Bayesian Experimental Design – Studies for Fusion Diagnostics

R. Fischer

Max-Planck-Institut für Plasmaphysik, EURATOM Association, Boltzmannstr. 2, 85 748 Garching, Germany

Abstract. The design of fusion diagnostics is essential for the physics program of future fusion devices. The goal is to maximize the information gain of a future experiment with respect to various constraints. A measure of information gain is the mutual information between the posterior and the prior distribution. The Kullback-Leibler distance is used as a utility function to calculate the expected information gain marginalizing over data and parameter space. The expected utility function is maximized with respect to the design parameters of the experiment. The method will be applied to the design of a Thomson scattering experiment.

INTRODUCTION

It is of major concern for any scientist planning experiments to optimize the design of a future experiment with respect to best performance within expected experimental scenarios. Best performance has to be defined with respect to the goals of the experiment and cannot be a derived from first principles. A utility function has to be defined specifying the desired benefit of the experimental outcome and to evaluate different designs:

$$\text{Utility} = U(D, \eta) \tag{1}$$

The utility function U depends on the data D of a future experiment with design parameters η. Experimental design is a decision theoretic problem based on optimizing expected utility functions with respect to η. An overview over classical and Bayesian utility functions can be found for example in [1].

The goal of this paper is to introduce a framework for quantified experimental design of fusion diagnostics. Plasma diagnostics are subject to significance requirements as well as various constraints. Constraints may arise from hardware restrictions (e.g. limited access), limited financial budget or soft constraints such as optimal performance under specific experimental scenarios or with respect to derived (model-) quantities. We want to have experiments which provide small estimation uncertainties as well as experiments which solve physical problems such as deciding about the validity of different physical models.

Any method for performing experimental design must alter either the experimental setup or the uncertainty in the data. The concept of information theoretic measures such as Shannon's information entropy is used to construct diagnostic measures. Entropies and Kullback-Leibler divergences are relevant to Bayesian inference and design problems [2].

BAYESIAN EXPERIMENTAL DESIGN

The data D are usually uncertain due to statistical and systematical uncertainties and due to the lack of knowledge about the behavior of the physical system to be studied. In addition, D is subject to changes in the physical scenarios such as parameter scans. Since the design has to be optimal with respect to all relevant future data, the quantity to be maximized in Bayesian experimental design is the expected (marginalized) utility

$$EU(\eta) = \int dD\, P(D|\eta)\; U(D,\eta). \tag{2}$$

Note that the weight $P(D|\eta)$ is not the frequency distribution of data to be expected in the future experiment but the distribution of data from scenarios relevant for an optimal design which is closely related to the pdf of parameters $P(\theta)$ (see below).

The utility function has to be chosen according to the desired benefit of the experiment to be designed. A trivial outcome is a parameter estimate with small estimation uncertainties (co-variates). For an experiment designed to work properly under different operational scenarios and for different physical quantities of interest small estimation uncertainties should be provided at best under all circumstances. Optimality criteria may compete if the reduction of estimation uncertainties in one scenario implicates larger uncertainties in another one or if the estimation uncertainties of single quantities compete with uncertainties of derived quantities such as gradients.

Closely correlated to the goal of small estimation uncertainties is the concept of information gain provided by an experiment. Information gain is the excess information provided by the measured data relative to the prior knowledge. A measure of information gain is given by the Kullback-Leibler (KL) distance (mutual information, negative information entropy) between the posterior pdf $P(\theta|D,\eta)$ and the prior pdf $P(\theta)$

$$U_{\mathrm{KL}}(D,\eta) = \int d\theta\; P(\theta|D,\eta)\; \log \frac{P(\theta|D,\eta)}{P(\theta)}. \tag{3}$$

Integrating over the parameter space, the KL distance is a measure of what we can learn from the experimental data D.

The KL distance provides an absolute measure of information gain. If we use the base-2 logarithm then the information gain is measured in bits. It is a distance in the space of probability distributions.

Using $U_{\mathrm{KL}}(D,\eta)$ we obtain the expected utility function

$$EU_{\mathrm{KL}}(\eta) = \iint dD\, d\theta\; P(D|\eta)\; P(\theta|D,\eta)\; \log \frac{P(\theta|D,\eta)}{P(\theta)} \tag{4}$$

which can be reformulated according to the Bayes theorem,

$$\begin{aligned} EU_{\mathrm{KL}}(\eta) &= \iint dD\, d\theta\; P(\theta)\; P(D|\theta,\eta)\; \log \frac{P(\theta|D,\eta)}{P(\theta)} && (5)\\ &= \iint dD\, d\theta\; P(\theta)\; P(D|\theta,\eta)\; \log \frac{P(D|\theta,\eta)}{P(D|\eta)} && (6) \end{aligned}$$

where

$$P(D|\eta) = \int d\theta \, P(D|\theta,\eta) \, P(\theta) \tag{7}$$

is the prior predictive value (ppv) or evidence of the data.

