Random sampling and validation of covariance matrices of resonance parameters

Lucijan Plevnik[a] and Gašper Žerovnik[b]

Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia

Abstract. Analytically exact methods for random sampling of arbitrary correlated parameters are presented. Emphasis is given on one hand on the possible inconsistencies in the covariance data, concentrating on the positive semi-definiteness and consistent sampling of correlated inherently positive parameters, and on the other hand on optimization of the implementation of the methods itself. The methods have been applied in the program ENDSAM, written in the Fortran language, which from a file from a nuclear data library of a chosen isotope in ENDF-6 format produces an arbitrary number of new files in ENDF-6 format which contain values of random samples of resonance parameters (in accordance with corresponding covariance matrices) in places of original values.

The source code for the program ENDSAM is available from the OECD/NEA Data Bank.

The program works in the following steps: reads resonance parameters and their covariance data from nuclear data library, checks whether the covariance data is consistent, and produces random samples of resonance parameters. The code has been validated with both realistic and artificial data to show that the produced samples are statistically consistent. Additionally, the code was used to validate covariance data in existing nuclear data libraries. A list of inconsistencies, observed in covariance data of resonance parameters in ENDF-VII.1, JEFF-3.2 and JENDL-4.0 is presented. For now, the work has been limited to resonance parameters, however the methods presented are general and can in principle be extended to sampling and validation of any nuclear data.

1. Introduction

Compared to deterministic uncertainty propagation [1], random sampling is less time efficient, however may be more accurate and less cumbersome.

In the covariances of existing nuclear data libraries, different mathematical inconsistencies are a common phenomenon. Some of the inconsistencies can be overcome while performing deterministic uncertainty propagation, but they completely disable random sampling. One of the requirements for random sampling is positive semi-definiteness of the covariance/correlation matrix [2,3]. When sampling inherently positive parameters (e.g. cross sections or resonance widths), their probability distributions is not governed by Gaussian function, but lognormal distribution can be used instead [4]. This distribution requires additional restrictions on correlation coefficient values [5] which are not taken into account in all libraries [2]. In practice, approximative methods avoiding lognormal distribution can be used, e.g. Gaussian distribution with truncated negative values. These methods generally work well with small uncertainties ($\lesssim 30\%$), however create significant biases when sampling parameters with larger relative uncertainties.

In this paper, analytically exact methods to randomly sample nuclear data are presented. The methods have been applied in the program ENDSAM, written in the Fortran 95 language, which receives

- a number of required samples n,
- a file from a nuclear data library of a certain isotope in ENDF-6 format [6],

as input and returns n new files in ENDF-6 format which contain values of random samples of resonance parameters in places of original values. The source code for the program ENDSAM is available from OECD/NEA Data Bank.

The code has been validated on three major contemporary nuclear data libraries: ENDF/B-VII.1 [7], JEFF-3.2 [8], and JENDL-4.0u2 [9]. In Sect. 4 the results given by ENDSAM are compared with expected values, while Sect. 5 contains the list of all isotopes with inconsistent data from aforementioned libraries.

More details about this work can be found in the paper [10].

2. Normal and lognormal parameters

In nuclear data libraries, physical parameters are represented by mean values and (possibly) the corresponding covariances. The usual assumption is that they are distributed according to the normal distribution. However, normally distributed parameters may attain positive or negative values, while in reality some parameters, e.g. neutron widths, are inherently positive. If a parameter has a small relative uncertainty ($< 30\%$), then normal

---

[a] e-mail: lucijan.plevnik@ijs.si

[b] Currently at EC-JRC Institute for Reference Materials and Measurements, Geel, Belgium; e-mail: gasper.zerovnik@ijs.si

distribution is a good approximation for most practical purposes. However, as emphasized in [2] and demonstrated with examples in Sect. 4, in some rare cases relative uncertainties of parameters given in nuclear data libraries may be extremely large (even 150000%). Therefore, we describe positive parameters with a lognormal distribution [4], which means that their logarithms are normally distributed.

We use $\langle x \rangle$, $\sigma(x)$, $\mathrm{ru}(x) = \frac{\sigma(x)}{\langle x \rangle}$, and $\mathrm{corr}(x, y)$ to denote mean of x, standard deviation of x, relative uncertainty of x, and correlation between x and y, respectively.

