Joint International Topical Meeting on Mathematics & Computation and Supercomputing in Nuclear Applications (M&C + SNA 2007)
Monterey, California, April 15-19, 2007, on CD-ROM, American Nuclear Society, LaGrange Park, IL (2007)

PARTICLE POPULATION DIAGNOSTICS IN ITERATED-SOURCE MONTE CARLO METHODS

Taro Ueki
The University of New Mexico, Department of Chemical and Nuclear Engineering
209 Farris Engineering Center, Albuquerque, NM 87131
tueki@unm.edu

ABSTRACT

On-the-fly diagnostics of the number of particles per iteration cycle in iterated-source Monte Carlo computation, which output diagnostic measures for a given spatial resolution of binning cells as iteration cycles progress, have been studied using relative entropy and chi-square distance. A source ratio vector is defined whose components are the ratio of the sources of adjacent iteration cycles at the individual binning cells. This enables one to define a problem-independent reference vector based on the integral equation representation of the static eigenvalue problem of particle multiplication. All vectors involved are normalized so that they represent discrete probability distribution. The weighted difference between the relative entropy and chi-square distance of the source ratio vector, both with respect to the reference vector, has been shown to be an effective measure of particle population. Numerical examples are presented for the initial core of a pressurized water reactor and the k-effective-of-the-world problem.

Key Words: Monte Carlo, iterated source, particle population, relative entropy, chi-square distance

1. INTRODUCTION

Recently, significant amount of work has been done on the convergence and stationarity diagnostics of source distribution in iterated-source Monte Carlo (MC) computation [1-4]. In each of these work, well-defined diagnostics were proposed and their performance is satisfactory on sound technical grounds. It is fair to say that their refinement appears to be an avenue that MC code developers can pursue concerning the reliability improvement of MC criticality/reactor calculation. On the other hand, the diagnostics of particle population, i.e., the number of particles per iteration cycle, have rarely been investigated except recent information theoretic work of this author [5]. Although the methodologies therein are a collection of well-defined diagnostics based on data compression, the asymptotic equipartition property and the concavity of Shannon entropy in information theory [6], their implementation is restricted to posterior checking as in the case of normality checking in some production MC code [7]. In other words, these information-theoretic diagnostics can be applied after all or significant portions of required iteration cycles were finished. Therefore, the on-the-fly diagnostics of particle population is strongly desired to prevent MC criticality/reactor analysis codes from producing potentially biased tallies in addition to the code's capability of rejecting such tallies in a posterior manner. Here, "on-the-fly" implies "application together with the progression of iteration cycles".

It is worth discussing if the normality test established in the statistics discipline by Shapiro and Wilk [8] and practiced in the nuclear engineering discipline by Burrows and MacMillan [9] can be extended to the diagnostics of particle population. A good example problem for this discussion is the k-effective-of-the-world problem in Fig. 1, which was originally proposed by Whitesides [10] and extensively investigated in the aforementioned information theoretic work [5]. Fig. 2 shows the 95% confidence interval of effective neutron multiplication factor (k_{eff}) for various numbers of particles per iteration cycle. It is clear that 800 particles per iteration cycle are too small to guarantee the estimation of k_{eff} free of bias. Table I summarizes the results of normality tests for the computation with 800 particles per iteration cycle. One can see that none of ten independent replicas of the computation led to the rejection of normality. Therefore, the normality test based on Shapiro and Wilk would not be suitable for particle population diagnostics even in a posterior sense.