The benefit of reformulation is that it is sufficient to specify the pdf of parameters $P(\theta)$ instead of $P(D|\eta)$. The "prior" $P(\theta)$ used in the EU is chosen to integrate all scenarios relevant for an optimal design. It does not depend on the design parameters η. Design relevant parameters θ generates statistical data according to the likelihood pdf $P(D|\theta,\eta)$, which does depend on the design parameters η.

For being a useful measure of information, the expected utility function must fulfill simple properties in the marginal cases (a) no data, (b) data totally uninformative about the parameters, and (c) exact knowledge about the parameters. If we have no data at all (case a) than the posterior pdf is equal to the prior pdf and, according to equation (5), $EU(\eta) = 0$. No data means no additional information and the information gain is correctly expected to be zero. If the experiment is totally ignorant about the chosen parameters (case b), which means that the data descriptive model is independent of the parameters of interest, than the likelihood and the ppv are equal and, hence, the expected utility function is zero again (6). The third case (c) for exact prior information about the parameters results, once again, in $EU(\eta) = 0$. This can easily be proven from (6) and (7) using a prior distribution which shrinks to a delta function around the parameter values assumed to be known exactly. All three cases do not only show the correct properties of the expected utility function chosen but are also useful for technical reasons: the numerical code evaluating the EU can be verified easily by checking the three cases (a)-(c). If Monte-Carlo methods are used to estimate EU the deviations of the estimates from zero are measures of the Monte-Carlo uncertainty.

Remark: Fraction of bits?

The results of the application presented below will show an information gain in terms of bits. What is the meaning of an information gain by a fraction of a bit? The information gain is intimately connected with an estimation uncertainty. The expected estimation uncertainty decreases with the increase of the information gain. Since we are integrating over data and parameter space the expected information gain is an average quantity for all possible experimental scenarios. Hence, for the general case we can not convert the number of bits of the expected information gain directly into estimation uncertainties. But for a simple example of a Gaussian distribution with standard deviation σ, the expected Shannon entropy is proportional to $-\log(\sigma)$ (neglecting a constant). It increases with decreasing σ as one would expect of a measure of information. Assuming we want to estimate the mean of N data values than the uncertainty of the mean estimate μ' is $\Delta\mu' = \sigma/\sqrt{N}$. Here an increase of the expected Shannon entropy by a fraction of a bit can be easily converted into a decrease of estimation uncertainty. Figure 1 shows the expected Shannon entropy difference (in bits) as a function of the relative change of σ. An increase of the expected Shannon entropy of 0.5 bits is achieved if σ is reduced by a factor of 0.7 which results in a 30% smaller estimation uncertainty for the mean value. Doubling the estimation precision results in an information gain of 1 bit!

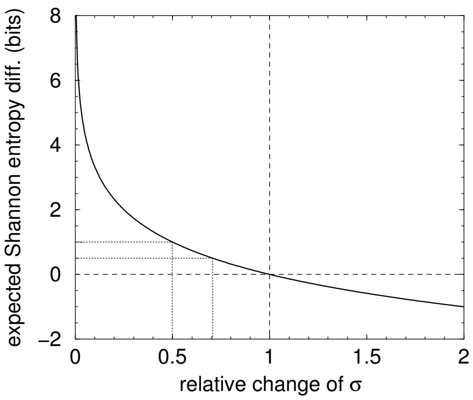

FIGURE 1. Expected Shannon entropy difference as a function of the relative change of the standard deviation of a mean estimate

Evaluating the expected utility

For finding the best experimental design $\hat{\eta}$ an effective algorithm to calculate the expected information gain $EU_{\text{KL}}(\eta)$ has to be used.

For problems where the distribution of the noise (likelihood pdf) is not dependent on the parameters and the design parameters, the Maximum-Entropy Sampling (MES) can be applied [3]. This means that instead of maximizing $EU_{\text{KL}}(\eta)$ of (6), we can maximize the information entropy IE (= negative Shannon entropy) of the ppv

$$IE(D|\eta) = -\int dD\, P(D|\eta)\ \log P(D|\eta). \tag{8}$$

Thus the best experimental design is the design for which the information entropy of the predictive distribution is maximized.

