \mu_k^{(1)}) \right\} \right) \\
& \sim 2c_3 c_4 \left( 2 \sum_{k=1}^{N-1} w_{j,k} w_{j,k+1} + \sum_{k=1}^{N-2} w_{j,k} w_{j-1,k+1} + \sum_{k=1}^{N-2} w_{j-1,k}^2 \right) \left( \frac{h}{n} \right) f(x) \\
& + \frac{200}{h^2} c_3 c_4 \left( 2 \sum_{k=1}^{N-1} w_{j,k} w_{j,k+1} - \sum_{k=1}^{N-2} w_{j,k} w_{j-1,k+1} - \sum_{k=1}^{N-2} w_{j-1,k}^2 \right) \left( \frac{h^3}{12n} \right) f(x) \\
& = \left( \frac{245\sqrt{2} - 352}{3} \right) c_3 c_4 \left( \frac{h}{n} \right) f(x). \tag{36}
\end{aligned}$$

(29) ~ (36) より、ビン  $B_j$  における漸近積分分散  $\text{AIV}\{\hat{f}_j(x)\}$  は、

$$\begin{aligned}
\text{AIV}\{\hat{f}_j(x)\} & = \left\{ \frac{6}{5h} + \frac{h^2}{12} \times \frac{120}{7h^3} + \frac{31\sqrt{2}}{6} \times \frac{3}{70h} + 2\sqrt{2} (2 - \sqrt{2}) \left( -\frac{1}{10h} \right) \right. \\
& \quad \left. + \frac{5\sqrt{2}}{3} (1 - \sqrt{2}) h \times \frac{3}{7h^2} + \left( \frac{245\sqrt{2} - 352}{3} \right) \frac{1}{280h} \right\} \frac{h}{n} f(x) \\
& = \left( \frac{992 + 695\sqrt{2}}{840} \right) \frac{1}{n} f(x). \tag{37}
\end{aligned}$$

したがって、全体での AIV  $\{\hat{f}(x)\}$  はリーマン積分近似より、

$$\text{AIV} \{\hat{f}(x)\} = \left( \frac{992 + 695\sqrt{2}}{840} \right) \frac{1}{nh}. \quad (38)$$

## Appendix 2 : $\hat{f}_j(x)$ の漸近正規性の証明

節点  $x_k = \frac{k}{N}$ ,  $k = 1, 2, \dots, N$  として、度数  $\nu_k$  を定義関数  $I(\cdot)$  を用いて表記すると、

$$\nu_k = \sum_{m=1}^n I_m \left( \frac{k-1}{N}, \frac{k}{N} \right), \quad (39)$$

ただし、

$$I_m(u, v) = \begin{cases} 1 & \text{if } u \leq X_m \leq v \\ 0 & \text{otherwise.} \end{cases} \quad (40)$$

各ビンの局所 1 次モーメント  $S_k$  を定義関数  $I(\cdot)$  を用いて表記すると、

$$S_k = \frac{1}{\nu_k} \sum_{m=1}^n (x_m - t_k) I_m \left( \frac{k-1}{N}, \frac{k}{N} \right), \quad (41)$$

ただし、 $I_m(u, v)$  は (40) と同様で、 $x_m (m = 1, 2, \dots, n)$  はデータ点である。

(16) を (39), (41) を用いて書き換えると、

$$\begin{aligned} \hat{f}_j(x) &= \frac{c_1}{n} \sum_{m=1}^n I_m \left( \frac{j-1}{N}, \frac{j}{N} \right) \\ &+ \frac{c_2}{n} \sum_{m=1}^n (x_m - t_j) I_m \left( \frac{j-1}{N}, \frac{j}{N} \right) \\ &+ \frac{c_3}{n} \sum_{k=1}^{N-1} w_{j,k} \left[ \sum_{m=1}^n I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) + \sum_{m=1}^n I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right. \\ &\quad \left. - \frac{10}{h} \left\{ \sum_{m=1}^n (x_m - t_{k+1}) I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) - \sum_{m=1}^n (x_m - t_k) I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right\} \right] \\ &+ \frac{c_4}{n} \sum_{k=1}^{N-1} w_{j-1,k} \left[ \sum_{m=1}^n I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) + \sum_{m=1}^n I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right. \\ &\quad \left. - \frac{10}{h} \left\{ \sum_{m=1}^n (x_m - t_{k+1}) I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) - \sum_{m=1}^n (x_m - t_k) I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right\} \right], \end{aligned} \quad (42)$$

ただし、 $c_1, \dots, c_4$  について  $x - t_j \in [-\frac{h}{2}, \frac{h}{2})$  より、有限な値を取る。

