NASA CR-122409

N72-24291

NUS-786/3
Volume 3

FINAL REPORT FOR MODIFICATIONS OF CODES NUALGAM AND BREMRAD

Volume 3: Statistical Considerations of the Monte Carlo Method

For

NATIONAL AERONAUTICS AND
SPACE ADMINISTRATION
Goddard Space Flight Center
Glenn Dale Road
Greenbelt, Maryland

NASA Contract Number NAS5-11781

By

H. Firstenberg

May 1971

NUS CORPORATION
4 Research Place
Rockville, Maryland

NUS CORPORATION

FINAL REPORT FOR
MODIFICATION OF CODES NUALGAM AND BREMRAD

Volume 3: Statistical Considerations of the
Monte Carlo Method

For

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
Goddard Space Flight Center
Glenn Dale Road
Greenbelt, Maryland

NASA Contract Number NAS5-11781

By

H. Firstenberg

May 1971

NUS CORPORATION
4 Research Place
Rockville, Maryland

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

This present Volume III of the final report NUS-786 to the National Aeronautics and Space Administration, Goddard Space Flight Center, was prepared by the NUS Corporation under contract NAS5-11781.

This volume considers the statistics of the Monte Carlo method relative to the interpretation of the NUGAM2 and NUGAM3 computer code results. A numerical experiment using the NUGAM2 code is presented and the results are statistically interpreted.

Section 2 presents the general theoretical considerations of the Monte Carlo method. Supporting theory is given in Appendices 1 and 2. Section 3 consists of an example application of the theory.

2. STATISTICAL CONSIDERATIONS OF THE MONTE CARLO METHOD

2.1 Introduction

The Monte Carlo method is a technique for the solution of physical and mathematical problems by the application of random sampling methods. If formulating a problem by this method, one constructs some random process in which the random variables assume the role of the physical quantities of interest. The procedure then consists of making observations on the random process and its statistical characteristics. In general, the random process need not bear any relation to the actual physical problem, and, consequently, the Monte Carlo method has found application in the solution of deterministic problems as, for example, the evaluation of multidimensional integrals . However, the most extensive and successful use of the method has been, and continues to be, for the solution of problems for which the inherent physics requires a statistical description.

In the case of radiation transport phenomena the analog random process can be quite straight forward; namely, one randomly introduces an elementary particle into a medium, allows the particle to interact with the medium in accord with the detailed microphysics of the problem, and then tallies the number of particles involved in particular types of interactions. This procedure yields a relative frequency for the interactions, which, in turn, give a measure of the intrinsic probability for the events. The accuracy of this prediction is intuitively related to the number of elementary particle histories used to interrogate the random process, but it is clearly desirable to quantify what is meant by the term accuracy in the context of a Monte Carlo solution. This report addresses the question of interpreting the results of the Monte Carlo method when a limit number of "plays" are made with the random process.

Generally texts on the Monte Carlo method [1,2,3] give a perfunctory
treatment to the topic under consideration, providing almost no back-
ground on mathematical statistics necessary for an understanding of
the problem. There are many good texts on probability theory and mathe-
matical statistics [4,5,6], but these offer far more information than is
necessary in a discussion of the Monte Carlo technique. Thus, it is felt
that some benefits might accrue by including in one report a discussion of
both the salient theory and its application to the Monte Carlo method.
Rather than unduly complicate the main text with laborious mathematical
derivations, all but the most essential mathematics have been confined
to the appendices. With this format, they provide a coherent and con-
venient reference for those readers who are unfamiliar with the theory
of random sampling by the Bernoulli scheme.