Suppose that x and y are lognormally distributed, that is $x = e^u$ and $y = e^v$ for normally distributed random parameters u and v, respectively. Then moments of normal parameters may be expressed in terms of moments of x and y in the following way:

$$\langle u \rangle = \log \frac{\langle x \rangle}{\sqrt{1 + \mathrm{ru}(x)^2}}, \quad \sigma(u) = \sqrt{\log \left(1 + \mathrm{ru}(x)^2\right)}, \tag{1}$$

$$\mathrm{corr}(u, v) = \mathrm{corr}(x, v) \frac{\mathrm{ru}(x)}{\sqrt{\log \left(1 + \mathrm{ru}(x)^2\right)}}, \tag{2}$$

$$\mathrm{corr}(u, v) = \frac{\log \left(1 + \mathrm{corr}(x, y)\, \mathrm{ru}(x)\, \mathrm{ru}(y)\right)}{\sqrt{\log \left(1 + \mathrm{ru}(x)^2\right) \log \left(1 + \mathrm{ru}(y)^2\right)}}, \tag{3}$$

see [10] and the references therein. We now deduce from $-1 \leq \mathrm{corr}(u, v) \leq 1$ and Eqs. (2) and (3) that correlations involving lognormal parameters should satisfy

$$|\mathrm{corr}(x, v)| \leq \frac{\sqrt{\log \left(1 + \mathrm{ru}(x)^2\right)}}{|\mathrm{ru}(x)|} \tag{4}$$

and

$$\frac{e^{-\sqrt{\log\left(1+\mathrm{ru}(x)^2\right)\log\left(1+\mathrm{ru}(y)^2\right)}} - 1}{\mathrm{ru}(x)\,\mathrm{ru}(y)} \leq \mathrm{corr}(x, y)$$

$$\leq \frac{e^{\sqrt{\log\left(1+\mathrm{ru}(x)^2\right)\log\left(1+\mathrm{ru}(y)^2\right)}} - 1}{\mathrm{ru}(x)\,\mathrm{ru}(y)}. \tag{5}$$
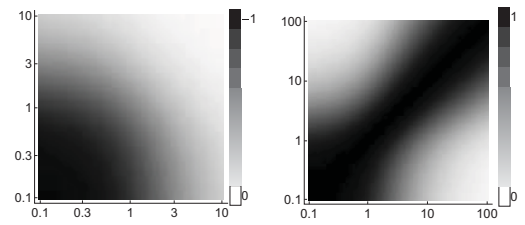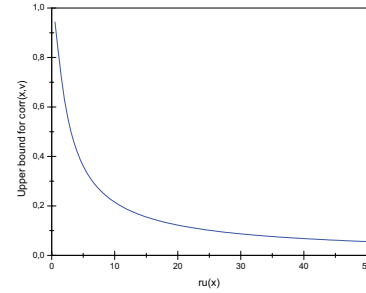
Figure 1 shows upper and lower bound for correlation between lognormal parameters in dependency of their relative uncertainties.

3. Random sampling method

Consider a combination of normal and lognormal parameters $x_1, \ldots, x_n$ and a collection of normal parameters

$$\widetilde{x}_j = \begin{cases} x_j \colon x_j \text{ is normal,} \\ \log x_j \colon x_j \text{ is lognormal,} \end{cases} \quad j = 1, \ldots, n. \tag{6}$$

If x_j is lognormal, then means and deviations of $\widetilde{x}_j$ can be computed by respectively using Eq. (1). Moreover, if x_i is another parameter, then depending on its type, correlation $\mathrm{corr}(x_i, x_j)$ has to satisfy either the condition in Eqs. (4) or (5). If it does, then $\mathrm{corr}(\widetilde{x}_i, \widetilde{x}_j)$ can be computed either by Eqs. (2) or (3), respectively.

Figure 1. Upper bound for correlation of lognormal and normal parameter (top); lower (bottom left) and upper (bottom right) bound for correlation of two lognormal parameters as functions of relative uncertainties.

However, it was pointed out in [2] and will be demonstrated with examples in Sect. 5 that conditions in Eqs. (4) and (5) are often not taken into account in nuclear data libraries. In such cases we assume that the correlation matrix from data library is given for a collection of normal parameters, so it applies to the collection $\widetilde{x}_1, \ldots, \widetilde{x}_n$ and not to the original collection of parameters.

