LA-UR-18-28764
Approved for public release; distribution is unlimited.

Title: Statistical Tests for Convergence in Monte Carlo Criticality Calculations

Author(s): Brown, Forrest B.
Martin, William Russell

Intended for: MCNP documentation

Issued: 2018-09-14

Statistical Tests for Convergence in Monte Carlo Criticality Calculations

Forrest B. Brown[1], William R. Martin[2]
[1] Monte Carlo Methods, Codes, & Applications Group, LANL
[2] NERS Department, University of Michigan

1.0 Introduction

This report describes methods for using statistical tests to automatically determine convergence of Monte Carlo (MC) iterations in criticality calculations. The methods described herein are intended to replace the traditional approach used for the past 60 years: make a trial run; determine convergence based on plots of k-effective and Shannon entropy vs cycle; adjust the input parameters for controlling the iterations; make a final run to obtain results and statistics.

The sections that follow provide background information on the MC iteration methods, describe 3 statistical tests on scalar metrics, and describe 3 statistical tests on fission distributions. The combination of all 6 statistical tests is used in mcnp6.2.1 to automatically diagnose convergence of the iterations to the stationary-state and begin the tallies for active cycles. The code requires all 6 statistical tests to pass before diagnosing convergence. What was previously left to user judgment (with varying degrees of success) can now be supplanted by a robust, "guaranteed" assurance with quantifiable evidence that a problem has converged.

The automatic diagnosis of convergence is part of a larger effort to thoroughly overhaul the calculational methods for criticality problems, providing for automated sampling of the initial fission distribution, acceleration of the iteration convergence process, automated detection of convergence in the iterations and starting of the active tally cycles, and the diagnosis of undersampling. All of these new methods have been prototyped in a local modified version of mcnp6.2 [1] and tested on a variety of critical systems.

2.0 Background on MC Criticality Calculations [2]

For the past 60 years, MC criticality calculations for k-effective and the fission distribution have been solved using the power method, also called the method of successive iterations [2-6].

$$\vec{S}^{(n)} = \frac{1}{k^{(n-1)}} \cdot \bar{F} \cdot \vec{S}^{(n-1)}, \qquad n = 1,2,\ldots$$

where $k^{(n)}$ and $\vec{S}^{(n)}$ are the k-effective eigenvalue and fission neutron distribution after *n* iterations, and $\bar{F}$ represents the MC random walk process for a single fission generation.

The MC iteration scheme begins with an initial guess for k-effective and the fission distribution. Iterations (called inactive cycles) are performed without tallies until k-effective and the fission distribution have converged to their stationary state. After convergence, tallies are turned on, and iterations (called active cycles) are continued until sufficiently small uncertainties are obtained for desired results.

Some straightforward analysis of the MC power iteration method yields the following behavior for the single-cycle k-effective and fission distribution [2] as a function of iteration, *n*:

$$\vec{S}^{(n)} \;=\; \vec{S}_0 \;+\; a_1 \rho^n \vec{S}_1 \;+\; \ldots$$

$$k^{(n)} = k_0 \cdot [\,1 \;+\; b_1 \rho^{n-1}(1-\rho) \;+\; \ldots]$$

where the $\vec{S}_j$ are eigenfunctions of the operator $\bar{F}$, k_j are the corresponding eigenvalues, terms a_1 and b_1 are constants related to an expansion of the initial fission source guess $\vec{S}^{(0)}$ in the basis $\vec{S}_j$, and ρ is the dominance ratio k_1/k_0 (which is always < 1). Only first-order terms are retained above. On each successive iteration, the higher-mode (noise) contributions to $\vec{S}^{(n)}$ and $k^{(n)}$ are reduced. After some number of iterations *n*, $\left\|\vec{S}^{(n)} - \vec{S}_0\right\|$ and $\left|k^{(n)} - k_0\right|$ both become small enough compared to statistical noise that the iteration process is said to have converged to the stationary state. At that point, tallies are started and iterations continue until sufficiently small statistics on tallies are attained. Figure 1 illustrates the k-effective results during the power iteration process, along with concerns for the different phases of the MC power iteration process [2].

