Peelle's Pertinent Puzzle and its Solution

R. Frühwirth^{1,a}, D. Neudecker^{2,b}, and H. Leeb^{2,c}

¹ Institut für Hochenergiephysik der ÖAW, Nikolsdorfer Gasse 18, 1050 Wien, Austria

² Atominstitut, Technische Universität Wien, Wiedner Hauptstraße 8–10, 1040 Wien, Austria

Abstract. Peelle's Pertinent Puzzle is a long-standing problem of nuclear data evaluation. In principle it is a phenomenon exhibiting unexpected mean values for experimental data affected by statistical and systematic errors. This occurs for non-linear functions of statistical quantities, e.g. for a product, but not for a sum. In the literature on nuclear data, this phenomenon was attributed to the underlying non-linearity of the relation between data. Here, we show in terms of Bayesian Statistics that Peelle's Pertinent Puzzle is primarily caused by improper estimates of covariance matrices of experiments and not exclusively by non-linearities. Applying the correct covariance matrix leads to the exact posterior expectation value and variance for an arbitrary number of uncorrelated measurement points which are normalized by the same quantity. It is also shown that the mean value converges in probability to zero with increasing number of observations, if the improper covariance matrix is applied.

1 Introduction

Peelle's Pertinent Puzzle (PPP) [1,2] denotes the occurrence of unexpected values of quantities that are estimated from experimental data which are affected by statistical and systematic errors. More specifically, the weighted mean is outside the range of the corresponding observations.

For highly correlated data, the occurrence of PPP might be reasonable. However, in some (non-linear) cases, PPP is caused by an improper construction of the experimental covariance matrix [3]. We show how to avoid PPP in this circumstance, in spite of the non-linear dependence of the estimated quantity on the observed data. Originally, PPP was formulated for the case of only two uncorrelated observations scaled by the same quantity. We extend the scope of our investigation to an arbitrary number of observations of the same physical variable scaled by the same scaling factor.

The puzzle is best illustrated by the example of two uncorrelated measurements q_1, q_2 of the same unknown quantity α , with standard errors σ_1, σ_2 . Both measurements are normalized by a factor N which is a stochastic quantity with expectation η and standard deviation σ_N . This results in a pair of correlated observations $r_i = Nq_i$, $i = 1, 2$, from which we wish to estimate the quantity $\rho = \eta\alpha$.

In the following discussion we will use an example with the following numerical values:

$$q_1 = 2.4, \sigma_1 = 0.12; q_2 = 2.0, \sigma_2 = 0.10;$$

$$N = 1.0, \sigma_N = 0.15, r_1 = Nq_1 = 2.4, r_2 = Nq_2 = 2.0.$$

2 Estimation

2.1 Least-squares Estimation

Obviously, r_1 and r_2 are correlated because of the scaling by the common factor N . Usually the covariance matrix C of r_1 and r_2 is approximated by linear error propagation, see [4,5]:

$$\hat{C}_{ij} = N^2 \delta_{ij} \sigma_i^2 + q_i q_j \sigma_N^2. \quad (1)$$

The least-squares estimator $\hat{\rho}$ of ρ is then given by [6]:

$$\hat{\rho} = (\mathbf{A}^T \hat{\mathbf{C}}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \hat{\mathbf{C}}^{-1} \mathbf{r},$$

with

$$\mathbf{A} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \mathbf{r} = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix}.$$

From this follows:

$$\hat{\rho} = \frac{\sum_{ij} \hat{C}_{ij}^{-1} r_j}{\sum_{ij} \hat{C}_{ij}^{-1}}, \quad \text{var}(\hat{\rho}) = \frac{1}{\sum_{ij} \hat{C}_{ij}^{-1}}.$$

In the Bayesian paradigm [7] the posterior of ρ is a Gaussian density, according to the principle of maximum entropy [8]:

$$\hat{P}(\rho | r_1, r_2) = \varphi(\rho; \hat{\rho}, \text{var}(\hat{\rho})). \quad (2)$$

In our example we obtain:

$$\hat{\mathbf{C}} = \begin{pmatrix} 0.144 & 0.108 \\ 0.108 & 0.100 \end{pmatrix}, \quad \hat{\mathbf{C}}^{-1} = \begin{pmatrix} 36.550 & -39.474 \\ -39.474 & 52.632 \end{pmatrix},$$

$$\hat{\rho} = \frac{-2.924 r_1 + 13.158 r_2}{10.234} = -0.286 r_1 + 1.286 r_2,$$

$$\hat{\rho} = 1.8857, \quad \text{var}(\hat{\rho}) = 0.0977.$$

The estimate $\hat{\rho}$ is outside the range of the observations. The reason for this strange result is the negative factor in front of r_1 , so that $\hat{\rho}$ is not a convex combination of r_1 and r_2 .

