CONF-9309131--3

LA-UR-93-1446

Los Alamos National Laboratory is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36

TITLE: A NEW METHOD TO ASSESS MONTE CARLO CONVERGENCE

AUTHOR(S): R. A. Forster
T. E. Booth
S. P. Pederson

SUBMITTED TO: Organization for Economic Cooperation and Development and Nuclear Energy Agency Seminar on Advanced Monte Carlo Computer Programs for Radiation Transport in Saclay, France

MASTER

By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes.

The Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

Los Alamos
Los Alamos National Laboratory
Los Alamos, New Mexico 87545

A NEW METHOD TO ASSESS MONTE CARLO CONVERGENCE

R. A. Forster, T. E. Booth, And S. P. Pederson
Los Alamos National Laboratory
Los Alamos, NM 87545

ABSTRACT

The central limit theorem can be applied to a Monte Carlo solution if the following two requirements are satisfied: 1) the random variable has a finite mean and a finite variance; and 2) the number N of independent observations grows large. When these are satisfied, a confidence interval based on the normal distribution with a specified coverage probability can be formed. The first requirement is generally satisfied by the knowledge of the type of Monte Carlo tally being used. The Monte Carlo practitioner has only a limited number of marginally quantifiable methods that use sampled values to assess the fulfillment of the second requirement; e.g., statistical error reduction proportional to $1/\sqrt{N}$ with error magnitude guidelines. No consideration is given to what has not yet been sampled.

A new method is presented here to assess the convergence of Monte Carlo solutions by analyzing the shape of the empirical probability density function (PDF) of history scores, $f(x)$, where the random variable x is the score from one particle history and $\int_{-\infty}^{\infty} f(x)dx = 1$. Since $f(x)$ is seldom known explicitly, Monte Carlo particle random walks sample $f(x)$ implicitly. Unless there is a largest possible history score, the empirical $f(x)$ must eventually decrease more steeply than $1/x^3$ for the second moment ($\int_{-\infty}^{\infty} x^2 f(x)dx$) to exist. It

is postulated that if such decreasing behavior in the empirical $f(x)$ has not been observed, then N is not large enough to satisfy the central limit theorem because $f(x)$ has not been completely sampled. Therefore, a larger N is required before a confidence interval should be formed.

The largest x's for an unbounded empirical history score PDF have been fit to a generalized Pareto function, which is a flexible two parameter distribution used to model rare events given by

$$Pareto \quad f(x) = a^{-1}(1 + kx/a)^{-1-1/k} \quad .$$

The two parameters a and k are determined by a maximum likelihood fit to the largest 5% of the x's up to 200 values using the downhill simplex method. The estimated slope of the largest history scores is then defined as

$$slope = 1 + 1/k \quad .$$

A slope value greater than 3 should be observed in the empirical unbounded $f(x)$ to satisfy the $1/x^3$ convergence criterion discussed above.

Several test problems were run with a modified version of MCNP that calculates $f(x)$ and the Pareto estimate of the slope for the user-defined tally fluctuation chart bin of each tally. One test problem was to calculate the surface neutron leakage flux above 12 MeV from an isotropic 14 MeV neutron point source of unit strength at the center of a 30 cm thick concrete shell with an outer radius of 390 cm. Point and ring detectors were deliberately used to estimate the surface neutron leakage flux with highly inefficient, long tailed $f(x)$s. A result from this calculation for a point detector appeared to be converged, but was a factor of 4 below the correct result. The empirical $f(x)$ clearly shows that a confidence interval should not be formed for this result because the estimated slope was only 1.4.

I. INTRODUCTION

Accurate confidence intervals can only be created when the number of Monte Carlo histories N becomes large enough such that the conditions of the central limit theorem (CLT) are met. The Monte Carlo user has a limited number of marginal methods to assess the fulfillment of this condition, such as statistical error reduction proportional to the $1/\sqrt{N}$ and error magnitude rules of thumb,[1] third moment estimators,[2] and fourth moment estimators such as the variance of the variance.[3,4] A new method is described in this paper that examines the probability density function (PDF) distribution of history scores.

