The Monte Carlo Method: Versatility Unbounded In A Dynamic Computing World
Chattanooga, Tennessee, April 17–21, 2005, on CD-ROM, American Nuclear Society, LaGrange Park, IL (2005)

Asymptotic Equipartition Property and Undersampling Diagnostics in Monte Carlo Criticality Calculation

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

ABSTRACT

Information-theoretic undersampling diagnostics have been proposed for the source distribution in Monte Carlo criticality calculations. The proposed diagnostics consist of the posterior checking of the number of neutron histories per cycle. First, a criterion based on a large sample property of particle population is defined within the framework of the asymptotic equipartition property and the method of types. Second, an auxiliary criterion is defined using the concave property of Shannon entropy. Numerical results are presented for the historical 'k-effective of the world' problem by Whitesides. The results indicate that the estimation bias of neutron effective multiplication factor will be reduced to a practically negligible level if the criteria are satisfied. The theoretical and numerical results combined indicate that equilibrium would be a stronger condition than stationarity.

Key Words: Nuclear criticality; asymptotic equipartition property; method of type; entropy, Monte Carlo

1 INTRODUCTION

Monte Carlo methods simulate the criticality of nuclear fission systems by source iterations. [1] After preliminary iterations, the simulation becomes stationary and various quantities of interest are computed at each iteration. These preliminary and stationary iterations are called inactive and active cycles, respectively, and the beginning of active cycles can be diagnosed using relative entropy [2, 3]. A realization of the source at each cycle can be expressed by the paired collection of fission sites and their corresponding statistical weights from the tracking of individual neutrons. This realization becomes the source distribution for the simulation in the next cycle. In order to keep neutron population constant through cycles, the source distribution is normalized in such a way that the ensemble average of the distributions of the initial sites of histories is unbiased conditional on the realization of fission sites at the previous cycle. Obviously, it is important to diagnose in a posterior manner whether the number of neutron histories per cycle has been sufficiently large to guarantee the reliable estimation of neutron effective multiplication factor (k_{eff}). However, a direct approach to this issue has rarely been attempted at in contrast to recent interest in stationarity and convergence diagnostics [2, 3, 4, 5].

In general, there are two forms of problems in the source convergence. [6] The first problem, deterministically slow convergence in both Monte Carlo and deterministic calculations, occurs in loosely coupled systems even if the number of neutron histories per cycle is infinite. This problem is characterized by a dominance ratio of nearly unity. Here the dominance ratio is the ratio of the second largest to the largest eigenvalues. The second problem, the 'k-effective of the world' problem by Whitesides [7], occurs in systems consisting of arrays of fissile components where only one component of the arrays dominates the criticality of the entire system. This problem afflicts only Monte Carlo methods and is a symptom of stochastic undersampling: Even after source convergence is achieved, the difficulty of keeping simulated neutrons in and around the dominant component persists and the source distribution fluctuates more than allowed in equilibrium. Thus the large departure of k_{eff} estimate from the true value may occur at a rate higher than dictated by statistics if the whole Monte Carlo simulation is replicated many times. However, this phenomenon will not occur in a hypothetical limit of the infinite number of neutron histories per cycle.

In this work, the diagnostic criterion for the number of neutron histories per cycle is investigated. The large sample property of neutron population is discussed based on the asymptotic equipartition property [8] in relation to thermodynamical equilibrium, and the method of type [8] is then applied to derive a criterion. The concave property of Shannon entropy is utilized to define an auxiliary criterion. Numerical results are presented to show the performance of these criteria.

2 ASYMPTOTIC EQUIPARTITION PROPERTY AND METHOD OF TYPE

The source distribution in Monte Carlo criticality calculation can be measured at the beginning of each cycle using given spatial meshes. Its observation is represented by the binned source $S^B(j)$ normalized to unity, where j corresponds to the bin numbers for the meshes and B stands for the number of the bins. The state of source distribution is characterized in terms of minimum descriptive length [8] using Shannon entropy:

$$H(S^B) = -\sum_{i=1}^{B} S^B(i)\log_2\left(S^B(i)\right). \tag{1}$$

Let Q be a discrete probability density function (pdf) defined in the same domain. The relative entropy of S^B with respect to Q is defined as

$$D(S^B \parallel Q) = \sum_{i=1}^{B} S^B(i)\log_2\left(S^B(i)/Q(i)\right). \tag{2}$$

These entropies are used to investigate a large sample property of particle population.

