

実践データ駆動科学オンラインセミナー

8th seminar データ駆動科学の最前線

複雑ネットワーク解析 / ベイズ統計

ベイズ推論に基づく 物理モデル構築過程の効率化

統計数理研究所 統計的機械学習研究センター

特任助教 本武 陽一

2021年7月30日

16:00 ~ 16:30

オンライン (ZOOM)

自己紹介

福岡県福岡市出身

学歴：

2008/3 東北大学 理学部 物理学科卒業

2010/3 北海道大学 大学院理学院 宇宙理学専攻

修士課程修了 指導教官：石川健三 教授（素粒子物理学、テーマ：波束のダイナミクス）

2013/3 東京大学 大学院総合文化研究科 広域科学専攻

修士課程修了 指導教官：植田一博 教授（認知科学、テーマ：脳活動を用いたHAI評価）

2016/3 東京大学 大学院総合文化研究科 広域科学専攻

博士課程修了 指導教官：池上高志 教授（複雑系、テーマ：深層学習、群れ）

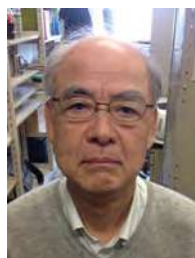
職歴：

2016/5 ～ 2019/3 東京大学 大学院新領域創成科学研究科 複雑理工学専攻

特任研究員 （岡田真人教授、テーマ：データ駆動科学）

2019/4 ～ 統計数理研究所 統計的機械学習研究センター

特任助教 （福水健次教授、テーマ：統計的機械学習）



自己紹介

武谷三段階論 [武谷 1942, 1966]

現象

機械学習が得意

実体論

人間が得意

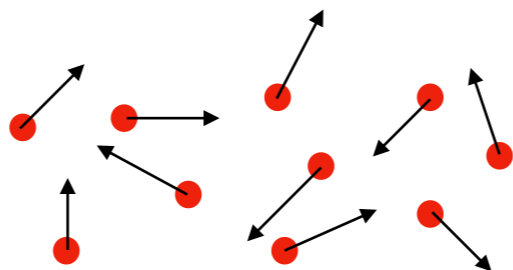
本質論

現象の内挿的モデル化と予測

抽象化による大胆な外挿

現象のモデル化

気体の分子運動論



一般原理の抽出

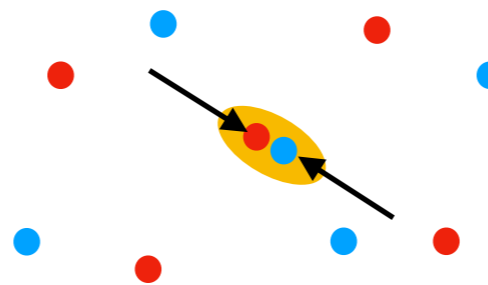
エントロピー増大則

$$dS = dQ/T$$

機械学習 ↔ 科学者

現象の予測

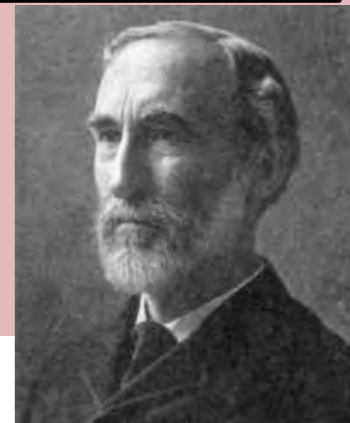
化学反応速度論



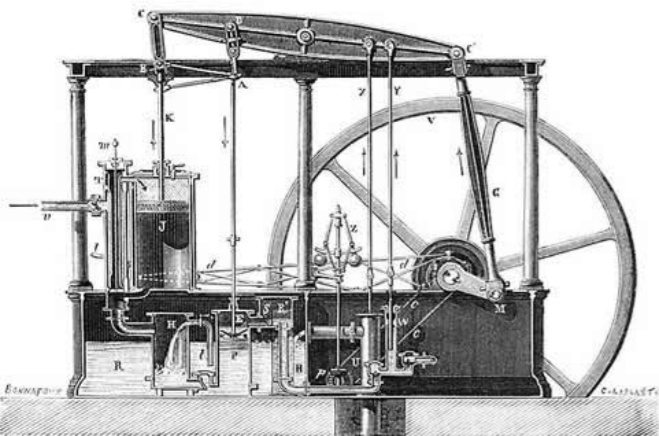
原理の転移

化学ポテンシャル

$$\mu = \partial F / \partial N$$



Willard Gibbs



蒸気機関



化学産業

自己紹介

武谷三段階論 [武谷 1942, 1966]

現象

大同特殊鋼 HPより



磁性材料

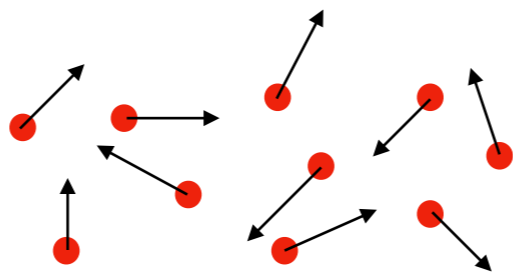
機械学習が得意

実体論

現象の内挿的モデル化と予測

現象のモデル化

気体の分子運動論



人間が得意

本質論

抽象化による大胆な外挿

一般原理の抽出

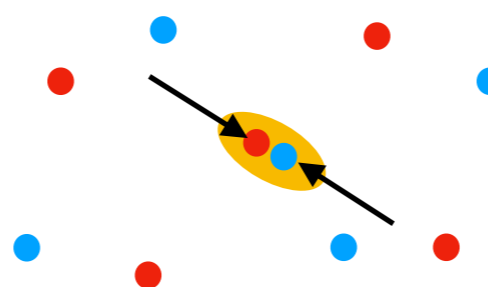
エントロピー増大則

$$dS = dQ/T$$

機械学習 ⇔ 科学者

現象の予測

化学反応速度論



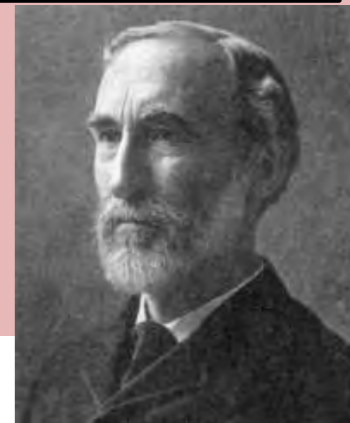
原理の転移

化学ポテンシャル

$$\mu = \partial F / \partial N$$



構造材料



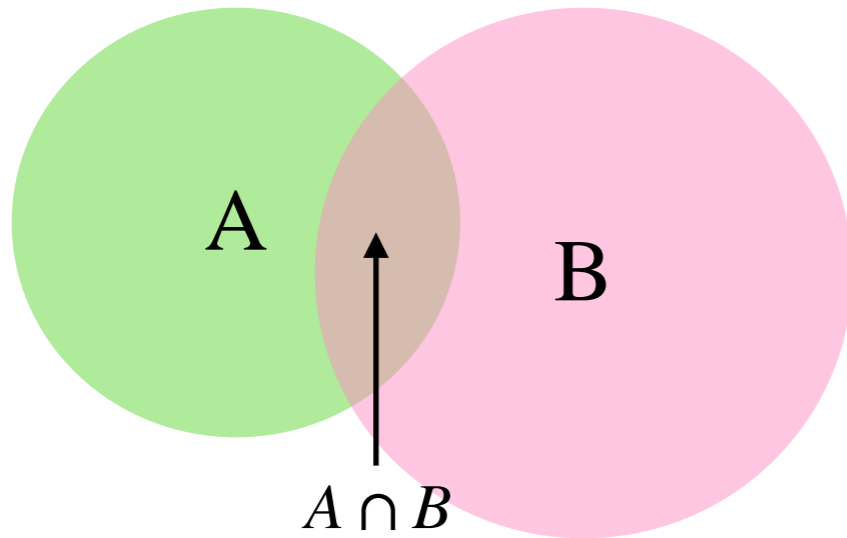
Willard Gibbs

本日の内容

1. 科学者の営みとベイズ推論
2. ベイズ的線形回帰
3. ベイズ推論の利点
4. ベイズ推論の応用事例紹介
5. まとめ

1. 科学者の営みとベイズ推論

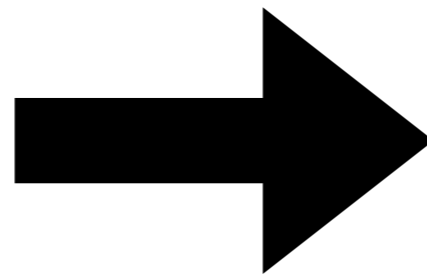
ベイズ統計



条件付き確率

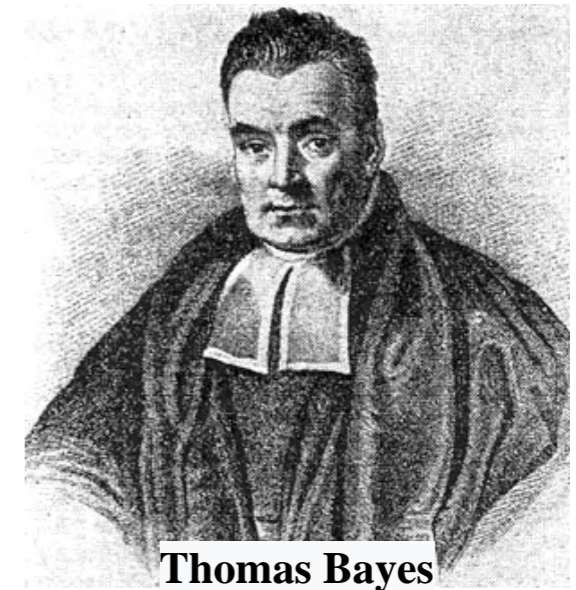
$$\begin{aligned} \bullet P(A|B) &= \frac{n(A \cap B)}{n(B)} \\ &= \frac{n(A \cap B)/N}{n(B)/N} = \frac{P(A \cap B)}{P(B)} \end{aligned}$$

$$\begin{aligned} \bullet P(B|A) &= \frac{n(B \cap A)}{n(A)} \\ &= \frac{n(B \cap A)/N}{n(A)/N} = \frac{P(B \cap A)}{P(A)} \end{aligned}$$



ベイズの定理

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$



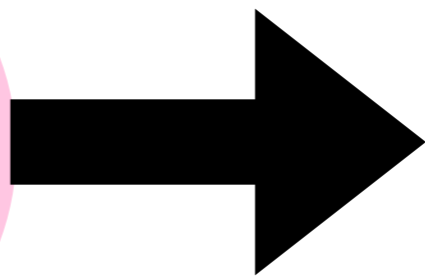
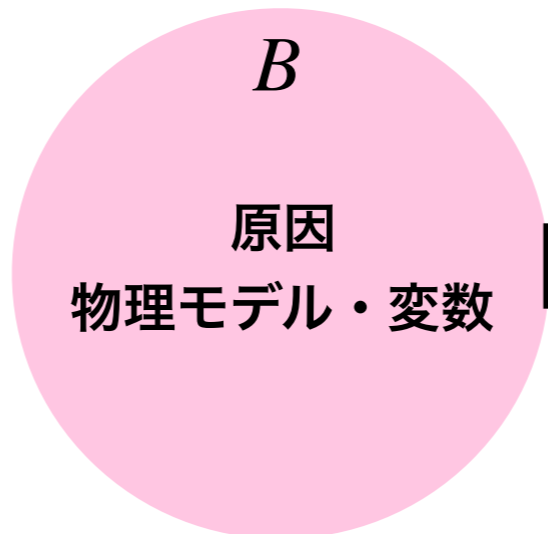
ベイズ統計

ベイズの定理 :

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

因果関係 : $B \rightarrow A$
 ➡逆転 : $A \rightarrow B$

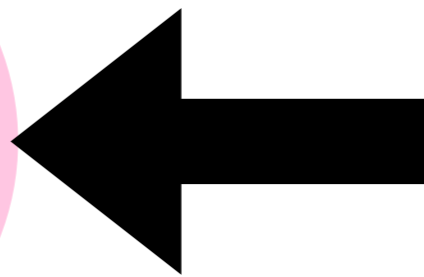
因果則
 $P(A|B)$



事前知識 $P(B)$

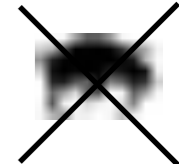


$P(B|A)$



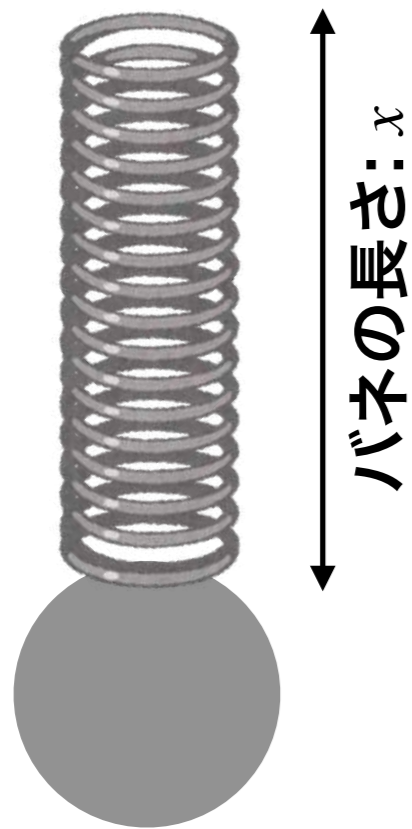
データから物理モデル・変数の尤もらしさを推定!

