Journal of the Arkansas Academy of Science

Volume 52 Article 12

1998

Applying Binomial Statistics to Weighted Monte Carlo

Sue Ellen McCloskey
University of Arkansas at Little Rock

William C. Hall
University of Arkansas at Little Rock

Wilfred J. Braithwaite
University of Arkansas at Little Rock

Applying Binomial Statistics to Weighted Monte Carlo

Sue Ellen McCloskey, William C. Hall and Wilfred J. Braithwaite*
Department of Physics and Astronomy,
University of Arkansas at Little Rock
Little Rock, AR 72204

*Corresponding Author

Abstract

Weighted Monte Carlo calculations requiring a uniform sampling of the problem-space can suffer from diminished statistical significance because many, if not most, of the randomly-chosen sampling points contribute only slightly to the desired result. Their contribution is reduced in size due the variable-size of the weighting terms. In contrast, none of the randomly-chosen points which are *favored* by variablesize weighting terms will have their statistical significance enhanced beyond that of just *one random point* in the Monte Carlo sampling. A Monte Carlo analysis was used in earlier work to verify both Gauss' Law and Newton's Shell Theorem. Both examples suffered from statistical difficulties since each Monte Carlo sampling point has a weight inversely proportional to the square of the distance between source and field points. The present work analyzes the diminished significance in weighted Monte Carlo for the specific example of Newton's Shell Theorem, describing the geometry in terms of closest approach distance of the spherical mass shell to the field point. Binomial Statistics is used to remedy this diminished statistical significance by providing a prescription for increasing the value of the Monte Carlo sample size needed to assure that the chosen precision remains invariant as the mass-shell geometry is changed.

Introduction

A weighted-form of Monte Carlo is applicable to many problems in physics and engineering. Newton's Shell Theorem and Gauss' Law are two famous examples which may be tested using a weighted-form of Monte Carlo analysis. The $1/\Delta R^2$ weighting terms in each of these examples can result in numerical infinities if the sampling allows any overlap between source and field points. This problem of singularities is a numerical nuisance for Monte Carlo, not a fundamental problem in the underlying formulation. These are specious infinities because singularities rarely occur in the analytical formulations. In each example, Newton's Shell Theorem and Gauss' Law, the offending $1/\Delta R^2$ is cured analytically by cancellations, resulting, at worst, in a predicted discontinuity in the observable (Halliday and Resnick, 1988).

Either a weighted-Monte Carlo problem must be reformulated to no longer require a uniform sampling of the entire space, a programmer-intensive task–not certain to succeed, or a large number of samples must be taken to remedy the diminished statistical significance of lower-weighted samples (Mikhailov, 1992). Figure 1 develops Binomial Statistics (Boaz, 1983) to provide an analysis for weighted Monte Carlo problems which works almost everywhere, even near Monte Carlo singularities. The present binomial analysis results in a rationale and a prescription for just how large the Monte Carlo sampling-size N must be as a function of ΔR_{CA} to assure the chosen level of precision. ΔR_{CA} is the closest approach distance between the uniformly sampled (source) points on the spherical shell of mass and the field point, just above the earth's surface.

Since the random events in Monte Carlo are independent of each other, Binomial Statistics is expected to apply to Monte Carlo. The binomial success fraction "p" is hypothesized to exist and have a unique value for each Monte Carlo geometry selected. Three variables (p, N, ε) are related in Fig. 1: $p = 1/[N\varepsilon^2 + 1]$, with N the number of independent event samples and ε the fractional variance in the observable.

The longitudinal force component of Newton's Shell Theorem was the chosen observable (Halliday and Resnick, 1988). A Monte Carlo analysis was used to test the hypothesis that the binomial probability of success "p" exists and has a unique value for any particular choice of geometry. Figure 2 shows an example of Monte Carlo geometry where the uniform-mass shell is located 90% of the distance to the earth's surface.

Each set of Monte Carlo calculations was repeated 1000 times for each of 24 distinct values of the parameter N. Each set of N Monte Carlo samplings of points on the mass shell provided a value for the longitudinal force component (F) at some precision σ_F. One Thousand repeated Monte Carlo samples were used to extract an estimate of this precision σ_F from the sum of the squares of the deviations of F from its average value $\langle F \rangle$ (Bevington and Robinson, 1992). This

Binomial Statistics

$\boxed{p+q=1}$

p = probability of **success, q= probability** of **failure.**
p & q are **unchanged** in **N repeated events.** N events unrelated $\Rightarrow$ outcome probability a simple product.

