

2023/3/7

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～」

X線小角散乱プロファイルの ベイズ計測

東京大学大学院 岡田研究室 修士課程1年

林 悠偉

共同研究者(敬称略)

- 片上 舜 東京大学大学院新領域創成科学研究科
- 桑本 滋生 公益財団法人高輝度光科学研究センター(JASRI)
- 永田 賢二 国立研究開発法人物質・材料研究機構(NIMS)
- 水牧 仁一郎 公益財団法人高輝度光科学研究センター(JASRI)
- 岡田 真人 東京大学大学院新領域創成科学研究科

自己紹介

メール : hayashi@mns.k.u-tokyo.ac.jp

- 2022/3
東京理科大学 理工学部 物理学科 卒業
卒業研究の題目：
「ニューラルネットワークを用いた中性子星内部の状態方程式推定」
- 2022/4～
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
研究題目：小角散乱法データのベイズ計測
論文：Bayesian Inference for Small-Angle Scattering Data
(J. Phys. Soc. Jpn. submitted)

アンケート

- スペクトルや画像データからフィッティングを行なっている。
- そのフィッティングの際に、パラメータを手打ちで決めている。最急降下法などを使っているが、うまくいかない。
- S/Nが悪いデータや欠損データをなんとかしたい。
- フィッティング用のモデルが複数あって、事前にどれを使うかを決めておかないといけない。

そのような方は、一度ベイズ計測をお試しく下さい。

目次

- X線小角散乱法の実験概要
- 従来 of 解析方法
- ベイズ推定
- 数値実験
- 今後の展望

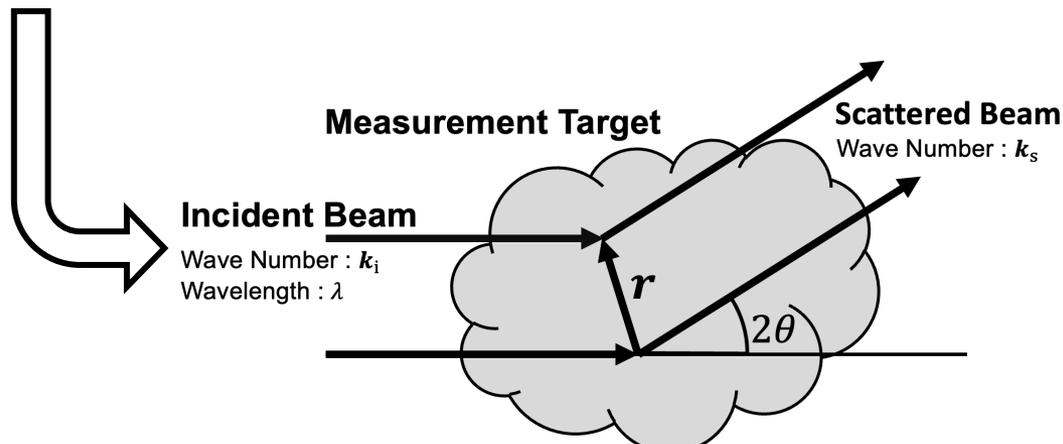
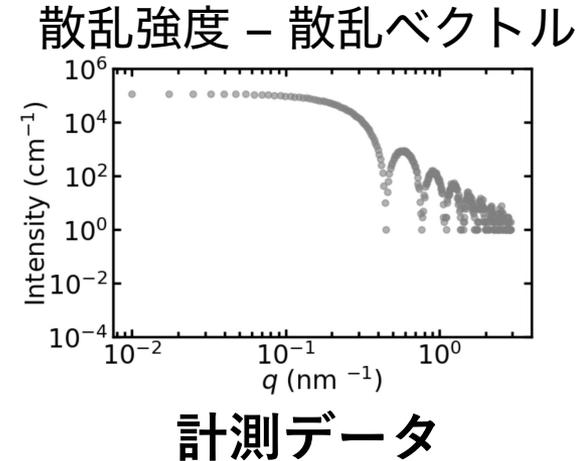
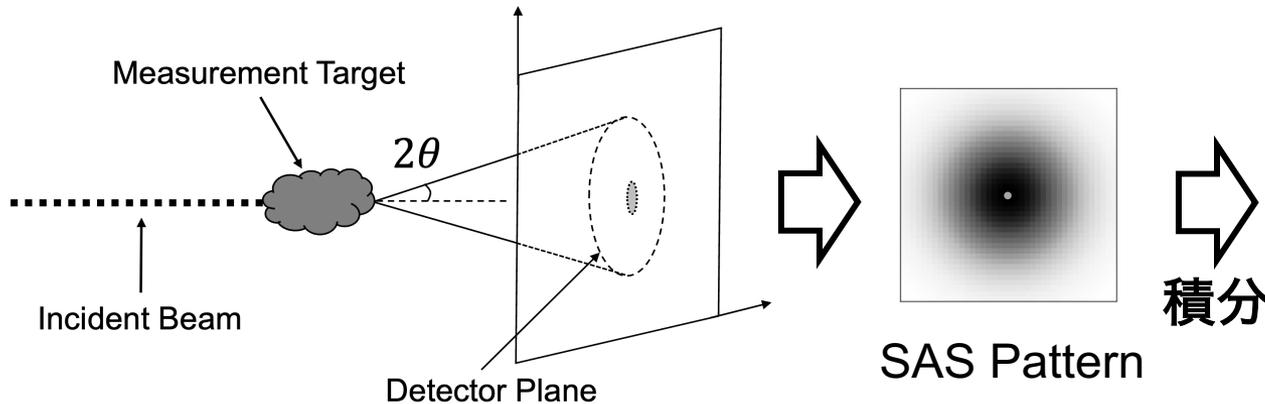
目次

