

令和 6 年 6 月 11 日現在

機関番号：62603

研究種目：若手研究

研究期間：2019～2023

課題番号：19K20222

研究課題名（和文）構造制約に着目した高次元カウントデータの未知母数推定法と不確実性評価法の構築

研究課題名（英文）Construction of estimation methods and their uncertainty quantification methods for high-dimensional count data focusing on structural constraints

研究代表者

矢野 恵佑 (Yano, Keisuke)

統計数理研究所・統計基盤数理研究系・准教授

研究者番号：20806070

交付決定額（研究期間全体）：（直接経費） 3,100,000円

研究成果の概要（和文）：高次元モデルやカウントデータモデルで利用可能なベイズ予測分布に基づく予測モデルの評価法を確立した。ベイズ予測分布に基づく予測モデルの評価ではWidely Applicable Information Criterion (WAIC)が広く活用されている。WAICの高次元モデルでの理論的な妥当性を示し、さらに深層学習を含む高次元モデルで効率的に計算する手法を確立した。さらに、WAICを観測の重みが存在する、予測と観測の評価関数が異なる、対数損失以外の予測評価関数を用いる、場合に拡張した Posterior Covariance Information Criterion (PCIC)を構築した。

研究成果の学術的意義や社会的意義

高次元モデルやカウントデータモデルは諸科学で広く現れる統計モデルである。しかし、その推論法は通常モデルと比べて十分に定まっているとはいえない。本研究では高次元モデルやカウントデータモデルで利用可能なベイズ予測分布に基づく予測モデルの評価法を確立した。これにより従来はできなかった予測評価（深層学習を含む高次元モデルでの予測評価・観測の重みが存在する場合の予測評価・予測と観測の評価関数が異なる場合の予測評価・対数損失以外の予測評価関数を用いた場合の評価）が可能となった。

研究成果の概要（英文）：This project has established evaluation methods for predictive models based on Bayesian predictive distributions that are applicable to high-dimensional models including count data models. The Widely Applicable Information Criterion (WAIC) has been extensively used for evaluating predictive models based on Bayesian predictive distributions. We demonstrated the theoretical validity of WAIC in high-dimensional models and established efficient computational methods for high-dimensional models, including deep learning. Furthermore, we established an extension of WAIC, the Posterior Covariance Information Criterion (PCIC), which accommodates cases where there are weights on observations, different evaluation functions for predictions and observations, and predictive evaluation functions other than the logarithmic loss.

研究分野：統計学

キーワード：情報量規準 ベイズ予測 MCMC

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

高次元カウントデータに関する高精度かつ高速な統計解析手法を構築する。高次元カウントデータは多岐にわたる学術分野や実社会で現れる。例えば、商品の購入者数、犯罪発生件数、地震の発生件数、遺伝子の発現数、太陽の黒点数などである。近年、擬疎性という実データに現れる構造制約に着目して高精度な未知母数推定法を構築されつつある。より幅広い高次元カウントデータに適用するためには「擬疎性以外の構造制約の考慮」が必要であり推定結果の公開には「実用的な不確実性評価法の構築」が不可欠である。本研究では「構造制約に着目した高次元カウントデータの未知母数推定法と不確実性評価法の構築」を目指す。

2. 研究の目的

高次元カウントデータに関する高精度かつ高速な統計解析手法を構築する。高次元カウントデータは様々な学術分野で現れる。申請者は擬疎性という実データに現れる構造制約に着目して高精度な未知母数推定法を構築した。より幅広い高次元カウントデータに適用するためには「擬疎性以外の構造制約の考慮」が必要であり推定結果の公開には「実用的な不確実性評価法の構築」が不可欠である。本研究では「構造制約に着目した高次元カウントデータの未知母数推定法と不確実性評価法の構築」を目指す。

3. 研究の方法

「構造制約に着目した高次元カウントデータの未知母数推定法と不確実性評価法」を構築することで上記の目的を達成する。

4. 研究成果

高次元カウントデータに関する高精度かつ高速な統計解析手法の構築に着手した。高次元カウントデータは多岐にわたる学術分野や実社会で現れる。例えば、商品の購入者数、犯罪発生件数、地震の発生件数、遺伝子の発現数、太陽の黒点数などである。

本年度はゼロ過剰性を持つ高次元カウントデータの高精度な母数推定法を構築した。高次元カウントデータは、ゼロあるいはゼロに近いカウントを過剰にもつというゼロ過剰という性質をしばしば持っている。本研究では、ゼロ過剰という性質を「疎性」の枠組みで捉え、スパイクスラブ型の事前分布を利用したベイズ推論法を提案した。ポアソン分布のガンマ分布混合が負の二項分布になることに着目し、提案手法からの高速な標本抽出法を開発した。提案手法の実データでの性能検証を、東京都のオープンデータ(東京都におけるスリの発生件数)と遺伝子発現データを用いて行った。