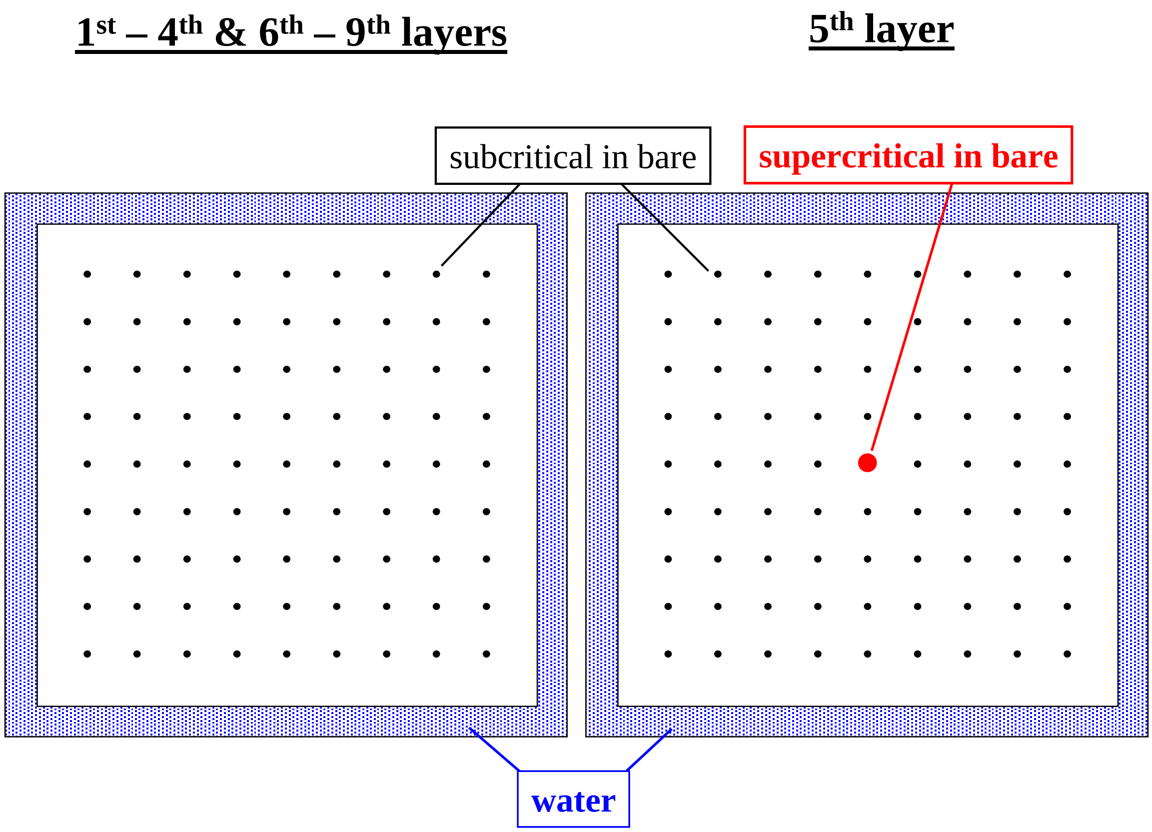

Figure 1: The k-effective-of-the-world problem (not scaled; 9 by 9 by 9 array of ^{239}Pu sphere; space between spheres is void)

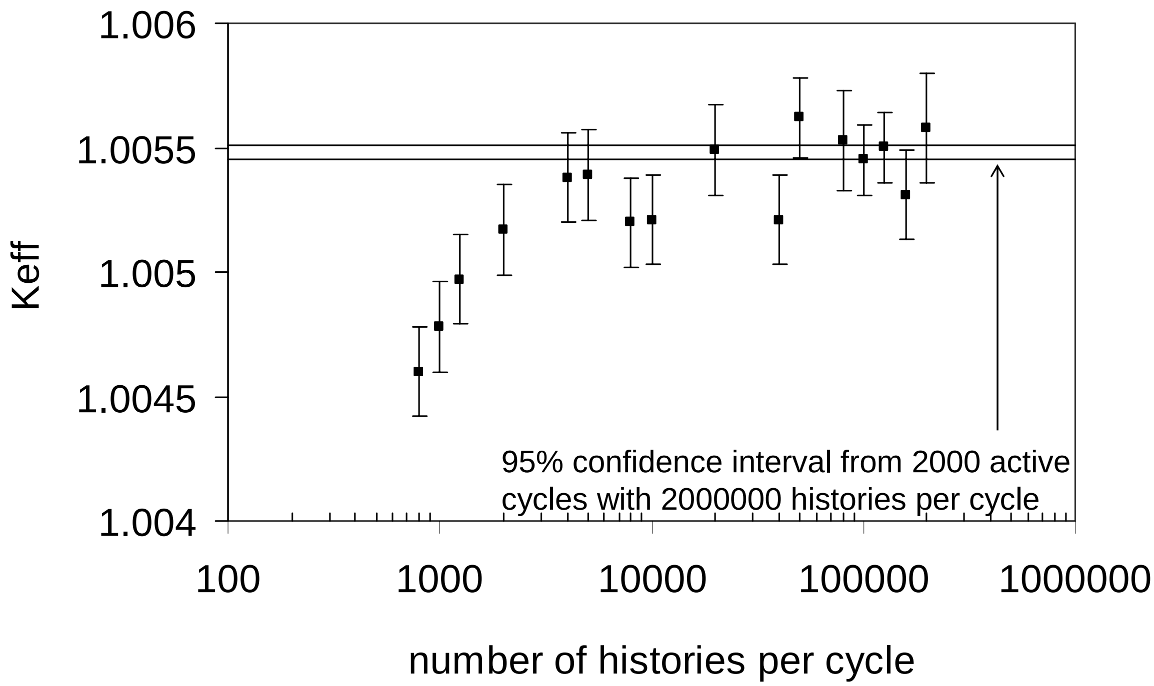

Figure 2: Confidence interval (95%) of the k-effective-of-the-world problem from the 200,000,000 histories through active cycles (variance computed via batch grouping (group average) of 50 cycles; taken from Ref. [5])

