Application of Approximate Bayesian Computation in Neutron Multiplicity Measurements: Providing Uncertainty Distributions and Correlations in all the Assay-Item Parameters

Tom Burr[1*], Stephen Croft[2], Daniella Henzlova[1], Brian Weaver[1], Andrea Favalli[1]
[1]Los Alamos National Laboratory
[2]Oak Ridge National Laboratory (retired)
LAUR21-23052

Uncertainty quantification (UQ) for safeguards can be approached from physical first principles ("bottom-up") or approached empirically by comparing measurements from different methods and/or laboratories ("top-down"). The two approaches can lead to different estimates of uncertainty, often with the bottom-up uncertainty estimate being smaller than the top-down uncertainty estimate; such a gap between the estimates is "dark uncertainty." Currently, one component of dark uncertainty in neutron multiplicity measurements arises due to item-specific effects, and another arises because uncertainty in nuclear data is ignored. One option to account for item-specific effects or for uncertainty in nuclear data is approximate Bayesian computation (ABC). This paper reviews ABC and illustrates how ABC can be applied in passive neutron multiplicity counting (PNMC) both with and without either item-specific effects or including the effects of errors in nuclear data. As a diagnostic, when an ABC-based interval for the true measurement error relative standard deviation (RSD) is constructed to contain approximately 95% of the true values, one can check whether the actual coverage is close to 95%. The performance of ABC illustrates potential advantages compared to current bottom-up and top-down approaches.

1. Introduction

All measurement data should be accompanied by a measurement uncertainty. As currently implemented, all measurements are characterized, calibrated, and applied using traditional analysis tools in a frequentist framework in which the measurand is regarded as a fixed and unknown constant while the measurements are regarded as random. In the passive neutron multiplicity counting (PNMC) bottom-up example used in this paper [1,2], point model multiplicity equations are solved (or inverted) to obtain a Pu mass estimate from three observed counting rates (singles (*S*), doubles (*D*), and triples (*T*)) assuming perfect (error free) nuclear data parameters, negligible item-specific effects, and error-free determined detector characteristics (efficiency and gate factors). A more modern uncertainty quantification (UQ) approach replaces the assumptions used in the traditional frequentist approach by a Bayesian framework in which both the measurand and the measurements (and even the forward model linking source terms such as item masses to measurements) are regarded as random. Bayesian methods are founded in probability theory, which is useful for constructing confidence in quantities and conclusions.

Historically, Bayesian techniques have been too computationally intensive to apply routinely, but this is no longer the case for a large class of practical problems [3-7]. Bayesian approaches can incorporate prior knowledge into the analysis in a natural way. Forward simulation linking physical forward model inputs to outputs is also intuitive and natural in the Bayesian paradigm to estimate a plausible posterior probability distribution of solutions. Bayesian methods, such as approximate Bayesian computation (ABC [1,2,4-7]), easily handle realistic probability density functions (pdfs), and so heavy-tailed and/or asymmetric distributions are readily accommodated. Forward modeling can create a distribution of plausible solutions as an alternative to algebraic inversion. This distribution of plausible solutions is being further developed both in likelihood-based and likelihood-free (ABC) Bayesian approaches for safeguards [1,2,4-7]. The term "approximate" in ABC means that an analytical likelihood (such as the normal, Poisson, or other familiar likelihood) is not available, but a forward model such as MCNP [8] is available.

It is common for the bottom-up approach to UQ to lead to a smaller estimate of errors relative standard deviation (RSDs) than the corresponding top-down approach; such a gap between the estimates is the so-called "dark uncertainty," and uncertainty in nuclear data used in the PNMC is currently not accounted for in bottom-up UQ and so currently contributes to dark uncertainty. Section 2 describes the PNMC. Section 3 describes ABC. Section 4 applies ABC to the PNMC with and without nuclear data uncertainty. Section 5 applies ABC to top-down UQ, including differences between operator and inspector measurements. Section 6 is a summary.

2. PNMC

The most common summary statistics used in PNMC to infer test item parameters are the so-called (*S*, *D*, *T)* values [8-14]. Assume that many ($N \approx 10^5$ or more) short-duration (approximately 10^{-5} sec) counting gates are used, one for each detected neutron; the counts in each of the *N* gates are the foreground (*f* in the notation below) counts. Each foreground gate also opens a random background gate following some time delay, so there are also *N* background gates (*b* in the notation below). Let F_j denote the number of neutrons detected in foreground gate *j* and let B_j denote the number of neutrons detected in background gate *j*. There are many gating schemes [9-15]; the scheme described here is a common one, in which the singles rate *S* is the rate of detected neutrons minus the rate of detected neutrons in the corresponding background measurement. The rate S can be calculated as $f_1 - b_1$, where $f_1 = \frac{1}{N}\sum_{j=1}^{N} F_j$ and $b_1 = \frac{1}{N}\sum_{j=1}^{N} B_j$ are the first moment (the mean) of the counts in the foreground and background gates. The doubles rate $D = S(f_1 - b_1)$, and the triples rate $T = S(f_2 - b_2 - 2\,b_1(f_1 - b_1))/2$ where $f_2 = \frac{1}{2N}\sum_{j=2}^{N} F_j\,(F_j - 1)$ and $b_2 = \frac{1}{2N}\sum_{j=2}^{N} B_j\,(B_j - 1)$ are the second order reduced factorial moments [9-15].

