The Concept of Intrinsic Discrepancy Applied to the Comparison of Experimental and Theoretical Data for Benchmarking Spallation Models

Rolf Michel
Zentrum für Strahlenschutz und Radioökologie
Leibniz Universität Hannover
michel@zsr.uni-hannover.de, www.zsr.uni-hannover.de

1. Introduction

Recent developments in metrology concerning measurement uncertainty were laid down in the ISO *Guide to the expression of uncertainty in measurement* (GUM) [1]. In the GUM, uncertainties are evaluated either by "statistical methods" (type A) or by "other means" (type B). Type A uncertainties can be evaluated from repeated or counting measurements, while Type B uncertainties cannot. They are, for instance, uncertainties given in certificates of standard reference materials or of calibration radiation sources which are used in the evaluation of a measurement.

It is the distinction between the two ways (Type A and Type B), by which uncertainties are evaluated, which causes the problem with the statistical foundation of the GUM, i.e. whether it can be based or not on Bayesian and/or conventional statistics. Conventional statistics can only handle Type A uncertainties, but not Type B ones. Only by Bayesian statistics, uncertainties of both types can be consistently determined. Both types of uncertainties express quantitatively the actual state of incomplete knowledge about the quantities involved.

At the time of the first publication of the GUM, the statistical foundation of the GUM was not clear though a Bayesian theory of measurement uncertainty already existed providing a Bayesian foundation of GUM [2, 3].

Meanwhile, a supplement [4] to GUM [1] has been published, dealing comprehensively with the treatment of measurement uncertainty using the Monte Carlo (MC) method in complex measurement evaluations. There it is stated for the first time that only Bayesian statistics is capable of providing the statistical basis. Furthermore, the Principle of Maximum (Information) Entropy [5] was explicitly applied in [4] to obtain the required probability density functions (PDFs) for the uncertainty analysis based on the constraints set by the available information. More details about this may be found in [4] or elsewhere [6].

Though many results of the conventional and the Bayesian approaches are numerically practically equal, they must not be confused with each other because the understanding of the term "*probability*" is completely different in both statistics. The conventional or frequentist view is "*Probability is the stochastic limit of relative frequencies*" while the Bayesian view is "*Probability is a measure of the degree of belief an individual has in an uncertain proposition*". But, there are frequencies which do not represent probabilities and there are probabilities which cannot be expressed as frequencies. Bayesian statistics provides a more intuitive assessment methodology than conventional statistics, closer to the scientific thinking than conventional ones. For more details of these questions see e.g. [7 - 14].

2. The Concept of Intrinsic Discrepancy

In Bayesian statistics, the **intrinsic discrepancy** $\delta\{p_1,p_2\}$ is a very general measure of the divergence between two distributions of the random vector x described by their density functions p_1 and p_2 [15]:

$$\delta\{p_1,p_2\} = \min\left\{\int p_1(x)\ln\frac{p_1(x)}{p_2(x)}\mathrm{d}x, \int p_2(x)\ln\frac{p_2(x)}{p_1(x)}\mathrm{d}x\right\} \tag{1}$$

The following characteristics hold for the intrinsic discrepancy [15]:

"If $p_1(x \mid \theta)$ and $p_2(x \mid \lambda)$ describe two alternative distributions for data $x \in X$, one of which is assumed to be true, their intrinsic discrepancy $\delta\{p_1, p_2\}$ is the minimum expected log-likelihood ratio in favour of the true sampling distribution.

It may be shown that the intrinsic divergence is symmetric, non-negative (and it is zero if, and only if, $p_1(x) = p_2(x)$ almost everywhere). The intrinsic discrepancy is invariant under one-to-one transformations of x. Besides, it is additive: if $x = \{x_1, \ldots, x_n\}$ and $p_i(\mathbf{x}) = \prod_{j=1}^{n} q_i(x_j)$, then

$$\delta\{p_1, p_2\} = n\delta\{q_1, q_2\}. \tag{2}$$

The intrinsic discrepancy serves to define a useful type of convergence; a sequence of PDFs $\{p_i(\mathbf{x})\}_{i=1}^{\infty}$ converges intrinsically to a PDF $p(x)$ if (and only if) $\lim_{i\to\infty} \delta(p_i, p) = 0$ i.e., if (and only if) the sequence of the corresponding intrinsic discrepancies converges to zero.