Another restriction which is sometimes not taken into account is that a correlation matrix must be positive semi-definite. Some cases appear when even a matrix given in a library has some negative eigenvalues and in some cases this happens to be the case after transformation of correlation coefficients. In this case we apply Higham's method [11] to find the nearest correlation matrix in Frobenius norm and work with the resulting matrix. This correction method may not be universally the best solution, however it is mathematically rigorous.

We summarize the following algorithm to randomly sample a collection $x_1, \ldots, x_n$.

Algorithm 3.1 *1) For $j = 1, \ldots, n$: if x_j is lognormal, then compute $\langle \widetilde{x}_j \rangle$ and $\sigma\left(\widetilde{x}_j\right)$ by respectively applying Eq. (1).*

2) Check whether all correlations satisfy theoretical bounds for lognormal parameters from Eqs. (4) and (5).

- *If not, then set $\widetilde{C} = C$.*
- *If yes, then compute new corrlelation matrix $\widetilde{C} = \left[\mathrm{corr}\left(\widetilde{x}_i, \widetilde{x}_j\right)\right]$ by applying Eqs. (2) or (3) when necessary.*

3) Check if $\widetilde{C}$ is positive definite by trying to compute its Cholesky factorization.

- *If it is, then save its lower Cholesky factor into L.*
- *If it is not, then compute nearest correlation matrix of $\widetilde{C}$ in Frobenius norm. Then compute lower Cholesky factor L of resulting matrix.*
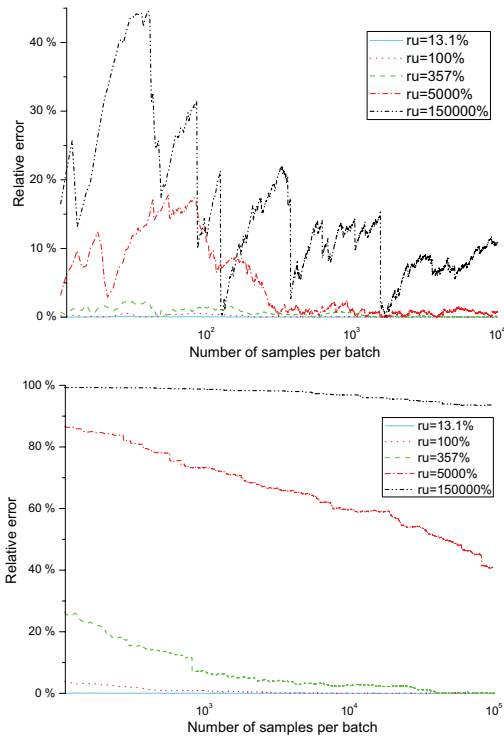
Figure 2. Relative error of samples of lognormal parameters with different values of relative uncertainty. Averaged over 1000 batches of samples. Up: sample means. Below: sample standard deviations. (Data: ENDF B/VII.1, ^{35}Cl.)

Figure 3. Absolute error of sample correlations for different expected values of correlation and relative uncertainty. Up: normal and lognormal parameter. Below: two lognormal parameters. Averaged over 1000 batches of samples. (Data are generated artificially.)

4) *Compute standard normal random vector $v = (v_1, \ldots, v_n)$ using the Box-Muller method [12].*

5) *Compute normal random vector $u = (u_1, \ldots, u_n)$ by formula*

$$u_j = \langle \widetilde{x}_j \rangle + \sigma\left(\widetilde{x}_j\right)(Lv)_j, \quad j = 1, \ldots, n. \quad (7)$$

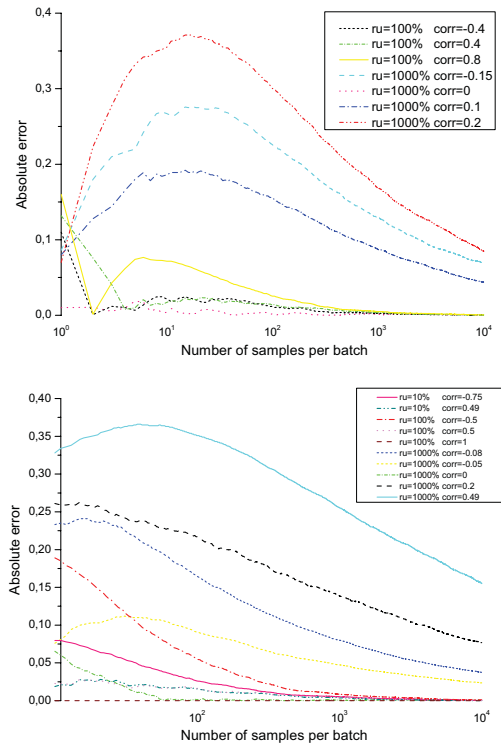6) *For $j = 1, \ldots, n$: sample value of x_j equals*

$$\begin{cases} u_j : x_j \text{ is normal}, \\ e^{u_j} : x_j \text{ is lognormal}. \end{cases} \quad (8)$$

