Integrated Data Analysis of the DIII-D Density Profile

L. Stagner & W.W. Heidbrink
University of California–Irvine, Irvine, California, USA

Abstract

1 Introduction

Introstuff

2 Integrated Data Analysis

As fusion experiments expand, so does the number of diagnostics used to probe the underlying physics. With the increase in diagnostics the number of interrelations among the diagnostics increases accordingly. While each diagnostic can act independently it is often advantageous to combine the data sets to acquire the most accurate result that is consistent with all the information available. This causes difficulties when trying to combine multiple data sets that cover complementary areas of interest. Bayesian data analysis give a comprehensive, scalable, and automated framework for combining complementary diagnostics.

2.1 The Bayesian perspective

Bayesian statistics differs from the prevailing frequentist statistics that is taught in most introductory courses. The difference between the two views is in the definition of probability. In the Bayesian perspective, probability represented a degree of belief or plausibility that an event was true, based on the information available. This school of thought is opposed to frequentist view that probability is a long run relative frequency with which the event occurs, given infinitely many repeated experimental trials.[1]

While the frequentist view is seemingly more objective, it is limited in its straightforward application since some things cannot have long run relative frequencies and, as a result, cannot be analysed by probability theory. For instance a constant, such as the mass of an object, cannot be analysed by a straightforward application of the frequentist definition of probability as a long run relative frequency. To use the frequentist view you first have to relate the mass to the data through some function called the *statistic*. Since statistic is subject to random noise it becomes the random variable to which the rules of probability can then be applied. The issue with this approach is that there is no natural way of choosing the best statistic.[2] The founders of the field have a plethora of ad hoc rules and tests to determine which statistic should be used in any situation. These rules and tests hides much of the mechanics and assumptions inherent in the analysis and the consequences can often turn up at the worst possible moments.

The bayesian definition has no such difficulty since nearly anything can be thought of as a probability and the framework explicitly formulates all assumptions. That is not to say that it is without its issues. The frequentist view is often much easier to use because it hides most of the difficulties from the user. In bayesian statistics you have to be very careful with the assumptions that are made since it can drastically affect the results. Also, the calculation of optimal parameters is also a challenging numerical exercise in Bayesian statistics. It was partially for this reason that the frequentist view arose into prominence. It was only in the mid 20^{th} century with the advent of computers and efficient sampling algorithms that Bayesian statistics has begun its resurgence. In recent years Bayesian analysis has become essential to many scientific fields such as cosmology and artificial intelligence.[1]

2.2 Mathematical Formulation

In 1946, Richard Cox began to formulate the rules for logical and consistent reasoning by considering how we might quantify our beliefs about the truth of an event or object. He concluded that the real numbers we attached to our beliefs followed the rules of probability:[2]

$$prob(X|I) + prob(\overline{X}|I) = 1 \tag{2.1}$$

and

$$prob(X, Y|I) = prob(X|Y, I) \times prob(Y|X, I) \tag{2.2}$$

Here $\overline{X}$ denotes that X is false and the vertical bar '|' means 'given'. All the probabilities are conditional on I to denote the relevant background information available, since there is no such thing as an absolute probability.

Equation 2.1 and 2.2 are called the sum and product rules, respectively. The sum and product rule form the basic equations of probability theory. From these two rules we can derive the core equations used in bayesian statistics: Bayes' theorem

$$prob(X|Y, I) = \frac{prob(Y|X, I) \times prob(X|I)}{prob(Y|I)} \tag{2.3}$$

and marginalization.

$$prob(X|I) = \int_{-\infty}^{\infty} prob(X, Y|I) dY \tag{2.4}$$

The power behind Bayes theorem can best be recognized with a change of notation.

$$prob(hypothesis|\{data\}, I) = \frac{prob(\{data\}|hypothesis, I) \times prob(hypothesis|I)}{prob(\{data\}|I)} \tag{2.5}$$

The terms in the above equation have formal names. $prob(hypothesis|I)$ is called the *prior* and is often denoted in parameter estimation problems as $\pi(\vec{\alpha}|I)$ where $\vec{\alpha}$ are the parameters to be inferred. The prior encodes the current state of knowledge about the parameters before we have analyzed any new data. When we have acquired new information the prior is modified by the first term, $prob(\{data\}|hypothesis, I)$ called the *likelihood* also denoted as $\mathcal{L}(\vec{d}|\vec{\alpha}, I)$. The product of the likelihood and the prior give $prob(hypothesis|\{data\}, I)$, which is called the *posterior* and is also denoted as $\mathcal{P}(\vec{\alpha}|\vec{d}, I)$. The posterior encodes our final state of knowledge after all the data has been incorporated. The maximum value of the posterior gives us the best estimate of the parameters we trying to infer. The last term,

$prob(\{data\}|I)$, is called the *evidence* or *marginal likelihood* and is denoted as $\mathcal{Z}$. In parameter estimation problems the evidence can be ignored since it is essentially a normalization constant. However, in model comparison problems it is of vital importance.