Table I: Number of rejecting the normality of a series of k-effective's (k_{eff}'s) in the k-effective-of-the-world problem; ten independent replicas of the 6400 active iteration cycles with 800 particles per iteration cycle preceded by the 100 inactive iteration cycles with the same number of particles per iteration cycle; the series of k_{eff}'s taken from the active iteration cycles

	estimator	number of cycles per batch (number of cycle-wise k_{eff}'s in group average)		
		1	8	20
number of rejecting normality	collision	**0**	**0**	**0**
	absorption	**0**	**0**	**0**
	track length	**0**	**0**	**0**

This work investigates the use of a source ratio vector for the on-the-fly diagnostics of particle population. Here, the source ratio vector is defined whose components are the ratio of the sources of adjacent iteration cycles at the individual binning cells. The integral equation representation of the static eigenvalue problem of particle multiplication requires that the source ratio vector must be parallel to a vector of the form $(1,1,\ldots,1)$. This is a desirable feature because the vector $(1,1,\ldots,1)$ is problem-independent. In this work, all vectors involved are normalized so that they represent discrete probability distribution, i.e., the sum of the components is unity. This enables one to characterize the descriptive difference between the source ratio vector and the vector $(1,1,\ldots,1)$ in terms of relative entropy and chi-square distance.

2. SOURCE RATIO VECTOR

In static analysis of particle (neutron) multiplication, MC iterated source methods are used to solve an integral equation of the form

$$S(\vec{r}) = \frac{1}{k}\int_{domain} F(\vec{r}\,' \to \vec{r})S(\vec{r}\,')dV' \tag{1}$$

where $\vec{r}$ stands for space coordinate vector, dV volume element, $S(\vec{r})$ source distribution, k eigenvalue, and $F(\vec{r}\,' \to \vec{r})$ an integral kernel that can be interpreted as the expected number of direct descendent source particles per unit volume at $\vec{r}$ resulting from a source particle at $\vec{r}\,'$. Suppose that one integrates Eq. (1) over part of the domain, say, cell *p*. Then, after rearranging, one obtains

$$\frac{\int_{cell\ j}\left[\int_{domain} F(\vec{r}\,' \to \vec{r})S(\vec{r}\,')dV'\right]dV}{\int_{cell\ j} S(\vec{r})dV} = k\,. \tag{2}$$

Taking into account the physical meaning of $F(\vec{r}\,' \to \vec{r})$ within the context of iterated-source MC computation, Eq. (2) can be interpreted as

$$\begin{aligned}&\frac{\text{number of source particles at next cycle in cell } j}{\text{number of source particles at current cycle in cell } j}\\ &= \text{constant through cells and stationary iteration cycles}\end{aligned} \quad . \tag{3}$$

Therefore, if one defines the source ratio vector of iteration cycle *p*, denoted $\mathbf{SR}(p)$, to be

$$\mathbf{SR}(p) \equiv \left(\frac{S(p+1,1)}{S(p,1)}, \frac{S(p+1,2)}{S(p,2)}, \cdots, \frac{S(p+1,B)}{S(p,B)} \right) \tag{4}$$

where

$$S(p,j) = \text{Monte Carlo estimate of } \int_{cell\ j} S(\vec{r})dV \text{ at iteration cycle } p\,, \tag{5}$$

and *B* is the total number of binning cells, then one should have

$$\mathbf{SR}(p) \,//\, (1,1,\ldots,1)\,. \tag{6}$$

3. RELATIVE ENTROPY AND CHI-SQUARE STATISTIC

Let $S^B(j)$ and $T^B(j)$ be discrete distribution normalized to unity, where $j = 1, \dots, B$ with B corresponding to the total number of binning cells as in the definition of $\mathbf{SR}(p)$ in (4). The relative entropy of S^B with respect to T^B is defined as

$$D(S^B \,\|\, T^B) \equiv \sum_{j=1}^{B} S^B(j) \log_2 \left(\frac{S^B(j)}{T^B(j)} \right) \tag{7}$$