スパイクスラブ型の事前分布は通常の高次元回帰の文脈で非常によく使われている。しかし、本研究の重要な理論的結果として、「事前分布の裾の重さはデータの分布の裾の重さよりも重くしなければ精度が悪化してしまう」ということが分かった。この結果はスパイクスラブ型の前分布の適用に重要な示唆を与えるものである。更に、驚くべき性質として、疎性に関する情報を解析者が知らなかったとしても、提案手法では精度を落とすことなく推論を行うことができる、すなわち、疎性に対する適応性をもつことが分かった。疎性に対する適用性は応用上非常に重要であるにも関わらず理論的にその性質が示されている手法はごく僅かである。特に高次元カウントデータで示したのは本研究が初である。

また、分散不均一性をもつ階層モデルでのベイズ予測分布の構築を議論した。実データに存在する観測値ごとの分散不均一性に対処した階層モデルでの分布予測法は未だ確立されていない。予測分布のロバスト性を議論するため損失として $-divergence$ を考え、一般の $-divergence$ のリスク不偏推定量を用い、 $divergence$ に対して適用的に決めた事前分布に基づくベイズ予測分布を提案した。提案予測分布について $-divergence$ の意味での最適性を示した。提案予測分布をウィルス感染のマーカーとなる遺伝子の同定問題に適用しその有用性を検証した。

最後に、高次元モデルやカウントデータモデルで利用可能なベイズ予測分布に基づく予測モデルの評価法を確立した。ベイズ予測分布に基づく予測モデルの評価では Widely Applicable Information Criterion (WAIC) が広く活用されている。WAIC の高次元モデルでの理論的な妥当性を示し、さらに深層学習を含む高次元モデルで効率的に計算する手法を確立した。さらに、WAIC を観測の重みが存在する、予測と観測の評価関数が異なる、対数損失以外の予測評価関数を用いる、場合に拡張した Posterior Covariance Information Criterion (PCIC) を構築した。また、以下の論文が出版された。

- [1] Tomonari Sei and Keisuke Yano, Minimum information dependence modeling, accepted at Bernoulli (arXiv:2206.06792)
- [2] Akifumi Okuno and Keisuke Yano, A generalization gap estimation for overparameterized models via Langevin functional variance, Journal of Computational and Graphical Statistics, vol. 32, 1287-1295, 2023 (<https://doi.org/10.1080/10618600.2023.2197488>; arXiv:2112.03660)
- [3] Yukito Iba and Keisuke Yano, Posterior Covariance Information Criterion for Weighted Inference, Neural Computation, vol. 35, 1340-1361, 2023 (https://doi.org/10.1162/neco_a_01592, arXiv:2106.13694)
- [4] Akifumi Okuno and Keisuke Yano, Dependence of variance on covariate design in nonparametric link regression, Statistics and Probability Letters, vol. 193, 2023, 109716 (<https://doi.org/10.1016/j.spl.2022.109716>; arXiv:2012.13106)
- [5] Keisuke Yano and Masayuki Kano, I1 Trend Filtering-based Detection of Short-term Slow Slip Events: Application to a GNSS Array in Southwest Japan, Journal of Geophysical Research: Solid Earth, vol. 127, 5, e2021JB023258, 2022 (<https://doi.org/10.1029/2021JB023258>; python code)
- [6] Yohta Yamanaka, Sumito Kurata, Keisuke Yano, Fumiyasu Komaki, Takahiro Shiina, Aitaro Kato Structured regularization based local earthquake tomography for the adaptation to velocity discontinuities, Earth, Planets and Space, vol. 74, 43, 2022 (<https://doi.org/10.1186/s40623-022-01600-x>)
- [7] Hidenobu Takahashi, Kazuya Tateiwa, Keisuke Yano, Masayuki Kano, A convolutional neural network-based classification of local earthquakes and tectonic tremors in Sanriku-oki, Japan, using S-net data, Earth, Planets and Space, vol. 73, 186, 2021 (<https://doi.org/10.1186/s40623-021-01524-y>)
- [8] Keisuke Yano, Takahiro Shiina, Sumito Kurata, Aitaro Kato, Fumiyasu Komaki, Shin'ichi Sakai, and Naoshi Hirata, Graph-partitioning based convolutional network for earthquake detection using a seismic array, Journal of Geophysical Research: Solid Earth, vol. 126, e2020JB020269, 2021 (<https://doi.org/10.1029/2020JB020269>) Top downloaded article (2023/3/30) among work published in an issue between 1 Jan. 2021-31 Dec. 2021.
- [9] Edward George, Gourab Mukherjee, and Keisuke Yano, Optimal Shrinkage Estimation of Predictive Densities under α -divergences, Bayesian Analysis, vol.16, 1139-1155, 2021 (<https://doi.org/10.1214/21-BA1264>)
- [10] Keisuke Yano, Ryoya Kaneko, and Fumiyasu Komaki, Minimax Predictive Density for Sparse Count Data, Bernoulli, vol. 27, 1212-1238, 2021 (doi:10.3150/20-BEJ1271 ; python code)
- [11] Kento Nakamura, Keisuke Yano, and Fumiyasu Komaki, Adjacency-based regularization for partially ranked data with non-ignorable missing, Computational Statistics & Data Analysis, vol. 145, 106905, 2020 (doi:10.1016/j.csa.2019.106905)
- [12] Keisuke Yano and Kengo Kato, On frequentist coverage errors of Bayesian credible sets in moderately high dimensions, Bernoulli, vol. 26, pp.616-641, 2020 (arXiv:1803.03450 ; doi: 10.3150/19-BEJ1142)