Measured *S*, *D*, and *T* values can be used to infer, for example, for point-like Pu mass standards, the spontaneous fission (SF) rate F_S, the ratio (denoted α) of the (α,n) reaction rate to the SF rate α, and sample multiplication M. Equations (1)-(3) below can be solved for F_S, α, M (using auxilliary data, usually by counting ^{252}Cf, to estimate sample/detector efficiency □) [9-15]. Another alternative is a version of time interval analysis (TIA) rather than Eqs. (1)-(3), where the estimated pdf of the time differences $X_i - X_{i-1}$between successive detected neutrons is used. Starting from the "list-mode" data consisting of each neutron detection time, other candidate summary statistics (instead of or in addition to *S*, *D*, *T* or $X_i - X_{i-1}$) might be considered for the goal of inferring item properties such as F_S . Section 2.1 describes the Hage-Cifarelli and Bohnel approach; Section 2.2 describes TIA.

2.1 Hage-Cifarelli and Bohnel approach

References [9, 11] showed that under "point model" assumptions, the expected values of *S*, *D*, and *T* (moments of the detected neutron distribution) denoted μ_S, μ_D, μ_T, respectively are:

$$\mu_S = F_S \varepsilon M \nu_{s1}(1+\alpha) \quad (1),$$

$$\mu_D = \frac{F_S \varepsilon^2 f_d M^2}{2}\left[\nu_{s2} + \left(\frac{M-1}{\nu_{i1}-1}\right)\nu_{s1}(1+\alpha)\nu_{i2}\right] \quad (2), \text{ and}$$

$$\mu_T = \frac{F_S \varepsilon^3 f_t M^3}{6}\left[\nu_{s3} + \left(\frac{M-1}{\nu_{i1}-1}\right)[3\nu_{s2}\nu_{i2} + \nu_{s1}(1+\alpha)\nu_{i3}] + 3\left(\frac{M-1}{\nu_{i1}-1}\right)^2 \nu_{s1}(1+\alpha)\nu_{i2}^2\right] (3).$$

The measured *S*, *D*, and *T* values can be used to infer, for example, for point-like Pu mass standards, the SF rate F_S, the ratio (denoted α) of the (α,n) reaction rate to the SF rate α, and sample multiplication M. The detector neutron efficiency ε and the doubles and triples gate fractions f_d and f_t can be estimated from measurements of ^{252}Cf reference neutron sources (that also obey the point model assumptions, with $\alpha = 0, M = 1$) and then efficiency is updated to the one of ^{240}Pu SF, the primary SF isotope in Pu items [1,10]. In addition, background measurements with no source items are made to adjust for background for both the Pu items and the ^{252}Cf items.

In the ^{252}Cf measurements, assume the only source of SF is ^{252}Cf; in the Pu item measurements, assume the only source of SF is Pu. In applying Equations (1-3) to the measurements of ^{252}Cf items, $M = 1$, $\alpha = 0$, both with zero error. The uncertainties in the nuclear data are described in [1,16], and then the efficiency ε can be estimated. Note that if the value of ε depends on the item, then it could be preferable (have smaller measurement error variance) to estimate ε separately for each item [13].

2.2 Time-interval analysis (TIA)

The main existing inverse methods use summary statistics such as *S*, *D*, *T* calculated from the observed histogram of counts in short randomly-triggered or neutron-triggered gates or both. Although TIA is rarely used in safeguards and other nonproliferation tasks, other applications of PNMC [12-14] provide an option to use TIA, using the pdf of the time lapse between detected neutrons as the summary statistic of the detected neutron times.

To apply TIA, a key derivation in [12-14] is Eq. (4) for the pdf of the time between detected neutron counts,

