Glen Cowan
RHUL Physics
28 July, 2008

Error analysis for efficiency

To estimate a selection efficiency using Monte Carlo one typically takes the number of events selected m divided by the number generated N. This note discusses how to account for the statistical uncertainty in the estimated efficiency. In Section 1 we model m as a binomial random variable and find the maximum-likelihood (ML) estimator for the efficiency and its variance. The issue of conditioning on the total number of events N is discussed. In Section 2 we show how to construct a confidence interval for the efficiency. Sometimes one wants to fit the efficiency as a function of some other variable, e.g., photon selection efficiency versus energy; this is examined in Section 3. Finally in Section 4 we discuss the Bayesian approach to the problem.

1 Binomial model

The usual approach when estimating a selection efficiency is to treat the number of selected events m as a binomially distributed variable, i.e., one finds m "successes" out of N independent trials, where the probability of success on each trial is the efficiency ε. That is, the probability to select m events is

$$P(m; N, \varepsilon) = \frac{N!}{m!(N-m)!}\varepsilon^m(1-\varepsilon)^{N-m} \,, \tag{1}$$

where the notation above indicates that the m listed before the semicolon is treated as the random variable and the number of trials N and efficiency ε are parameters. We take N as known and our goal is to estimate ε.

The log-likelihood function for the unknown parameter ε is

$$\ln L(\varepsilon) = m \ln \varepsilon + (N-m)\ln(1-\varepsilon) + C \,, \tag{2}$$

where C represents terms that do not depend on ε and can therefore be dropped. Setting the derivative of $\ln L(\varepsilon)$ equal to zero gives

$$\hat{\varepsilon} = \frac{m}{N} \,, \tag{3}$$

where the hat is used to denote the *estimator* of the true parameter ε.

The variance of the binomially distributed m is

$$V[m] = N\varepsilon(1-\varepsilon) \,, \tag{4}$$

and so the variance of the estimator $\hat{\varepsilon}$ is

$$V[\hat{\varepsilon}] = V\left[\frac{m}{N}\right] = \frac{1}{N^2}V[m] = \frac{\varepsilon(1-\varepsilon)}{N} \,. \tag{5}$$

Of course to evaluate this numerically we need an *estimator* of the variance of the estimator, and we can take this to be

$$\widehat{V}[\hat{\varepsilon}] = \frac{\hat{\varepsilon}(1-\hat{\varepsilon})}{N} = \frac{m(1-m/N)}{N^2} \,. \tag{6}$$

The estimate of the standard deviation is simply the square root,

$$\hat{\sigma}[\hat{\varepsilon}] = \sqrt{\frac{\hat{\varepsilon}(1-\hat{\varepsilon})}{N}} = \frac{\sqrt{m(1-m/N)}}{N} \,. \tag{7}$$

One can show that with equation (7), the following inequalities hold:

$$\hat{\varepsilon} - \hat{\sigma}[\hat{\varepsilon}] \;\; \geq \;\; 0 \,, \tag{8}$$

$$\hat{\varepsilon} + \hat{\sigma}[\hat{\varepsilon}] \;\; \leq \;\; 1 \,. \tag{9}$$

It may be of some comfort to know that the estimate plus or minus one standard deviation stays with in the allowed range of $0 \leq \varepsilon \leq 1$, but in fact if it were not true this would not imply any particular contradiction. Using one standard deviation to represent the statistical error is after all a matter of convention, and if we had taken 2σ as the conventional error bar then the corresponding inequalities would not hold.

Another interesting special case is $m = N$, i.e., $\hat{\varepsilon} = 1$. Then from equation (7) one has $\hat{\sigma}[\hat{\varepsilon}] = 0$. This of course does not necessarily mean that the true efficiency is equal to unity nor that the true standard deviation of the estimator $\hat{\varepsilon}$ is zero, but rather simply that our estimates of these quantities have come out this way for the given MC data set. Similar considerations apply of course to the case $m = 0$. In both cases, the estimate $\hat{\sigma}[\hat{\varepsilon}] = 0$ is essentially useless. In particular it cannot be used in the method of least squares (see Section 3.1). However, rather than modifying the estimated error, it is often better to proceed in a way that sidesteps completely the need for $\hat{\sigma}[\hat{\varepsilon}]$, as we show in Section 3.2.

