



KfK 4099  
August 1986

# Principles and Techniques of Data Evaluation

F. H. Fröhner  
Institut für Neutronenphysik und Reaktortechnik  
Projekt Schneller Brüter

**Kernforschungszentrum Karlsruhe**



KERNFORSCHUNGSZENTRUM KARLSRUHE

Institut für Neutronenphysik und Reaktortechnik  
Projekt Schneller Brüter

KfK 4099

Principles and Techniques of Data Evaluation

F.H. Fröhner

Kernforschungszentrum Karlsruhe GmbH, Karlsruhe

Als Manuskript vervielfältigt  
Für diesen Bericht behalten wir uns alle Rechte vor

Kernforschungszentrum Karlsruhe GmbH  
Postfach 3640, 7500 Karlsruhe 1

ISSN 0303-4003

## PRINCIPLES AND TECHNIQUES OF DATA EVALUATION \*)

### Abstract

The probabilistic foundations of data evaluation are reviewed, with special emphasis on parameter estimation based on Bayes' theorem and modern methods concerning prior probabilities. The process leading from raw data to evaluated data files is briefly explained for the example of nuclear reaction cross sections, with a discussion of systematic and statistical errors and their propagation. It is shown how evaluators can establish data covariances if error components are properly specified by experimentalists, and how correlated data uncertainties in covariance files are utilised in sensitivity studies and accuracy assessments. The problem of inconsistent data is also addressed briefly.

## GRUNDLAGEN UND METHODEN DER MESSDATEN-AUSWERTUNG

### Zusammenfassung

Ein Überblick wird gegeben über die wahrscheinlichkeitstheoretischen Grundlagen der Meßdaten-Auswertung, mit besonderer Betonung der Parameterschätzung mit Hilfe des Satzes von Bayes und moderner Erkenntnisse über a-priori-Wahrscheinlichkeiten. Der Weg von Rohdaten bis hin zu evaluierten Dateien wird kurz erläutert am Beispiel der Wirkungsquerschnitte für Kernreaktionen, mit Diskussion der statistischen und systematischen Fehler und ihrer Fortpflanzung. Es wird gezeigt, wie die Auswerter Datenkovarianzen ermitteln können, wenn die Fehlerkomponenten von den Experimentatoren angemessen dokumentiert sind, und wie die korrelierten Datenunsicherheiten aus Kovarianz-Dateien bei Empfindlichkeitsstudien und Fehlerabschätzungen verwendet werden. Das Problem der Unverträglichkeit von Daten wird ebenfalls kurz behandelt.

---

\*) This report is based on lecture notes for the Ispra Course on Data Uncertainty, Sensitivities, Consistency and Adjustment, JRC Ispra, 14-18 April 1986

## TABLE OF CONTENTS

1. THE MATHEMATICS OF DATA EVALUATION	1
1.1 Probability as a Numerical Scale for Rational Expectation	1
1.2 Bayes' Theorem as the Rule for Updating Knowledge with New Data	2
1.3 Generalisation to More Observations and More Parameters	5
1.4 Closer Look at Prior Probabilities, Group-Theoretical Assignment	7
1.5 Bayesian Parameter Estimation for the Univariate Gaussian	8
1.6 Assignment of Priors by Entropy Maximisation	12
1.7 Approximations: Maximum Likelihood and Least Squares	14
2. EVALUATION OF NUCLEAR DATA FOR APPLICATIONS	16
2.1 Steps in the Preparation of Nuclear Data for Applications	16
2.2 International Cooperation	19
2.3 Iterative Least-Squares Fitting	20
2.4 Statistical Errors: the Poisson Distribution	22
2.5 Systematic Errors: Correlated Uncertainties and their Propagation	23
2.6 Inconsistent Data	26
3. CONCLUDING REMARKS	29
ACKNOWLEDGEMENT	30
FIGURES	31
REFERENCES	35

## 1. THE MATHEMATICS OF DATA EVALUATION

We begin with a brief review of the probability-theoretical foundations of data evaluation. This will help to tie together various rules for the extraction of "best" values and their uncertainties from experimental data, and recipes for data fitting and adjustment. Physicists have to learn these recipes usually during lab courses and on the job, and most of the textbooks on probability theory are not very helpful, being full of intimidating jargon and complicated "ad-hockeries" originating from desperate attempts to avoid Bayes' theorem with its a-priori probabilities. The following presentation, which (i) is squarely based on Bayes' theorem and (ii) utilises recent progress concerning prior probabilities, will be found to provide a concise and mathematically simple treatment of parameter estimation and data adjustment in the general context of inductive inference, or learning from observations.

### 1.1 Probability as a Numerical Scale for Rational Expectation

All our results will be fairly direct consequences of the basic product and sum rules of probability theory,

$$p(AB|C) = p(A|BC)p(B|C) = p(B|AC)p(A|C) , \quad (1)$$

$$p(A|B) + p(\bar{A}|B) = 1 , \quad (2)$$

where A, B, C denote various propositions such as "the coin shows head" or "the cross section is larger than 200 b", AB means "both A and B are true",  $\bar{A}$  means "A is false", and  $p(A|B)$  is the probability of A given B. This latter notation indicates that all probability assignments are conditional, based on either empirical or theoretical information or on assumptions. Following Laplace we shall interpret these probabilities as a numerical scale for degrees of rational expectation, ranging from 0 (impossibility) to 1 (certainty), intermediate values indicating intermediate degrees of plausibility.

This interpretation (not the equations themselves) is often criticised by statisticians who insist that by "probability" one must mean only "frequency in a random experiment" such as coin tossing, in the limit of very many repetitions, and that one can assign the "direct" probabilities of effects (observations) if the causes (natural laws and natural constants) are given, but never the "inverse" probabilities of causes if the observations are given. For scientists in general, and data evaluators in particular, this viewpoint is much too narrow. It would not permit them to say that a physical constant has, according to measured data, such and such a probability to lie within given confidence limits. The job to infer the values of natural constants, half-lives, reaction cross sections etc. from error-affected and incomplete experimental data is not a random experiment that can be repeated at will, but rather an exercise in inductive inference (reasoning in the face of uncertainty). For evaluation work Laplace's probability concept is, therefore, the appropriate one. Incidentally, R.T. Cox [1] proved in 1946 that degrees of plausibility can, in fact, always be represented by real numbers, within any consistent system of inductive reasoning. The most general consistency conditions take the form of functional equations whose solution is found to satisfy our basic two equations. Thus the mathematics is exactly the same for Laplacean and frequentist probabilities and the whole controversy has become rather academic.

## 1.2 Bayes' Theorem as the Rule for Updating Knowledge with New Data

A scientific theory enables us to calculate the "direct" probability of some observation if the physical quantities occurring in the theory are known. If they are unknown, but observations have been made, the situation is reversed. We must find the "inverse" probabilities for the various possible values of the physical constants from the given empirical data. This inversion is accomplished by Bayes' theorem. Its simplest form,

$$p(A|BC) = \frac{p(B|AC)p(A|C)}{p(B|C)} \quad (3)$$



is a direct consequence of the symmetry of the product rule (1) with respect to A and B. Laplace generalised this to the case of several distinct alternatives  $A_i$ ,

$$p(A_i|BC) = \frac{p(B|A_iC)p(A_i|C)}{\sum_i p(B|A_iC)p(A_i|C)}, \quad i=1,2,\dots,n. \quad (4)$$

With (1) it is easily verified that the normalising denominator is, in fact, equal to  $p(B|C)$  while the normalisation itself satisfies the sum rule (2). Bayes' theorem follows thus directly from the basic equations of probability theory. Especially in its form for continuous alternatives,

$$p(A|BC)dA = \frac{p(B|AC)p(A|C)dA}{\int p(B|AC)p(A|C)dA}, \quad A_L < A < A_H, \quad (5)$$

it can be considered as the very cornerstone of data evaluation. The usual situation is that we have data B which depend on the value of an unknown physical quantity A and on a theoretical model C of the experiment. If we know the mathematical form of the "likelihood function"  $p(B|AC)$ , in which the unknown quantity A appears as parameter, and also the a-priori probability ("prior" for short)  $p(A|C)dA$ , we can calculate the a-posteriori probability (or "posterior")  $p(A|BC)dA$  that the unknown parameter lies in the infinitesimal increment  $dA$  at A.

As a fairly realistic illustration let us consider the determination of the decay constant  $\lambda$  of some short-lived radioisotope from decays registered at times  $t_1, t_2, \dots, t_n$ . Obviously we must identify  $\lambda$  with A, and the data  $t_1, \dots, t_n$  with B, while C consists of all other information we have about the experiment such as applicability of the exponential decay law, purity of the sample, reliability of the recording apparatus, sufficiently long observation time for all observable decays to be recorded, etc. The so-called sampling distribution, i.e. the probability that a decay is recorded in a time interval  $dt$  at  $t$ , given  $\lambda$ , is

$$p(t|\lambda)dt = e^{-\lambda t} \lambda dt, \quad 0 < t < \infty. \quad (6)$$

(We simplify the notation by omitting explicit reference to C). The joint probability of observing the data, given  $\lambda$ , is then obtained as

$$p(t_1, \dots, t_n | \lambda) dt_1 \dots dt_n = \exp\left(-\lambda \sum_{i=1}^n t_i\right) \lambda^n dt_1 \dots dt_n, \quad 0 < t_i < \infty, \quad (7)$$

according to the product rule. This corresponds to  $p(B|AC)$  above. The probability density  $p(t_1, \dots, t_n | \lambda)$  is called the likelihood function. As the increments  $dt_i$  cancel upon normalisation, we get Bayes' theorem in the form

$$p(\lambda | t_1, \dots, t_n) d\lambda \propto p(t_1, \dots, t_n | \lambda) p(\lambda) d\lambda, \quad 0 < \lambda < \infty, \quad (8)$$

which illustrates the fundamental rule: The posterior probability is the product of likelihood function and prior probability (apart from a trivial normalisation constant). The likelihood function transmits the impact of the new data. In our example the data appear only in the form  $\sum_i t_i \equiv n\bar{t}$  so that for given  $n$  the sample average  $\bar{t}$  conveys all the information contained in the data. In statistical terminology  $\bar{t}$  is a "sufficient statistic",  $n$  an "ancillary statistic", statistic meaning any function of the "sample" (i.e. of the data).