^a e-mail: rudolf.fruhwirth@oeaw.ac.at

^b e-mail: dn@kph.tuwien.ac.at

^c e-mail: leeb@kph.tuwien.ac.at

Fig. 1. The Gaussian posterior in Eq. (2) and the exact posterior in Eq. (3).

2.2 Exact Posterior

If we use the full information (q_1, q_2, N) and assume Gaussian distributions for α and η according to the principle of maximum entropy, we can construct the exact posterior density of ρ :

$$P(\rho|q_1, q_2, N) = \iint \delta(\rho - \eta\alpha) P(\alpha|q_1, q_2) P(\eta|N) d\alpha d\eta \quad (3)$$

with

$$P(\eta|N) = \varphi(\eta; N, \sigma_N^2), \quad P(\alpha|q_1, q_2) = \varphi(\alpha; \bar{q}, \bar{\sigma}^2) \quad (3a)$$

and

$$\bar{q} = \frac{q_1/\sigma_1^2 + q_2/\sigma_2^2}{1/\sigma_1^2 + 1/\sigma_2^2}, \quad \bar{\sigma}^2 = \frac{1}{1/\sigma_1^2 + 1/\sigma_2^2}. \quad (3b)$$

The posterior mean and the posterior variance of this density are

$$\begin{aligned} E(\rho) &= E(\eta\alpha) = E(\eta)E(\alpha) = N\bar{q}, \\ \text{var}(\rho) &= E(\eta^2\alpha^2) - E(\eta\alpha)^2 = \bar{q}^2\sigma_N^2 + N^2\bar{\sigma}^2 + \bar{\sigma}^2\sigma_N^2. \end{aligned}$$

In our example this evaluates to

$$E(\rho) = 2.1639, \quad \text{var}(\rho) = 0.1114.$$

The mean $E(\rho)$ now lies between r_1 and r_2 , conforming to our expectations.

The posterior is not Gaussian, but has very small skewness in our example:

$$\gamma_1(\rho) = \frac{6N\bar{q}\bar{\sigma}^2\sigma_N^2}{\text{var}(\rho)^{3/2}} = 0.04637.$$

Fig. 1 shows that the true posterior is very close to a Gaussian, therefore the non-linearity of the function

$$\rho = \eta\alpha$$

cannot be the only source of the pathologically low estimate $\hat{\rho}$, contrary to some of the pertinent literature [4,5].

2.3 The Proper Covariance Matrix

The linear error propagation in Eq. (1) requires the partial derivative $\partial\rho/\partial\eta = \alpha$. As α is unknown, it has to be approximated by a value a computed from the data. Before we specify a we note that using the same a for all matrix elements gives the following covariance matrix [3]:

$$\mathbf{C}_{ij}^a = N^2\delta_{ij}\sigma_i^2 + a^2\sigma_N^2.$$

Inserting this covariance matrix into the Gaussian density postulated by the principle of maximum entropy,

$$P^a(\rho|r_1, r_2) = \varphi(\rho; \rho^a, \text{var}(\rho^a)),$$

results in:

$$\rho^a = N\bar{q}, \quad \text{var}(\rho^a) = N^2\bar{\sigma}^2 + a^2\sigma_N^2.$$

Remarkably, the mean value ρ^a does not depend on the choice of a in our setting. The obvious candidate for a is the experimental mean value $\bar{q}$. Using this value gives the covariance matrix

$$\bar{\mathbf{C}}_{ij} = N^2\delta_{ij}\sigma_i^2 + \bar{q}^2\sigma_N^2,$$

and the posterior

$$\bar{P}(\rho|r_1, r_2) = \varphi(\rho; \bar{\rho}, \text{var}(\bar{\rho})), \quad (4)$$

with

$$\bar{\rho} = N\bar{q}, \quad \text{var}(\bar{\rho}) = N^2\bar{\sigma}^2 + \bar{q}^2\sigma_N^2. \quad (4a)$$