Often one wants to compute a set of efficiencies, e.g., one for each bin of a histogram, $\varepsilon_1, \ldots \varepsilon_{N_{\rm bin}}$. Suppose the true number of events in each bin is $N_1, \ldots, N_{N_{\rm bin}}$, and the corresponding numbers of events selected are $m_1, \ldots, m_{N_{\rm bin}}$. If these numbers are put into ROOT histograms, then the routine `TH1::Divide` can be used to compute the corresponding efficiencies. If the option `"B"` is used, then this routine will compute the errors according to the binomial model described above. Note that here m_i represents the total number of events selected out of the N_i generated. If the events are subject to some smearing that results in migration between bins, then the number selected in the same bin as where they were generated will be less. Here, however, m_i means the number selected anywhere out of the N_i events generated in bin i (see also Section 4 concerning the routine `TGraphAsymmErrors::BayesDivide`).

1.1 Conditioning on the total number of events

There may be cases where the number of generated events N is also random. For example, one may first select the N events from a larger sample, and so N is itself could be binomially distributed. One could choose to model m/N as a ratio of two binomially distributed variables, in which case the variance of the ratio would be larger than in the case of treating N as a constant. This, however, does not in general provide the best way to summarize the statistical uncertainty in the estimate of the efficiency.

For purposes of estimating the statistical uncertainty of $\hat{\varepsilon} = m/N$, it is almost always best to treat N as constant. This is an example of *conditioning on an ancillary statistic.* That is, the estimate of of the efficiency ε is conditional upon having obtained N generated events from which one finds m that are selected. An ancillary statistic, here N, is a random quantity whose distribution does not depend on the parameter of interest, in this case ε. That is, if we were to repeat the entire estimation procedure many times with statistically independent data, then sometimes N would be higher, sometimes lower. If N comes out high, we obtain a more accurate estimate of ε, i.e., we take the reported accuracy to be conditional upon having found the given N.

The idea of conditioning on the total number of events arises in a similar way when one estimates a branching fraction B from the observation of m decays to a final state X out of N total events. Suppose experiment A finds 5 events, and 2 of them go to final state X. In a run of the same integrated luminosity, competitor experiment B happens to find 10 events, and 4 of them decay to X. Experiment B's estimate of the branching ratio really is more precise, because they have been lucky to obtain more events. There are theories, however, (e.g., the Standard Model), where branching ratios and event production rates are related to more fundamental parameters. If one were to insist on the model relation between the production rate and branching ratio, then the number of events N is no longer an ancillary statistic. If one were to estimate directly the more fundamental model parameters, then both m and N would be treated as random quantities. A branching ratio or an efficiency, however, would usually be regarded as a separate parameter, and one would condition on N. More discussion on conditioning can be found in the book by Cox [1].

2 Confidence intervals

The standard deviation $\sigma[\hat{\varepsilon}]$ is a single parameter that characterizes the width of the sampling distribution of the estimator $\hat{\varepsilon}$, i.e., the distribution of $\hat{\varepsilon}$ values that one would obtain from a large number of repetitions of the entire procedure. This distribution, $f(\hat{\varepsilon}; \varepsilon, N)$, which depends on ε and N as parameters, is not in general a symmetric function, i.e., upwards and downwards fluctuations of $\hat{\varepsilon}$ may not be equally likely. In some circumstances one may want to quantify the statistical precision of the efficiency by constructing a *confidence interval* $[\varepsilon_{\rm lo}, \varepsilon_{\rm up}]$. This interval can be asymmetric about the point estimate. General procedures for constructing confidence intervals can be found in, for example, [2, 3].

We can specify separate probabilities for the upper and lower edges of the confidence interval to be above or below the true parameter:

$$P(\varepsilon_{\rm lo} > \varepsilon) \;<\; \alpha_{\rm lo} \,, \tag{10}$$

$$P(\varepsilon_{\rm up} < \varepsilon) \;<\; \alpha_{\rm up} \,. \tag{11}$$

Taken together these equations say that the interval $[\varepsilon_{\rm lo}, \varepsilon_{\rm up}]$ covers the true parameter with probability

$$P(\varepsilon_{\rm lo} \le \varepsilon \le \varepsilon_{\rm up}) \ge 1 - \alpha_{\rm lo} - \alpha_{\rm up} \,. \tag{12}$$

Inequalities are used in the equations above because of the discrete nature of the binomial data, e.g., the probability for the interval $[\varepsilon_{\rm lo}, \varepsilon_{\rm up}]$ to cover ε is *greater than* or equal to

$1 - \alpha_{\rm lo} - \alpha_{\rm up}$. One constructs the intervals so that the probabilities are as close as possible to the desired values on the right-hand sides of (10)—(12).