If we consider all  $d\lambda$  for  $0 < \lambda < \infty$  as equally probable a priori, so that the prior is  $p(\lambda)d\lambda \propto d\lambda$ , we get as the final result of our Bayesian parameter estimation

$$p(\lambda | t_1, \dots, t_n) d\lambda = \Gamma(n+1)^{-1} e^{-x} x^n dx, \quad 0 < x \equiv \lambda n \bar{t} < \infty, \quad (9)$$

the gamma function  $\Gamma(n+1)$  ensuring proper normalisation. This posterior distribution, known as gamma distribution or as chi-square distribution with  $2n+2$  degrees of freedom, represents the complete information about  $\lambda$  which can be extracted from the data and the prior.

Now most users of nuclear decay data do not want to be bothered by the details of an a-posteriori distribution. What they want is a recom-

mended decay constant and its uncertainty, and nothing else. So we calculate the expectation value,  $\langle \lambda \rangle$ , and the square root of the variance,  $\delta \lambda$ ,

$$\langle \lambda \rangle = \int_0^{\infty} d\lambda p(\lambda | t_1, \dots, t_n) \lambda = \frac{n+1}{\bar{t}}, \quad (10)$$

$$\delta \lambda = \left[ \int_0^{\infty} d\lambda p(\lambda | t_1, \dots, t_n) (\lambda - \langle \lambda \rangle)^2 \right]^{1/2} = \frac{\sqrt{n+1}}{\bar{t}}, \quad (11)$$

and state the result summarily as  $\lambda = \langle \lambda \rangle \pm \delta \lambda$ . The choice of  $\langle \lambda \rangle$  as the recommended value can be justified by the following least-square argument: The point estimate  $\lambda_0$  which minimises the expected squared error, so that

$$\int d\lambda p(\lambda | t_1, \dots, t_n) (\lambda - \lambda_0)^2 = \min, \quad (12)$$

is  $\lambda_0 = \langle \lambda \rangle$ , as one readily verifies by differentiation with respect to  $\lambda_0$  and equating to zero. With this point estimate the expected squared error is just the variance,  $\text{var } \lambda = (\delta \lambda)^2$ , which justifies also our uncertainty specification. The notation  $\lambda = \langle \lambda \rangle \pm \delta \lambda$ , however, obscures the fact that especially for small  $n$  the gamma distribution is quite asymmetric. If such details are important one must return to the full posterior distribution. Our result, Eqs. 10-11, looks reasonable enough, but we shall see that there is a problem caused by the rather cavalier fashion in which we assigned the prior probability.

### 1.3 Generalisation to More Observations and More Parameters

Before we deal more carefully with priors let us see what impact a second measurement (with a fresh radioactive sample) would have on our knowledge of the decay constant. Using the posterior distribution of the first measurement as the prior for the second one we find the new posterior distribution

$$p(\lambda | t_1, \dots, t_n, t'_1, \dots, t'_m) d\lambda \propto p(t'_1, \dots, t'_m | \lambda) p(t_1, \dots, t_n | \lambda) p(\lambda) d\lambda, \quad (13)$$

where  $t'_1, \dots, t'_m$  are the new data. More generally, if there are  $k$  measurements, with associated data sets  $D_1, \dots, D_k$  and likelihood functions  $L_1, \dots, L_k$ , one gets

$$p(\lambda | D_1, \dots, D_k) d\lambda \propto \left\{ \prod_{j=1}^k L_j(D_j | \lambda) \right\} p(\lambda) d\lambda, \quad (14)$$

which shows quite nicely how Bayes' theorem models the process of learning by experience: Each new experimental result can be incorporated into the existing body of knowledge simply by multiplication of the associated likelihood function into the probability distribution (and renormalisation). It is by no means necessary that all experiments are of the same type. In resonance analysis, for instance, one usually combines likelihood functions from transmission, capture, fission and scattering experiments involving all kinds of detectors and samples in order to obtain best values of resonance energies and partial widths. With each added data set the posterior distribution becomes narrower, so that the uncertainty of the estimated parameter becomes smaller. We see this explicitly in our example. The relative uncertainty of  $\lambda$  goes to zero as  $1/\sqrt{n}$  for large  $n$ .

A last generalisation concerns the estimated parameters. In data evaluation and adjustment one has not only to deal with large bodies of data from many different experiments but also with many correlated parameters that must be estimated simultaneously. Instead of one parameter  $\lambda$  one has then a parameter vector  $\lambda$  in the equations, instead of the increment  $d\lambda$  one has a volume element  $d^N\lambda$  in the parameter space, and the prior and posterior distributions represent joint probabilities for all  $N$  parameters complete with correlations. Again resonance analysis provides an example. With modern shape analysis codes one can estimate the resonance energies and widths of many resonances simultaneously by fitting data from many types of resonance measurements (see e. g. [13]).

#### 1.4 Closer Look at Prior Probabilities, Group-Theoretical Assignment

We must now deal more thoroughly with prior probabilities. In our example we used the prior  $p(\lambda)d\lambda \propto d\lambda$  which in terms of the mean life  $\tau \equiv 1/\lambda$  can be rewritten as  $p(1/\tau)d\tau/\tau^2 \equiv p(\tau)d\tau \propto d\tau/\tau^2$ . It seems that we could have equally well estimated  $\tau$  instead of  $\lambda$ , and assumed all  $d\tau$  equally probable, so that  $p(\tau)d\tau \propto d\tau$ . This, however, would have resulted in a different posterior distribution. From a principal viewpoint it is no consolation that the dependence on the prior is weak if data are abundant. There seems to be a fundamental arbitrariness about priors, especially for continuous parameters.

For more than a century this seeming arbitrariness has caused many statisticians to repudiate the Bayesian approach to parameter estimation and to try alternative methods that circumvent priors. Others, comparing these efforts to an attempt to do arithmetic without zero, defended Bayes' theorem as derivable in a few lines from the basic equations and used "subjective" priors or, as H. Jeffreys [2], invoked invariance arguments to find priors which avoided ambiguities. A major step forward was the realisation by A. Wald [3] that the optimal strategies for making decisions in the face of uncertainty, as derived without priors, were just the Bayesian rules. Even more important was the application of group theory and information theory to the problem of priors by E.T. Jaynes [4, 5] in 1968. He demonstrated that in a number of simple but practically important cases the symmetry of the problem determines the prior unambiguously. If a so-called location parameter is estimated, for instance the mean  $\mu$  of a Gaussian, the form of the prior must be invariant under a shift in location,  $\mu \rightarrow \mu + c$ . This implies

$$p(\mu)d\mu \propto d\mu, \quad -\infty < \mu < \infty, \quad (15)$$

a result that was never controversial. For a scale parameter such as the standard deviation  $\sigma$  of a Gaussian, the form of the prior must be invariant under rescaling,  $\sigma \rightarrow c\sigma$ , and this leads to

$$p(\sigma)d\sigma \propto d\sigma/\sigma, \quad 0 < \sigma < \infty, \quad (16)$$

as already suggested in 1939 by H. Jeffreys [2]. Despite its importance and simplicity Jaynes' proof seems so little known that we quote it here almost verbatim for the case of a rate constant which scales all times in a problem (as  $\lambda$  in Eq. 6 does).

Suppose that two observers, Mr. X and Mr. X', wish to estimate a rate constant from a number of events. If their watches run at different rates so that their measurements of a given time interval are related by  $t = ct'$ , their rate or scale parameters will be related by  $\lambda' = c\lambda$ . They assign prior probabilities  $p(\lambda)d\lambda$  and  $q(\lambda')d\lambda'$ , and if these are to represent the same state of ignorance,  $p$  and  $q$  must be the same function so that  $p(\lambda)d\lambda = p(\lambda')d\lambda'$ . From the two equations for  $\lambda$  and  $\lambda'$  one gets the functional equation  $p(\lambda)d\lambda = p(c\lambda)cd\lambda$ . Its unique solution is Jeffreys' prior,

$$p(\lambda)d\lambda \propto d\lambda/\lambda, \quad 0 < \lambda < \infty. \quad (17)$$

Obviously this is the correct prior for our example, since our decay constant is just such a rate or scale parameter, multiplying (scaling) all times and time increments in our equations. With this prior we get

$$p(\lambda | t_1, \dots, t_n)d\lambda = \Gamma(n)^{-1} e^{-x} x^{n-1} dx, \quad 0 < x \equiv \lambda \bar{t} < \infty, \quad (18)$$

$$\langle \lambda \rangle = \frac{1}{\bar{t}}, \quad (19)$$

$$\delta\lambda = \frac{1}{\bar{t}\sqrt{n}}. \quad (20)$$

This looks neater than the result we had before, illustrating Ockham's wisdom that the simpler result is usually the more correct one. Further examples of priors derived from group-theoretical invariances can be found in Refs. 4 - 7.

### 1.5 Bayesian Parameter Estimation for the Univariate Gaussian

Let us apply Jaynes' results also to the the principally and practically important univariate Gaussian distribution. Suppose a repeated measurement of the same physical quantity  $\mu$  has produced the

results  $x_1, \dots, x_n$ , with experimental errors that can be assumed to be normally distributed. Then the sampling distribution is

$$p(x|\mu, \sigma) dx = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx, \quad -\infty < x < \infty \quad (21)$$

where the error dispersion  $\sigma$  is unknown, and the prior describing complete ignorance of location and scale (width) of the Gaussian is

$$p(\mu, \sigma) d\mu d\sigma \propto d\mu d\sigma / \sigma, \quad -\infty < \mu < \infty, \quad 0 < \sigma < \infty. \quad (22)$$

The posterior is thus

$$p(\mu, \sigma | x_1, \dots, x_n) d\mu d\sigma \propto \frac{1}{\sigma^n} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right] \frac{d\mu d\sigma}{\sigma}. \quad (23)$$

In terms of sample mean and sample variance,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad s'^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2, \quad (24)(25)$$

we can write

$$\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 = \frac{ns'^2}{2\sigma^2} + \frac{ns'^2}{2\sigma^2} \left(\frac{\bar{x} - \mu}{s'}\right)^2 \equiv \eta + \eta\xi \quad (26)$$

so that, properly normalised, the posterior joint probability for  $\mu$  and  $\sigma$  becomes

$$p(\mu, \sigma | x_1, \dots, x_n) d\mu d\sigma = \frac{1}{\sqrt{\pi}} e^{-\eta\xi^2} \frac{1}{\Gamma\left(\frac{n-1}{2}\right)} e^{-\eta} \eta^{(n-1)/2-1} d\eta d\xi, \\ -\infty < \xi \equiv \frac{\mu - \bar{x}}{s'} < \infty, \quad 0 < \eta \equiv \frac{ns'^2}{2\sigma^2} < \infty. \quad (27)$$

The probability distribution of  $\xi$  given  $\eta$  is Gaussian, and that of  $\eta$  given  $\xi$  is a gamma distribution, but in general the two estimated parameters are correlated.