$n_1 = \#$ of successes $\Rightarrow p^{n_1}$; $n_2 = \#$ of failures $\Rightarrow q^{n_2}$.

$\frac{N!}{n_1!n_2!}$ =*multiplicity* for $\boxed{p^{n_1}q^{n_2}}$ term, $\boxed{n_1+n_2=N}$.

$\boxed{p\frac{\partial}{\partial p}}$ & $\boxed{p\frac{\partial}{\partial p}p\frac{\partial}{\partial p}}$ provide $<n_1>$ and $<n_1^2>$ below.

$$\boxed{(p+q)^N \equiv \sum_{n_1=0}^{N} \frac{N!}{n_1!n_2!} p^{n_1} q^{n_2} \rightarrow 1, \quad \boxed{n_1+n_2=N}.}$$

$$<n_1> \equiv \sum_{n_1=0}^{N} n_1 \frac{N!}{n_1!n_2!} p^{n_1}q^{n_2} \equiv p\frac{\partial}{\partial p} \sum_{n_1=0}^{N} \frac{N!}{n_1!n_2!} p^{n_1}q^{n_2}$$

$$<n_1^2> \equiv \sum_{n_1=0}^{N} n_1^2 \frac{N!}{n_1!n_2!} p^{n_1}q^{n_2} \equiv p\frac{\partial}{\partial p} p\frac{\partial}{\partial p} \sum_{n_1=0}^{N} \frac{N!}{n_1!n_2!} p^{n_1}q^{n_2}$$

$$p\frac{\partial}{\partial p}(p+q)^N \equiv Np(p+q)^{N-1}, \ (p+q=1) \Rightarrow \boxed{<n_1> = Np}$$

$$\left[p\frac{\partial}{\partial p}p\frac{\partial}{\partial p}\right] = \left[p\frac{\partial}{\partial p} + p^2\frac{\partial^2}{\partial p^2}\right], \left[p\frac{\partial}{\partial p} + p^2\frac{\partial^2}{\partial p^2}\right](p+q)^N \equiv$$

$$<n_1^2> = Np(p+q)^{N-1} + N(N-1)p^2(p+q)^{N-2}, \ (p+q=1) \Rightarrow$$

$$\boxed{<n_1^2> = Np+N(N-1)p^2}. \quad \boxed{\begin{array}{l}\text{Fractional variance}\\ \text{(in the observable)}\end{array}} \equiv \varepsilon \equiv \frac{\sigma_1}{<n_1>} \equiv$$

$$\frac{\sqrt{<n_1^2> - <n_1>^2}}{<n_1>} = \frac{\sqrt{Np+N(N-1)p^2 - N^2p^2}}{Np} = \frac{\sqrt{Np - Np^2}}{Np}$$

$$\frac{\sigma_1}{<n_1>} \equiv \boxed{\varepsilon = \frac{\sqrt{(1-p)/p}}{\sqrt{N}}} \Rightarrow \boxed{p = \frac{1}{N\varepsilon^2+1}} \text{ and } \boxed{N = \frac{1-p}{p\varepsilon^2}}.$$

Fig 1. Binomial Statistics is outlined for an analysis of Monte Carlo, providing three (equivalent) relationships between three variables: ε, P and N.

estimate of σ_F is used to approximate the fractional variance in the observable, $\varepsilon = \sigma_F/\langle F\rangle$; ε is used with the parameter N to calculate a value for the binomial probability of success: $p = 1/[N\varepsilon^2 + 1]$. It is interesting and obvious that $N\varepsilon^2$ must form an invariant product for any particular (chosen) geometry or the binomial success fraction "p" would not exhibit a constant (unique) value, independent of N. for the geometry selected in Fig. 2. If the value of "p" did not exhibit a unique value (i.e., if "p" were not independent of N) then Binomial Statistics would not provide a useful analysis of weighted Monte Carlo problems.

