Statistical significance of theoretical predictions: A new dimension in nuclear structure predictions (I)

J DUDEK¹, B SZPAK², M-G PORQUET³ and B FORNAL²

¹Institut Pluridisciplinaire Hubert Curien Département de Recherches Subatomiques UMR7178, IN2P3-CNRS and Université de Strasbourg, 23 rue du Loess B.P. 28, F-67037 Strasbourg Cedex 2, France
²The Niewodnicza´ nski Institute of Nuclear Physics, Polish Academy of Sciences, ul. Radzikowskiego 152, Pl-31-342 Kraków, Poland
³Centre de Spectrométrie Nucléaire et de Spectrométrie de Masse, IN2P3-CNRS and Université Paris-Sud, Bâtiments 104 et 108, F-91405 Orsay Campus, France

E-mail: Jerzy.Dudek@IReS.in2p3.fr, Bartlomiej.Szpak@ifj.edu.pl, Marie-Genevieve.Porquet@csnsm.in2p3.fr, Bogdan.Fornal@ifj.edu.pl

Abstract.

In this and the follow-up article we briefly discuss what we believe represents one of the most serious problems in contemporary nuclear structure: the question of statistical significance of parametrizations of nuclear microscopic Hamiltonians and the implied predictive power of the underlying theories. In the present Part I, we introduce the main lines of reasoning of the so-called Inverse Problem Theory, an important sub-field in the contemporary Applied Mathematics, here illustrated on the example of the Nuclear Mean-Field Approach.

1. Introduction

In societies with long history, within their intellectual or political development there arise issues that become a common secret: subjects whose discussing is rather avoided despite the fact that the problems behind are both important and disturbing and everybody knows about their existence. Similarly, in scientific communities certain themes become kind-of-embarrassing and we may even develop a tendency of systematically avoiding the issue. The nuclear physics community with over 100 years of existence has produced already some such delicate subjects. One of them is the actual subject of this article: it has to do with (possibly un-) stable parametrizations of the nuclear structure Hamiltonians, the avoided issue being parametrization’s statistical significance and the Hamiltonian’s (likely missing) predictive power.

The first signs of theory’s missing predictive power are: a) Quickly growing in the literature a number of alternative (thus generally incompatible) parametrizations of the same Hamiltonian and b) Diverging predictions from various parametrizations when extrapolating to nuclear areas that are experimentally unknown at a given time. To illustrate the item a) above: Our community has produced to date over 130 various parametrizations of the effective Hamiltonian of Skyrme Hartree-Fock mean field theory. Moreover, in other microscopic theories parametrizations are sometimes accepted that vary from one narrow nuclear range to another or even from one nucleus to the next. Manifestations of the same problem can be found in yet
another important microscopic method: the nuclear shell-model. To illustrate the origin of the problem b) above, let us quote the authors [1] who remark with strong criticism:

"Unfortunately, many practitioners of parameter estimation never proceed beyond determining the numerical values of the parameter fit. They deem a fit acceptable if a graph of data and model 'looks good'. This approach is known as chi-by-the-eye. Luckily, its practitioners get what they deserve"

[i.e. - 'they' get a 'statistical nonsense', see the above reference for explanatory comments].

The question of predictive power of physical theories and statistical significance of theoretical predictions is a vast and quickly growing field of mathematics, so far applied mainly in domains of science other than nuclear physics. These applications range from geology and geophysics to social and political sciences, image modelling, medicine etc., etc. Within the volume of this presentation we will be able to touch upon a small number of selected issues and illustrations belonging to the underlying sub-field of Applied Mathematics: the Inverse Problem Theory1 one of the most important technical tools of the general Inference Theory. The former deals with the question of a mathematically correct extracting the mathematical model’s adjustable parameters and determining their statistical significance. As for the latter, it has to do with the mathematical treatment necessary to obtain the statistically significant information about all the data through theoretical modelling - while having only very limited sample of experimental data at our disposal - as an actual information-input.

2. Statistical significance of a Hamiltonian and predictive power of the theory
Examine a physical system we usually wish to obtain ‘the full truth’, the latter mathematically represented by the solutions of the eigen-value problem \( H \psi_n = e_n \psi_n \) in which \( H = H^{true} \) is supposed to contain the complete information. Since we believe in the objective existence of studied physical objects and in an ever-lasting improvement of our understanding, the adjective ‘complete’ signifies here both: already known and yet undiscovered. It then follows that the true solutions \( \{ e_n, \psi_n \} \) remain asymptotic in time devoted to the research, rather fictitious, ideal and to an extent unknown objects and that, to the same extent, the desired full truth remains unknown to us. To say: unknown - is an expression of the lack of knowledge synonymous with ignorance, here imposed by nature rather than human laziness. Humans do not know any efficient cure against deliberate laziness, but they did invent a cure against forced ignorance. The solution: Parametrize our ignorance; more precisely: estimate by all available means which answer is more-, and which less-likely the right solution. In other words we will need to find relative probability of what we think the right answer is! In Applied Mathematics, the ignorance is usually represented by a random variable \( X \), quantified by its probability distribution \( P_X = P_X(x) \), \( x \) called ‘realisation’ of the random variable variable \( X \). In short: The way to progress will be to learn diminishing the effects of ignorance through parametrizing it in terms of probabilities.