If only  $\mu$  is of interest, regardless of  $\sigma$ , one obtains the marginal distribution by integration over all  $\sigma$  as

$$p(\mu | x_1, \dots, x_n) d\mu = \frac{\Gamma(\frac{n}{2})}{\sqrt{\pi} \Gamma(\frac{n-1}{2})} \frac{d\xi}{(1+\xi^2)^{n/2}}, \quad -\infty < \xi \equiv \frac{\mu - \bar{x}}{s'} < \infty. \quad (28)$$

This is known as Student's t-distribution with  $n-1$  degrees of freedom. Obviously its mean is  $\langle \xi \rangle = 0$  which implies

$$\langle \mu \rangle = \bar{x}.$$

This is the familiar, plausible rule to use the sample mean as point estimate of the population mean. The variance is  $\langle \xi^2 \rangle = 1/(n-3)$  so that no finite real  $\delta\mu$  can be calculated from it for  $n < 4$ . On the other hand the half width is always well defined and can be used to indicate the uncertainty of  $\mu$  as is familiar from the case  $n = 2$ , the Cauchy distribution (known to physicists also as Lorentzian or as symmetric Breit-Wigner profile). If only  $\sigma$  is of interest, one finds

$$p(\sigma | x_1, \dots, x_n) d\sigma = \frac{1}{\Gamma(\frac{n-1}{2})} e^{-\eta} \eta^{(n-3)/2} d\eta, \quad 0 < \eta \equiv \frac{ns'^2}{2\sigma^2} < \infty, \quad (29)$$

a chi-square distribution with  $n-1$  degrees of freedom for  $\eta$ . Its mean and variance are  $\langle \eta \rangle = \text{var } \eta = (n-1)/2$ . The recommended value for  $1/\sigma^2$  is therefore

$$\langle \sigma^{-2} \rangle = \frac{n-1}{n} s'^{-2} \equiv s^{-2} = \frac{n-1}{\sum_i (x_i - \bar{x})^2}. \quad (30)$$

This is the precise formulation of the familiar, less plausible rule to use  $s^2$  instead of the the sample variance  $s'^2$  as a (biased) estimate for  $\sigma^2$ . We recognise that  $s^{-2}$  is actually an unbiased estimate of  $\sigma^{-2}$ , the so-called precision.

The case with one datum only,  $n = 1$ , must be treated separately because  $s' = 0$ , but this is easy. One finds



$$p(\mu, \sigma | x_1) d\mu d\sigma \propto \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x_1 - \mu}{\sigma}\right)^2\right] \frac{d\mu d\sigma}{\sigma}, \quad (32)$$

$$p(\mu | x_1) d\mu \propto \frac{d\mu}{|\mu - x_1|}, \quad (33)$$

$$p(\sigma | x_1) d\sigma \propto \frac{d\sigma}{\sigma}. \quad (34)$$

The marginal distribution of  $\mu$  has a sharp maximum at the observed value but that of  $\sigma$  is seen to be still equal to the uninformative prior. This makes sense because from a sample of 1 one can learn something about the location, but nothing whatsoever about the spread of a distribution. Obviously the Bayesian method is consistent with common sense even in this extreme case. We mention that the posterior (27) for  $n \geq 2$  was found long before the Jeffreys-Jaynes prior became available, but those who know e.g. R.A. Fisher's fiducial approach [10] will appreciate how much simpler the present (Jeffreys', [2]) derivation is and how easily it is extended to the case  $n = 1$ . Quite generally the Bayesian approach is simpler than alternative methods. Concepts like bias, efficiency, sufficiency, admissibility, James-Stein shrinking etc., on which other methods are based, need not be introduced at all since they appear only as more or less incidental aspects of the posterior distribution and its mean and variance.

The non-normalisability of the uninformative priors is sometimes criticised. Now one can employ also a broad normalisable prior of convenient ("conjugate") mathematical form. In our first example this would be a gamma function. The posterior will then, of course, depend on the width of this prior. If one lets the width grow indefinitely one finds invariably that the posterior tends towards the posterior derived from the uninformative priors. These can therefore be considered as limiting cases of extremely broad, normalised distributions on the linear ( $d\mu$ ) and logarithmic ( $d\sigma/\sigma = d \ln \sigma$ ) scale, just as Dirac's delta function is the limiting case of an extremely narrow, normalised distribution. There are no conceptual or mathematical difficulties if one keeps in mind that both the completely uninformative priors and the "completely

informative" delta function are, in this sense, nothing but convenient shorthand notations, meaningful only in convolution with other distributions.

### 1.6 Assignment of Priors by Entropy Maximisation

Jaynes [4-7] also considered the case that one is not completely ignorant a priori. He showed how prior probabilities can be assigned in a well defined way if at least vague information is available about average quantities, e. g. order-of-magnitude estimates of mean values such as first and second moments. The key concept is that of information entropy, introduced by C.E. Shannon in 1948 [11] as the unique measure of the indeterminacy or missing information implied by a given probability distribution. The information entropy of a continuous distribution  $p(x)dx$ , with a-priori equivalent increments  $dx$ , is

$$S = - \int dx p(x) \ln p(x) . \quad (35)$$

Let us now assume that information about  $p(x)$  is given in the form of expectation values for known functions  $f_k(x)$ ,

$$\langle f_k \rangle = \int dx p(x) f_k(x) , \quad k=1,2,\dots,K. \quad (36)$$

What is the probability density  $p(x)$  that satisfies these  $K$  equations but does not imply any other information or hidden assumptions? The last requirement in fact means maximal indeterminacy apart from the conditions (36), i.e. we must solve the variational problem  $S = \max$  with the  $K$  constraints (36) (and the additional constraints that  $p(x)$  is nonnegative and normalised to unity). The well-known solution, obtained by the method of Lagrange multipliers, is

$$p(x) = \frac{1}{Z} \exp \left[ - \sum_{k=1}^K \lambda_k f_k(x) \right] . \quad (37)$$

This is manifestly positive for real  $\lambda_k$ , and properly normalised to unity if

$$Z = \int dx \exp \left[ - \sum_{k=1}^K \lambda_k f_k(x) \right] . \quad (38)$$

The Lagrange multipliers can be found from the K equations

$$\langle f_k \rangle = - \frac{\partial}{\partial \lambda_k} \ln Z , \quad k=1,2,\dots,K \quad (39)$$

which are equivalent to the constraints (36). If, for instance, we know only the first and second moment of the distribution  $p(x)dx$ , with  $-\infty < x < \infty$ , we readily obtain a Gaussian with these moments. If  $x$  is inherently positive,  $0 < x < \infty$ , we can substitute  $y = \ln x$ . With known first and second moments on the log scale we get a Gaussian on the log scale, i.e. a lognormal distribution on the linear ( $x$ -)scale. This is one of the reasons for the ubiquity of these distributions in statistics and data analysis. Moreover, we have now a recipe to set up a complete (though approximate) probability distribution from information given in the form  $x = \langle x \rangle \pm \delta x$ .

The maximum-entropy algorithm (37)-(39) ought to look quite familiar to physicists, since it represents nothing less than Gibbs' axiomatic approach to thermodynamics. Clausius' thermodynamic entropy is the maximised information entropy multiplied by Boltzmann's constant, and  $Z$  is the partition function from which all macroscopically observable ensemble averages can be found by suitable differentiation. For instance, if  $x$  is interpreted as the (positive) energy of a thermodynamic system, about which nothing is known except its mean energy, one gets the canonical distribution,  $p(x) \propto \exp(-\lambda x)$ , i. e. a Boltzmann factor with the inverse temperature appearing as Lagrange parameter. If we know in addition the average particle number we obtain the grand-canonical distribution, with the chemical potential as a second Lagrange parameter, etc. A highly informative and readable review of the maximum-entropy formalism, including a wide variety of applications from hypothesis testing to non-equilibrium thermodynamics, can be found in Ref. 8.

### 1.7 Approximations: Maximum Likelihood and Least Squares

The more abundant the data are, the less important is the prior and vice versa. In many cases it is therefore a reasonable approximation to use a constant prior, as we did initially in our first (decay constant) example. This means that the posterior probability density becomes equal to the likelihood function. The maximum likelihood method consists essentially in the rule to recommend that parameter value or parameter vector which maximises the likelihood function. In many simple problems this point estimate turns out to be a function of one or a few "statistics" (functions of the sample values). Those are then said to be "sufficient", and their sampling distribution provides an indication of the uncertainty of the point estimate. If sufficient statistics exist and their probability distribution can be found, it turns out that the maximum-likelihood result coincides with the Bayesian result. We can illustrate this with our decay constant example. The likelihood function in (7) becomes maximal for  $\lambda = 1/\bar{t}$ , where  $\bar{t}$  is a sufficient statistic. One finds the distribution of  $\bar{t}$  by integrating the likelihood function over all  $t_i$  with  $\bar{t}$  kept constant, i. e. over a spherical surface in the space of the  $dt_i$ . Using polar coordinates so that  $\bar{t} = r^2$ ,  $d\bar{t} = 2rdr$ ,  $d^n t \propto r^{2n-1} dr d\Omega$  one gets, upon integration over  $\Omega$ ,

$$p(\bar{t}|\lambda)d\bar{t} \propto e^{-\lambda n r^2} r^{2n-1} dr, \quad 0 < r \equiv \sqrt{\bar{t}} < \infty. \quad (40)$$

After normalisation the right-hand side is the same as the Bayesian posterior (18) obtained with Jeffreys' prior. Since this is actually the probability distribution of  $x = \lambda n \bar{t}$  it can either be interpreted as the probability of  $\lambda$  given  $n$  and  $\bar{t}$  (one sample, various possible decay constants), or, equally well, as that of  $\bar{t}$  given  $n$  and  $\lambda$  (one decay constant, various possible samples of size  $n$ ).