- X線小角散乱法の実験概要
- 従来解析方法
- ベイズ推定
- 数値実験
- 今後の展望

X線小角散乱 (SAXS)

試料にX線を照射して生じる散乱のうち小さい角度領域の散乱強度を調べ、試料の周期構造やナノ粒子の大きさ等のナノスケールの構造情報を非破壊的に解析する。

実験の概略図



Scattering Vector :

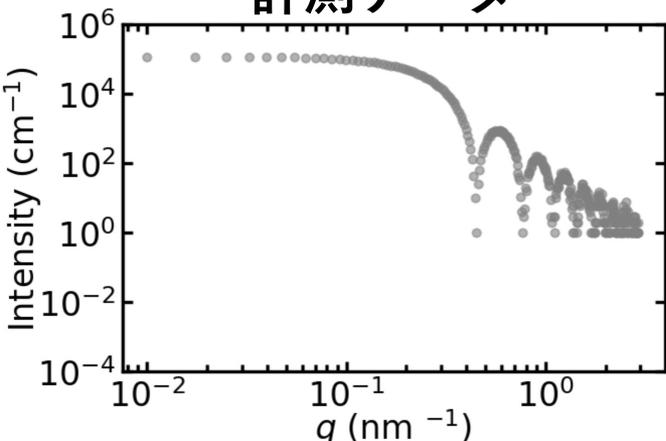
$$q = |\mathbf{q}| = |\mathbf{k}_s - \mathbf{k}_i| = \frac{4\pi\sin\theta}{\lambda}$$

SAXSデータ解析

計測データに対してフィッティングを行い、試料の大きさなどのパラメータを推定する。

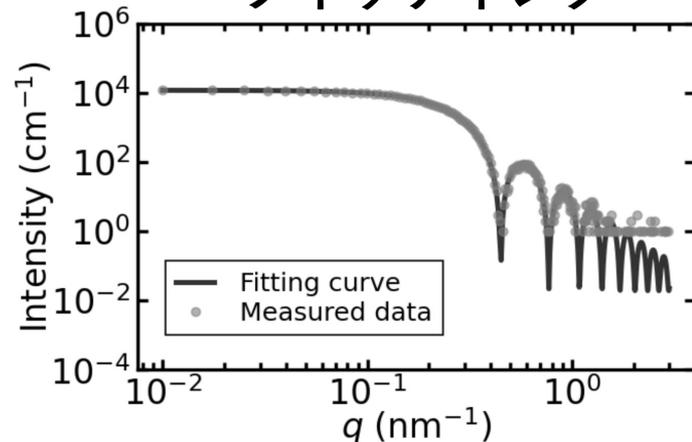
例：単分散球の場合

計測データ



理論式(モデル) →

フィッティング



モデル

散乱強度：

$$I_M(q; \Theta) = \left[\frac{\phi}{V} \left\{ \frac{3\Delta\rho V [\sin(qR_M) - qR_M \cos(qR_M)]}{(qR_M)^3} \right\}^2 + b \right] \times t$$

パラメータ： $\Theta = \{R_M, b, t\}$



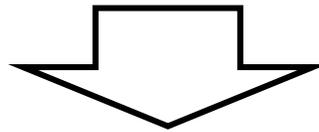
パラメータ Θ を推定

目次

- X線小角散乱法の実験概要
- 従来 of 解析方法
- ベイズ推定
- 数値実験
- 今後の展望

従来のフィッティング方法

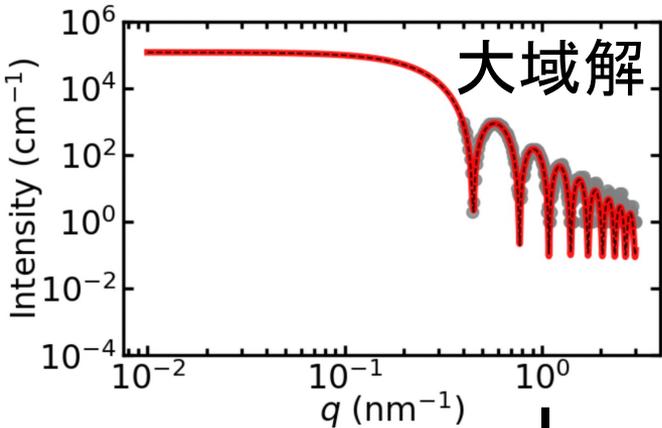
- 非線形最小二乗法(勾配法, Levenberg-Marquardt法)
- ハンドフィッティング