Last, but not least, it is defined even if the support of one of the densities is strictly contained in the support of the other."

3. Intrinsic Discrepancy for Benchmarking Spallation Codes

In the former model and code intercomparisons [16, 17] methods of descriptive statistics were used to calculate global numbers for the agreement between experimental and calculated data. A metrological foundation of these general measures, i.e. figures of merit [16] and deviation factors [17], was not given and was also not intended. In order to put such measures on a actual metrological basis, here it is proposed to test the applicability of the concept of intrinsic discrepancy as a method of judgement about the agreement of experimental and theoretical data for benchmarking spallation codes,

To this end, a cross section as a function of energy or of mass or charge of residuals is understood as a probability density function (PDF), i.e. the probability of a particular reaction to occur as a function of energy, mass or charge. Any such PDF may be looked at with or without normalization in the region of interest of the independent variable.

That was indicated without normalization in my talk as Concept I. Then the concept of intrinsic discrepancy reads for the cross sections:

$$\delta\{\sigma_{\mathrm{exp}}(x),\sigma_{\mathrm{calc}}(x)\}=\min\left\{\int\ \sigma_{\mathrm{exp}}(x)\ln\frac{\sigma_{\mathrm{exp}}(x)}{\sigma_{\mathrm{calc}}(x)}\mathrm{d}x,\int\ \sigma_{\mathrm{calc}}(x)\ln\frac{\sigma_{\mathrm{calc}}(x)}{\sigma_{\mathrm{exp}}(x)}\mathrm{d}x\right\} \tag{3}$$

with $x=E,A,Z,\ldots$.

If the intrinsic discrepancy in form of equation 3 is used to compare different reactions, the numbers obtained are biased in the sense that more probable reactions are yielding higher numbers. Moreover, neither $\sigma_{\mathrm{exp}}(x)$ nor $\sigma_{\mathrm{calc}}(x)$ fulfill the necessary condition of a PDF, namely that $\int_{x_{\min}}^{x_{\max}}\sigma(x)\,\mathrm{d}x=1$ should hold. After some discussions with Alexandre Konobeev we came to the conclusion that what I called Concept II in my viewgraphs should be used.

The problems can be avoided by normalizing the experimental and calculated cross sections to the integral cross section of the experimental respectively calculated data in the region of interest (Concept II):

$$\sigma'_{\mathrm{exp}}(x)=\frac{\sigma_{\mathrm{exp}}(x)}{\int_{x_{\min}}^{x_{\max}}\sigma_{\mathrm{exp}}(x)\,\mathrm{d}x}\quad\text{and}\quad\sigma'_{\mathrm{calc}}(x)=\frac{\sigma_{\mathrm{calc}}(x)}{\int_{x_{\min}}^{x_{\max}}\sigma_{\mathrm{calc}}(x)\,\mathrm{d}x} \tag{4}$$

Then the intrinsic discrepancy reads:

$$\delta\{\sigma'_{\mathrm{exp}}(x),\sigma'_{\mathrm{calc}}(x)\}=\min\left\{\int\ \sigma'_{\mathrm{exp}}(x)\ln\frac{\sigma'_{\mathrm{exp}}(x)}{\sigma'_{\mathrm{calc}}(x)}\mathrm{d}x,\int\ \sigma'_{\mathrm{calc}}(x)\ln\frac{\sigma'_{\mathrm{calc}}(x)}{\sigma'_{\mathrm{exp}}(x)}\mathrm{d}x\right\} \tag{5}$$

with $x=E,A,Z,\ldots$.

