

九州シンクロトン光研究センター
シンポジウム「ベイズ計測～現状と展望～」

2024年10月16日

ベイズ計測による小角散乱法の 試料モデル選択とパラメータ推定

東京大学大学院 新領域創成科学研究科 複雑理工学専攻

岡田研究室 博士課程1年

林 悠偉

共同研究者(五十音順, 敬称略)



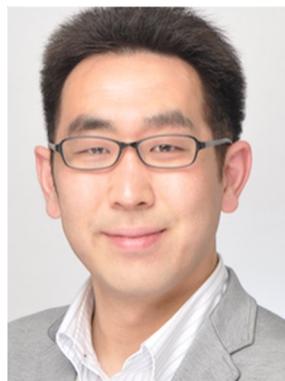
岡田真人
東京大学



片上舜
東京大学



桑本滋生
SPring-8/JASRI



永田賢二
物材研(NIMS)



水牧仁一郎
熊本大学

目次

1. 自己紹介
2. 小角散乱のベイズ計測
 - 研究背景
 - 定式化
 - 事後確率推定：試料パラメータ推定
 - ベイズモデル選択：散乱強度モデルの選択
 - まとめ
3. ベイズ計測の社会実装

目次

1. 自己紹介

2. 小角散乱のベイズ計測

- 研究背景
- 定式化
- 事後確率推定：試料パラメータ推定
- ベイズモデル選択：散乱強度モデルの選択
- まとめ

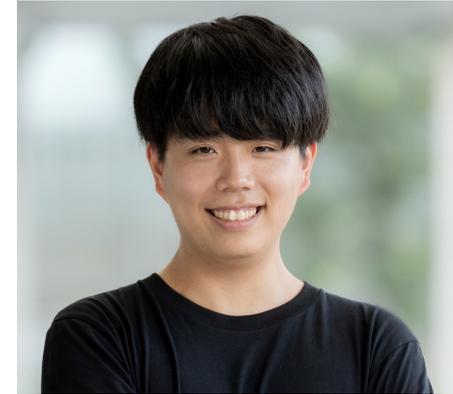
3. ベイズ計測の社会実装

自己紹介

名前：林 悠偉 (はやし ゆい)

所属：東京大学大学院 新領域創成科学研究科

研究領域：小角散乱法のベイズ計測



学術論文(抜粋)

- Hayashi, Y., Katakami, S., Kuwamoto, S., Nagata, K., Mizumaki, M., & Okada, M. (2023). Journal of the Physical Society of Japan, 92(9), 094002.
- Hayashi, Y., Katakami, S., Kuwamoto, S., Nagata, K., Mizumaki, M., & Okada, M. (2024). Journal of Applied Crystallography, 57(4), 955-965.

招待講演

- 林悠偉, 片上舜, 桑本滋生, 永田賢二, 水牧仁一郎, 岡田真人, 「小角散乱法の計測データを用いた試料パラメータのベイズ推定」, 『第88回SPring-8先端利用技術ワークショップ/第4回データ駆動科学によるデータ解析高度化～ベイズ計測～』, SACLA実験研究棟大会議室 及びwebハイブリッド開催, 2023年 3月.
- 林悠偉, 『小角散乱データに基づく試料パラメータとモデルのベイズ推論』, 東京大学物性研究所 中性子セミナー, 機能物性セミナー, 東京大学物性研究所 6階大講義室及びWeb ハイブリッドウェブイベント, 2024年4月.

研究プロジェクト

- 2024/04～2027/03：国立研究開発法人科学技術振興機構(JST)の国家戦略分野の若手研究者及び博士後期課程学生の育成事業(BOOST)事業が支援する、東京大学「次世代知能社会を先導する高度AI人材育成(BOOST NAIS)」プロジェクト.

目次

1. 自己紹介

2. 小角散乱のベイズ計測

- 研究背景
- 定式化
- 事後確率推定：試料パラメータ推定
- ベイズモデル選択：散乱強度モデルの選択
- まとめ

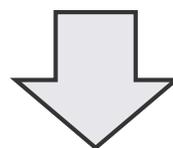
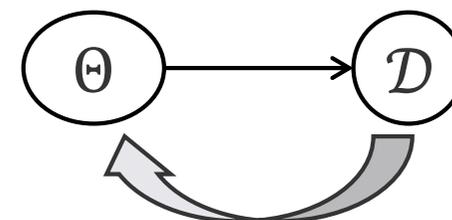
3. ベイズ計測の社会実装

自然科学的視点からのベイズ推論

- データ : \mathcal{D}
- モデル・パラメータ : Θ

$$p(\mathcal{D}, \Theta) = p(\mathcal{D}|\Theta)p(\Theta) = \underline{p(\Theta|\mathcal{D})p(\mathcal{D})}$$

生成(因果律)



$$\text{ベイズの定理} : p(\Theta|\mathcal{D}) = \frac{p(\mathcal{D}|\Theta)p(\Theta)}{p(\mathcal{D})}$$

- $p(\Theta|\mathcal{D})$: 事後確率. データが与えられたもとでの, パラメータの確率.
- $p(\Theta)$: 事前確率. 科学的知見などをもとに, あらかじめ設定する.

ベイズ計測の定義

ベイズ計測は、計測科学に必要なベイズ推論の以下の3つの枠組みである。

■ 物理パラメータの事後確率推定

従来の最小二乗法とは異なり物理パラメータを誤差評価付きで推定できる。評価関数の局所解を避けることができる。

■ モデル選択

実験データを説明する複数の数理モデルの候補の内、どれが適切かを数理的に選択できる。

■ 複数データの統合

比熱と磁化率のように一つの物質を異なる実験で計測した場合のデータを数理的に統合し解析する。

小角散乱法におけるベイズ計測

■物理パラメータの事後確率推定

- 大域的最適解の推定, 推定結果の信頼度を評価可能

■モデル選択

- データに対する数理モデルの妥当性を定量化
- 適切なモデル化によって精度改善

■複数の計測データを統合解析 解析精度の向上の可能性

目次

1. 自己紹介

2. 小角散乱のベイズ計測

- 研究背景

- 定式化

- 事後確率推定：試料パラメータ推定

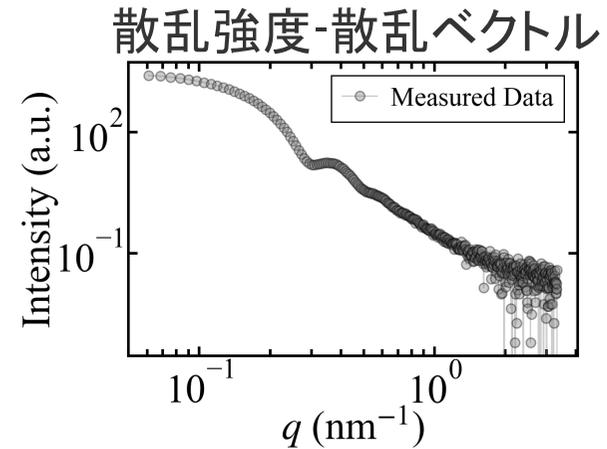
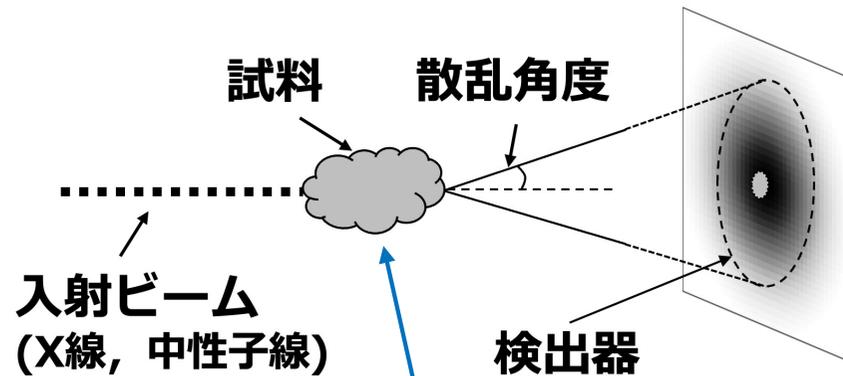
- ベイズモデル選択：散乱強度モデルの選択