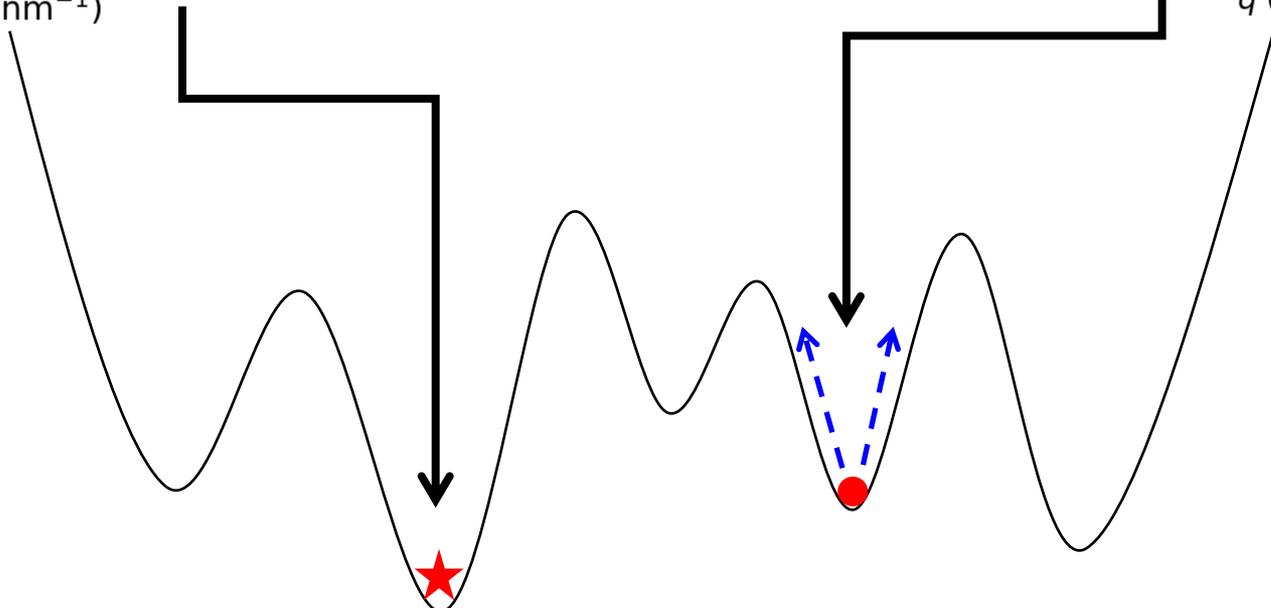
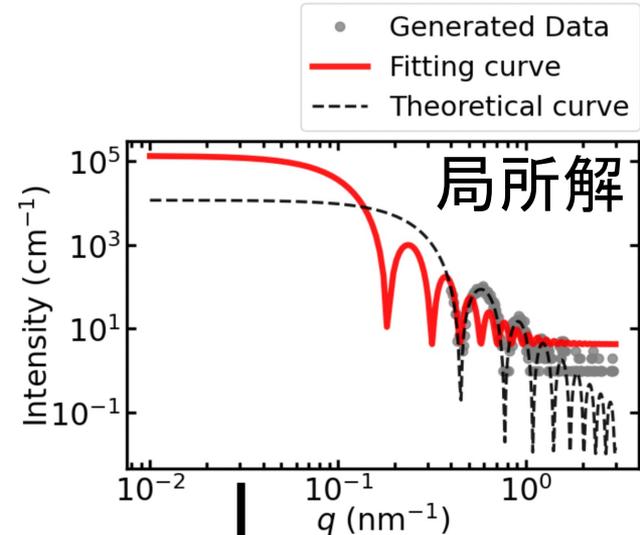
問題点

- 局所解に陥りやすく, 逃れる機構を持たない.
- 推定結果の信頼度を調べるのが難しい.
- 解析過程で用いた事前知識が客観的に分かり難い.