For problems where MES applies nested Monte Carlo methods can be used to calculate $IE(D|\eta)$ as a function of the design parameters η [4]. For each η, a data vector is sampled from $P(D|\eta)$ by first drawing a set of parameter values from $P(\theta)$, and then drawing a data vector from the sampling distribution $P(D|\theta,\eta)$ with those parameters. $P(D|\eta)$ (7) can be calculated for that data vector by quadrature (for low-dimensional parameter spaces) or by Monte Carlo integration. The author made good experience with the VEGAS algorithm which could be modified for a reasonable first guess for the initial grid [5]. Repeating this process and averaging the negative logarithm of $P(D|\eta)$ provides a Monte Carlo estimate of equation (8).

For problems where the likelihood depends explicitely on the design η we have to calculate $EU_{\text{KL}}(\eta)$ of (6). Such problems arise for example when the number of measured data enters η, when the length of a measurement is a design variable which enters the measurement uncertainty, or when the measurement uncertainty explicitely depends on the parameters, e.g. for Poisson distributions. The method to calculate $EU_{\text{KL}}(\eta)$ from (6) is similar to MES. First we draw a pair of values (D,θ) from the joint distribution $P(D,\theta|\eta)$ by drawing subsequently a set of parameter values from $P(\theta)$ and

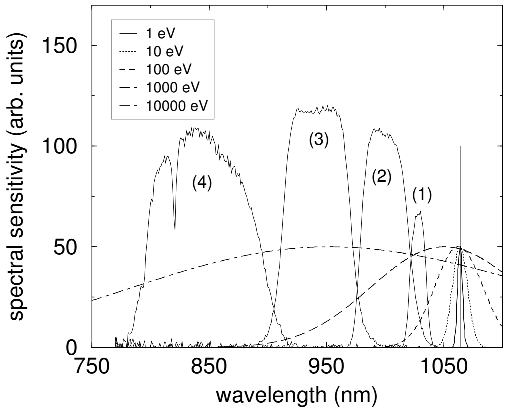

FIGURE 2. 4 spectral filters as a function of the wavelength and 5 different Thomson scattering functions for temperatures $T_e = 1\text{eV} - 10\text{keV}$.

a data vector from $P(D|\theta,\eta)$ using the sampled parameter. Then, the logarithm of the ratio of the likelihood and the ppv, $\log[P(D|\theta,\eta)/P(D|\eta)]$, is averaged.

EXAMPLE: THOMSON SCATTERING

Our work on experimental design is motivated by the Wendelstein 7-X stellarator fusion device which is in its construction phase. The fusion device was designed to study advanced stellarator concepts for a future fusion power plant. The physical scenarios to be studied require elaborate experimental diagnostics capabilities to be designed for best performance under a set of wide plasma parameters.

One of the most important ("level 0") plasma diagnostics is the Thomson scattering (TS) diagnostics for measuring the electron temperature T_e and electron density n_e profiles by means of laser scattering on electrons in a hot plasma. TS provides spatially and temporally resolved T_e and n_e estimates in the plasma core as well as in the edge [6]. We will focus in this paper on a single spatial channel close to the center of the plasma.

The present example of a simplified future TS experiment has a two-dimensional parameter vector $\theta = (T_e, n_e)$. The polychromator equipped with interference filters allows one to obtain spectrally resolved information. An abbreviated physical model of the measurement of the TS light is given by

$$D_j \propto n_e \int \tau_j(\lambda)\, S(\lambda, T_e)\, d\lambda, \tag{9}$$

where spectral filters $\tau_j(\lambda)$, $j = 1 \ldots 4$, cut out bands of the scattering function $S(\lambda, T_e)$ of the spectral shape of the scattered light of a laser beam. The likelihood is assumed to be a Gaussian with variance $\sigma^2 = \sigma_0^2 + \xi D$, where σ_0^2 is an offset and ξ is a scale. Both quantities are determined experimentally. MES does not apply for this likelihood pdf because the variance depends on the data, and, thereby, implicitly on the design parameters. The relative intensities of the spectral channels j allow to infer T_e and the absolute intensities allow to infer n_e. The absolute calibration is done with Raman