- まとめ

3. ベイズ計測の社会実装

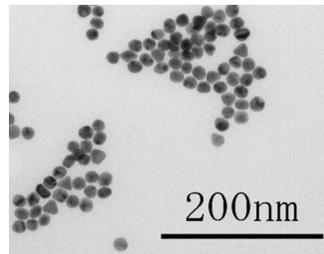
小角散乱法(SAS)

試料にビームを照射し、小さい角度の散乱強度からナノ構造解析を行う。



不均一系試料

- 高分子溶液
- タンパク質
- コロイド粒子
- 金属, セラミック
- 繊維 ...



出典: [1]

データ解析

1. 散乱強度モデルを設定

$$\text{球粒子: } I(q, \Theta) = \left(\frac{\phi}{V} F(q, R_M)^2 + b \right) \times t$$

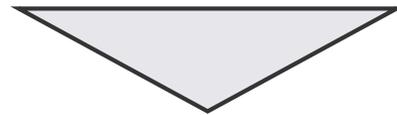
2. モデルパラメータ推定

粒径 $R_M \rightarrow 10 \text{ nm}$

[1] "X線小角散乱法 (SAXS) による液体中に分散したナノ粒子の構造解析", NISSAN ARC,
URL: <https://www.nissan-arc.co.jp/services/p088/>

SASデータ解析の従来法の課題

- 試料パラメータ推定に勾配法などを用いているために、局所解が得られることが多い。
→フィットが良くなるまで、何度もパラメータの初期値を変えて対処している。
- 解析に用いる散乱強度モデルを、専門家の経験則を基に選択している。
→深い知見が必要。属人的な結果が生まれることもある。



本研究の目的

データドリブンな以上2つに対応する解析方法を構築する。

目次

1. 自己紹介

2. 小角散乱のベイズ計測

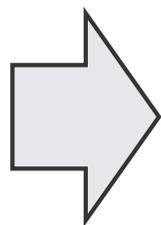
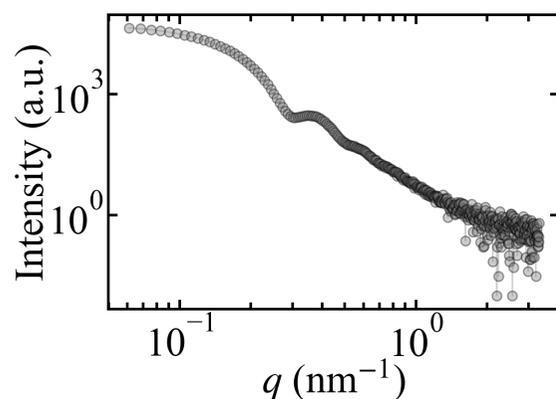
- 研究背景
- **定式化**
- 事後確率推定：試料パラメータ推定
- ベイズモデル選択：散乱強度モデルの選択
- まとめ

3. ベイズ計測の社会実装

小角散乱法におけるベイズ計測

散乱強度モデルの選択と試料パラメータ推定に適用する。

散乱強度-散乱ベクトル : D



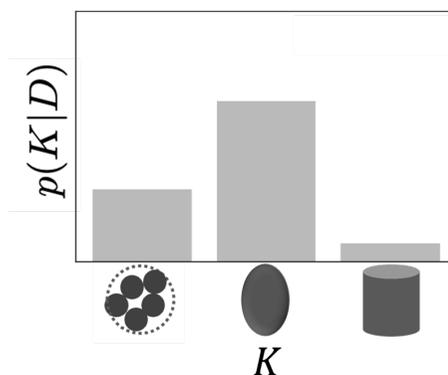
目的

データから試料構造を解析する。

モデル選択

$$p(K|D) = \frac{\int p(D, \Theta, K) d\Theta}{\sum_K \int p(D, \Theta, K) d\Theta}$$

試料構造を表す
散乱強度モデルの選択

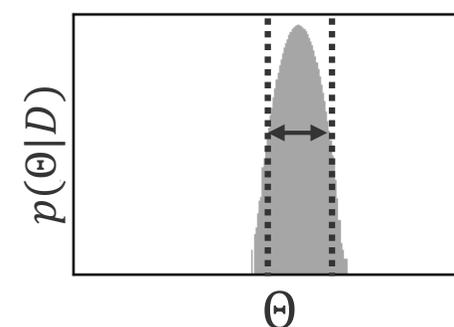


データ駆動な
散乱強度の選択

パラメータ推定

$$p(\Theta|D) = \frac{p(D|\Theta)p(\Theta)}{\int p(D, \Theta) d\Theta}$$

試料の大きさや密度に
関するパラメータを推定



大域解の推定と
信頼度を定量評価

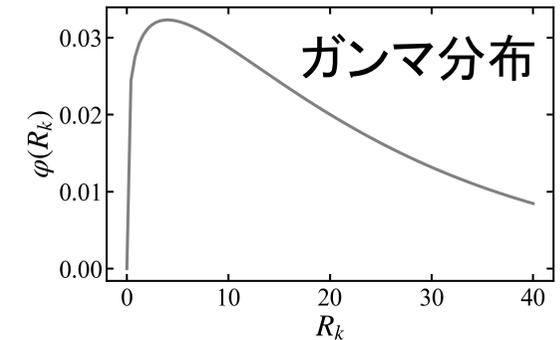
小角散乱法のベイズ計測 - 定式化

■ データ y_i の確率分布 : $p(y|q, \Theta, K) = \frac{I_K(q, \Theta)^y \exp(-I_K(q, \Theta))}{y!}$

(∵ 散乱強度は, 光子のカウントデータとして計測)

散乱強度の計測値 : y_i
散乱ベクトル : q_i
試料モデル : $I(q_i; \Theta)$

■ 事前分布 $\varphi(\Theta)$: 事前知識を基に設定



■ パラメータ Θ の事後確率 : $p(\Theta|\mathcal{D}) = \frac{p(\mathcal{D}|\Theta)\varphi(\Theta)}{\int p(\mathcal{D}, \Theta)d\Theta}$
 $\propto \exp(-NE(\Theta))\varphi(\Theta)$