The special case of $\alpha_{\rm lo} = \alpha_{\rm up} \equiv \gamma/2$ gives a *central* confidence interval with confidence level CL $= 1 - \gamma$. Note that this does not necessarily mean that the interval is symmetric about the point estimate $\hat{\varepsilon}$, rather only that there are equal probabilities for the interval to miss the true value from above and from below. For the case where the estimator $\hat{\varepsilon}$ is Gaussian distributed, e.g., when N is very large and ε is sufficiently far from both 0 and 1, then the interval found from plus or minus one standard deviation about the estimator $\hat{\varepsilon}$ is central confidence interval with confidence level CL $= 1 - \gamma = 0.683$. For this reason one often chooses as a convention to construct a central confidence interval with confidence level CL $= 0.683$ also for those cases where the estimator does not follow a Gaussian distribution.

For the case of binomially distributed m out of N trials with probability of success ε, the upper and lower limits on ε are found to be [3],

$$\varepsilon_{\rm lo} = \frac{mF_F^{-1}[\alpha_{\rm lo}; 2m, 2(N-m+1)]}{N-m+1 \, + \, mF_F^{-1}[\alpha_{\rm lo}; 2m, 2(N-m+1)]} \, , \tag{13}$$

$$\varepsilon_{\rm up} = \frac{(m+1)F_F^{-1}[1-\alpha_{\rm up}; 2(m+1), 2(N-m)]}{(N-m) \, + \, (m+1)F_F^{-1}[1-\alpha_{\rm up}; 2(m+1), 2(N-m)]} \, . \tag{14}$$

Here F_F^{-1} is the quantile of the F distribution (also called the Fisher–Snedecor distribution; see [4]). The function F_F^{-1} can be obtained using `ROOT::Math::fdistribution_quantile` from the ROOT's MathCore library [5].

Instead of constructing a central confidence interval, one can instead use equations (13) and (14) to set upper or lower limits, i.e., one of either $\alpha_{\rm lo}$ or $\alpha_{\rm up}$ is taken to be zero. For example, one could take $\alpha_{\rm lo} = 0.9$ and use (13) to obtain a 90% CL lower limit on the efficiency, which is of particular interest if one finds $m = N$.

3 Parametric approach

Suppose that the events are generated as a function of some other variable x, e.g., photons are generated as a function of energy. In this section we consider how to estimate the efficiency as a function of x using a parametric function.

The Monte Carlo produces two sets of numbers: the number of true events N_i in bin i, with $i = 1, \ldots, N_{\rm bin}$, in each bin of the variable x, and the number of those events that satisfy the selection criteria, m_i. Note that being selected does not necessarily imply that the reconstructed value of the variable x is in the same bin as that of the original event. If the event whose true value of x is in bin i is accepted anywhere, it counts towards m_i. The question of the migration of events between bins can be treated by defining a *response matrix*

$$R_{ij} = P(\text{event found in bin } i | \text{true value in bin } j) \, . \tag{15}$$

What is meant in this note by the efficiency is

$$\varepsilon_j = \sum_i R_{ij} = P(\text{event found anywhere} | \text{true value in bin } j) \, . \tag{16}$$

In cases where one needs to include effects of migration between bins one can easily generalize the ideas presented in this note to the estimation of the response matrix R_{ij}. Here we will only discuss the efficiency.

The m_i are all independent, binomial variables and therefore the estimate of the efficiency ε_i is as before

$$\hat{\varepsilon}_i = \frac{m_i}{N_i} \; . \tag{17}$$

Now suppose we have a parametric function that gives the efficiency as a function of the variable x, $\varepsilon(x;\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ represents a set of parameters whose values are not known. Our goal is to estimate the parameters, and then the estimated efficiency at any value of x can be found from the function $\hat{\varepsilon}(x) = \varepsilon(x;\hat{\boldsymbol{\theta}})$.

