AAPPS-DPP2025
@Fukuoka, Japan
September 26, 2025

Application and progress of Bayesian statistics in plasma physics

Satoru Tokuda

Institute of Mathematics for Industry (IMI), Kyushu University

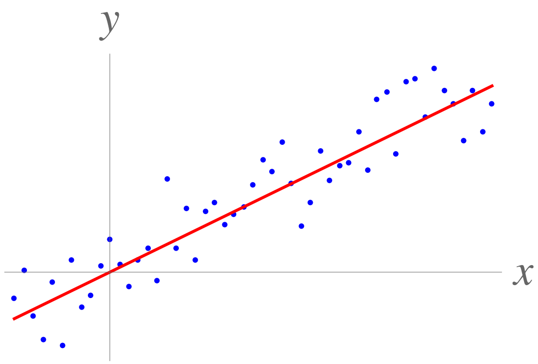
tokuda@imi.kyushu-u.ac.jp

Collaborators (affiliation at that time):
H. Arakawa (Shimane Univ.), K. Ida (NIFS), S. Inagaki (Kyushu Univ.), H. Uehara (NIFS), M. Okada (UTokyo), S. Kado (Kyoto Univ.) , Y. Kawachi (Kyushu Univ.), T. Kobayashi (NIFS), M. Sasaki (Nihon Univ.), A. Sanpei (KIT), H. Takahashi (NIFS), Y. Takemura (NIFS), K. Tanaka (NIFS), T. Tokuzawa (NIFS), T. Nishizawa (Kyushu Univ.), K. Terasaka (Kyushu Univ.), K. Nagata (UTokyo), M. Hasegawa (Kyushu Univ.), A. Fujisawa (Kyushu Univ.), H. Funaba (NIFS), R. Yasuhara (NIFS), K. Yamasaki (Hiroshima Univ.), I. Yamada (NIFS), G. Grenfell (MPI), P. Manz (Univ. of Greifswald)

Encounter of Statistics and Physics

Error theory

(since the 1800s)

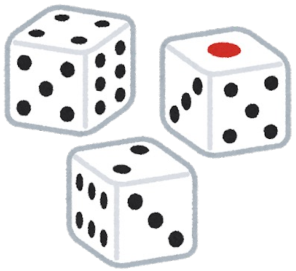


Sample (Data)

Observable

incomplete measurements

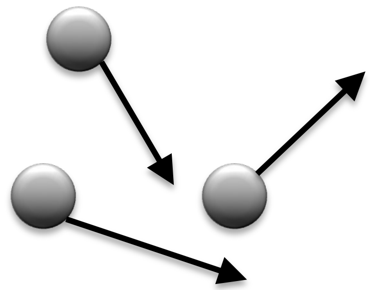
【Epistemology】



Random variable

Statistical mechanics

(since the 1850s)



State (Particle motion)

Unobservable

Invisible but existing truth

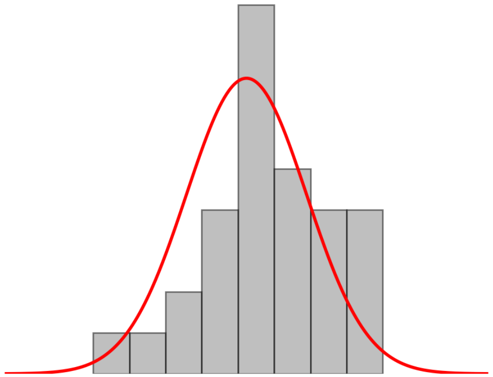
【Ontology】

Share the same concept, while the underlying philosophies are distinct.

Parting of Statistics and Physics

Statistical inference

(since the 1910s)

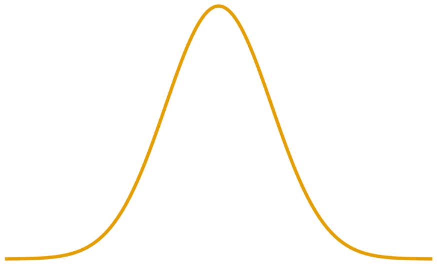


Density estimation

"Indirect" evaluation

Sample → Inference

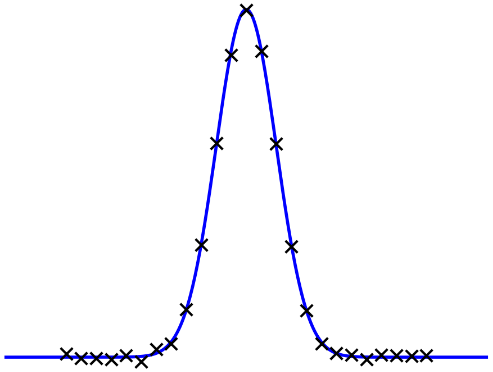
【Epistemology】



Probability density function (PDF)

Quantum mechanics

(since the 1920s)



Doppler spectroscopy

"Direct" measurement

Observation → Reality

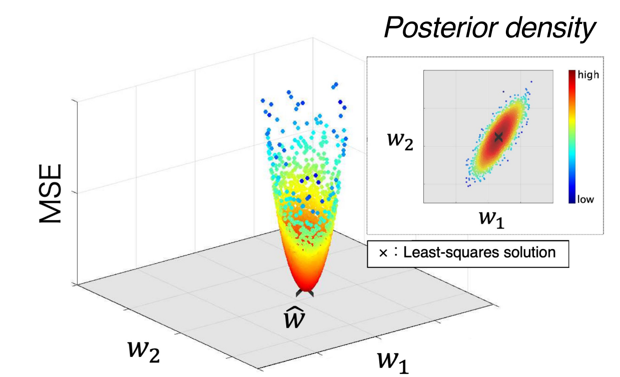
【Ontology】

Passed each other due to differences in direction in the 20th Century.

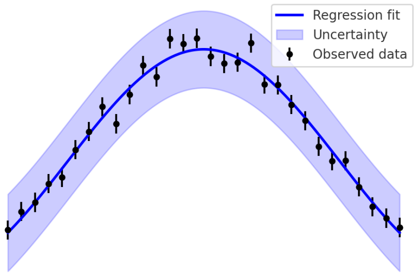
Reunion of Statistics and Physics

Bayesian statistics

(since the 1920s)



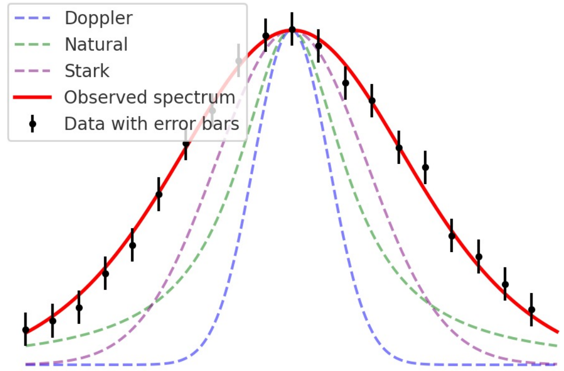
Uncertainty of parameters



Regression problems

Plasma physics

(since the 1920s)



broadenings & noise

Inverse problems with uncertainty

Direct problems of measurement

Sample & Direct problem → Inversion

【Epistemology】

Underlying function → Convolution

【Ontology】

Regression problems in hierarchical generative models are a meeting ground of the two fields — and also a new frontier.

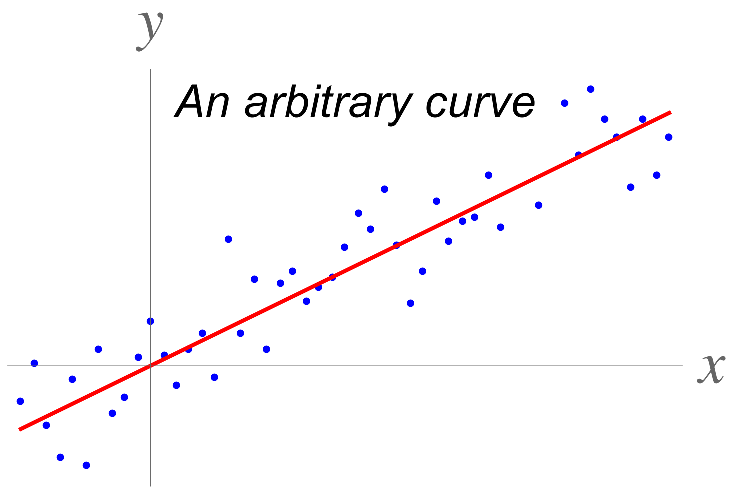
Preliminaries: Math of Bayesian regression

– On parametric regression –

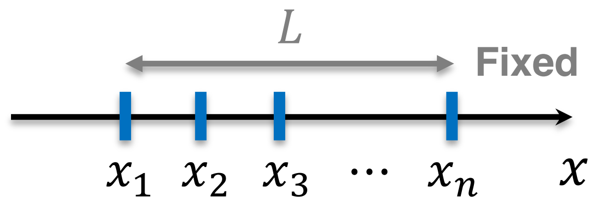
Setup: Observations adding Gaussian noise

$$\boxed{Y_i = f(x_i; w) + N_i} \quad (i = 1, 2, \cdots, n)$$

- $\{Y_i\}_{1:n}$: Sample
  $Y_i = y_i$: Realization (datum)
- $x = x_i$: Observation point
  $x_i = x_{i-1} + L/(n-1)$
  ($L \coloneqq x_n - x_1$; Fixed interval)
- $f \in \{f_K\}$: Arbitrary function of x
- w: Parameter set of f
- N_i: Observation noise
  $N_i \sim \mathcal{N}(0, \beta^{-1})$ *Gaussian noise*
  i.e., $Y_i \sim \mathcal{N}(f(x_i; w), \beta^{-1})$
  $\beta^{-1} > 0$: Noise level (variance)



