Approximate Bayesian Computation applied to Metrology for Nuclear Safeguards

Journal of Physics: Conference Series

Tom Burr¹, Thomas Krieger², Claude Norman¹

Production Editor, Journal of Physics: Conference Series,
IOP Publishing, Dirac House, Temple Back, Bristol BS1 6BE, UK
¹International Atomic Energy Agency, Vienna Austria
²Forschungszentrum Jülich GmbH, Inst. Energy and Climate Research, Jülich, Germany

E-mail: t.burr@iaea.org

Abstract. Approximate Bayesian Computation (ABC) is an inference option if a likelihood
for the measurement data is not available, but a forward model is available that outputs
predicted observables for specified input parameters. This paper applies ABC to the
enrichment meter principle (EMP) method, which is used in nuclear safeguards and does not
have an explicit likelihood. A key aspect of metrology is uncertainty quantification (UQ),
approached from physical first principles (“bottom-up”) or approached empirically by
comparing measurements from different methods and/or laboratories (“top-down”). Although
ABC is not commonly used in metrology, the EMP example illustrates advantages in ABC
compared to current bottom-up approaches. ABC is also shown to be useful in top-down UQ.

As a diagnostic, in bottom-up and top-down applications of ABC, the actual coverages of
probability intervals are compared to the true coverages. If an ABC-based interval for a
parameter is constructed to contain approximately 95% of the true parameter values, then it is
important to verify that the actual coverage is close to 95%. It is shown that one advantage of
ABC compared to other Bayesian approaches is its apparent robustness to miss-specifying the
likelihood while maintaining good agreement between nominal and actual coverage.

1. Metrology for Nuclear Safeguards

Nuclear safeguards aim to verify that nuclear materials are used exclusively for peaceful purposes.
To ensure that States are honouring their safeguards obligations, measurements of nuclear material
inventories and flows are needed. Statistical analyses used to support conclusions require uncertainty
quantification (UQ), usually by estimating the relative standard deviation (RSD) in random and
systematic errors associated with each measurement method [1-2].

To monitor for possible data falsification by the operator that could mask nuclear material
diversion, paired (operator, inspector) data are assessed. These paired data are declarations usually
based on measurements by the operator, often using destructive assay, and measurements by the
inspector, often using non-destructive assay (NDA). Paired (operator, inspector) data is used for top-
down UQ, applying analysis of variance (ANOVA), to estimate RSDs. Bottom-up UQ propagates
errors in all key assay steps of to predict the RSD in the estimated nuclear material mass; this error

¹ t.burr@iaea.org
propagation is similar to that used in the guide to expression of uncertainty in measurements [GUM, 3]. It is common for bottom-up RSD estimates to be smaller than those from top-down UQ [2].

This paper is organized as follows. Sections 2 and 3 describe top-down and bottom-up UQ, respectively. Section 4 applies approximate Bayesian computation (ABC [4]) to top-down and bottom-up UQ for safeguards for NDA using the enrichment meter principle (EMP [5]).

2. Top-down UQ applied to paired (operator, inspector) data

An effective measurement error model must account for variation within and between groups, where a group is, for example, a calibration or inspection period. A typical model for relative errors for the inspector (I) and similarly for the operator (O) is

\[ I_{ij} = \mu_{ij} (1 + S_{ij} + R_{ij}) \]

where \( I_{ij} \) is the inspector’s measured value of item \( j \) in group \( i \), \( \mu_{ij} \) is the true but unknown value of item \( j \) from group \( i \), \( S_{ij} \sim N\left(0, \delta_{si}^2\right) \) is a short-term systematic error in group \( i \), \( R_{ij} \sim N\left(0, \delta_{ri}^2\right) \) is a random error of item \( j \) from group \( i \). Fig. 1 plots 10 simulated values in each of 3 groups of \( d = (O - I)/O \) (which by error variance propagation has approximate variance \( \delta_{so}^2 + \delta_{ro}^2 + \delta_{si}^2 + \delta_{ri}^2 = \delta_d^2 \), and see Section 5.1 for more information regarding Fig. 1). The measurement error model sets the stage for applying ANOVA with random effects [1,2,6].