真実はひとつではない

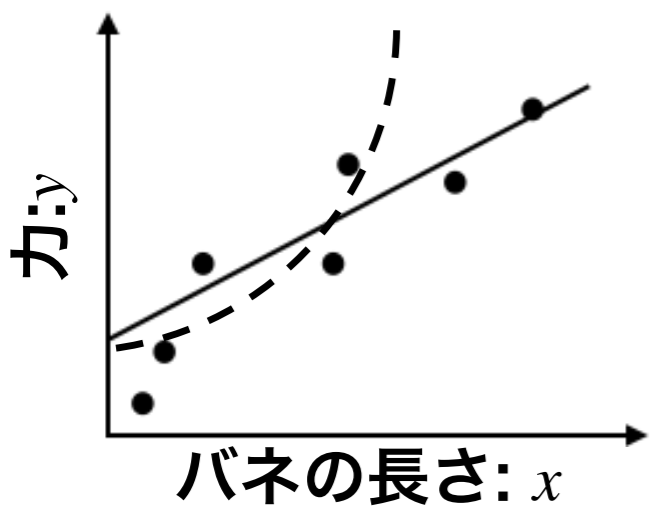


- B: 原因
- A: 結果
- $P(B)$: 原因の事前知識
- $P(A|B)$: 因果モデル
- $P(B|A)$: 原因の推定結果

ベイズ統計



バネの初期長さ x_0



因果則
 $P(A | B)$

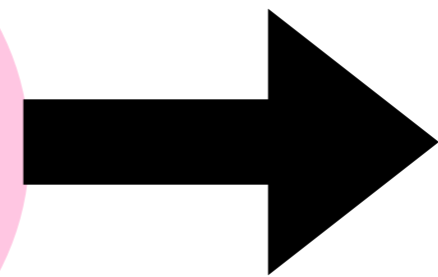
B

$M_1 : y = k(x - x_0)$
 $\theta_1 = \{k, x_0\}$

$M_2 : y = k(x - x_0)^2$
 $\theta_2 = \{k, x_0\}$

A

$D = \{x_i, y_i\}_{i=1}^N$



$P(B | A)$

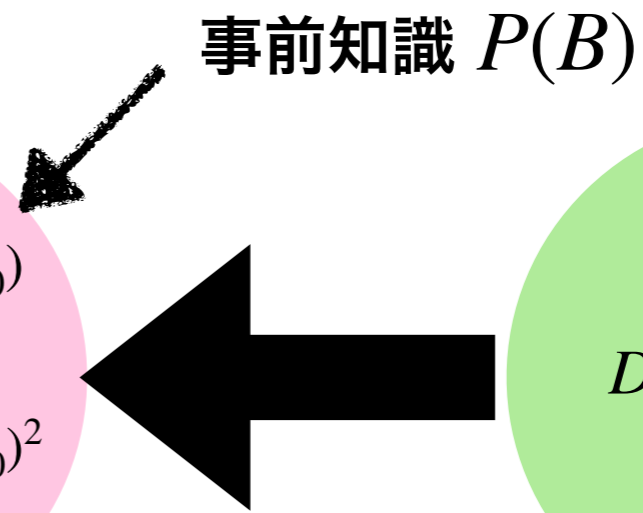
B

$M_1 : y = k(x - x_0)$
 $\theta_1 = \{k, x_0\}$

$M_2 : y = k(x - x_0)^2$
 $\theta_2 = \{k, x_0\}$

A

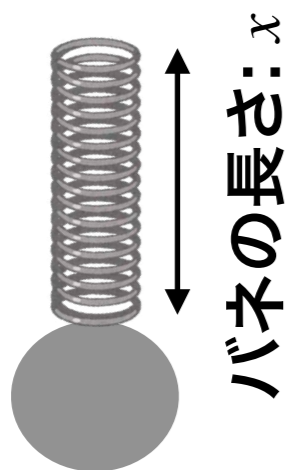
$D = \{x_i, y_i\}_{i=1}^N$



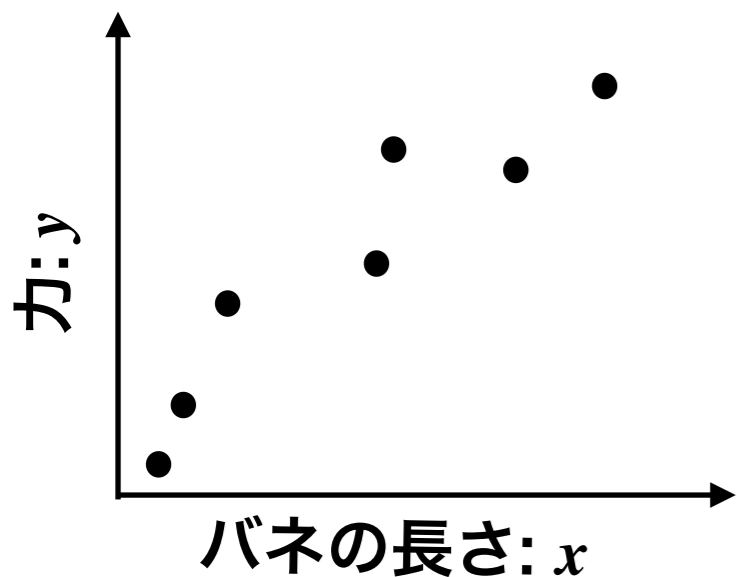
物理モデルや物理変数は一意に定まらない。
これが自然の摂理と考えるのは気持ち悪い！

科学者（人間）の営みとベイズ推論

観測・測定



バネの初期長さ x_0



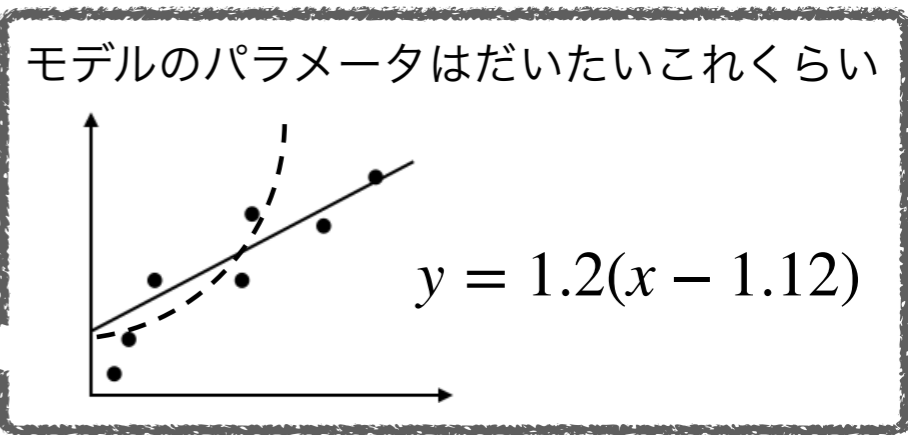
科学者

モデル候補
 $y = k(x - x_0)$
 $y = k(x - x_0)^2$



事前知識

測定装置はあまり信頼できない
 実験者が初心者
 信頼できる先生の発言
 シンプルなモデルが良いモデル
 エントロピー増大則
 太陽は東から登る
 地震が起こった
 湿度が高かった
 暑かった・寒かった
 宇宙線が飛んできた



「真実（犯人）は **(80%の確率で)** これだ！ $\rightarrow y = 1.2(x - 1.12)$ 」

科学者の営みをベイズ推論としてモデル化すると考える

2. フックの法則のベイズ推論

バネモデルの構築

【知っていること】

因果則 $P(Y | k, x_0, M_k)$

因果則

$P(A | B)$

B

$$M_1 : y = k(x - x_0)$$

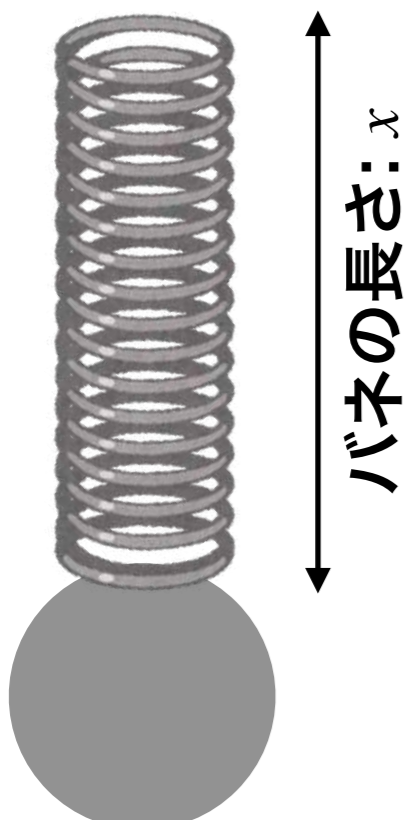
$$\theta_1 = \{k, x_0\}$$

$$M_2 : y = k(x - x_0)^2$$

$$\theta_2 = \{k, x_0\}$$

A

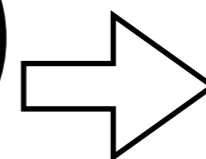
$$D = \{x_i, y_i\}_{i=1}^N$$



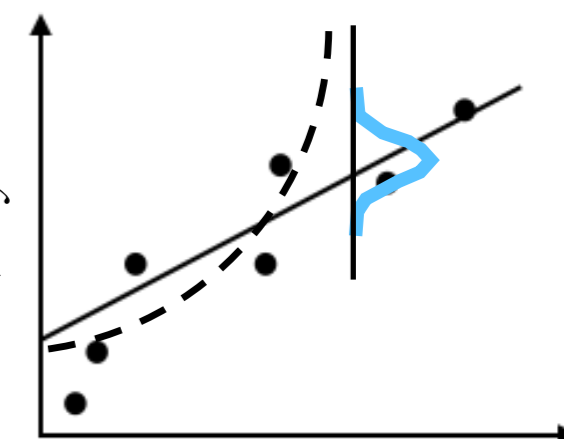
バネの初期長さ x_0

$$P(y_i | k, x_0, M_1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - k(x_i - x_0))^2}{2\sigma^2}\right)$$

回帰式を中心に測定誤差 (ガウスノイズ) が乗る



力: y



バネの長さ: x

$$P(A | B) = P(y_1, y_2, \dots, y_N | k, x_0, M_1)$$

$$= \prod_{i=1}^N P(y_i | k, x_0, M_1) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - k(x_i - x_0))^2}{2\sigma^2}\right)$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{N}{2}} \exp\left[-\frac{\sum_{i=1}^N (y_i - k(x_i - x_0))^2}{2\sigma^2}\right] \rightarrow \text{最小二乗法}$$

ベイズ推論によるバネ定数推定

【知りたいこと】

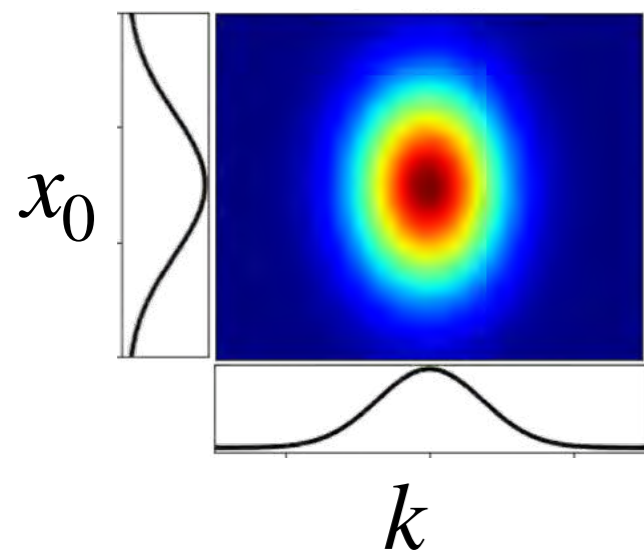
事後分布 $P(k, x_0 | \mathbf{Y})$

ベイズの定理 :

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

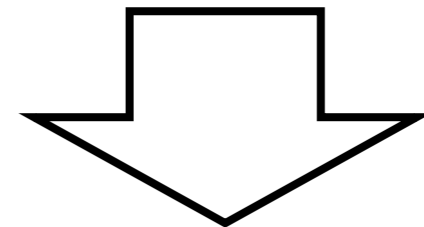
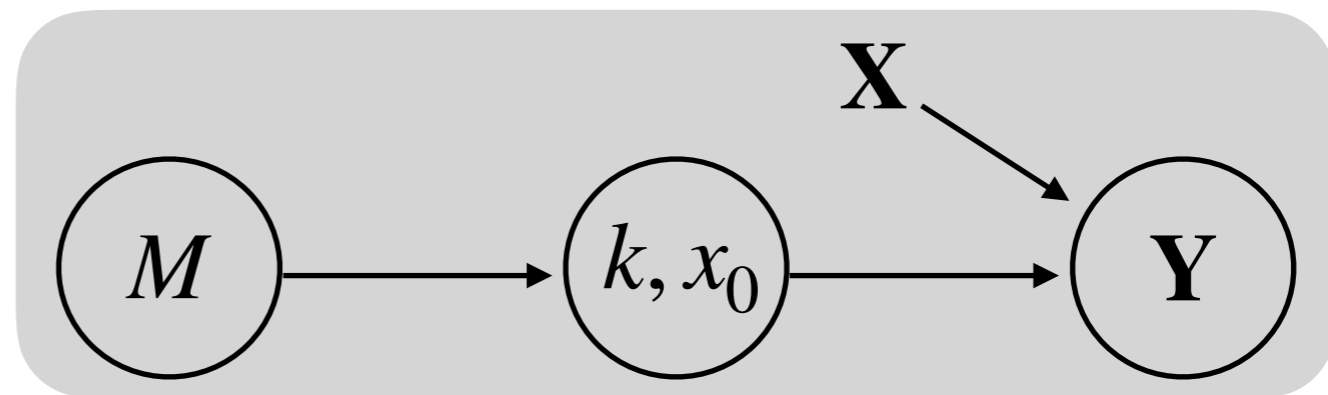
因果関係 : $B \rightarrow A$

➡逆転 : $A \rightarrow B$



「知ってること」 : 因果則 (順モデル)

$$P(y_1, y_2, \dots, y_N | k, x_0, M_1)$$



「知りたいこと」 : 尤もらしいバネ定数 (パラメータ)

$$P(k, x_0 | y_1, y_2, \dots, y_N, M_k) = \frac{P(y_1, y_2, \dots, y_N | k, x_0, M_k) P(k, x_0, M_k)}{P(y_1, y_2, \dots, y_N)}$$

ベイズ推論によるバネモデル評価

【知りたいこと】

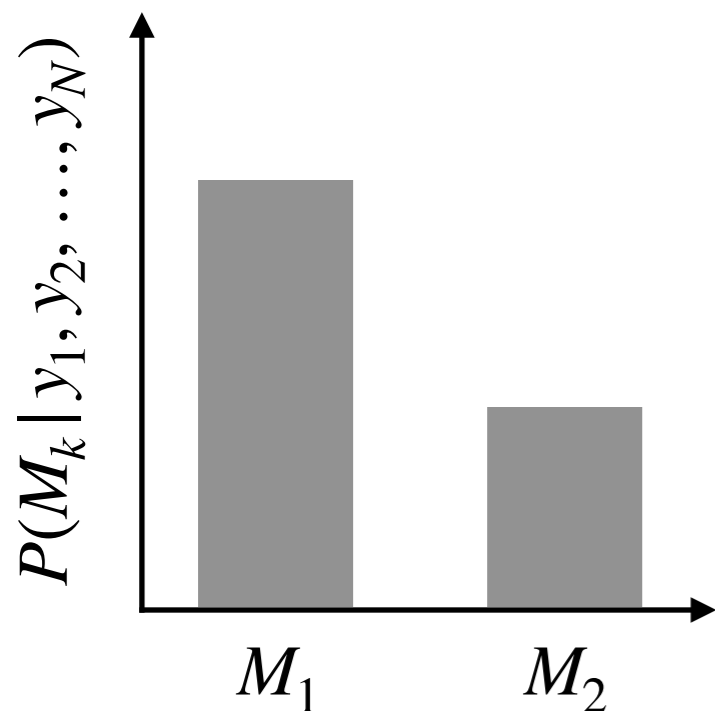
事後分布 $P(M_k | \mathbf{Y})$

ベイズの定理：

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

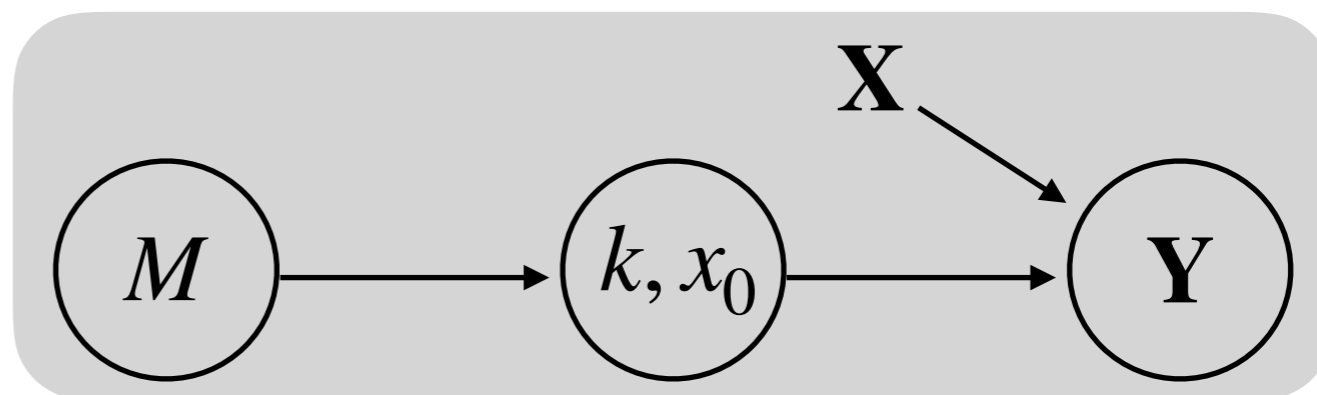
因果関係：B → A

→ 逆転：A → B



「知ってること」：因果則（順モデル）

$$P(y_1, y_2, \dots, y_N | k, x_0, M_1)$$



「知りたいこと」：モデルの尤もらしさ

$$P(M_k | y_1, y_2, \dots, y_N) = \frac{P(y_1, y_2, \dots, y_N | M_k)P(M_k)}{P(y_1, y_2, \dots, y_N)}$$