$Y_i | x_i$: **conditionally independent**

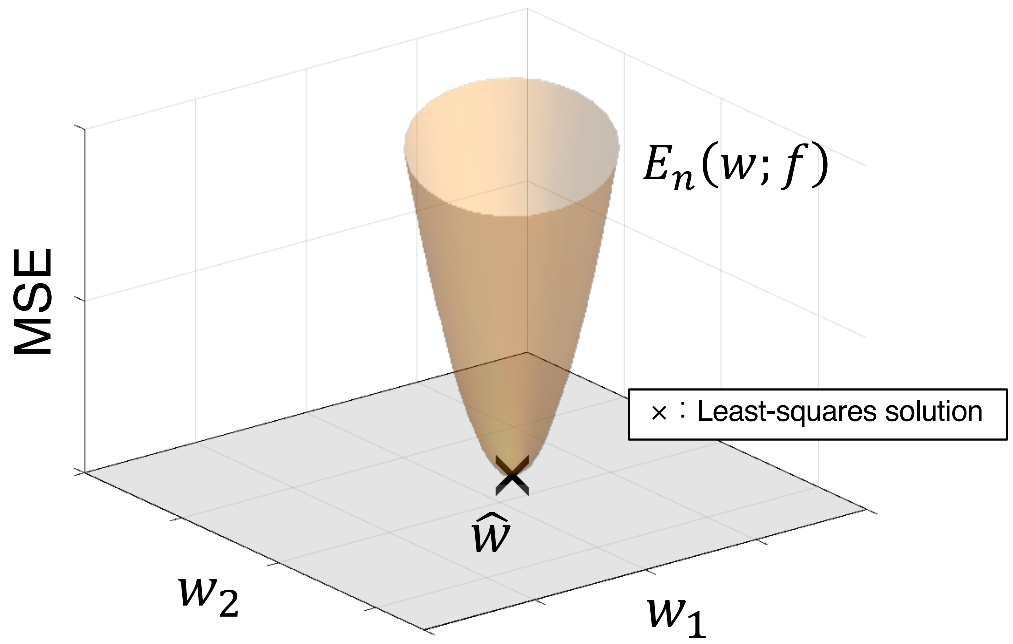


Data are realizations of random variables following a Gaussian distribution

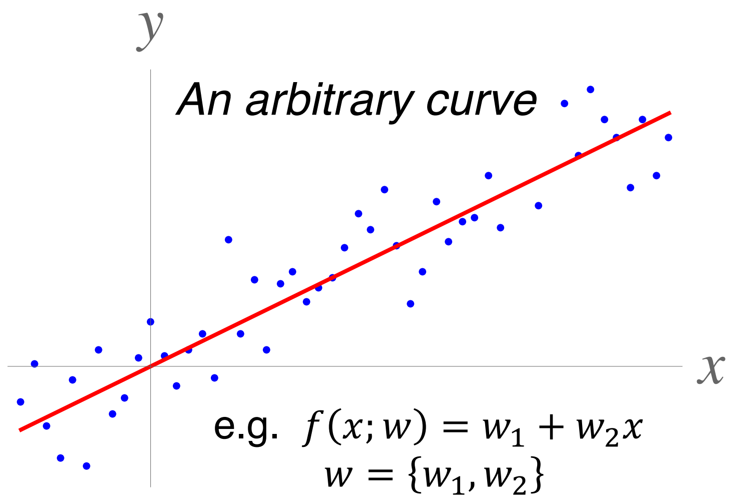
Revisiting ordinary least squares (OLS)

$$E_n(w;f) := \frac{1}{n}\sum_{i=1}^{n}\{Y_i - f(x_i;w)\}^2$$

Mean squared error (MSE)



“Ground state” of parameters



$$\hat{w} := \underset{w}{\operatorname{argmin}}\, E_n(w;f)$$

A closed-form solution is available only for linear regression

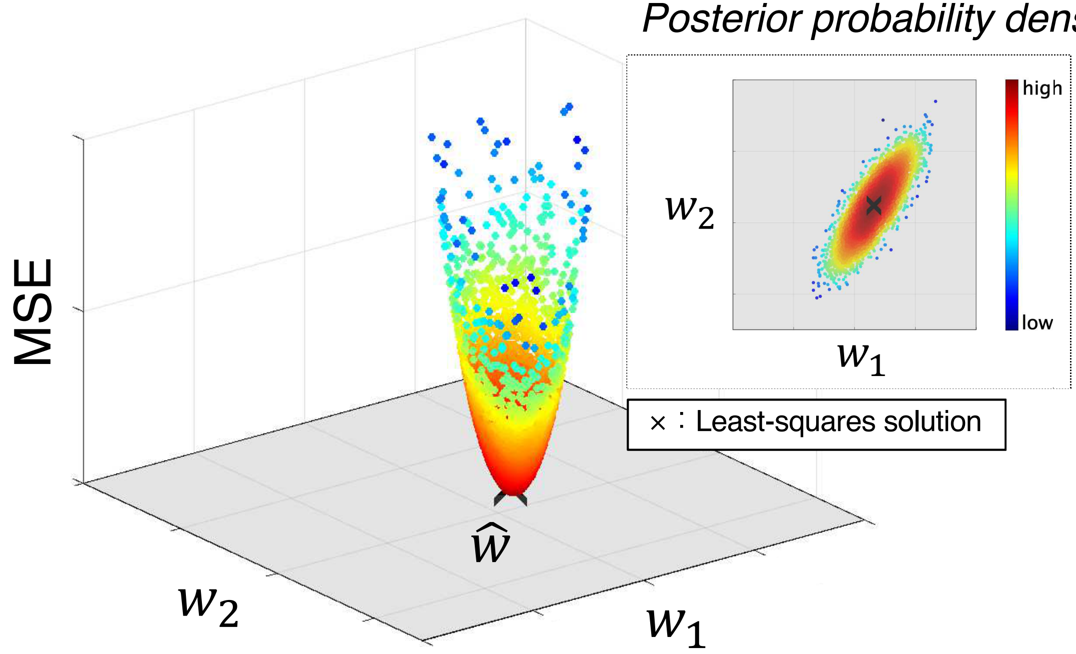
How certain are the “optimal” parameter values obtained?

Bayesian regression as an extension of OLS

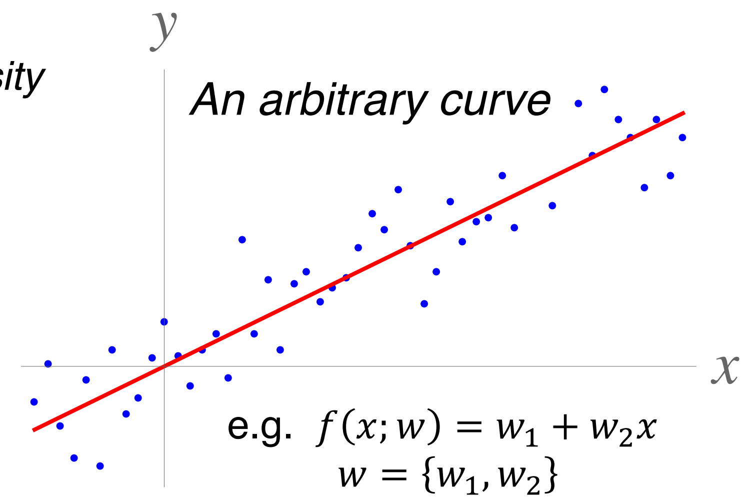
$$E_n(w;f) := \frac{1}{n}\sum_{i=1}^{n}\{Y_i - f(x_i;w)\}^2$$

Mean squared error (MSE)

Posterior probability density



"Statistical ensemble" of parameters



$$\hat{w} := \underset{w}{\operatorname{argmin}}\, E_n(w;f)$$

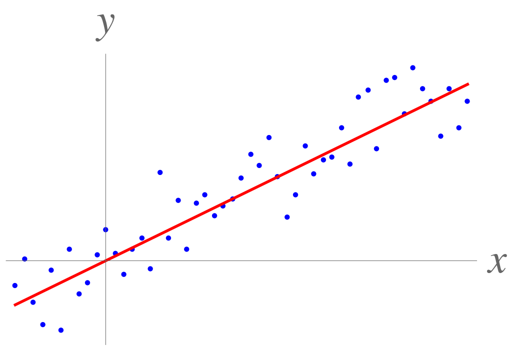
$$w \sim p(w|\{x_i, Y_i\}_{1:n}, f)$$ Posterior distribution

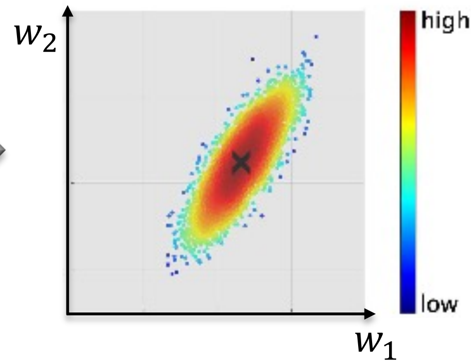
Regard parameters as random variables

Uncertainty quantification of distributed parameter values

Formulation based on a probabilistic model

Chain rule $\quad p(\{x_i, Y_i\}_{1:n}|w)p(w) = p(w|\{x_i, Y_i\}_{1:n})p(\{x_i, Y_i\}_{1:n})$





Distribution of observed data

$$Y_i \sim p(Y_i|x_i, w)$$
$$= \mathcal{N}(f(x_i; w), \beta^{-1})$$

$f(x_i; w)$: Regression function

β^{-1}: Noise level (variance)

n: Sample size

(Conditional) independence

$$p(\{x_i, Y_i\}_{1:n}|w) = \prod_{i=1}^{n} p(Y_i|x_i, w)$$

+

Normality of observation noise

$$\prod_{i=1}^{n} p(Y_i|x_i, w) = \left(\frac{\beta}{2\pi}\right)^{\frac{n}{2}} \exp\left[-\frac{n\beta}{2} E_n(w; f)\right]$$

Distribution of parameter values

$$w \sim p(w|\{x_i, Y_i\}_{1:n})$$
$$\propto \exp\left[-\frac{n\beta}{2} E_n(w; f)\right] p(w)$$

$p(w)$: well-behaved function
[almost flat prior distribution]

$E_n(w; f)$: MSE

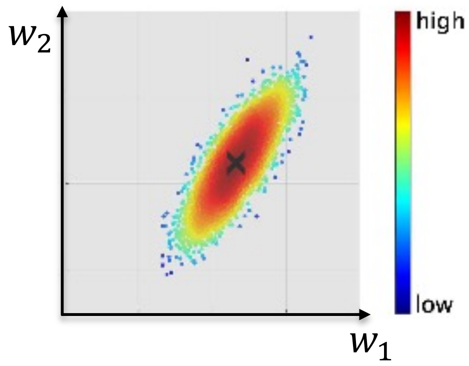
"Error propagation" from observed data to parameter values

From regression to model selection

Regression: To estimate w while $D^n \coloneqq \{x_i, Y_i\}_{1:n}$ and f are fixed

Distribution of parameter values

$$p(w|D^n, f) \propto p(D^n|w, f)p(w|f) \quad (*)$$



Likelihood

$$p(D^n|w, f) = \left(\frac{\beta}{2\pi}\right)^{\frac{n}{2}} \exp\left[-\frac{n\beta}{2} E_n(w; f)\right]$$

$p(w|f)$: well-behaved function $\quad E_n(w; f)$: MSE $\quad \beta^{-1}$: noise level

