

令和 3 年 6 月 7 日現在

機関番号：12102

研究種目：若手研究(B)

研究期間：2017～2020

課題番号：17K13714

研究課題名（和文）裾・境界バイアス問題のないノンパラメトリック直接型密度比推定量とその応用について

研究課題名（英文）Tail- or boundary-bias-free nonparametric direct density ratio estimation and its application

研究代表者

五十嵐 岳 (Igarashi, Gaku)

筑波大学・システム情報系・助教

研究者番号：40759346

交付決定額（研究期間全体）：（直接経費） 2,000,000円

研究成果の概要（和文）：確率密度関数の比（確率密度比）をノンパラメトリックに直接推定する、直接型確率密度比推定に焦点を置いて研究を行い、ベータカーネルを用いた裾・境界バイアスのない確率密度比推定量を提案した。さらに、推定量の滑らかさを制御する平滑化パラメータの選択方法についても研究を行い、leave-one-out交差検証法による平滑化パラメータ選択法を考案した。また、直接型ベータカーネル確率密度比推定量を検定統計量に用いた確率密度の不連続性検定を提案し、検定のための平滑化パラメータ選択法を考案、片側検定の際に検出力が大きくなることを明らかにした。

研究成果の学術的意義や社会的意義

確率密度関数の比（確率密度比）は、条件付き確率密度推定や、異常値検出などの2標本検定に用いられる。裾・境界バイアスのない確率密度比推定は、裾・境界における推定・検定精度の向上に役立つと考えられる。また、確率密度の不連続性検定は回帰不連続デザインの条件確認等に用いられ、直接型ベータカーネル確率密度比推定量を用いた検定は、片側検定の際に精度が向上が期待できる。

研究成果の概要（英文）：We focused on direct probability density ratio estimation, which directly estimates the ratio of probability density function nonparametrically. We proposed a probability density ratio estimator, which is free of tail or boundary bias, using the beta kernel. We also studied methods of selecting smoothing parameter which control the smoothness of the estimator, and proposed a method by the leave-one-out cross-validation. Furthermore, we also proposed a nonparametric discontinuity test of probability density using the direct beta kernel probability density ratio estimator as the test statistic, and a smoothing parameter selection method for the test. We showed that the power of the proposed test is large when used as a one-tailed test.

研究分野：統計科学

キーワード：ノンパラメトリック密度比推定

1. 研究開始当初の背景

統計科学における確率密度関数の比の推定方法としては、母数モデルを仮定し、母数を推定するパラメトリック法と、母数モデルを仮定しないノンパラメトリック法が主流である。ノンパラメトリックな密度比推定量では、2つの密度を個別に推定してその比を取る間接型密度比推定量があるが、密度比そのものを直接推定する直接型密度比推定量がĆwik and Mielniczuk (1989) により提案されている。この直接型密度比推定の手法は、ハザード比の直接型推定や、条件付き密度の直接型推定にも発展可能で、ノンパラメトリック回帰にも応用できる。

密度比をデータから推定するとき、性能を『平均2乗誤差』(バイアスの2乗と分散の和に等しい)や『平均積分2乗誤差』(平均2乗誤差を積分した基準)で測る。統計科学においては、データのサイズ n が大きい場合の『漸近理論』を考え、平均2乗誤差や平均積分2乗誤差が速く0に収束し、より小さい方が望ましい(以降、この収束の速さを『収束比』と呼ぶ)。

直接型密度比推定量は、代表的なノンパラメトリック密度推定量であるカーネル推定量(Rosenblatt (1956))を応用したものである。ところで、カーネル推定量は、推定対象の密度の台が有界区間や半無限区間 $([0, 1]$ や $[0, \infty)$ など)の場合、境界付近でバイアスがあって、平均2乗誤差も0に収束しない問題(以降、『境界バイアス問題』と呼ぶ)がある。境界バイアス問題に対して、その対処法がいくつか議論されている(Jones (1993)など)。また、ベータカーネル推定量(Chen (1999))などの境界バイアス問題のない推定量も提案されている。さらには、これら境界バイアス問題のない推定量の収束比を改良する議論もなされている(Hirukawa (2010)など)。

カーネル推定量の手法を応用した直接型密度比推定量にも境界バイアス問題が存在しており、さらに、推定対象の密度の台が $(-\infty, \infty)$ で境界がない場合でも、直接型密度比推定量では裾でバイアスがあり、境界がある場合と同様の問題(以降、『裾・境界バイアス問題』と呼ぶ)が存在する。Mielniczuk (1991)は、カーネル推定量における境界バイアス問題の対処法の1つであるリフレクション法を応用し、裾・境界バイアス問題を軽減する直接型密度比推定量について議論している。