Little work appears to have been done in this area. Score distribution histograms have been generated[4] with an emphasis on determining if a small number of histories account for a large fraction of the total result. Score distribution histograms can also be used to create and study both analytic[5,6] and empirical[7] PDFs for history scores. These are useful in determining if the distribution has been sampled well enough to expect a well converged solution. This paper focuses on initial studies of Monte Carlo history score distributions from particle transport problems. The test calculations have been performed using a modified version of MCNP.[1]

II. THE HISTORY SCORE PROBABILITY DENSITY FUNCTION $f(x)$

A history score posted to a tally bin by MCNP can be thought of as having been sampled from an underlying and generally unknown history score PDF $f(x)$, where the random variable x is the score from one complete particle history to a tally bin. The history score can be either positive or negative. The quantity $f(x)dx$ is the probability of selecting a history score between x and $x + dx$ for the tally bin. Each tally bin will have its own $f(x)$.

The most general form for expressing $f(x)$ mathematically is

$$f(x) = f_c(x) + \sum_i p_i \delta(x - x_i) \quad , \tag{1}$$

where $f_c(x)$ is the continuous part and $\sum_i p_i \delta(x - x_i)$ represents the different discrete components occurring at x_i with probability p_i. An $f(x)$ could be composed of either or both parts of the distribution. A history score of zero is included in $f(x)$ as the discrete component $\delta(x - 0)$. By the definition of a PDF,

$$\int_{-\infty}^{\infty} f(x)dx \equiv 1 \quad . \tag{2}$$

The mean or expected value $< x >$ of the history scores is the first moment of the history scores distribution; namely,

$$< x > = \int_{-\infty}^{\infty} x f(x)dx \quad . \tag{3}$$

The n^{th} central moment of x is defined to be

$$n^{th} \ central \ moment = \int_{-\infty}^{\infty} f(x)(x - < x >)^n dx \quad . \tag{4}$$

The variance σ^2 of the population of x's is defined to be the second central moment of x, which is equal to

$$\sigma^2 = < x^2 > - < x >^2 \quad , \tag{5}$$

where the expected value of the second moment $< x^2 > = \int_{-\infty}^{\infty} x^2 f(x)dx$. The square root of the variance is called the standard deviation σ and is a measure of the dispersion in the values of x.

Since $f(x)$ is seldom explicitly known, Monte Carlo samples the history score PDF implicitly using particle random walks. The true mean $< x >$ is estimated by

$$\overline{X} = \sum_{i=1}^{N} x_i / N \quad , \tag{6}$$

where x_i is the i^{th} history score to a tally bin. The estimated mean $\overline{X}$ is the sum of all of the history scores normed by the number of histories run; i.e., $\overline{X}$ is the average score per history for each tally bin.

The standard deviation σ is estimated by S where S is given by (and dropping the i subscript from the $\sum$ and assuming N is much greater than one)

$$S = \sqrt{\sum x_i^2 / N - (\sum x_i)^2 / N^2} \quad . \tag{7}$$

The estimated standard deviation of the mean is

$$S_{\overline{X}} \quad S / \sqrt{N} \quad . \tag{8}$$

This implies that $S_{\overline{X}}$ should decrease as the inverse of the square root of N as long as the estimate of σ, S, does not change markedly during the calculation.

For each tally bin, MCNP prints the estimated mean (Eqn. (6)) and the estimated relative error RE, where the RE is

$$RE = S_{\overline{X}}/\overline{X} \tag{9}$$

$$= \sqrt{\sum x_i^2/(\sum x_i)^2 - 1/N} \quad . \tag{10}$$

This quantity is convenient because the estimated statistical uncertainty in the result is given as a fraction of the result.

III. THE CENTRAL LIMIT THEOREM AND $f(x)$

The CLT states that the estimated mean will appear to be sampled from a normal distribution with a KNOWN standard deviation $\sigma/\sqrt{N}$ when N approaches infinity. In practice, σ is NOT known and must be approximated by the estimated standard deviation S. The major difficulty in applying the CLT correctly to a Monte Carlo result to form a confidence interval is knowing when N is large enough.