3.1 Least squares

One way of estimating the parameters $\boldsymbol{\theta}$ is to use the method of least squares in conjunction with the estimates of the efficiency and their standard deviations obtained as described in Section 1. That is, we define the estimator $\hat{\boldsymbol{\theta}}$ to be the value of $\hat{\boldsymbol{\theta}}$ that minimizes the quantity

$$\chi^2(\boldsymbol{\theta}) = \sum_{i=1}^{N_{\rm bin}} \frac{(\hat{\varepsilon}_i - \varepsilon_i(\boldsymbol{\theta}))^2}{\hat{\sigma}[\hat{\varepsilon}_i]^2} \; . \tag{18}$$

Here $\varepsilon_i(\boldsymbol{\theta})$ denotes the predicted efficiency averaged over the ith bin. As long as the variation of the efficiency is not too nonlinear over an individual bin, one can use the approximation

$$\varepsilon_i(\boldsymbol{\theta}) \approx \varepsilon(\langle x\rangle_i;\boldsymbol{\theta}) \; , \tag{19}$$

where $\langle x\rangle_i$ is the average value of x in the ith bin. Note that the least-squares method requires the standard deviations $\hat{\sigma}[\hat{\varepsilon}_i]$ of the estimated values $\hat{\varepsilon}_i$. The confidence interval from Section 2 plays no role in this procedure.

Typically one will use a program such as MINUIT to minimize $\chi^2(\boldsymbol{\theta})$ thus determine the estimators $\hat{\boldsymbol{\theta}}$. The program will also deliver an estimate of the covariance matrix of the estimators

$$V_{ij} = \mathrm{cov}[\hat{\theta}_i, \hat{\theta}_j] \; . \tag{20}$$

Usually this will be obtained by using the approximate relation between the inverse covariance matrix and the second derivative of $\chi^2(\boldsymbol{\theta})$ at its minimum,

$$V_{ij}^{-1} \approx -\frac{1}{2}\left[\frac{\partial^2\chi^2(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \; . \tag{21}$$

This is a special case of the procedure used in connection with Maximum Likelihood and for the approximation to hold the data $\hat{\varepsilon}_i$ should be Gaussian distributed (see below).

The minimized value of $\chi^2(\boldsymbol{\theta})$ can as usual be used to assess the goodness-of-fit.

3.2 Maximum likelihood

If the estimates $\hat{\varepsilon}_i$ are treated as being Gaussian distributed, then one can easily show that $\chi^2(\boldsymbol{\theta}) = -2\ln L(\boldsymbol{\theta}) + C$, where $L(\boldsymbol{\theta})$ is the likelihood function and C does not depend on $\boldsymbol{\theta}$. If this holds then the method of least squares and the method of maximum likelihood are equivalent. But the Gaussian approximation for the $\hat{\varepsilon}_i$ only holds when N, m and $N - m$ are all large, i.e, one needs lots of events and the efficiency cannot be too close to zero or unity. If this is not the case one should construct the ML estimators using the correct sampling distribution of the measurements. Since we have $\hat{\varepsilon}_i = m_i/N_i$ where m_i is binomially distributed, we can construct the likelihood function using the m_i directly; the $\hat{\varepsilon}_i$ and $\hat{\sigma}[\hat{\varepsilon}_i]$ are not required explicitly.

The likelihood function is given by the joint probability for the set of measured values m_i, $i = 1, \ldots, N_{\text{bin}}$. Since they are independent and binomially distributed, we have

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N_{\text{bin}}} P(m_i; N_i, \varepsilon_i(\boldsymbol{\theta})) = \prod_{i=1}^{N_{\text{bin}}} P(m_i; N_i, \varepsilon(x_i; \boldsymbol{\theta})) \tag{22}$$

where in the final part of (22) the relation for $\varepsilon_i(\boldsymbol{\theta})$ given by equation (19) was used. Using the binomial law (1) and taking the logarithm gives the log-likelihood function,

$$\ln L(\boldsymbol{\theta}) = \sum_{i=1}^{N_{\text{bin}}} [m_i \ln \varepsilon(x_i; \boldsymbol{\theta}) + (N_i - m_i) \ln(1 - \varepsilon(x_i; \boldsymbol{\theta}))] + C \;, \tag{23}$$

where C represents terms that do not depend on $\boldsymbol{\theta}$ and therefore can be dropped.

Typically one will use a program such as MINUIT to maximize the log-likelihood and thus determine the estimators $\hat{\boldsymbol{\theta}}$. As in the least-squares case, the program will also deliver an estimate of the covariance matrix of the estimators, usually obtained using the approximate relation between the inverse covariance matrix and the second derivative of the log-likelihood at its maximum,

$$V_{ij}^{-1} \approx - \left[\frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j} \right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \;. \tag{24}$$

This approximation holds for ML estimators in the large sample limit and is usually regarded as accurate enough for most problems.