なお、ノンパラメトリックな密度比推定法には、バイアスと分散のトレードオフを制御するパラメータ(以降、『平滑化パラメータ』と呼ぶ)が存在し、実際の応用の際には平滑化パラメータを選択する問題があり、Ćwik and Mielniczuk (1993)が直接型密度比推定量の平滑化パラメータ選択法について議論している。

2. 研究の目的

境界バイアスのない密度推定量の研究を行っていた経緯から、直接型密度比推定量の裾・境界バイアス問題には、密度推定量と同様の手法が有効であると考えられる。従って、まず、ベータカーネル推定量を応用して、裾・境界バイアス問題のない直接型密度比推定量を開発しその数理を明らかにすることを目的とする。先行研究では裾・境界バイアス問題を軽減するに留まっており、それが一番の特色である。裾においても直接型密度比推定量により密度比を推定可能になることは、大変意義がある。

さらに、裾・境界バイアス問題のない直接型密度比推定量について収束比を改良することをめざす。一般に、ノンパラメトリック法はパラメトリック法に比べて収束比が遅いため、直接

型密度比推定量の収束比を改良する研究は、非常に意義がある。

次に、平滑化パラメータの選択は推定量の性能に大きく影響するため、推定量の性能を最大限に発揮させる平滑化パラメータの選択法を開発する。

そして、直接型密度比推定量を密度の不連続性の検定 (以降、『不連続性検定』と呼ぶ) に応用する。McCrary (2008) は、間接型密度比推定量を用いて不連続性検定を行うことを提案している。この検定では、不連続点で密度を分割して、それぞれを分子、分母の密度としてその点で密度比を推定し、密度が連続であれば分子、分母の密度の値は等しくなり密度比が1になることを利用して、不連続性を検定している。この手法を応用して、直接型密度比推定量による、密度の不連続性の検定を開発する。ところで、間接型密度比推定量の漸近分散は、推定対象の、本来未知である密度に依存することが知られている。したがって、間接型密度比推定量による不連続性検定では、その分散を推定して検定統計量を構成している。直接型密度比推定量の漸近分散も通常は未知の密度に依存するが、分子、分母の密度が等しい場合には、その密度に依存しなくなるという性質がある。そのため、不連続性の検定においては、漸近分散を推定する必要がないという特色があり、間接型密度比推定量による検定に比べて、利点がある。

3. 研究の方法

(1) 裾・境界バイアス問題のない直接型密度比推定量を構成して、その漸近的性質を導出する。具体的には、境界バイアス問題のないノンパラメトリック密度推定量であるベータカーネル推定量 (Chen (1999)) を応用して推定量を構成する。このとき、以下のことに注意する必要がある。

- ① 不連続性検定では、密度を分割してその端点に注目するため、推定量の境界でのふるまいが重要になる。ベータカーネル推定量は、境界バイアス問題はないものの、境界付近で分散が大きくなるという問題がある。分散の大きさは推定、検定の性能に関わり、小さいほうが好ましい。単に応用しただけでは、その問題を引き継ぐ可能性があるため、その問題が起こらないように推定量を構成する。
- ② Ćwik and Mielniczuk (1989) の直接型密度比推定量では、分子、分母の密度が等しい場合には、漸近分散がその密度に依存しなくなり、安定するという、不連続性検定において望ましい性質が存在する。ベータカーネル推定量を応用した直接型密度比推定量も、同様の性質を持つよう推定量を構成する。

このような①, ②の点に注意しながら、推定量の漸近性能を数学的基礎研究で明らかにする。具体的には、裾・境界バイアス問題のない直接型密度比推定量の漸近的性質 (バイアス, 分散, 平均2乗誤差, 漸近正規性など) を導出していく。

乱数によるシミュレーション実験を実施して、得られた諸結果の理論を検証する。

(2) (1) で構成した裾・境界バイアス問題のない直接型密度比推定量について収束比を改良することをめざす。具体的には、密度推定量の収束比の改良で用いられている、Schucany and Sommers (1977) の加法型バイアス修正法を直接型密度比推定量に応用する。ここで、SS型バイアス修正法は、異なる平滑化パラメータを持つ改良前の2つの推定量をそれぞれ加法、乗法で組み合わせることにより収束比を改良するというアイディアに基づく。改良した裾・

境界バイアス問題のない直接型密度比推定量の漸近的性質を導出して、その漸近性能を明らかにする。

(3) ノンパラメトリックな密度比推定量には、実際の応用に際して平滑化パラメータを選択する問題がある。平滑化パラメータの選択は、推定量の性能に大きく影響するため、平滑化パラメータ選択法の研究を行う。平滑化パラメータは、通常、平均積分2乗誤差を最小にして選ばれる。従って、研究の初期段階において平均積分2乗誤差の漸近公式が導出できれば、最適な平滑化パラメータも理論的に導出することができる。しかし、そのような最適な平滑化パラメータは、推定対象となる、本来未知である密度やその導関数に依存してしまうため、実際には使用できない。そこで、従来の直接型密度比推定量やカーネル密度推定量の平滑化パラメータ選択法を詳細に文献調査及び再検討して、裾・境界バイアス問題のない直接型密度比推定量へ応用し、乱数によるシミュレーション実験を実施する。