The maximum-likelihood method, one of the techniques invented to circumvent priors, is thus in favourable cases as rigorous as the Bayesian approach, but even then it is more cumbersome, requiring identification of sufficient statistics and calculation of their probability distribution. The basic simplicity and superiority of the Bayesian approach as compared to other estimation techniques is quite forcefully demonstrated with a whole series of examples in Ref. 6.

The next approximation to be discussed, the least-squares method, is the most important one practically in data evaluation and adjustment. Let us consider

- observables  $y_i$ ,  $i = 1, 2, \dots, I$  (e.g. neutron transmission data)
- parameters  $x_\mu$ ,  $\mu = 1, 2, \dots, M$  (e.g. resonance parameters)
- a model  $y = y(x)$  (e.g. R-matrix theory)

where  $y = (y_1, \dots, y_I)$ ,  $x = (x_1, \dots, x_M)$  are vectors in the data and parameter spaces, respectively, and  $I > M$ . Suppose

- (a) that before the data became available one had prior knowledge about the parameters  $x_\mu$ , namely estimates  $\xi_\mu$  and correlated uncertainties  $A_{\mu\nu} = \langle \delta\xi_\mu \delta\xi_\nu \rangle$ , or at least variances  $A_{\mu\mu}$ , so that the probability of  $x$  given  $\xi$  can be taken as

$$p(x|\xi) \propto \exp\left[-\frac{1}{2} (x-\xi)^+ A^{-1} (x-\xi)\right], \quad (41)$$

where  $^+$  denotes the transpose;

- (b) that new measurements yielded values  $\eta_i$  and correlated uncertainties  $B_{ik} = \langle \delta\eta_i \delta\eta_k \rangle$  for the observables  $y_i$ , so that the likelihood to obtain these values, provided the true parameter vector is  $x$ , can be taken as

$$p(\eta|y) \propto \exp\left[-\frac{1}{2} (y-\eta)^+ B^{-1} (y-\eta)\right]. \quad (42)$$

The assumption of multivariate Gaussians is the only approximation invoked. For relatively small distances  $|x-\xi|$  and  $|\eta-y(x)|$  it is expected to be reasonable and adequate for parameter estimation purposes. Bayes' theorem yields now

$$\begin{aligned} p(x|\xi\eta) &\propto p(\eta|x)p(x|\xi) \\ &\propto \exp\left[-\frac{1}{2} (x-\xi)^+ A^{-1} (x-\xi) - \frac{1}{2} (\eta-y)^+ B^{-1} (\eta-y)\right]. \end{aligned} \quad (43)$$

The most probable vector  $x$  is the one that minimises the exponent,

$$\chi^2 \equiv (x-\xi)^+ A^{-1} (x-\xi) + (\eta-y(x))^+ B^{-1} (\eta-y(x)) = \min .$$

This is the formal statement of the principle of least squares in its most general form, involving not only data errors but also uncertain prior knowledge about the estimated parameters and possible nonlinearity of the model  $y(x)$ . We shall come back to this equation in the following more technical sections.

## 2. EVALUATION OF NUCLEAR DATA FOR APPLICATIONS

In the following sections we shall discuss some of the more practical aspects of nuclear data evaluation. The progress from experimental raw data to evaluated files of cross sections and cross section covariances will be outlined, with more detailed reference to the general least-squares formalism, to statistical versus systematic errors and how the latter cause correlations among data points and data sets.

### 2.1 Steps in the Preparation of Nuclear Data for Applications

Nuclear (and other scientific) data for technological applications are usually prepared in several steps. We illustrate these steps with neutron cross section measurements.

- (1) Measurement: Experimenters take data, typically at steady-state or pulsed accelerators, the latter permitting use of the time-of-flight technique which covers wide energy ranges with high resolution. The simplest measurement is that of the total cross section  $\sigma_t$ . One measures that fraction of a beam of particles of given energy (or flight time between accelerator pulse and detector response) which traverses without interaction a sample of given thickness  $n$  (atoms/b). This fraction, the transmission, is  $1 - \sigma_t \Delta n$  for a very thin layer of material. For a sample of finite thickness it is

$$T = \lim_{\Delta n \rightarrow 0} (1 - \sigma_t \Delta n)^{n/\Delta n} = \exp(-n\sigma_t) . \quad (45)$$

In practice T is obtained as the ratio of two count rates from a "sample-in" and a "sample-out" run. The incoming flux and the detector efficiency cancel out, so there is no calibration uncertainty. Background noise, however, requires corrections. Usually one needs "thin" and "thick" sample data in order to obtain optimal accuracy of the extracted cross sections or resonance parameters. Fig. 1 shows four sets of transmission data obtained at ORNL across two resonances of  $^{238}\text{U}$  [12].

Partial cross section are more difficult to measure. Experimentally one obtains a reaction yield, for example of fission products or capture gamma rays which are emitted from a thin sample upon bombardment with the beam particles and subsequently recorded by a detector. The yield is defined as the fraction of beam particles undergoing a reaction of the type measured in the sample. It is a sum of contributions from multiple-collision events with zero, one, two etc. preceding scattering collisions,

$$Y = Y_0 + Y_1 + \dots , \quad (46)$$

$$Y_0 = (1 - T)\sigma_x/\sigma_t , \quad (47)$$

where  $\sigma_x$  is the partial cross section for the (n,x) reaction considered. The multiple-collision terms must usually be obtained by Monte Carlo simulation [13]. The count rate observed is the product of Y, the flux and the detector efficiency. These latter two quantities must be measured separately and introduce correlated normalisation errors. Fig. 2 shows recent neutron capture data across the 1.15 keV resonance of  $^{56}\text{Fe}$  which is responsible for most of the iron contribution to the Doppler coefficient in fast reactors [14].

- (2) Reduction of raw data: Constant and time-dependent backgrounds are subtracted, sample impurities are corrected for, and, in the case

of partial cross section data (yields), flux and detector efficiency are factorised out. Corrections for multiple scattering and instrumental resolution are usually deferred to the next step if they require e.g. resonance theory, and due to the complicated relationship between observables and cross section (see Eqs. 45-47) this means that at this stage cross sections are still unavailable, except for thin-sample measurements.

(3) Analysis of clean data: Whenever possible, nuclear reaction theory is used to parametrise the cross sections. In order of increasing energy the following theories and models are used:

- R-matrix theory for the thermal and resolved resonance region,
- level-statistical (Hauser-Feshbach) theory in the unresolved resonance region,
- the optical model at higher energies where levels overlap strongly but compound reactions still dominate,
- precompound, direct and multistep theories at still higher energies where direct and pre-equilibrium processes are important.

This is complemented by the giant-dipole resonance model for photon reactions (and beta decay), fission barrier models for fission reactions, etc. At this stage parameter estimation techniques are used extensively, for instance to extract resonance parameters or the parameters of an optical potential. From the estimated parameters and their uncertainties one calculates the cross sections and their uncertainties which the user finds in files of evaluated nuclear ("microscopic") data. Where suitable models are unavailable polynomial or similar fits are used to interpolate between evaluated data points. Unmeasured data are generated with the help of nuclear models or from systematics. Fig. 3 shows neutron scattering data for  $^{93}\text{Nb}$  and curves fitted to them by adjusting the parameters of an optical model [15].

(4) Generation of group constants: Doppler-broadened point cross sections for all open reaction channels and for various temperatures can now



be calculated and averaged suitably for reactor or shielding applications over relatively large energy intervals. The result is a set of group constants, usually including so-called self-shielding factors for specified temperatures and "dilutions" (i.e. admixtures of other nuclides), as needed e.g. for transport calculations in nuclear technology.

## 2.2 International Cooperation

All these steps require time and many years may pass before nuclear data needed for technological or scientific applications become available in the form of machine-readable computer files. This motivated efforts to coordinate the work on an international scale, with OECD (Organisation for Economic Cooperation and Development of the Western industrialised countries) and IAEA (International Atomic Energy Agency of the United Nations) playing the leading parts. Two cooperating nuclear data committees, NEANDC (advising the Nuclear Energy Agency of OECD) and INDC (advising the Nuclear Data Section of IAEA), collect and screen formal requests for nuclear data which are periodically published by IAEA in WRENDATA, the Worldwide Requests for Nuclear Data. Measured Data are collected by a network of data centres, each operating within its agreed service area:

- NNDC (National Nuclear Data Center)  
at Brookhaven, USA, servicing the US and Canada;
- NEADB (NEA Data Bank, OECD)  
at Saclay, France, servicing the non-American OECD countries;
- CJD (Centr po Jadernym Dannym)  
at Obninsk, USSR, servicing the Soviet Union;
- NDS (Nuclear Data Section, IAEA)  
at Vienna, Austria, servicing all other countries.

Regular data exchange in a special format (EXFOR) ensures that the data base is essentially the same at all four centres. Evaluated data are also collected, notably the files ENDF (US), JEF (NEA member countries), JENDL (Japan), SOKRATOR (USSR), KEDAK (Germany), the first two only available to restricted user communities. The four centres produce

periodically also the widely used Computer Index to Neutron Data (CINDA). The well known "barn book", BNL 325, containing resonance parameter tables and cross section plots, is a product of NNDC. Computer programs for cross section analysis and fitting or for group constant generation are also collected and distributed by the neutron data centres. Similar networks of data centres compile and distribute charged-particle data and nuclear structure and decay data. The ENSDF file contains evaluated data of the latter type, it is the machine-readable offspring of the well known Table of Isotopes and the Nuclear Data Sheets.

Comparable international cooperation exists in meteorology, aerospace, high-energy physics and other scientific and technological areas.

### 2.3 Iterative Least-Squares Fitting.

Most of the parameter estimation work in step (3), analysis of clean data, employs the least-squares method. We return therefore to the general least-squares equation,

$$\chi^2 \equiv (x-\xi)^+ A^{-1} (x-\xi) + (\eta-y(x))^+ B^{-1} (\eta-y(x)) = \min . \quad (44)$$

We stress again that the data vector  $\eta$  may contain data from quite different types of measurements which, of course, must be mathematically described by the corresponding components of the modelling vector  $y(x)$ . We shall consider the solution vector  $x$  as the improved estimate and denote it by  $\xi'$ . Note that without prior knowledge  $A^{-1}$  vanishes and with it the whole first term. Neglecting also the off-diagonal elements of the matrix  $B^{-1}$  one gets the starting condition for "primitive" least-squares fitting which is used in many computer codes. It utilises only the data and their uncertainties, but the resulting parameters and their uncertainties must be combined with previous parameter estimates, derived from other data, by some kind of weighted averaging after the fit. It is much more practical to utilise existing values and their uncertainties right from the start in a prior (cf. Eq. 41). In most cases the off-diagonal elements of the matrix  $A$  are unknown. They are then set equal to zero, i. e.  $A^{-1}$  is diagonal, with the inverse squared uncertainties of the existing or estimated parameters as elements. This is easy to

do and not only ensures that existing knowledge is automatically taken into account but also improves convergence, as we shall see.