3. Bottom-up UQ

NDA uses calibration and/or modelling to infer nuclear material mass using detected radiation such as neutron and gamma emissions. Two issues in UQ for NDA are: (1) the detector is brought to the facility where ambient conditions can vary over time, and the items are often heterogeneous, so dark uncertainty [7] can be large, as is evident whenever bottom-up UQ predicts smaller RSD than is observed in top-down UQ, and (2) NDA is often used when test items differ substantially from calibration items; therefore, the concept of item-specific bias is important (see Section 5). Bottom-up UQ is often approached by using the GUM’s measurement equation, expressed as

\[ Y = f(X_1, X_2, ..., X_N) \]

for measurand \( Y \) and inputs \( X_1, X_2, ..., X_N \). The GUM applies the delta method to Eq. (2) to propagate error variances in the \( X_i \) to estimate the standard deviation in \( Y \). The input quantities can include measured count rates and estimates of calibration parameters or other measurands, such as measured values in steps an assay method. Eq. (2) implies that \( Y \) is random, so the GUM implicitly adopts a Bayesian viewpoint (Section 4) without explicitly stating a prior distribution for \( Y \) [8].

4. ABC

Bayesian ANOVA such as could be applied to data generated from Eq. (1) has been studied [6,8,9]; however, Bayesian ANOVA using ABC has not been well studied. In a traditional Bayesian approach, the user specifies prior information regarding the magnitudes or relative magnitudes of \( \delta_{RI}^2 \) and \( \delta_{SI}^2 \) and a likelihood that describes the probability of obtaining data values for given parameter values (such as the Gaussian probability density function for specified mean and variance). ABC does not require a likelihood for the data (but this section provides clarification regarding the need for a likelihood in this NDA context), and, as in any Bayesian approach, ABC accommodates constraints on variances through prior distributions [8,9].
The “output” of any Bayesian analysis is the posterior distribution for each model parameter, and so the output of ABC for data generated from Eq. (1) is an estimate of the posterior distributions of $\delta^2_{RI}$ and $\delta^2_{SI}$. Inference using standard ABC can be briefly summarized in 3 steps: For $i$ in 1, 2, ..., $N$

1. Sample $\theta$ from the prior, $\theta \sim f_{prior}(\theta)$.
2. Simulate data $y'$ from the model $y' \sim P(y | \theta)$
3. Denote the real data as $y$. If the distance $\text{dist}(S(y'), S(y)) \leq \epsilon$, accept $\theta$ as an observation from $f_{posterior}(\theta | y)$ (weights are allowed).

In ABC, the model has input parameters $\theta$ and outputs data $y(\theta)$ and there is corresponding real data $y_{obs}$. For example, the model could be Eq. (1), which specifies how to generate synthetic $I$ (or $O$) data, and does require a likelihood; however, 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 $\theta$, and trial $\theta$ values are accepted as contributing to the estimated posterior distribution for $\theta$ if $\text{dist}(S(y_{obs}), S(y(\theta)))$ between $y_{obs}$ and $y(\theta)$ is reasonably small. Alternatively, for most applications, it is necessary to reduce the dimension of $y_{obs}$ to a small set of summary statistics $S$ and accept trial values of $\theta$ if $\text{dist}(S(y_{obs}), S(y(\theta))) < \epsilon$, where $\epsilon$ is a user-chosen threshold. Here, for example, $y_{obs} = d = (O - I)/O$ data in each inspection group, and $(y_{obs})$ includes within and between groups sums of squares. Specifically, the ANOVA-based estimator of $\delta^2_{RI}$ is $\hat{\delta}^2_{RI} = \sum_{i=1}^{g} \sum_{j=1}^{n_i} (d_{ij} - \bar{d}_i)^2 / (n - g)$ and the estimate of $\delta^2_{SI}$ is $\hat{\delta}^2_{SI} = \sum_{j=1}^{g} (\bar{d}_j - \bar{d})^2 / (g - 1) - \hat{\delta}^2_{RI}/n$, and $\hat{\delta}^2_{RI}$ and $\hat{\delta}^2_{SI}$ are good summary statistics. Recall that because trial values of $\theta$ are accepted if $\text{dist}(S(y_{obs}), S(y(\theta))) < \epsilon$, an approximation error to the posterior distribution arises that several ABC options attempt to mitigate by weighting the accepted $\theta$ values by the actual distance $\text{dist}(S(y_{obs}), S(y(\theta)))$ (abctools in R [10]).
5. EMP Example

Enrichment (the ratio $^{235}\text{U}/\text{U}$) can be measured using the full-energy 185.7 keV gamma-rays emitted from $^{235}\text{U}$ by applying the EMP [5]. The EMP assumes that the detector field of view into each item is identical to that in the calibration items (the "infinite thickness" assumption), that the item must be homogeneous with respect to both the $^{235}\text{U}$ enrichment and chemical composition, and that the container attenuation of gamma-rays is that same as or similar to that in the calibration items so that empirical correction factors have modest impact and are reasonably effective. If these three assumptions are met, the known physics implies that the enrichment of $^{235}\text{U}$ in the U is directly proportional to the count rate of the 185.7 keV gamma-rays emitted from the item.