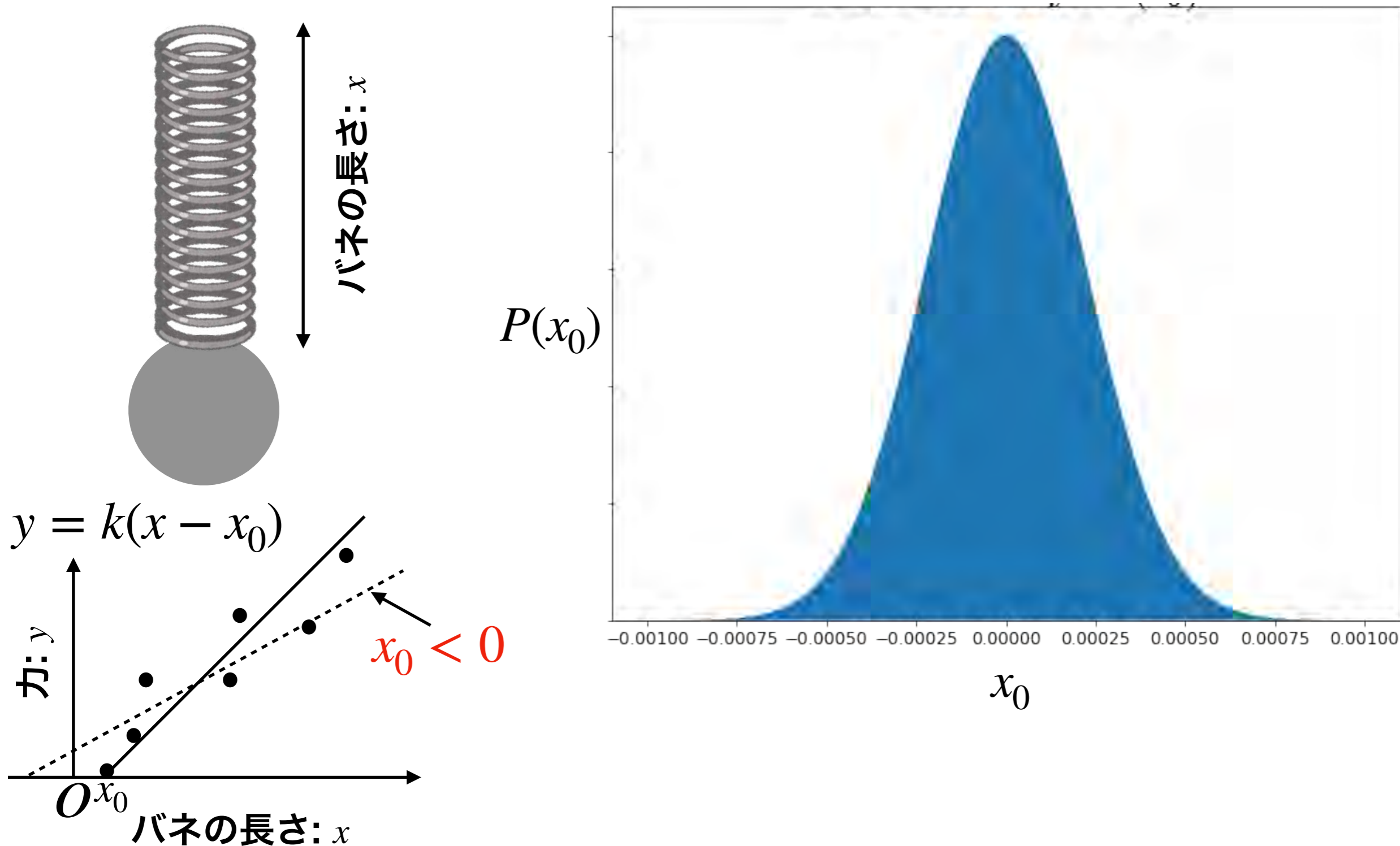
$$= \int dk dx_0 \frac{P(y_1, y_2, \dots, y_N | k, x_0, M_k)P(k, x_0 | M_k)P(M_k)}{P(y_1, y_2, \dots, y_N)}$$

3. ベイズ推論の利点

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

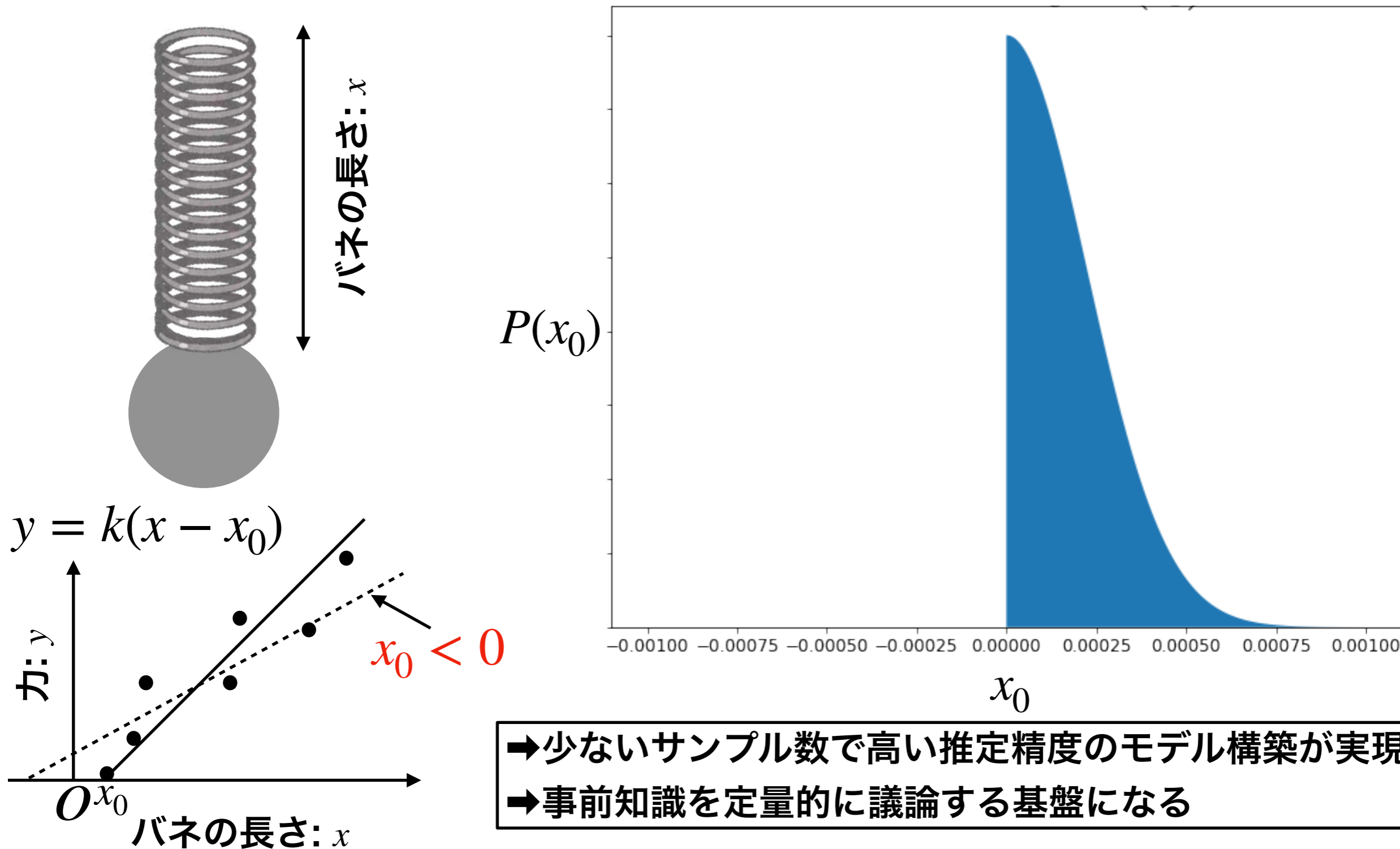
ベイズ推論の利点1

データにない物理学者の知識を事前分布として導入できる



$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$ ベイズ推論の利点1

データにない物理学者の知識を事前分布として導入できる



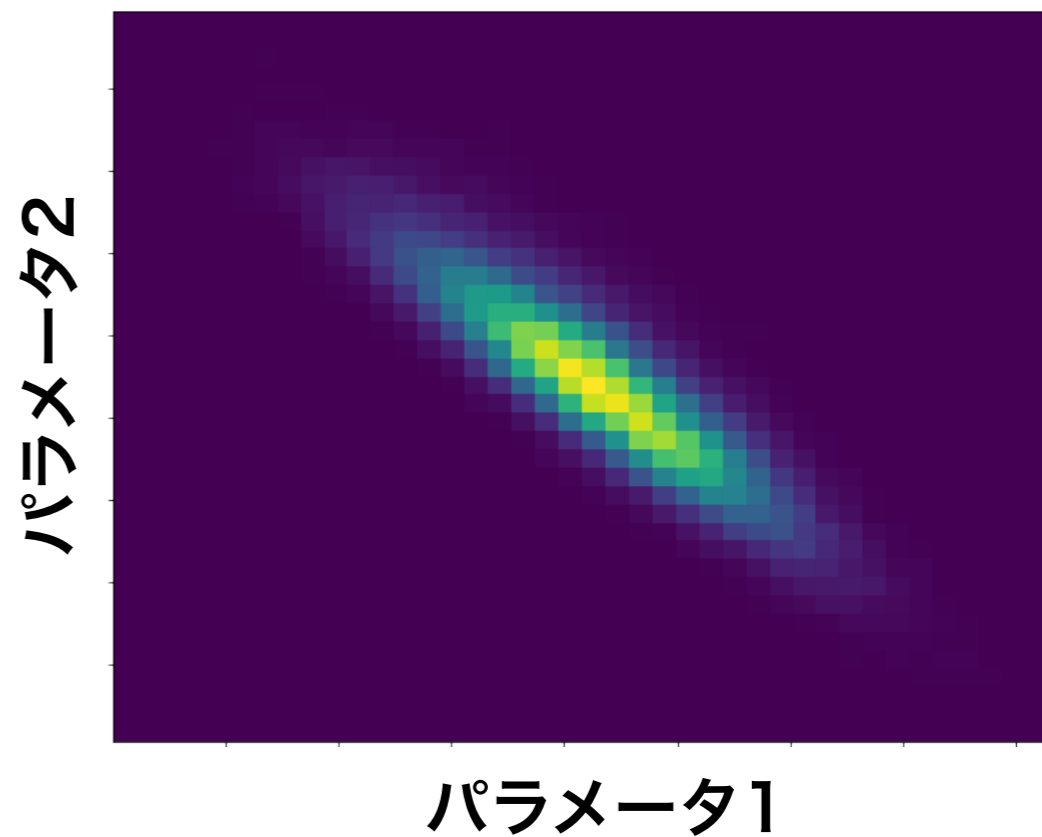
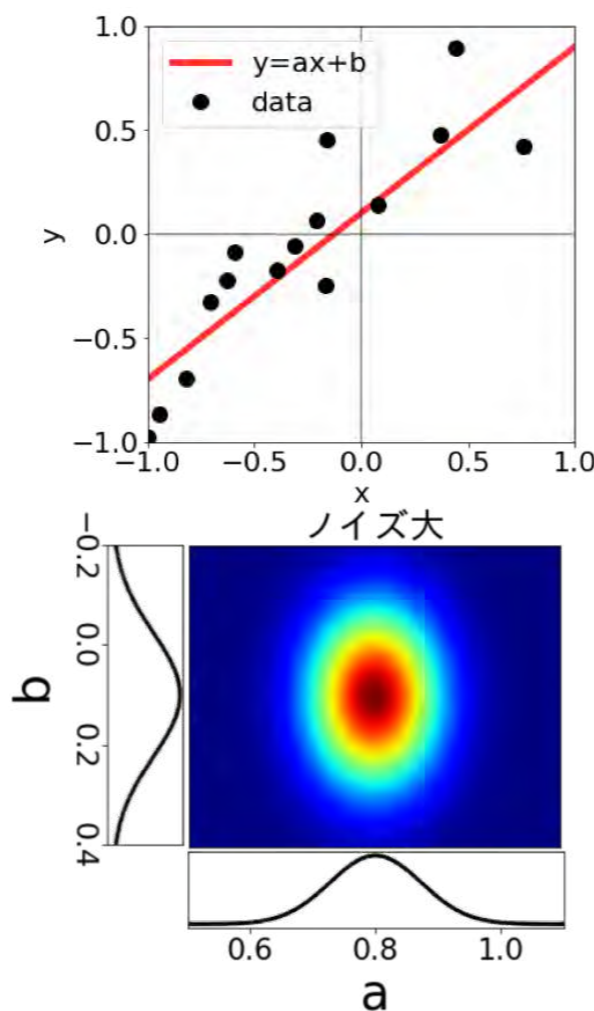
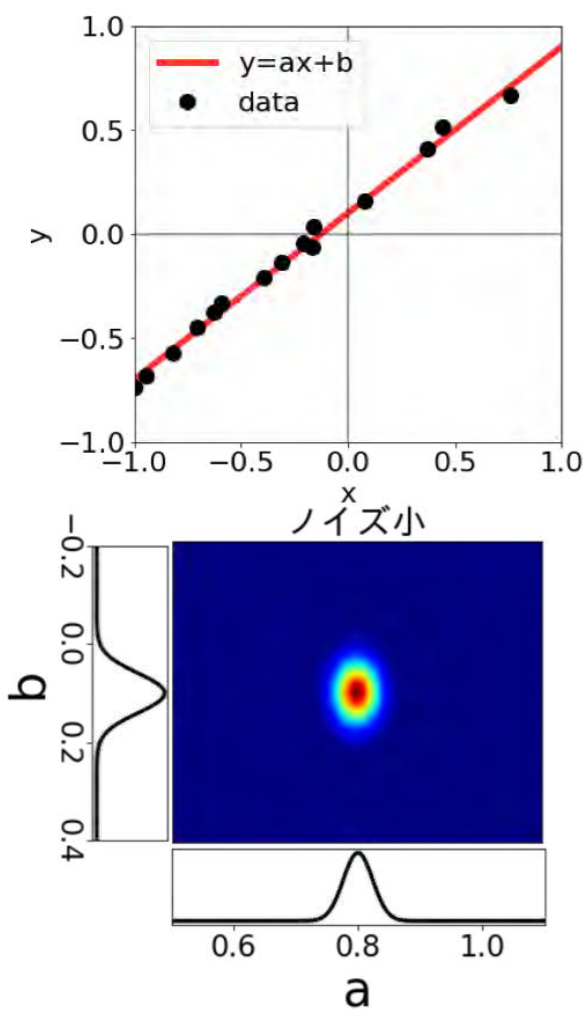
- ➡ 少ないサンプル数で高い推定精度のモデル構築が実現できる
- ➡ 事前知識を定量的に議論する基盤になる