(4) 裾・境界バイアス問題のない直接型密度比推定量の応用として不連続性検定を念頭に置いている。McCrary (2008) による間接型密度比推定量を用いた不連続性検定では、不連続性を検定する点で密度を分割し、それぞれを分母と分子の密度として比を取り、境界バイアスのない密度推定量により、その点の密度比を間接的に推定する。そして、連続であるならば、つまり、分母と分子の密度が等しければ密度比は1になるという性質を利用して検定を行っている。従って、裾・境界バイアス問題のない直接型密度比推定量により、分割した密度の比を推定することを考えて、不連続性検定へ応用する。さらに、シミュレーション実験により、その検定の性能を検討する。

4. 研究成果

(1) ガンマ密度の一般化である、一般化ガンマ(アモロソ)密度をカーネルに用いた、台が非負である密度を対象とする境界バイアスのない一般化ガンマ(アモロソ)カーネル密度推定量の漸近的性質(バイアス, 分散, 強一貫性, 漸近正規性, 平均積分2乗誤差)を示した。さらに、その平均積分2乗誤差の収束比を改良する、バイアス修正を行った。また、数値実験を実施して、有限標本においてバイアス修正が機能していることを示した。

(2) 非対称カーネルを用いた境界バイアスのないノンパラメトリック密度推定量について、既存研究で行われた3つのバイアス修正法をそれぞれ一般化することにより、高次のバイアス修正法を提案して、その漸近的性質(バイアス, 分散, 一貫性, 漸近正規性, 平均2乗誤差)を導出した。また、数値実験により、有限標本における性能を検証した。

(3) 従来のノンパラメトリックな直接型カーネル密度比推定量には、裾や境界での推定に問題があり、実際に裾や境界で0に収束しないバイアスがあり、したがって、裾や境界では平均2乗誤差も0に収束しないことを示した。また、数値実験を実施して、有限標本でも裾・境界付近での推定に問題があることを示した。

さらに、既存研究であるリフレクション法を用いて裾・境界バイアスを補正した直接型カーネル密度比推定量について、裾・境界バイアスを導出し、一貫性はあるものの、裾・境界では内部のバイアスよりも収束が遅いことを示した。

(4) ベータカーネルを直接型密度比推定に応用した、直接型ベータカーネル密度比推定量の漸近的性質(バイアス, 分散, 平均2乗誤差, 漸近正規性)を導出し、裾・境界バイアスがなないことを示した。

また、実際の利用には平滑化パラメータを決める問題が存在するため、leave-one-out 交差検証法を用いた平滑化パラメータ選択法を考案した。さらに、その leave-one-out 交差検証法による平滑化パラメータ選択法を裾・境界バイアス問題のある従来の直接型カーネル密度比推定量及び、リフレクション法を用いて裾・境界バイアスを補正した直接型カーネル密度比推定量にも応用し、密度比推定の数値実験によって、考案した平滑化パラメータ選択法が正常に働くこと、直接型ベータカーネル密度比推定量が有限標本でも裾や境界で上手く推定できること、リフレクションによる直接型密度比推定量に匹敵する性能を持つことを示した。

さらに、実際のデータに適用した際のふるまいを確認するため、アメリカにあるイエローストーン国立公園のオールドフェイスフル間欠泉の2つの標本を用いて確率密度関数の比を推定した。

(5) 直接型ベータカーネル密度比推定量について、異なる平滑化パラメータを持つ2つの推定量の線形結合によってバイアスを削減する、ジャックナイフ法によるバイアス修正(加法型バイアス修正)を行い、推定精度を改善する推定量を提案し、漸近的性質(バイアス, 分散, 平均2乗誤差, 漸近正規性)を導出した。さらに、そのバイアス修正推定量を用いて確率密度関数の不連続性を仮説検定するための検定統計量を考案した。

また、平均2乗誤差を最小にするような理論的に最適な平滑化パラメータには、仮説検定の対象である未知の確率密度関数やその導関数に関する情報が必要なことから実際には実行不可能であるが、それらの確率密度関数・導関数について推定量を考案して置き換えたり、簡単な仮定を置くことによる平滑化パラメータ選択法を考案した。