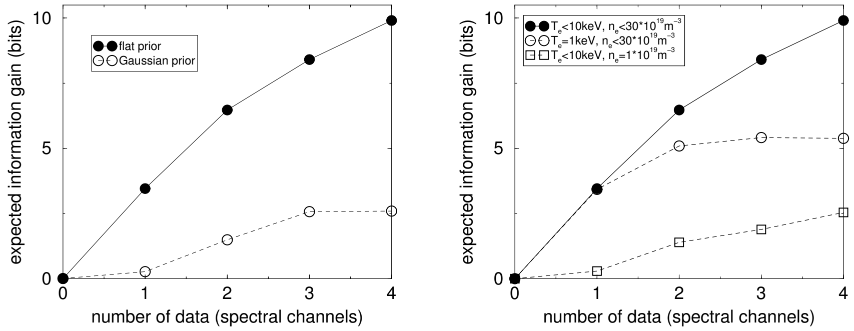

FIGURE 3. Expected information gain in units of bits as a function of the number of spectral channels used to measure the Thomson scattering light. Left: The full circles (open circles) depict the expected information gain for a prior distribution constant in the full parameter range (Gaussian in a small parameter space), respectively. Right: Decomposition of the information gain assuming either T_e (open circles) or n_e (open squares) to be known exactly.

scattering. Figure 2 shows 4 typical spectral filters as a function of the wavelength λ and 5 different scattering functions for temperatures $T_e = 1\text{eV} - 10\text{keV}$. At low temperatures the scattering function is centered at the laser wavelength (1064 nm) whereas with increasing temperatures the scattering function broadens and the center shifts to smaller wavelengths due to relativistic effects. The spectral channels close to the laser line provide information about the small temperatures whereas those far from the laser line are necessary for the large temperatures.

An optimal diagnostics consists of a small number of spectral filters (financial and spatial limitations) and a sufficient overlap of the spectral bands with the scattering function for all physical scenarios. Design parameters η addressed in this work are given by the number of spectral filters, the wavelength position and band width of the filters.

The parameters envisaged for the W7-X stellarator are $T_e \leq 10$ keV and $n_e \leq 3 \times 10^{20}$ m^{-3} [7]. The maximum values of T_e and n_e can not be reached simultaneously because the plasma pressure ($\propto T_e \times n_e$) is limited. To simulate different scenarios relevant for an optimal design two extreme prior distributions $P(\theta)$ are chosen. The first prior is a Gaussian distribution centered at $T_e = 1$ keV and $n_e = 1 \times 10^{19}$ m^{-3}, with a standard deviation of 20% of the mean values, to simulate a dedicated experiment where we aim at an optimal design in a very narrow parameter region. The second, more realistic, prior is a constant in the full parameter range ($T_e \leq 10$ keV, $n_e \leq 3 \times 10^{20}$ m^{-3}). A comprehensive design will use refined prior distributions considering the correlated nature of T_e and n_e and considering small T_e and n_e values to be of minor interest depending on the spatial position in the plasma.

The first quantity of interest is a measure of information provided by a single spectral channel. We start with no data and add spectral channels in the order shown in figure 2, although there exists a large number of combinatorial sets of channels out of 4. Figure 3 shows the expected information gain in units of bits as a function of the number of spectral channels. The full (open) circles in the left panel depict the expected information gain for the constant (Gaussian) prior pdf, respectively. The expected information gain without a single measurement is zero as expected. The expected information gain for

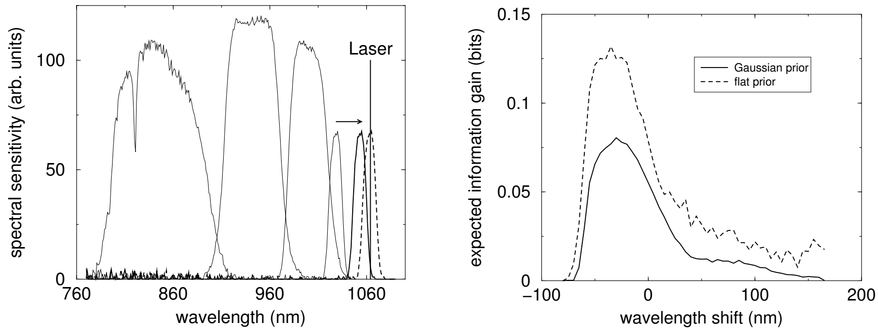

FIGURE 4. Information gain (right panel) as a function of the shift of one spectral curve (left)

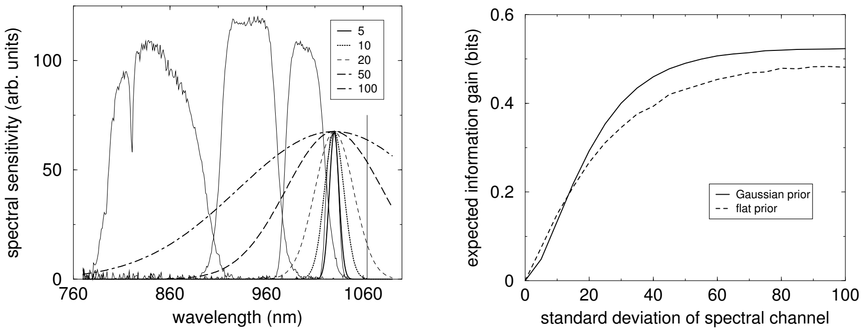

FIGURE 5. Information gain (right panel) as a function of the width of one spectral curve (left)