2.2   General Comments on the Accuracy of the Monte Carlo Method

From a practical point of view the Monte Carlo method can be considered
a numerical experiment performed on a high-speed digital computer in
which the outcome of repetitive Bernoulli trials are used to simulate a
physical problem. In such an experiment, the Bernoulli law of large
numbers (Appendix 1, page 25) states that the frequency of occurrence
of an arithmetic mean, Y_n, in the interval $|Y_n - p| > \epsilon$ approaches zero
as one increases the number, n, of the Bernoulli trials. This can be
expressed mathematically in the form:

$$P\left(|Y_n - p| \geq \epsilon\right) \leq (n\epsilon^2)^{-1} \tag{1*}$$

where ϵ is any number greater than zero and p is the probability for the
"successful" event. Strictly speaking this must be interpreted in a pro-
babilistic manner, since the possibility must be admitted for the realization

---

*Since all possible events within a sample space must occur with a proba-
bility of unity, then equation (1) can also be written

$$P\left(|Y_n - p| < \epsilon\right) = 1 - P\left(|Y_n - p| \geq \epsilon\right) \tag{1a}$$

of the same outcome (i.e., $Y_n = 0$ or $Y_n = 1$) for any arbitrary n. The law of large numbers merely insures us that such "pathological results" are expected to occur with a vanishingly small frequency as n becomes sufficiently large. Thus, equation (1) or (1a) gives a measure of the confidence one can place on an experimentally determined Y_n being within $\pm\epsilon$ of the intrinsic probability p; it does not preclude the possibility that the Y_n in a given experiment exceeds p by more than $\pm\epsilon$ no matter how many Bernoulli trials are performed.

Equation (1) has important consequences with respect to the accuracy of data obtained by the Monte Carlo method. Obviously ϵ is a measure of the error in the experiment. If the average time to perform one trial in the experiment is denoted by τ and if $\epsilon \propto n^{-1/2}$, then the total time to perform n trials is approximately:

$$T = n\,\tau \propto \frac{\tau}{\epsilon^2} \qquad . \tag{2}$$

Equation (2) shows that an improvement of one order of magnitude in the accuracy of the experiment is purchased at the expense of a hundred-fold increase in the time for the experiment. Furthermore the increase in accuracy is obtained without any increase in the confidence; i.e., $P(|Y_n - p| > \epsilon)$ remains unchanged although the number of Bernoulli trials is increased*. To bring this into better focus, the NUS photon transport code NUGAM2 may require about 1 second to perform 150 photon history calculations. Typically a Monte Carlo calculation is performed with 10^4 photon histories thereby requiring approximately 50 seconds on an IBM-360/91 (not including compilation time). If it were desired to increase the accuracy of the calculation by one-order-of-magnitude, it would be necessary to perform 10^6 photon history calculations and computer running time would be increased

---

*This can be seen by substituting $\alpha \sqrt{n} = \epsilon^{-1}$ in equation (1):
$$P(|Y_n - p| > (\alpha \sqrt{n})^{-1}) \leq \alpha^2$$
so that the relative frequency remains less than or equal to α^2 for $|Y_n - p| \geq \epsilon$

4

to about 1.4 hours.

The absolute error in a calculation is less meaningful than the relative error; i.e.,

$$\left| \frac{Y_n}{p} - 1 \right| < \epsilon_r \propto \frac{1 - p}{np} \tag{3}$$

In terms of the relative error ϵ_r, the number of trials required to achieve a given accuracy will vary inversely as the "intrinsic probability for the event" p. Obviously, a direct modeling of the problem becomes impracticable when the event being investigated has a small probability thus, one must be willing to accept less accuracy in the result or change the strategy of the experiment to enhance the probability of success. For example, if the probability for the transmission of photons through a shield is of the order of 10^{-4}, then a modeling of "physical trajectories" for 10^6 photons would be required for about 100 photons to penetrate the shield. A large relative error can be expected for the number of photons that penetrate the shield, and information on the direction and energy of the emerging photons would be less accurately known.