Marginalize w

$$p(f|D^n) = \int p(w|D^n, f)p(f)dw$$

Analytically intractable in many cases

Model selection: To estimate $f \in \{f_1, f_2, \cdots\}$ while D^n is fixed

Distribution of models

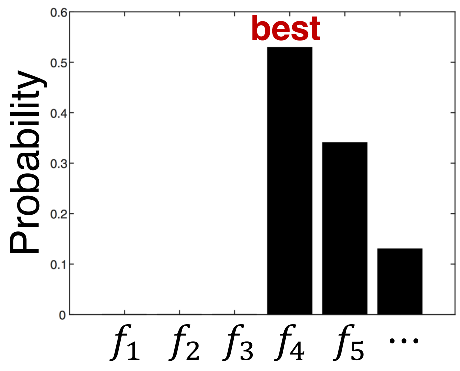
$$p(f|D^n) \propto p(D^n|f)p(f)$$

Marginal likelihood

$$p(D^n|f) = \int p(D^n|w, f)p(w|f)dw$$

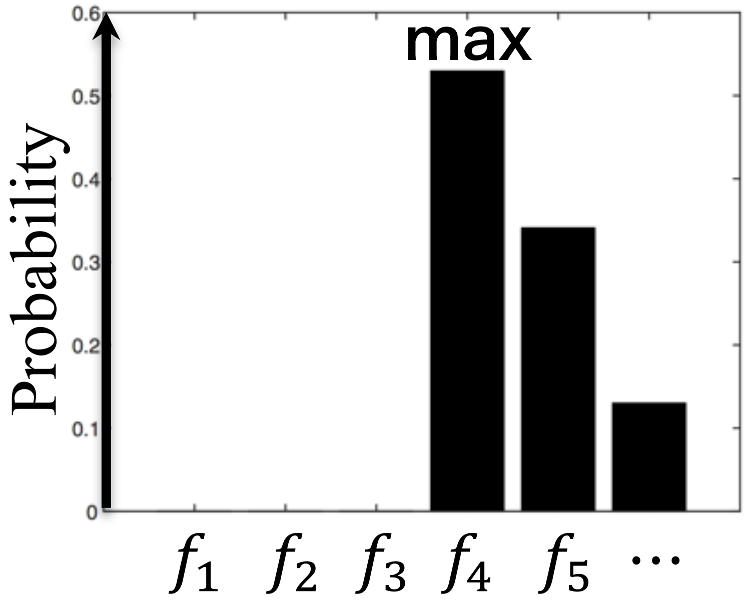
$p(f)$: unform distribution of $f \in \{f_K\}$

proportional const. of $(*)$
(normalizing)



Bayesian model selection (BMS)

$$p(f|D^n) \propto p(D^n|f)p(f)$$



D^n: Observed data

f: Model

$p(f)$: Unform distribution of $f \in \{f_1, f_2, \cdots\}$

"equally likely"

The maximizer of $p(f|D^n)$ is a valid model.
(**choose from candidates** with their uncertainties)

Probabilities of models outside the candidates are 0.

Model selection of ion velocity distributions

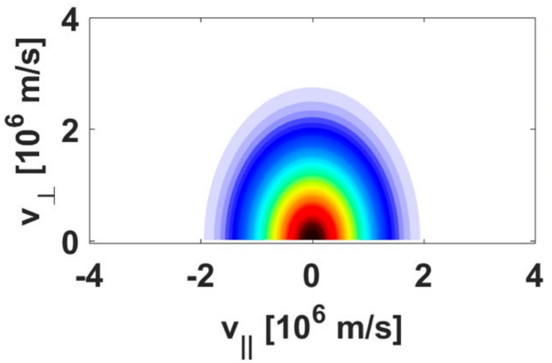
– Revealing the statistical nature of a linear magnetized plasma –

1. ST, et al, “Bayesian inference of ion velocity distribution function from laser-induced fluorescence spectra”, *Sci. Rep.*, 11, 20810 (2021).
    - Press release: ”ベイズ推定を用いた速度分布関数の解析法を開発” (in Japanese)
    - Grants: JSPS KAKENHI (20K19889), etc.

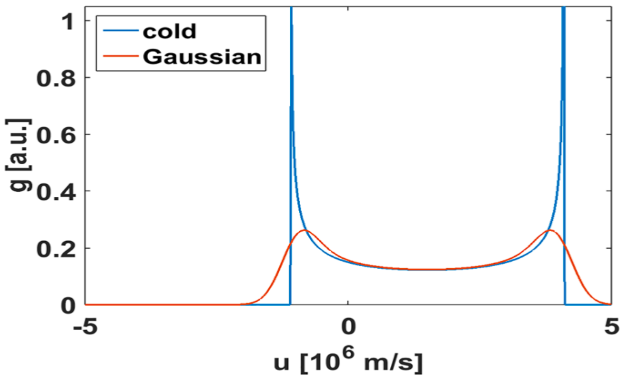
Ion velocity distribution in magnetized plasmas

[Moseev & Salewski, *Phys. Plasmas*, 2019]

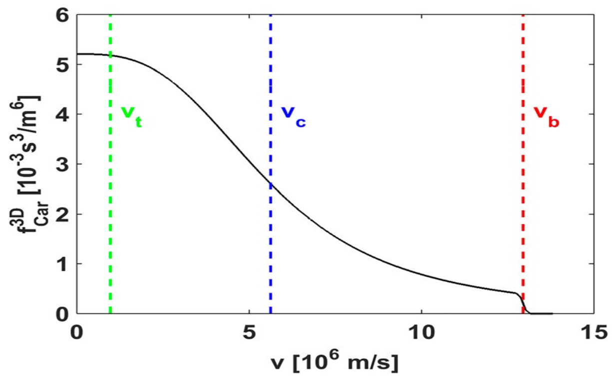
bi-Maxwellian



cold/warm ring



slowing-down



Electrons

$T_{e\,\|} \neq T_{e,\perp}$

Ions

$T_{i,\|} \neq T_{i,\perp}$

Electrons

Ions

Fast ions

Electrons

Ions

T_e

Fast ions

T_i

The velocity distribution function reflects the macroscopic state, while determining it experimentally is a nontrivial problem.

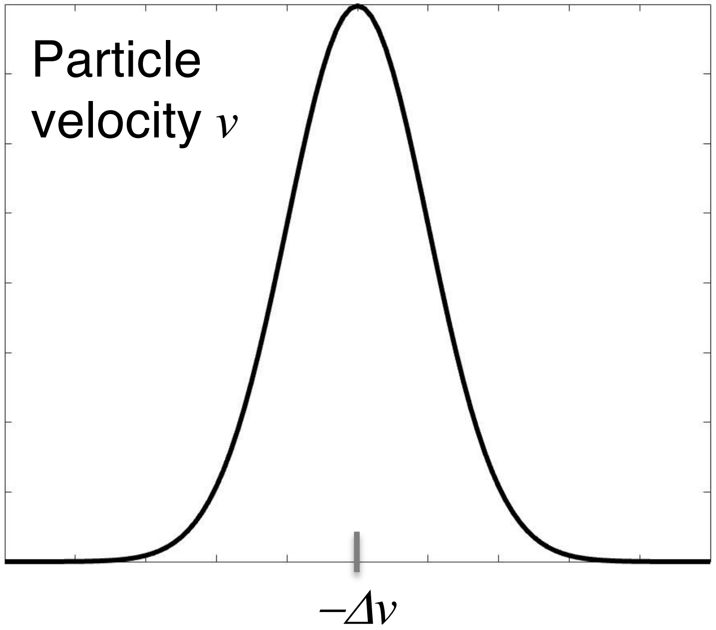
(It is often "*assumed*" that each species follows a bi-Maxwellian distribution.)

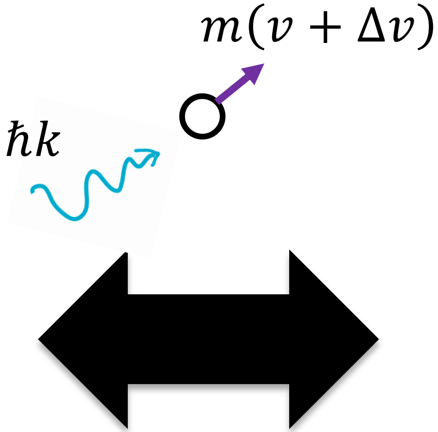
Observation via Doppler spectroscopy

E.g., Maxwellian distribution

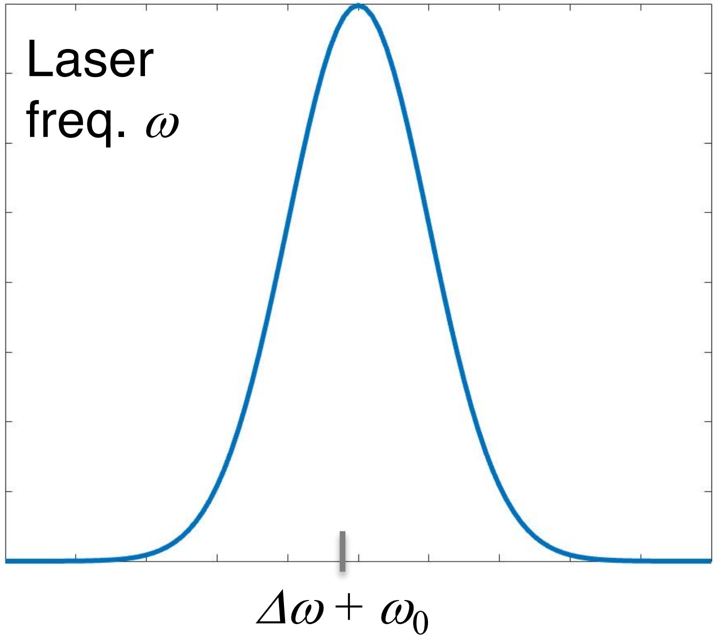
$$f(v) := n\sqrt{\frac{m}{2\pi k_B T}}\exp\left(-\frac{m(v+\Delta v)^2}{2k_B T}\right)$$

$$f(\omega) := n\sqrt{\frac{mc^2}{2\pi k_B T\omega_0^2}}\exp\left(-\frac{mc^2(\omega-\Delta\omega-\omega_0)^2}{2k_B T\omega_0^2}\right)$$





$$\omega = \omega_0\left(1+\frac{v+\Delta v}{c}\right)$$



Δv: Bulk fluid flow velocity
(0 if in thermal equilibrium)

