The Journal of Undergraduate Research

Volume 15 *The Journal of Undergraduate Research*

Article 4

2017

Approximate Bayesian Computation in Forensic Science

Jessie H. Hendricks
South Dakota State University

Recommended Citation

Hendricks, Jessie H. (2017) "Approximate Bayesian Computation in Forensic Science," *The Journal of Undergraduate Research*: Vol. 15 , Article 4.
Available at: https://openprairie.sdstate.edu/jur/vol15/iss1/4

Approximate Bayesian Computation in Forensic Science

Cover Page Footnote

This research was supported by the SDSU Scholarly Excellence Fund, the Department of Mathematics and Statistics, and the Jerome J. Lohr College of Engineering.

Approximate Bayesian Computation in Forensic Science

Author: Jessie Hendricks
Faculty Sponsors: Danica M. Ommen, MS; Cedric Neumann, PhD; and Christopher P. Saunders, PhD
Department: Mathematics and Statistics

ABSTRACT

Forensic evidence is often an important factor in criminal investigations. Analyzing evidence in an objective way involves the use of statistics. However, many evidence types (i.e., glass fragments, fingerprints, shoe impressions) are very complex. This makes the use of statistical methods, such as model selection in Bayesian inference, extremely difficult.

Approximate Bayesian Computation is an algorithmic method in Bayesian analysis that can be used for model selection. It is especially useful because it can be used to assign a Bayes Factor without the need to directly evaluate the exact likelihood function - a difficult task for complex data. Several criticisms of ABC (specifically when used for model choice) can be found in the literature, yet the potential benefits of the method warrant study of the performance under controlled situations. This paper explores the use of ABC in forensic science through use of a glass example. The ABC approximation of the Bayes Factor provided the correct direction of support in the majority of known cases in the glass example. However, the approximations did not agree with the weight of support given by analytical Bayes Factors. Exploration of the inconsistent portion of results uncovered an issue with the method used to choose a threshold value in ABC. A novel method, based on a relationship between the ABC Bayes Factor and the Receiver Operating Characteristic curve, was developed to address this issue. This method produced more stable results than the traditional ABC methodology.

INTRODUCTION

The analysis of forensic evidence is often an important factor in forensic investigations. Tools in Bayesian statistics can be used to provide a quantitative value of the evidence that can lend support to the prosecution or the defense. Bayesian modeling allows models to be built under the prosecution and defense hypotheses to compare the probabilities of observing the collected evidence under each model.

Modeling involves building a description of a population using parametric probability density functions. Knowledge about the value of the parameters is gained through the observation of sample data sets from the population [5]. Models will be denoted with M_i, where i indexes the model number for the case when multiple models are being considered.

Several terms and definitions in Bayesian statistics from *Bayesian Essentials with R* [5] will be discussed next for use throughout the remainder of the paper. A complete list of symbols and definitions is included in Appendix A. The prior density, the likelihood function, the marginal density, the posterior density, and the Bayes Factor are several fundamental tools used in Bayesian inference. Note that Bayesian probability is described as a subjective belief, which is different than the objective, standard definition of a Frequentist probability [5].

The **prior density** on the value of parameters of a population, $\pi(\theta)$, represents a belief about the value of these parameters before observing the sample data set, D_n. When multiple models are being considered, $\pi(\theta|M_i)$ will be used to distinguish between the priors used for different models. The **likelihood function**, $f(D_n|\theta)$, gives the probability of observing the sample data set D_n from a population as a function of the set of parameters, θ. The **marginal density**, $\pi(D_n|M_i) = \int f(D_n|\theta) \cdot \pi(\theta|M_i)\, d\theta$, is the density of D_n under model M_i, which is not dependent upon θ. The **posterior**

density on the value of parameters of a population, $\pi(\theta|D_n)$, combines the information from the likelihood function and the prior density, and represents the knowledge about the parameters after additional information has been gathered from the observation of D_n. The **posterior odds** is the ratio of two posterior probabilities. The posterior odds, given by $\frac{\pi(M_1|D_n)}{\pi(M_2|D_n)}$, compares the probability of one model, M_1, given the observed data D_n to the probability of a competing model, M_2, given the same observed data.

Bayes' Theorem provides a way to calculate the posterior probability indirectly by using the prior density and the likelihood function.

Theorem 1 (Bayes' Theorem [6])**.** *Let P be a general probability operator. Let* A_1, A_2, A_3, ... A_n *be a set of mutually exclusive events and B be an event such that* $P(B) \neq 0$. *Then*

$$P(A_i|B) = \frac{P(B|A_i)\,P(A_i)}{\sum_{i=1}^{n} P(B|A_i)\,P(A_i)}$$

for any A_i, *where* $i = 1,2,3,...n$.

See the example below for how the posterior odds can be decomposed using Bayes' Theorem in both the numerator and denominator.

$$\begin{aligned}\frac{\pi(M_1|D_n)}{\pi(M_2|D_n)} &= \frac{\left(\dfrac{\pi(D_n|M_1)\cdot\pi(M_1)}{\pi(D_n|M_1)\cdot\pi(M_1)+\pi(D_n|M_2)\cdot\pi(M_2)}\right)}{\left(\dfrac{\pi(D_n|M_2)\cdot\pi(M_1)}{\pi(D_n|M_1)\cdot\pi(M_1)+\pi(D_n|M_2)\cdot\pi(M_2)}\right)} \\ &= \frac{\left(\dfrac{\pi(D_n|M_1)\cdot\pi(M_1)}{\pi(D_n)}\right)}{\left(\dfrac{\pi(D_n|M_2)\cdot\pi(M_2)}{\pi(D_n)}\right)} \\ &= \frac{\pi(D_n|M_1)}{\pi(D_n|M_2)}\cdot\frac{\pi(M_1)}{\pi(M_2)}\end{aligned}$$

The result of the decomposed posterior odds is a ratio called the Bayes Factor, $\frac{\pi(D_n|M_1)}{\pi(D_n|M_2)}$, multiplied by the **prior odds**, $\frac{\pi(M_1)}{\pi(M_2)}$. The **Bayes Factor**, given below in Equations 1 and 2, is defined as the ratio of the probability of the data under each model framework.

$$BF = \frac{\pi(M_1|D_n)}{\pi(M_2|D_n)}\cdot\frac{\pi(M_2)}{\pi(M_1)} \tag{1}$$

$$= \frac{\pi(D_n|M_1)}{\pi(D_n|M_2)} = \frac{\int f(D_n|\theta)\cdot\pi(\theta|M_1)\,d\theta}{\int f(D_n|\theta)\cdot\pi(\theta|M_2)\,d\theta}. \tag{2}$$

A Bayes Factor greater than one gives support to M_1 because the probability of observing the data under M_1 is greater than the probability under M_2. Conversely, a Bayes Factor less than one gives support to M_2 because the probability of observing the data under M_2 is greater than the probability under M_1. The magnitude of the Bayes Factor corresponds to the strength of the support. Large positive values provide more support to M_1 than values close to one. Values close to zero provide strong support for M_2.

The Bayes Factor can be used in forensic science to quantify the probative value of forensic evidence [2]. This is useful when inferring the source of trace evidence (recovered evidence materials whose original source is unknown) in forensic cases because the trace may appear to correspond to several possible reference control samples (evidence materials taken from a known source) by chance. The prosecution believes the trace evidence originated from a specific source (a known, possible source for the origin of the trace evidence), while the defense believes the trace evidence originated from an unknown alternative source. The value of the Bayes Factor allows one conclusion to be drawn, which can lend support to the prosecution or the defense.

However, data collected for most forensic evidence types are very complex, so the likelihood function in Equation 2 is too difficult to evaluate. In addition, the posterior density on the parameters in Equation 1 cannot be directly evaluated, so the Bayes Factor cannot be calculated. Approximate Bayesian Computation offers a possible solution.

The ABC algorithm

The Approximate Bayesian Computation method bypasses the need for the actual, complex likelihood function. Instead, ABC uses a simpler likelihood function $f_m(D_n^*|\theta_i)$ along with the original sample data set, D_n. Robert et al. [10] explain the ABC algorithm as follows. First, model indexes are treated as parameters and sampled from $\pi(M_i = m)$, a Bernoulli distribution with the probability of success (sampling M_2) equal to 1 minus the probability of sampling M_1. Next, sets of parameters are sampled from their prior densities, $\pi(\theta|M_i)$. The sampled parameters are used along with the simpler chosen likelihood function to generate new data sets, denoted D_n^*, of the same dimension as D_n. A summary statistic is computed for D_n, denoted $\eta(D_n)$, and for each of the sampled data sets, denoted $\eta(D_n^*)$. The distance between $\eta(D_n)$ and $\eta(D_n^*)$ is calculated for each generated data set. Parameters corresponding to the D_n^*'s which are satisfactorily close (by the use of a threshold, t) to the original data set D_n are retained, and used to compose a sample from the approximate posterior density, denoted $\pi_t(\theta|D_n)$. The Approximate Bayesian Computation method results in $\pi_t(\theta|D_n)$, an approximation of the posterior density $\pi(\theta|D_n)$ on the value of parameters of a population, given the data set D_n [10].