$$I_0(t)\Delta T = Term1 + Term2 \qquad (4),$$

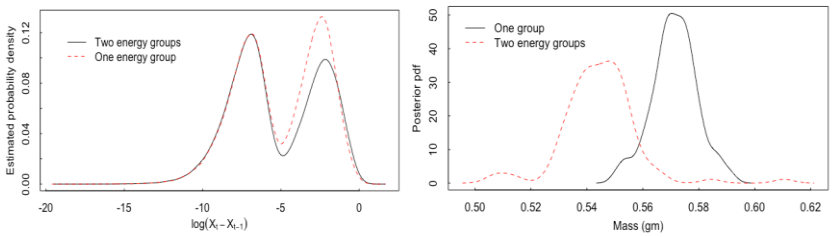
where $Term\ 1 = R_1 r_0 n_0 \Delta T$, and $Term\ 2 = \frac{F_S}{R_1}\sum_{n=2}^{N_{max}} e_n(\varepsilon)\left\{\sum_{k=1}^{n-1} k e^{-k\lambda T}\right\} b_0(T)\lambda\Delta T$, ΔT is a small time interval such that $I_0(t)\Delta T$ can be regarded as an approximation of the integral $\int_t^{t+\Delta T} I_0(t)dt$, $R_1 = \varepsilon q M \bar{\nu}_S F_S$ is the count rate, F_S is the SF rate, M is the sample multiplication, $q = 1 - p_{IF}$ where p_{IF} is the probability that a neutron induces fission (IF), and $\bar{\nu}_S$ is the average number of neutrons emitted per SF. The probability to detect *n* neutrons from a fission chain is $e_n = \sum_{\nu=n}^{\infty} P_\nu \binom{\nu}{n} \varepsilon^n (1-\varepsilon)^{\nu-n}$, where ε is the probability that a neutron is detected (the overall detection efficiency). The term r_0 is the probability that no neutrons are detected within a time *T* of the initial neutron trigger in a chain. The term b_0 is the probability that no neutrons are detected within a random time interval *T*. The term $n_0 = r_0 b_0$. The P_ν values are probability of emitting ν neutron from the fission and λ is the inverse of the die-away time. The point model assumes only one neutron energy group, and so also assumes there is only one probability of inducted fission. A two-energy model allows for differing neutron energies arising, for example due to having a large fraction of (α,n) neutrons compared to SF neutrons.

Fig. 1a is the pdf (on $\log_e$ = ln scale) for successive differences $X_i - X_{i-1}$ in neutron detection times for a simulated example that use either one or two neutron energy groups. Fig. 1b is the ABC-based posterior distribution for one real item based on using either one or two energy groups. The statistical programming language R [17] is used for all analyses; see Section 3.

Some neutrons arrive in "chains," in which the neutron birth times are essentially all the same (for the initiating neutron(s) and the neutrons from induced fissions). Therefore, some of the $X_i - X_{i-1}$ time differences tend to be small (as in the small peak in the estimated pdf at approximately 10^{-7} sec), arising mostly from the time to slow down and then detect both neutrons. Even if there is no induced fission, there can be multiple neutrons released in SF, so there are bursts of neutrons as seen by the pdf of $\ln(X_i - X_{i-1})$ in Fig.1a.

3. ABC

In any Bayesian approach, prior information regarding the magnitudes or relative magnitudes of any model parameter(s) can be provided. If the prior is "conjugate" for the likelihood, then the posterior is in the same likelihood family as the prior, in which case analytical methods are available to compute posterior prediction intervals for quantities of interest. So that a wide



(a) **(b)**

Fig. 1. (a) The pdf of log($X_i - X_{i-1}$)with one and two energy groups. (b) ABC-based estimate of the posterior pdf using one energy group and two energy groups for real item for which the Hage-Cifarelli approach gives an estimate of 0.56 to 0.58 g (the nominal value is 0.56 gm, and the α rate is approximately 6, which is quite large). The posterior is wider (and also shifted lower) when the forward model assumes two energy groups (approximately 2% RSD compared to 1% RSD for one energy group). Some of the dark uncertainty is exposed.

variety of priors and likelihoods can be accommodated, modern Bayesian methods do not rely on conjugate priors, but use numerical methods to obtain samples from approximate posterior distributions [3]. For numerical methods such as Markov Chain Monte Carlo (MCMC) [3], the user specifies a prior distribution (which need not be normal), for the parameters to be inferred, such as F_S, α, *M* and a likelihood for (*S*, *D*, *T*) (see [1]) or for $X_i - X_{i-1}$ ([1,2,4]) . In contrast, ABC does not require a likelihood for the data; and, as in any Bayesian approach, ABC accommodates constraints on variances through prior distributions [1,2,4-7].

ABC is ideally suited for extending PNMC to items that violate the point-model assumptions. The "output" of any Bayesian analysis is the posterior distribution for each model parameter, and so the output of ABC for data generated assuming that *S*, *D*, *T* have mean values given in Equations 1-3, and a multivariate normal covariance that is estimated from data [1,2]. Or, if TIA is used, then Eq. (4) can be used as the likelihood to estimate the posterior probabilities for F_S, α, *M*. ABC is also well suited for comparing the posterior width for different choices of summary statistics. Posterior width can be used to make a data-driven decision whether an item has been counted for a long enough time.

A well-calibrated Bayesian approach is an approach that satisfies several requirements. Requirement (1) is that in repeated applications of ABC, approximately 95% of the middle 95% of the posterior distribution for F_S (for example) should contain the true value. That is, the actual coverage should be closely approximated by the nominal coverage. The nominal coverage is obtained by using the quantiles of the estimated posterior. For example, the 0.005 and 0.995 quantiles define the lower and upper limits of a 99% probability interval. Requirement (2) is that the true mean squared error (MSE) of the ABC-based estimate of F_S should be closely approximated by the variance of the ABC-based posterior distribution of F_S. For data such as that simulated in Figure 1, using $\varepsilon = 0.01$, both requirements (1) and (2) are met [2].