the constant prior is larger than for the Gaussian prior since we can learn more if the prior knowledge is less precise. Additionally, the shape of the curves differs for the two prior pdfs. The right panel of Figure 3 shows the contributions of T_e and n_e to the information gain for the constant prior only. We assume either T_e (open circles) or n_e (open squares) to be known exactly at the mean values of the Gaussian prior. If T_e is known precisely than a single channel provides significant information to estimate n_e. Additional channels increase the estimation precision but the information gain saturates already for 3 spectral filters. If n_e is known precisely than a single channel (here the one close to the laser line) provides poor information about T_e. At least a second channel is necessary for significant T_e estimation. Additional channels are useful and saturation is not observed within 4 spectral channels. Note that the information gain for single parameter estimates do not add to the value for joint parameter estimates.

Another design parameter is the position of the spectral bands. The left panel of figure 4 shows shifts of the filter closest to the laser line in the direction of the laser line. The two shifted spectral bands result from the maximum of the information gain as a function of the shift parameter. For both prior pdfs it is favorable to chose one spectral filter close to the laser line. In order to obtain reliable results for small T_e values (constant prior) one spectral band should be very close to the laser line. A detailed study will require simultaneous optimization of filter positions which is beyond the scope of this paper.

The third quantity of interest in TS design is the width of the spectral bands. The left panel of figure 5 shows the replacement of the filter close to the laser line with a Gaussian function of varying width. The right panel depicts the resulting information gain which is a monotonic function of the filter band width (standard deviation). The original band has a standard deviation of about 5 nm. The information gain can be increased by a factor of 5 if one would use a standard deviation of about 50 nm. This is mainly due to the increase of measurement intensity and decrease of relative measurement uncertainty.

CONCLUSION AND OUTLOOK

We introduced a framework for quantified experimental design of fusion diagnostics and applied it to a Thomson scattering experiment. A comprehensive design of a future TS system in Wendelstein 7-X requires elaborate prior distributions and hardware constraints. Further design criteria will arise from the inclusion of calibration nuisance parameters, from robustness requirements dealing with systematic uncertainties, geometric considerations of optimal profile reconstructions, and from transport analyses which require accurate measurements on $\nabla T/T$, $\nabla n/n$ and E_r [8]. The design of meta-diagnostics (combination of heterogeneous and complementary diagnostics) will benefit from the use of interdependencies which compensate single-diagnostics drawbacks.

ACKNOWLEDGMENTS

Helpful discussions with A. Dinklage, A. van der Linde and J. Skilling are gratefully acknowledged. The author is indebted to V. Dose for his continuous support and stimulating discussions. This work would not be realized without V. Dose's inspiration and pioneering work for data analysis.

REFERENCES

1. Chaloner, K., and Verdinelli, I., *Stat. Sci.*, **10**, 273–304 (1995).
2. Müller, P., and Parmigiani, G., "Numerical Evaluation of Information Theoretic Measures," in *Bayesian Statistics and Econometrics: Essays in Honor of A. Zellner*, edited by D. A. Berry, K. M. Chaloner, and J. F. Geweke, Wiley, New York, 1995, pp. 397–406.
3. Sebastiani, P., and Wynn, H. P., *J. Roy. Stat. Soc. B*, **62**, 145–157 (2000).
4. Loredo, T. J., "Bayesian Adaptive Exploration," in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, edited by G. Erickson and Y. Zhai, AIP, Melville, NY, 2004, pp. 330–346.
5. Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P., *Numerical Recipes*, Cambridge University Press, Cambridge, 1992, 2nd edn.
6. Fischer, R., Wendland, C., Dinklage, A., Gori, S., Dose, V., and the W7-AS team, *Plasma Phys. Control. Fusion*, **44**, 1501–1519 (2002).
7. Hartfuss, H. J., Brakel, R., Endler, M., Geist, T., Grigull, P., Hofmann, J. V., Junker, J., Kick, M., Kühner, G., Niedermeyer, H., Ringler, H., Teubel, A., Wagner, F., and Weller, A., *Rev. Sci. Instrum.*, **68**, 1244–1249 (1997).
8. Dinklage, A., Fischer, R., Dreier, H., Svensson, J., and Turkin, Y., "Integrated Approaches in Fusion Data Analysis," in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, edited by R. Fischer, R. Preuss, and U. von Toussaint, AIP, Melville, NY, 2004, in this proceedings.