In our example:

$$\bar{\mathbf{C}} = \begin{pmatrix} 0.1198 & 0.1054 \\ 0.1054 & 0.1154 \end{pmatrix}, \quad \bar{\mathbf{C}}^{-1} = \begin{pmatrix} 42.493 & -38.810 \\ -38.810 & 44.114 \end{pmatrix},$$

$$\bar{\rho} = \frac{3.684 r_1 + 5.304 r_2}{8.988} = 0.410 r_1 + 0.590 r_2,$$

$$\bar{\rho} = 2.1639, \quad \text{var}(\bar{\rho}) = 0.1113.$$

The estimate $\bar{\rho}$ is now a convex combination of r_1 and r_2 and therefore inside the range of observations. The posterior mean of the density in Eq. (4) coincides with the true one, and the posterior variance differs only by the higher order term. The exact posterior variance can be obtained by choosing a such that $a^2 = \bar{q}^2 + \bar{\sigma}^2$. Fig. 2 shows that the posterior $\bar{P}$ is nearly identical to the true one.

2.4 Comparison of the Estimates

It is possible to rewrite $\hat{\rho}$ in Eq. (2) as

$$\hat{\rho} = N\bar{q} \left(1 + \frac{\sigma_N^2(q_1 - q_2)^2}{N^2(\sigma_1^2 + \sigma_2^2)} \right)^{-1}.$$

Thus we always have:

$$\hat{\rho}/\bar{\rho} \leq 1.$$

Equality holds if and only if $q_1 = q_2$. The various estimates are summarized in Table 1 and Fig. 3.

Fig. 2. The Gaussian posterior in Eq. (4) and the exact posterior in Eq. (3).

Table 1. Summary of the estimates $\bar{\rho}$, $\hat{\rho}$, their variances and the exact posterior moments.

Posterior	mean	variance	skewness
True posterior pdf	2.1639	0.1114	0.0464
Gaussian pdf with $\bar{\mathbf{C}}$	2.1639	0.1113	0
Gaussian pdf with $\hat{\mathbf{C}}$	1.8857	0.0977	0

Fig. 3. Graphical representation of the observations r_1, r_2 , their standard errors, and the estimates $\hat{\rho}$ and $\bar{\rho}$.

3 The Geometry of PPP

According to a result by Sivia [9], PPP occurs if and only if the covariance between r_1 and r_2 is larger than the smaller of the two variances:

$$\text{cov}(r_1, r_2) > \min(\text{var}(r_1), \text{var}(r_2)).$$

The underlying reason is that in this case the least-squares estimate is not a convex combination of r_1 and r_2 . Because of this the estimate is outside the range of the observations.

The covariance matrix $\hat{\mathbf{C}}$ in our example,

$$\hat{\mathbf{C}} = \begin{pmatrix} 0.144 & 0.108 \\ 0.108 & 0.100 \end{pmatrix},$$

satisfies Sivia's criterion:

$$\hat{\mathbf{C}}_{12} = 0.108 > 0.100 = \min(\hat{\mathbf{C}}_{11}, \hat{\mathbf{C}}_{22}) \Rightarrow \text{PPP!}$$

On the other hand, with the proper covariance matrix $\bar{\mathbf{C}}$ we have:

$$\bar{\mathbf{C}} = \begin{pmatrix} 0.1198 & 0.1054 \\ 0.1054 & 0.1154 \end{pmatrix},$$

$$\bar{\mathbf{C}}_{12} = 0.1054 < 0.1154 = \min(\bar{\mathbf{C}}_{11}, \bar{\mathbf{C}}_{22}) \Rightarrow \text{no PPP!}$$

In general, PPP cannot occur if we use $\bar{\mathbf{C}}$, because the covariance never exceeds the smaller variance:

$$\bar{\mathbf{C}}_{12} = \bar{q}^2 \sigma_N^2, \bar{\mathbf{C}}_{ii} = N^2 \sigma_i^2 + \bar{q}^2 \sigma_N^2 \Rightarrow \bar{\mathbf{C}}_{12} \leq \min(\bar{\mathbf{C}}_{11}, \bar{\mathbf{C}}_{22}).$$

As a consequence, the estimate is a convex combination of r_1 and r_2 and always lies between the two.

We now ask how frequently PPP occurs for a given experimental setup, when $\hat{\mathbf{C}}$ is used in the estimator. To answer this question, we have performed a simulation experiment using the following values:

$$\alpha = 2, \eta = 1, \sigma_1 = 0.12, \sigma_2 = 0.1, \sigma_N = 0.15, 0.20, 0.25.$$

The results from 5000 pairs of observations are shown in Fig. 4, for different values of σ_N . If the ratio of σ_N to σ_i rises, the correlation between r_1 and r_2 increases, and along with it the frequency of PPP.