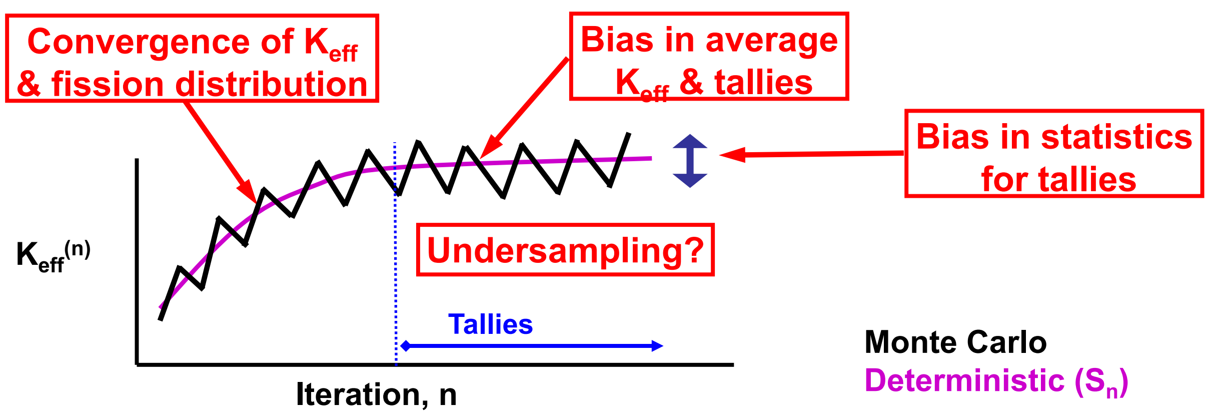

Figure 1. Concerns in MC Criticality Calculations

During the past 60 years of MC criticality calculations, determining convergence of the iteration process has been problematic. Until 2003, the principal means of assessing convergence was to make a trial run, examine the plot of k-effective vs cycle (similar to Figure 1), and determine the first cycle where k-effective appears to have reached its asymptotic value, considering the statistical fluctuations. Unfortunately, that procedure is flawed, since possibly many more iterations are required to converge the shape of the fission distribution. (As discussed in [9], k-effective is an integral quantity and appears to converge in fewer iterations than the fission distribution.)

In 2003, the metric called Shannon entropy, *H*, of the fission source distribution was introduced into mcnp5 [7-9]. *H* is computed at the end of each iteration and plotted vs iteration; when *H* has converged, the fission distribution has converged and active cycles can begin. (k-effective converges in fewer iterations, as discussed in [9].) To compute *H*, a grid covering all fissile regions is superimposed on the problem, and the number of source neutrons from fission in cycle *n* are counted in each of the grid regions. After normalization to probabilities, $p_i^{(n)}$, $H^{(n)}$ is computed as

$$H^{(n)} = -\sum_{i=1}^{B} p_i^{(n)} \log_2 p_i^{(n)}$$

where B is the number of bins in the mesh, and $p \log_2 p = 0$ when $p=0$. $H^{(n)}$ is a convex function, with a minimum value of 0 (for all neutrons in the same bin) and a maximum value of $\log_2 B$ (for neutrons uniformly distributed among bins).

3.0 Reference Solution for Statistical Tests on the Fission Neutron Distribution

In the following sections, 3 statistical tests involving metrics are described, and then 3 statistical tests involving the fission neutron distribution are described. For statistical tests on the fission neutron distribution, it is presumed that an independent, accurate reference solution for the fission source distribution is available. For mcnp6.2.1, that independent estimate of the fission source distribution is provided by the sparse-storage fission matrix method.

The fission matrix method using sparse-storage was previously described in detail in references [10, 11], and an update on the current status is in preparation [12]. References [10,11] have shown that the fission matrix method provides an accurate solution to the k-eigenvalue criticality problem if the mesh spacing is fine enough that the “flat-source” approximation is valid. Importantly, it is not subject to the source renormalization bias that can significantly affect the neutron distribution if the number of neutrons/cycle is too small [2]. Reference [12] provides details on the current status, including the physics-based method for assuring that the mesh is chosen appropriately – fine enough to provide accuracy, but not so fine that statistical tallies for the matrix elements introduce too much noise.

For the present work, we assume that the fission matrix eigenfunction is a nearly “exact” representation of the stationary-state fission neutron distribution. The fission matrix is accumulated over all cycles (except the first), while the fission neutron distribution is not accumulated during inactive cycles. The eigenfunction for the accumulated fission matrix is approximate at first, but converges to the stationary-state distribution well before the single-cycle fission neutron distribution converges.