The extremum condition (44) is equivalent to

$$A^{-1}(x-\xi) - \dot{y}(x)^+ B^{-1}(\eta-y(x)) = 0, \quad (48)$$

where  $\dot{y}$  is the rectangular matrix of sensitivity coefficients,

$$\dot{y}_{i\mu} \equiv \frac{\partial y_i}{\partial x_\mu}. \quad (49)$$

Eq. 48 is easily solved for  $x$  if  $y$  is a linear function of  $x$ . In nuclear data work, however,  $y(x)$  is nonlinear and one must iterate, e.g. with the Newton-Raphson method (in  $M$  dimensions). Fig. 4 shows a typical example [16] of the nonlinearities and correlations induced by measurements of various types for two estimated parameters (partial widths of a  $^{232}\text{Th}+n$  resonance). Starting the iterative process with the a priori most probable value,  $x_0 = \xi$ , one finds after  $n$  steps

$$x_{n+1} = \xi + \left[ A^{-1} + \dot{y}(x_n)^+ B^{-1} \dot{y}(x_n) \right]^{-1} \dot{y}(x_n)^+ B^{-1} \left[ \eta - y(x_n) - \dot{y}(x_n)(\xi - x_n) \right] \quad (50)$$

and finally, after convergence, the new estimate

$$\xi' = \xi + \left[ A^{-1} + \dot{y}(x_\infty)^+ B^{-1} \dot{y}(x_\infty) \right]^{-1} \dot{y}(x_\infty)^+ B^{-1} \left[ \eta - y(x_\infty) - \dot{y}(x_\infty)(\xi - x_\infty) \right]. \quad (51)$$

The new correlated errors are obtained as follows: We consider a small domain around  $x = \xi'$  where  $y(x)$  can be considered as linear. The right-hand side of (43) reduces then to a product of two multi-variate Gaussians which is equivalent to another multivariate Gaussian with the most probable value (mean)  $\langle x \rangle = \xi'$  and correlated errors given by

$$A'^{-1} = A^{-1} + \dot{y}(x_\infty)^+ B^{-1} \dot{y}(x_\infty). \quad (52)$$

In practice, of course, one does not need infinitely many steps as our notation seems to imply. Usually three or four steps are quite enough for practical convergence (i.e. stationarity within single precision of the computer).

Sometimes it is better to express everything in terms of the covariance matrices A and B instead of their inverses. For instance, a common (systematic) background uncertainty in the data,  $\delta\eta_i = b$ , results in  $B_{ik} = \langle \delta\eta_i \delta\eta_k \rangle = b^2$ . The matrix B is then singular and  $B^{-1}$  is undefined. It can be shown that Eqs. 51 and 52 are equivalent to

$$\xi' = \xi + A\dot{y}(x_\infty) \left[ B + \dot{y}(x_\infty) A \dot{y}(x_\infty) \right]^{-1} \left[ \eta - y(x_\infty) - \dot{y}(x_\infty) (\xi - x_\infty) \right], \quad (53)$$

$$A' = A - A\dot{y}(x_\infty) \left[ B + \dot{y}(x_\infty) A \dot{y}(x_\infty) \right]^{-1} \dot{y}(x_\infty) A. \quad (54)$$

The pairs of Eqs. 51, 52 and 53, 54 show explicitly how the prior estimates and uncertainties  $\xi$ , A are updated by the new data  $\eta$ , B resulting in posterior estimates  $\xi'$ , A'. The minus sign in Eq. 54 indicates that new data in fact reduce the uncertainties. The changes are small if the sensitivity coefficients  $\dot{y}_{i\mu}$  are small and vice versa.

This iterative least-squares method [17] is employed in the resonance analysis code SAMMY [18] and in the Hauser-Feshbach code FITACS [19]. Experience with these codes has clearly shown the advantage of formalised inclusion of a-priori information. Since the parameter search is constrained smoothly to a reasonable domain, the (linear programming) problems encountered with sharp limits are avoided, and convergence is dramatically improved compared with earlier "primitive" least-squares versions of these codes which did not utilise prior uncertainties.

#### 2.4 Statistical Errors: the Poisson Distribution

We must now discuss the error information which is needed for the construction of the covariance matrix B describing the data uncertainties and their correlations. In practically all nuclear data measurements particles of a certain type are detected and counted, for instance fission fragments signalling nuclear fission, or gamma quanta signalling radiative capture. The counts are a measure for the corresponding fission or capture probabilities (conventionally expressed as fission or capture cross sections). In the limit of infinite counting time, and in the

absence of other errors, one would measure the probabilities (in the frequentist sense) directly, but in practice there is always some statistical uncertainty as to the limiting count rate (or cross section) due to the finite number of counts obtained within a finite time span. What can we say about the true rate  $\lambda$  if  $n$  events were registered during a time  $t$ ? Counting statistics is governed by the Poisson distribution,

$$p(n|\lambda t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \quad n=0,1,\dots \quad (55)$$

in which  $\lambda$  is seen to be a scale parameter. Bayes' theorem with Jeffreys' prior yields immediately the inverse probability

$$p(\lambda|t,n) = \Gamma(n)^{-1} e^{-x} x^{n-1} dx, \quad 0 < x \equiv \lambda t < \infty, \quad (56)$$

whence

$$\langle \lambda \rangle = \frac{n}{t}, \quad \delta \lambda = \frac{\sqrt{n}}{t} \quad (57)(58)$$

The relative uncertainty is

$$\frac{\delta \lambda}{\langle \lambda \rangle} = \frac{1}{\sqrt{n}} \quad (59)$$

which, of course, is the familiar rule for the assessment of statistical errors, widely used not only in counting statistics but also in Monte Carlo calculations.

## 2.5 Systematic Errors: Correlated Uncertainties and their Propagation

We shall now briefly discuss a few basic types of systematic errors and how they cause correlations between data (or parameters). Above we denoted the unknown errors of the  $\eta_i$  by  $\delta\eta_i$ . If they were purely statistical they would be uncorrelated and one would have

$$\langle \delta\eta_i \delta\eta_j \rangle = \text{var}(\eta_i) \delta_{ij} \equiv \sigma_i^2 \delta_{ij}. \quad (60)$$

This is assumed in many primitive least-squares codes where B is a diagonal matrix with elements  $B_{ii} = \sigma_i^2$ , so that the i-th data point is weighted by  $1/\sigma_i^2$ . Besides the statistical errors from counting statistics there are, however, always errors from flux determination, detector calibration, timing uncertainty etc. In contrast to the statistical errors these so-called systematic errors are common to a whole set of data, for instance to the data obtained from all the time channels in a time-of-flight measurement. Now common errors are, quite generally, the cause of correlations between data. To see this we express the unknown total errors in the form

$$\delta\eta_i = \delta\eta_i' + \delta\eta' , \quad (61)$$

where  $\delta\eta_i'$  is the statistical error, and the systematic error  $\delta\eta'$  has no subscript since it is the same for all data points. The elements of the covariance matrix B are now readily obtained as

$$\langle \delta\eta_i \delta\eta_j \rangle = \langle (\delta\eta_i')^2 \rangle \delta_{ij} + \langle (\delta\eta')^2 \rangle \quad (62)$$

if one uses the fact that statistical errors of different data points are uncorrelated,  $\langle \delta\eta_i' \delta\eta_j' \rangle = \langle (\delta\eta_i')^2 \rangle \delta_{ij}$ , with zero mean,  $\langle \delta\eta_i' \rangle = 0$ , and that there is no correlation between statistical and systematic errors,  $\langle \delta\eta_i' \delta\eta' \rangle = 0$ . Eq. 62 shows that common, i.e. systematic, errors automatically produce correlations between the elements of a data set. In case there is a common background error  $\delta b$  we have  $\delta\eta' = \delta b$ . If there is a calibration error  $\delta c$  the resulting relative error is  $\delta\eta'/\eta_i = \delta c/c$ . If both error types are present we have  $\delta\eta' = \delta b + \eta_i \delta c/c$  and thus

$$\langle \delta\eta_i \delta\eta_j \rangle = \langle (\delta\eta_i')^2 \rangle \delta_{ij} + \langle (\delta b)^2 \rangle + \eta_i \eta_j \langle (\delta c/c)^2 \rangle . \quad (63)$$

This should suffice to illustrate how one can generate the elements of the matrix B if enough information about the various error sources is available. It is important to realise that only the numerical values of expected (absolute or relative) errors are needed for this purpose, but not their correlations. Those are automatically obtained once the split of the total error into components has been made. The addition of squared statistical and systematic errors is sometimes criticised as incorrect.

Our equations show, however, that it is perfectly alright. The only problem with the sum of squared errors is that it does not reveal how much of it is systematic and how much statistical, i.e. how much correlation there is.

This demonstrates how important it is that experimentalists state clearly and in as much detail as possible the statistical and systematic error components when they report their data, whereas they need not worry about correlations or covariance matrices since these can be constructed easily from the error components. For an instructive example of correlated data errors and their impact on estimated resonance parameters see the discussion of resonance energy standards by F.G. Perey [20].

The statement that experimentalists need not worry about correlations between their data should, by no means, be misunderstood as meaning that correlations are unimportant. The remark referred only to the data uncertainties required for construction of the covariance matrix B. The correlated uncertainties of cross sections or of cross section parameters in the posterior covariance matrix A' are highly relevant information for users of the data. The uncertainty of any function y of the cross section parameters  $x_\mu$ , for example the calculated criticality of a nuclear reactor, is given in linear approximation by the square root of

$$\langle (\delta y)^2 \rangle = \sum_{\mu} \sum_{\nu} \frac{\partial y}{\partial x_{\mu}} \langle \delta x_{\mu} \delta x_{\nu} \rangle \frac{\partial y}{\partial x_{\nu}} \quad (64)$$

where  $\langle \delta x_{\mu} \delta x_{\nu} \rangle$  is the element of the covariance matrix A' obtained in the least-squares fit. It is obvious that a good sensitivity study is not possible without the covariance matrix or at least its more important elements. In the past it often seemed that nuclear data are not accurate enough for certain applications when covariance information was ignored, whereas their accuracies were quite acceptable, due to many negative sum terms in the last equation, when the correlations were properly taken into account. Hence those who extract cross section parameters from experimental data should not just state the parameters and their uncertainties, but also at least the more important elements of the covariance matrix.