The asymptotic equipartition property is stated as follows [8]:

Theorem 1 (asymptotic equipartition property (AEP)): Let $X_{1,}\ X_2, ..., X_n$ be independent and identically distributed random variables taking finite values drawn from a pdf $p(x)$. Then

$$-\frac{1}{n}\log p^n(X_1, X_2, \ldots, X_n) \to H(p) \quad \textit{in probability},$$

where $p^n(X_1, X_2, \ldots, X_n)$ is the joint pdf of $X_1, X_2, \ldots, X_n$.

Next, the concept of a typical set is introduced:

Definition: The typical set $A_\varepsilon^{(n)}$ with respect to $p(x)$ is the set of sequences $(x_1, x_2, \ldots, x_n)$ with the following property:

$$2^{-n(H(p)+\varepsilon)} \le p^n(x_1, x_2, \ldots, x_n) \le 2^{-n(H(p)-\varepsilon)}, \tag{3}$$

where ε is a positive number.

The typical set has the following property [8]:

Theorem 2:

1. If $(x_1, x_2, \ldots, x_n) \in A_\varepsilon^{(n)}$, then $H(p) - \varepsilon \le -\frac{1}{n}\log_2 p^n(x_1, x_2, \ldots, x_n) \le H(p) + \varepsilon$.
2. Probability $\left\{(x_1, x_2, \ldots, x_n) \in A_\varepsilon^{(n)}\right\} \ge 1 - \varepsilon$ for n sufficiently large.
3. $|A_\varepsilon^{(n)}| \le 2^{n(H(p)+\varepsilon)}$, where $|A|$ denotes the number of elements in the set A.
4. $|A_\varepsilon^{(n)}| \ge (1-\varepsilon)2^{n(H(p)-\varepsilon)}$ for n sufficiently large.

Suppose that n is sufficiently large such that Probability $\left\{(x_1, x_2, \ldots, x_n) \in A_\varepsilon^{(n)}\right\} \ge 1 - \varepsilon$ for a $\varepsilon\,(\ll 1)$ as in the item 2 in Theorem 2. Using independence and the item 3 and 4 in Theorem 2, one can show

$$\begin{aligned} &H(p^n(X_1, X_2, \ldots, X_n)) = -E(\log_2 p^n(X_1, X_2, \ldots, X_n)) \\ &= -nE(\log_2 p(X)) = nH(p) \approx \log_2 |A_\varepsilon^{(n)}|. \end{aligned} \tag{4}$$

If X_i can assume only integer values like bin numbers, the Shannon entropy of $(X_1, \ldots, X_n)$ can be interpreted as

$$H(p^n(X_1, X_2, \ldots, X_n)) \approx \log_2 |A_\varepsilon^{(n)}| = \log_2(\mathit{Volume\ of}\ A_\varepsilon^{(n)}). \tag{5}$$

This is essentially equivalent to the Boltzmann entropy [9]:

$$\mathit{Boltzmann\ entropy\ of\ macroscopic\ description}\ x = (k_B \ln 2)\log_2 L(\Gamma_x), \tag{6}$$

where $k_B = 1.38 \times 10^{-23}$ joules/$^\circ$kelvin (Boltzmann constant) and Γ_x is the cell in state space corresponding to the macroscopic description x and $L(\Gamma_x)$ is the volume of that cell. The bridge between (4) and (6) is the interpretation of the typical set $A_\varepsilon^{(n)}$ as the description of equilibrium for large sample cases. In a thermo-dynamical sense, the typical set corresponds to the macrostate that contains most microstates observable in equilibrium. Thus one can seek a condition on the number of samples that makes the macroscopic description of observations stay the same over time. To this end, the method of type in information theory is introduced [8]:

Definition 2: The type P_{X^n} (or empirical pdf) of a sequence $X_1, X_2, \ldots, X_n$ is the relative proportion of occurrences of each discrete value, i.e., $P_{X^n}(a) = N(a \mid X^n)/n$, where $N(a \mid X^n)$ is the number of times the value a occurs in the sequence.