To be able to illustrate the formalism with the realistic situations, we will focus on the simple case of nuclear phenomenological mean-field theory, otherwise our considerations are general.

---

1. The term ‘Inverse Problem Theory’ of Applied Mathematics has not much in common with its traditional meaning in quantum mechanics, where for a long time it referred to finding the form of the potential in the Schrödinger equation out of bound states and scattering cross-sections. When we took as a somewhat arbitrary measure of intellectual activity the number of hits in Google search to compare two fields, the nuclear physics and the inverse problem, on Sunday, May 16, 2010, at 10:30 AM, we found

"Nuclear Physics" → 3,200,000 results (about 100 years of active history)
"Inverse Problem" → 510,000 results (about 20 years of active history)
2.1. Theoretical and experimental errors as sources of uncertainties: Probability of the truth

At the origin of our ignorance lie errors and here we distinguish between their two sources:

a) Often discussed experimental errors, and, much more important and almost never examined:
b) The theory errors originating from the fact that our many-body theories are incomplete:

\[
\text{Theory} \rightarrow e_n = e_n^{\text{true}} + \delta e_n^{\text{error}}(p) \quad \& \quad \varepsilon_n = \varepsilon_n^{\text{true}} + \delta \varepsilon_n^{\text{error}} \rightarrow \text{Experiment}
\]

Above, both the theoretical and experimental energies, \(e_n\) and \(\varepsilon_n\), respectively, are random variables characterised by probability distributions \(P_n^{\text{th}}(e_n)\) and \(P_n^{\exp}(\varepsilon_n)\). Errors propagate to the theory predictions through parameter fits

\[
\chi^2(p) \sim \sum w_n \left[ (\varepsilon_n^{\text{true}} + \delta \varepsilon_n^{\text{error}} - e_n^{\text{true}}) \right]^2 \quad \text{Theory} \nonumber \quad \frac{\partial \chi^2}{\partial p} = 0.
\]

The final theory predictions will depend on an appropriate convolution between the error probabilities \(P_n^{\text{th}}(e)\) and \(P_n^{\exp}(\varepsilon)\) leading to the Hamiltonian-parameter probability distributions \(P(p)\) and, through the latter, to the error probability distribution of the final predictions for observables. In other words, we suggest an extended strategy in constructing theoretical models:

*Physical theories should provide not only the numerical predictions but also probabilities with which they are believed to give right answers when compared to experimental data.*

*Studying these probabilities opens a new dimension*\(^2\) *in modelling nuclear phenomena.*

A new formulation (or re-formulation) of theories will require a thorough investigation of the exact forms of the theory’s error probability distributions \(P_n^{\text{th}}(e)\). We cannot offer such a full solution at this time, the whole field of research being at its initial stage. Instead, we will focus in this text on the underlying mathematical aspects and in the following one \([3]\) on the Hamiltonian represents only a part of physical reality whereas the true masses are contributed by the complete, true Hamiltonian. Consequently, in the spirit of the ‘chi-by-the-eye method’, the Hamiltonian’s parameters are forced to compensate for the absence of certain unknown fragments of this true Hamiltonian - the task that is impossible to keep mathematically under control since the true nuclear Hamiltonian is, as usual, not known.

\(^2\) Obviously, the word *dimension* should not be taken literally but rather as a nickname for a new direction of research combining the application of advanced methods of probability theories with a new type of physics tasks: modelling the impact of errors caused by missing elements of theories, for instance missing terms in the underlying Hamiltonians - as well as constructing the underlying probability distributions (‘theory of theoretical errors’).

\(^3\) It is well known that some authors include the nuclear masses to fit the HF Hamiltonian parameters. Such a choice has necessarily a negative impact on the predictive power of the model. This is obvious since the mean-field Hamiltonian represents only a part of physical reality whereas the true masses are contributed by the complete, true Hamiltonian.
2.2. Parameters of Hamiltonians and ill-posedeness of the $\chi^2$-minimisation problem

The Hamiltonian’s parameter adjustment usually corresponds to a $\chi^2$-minimisation

$$\min_{\{p_k\}} \{ \chi^2(\{p\}) = \sum_{j=1}^{n_d} W_j \left[ e^{\text{exp}}_j - e^{\text{th}}(\{p\}) \right]^2 \} \rightarrow \frac{\partial \chi^2}{\partial p_k} = 0, \text{ for } k = 1 \ldots n_p, \tag{4}$$

where $n_d$ denotes the number of data points, $n_p$ - number of model parameters and $\{W_i\}$ - the physicist-defined set of weight factors. In practical realisations of the inverse problem theory one usually introduces a linearization procedure after simplifying the notation $e^{\text{th}}(\{p\}) \rightarrow f_j(\{p\})$.