In mcnp6.2.1, tallies for the fission matrix are not made during the first cycle, since the neutrons starting cycle 1 have energies sampled from an assumed fission spectrum and may or may not have started in valid fissile regions. In addition, physics information is obtained during cycle 1 that is used to set an appropriate mesh spacing for accurate fission matrix tallies. Tallies of the fission matrix are accumulated starting with cycle 2. Since the fission matrix represents a collection of point-to-point Green’s functions, it may be tallied directly, even if the fission neutron distribution has not converged.

4.0 Metric-Based Statistical Tests for Convergence

Based on the above discussion, the following are proposed as tests for convergence in MCNP criticality problems. Beginning with cycle *n=2*, each block of *L* consecutive cycles is examined. Currently, *L=10* is recommended.

At the end of each block of *L* cycles:

> **Solve the fission matrix equations for the fundamental mode eigenvalue and eigenfunction. Then compute H_{FM}, the Shannon entropy of the fission matrix eigenfunction.**
>
> **Test 1 – H-slope:**
> Compute the least-squares slope of $H^{(n)}$ vs *n* for the cycles in the block, and compute the standard deviation of that slope. If the magnitude of the slope is less than the standard deviation of the slope, then the slope is 0.0 within statistics, and the test is passed. Otherwise, the test fails.
>
> **Test 2 – k-slope:**
> Compute the least-squares slope of $k^{(n)}$ vs *n* for the cycles in the block, and compute the standard deviation of that slope. If the magnitude of the slope is less than two standard deviations of the slope, then the slope is 0.0 within statistics, and the test is passed. Otherwise, the test fails.
>
> **Test 3 – H-block:**
> For the fission neutron distribution accumulated over all cycles in the block, compute H_{block} using the same mesh as used for the fission matrix. If $|(H_{block} - H_{FM})/H_{FM}| < 1\%$, the test passes. Otherwise, the test fails.
>
> **If any of the 3 tests fails, then the fission neutron distribution has not converged. All results from the current block of cycles are discarded, and a new block of cycles is begun.**

Some comments follow:

- Larger block sizes (*L>10*) would make the individual tests more robust and reliable, at the expense of less-frequent convergence checking and possibly a number of additional cycles. That is, choosing too large a value for *L* will delay the testing, since it is performed only at the end of a block of *L* cycles, and some computer time may be wasted. Smaller block sizes (e.g., *L=5*) lead to increased statistical noise for the testing. In nearly all testing to date on a variety of problems, choosing *L=10* has proved reliable.
- For a block of *L* cycles, the H-slope and k-slope tests are rather obvious tests for stationarity. The H-slope test uses statistics of $\pm 1\sigma$, while the k-slope test uses $\pm 2\sigma$. The difference is simply practical – the variation in $H^{(n)}$ vs *n* is smaller and smoother than the variation in $k^{(n)}$ over a block of successive cycles. A more stringent test for zero-slope is used for the H-slope.

- For the H-block test, note that the fission neutron source for all cycles in the block is accumulated first, and then the Shannon entropy is computed once for the block. (The average of the $H^{(n)}$ values for cycles in the block may differ significantly from H_{block} if undersampling or clustering is present [13].)
- For the H-block test, the tolerance of 1% is arbitrary, but has proved suitable in all testing to date. Note that using a smaller tolerance may result in never passing the test, due to possible renormalization bias in the fission neutron source distribution if too few neutrons/cycle are used. The use of a 1% tolerance appears to be consistent with the strong recommendation to use 10,000 or more neutrons/cycle (or 100,000 for reactors, large solution tanks, and loosely-coupled systems) [14].
- For the H-block test, the same 3D mesh must be used for the fission matrix tallies and the computation of H_{block}.

5.0 Distribution-based Statistical Tests for Convergence

These statistical tests involve comparing the fission neutron distribution to a reference distribution (the eigenfunction of the fission matrix). Beginning with cycle $n=2$, each block of L consecutive cycles is examined. Currently, $L=10$ is recommended.

At the end of each block of *L* cycles:

Solve the fission matrix equations for the fundamental mode eigenfunction, $\vec{S}_{FM}$.