The relative entropy is non-negative, zero if and only if $S^B = T^B$, and can be interpreted as the penalty measured in bits of the binary description of the output/observed distribution S^B assuming that the true/ideal distribution is T^B [6]. Let the normalization of $(1, 1, \dots, 1)$ and $\mathbf{SR}(p)$ be

$$\mathbf{R}^B \equiv \frac{1}{B} \overbrace{(1, 1, \dots, 1)}^{B \text{ components}} \tag{8}$$

and

$$\mathbf{SR}(p) \leftarrow \frac{\left(\frac{S(p+1,1)}{S(p,1)}, \frac{S(p+1,2)}{S(p,2)}, \dots, \frac{S(p+1,B)}{S(p,B)} \right)}{\sum_{j=1}^{B} \frac{S(p+1,j)}{S(p,j)}} \tag{9}$$

where $\leftarrow$ denotes assignment. Ideally, $D(\mathbf{SR}(p) \,\|\, \mathbf{R}^B)$ should be zero due to the requirement in (6) with the original definition of $\mathbf{SR}(p)$ in (4) and the normalization in (8) and (9). Therefore, one has to argue an acceptable level of $D(\mathbf{SR}(p) \,\|\, \mathbf{R}^B)$.

In information theory [6], the binary encoding for the description of discrete probability distribution is not performed without restriction. It is governed by prefix-free encoding where none of the binary encodings of discrete values is a prefix of the binary encoding of any other discrete value. Thus, a sequence of encoded discrete variables is instantaneously and uniquely decodable. Suppose that one performs sampling from $\mathbf{SR}(p)$ and encodes each sampled value in a prefix-free manner. Then, the average of the lengths of encoded values is bounded below by the Shannon entropy of $\mathbf{SR}(p)$. Also, there always exists a prefix-free encoding scheme which yields the average of the lengths of encoded values bounded above by the Shannon entropy of $\mathbf{SR}(p)$ plus unity. Suppose further that one pretends that the above values are obtained via sampling from $\mathbf{R}^B$; one encodes based on an optimum scheme dictated by $\mathbf{R}^B$ while performing sampling from $\mathbf{SR}(p)$. In this case, the aforementioned lower and upper bounds become the Shannon entropy of $\mathbf{SR}(p)$ plus $D(\mathbf{SR}(p) \,\|\, \mathbf{R}^B)$ and the sum of the Shannon entropy of $\mathbf{SR}(p)$, unity, and $D(\mathbf{SR}(p) \,\|\, \mathbf{R}^B)$. In other words, the original lower and upper bounds are shifted upward by $D(\mathbf{SR}(p) \,\|\, \mathbf{R}^B)$. Therefore, $D(\mathbf{SR}(p) \,\|\, \mathbf{R}^B)$ can be regarded as the penalty incurred by the assumption that the values were sampled from $\mathbf{R}^B$. Such a penalty should be negligible due to Eq. (6) and the normalization in (8) and (9), which demands that $D(\mathbf{SR}(p) \| \mathbf{R}^B)$ be much smaller than the smaller of unity and the Shannon entropy of $\mathbf{SR}(p)$:

$$\textbf{Weak Criterion}: \quad D\left(\mathbf{SR}(p) \| \mathbf{R}^B\right) \le 0.1 \times \min\left(1, \text{Shannon entropy of } \mathbf{SR}(p)\right) \tag{10}$$

In the above arguments, Shannon entropy is defined in base 2:

$$\text{Shannon entropy of } S^B \equiv -\sum_{j=1}^{B} S^B(j) \log_2 \left[S^B(j) \right] \tag{11}$$

As a distance to some assumed probability distribution, one can introduce the relative L^α distance:

$$L^{r,\alpha}(S^B \| T^B) \equiv \sum_{j=1}^{B} \frac{\left| S^B(j) - T^B(j) \right|^\alpha}{\left[S^B(j) \right]^\alpha} S^B(j) = \sum_{j=1}^{B} \frac{\left| S^B(j) - T^B(j) \right|^\alpha}{\left[S^B(j) \right]^{\alpha-1}} \tag{12}$$

Qualitatively, one can interpret $L^{r,\alpha}(S^B \| T^B)$ to be a measure of how far/close the assumed true/ideal distribution T^B is from the observed/output distribution S^B. Note that the weighting in (12) is in line with the weighting in the relative entropy $D\left(S^B \| T^B\right)$. The definition in (12) yields two familiar distances:

$$L^{r,1} = \sum_{j=1}^{B} \left| S^B(j) - T^B(j) \right| = L^1 \text{ distance} \tag{13}$$

$$\begin{aligned} L^{r,2}(S^B \| T^B) &= \sum_{j=1}^{B} \frac{\left[S^B(j) - T^B(j) \right]^2}{S^B(j)} \\ &\equiv \chi(S^B \| T^B) \quad \left[\text{chi-square distance (statistic)}\right] \end{aligned} \tag{14}$$