(42) より、

$$\begin{aligned}
Z_m = & \frac{1}{n} \left( c_1 I_m \left( \frac{j-1}{N}, \frac{j}{N} \right) \right. \\
& + c_2 (x_m - t_j) I_m \left( \frac{j-1}{N}, \frac{j}{N} \right) \\
& + c_3 \sum_{k=1}^{N-1} w_{j,k} \left[ I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) + I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right. \\
& \left. - \frac{10}{h} \left\{ (x_m - t_{k+1}) I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) - (x_m - t_k) I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right\} \right] \\
& + c_4 \sum_{k=1}^{N-1} w_{j-1,k} \left[ I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) + I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right. \\
& \left. - \frac{10}{h} \left\{ (x_m - t_{k+1}) I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) - (x_m - t_k) I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right\} \right] \Big). \quad (43)
\end{aligned}$$

このとき、

$$\hat{f}_j(x) \simeq \sum_{m=1}^n Z_m, \quad (44)$$

で、 $Z_m$  は独立同一分布に従う。

ここで、

$$E |Z_m - E[Z_m]|^3 \leq E |Z_m|^3 + 3E [Z_m^2] |E[Z_m]| + 4|E[Z_m]|^3. \quad (45)$$

$Z_m^2$  についてシュワルツの不等式より、

$$\begin{aligned}
Z_m^2 \leq & 6 \left( \frac{1}{n} \right)^2 \left[ c_1^2 \left\{ I_m \left( \frac{j-1}{N}, \frac{j}{N} \right) \right\}^2 \right. \\
& + c_2^2 \left\{ (x_m - t_j) I_m \left( \frac{j-1}{N}, \frac{j}{N} \right) \right\}^2 \\
& + c_3^2 \sum_{k=1}^{N-1} w_{j,k}^2 \left\{ I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) + I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right\}^2 \\
& + \left( \frac{100c_3^2}{h^2} \right) \sum_{k=1}^{N-1} w_{j,k}^2 \left\{ (x_m - t_{k+1}) I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) - (x_m - t_k) I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right\}^2 \\
& + c_4^2 \sum_{k=1}^{N-1} w_{j-1,k}^2 \left\{ I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) + I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right\}^2 \\
& \left. + \left( \frac{100c_4^2}{h^2} \right) \sum_{k=1}^{N-1} w_{j-1,k}^2 \left\{ (x_m - t_{k+1}) I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) - (x_m - t_k) I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right\}^2 \right]. \quad (46)
\end{aligned}$$

(46) について期待値を取ると、

$$\begin{aligned}
E[Z_m^2] \leq & 6 \left(\frac{1}{n}\right)^2 \left( c_1^2 E \left[ \left\{ I_m \left( \frac{j-1}{N}, \frac{j}{N} \right) \right\}^2 \right] \right. \\
& + c_2^2 E \left[ \left\{ (x_m - t_j) I_m \left( \frac{j-1}{N}, \frac{j}{N} \right) \right\}^2 \right] \\
& + c_3^2 E \left[ \sum_{k=1}^{N-1} w_{j,k}^2 \left\{ I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) + I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right\}^2 \right] \\
& + \left( \frac{100c_3^2}{h^2} \right) E \left[ \sum_{k=1}^{N-1} w_{j,k}^2 \left\{ (x_m - t_{k+1}) I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) - (x_m - t_k) I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right\}^2 \right] \\
& + c_4^2 E \left[ \sum_{k=1}^{N-1} w_{j-1,k}^2 \left\{ I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) + I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right\}^2 \right] \\
& \left. + \left( \frac{100c_4^2}{h^2} \right) E \left[ \sum_{k=1}^{N-1} w_{j-1,k}^2 \left\{ (x_m - t_{k+1}) I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) - (x_m - t_k) I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right\}^2 \right] \right). \tag{47}
\end{aligned}$$

(47) の大カッコ内の第 1 項目について、 $I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \stackrel{i.i.d.}{\sim} B(1, p_k)$  ( $k = 1, 2, \dots, N$ ) より、

$$\begin{aligned}
E \left[ \left\{ I_m \left( \frac{j-1}{N}, \frac{j}{N} \right) \right\}^2 \right] &= \text{Var} \left[ I_m \left( \frac{j-1}{N}, \frac{j}{N} \right) \right] + E \left[ I_m \left( \frac{j-1}{N}, \frac{j}{N} \right) \right]^2 \\
&= p_j (1 - p_j) + p_j^2 \\
&= p_j \\
&= \int_{B_j} f(t) dt \\
&\sim \int_{B_j} \{ f(x) + (t-x)f'(x) + \dots \} dt \\
&\sim hf(x). \tag{48}
\end{aligned}$$

したがって、前述の正則条件から  $f^{(4)}(x) < \infty$  より、

$$c_1^2 E \left[ \left\{ I_m \left( \frac{j-1}{N}, \frac{j}{N} \right) \right\}^2 \right] < \frac{C_1}{h}. \tag{49}$$