■ モデル K の事後確率 : $p(K|\mathcal{D}) = \frac{\int p(\mathcal{D}, \Theta, K)d\Theta}{\sum_K \int p(\mathcal{D}, \Theta, K)d\Theta}$
 $= \frac{\exp(-F(K))\varphi(K)}{\sum_K \exp(-F(K))\varphi(K)}$

目次

1. 自己紹介

2. 小角散乱のベイズ計測

- 研究背景
- 定式化
- 事後確率推定：試料パラメータ推定
- ベイズモデル選択：散乱強度モデルの選択
- まとめ

3. ベイズ計測の社会実装

試料パラメータ推定

従来法

勾配法や準ニュートン法によるパラメータフィッティング。

課題

- 局所解にトラップされる
- 結果の信頼度が分からない

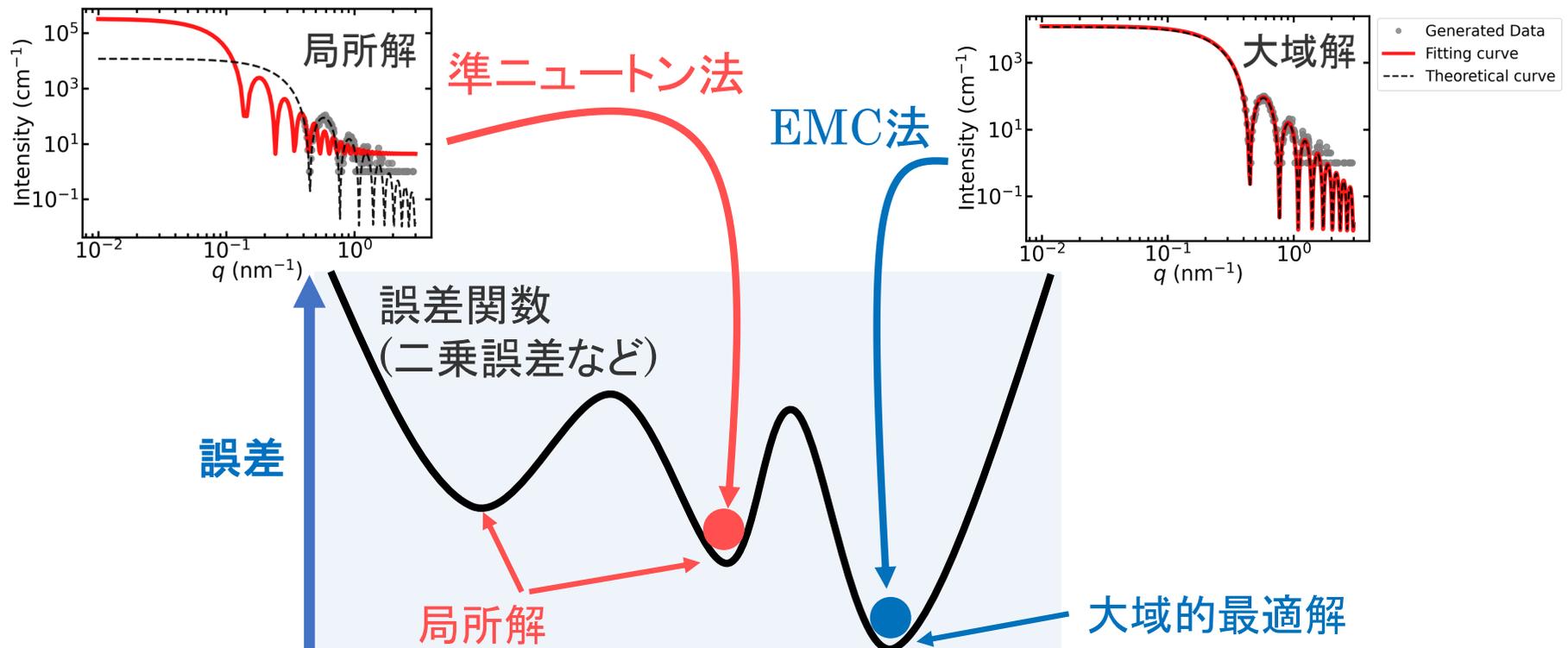
ベイズ計測

交換モンテカルロ(EMC)法[2]により事後分布をサンプリング。

課題の解決

- 大域的最適解の推定
- 事後分布から信頼度評価

Yui Hayashi et al., J. Phys. Soc. Jpn. 92, 094002 (2023).



交換モンテカルロ(EMC)法

メトロポリス法

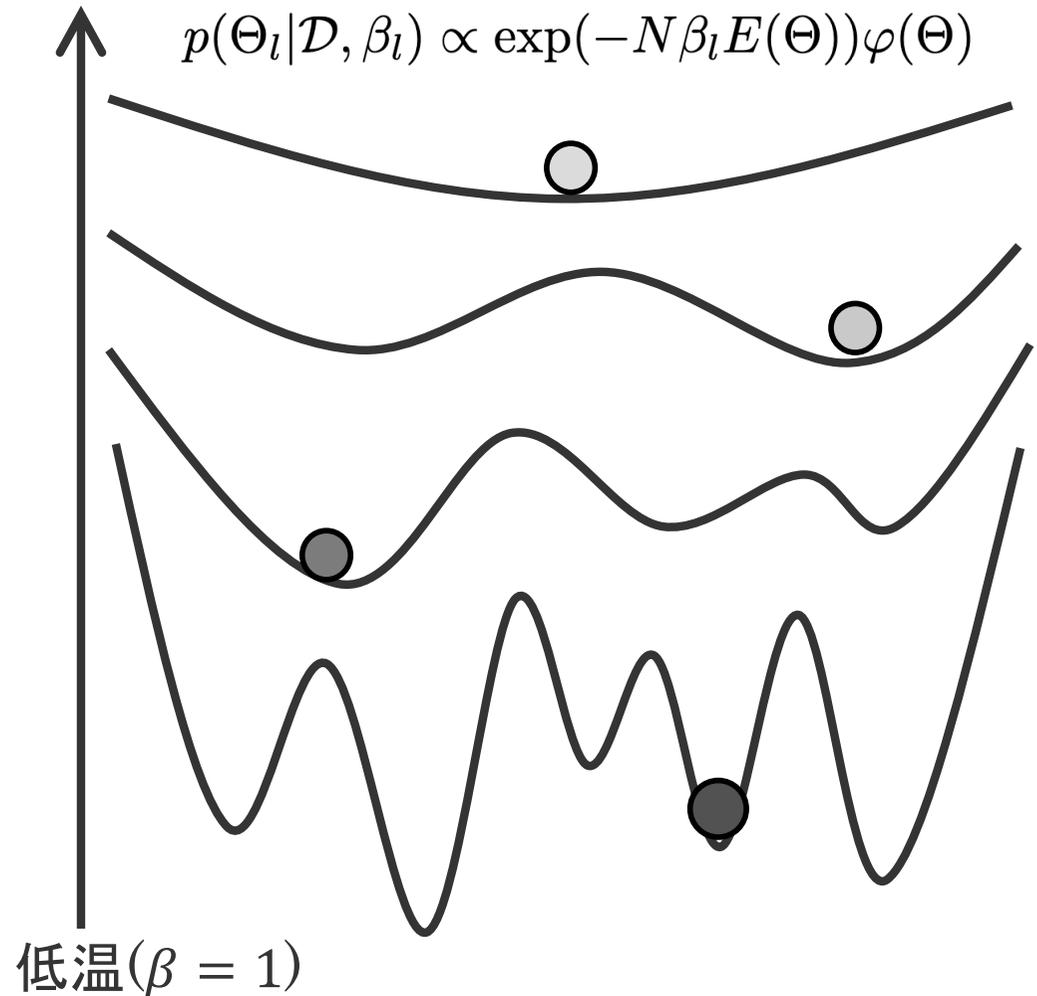
$$p(\Theta|\mathcal{D}) \propto \exp(-NE(\Theta))\varphi(\Theta)$$