In most statistical literatures, the chi-square distance with $T^B(j)$ in the denominator instead of $S^B(j)$ is introduced in the goodness-of-fits test for independent multinomial experiments [11]. However, Eq. (14) implies that the chi-square distance $\chi(S^B \| T^B)$ is simply an observation-based relative L^2 distance of S^B and T^B regardless of how the data for computing S^B were obtained.

There is a relation between the relative entropy and chi-square distance. Using

$$\log_e\left(1 + x\right) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \cdots \tag{15}$$

one can rewrite the relative entropy as

$$
\begin{aligned}
D(S^B \| T^B) &= -\frac{1}{\log_e 2}\sum_{j=1}^{B} S^B(j)\log_e\left(\frac{T^B(j)}{S^B(j)}\right) \\
&= -\frac{1}{\log_e 2}\sum_{j=1}^{B} S^B(j)\log_e\left(1+\frac{T^B(j)-S^B(j)}{S^B(j)}\right) \\
&= -\frac{1}{\log_e 2}\sum_{j=1}^{B} S^B(j)\left[\frac{T^B(j)-S^B(j)}{S^B(j)} - \frac{1}{2}\left(\frac{T^B(j)-S^B(j)}{S^B(j)}\right)^2 + \frac{1}{3}\left(\frac{T^B(j)-S^B(j)}{S^B(j)}\right)^3 - \cdots\right] \quad (16) \\
&= -\frac{1}{\log_e 2}\sum_{j=1}^{B}\left[T^B(j)-S^B(j) - \frac{1}{2}\frac{\left(T^B(j)-S^B(j)\right)^2}{S^B(j)} + \frac{1}{3}\frac{\left(T^B(j)-S^B(j)\right)^3}{\left(S^B(j)\right)^2} - \cdots\right] \\
&= \frac{1}{2\log_e 2}\chi\left(S^B \| T^B\right) - \frac{1}{3\log_e 2}\sum_{j=1}^{B}\frac{\left(T^B(j)-S^B(j)\right)^3}{\left(S^B(j)\right)^2} + \cdots
\end{aligned}
$$

($\sum_{j=1}^{B} T^B(j) = \sum_{j=1}^{B} S^B(j) = 1$ and definition in (14) are used.)

This implies

$$
\left| D(S^B \| T^B) - \frac{1}{2\log_e 2}\chi\left(S^B \| T^B\right)\right| < \text{ second order quantity of } \max_{1\le j\le B}\left[\left|\frac{T^B(j)-S^B(j)}{S^B(j)}\right|\right]. \quad (17)
$$

Since the right side is higher order, the left side must be very small when the difference between S^B and T^B is negligible. Accordingly, one should impose a stronger condition on the use of the left side of (17) than Weak Criterion in (10):

$$
\textbf{Criterion}: \left| D\left(\mathbf{SR}(p) \| \mathbf{R}^B\right) - \frac{1}{2\log_e 2}\chi\left(\mathbf{SR}(p) \| \mathbf{R}^B\right)\right| < 0.025\times\min\left(1, \text{Shannon entropy of } \mathbf{SR}(p)\right). \quad (18)
$$

One might suspect what the roles of two criteria were. This can be discussed by the analogy to the judgment of whether the correlation of two random variables is negligible. If the correlation coefficient of a random variable *X* and *Y*, denoted $CC(X,Y)$, is between –0.1 and 0.1, one can say that the correlation of *X* and *Y* is weak. If $|CC(X,Y)| < 0.05$, some say that the correlation of *X* and *Y* is negligible, but others say that $|CC(X,Y)|$ must be as small as possible within a computational/experimental uncertainty limit in order for *X* and *Y* to be regarded uncorrelated. Here, the middle ground would be $|CC(X,Y)| < 0.025$. Going back to Criterion and Weak Criterion, the relative entropy $D(\mathbf{SR}(p) \| \mathbf{R}^B)$ is the penalty of the descriptive length of $\mathbf{SR}(p)$ resulting from assuming that $\mathbf{R}^B$ is the ideal value of $\mathbf{SR}(p)$. This penalty appears in the fundamental relation for descriptive length with an uncertainty of unity (one bit) as mentioned in the discussion leading to Weak criterion. Simply put, the penalty is negligible if it is much smaller than an uncertainty of unity. Therefore, taking into account that Shannon entropy can be smaller than unity for a system of a few components, Weak Criterion corresponds to the weaker requirement of uncorrelatedness stated above. For the corresponding stronger requirement, a second order measure is chosen (left side in inequality (17)) and the middle ground stated above is adopted (factor 0.025 in Criterion).