$$f_j(\{p\}) \approx f_j(p_0) + \sum_{k=1}^{n_p} \left( \frac{\partial f_j}{\partial p_k} \right)_{p=p_0} (p_k - p_{0,k}). \tag{5}$$

Introducing further abbreviations

$$J_{jk} \equiv \sqrt{W_j} \left( \frac{\partial f_j}{\partial p_k} \right)_{p=p_0} \text{ and } b_j = \sqrt{W_j} \left[ e^{\text{exp}}_j - f_j(p_0) \right] \tag{6}$$

one may show that the minimised function takes the form

$$\chi^2(\{p\}) = \sum_{j=1}^{n_d} \left[ \sum_{k=1}^{n_p} (p_k - p_{0,k}) - b_j \right]^2. \tag{7}$$

It then follows that the relation within the sought parameters $p$ and the input data $b$ is linear

$$\frac{\partial \chi^2}{\partial p_k} = 0 \iff (J^T J) \cdot (p - p_0) = J^T b \rightarrow (p - p_0) = (J^T J)^{-1} J^T b. \tag{8}$$

In Applied Mathematics one usually changes the wording and notation: the set of parameters $\{p\}$ being called ‘causes’ will be denoted $x$ and the resulting quantities, in our case a set of energies $\{e\}$, called ‘effects’, is equivalently hidden in $b$. Observe that generally $J$ is an $n_d \times n_p$ dimensional, i.e., a non-trivially rectangular rather than square matrix since we tend to make the numbers of parameters as small as possible and the number of data points as large as possible and thus $x \in \mathbb{R}^{n_d}$, and $b \in \mathbb{R}^{n_p}$. This is also why it is convenient to use a non-typical notation in (8) that allows to involve explicitly the square matrix $(J^T J)$.

Thanks to the linearized representation in (8) the problem of instabilities of the solutions to the $\chi^2$-minimisation, also called ill-conditioning, can be very conveniently posed using the so-called Singular-Value Decomposition theorem applied to $J$. The latter can always be decomposed as:

$$J = U \cdot D \cdot V^T \text{ with } U \in \mathbb{R}^{n_d \times n_d}, \ V \in \mathbb{R}^{n_p \times n_p}, \ D \in \mathbb{R}^{n_d \times n_p}, \tag{9}$$

where diagonal matrix has a form $D = \text{diag}\{d_1, d_2, \ldots d_{\min(n_d, n_p)}\}$. Formally the solution $x$ is

$$x = \left( J^T J \right)^{-1} J^T b \text{ with } \left( J^T J \right)^{-1} J^T = V \cdot D^{-1} \cdot U^T \tag{10}$$

where

$$D^{-1} = \text{diag}\{\frac{1}{d_1}, \frac{1}{d_2}, \ldots \frac{1}{d_p}, 0, 0, \ldots 0\}. \tag{11}$$

Using the Singular-Value Decomposition one shows, within schematic but an obvious notation, that $(J^T J)^{-1} = V \frac{1}{d^2} V^T$. Independently one can derive the expression for the correlation matrix

$$\langle (p_i - \langle p_i \rangle) \cdot (p_j - \langle p_j \rangle) \rangle = \chi^2(p) \frac{1}{\alpha_{2, n_d-n_p}} \left( J^T J \right)_{ij}^{-1} \tag{12}$$
where \( t_{\alpha/2,n_d-n_p} \) is the usual Student t-probability distribution function. After having defined next the two standard quantities, viz. the so-called unbiased estimator \( \Delta_{\text{rms}} \) for the root-mean-square deviation and the variance-covariance matrix \( M \), respectively, by

\[
\Delta_{\text{rms}}^2 \equiv \frac{\chi^2}{n_d - n_p} \quad \text{and} \quad M \equiv \Delta_{\text{rms}}^2 (J^T J)^{-1} \tag{13}
\]

we obtain the estimates for the 100\%-\( \alpha \) confidence interval for the least-square estimator \( \hat{p}_i \) of the parameter \( p_i \)

\[
p_i \in [\hat{p}_i - t_{\alpha/2,\nu} \sqrt{M_{ii}}, \hat{p}_i + t_{\alpha/2,\nu} \sqrt{M_{ii}}] \quad \text{with} \quad \nu \equiv n_d - n_p \tag{14}
\]

As it is well known, some diagonal elements of the \( D \)-matrix can become close to zero or actually vanish - and this property depends generally both on the properties of the actual Hamiltonian as well as on the selection of the data points (in other words: on the experimentalist). We thus arrive at the following important conclusion:

> If one or more diagonal values \( d_k \to 0 \), it follows that \( (J^T J)^{-1} \) tends to infinity and consequently the correlation matrix and the confidence intervals of all parameters diverge.