3.3 Error propagation

Using the methods of least squares or maximum likelihood provides estimators for the parameters $\boldsymbol{\theta}$ that enter into the function $\varepsilon(x; \boldsymbol{\theta})$. The uncertainty in the efficiency at an arbitrary value of x can be obtained using error propagation. The variance of $\varepsilon(x; \hat{\boldsymbol{\theta}})$ is (see e.g. [2, 3]),

$$V[\varepsilon(x; \hat{\boldsymbol{\theta}})] = \sum_{i,j} \left[\frac{\partial\varepsilon(x; \boldsymbol{\theta})}{\partial\theta_i} \frac{\partial\varepsilon(x; \boldsymbol{\theta})}{\partial\theta_j} \right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} V_{ij} \;, \tag{25}$$

where the sum over i and j runs over all of the components of $\boldsymbol{\theta}$ and V_{ij} represents the estimate of the covariance matrix $\text{cov}[\hat{\theta}_i, \hat{\theta}_j]$. The estimated standard deviation $\hat{\sigma}[\varepsilon(x; \hat{\boldsymbol{\theta}})]$ is of course the square root of the variance (25).

If one needs the covariance between estimates of the efficiency at two different values, say, x and x', then from error propagation one has

$$\mathrm{cov}[\varepsilon(x;\hat{\boldsymbol{\theta}}),\varepsilon(x';\hat{\boldsymbol{\theta}})] = \sum_{i,j}\left[\frac{\partial\varepsilon(x;\boldsymbol{\theta})}{\partial\theta_i}\frac{\partial\varepsilon(x';\boldsymbol{\theta})}{\partial\theta_j}\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} V_{ij} \ . \tag{26}$$

In some cases it may be possible to compute the required derivatives in closed form, in others one may need to estimate them numerically using small steps $\Delta\theta_i$ in the parameters. If this is done, then these steps need to be sufficiently large so as to avoid rounding errors, but not so large so as to give a poor estimate of the derivative. As a rule of thumb one could take the $\Delta\theta_i$ to be somewhere in the range 0.1 to 0.5 times the standard deviation $\sigma[\hat{\theta}_i]$. This is then valid as long as a the efficiency as a function of the parameter is sufficiently linear over a range comparable to $\sigma[\hat{\theta}_i]$. This must be true in any case for the error propagation formula to be a valid approximation.

4 Bayesian approach

The methods described above are all in the framework of *frequentist statistics*, i.e., probabilities are only associated with the outcomes of repeatable measurements, and not with the value of a parameter such as the efficiency ε. In *Bayesian statistics*, one can associate a probability with a parameter and treat it as a degree of belief about where the parameter's true value lies. The Bayesian approach for estimating efficiencies has been described in [6].

Suppose we generate N events of which m are selected. We treat m as a binomially distributed variable with selection probability of each event ε, and our goal is to determine ε. Using Bayes' theorem we can write the *posterior probability density* for ε as

$$p(\varepsilon|m,N) = \frac{P(m|\varepsilon,N)\pi(\varepsilon)}{\int P(m|\varepsilon,N)\pi(\varepsilon)\,d\varepsilon} \ . \tag{27}$$

Here $P(m|\varepsilon,N)$ is the likelihood, here written as the conditional probability of the data m given the parameters ε and N. The probability density function (pdf) $\pi(\varepsilon)$ is the *prior* probability, i.e., it reflects our degree of belief about ε prior to observing m. The denominator in (27) serves to normalize the posterior density $p(\varepsilon|m,N)$ to unity.

Bayes' theorem effectively states how one's knowledge about ε should be updated in the light of the observation m. Bayesian statistics provides no unique recipe, however, for writing down the prior pdf $\pi(\varepsilon)$. This could depend on theoretical prejudice, previous measurements, symmetries, and so forth. For the case of the efficiency, for example, we would clearly set the prior equal to zero outside the allowed region $0 \le \varepsilon \le 1$. If we have no prior knowledge of the efficiency beyond that, one could choose a uniform prior between zero and one:

$$\pi(\varepsilon) = \begin{cases} 1 & 0 \le \varepsilon \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{28}$$

If we use this prior together with the binomial probability (1) for the likelihood, Bayes' theorem gives for the posterior pdf of ε (see [6]),

$$p(\varepsilon|m,N) = \frac{\Gamma(N+2)}{\Gamma(m+1)\Gamma(N-m+1)}\varepsilon^m(1-\varepsilon)^{N-m} \ , \tag{29}$$

where Γ is the Euler gamma function (`TMath::Gamma`)[1], which is a special case of the beta distribution. Figure 1 shows the posterior density obtained using the uniform prior with $N = 100$ events for several values of m.