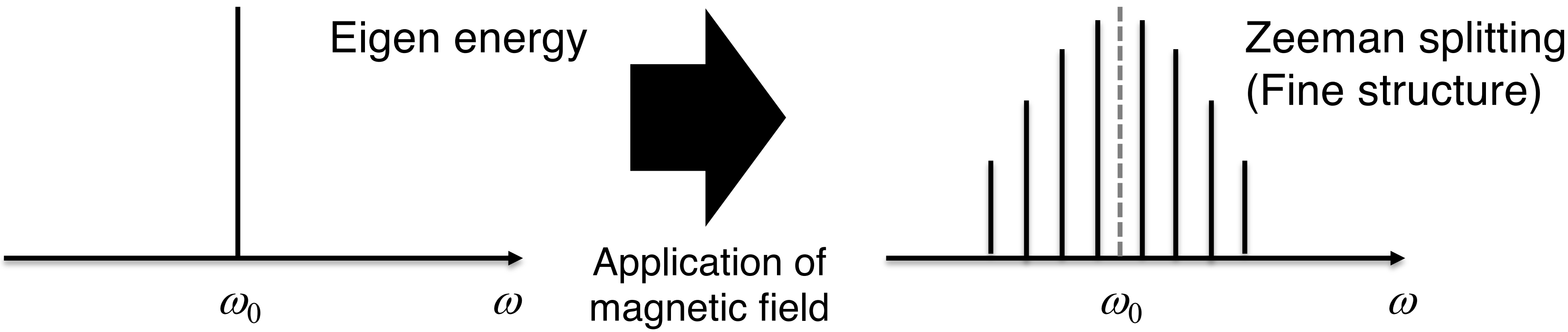
$\Delta\omega$: Frequency shift reflecting Δv
$\hbar\omega_0$: Eigen energy of the particle

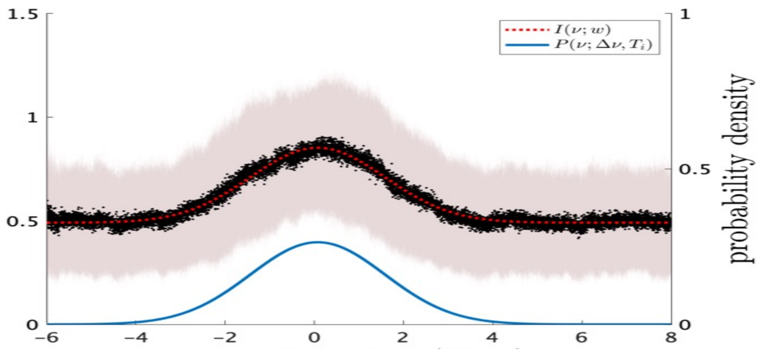
The velocity distribution function can be observed as absorption bands.

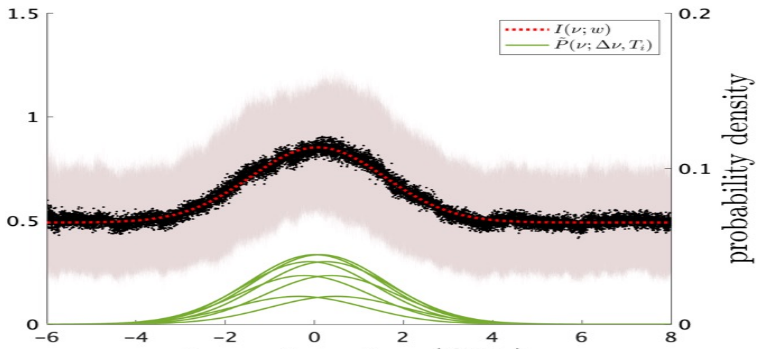
Spectral changes due to quantum effects

[Boivin & Scime, *Rev. Sci. Instrum.*, 2003]

E.g., Zeeman effect





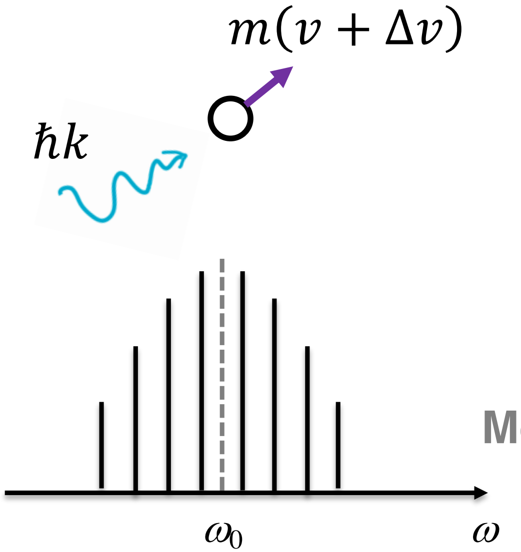


The spectral shape reflects not only the velocity distribution function, i.e., atomic motion, but also quantum effects, i.e., internal degrees of freedom.

(It is often *"determined"* by the order estimation whether such an effect is negligible.)

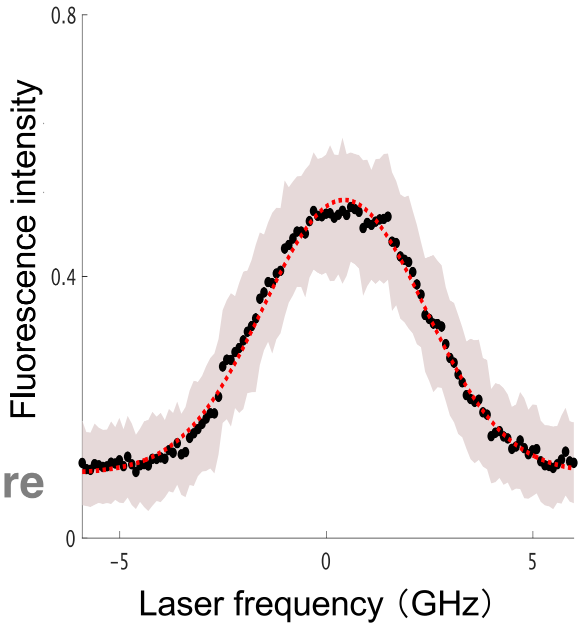
Realities in studies of magnetized plasmas

Experiment
(Doppler & Zeeman)



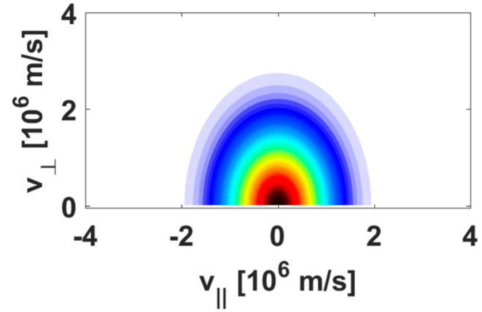
Measure

Observed data
(frequency spectra)



Interpret

Theory
(velocity distribution)



Electrons

$T_{e\,\parallel} \neq T_{e,\perp}$

Ions

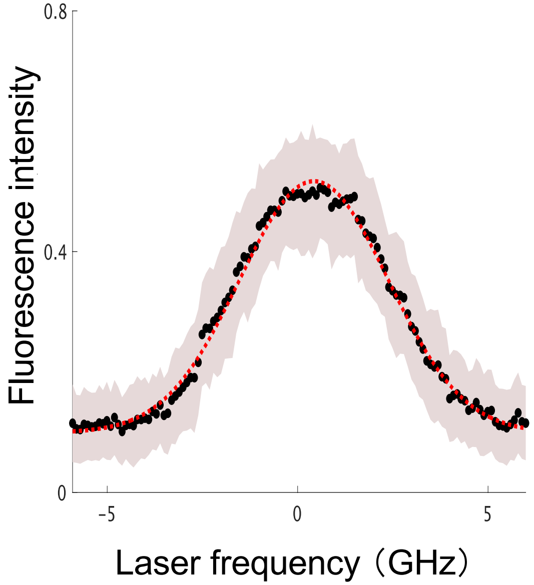
$T_{i,\parallel} \neq T_{i,\perp}$

Previous studies: Often assuming a Maxwellian distribution and fitting via OLS.

Our approach:
Selecting a valid model of observed data from among candidates, including non-Maxwellian distributions, also divided into those w/ and w/o quantum effects.

Model selection of ion velocity distributions

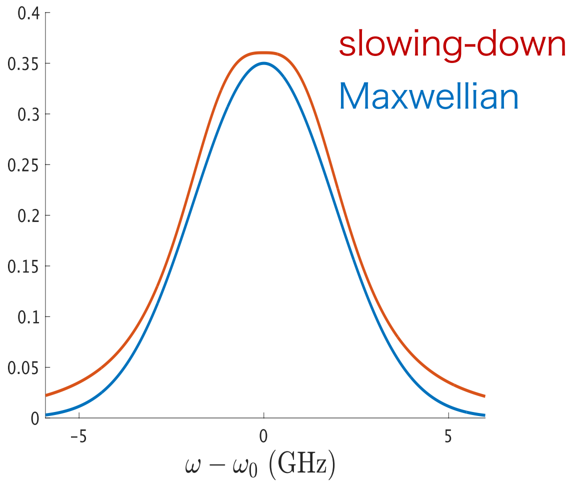
[ST, et al,, *Sci. Rep.*, 2021]



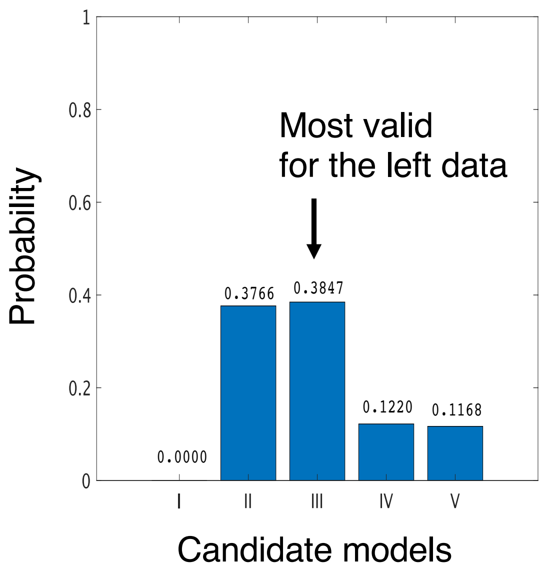
We developed a methodology to select a statistically valid model of laser-induced fluorescence data.

- Absorption band of Ar ions in magnetized plasma

  Modeled by two types of ion velocity distribution functions, also classified as w/ or w/o Zeeman effect



Posterior model probabilities



Candidate models

	Velocity distributions	Zeeman splitting
Model I	no ions	-
Model II	Maxwellian	without
Model III	Maxwellian	with
Model IV	slowing-down	without
Model V	slowing-down	with

Although Zeeman splitting must exist, only a slight difference in probabilities.