How can this be assessed? Several marginally quantifiable methods were mentioned in the introduction. A new method described in this paper involves examining the behavior of $f(x)$ for large history scores to attempt to assess if $f(x)$ appears to have been "completely" sampled. If "complete" sampling has occurred, the largest values of the sampled x's should have either nearly reached the upper limit (if such a limit exists) or should decrease faster than $1/x^3$ so that $< x^2 >= \int_{-\infty}^{\infty} x^2 f(x)dx$ exists (σ is assumed to be known in the CLT). Otherwise, N is assumed not to have approached infinity in the sense of the CLT. This is the basis of the proposed use of the empirical $f(x)$ to assess Monte Carlo tally convergence.

The argument should be made that since S must be a good estimate of σ, the expected value of the fourth history score moment $< x^4 >= \int_{-\infty}^{\infty} x^4 f(x)dx$ should exist. In the paper, we will assume that only the second moment needs to exist so that the $f(x)$ convergence criterion will be relaxed somewhat. Nevertheless, this assumption should not be forgotten.

IV. ANALYTIC STUDY OF $f(x)$ FOR TWO-STATE MONTE CARLO PROBLEMS

This project was divided into two parts: 1) derive and examine analytic Monte Carlo score distributions both to assess their general nature and to use for statistical confidence interval studies; and 2) create and examine empirical history score PDFs for real transport problems to assess their usefulness in predicting Monte Carlo convergence. For the first part, the tally distribution of history scores has been examined analytically[5,6] for both an analog two-state splitting and exponential transform problem. This work provides the theoretical foundation for statistical studies on relevant analytic functions to increase understanding of confidence interval coverage rates for Monte Carlo calculations.

It was found that the splitting problem history score PDF decreases geometrically as the score increases by a constant increment (this is equivalent to a negative exponential behavior for a continuous score PDF). The history score PDF for the exponential transform problem decreases geometrically with geometrically increasing x. Therefore, the splitting problem produces a linearly decreasing PDF for the history score on a linear-log plot of the score probability versus score. The exponential transform problem generates a linearly decreasing score behavior (with high score negative exponential roll off) on a log-log plot of the score probability versus score plot. In general, the exponential transform problem is the more difficult to sample because of the larger impact of the low probability high scores.

The analytic shapes were compared with comparable problems calculated with a modified version of MCNP. These shapes of the analytic and empirical $f(x)$s were in excellent agreement.[6]

V. PROPOSED USE FOR THE EMPIRICAL $f(x)$

In general, there has been very little discussion about the underlying or empirical $f(x)$ for Monte Carlo transport problems.[7] A new capability has been added to MCNP to allow inspection and analysis of the empirical $f(x)$ for the tally fluctuation chart bin of each tally. This should be important for assessing if there are any unexpected unsampled regions in the empirical history score PDF $f(x)$.

The most important proposed use for the empirical $f(x)$ is to attempt to determine if N has approached infinity in the sense of the CLT so that valid confidence intervals may be formed. The application of the CLT to Monte Carlo results to form a valid confidence interval requires the existence of the first two moments of $f(x)$. The history score PDF is assumed to have these properties for all of our current estimators and variance reduction techniques. (Point detectors with no constant flux neighborhoods in scattering materials and the exponential transform with a "large" [problem dependent] stretching parameter and no weight window are two known exceptions.)

It is assumed that the underlying $f(x)$ satisfies the CLT requirements. Therefore, so should the empirical $f(x)$. Unless there is a largest possible history score, the empirical $f(x)$ must eventually decrease more steeply than x^{-3} for the second moment ($\int_{-\infty}^{\infty} x^2 f(x)dx$) to exist. It is postulated[7] that if such decreasing behavior in the empirical $f(x)$ with no upper bound has not been observed, then N is not large enough to satisfy the CLT because $f(x)$ has not been completely sampled. Therefore, a larger N is required before a confidence interval can be formed. It is important to note that this convergence criterion is NOT affected by any correlations that may exist between the estimated mean and the estimated RE. In principle, this should make this $f(x)$ diagnostic robust in assessing "complete" sampling.

As mentioned in Section 3, the argument could be made that the fourth moment of $f(x)$ should exist so that the estimate of $S_{\overline{X}}$ in Eqn. (8) is valid for use in the CLT in the place if $\sigma/\sqrt{N}$. This would increase the decreasing behavior requirement to x^{-5} so that $\int_{-\infty}^{\infty} x^4 f(x)dx$ exists. We have elected to use the x^{-3} requirement because it is less stringent, but will still detect the many important cases of incomplete sampling. This loosened $f(x)$ criterion for convergence could be modified in the future as experience with the method increases.