誤差関数は局所解を持つ



＜通常の最適化法＞
e.g., 勾配法, LM法



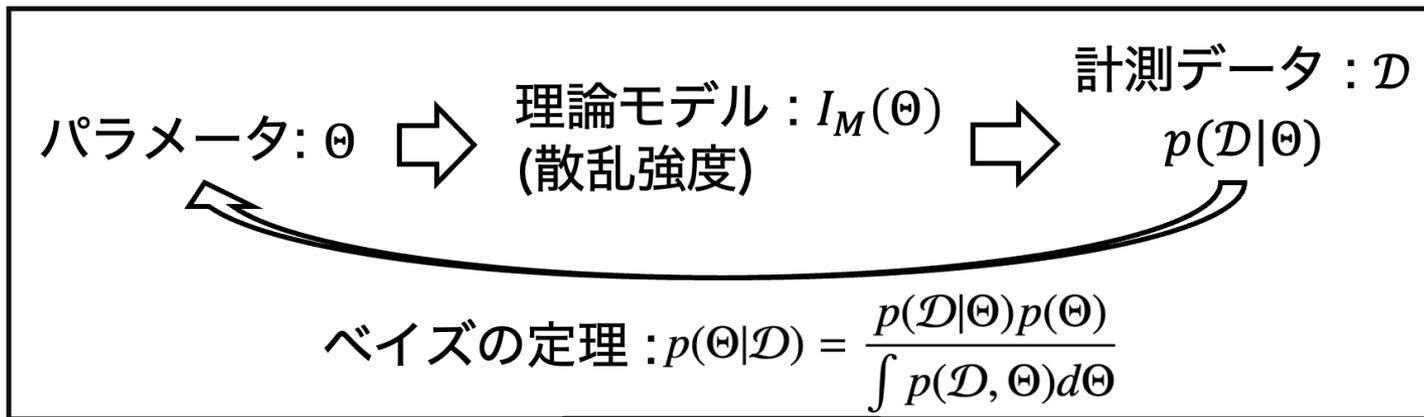
局所解から逃れることが可能で、
推定の信頼度を解釈することが出来る解析手法が必要

目次

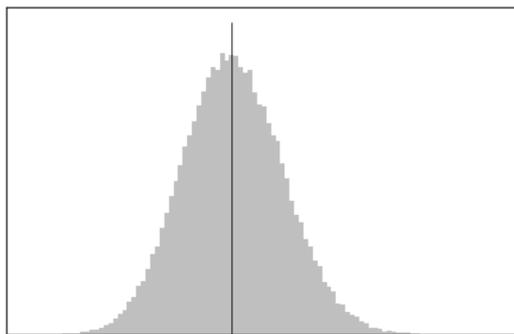
- X線小角散乱法の実験概要
- 従来の解析方法
- **ベイズ推定**
- 数値実験
- 今後の展望

ベイズ推定

計測データからパラメータを事後分布として分布推定する。



ベイズ推定



パラメータの推定結果
(事後分布)

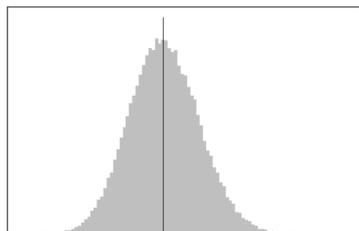


- ピーク形状
- 統計量 (分布の分散, 信用区間, ...)

信頼度を解釈

ベイズ推定の要点

- パラメータのサンプリングに交換モンテカルロ法(EMC)を用いることで局所解から逃れることが可能.
- 分布推定することで形状や統計量から信頼度を解釈することが可能.



パラメータの推定結果
(事後分布)

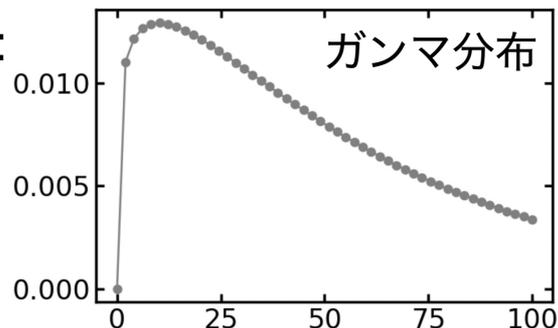


- ピーク形状
- 統計量 (分布の分散, 信用区間, ...)

信頼度を解釈

- 事前知識を事前分布として、陽に用いることが可能.