Accumulate the fission neutron source distribution for all cycles in the block, binned using the same 3D mesh as for the fission matrix, $\vec{S}_{neut}$

Test 4 – KL-test:

Compute the Kullback-Liebler divergence [15] between $\vec{S}_{FM}$ and $\vec{S}_{neut}$, $D_{KL}(\vec{S}_{neut} | \vec{S}_{FM})$. If D_{KL} is less than the critical value for 95% confidence, then the test passes. Otherwise the test fails.

Test 5 – KS-test:

Compute the Kolmogorov-Smirnov statistic [16] for the paired distributions $\vec{S}_{FM}$ and $\vec{S}_{neut}$, $D_{KS}(\vec{S}_{neut}, \vec{S}_{FM})$. If D_{KS} is less than the critical value for 95% confidence, then the test passes. Otherwise the test fails.

Test 6 – ChiSq-test:

Compute the chi-square statistic [17] for goodness of fit between $\vec{S}_{FM}$ and $\vec{S}_{neut}$, χ^2. If χ^2 is less than the critical value for 95% confidence, then the test passes. Otherwise the test fails.

If any of the 3 tests fails, then the fission neutron distribution has not converged. All results from the current block of cycles are discarded, and a new block of cycles is begun.

Some comments follow:

- The blocksize *L* for these tests is the same as for the metric-based tests, and all of the same comments apply here. The default blocksize of *L=10* appears to be a good choice.
- The Kullback-Liebler divergence, also called relative entropy, is a metric for comparing distributions and is computed by:

$$\boldsymbol{D_{KL}}\left(\vec{\boldsymbol{S}}_{\boldsymbol{neut}} \middle| \vec{\boldsymbol{S}}_{\boldsymbol{FM}} \right) = \sum_{i=1}^{B} \boldsymbol{S_{neut,i}} \cdot \log_2\left(\frac{\boldsymbol{S_{neut,i}}}{\boldsymbol{S_{FM,i}}}\right)$$

  where *B* is the number of bins in the mesh, and the terms in the summation are included only if $\boldsymbol{S_{neut,i}} > \boldsymbol{0}$ and $\boldsymbol{S_{neut,i}} > \boldsymbol{0}$.

- The Kullback-Liebler divergence is found in the information theory literature, and is not associated in that literature with statistical testing. However, in the statistics literature there is a statistical test for comparing 2 distributions called the “G test,” also known as the likelihood ratio test [18]. The G test statistic is related to D_{KL} by: $\boldsymbol{G} = \boldsymbol{2N} \cdot \boldsymbol{D_{KL}}\left(\vec{\boldsymbol{S}}_{\boldsymbol{neut}} \middle| \vec{\boldsymbol{S}}_{\boldsymbol{FM}} \right)$, where *N* is the number of neutrons/cycle used in computing $\vec{\boldsymbol{S}}_{\boldsymbol{neut}}$. Accordingly, the critical value for G at the 95% confidence limit may be found, and then “converted” to a corresponding target test statistic for D_{KL} by $G_{0.05}/2N$.
- For large *N*, the $G_{0.05}$ statistic for 95% confidence approaches the $\chi^2_{0.05}$ value. For the applications here, where *N* is always in the range $10^4 – 10^8$, use of the chi-square statistic (in place of the G value) is entirely justified.
- The KS-test strictly applies to comparing 1D distributions. For the 3D distributions in criticality problems, the proper ordering of the paired entries in $\vec{\boldsymbol{S}}_{\boldsymbol{FM}}$ and $\vec{\boldsymbol{S}}_{\boldsymbol{neut}}$ is handled pragmatically: Initially, the pairs of entries in $\vec{\boldsymbol{S}}_{\boldsymbol{FM}}$ and $\vec{\boldsymbol{S}}_{\boldsymbol{neut}}$ are sorted such that $\vec{\boldsymbol{S}}_{\boldsymbol{FM}}$ values are in increasing order (as for the 1D KS-test), and then D_{KS} is computed. Then the entries are randomly permuted and D_{KS} is determined again. The permute/recompute process is repeated a number of times, and the maximum of all the D_{KS} values is retained. In very many cases, the first D_{KS} computed with increasing ordering of $\vec{\boldsymbol{S}}_{\boldsymbol{FM}}$ has been the maximum, but that is not guaranteed.
- The critical value for D_{KS} at the 95% confidence level is computed for large degrees-of-freedom by $D_{KS,0.05} = \sqrt{-\frac{1}{2} \cdot ln\left(\frac{0.05}{2}\right) \cdot \frac{2}{\nu}}$, where ν=*B-1* and *B* is the number of nonzero bins in $\vec{\boldsymbol{S}}_{\boldsymbol{FM}}$.
- The critical value $\chi^2_{0.05}$ for the chi-square test for large degrees of freedom is obtained from the critical value for a normal distribution by [19]