When ABC is used for model selection, the posterior odds can be approximated by the frequency of retained parameter sets, θ_i, under each model. The approximation of the posterior odds is divided by the prior odds $\dfrac{\pi(M_1)}{\pi(M_2)}$, resulting in the ABC approximation of the Bayes Factor, BF_{abc}. The ABC Bayes factor can then be used to provide support to one of the two models (e.g., M_1 if $BF_{abc} > 1$ or M_2 if $BF_{abc} < 1$).

Algorithm 1: ABC model selection algorithm [10]

```
Initialize i=0;
while $i < N_{acc}$ do
    Sample a model index $m$ from the model prior $\pi(M_1 = m)$;
    Generate the parameters $\theta_i$ from the prior density given the model index $\pi(\theta|m)$;
    Generate a sample data set, $D_n^*$, from the approximate likelihood $f_m(D_n^*|\theta_i)$;
    Compute $\Delta(\eta(D_n), \eta(D_n^*))$, the distance between $\eta(D_n)$ and $\eta(D_n^*)$;
    Compare $\Delta(\eta(D_n), \eta(D_n^*))$ to a pre-determined value for $t$;
    if $\Delta(\eta(D_n), \eta(D_n^*)) \leq t$ then
        Accept $\theta_i$ and corresponding model index $m$;
        i = i+1;
    end
end
```

The formal ABC model selection method, described in Algorithm 1 above, results in a pre-determined number N_{acc} of accepted parameter sets. A less computationally intensive modified version is described in Algorithm 2. This method repeats the sampling processes a fixed number of times, N_{sim}, but does not guarantee a specific number of accepted parameter sets.

Algorithm 2: Modified ABC model selection algorithm [5]

```
for $i$ = $1$ to $N_{sim}$ do
    Sample a model index $m$ from the model prior $\pi(M_1 = m)$;
    Generate the parameters $\theta_i$ from the prior density given the model index $\pi(\theta|m)$;
    Generate a sample data set, $D_n^*$ from the approximate likelihood $f_m(D_n^*|\theta_i)$;
    Compute $\Delta(\eta(D_n), \eta(D_n^*))$, the distance between $\eta(D_n)$ and $\eta(D_n^*)$;
end
Use a pre-determined quantile of the set of $\Delta(\eta(D_n), \eta(D_n^*))$'s to choose a value for the threshold $t$;
Compare $\Delta(\eta(D_n), \eta(D_n^*))$ to the value chosen for $t$;
Accept $\theta_i$'s and corresponding model index $m$, that generated $D_n^*$'s with $\Delta(\eta(D_n), \eta(D_n^*)) \leq t$.
```

At the completion of both algorithms, $\frac{\pi_t(M_1|D_n)}{\pi_t(M_2|D_n)}$ is estimated by the ratio of the frequency of retained θ_i's under each model (M_1 and M_2), which can be written as $\frac{\sum_{i=1}^{N} \mathbb{I}_{m^i=1} \mathbb{I}_{\Delta(\eta(D_n),\eta(D_n^{*,i}))\leq t}}{\sum_{i=1}^{N} \mathbb{I}_{m^i=2} \mathbb{I}_{\Delta(\eta(D_n),\eta(D_n^{*,i}))\leq t}}$, where $\mathbb{I}_B(\cdot)$ is the indicator function of B. Finally, the Bayes Factor, $BF = \frac{\pi(D_n|M_1)}{\pi(D_n|M_2)}$, is approximated by $BF_{abc} = \frac{\pi_t(M_1|D_n)}{\pi_t(M_2|D_n)} \cdot \frac{\pi(M_2)}{\pi(M_1)}$.

Literature review

Although ABC has the ability to bypass the need for evaluating a complex likelihood function, it has several limitations when used for model choice. As the value of the threshold t approaches zero, the ABC approximation of the posterior density, $\pi_t \theta|D_n)$, converges to $\pi(\theta|D_n)$, as stated by Robert et al. [10]. However, they also showed that convergence only holds when the summary statistic, $\eta(D_n)$, used in the algorithm is a sufficient statistic. In the case where $\eta(D_n)$ is an insufficient statistic, $\pi_t(\theta|D_n)$ converges to $\pi(\theta|\eta(D_n))$, the posterior density given the summary statistic [10].

A sufficient statistic contains all the information contained in the sample data set to estimate the value of a parameter of the population, meaning that no information is lost when the data is condensed into the form of the sufficient statistic [10]. Sufficient statistics are available for distributions in any exponential family [10]. One example of a sufficient statistic is the sample mean, when used to estimate the population mean for a normally distributed population [5].

When a sufficient statistic does exist, a factorization theorem allows the likelihood $f_i(D_n|\theta)$ to be factored, resulting in the likelihood for the summary statistic of the data $f_i(\eta(D_n)|\theta)$ multiplied by a factor $g_i(D_n)$, which can be considered the measure of error between the two likelihoods [10]. However, there is no standard method to calculate $g_i(D_n)$. As pointed out by [10], this causes an issue when ABC is used for model selection to approximate the Bayes Factor because there is no way of comparison between the error of approximation across models, which can be seen in Equation 3.

$$\begin{aligned}
BF &= \frac{\int g_1(D_n) \cdot f_1(\eta(D_n)|\theta) \cdot \pi(\theta|M_1)\, d\theta}{\int g_2(D_n) \cdot f_2(\eta(D_n)|\theta) \cdot \pi(\theta|M_2)\, d\theta} \\
&= \frac{g_1(D_n)}{g_2(D_n)} \cdot \frac{\int f_1(\eta(D_n)|\theta) \cdot \pi(\theta|M_1)\, d\theta}{\int f_2(\eta(D_n)|\theta) \cdot \pi(\theta|M_2)\, d\theta} \\
&= \frac{g_1(D_n)}{g_2(D_n)} \cdot BF_{abc}
\end{aligned} \tag{3}$$

Because of this, two major errors are possible: the algorithm supports the wrong model (direction of support), or the algorithm supports the correct model but with the wrong magnitude (strength of support). The occurrence of either or both of these errors would cause the ABC Bayes Factor approximation to represent an incorrect value of the evidence. Only in a very small number of known cases when $\frac{g_1(D_n)}{g_2(D_n)} = 1$, does the ABC Bayes Factor converge to the actual Bayes Factor [10]. Despite these criticisms from Robert, the benefit of using simpler likelihood functions made possible by ABC may allow for the quantification of the value of forensic evidence that is not otherwise possible. The potential of ABC warrants study of the algorithm's behavior under controlled example environments.

ABC appears several more times in the literature. In "Reliable ABC model choice via random forests," [9] Pudlo et al. discuss a method for use in high dimension settings which uses random forest techniques as the distance metric for use in the ABC algorithm. The description of random forest techniques is outside the scope of this paper, but can be reviewed in [9]. Although Pudlo et al.'s paper uses an example in population genetics, this method would be especially helpful in forensic science for fingerprint evidence because there is currently no known likelihood structure for this high dimensional evidence type.

ABC has been successfully used in other disciplines. Lombaert et al. [4] used ABC to study the path of invasion of the Asian ladybird *Harmonia axyridis*. The use of ABC in this situation allowed more exhaustive tests than had been possible in previous studies, while using a greater number of invasive site samples. The results obtained using ABC were consistent with results of previous studies. In addition, new detail about invasion routes was gained through use of ABC.

METHODOLOGY

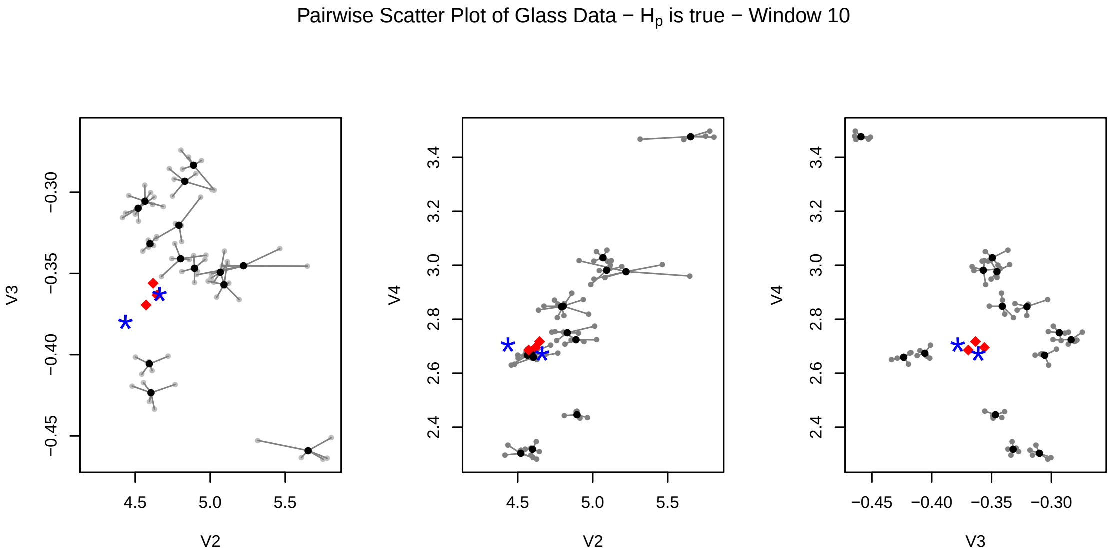
In forensic investigations, the prosecution and the defense often have different theories on the origin of trace evidence. The evidence is assessed to determine the strength at which it favors the prosecution or the defense. Some terminology and notation in forensic science from Ommen [7] will be used to discuss the example. The specific source problem is concerned with the question of whether trace evidence e_u originated from a known specific source, or instead from an unknown source in a population of alternative sources. Collected evidence from the specific source is denoted e_s. Observations characterizing the alternative source population is denoted e_a. The prosecution asserts that e_u originated from the specific source. The prosecution hypothesis is denoted H_p. The defense asserts that e_u originated from an unknown alternative source. The defense hypothesis is denoted H_d.