Inference using ABC can be summarized as follows. For *i* in 1, 2, ..., N, do these 3 steps: (1) Sample θ from the prior,(2) Simulate data x' from the model $x'' \sim P(\theta|x)$, (3) Denote the real data as x. If distance $d(S(x'), S(X)) \leq \varepsilon$ then accept θ from $f_{posterior}(\theta|x)$. Experience with ABC suggests that the ABC approximation to $f_{posterior}(\theta|x)$ improves if step (3) is modified to include a weighting factor, so that trial values of θ simulated from $f_{prior}(\theta)$ that lead to very small distance $d(S(x'), S(x))$ are more heavily weighted in the estimated posterior [6,7]. In step (2), the model can be analytical or any forward model. The model in this paper has input parameters □={α, M, F_S}, has output data $x' \sim P(x|\theta)$, and there is corresponding real data x_{obs}. In the Hage-Cifarelli approach, the model is Eqs. (1-3), plus specification of the distribution of *S*, *D*, *T* given μ_S, μ_D, μ_T (multivariate normal; see [1]).

Eqs. (1-3) specify how to generate synthetic *S*, *D*, *T* data, and Eq. (4) specifies how to generate synthetic $X_i - X_{i-1}$ data, but ABC does not require one to explicitly specify a likelihood. So, the true likelihood used to generate the data need not be known to the user. Synthetic data is generated from the model for many trial values of □=$\{\alpha, M, F_s\}$ (with or without including the effects of errors in nuclear data [1-3]), and trial □ values are accepted as contributing to the estimated posterior distribution for □| x_{obs} if the distance $d(x_{obs}, x(\theta))$ between x_{obs} and $x(\theta)$ is reasonably small. Alternatively, for most applications, it is necessary to reduce the dimension of x_{obs} to a small set of summary statistics □ and accept trial values of □ if $d\big(S(x_{obs}), S(x(\theta))\big) < \varepsilon$, where ε is a user-chosen threshold (not the efficiency parameter in NMC). Here, for example, x_{obs}= S, D, T (adjusted for the background data) or is a histogram-based summary of the pdf of $\ln(X_i - X_{i-1})$.

Recall that, because trial values of □ are accepted if $d\big(S(x_{obs}), S(x(\theta))\big) < \varepsilon$, an approximation error to the posterior distribution arises that several ABC options attempt to mitigate. Additionally, recall that such options weight the accepted □ values by the actual distance $d\big(S(x_{obs}), S(x(\theta))\big)$ (abctools package in R [17]). In practice, the true posterior distribution is not known, so a method such as that in [5] using the calibration checks described above, is needed in order to choose an effective value of the threshold □.

ABC can evaluate candidate summary statistics. The estimated distribution of plausible nuclear material mass values (or, for example, α, M, F_s values) can be assessed by checking at least the two calibration requirements listed previously. ABC is used here for two main reasons:

A) many real items are non point-like and violate point-model assumptions to varying extents (such as a single energy-independent efficiency and a single multiplication), but ABC can be used with realistic forward models with or without an explicit likelihood [2] with and without errors in nuclear data, and
B) ABC allows comparison of the performance of candidate summary statistics such as *S*, *D*, *T*; the histogram of detected numbers of neutrons in gates, the pdf of $X_i - X_{i-1}$; and the pdf of $X_i - X_{i-2}$.

4. Item-specific Bias in NMC

ABC was implemented using either of 4 cases: : (1) point model, (2) point model but two energy groups, (3) point model but dead time effects resulting in approximately 15% loss of counts, and (4) point model but two energy groups and dead time losses of approximately 15%. The 2-group model assumes two groups of neutron energies, resulting in two probabilities of induced fission.

Choice A for summary statistics was {*S*, *D*, *T*}. Choice B for summary statistics was 13 bins of a histogram-based summary of the pdf for log($X_i - X_{i-1}$). Recall that ABC s users to experiment with candidate summary statistics. ABC results for choice A will be summarized here. A separate training set of 10^4 ABC simulations was computed for each of the 4 cases. The two main calibration checks for ABC indicated good calibration, so the ABC-based posterior is a good approximation. Figure 2 plots the ABC-based posterior pdf for a simulated test case that had a true SF rate of 258.2 for each of 4 training sets, one for each of the 4 cases. Note that 258.2/475.3 = 0.54 gms, where 475.3 is the number of fissions per gm per second, the specific SF rate of Pu. As expected, the true value of 0.54 is within the middle of the posterior pdf for case 1 (with a mean of 0.55 and standard deviation of 0.01). However, the true value of 0.54 is not within the middle of the posterior pdf for cases 2, 3, or 4, (with means 0.59, 0.57, 0.59 and standard deviations 0.01, 0.01, and 0.005, respectively) which indicates non-negligible departure from the point model. ABC results for choice B were more dependent on how many bins were used in the histogram-based summary of the pdf for log($X_i - X_{i-1}$). Interestingly, the item-specific biases shifted from being positive when using choice A (0.59, 0.57, and 0.59 are all a few multiple of the posterior standard deviation of 0.01 above the true value of 0.54) to negative for

choice B, also by a few multiples of the posterior standard deviation, which was also approximately 0.01 in all four cases. ABC performed will in the two main calibration checks so the posterior pdfs are good approximation.