$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$ ベイズ推論の利点2

事後分布を通してパラメータ推定精度や物理モデルの性質がわかる

●推定精度

●物理モデルの理解



➡パラメータ1とパラメータ2は相関関係

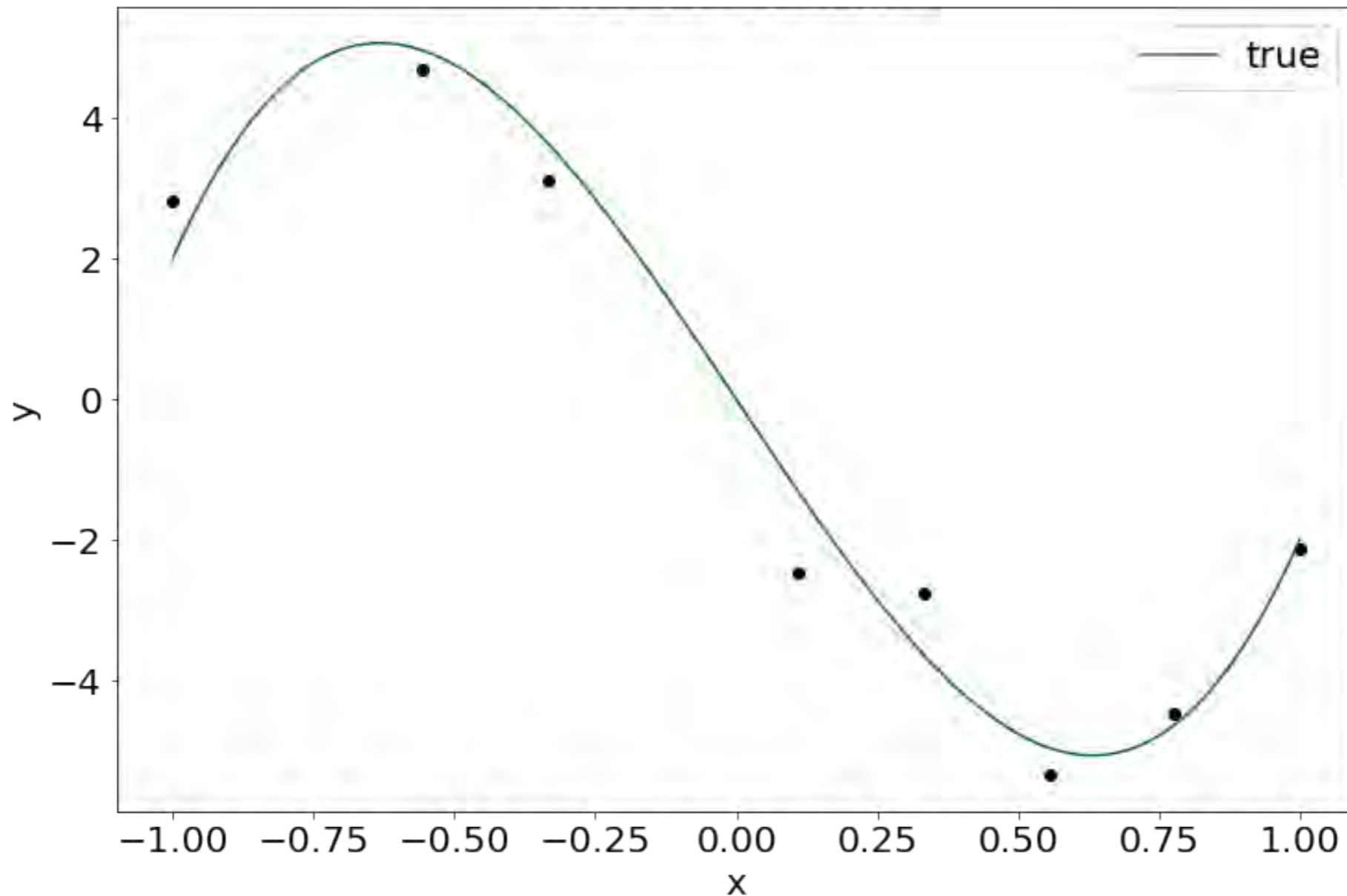
ベイズ推論の利点3

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

このデータは何次関数？→3次関数

$$y = 10x^3 - 12x$$

8th order function

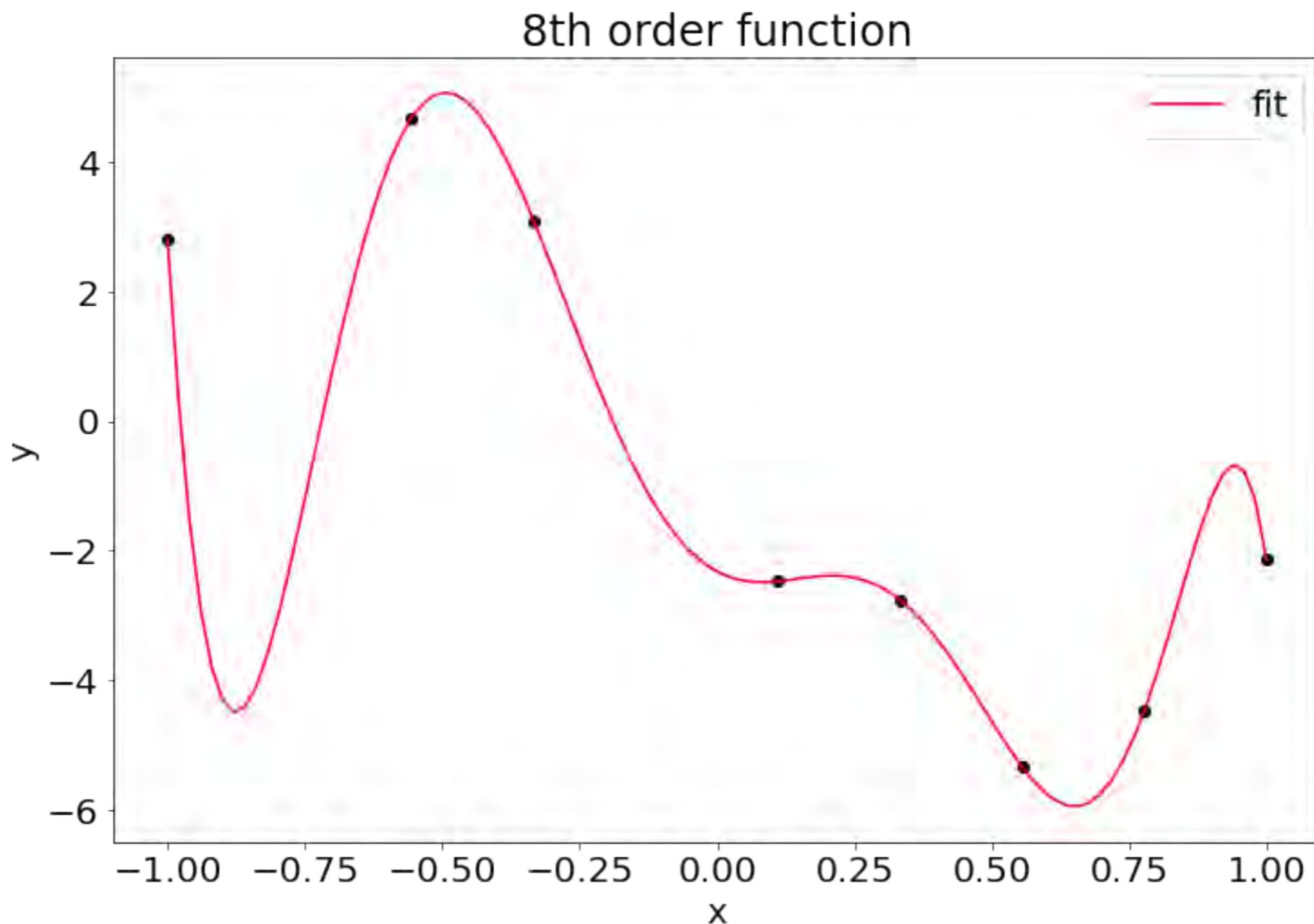


ベイズ推論の利点3

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

●線形回帰

$$f(x) = w_0 + w_1x + w_2x^2 + w_3x^3 + w_4x^4 + w_5x^5 + w_6x^6 + w_7x^7 + w_8x^8$$

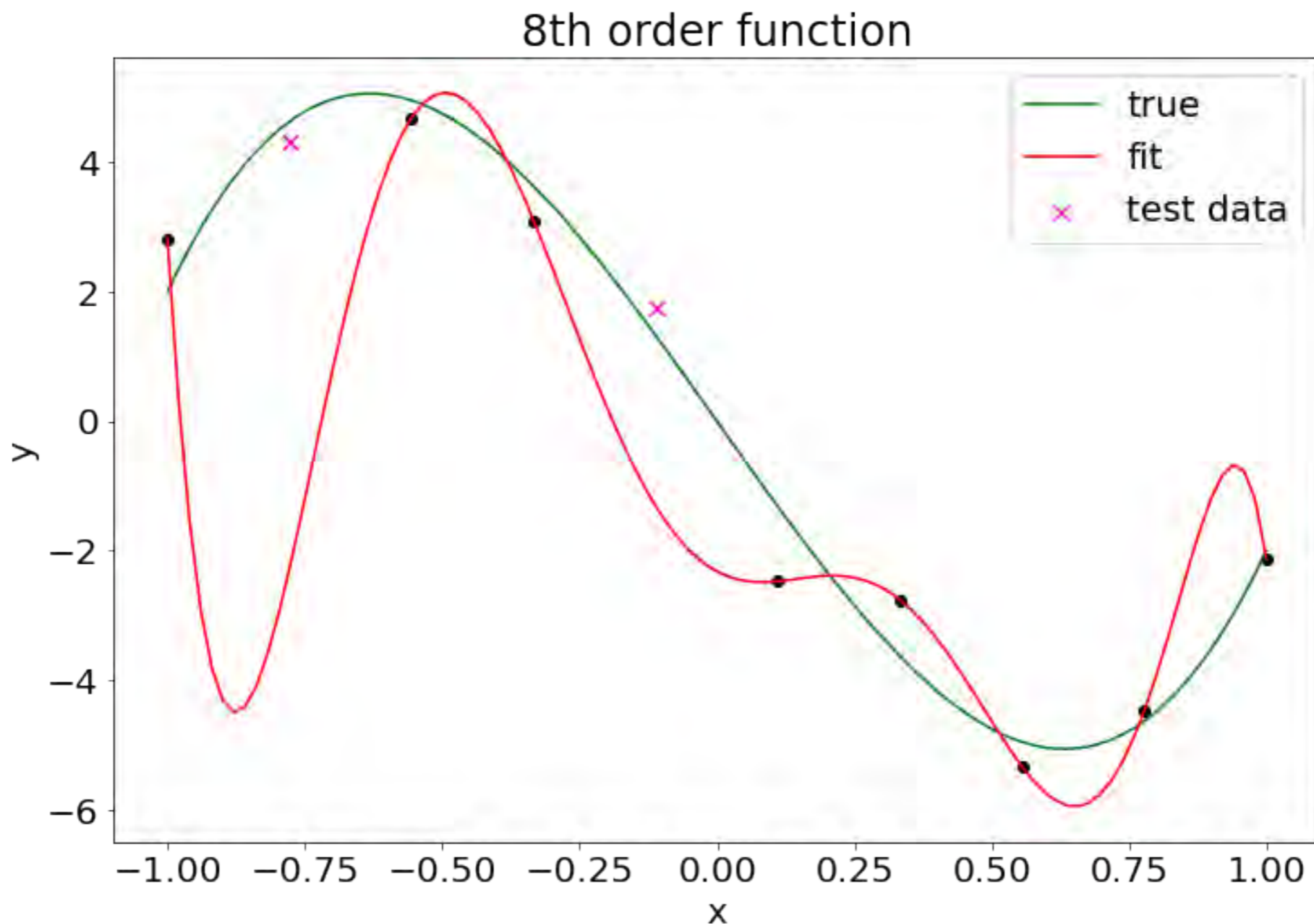


ベイズ推論の利点3

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

●線形回帰

$$f(x) = w_0 + w_1x + w_2x^2 + w_3x^3 + w_4x^4 + w_5x^5 + w_6x^6 + w_7x^7 + w_8x^8$$



$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$ ベイズ推論の利点3

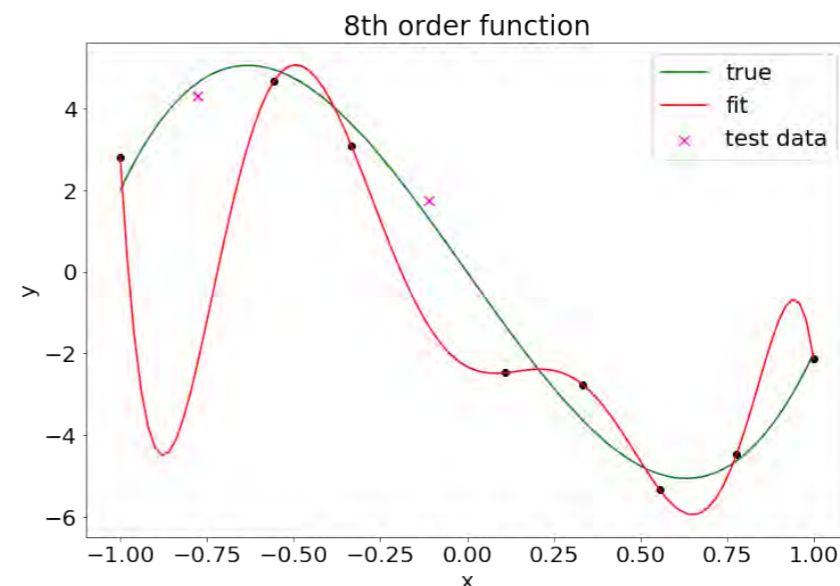
過適合を考慮したモデル評価ができる

最小二乗法 (最尤法)

→ 二乗誤差を最小化

$$E = \frac{1}{2N} \sum_{i=1}^N (y_i - f(x_i; \mathbf{w}))^2$$

\mathbf{w} : モデルパラメータ



過適合を防ぐ

$$-\log P(M_k | y_1, y_2, \dots, y_N) \sim \underbrace{\sum_{i=1}^N (y_i - f(x_i; \mathbf{w}))^2}_{NE \text{ (エネルギー)}} + \frac{d}{2} \log N$$

モデルの複雑さ
に対する罰則項
 d : パラメータ数

※鞍点近似が成り立つ場合

ベイズ推論の利点のまとめ

利点1：データにない科学者の知識を事前分布として導入可能

利点2：事後分布から物理量の推定精度や物理モデルの性質がわかる

利点3：過適合を考慮したモデル評価が可能

4. ベイズ推論の応用事例紹介

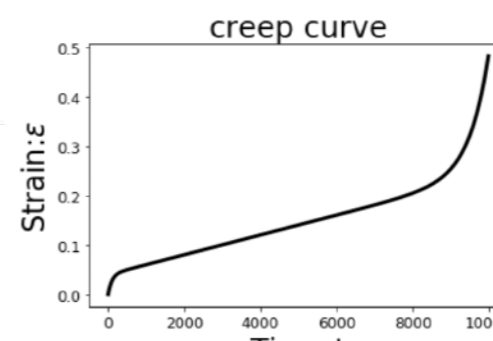
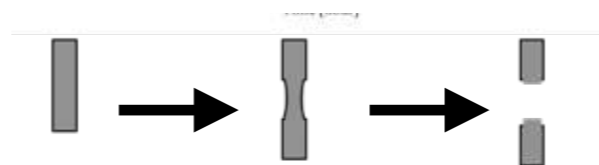
ベイズ推論による 物理モデル構築の支援

同一の現象に対して複数の縮約モデル候補
 → ベイズ推論を用いて各モデルを評価・解釈

金属のクリープ過程の縮約モデル



NEDO Project Success Story

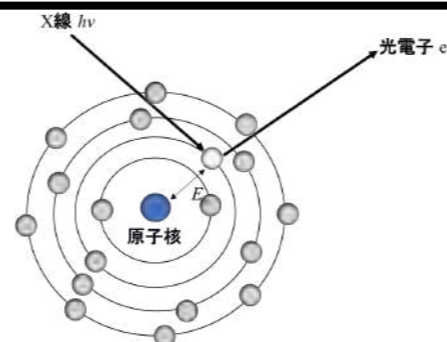


[Mototake et al., Scientific Reports, 2020]

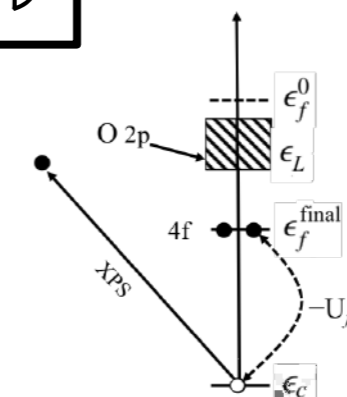
3d-X線光電子分光過程の縮約モデル



SPring-8 [SPring-8 HPより]



日本分析機器工業会 HPより



[Mototake et al., JPSJ, 88(3), 034004, 2019]

A universal Bayesian inference framework for complicated creep constitutive equations

Yoh-ichi Mototake¹, Hitoshi Izuno², Kenji Nagata²,
Masahiko Demura², and Masato Okada³

¹The Institute of Statistical Mathematics

²Research and Services Division of Materials Data and Integrated System, National Institute for Materials Science

³Graduate School of Frontier Sciences, University of Tokyo

Scientific Reports 10, 1, (2020).