4 PPP with Several Observations

We now generalize the problem to an arbitrary number m of observations $\mathbf{q} = (q_1, \dots, q_m)$, with the joint covariance matrix $\mathbf{\Sigma} = \text{diag}(\sigma_1^2, \dots, \sigma_m^2)$. The reduced information is then given by $\mathbf{r} = (r_1, \dots, r_m)$ with $r_i = Nq_i$, $i = 1, \dots, m$.

Linear error propagation as applied in [4,5] gives the following joint covariance matrix of $\mathbf{r}$:

$$\hat{\mathbf{C}} = N^2 \mathbf{\Sigma} + \sigma_N^2 \mathbf{q} \mathbf{q}^T.$$

Using the Sherman-Morrison identity to invert $\hat{\mathbf{C}}$, we can give an explicit expression for the least-squares estimator:

$$\hat{\rho} = N \bar{q} \frac{N^2}{N^2 + \sigma_N^2 \bar{\sigma}^2 (s_0 s_2 - s_1^2)},$$

with

$$s_0 = \sum_{i=1}^m \sigma_i^{-2}, \quad s_1 = \sum_{i=1}^m q_i \sigma_i^{-2}, \quad s_2 = \sum_{i=1}^m q_i^2 \sigma_i^{-2},$$

$$\bar{q} = \frac{s_1}{s_0}, \quad \bar{\sigma}^2 = \frac{1}{s_0}.$$

Alternatively, linear error propagation using $\bar{q}$ yields:

$$\bar{\mathbf{C}} = N^2 \mathbf{\Sigma} + \sigma_N^2 \mathbf{w} \mathbf{w}^T, \quad \mathbf{w} = (\bar{q}, \dots, \bar{q}).$$

The corresponding estimator reads:

$$\bar{\rho} = N \bar{q}, \quad \text{var}(\bar{\rho}) = \bar{q}^2 \sigma_N^2 + N^2 \bar{\sigma}^2.$$

The Cauchy-Schwarz inequality implies $s_1^2 \leq s_0 s_2$, so that we have proved for all m :

$$\hat{\rho} / \bar{\rho} \leq 1.$$

Equality holds only if all q_i are the same as already shown in [3].

We now proceed to give an estimate for the bias of $\hat{\rho}$. We can write $\hat{\rho}$ as

$$\hat{\rho} = N \bar{q} \frac{N^2}{N^2 + \sigma_N^2 (s_2 - s_1^2 / s_0)} = \frac{N^3 \bar{q}}{N^2 + \xi \sigma_N^2}, \quad (5)$$

Fig. 4. Frequency of PPP for various values of σ_N .

with $\xi = (s_2 - s_1^2/s_0)$. It is easy to show that ξ is invariant with respect to α . Without loss of generality we can therefore assume that $\alpha = 0$. In this case s_2 is χ^2 distributed with m degrees of freedom, and s_1^2/s_0 is χ^2 distributed with one degree of freedom. It follows from Cochran's theorem that ξ is χ^2 distributed with $m - 1$ degrees of freedom, and that ξ and $\bar{q}$ are independent. We therefore have:

$$E(\xi) = m - 1, \quad \text{var}(\xi) = 2(m - 1).$$

Table 2. Comparison of the bias $\hat{b}$ of $\hat{\rho}$ obtained from the simulation experiment with the approximate bias $\hat{b}_1$, and of the variance $\hat{\sigma}^2$ of $\hat{\rho}$ obtained from the simulation experiment with the approximate variance $\hat{\sigma}_1^2$, for $m = 2, 5, 10, 20$.

m	$\hat{b}$	$\hat{b}_1$	$\hat{\sigma}^2$	$\hat{\sigma}_1^2$
2	-0.0388	-0.0449	0.1058	0.1043
5	-0.1619	-0.1676	0.1162	0.1174
10	-0.3286	-0.3384	0.1239	0.1289
20	-0.5848	-0.5967	0.1261	0.1291

Linear error propagation gives the following approximate expressions for the mean, the bias and the variance of $\hat{\rho}$:

$$\begin{aligned}
 E(\hat{\rho}) &\approx \frac{N^3 \bar{q}}{N^2 + (m-1)\sigma_N^2}, \\
 \hat{b}_1 &\approx \frac{-N \bar{q}(m-1)\sigma_N^2}{N^2 + (m-1)\sigma_N^2}, \\
 \hat{\sigma}_1^2 &\approx \frac{N^6 \bar{\sigma}^2}{(N^2 + (m-1)\sigma_N^2)^4} + \\
 &+ \frac{\bar{q}^2 N^8 \sigma_N^2 + 8(m-1)\bar{q}^2 N^6 \sigma_N^4 + 9(m-1)^2 \bar{q}^2 N^4 \sigma_N^6}{(N^2 + (m-1)\sigma_N^2)^4}.
 \end{aligned}$$

