Statistics to Identify and Quantify Pure Radionuclides by Their Neutron Emission Patterns

Jacqueline Kiffe and James C. Rock, Department of Nuclear Engineering, Texas A&M University, College Station, TX 77843

July 27, 1999

Abstract

Perfect knowledge of emission time for every neutron from a radionuclide allows identification of the material and estimation of the quantity present via nondestructive assay. In practice, we lack this perfect knowledge. Detector efficiency is less than unity, thermalized neutron arrival is delayed randomly, and neutron showers triggered by cosmic rays in our atmosphere add noise to the process. We compare two classes of statistical estimators capable of recovering characteristic parameters for radionuclides from this imperfect information: method of moments (the current standard) and maximum likelihood, which holds the advantage as the number of parameters to be estimated simultaneously increases.

1 Introduction

Passive neutron counting is frequently used in the nondestructive assay of fissile materials. The current methods of analysis as developed by Böhnle [1] and Hage and Cifarelli [2, 3] and extended by Ensslin, et al. [4] are known as method-of-moments procedures; that is, they depend upon knowing the way various moments depend functionally upon the parameter(s) of interest. Method-of-moments estimators (MOMEs) have two chief advantages. First, they are created in a comprehensible manner. One calculates the moments (either ordinary, central, or factorial) as functions of the population parameters of interest and then solves for those parameters in terms of the moments. Sample moments from the data are then substituted for the theoretical moments to obtain the estimators. Second, when they are available, it is generally in closed form. However, problems arise when one wishes to simultaneously estimate several parameters. The variance of the $k^{th}$ sample moment (the expected value of the random variable to the $k^{th}$ power) is a function of the $2k^{th}$ population moment, so that by the time one progresses to estimating four or five parameters simultaneously, the natural variability of the estimators — regardless of how good the data are — dominates the problem. If the estimators are nonlinear functions of the moments, the problem can be further compounded.
DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, make any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.
DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.
Maximum Likelihood Estimation — A Brief Overview

Maximum likelihood estimation presents numerous advantages over the method of moments for fitting dynamic models (that is, models of processes which change over time) which are mechanistic — as opposed to heuristic — constructs. That is, we assume that the model has been designed explicitly to mathematically mimic physical processes, as opposed to merely attempting to fit data with an arbitrary curve, the parameters of which have no specific physical meaning.

2.1 How is maximum likelihood estimation performed?

The probability function for a distribution is (loosely speaking) the relative likelihood of the occurrence of a specific value of the random variable. The exponential distribution represents a simple example. For an exponential random variable $X$ with rate parameter $\lambda$, we have the probability density function (pdf)

$$f(x|\lambda) = \lambda e^{-\lambda x} \cdot I_{(0,\infty)}(x)$$

where $I_{(a,b)}(x) = 1$ if $x \in (a, b)$ and is zero otherwise. The population parameter $\lambda$ is treated as a fixed constant. The likelihood function is this same function, except that the parameter $\lambda$ becomes the variable quantity and an observed value for $X$ is substituted for $x$. Thus, for this distribution,

$$L(\lambda|X = x_0) = \lambda e^{-\lambda x_0} \cdot I_{(0,\infty)}(x_0),$$

where $x_0$ is some observed value of $X$. The maximum likelihood estimator (MLE) in this simple case is the value of $\lambda$ which maximizes the above function. This can be shown through simple calculus to be $\hat{\lambda} = \frac{1}{x_0}$. This does make intuitive sense: We would not choose a value for $\lambda$ that makes the value of $X$ we observed — $x_0$ — a rare (unlikely) occurrence.

This concept can be extended to multiple observations: The observed values are substituted into the joint distribution function. The parameter value which maximizes this likelihood function is the maximum likelihood estimate. In the above sample, the joint likelihood function for a simple random sample of size $n$ is the product of the individual likelihood functions:

$$L(\lambda|X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} \cdot I_{(0,\infty)}(x_i)$$

We find that the MLE for this example is $\hat{\lambda} = \frac{1}{n}$, which in this case is identical to the method of moments estimator. This is often the case for one-parameter distributions.

We can perform maximum likelihood estimation for $r$ parameters simultaneously by maximizing the likelihood function over $r$-space. In practice, we use the log-likelihood, which is the natural log of the likelihood function, and maximize numerically.
2.2 What are its specific advantages?