## 2.6 Inconsistent Data

One of the thorniest problems in data evaluation is that of inconsistent data. Suppose we are given the results of  $n$  completely independent and experimentally different measurements of the same physical quantity,  $\mu$ , in the form  $x_i \pm \sigma_i$ ,  $i = 1, 2, \dots, n$ . If the separation of any two values,  $|x_i - x_j|$ , is smaller or at least not much larger than the sum of the corresponding uncertainties,  $\sigma_i + \sigma_j$ , the data are said to be consistent or to agree "within error bars". (The probability that two equally precise measurements yield a separation greater than  $\sigma_i + \sigma_j = 2\sigma_i$  is only  $\text{erfc } 1 \approx 15.7\%$ , provided the sampling distribution is Gaussian with standard deviation  $\sigma_i$ ). If some or all separations are much larger, the data are not consistent with the assumptions implied by the stated errors. Inconsistencies are caused by unrecognised or malcorrected experimental effects such as backgrounds, dead time of the electronics, instrumental resolution, sample impurities, calibration errors, etc. The data are then not properly corrected for these effects, and the given uncertainties tend to be too small on their account.

What can we say about the unrecognised errors? If we have no other information but the data, and know nothing about the experiments that yielded them, positive and negative errors are equally probable, hence the probability distribution for the unrecognised error  $\epsilon_i$  of the  $i$ -th experiment should be symmetric about zero, and the same distribution should apply to all experiments. Let us therefore assume identical Gaussians for all  $\epsilon_i$ ,

$$p(\epsilon_i | \tau) d\epsilon_i = \frac{1}{\sqrt{2\pi\tau^2}} \exp\left[-\frac{1}{2} \left(\frac{\epsilon_i}{\tau}\right)^2\right] d\epsilon_i, \quad -\infty < \epsilon_i < \infty. \quad (65)$$

The probability to measure the value  $x_i$ , given the true value  $\mu$ , the unrecognised error  $\epsilon_i$  and the uncertainty  $\sigma_i$  due to all recognised error sources, is now given by

$$p(x_i | \mu, \epsilon_i, \sigma_i) dx_i = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{1}{2} \left(\frac{x_i - \mu - \epsilon_i}{\sigma_i}\right)^2\right] dx_i, \quad -\infty < x_i < \infty. \quad (66)$$



The joint posterior distribution for  $\mu$  and the  $\epsilon_i$  is

$$p(\mu, \epsilon_1, \dots | x_1, \dots, \sigma_1, \dots, \tau) d\mu d^n \epsilon \propto d\mu \prod_{i=1}^n d\epsilon_i \exp \left[ - \frac{(x_i - \mu - \epsilon_i)^2}{2\sigma_i^2} - \frac{\epsilon_i^2}{2\tau^2} \right] \quad (67)$$

if the dispersion  $\tau$  of unrecognised errors is known. Completing squares in the exponent we can easily integrate over the  $\epsilon_i$ . The resulting posterior distribution for  $\mu$  can be written in the form

$$p(\mu | x_1, \dots, \sigma_1, \dots, \tau) d\mu = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ - \frac{1}{2} \left( \frac{\mu - \bar{x}}{\sigma} \right)^2 \right] d\mu, \quad -\infty < \mu < \infty \quad (68)$$

with

$$\langle \mu \rangle = \bar{x}, \quad \text{var} \mu = \sigma^2, \quad (69)(70)$$

where

$$\bar{x} \equiv \frac{\sum_i (\sigma_i^2 + \tau^2)^{-1} x_i}{\sum_i (\sigma_i^2 + \tau^2)^{-1}}, \quad \sigma^{-2} \equiv \sum_i (\sigma_i^2 + \tau^2)^{-1}. \quad (71)(72)$$

For consistent data one can put  $\tau$  equal to zero so that  $\bar{x}$  becomes just the familiar weighted average with weights equal to  $1/\sigma_i^2$ . In the other limiting case of extreme inconsistency one can neglect the given errors  $\sigma_i$ , whereupon  $\bar{x}$  becomes the unweighted average and  $\sigma$  becomes equal to  $\tau/\sqrt{n}$ , as might have been expected.

If we have no information about the scale parameter  $\tau$  we take Jeffreys' prior. The joint posterior probability for  $\mu$ ,  $\tau$  and all  $\epsilon_i$  is then the right-hand side of Eq. 67 multiplied by  $d\tau/\tau$ . Integrating out the  $\epsilon_i$  one finds

$$p(\mu | x_1, \dots, \sigma_1, \dots) d\mu \propto d\mu \int_0^\infty \frac{d\tau}{\tau} \exp \left[ - \frac{(\mu - \bar{x})^2 + s^2}{2\sigma^2} \right] \prod_{i=1}^n \frac{1}{\sqrt{\sigma_i^2 + \tau^2}} \quad (73)$$

where  $\bar{x}$  and  $\sigma$  are defined as before (Eqs. 71, 72) and

$$s'^2 = \overline{x^2} - \bar{x}^2, \quad \bar{x}^2 = \frac{\sum_i (\sigma_i^2 + \tau^2)^{-1} x_i^2}{\sum_i (\sigma_i^2 + \tau^2)^{-1}}. \quad (74)(75)$$

The integration over  $\tau$  is easy only in the case of negligible known errors,  $\sigma_i \ll \tau$ . With  $\sigma_i = 0$  for all  $i$  the integrand reduces to a gamma distribution, integration of which yields the Student distribution

$$p(\mu | x_1, \dots, \sigma_1, \dots) d\mu = \frac{\Gamma(\frac{n}{2})}{\sqrt{\pi} \Gamma(\frac{n-1}{2})} \frac{d\xi}{(1+\xi^2)^{n/2}}, \quad -\infty < \xi \equiv \frac{\mu - \bar{x}}{s'} < \infty \quad (76)$$

with

$$\langle \mu \rangle = \bar{x}, \quad \text{var } \mu = \frac{s'^2}{n-3}, \quad (77)(78)$$

where  $\bar{x}$  and  $\overline{x^2}$  are now unweighted averages. Thus the uncertainty of  $\mu$  in this extreme case is determined by the sample variance  $s'^2$ , i. e. by the scatter of the data  $x_i$  (sometimes called the "external error"). This, of course, is just what we had before, when we discussed estimation of  $\mu$  from a given sample drawn from a Gaussian with unknown standard deviation (compare Eq. 28). For large  $n$  the distribution of  $\mu$  is practically Gaussian.

In general the integration over  $\tau$  must be performed numerically for a number of  $\mu$ -values chosen in such a way (in the vicinity of the measured values) that the  $\mu$ -distribution can be established with sufficient accuracy. If information on the relative reliability of the  $n$  measurements is available, one can introduce, instead of the common precision parameter  $1/\tau^2$  in Eq. 6 above, individual precision parameters  $w_i/\tau^2$ , where  $w_i$  is the relative precision (with respect to unrecognised errors) of the  $i$ -th experiment. It is straightforward to repeat the calculations with this modification. The only change in the results is that the weights  $1/(\sigma_i^2 + \tau^2)$  are replaced by  $1/(\sigma_i^2 + \tau^2/w_i)$ , and the unweighted averages  $\bar{x}$  and  $\overline{x^2}$  in the last equations become averages weighted by  $w_i$ .

### 3. CONCLUDING REMARKS

We reviewed briefly the probabilistic foundations of data evaluation, with special emphasis on Bayes' theorem and its consequences. The state of the art concerning prior probabilities was reported, in particular the group-theoretical approach and the technique of entropy maximisation. Parameter estimation was discussed in some detail, the exponential, Gaussian and Poisson distributions and the generalised least-squares problem providing practically important examples. The process leading from raw experimental data to evaluated nuclear cross section libraries including covariance files was outlined. The role of nuclear models and reaction theory and their use in least-squares fitting was illustrated by recent examples. It was explained how data correlations are induced by common errors and how correlated error input for least-squares adjustment can be set up if error sources and error estimates are known in sufficient detail. The utilisation of correlated output uncertainties, e. g. of cross section covariance matrices, was briefly indicated. Finally it was shown how the problem of inconsistent data can be tackled with the methods of Bayesian parameter estimation and group-theoretical priors.

The whole presentation was, by necessity, very short. The interested reader will find much more detailed material on data reduction and evaluation methods involving nuclear models and reaction theory in Ref. 21, for instance in the papers by W.P. Poenitz and M.R. Bhat there. Special experimental and analytical techniques for resonance cross sections are reviewed in Refs. 13 and 17, while Ref. 22 treats estimation of level-statistical parameters such as strength functions and mean level densities from resonance parameters, with account of missing weak levels.

A final remark concerns future work. Although data evaluation has reached a high level of sophistication and organisation in response to the demands of applied science and technology, it should be understood that there are unsolved problems, some of them quite basic. We mention four: (1) general recipes for the assignment of prior probabilities in multiparameter estimation problems, (2) generalisation of the least-squares formalism to discrete parameters such as resonance spins, (3) assessment of information entropy maximisation as a potential competitor

for least-squares fitting, and (4) practical methods for evaluation of inconsistent data.

ACKNOWLEDGEMENT

It is a pleasure to thank Dr. H. Rief (Euratom Ispra) and Prof. Y. Yeivin (Hebrew University, Jerusalem) for their encouragement to lecture on data evaluation at the Ispra Course on Data Uncertainty, Sensitivities, Consistency and Adjustment, April 1986. The present paper is an expanded version of the Ispra lecture notes.