Both the analytic and empirical history score distributions suggest that large score fill-in and one or more extrapolation schemes for the high score tail of the history score PDF could provide a meaningful estimate of scores not yet sampled to help assess the impact of unsampled history scores on the mean and confidence interval. This has not yet been considered.

VI. CREATION OF $f(x)$ IN MCNP

We wanted the creation of empirical $f(x)$s in MCNP to automatically cover nearly all tally fluctuation chart bin tallies that a user might reasonably be expected to make. We finally settled on using a logarithmically spaced history score grid for $f(x)$ because the tail behavior is assumed to be of the form $1/x^n, n > 3$ (unless an upper bound for the history scores exists). This would produce an equal bin width histogram of a straight line for $f(x)$ on a log-log plot that decreases n decades in $f(x)$ per decade increase in x.

We used 10 bins per x decade and covered the unnormalized tally range from 10^{-30} to 10^{30}. The term "unnormalized" indicates that normalizations that are not performed until the end of the problem, such as cell volume or surface area, are not included in $f(x)$. The user can multiply this range using MCNP input when the range is not sufficient. We keep track of both history score number and history score in the x grid to examine the cumulative number and score distributions. We use a linear fit to the logarithmically spaced x grid of the form $a + b \cdot ln(x)$ to find the grid location without a time-consuming search.

With this x grid in place, the average empirical $f(\overline{x_i})$ between x_i and x_{i+1} is defined to be

$$f(\overline{x_i}) = (number\ of\ history\ scores\ in\ i^{th}\ score\ bin)/(N(x^{i+1} - x^{i})) \quad , \tag{11}$$

where $x^{i+1} = 1.2589x^{i}$. The quantity 1.2589 is $10^{0.1}$ and comes from 10 equally spaced log bins per decade. The calculated $f(\overline{x_i})$s are available on printed plots in the output or using the new tally graphics commands. Any history scores that are outside the x grid are counted as either above or below to provide this information to the user.

Negative history scores are a possibility for some charge deposition tallies. The default procedure lumps any negative history score into the one bin below the lowest history score in the built-in grid. If desired, a $f(-x)$ can be created for negative scores. Positive history scores will then be lumped into the highest bin in this case.

VII. PARETO FIT TO THE LARGEST HISTORY SCORES

We estimate the slope n in $1/x^n$ of the largest history tallies x to determine if and when they decrease faster than $1/x^3$. This requires saving and sorting the largest history tallies at various points during the calculation. We use the "heapsort" sorting algorithm because it is reasonably fast and robust.[8]

We used the generalized Pareto function[9]

$$Pareto\ \ f(x) = a^{-1}(1 + kx/a)^{-1-1/k} \tag{12}$$

to fit the largest x's. This function fits a number of extreme value distributions including $1/x^n$, negative exponential ($k = 0$), and constant ($k = 1$). We developed a large history tally tail fitting technique using the "simplex" algorithm,[8] which finds the values of a and k that best fit the largest history scores by maximum likelihood estimation. Other algorithms, such as those using the derivative of the Pareto, could be used for increased speed, but the simplex method appears to be more robust and not that time consuming for this MCNP application.

The number of largest history score values to use for the fit is a variable that was investigated. We settled on a maximum of 201 points because this would provide about 10% precision[9] on the slope estimator at $n = 3$. The precision increases for smaller values of n and vice versa. The number of values actually used in the fit is the lesser of 5% of the nonzero history scores or 201. The minimum number of values used for a Pareto fit is 25 with at least two "different" values, which requires at least 500 nonzero history scores with the 5% criterion. We were careful about the implementation to be sure the technique would multitask and be reasonably efficient.

From the Pareto fit, we defined the slope of $f(x_{large})$ to be

$$slope \equiv 1 + 1/k \quad . \tag{13}$$

A slope value of zero is defined to indicate that not enough $f(x_{large})$ tail information exists for a slope estimate. The slope is not allowed to exceed a value of 10 (defined to be a "perfect score"), which would indicate an essentially negative exponential decrease. If the 100 largest history scores all have values with a spread of less than 1%, an upper limit is assumed to have been reached and the slope is set to 10. The slope should be greater than 3 to satisfy the second moment existence requirement. Then, $f(x)$ will appear to be "completely" sampled and hence N will appear to be large enough to satisfy the CLT.