EMC法

高温($\beta = 0$)

$$p(\Theta_l|\mathcal{D}, \beta_l) \propto \exp(-N\beta_l E(\Theta))\varphi(\Theta)$$

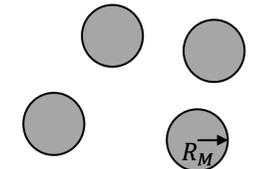


パラメータ推定の数値実験 - 設定

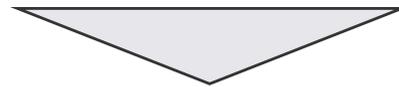
単分散球試料のパラメータを推定する.

$$\text{散乱強度モデル} : I_M(q; \Theta) = \left[\frac{\phi}{V} \left\{ \frac{3\Delta\rho V [\sin(qR_M) - qR_M \cos(qR_M)]}{(qR_M)^3} \right\}^2 + b \right] \times t$$

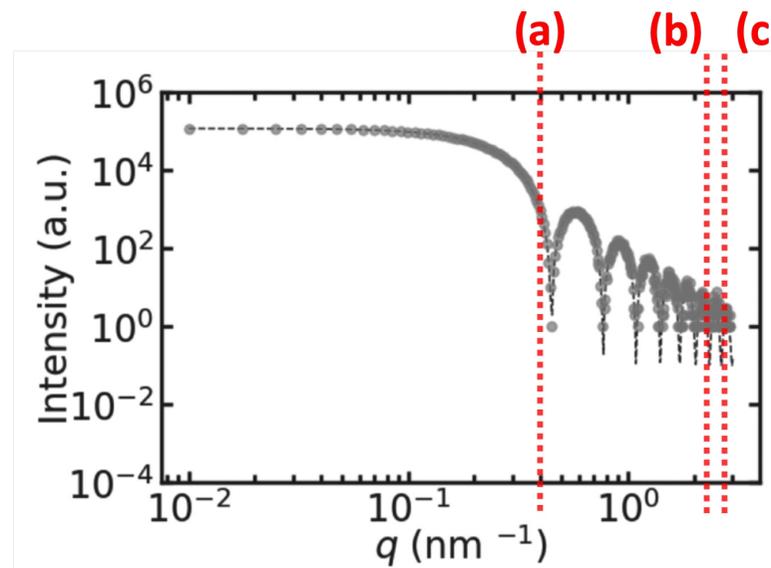
パラメータ: $\Theta = \{R_M, b, t\}$ (粒径: R_M , バックグラウンド: b , 計測時間: t)



例: シリカ粒子など
単分散球



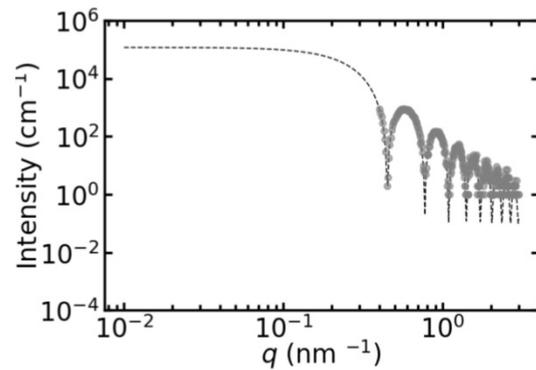
角度領域の下限値の異なる3種類の人エデータ実験を行う.