大カッコ内の第 2 項目について、 $x_m - t_k \in [-\frac{h}{2}, \frac{h}{2})$  ( $k = 1, 2, \dots, N$ ) を用いて、

$$\begin{aligned}
E \left[ \left\{ (x_m - t_j) I_m \left( \frac{j-1}{N}, \frac{j}{N} \right) \right\}^2 \right] &= \text{Var} \left[ (x_m - t_j) I_m \left( \frac{j-1}{N}, \frac{j}{N} \right) \right] + E \left[ (x_m - t_j) I_m \left( \frac{j-1}{N}, \frac{j}{N} \right) \right]^2 \\
&\leq h^2 \left( \text{Var} \left[ I_m \left( \frac{j-1}{N}, \frac{j}{N} \right) \right] + E \left[ I_m \left( \frac{j-1}{N}, \frac{j}{N} \right) \right]^2 \right) \\
&= h^2 \{ p_j (1 - p_j) + p_j^2 \} \\
&= h^2 p_j \\
&\sim h^2 f(x). \tag{50}
\end{aligned}$$

したがって、

$$c_2^2 E \left[ \left\{ (x_m - t_j) I_m \left( \frac{j-1}{N}, \frac{j}{N} \right) \right\}^2 \right] < \frac{C_2}{h}. \tag{51}$$

大カッコ内の第3項目について、

$$\begin{aligned}
& E \left[ \sum_{k=1}^{N-1} w_{j,k}^2 \left\{ I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) + I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right\}^2 \right] \\
&= \text{Var} \left[ \sum_{k=1}^{N-1} w_{j,k} \left\{ I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) + I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right\} \right] \\
&\quad + E \left[ \sum_{k=1}^{N-1} w_{j,k} \left\{ I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) + I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right\} \right]^2 \\
&= \sum_{k=1}^{N-1} w_{j,k}^2 \left( \text{Var} \left[ I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) \right] + \text{Var} \left[ I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right] \right) \\
&\quad + 2\text{Cov} \left( \sum_{k=1}^{N-1} w_{j,k} I_m \left( \frac{k}{N}, \frac{k+1}{N} \right), \sum_{k=1}^{N-1} w_{j,k} I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right) \\
&\quad + \sum_{k=1}^{N-1} w_{j,k}^2 \left( E \left[ I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) \right]^2 + E \left[ I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right]^2 \right) \\
&\quad + 2 \sum_{k=1}^{N-1} w_{j,k}^2 E \left[ I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) \right] E \left[ I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right] \\
&= \sum_{k=1}^{N-1} w_{j,k}^2 \{ p_{k+1} (1 - p_{k+1}) + p_k (1 - p_k) \} - 2 \sum_{k=1}^{N-1} w_{j,k}^2 p_{k+1} p_k \\
&\quad + \sum_{k=1}^{N-1} w_{j,k}^2 (p_{k+1}^2 + p_k^2) + 2 \sum_{k=1}^{N-1} w_{j,k}^2 p_{k+1} p_k \\
&= \sum_{k=1}^{N-1} w_{j,k}^2 (p_{k+1} + p_k) \\
&\sim 2 \sum_{k=1}^{N-1} w_{j,k}^2 h f(x). \tag{52}
\end{aligned}$$

したがって、

$$c_3^2 E \left[ \sum_{k=1}^{N-1} w_{j,k}^2 \left\{ I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) + I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right\}^2 \right] < \frac{C_3}{h}. \tag{53}$$

大カッコ内の第 4 項目について、

$$\begin{aligned}
& E \left[ \sum_{k=1}^{N-1} w_{j,k}^2 \left\{ (x_m - t_{k+1}) I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) - (x_m - t_k) I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right\}^2 \right] \\
&= \text{Var} \left[ \sum_{k=1}^{N-1} w_{j,k} \left\{ (x_m - t_{k+1}) I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) - (x_m - t_k) I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right\} \right] \\
&\quad + E \left[ \sum_{k=1}^{N-1} w_{j,k} \left\{ (x_m - t_{k+1}) I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) - (x_m - t_k) I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right\} \right]^2 \\
&\leq h^2 \sum_{k=1}^{N-1} w_{j,k}^2 \left( \text{Var} \left[ I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) \right] + \text{Var} \left[ I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right] \right) \\
&\quad - 2h^2 \text{Cov} \left( \sum_{k=1}^{N-1} w_{j,k} I_m \left( \frac{k}{N}, \frac{k+1}{N} \right), \sum_{k=1}^{N-1} w_{j,k} I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right) \\
&\quad + h^2 \sum_{k=1}^{N-1} w_{j,k}^2 \left( E \left[ I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) \right] + E \left[ I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right] \right)^2 \\
&= \sum_{k=1}^{N-1} w_{j,k}^2 (p_{k+1} + p_k) \\
&\sim 2 \sum_{k=1}^{N-1} w_{j,k}^2 h f(x). \tag{54}
\end{aligned}$$