- No moments need be available, in closed form or otherwise. This means that we can construct a distribution to echo reality as closely as is feasible, even if it means that the moments of the distribution are not accessible.

- MLEs are not nearly as sensitive to smaller sample sizes as MOMEs. For estimating three or more parameters simultaneously, very large sample sizes are needed with MOMEs.

- If \( \theta = \psi^2 \), then the MLE for \( \theta \) is the square of the MLE for \( \psi \). In general, we can find the maximum likelihood estimates for the most convenient parameterization and then find the MLEs for the desired parameters from those. This is not true for MOMEs.

- Under reasonable regularity conditions, MLEs are consistent. i.e., they converge in probability to the parameter which they are estimating. (If \( \hat{\gamma} \) is a consistent estimator for \( \gamma \), then for any \( \varepsilon > 0 \), the probability that \( \hat{\gamma} \) differs from \( \gamma \) by more than \( \varepsilon \) goes to zero in the limit as sample size increases.)

- Under these same regularity conditions, MLEs are asymptotically efficient; that is, they approach the Cramér-Rao Lower Bound (the absolute lower bound for the variance of estimators in the class containing most reasonable estimators) in the limit as sample size increases.

- No closed-form expression is needed for the estimator. One can maximize the likelihood (or better, the log likelihood) directly, either analytically or numerically. If the derivatives are available in closed form, one may instead solve the resulting set of simultaneous equations, either analytically or numerically.

- Maximum likelihood estimation takes into account all the data, not just the summarization of it included in the means. In the simple example above, the mean was sufficient for \( \phi \), that is, included all of the information about \( \phi \) to be had in the sample. It is fairly easy to construct distributions where it is impossible to sum up the information about a parameter in just one or a few values. In those cases, the entire collection of observations is used in original form in solving for the estimator.


2.3 When might the method of moments be superior?

Statistically, maximum likelihood estimators are virtually always preferable. However, in practice, obtaining an MLE is equal parts statistics, numerical analysis, and programming. The main drawback to maximum likelihood estimation is that it requires excellent numerical algorithms to search for the optimal surface point in hyperspace when one is simultaneously estimating several parameters. If the likelihood function is badly behaved, either very flat over a large expanse or with severely pathological dips and saddle points, finding the best solution can be a challenging combination of numerical analysis and programming. If the problem has
high dimension (four or more parameters being simultaneously estimated) and the likelihood function is complicated, one must take care to ensure that results are correct. Most computer routines equate convergence with failure to improve an estimate, and it requires some examination of the results to ensure that it is not a local maximum which the routine has selected. The algorithms also all require suitable starting values, but this is usually not a problem. If quality software and appropriately powerful computers are not available, closed-form estimators (such as method-of-moments estimators) may be preferable.

3 Simulation Example

A Monte Carlo simulation of a calibration experiment was designed involving a small amount of a pure (theoretical) isotope. The experiment was designed to be illustrative, and thus was kept as simple as possible, in order to be able to calculate the appropriate method-of-moments estimators in closed form. Even with a noise element added (as there is in this experiment), such synthetic experiments give impossibly optimistic results. However, it is the relative performance of the two estimation techniques that we are concerned with, and there is no a priori reason to suspect that the increased problems of real-life data will disproportionately favor one method over another.

The theoretical isotope is assumed to have a neutron emission profile (the distribution of the number of neutrons in a burst) which is distributed as a Poisson($\gamma$) random variable. That is, if $N_b$ represents the number of neutrons emitted during a single fission event (burst), then

$$P(N_b = k) = \frac{\gamma^k e^{-\gamma}}{k!}, k = 0, 1, 2, \ldots \tag{3.5}$$

Fission is assumed to be a Poisson($\lambda_s$) process, where $\lambda_s$ is a function of the mass of isotope present. That is, the number of fission events $N_s$ occurring in a time interval of length $t$ has the distribution

$$P(N_s(t) = k) = \frac{(\lambda_s t)^k e^{-\lambda_s t}}{k!}, k = 0, 1, 2, \ldots \tag{3.6}$$

(Note the distinction between the Poisson random variable in Equation 3.5 and the Poisson process in Equation 3.6.) There is another Poisson process with parameter $\lambda_n$, which represents a noise element. Noise consists of single pulses only, and the number of noise events in an interval of length $t$ has the distribution

$$P(N_n(t) = k) = \frac{(\lambda_n t)^k e^{-\lambda_n t}}{k!}, k = 0, 1, 2, \ldots \tag{3.7}$$

Over all this, an efficiency rate of $\pi$ was applied, where $0 < \pi < 1$. That is, the probability of any single neutron being captured and recorded was $\pi$.