2.3   The Method of Statistical Weights

A procedure for circumventing the above metioned difficulties is to discard the modeling of the physical trajectory and to introduce the artifice of statistical weights. This technique increases the probability for the successful outcome in an experiment by application of some statistical weight which permits the continuance of the trajectory after each event rather than terminate the history by an unsuccessful event. Thus, if the probability for

the i-th event which favors a successful outcome is denoted by w_i ($i = 1, 2, 3, \ldots$), then weighting the event by w_i will permit continuance the history. When the outcomes of the event are mutually exclusive, the weighting factor for terminating the history is $1 - w_i$. Finally, if independent events form the elements of a trajectory, then the probability for the continuance of a trajectory after m events is given by the product of the probabilities for each event:

$$W_m^{(k)} = \prod_{i=1}^{m} w_i^{(k)} \tag{4}$$

where the superscript (k) denotes the weighting for the k-th history. When the successful outcome occurs for the experiment, the result is weighted by equation (4) for the number of events along the trajectory. In the case of radiation transport the k-th elementary particle is preserved in each interaction, and it is said to carry a weight of $W_m^{(k)}$ if the successful outcome of interest occurs after m interactions. In effect, equation (4) applies à priori information on the statistics of the events to avoid the termination of a history before the trajectory can lead to a successful outcome.

The advantage of this method rests on the fact that more information can be extracted from a given Monte Carlo experiment with fewer histories than would be possible with the use of a model based on physical trajectories. In addition, the errors associated with a given calculation can also be appreciably reduced by the method of statistical weights. To properly focus on this point, it should be recalled from equation (3) that the relative error for n histories is:

$$\left| \frac{Y_n}{p} - 1 \right| < \epsilon_r \propto \sqrt{\frac{1-p}{np}} .$$

Using the method of statistical weights, each trajectory has an associated weight, $W^{(K)}$, for a successful outcome* and the relative frequency for this model is:

$$Y_n = \frac{1}{n} \sum_{k=1}^{n} W^{(k)} \qquad (5)$$

where n is the number of histories. Unlike the use of the physical trajectories, the error in the determination of Y_n is now associated with the sum of the statistical weights in equation (4), which, in turn, depends on the number of successful outcomes tallied in the experiment. Equation (3) still applies to the estimate of the relative error, but the correct probability to use for this purpose would be the value obtained when the statistical weighting procedure is used to bias the successful outcome.

By way of illustration, reconsider the problem of the transmission of photons through a thick shield outlined briefly at the end of Section 2.0. Clearly, the possibility of a photon penetrating the shield is enhanced if each interaction of the photon is considered a scattering event. The ratio of the macroscopic scattering cross-section $\Sigma_s(E)$ to the total macroscopic cross-section, $\Sigma_t(E)$ is the probability that a given interaction will result in a scattering event. Thus, the appropriate statistical weight is $w(E_i) = \Sigma_s(E_i)/\Sigma_t(E_i)$, where E_i is the energy of the incident photon for the i-<u>th</u> interaction along the trajectory. Whereas the transmission probability for a photon might be 10^{-4}, the statistically weighted transmission probability might be 10^{-2}. Consequently, 10^4 statistically weighted histories would

---

*Since the computational time depends on the number of events along the trajectory, it is customary to terminate a history when the statistical weight for the trajectory becomes less than some prescribed value. Under these circumstances the statistical weight for the successful outcome is set equal to zero, and, conversely, the statistical weight to the unsuccessful outcome is set equal to unity. Since the statistical weight for a trajectory is merely the probability of any given trajectory ending in a successful outcome, it is possible to interpret this as the relative frequency of a large number of similar trajectories when a 0 or 1 count is given to the individual outcomes.

achieve the same accuracy as 10^6 histories with a physical trajectory
model. This procedure is a simple statistical weighting method, but it
is possible to further increase the probability for a successful outcome by
biasing the history for the location of the first interaction or the direction
of scattering from the first interaction. Both methods reflect an insight
into the physical problem which recognizes that the occurrence of a
photon interaction within the first few mean free paths of a "thick" shield
will not contribute appreciably to the number of transmitted photons.
Following such trajectories would be wasteful of computer time. A simple
weighting scheme might involve only consideration of those trajectories
which have their first event some number of mean free paths from the backface
of the shield $(\mu x_1 \leq \lambda \leq \mu \ell)$. A statistical weight of $e^{-\mu x_1}$ is then applied
to the trajectory to account for the probability of traversing the thickness
$0 \leq \lambda < \mu x_1$ without an interaction. In effect, this method of statistical
weighting artificially increases the probability of a successful outcome by
analyzing a thinner shield $(\mu x_1 \leq \lambda \leq \mu \ell)$, since there is a tacit assumption
that first interactions in the region $0 \leq \lambda < \mu x_1$ have a zero transmission
probability. In this case the precision of the result is increased with a
loss of accuracy.