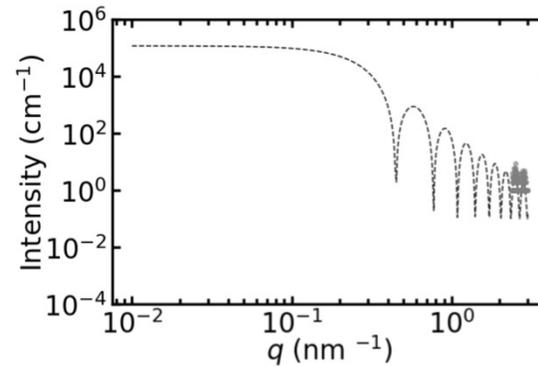
パラメータ推定の数値実験 - 結果

q_{\min} : 散乱ベクトル q の下限

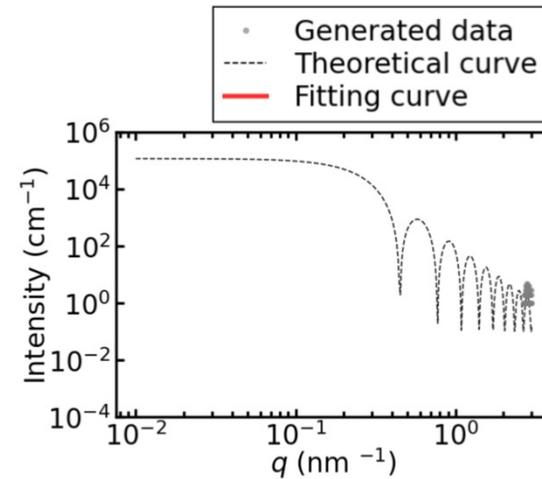
人工データ



(a) $q_{\min} = 0.4 \text{ nm}^{-1}$

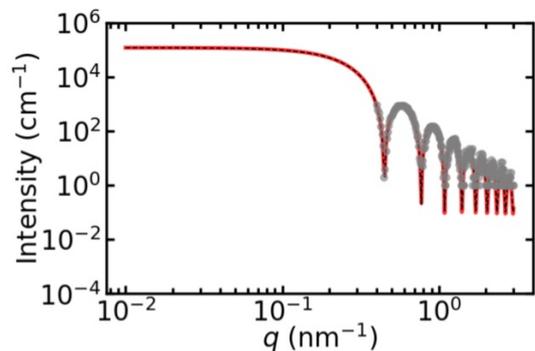


(b) $q_{\min} = 2.35 \text{ nm}^{-1}$

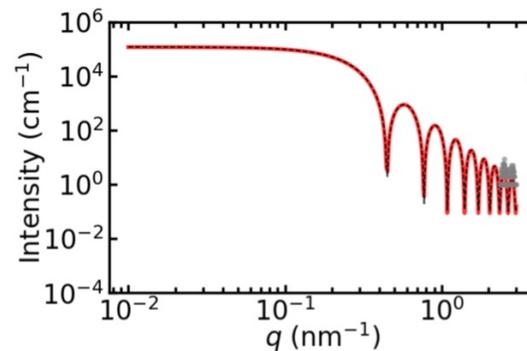


(c) $q_{\min} = 2.65 \text{ nm}^{-1}$

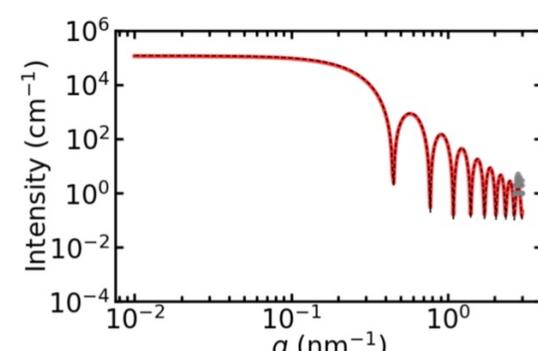
フィッティング



(a) $q_{\min} = 0.4 \text{ nm}^{-1}$



(b) $q_{\min} = 2.35 \text{ nm}^{-1}$

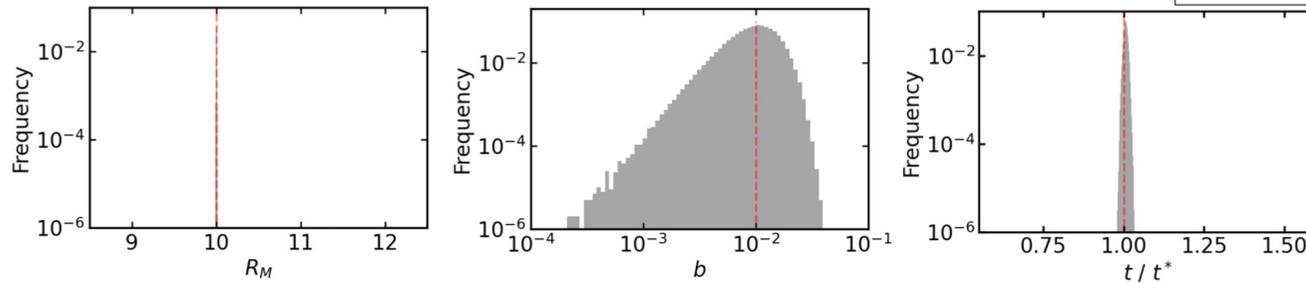


(c) $q_{\min} = 2.65 \text{ nm}^{-1}$

パラメータの事後分布

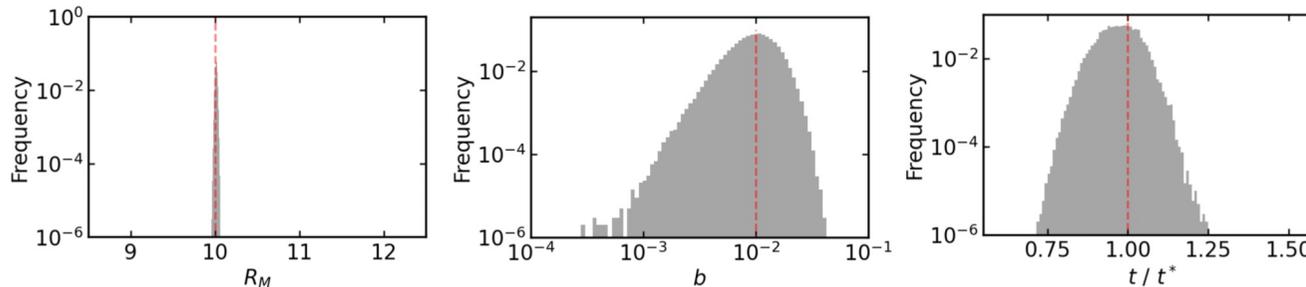
角度領域
 広
 ↓
 狭

(a) $q_{\min} = 0.4$

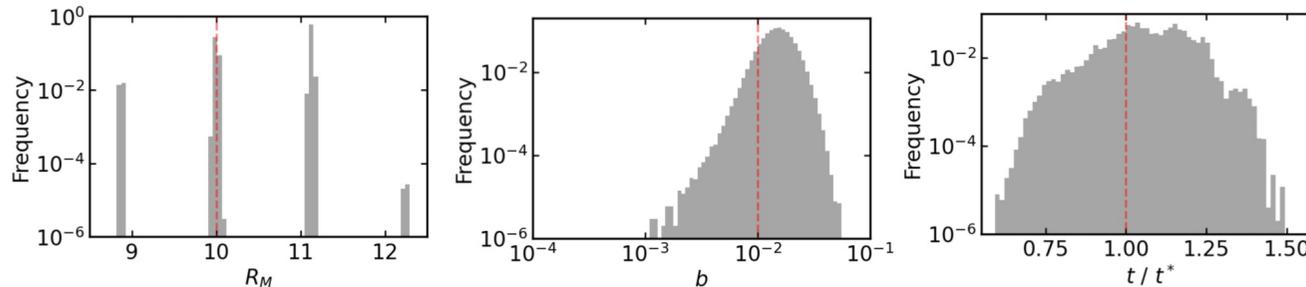


t の真値： t^*

(b) $q_{\min} = 2.35$



(c) $q_{\min} = 2.65$



(c)のデータに対する R_M の事後分布に複数のピークが見られる。

→ $q_{\min} = 2.35 - 2.65 \text{ nm}^{-1}$ で推定限界