$$\chi^2_{0.05} \approx \nu \cdot \left(1 - \frac{2}{9\nu} + \sqrt{\frac{2}{9\nu}} \cdot z_{0.05} \right)^3$$

- For all of the distribution-based statistical tests, there is a concern regarding the mesh spacing used for computing $\vec{\boldsymbol{S}}_{\boldsymbol{FM}}$ and $\vec{\boldsymbol{S}}_{\boldsymbol{neut}}$. If the mesh spacing is too fine, then there may not be sufficient counts of the fission neutrons in the mesh bins for the statistical tests. (Bin counts of

5 or more in each bin are typically recommended for the distribution-based tests.) For the present application, the first concern in choosing the mesh spacing is that it be sufficiently fine to permit an accurate solution of the fission matrix equations. There must be a balance, however, between the fission matrix solution accuracy and the statistical tests, such that the mesh is not refined more than necessary. These issues are discussed in Reference [12]. For the present applications, it is reasonable to increase the number of neutrons/cycle to a value that is 10-15 times larger than the number of mesh bins that contain fissile material, in order to provide sufficient counts for the statistical tests. At present, that condition is not enforced, but is under consideration for future improvements.

6.0 Conclusions and Further Work

The 6 statistical tests described in Sections 4 and 5 are performed at the end of each block of cycles in an MCNP criticality problem, as implemented in a local version of MCNP6.2. If all 6 tests pass, then convergence of the iteration process is asserted and “locked in” for the duration of the calculation. Active cycles and tallies begin with the following cycle. During subsequent active cycles, some of the convergence tests may occasionally fail (after all, statistical fluctuations happen), but will pass again on later cycles. Once an entire block of active cycles is complete, the population size testing described in [13] is performed to diagnose whether the number of neutrons/cycle is sufficient to avoid undersampling of the fission neutron source.

These convergence tests have been applied to an assortment of criticality problems, including: a 2D PWR model, a 3D model of the Sandia ACRR burst reactor (with FREC), a 3D model of the ATR (advanced test reactor), the 3D Kord Smith Challenge problem (OECD-NEA 3D reactor computer-performance benchmark), the August Winkelman research reactor, the ICSBEP benchmark case LEU-COMP-THERM-078 (a Sandia experiment), a large 3D storage pool with checkerboard arrangement (OECD-NEA EG on source convergence benchmark), a 400 cm tall single reactor fuel-pin unit cell with reflecting boundary conditions, and the Whitesides problem (k-effective of the world).

In all cases tested, there was not a single instance of passing the convergence tests prematurely (before convergence was actually achieved). That is, the combination of the 6 statistical tests for convergence did not yield any “false positive” results for convergence.

The other, less important, concern is whether the statistical testing was too demanding, requiring many extra cycles and computer time. In some of cases tested, 1 or 2 extra blocks of cycles were needed beyond traditional estimates of convergence (i.e., eyeball the H vs cycle plots to determine when the asymptotic level is reached). Since the statistical tests examine more quantities in more detail than the conventional method, there is no definitive evidence of “false negatives.” That is, in the cases where an extra 1 or 2 blocks of cycles were needed beyond conventional estimates, it is likely that the conventional estimates were overly optimistic.

To summarize, the combination of the 6 statistical tests for convergence appears to be robust, reliable, and a definitive means of automatically determining convergence. Quantified evidence of convergence is provided by the statistical metrics that are calculated and printed.