4. NUMERICAL EXAMPLES

A numerical example is presented for a modeling of the initial core of a pressurized water reactor (PWR) in Fig. 3. More detailed specifications than those in Fig. 3 are available elsewhere [12]. The following two cases were analyzed:

Case 1: Two-dimensional geometry with reflecting boundaries on top and bottom; 200000 particles per iteration cycle; initial guess of uniform fission source distribution; one binning cell per fuel bundle (193 binning cells).

Case 2: Three-dimensional geometry with upper plenum, top and bottom end plugs, and top and bottom supports; initial guess of uniform fission source distribution; 10000 particles per iteration cycle; four binning cells per fuel bundle and twenty four cells vertically (18528 binning cells).

Fig. 4 shows the diagnostic results of Case 1 and 2. It is observed that Criterion is satisfied with large margin for Case 1 but is not satisfied for Case 2.

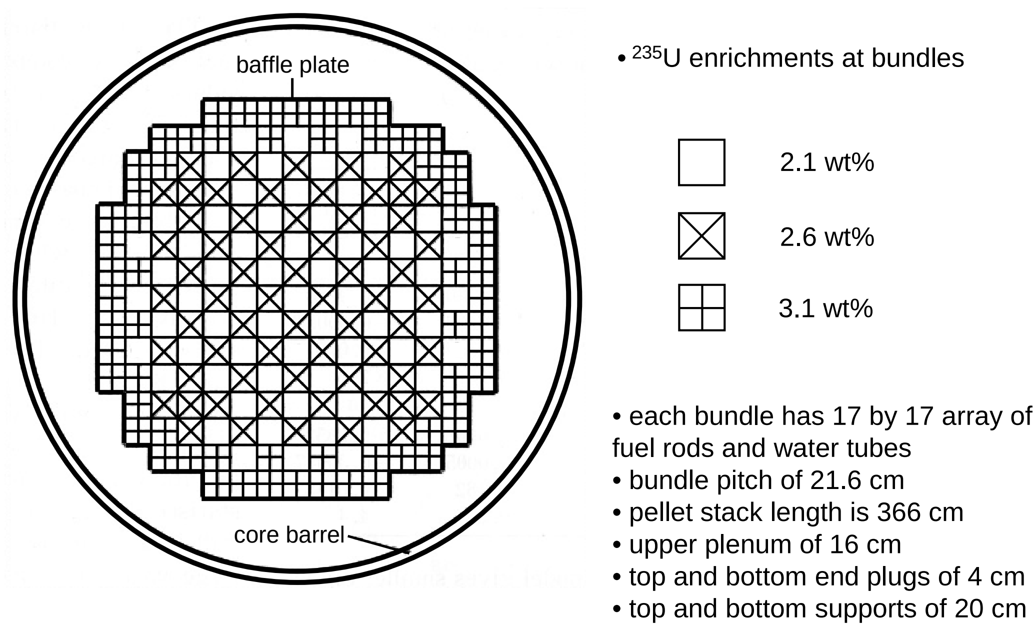

Figure 3: PWR initial core model (Case 1 and 2)