3.1 Practical Application to Mass and Charge Distributions

Using the concept of intrinsic discrepancy for comparison of experimental and calculated mass and charge distributions is simple and straight forward. Given $\sigma_{\mathrm{exp}}(A_i)$ $(i=1,\ldots,n_{\mathrm{exp}})$ in a mass interval of interest $[A_{\min},A_{\max}]$ and calculated ones $\sigma_{\mathrm{calc}}(A_i)$ for the same masses. Then one normalizes both, the experimental and calculated cross sections to the integral experimental cross section:

$$\sigma'_{\exp}(A_i) = \frac{\sigma_{\exp}(A_i)}{\int\limits_{E_{\min}}^{E_{\max}} \sigma_{\exp}(A)\,\mathrm{d}A} = \frac{\sigma_{\exp}(A_i)}{\sum\limits_{k=1}^{n_{\exp}} \sigma_{\exp}(A_k)\,\Delta A_k} \quad \text{with} \quad \Delta A_k = 1 \tag{6}$$

$$\sigma'_{\text{calc}}(A_i) = \frac{\sigma_{\text{calc}}(A_i)}{\int\limits_{E_{\min}}^{E_{\max}} \sigma_{\text{calc}}(A)\,\mathrm{d}A} = \frac{\sigma_{\text{calc}}(A_i)}{\sum\limits_{k=1}^{n_{\exp}} \sigma_{\text{calc}}(A_k)\,\Delta A_k} \quad \text{with} \quad \Delta A_k = 1 \tag{7}$$

Then, the intrinsic discrepancy between the experimental and calculated cross sections is calculated by:

$$\delta\{\sigma'_{\exp}(A), \sigma'_{\text{calc}}(A)\} = \min\left\{ \sum_{i=1}^{n_{\exp}} \sigma'_{\exp}(A_i) \ln\frac{\sigma'_{\exp}(A_i)}{\sigma'_{\text{calc}}(A_i)}, \sum_{i=1}^{n_{\exp}} \sigma'_{\text{calc}}(A_i) \ln\frac{\sigma'_{\text{calc}}(A_i)}{\sigma'_{\exp}(A_i)} \right\}. \tag{8}$$

For charge distributions replace A by Z in equations 6 - 8.

3.2 Practical Application to Excitation Functions

The application of the concept of intrinsic discrepancy to excitation functions is a little bit more difficult since some integrals have to be evaluated. This may cause problems if applied also in cases where only a few experimental data exist. For experimentally well established excitation functions, i.e. many experimental cross sections in the energy region of interest, it can be more easily applied. The following procedure is proposed.

Let be $\sigma_{\exp}(E_i)$ $(i = 1,...,n_{\exp})$ the experimental cross sections in an interval of interest $[E_{\min}, E_{\max}]$ and $\sigma_{\text{calc}}(E_j)$ $(j = 1,...,n_{\text{calc}})$ a set of calculated data, then one normalizes both, the experimental and calculated cross sections, by the integral experimental cross section $\int_{E_{\min}}^{E_{\max}} \sigma_{\exp}(E)\,\mathrm{d}E$ in the energy region of interest:

$$\sigma'_{\exp}(E_i) = \frac{\sigma_{\exp}(E_i)}{\int\limits_{E_{\min}}^{E_{\max}} \sigma_{\exp}(E)\,\mathrm{d}E} \quad \text{and} \quad \sigma'_{\text{calc}}(E_j) = \frac{\sigma_{\text{calc}}(E_j)}{\int\limits_{E_{\min}}^{E_{\max}} \sigma_{\text{calc}}(E)\,\mathrm{d}E} \tag{9}$$

The intrinsic discrepancy between the experimental and calculated cross sections is defined by:

$$\delta\{\sigma'_{\exp}(E), \sigma'_{\text{calc}}(E)\} =$$

$$\min\left\{ \int_{E_{\min}}^{E_{\max}} \sigma'_{\exp}(E) \ln\frac{\sigma'_{\exp}(E)}{\sigma'_{\text{calc}}(E)}\,\mathrm{d}E, \int_{E_{\min}}^{E_{\max}} \sigma'_{\text{calc}}(E) \ln\frac{\sigma'_{\text{calc}}(E)}{\sigma'_{\exp}(E)}\,\mathrm{d}E \right\} \tag{10}$$