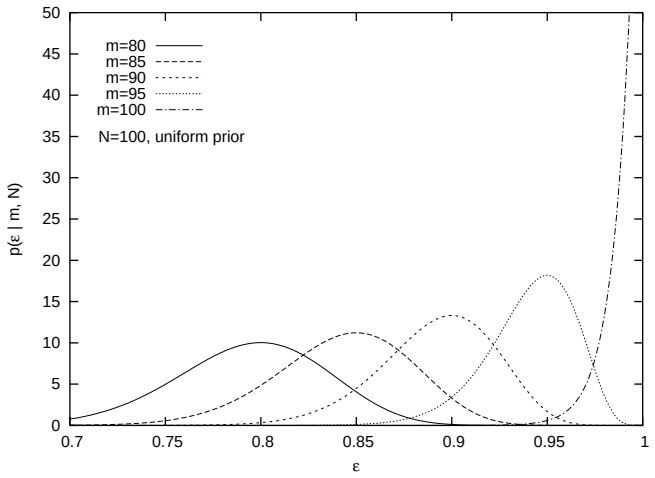

Figure 1: The posterior pdf $p(\varepsilon|m, N)$ obtained using a uniform prior with $N = 100$ events for several values of m.

In the Bayesian approach, all of the information about the parameter ε is encapsulated in the posterior pdf $p(\varepsilon|m, N)$. Rather than reporting the entire curve, however, it can be summarized by giving, for example, the mode (value of the peak). From Bayes' theorem (27) one can see that if the prior $\pi(\varepsilon)$ is a constant, then the posterior, $p(\varepsilon|m, N)$, is directly proportional to the likelihood, $P(m|\varepsilon, N)$, and therefore for this special case the mode of the posterior coincides with the maximum-likelihood estimator. Comparing with equation (3) we therefore have

$$\text{mode}[\varepsilon] = \frac{m}{N} \, . \tag{30}$$

Alternatively one could give the posterior expectation (mean) value of ε,

$$E[\varepsilon] = \int_0^1 \varepsilon p(\varepsilon|m, N) \, d\varepsilon \, . \tag{31}$$

One finds

$$E[\varepsilon] = \frac{B(m+2, N-m+1)}{B(m+1, N-m+1)} \, , \tag{32}$$

where $B(a, b)$ is the Euler Beta function (`TMath::Beta`),

$$B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1} \, dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \, . \tag{33}$$

To quantify the width of the posterior density one can compute its variance (standard deviation squared), which is

$$V[\varepsilon] = E[\varepsilon^2] - (E[\varepsilon])^2 = \frac{B(m+3, N-m+1)}{B(m+1, N-m+1)} - \left(\frac{B(m+2, N-m+1)}{B(m+1, N-m+1)}\right)^2 \, . \tag{34}$$

[1]For even moderately large N and m the gamma functions can give numerical overflow. It is better to compute the logarithm of the combination of gamma functions using, e.g., the routine `TMath::LnGamma` and then exponentiate the result.

Alternatively one can find the shortest interval containing 68.3% of the probability. Software to do this numerically can be obtained from [6]. This is also implemented in the ROOT routine `TGraphAsymmErrors::BayesDivide`. This routine takes the values for the numbers of selected events $m_1, \ldots, m_{N_{\rm bin}}$ and the number of generated events $N_1, \ldots, N_{N_{\rm bin}}$, both in the form of histograms and computes for each bin the corresponding efficiency and shortest 68.3% Bayesian intervals. Note that with `TGraphAsymmErrors::BayesDivide` one implicitly assumes that an event generated in a given bin is either found in that bin or not found at all; migration between bins is assumed not to occur. The routine simply takes the value in bin i of one of the histograms as m_i and in that of the other as N_i, and carries out the Bayesian calculation assuming a binomial model with uniform prior for these quantities.

References

[1] D.R. Cox, *Principles of Statistical Inference*, Cambridge University Press, 2006.

[2] G. Cowan, *Statistical Data Analysis*, Oxford University Press, 1998.

[3] W.-M. Yao et al., Journal of Physics, G 33, 1 (2006); see also `pdg.lbl.gov`.

[4] F.E. James, *Statistical Methods in Experimental Physics*, 2nd ed., World Scientific, 2007.

[5] The ROOT MathCore library, `project-mathlibs.web.cern.ch/project-mathlibs/sw/html/MathCore.html`.

[6] Marc Paterno, *Calculating Efficiencies and their Uncertainties*, 2003; note and related software available from `home.fnal.gov/~paterno/images/effic.pdf`.