H_p: The trace evidence e_u came from the specific source.

H_d: The trace evidence e_u did not come from the specific source.

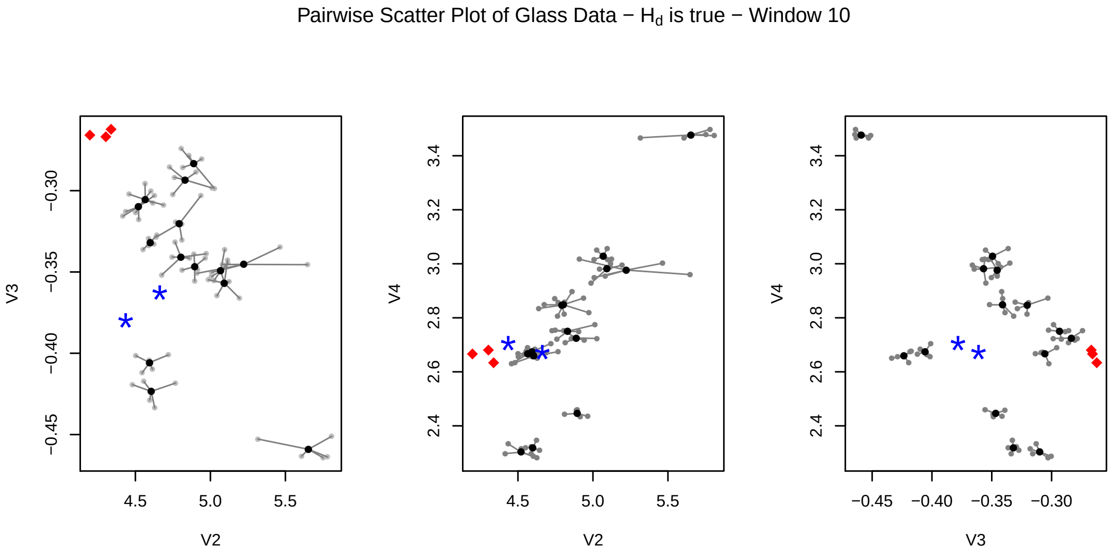
In this paper, the case is considered that two glass fragments (e_u) have been collected from a suspect's clothing, and are compared by measurements of elemental composition to a set of three fragments (e_s) from a broken window at the scene of a crime, and the relevant alternative population of windows (e_a).

A data set of measurements taken on the elemental composition of the glass fragments was used to set up this example. Three statistics, $V_2 = log\left(\frac{Ca}{K}\right)$, $V_3 = log\left(\frac{Ca}{Si}\right)$, and $V_4 = log\left(\frac{Ca}{Fe}\right)$ were used to compose e_u, e_s, and e_a. The statistic V_2 gives the ratio between the amounts of calcium (Ca) and potassium (K). The statistic V_3 represents the ratio between the amounts of calcium and silicon (Si). The statistic V_4 represents the ratio between the amounts of calcium and iron (Fe). The complete data set is fully described by Aitken and Lucy [1]. The subset of the elemental glass composition data used in this example is given in Appendix B.



Pairwise Scatter Plot of Glass Data − H_p is true − Window 10

Figure 1: The blue stars represent the trace evidence e_u, taken from window 10. The red diamonds represent the specific source evidence e_s, also taken from window 10. In this scenario, H_p is true. The black dots illustrate the mean of each window in the alternative source population, while the grey dots represent measurements on individual fragments in the alternative source population e_a.



Pairwise Scatter Plot of Glass Data − H_d is true − Window 10

Figure 2: The blue stars represent the trace evidence e_u, taken from window 10. The red diamonds represent the specific source evidence e_s, taken from window 5. In this scenario, H_d is true. The black dots illustrate the mean of each window in the alternative source population, while the grey dots represent measurements on individual fragments in the alternative source population, e_a.

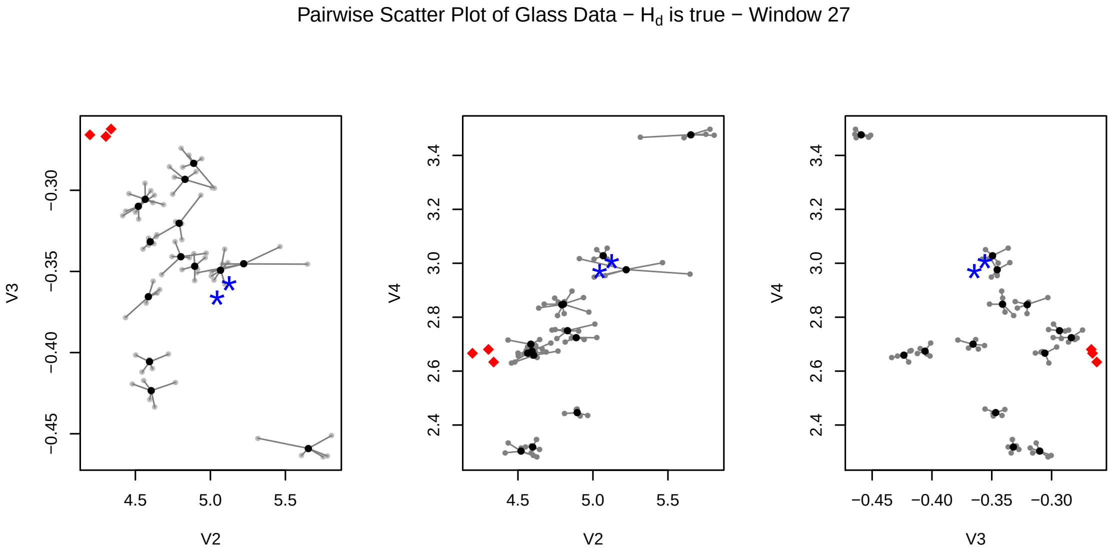

Pairwise Scatter Plot of Glass Data − H_d is true − Window 27

Figure 4: The blue stars represent the trace evidence e_u, taken from window 27. The red diamonds represent the specific source evidence e_s, taken from window 5. In this scenario, H_d is true. The black dots illustrate the mean of each window in the alternative source population, while the grey dots represent measurements on individual fragments in the alternative source population, e_a.

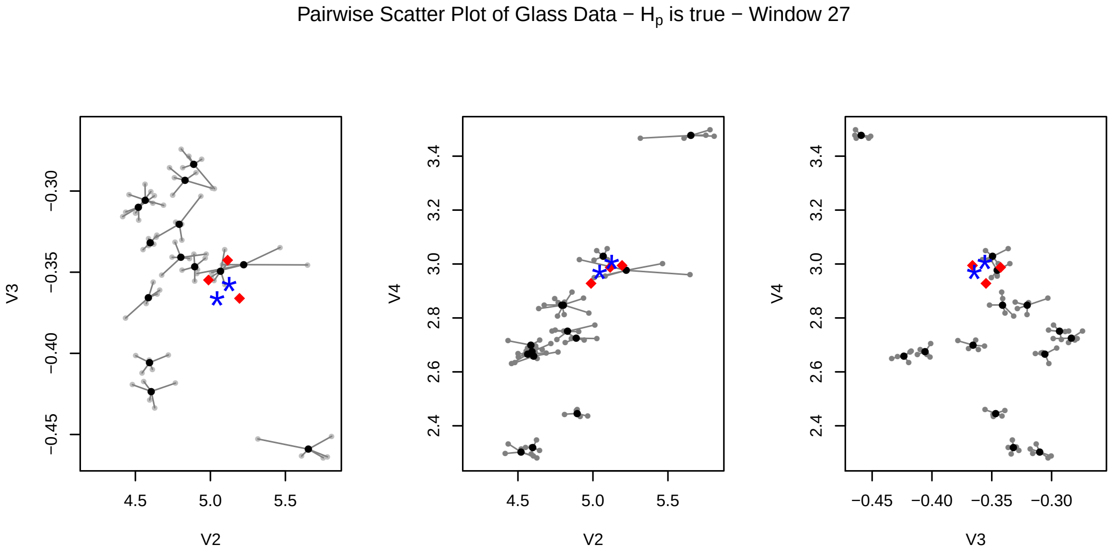

Pairwise Scatter Plot of Glass Data − H_p is true − Window 27

Figure 3: The blue stars represent the trace evidence e_u, taken from window 27. The red diamonds represent the specific source evidence e_s, also taken from window 27. In this scenario, H_p is true. The black dots illustrate the mean of each window in the alternative source population, while the grey dots represent measurements on individual fragments in the alternative source population, e_a.