5. 主な発表論文等

〔雑誌論文〕 計12件（うち査読付論文 11件 / うち国際共著 3件 / うちオープンアクセス 7件）

1. 著者名 Iba Yukito, Yano Keisuke	4. 巻 35
2. 論文標題 Posterior Covariance Information Criterion for Weighted Inference	5. 発行年 2023年
3. 雑誌名 Neural Computation	6. 最初と最後の頁 1340 ~ 1361
掲載論文のDOI (デジタルオブジェクト識別子) 10.1162/neco_a_01592	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 Okuno Akifumi, Yano Keisuke	4. 巻 32
2. 論文標題 A Generalization Gap Estimation for Overparameterized Models via the Langevin Functional Variance	5. 発行年 2023年
3. 雑誌名 Journal of Computational and Graphical Statistics	6. 最初と最後の頁 1287 ~ 1295
掲載論文のDOI (デジタルオブジェクト識別子) 10.1080/10618600.2023.2197488	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 Tomonari Sei and Keisuke Yano	4. 巻 -
2. 論文標題 Minimum information dependence modeling	5. 発行年 2024年
3. 雑誌名 Bernoulli	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) なし	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 Okuno Akifumi, Yano Keisuke	4. 巻 193
2. 論文標題 Dependence of variance on covariate design in nonparametric link regression	5. 発行年 2023年
3. 雑誌名 Statistics & Probability Letters	6. 最初と最後の頁 109716-109716
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.spl.2022.109716	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Iba Yukito, Yano Keisuke	4. 巻 35
2. 論文標題 Posterior Covariance Information Criterion for Weighted Inference	5. 発行年 2023年
3. 雑誌名 Neural Computation	6. 最初と最後の頁 1-22
掲載論文のDOI (デジタルオブジェクト識別子) 10.1162/neco_a_01592	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Okuno Akifumi, Yano Keisuke	4. 巻 1
2. 論文標題 A generalization gap estimation for overparameterized models via the Langevin functional variance	5. 発行年 2023年
3. 雑誌名 Journal of Computational and Graphical Statistics	6. 最初と最後の頁 1-20
掲載論文のDOI (デジタルオブジェクト識別子) 10.1080/10618600.2023.2197488	査読の有無 無
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 George Edward, Mukherjee Gourab, Yano Keisuke	4. 巻 16
2. 論文標題 Optimal Shrinkage Estimation of Predictive Densities Under ϕ -Divergences	5. 発行年 2021年
3. 雑誌名 Bayesian Analysis	6. 最初と最後の頁 1139-1155
掲載論文のDOI (デジタルオブジェクト識別子) 10.1214/21-BA1264	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 Yano Keisuke, Kaneko Ryoya, Komaki Fumiyasu	4. 巻 27
2. 論文標題 Minimax predictive density for sparse count data	5. 発行年 2021年
3. 雑誌名 Bernoulli	6. 最初と最後の頁 1212-1238
掲載論文のDOI (デジタルオブジェクト識別子) 10.3150/20-BEJ1271	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Yano Keisuke, Kaneko Ryoya, Komaki Fumiyasu	4. 巻 27
2. 論文標題 Minimax predictive density for sparse count data	5. 発行年 2021年
3. 雑誌名 Bernoulli	6. 最初と最後の頁 1212-1238
掲載論文のDOI (デジタルオブジェクト識別子) 10.3150/20-BEJ1271	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 George Edward, Mukherjee Gourab, Yano Keisuke	4. 巻 -1
2. 論文標題 Optimal Shrinkage Estimation of Predictive Densities Under ϕ -Divergences	5. 発行年 2021年
3. 雑誌名 Bayesian Analysis	6. 最初と最後の頁 0
掲載論文のDOI (デジタルオブジェクト識別子) 10.1214/21-BA1264	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 Keisuke Yano and Kengo Kato	4. 巻 26
2. 論文標題 On frequentist coverage errors of Bayesian credible sets in moderately high dimensions	5. 発行年 2020年
3. 雑誌名 Bernoulli	6. 最初と最後の頁 616-641
掲載論文のDOI (デジタルオブジェクト識別子) 10.3150/19-BEJ1142	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 該当する