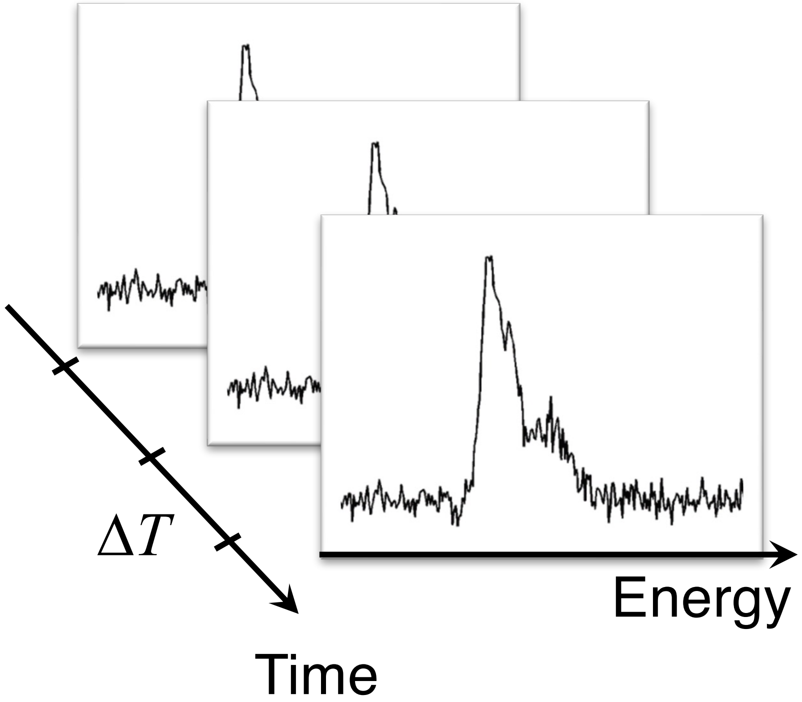
Scaling behavior of Bayesian regression

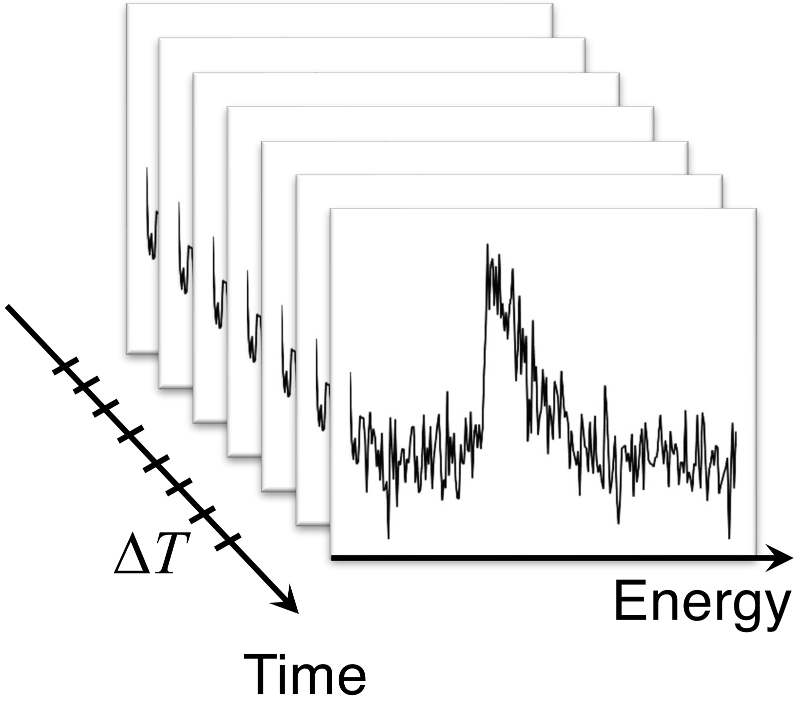
– Dependence of model selection on the quality and quantity of data –

2. ST, et al., "Intrinsic regularization effect in Bayesian nonlinear regression scaled by observed data", *Phys. Rev. Res.* 4, 043165 (2022).
    - Press release:：”計測データの量や質に対するベイズ推定のスケーリング則を解明” (in Japanese)
    - Grants: JSPS KAKENHI (20K19889), etc.

Trade-off between resolution and sensitivity

E.g., Time-resolved spectroscopy





Framerate ΔT^{-1}: Low
SN ratio: High

Framerate ΔT^{-1}: High
SN ratio: Low

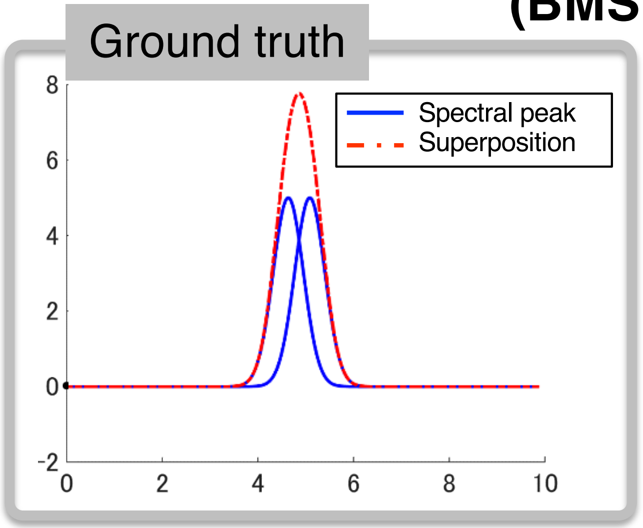
Framerate (time resolution) and SN ratio (sensitivity) are incompatible.
→ There exist noisy data whose sample size is large.

Mystery of "phase transitions" in BMS

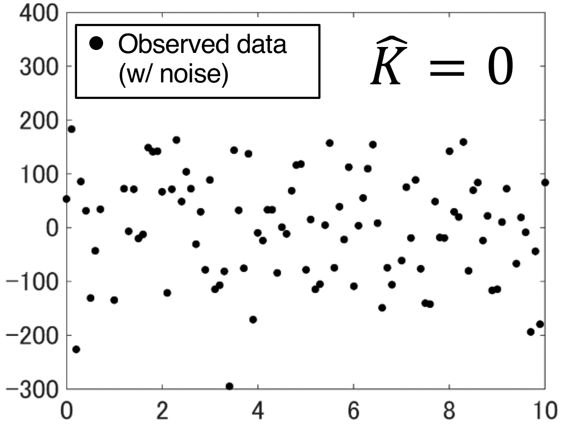
(**BMS** = Bayesian model selection)

E.g., Peak separation

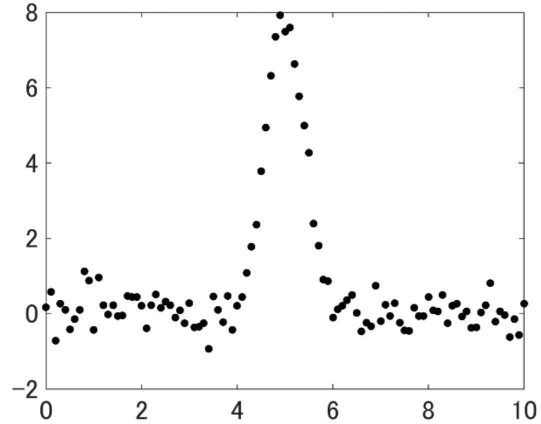
$\widehat{K}$: # of spectral peaks
chosen by BMS

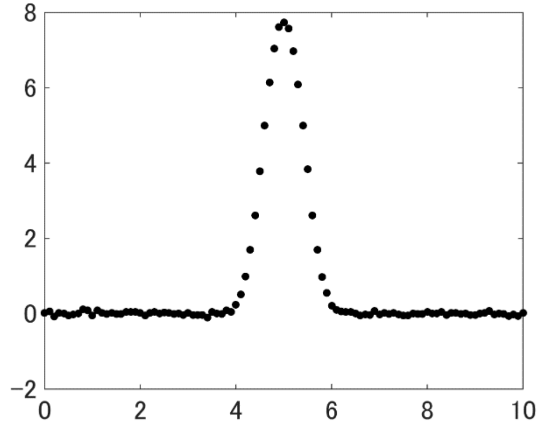


High ← Noise level → Low



"Uninformative"



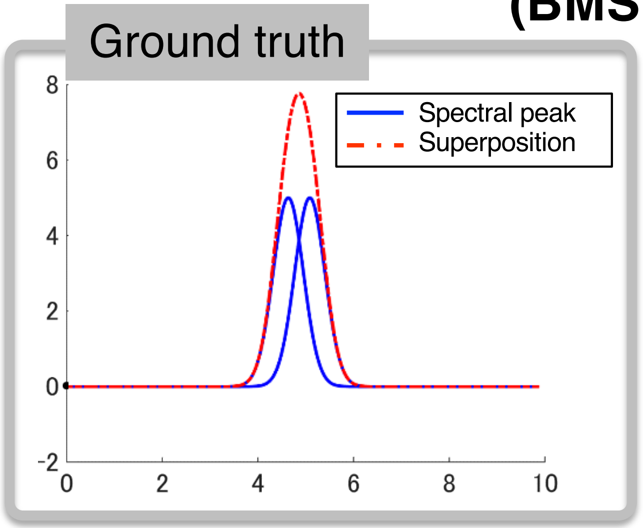


Mystery of “phase transitions” in BMS

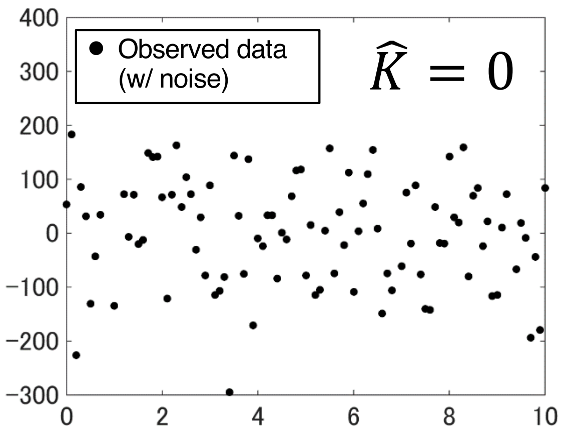
(**BMS** = Bayesian model selection)

E.g., Peak separation

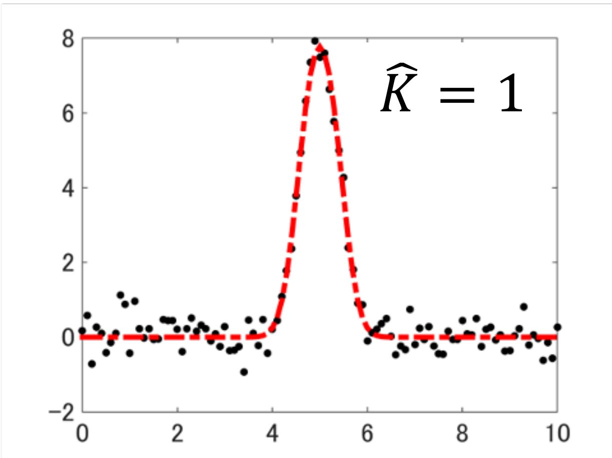
$\widehat{K}$: # of spectral peaks chosen by BMS