金属のクリープ破壊

●金属のクリープ現象

高温・高圧下での構造材料の破壊現象

●クリープ試験



NEDO Project Success Story

Acquisition, Technology & Logistics Agency

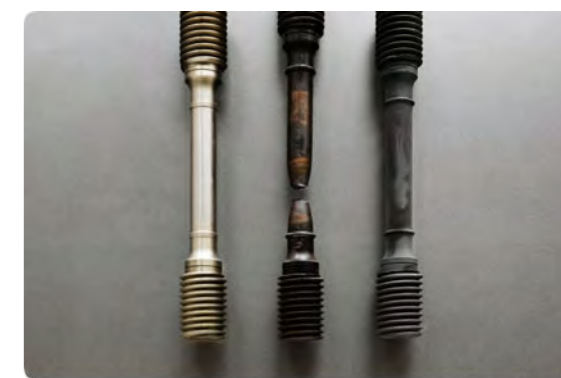


Photo by Nacása & Partners Inc.

<http://fykai.blog.fc2.com/blog-entry-140.html>

NIMSが1969年6月19日に開始したクリープ試験が、2011年3月14日時点で試験開始からの総試験日数が14,868日に到達し、最長クリープ試験のギネス世界記録™として認定されました。

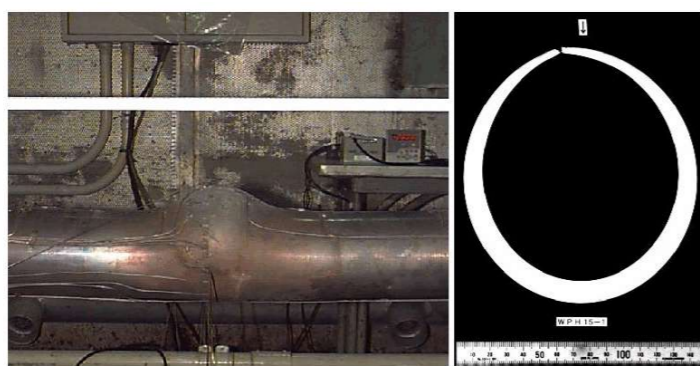


図8 配管高温負荷試験後の小口径直管試験部外観及び加熱領域断面形状
[出典]Proc.ICONE-7, ICONE-7255, Tokyo, Japan, April 19-23, 1999

<http://fykai.blog.fc2.com/blog-entry-140.html>

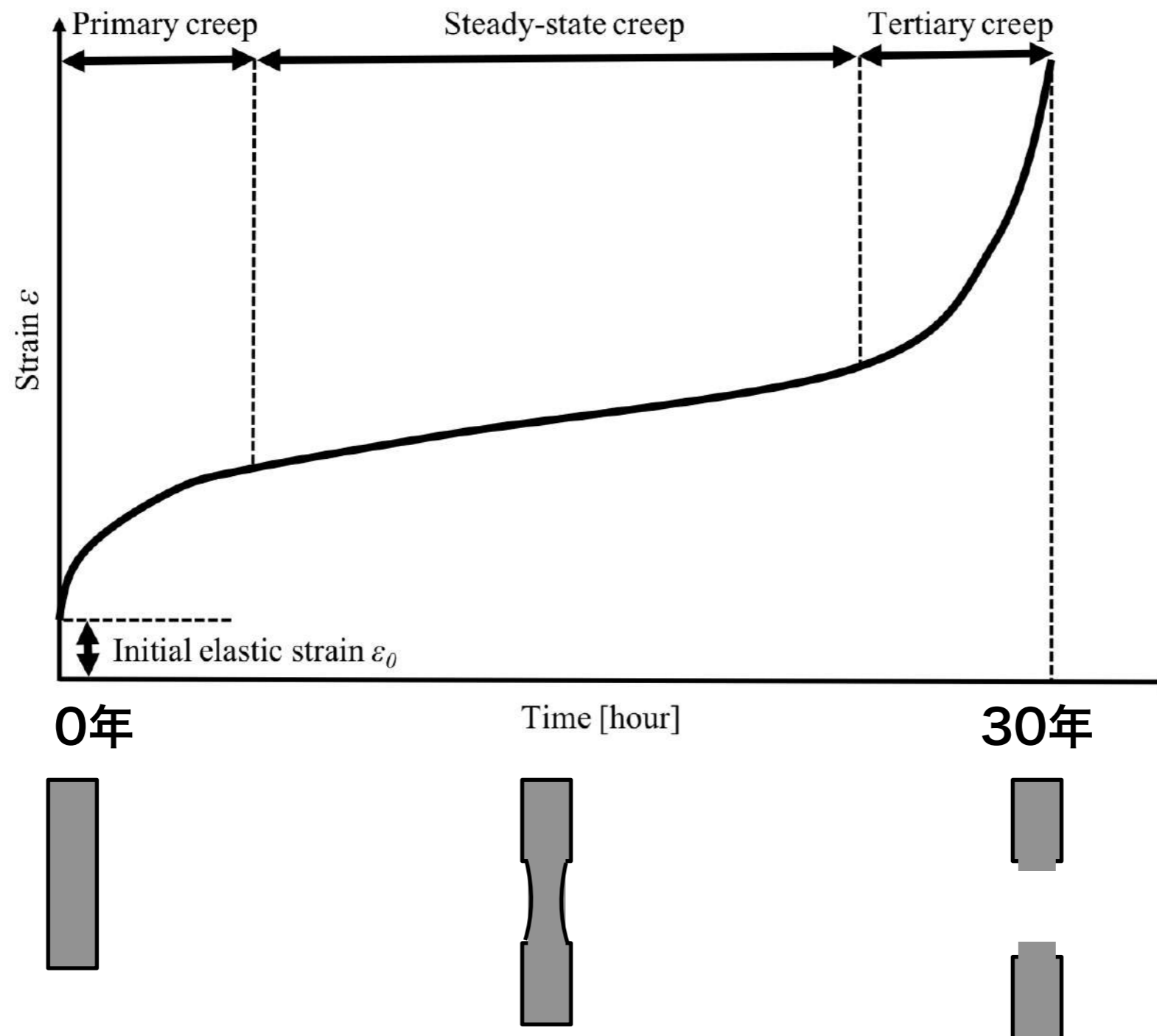


NIMS HPより

<https://www.nims.go.jp/news/archive/2020/02/202002280.html>

金属のクリープ破壊

●クリープ試験



クリープ構成式によるアプローチ

複数提案されているクリープモデル

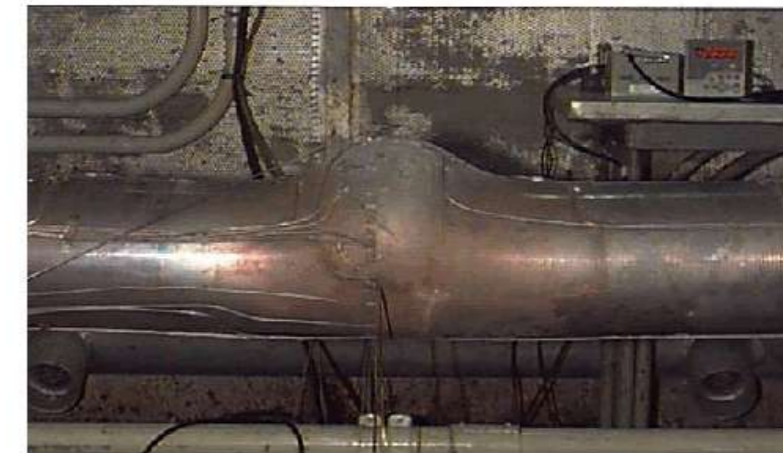
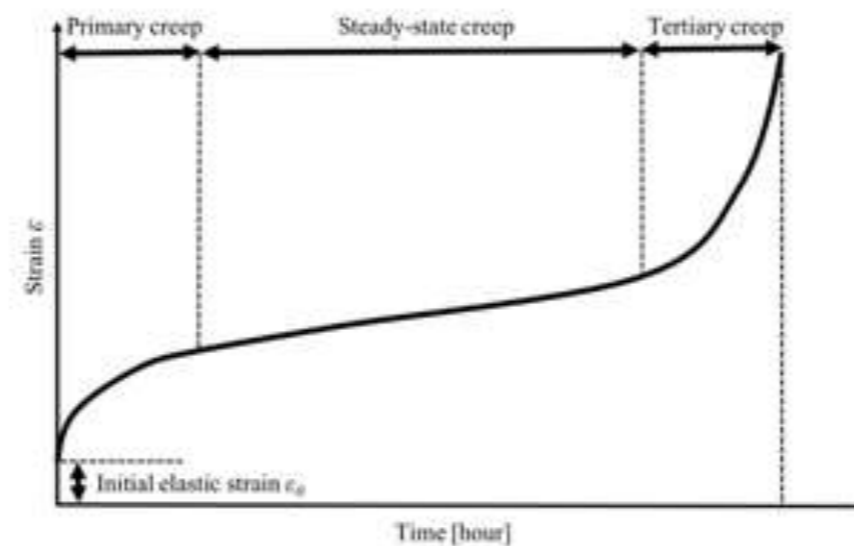
$$\varepsilon = \varepsilon_0 + w_1(1 - \exp(-\alpha t)) + w_2 t + w_3(\exp(\beta t) - 1)$$

$$\varepsilon = \varepsilon_0 + w_1(1 - \exp(-\alpha t)) + w_3(\exp(\beta t) - 1)$$

$$\varepsilon = \varepsilon_0 + a_1 \times t^{b_1} + a_2 \times t^{b_2} + c_1 \exp(d_1 \times t) + c_2 \exp(d_2 \times t)$$

⋮

どれが適切なモデルなのか？



(a) 外観

図8 配管高温負荷試験後の小口径直管試験部

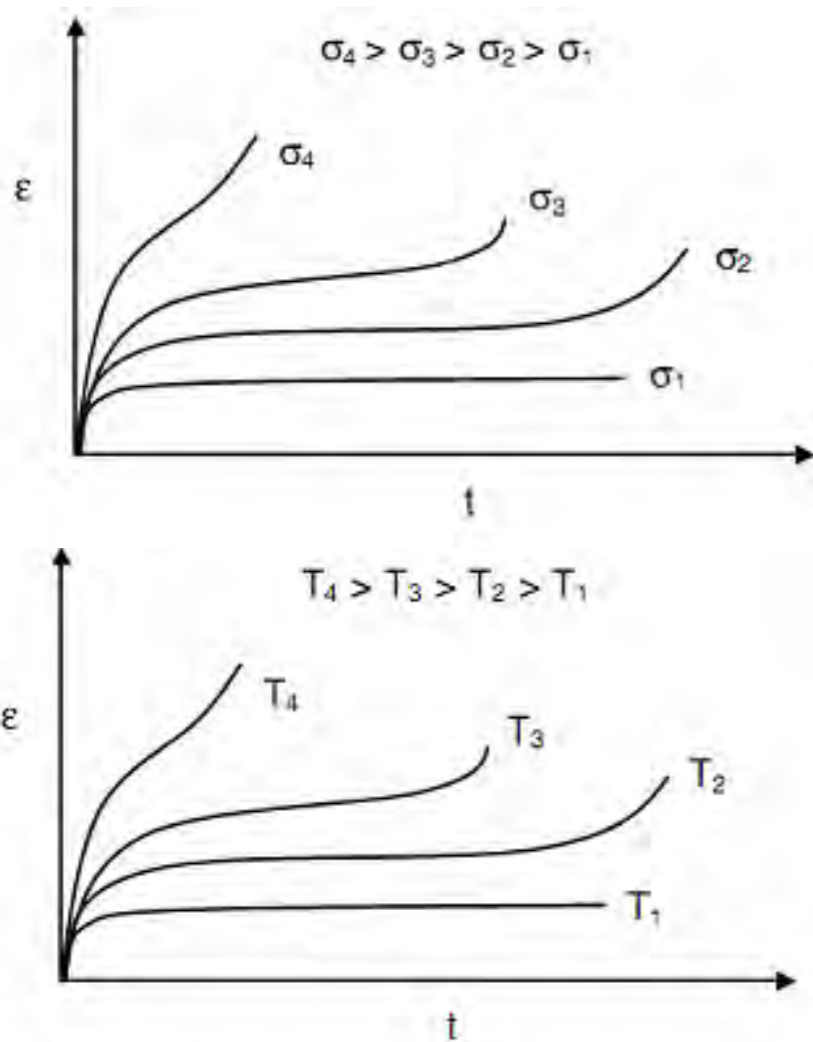
[出典]Proc.ICONE-7,ICONE-7255,Tokyo,Japan,2005

<http://fykai.blog.fc2.com/blog-entry-140.html>

クリープ構成式によるアプローチ

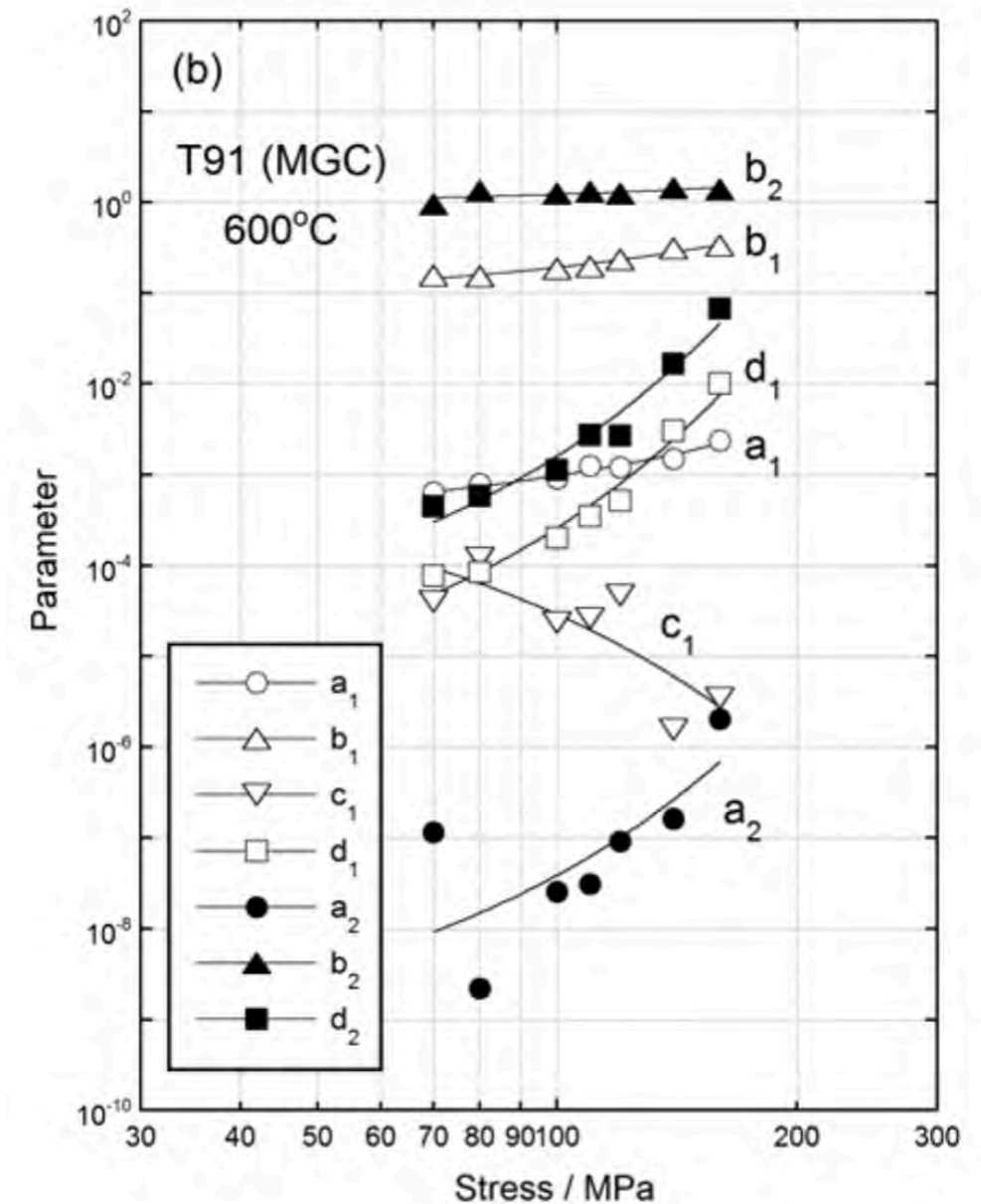
1. クリープ曲線を構成式で最小二乗回帰

$$\epsilon_{kim}(t; \theta_{kim}) = \epsilon_0 + a_1 t^{b_1} + a_2 t^{b_2} + c_1 \exp(d_1 t) + c_2 \exp(d_2 t)$$