2.3 Model Comparison

2.4 Choosing Priors

It was noted earlier that one has to be extra careful with the assumptions that are made prior to incorporating data. Quite fittingly, these assumptions are encoded in the choice of prior. This process can be difficult and if often done incorrectly. In order to choose the best prior three main schools of thought for choosing priors have been developed.

The first school of thought is to choose a prior that expresses specific, definite information about a parameter. If, for instance, a parameter is known, through past experience or expert testimony, to have a value around 1 ± 0.2 with upper bound of 2. It would not be uncommon for someone to assign a prior that is a Gaussian with a mean of 1 and standard deviation of 0.2 that is zero for values greater than 2. Informative priors are both a major advantage and can also be a major pitfall. If our prior knowledge was incorrect and the upper bound is around 5, depending on the amount of data, we may have biased the posterior such that it doesn't reflect the truth. We must then be careful not to bias our results without proper justification. Informative priors also suffer from the fact that there is not a well defined procedure for choosing them.

The second school of thought is to be as non-informative as possible. This method should be used when we have no prior information about the parameter in question. The most common type of non-informative prior is called the Jeffreys' Prior. The key property of the Jeffreys' Prior is that it is invariant under re-parameterization. The Jeffreys' prior is defined as

$$\pi(\vec{\alpha}) \propto \sqrt{det\,\mathcal{I}(\vec{\alpha})} \tag{2.6}$$

where $\mathcal{I}$ is the Fisher information. One of the main problems with the Jefferys' Prior is that it can be improper in the sense that it cannot normalizable to one (although there are ways around this). Improper priors should not be used since it can lead to paradoxes.[1] The Jefferys' Prior is also sometimes impossible to calculate and is therefore of limited use.

The third school of thought is to choose the prior that has the largest Shannon information entropy that is still consistent with the given testable information. The idea is that entropy is a measure of "uninformativeness" so picking the most 'uninformative' prior that is still consistent with testable information would be the best choice. Mathematically, this the constrained optimization problem:

$$Q = -\int \pi(x) log(\frac{\pi(x)}{m(x)}) dx - \lambda C(x) \tag{2.7}$$

where $p(x)$ is the prior, $m(x)$ is the Lebesgue measure which ensures invariance under transformation, λ is the Lagrange multiplier, and $C(x)$ is the constraint function. For example, if we knew the mean of a parameter, using the principle of maximum entropy the best prior would be a Poisson distribution. Likewise, if we knew both the mean and the variance of the parameter the principle would yield a gaussian. The principle of maximum entropy provides a balanced and systematic approach to problem of picking the right prior. It is for these reasons that we will endeavor to apply the principle of maximum entropy whenever possible.

2.5 Combining multiple diagnostics

One of the advantages of bayesian statitics is that it provides a systematic framework for combining data sets from multiple sources. This can be seen from Eq. 2.5. Consider just two data points D_1 and D_2. Equation 2.5 would yield:

$$prob(H|D_1, D_2, I) \propto prob(D_1, D_2|H, I) \times prob(H|I) \tag{2.8}$$

We can use Bayes theorem to express the posterior to be conditional on D_1.

$$prob(H|D_1, D_2, I) \propto prob(D_2|H, D_1, I) \times prob(H|D_1, I) \tag{2.9}$$

This shows that the prior in Eq. 2.8 can be replaced by the posterior based on D_1.

If D_1 and D_2 are independent then

$$prob(D_1|H, D_2, I) = prob(D_1|H, I) \tag{2.10}$$

and

$$prob(D_2|H, D_1, I) = prob(D_2|H, I) \tag{2.11}$$

Substituting Eq. 2.11 into Eq. 2.9 and applying Bayes theorem to the prior yields:

$$prob(H|D_1, D_2, I) \propto prob(D_2|H, I) \times prob(D_1|H, I) \times prob(H|I) \quad (2.12)$$

This result can be generalized for K data points as

$$prob(H|\{D_k\}, I) \propto \left(\prod_{k=0}^{K} prob(D_k|H, I) \right) \times prob(H|I) \quad (2.13)$$

This result forms the basis for integrated data analysis. So long as all the data is independent from each other, we can assign different likelihoods for each data point. This allows for the data to have different types of errors. For instance, we could combine a diagnostic that is subject to systematic error with a diagnostic whose error is distributed according to a Poisson or Gaussian distribution. Using more orthodox methods this could not be done easily, but within a bayesian framework it flows naturally from the basic equations.

