

開催趣旨: 計算技術と統計手法研究の融和を目指して

奥野彰文<sup>1,2</sup>

<https://okuno.net>

<sup>1</sup>統計数理研究所, <sup>2</sup>理化学研究所

# はじめに

ようこそ立川へ!

## 参加者に関する統計

- ▶ 対面参加：36名（内・28名が懇親会も参加）
    - ▶ 口頭発表（招待講演）：9名
    - ▶ ポスター発表：11名+ $\alpha$
    - ▶ 学生（学部～博士）：11名
  - ▶ オンライン参加：26名
- 
- ▶ 当初想定の2倍の規模となりました。

# ワークショップの趣旨

- ▶ “計算”と“統計”に焦点を当てた研究会です。
- ▶ 統計研究で、「計算機で解決できる問題」は結構多い気がする...
- ▶ 似たような問題をやっている領域も多い気がする...

計算代数

数値解析

統計学

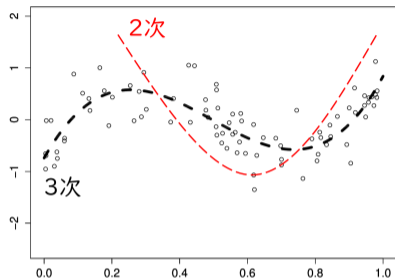
高性能計算

最適化

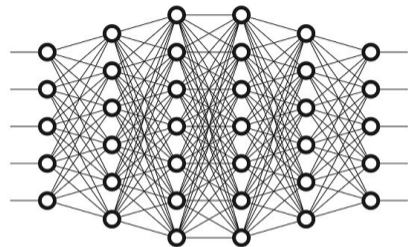
本音) 「統計を応用」ではなく、「統計に応用」をやりたい。

今までやっていたもの

## Okuno and Yano (JCGS2023)



多項式モデルの選択 (AIC)



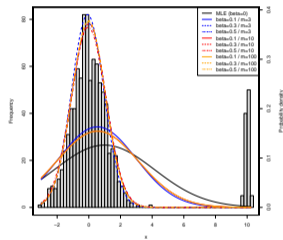
巨大モデル (ニューラルネットなど)

- ▶ 小規模だが特異モデルも選択できるWAIC (Watanabe, 2010).
- ▶ WAICが巨大モデルでも利用できることを証明.
- ▶ ランジュバン動力学による効率的な計算法の提案<sup>1</sup>.

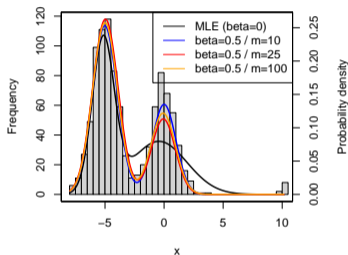
<sup>1</sup>Okuno and Yano (JCGS2023). 日本語での解説 : <https://doi.org/10.51094/jxiv.537>

# Okuno (AISM2024)

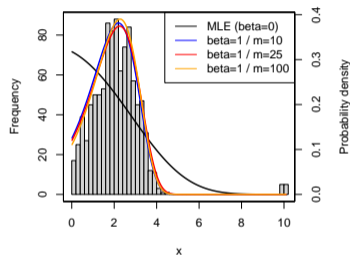
- ▶ 統計学での伝統的なロバスト推定法は、正規分布くらいでしか使えなかった。
- ▶ 任意の分布でロバスト推定する効率的な最適化法を提案し、収束を証明<sup>2</sup>。



(これまで) 正規分布



(提案法) 多峰や歪んだ分布など、任意の分布が推定できる。



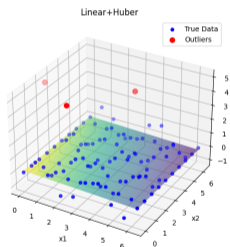
- ▶ 3次元の分布推定では、数値積分と比べて数万倍の高速化<sup>3</sup>。

<sup>2</sup>Okuno (AISM2024), 日本語での解説 : <https://doi.org/10.51094/jxiv.642>

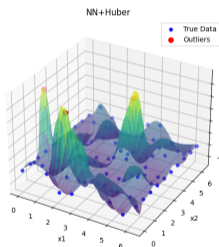
<sup>3</sup>sgdnpパッケージ : <https://github.com/oknakfm/sgdnp>

▶ ニューラルネット + トリム損失

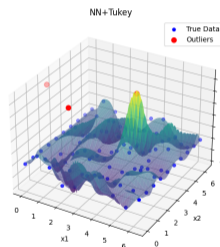
+ 高次変動正則化  $\int \left( \frac{\partial^k f_\theta(x)}{\partial x^k} \right)^2 dx$



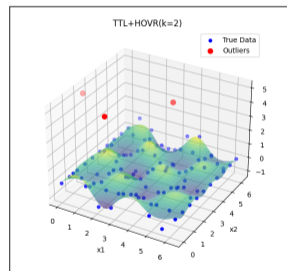
(a) Linear+Huber



(b) NN+Huber



(c) NN+Tukey

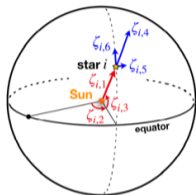


(d) **Ours**: NN+ARTL

▶ PyTorchなど使えば実装は簡単。 確率的な最適化。

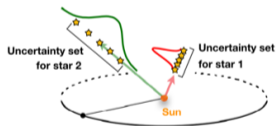
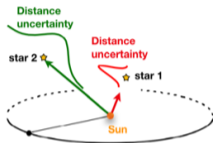
# Hattori, Okuno, and Roederer (Astrophysical Journal 2023)

- ▶ 天体をクラスタリングして同一起源の天体を探す。



Position (relative to the Sun)	Velocity (relative to the Sun)
$\zeta_{i,1}$ Parallax = $1/\text{Distance}$	$\zeta_{i,4} = d(1/\zeta_{i,1})/dt$
$\zeta_{i,2}$ Azimuthal angle	$\zeta_{i,5} = d\zeta_{i,2}/dt$
$\zeta_{i,3}$ Polar angle	$\zeta_{i,6} = d\zeta_{i,3}/dt$

(a) Stellar position and velocity observed from the Sun.