提案した検定統計量は、不連続点で下側(左側)の値が上側(右側)の値より大きいときに分散が小さくなることが明らかになった。そのため、その条件が満たされないときは用いる標本を反射させるといった変換を行うことにより、片側検定の際に検出力を大きくすることを提案した。さらに数値実験により、片側検定の際には先行研究の検定統計量より概ね大きな検出力が得られることを確認した。

参考文献

- Chen, S. X. (1999) "Beta kernel estimators for density functions," *Computational Statistics and Data Analysis*, 31, 131–145.
- Ćwik, J. and Mielniczuk, J. (1989) "Estimating density ratio with application to discriminant analysis," *Communications in Statistics - Theory and Methods*, 18, 3057–3069.
- Ćwik, J. and Mielniczuk, J. (1993) "Data-dependent bandwidth choice for a grade density kernel estimate," *Statistics and Probability Letters*, 16, 397–405.
- Hirukawa, M. (2010) "Nonparametric multiplicative bias correction for kernel-type density estimation on the unit interval," *Computational Statistics and Data Analysis*, 54, 473–495.
- Jones, M. C. (1993) "Simple boundary correction for kernel density estimation," *Statistics and Computing*, 3, 135–146.
- McCrary, J. (2008) "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, 142, 698–714.
- Mielniczuk, J. (1991) "Grade estimation of chi-square divergence," *Communications in Statistics - Theory and Methods*, 20, 4021–4041.
- Rosenblatt, M. (1956) "Remarks on some nonparametric estimates of a density function," *The Annals of Mathematical Statistics*, 27, 832–837.
- Schucany, W. R. and Sommers, J. P. (1977) "Improvement of kernel type density estimators," *Journal of the American Statistical Association*, 72, 420–423.

5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 4件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 Igarashi Gaku, Kakizawa Yoshihide	4. 巻 141
2. 論文標題 Multiplicative bias correction for asymmetric kernel density estimators revisited	5. 発行年 2020年
3. 雑誌名 Computational Statistics & Data Analysis	6. 最初と最後の頁 40 ~ 61
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.csda.2019.06.010	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Igarashi Gaku	4. 巻 54
2. 論文標題 Nonparametric direct density ratio estimation using beta kernel	5. 発行年 2020年
3. 雑誌名 Statistics	6. 最初と最後の頁 257 ~ 280
掲載論文のDOI (デジタルオブジェクト識別子) 10.1080/02331888.2020.1722671	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Igarashi Gaku, Kakizawa Yoshihide	4. 巻 30
2. 論文標題 Generalized gamma kernel density estimation for nonnegative data and its bias reduction	5. 発行年 2018年
3. 雑誌名 Journal of Nonparametric Statistics	6. 最初と最後の頁 598 ~ 639
掲載論文のDOI (デジタルオブジェクト識別子) 10.1080/10485252.2018.1457791	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Igarashi Gaku, Kakizawa Yoshihide	4. 巻 47
2. 論文標題 Limiting bias-reduced Amoroso kernel density estimators for non-negative data	5. 発行年 2018年
3. 雑誌名 Communications in Statistics - Theory and Methods	6. 最初と最後の頁 4905 ~ 4937
掲載論文のDOI (デジタルオブジェクト識別子) 10.1080/03610926.2017.1380832	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計7件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 五十嵐岳
2. 発表標題 ベータカーネルを用いた密度比の直接型推定量について
3. 学会等名 2018年度統計関連学会連合大会
4. 発表年 2018年

1. 発表者名 五十嵐岳
2. 発表標題 ベータカーネルを用いた境界バイアスのない直接型密度比推定
3. 学会等名 日本数学会2019年度年会
4. 発表年 2019年

1. 発表者名 五十嵐岳、柿沢佳秀
2. 発表標題 Amorosoカーネル密度推定量に対する収束比の改良について
3. 学会等名 2017年度統計関連学会連合大会
4. 発表年 2017年

1. 発表者名 五十嵐岳
2. 発表標題 ベータカーネルを用いた境界問題のない密度比推定
3. 学会等名 研究集会「第19回ノンパラメトリック統計解析とベイズ統計」
4. 発表年 2018年

1. 発表者名 五十嵐岳
2. 発表標題 ノンパラメトリックベータカーネル密度比推定による確率密度の不連続性検定
3. 学会等名 2020年度統計関連学会連合大会
4. 発表年 2020年

1. 発表者名 五十嵐岳
2. 発表標題 ノンパラメトリックベータカーネル密度比推定による確率密度の不連続性検定
3. 学会等名 2020年度日本数学会秋季総合分科会
4. 発表年 2020年

1. 発表者名 五十嵐岳
2. 発表標題 ベータカーネルを用いたノンパラメトリックな確率密度の不連続性検定
3. 学会等名 研究集会「第21回ノンパラメトリック統計解析とベイズ統計」
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------