Equation 10 can be further evaluated and one obtains (leaving away the limits of the integration):

$$\delta\{\sigma'_{\mathrm{exp}}(E),\sigma'_{\mathrm{calc}}(E)\} = \min\left\{\begin{array}{l}\left(\int\sigma'_{\mathrm{exp}}(E)\ln\dfrac{\sigma_{\mathrm{exp}}(E)}{\sigma_{\mathrm{calc}}(E)}\,\mathrm{d}E + \ln\dfrac{\int\sigma_{\mathrm{calc}}(E)\mathrm{d}E}{\int\sigma_{\mathrm{exp}}(E)\mathrm{d}E}\right), \\ \left(\int\sigma'_{\mathrm{calc}}(E)\ln\dfrac{\sigma_{\mathrm{calc}}(E)}{\sigma_{\mathrm{exp}}(E)}\,\mathrm{d}E + \ln\dfrac{\int\sigma_{\mathrm{exp}}(E)\mathrm{d}E}{\int\sigma_{\mathrm{calc}}(E)\mathrm{d}E}\right)\end{array}\right\} \tag{11}$$

The problem of practical calculation is to evaluate the integrals given the discrete experimental and calculated cross sections at certain energy points. This problem can, however, be solved by interpolation and averaging.

4. Conclusion

The intrinsic discrepancy provides a useful measure for comparing experimental and theoretical data for benchmarking spallation codes which is well founded on metrological methods used today for uncertainty analysis. It has favorable characteristics, u. o. an appropriate convergence of $\delta\{\sigma'_{\mathrm{exp}}(E),\sigma'_{\mathrm{calc}}(E)\}$ which becomes zero if the distributions are identical. It also allows for the global comparison of many different reactions.

5. References:

[1] ISO, ISO Guide to the Expression of Uncertainty in Measurement, ISO, corrected reprint, Geneve, ISO1995.

[2] Weise, K., Wöger, W.; A Bayesian theory of measurement uncertainty. Meas. Sci. Technol. 4, 1-11 (1993).

[3] Weise, K., Wöger, W.; Meßunsicherheit und Meßdatenauswertung. (Berlin: Wiley-VCH) (1999).

[4] ISO, Guide to the expression of uncertainty in measurement (GUM) - Supplement 1: Propagation of distributions using a Monte Carlo method , JCGM 101:2008.

[5] Jaynes, E.T.; Probability Theory - The Logic of Science, edited by G.L. Bretthorst. (Cambridge University Press, Cambridge) (2003).

[6] Weise, K. Kanisch, G., Michel, R., Schläger, M., Schrammel, D., Täschner, M.; Monte Carlo determination of the characteristic limits in measurement of ionising radiation: Fundamentals and numerics, Radiation Protection Dosimetry 135 No. 3 (2009) 169 – 196; doi:10.1093/rpd/ncp105.

[7] Berger, J. O.; Statistical Decision Theory and Bayesian Analysis. (New York, Springer) (1985).

[8] Lee, P.M.; Bayesian Statistics: An Introduction. (Oxford University Press, New York) (1989).

[9] Bernardo, J. M., Smith, A.F.M.; Bayesian Theory. (Chicester, UK: Wiley) (1994).

[10] Bernardo, J.M.; Bayesian Statistics, in: Probability and Statistics (R.Viertl, ed.) of the Encyclopedia of Life Support Systems (EOLSS). (Oxford, UK, UNESCO) (2003).

[11] Jaynes, E.T.; Papers on probability, statistics and statistical physics. ed. R.D. Rosenkrantz, (Boston, New York, Dordrecht: Kluwe Publ.) (1989).

[12] Robert, Ch.P.; The Bayesian Choice (2nd. ed.). (New York: Springer) (2001).

[13] Gelman, A., Carlin, J. B., Stern, H. and Rubin, D.B.; Bayesian Data Analysis (2^{nd} edition). (London, Chapman) (2003).

[14] Gregory, P.C.; Bayesian logical data analysis for the physical sciences. (Cambridge University Press, Cambridge) (2005).

[15] Bernardo, J.M., Bayesian Statistics, in: *Probability and Statistics* (R. Viertl, ed.) *Encyclopedia of Life Support Systems* (EOLSS). Oxford, UK: UNESCO, 2003.

[16] Blann, M., Gruppelaar, H., Nagel, P., Rodens, J.; International Code Comparison for Intermediate Energy Nuclear Data, NEA/OECD, NSC/DOC(94)-2, Paris 1993.

[17] Michel, R., Nagel, P.; International Codes and Model Intercomparison for Intermediate Energy Activation Yields, NEA/OECD, NSC/DOC(97)-1, Paris 1997.