In this model (with not only a theoretical isotope but a theoretical instrument as well), we record the number of detected pulses (which may be either captured neutrons or noise) in each detected burst and the time at which they occur. It can be shown that the detected process (that is, the bursts for which at least one pulse is detected) is also a Poisson process with rate
parameter $\phi$.

$$\phi = (1 - e^{-\pi\gamma})\lambda_s + \pi\lambda_n$$

(3.8)

$$P(N_{dp}(t) = k) = \frac{(\phi t)^k e^{-\phi t}}{k!}, k = 0, 1, 2, \ldots$$

(3.9)

That is, the inter-burst time periods are distributed as an exponential random variable with rate parameter $\phi$. In order to construct the moments and the likelihood function, we must also have the conditional distribution of detected pulses in a burst in which one or more pulses are detected ($N_c$). This is an important, if conceptually difficult, distinction: We must take into account the fact that we have no way to perceive bursts in which zero pulses are detected, which would imply that many of the observed intervals we measure are in reality two or more inter-burst intervals. This can be shown\(^1\) to be given by

$$P(N_c = 1) = \frac{\pi\gamma e^{-\pi\gamma}\lambda_s + \pi\lambda_n}{(1 - e^{-\pi\gamma})\lambda_s + \pi\lambda_n};$$

(3.10)

$$P(N_c = k; k \neq 1) = \frac{(\pi\gamma)^k e^{-\pi\gamma}\lambda_s}{k![(1 - e^{-\pi\gamma})\lambda_s + \pi\lambda_n]}.$$  

(3.11)

In ordinary English, this simple model assumes that we have a point mass and a flat efficiency profile across the detector. Fission events are randomly occurring in time at a rate dependent upon the sample mass. The number of neutrons emitted in a given fission event is dependent upon the isotope, here a pure metal. Noise, which consists only of single pulses (never multiples), occurs randomly in time.

### 3.1 Constructing the Method-of-Moments Estimators

Given these distributions, we can construct the moments and solve. Assume $\phi$ is defined as in Equation 3.8. For the sake of making our equations more readable, define $\bar{t}$ to be the mean interburst time period for a sample and $m_i$ to be the $i^{th}$ sample moment of $N_c$. If $\gamma$ and $\pi$ are assumed known, our estimators are

$$\hat{\lambda}_s = \frac{m_2 - m_1}{\pi^2\gamma\bar{t}};$$

(3.12)

$$\hat{\lambda}_n = \frac{m_1}{\bar{t}} - \frac{m_2 - m_1}{\pi^2\gamma\bar{t}}.$$  

(3.13)

If $\gamma$ is assumed known, our estimators for the remaining three parameters are

$$\hat{\lambda}_s = \frac{(m_2 - m_1)^3}{\bar{t}(m_3 - 3m_2 + 2m_1)^2};$$

(3.14)

$$\hat{\lambda}_n = \gamma \frac{m_2 - m_1}{\bar{t} m_3 - 3m_2 + 2m_1} \times \ldots$$

\(^1\)The derivation, when enough detail is given to make it readable by nonstatisticians, is too lengthy to be suitable for a proceedings paper, and can be obtained directly from the first listed author at j@kiffe.com.
We can also assume that \( \pi \) is known and solve for the remaining three parameters. \( \hat{\lambda}_s \) remains as in Equation 3.14, and for the others we have

\[
\hat{\gamma} = \frac{1}{\pi} \frac{m_3 - 3m_2 + 2m_1}{m_2 - m_1};
\]
\[
\hat{\lambda}_n = \frac{1}{\pi \tilde{t}} \left[ m_1 - \frac{(m_2 - m_1)^2}{m_3 - 3m_2 + 2m_1} \right].
\]