Among resonance parameters in nuclear data libraries, resonance energy ER is the only parameter which is assumed to be normally distributed by ENDSAM, since its relative uncertainty is always very small.

4. Validation of the sampling code

The sampling code ENDSAM was validated by sampling different examples of resonance parameters from existing nuclear data libraries. Figure 2 shows relative errors of the first two moments. A "batch of samples" is referred to as a sequence of generated samples, e.g. 10^5 samples in the graph below in Fig. 2. In each batch, distribution moments of the first i samples were calculated, averaged over all batches, and plotted as a function of i. We considered the following resonance parameters of ^{35}Cl with different values of relative uncertainties from the ENDF B/VII.1 library:

13.1%: neutron width with resonance spin $J = 1$ at energy around 55 keV, and orbital angular momentum $l = 0$,

100%: proton width with $J = 1$ at around 183 keV, $l = 0$,
357%: proton width with $J = 2$ at around 15 keV, $l = 0$,
5000%: proton width with $J = 2$ at around 7.6 MeV, $l = 0$,
150000%: proton width with $J = 1$ at around 68 keV, $l = 0$.

Sample means of a parameter with relative uncertainty 100% converge to relative error less than 1% after less than 10 samples per batch, while error is still around 11% after 10^4 samples per batch for a parameter with relative uncertainty 150000%. Convergence of standard deviations is slower. For the first parameter, the relative error falls within 1% after 500 samples per batch, while it is nowhere near required value after 10^5 samples per batch for the second parameter.

On the other hand, convergence of correlations depends on more factors. We give analysis for two artificially chosen parameters x and y with different expected values of moments.

Convergence depends on $\text{corr}(x, y)$ and $\text{ru}(x)$ when x is lognormal and y is normal. If $\text{corr}(x, y)$ has zero value, then absolute error falls within 0.01 almost immediately and does not vary significantly for different values of $\text{ru}(x)$. If correlation is nonzero, then larger value of $|\text{ru}(x)|$ results in slower convergence, see Fig. 3.

Finally, correlation and relative uncertainties influence rate of convergence of correlation between two lognormal parameters x and y. Again, zero value of $\text{corr}(x, y)$ yields quick convergence independently of relative uncertainties and larger values of relative uncertainties imply slower convergence. Figure 3 shows absolute error as a function of number of samples for $\text{ru}(x) = 100\%$ and different values of $\text{ru}(y)$ and $\text{corr}(x, y)$.

Table 1. Table of invalid correlation matrices from major evaluated nuclear data libraries.

ISSUE \ LIBRARY	ENDF/B-VII.1	JEFF 3.2	JENDL 4.0
COMPATIBILITY WITH LOG-NORMAL DISTRIBUTION	^{35,37}Cl, ^{41}K, $^{46-50}$Ti, ^{50,53}Cr, ^{152,154}Gd, ^{155,157}Gd, ^{183,184,186}W, ^{232}Th, ^{236}Np	^{32,34}S, ^{37}Cl, ^{39}Ar, ^{50,53}Cr, ^{63,65}Cu, $^{64,66-68,70}$Zn, ^{69,71}Ga, ^{70,73}Ge, ^{75}As, ^{76,77,80}Se, ^{79,81}Br, ^{86}Kr, ^{85,87}Rb, $^{86-88}$Sr, ^{90,94}Zr, ^{122}Sn, ^{122,124}Te, ^{138}La, ^{136}Ce, $^{152-154}$Gd, ^{156,158,160}Dy, $^{166-170}$Er, ^{169}Tm, $^{170-174,176}$Yb, ^{175,176}Lu, ^{181}Hf, ^{185,187}Re, $^{186-189}$Os, ^{195}Pt, ^{197}Au, ^{203}Tl, $^{204-208}$Pb, ^{209}Bi, ^{232}Th, ^{231}Pa, ^{241}Am, ^{242,244}Cm,	^{53}Cr, ^{55}Mn, ^{236}Np
POSITIVE SEMI-DEFINITENESS	^{39}K, ^{52}Cr, ^{55}Mn, ^{60}Ni, ^{182}W, ^{232}Th, ^{238}Np	^{34}S, ^{39}Ar, ^{52}Cr, ^{55}Mn, ^{66,68,70}Zn, ^{69,71}Ga, ^{73}Ge, ^{75}As, ^{76,77}Se, ^{79,81}Br, ^{85}Rb, ^{86,87}Sr, ^{122}Te, ^{148}Gd, ^{166,168,170}Er, ^{169}Tm, ^{172,176}Yb, ^{176}Lu, ^{197}Au, ^{203}Tl, $^{204-208}$Pb, ^{209}Bi, ^{232}Th, ^{231}Pa, ^{241}Am	^{52}Cr, ^{238}Np