Theorem 3: Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed discrete random variables drawn from a pdf $Q(x)$. Suppose that X can take finite values. Then the joint probability of $X_1, \ldots, X_n$, denoted by $Q^n(X^n)$, is given by

$$Q^n(X^n) = 2^{-n\left(H(P_{X^n}) + D(P_{X^n} \| Q)\right)}. \tag{7}$$

3 UNDERSAMPLING DIAGNOSTICS

Let *N* be the number of neutron histories per cycle and X_i the bin number of *i*-th source neutron. Then P_{X^N} corresponds to the normalized binned source S_j^B at cycle *j*:

$$P_{X^N} \leftarrow S_j^B . \tag{8}$$

When *N* is sufficiently large, Theorem 2 states that $(X_1,\ldots,X_N)$ belongs to the typical set $A_\varepsilon^{(N)}$ with a probability larger than $1-\varepsilon$. Therefore, by the definition of the typical set and Theorem 3, the following inequality holds with a probability larger than $1-\varepsilon$:

$$2^{-N(H(Q)+\varepsilon)} \le Q^N(X^N) = 2^{-N\left(D(S_j^B \| Q)+H(S_j^B)\right)} \le 2^{-N(H(Q)-\varepsilon)} , \tag{9}$$

where *Q* introduced in Sec. 2 corresponds to the true distribution of X_i. Hence one obtains the bounds of the relative entropy $D\left(S_j^B \| Q\right)$:

$$H(Q)-H(S_j^B)-\varepsilon \le D(S_j^B \| Q) \le H(Q)-H(S_j^B)+\varepsilon \tag{10}$$

Based on this, one can propose the following diagnostics:

<u>Principal undersampling diagnostic criterion</u>

> For sufficiently small ε and sufficiently large *N*, $D\left(S_j^B \| Q\right)$ stays in $\left[H(Q)-H(S_j^B)-\varepsilon, H(Q)-H(S_j^B)+\varepsilon\right]$ through active cycles.

As in previous work [2, 3], *Q* is estimated as the average of S_j^B over the second half of active cycles:

$$Q \leftarrow \frac{2}{M} \sum_{j=M/2+1}^{M} S_j^B , \tag{11}$$

where *M* is the total number of active cycles. In this case, $D\left(S_j^B \| Q\right)$ is called posterior relative entropy.

Now, the independent assumption made in the previous definitions and theorems is examined. Let $\vec{r}_l$ be the spatial coordinate vector of the *l*-th fission site in cycle *j* and w_l be the expected total statistical weight of the neutrons to be born at $\vec{r}_l$. The source distribution for the next cycle is expressed as

$$\frac{\sum_{i=1}^{L(j)} w_i \delta(\vec{r}-\vec{r}_i)}{\sum_{l=1}^{L(j)} w_l} , \tag{12}$$

where $L(j)$ is the total number of fission events in cycle *j* and $\delta(\vec{r})$ is the Dirac delta function (point mass). The criterion defined above is valid if the position coordinates of *N* source neutrons are sampled independently with a probability of $w_i / \sum_{l=1}^{N(j)} w_l$ for $\vec{r}_i$. However, this is not an efficient process and some sampling scheme partly similar to stratified sampling is generally employed in such a way that the expected number of source neutrons at $\vec{r}_i$ is $Nw_i / \sum_{l=1}^{N(j)} w_l$. See elsewhere for a particular example. [10] Therefore, the criterion defined above is the zero-th order method for dealing with an equilibrium property of large neutron population.

Since Shannon entropy characterizes discrete random variables in terms of the minimum descriptive length [8], diagnostics using only Shannon entropy may be sought. A unique geometric property of Shannon entropy is concavity [8]:

$$H\left(\lambda p_1 + (1-\lambda) p_2\right) \geq \lambda H\left(p_1\right) + (1-\lambda) H(p_2), \quad 0 \leq \lambda \leq 1, \tag{13}$$

where p_1 and p_2 are discrete pdf. The inequality in (13) can be easily extended to *n* pdf $p_1, \ldots, p_n$ by induction:

$$\begin{aligned} H(\lambda_1 p_1 + \lambda_2 p_2 + \cdots + \lambda_n p_n) &\geq \lambda_1 H(p_1) + (1-\lambda_1) H\left(\frac{\lambda_2 p_2 + \cdots + \lambda_n p_n}{1-\lambda_1}\right) \\ &= \lambda_1 H(p_1) + (1-\lambda_1) H\left(\frac{\lambda_2}{1-\lambda_1} p_2 + \frac{1-\lambda_1-\lambda_2}{1-\lambda_1} \frac{\lambda_3 p_3 + \cdots + \lambda_n p_n}{1-\lambda_1-\lambda_2}\right) \\ &\geq \lambda_1 H(p_1) + \lambda_2 H(p_2) + (1-\lambda_1-\lambda_2) H\left(\frac{\lambda_3 p_3 + \cdots + \lambda_n p_n}{1-\lambda_1-\lambda_2}\right) \\ &\vdots \\ &\geq \lambda_1 H(p_1) + \lambda_2 H(p_2) + \cdots + \lambda_n H(p_n), \end{aligned} \tag{14}$$

where $\sum_{i=1}^{n} \lambda_i = 1$ and $\lambda_i > 0$. Letting *M* be the number of active cycles as before and applying (14) to the binned sources over the second half of active cycles, $S_{M/2+1}^B, \ldots, S_M^B$, with $\lambda_i = 2/M$, one obtains

$$H(Q) \simeq H\left(\frac{2}{M} \sum_{j=M/2+1}^{M} S_j^B\right) > \bar{H} = \frac{2}{M} \sum_{j=M/2+1}^{M} H(S_j^B) \ . \tag{15}$$

where the equality holds if and only if all S_j^B through the second half of active cycles are equal. If an acceptable difference between $H(Q)$ and $\bar{H}$ is specified, a useful diagnostic criterion will be obtained. Since Shannon entropy is the lower bound of minimum descriptive length with one bit uncertainty [8], the following may be tested as an auxiliary criterion:

$$H(Q) - \bar{H} < \eta = 0.05 \, . \tag{16}$$

Although the criterion (16) appears to lack firm theoretical basis on the magnitude of η, it does not assume independence among the bins of source neutrons. Therefore, the criterion (16) complements the principal undersampling diagnostic criterion defined before.

4 NUMERICAL RESULTS

Simulations of Whitesides' k_{eff}-of-the-world problem [7] were conducted with various numbers of histories per cycle. The detailed specification of Whitesides' k_{eff}-of-the-world problem is as follows. Pu 239 metal spheres with a radius of 3.7819 *cm* are placed at an interval of 60 *cm* to form a 9 by 9 by 9 array except the central sphere; it has a radius of 4.968 cm. The space between the spheres is void. The array is reflected on all six sides by a water reflector of a thickness of 30 cm, whose interior surface is 60 cm away from the center of the sphere at the exterior array. A schematic drawing of the geometry is shown in Figure 1. As shown in Table I, the dominance ratio has been computed to be 0.74 ± 0.04 (two σ) using time series analysis [11]. Therefore, as pointed out by Blomquist [6], Whitesides' k_{eff}-of-the-world problem is not a slow convergence problem but a stochastic undersampling problem. MCNP5 [10] and its optional linear congruential generators with 63 bit arithmetic recommended in recent literature [12] were used to evaluate the performance of the diagnostics. In all simulations, the source convergence to stationarity was judged based on whether the posterior relative entropy $D(S_j^B \| Q)$ crosses downward its average over the second half of active cycles before the first active cycle [2,3], and one bin was assigned to each spherical unit in Figure 1, i.e., the total number of bins being $729 = 9 \times 9 \times 9$.