例：正の領域で定義される θ の事前分布 $p(\theta)$ ：

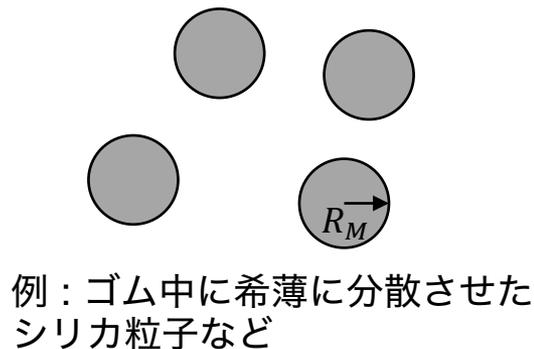


目次

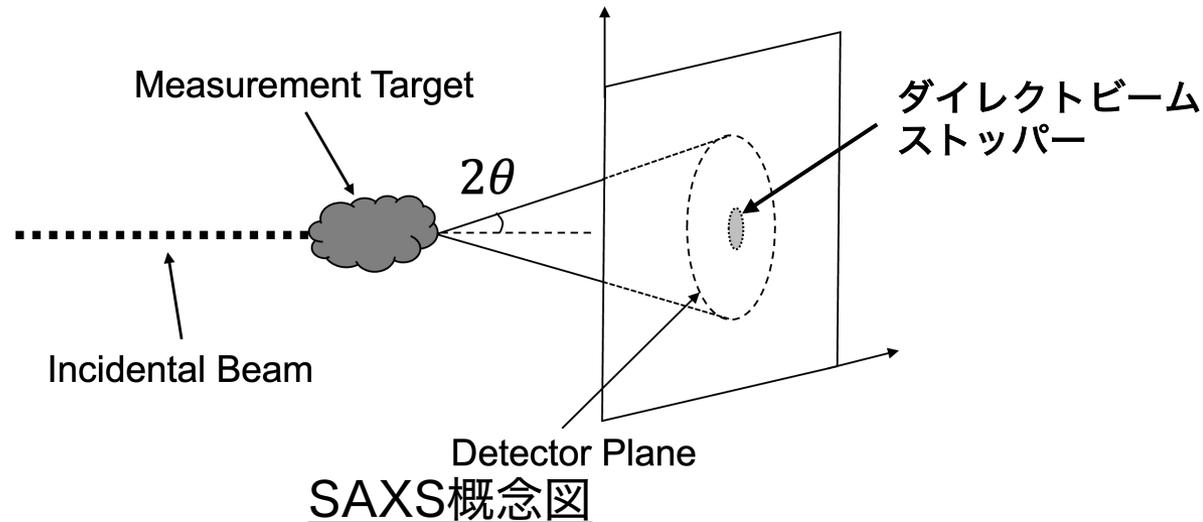
- X線小角散乱法の実験概要
- 従来 of 解析方法
- ベイズ推定
- **数値実験**
- 今後の展望

数値実験：モデルパラメータの推定

単分散球から成る試料の人工データから、パラメータ θ を推定する。



単分散球



目的

データの小角領域の欠損に対してパラメータ推定の限界を調べる。

パラメータ推定の手順

1. 理論モデルとパラメータ θ の事前分布を設定する。
2. 事後分布からパラメータ θ をEMCでサンプリングする。

数値実験 - 設定

理論モデル

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

$$\text{パラメータ} : \Theta = \{R_M, b, t\}$$

R_M : 粒子の半径
 b : バックグラウンドノイズ
 t : 計測時間

コスト関数

$$\text{ポアソンコスト関数} : E(\Theta) = \frac{1}{N} \sum_{i=1}^N \left\{ I(q_i; \Theta) - y_i \log I(q_i; \Theta) + \sum_{j=1}^{y_j} \log j \right\}$$

事前分布

$$\text{パラメータ}\Theta\text{の事前分布} : p(\Theta) = p(R_M)p(b)p(t)$$

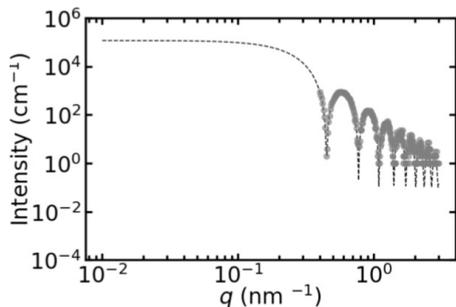
$$\text{粒子の半径}R_M : p(R_M) = \text{Gamma}(1.5, 100)$$

$$\text{バックグラウンドノイズ}b : p(b) = \text{Gamma}(1.8, 1)$$

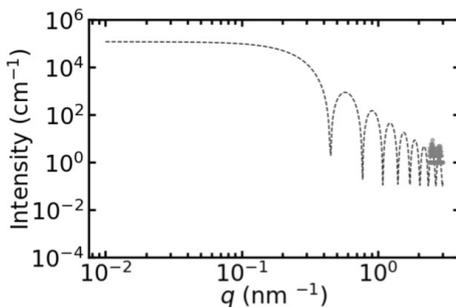
$$\text{計測時間}t : p(t) = \text{Gamma}(1.1, 1)$$