最小二乗法 ≡ 測定ノイズは考慮していない
(そもそも計測系の測定ノイズは不明)

2. 回帰パラメータの温度・応力依存性を回帰



[Kimura et al. 2014, 2015]

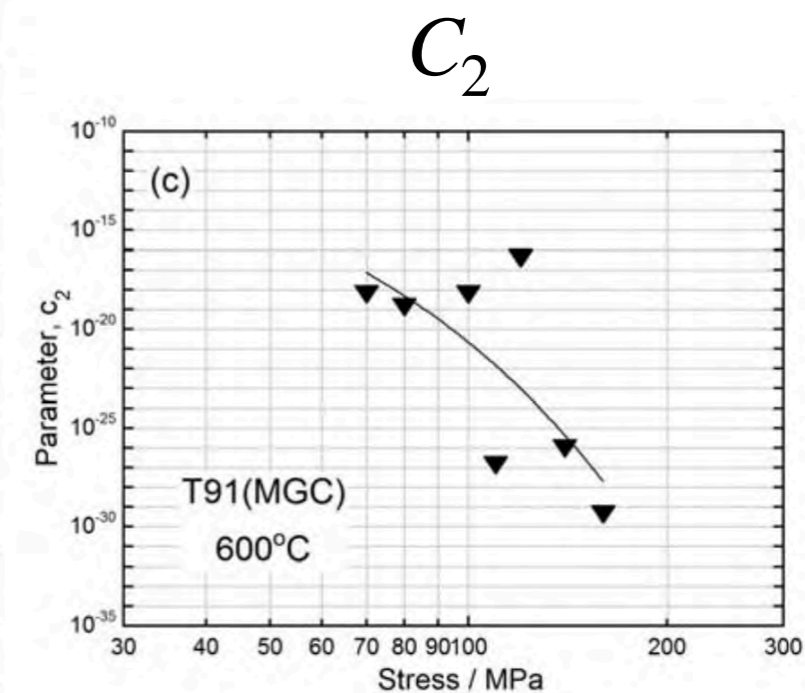
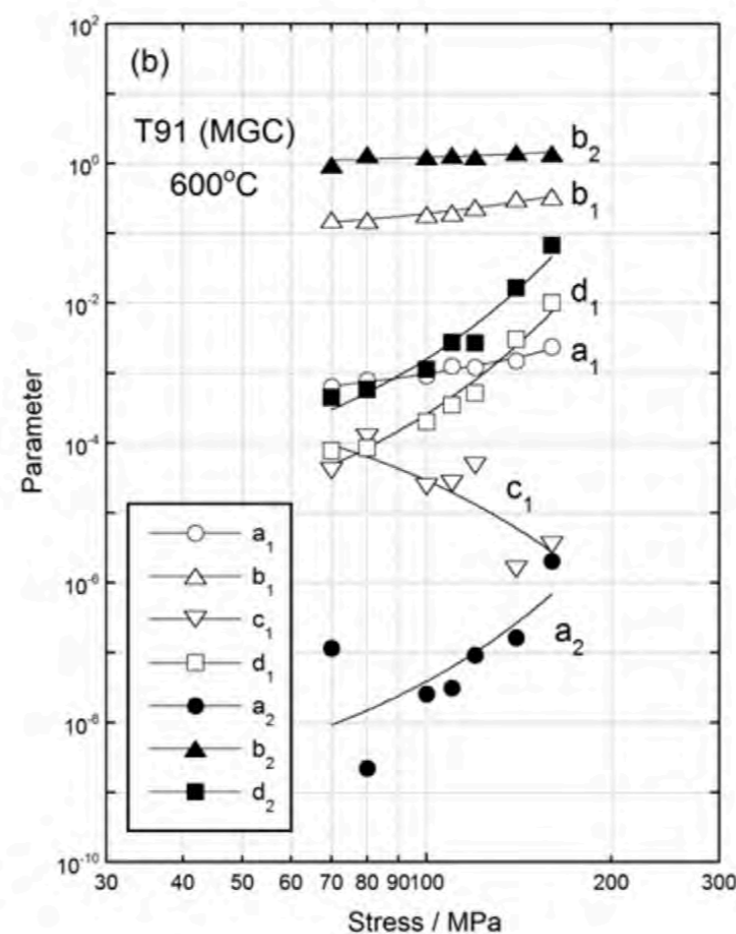
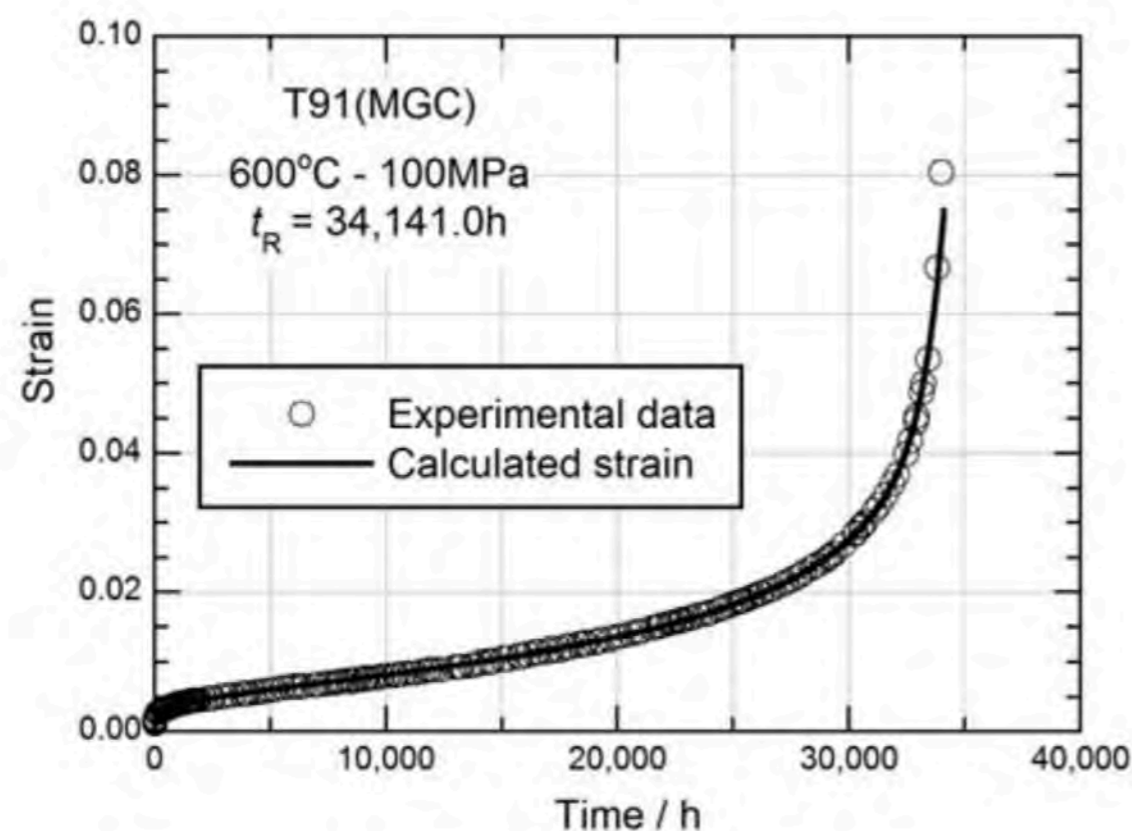
木村モデル

木村モデル：state of the artsなモデル

$$\epsilon_{\text{kim}}(t; \theta_{\text{kim}}) = \epsilon_0 + a_1 t^{b_1} + a_2 t^{b_2} + c_1 \exp(d_1 t) + c_2 \exp(d_2 t)$$

[Kimura et al. 2014, 2015]

b_2 はほぼ1?



[Kimura et al. 2014]

木村モデル

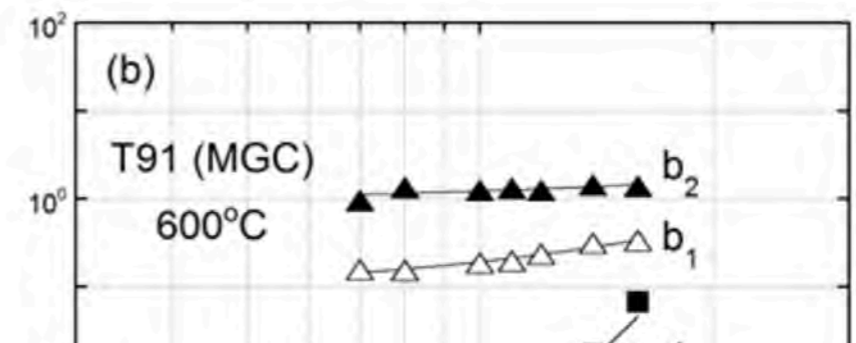
木村モデル：state of the artsなモデル

$$\epsilon_{\text{kim}}(t; \theta_{\text{kim}}) = \epsilon_0 + a_1 t^{b_1} + a_2 t^{b_2} + c_1 \exp(d_1 t) + c_2 \exp(d_2 t)$$

$$\theta_{\text{kim}} = \{a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2\}$$

[Kimura et al. 2014, 2015]

モデル比較



b_2 はほぼ1?

修正木村モデル

$$\epsilon_{\text{mod}}(t; \theta_{\text{mod}}) = \epsilon_0 + a_1 t^{b_1} + a_2 t + c_1 \exp(d_1 t) + c_2 \exp(d_2 t)$$

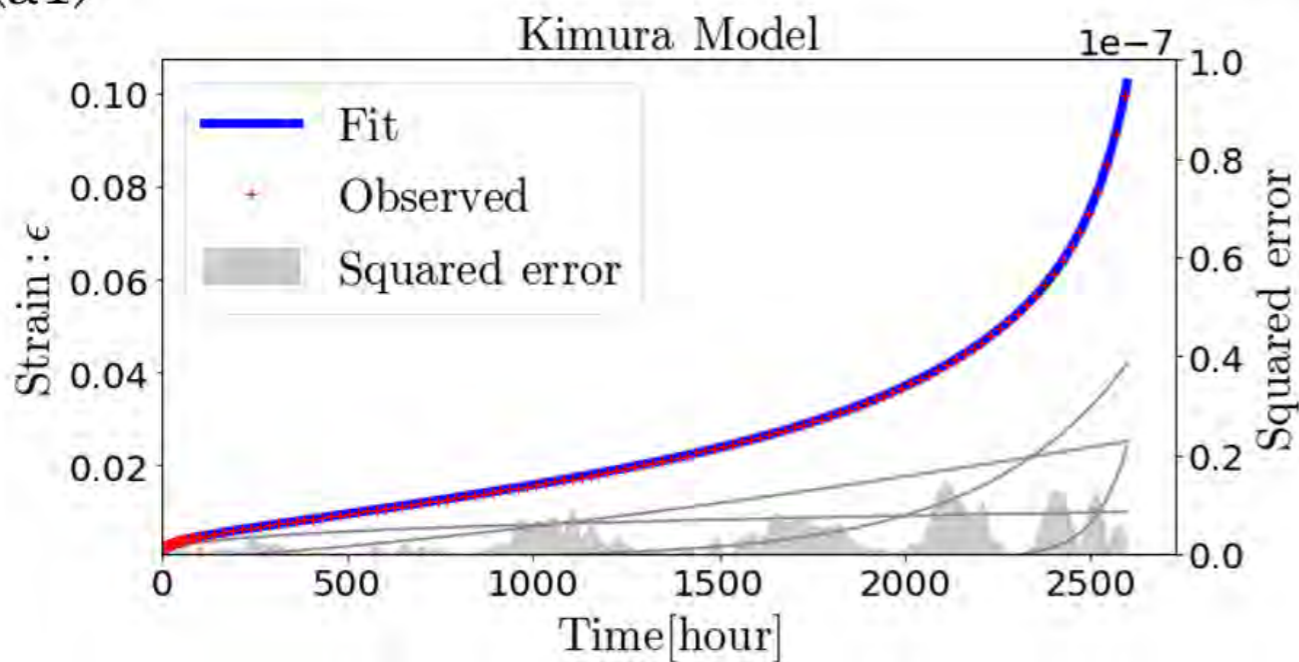
$$\theta_{\text{mod}} = \{a_1, a_2, b_1, c_1, c_2, d_1, d_2\}$$

提案モデルによる回帰結果

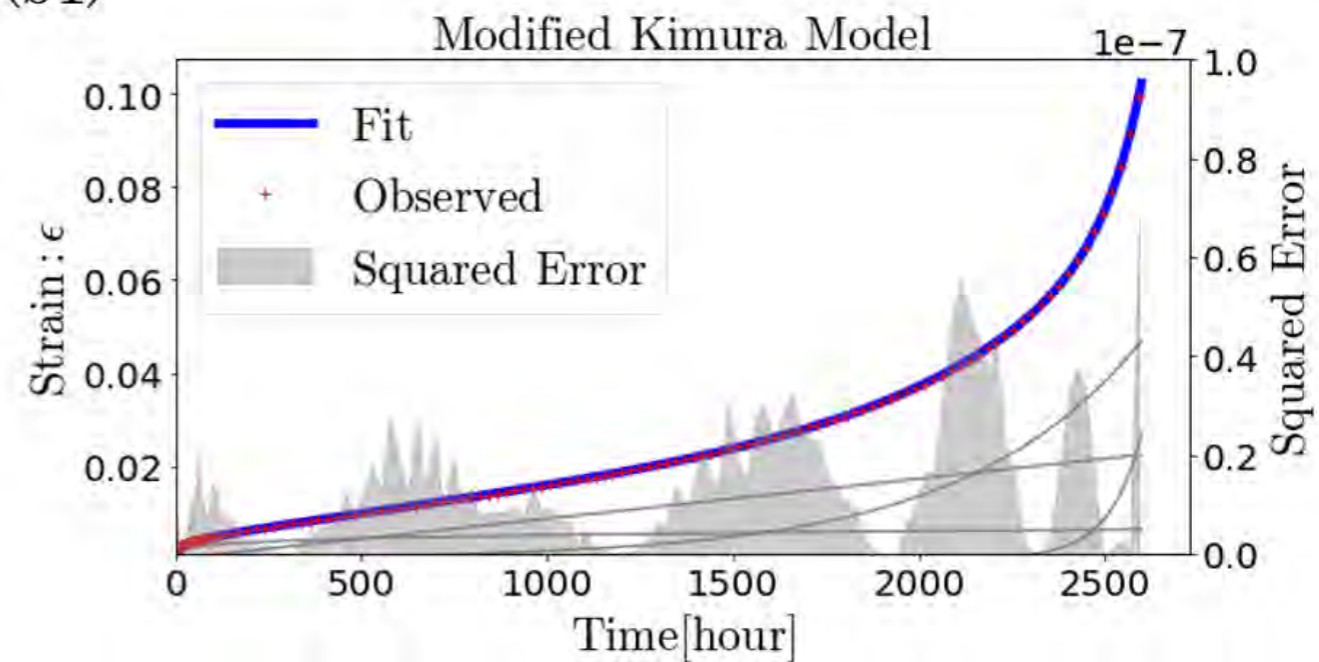
木村モデル

修正木村モデル

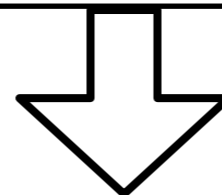
(a1)



(b1)