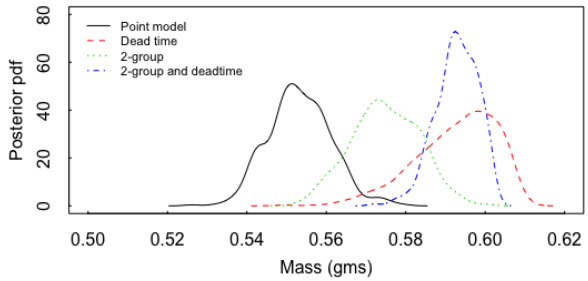


Fig. 2. The ABC-based posterior pdf for a simulated test case with true mass 0.54 gms for each of 4 training sets, one for each of the 4 cases using *S*, *D*, *T* (choice A summary statistics). The simulated test case obeyed the point model (case 1); the other 3 cases had point model violations.

Another check that ABC is well calibrated is to match the simulated test case with the corresponding training data (as in Fig. 2 with case 1, the point model for which the posterior includes the true mass of 0.54 gms). In this example in Fig. 3, when the test case was simulated using the same model (case 1, 2, 3, or 4 model) as the 10^4 simulated training observations, then the resulting posteriors all easily include the true value of 0.54 gms, all with mean values of approximately 0.55 and standard deviations of approximately 0.01.

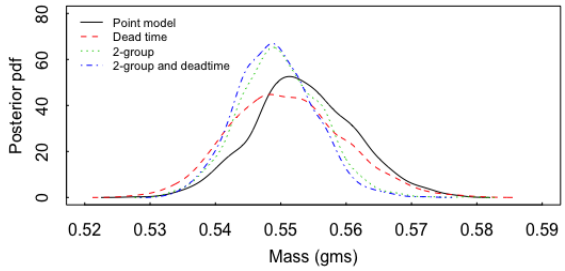


Fig. 3. The ABC-based posterior pdf for a simulated test case that had a true mass of 0.54 gms for each of 4 training sets, one for each of the 4 cases. The simulated test case obeyed the point model (case 1), and the other 3 cases had point model violations, but the simulated training observations used by ABC matched those of the corresponding test case, so this is a calibration check on ABC using *S*, *D*, *T* (choice A summary statistics).

4. Application of ABC to top-down UQ

Section 4 illustrates ABC in three top-down analyses.

4.1 Normal data with unknown mean and known variance

In this example the scalar-valued data has an unknown mean value μ, and conditional on μ it is assumed that X has a normal distribution with mean μ and variance σ_x^2, denoted $X|\mu \sim N(\mu, \sigma_x^2)$. The unknown mean μ is assigned a prior $\mu \sim N(0, \sigma_{prior}^2)$, where σ_{prior}^2 can be large if little is known about μ prior to collecting data. Because the mean of μ is zero, X could be difference data, such as the difference between a measured and declared value. The normal prior for μ is conjugate for the normal likelihood for X, so the resulting posterior is also normal,

$\mu|x \sim N(\mu_{posterior}, \sigma^2_{posterior})$ where $\mu_{posterior} = \tau\sigma_x^2$ and $\sigma^2_{posterior} = \tau\sigma_x^2$ where $\tau = \frac{\sigma^2_{prior}}{\sigma^2_{prior}+\sigma_x^2}$ and $\mu \sim N(0, \sigma^2_{prior})$ and $X|\mu \sim N(\mu, \sigma_x^2)$. Because the posterior is known exactly [3], it is possible to assess how well ABC works in this example. Fig. 4 plots the ABC-based posterior using $\varepsilon = 0.001, 0.01, 0.1$ using one observation of *X* as the data (and summary startstic). The true posterior is also shown.

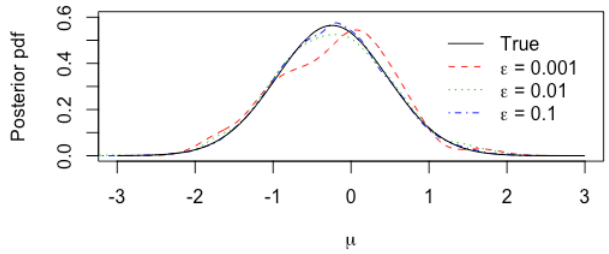


Fig. 4. The posterior pdf for Example 4.1 with known variance and normal data.

Regarding the two calibration requirements, the true posterior standard deviation is $\sqrt{\tau} = 0.71$, and the estimated RMSE is 0.72 using $\varepsilon = 0.001,\ 0.01,$ or 0.1. And the actual 90%, 95%, and 99% coverages are 0.99, 0.95, 0.90 for $\varepsilon = 0.01$ or 0.1 and 0.97, 0.93, 0.89 for $\varepsilon = 0.001$. Figure 5 plots the actual PI coverage probability versus the nominal PI coverage probability (using quantiles of the posterior) for $\varepsilon = 0.01$. Clearly, ABC is well calibrated.

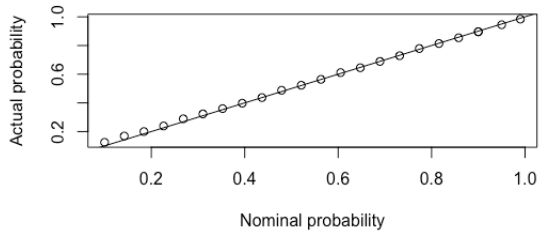


Fig. 5. ABC calibration check for Example 4.1 with known variance and normal data.