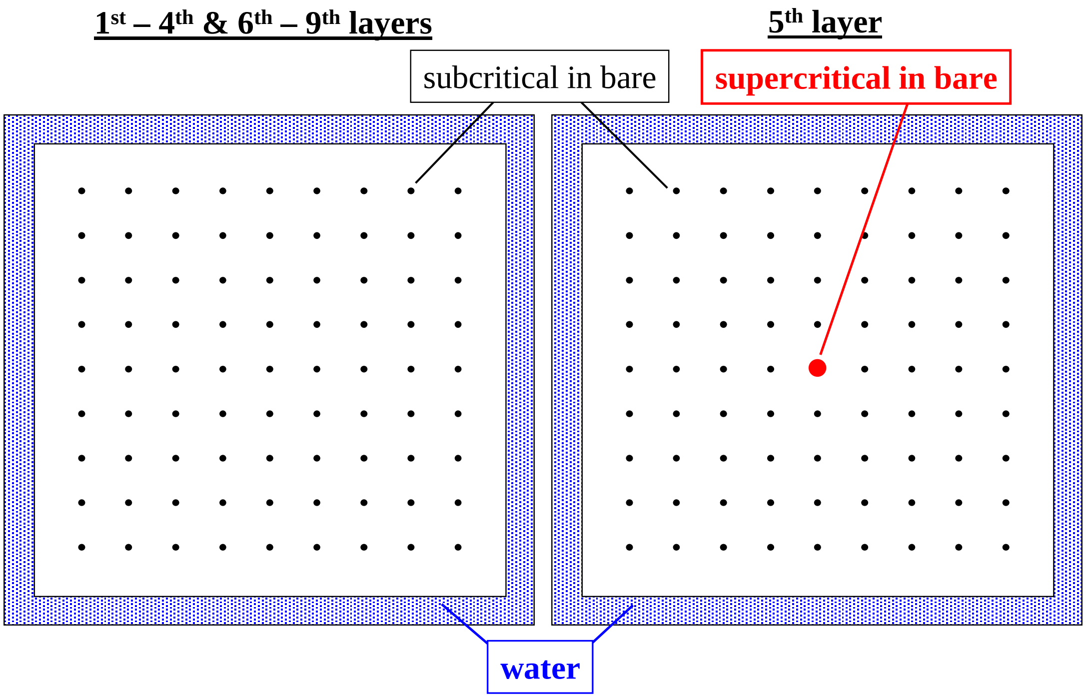

Figure 1: Whitesides k-effective-of-the-world problem (space between spheres is void; not scaled)

Table I: Dominance ratio computed by time series analysis

Whitesides keff-of-the-world	9 by 9 by 9 arrays of all identical spheres
$0.73 \pm 0.04\,(2\sigma)$	$0.90 \pm 0.01\,(2\sigma)$

Figures 2 and 3 show diagnostic results using the principal undersampling diagnostic criterion with $\varepsilon = 0.1$. It is clear that the criterion is and is not satisfied in simulations of 60000 and 1000 histories per cycle, respectively. Figure 4 shows $H(Q)$ and $\bar{H}$. It is clear that the auxiliary criterion in (16) is satisfied if the number of histories per cycle is larger than 40000.

To investigate how the criteria affect the estimation of k_{eff}, its bias is numerically evaluated. Figure 5 shows the 95% confidence intervals of k_{eff} for various numbers of histories per cycle with the total number of histories through active cycles being 200,000,000. The true value of k_{eff} was estimated by 1800 active cycles with 2,000,000 histories per cycle. One can observe that any confidence interval does not contain the true value estimate for the number of histories per cycle less than 3000, while all confidence intervals for the number of histories larger than 50000 contain part or all of the confidence interval of the true value estimate. One can safely say that the bias is negligible if the number of histories per cycle is larger than 50000 and three fractional digit accuracy is required. This observation is consistent with the diagnostic results shown previously. All the results combined indicate that equilibrium would be a stronger condition than stationarity.