したがって、

$$\left( \frac{100c_3^2}{h^2} \right) E \left[ \sum_{k=1}^{N-1} w_{j,k}^2 \left\{ (x_m - t_{k+1}) I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) - (x_m - t_k) I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right\}^2 \right] < \frac{C_4}{h}. \tag{55}$$

大カッコ内第 5, 6 項目は (53), (55) と同様にして、

$$c_4^2 E \left[ \sum_{k=1}^{N-1} w_{j-1,k}^2 \left\{ I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) + I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right\}^2 \right] < \frac{C_5}{h}. \tag{56}$$

$$\left( \frac{100c_4^2}{h^2} \right) E \left[ \sum_{k=1}^{N-1} w_{j-1,k}^2 \left\{ (x_m - t_{k+1}) I_m \left( \frac{k}{N}, \frac{k+1}{N} \right) - (x_m - t_k) I_m \left( \frac{k-1}{N}, \frac{k}{N} \right) \right\}^2 \right] < \frac{C_6}{h}. \tag{57}$$

以上より、(49)~(57) から  $E [Z_m^2]$  は、

$$E [Z_m^2] < C \left( \frac{1}{n} \right)^2 \times O \left( \frac{1}{h} \right) \sim O \left( \frac{1}{n^2 h} \right). \tag{58}$$

また、 $|Z_m| = O\left(\frac{1}{nh}\right)$  を用いて、

$$|Z_m|^3 < Z_m^2 O\left(\frac{1}{nh}\right), \quad (59)$$

であることから、

$$E|Z_m|^3 = O\left(\frac{1}{n^2h}\right) \times O\left(\frac{1}{nh}\right) = O\left(\frac{1}{n^3h^2}\right). \quad (60)$$

したがって、(45)、(58) より、

$$\begin{aligned} \sum_{m=1}^n E|Z_m - E[Z_m]|^3 &= O\left(\frac{1}{n^2h^2}\right) + o\left(\frac{1}{n^2h}\right) + o\left(\frac{1}{n^2}\right) \\ &= O\left(\frac{1}{n^2h^2}\right). \end{aligned} \quad (61)$$

$\hat{f}_j(x)$  の分散を  $\sigma[\hat{f}_j(x)]^2$  とすると、 $\sigma[\hat{f}_j(x)]^2 \sim \frac{C}{nh}$  より、

$$\frac{\sum_{m=1}^n E|Z_m - E[Z_m]|^3}{\sigma[\hat{f}_j(x)]^3} = \frac{O\left(\frac{1}{n^2h^2}\right)}{O\left(\frac{1}{n^{\frac{3}{2}}h^{\frac{3}{2}}}\right)} = O\left(\frac{1}{\sqrt{nh}}\right) = o(1). \quad (62)$$

## 参考文献

- [1] L. I. Boneva, D. Kendall, and I. Stefanov(1971), "Spline Transformations: Three New Diagnostic Aids for the Statistical Data-Analyst." *Journal of the Royal Statistical Society. Series B (Methodological)*, 33.1, 1-71.
- [2] M. C. Jones, M. Samiuddin, A. H. AL-Harbey and T. A. H. Maatouk(1998), "The edge frequency polygon", *Biometrika*, Vol.85, Issue 1, 235-239.
- [3] Keh-Shin Lii, and M. Rosenblatt(1975), "Asymptotic behavior of a spline estimate of a density function", *Computers & Mathematics with Applications*, 1(2), 223-235.
- [4] M. C. Minnotte(1996), "The Bias-Optimized Frequency Polygon", *Computational Statistics*, 11, 35-48.
- [5] M. C. Minnotte(1998), "Achieving Higher-Order Convergence Rates for Density Estimation with Binned Data", *Journal of the American Statistical Association*, Vol.93, No.442, 663-672.
- [6] M. Sagae and D. W. Scott(1997a), "Bin Interval Method of Locally Adaptive Nonparametric Density Estimation", Joint Statistical Meetings 1997.
- [7] I. J. Schoenberg(1973), "Splines and Histograms", *Spline Functions and Approximation Theory*, Birkhauser, Basel, 277-327.
- [8] D. W. Scott(1985), "Frequency Polygons: Theory and Application", *Journal of the American Statistical Association*, Vol. 80, No. 390, 348-354.

- [9] D. W. Scott and M. Sagae(1997b), "Adaptive Density Estimation with Massive Data Sets", Joint Statistical Meetings 1997.
- [10] 齊藤, 寒河江 (2021), "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～", 人間社会環境研究, 第 41 号, 49-62.