Bayesian Inference for the LHD Experiment Data

Keisuke Fujii
Bayes rule

\[ p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)} \]

Posterior
Full knowledge of \( \theta \)
Incl. mean and standard deviation
Bayes rule

**Likelihood**
How data (or noise) behaves.
e.g. Gaussian with mean $\theta$

**Posterior**
Full knowledge of $\theta$
Incl. mean and standard deviation

$$p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)}$$
Bayes rule

\[ p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)} \]

**Posterior**
Full knowledge of \( \theta \)
Incl. mean and standard deviation

**Likelihood**
How data (or noise) behaves.
e.g. Gaussian with mean \( \theta \)

**Prior**
Our assumption on data.
Bayes rule

Posterior
Full knowledge of $\theta$
Incl. mean and standard deviation

Likelihood
How data (or noise) behaves.
e.g. Gaussian with mean $\theta$

Prior
Our assumption on data.

$$p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)}$$

Probabilistic modeling
• Quantify what we assume.

Advantage
• Uncertainty quantification
• Assumption selection (model selection)
Outline

• Brief introduction

• Evaluation of fractional abundance data for W
  • Avoiding over and under fitting
    -model selection-

• Evaluation of systematic noise of
  LHD Thomson scattering system.

• Summary
Outline

• Brief introduction

• Evaluation of fractional abundance data for W
  • Avoiding over and under fitting
    -model selection-

• Evaluation of systematic noise of
  LHD Thomson scattering system.

• Summary
Fractional abundance of W is an important data essential to the tungsten transport diagnostics. Significant disagreement has been reported among the results by different groups, in particular $q < 30$. 

\[ d_{\text{model}} = \int_{\text{LOS}} \varepsilon_i(n_e, T_e)n_e \xi_q(T_e)n_W \,dx \]
Measurement

(a) $T_e$ (keV) vs. $r$ (m) at $t = 0.0$ s

(b) $T_e$ (keV) vs. $r$ (m) at $t = 4.22$ s, $6.32$ s, $7.52$ s

(c) $n_e$ ($x10^{19}$ m$^{-3}$) vs. $t$ (s) at $Z = 0.026$ m

(d) Intensity (arb. units) vs. $z$ (m) at $t = 4.22$ s, $6.32$ s, $7.52$ s
Objective: Inference of $\xi_q$ from the experimental data

Parameters:

- $\epsilon_i n_e$: Emission rate per 1 ground state ion.
- $\xi_q$: Fractional abundance
- $n_w$: Total tungsten density distribution

Assumptions:

- Independent of $n_e$ and $T_e$
- Smooth function of $T_e$
- Smooth function of $r$ and $t$

\[ d_{\text{model}} = \int_{\text{LOS}} \epsilon_i(n_e, T_e)n_e\xi_q(T_e)n_w\,dx \]
# How much we should assume

## Parameters:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varepsilon_i n_e$</td>
<td>Emission rate per 1 ground state ion.</td>
</tr>
<tr>
<td>$\xi_q$</td>
<td>Fractional abundance</td>
</tr>
<tr>
<td>$n_w$</td>
<td>Total tungsten density distribution</td>
</tr>
</tbody>
</table>

## Assumptions

- Independent of $n_e$ and $T_e$
- Smooth function of $T_e$
- Smooth function of $r$ and $t$

### How smooth profile we should assume?

- Too strong assumption.
- Too weak assumption.

Introduce hyperparameter

Parameters:
- $\xi_q$: Fractional abundance
- $n_w$: Total tungsten density distribution

Assumptions:
- Smooth function of $T_e$
- Smooth function of $r$ and $t$

It is necessary to quantify the smoothness.

- Discretize the profile into finite number points
- Apply prior distribution for the difference $\Delta \xi$
- Parameterize the prior by hyperparameter

Prior (assumption)

$$p(\xi_q | \lambda_\xi) = \mathcal{N}(\Delta \xi_q | 0, \lambda_\xi)$$

How strong we assume.

$$p(\theta | d) = \frac{p(d | \theta)p(\theta)}{p(d)}$$
Choose how much we should assume from data

Prior (assumption)

\[ p(\xi_q | \lambda \xi) = \mathcal{N}(\Delta \xi_q | 0, \lambda \xi) \]

How strong we assume.

Too strong assumption.

Too weak assumption.

\[ p(\xi_q | \lambda \xi) = \mathcal{N}(\Delta \xi_q | 0, \lambda \xi) \]
Choose how much we should assume from data.

Too strong assumption.

How should we remove the dependence on $\lambda_\xi$?

Marginalization (apply prior for $\lambda_\xi$ and integrate out)

$$p(\xi_q | d) \propto \int p(d | \xi_q) p(\xi_q | \lambda_\xi) p(\lambda_\xi) d\lambda_\xi$$

This avoids the under and over-fitting.
Our model well represents the measured data.

No under-fitting
Inferred $\xi_q$ profiles are smooth enough.

No over-fitting

Our results are close to those by Putterich et al, but our peak positions locate at the smaller $T_e$ side.

Our results may be used as benchmark for future theoretical works.
Outline

- Brief introduction

- Evaluation of fractional abundance data for W
  - Avoiding over and under fitting -model selection-

- Evaluation of systematic noise of LHD-TS diagnostic system.
  arXiv:1607.05380

- Summary
**Systematic noise in LHD-TS system**


Systematic noise: Has large correlation. Inaccurate calibration, model, ...

Can be analyzed by legacy statistic.