Using this two-level Monte Carlo analysis, a value was extracted for "p" for each value of the parameter N and displayed in Fig. 2. The constancy of "p" over a dynamic range in N of 1000:1 empirically verifies it to be unique for the geometry specified in Fig. 2. A value for "p" was calculated for each of the 24 different values between 1,000 and 1,000,000 chosen for N using the numerically extracted (fractional variance) ε calculated for each value of the parameter N. The resulting "p" is seen to have a constant value independent of N. This result means that $N\varepsilon^2$ does indeed form an invariant product for the chosen geometry (see Fig. 2) or the value of "p" would not be constant and independent of N.

Materials and Methods

The present work examines the effort to verify Sir Isaac Newton's Shell Theorem using weighted Monte Carlo. One version of Newton's Shell Theorem states (1) For any field point *outside* a uniform spherical mass shell, the shell acts as though all its mass were concentrated at its symmetry center. [Only longitudinal forces survive as transverse forces vanish by symmetry.] (2) For any field point located *inside* a uniform spherical mass shell, the shell acts as though its mass were zero (i.e., the mass shell provides no net gravitational force on an enclosed field point).

Binomial Statistics was chosen as the vehicle for studying the onset of diminished statistical significance in this weighted Monte Carlo problem. Earlier work (McCloskey and Braithwaite, 1995) used Monte Carlo to verify both Gauss' Law and Newton's Shell Theorem. Both these efforts are examples of the use of weighted Monte Carlo calculations since each sampling point is weighted inversely by the square of the distance between source and field points.

Binomial Statistics is relevant to this class of weighted Monte Carlo problems since for each geometrical situation there is a unique value for the sampling probability (interpreted as the binomial probability of success). As required for binomial statistics to be valid, each sampling probability is independent of all prior and future samplings. Further, this probability of success is $p \rightarrow 1$ when all events are weighted equally.

In verifying Newton's Shell Theorem using Monte Carlo, $p \approx 1$ when the distance between source and field points is much larger than the extended size of the source-distributions of mass (as in the earth-moon system). [The Monte Carlo points plotted in Fig. 5 show $p \approx 1$ for $R \approx 0$ (since the distance of closest approach is $\Delta R_{ca} \approx 1$); all Monte Carlo point events on the spherical mass shell are weighted approximately equally.] Newton developed his Shell Theorem to determine the force of gravity at the earth's surface by viewing the mass of the earth as distributed in uniform, concentric spherical shells arranged much like the layers of an onion (Halliday and Resnick, 1988).

When the closest-approach distance is small compared to size-distribution of source-masses, only the closest points on the mass shell contribute significantly to the Monte Carlo sampling of force components at the (nearby) field point. The farther-away points are relegated, progressively, to statistical insignificance by the $(\Delta R)^{-2}$ weighting of the Monte Carlo samples. The size of binomial probability of success "p" is expected to fall monotonically from unity for each mass-shell geometry, as the distance of closest approach ΔR_{ca} progressively decreases.

Figure 2 shows a spherical mass shell for $\Delta R_{ca} = 0.1$ (R = 0.9 and the earth's radius = 1). Mass points near the field point on the earth's surface are favored with a weighting of $\approx 1/0.1^2 = 100$ in contrast to mass points on the far side of the shell which are disfavored with a weighting of $\approx 1/2^2 = 1/4$. This means these "near points" are weighted by about a factor of ≈ 400 over the "far points."

For $\Delta R_{ca} = 0.01$ (and $R = 0.99$) "near points" are favored by $\approx 40{,}000$ over the "far points." As ΔR_{ca} approaches zero, the probability of success for any particular sample "p" also approaches zero. The radius of the contributing part of the shell (associated with "near points") scales with ΔR_{ca}, so the area of the contributing part of the shell (associated with "near points") scales as $(\Delta R_{ca})^2$, and asymptotically, one might expect $p \propto (\Delta R_{ca})^2$. It is well known that a power law graphs as a straight line on a log-log plot with the slope of the graph being the "power." The asymptotic variation of "p" is shown in the final analytical calculation of Fig. 4 as well as in the log-log graph of Fig. 5. Asymptotically both analytical and graphical presentations show $p \propto (\Delta R_{ca})^2$.