We have checked the validity of the approximation by a simulation experiment with the following assumptions:

$$\alpha = 2, \quad \eta = 1, \quad \sigma_i \sim \text{Un}(0.1, 0.12), \quad \sigma_N = 0.15.$$

We have simulated 5000 data sets for $m = 2, 5, 10, 20$. Table 2 shows a comparison of the bias $\hat{b} = \langle \hat{\rho} \rangle - \eta\alpha$ with the average approximate bias $\langle \hat{b}_1 \rangle$, where the angular brackets denote the average over the simulated samples, and the analogous comparison of the variances. There is fairly good agreement, an indication that the approximation is adequate.

The asymptotic distribution of $\hat{\rho}$ for $m \rightarrow \infty$ is given by the following theorem.

Theorem 1 *The random variable $\hat{\rho}$ converges in probability to 0 as $m \rightarrow \infty$, provided that the σ_i stay bounded.*

Proof Let σ denote an upper bound of all σ_i . The variance of $\bar{q}$ is not larger than σ^2/m for all m and converges to 0. The mean of the numerator $n = N^3 \bar{q}$ in Eq. (5) does not depend on m and its variance converges to $\text{var}(N^3) \alpha^2$ as $m \rightarrow \infty$. Hence, for every $\epsilon > 0$ we can find a number $M > 0$ so that for all m the absolute value $|n|$ is smaller than M with a probability close to 1:

$$\Pr(|n| < M) > 1 - \epsilon/2.$$

Now consider the denominator $d = N^2 + \xi \sigma_N^2$ in Eq. (5). By choosing m sufficiently large, we can ensure for every $D > 0$ that

$$\Pr(d > D) > 1 - \epsilon/2.$$

To this end, we increase m until the $\epsilon/2$ -quantile of the distribution of d exceeds D . This is possible because for

large m the denominator is dominated by ξ , and because the $\epsilon/2$ -quantile of the χ^2 distribution with $m - 1$ degrees of freedom exceeds any bound as $m \rightarrow \infty$.

If $|n| < M$ and $d > D$ then certainly $|\hat{\rho}| = |n|/d < M/D$. From this follows:

$$\begin{aligned} \Pr(|\hat{\rho}| < M/D) &\geq \Pr(|n| < M \cap d > D) \\ &= \Pr(|n| < M) + \Pr(d > D) \\ &\quad - \Pr(n < M \cup d > D) \\ &\geq \Pr(|n| < M) + \Pr(d > D) - 1 \\ &> 1 - \epsilon/2 + 1 - \epsilon/2 - 1 \\ &= 1 - \epsilon. \end{aligned}$$

Now, for every $\delta > 0$ we can choose D such that $M/D < \delta$. Thus, for every $\delta > 0$ and every $\epsilon > 0$ we can find an m such that

$$\Pr(|\hat{\rho}| > \delta) < \epsilon. \quad \square$$

It is natural to ask whether for $m > 2$ there is a simple criterion for the occurrence of PPP, similar to the one found by Sivia for $m = 2$ [9]. If $\hat{\rho}$ is a convex combination of $r_1, \dots, r_m$, then PPP is excluded. The converse, however, is no longer true in general if $m > 2$. If PPP does not occur, it does not follow that $\hat{\rho}$ is necessarily a convex combination of $r_1, \dots, r_m$, as can be ascertained by the following counterexample:

$$\mathbf{r} = (2.0, 2.2, 2.4), \quad \hat{\rho} = -0.2 r_1 + 1.4 r_2 - 0.2 r_3 = 2.2.$$

$\hat{\rho}$ is always an affine combination of $r_1, \dots, r_m$, i.e., the coefficients sum to 1, so a general criterion would be required to determine whether an affine combination lies in the interval $[r_{\min}, r_{\max}]$. We are not aware of such a criterion, and it is doubtful whether one exists.

We now investigate the probability of PPP as a function of m , the experimental conditions staying the same. As we have not been able to find an explicit expression for the joint distribution of $\hat{\rho}$ and $r_{\min}$, we have resorted to the simulation experiment described above. Fig. 5 shows the empirical distribution of $\hat{\rho}$ vs. $r_{\min}$ for various values of m . PPP occurs whenever $\hat{\rho} < r_{\min}$. Clearly the frequency of PPP rises with m . For $m = 20$ only a single sample out of 5000 does not show PPP, so the probability of the latter must already be very close to 1.