Two scenarios were set up to observe the behavior of ABC in each case. In the first scenario, e_u and e_s were chosen from the same window, making the prosecution hypothesis H_p true. The ABC approximation of the Bayes Factor should be greater than one, supporting the prosecution. In the second scenario, e_u and e_s were chosen from different windows, making the defense hypothesis H_d true. The ABC approximation of the Bayes Factor should be less than one, supporting the defense. In both scenarios, measurements on fragments from the remaining windows in the data set were used to compose e_a. Each scenario was set up using two different window choices (window 10 and window 27 from the data) for e_s in order to observe the behavior of ABC under a variety of evidence.

Figures 1-4 contain pairwise scatter plots of the statistics, V_2, V_3, and V_4. Evidence from the first scenario, when H_p is true, is illustrated in Figures 1 and 3. Evidence from the second scenario, when H_d is true, is illustrated in Figures 2 and 4.

Algorithm 2 was implemented in R programming statistical software and simulations for this example were run to get an approximate Bayes Factor for each of several situations. When H_p is sampled in the ABC algorithm, the sampling model used for the vector of measurements on the glass fragments follows a multivariate normal model. When H_d is sampled in the ABC algorithm, the sampling model used for the vector of measurements on the glass fragments follows a simple random effects model, defined by Ommen [7]. Typical non-informative priors were used for both models [5]. For a detailed description of these models and priors, reference Ommen [7].

This example was used to observe the behavior of ABC in forensic science under a controlled setting. Each scenario was run using several different values for model priors. The results of each scenario were compared to a Monte Carlo (MC) Standard Mean approximation of the Bayes Factor [7]. The Standard Mean method uses Monte Carlo integration and approximate samples from the posterior density of the parameters through Gibbs sampling algorithms to approximate the Bayes Factor. For a full derivation and justification of the Standard Mean approximation, reference Ommen [7].

Algorithm 3: Glass Simulations

```
Initialize i=0;
for $e_u$ from {window 10, window 27} do
    for {$H_p$, $H_d$} is true do
        for i in 1:10 do
            ABC with $\pi(H_d) = 0.25$;
            ABC with $\pi(H_d) = 0.50$;
            ABC with $\pi(H_d) = 0.75$;
            MC Standard Mean method;
            i=i+1;
        end
        i=0.
    end
end
```

Ten simulations were completed for each of a total of 16 cases, with eight variations under each of the two main scenarios: H_p is true (e_s taken from the same window as e_u), and H_d is true (e_s taken from window 5). The simulations were completed according to Algorithm 3, above.

RESULTS AND DISCUSSION

Results from the glass example using traditional ABC

The Bayes Factor approximations made by the Monte Carlo Standard Mean method produced stable results under each scenario, as expected. The ABC approximations of the Bayes Factor were unstable in comparison to the Standard Mean approximations. In most cases, the ABC approximation provided less strength of support (a less extreme Bayes Factor) than the Standard Mean approximation. The two approximations agreed in the direction of support in most cases; however in the scenario where H_d was true and the measurements composing e_u were taken from window 10, the ABC approximation provided support to the incorrect proposition, H_p. Results of the glass example are illustrated in Figures 5 and 6. The numerical results are given in Appendix C in Tables 2-5.

For the scenario when H_p was true and both e_u and e_s originated from window 10, the ABC approximations of the Bayes Factor were smaller in magnitude in comparison to the Standard Mean, but favored the correct model.

In the scenario when H_d was true and only e_u originated from window 10 (e_s was taken from window 5), the ABC estimates did not favor the correct model. One possibility for the inconsistency can be seen in Figure 2. Several values of the statistics used for comparison of the glass fragments in ABC are very similar between e_u and e_s. By random selection of the windows, a window that was similar to the trace evidence fragments was chosen for the specific source window to set up this scenario.

For the scenario when H_p was true and both e_u and e_s originated from window 27, most ABC estimates of the Bayes Factor were smaller in magnitude in comparison to the Standard Mean, and all favored the correct model.

When H_d was true and only e_u originated from window 10 (e_s taken from window 5), the ABC estimates favored the correct model, but were not as close to zero as the Standard Mean estimates.

The density plots in Figures 7-8 of the distance metrics calculated in the ABC algorithm, $\Delta(\eta(D_n), \eta(D_n^*))$, separated by the sampled model, H_p and H_d, illustrate a possible explanation of the instability observed in the ABC Bayes Factor approximations. The distributions of $\Delta(\eta(D_n), \eta(D_n^*))$ shift slightly between completions of the algorithm due to slight differences in the set of generated sample data sets. The ratio of the left tail areas (contained by the vertical line representing the threshold t) in the distributions of $\Delta(\eta(D_n), \eta(D_n^*))$ for each sampled model (H_p and H_d) represents the approximation of the posterior odds, or the Bayes Factor when equal model priors are used ($\pi(H_p) = \pi(H_d) = 0.50$) as in Figures 7 and 8. The slight shifts in the distributions cause the ratio of areas to change, resulting in the unstable results of the ABC Bayes Factor approximation between completions of the algorithm.

A similar effect is also illustrated in Figures 9a and 9b. As the value of the quantile used for threshold choice in ABC decreases, the approximate Bayes Factors become more extreme, providing more strength of support. However, as the quantile approaches zero, the Bayes Factor approximation becomes very unstable, instead of stabilizing as originally expected.

As the quantile (and threshold t) approaches zero, the ratio of the accepted data (left tail of each density) generated under each of the models varies in magnitude and direction. Further study is warranted to investigate alternative methods of choosing a value for the threshold in order to get a more stable approximation of the Bayes Factor.

Other sources of error in the example include using an insufficient statistic for the summary statistic in the implemented ABC algorithm, and the selected number of repetitions in the algorithm, N_{sim}. One hundred thousand repetitions (N_{sim}) were used in the algorithm due to the amount of available computational power. An increase in the number of repetitions may have increased the accuracy of Bayes Factor approximation.

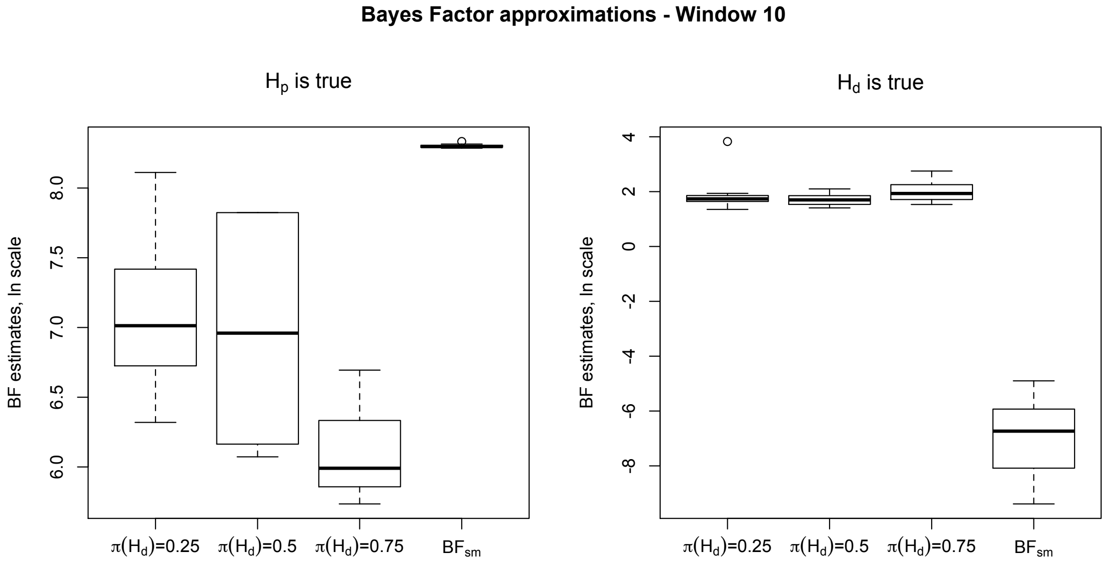


Figure 5: Results of the ABC Bayes Factor approximations in comparison to the Standard Mean approximations, given in natural log scale, for all cases using window 10 measurements. Cases are from left to right: ABC algorithm with $\pi(H_d) = 0.25$; ABC algorithm with $\pi(H_d) = 0.50$; ABC algorithm with $\pi(H_d) = 0.75$; MC Standard Mean method; ABC algorithm with $\pi(H_d) = 0.25$; ABC algorithm with $\pi(H_d) = 0.50$; ABC algorithm with $\pi(H_d) = 0.75$; MC Standard Mean method. Reference Tables 2 and 3 in Appendix C for numerical results.