5. Validation of nuclear data libraries

The correlation matrices of resonance parameters in ENDF/B-VII.1, JEFF 3.2 and JENDL 4.0 libraries have been analyzed and the list containing information of possible faults mentioned in Sect. 3 in the correlation/covariance matrices is given in Table 1.

For example ^{232}Th from ENDF/B-VII.1 and JEFF 3.2 is incompatible with lognormal distribution. The resonance covariance file for resolved energy range contains 2781 resonance parameters. Parameter No. 36 corresponds to radiation width of 12th resonance and its relative uncertainty equals 0.13. On the other hand, parameter No. 797 corresponds to neutron width of 266th resonance with (extremely large) relative uncertainty of 27.66. Both parameters are sampled according to lognormal distribution. Their correlation equals -0.205 which is smaller than required lower bound from Eq. (5):

$$\frac{e^{-\sqrt{\log\left(1+\text{ru}(x)^2\right)\log\left(1+\text{ru}(y)^2\right)}}-1}{\text{ru}(x)\text{ru}(y)} = -0.08.$$

6. Conclusions

An open source code, named ENDSAM, has been implemented, which:

- reads resonance parameters and their covariance data from nuclear data libraries,
- checks whether the covariance data is consistent,
- produces random samples of resonance parameters.

In the future, the code is planned to be extended to sampling of other nuclear data, such as cross sections, angular and energy distributions, and other. The code has been validated against realistic and artificial data to show that the produced samples meet statistical expectations. Additionally, the code can be used to validate existing nuclear data libraries. A list of inconsistencies, spotted in covariance data of resonance parameters in ENDF-VII.1, JEFF-3.2 and JENDL-4.0 is presented.

References

[1] I. Kodeli, Nucl. Sci. Eng. **138**, 45–66 (2001)
[2] G. Žerovnik, A. Trkov, L.C. Leal, Nucl. Instr. Meth. A **743**, 39–43 (2014)
[3] R. Vanhanen, Nucl. Sci. Eng. **179**, 411–422 (2015)
[4] E.T. Jaynes, IEEE T. Syst. Sci. Cyb. **4**, 227–241 (1968)
[5] G. Žerovnik, A. Trkov, D.L. Smith, R. Capote, Nucl. Instr. Meth. A **727**, 33–39 (2013)
[6] Cross Sections Evaluation Working Group (Edited by Herman, M. and Trkov, A.), *ENDF-6 Formats Manual, Data formats and procedures for the Evaluated Nuclear Data File ENDF/B-VI and ENDF/B-VII* (National Nuclear Data Center (NNDC), Brookhaven National Laboratory, Upton, USA, 2011)
[7] M.B. Chadwick, et. al., Nucl. Data Sheets (N.Y. N.Y.) **112**, 2887–2996 (2011)
[8] OECD/NEA, NEA Data Bank, (2014), Available online: `https://www.oecd-nea.org/dbforms/data/eva/evatapes/jeff_32/`
[9] K. Shibata, et. al., J. Nucl. Sci. Technol. **48**, 1–30 (2011)
[10] L. Plevnik, G. Žerovnik, Ann. Nucl. Energy **94**, 510–517 (2016)
[11] N.J. Higham, IMA J. Numer. Anal. **22**, 329–343 (2002)
[12] G.E.P. Box, M.E. Muller, Ann. Math. Stat. **29**, 610–611 (1958)