Results and Discussion

Monte Carlo was used (above) to verify the hypothesis that the binomial success probability "p" exists and has a unique value for $\Delta R_{ca} = 0.1$ (one particular geometry). The fractional variance ε in the observable was determined for each value of the parameter N, and "p" was calculated using the formula: $p = 1/(N\varepsilon^2 + 1)$. Results of these calculations

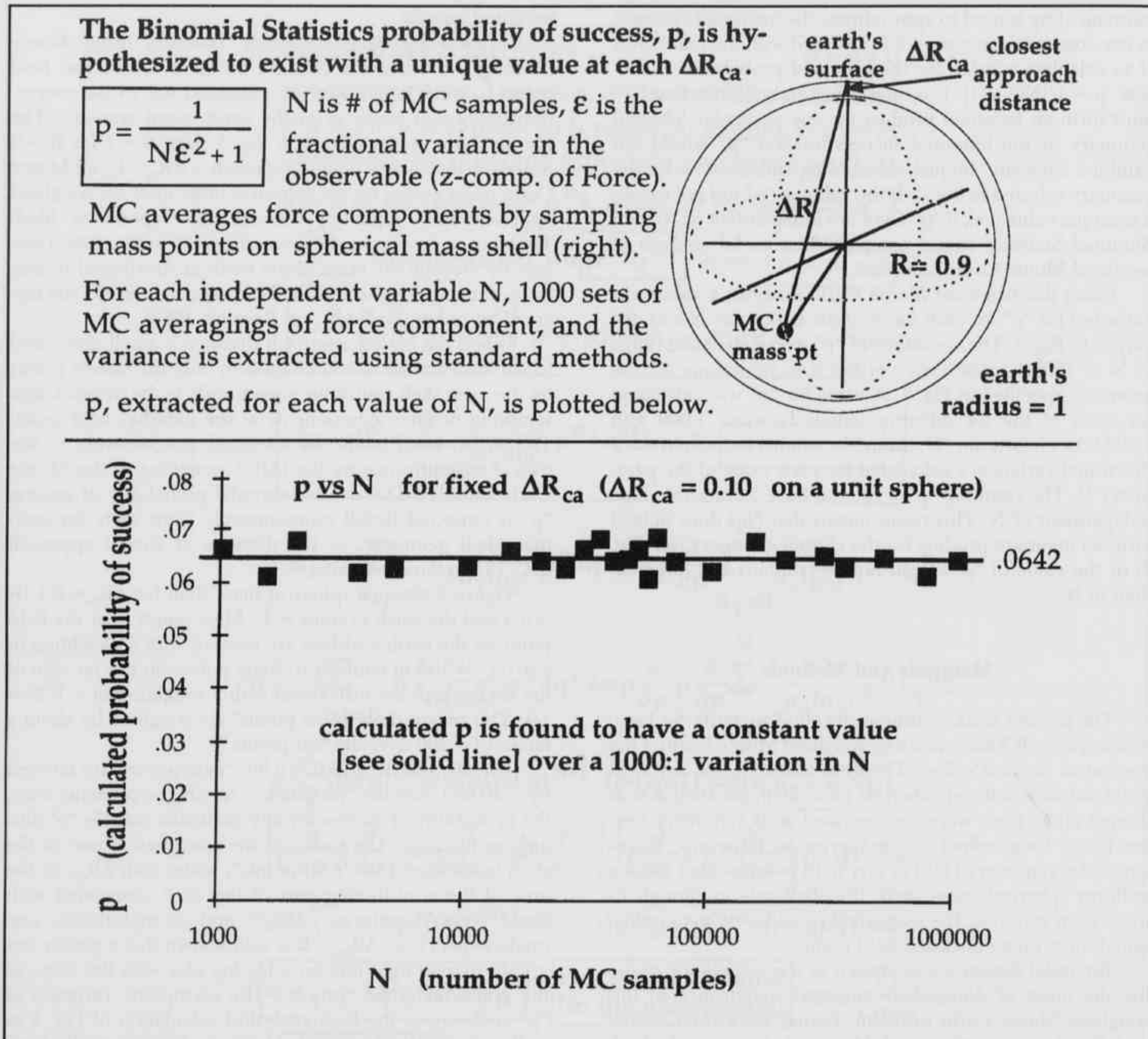

Fig. 2. p is extracted using two-level Monte Carlo analysis: N is chosen as an independent parameter in each set, with ε determined by standard methods from 1000 sets of averagings.

are shown in Fig. 2. "p" was seen to have a unique value for a 1000:1 variation in the parameter N. Flutter in these Monte Carlo points (plotted as squares) is consistent with a fractional uncertainty of ≈3%, about the size of the squares. (A unique value for "p" for the chosen geometry occurs only if $N\varepsilon^2$ ≈ constant.)