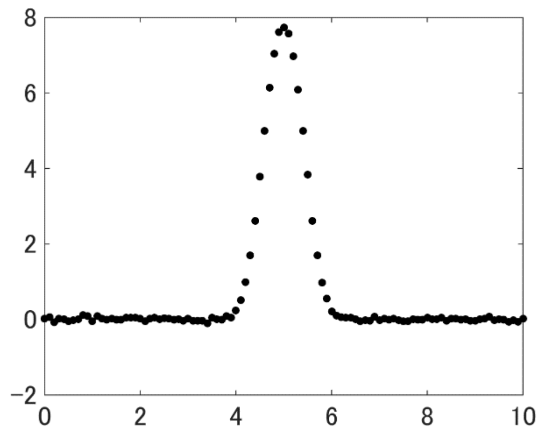
High ⟵ Noise level ⟶ Low



“Uninformative”



“Gross structure”

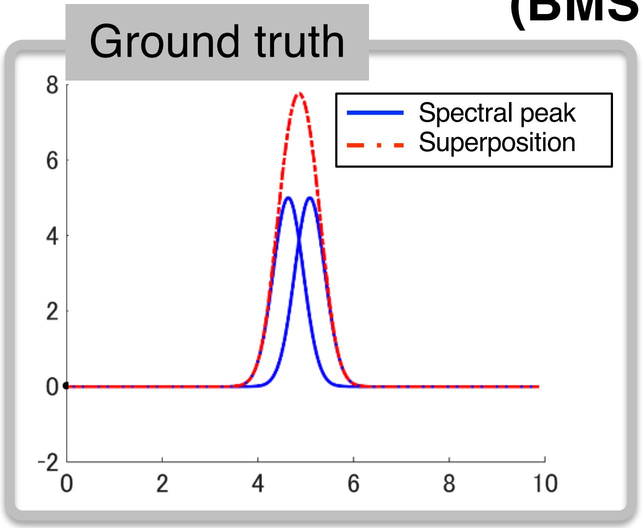


Mystery of "phase transitions" in BMS

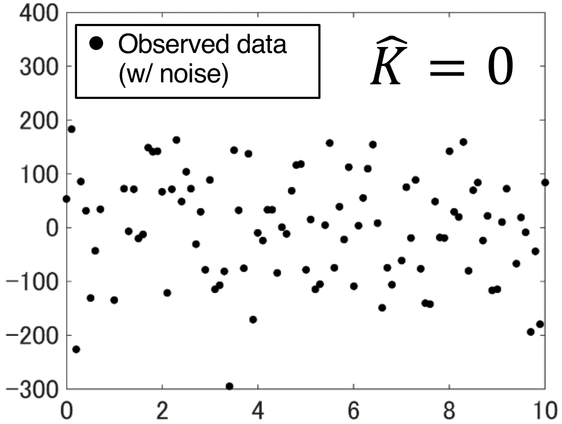
(**BMS** = Bayesian model selection)

E.g., Peak separation

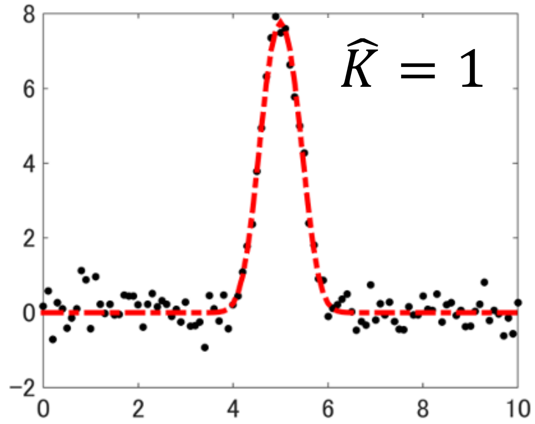
$\widehat{K}$: # of spectral peaks chosen by BMS



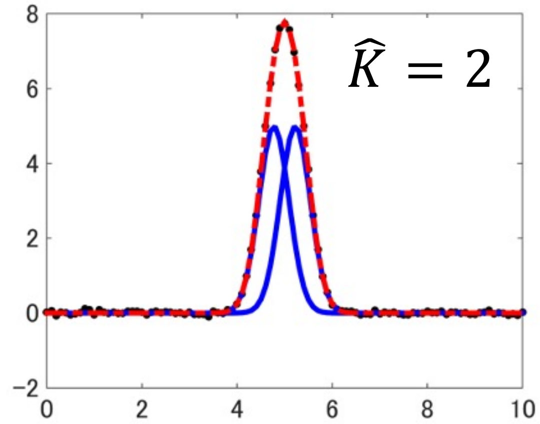
High ← Noise level → Low



"Uninformative"



"Gross structure"



"Fine structure"

Mystery of "phase transitions" in BMS

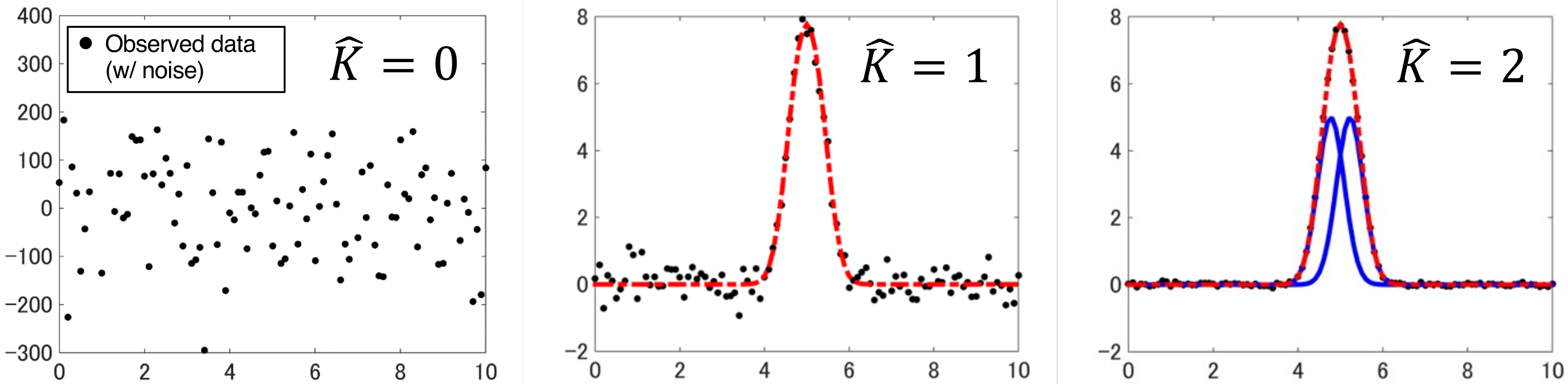
(**BMS** = Bayesian model selection)

How do the quality (and quantity) of data affect BMS?

Conventional asymptotic theories ignore such effects.

(Low) (Data quality) (High)

High ⟵ Noise level ⟶ Low



"Uninformative" "Gross structure" "Fine structure"

Theory of "proportional high-noise regime"

[ST et al., *Phys. Rev. Res.*, 2022]

Fineness of observations $\kappa := n\beta$

$n \to \infty$

$\beta = O\left(\frac{1}{n}\right)$

Finite-size scaling

$$F_n(\beta) = F(\kappa) + \frac{R_n}{2n\beta^*}$$

$\kappa \to \infty$

Conventional theory

$$\kappa F_n(\beta) := -\log Z_n(\beta)$$

Bayes free energy

$$\xrightarrow[\beta = O(1)]{n \to \infty}$$

$$\kappa F_n(\beta) = \kappa E_n(w_0) + \lambda \log n + \cdots$$

[S. Watanabe, *Neural Comput.*, 2001] λ: RLCT

$-\beta^2 \frac{\partial^2}{\partial^2 \beta}$

Estimator of RLCT
(Bayes specific heat)

$$C_n(\beta) := \left(\frac{\kappa}{2}\right)^2 \{\langle E_n(w)^2 \rangle - \langle E_n(w) \rangle^2\}$$

$$\xrightarrow[\beta = O(1/\log n)]{n \to \infty}$$

$$C_n(\beta) \to \lambda$$

$n \to \infty$

$\beta = O\left(\frac{1}{n}\right)$

$$C_n(\beta) = \Lambda(\kappa) + O_p\left(\frac{1}{\log \kappa} + \frac{\beta}{\beta^*} + \frac{1}{n\sqrt{\beta^*}}\right)$$