どちらも高精度の回帰結果かつ、わずかな差しかない
→測定誤差の値が重要になる



しかし、測定誤差はレンジでしかわからない
→測定誤差を確率変数としたベイズ推論

ベイズ推論に期待すること

利点1：データにない科学者の知識を事前分布として導入可能

➡レンジで与えられた測定誤差の情報をモデル評価に取り込む

利点2：事後分布から物理量の推定精度や物理モデルの性質がわかる

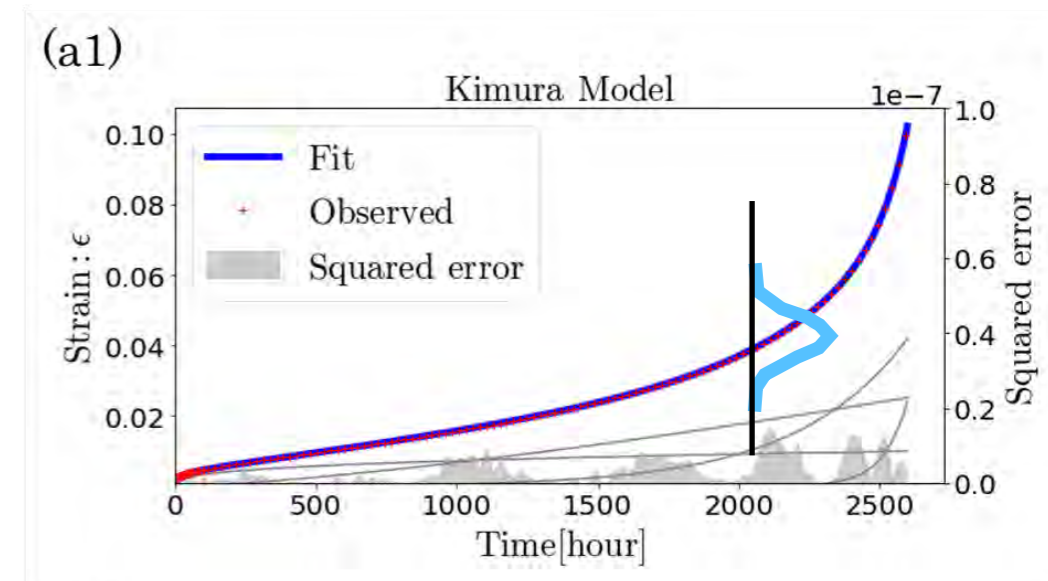
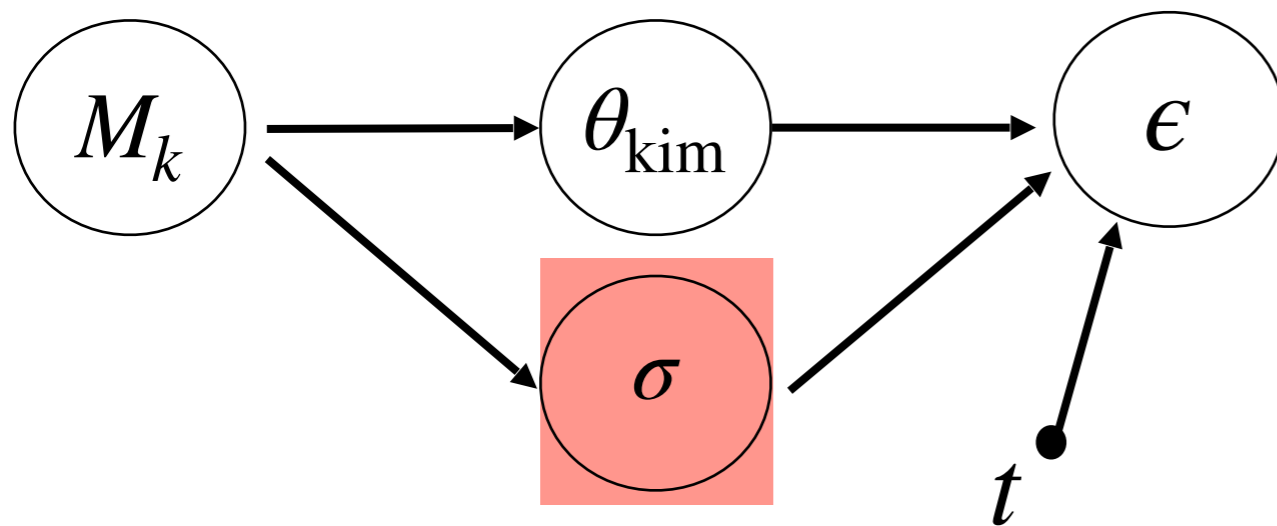
➡木村モデルでなぜ C_2 の温度・応力依存性が複雑化するのか解明する

利点3：過適合を考慮したモデル評価が可能

➡修正木村モデルが木村モデルより良い物理モデルか比較する

ベイズ的クリープモデル選択

$$\epsilon_{\text{kim}}(t; \theta_{\text{kim}}) = \epsilon_0 + a_1 t^{b_1} + a_2 t^{b_2} + c_1 \exp(d_1 t) + c_2 \exp(d_2 t)$$



【知っていること】
因果則 (順モデル)

$$P(\epsilon_1, \epsilon_2, \dots, \epsilon_N | \theta_{\text{kim}}, \sigma, M_{\text{kimu}}) = \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{N}{2}} \exp \left[- \frac{\sum_{i=1}^N [\epsilon_i - \epsilon_{\text{kim}}(t_i; \theta_{\text{kim}})]^2}{2\sigma^2} \right]$$

【知りたいこと】
 θ_{kimu} の事後分布

$$P(\theta_{\text{kimu}}, \sigma | \epsilon_1, \epsilon_2, \dots, \epsilon_N, M_{\text{kimu}}) = \frac{P(\epsilon_1, \epsilon_2, \dots, \epsilon_N | \theta_{\text{kimu}}, \sigma, M_{\text{kimu}}) P(\theta_{\text{kimu}} | M_{\text{kimu}}) P(\sigma | M_{\text{kimu}})}{P(\epsilon_1, \epsilon_2, \dots, \epsilon_N)}$$

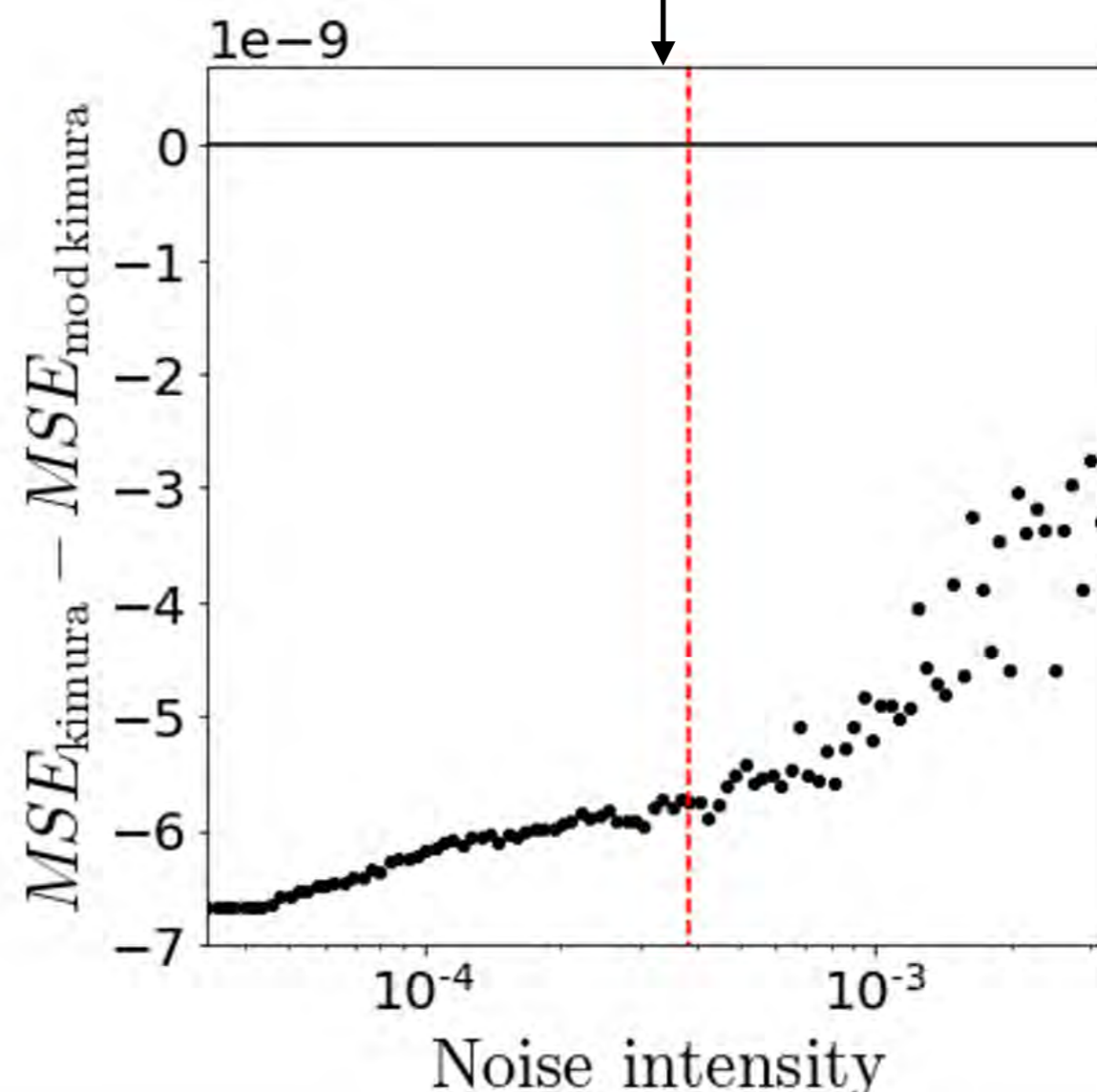
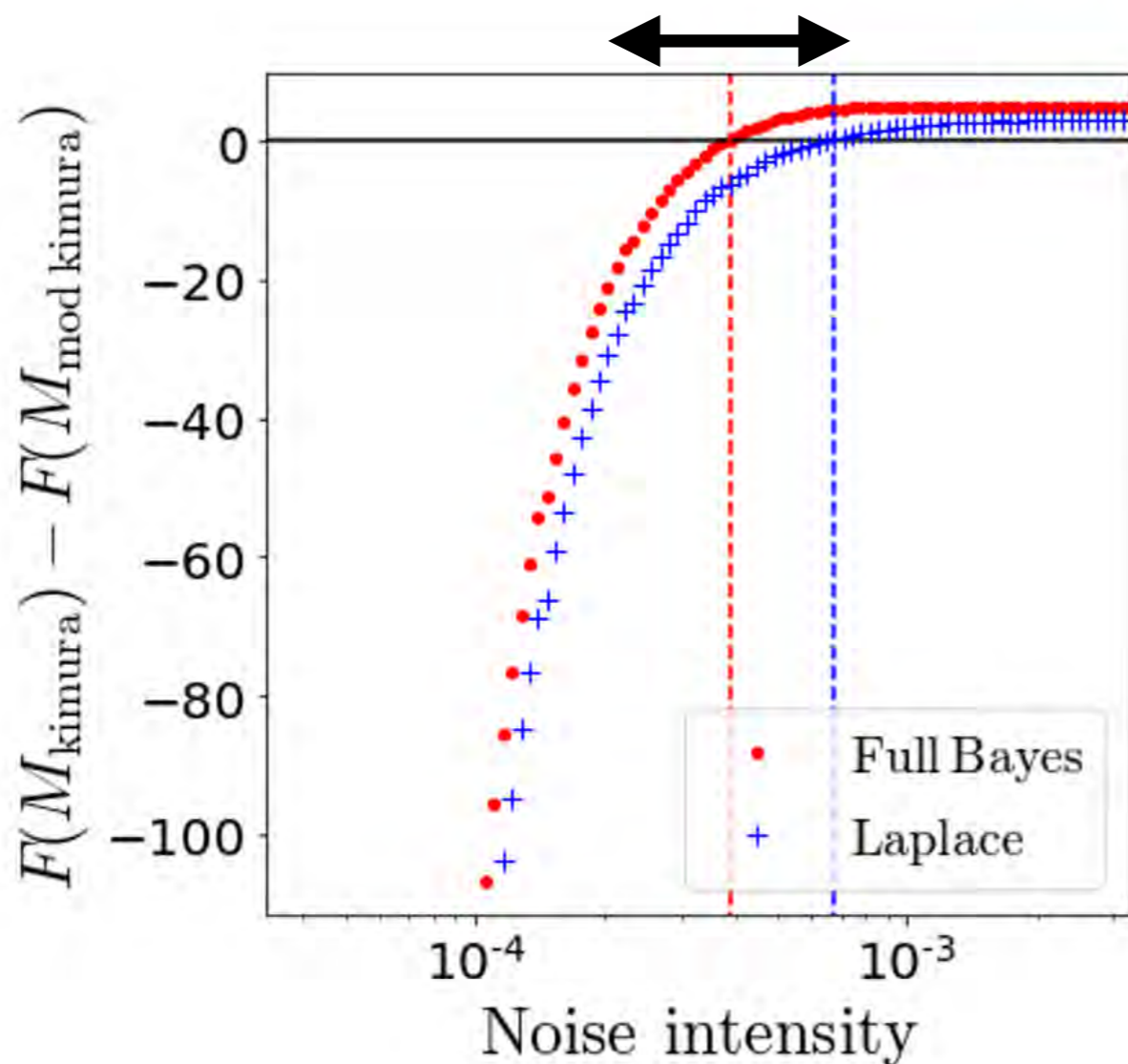
【知りたいこと】
モデルの事後分布

$$P(M_{\text{kimu}} | \epsilon_1, \epsilon_2, \dots, \epsilon_N) = \int d\theta_{\text{kimu}} d\sigma \frac{P(\epsilon_1, \epsilon_2, \dots, \epsilon_N | \theta_{\text{kimu}}, \sigma, M_k) P(\theta_{\text{kimu}} | M_k) P(\sigma | M_k) P(M_k)}{P(y_1, y_2, \dots, y_N)}$$