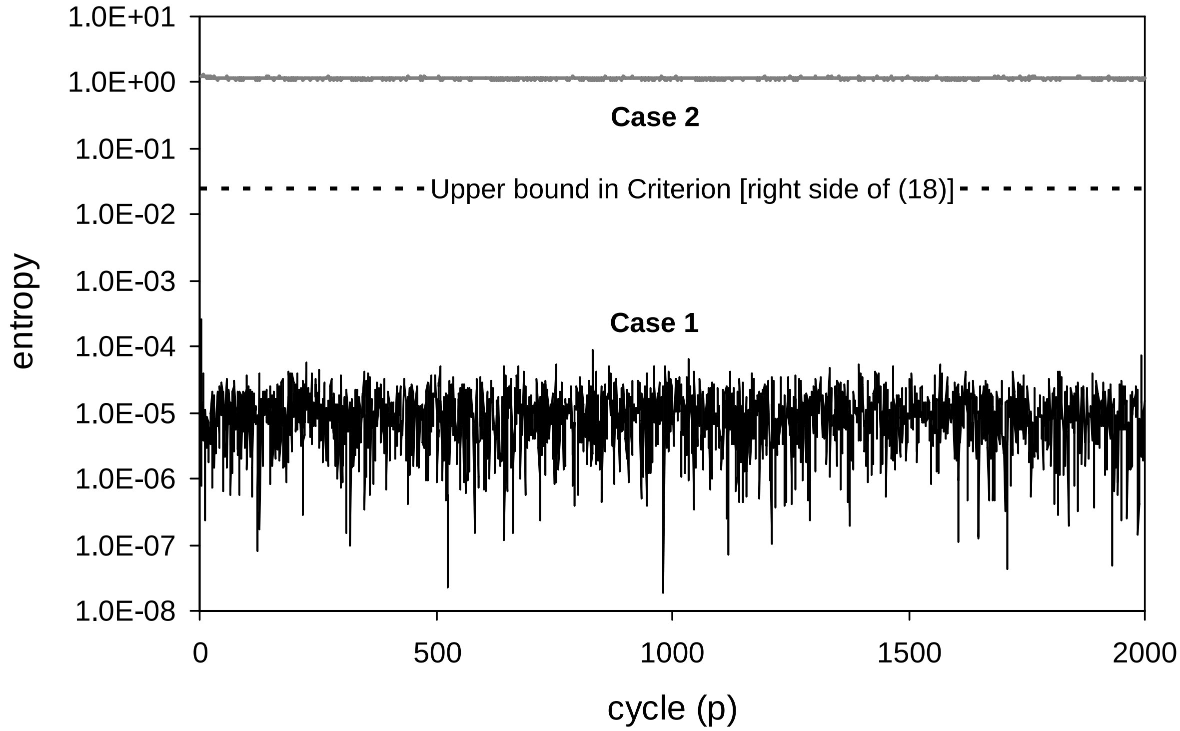

Figure 4. Diagnostic Results of Case 1 and 2

Another numerical example is presented for the k-effective-of-the-world problem discussed in 1. INTRODUCTION. Three cases were analyzed:

Case 3: 1000 particles per iteration cycle; initial guess of uniform fission source distribution; one binning cells per sphere.
Case 4: 10000 particles per iteration cycle; initial guess of uniform fission source distribution; one binning cells per sphere.
Case 5: 100000 particles per iteration cycle; initial guess of uniform fission source distribution; one binning cells per sphere.

Fig. 5 shows the diagnostics of these cases. As expected, Criterion is not satisfied in Case 3 while it is satisfied in Case 5 with large margin. These diagnostics correctly reflect the results in Fig. 2. On the other hand, in Case 4, the violation of Criterion is not negligible although Criterion appears to be satisfied through the majority of iteration cycles. This observation is consistent with Fig. 2 where a clear cut judgment on the estimation bias of k_{eff} appears to be impossible for 10000 particles per iteration cycle.