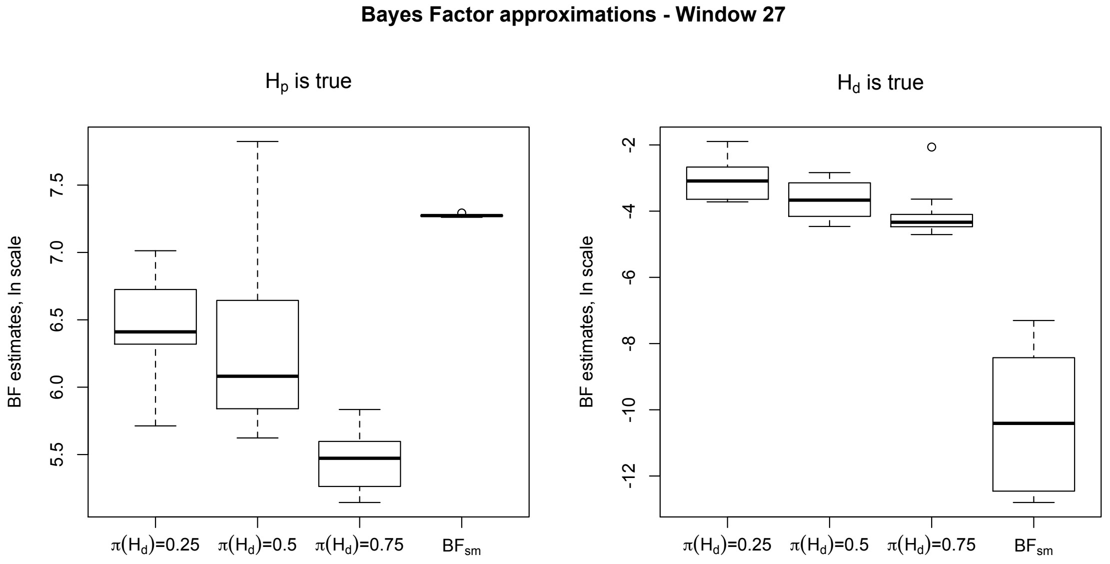
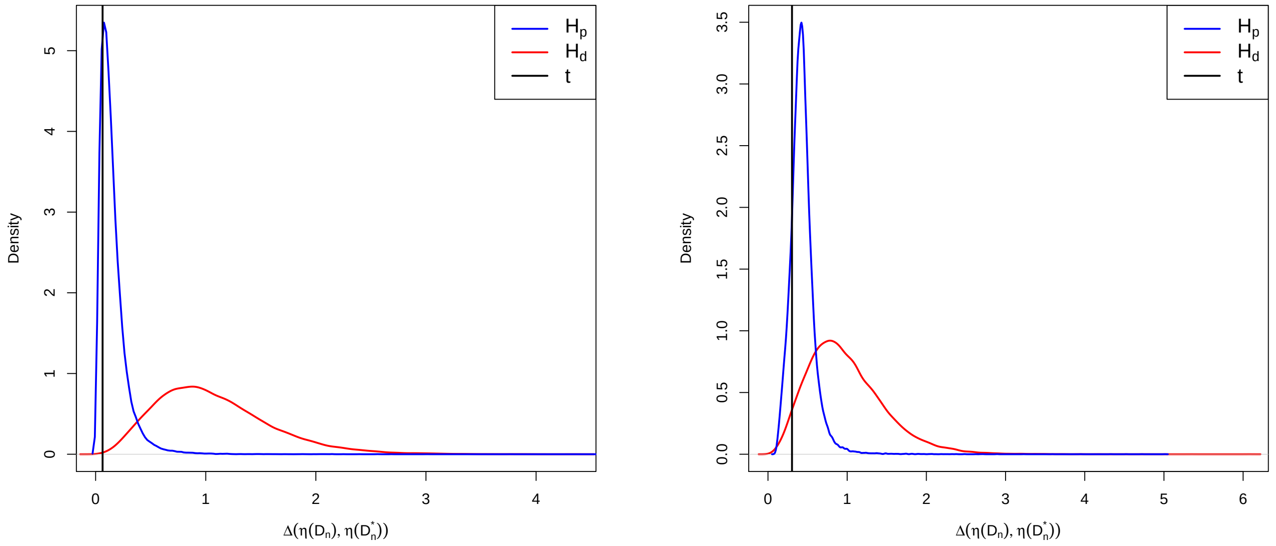


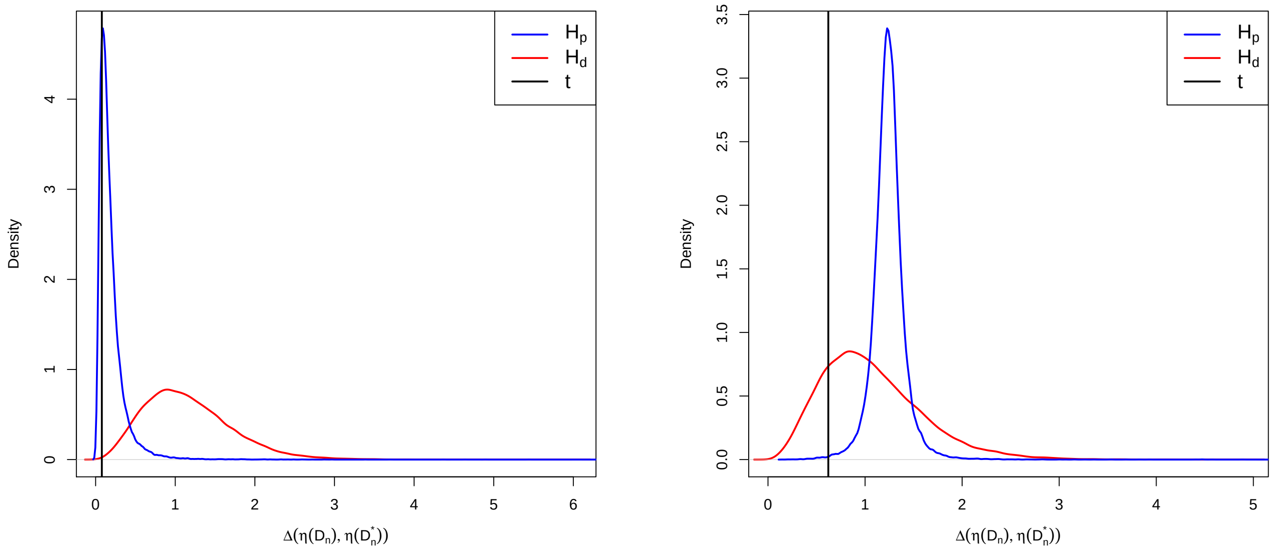
Figure 6: Results of the ABC Bayes Factor approximations in comparison to the Standard Mean approximations, given in natural log scale, for all cases using window 27 measurements. Cases are from left to right: ABC algorithm with $\pi(H_d) = 0.25$; ABC algorithm with $\pi(H_d) = 0.50$; ABC algorithm with $\pi(H_d) = 0.75$; MC Standard Mean method; ABC algorithm with $\pi(H_d) = 0.25$; ABC algorithm with $\pi(H_d) = 0.50$; ABC algorithm with $\pi(H_d) = 0.75$; MC Standard Mean method. Reference Tables 4 and 5 in Appendix C for numerical results.