The known nominal enrichment in each of several standards can be fit to observed counts in a few energy channels near the 185.7 keV energy as the "peak" region and to the counts in a few energy channels somewhere below and above the 185.7 keV energy but outside the peak area to estimate background (two-region EMP method), expressed as

$$ Y = \beta_i \text{(item)} N + R_e $$

where $N$ is the net peak region count rate, and the calibration parameter $\beta_i \text{(item)}$ varies across items because of item-specific effects such as container thickness.

Figure 2 is an example low-resolution (NaI detector) gamma spectrum near the 185.6 keV. The gross count and the two background ROI counts can be combined into one net count, resulting in one predictor as in Eq. (3), which can be fit using least squares regression. There will be measurement errors in the gross and background count rates and there will often be correction factors applied, for example, to adjust test item container thickness to calibration item container thickness. There is literature regarding errors in predictors and whether to fit $Y$ as a function of $N$ (reverse calibration) or to fit $N$ as a function of $Y$ and invert to solve for $Y$ (inverse calibration) [2].

5.1 ABC applied to the EMP

ABC applied to the EMP can be implemented in the following seven steps. (1) Estimate the average regression coefficient in Eq. (6) using available real calibration data, typically consisting of approximately 3 to 5 $(Y,N)$ pairs. The real calibration data used here are $Y = 0.355, 0.80, 2.175, 0.305, 5.0 \ (^{235}\text{U} \text{ enrichments of 5 standards})$ and the corresponding $N = 0.062, 0.139, 0.37, 0.575, 0.866 \text{ net count rates}$. (2) Use the estimate $\hat{\beta}_i$ from (1) to generate many ($S = 10^5$ or more) synthetic calibration runs using $Y = \beta_i \text{(item)} N + R_e$ to generate synthetic sets of 5 paired $(Y,N)$ values, with run $i$ producing the estimate $\hat{\beta}_i$. This example generated the $\beta_i \text{(item)}$ values randomly and uniformly from 0.85 to 0.95. (3) Specify a prior distribution for the true enrichment $\mu_Y$. If little is known about the true enrichment values, then, for example, specify a uniform prior ranging from the lowest possible true enrichment to the highest possible true enrichment. This example used wide a uniform distribution from 0.355 to 5.0, which avoids extrapolating outside the range of the true enrichments. (4) Specify a background count rate $\mu_B$ (this example used $\mu_B = 0.05$) and use the estimated regression coefficient $\hat{\beta}_i$ from the regression equation $N = \alpha_i \text{(item)} Y + R_N$ to generate a net count rate. $\mu_N$ that corresponds to a $\mu_Y$ sampled from its prior distribution. This example used an RSD in $Y$ of 0.1% and in $R_N$ of 5%. (5) Specify a count time (this example used 600 seconds) $t$, simulate $B \sim \text{Poisson}(\mu_B t)$, $G \sim \text{Poisson}(\mu_G t)$, and compute a net count rate (assuming the same number of energy channels for the peak and background ROIs) $N = G/t - B/t$ (6) Repeat (4) and (5) many ($10^5$ or more) times to construct a large collection of simulated true enrichments $\mu_Y$ and corresponding net
count rates \( N \), which is an effective summary statistic. (7) For each simulated test case, simulate a value of \( \mu_y \) from its prior, use steps (4) and (5) to generate \( N_{\text{test}} \), and compute the distance \( d(N_{\text{test}}, N_i) = |N_{\text{test}} - N_i| \) from \( N_{\text{test}} \) to each of the \( i = 1, 2, \ldots, 10^5 \) realizations from step (6), and accept those \( \mu_y \) generated in step (6) that correspond to \(|N_{\text{test}} - N_i| \leq \varepsilon\) as observations from the posterior \( \mu_y | N \) (which in this case is somewhat complicated to specify analytically) weighting inversely by the distance \(|N_{\text{test}} - N_i| \) if desired. Linear regression was not used in this ABC implementation for predicting \( \mu_y \) for each simulated test value of \( N \), although it could have been, and note that regression is used in step (2) to generate the \( 10^5 \) pairs of \(( \mu_y, N) \) in the training data for ABC.