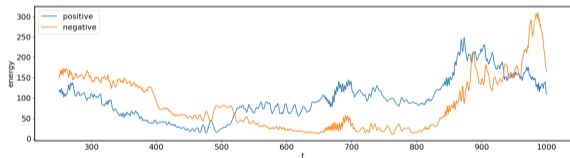
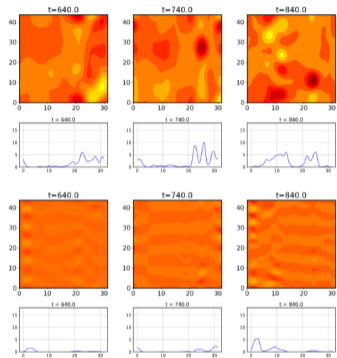


(b) Uncertainty sets for individual stars.

- ▶ 計算量はざっくり  $O(n^2)$  くらいあり, どんどん辛くなる。
- ▶ (既存) 35天体  $\Rightarrow$  (Ours) 161天体  $\Rightarrow$  (次の目標) 数十万天体 (w/寺田さん)



- ▶ 複雑なプラズマ乱流を，人間が理解できる流れに分解する。

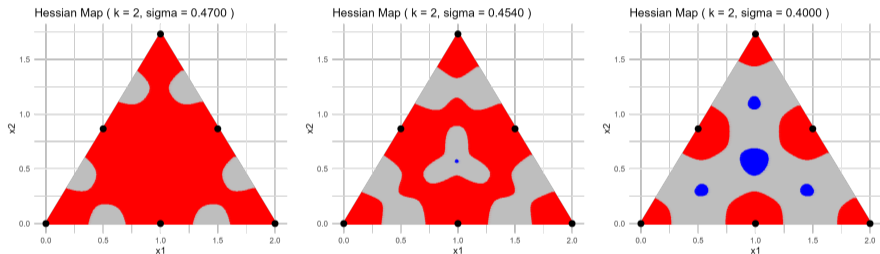


- ▶ (現在) 数値乱流場  $\phi(x, y, t)$  の部分的な離散フーリエ展開  
⇒ (次にやること) 直交テンソル分解? : 時点数 $\times$ 2次元 (or 3次元) の解像度

最近かんがえているもの

## Okuno and Kabata (ongoing)

- ▶ 混合正規分布の峰の検出. 様々なパラメータで全通りやる.
- ▶ 画像処理では昔からある問題らしいが, 実は非常に難しい . . .



通称：ミャクミャクMap

- ▶ 2次元で3つの混合だけでも, パラメータ数15. 一瞬で組み合わせ爆発する.  
(予想に失敗している論文もある)

## Fukasaku and Okuno (ongoing)

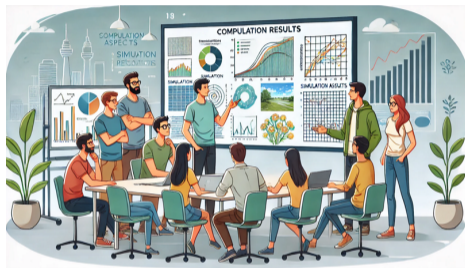
- ▶ 因子分析の解の性質を、数値的でなく、代数的に調べる：  
Fukasaku, Hirose, Kabata, and Teramoto (arXiv:2402.08181)
- ▶ (超小規模なら) ニューラルネットも解けるのでは…？  
Okuno単体では解けないが、Fukasakuさんならコーディングが可能…！

```
f[x_, a_, b_, c_, d_] = a + b * ReLU[c * x + d];  
Solve[Sum[(y[i] - f[x[i], a, b, c, d])^2, {i, 1, n}] == Min, {a, b, c, d}]
```

- ▶ 関連して声をかけている深作さん、高畠さん、矢野さん、松田さん、etc...

# というわけで

やりたいことがたくさんある。



GPT (DALL-E) が書いた謎のイラスト

- ▶ ChatGPTのおかげで、コーディングスキルの不足がどうにかなりそう。
- ▶ 切り口を変えていきたい。興味ある人はぜひお声がけください！

## 招待講演者9名

統計学・最適化・数値解析・計算代数・その他（特異点）から呼びました。

- ▶ 寺田吉壺（大阪大学）
- ▶ 相島健助（法政大学）
- ▶ 柳下翔太郎（統計数理研究所）
- ▶ 廣瀬慧（九州大学）
- ▶ 深作亮也（九州大学）
- ▶ 今倉暁（筑波大学）
- ▶ 佐藤峻（東京大学）
- ▶ 加葉田雄太郎（長崎大学）
- ▶ 松田孟留（東京大学）