(a) H_p is true - window 10

(b) H_d is true - window 10

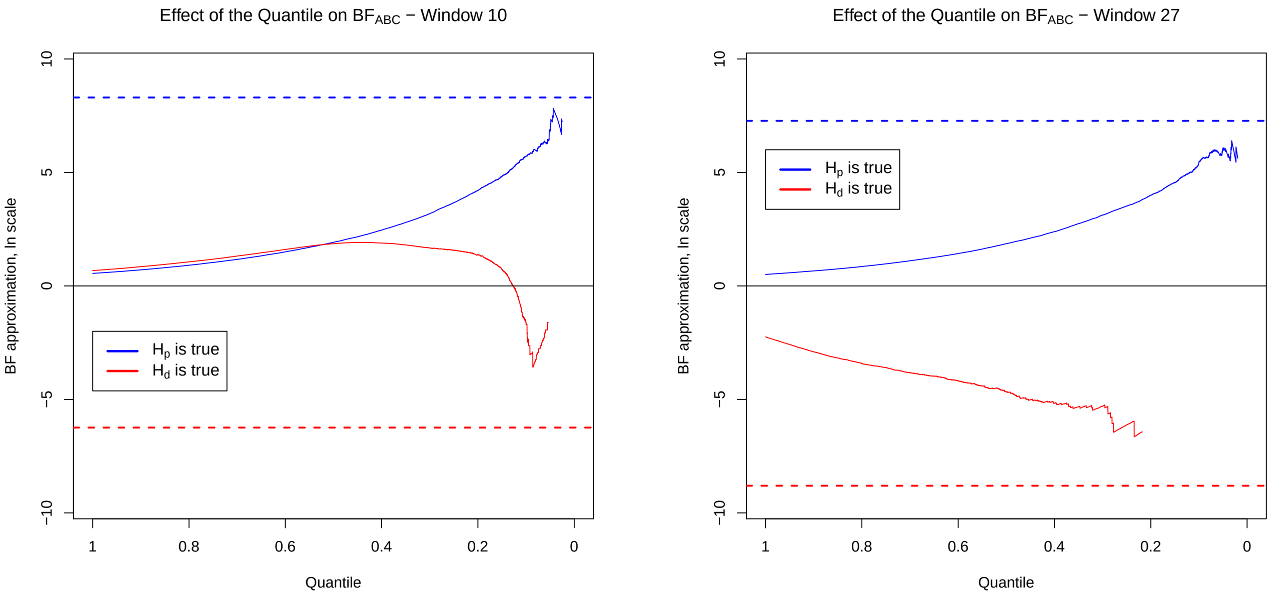
Figure 7: The blue density curve gives the distribution of $\Delta(\eta(D_n),\eta(D_n^*))$ when H_p was sampled. The red density curve gives the distribution of $\Delta(\eta(D_n),\eta(D_n^*))$ when H_d was sampled. The black vertical line gives the value of the threshold t, which corresponds to the 0.10 quantile of the set of $\Delta(\eta(D_n),\eta(D_n^*))$. These plots were created using results from one completion of the algorithm in the case where $\pi(H_d) = 0.50$.



(a) H_p is true - window 27

(b) H_d is true - window 27

Figure 8: The blue density curve gives the distribution of $\Delta(\eta(D_n),\eta(D_n^*))$ when H_p was sampled. The red density curve gives the distribution of $\Delta(\eta(D_n),\eta(D_n^*))$ when H_d was sampled. The black vertical line gives the value of the threshold t, which corresponds to the 0.10 quantile of the set of $\Delta(\eta(D_n),\eta(D_n^*))$. These plots were created using results from one completion of the algorithm in the case where $\pi(H_d) = 0.50$.



(a) window 10

(b) window 27

Figure 9: The blue curve shows how the ABC approximation of the Bayes Factor changes as the quantile in the algorithm changes, in the scenario when H_p is true. Similarly, the red curve shows how the ABC approximation of the Bayes Factor changes as the quantile in the algorithm changes, in the scenario when H_d is true. The blue and red dotted lines give the average value of the 10 Standard Mean approximations in each respective scenario.

A more stable approximation could be produced by reducing its strong dependency on t. This is done in the next section by proposing an alternative approach to assigning an ABC Bayes Factor that relies on a relationship with the receiver operating characteristic.

Modification of ABC methodology using the Receiver Operating Characteristic

The first portion of the ABC algorithm results in two different distributions of distances: the distribution of $\Delta(\eta(D_n),\eta(D_n^*))$'s resulting from D_n^*'s generated under model M_1 and the distribution of $\Delta(\eta(D_n),\eta(D_n^*))$'s resulting from D_n^*'s generated under model M_2. A threshold, t, is used to discriminate among the combined group of distances. Distances less than or equal to t are accepted, while distances greater than t are rejected. The number of acceptances under model M_1 is part of the numerator of BF_{abc}, while the number of acceptances under model M_2 is part of the denominator of BF_{abc}. As illustrated in the glass example in Figure 9, this is very dependent on the quantile-selected value of t from Algorithm 2, and is especially apparent as the quantile approaches zero.

This concept used to arrive at the ABC approximation of the Bayes Factor can be compared to assessment of the performances of classification systems using receiver operating characteristic (ROC) curves. A ROC curve provides a comparison between two distributions of scores. The first distribution is composed of scores obtained by comparing pairs of objects generated under M_1, while the second distribution is composed of scores obtained by comparing pairs of objects obtained under M_2. A ROC curve represents the cumulative distribution of the scores less than a threshold t under the first distribution against the cumulative distribution of scores less than t under the second distribution across all possible values of t [8]. To do this, the two cumulative distributions, denoted $G(t)$ and $F(t)$, are plotted as functions of t on the x- and y-axes, respectively, resulting in the ROC curve [8]. If p is defined by $p = G(t)$, then $t = G^{-1}(p)$. Substituting this expression for t into $F(t)$ gives $F(t) = F(G^{-1}(p))$, or $F(t) = \mathrm{ROC}(p)$. Using this notation, $\mathrm{ROC}(p)$ is plotted on the y-axis while p is plotted on the x-axis [8]. Note that $G^{-1}(p)$ denotes the quantile function, defined by van der Vaart [11].

The ROC can be related back to the ABC Bayes Factor by recalling from Section 2:

$$BF_{abc} = \frac{\sum_{i=1}^{N} \mathbb{I}_{m^i=1}\mathbb{I}_{\Delta(\eta(D_n),\eta(D_n^{*,i}))\le t}}{\sum_{i=1}^{N} \mathbb{I}_{m^i=2}\mathbb{I}_{\Delta(\eta(D_n),\eta(D_n^{*,i}))\le t}} \cdot \frac{\pi(M_2)}{\pi(M_1)}.$$

Defining $k = \sum_{i=1}^{N} \mathbb{I}_{m^i=1}$ and $l = \sum_{i=1}^{N} \mathbb{I}_{m^i=2}$, along with letting $\pi(M_1) = \pi(M_2)$, results in

$$\begin{aligned} BF_{abc} &= \frac{k}{l}\frac{F_k(t)}{G_l(t)} \\ &= \frac{k}{l}\frac{F_k(G_l^{-1}(G_l(t)))}{G_l(t)} \\ &= \frac{k}{l}\frac{F_k(G_l^{-1}(p))}{p} \\ &= \frac{k}{l}\frac{\mathrm{ROC}_{kl}(p)}{p} \\ &= \frac{\mathrm{ROC}_{kl}(p)}{p}. \end{aligned} \tag{4}$$

The last step in Equation 4 assumes that $k = l$ for a large number of simulations with $\pi(M_1) = \pi(M_2)$. Also note that $F_k(t)$ and $G_l(t)$ denote the empirical distribution functions as defined by van der Vaart [11].

Due to the construction of the ABC algorithm, the limit of BF_{abc} when $p \to 0$ is of interest.

$$\begin{aligned} BF &= \lim_{t\to 0} BF_{abc} \\ &= \lim_{p\to 0} BF_{abc} \\ &= \lim_{p\to 0} \frac{\mathrm{ROC}_{kl}(p)}{p} \\ &= \lim_{p\to 0} \frac{\frac{d}{dp}\big(\mathrm{ROC}_{kl}(p)\big)}{\frac{d}{dp}(p)} \\ &= \lim_{p\to 0} \frac{d}{dp}\big(\mathrm{ROC}_{kl}(p)\big) \end{aligned}$$

By the chain rule,

$$\begin{aligned} \frac{d}{dp}\mathrm{ROC}_{kl}(p) &= \frac{d}{dp}F(G^{-1}(p)) \\ &= f(G^{-1}(p)) \cdot \frac{1}{g(G^{-1}(p))} \end{aligned}$$

The empirical ROC can be obtained from samples of scores from the two distributions of $\Delta(\eta(D_n), \eta(D_n^*))$ under M_1 and M_2. The empirical curve can be modeled using a parametric function, resulting in a functional form that can facilitate the Monte-Carlo estimation of the ABC Bayes Factor for t close to 0.

In particular, the binormal representation of the ROC

$$ROC(p) = \Phi(a + b\Phi^{-1}(p)) \tag{5}$$

is a common functional form of the ROC that involves two parameters, $a = \frac{\mu_F - \mu_G}{\sigma_G}$ and $b = \frac{\sigma_F}{\sigma_G}$, where μ_G and σ_G are parameters of one distribution of scores and μ_F and σ_F are parameters of the other distribution of scores[8]. Note that Φ in Equation 5 represents the standard normal CDF.