To assess ABC performance, two criteria can be used: the estimated standard deviation of the posterior should be in good agreement with the observed standard deviation across test items, and the nominal probability interval coverage should be in good agreement with the actual coverage. The data plotted in Fig. 1 were generated using the steps just given to apply ABC for both operator and inspector data, assuming for simplicity that both used the EMP and both recalibrated at the beginning of periods 1, 2, and 3. The estimated standard deviation of \( d_g \) from top-down data such as that in Fig. 1 (also using ABC as outlined in Section 4) is 0.11, which is very close to that predicted from the bottom-up ABC based (0.12 as explained in the next paragraph) posterior standard deviations for \( O \) and \( I \). Recall from Section 4 that the usual ANOVA-based estimator of \( \delta_{rd}^2 \) is

\[
\hat{\delta}_{rd}^2 = \frac{\sum_{j=1}^{g} \sum_{j=1}^{n} (d_{ij} - \bar{d})^2}{(n - g)}
\]

and the usual estimate of \( \delta_{rd}^2 \) is

\[
\hat{\delta}_{rd}^2 = \frac{\sum_{j=1}^{g} (\bar{d}_{j} - \bar{d})^2}{(g - 1) - \hat{\sigma}_R^2/n}
\]

The quantities, \( \hat{\delta}_{rd}^2 \) and \( \hat{\delta}_{rd}^2 \) were used to implement ABC for the top-down analysis of data such as that in Fig. 1.

The 0.12 bottom-up prediction for the standard deviation of \( d \) (denoted as \( \delta_d \) here, because \( d = (O - I)/O \) is a relative difference) is illustrated by plotting the posterior for \( O \) for a particular \( N \) value in Fig. 3, which has a total (random plus systematic) RSD of 0.08 (from the 7-step procedure). Because this example assumes \( O \) and \( I \) made the same type of EMP measurements, the bottom-up prediction of \( \delta_d \) is \( \sqrt{(0.08^2 + 0.08^2)} = 0.11 \). The 0.12 top-down estimate of \( \delta_d \) (see Fig. 4, using data such as the data in Fig. 1) is the standard deviation of the ABC-based posterior distribution for \( \delta_d \) from top-down UQ, with \( g = 3 \) groups and \( n = 10 \) paired measurements per group (as in Fig. 1). The 0.12 estimate has a standard deviation of 0.03, and an approximate 95% probability interval for \( \delta_d \) is 0.07 to 0.21. In this example, bottom-up UQ using ABC agrees well with corresponding top-down UQ using ABC that used simulated \( O \) and \( I \) values as in Fig. 1. The 99%, 95%, and 90% probability intervals contained approximately 99%, 95%, and 90%, respectively of the true values of \( \mu_y \).

A normal distribution is not always a good approximation for the actual distribution of \( d = (O - I)/O \) values. So, regarding robustness of ABC in top-down UQ, it has been found that the actual coverages are essentially the same (to within simulation uncertainty) as the nominal coverages, at 90%, 95%, and 99% probabilities, for a normal distribution and all of the non-normal distributions investigated (uniform, gamma, lognormal, beta, t, and generalized lamba with thick or thin tails) for the distribution of the random error term \( R_e \) in Eq. (3). Regarding robustness of ABC in the bottom-up context, a key aspect of ABC is the ease with which different forward models linking model parameters (such as the true RSDs in Eq. (1)) to model output and corresponding summary statistics.
For example, the Poisson model used in the ABC implementation for the EMP can be easily replaced with an overdispersed Poisson model if exploratory analysis of real data suggests overdispersion.

6. Discussion and Summary

ABC was used for both bottom-up and top-down RSD estimation in simulated EMP data (using a small calibration set of 5 real EMP data pairs). ABC provided robust estimates of the posteriors for model parameters (the RSD values), so bottom-up RSD estimates could be compared to top-down estimates while accounting for parameter uncertainty (as defined by the width of the posterior).

7. References

[1] Walsh, S., Burr, T., Martin, K., The IAEA error approach to variance estimation 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.
[2] 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.
[3] JCGM 104:2009, Evaluation of measurement data- an introduction to the “Guide to the Expression of Uncertainty in Measurement,” 2009.
[4] Joyce, P., Marjoram, P., Approximately sufficient statistics and Bayesian computation, Statistical Applications in Genetics and Molecular Biology, 7(1), article 26, 2008.
[5] ASTM C1514, Standard test method for measurement of $^{235}$U fraction using the enrichment meter principle, 2008.
[6] Miller, R., Beyond ANOVA: basics of applied statistics, Chapman & Hall, 1998.
[7] Thompson, M. and Ellison, S., Dark uncertainty, Accreditation and Quality Assurance 16, 483–487, 2011.
[8] Carlin, B., John, B., Stern, H., Rubin, D., Bayesian Data Analysis (1st ed), Chapman and Happ, 1995.
[9] Blum, M., Nunes, M., Prangle, D. Sisson, S., A comparative review of dimension reduction methods in approximate Bayesian computation. Statistical Science 28(2), 189–208, 2013.
[10] R: A language and environment for statistical Computing, R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria, 2008.