### 3.2 Constructing the Maximum Likelihood Estimators

Define \( p_k \) to be the proportion of observed bursts of size \( k \), and define \( m_1 = \sum_{k=1}^{\infty} k p_k \), the first sample moment of observed burst size. Then it can be shown that the log likelihood of interest is given by

\[
\frac{1}{n} \ell(\lambda_s, \lambda_n, \pi, \gamma) = p_1 \ln \left( \gamma e^{-\pi \gamma} \lambda_s + \lambda_n \right) - (1 - p_1) \pi \gamma
\]
\[
+ (1 - p_1) \ln(\lambda_s) + m_1 \ln(\pi) + (m_1 - p_1) \ln(\gamma)
\]
\[
+ \sum_{k=2}^{\infty} p_k \ln(k!) - \left( (1 - e^{-\pi \gamma}) \lambda_s + \pi \lambda_n \right) \tilde{t}.
\]

The multiplicative constant does not matter, since we wish to maximize this function of the data over the parameter space. We can fix any of the parameters and maximize over the resulting subspace, or solve for all four parameters simultaneously.

### 3.3 Results

One useful single quantity for comparing results across several simulated estimation methods is the relative mean squared error (relative MSE), the ratio of the mean squared error to the square of the true parameter value. The term mean squared error here refers to the mean squared difference between an estimator and the target parameter it is estimating. That quantity is scale-dependent, and dividing by the square of the target parameter removes scale effects. This measure is portable, meaning that it can be used to compare the accuracy of estimates for different parameters in different scales.

In each simulation, the efficiency (\( \tau \)) was set at .5 and \( \gamma \) was fixed at 3.5. Because it is only the relative rates of the signal and noise processes that are critical, \( \lambda_s \) was fixed at 1000 and \( \lambda_n \) ranged between 1 and 1000.

A stream of fission events was generated synthetically. Each data set was fitted a total of four times: Two combinations of three of the parameters (\( \tau, \lambda_s, \) and \( \lambda_n \) or \( \gamma, \lambda_s, \) and \( \lambda_n \)) were estimated each by method of moments or maximum likelihood. This simulation-and-estimation
was repeated 500 times, and statistics on the estimators were calculated. This entire process
was repeated several hundred times for varying values of \( \lambda_n \) and varying sample sizes (run
times). This was done not with the goal of testing any specific theory, but of gaining a feel for
the relative behavior of the two methods under a variety of conditions.

The relative MSE for the MLEs ranged between approximately one-tenth and one-half of
the values for the MOMES for both smaller (Table 1) and larger (Table 2). Neither method
was able to adequately estimate \( \lambda_n \) when \( \lambda_n \leq 10 \). For \( \lambda_n \geq 100 \), the MLE was appropriate.
However, the MOME was unable to estimate \( \lambda_n \) adequately except when \( \lambda_n > 100 \).

To see if the method of moments gave comparable performance when fewer parameters
are being estimated simultaneously (so that fewer moments are required), a series of smaller-
scale simulations (\( n=1,000 \)) were performed. Both \( \pi \) and \( \gamma \) were assumed known, and \( \lambda_s \) and
\( \lambda_n \) were estimated for several values of \( \lambda_n \). The maximum likelihood estimators had smaller
relative MSE across the board. However, as before, the MLE was inadequate to the task of
estimating \( \lambda_n \) for \( \lambda_n \leq 10 \). The method-of-moments estimator performed unacceptably for all
values of \( \lambda_n \) except \( \lambda_n = 1000 \), and even then it performed poorly. (See Table 3.)