Alternatively, it is possible to bias the trajectory by the scattering direction
and thereby allow first interactions within the full thickness of the shield.
Thus, a trajectory which at its first event near the front face of the shield
might be weighted for a forward scatter in a small solid angle centered on the
perpendicular to the face of the shield. The solid angle would increase as
the first interaction occurs deeper in the shield. For isotropic scattering,
the weighting on the first interaction would be

$$w(E) = \frac{\Omega(\mu x)}{4\pi} * \frac{\Sigma_s(E)}{\Sigma_t(E)} \tag{6}$$

8

and subsequent weights applied to events along the trajectory would be $w(E_i) = \Sigma_s(E_i)/\Sigma_t(E_i)$, $E_i \leq E$. If $\Omega(\mu x) = 4\pi$ for $\lambda \geq \mu x$, then this technique for statistical weighting is an improved version of confining first interactions to $\mu x_1 < \lambda < \mu \ell$, since a finite transmission probability is allowed first interactions in $0 \leq \lambda < \mu x_1$.

Either of these approaches to the solution of photon transmission through a thick shield conserves histories by increasing the likelihood of a successful outcome, albeit there might be some small systematic error introduced by the biasing scheme. The second approach is obviously more desirable, since the systematic error should be smaller. Equally clear from this illustrative example is the fact that the optimization of a Monte Carlo solution rests on the ingenuity of the programmer for the selection of a statistical weighting scheme which increases the probability for a successful outcome of interest without a sacrifice in accuracy.

3.   <u>NUMERICAL ESTIMATES OF ERRORS</u>

The error introduced in the Monte Carlo method was considered in general terms in Sections 2.0 and 3.0. In this section the errors in an actual Monte Carlo calculation are considered for the albedo of a parallel beam of 0.662 MeV photons incident on a cylinder of aluminum (R = H = 10cm). Specifically, this section will consider the interpretation of the errors associated with experimental results.

By way of a preliminary introduction, the usual statement of the error in a Monte Carlo calculation is usually assumed to be taken as $n^{-1/2}$, where n is the number of histories. This can be understood in terms

of standard deviation for n Bernoulli trials (see equation (1.14), Appendix I):

$$\sigma_n \doteq \sqrt{\frac{p(1-p)}{n}} \tag{7}$$

Equation (7) has a maximum for $p = 0.5$ so that

$$\sigma_n \leq \frac{1}{2 n} < \frac{1}{n}$$

and, therefore, the use of $n^{-1/2}$ as a measure of the error tends to bound the standard deviation from above. However, a more significant question remains; namely, what interpretation to give the value of σ_n even if it could be determined in a Monte Carlo experiment?

Table 1 summarizes the results for the albedo calculations as described above. These calculations were performed with the NUGAM2 code, which employs a method of statistical weights. In the first part of Table 1 the results are given for 15,000 histories with a printout obtained at the conclusion of every 500 histories. Because of the particular random number generating scheme, it is possible for each of the 500 histories to be coupled, i.e., the albedos in the first part of Table 1 are correlated. To avoid this possible source of uncertainty the problem was re-run using a different random number every 500 histories. These results are presented in the second portion of Table 1. Unless otherwise demonstrated, the first set of data correspond to a single Monte Carlo experiment, whereas the second set consists of 30 independent experiments performed with 500 photon histories.