4.2 Normal data with unknown mean and unknown variance

Example 4.2 is the same as Example 4.1, but the mean μ and variance σ_x^2 are unknown [3]. One way to specify the conjugate prior specifies an inverse gamma prior Invgamma(α, β), for σ_x^2, $\frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{\frac{-\beta}{x}}$ (with x denoting the possible values of σ_x^2). Conditional on σ_x^2, $\mu|\sigma_x^2 \sim N(0, \nu\sigma^2)$ and $X|\mu \sim N(\mu, \sigma_x^2)$. The parameter ν can be related to an effective degrees of freedom in the prior information. The resulting posterior is also normal for μ, and is gamma$(\alpha_{posterior}, \beta_{posterior})$ for σ_x^2, where $\alpha_{posterior} = \alpha + n$, $\beta_{posterior} = \beta + \frac{1}{2}\sum_{i=1}^{n}(x_i - \bar{x})^2$ and the data are $x_1, x_2, \ldots, x_n$ As in Example 4.1, the posterior is known exactly, so it is possible to assess how well ABC works. Figure 6 plots the ABC-based posterior of σ_x using $\varepsilon = 0.001, 0.01, 0.1$ using the sample mean and variance of n =5 observations of *X* as the data (and summary statistic). The true posterior is also shown. Regarding the two calibration

requirements, for $\varepsilon = 0.01$, the predicted RMSE is 0.29 and the observed average standard deviation is 0.31, and the actual PI coverages are 0.99, 0.96, 0.91 (nominal are 0.99, 0.95, 0.90).

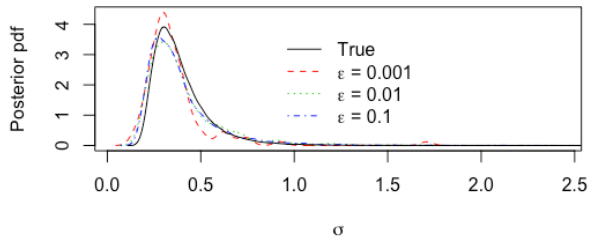


Fig. 6. The ABC-based posterior of σ_x.

4.3 Operator – Inspector Data

The error model used in the IAEA's statistical methodologies for safeguards sets the stage for applying one-way analysis of variance (ANOVA) with random effects [18-21]. The error model accounts for variation within and between groups, where a group is, for example, a calibration or inspection period. A typical model for multiplicative errors for the inspector (*I*) is

$$I_{jk} = \mu_k(1 + S_{Ij} + R_{Ijk}), \tag{5}$$

where I_{jk} is the inspector's measured value of item k, μ_k is the true but unknown value of item k, $S_{IJ} \sim N(0, \delta_{SI}^2)$ is a short-term systematic error in inspection period j, $R_{Ijk} \sim N(0, \delta_{RI}^2)$ is a random error of item k, and the variance of I_{jk} is $\sigma_I^2 \approx \mu_k^2(\delta_\mu^2 + \delta_{SI}^2 + \delta_{RI}^2)$, where the item variability δ_μ^2 is the relative variance of the of random variable μ_k (the true item values). The δ symbol denotes a relative standard deviation and the σ symbol denotes an absolute standard deviation. Figure 7 plots n = 10 (*O*-*I*)/*O* simulated values for each of g = 5 groups with parameters $\delta_{SO} = 0.001, \delta_{RO\,=0.001,}, \delta_{SI} = 0.027, \delta_{RI} = 0.027, \delta_\mu = 0.003$.

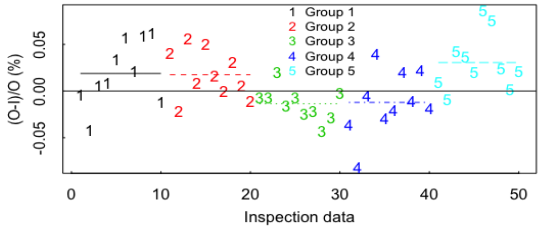


Fig. 7. Ten simulated values of (*O*-*I*)/*O* for each of g = 5 groups.

One-way ANOVA based on paired data allows us to estimate the measurement error variances of plant operators and inspectors. ANOVA requires the data to fall in groups, so that within-group and between-groups variances can be defined and estimated. In this example, the groups are the inspection periods. The basis of a Grubbs-based estimator [3-6,22] as applied to data assumed be generated according to Eq. (5) in order to estimate δ_{RI}^2 (all five terms $\delta_\mu^2, \delta_{SI}^2, \delta_{RI}^2, \delta_{SO}^2, \delta_{RO}^2$ can be estimated but for brevity here, the example considers estimating δ_{RI}^2) is that the covariance between operator and inspector equals σ_μ^2, while the variance σ_I^2 conditional on the value of S_{Ij} is given by $\sigma_I^2 = \mu_i^2(\delta_\mu^2\delta_{RI}^2 + \delta_\mu^2(1 + S)^2 + \delta_{RI}^2)$, which has an expected value over inspection