Figure 3 begins the next step, which is to calculate the fractional variance in the transverse force component analytically, to calculate a unique "p" as a function of the closest approach distance ΔR_{ca}. ΔR_{ca} is sufficient to specify the geometry of each mass shell in its relationship to the field point (just above the surface of the earth). Figure 3 provides a diagram and outlines a method for calculating the fractional variance in the observable, by analytically averaging

Using Binomial Statistics in an Analysis of Newton's Shell Theorem for Monte Carlo

The longitudinal component of the Force is:

$$F_{sz} = \frac{\cos\phi}{s^2}$$

Find $\langle F_{sz} \rangle$ by averaging this force component over the spherical shell (below).

Then find $\langle F_{sz}^2 \rangle$ by averaging the square of the force component over the spherical shell.

The goal is to find the fractional variance in this observable to obtain N needed for MC.

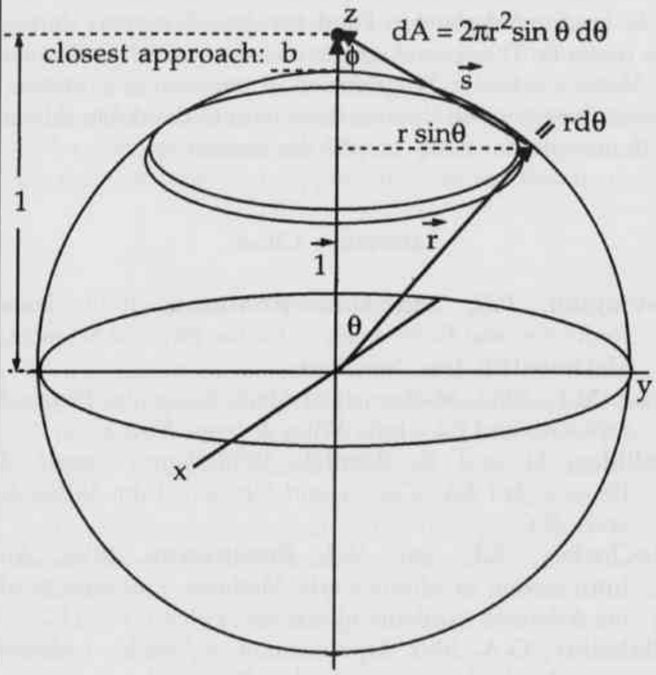

Fig. 3. The longitudinal component of the gravitational force (from a spherical shell) is selected as the observable whose fractional variance is needed in the analysis of Newton's Shell Theorem to determine the Monte Carlo sample size, N=N(p, ε), to assure the precision is unaffected by geometry.

Binomial probability of success, p, for a single MC event

Force (longitudinal component): $F_{sz} = \frac{\cos\phi}{s^2}$

$$\langle F_{sz} \rangle = \frac{1}{4\pi r^2}\int_0^\pi \frac{\cos\phi}{s^2}\, 2\pi r^2 \sin\theta\, d\theta \quad \text{and} \quad \langle F_{sz}^2 \rangle = \frac{1}{2}\int_0^\pi \frac{\cos^2\phi}{s^4} \sin\theta\, d\theta$$

vectors give geometrical relations:

$$\vec{r} = \vec{1} - \vec{s} \Rightarrow r^2 = 1 + s^2 - 2s\cos\phi \Rightarrow \cos\phi = \frac{1 - r^2 + s^2}{2s}$$

$$\vec{s} = \vec{1} - \vec{r} \Rightarrow s^2 = 1 + r^2 - 2r\cos\theta \Rightarrow 2s\, ds = -2r(-\sin\theta)\, d\theta$$

substitute:

$$\langle F_{sz} \rangle = \frac{1}{2}\int_0^\pi \frac{\cos\phi}{s^2} \sin\theta\, d\theta = \frac{1}{2r}\int_{1-r}^{1+r} \frac{1}{s^2}\frac{1 - r^2 + s^2}{2s}\, s\, ds =$$

$$\frac{1}{4r}\int_{1-r}^{1+r} \frac{1 - r^2 + s^2}{s^2}\, ds = \frac{1}{4r}\left[-\frac{(1-r)(1+r)}{(1+r)} + \frac{(1-r)(1+r)}{(1-r)} + 1 + r - (1-r)\right]$$