We can explain this effect by the following considerations. For small m the distribution of $r_{\min}$ is centered somewhat below $\eta\alpha$. With rising m the distribution is shifted toward 0. If $\eta\alpha$ is sufficiently large this shift is slower than the shrinking of $\hat{\rho}$ toward 0. For larger m $\hat{\rho}$ is therefore virtually always smaller than $r_{\min}$. It has to be noted, however, that for extremely large values of m , many orders of magnitude beyond the range of practical relevance, negative values of $r_{\min}$ are possible, in which case PPP does not occur.

The situation is different if $\eta\alpha$ is small compared to σ_N and the σ_i . In the extreme case of $\alpha = 0$, $r_{\min}$ is virtually always negative, while $\hat{\rho}$ shrinks to zero, so that PPP is increasingly unlikely. As an illustration, Fig. 6 shows the distribution of $\hat{\rho}$ vs. $r_{\min}$ for $\alpha = 0.2$ and $\eta = 1$, with the same σ_N and σ_i as in Fig. 5. The mean value of $r_{\min}$ shifts at the same rate as before, but the shrinking of $\hat{\rho}$ towards 0 is now slower than before, because the bias of $\hat{\rho}$ is proportional to $\eta\alpha$. Thus PPP does not occur at all.

Fig. 5. Occurrence of PPP for $\alpha = 2$ and $m = 2, 5, 10, 20$.

Fig. 6. No occurrence of PPP for $\alpha = 0.2$ and $m = 2, 5, 10, 20$.

5 Conclusions

We have shown that the occurrence of unexpectedly small mean values, termed Peelle's Pertinent Puzzle, is caused by improper construction of the covariance matrix of the reduced observations $r_i = Nq_i$, and not, as widely believed, by the non-linearity of the underlying functional relationship $\rho = \eta\alpha$, see also [3]. If the proper value of the derivative $\partial\rho/\partial\eta$ is used in the linear error propagation, the estimate is always in the range of the observations, and the Gaussian posterior is nearly indistinguishable from the exact posterior. As the frequency of PPP rises with the number of observations for sufficiently large α , the proper construction of the joint covariance matrix is absolutely essential.

In the case of more complex non-linear relationships, one should check whether the exact posterior density can be well approximated by a Gaussian density based on the reduced observations. Otherwise, more moments or even the entire posterior density have to be put at the disposal of the subsequent analysis.

We thank the American Nuclear Society for the permission to reproduce Fig. 1, Fig. 3, part of Eqs. (6) and (17), as well as Eqs. (9), (13) and (20) from the paper "Peelle's Pertinent Puzzle: A Fake Due to Improper Analysis", by Denise Neudecker et al., published in Nuclear Science and Engineering, Vol. 170, Issue 1 (2012) p. 54-60; Copyright January 2012 by the American Nuclear Society, La Grange Park, Illinois, USA. This work was partly supported by the EURATOM project ANDES. The views and opinions expressed herein do not necessarily reflect those of the European Commission. We thank Peter Schillebeeckx for initial discussions.

References

1. R.W. Peelle, *Peelle's Pertinent Puzzle* (Informal ORNL memorandum, Oak Ridge Oct. 13, 1987)
2. D.L. Smith, *Probability, Statistics and Data Uncertainties in Nuclear Science and Technology* (American Nuclear Society, La Grange Park, Illinois 1991)
3. D. Neudecker, R. Frühwirth, H. Leeb, Nucl. Sci. Eng. **170** Issue 1, (2012) 54-60
4. F.H. Fröhner, *Reactor Physics and Reactor Computation* (Ben-Gurion University of the Negev Press, Beer-Sheva, Israel 1994) 287
5. G. D'Agostini, Nucl. Instr. Meth. **A346**, (1994) 306
6. T. Burr, T. Kawano, P. Talou, F. Pan, N. Hengartner, Algorithms **4**, (2011) 28
7. Th. Bayes Rev., Phil. Trans. Roy. Soc. **53**, (1763) 370
8. D.S. Sivia, *Data Analysis A Bayesian Tutorial* (Clarendon Press, Oxford 1996)
9. D.S. Sivia, *Advanced Mathematical and Computational Tools in Metrology VII* (World Scientific Publishing Company, Lisbon, Portugal 2006)