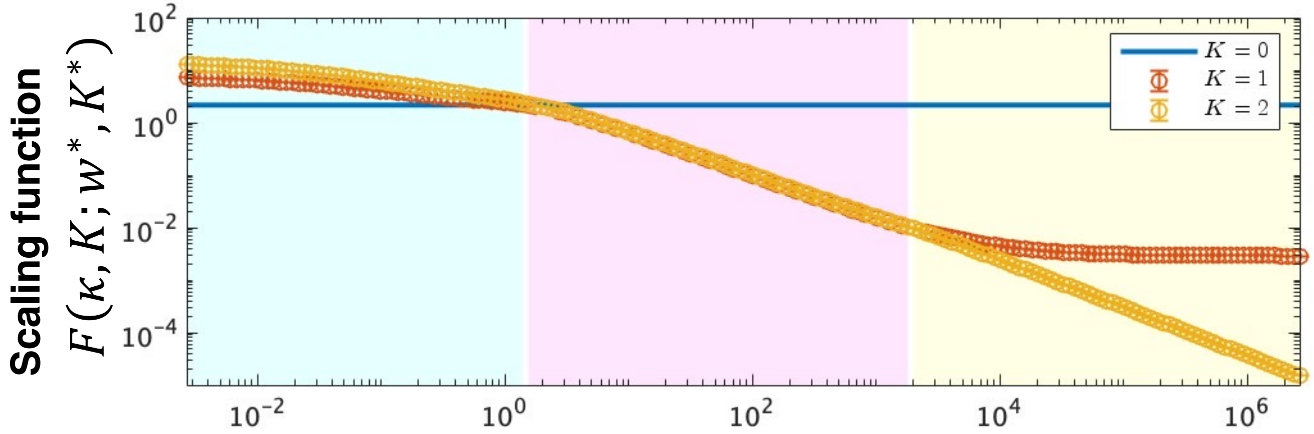
$\kappa, \beta^* \to \infty$

n: Sample size

β^{-1}: Noise level

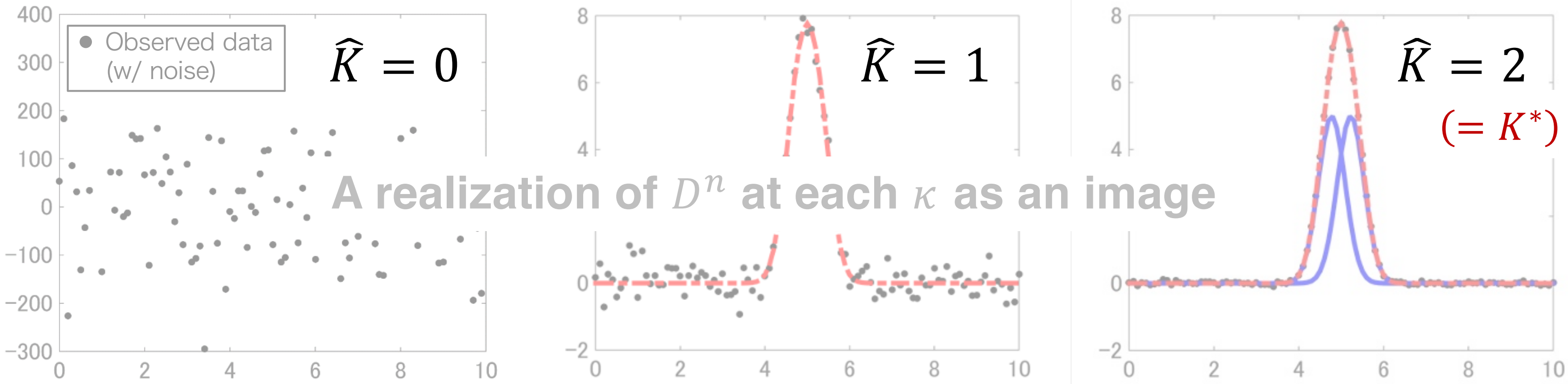
$(\cdots)^*$: Ground truth of $(\cdots)$

Scaling behavior of Bayes free energy



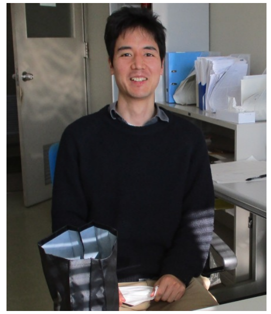
n: Sample size
β^{-1}: Noise level

Low ⟵ **Fineness of observations $\kappa := n\beta$** ⟶ High



A realization of D^n at each κ as an image

- The minimizer $\widehat{K}$ of $F(\kappa, K; w^*, K^*)$, the solution of BMS, does not always coincide with the ground truth K^*, and changes at transitions points of κ.
  - As κ increases, the models selected shift from simpler to more complex ones.



Takashi Nishizawa
(*Kyushu Univ.*)

Application of Gaussian process regression

– On nonparametric regression –

3. T. Nishizawa, ST, et al., "Estimation of plasma parameter profiles and their derivatives from linear observations by using Gaussian processes", *Plasma Phys. Control. Fusion* 65, 125006 (2023).
    - Press release: 機械学習を用いたプラズマ分布解析手法を開発 (in Japanese)
    - Grants: JSPS KAKENHI (20K19889, 22K20358), etc.

Setup: Observations adding Gaussian noise

$$\boxed{Y_i = f(x_i; w) + N_i} \quad (i = 1, 2, \cdots, n)$$

- $\{Y_i\}_{1:n}$: Sample

  $Y_i = y_i$: Realization (datum)
- $x = x_i$: Observation point

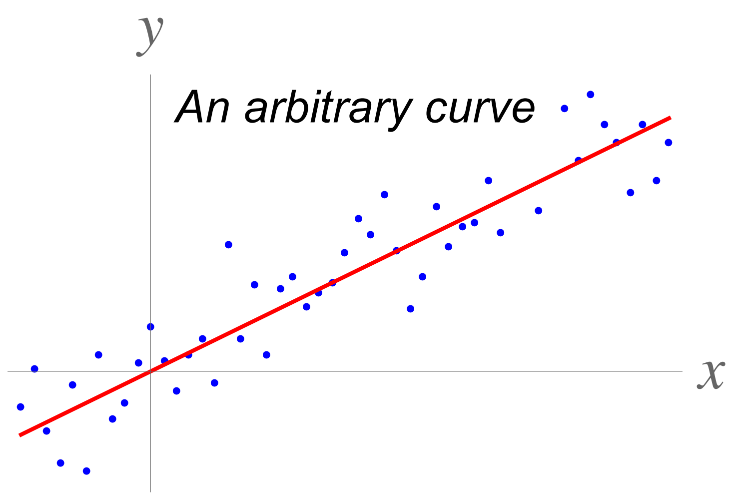
  $x_i = x_{i-1} + L/(n-1)$

  ($L \coloneqq x_n - x_1$; Fixed interval)
- $f \in \{f_K\}$: Arbitrary function of x
- w: Parameter set of f
- N_i: Observation noise

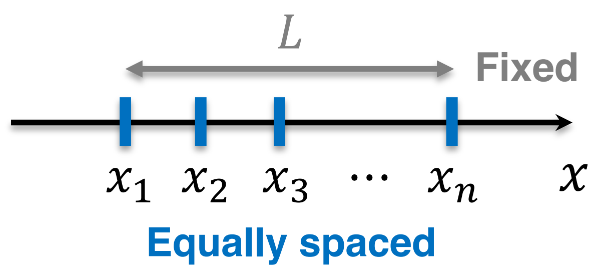
  $N_i \sim \mathcal{N}(0, \beta^{-1})$ *Gaussian noise*

  i.e., $Y_i \sim \mathcal{N}(f(x_i; w), \beta^{-1})$

  $\beta^{-1} > 0$: Noise level (variance)



$Y_i | x_i$: **conditionally independent**

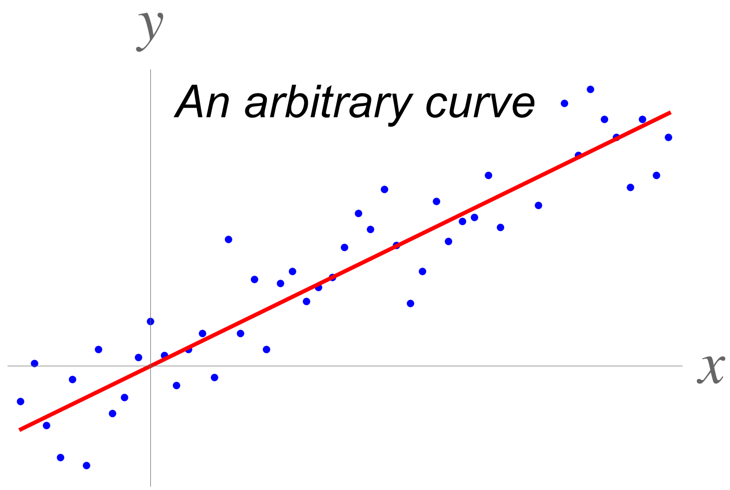


Data are realizations of random variables following a Gaussian distribution

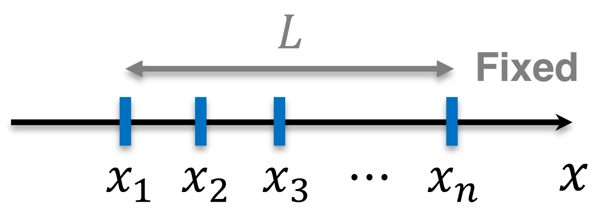
From parametric to nonparametric model

$$\boxed{Y_i = f(x_i \cancel{; w}) + N_i} \quad (i = 1, 2, \cdots, n)$$

- $\{Y_i\}_{1:n}$: Sample
  $Y_i = y_i$: Realization (datum)
- $x = x_i$: Observation point
  $x_i = x_{i-1} + L/(n-1)$
  ($L \coloneqq x_n - x_1$; Fixed interval)
- $f \in \{f_K\}$: Arbitrary function of x
- ~~w: Parameter set of f~~
- N_i: Observation noise
  $N_i \sim \mathcal{N}(0, \beta^{-1})$ *Gaussian noise*
  i.e., $Y_i \sim \mathcal{N}(f(x_i \cancel{; w}), \beta^{-1})$
  $\beta^{-1} > 0$: Noise level (variance)



$Y_i | x_i$: **conditionally independent**



Equally spaced

Data are realizations of random variables following a Gaussian distribution

Nonparametric model with resolution matrix

$$\boxed{Y_i = \sum_j M_{ij} f(x_j) + N_i} \quad \begin{matrix}(i = 1, 2, \cdots, m) \\ (j = 1, 2, \cdots, n)\end{matrix}$$

- $\{Y_i\}_{1:n}$: Sample

  $Y_i = y_i$: Realization (datum)
- $x = x_i$: Observation point

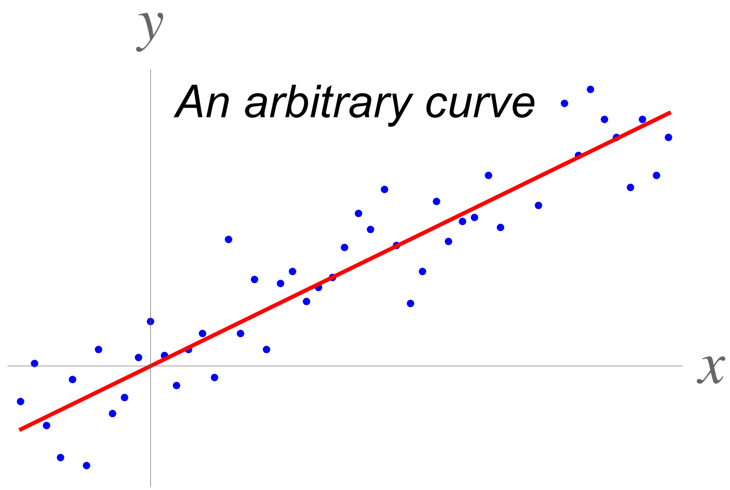
  $x_i = x_{i-1} + L/(n-1)$

  ($L := x_n - x_1$; Fixed interval)
- $f \in \{f_K\}$: Arbitrary function of x
- M_{ij}: Spatial (or temporal) resolution
- N_i: Observation noise