Resolution:	Doppler	2.2 eV
	Pulse Width	0.5 eV
	Moderator etc.	1.7 eV
	Total (FWHM)	2.8 eV

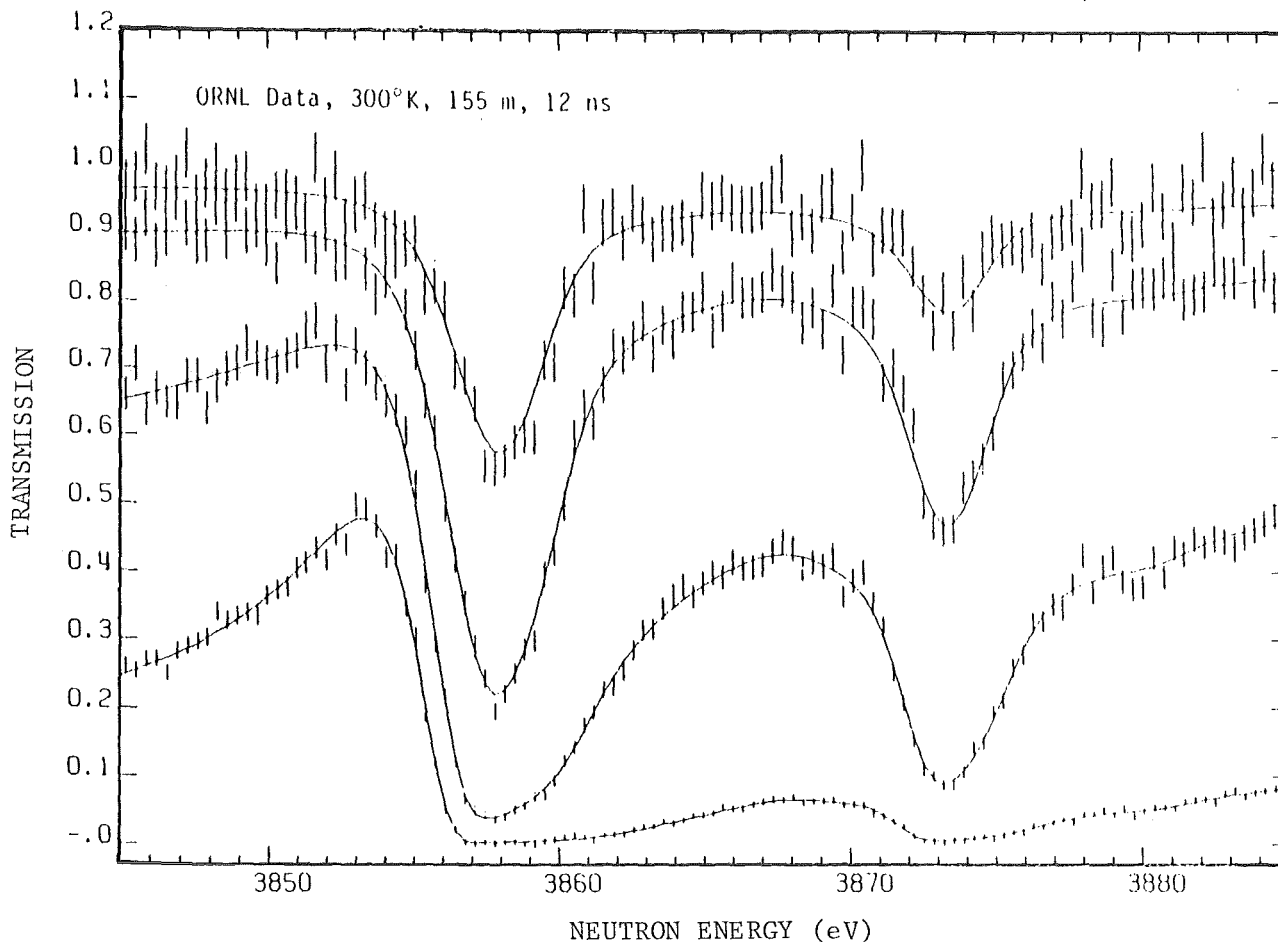


Fig. 1 - Simultaneous least-squares fit to four sets of transmission data obtained with four different  $^{238}\text{U}$  samples by D. Olsen, G. de Saussure et al. at the Oak Ridge Electron Linear Accelerator (ORELA) over the 3858 and 3873 eV resonances of  $^{238}\text{U}+n$ . Time-of-flight data are represented by bars, fitted curves are based on resonance theory including Doppler and resolution broadening (code SIOB) [12].

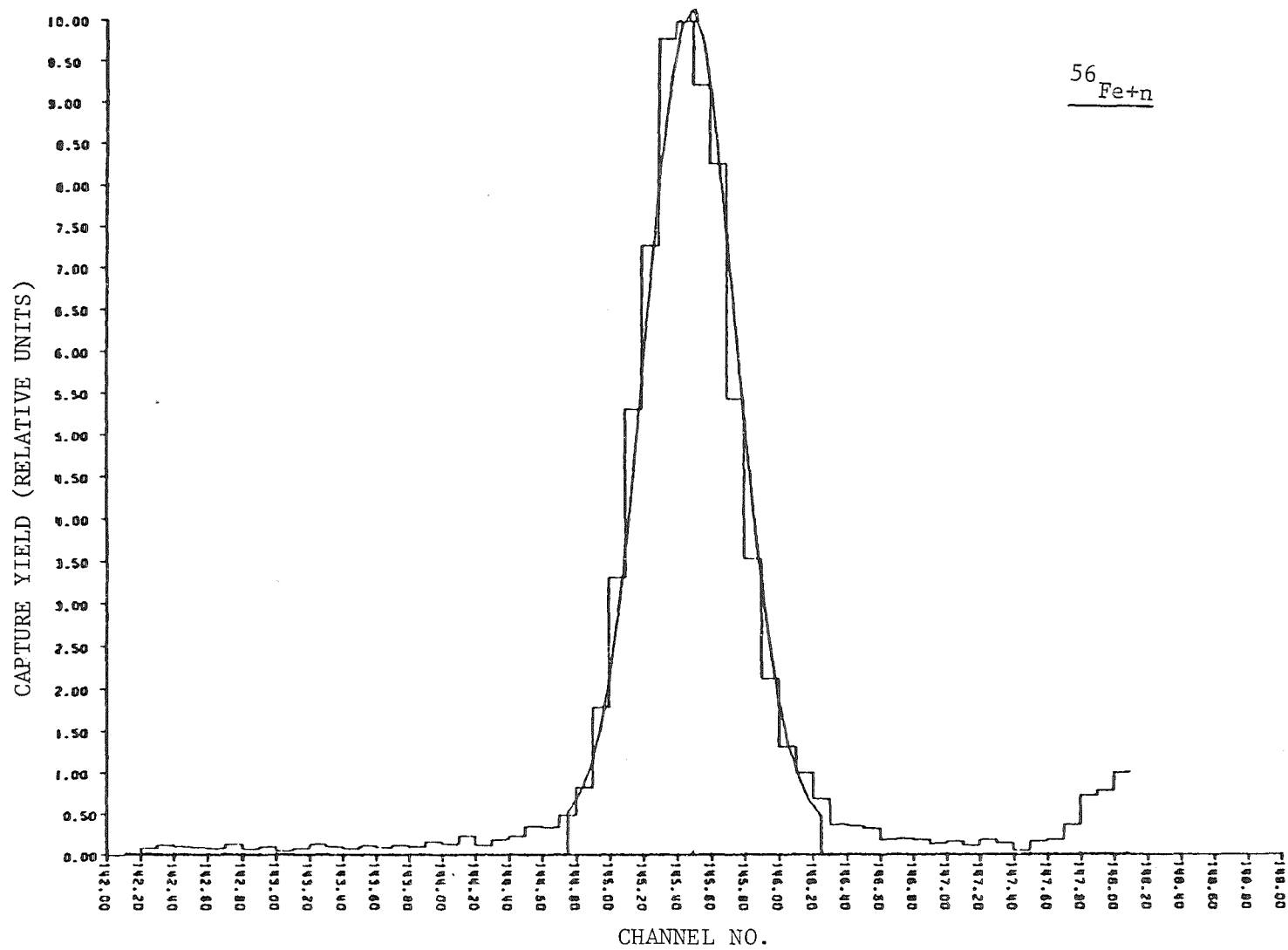


Fig. 2 - Neutron capture yield data measured at Geel by Corvi et al. across the 1.15 keV resonance of  $^{56}\text{Fe}+n$  (histogram) and R-matrix calculation including instrumental and Doppler broadening as well as multiple scattering (smooth curve, TACASI code [16]), from [14].