The slope of the binormal ROC is given below.

$$\frac{\partial(ROC(p))}{\partial p} = \frac{\partial(\Phi(a+b\Phi^{-1}(p)))}{\partial p} = \frac{\phi(a+b\Phi^{-1}(p))\cdot b}{\phi(\Phi^{-1}(p))}$$

The behavior of the first derivative of the ROC as the threshold t approaches zero can be examined by evaluating the limit of the derivative of the binormal ROC as p approaches zero, by utilizing a change of variable from p to the standard normal z-score ($z = \Phi^{-1}(p)$). Note that ϕ as used below denotes the standard normal probability density function (PDF) and Φ denotes the standard normal CDF.

$$\begin{aligned}\lim_{p\to 0}\frac{\phi(a+b\Phi^{-1}(p))\cdot b}{\phi(\Phi^{-1}(p))} &= \lim_{z\to-\infty}\frac{\phi(a+b\cdot z)\cdot b}{\phi(z)}\\ &= \lim_{z\to-\infty}\frac{\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(a+bz)^2}\cdot b}{\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(z)^2}}\\ &= \lim_{z\to-\infty} b\cdot e^{-\frac{1}{2}(a^2+2abz+b^2z^2-z^2)}\end{aligned} \tag{6}$$

The remaining portion of the limit will be evaluated by cases.

When $b = 1$, substitution provides for the following simplifications:

$$\ldots = \lim_{z\to-\infty} e^{-\frac{1}{2}(a^2+2az)}. \tag{7}$$

When $b = 1$ with $a < 0$ it follows that $lim \to 0$.

When $b = 1$ with $a = 0$ it follows that $lim \to 1$.

When $b = 1$ with $a > 0$ it follows that $lim \to \infty$.

When $b > 1$ it follows that $lim \to 0$.

When $0 < b < 1$ it follows that $lim \to \infty$.

When $b = 0$ it follows that $lim \to 0$.

When $b < 0$ it follows that $lim \to 0$.

Thus, the first derivative of the binormal ROC as p (and the threshold t) approaches zero does not produce a stable result. Since BF_{abc} when $t = 0$ is of interest (and thus, the likelihood ratio at $t = 0$), this is not ideal. However, the first derivative of the binormal ROC curve can still be observed at a neighborhood near zero. For this paper, the binormal model will still be used to model the ROC for BF_{abc}, and the first derivative will be evaluated at a neighborhood near zero. Future steps will involve using a mixture of noncentral beta distributions to model the ROC curve near zero.

Results from the glass example using ROC modified ABC

A binormal fit was performed on a set of $\Delta(\eta(D_n), \eta(D_n^*))$'s generated by the ABC algorithm in the case where equal priors were used, H_p was true, and both e_u and e_s originated from window 10, and resulted in a functional form of the ROC. The binormal ROC curve exhibited a close fit to the empirical ABC ROC curve, especially in the neighborhood near $t = 0$. This can be seen in Figure 10. From the fitted binormal function, the first derivative could be examined and evaluated near zero to produce a more stable value of the Bayes Factor than the original ABC algorithm, as illustrated in Figure 11.

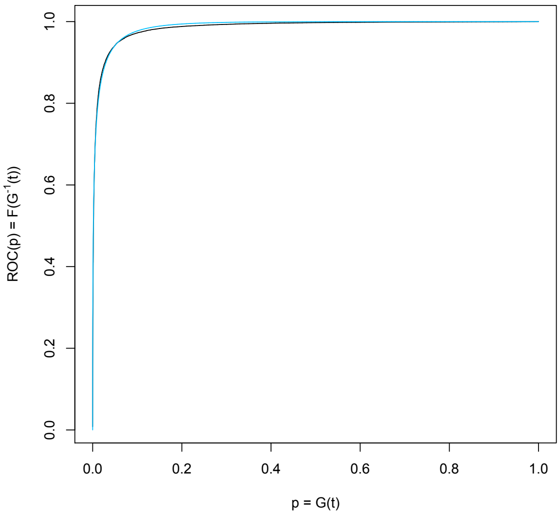


Figure 10: The black curve gives the empirical ABC ROC and the blue curve gives the fitted binormal ROC.

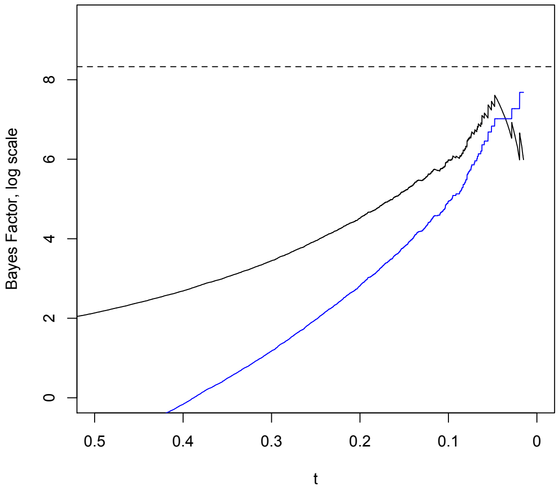


Figure 11: The black curve represents the ABC Bayes Factor (becoming very unstable as it approaches zero) and the blue curve gives the stabilized binormal version of the ABC Bayes Factor. The dashed, black line represents the value of the Standard Mean Bayes Factor.

CONCLUSION

Approximate Bayesian Computation provides a method to approximate the Bayes Factor in cases where the exact likelihood function is not available in a usable form. This is extremely useful in forensic science where complex and high-dimension evidence forms are often encountered, and exact likelihood functions cannot be evaluated. A traditional ABC Bayes Factor could be used in these settings to quantify the probative value of forensic evidence.

Although ABC has received criticism when used for model selection, there is still merit in its abilities as demonstrated by the simple glass example. ABC Bayes Factors provided the correct direction of support in the majority of test cases in the glass example. However, instability in the results indicated an issue relating to the choice of threshold in a traditional ABC model selection algorithm. After addressing this issue by using a Receiver Operating Characteristic curve, updated results of the glass example showed more stability.

In conclusion, results from the controlled glass example show that the ROC modified ABC approach has promising potential in the field of forensic science. This approach also has the potential to be applied to other complex evidence types. More testing would be needed. Further work could also be done using a mixture of noncentral beta distributions to model the ROC curve near zero. Anticipated results would be greater reliability and more stability than the currently proposed modification to ABC provides.

ACKNOWLEDGEMENTS

This research was supported by the SDSU Scholarly Excellence Fund, the Department of Mathematics and Statistics, and the Jerome J. Lohr College of Engineering.

REFERENCES

[1] Aitken, C.G.G., Lucy, D., “Evaluation of trace evidence in the form of multivariate data.” *Appl. Statist.*(53) (2004) 109-122.

[2] Aitken, C.G.G., and Taroni, F., *Statistics and the Evaluation of Evidence for Forensic Scientists, Second Ed.*. West Sussex, England: John Wiley & Sons Ltd, 2004.

[3] Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., *Bayesian Data Analysis, Third Ed.*. Boca Raton, FL: CRC Press, 2014.

[4] Lombaert, E., Guillemaud, T., Lundgren, J., Koch, R., Facon, B., Grez, A., Loomans, A., Malausa, T., Nedved, O., Rhule, E., Staverlokk, A., Steenberg, T., and Estoup, A., “Complementarity of statistical treatments to reconstruct worldwide routes of invasion: the case of the Asian ladybird *Harmonia axyridis*.” *Molecular Ecology*(23) (2014) 5979-5997.

[5] Marin, J-M. and Robert, C.P., *Bayesian Essentials with R, Second Ed.*. New York: Springer Science + Business Media, 2014.

[6] Milton, J.S. and Arnold, J.C., *Introduction to Probability and Statistics*. New York: McGraw-Hill, 2003.

[7] Ommen, Danica (2014) “Computational Aspects of Forensic Evidence Interpretation: Paper for Master’s of Mathematics Degree.” Unpublished Manuscript.

[8] Pepe, Margaret S., *The Statistical Evaluation of Medical Tests for Classification and Prediction*. New York: Oxford University Press Inc., 2003.

[9] Pudlo, P., Marin, J-M., Estoup, A., Cornuet, J-M., Gautier, M., Robert, CP., “Reliable ABC model choice via random forests.” *Technical Report arXiv:1406.6288, ArXiv* (2015) Submitted to PNAS.

[10] Robert, C.P., Cornuet, J-M., Marin J-M., and Pillai, N.S., “Lack of confidence in ABC model choice.” *Proceedings of the National Academy of Sciences of the United States of America*(108) (2011) 5112-5117.

[11] Vaart, A.W. van der, *Cambridge Series in Statistical and Probabilistic Mathematics: Asymptotic Statistics*. Cambridge, UK: Cambridge University Press, 1998.