The theory predictions lose stability - thus the theory in question loses its all predictive power - and yet ‘the graph of the solution will usually still look good’ - except that the latter property is now dependent on the experimentalist who measured the data or on the theorist who selected some data points for his/her tests.

In short: full statistical in-significance referred to as statistical nonsense [and yet - cf. footnote 4].

[The reader is here referred to the recent publications [4, 5] addressing very similar problems.]

2.3. Elementary tests of stability: Single-point vs. Monte-Carlo approaches, Overfitting

The calculation schematised in (12)-(14) can be performed using just one single minimisation run - and information of this type is referred to as ‘single point’ or local. Its advantage is relative simplicity but it has also a disadvantage of not representing any global features of the problem as for instance the sensitivity of various observables (such as e.g. single-particle energies) with respect to specific uncertainties in function of the Hamiltonian parameters or, what is related, the fact that some calculated levels may be much more sensitive to uncertainties in inverse-problem input data points as compared to some other levels. Information of this type is illustrated in figure 1 showing the positions of various levels varying in function of the fit-input energy of a given level: here \( e_{2p_{9/2}} \in [-2, +2] \) MeV. Let us emphasise that the information of this type may be very precious if we wish to model with precision the positions of certain levels: illustrations like the one in figure 1 show the ‘strength of the coupling’ thus indicating which levels will need to be measured with precision if we wish to model a particular class of other levels with precision.

Another way of testing the global features of the model with respect to the data sample at hand involves the Monte-Carlo simulations. Within the latter approach we assume a certain probability distribution representing an uncertainty of the input data, for instance we suppose that each input level has e.g., Gaussian energy-position probability \( P_i(e) \sim \exp[-(e-e_i)^2/2\sigma_i^2] \) (or impose, deduce, derive ... an alternative probability distribution). One of the first possible applications is to test for an over-fitting - a mechanism influencing the model’s predictive power.

---

4 This does not imply that the theory is in any fundamental trouble since according to the existing methods the negative effects of small or vanishing diagonal (singular) values can be kept easily under control. The problem is: To bother doing the necessary in place of the chi-by-the-eye approach.
Figure 1. Global correlations in response of the model to various minimisation-input data, are represented by the variation of several (in principle all) single particle levels of the system as function of a single input point, in this example - the experimental $e_{2g9/2}$-level energy. At each of the latter energies the fit of the Hamiltonian parameters is repeated and new parameter set and thus new energies are obtained. Observe a presence of levels nearly insensitive to the variation in input chosen (here: $e_{2g9/2}$) - but there is a richness of possibilities varying from level to level.

Figure 2. An illustration of the over-fitting of a Hamiltonian studied using Monte-Carlo approach. With the fit-input spectra sampled according to the Gaussian distributions (referred to as ‘noise’) discussed in the text, here: $\sigma_i = \sigma$, the fitting is repeated 10 000 times, the histograms giving occurrences of any given $\chi$ are produced representing related probabilities. The over-fitting takes place when the data points are not sufficiently constraining and it is possible to find as most probable the $\chi^2 = 0$ solution starting from any input.

3. Summary and Conclusions
We have discussed statistical significance of theory predictions and of the parametrizations of the underlying Hamiltonians. We introduced selected mathematical elements of the Inverse Problem Theory adapted to the application to nuclear Hamiltonians. We formulated the problem of instability of the Hamiltonian’s parameter adjustment related to the loss of model’s predictive power [for illustrations cf. follow-up in [3]]. As an important element of the future evolution in nuclear theories we suggest a reformulation of their strategic constructions such that the theory will provide also the probability of the validity of its actual prediction.

Acknowledgements
This work is supported by the Polish Ministry of Science and Higher Education, under contract N-N202-263238, by LEA COPIGAL and by COPIN-IN2P3 collaboration agreement No.06-126.

References
[1] Press W H, Teukolsky S A, Vetterling W T and Flannery B P 1992 Numerical Recipes in FORTRAN 77: The Art of Scientific Computing (Cambridge: Cambridge University Press)
[2] Dudek J et al., 2010 J. Phys. G: Nucl. Part. Phys. 37 064031
[3] Szpak B, Dudek J, Porquet M-G and Fornal B, 2010 J. Phys.: Conference Series This Volume
[4] Toivanen J et al., 2008 Phys. Rev. C 78, 054306
[5] Reinhard P-G and Nazarewicz W, 2010 Phys. Rev. C 81, 051303(R); Kortelainen M et al., 2010 Phys. Rev. C 82, 024313