数値実験 - 結果

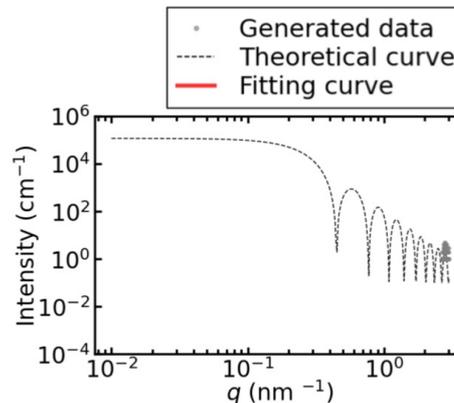
人工データ



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

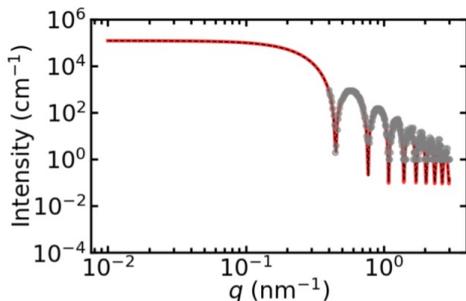


(b) $q_{\min} = 2.35 \text{ nm}^{-1}$

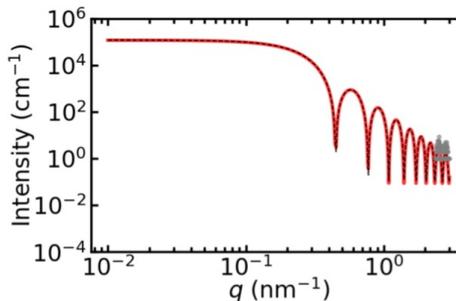


(c) $q_{\min} = 2.65 \text{ nm}^{-1}$

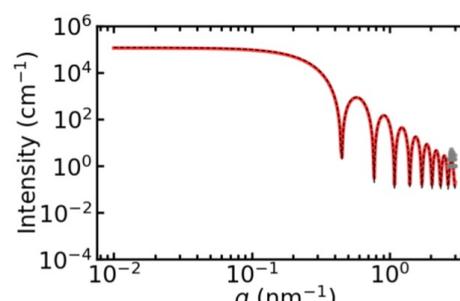
フィッティング結果



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



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



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

MAP推定解

真値	$q_{\min} [\text{nm}^{-1}]$	$R_M [\text{nm}]$	$b [\text{cm}^{-1}\text{sr}^{-1}]$	t
$R_M = 10$	0.4	$10.00^{+0.01}_{-0.01}$	$(8.71^{+15.10}_{-7.01}) \times 10^{-3}$	$10.06^{+0.15}_{-0.16}$
$b = 0.01$	2.35	$10.02^{+0.03}_{-0.03}$	$(8.74^{+15.03}_{-6.15}) \times 10^{-3}$	$9.85^{+1.25}_{-1.55}$
$t = 10$	2.65	$10.01^{+1.18}_{-1.16}$	$(1.60^{+1.49}_{-1.01}) \times 10^{-2}$	$9.51^{+4.05}_{-2.06}$

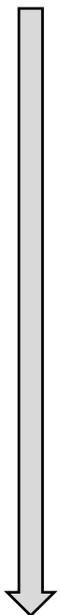
±(99%信用区間)