目次

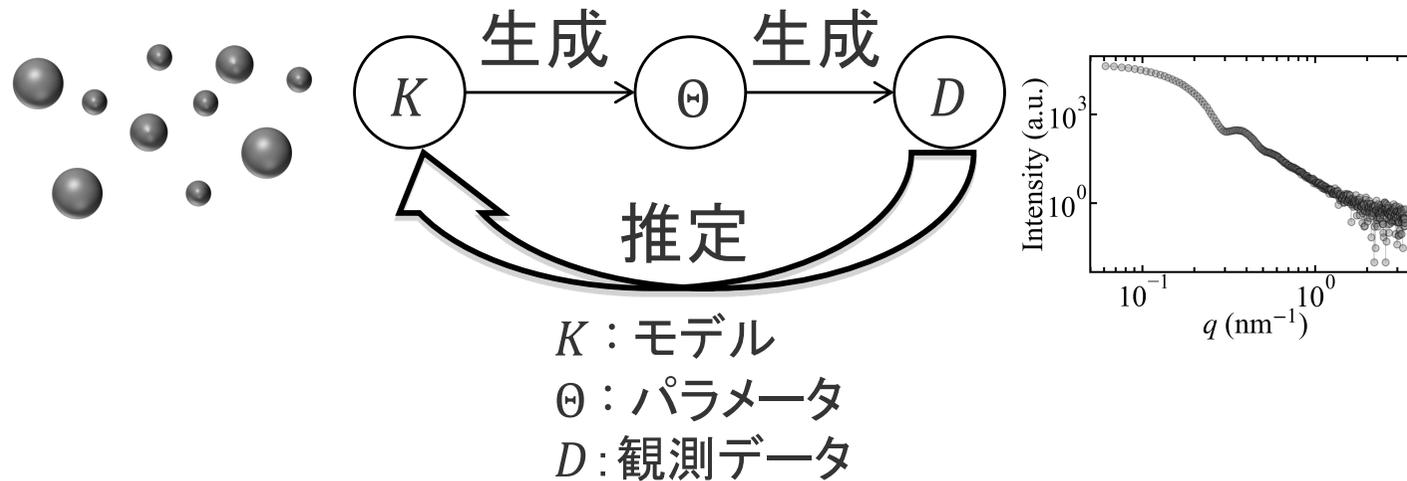
1. 自己紹介

2. 小角散乱のベイズ計測

- 研究背景
- 定式化
- 事後確率推定：試料パラメータ推定
- **ベイズモデル選択：散乱強度モデルの選択**
- まとめ

3. ベイズ計測の社会実装

数理モデルの事後確率の計算



$$\begin{aligned} \text{モデル } K \text{ の事後確率 : } p(K|D) &= \frac{\int p(D, \Theta, K) d\Theta}{\sum_K \int p(D, \Theta, K) d\Theta} \\ &= \frac{\exp(-F(K)) \varphi(K)}{\sum_K \exp(-F(K)) \varphi(K)} \end{aligned}$$

$$\text{自由エネルギー : } F(K) = -\log \int p(D, \Theta|K) d\Theta$$

$F(K)$ をEMC法で数値的に計算する.

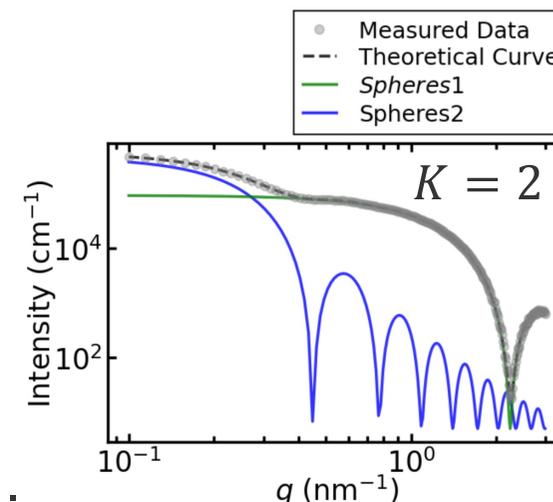
散乱強度モデル選択の数値実験 – 設定

2成分の単分散球が混合しているデータから、成分数 K を選択する.

K 成分の単分散球モデル:

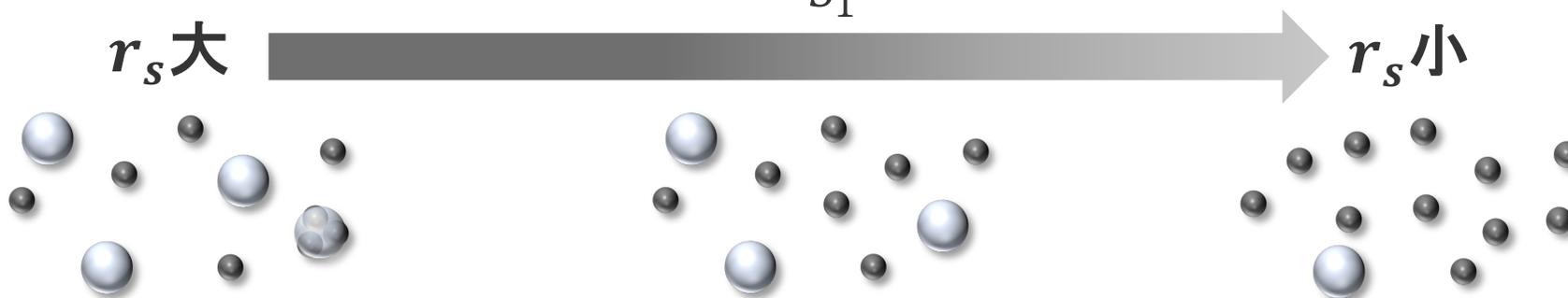
$$I_K(q, \Theta) = \sum_{k=1}^K S_k V_k \left(\frac{(\sin(qR_k) - qR_k \cos(qR_k))}{(qR_k)^3} \right)^2 + B$$

$$\Theta = \{ \{R_k, S_k\}_{k=1}^K, B \}$$



混合比 r_s を変化させて6種のデータを生成した.