Looking more closely at Eq. 2.13 you will notice that each likelihood is dependent of H. This requires that we have a forward model for the diagnostic that maps the parameters we are trying to infer to the data. To put Eq. 2.13 into a more explicit form:

$$prob(\vec{\alpha}|\{D_k\}, I) \propto \left(\prod_{k=0}^{K} prob(D_k|H, I) \right) \times prob(\vec{\alpha}|I)$$

3 Forward modeling of density diagnostics

stuff

3.1 Functional form of the density profile

stuff

3.2 Interferometry

stuff

3.3 Reflectometry

stuff

3.4 Thomson scattering

stuff

3.5 Beam emission spectroscopy

stuff

4 Likelihood and prior probabilities

stuff

4.1 Priors

stuff

4.1.1 Parameter space priors

stuff

4.1.2 Data space priors

stuff

4.2 Likelihoods

stuff

4.2.1 Interferometry

stuff

4.2.2 Reflectometry

stuff

4.2.3 Thomson scattering

stuff

4.2.4 Beam emission spectroscopy

stuff

4.2.5 Combined likelihood probability

stuff

5 Total posterior probability

stuff

5.1 Exploring the posterior

stuff

6 Representation of Error

stuff

7 Density profile reconstructions

stuff

7.1 H-mode reconstruction

stuff otherstuff

7.2 L-mode reconstruction

stuff

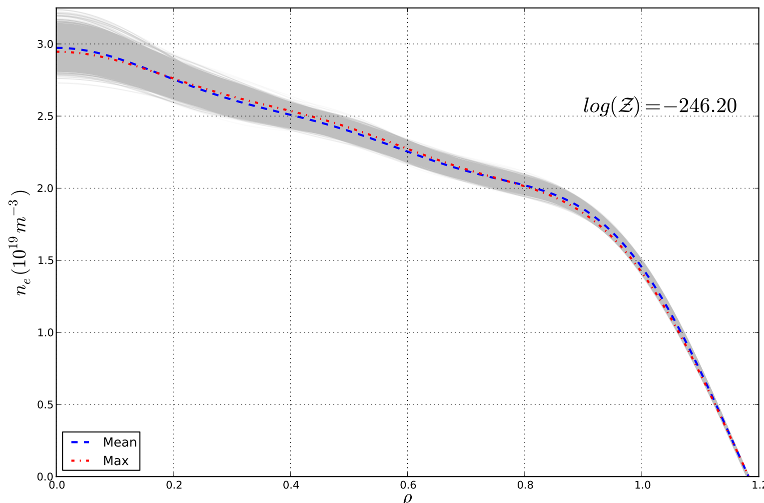


Figure 1: Reconstruction

8 Determining discrepant diagnostics

stuff

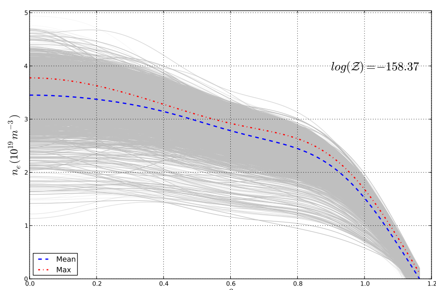
9 Conclusions and future work

stuff

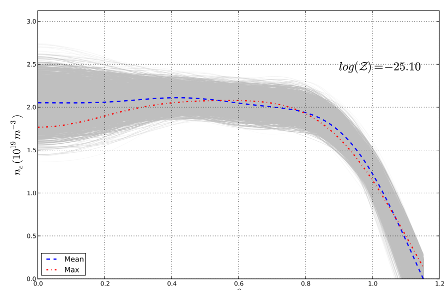
References

[1] Udo von Toussaint. Bayesian inference in physics. *Reviews of Modern Physics*, 83(3):943, 2011.

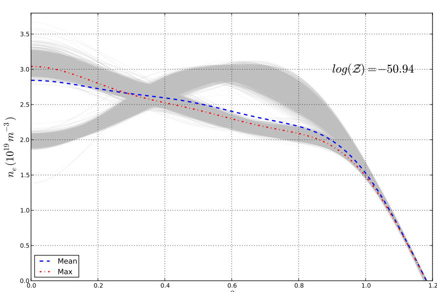
[2] Devinderjit Sivia and John Skilling. *Data Analysis: A Bayesian Tutorial*. Oxford University Press, USA, 2006.



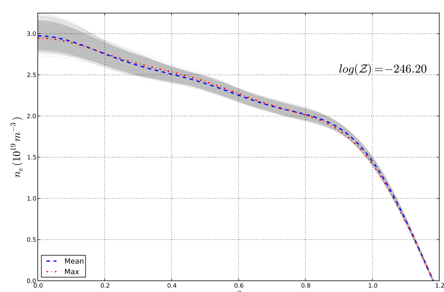
(a) Thomson scattering



(b) Interferometer



(c) Reflectometry



(d) Combined

Figure 2: The l-o-n-g caption for all the subfigures (FirstFigure through FourthFigure) goes here.