VIII. MCNP TEST PROBLEM RESULTS AND ANALYSIS

Several test problems were run with a modified version of MCNP that calculates the empirical $f(x)$ and the Pareto estimate of the slope for the user-defined tally fluctuation chart bin of each tally. One test problem[3,7] was to calculate the surface neutron leakage flux above 12 MeV from an isotropic 14 MeV neutron point source of unit strength at the center of a 30 cm thick concrete shell with an outer radius of 390 cm. Point and ring detectors[1] were deliberately used to estimate the surface neutron leakage flux with highly inefficient, long tailed $f(x)$s.

The variance reduction methods employed[1] were implicit capture with weight cutoff, low-score point detector Russian roulette, and a 0.5 mean free path (4 cm) neighborhood around the detectors to produce large, but finite, higher moments. Other tallies or variance reduction methods could be used to make this calculation much more efficient, but that was not the object of this calculation.

One point detector result at 14,000 histories was 1.41×10^{-8} n/cm^2/sec with an estimated relative error of 0.041. The estimated $S_{\overline{X}}$ was decreasing as approximately $1/\sqrt{N}$ for the last half of the problem. Using the currently accepted rules-of-thumb for a valid Monte Carlo result, the confidence interval formed from this tally could easily be accepted by even a careful Monte Carlo practitioner.

The correct detector result, obtained from a 50 million history ring detector tally, is 5.68×10^{-8} n/cm^2/sec with an estimated relative error of 0.0054 and a slope of 4.9. (The ring detector result for 14,000 histories was 4.60×10^{-8} n/cm^2/sec with an estimated relative error of 0.17 and a slope of 2.1.) The apparently converged 14,000 history point detector result is a factor of four below the correct result!

The new convergence criterion proposed here provides compelling evidence that an N of 14,000 is not large enough and a confidence interval should NOT be formed. The large x slope is a very shallow 1.4 (i.e., $1/x^{1.4}$), which could not continue indefinitely because the mean would not exist. This is a clear indicator that the unbounded $f(x)$ has not yet been completely sampled.

Figure 1 compares the empirical point detector $f(x)$s for 14,000 and 200 million histories. The 14,000 history $f(x)$ clearly has unsampled regions in the tail, which indicate incomplete $f(x)$ sampling. For the point detector, over seven decades of x have been sampled by 200 million histories compared to only three decades for 14,000 histories. The largest x's occur from the extremely difficult to sample histories that have multiple small energy loss collisions close to the detector. The 200 million history point detector result is 5.44×10^{-8} n/cm^2/sec with an estimated relative error of 0.036 and an estimated slope of 2.4. The point detector $f(x)$ slope is increasing, but still does not yet appear completely sampled. The more compact empirical ring $f(x)$ for 50 million histories, shown in Fig. 1, appears to be completely sampled because of the large slope.