σ の事前分布とモデル選択結果

ゲージ目盛りから見積もられた
おおよその測定誤差レンジ

σ の事前分布によらず誤差は木村モデルが常に小さい
→誤差の大小ではないモデル比較が実現



- 測定誤差の事前知識に応じて選択されるモデルが切り替わる
- モデルの正確な評価には測定誤差の正確な見積もりが必要

事後分布の分析

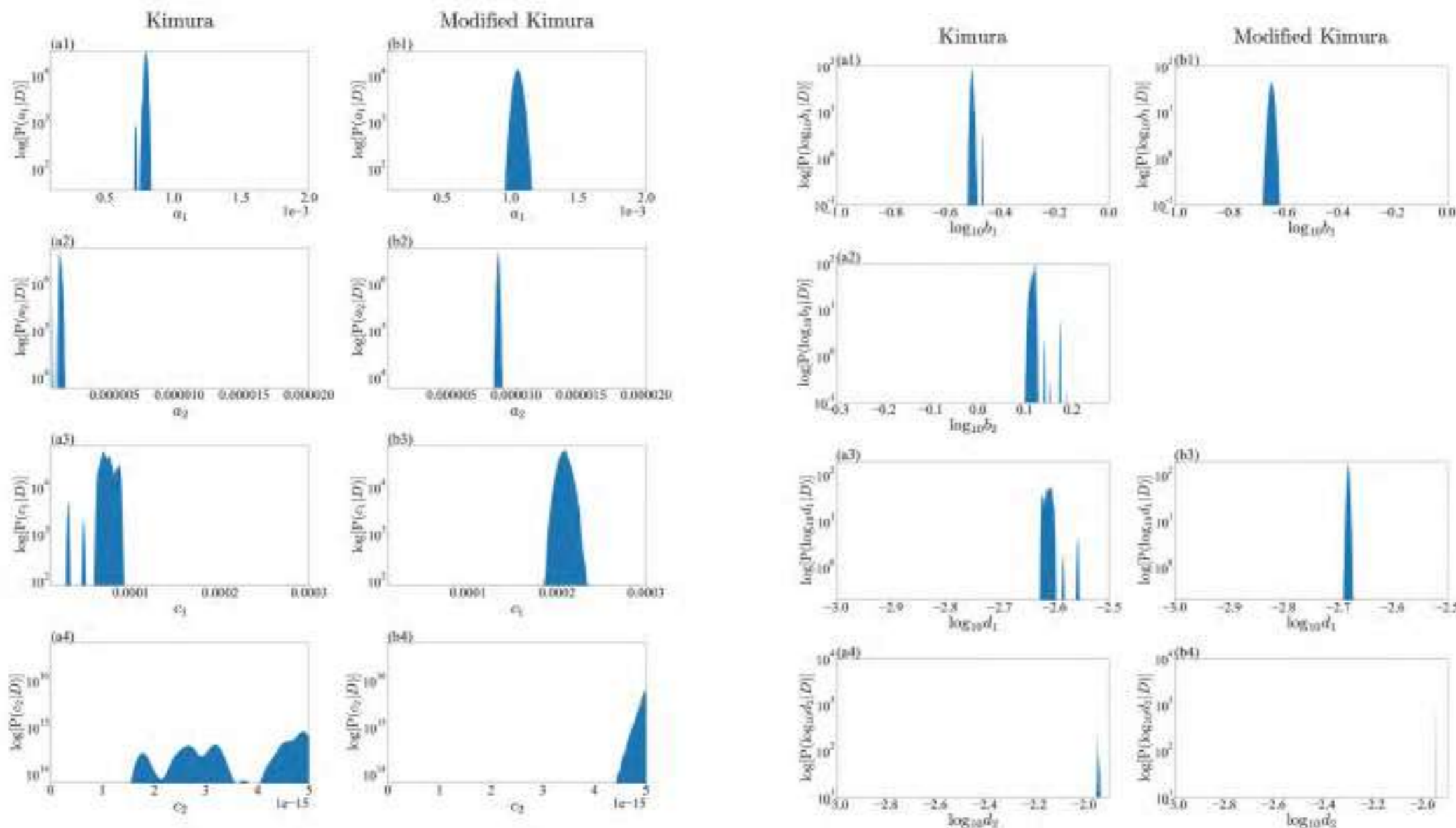
$$\epsilon_{\text{kim}}(t; \theta_{\text{kim}}) = \epsilon_0 + a_1 t^{b_1} + a_2 t^{b_2} + c_1 \exp(d_1 t) + c_2 \exp(d_2 t)$$

木村モデル

修正木村モデル

木村モデル

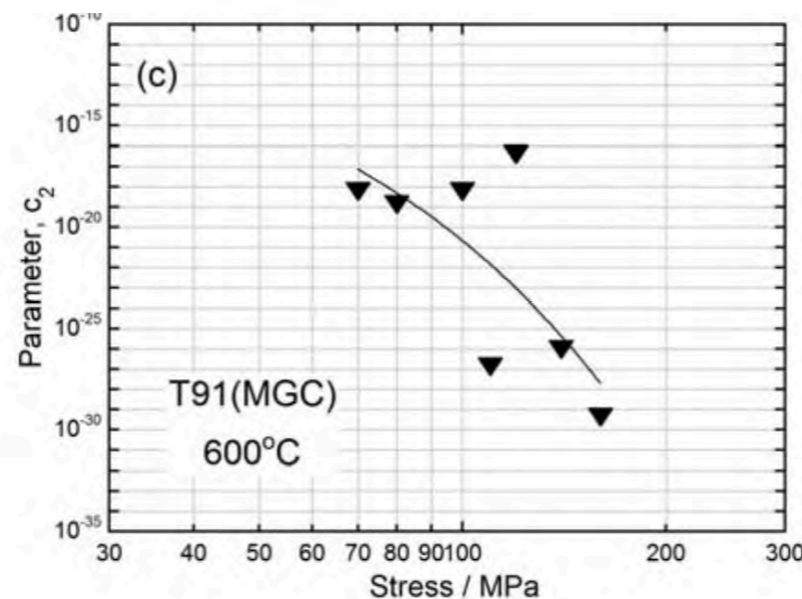
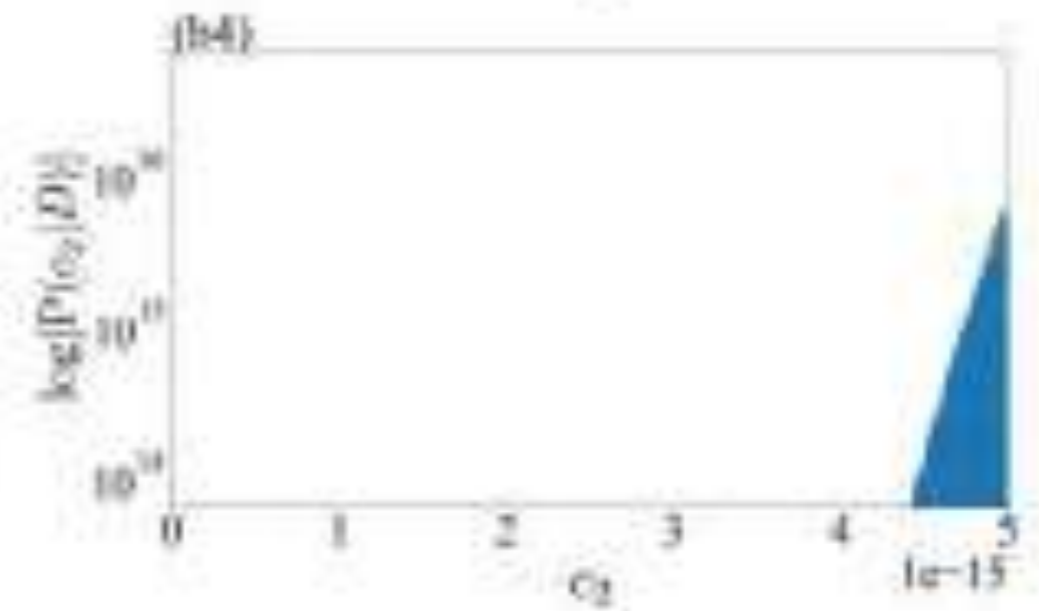
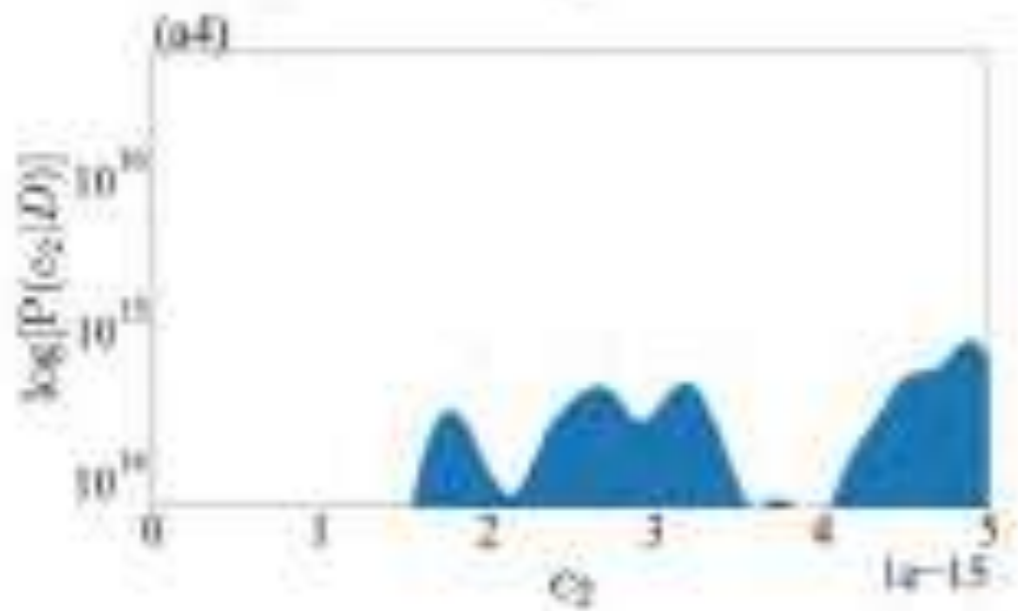
修正木村モデル



→事後分布は、木村モデルが過学習を行っていることを示唆

C_2 の温度応力依存性が複雑な要因

$$\epsilon_{\text{kim}}(t; \theta_{\text{kim}}) = \epsilon_0 + a_1 t^{b_1} + a_2 t^{b_2} + c_1 \exp(d_1 t) + c_2 \exp(d_2 t)$$

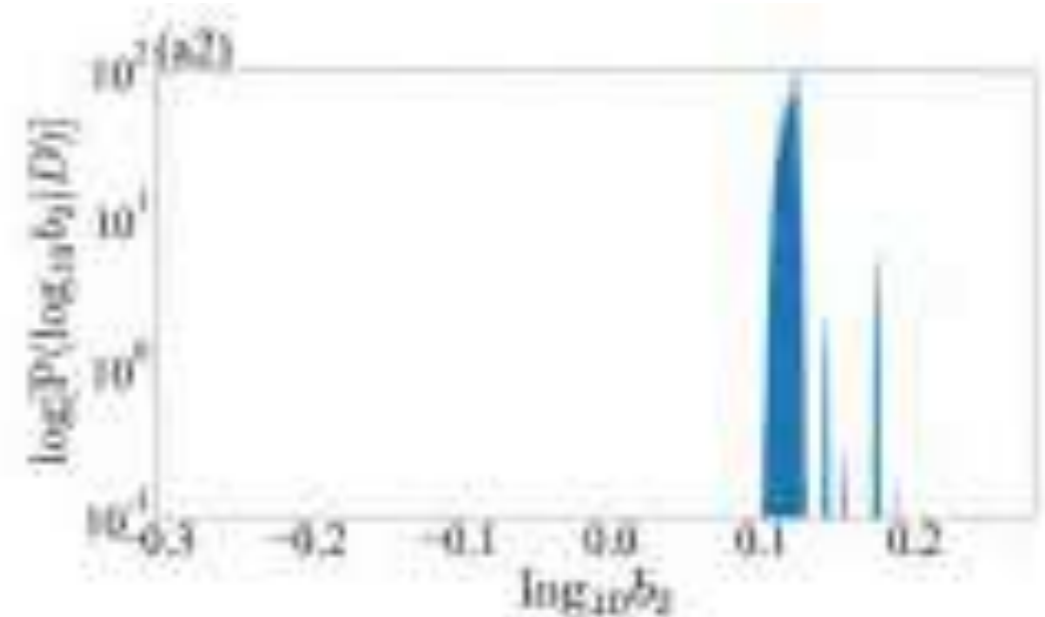
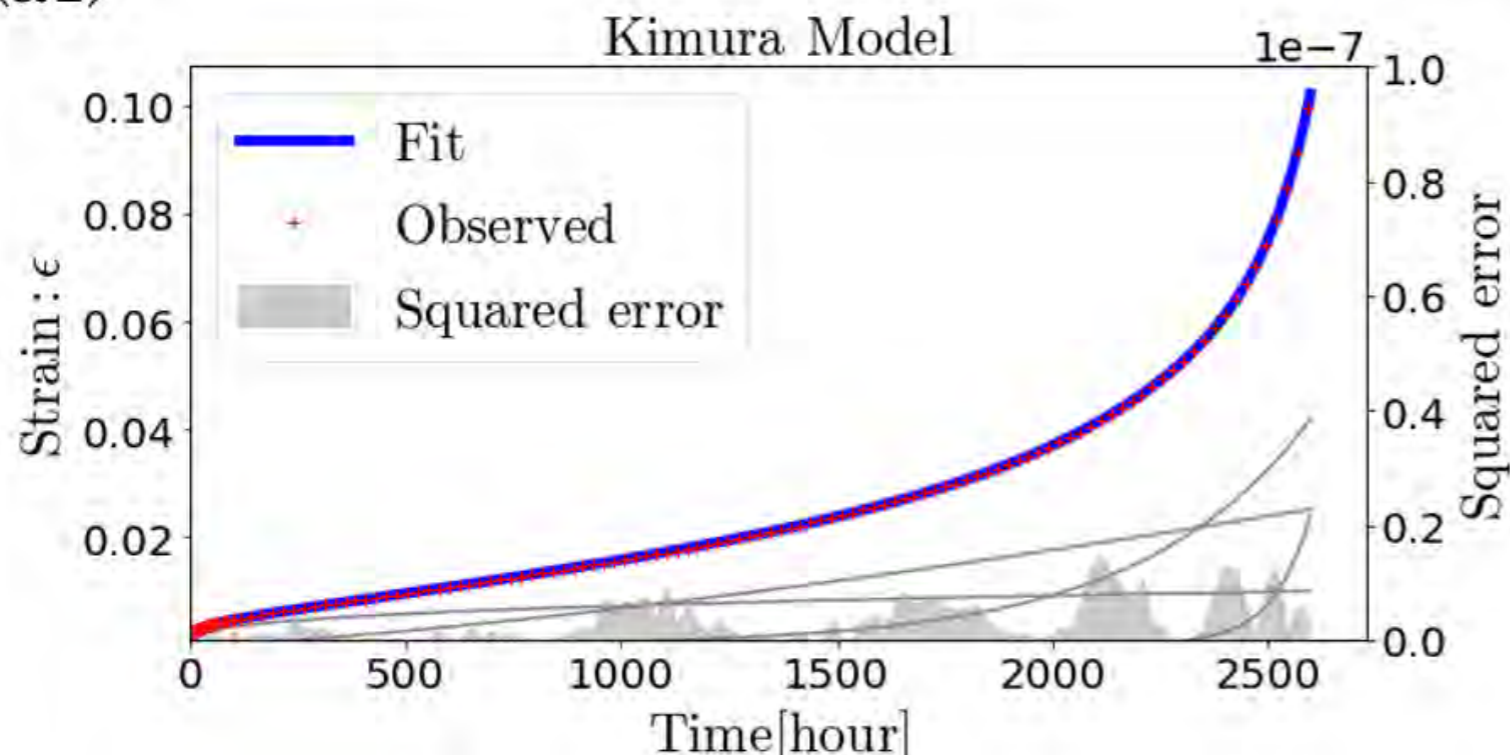


→修正木村モデルが温度・応力方向の推定において、木村モデルを改善できそう

C_2 の温度応力依存性が複雑な要因

$$\epsilon_{\text{kim}}(t; \theta_{\text{kim}}) = \epsilon_0 + a_1 t^{b_1} + a_2 t^{b_2} + c_1 \exp(d_1 t) + c_2 \exp(d_2 t)$$

(a1)



第2項の冪が0より大きくても、2次クリープ領域でほぼ線形になる

➡加速クリープ領域を3つの基底で回帰した結果過学習している可能性

5. まとめ

まとめ

- ベイズ推論は、科学者の知的活動を自然にモデル化する。
- ベイズ推論には、科学者にとって有用な特性がある。
- ベイズ推論は我々が現象や物理モデルを理解することを助けてくれるツールとなる。

一方で、、、

$$P(\theta_{\text{kimu}}, \sigma | \epsilon_1, \epsilon_2, \dots, \epsilon_N, M_{\text{kimu}}) = \frac{P(\epsilon_1, \epsilon_2, \dots, \epsilon_N | \theta_{\text{kimu}}, \sigma, M_{\text{kimu}})P(\theta_{\text{kimu}} | M_{\text{kimu}})P(\sigma | M_{\text{kimu}})}{P(\epsilon_1, \epsilon_2, \dots, \epsilon_N)} \quad (1)$$

$$P(M_{\text{kimu}} | \epsilon_1, \epsilon_2, \dots, \epsilon_N) = \int d\theta_{\text{kimu}} d\sigma \frac{P(\epsilon_1, \epsilon_2, \dots, \epsilon_N | \theta_{\text{kimu}}, \sigma, M_k)P(\theta_{\text{kimu}} | M_k)P(\sigma | M_k)P(M_k)}{P(y_1, y_2, \dots, y_N)} \quad (2)$$

複雑なモデルの場合、(1)事後分布の関数形を調べたり、(2)周辺化の積分を実行するのに**数値計算（モンテカルロサンプリング）**が必要になる。

➡**職人技的な技術が必要になる & 真実はサンプリングだけからは確定できない**

➡お気軽にご相談ください（@営業） mototake@ism.ac.jp