Obviously further work is needed, including:

- Apply the tests to many more criticality problems. Many more ordinary cases as well as odd-ball cases need to be tested.
- A full report on the testing process will be completed.
- For some problems – large reactors, fuel storage pools, very loosely-coupled problems – there are concerns about how many cycles are needed to get an accurate representation of the fission matrix, especially when the number of neutrons/cycle is small. These concerns are being investigated, and it is likely that some additional requirements will be placed on solving the fission matrix equations. That issue is separate from the statistical testing for convergence.
- For some problems – large reactors, fuel storage pools, very loosely-coupled problems – there are concerns about how many neutrons/cycle are needed to provide sufficient counts in the mesh bins so that the distribution-based statistical tests are reliable. These concerns are also being investigated, and it is likely that some additional requirements may be imposed on neutrons/cycle relative to the number of mesh bins.
- While the use of 6 statistical tests for convergence has been robust and reliable, it is possible that additional tests will be included.

Acknowledgements

This work was supported by the US-DOE-NNSA Nuclear Criticality Safety Program and by the US-DOE-NNSA Advanced Simulation & Computing Program.

References

1 C.J. Werner, et al., "MCNP6.2 Release Notes", Los Alamos National Laboratory, report LA-UR-18-20808 (2018)

2 F.B. Brown, “Monte Carlo Techniques for Nuclear Systems”, LA-UR-16-29043 (2016).

3 W. Goad and R. Johnston, "A Monte Carlo Method for Criticality Problems," *Nucl. Sci. Eng.* 5, 371-375 (1959).

4 J. Lieberoth, "A Monte Carlo Technique to Solve the Static Eigenvalue Problem of the Boltzmann Transport Equation," *Nukleonik* 11,213 (1968).

5 H. Rief and H. Kschwendt, "Reactor Analysis by Monte Carlo," *Nucl. Sci. Eng.*, 30, 395 (1967).

6 M. R. Mendelson, "Monte Carlo Criticality Calculations for Thermal Reactors," *Nucl. Sci Eng.* 32, 319-331 (1968).

7 T. Ueki, F.B. Brown, “Stationarity and Source Convergence in Monte Carlo Criticality Calculations”, ANS Topical Meeting on Mathematics & Computation, Gatlinburg, TN April 6-11, 2003 LA-UR-02-6228 (September, 2002).

8 T. Ueki, F.B. Brown, “Stationarity Modeling and Information-Based Diagnostics in Monte Carlo Criticality Calculations”, *Nucl. Sci. Eng.*, **149**, 38-50 (2005).

9 F.B. Brown, "On the Use of Shannon Entropy of the Fission Distribution for Assessing Convergence of Monte Carlo Criticality Calculations", PHYSOR-2006, Vancouver, LA-UR-06-3737 (2006).

10 F.B. Brown, S.E. Carney, B.C. Kiedrowski, W.R. Martin, "Fission Matrix Capability for MCNP, Part I - Theory", Mathematics & Computation 2013, Sun Valley, ID, LA-UR-13-20429 (2013)

11 S.E. Carney, F.B. Brown, B.C. Kiedrowski, W.R. Martin, "Fission Matrix Capability for MCNP, Part II - Applications", Mathematics & Computation 2013, Sun Valley, ID, report LA-UR-13-20454 (2013)

12 F.B. Brown, “Current Status of the Fission Matrix Method in MCNP6”, in preparation (2018)

13 F.B. Brown, C.J. Josey, “Diagnostics for Undersampling and Clustering in Monte Carlo Criticality Calculations”, LA-UR-18-27656 (2018).

14 F.B. Brown, “A Review of Best Practices for Monte Carlo Criticality Calculations”, ANS NCSD 2009, Hanford WA, LA-UR-09-03136 (2009)

15 T.M. Cover, J.A. Thomas, Elements of Information Theory, 2nd Edition, Wiley (2006)

16 F.J. Massey, “The Kolmogorov-Smirnov Test for Goodness of Fit”, *J. American Statistical Assoc.* 46:253, 68-78 (1951)

17 P.R. Bevington, D.K. Robinson, Data Reduction and Error Analysis for the Physical Sciences, 2nd Edition, McGraw-Hill (1992)

18 J.H. McDonald, “G-test of Goodness-of-Fit”, Handbook of Biological Statistics, 3rd Edition, Sparky House Pub. (2014)

19 M. Evans, N. Hastings, B. Peacock, Statistical Distributions, 2nd Edition, Wiley (1975)