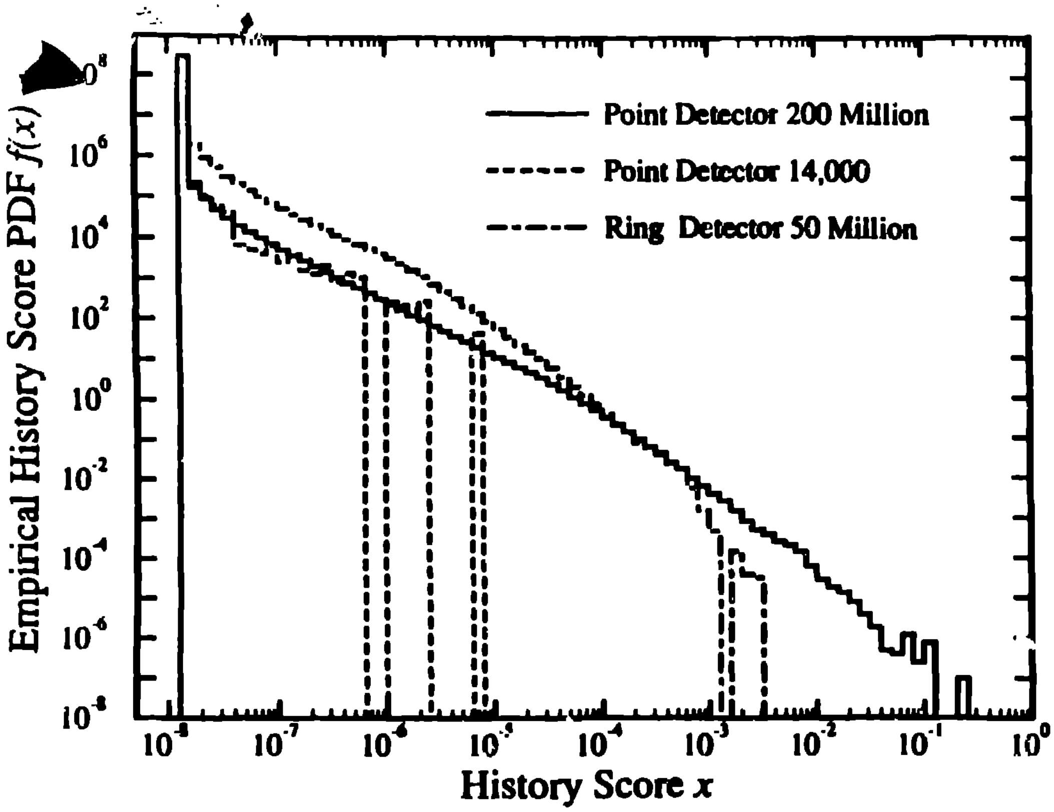

Figure 1. Empirical point detector $f(x)$s for 14,000 and 200 million neutron histories and the empirical ring detector $f(x)$ for 50 million neutron histories versus the history score x.

IX. CONCLUDING REMARKS

The proposed $f(x)$ convergence criteria should provide new insight into the quality of a Monte Carlo result. The empirical $f(x)$ is available for the first time for user to examine. The Pareto fit for the slope is hoped to be a useful new diagnostic for predicting the importance of large scores that have not yet been sampled. The pathological example discussed above was easily identified as a result for which a confidence interval should NOT be formed.

This new capability has been incorporated into the latest version of MCNP. Additional experience with this diagnostic is required to determine its effectiveness in assessing the fulfillment of the central limit theorem requirements for Monte Carlo solutions. It is emphasized that even an apparently well sampled $f(x)$ may have an important, but improbable, portion that has not yet produced a large history score. *CAVEAT EMPTOR*.

REFERENCES

1. J. F. Briesmeister, Ed., "MCNP – A General Monte Carlo Code for Neutron and Photon Transport," LA–7396–M, Rev. 2, Los Alamos National Laboratory (1986).

2. M. H. Kalos, P. A. Whitlock, ***Monte Carlo Methods, Volume I: Basics***, John Wiley & Sons, p. 27 (1986).

3. G. P. Estes, E. D. Cashwell, "MCNP1B Variance Error Estimator," Los Alamos National Laboratory memo TD-6-27-78 (1978).

4. W. J. Vieira, "A General Study of Undersampling Problems in Monte Carlo Calculations," Doctoral Dissertation, University of Tennessee (1989).

5. T. E. Booth, "Analytic Comparison of Monte Carlo Geometry Splitting and Exponential Transform," *Trans. Am. Nucl. Soc.*, **64**, 303 (1991).

6. T. E. Booth, "Analytic Monte Carlo Score Distributions for Future Statistical Confidence Interval Studies," *Nucl. Sci. Eng.*, **112**, 159 (1992).

7. R. A. Forster, "A New Method of Assessing the Statistical Convergence of Monte Carlo Solutions," *Trans. Am. Nucl. Soc.*, **64**, 305 (1991).

8. W. H. Press B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, ***Numerical Recipes (FORTRAN Version): The Art of Scientific Computing***, p. 229 and 289, Cambridge University Press (1990).

9. J. R. M. Hosking and J. R. Wallis, "Parameter and Quantile Estimation for the Generalized Pareto Distribution," *Technometrics*, **29**, 339 (1987).