APPENDIX A: Notation from ABC and Bayesian Statistics

symbol	description	page
θ	a set of parameters describing a population	1
$\pi(\theta)$	the **prior density** represents a belief about the value of population parameters prior to observing data	1
D_n	a sample data set with n observations	1
$f(D_n\|\theta)$	the **likelihood function** gives the probability of observing the data set D_n as a function of the parameters θ	1
$\pi(\theta\|D_n)$	the **posterior density** represents a belief about the values of the population parameters after observation of the data	1
M_1 and M_2	two defined, competing models	1
$\pi(D_n\|M_i)$	the **marginal density** is the density of D_n under M_i, not dependent on θ	1
$\frac{\pi(M_1\|D_n)}{\pi(M_2\|D_n)}$	the **posterior odds** is the ratio of probabilities for two competing models given a common data set	1
$\frac{\pi(M_1)}{\pi(M_2)}$	the **prior odds** is the ratio of model priors for two competing models	2
$BF = \frac{\pi(D_n \mid M_1)}{\pi(D_n \mid M_2)}$	the **Bayes Factor** gives the quantitative value of evidence	2
D_n^*	a sampled data set with n observations	3
$f_m(D_n^*\|\theta_i)$	the simpler likelihood function chosen for use in the ABC algorithm	3
$\eta(D_n)$	summary statistic on D_n	3
$\eta(D_n^*)$	summary statistic on D_n^*	3
t	the threshold used in the ABC algorithm	3
$\pi_t(\theta \mid D_n)$	the ABC approximation of the posterior density	3
N_{acc}	the predetermined number of accepted parameters in the ABC algorithm.	4
N_{sim}	the fixed number of repetitions for the modified ABC algorithm	4
$\Delta(\eta(D_n),\eta(D_n^*))$	the distance metric computed in the ABC algorithm	4
BF_{abc}	the ABC approximation of the Bayes Factor	4
e_u	observations characterizing the trace evidence	6
e_s	observations characterizing the evidence from the specific source	6
e_a	observations characterizing the alternative population	6
H_p	the prosecution hypothesis	6
H_d	the defense hypothesis	6

APPENDIX B: Glass data

The subset of the data used for the glass example is provided in the table below. The statistics on the fragments plotted in the pairwise scatter plots (Figures 1 - 4) are given in the first three columns.

Table 1: Glass data, Group 1

logCaK	logCaSi	logCaFe	window
4.91496	-0.34855	2.43375	2
4.8904	-0.33896	2.45729	2
4.811	-0.34886	2.44283	2
4.96556	-0.3414	2.43553	2
4.89558	-0.35564	2.45974	2
4.3378	-0.26225	2.63358	5
4.19659	-0.26582	2.66615	5
4.30314	-0.26684	2.68043	5
4.34333	-0.27339	2.67434	5
4.2737	-0.26973	2.67515	5
4.50285	-0.4015	2.65562	6
4.71998	-0.40101	2.70488	6
4.59287	-0.40417	2.66377	6
4.54464	-0.41209	2.66506	6
4.61372	-0.40982	2.6835	6
4.61924	-0.35607	2.69542	10
4.64588	-0.36343	2.71724	10
4.57294	-0.36936	2.68579	10
4.43477	-0.37827	2.71534	10
4.66207	-0.36113	2.68219	10
4.55597	-0.41731	2.67724	11
4.62876	-0.43356	2.65036	11
4.59576	-0.42875	2.65662	11
4.48033	-0.41931	2.63373	11
4.76723	-0.41832	2.67436	11
4.72771	-0.28559	2.75208	26
5.01285	-0.29837	2.7742	26
4.90503	-0.28864	2.74915	26
4.76014	-0.29178	2.72062	26
4.74913	-0.30246	2.75422	26
5.11487	-0.34261	2.9874	27

Table 1: Glass data, Group 1

logCaK	logCaSi	logCaFe	window
4.9873	-0.35484	2.92806	27
5.19403	-0.36614	2.99461	27
5.12464	-0.35581	3.01663	27
5.04411	-0.36464	2.97961	27
5.00871	-0.35024	2.94909	28
5.64772	-0.34551	2.96003	28
5.08184	-0.34541	2.95508	28
5.46454	-0.3348	3.00193	28
4.91008	-0.35085	3.0165	28
4.85665	-0.27848	2.72331	31
4.80565	-0.27413	2.75176	31
4.81562	-0.28573	2.70856	31
4.94187	-0.28045	2.71725	31
5.02693	-0.29857	2.72357	31
5.11738	-0.34456	3.00107	33
5.09724	-0.35778	3.01569	33
5.02488	-0.3552	3.05047	33
5.09537	-0.33615	3.05705	33
5.00744	-0.35284	3.01518	33
4.58919	-0.33376	2.29581	37
4.55096	-0.33616	2.31947	37
4.6238	-0.33274	2.34727	37
4.58853	-0.32933	2.32211	37
4.64406	-0.32723	2.30886	37
4.41554	-0.31573	2.29715	42
4.43597	-0.31287	2.3331	42
4.62635	-0.30298	2.28141	42
4.52247	-0.31785	2.31428	42
4.60349	-0.30033	2.28786	42
4.76416	-0.33157	2.80666	46
4.86111	-0.34159	2.89627	46
4.97293	-0.33882	2.81864	46
4.74509	-0.34084	2.871	46
4.67535	-0.35184	2.84802	46
4.45784	-0.30217	2.63046	47

Table 1: Glass data, Group 1

logCaK	logCaSi	logCaFe	window
4.56416	-0.29569	2.68883	47
4.61673	-0.30759	2.67237	47
4.68832	-0.30885	2.67029	47
4.50146	-0.31348	2.66752	47
4.81052	-0.33034	2.85795	48
4.93715	-0.30307	2.87275	48
4.63925	-0.32851	2.83454	48
4.80885	-0.32053	2.8127	48
4.76915	-0.31918	2.8568	48
5.80845	-0.45109	3.4743	58
5.31633	-0.45276	3.46656	58
5.75272	-0.46434	3.47846	58
5.60795	-0.46332	3.46605	58
5.78077	-0.46379	3.49782	58

APPENDIX C: Glass example results

Numerical results obtained from the simulations described in Algorithm 3 are given below in Tables 2 through 5.

Table 2: Approximate Bayes Factors when H_p is true, with trace evidence e_u from window 10.

	$\pi(H_d) = 0.25$	$\pi(H_d) = 0.50$	$\pi(H_d) = 0.75$	Standard Mean
	1110.778	2499	349.9412	4005.284
	3333	1110.111	326.6703	4051.384
	833	1110.111	309.5	4045.92
	1666.333	768.2308	386.6104	4018.659
	1110.778	433.7826	402.4054	3984.788
	833	475.1905	413.6667	3993.366
	555.2222	433.7826	807.8108	4085.482
	3333	999	397	3966.998
	1110.778	2499	573.9231	4161.404
	555.2222	2499	563.0377	4006.677
AVG	1444.11114	1282.72085	453.05656	4031.9962

Table 3: Approximate Bayes Factors when H_d is true, with trace evidence e_u from window 10.

	$\pi(H_d) = 0.25$	$\pi(H_d) = 0.50$	$\pi(H_d) = 0.75$	Standard Mean
	6.944687	4.144033	7.387812	0.001580734
	5.240803	6.262164	4.82881	0.000851726
	6.414305	4.640158	12.51992	0.000896609
	5.979798	5.968641	9.526096	0.003657319
	45.96296	8.157509	15.65672	0.000308687
	5.167217	5.006006	5.542141	0.007451877
	5.683514	7.748906	6.454775	0.000162563
	4.779141	6.380074	4.625826	8.38E-05
	3.864819	4.991612	5.862629	0.001812223
	5.672673	4.081301	8.843664	0.002659687
AVG	9.5709917	5.7380404	8.1248393	0.001946519

Table 4: Approximate Bayes Factors when H_p is true, with trace evidence e_u from window 27.

	$\pi(H_d) = 0.25$	$\pi(H_d) = 0.50$	$\pi(H_d) = 0.75$	Standard Mean
	555.2222	369.3704	341.8276	1437.623
	833	276.7778	212.8273	1434.838
	555.2222	293.1176	242.9016	1423.567
	555.2222	2499	269.7273	1444.326
	666.3333	768.2308	269.7273	1447.91
	1110.778	383.6154	312.7895	1469.986
	833	624	188.0828	1439.504
	302.697	343.8276	171.4186	1426.103
	666.3333	768.2308	193.0784	1445.715
	555.2222	499	233.2205	1447.562
AVG	663.30304	682.51704	243.56009	1441.7134

Table 5: Approximate Bayes Factors when H_d is true, with trace evidence e_u from window 27.

	$\pi(H_d) = 0.25$	$\pi(H_d) = 0.50$	$\pi(H_d) = 0.75$	Standard Mean
	0.05100119	0.01317123	0.009027081	3.87E-06
	0.02621076	0.02848915	0.01144349	0.000218999
	0.02424372	0.01564087	0.009630819	2.76E-06
	0.05827068	0.04307917	0.01265314	3.89E-06
	0.02470462	0.04014978	0.01356102	4.43E-05
	0.06938907	0.05385183	0.01416658	2.05E-05
	0.0406942	0.05864916	0.01265314	0.000470388
	0.1501789	0.02291326	0.02632906	0.000673247
	0.07626772	0.01698363	0.1269543	5.92E-05
	0.03446247	0.01153146	0.01659125	4.10E-06
AVG	0.055542333	0.030445954	0.025300988	0.000150121