periods of $\mu_i^2(\delta_\mu^2\delta_{RI}^2 + \delta_\mu^2(1+\delta_{SI}^2)+\delta_{RI}^2)$. Therefore, the expected between-group and within-group sums of squares involve both δ_{SI}^2 and δ_{RI}^2. Provided δ_{SI}, δ_{RI}, and δ_μ are each less than approximately 0.20, the approximation $\sigma_I^2 \approx \mu_i^2(\delta_\mu^2 + \delta_{RI}^2)$ is adequate, so that the sample covariance between operator and inspector measurements can be subtracted from the sample variance of the operator measurements to estimate δ_{RI}^2 (and similarly for estimating δ_{RO}^2). That is, within a single inspection period (group) (lower-case *i* (*o*) for numerical values of *I* (*O*)), an effective estimate of δ_{RI}^2 is (with $\bar{o}$ used to estimate the average true value μ)

$$\hat{\delta}_{RI}^2 = \frac{1}{\bar{o}^2(n-1)}\{\sum_{j=1}^n (i_j-\bar{\imath})^2 - \sum_{j=1}^n (o_j-\bar{o})(i_j-\bar{\imath})\} \quad (6).$$

The original Grubbs' estimate [22-24] is for additive error models. The ABC framework [1, 4,23,24] makes Grubbs' type estimation straightforward for either additive or multiplicative models. The five summary statistics used in this application of ABC for *n* (*O*,*I*) pairs in each of *g* groups are the average over *g* groups of the estimates $\{\hat{\delta}_\mu^2 = \frac{\sum_{j=1}^n (o_j-\bar{o})(i_j-\bar{\imath})}{(n-1)\bar{o}^2}, \hat{\delta}_{RI}^2, \hat{\delta}_{RO}^2\}$ and $\{\frac{\sum_{i=1}^g (\bar{I}_j-\bar{I})^2}{(g-1)\bar{o}^2} - \frac{\hat{\delta}_\mu^2}{n} - \frac{\hat{\delta}_{RI}^2}{n}, \frac{\sum_{i=1}^g (o_i-\bar{o})^2}{(g-1)\bar{o}^2} - \frac{\hat{\delta}_\mu^2}{n} - \frac{\hat{\delta}_{RO}^2}{n}\}$. Note that $\sigma_I^2 = \mu_i^2(\delta_\mu^2 + \delta_{RI}^2 + \delta_{SI}^2)$ so the variance of the between group means (equal sample size with *n* observations per group) is $\sigma_{\text{Between},I}^2 = \mu_i^2(\frac{\delta_\mu^2}{n} + \frac{\delta_{RI}^2}{n} + \delta_{SI}^2)$, which is the basis for the fourth and fifth summary statistics.

As an example, Figure 8 plots the ABC-based posterior pdf for δ_{RI} using $\varepsilon = 0.001, 0.02, 0.1$. Regarding the two calibration requirements, for $\varepsilon = 0.005, 0.01$ or 0.1, the predicted RMSE is 0.027 and the observed average standard deviation is 0.028, and the actual PI coverages are 0.99, 0.96, 0.91 (nominal are 0.99, 0.95, 0.90). However, if $\varepsilon = 0.001$ then outliers impact the posterior mean leading to poor calibration.

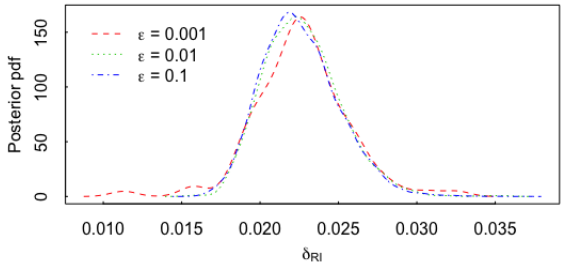


Fig. 8. The posterior pdf for Grubbs'-like estimation of δ_{RI}**.** The posterior mean is 0.023 and the true value of δ_{RI} is 0.027. The threshold $\varepsilon = 0.001$ does not lead to well-calibrated ABC.

5. Summary

This paper reviewed material from recent papers that have implemented ABC for safeguards measurent UQ, both bottom-up and top-down. The new emphasis in this paper is a strategy to assess whether ABC is well calibrated. In this context, well calibrated means that at least two key properties are met by the ABC-based estimate of posterior pdfs for parameters such as an item's SF rate F_S or mass. Property 1 is whether the posterior width as measured by the posterior standard deviation provides a good estimate of the root mean squared error of the estimate (and the estimate is the posterior mean). Property 2 is whether quantiles of the posterior such as those based on 90%, 95%, and 99% PIs actually lead to PI coverages of approximately 90%, 95%, and 99%, respectively. Because of the simulation-based nature of ABC, is it straightforward to assess

whether properties 1 and 2 are approximately met. In all examples presented, ABC is well calibrated, provided the acceptance threshold ε is chosen appropriately.

Acknowledgments. The work presented in this paper was funded by the National Nuclear Security Administration of the Department of Energy, Office of International Nuclear Safeguards.