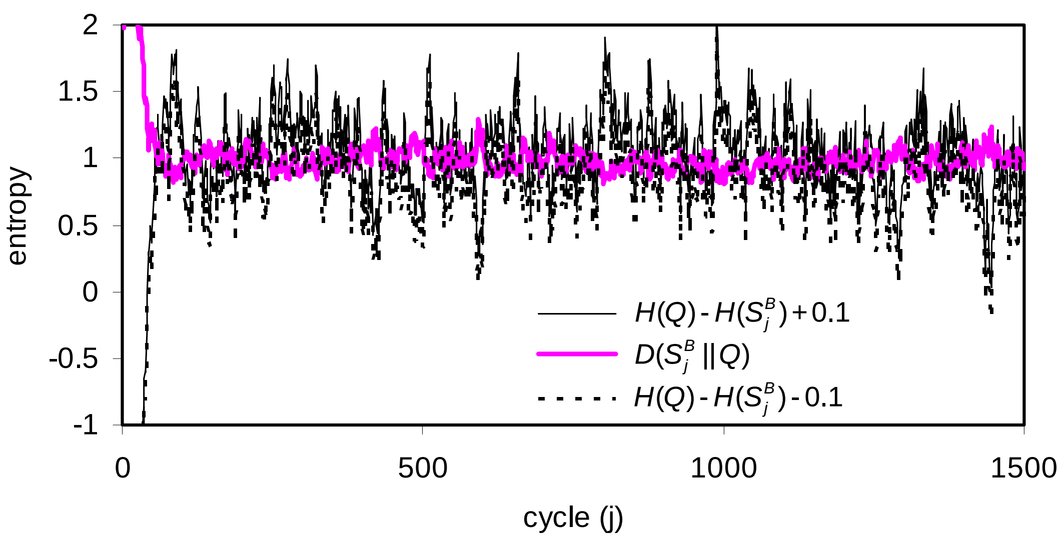

Figure 2: Posterior relative entropy and asymptotic equipartition bounds for a 1000 histories per cycle simulation with 100 inactive cycles (ε=0.1).

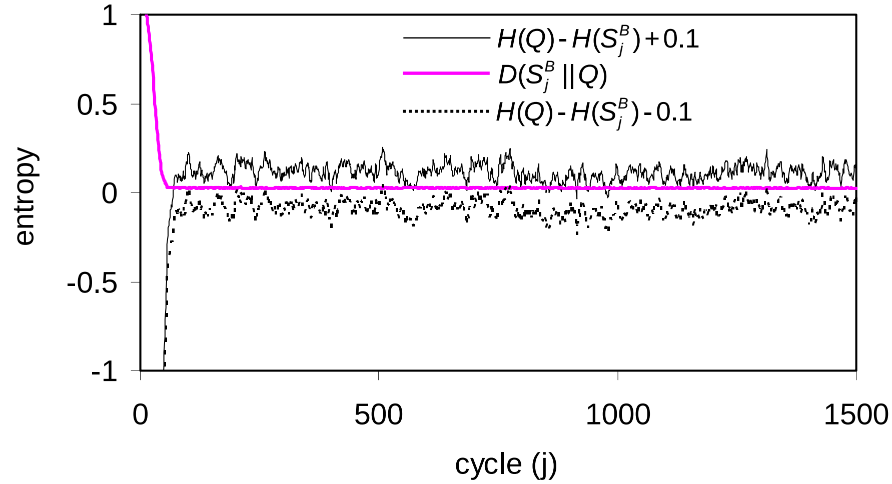

Figure 3: Posterior relative entropy and asymptotic equipartition bounds for a 60000 histories per cycle simulation with 100 inactive cycles (ε=0.1).