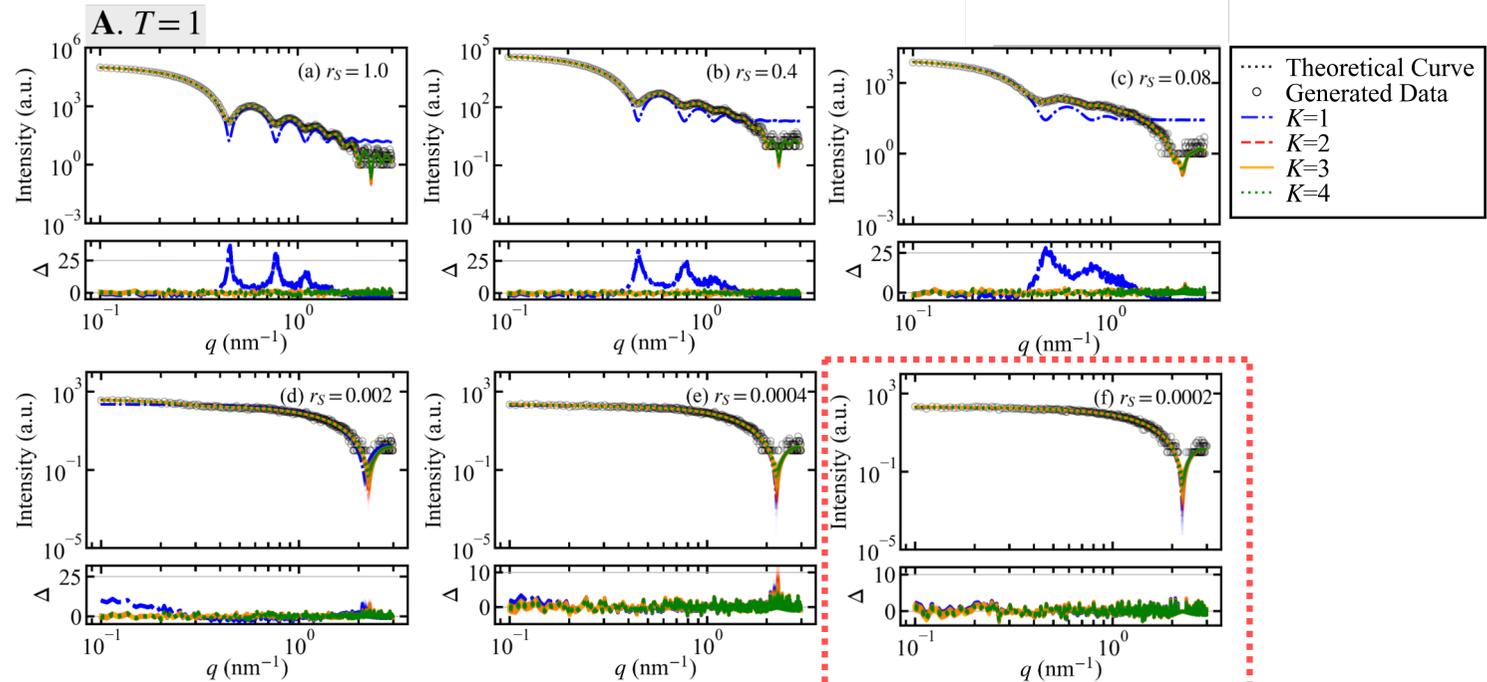
$$r_s = \frac{S_2}{S_1}$$



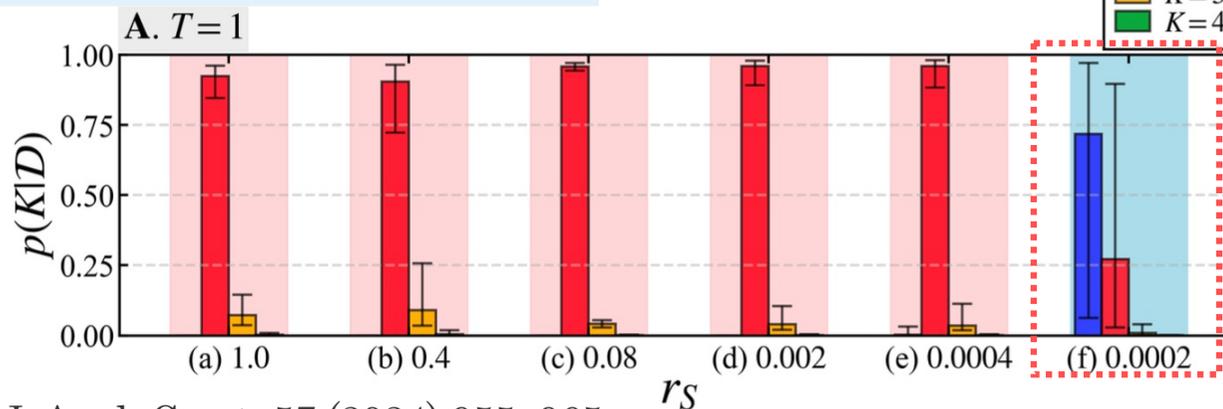
$K = 1, 2, 3, 4$ に対して事後確率 $p(K|D)$ を計算し比較する.

モデル選択の数値実験 - 結果(1/2)

データと
フィッティング



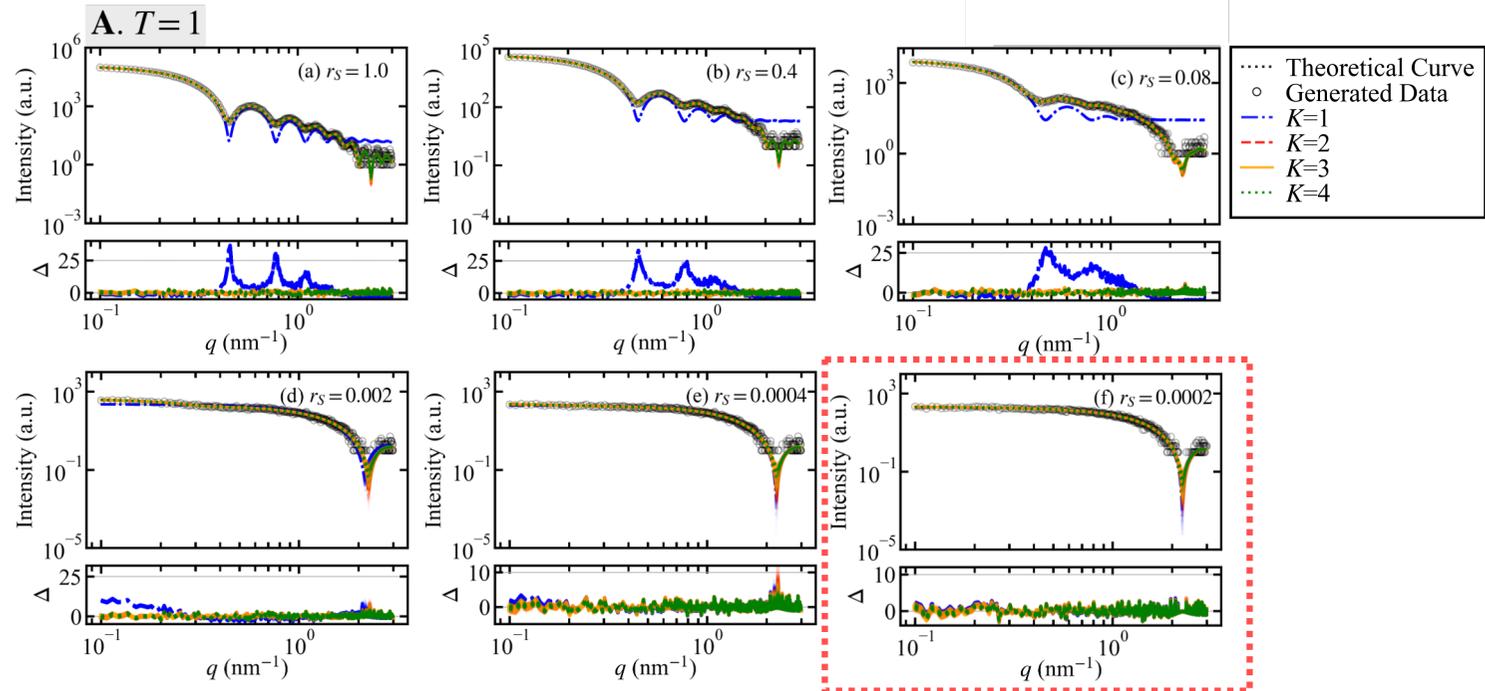
データ生成時の乱数シードを変えて10回
の実験を行った結果. (真値: $K = 2$)



解析限界

モデル選択の数値実験 - 結果(2/2)

データと
フィッティング



データ生成の乱数シードを変化させた場合

10データ中で, $K = 1 \sim 4$
のモデルが支持された回数

$r_s \backslash K$	1	2	3	4
(a) 1.0	0	10	0	0
(b) 0.4	0	10	0	0
(c) 0.08	0	10	0	0
(d) 0.002	0	10	0	0
(e) 0.0004	0	10	0	0
(f) 0.0002	8	2	0	0

解析限界

目次

1. 自己紹介

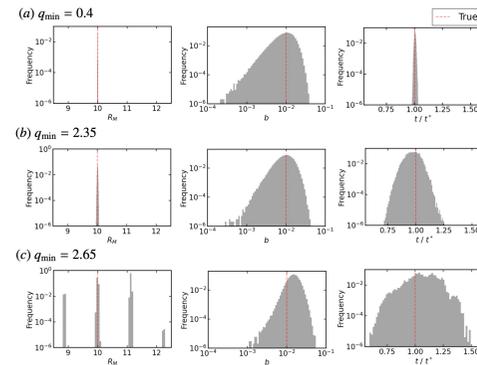
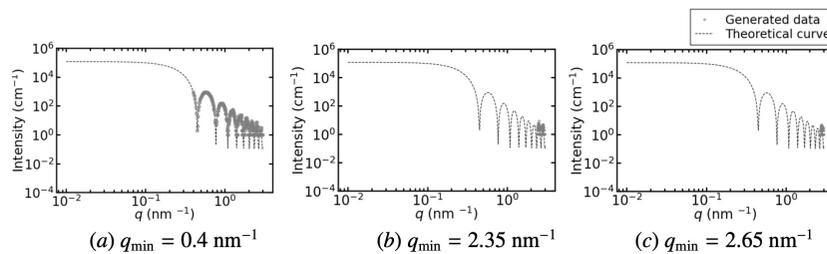
2. 小角散乱のベイズ計測