数値実験 - 結果

パラメータの事後分布

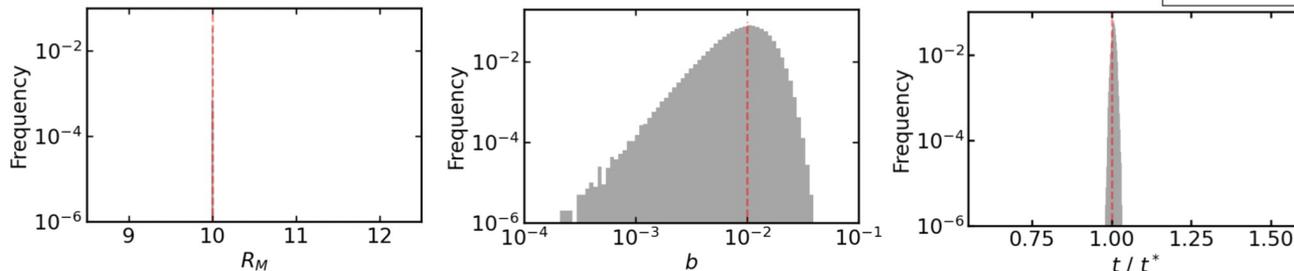
データ欠損

小

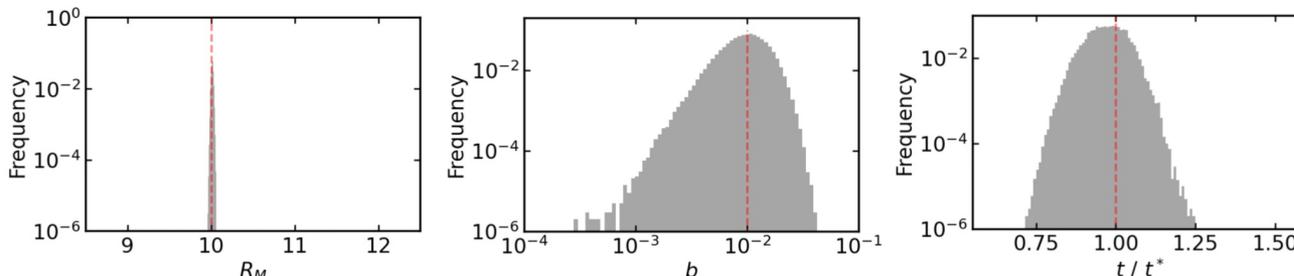


大

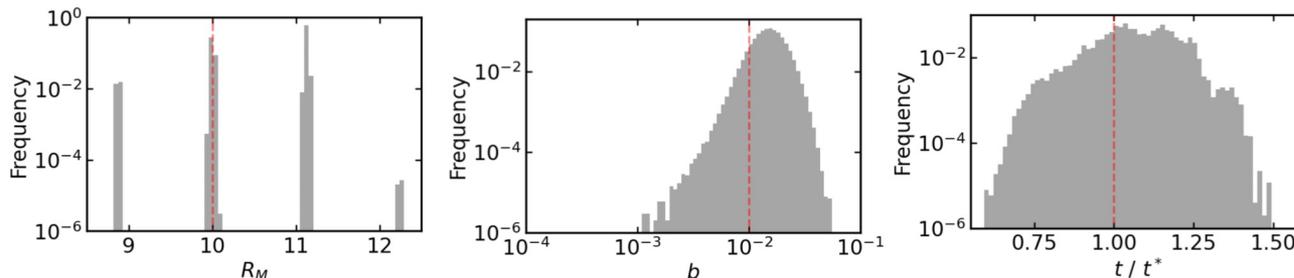
(a) $q_{\min} = 0.4$



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



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



t の真値: t^*

$q_{\min} = 2.65$ のとき R_M のヒストグラム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

まとめ

- 交換モンテカルロ法を用いることで局所解を避けてパラメータ推定することが出来た.
- パラメータを分布推定し, 推定の信頼度を解釈した.
→データの角度領域の欠損に対する推定限界を求めた.