References

[1] Burr, T., Croft, S., Henzlova, D., Weaver, B., Favalli, A., Bayesian Bottom-up Uncertainty Quantification in Neutron Multiplicity Measurements: Providing Uncertainty Distributions and Correlations in all the Assay-Item Parameters, Proc. 61st Annual INMM Meeting, 2020
[2] Burr, T., Croft, S., Henzlova, D., Weaver, B., Favalli, A., Comprehensive Bayesian Uncertainty Quantification for Neutron Correlation Counting, LAUR20-27531, 2020. Submitted Feb 2021
[3] Carlin, B., John, B., Stern, H., Rubin, D., Bayesian Data Analysis, Chapman and Hall, 1995.
[4] Burr, T., Krieger, T., Norman, C., Approximate Bayesian Computation Applied to Metrology for Nuclear Safeguards ESARDA bulletin 57, 50-59, 2018.
[5] Burr, T., Skurikhin, A., Selecting Summary Statistics in Approximate Bayesian Computation for Calibrating Stochastic Models, BioMed Research International, vol. 2013, Article ID 210646, 10 pages, 2013. doi:10.1155/2013/210646, 2013.
[6] Blum, M., Nunes, M., Prangle, D. Sisson, S., A Review of Dimension Reduction Methods in Approximate Bayesian Computation, Statistical Science 28(2), 189–208, 2013.
[7] Nunes, M. Prangle, D., abctools: an R package for Tuning Approximate Bayesian Computation Analyses, The R Journal 7(2), 189–205, 2016.
[8] Werner, C.J.(editor), "MCNP Users Manual - Code Version 6.2", Los Alamos National Laboratory, report LA-UR-17-29981 (2017)
[9] Cifarelli,D.M., Hage, W.,Models for a Three-Parameter Analysis o Neutron Signal Correlation Measurements For Fissile Material Assay, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, Volume 251(3) 1986,
[10] Ensslin, N., Harcker, W., Krick, M., Langner, D., Pickrell M., Stewart J., Application Guide to Neutron Multiplicity Counting, LA-13422-M, 1998, permalink.lanl.gov/object/tr?what=info:lanl-repo/lareport/LA-13422-M.
[11] Favalli, A., Croft S., Santi P., Point Model Equations for Neutron Correlation Counting: Extension of Bohnel's Equations to any Order, Nuclear Instruments and Methods in Physics Research A 795, 370-375, 2015.
[12] Bruggeman,M., Baeten, P., De Boeck, W., Carchon, R., Neutron Coincidence Counting Based on Time Interval Analysis with One- and Two- Dimensional Rossi-alpha Distributions: an Application for Passive Neutron Waste Assay, Nuclear Instr. Methods, A382 511-518, 1996.
[13] Walston S., Candy J., Chambers D., Chandrasekaran H., Snyderman, N. Real-time Characterization of Special Nuclear Material LLNL-TR-676944, 2015.
[14] Prasad, M., Snyderman, N., Verbeke, J., Wurtz, R., Time Interval Distributions and the Rossi Correlation Function, Nuclear Science and Engineering 174, 1, 1-29, 2013.
[15] Prasad M., Snyderman N., Walston S., Neutron Time Interval Distributions with Background Neutrons, Nuclear Science and Engineering 186, 277-292, 2017.
[16] Croft S., Favalli A., Henzlova D., Burr T., Weaver B., Nuclear Data Uncertainties in Neutron Multiplicity Counting, LAUR19-29477, 2019.
[17] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org, 2017.
[18] Zhao, K., et al., International Target Values 2010 for Measurement Uncertainties for Safeguarding Nuclear Safeguards, STR 368, IAEA, 2010.
[19] Burr, T., Sampson, T, Vo, D., Statistical Evaluation of FRAM γ-ray Isotopic Analysis Data, Applied Radiation and Isotopes 62**,** 931-940, 2005.
[20] Burr, T., Hemphill, G., Multi-Component Radiation Measurement Error Models, Applied Radiation and Isotope**s** 64(3), 379-385, 2006.
[21] Burr T., Croft,S., Jarman K., Nicholson A., Norman C., Walsh S., Improved Uncertainty Quantification in NonDestructive Assay for Nonproliferation, Chemometrics, 159, 164-173, 2016.
[22] Walsh, T., Burr, T., Martin, K., Discussion of the IAEA Error Approach to Producing Variance Estimates for use in Material Balance Evaluation and the International Target Values, and Comparison to Metrological Definitions of Precision, Journal of Nuclear Materials Management, 45(2),4-14, 2017.
[23] Bonner, E., Burr, T., Guzzardo, T., Krieger, T., Norman, C., Zhao, K., Beddingfield, D., Geist, W., Laughter, M., Lee, T., Ensuring the Effectiveness of Safeguards through Comprehensive Uncertainty Quantification, Journal of Nuclear Materials Management 44(2), 53-61, 2016.
[24] Bonner, E., Burr, T., Krieger, T., Martin, K., Norman, C., Comprehensive Uncertainty Quantification in Nuclear Safeguards, Science and Technology of Nuclear Installations, 1-16, 10.1155/2017/2679243, 2017