Bayesian statistics

![Graph showing original data with significant dependent noise due to mis-calibration.](image)
Objective:
Machine learning of the mis-calibration noise

(a) original data

(b) post calibration data

Systematic noise model

Current calibration factor for channel $i$

$$R_i = R_i^0 (1 + \Delta_i)$$

True calibration factor for channel $i$

Mis-calibration noise (to be estimated)
Probabilistic modeling

Experiment 1

Experiment 2

Experiment M

data
true values
systematic noise

\[ y_1 = f_1 + n_1 + n_\Delta \]
\[ y_2 = f_2 + n_2 + n_\Delta \]

\[ y_M = f_M + n_M + n_\Delta \]
Gaussian Process for multiple frame data

\[
\begin{align*}
\mathbf{y}_1 &= \mathbf{f}_1 + \mathbf{n}_1 + \mathbf{n}_\Delta \\
\mathbf{y}_2 &= \mathbf{f}_2 + \mathbf{n}_2 + \mathbf{n}_\Delta \\
&\quad \quad \vdots \\
\mathbf{y}_M &= \mathbf{f}_M + \mathbf{n}_M + \mathbf{n}_\Delta \\
\end{align*}
\]

Prior:
\[
p(\mathbf{f}) = N(0,K_f)
\]

Prior:
\[
p(\mathbf{n}) = N(0,K_n)
\]
Gaussian Process for multiple frame data

\[
\begin{align*}
    y_1 &= f_1 + n_1 + n_\Delta \\
    y_2 &= f_2 + n_2 + n_\Delta \\
    &\vdots \\
    y_M &= f_M + n_M + n_\Delta
\end{align*}
\]

prior
\[
p(f) = N(0,K_f)
\]

prior
\[
p(n) = N(0,K_n)
\]

prior
\[
p(\Delta) = N(0,K_\Delta)
\]
\[ R_i = R_i^0 (1 + \Delta_i) \]

Current calibration factor for channel \(i\)

True calibration factor for channel \(i\)

Mis-calibration noise (to be estimated)

210 frames of the new observation data by LHD Thomson system.
Application to the derivative inference.
Outline

• Brief introduction

• Evaluation of fractional abundance data for W
  • Avoiding over and under fitting
    -model selection-

• Evaluation of systematic noise of LHD-TS diagnostic system.

• Summary
We inferred

• the fractional abundance of W ions from LHD experimental data
• The systematic noise for the LHD-TS system by applying the Bayesian inference.
Bayesian statistics states “the importance of the assumption”.

The main challenge in Bayesian statistics is how we quantify our assumption.

There is no super-tool that is used for all the purposes. We A.M. data unit may need to develop our own statistical models to
- model the theoretical results
- update the data with experimental data
Our second attempt is to infer the systematic noise for LHD-TS system from a large amount of LHD experiment data. (data-driven science)

Revealed more detailed structure of $n_e$. 
Details 1: Additive approximation

\[ R_i = R_i^0 (1 + \Delta_i) \]

True calibration factor for channel \( i \)

Current calibration factor for channel \( i \)

Mis-calibration noise (to be estimated)

\[ y = f + n + n_{\Delta} \]

Mis-calibration noise is not additive.

\[ y_j = f_j + n_j + f_j \Delta \]

Additive approximation with iteration.
The distribution of $\Delta$ may not be Gaussian.

We adopt a Cauchy distribution for $\Delta$.

Hierarchical model

\[
\begin{aligned}
\rho(\Delta_i) &= \mathcal{N}(0, \sigma_i^2) \\
\rho(\sigma_i^2) &= \mathcal{IG}\left(\frac{1}{2}, \frac{\sigma^2_{\Delta}}{2}\right)
\end{aligned}
\]
There are some outliers. The distribution of $n$ may not be Gaussian.

We adopt a Cauchy distribution also for $n$. 
Application to the derivative inference.

(a) 

![Graph showing original data and post-calibrated data for $n_e$ vs. $|r|$ (m)].

(b) 

![Graph showing $dn_e/dr$ vs. $|r|$ (m)].

(c) 

![Graph showing $d^2n_e/dr^2$ vs. $|r|$ (m)].
Inference for the training data

\[
\frac{y_{i,j}}{1 + \Delta_i} = f_{i,j} + \frac{n_{Di,j} + n_{Pi,j}}{1 + \Delta_i}
\]
Inference for the test data

We made this post-calibration for test data that are NOT used for the $\Delta$ inference.

Detailed structures become apparent, suggesting no over-fitting.
Application to the derivative inference.
FIG. 5. (a) Two-dimensional image of the spectrum observed for the discharge #121534 at $t=4.22$ s as a function of the wavelength (horizontal axis), height $z$ (vertical axis) and intensity (by false color). (b) The spectrum observed at $t=4.22$ s for the LOS with $z=0.026$ m. The central wavelengths for the highly charged tungsten ion emission lines are indicated in the figure.
FIG. A.1. $n_e$ dependence of $\eta_i$ values for 333.71-., 335.75-, 389.40- and 337.74-nm lines estimated by collisional-radiative model [30]. The calculations were made with the assumption of $T_e = 0.8$ keV for the $q = 26$ lines, while $T_e = 1.0$ keV is assumed for the $q = 27$ lines. $\eta_i$ linearly increases in $n_e < 10^{17}$ m$^{-3}$, while it becomes saturated in $n_e > 10^{18}$ m$^{-3}$. The $n_e$ range considered in this work ($n_e = 1 - 5 \times 10^{19}$ m$^{-3}$) are indicated by shadows. Note that this calculation does not contain the ion-collision effect.