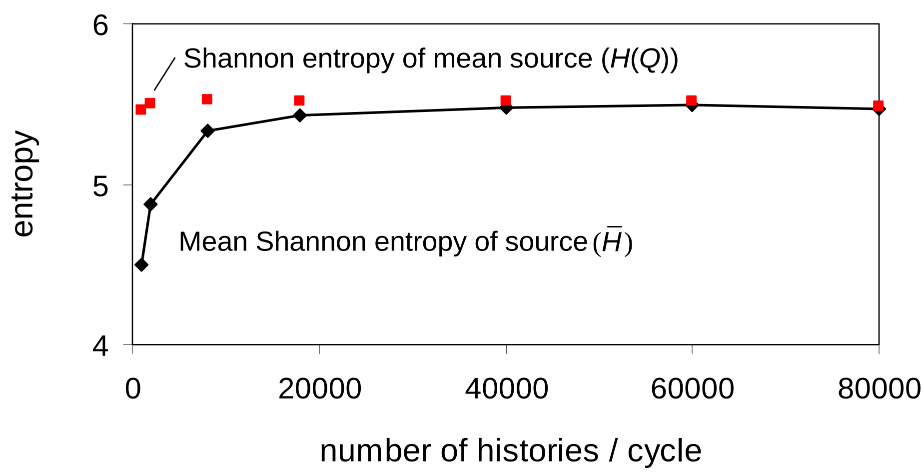

Figure 4: Shannon Entropies at Stationarity

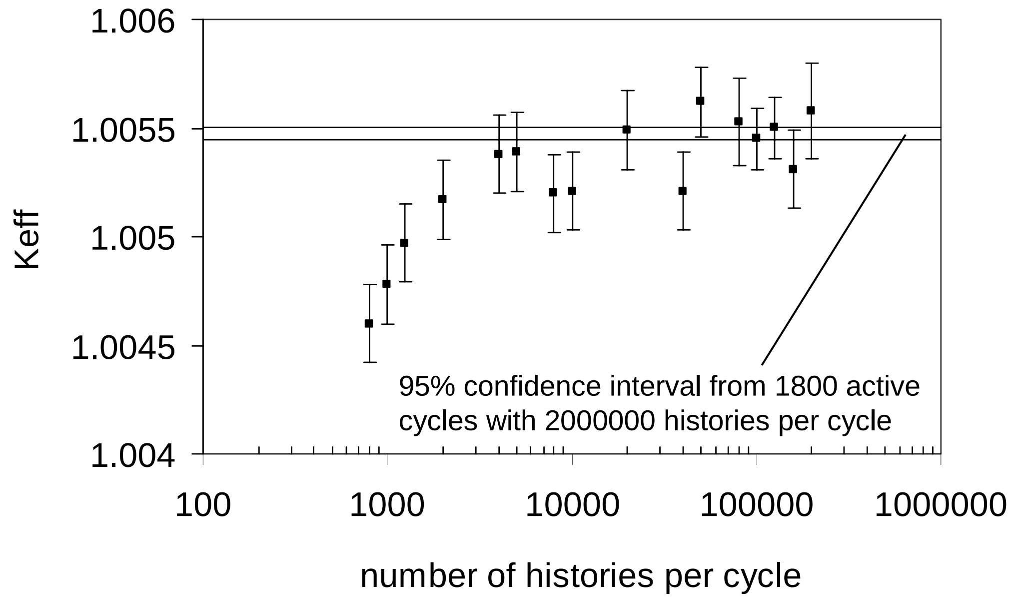

Figure 5: Confidence interval (95%) from 200,000,000 histories in active cycles (variance computed by batch grouping of 50 cycles)

5 CONCLUSIONS

The present work has shown that the potential exists to diagnose the number of histories per cycle in Monte Carlo criticality calculations using information theory. The numerical results show that the two diagnostics proposed will improve reliability in k_{eff} estimation. However, the local quantity estimation in a global simulation was not analyzed. Therefore, effort in future work should be directed toward reactor analysis emphasizing the estimation of reaction rate per fuel bundle or per fuel pin in a whole reactor core simulation.

6 REFERENCES

1. J. Lieberoth, “A Monte Carlo technique to solve the static eigenvalue problem of the Bolzmann transport equation,” *Nucleonik*, **Band 11**, Heft 5, 213-219 (1968).
2. T. Ueki and F.B. Brown, “Informatics approach to stationarity diagnostics of the Monte Carlo fission source distribution,” *Transactions of American Nuclear Society* **89**, 458-461 (2003).
3. T. Ueki and F.B. Brown, “Stationarity modeling and informatics-based diagnostics in Monte Carlo criticality calculations,” *Nucl. Sci. Eng.*, **149**, 38-50 (2005).
4. Y. Richet, O. Jacquet, and X. Bay, “Automated suppression of the initial transient in Monte Carlo calculations based on stationarity detection using the Brownian bridge theory,” in *Proc. ICNC 2003, 7th International Conference on Nuclear Criticality Safety*, Tokai-Mura, Japan, October 2003.

5. Y. Naito and J. Yang, "The sandwich method for determining source convergence in Monte Carlo calculation," *Journal of Nuclear Science and Technology*, **41**, 5, 559-568 (2004).

6. R.N. Blomquist, "Monte Carlo source convergence and the Whitesides problem," in *Proc. PHYSOR 2000, ANS International Topical Meeting on Advances in Reactor Physics and Mathematics and Computation into the Next Millennium*, Pittsburgh, PA, May 2000.

7. G.E. Whitesides, "A difficulty in computing the k-effective of the World," *Trans. Am. Nucl. Soc.*, **14**, 680 (1971).

8. T. M. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, NY USA 1991.

9. M. Li and P. Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications*, Second Edition, Springer-Verlag, New York, NY USA 1997.

10. 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.

11. T. Ueki, F.B. Brown, D.K. Parsons and J.S. Warsa "Time Series Analysis of Monte Carlo Fission Sources: I. Dominance Ratio Computation," *Nucl. Sci. Eng*, **148**, 374 -390 (2004).

12. F. B. Brown and Y. Nagaya, "The MCNP5 Random Number Generator," *Transactions of American Nuclear Society* **87**, 230-232 (2002).