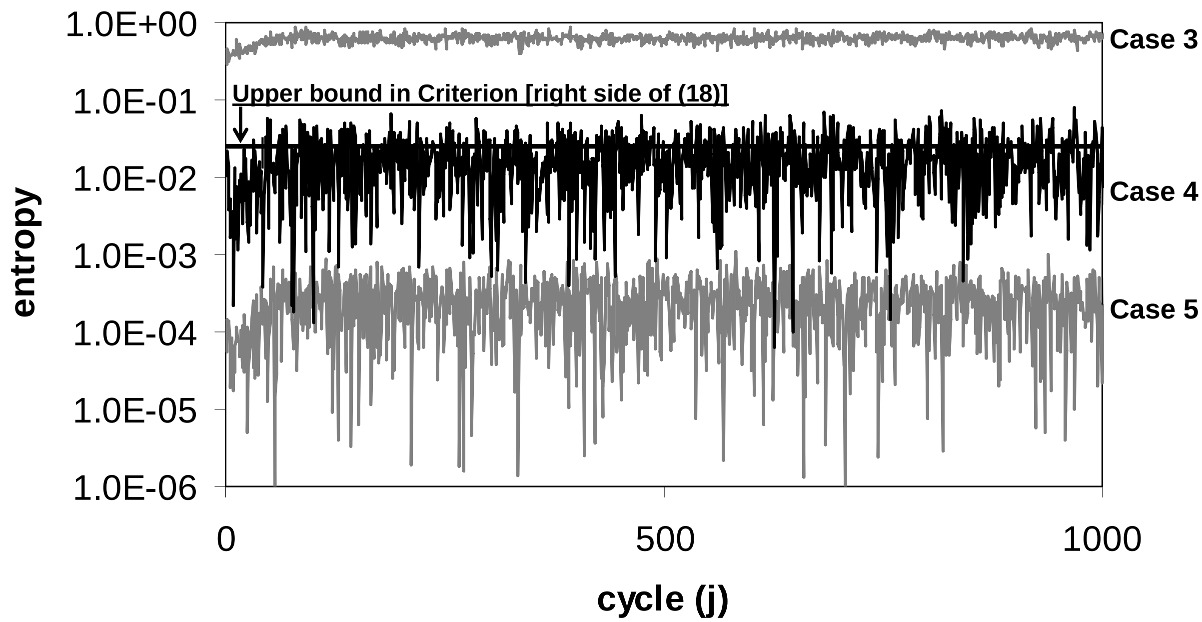

Figure 5: Diagnostic Results of Case 3, 4 & 5

Some considerations and speculations are worth mentioning:

A. The appearances of spike-like local maxima indicate that the diagnostic measure in Criterion, i.e., left side in (18), would behave like non-linear time series. It will not be straightforward to devise a statistical test for the judgment of whether the number of particles per iteration cycle should be increased.

B. The best that one can do at the present moment is the pursuit of an empirical rule like the doubling of the number of particles per iteration cycle in the case of criterion violation through the majority of the present 20 consecutive cycles.

C. The adaptive increase of the number of particles per iteration cycle, if necessary, should be conducted after on-the-fly stationarity diagnostics of source distribution issued a positive flag.

D. For further study, data equivalent to Fig. 2 should be collected for flux and reaction rates at the representative fuel bundle locations of a whole core modeling.

5. CONCLUSIONS

Work in this article has shown that the diagnostics of particle population for a given spatial resolution of binning cells can be implemented in an on-the-fly manner. The source ratio vector whose components are the ratio of the sources of adjacent iteration cycles at the individual binning cells plays a critical role. The distance from the source ratio vector to the problem-independent reference vector $(1, 1, \ldots, 1)$ is characterized using relative entropy and chi-square distance under the normalization condition of a total mass of unity. Future work will include the refinement of diagnostic criteria depending on whether the users of MC criticality/eigenvalue codes need to evaluate effective neutron multiplication factor (k_{eff}) or power and reaction rate distribution.

REFERENCES

1. Y. Naito and J. Yang, "The Sandwich Method for Determining Source Convergence in Monte Carlo Calculation," *Journal of Nuclear Science and Technology*, **41**, 559-568 (2004).
2. T. Ueki and F.B. Brown, "Stationarity Modeling and Informatics-Based Diagnostics in Monte Carlo Criticality Calculations," *Nuclear Science and Engineering*, **149**, 38-50 (2005).
3. Y. Richet, O. Jacquet, and X. Bay, "Initialization Bias Suppression of an Iterative Monte Carlo Calculation," CD-ROM, Monte Carlo 2005 Topical Meeting, Chattanooga, Tennessee, April 17-21, 2005.
4. H.J. Shim and C.H. Kim, "Convergence Criteria of Inactive Cycle Monte Carlo Calculations," *Transactions of American Nuclear Society*, **93**, 557-559 (2005).
5. T. Ueki, "Information Theory and Undersampling Diagnostics for Monte Carlo Simulation of Nuclear Criticality," *Nuclear Science and Engineering*, **151**, 283-292 (2005).
6. T.M. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, NY 1991.
7. X-5 Monte Carlo Team, "MCNP – A General Purpose Monte Carlo N-Particle Transport Code, Version 5," LA-UR-03-1987, Los Alamos National Laboratory, Los Alamos, NM, April, 2003.
8. S.S. Shapiro and M.B. Wilk, "An Analysis of Variance Test for Normality," *Biometrika*, **52**, 3 and 4, 591-611 (1965).
9. G.L. Burrows and D.B. MacMillan, "Confidence Limts for Monte Carlo Calculations," *Nuclear Science and Engineering*, **22**, 384-385 (1965).
10. G.E. Whitesides, "A Difficulty in Computing the K-effective of the World," *Transactions of American Nuclear Society*, **14**, 680 (1971).
11. R.J. Larsen and M.L. Marx, *An Introduction to Mathematical Statistics and Its Applications*, Prentice-Hall, Englewood Cliffs, NJ 1986.
12. M. Nakagawa and T. Mori, "Whole Core Calculations of Power Reactors by Use of Monte Carlo Method," *Journal of Nuclear Science and Technology*, **30**, 692-701 (1993).