$$\boxed{\langle F_{sz} \rangle = 1} = \frac{4r}{4r}.$$

$$\langle F_{sz}^2 \rangle = \frac{1}{2r}\int_{1-r}^{1+r} \left[\frac{1}{s^2}\frac{1 - r^2 + s^2}{2s}\right]^2 s\, ds =$$

$$\frac{1}{8r}\int_{1-r}^{1+r} \left[\frac{(1-r)^2(1+r)^2}{s^6} + 2\frac{(1-r)(1+r)}{s^4} + \frac{1}{s^2}\right] s\, ds = \frac{1}{8r}\left[-\frac{(1-r)^2(1+r)^2}{4s^4} - 2\frac{(1-r)(1+r)}{2s^2} + \ln(s)\right]_{1-r}^{1+r}$$

$$= \frac{1}{8r}\left[-\frac{(1-r)^2(1+r)^2}{4(1+r)^4} + \frac{(1-r)^2(1+r)^2}{4(1-r)^4} - \frac{(1-r)(1+r)}{(1+r)^2} + \frac{(1-r)(1+r)}{(1-r)^2} + \ln\left\{\frac{1+r}{1-r}\right\}\right] \cong \frac{1}{8b^2}$$

$$\boxed{\langle F_{sz}^2 \rangle = \frac{1}{8r}\left[\frac{(1+r)^2}{4(1-r)^2} - \frac{(1-r)^2}{4(1+r)^2} + \frac{1+r}{1-r} - \frac{1-r}{1+r} + \ln\left\{\frac{1+r}{1-r}\right\}\right] \cong \frac{1}{8b^2}}$$

closest approach: $b \equiv 1-r$, $r \cong 1$ $1+r \cong 2$, ε = chosen precision

$$\boxed{\sigma^2(F_{sz}) = \varepsilon^2(F_{sz}) = \langle F_{sz}^2 \rangle - 1 \cong \frac{1}{8r\, b^2} - 1} \quad \boxed{N = \frac{1-p}{p\,\varepsilon^2} \cong \frac{1}{8b^2\varepsilon^2}}$$

Binomial probability of success, p, for a single MC event is:

$$\boxed{p = \frac{1}{\varepsilon^2(F_{sz}) + 1} = \frac{1}{\frac{1}{8r}\left[\frac{(1+r)^2}{4(1-r)^2} - \frac{(1-r)^2}{4(1+r)^2} + \frac{1+r}{1-r} - \frac{1-r}{1+r} + \ln\left\{\frac{1+r}{1-r}\right\}\right]} \cong 8b^2}$$

Fig. 4. The fractional probability of success (p) for fixed geometry is calculated via Binomial Statistics from the fractional variance in the longitudinal component of the gravitational force. Approximate values of p and N are also provided.

F_{sz} and $(F_{sz})^2$ over a spherical shell of radius r.

An expression which works for almost all spherical shells of radii (i.e., $0 \leq R < 1$) was calculated for the binomial probability of success "p" in Fig. 4 and plotted as a solid line in Fig. 5. The analytical calculation of Fig. 4 is seen in Fig. 5 to agree well with a collection of individual Monte Carlo calculations for "p," which are represented in Fig. 5 by squares, with each repeated Monte Carlo calculation represented by a diamond. In addition, a dotted line shows the asymptotic prediction for "p" which is seen to improve progressively for decreasing values of closest approach ($b \equiv \Delta R_{ca}$) in the range $0 < b < 0.1$.

Figure 4 outlines the calculation that provides an exact analytical prediction for the binomial probability of success "p," as well as providing approximate values for sufficiently