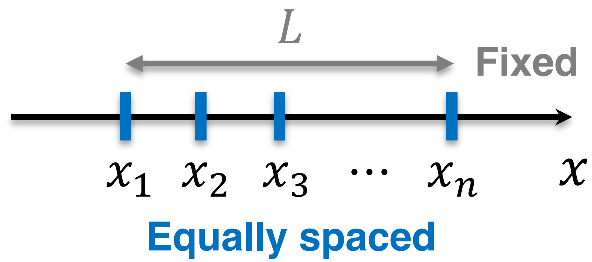
  $N_i \sim \mathcal{N}(0, \beta^{-1})$ *Gaussian noise*

  i.e., $Y_i \sim \mathcal{N}(f(x_i), \beta^{-1})$

  $\beta^{-1} > 0$: Noise level (variance)



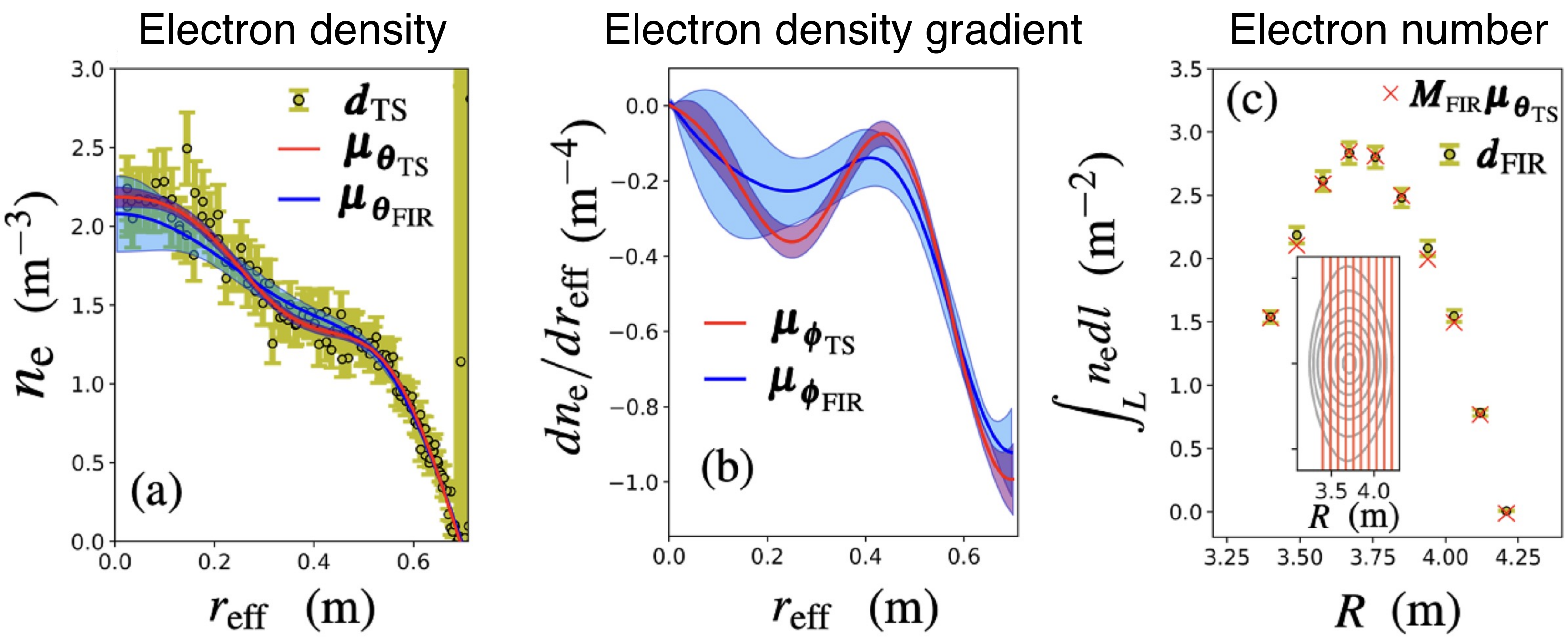
$Y_i \,\middle|\, \{M_{ij}, x_j\}_{1:n}$: **conditionally independent**

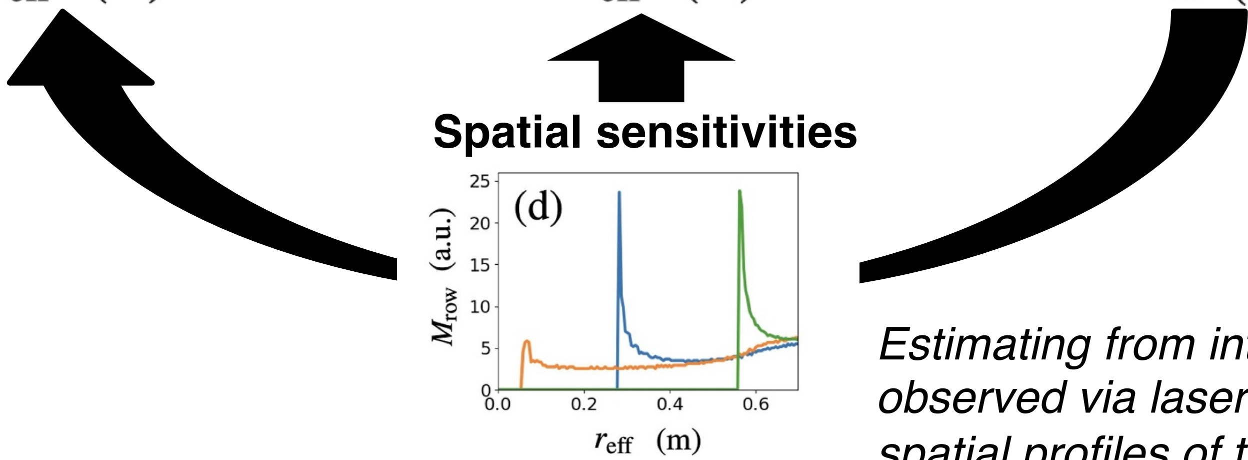


Data are realizations of random variables following a Gaussian distribution

Gaussian process for line-integrated quantities

[T. Nishizawa, ST, et al., *Plasma Phys. Control. Fusion*, 2023]





Estimating from integral quantities observed via laser interferometers spatial profiles of their derivatives

Physics-informed Gaussian process regression

Incorporation of physical laws

[T. Nishizawa, ST, et al., *PPCF*, 2024]

- Extension for axisymmetric plasma

Observations (via magnetometry etc.)

↓ ← Spatial smoothness

Gaussian process (closed-form)

↓ Plasma current distribution (Prior distribution)

MCMC sampling ← Grad-Shafranov eq. (Likelihood)

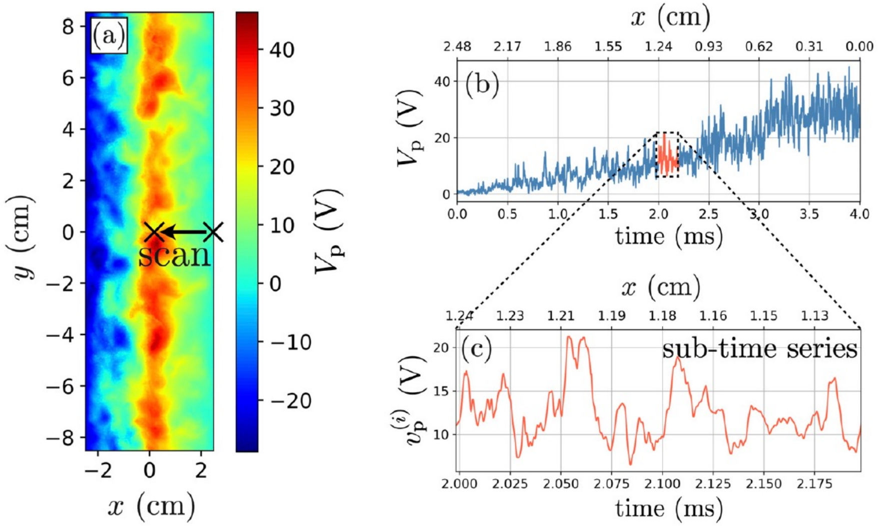
↓

Solution space of GS eq.

Quantification of plasma fluctuation

[T. Nishizawa, P. Manz, ST, et al., *PoP*, 2025]

- Extension for spatial scanning measurements by moving the probe

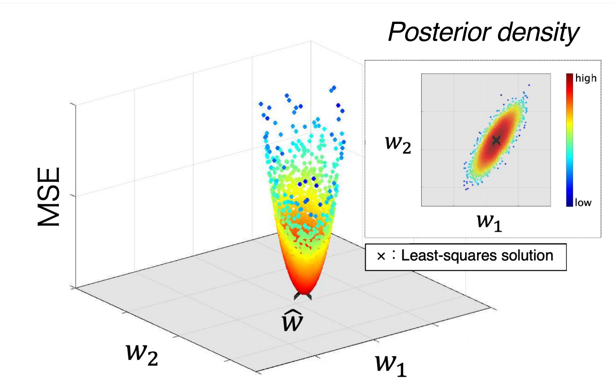


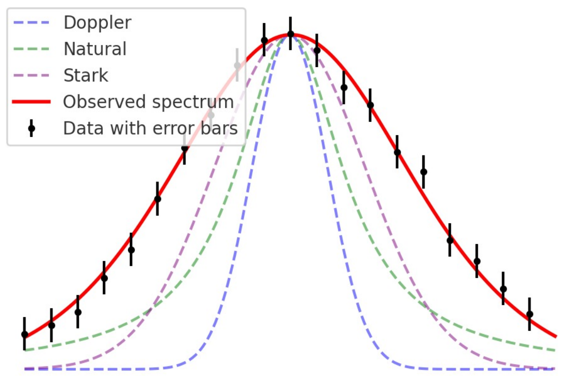
Autocorrelation in a sub-time series as an error

Bridges between Bayesian statistics and plasma physics

Summary

- Relationship between statistics and physics
  - Error theory and statistical mechanics
  - Epistemology and ontology
  - Regression of hierarchical generative models
- Parametric Bayesian regression
  - Model selection of ion velocity distributions
  - Theory of “proportional high-noise regime”
- Nonparametric Bayesian regression
  - Gaussian process with resolution matrix
  - Gaussian process with physical properties





1. ST, et al, “Bayesian inference of ion velocity distribution function from laser-induced fluorescence spectra”, *Sci. Rep.*, 11, 20810 (2021).
2. ST, et al., "Intrinsic regularization effect in Bayesian nonlinear regression scaled by observed data", *Phys. Rev. Res.* 4, 043165 (2022).
3. T. Nishizawa, ST, et al., "Estimation of plasma parameter profiles and their derivatives from linear observations by using Gaussian processes", *Plasma Phys. Control. Fusion* 65, 125006 (2023).