Table 1: Simulation results: Three-parameter simultaneous estimation with shorter run time.
The relative MSE for the MLEs ranges between 10% and 52% of that for the MOMES, indicating
superior performance.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Rel. MSE (%)</th>
<th>MLE</th>
<th>MLE MSE/MOME MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MOME</td>
<td>MLE</td>
<td></td>
</tr>
<tr>
<td>( \pi )</td>
<td>.5</td>
<td>42–62</td>
<td>21–29</td>
<td>44–52%</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>3.5</td>
<td>42–62</td>
<td>20–31</td>
<td>64–71%</td>
</tr>
<tr>
<td>( \lambda_s )</td>
<td>1000</td>
<td>121–196</td>
<td>42–79</td>
<td>26–40%</td>
</tr>
<tr>
<td>( \lambda_n )</td>
<td>1</td>
<td>270,000,000–21,000,000</td>
<td>22,000–27,100,000</td>
<td>10–12%</td>
</tr>
<tr>
<td>( \lambda_n )</td>
<td>10</td>
<td>232,000–233,000</td>
<td>21,000–260,000,000</td>
<td>10–12%</td>
</tr>
<tr>
<td>( \lambda_n )</td>
<td>100</td>
<td>2200–22,800</td>
<td>1600–1680</td>
<td>10–12%</td>
</tr>
<tr>
<td>( \lambda_n )</td>
<td>1000</td>
<td>196–361</td>
<td>71–121</td>
<td>35–38%</td>
</tr>
</tbody>
</table>

Table 2: Simulation results: Three-parameter simultaneous estimation with longer run time.
The relative MSE for the MLEs ranges between 8% and 49% of that for the MOMES, main-
taining the relative relationship even for larger sample sizes.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Rel. MSE (%)</th>
<th>MLE</th>
<th>MLE MSE/MOME MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MOME</td>
<td>MLE</td>
<td></td>
</tr>
<tr>
<td>( \pi )</td>
<td>.5</td>
<td>4.4–4.8</td>
<td>2.3</td>
<td>29–49%</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>3.5</td>
<td>4.4–4.8</td>
<td>2.3</td>
<td>37–49%</td>
</tr>
<tr>
<td>( \lambda_s )</td>
<td>1000</td>
<td>12</td>
<td>3.6–4.4</td>
<td>22–38%</td>
</tr>
<tr>
<td>( \lambda_n )</td>
<td>1</td>
<td>21,600,000–21,800,000</td>
<td>1,850,000–1,930,000</td>
<td>8–9%</td>
</tr>
<tr>
<td>( \lambda_n )</td>
<td>10</td>
<td>232,000–233,000</td>
<td>21,000–22,500</td>
<td>10%</td>
</tr>
<tr>
<td>( \lambda_n )</td>
<td>100</td>
<td>2200–2300</td>
<td>400–441</td>
<td>19–20%</td>
</tr>
<tr>
<td>( \lambda_n )</td>
<td>1000</td>
<td>44–11.6</td>
<td>4.4–11.6</td>
<td>27–29%</td>
</tr>
<tr>
<td>Parameter</td>
<td>Value</td>
<td>Rel. MSE (%)</td>
<td>MLE Rel. MSE</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-------</td>
<td>--------------</td>
<td>--------------</td>
<td></td>
</tr>
<tr>
<td>( \lambda_s )</td>
<td>1000</td>
<td>26-77</td>
<td>1.4-4</td>
<td>2-12%</td>
</tr>
<tr>
<td>( \lambda_n )</td>
<td>1</td>
<td>108,000,000</td>
<td>220,000</td>
<td>.2%</td>
</tr>
<tr>
<td>( \lambda_n )</td>
<td>10</td>
<td>1,350,000</td>
<td>24,600</td>
<td>2%</td>
</tr>
<tr>
<td>( \lambda_n )</td>
<td>100</td>
<td>17,000</td>
<td>440</td>
<td>3%</td>
</tr>
<tr>
<td>( \lambda_n )</td>
<td>1000</td>
<td>361</td>
<td>1.4</td>
<td>.4%</td>
</tr>
</tbody>
</table>

Table 3: Simulation results: Two-parameter simultaneous estimation with short run times. The relative MSE for the MLEs range between less than 1% and 12% of that for the MOMES, making them more desirable even for simultaneous estimation of fewer parameters.

4 Conclusion

Maximum likelihood estimation holds a number of advantages over the method of moments, especially as the number of parameters to be estimated and the complexity of the distribution function increase. The power of available computing resources has increased greatly since the first method-of-moments estimators were introduced to deal with neutron multiplicity problems, increasing the potential usefulness of the maximum likelihood methods in this type of problem.

References


This report was prepared with the support of the US Department of Energy (DOE) Cooperative Agreement No. DE-FC04-5AL85832. Opinions, findings, conclusions or recommendations expressed herein are those of the authors and do not necessarily reflect the views of the DOE. This work was conducted through the Amarillo National Resource Center for Plutonium.