目次

- X線小角散乱法の実験概要
- 従来 of 解析方法
- ベイズ推定
- 数値実験
- 今後の展望

今後の展望

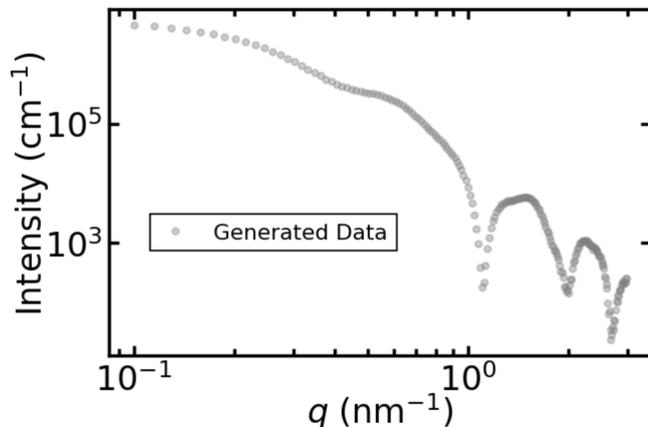
① 現実のデータの解析

現実のSAXSデータを用いたベイズ計測の評価を行う。

② ベイズモデル選択の応用

モデル選択によって、計測データから試料の構造を求める。

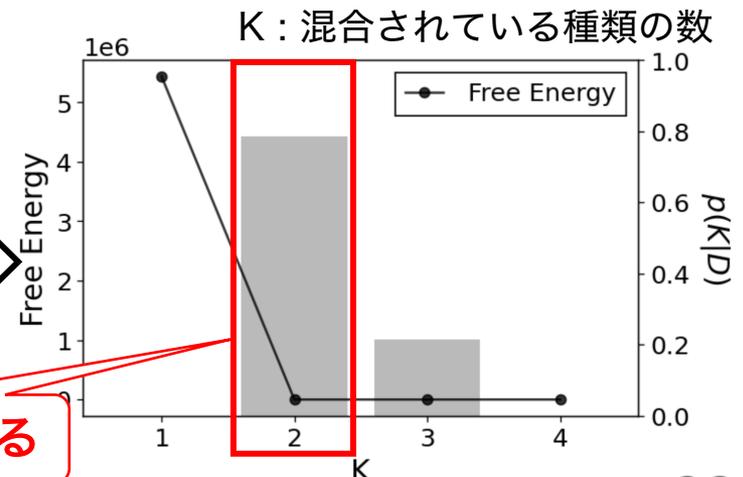
例：何種類の単分散球が混合されているかを求める。



計測データ

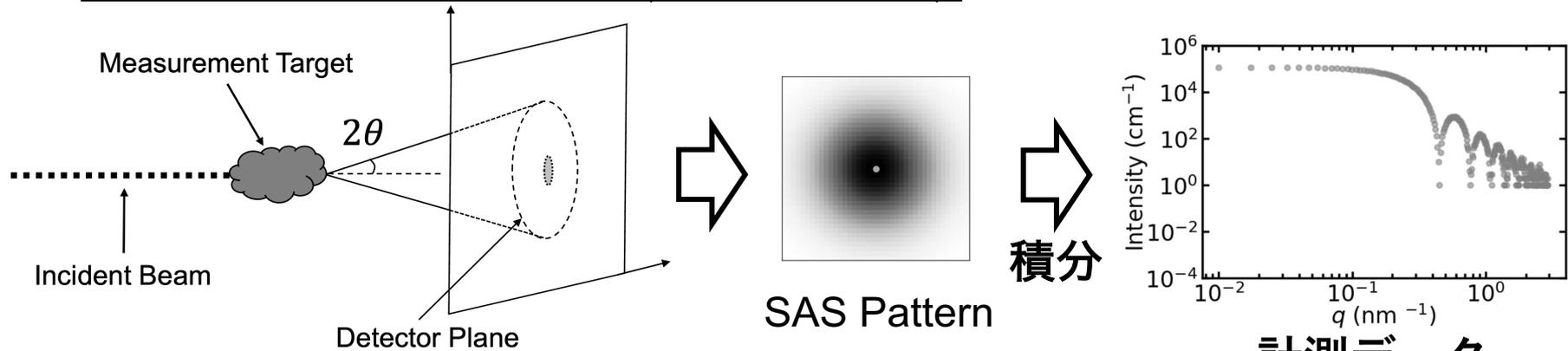


2種類と分かる



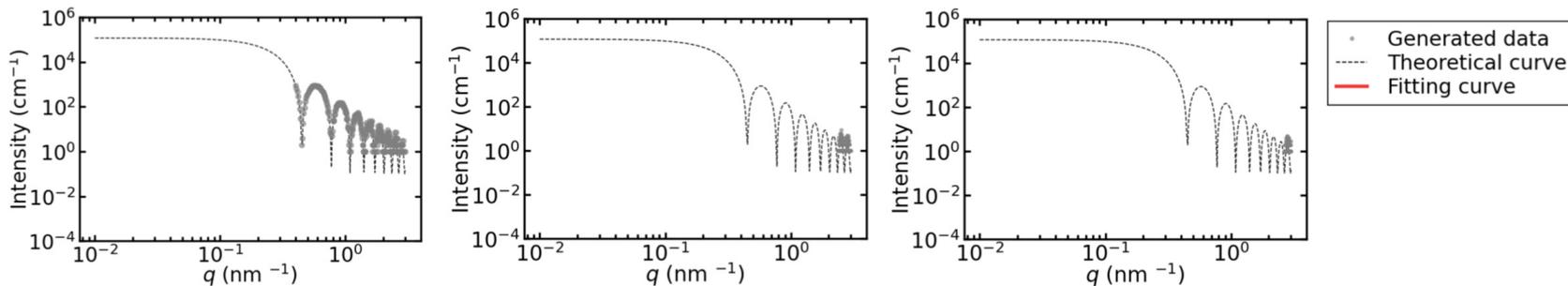
ベイズ自由エネルギー

X線小角散乱(SAXS)

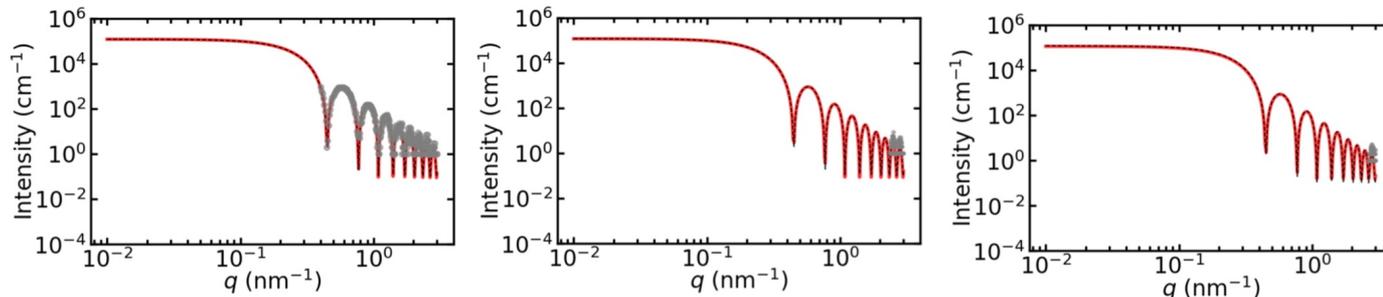


ベイズ計測

人工データ



フィッティング



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

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

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