1. 著者名 Kento Nakamura, Keisuke Yano, and Fumiyasu Komaki	4. 巻 145
2. 論文標題 Adjacency-based regularization for partially ranked data with non-ignorable missing	5. 発行年 2020年
3. 雑誌名 Computational Statistics & Data Analysis	6. 最初と最後の頁 106905
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.csda.2019.106905	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計16件（うち招待講演 14件 / うち国際学会 8件）

1. 発表者名 Keisuke Yano
2. 発表標題 A new approach to mixed-domain and higher-order dependence modeling
3. 学会等名 Global Plasma Forum in Aomori (招待講演) (国際学会)
4. 発表年 2023年

1. 発表者名 Keisuke Yano
2. 発表標題 最小情報従属モデルを用いた混合ドメイン多変量解析
3. 学会等名 令和5年度第2回 日本大学生産工学部人工知能リサーチセンター講演会 (招待講演)
4. 発表年 2023年

1. 発表者名 Keisuke Yano
2. 発表標題 重み付き推論における汎化性能推定のための事後共分散型情報量規準
3. 学会等名 大阪大学 数理・データ科学セミナー データ科学セミナーシリーズ (招待講演)
4. 発表年 2023年

1. 発表者名 矢野 恵佑
2. 発表標題 高次元・無限次元モデルにおける予測分布
3. 学会等名 日本統計学会各賞受賞者記念講演 (招待講演)
4. 発表年 2022年

1. 発表者名 矢野 恵佑
2. 発表標題 予測分布論の最近の進展
3. 学会等名 日本数学会 (招待講演)
4. 発表年 2021年

1. 発表者名 Keisuke Yano
2. 発表標題 On estimating generalization gaps via the functional variance in overparameterized models
3. 学会等名 CMStatistics2021 (招待講演)
4. 発表年 2021年

1. 発表者名 矢野 恵佑
2. 発表標題 予測の情報量規準
3. 学会等名 情報計測オンラインセミナー (招待講演)
4. 発表年 2021年

1. 発表者名 矢野恵佑, 伊庭幸人
2. 発表標題 擬ベイズ事後分布に基づく予測評価のための情報量規準
3. 学会等名 2021年度統計関連学会連合大会
4. 発表年 2021年

1. 発表者名 Keisuke Yano
2. 発表標題 Risk-estimation based predictive densities for heteroskedastic hierarchical models
3. 学会等名 ICSA 2019, China (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 Keisuke Yano
2. 発表標題 On estimation and prediction for high-dimensional Poisson models with quasi-zero inflation
3. 学会等名 CMStatistics 2019, UK (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 Keisuke Yano
2. 発表標題 On frequentist coverage errors of Bayesian credible sets in moderately high dimensions
3. 学会等名 European Meeting of Statisticians (EMS 2019), Italy (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 Keisuke Yano
2. 発表標題 The Berry--Esseen type bound for the Bernstein--von Mises theorem in moderately high dimensions
3. 学会等名 EAC-ISBA 2019, Japan (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 Keisuke Yano
2. 発表標題 On the construction of adaptive predictive densities for sparse count data
3. 学会等名 EcoSta 2019, Taiwan (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 Keisuke Yano
2. 発表標題 Adaptive minimax predictive density for sparse Poisson models
3. 学会等名 Banff workshop "New and Evolving Roles of Shrinkage in Large-Scale Prediction and Inference (19w5188)", Canada (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 Keisuke Yano, Takahiro Shiina, Sumito Kurata, Aitaro Kato, Fumiyasu Komaki, Shin'ichi Sakai, Naoshi Hirata
2. 発表標題 Earthquake detection using deep learning for continuous seismic network records
3. 学会等名 StatSei11, Japan (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 矢野恵佑, 加納将行
2. 発表標題 ベイズI1トレンドフィルタリングによるスロースリップ自動検知
3. 学会等名 固体地球科学データ同化に関する研究会
4. 発表年 2019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		
米国	University of Southern California	Data Sciences and Operations	