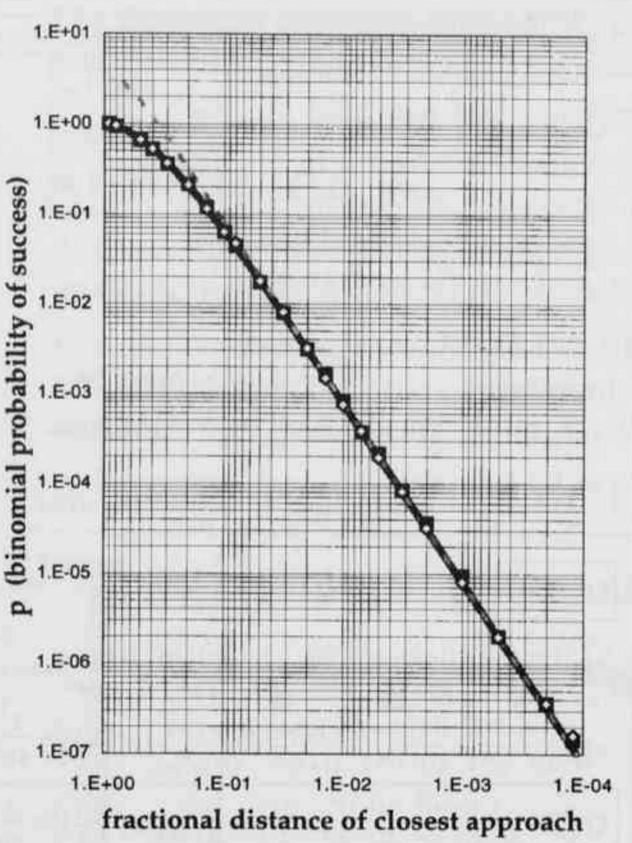

Fig. 5. The calculated probability of success (p) is plotted versus fractional distance of closest approach. Square and diamond data points are from repeated calculations. The dotted line is asymptotically correct for p as r approaches l.

small values of the closest approach (i.e., for $b \ll r \approx 1$: $p \approx 8b^2$). The dotted line in Fig. 5 shows the asymptotic prediction for "p" [$p = 8(\Delta R_{ca})^2 = 8b^2$] is an excellent approximation for $b \equiv \Delta R_{ca} < 0.05$, as the dotted graph lies exactly on top of the solid line (representing the exact prediction) to within the accuracy of the display.

Figure 5 is a log-log plot of the predictions for "p" from the exact calculation (solid), from the asymptotic calculation (dotted) and from the repeated Monte Carlo calculations (squares and diamonds). The exact and Monte Carlo predictions are seen to be in excellent agreement, with the dotted line showing the expected asymptotic variation: $p = 8(\Delta R_{ca})^2 \propto (\Delta R_{ca})^2$ for small $b \equiv \Delta R_{ca}$. In terms of the calculations of Fig. 4, only the first term in the denominator for "p" is retained to provide the asymptotic prediction. A straight-line variation on a log-log plot is the result of a power law, where the *slope* of the plot provides the value of the exponent. Note the graphical agreement in the asymptotic region, where "p" drops two orders of magnitude for every one order of magnitude decrease in the fractional distance of closest approach.

Since "p" is known it may be used to the remedy the diminished statistical significance in the Monte Carlo by providing the desired prescription for increasing the sample size (N) needed to assure that the chosen precision (ε) remains invariant as mass-shell geometry is changed. $N = (1 - p)/(p\varepsilon^2)$ is the exact prescription using values for p from Fig. 5. The asymptotic $p \approx 8b^2$ provides an asymptotic prescription for $N \approx .125/(b\varepsilon)^2$, where $b \equiv \Delta R_{ca}$.

ACKNOWLEDGMENTS.—This research was supported in part by the U. S. Department of Energy through Grant DE-FG05-92ER-40753. The first author wishes to thank the Rosa Isacson Scholarship Fund for critical support during this research. The second author wishes to thank the Ronald E. McNair Scholars Program for its support as a McNair Research Fellow. All three authors wish to thank Dr. Edwin S. Braithwaite for many helpful discussions.

Literature Cited

Bevington, P.R, and **D.K. Robinson.** 1992. Data Reduction and Error Analysis for the Physical Sciences. McGraw Hill, Inc., New York.

Boaz, M.L. 1983. Mathematical Methods for the Physical Sciences (2nd Ed.). John Wiley & Sons, New York.

Halliday, D. and **R. Resnick,** 1988. Fundamentals of Physics. 3rd Ed. (Ch. 15 and Ch. 25). John Wiley & Sons, NY.

McCloskey, S.E. and **W.J. Braithwaite.** 1995. An Introduction to Monte Carlo Methods. Proceedings of the Arkansas Academy of Science, Vol. 49, 109-114.

Mikhailov, G.A. 1992. Optimization of Weighted Monte Carlo Methods. Springer-Verlag, New York.