- 研究背景
- 定式化
- 事後確率推定：試料パラメータ推定
- ベイズモデル選択：散乱強度モデルの選択
- まとめ

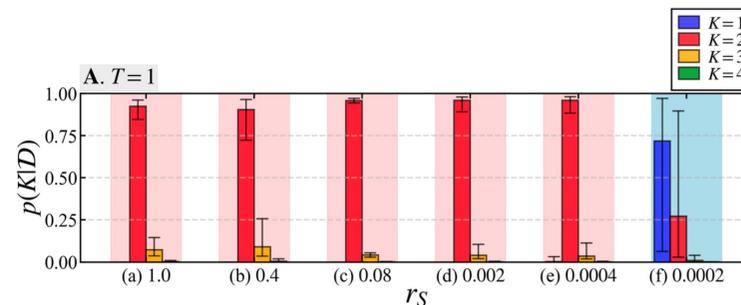
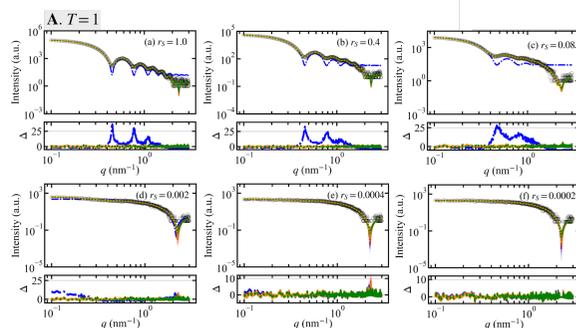
3. ベイズ計測の社会実装

小角散乱法のベイズ計測のまとめ

- 試料パラメータ推定に勾配法などを用いているために、局所解が得られることが多い。
→ベイズ計測により大域的最適解の信頼度付き推定



- 解析に用いる散乱強度モデルを，専門家の経験則を基に選択している。
→事後確率を基にデータ駆動に散乱強度モデル選択



目次

1. 自己紹介
2. 小角散乱のベイズ計測
 - 研究背景
 - 定式化
 - 事後確率推定：試料パラメータ推定
 - ベイズモデル選択：散乱強度モデルの選択
 - まとめ
3. **ベイズ計測の社会実装**

ベイズ計測の社会実装 株式会社 a.s.ist

会社名	株式会社 a.s.ist (読み:アシスト)
所在地	東京都足立区伊興 5-17-24
Webサイト	https://www.a-s-ist.com/
事業内容	<ul style="list-style-type: none">• データ解析コンサルティング• 受託データ解析• ライブラリ, ソフトウェア開発

会社Webサイト



■ 技術顧問：東京大学 岡田真人教授

■ ご連絡先：sales@a-s-ist.com

ベイズ計測を導入したい企業様・研究者の方は、
お気軽にご連絡ください。

自由エネルギーの数値的計算法 レプリカ交換モンテカルロ法を使う

たくさんの温度でのパラメータの事後確率 $p(\Theta|D, K, \beta)$ のサンプルを用いて計算する($\because 0 = \beta_1 < \dots < \beta_L = 1$):

$$\begin{aligned} Z(K, \beta = 1) &= \prod_{l=1}^{L-1} \frac{Z(K, \beta_{l+1})}{Z(K, \beta_l)} \\ &= \prod_{l=1}^{L-1} \langle \exp(-N(\beta_{l+1} - \beta_l)E(\Theta, K)) \rangle_{p(\Theta_l|D, K, \beta_l)} \end{aligned}$$

→ $F(K)$, $p(K|D)$ が計算できる.

$$\text{ベイズ自由エネルギー} : F(K) = -\log Z(K)$$

$$\begin{aligned} \text{モデル}K\text{の事後分布} : p(K|D) &= \frac{\int p(D, \Theta, K) d\Theta}{\sum_K \int p(D, \Theta, K) d\Theta} \\ &= \frac{\exp(-F(K))\varphi(K)}{\sum_K \exp(-F(K))\varphi(K)} \end{aligned}$$

ポアソンノイズ

計測時間を T として、仮想的に n 分割した微小時間 Δt を導入し、 $T = n\Delta t$ とする。

単位時間あたりのカウント数を m とする。

ここで $n \rightarrow \infty$ を考えると、微小時間 Δt には光子が1カウントされるか、されないかのどちらかとなる。

微小時間に光子がカウントされる確率は、 $p = m\Delta t = \frac{mT}{n}$ となる。

有限時間 T の間に j 回カウントされる確率 $P(j)$ は独立な n 回の試行における j 回の実現確率となり、次式を得る。

$$P_n(j) = \frac{n!}{(n-j)! j!} \left(1 - \frac{mT}{n}\right)^{n-j} \left(\frac{mT}{n}\right)^j$$

$n \rightarrow \infty$ を考えて、 $n \gg j$ であることを用いると、次式が得られる。

$$\begin{aligned} P(j) &= \lim_{n \rightarrow \infty} \frac{n!}{(n-j)! j!} e^{-mT} \left(1 - \frac{mT}{n}\right)^{-j} \frac{(mT)^j}{n^j} \\ &= \frac{(mT)^j}{j!} e^{-mT} \lim_{n \rightarrow \infty} \frac{n(n-1)\cdots(n-j+1)}{n^j} \\ &= \frac{(mT)^j}{j!} e^{-mT} \end{aligned}$$

事後分布のサンプリング

交換モンテカルロ法(EMC)のアルゴリズム

1. パラメータ Θ を初期化する.
2. 逆温度 $0 = \beta_1 < \dots < \beta_l < \dots < \beta_L = 1$ を用意する.
3. L 個の各逆温度で, MCMCにより事後分布のサンプリングを行う.

$$\beta_l \text{での事後分布} : p(\Theta_l | \mathcal{D}, \beta_l) \propto \exp(-N\beta_l E(\Theta_l)) p(\Theta_l)$$

4. 任意の時点で隣り合う逆温度での状態を確率 W で交換する.

$$W(\Theta_l | \Theta_{l+1}) = \min \left[1, \frac{p(\Theta_{l+1} | \mathcal{D}, \beta_l) p(\Theta_l | \mathcal{D}, \beta_{l+1})}{p(\Theta_l | \mathcal{D}, \beta_l) p(\Theta_{l+1} | \mathcal{D}, \beta_{l+1})} \right]$$

5. 3~4の手順を繰り返し, 目的の分布 $p(\Theta_L | \mathcal{D})$ を得る.