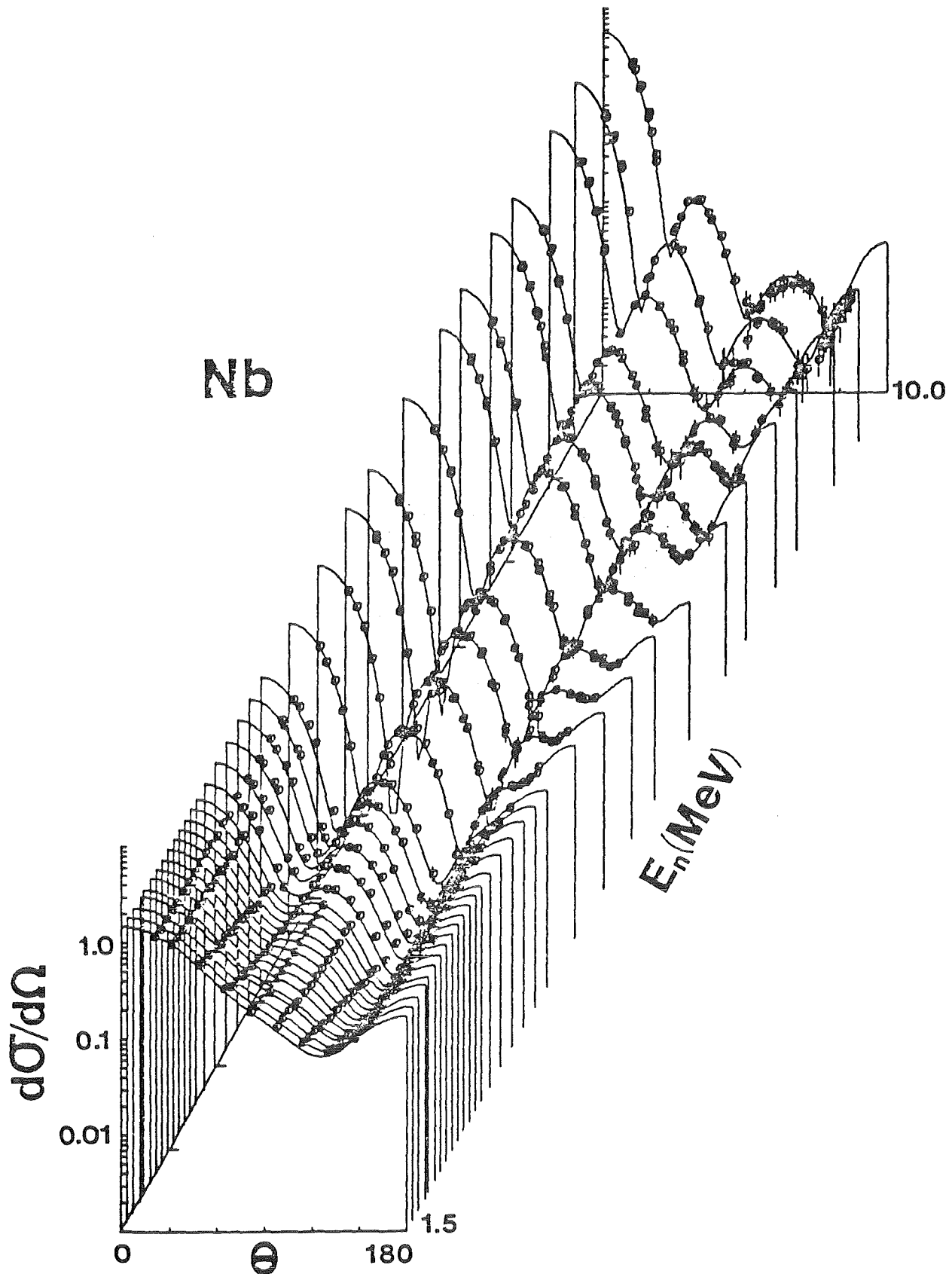


Fig. 3 - Measured differential-elastic-scattering cross sections of niobium. The measured values are indicated by circular data symbols and the results of optical-model calculations by curves. Cross sections are in b/sr and scattering angle ( $\theta$ ) in laboratory-system degrees. (from [15])

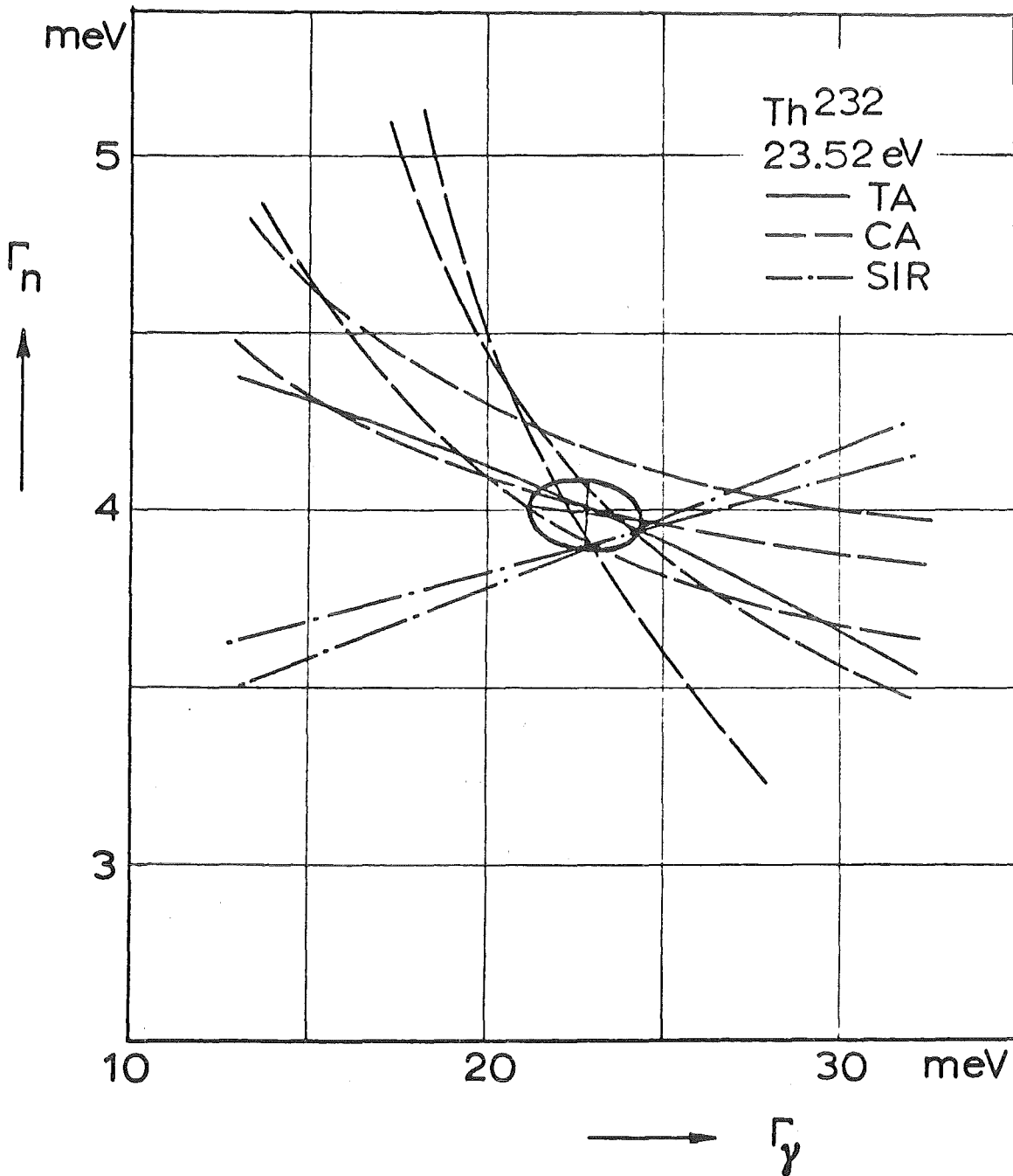


Fig. 4 - Illustration of nonlinearities and correlations in least-squares fitting: neutron width  $\Gamma_n$  and radiation width  $\Gamma_\gamma$  of the 23.52 eV resonance of  $^{232}\text{Th}$  were estimated by simultaneous adjustment to two measured transmission dip areas (TA), five capture peak areas (CA) and two self-indication ratios (SIR). The curves correspond to  $y_i(x_\infty)$  in Section 2.3 ( $i=1,2,\dots,8$ ), the "error ellipse" shows the best values and their correlated uncertainties (code TACASI). From [16].



REFERENCES

- [1] R.T. Cox, "Probability, Frequency and Reasonable Expectation",  
Am. J. Phys. 14 (1946) 1
- [2] H. Jeffreys, "Theory of Probability", Oxford (1939)
- [3] A. Wald, "Statistical Decision Functions", J. Wiley & Sons,  
New York (1950)
- [4] E.T. Jaynes, "Prior Probabilities", IEEE Trans. Syst. Cybern.  
SSC-4 (1968) 227; reprinted in [9] p. 114
- [5] E.T. Jaynes, "The Well-Posed Problem", Foundations of Physics  
3 (1973) 477; reprinted in [9] p. 131
- [6] E.T. Jaynes, "Confidence Intervals vs. Bayesian Intervals",  
in "Foundations of Probability Theory, Statistical Inference and  
Statistical Theories of Science", W.L. Harper and C.A. Hooker, Eds.,  
Reidel, Dordrecht (1976); reprinted in [9] p. 149
- [7] E.T. Jaynes, "Marginalization and Prior Probabilities", in  
"Bayesian Analysis in Econometrics and Statistics", A. Zellner, Ed.,  
North-Holland, Amsterdam (1980); reprinted in [9] p. 337
- [8] E.T. Jaynes, "Where Do We Stand on Maximum Entropy", in  
"The Maximum Entropy Formalism", R.D. Levine and M. Tribus, Eds.,  
M.I.T. Press, Cambridge, Massachusetts (1978), reprinted in [9] p. 210
- [9] E.T. Jaynes, "Papers on Probability, Statistics and Statistical  
Physics", R.D. Rosenkrantz, Ed., Dordrecht-Boston-London (1983)
- [10] R.A. Fisher, "The Fiducial Argument in Statistical Inference",  
Annals of Eugenics 6 (1935) 391, reprinted in  
R.A. Fisher, "Contributions to Mathematical Statistics",  
J. Tukey, Ed., New York - London (1950)
- [11] C.E. Shannon, Bell Syst. Tech. J. 27 (1948) 379 and 623,  
C.E. Shannon and W. Weaver, "The Mathematical Theory of Communi-  
cation", U. of Illinois Press, Urbana (1949)
- [12] F.G. Perey, private communication (1985)
- [13] F.H. Fröhner, "Applied Neutron Resonance Theory", Report KfK-2669  
(1978); reprinted in "Nuclear Theory for Applications", IAEA-SMR-43,  
ICTP Trieste (1980) p. 59
- [14] F. Corvi, private communication (1985)
- [15] A.B. Smith, D.L. Smith, R.J. Howerton, Report ANL/NDM-88 (1985)
- [16] F.H. Fröhner, Report GA-6906 (1966)

- [17] F.H. Fröhner, "New Techniques for Multi-Level Cross Section Calculation and Fitting", Report KfK 3081 (1980);  
reprinted in [21] vol. I, p. 375
- [18] Nancy M. Larson, F.G. Perey, Report ORNL/TM-7485, Oak Ridge (1980)  
cf. also N.M. Larson and F.G. Perey, Proc. Internat. Conf. on Nucl.  
Data for Basic and Appl. Sci., Santa Fe, May 1985 (in print)
- [19] F.H. Fröhner, B. Goel and U. Fischer, Proc. Meeting on Fast  
Neutron Capture Cross Sections, Argonne, ANL-83-4 (1982) p. 116
- [20] F.G. Perey, "Covariance Matrices of Experimental Data",  
Proc. Int. Conf. Neutron Physics and Nuclear Data, Harwell (1978)  
p. 104
- [21] Proc. Conf. on Nucl. Data Eval. Meth. and Proced., Brookhaven,  
Sept. 1980, (Ed. B.A. Magurno and S. Pearlstein), BNL-NCS-51363,  
vol. 1 and 2 (1981)
- [22] F.H. Fröhner, "Statistical Inference of Level Densities from  
Resolved Resonance Parameters", Report KfK 3553 (1983);  
reprinted in Proc. Meet. on Basic and Applied Problems of  
Nuclear Level Densities, Brookhaven, April 1983 (ed. M.R. Bhat),  
BNL-NCS-51694 (1983)