Because the NUGAM2 code employs the method of statistical weights, each photon history gives a value for the albedo $0 \le A_i \le 1$. In any such finite series of measurements the best approximation to the true mean value for the distribution is:

$$\bar{A} = \frac{1}{N} \sum_{i=1}^{N} A_i \qquad . \qquad (7)$$

For 15,000 photon histories, the best approximation for the mean value is 0.245545 as given at the bottom of the first tabulation. Aside from this estimate of the mean value, very little else is known about the distribution of the albedo. The standard deviation of this experiment can only be estimated by the usual approximation; namely,

$$\sigma \approx \frac{1}{\sqrt{N}} = 0.00816 \qquad .$$

However, this quantity has very little meaning as a measure of the error unless the distribution is known. A more meaningful measure of the standard deviation is

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{n} (A_i - \bar{A})^2} \qquad ,$$

but this quantity is not presently computed in the NUGAM2 code.

Since the albedo is a statistically distributed quantity with a true mean value, the repetitive performance of the experiment will lead to mean values which are distributed about the true mean. In the second portion of Table 1, the 30 average values, each based on 500 histories, gives a mean value of 0.250904 which is about 2% higher than the estimate based on 15,000 photon histories. It is well known that a series of

11

k mean values, each based on n observations, will tend to exhibit a normal distribution about their grand average. Appendix 2 gives the mathematical derivation for the normal distribution as a limiting form of the binomial distribution. Although the DeMoivre-Laplace Limit Theorem is a more restrictive case of the Central Limit Theorem it is sufficient for the present purpose. Suffice it to state that the distribution of the mean values tends to be normal irrespective of the distribution from which the observations are made. Thus, it is possible to obtain an estimate of the true mean value and also an estimate of the error.

The data in the second portion of Table 1 are displayed in Figure 1 in the form of a cumulative frequency distribution; i.e., $F(A < A_o)$ where A_o is the abscissa (albedo). These data have the characteristic S-shape of the normal distribution. The data set has a variance

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (\bar{A}_i - \bar{\bar{A}})^2 = 0.01438 \qquad ,$$

where $\bar{A}_i$ is the mean obtained in the i-th experiment and $\bar{\bar{A}}$ is the grand average of all the experiments. Also shown on Figure 1 is the normal cumulative frequency distribution based on the grand average mean and the variance. The fit of the normal distribution to these data is quite good, and it tends to confirm the theory.

Remembering that this second set of data is equivalent to making 15,000 observations in the sample space of the albedo, the accuracy with which we can know the true mean value should be no better in the first as in the second experiment. However, the second experiment permits one to quantify the error in the mean value, i.e., $\bar{\bar{A}} \pm \sigma$, where σ is the standard

deviation for the distribution of mean values. Since the distribution of mean values will approximate a normal distribution, it is possible to state that the probability is 68% for the true mean to be within $\pm \sigma$ of the grand average mean. The best estimate of the mean value from 15,000 photon histories (data set #1, Table 1) is in agreement with this interpretation.

TABLE 1

RESULTS OF ALBEDO CALCULATIONS FOR A
PARALLEL BEAM OF 0.662 MeV INCIDENT
PHOTONS ON A 10cm X 10cm ALUMINUM CYLINDER

Case I: 15,000 Photon Histories with Printout Every 500 Histories

Numbers of Albedo Photons	Albedo	Number of Albedo Photons	Albedo
113.17	0.22634	109.23	0.21846
118.45	0.23690	126.17	0.25234
114.78	0.22956	118.29	0.23654
121.99	0.24398	133.03	0.26606
109.50	0.21900	110.83	0.22166
109.95	0.21990	113.24	0.22648
116.89	0.23378	125.50	0.25100
126.66	0.25332	126.60	0.25320
113.14	0.22628	137.84	0.27568
133.85	0.26770	123.11	0.24622
130.85	0.26170	132.19	0.26438
138.54	0.27708	131.70	0.26340
132.32	0.26464	125.42	0.25084
128.82	0.25764	127.27	0.25454
120.29	0.24058	113.56	0.22712

Average Albedo = 0.245545

Case II: 500 Histories

| Number of Albedo Photons	Albedo	Number of Albedo Photons	Albedo
131.99	0.26398	137.99	0.27598
140.33	0.28066	111.25	0.22250
118.41	0.23682	128.77	0.25754
113.17	0.22634	129.22	0.25844
114.07	0.22814	124.54	0.24908
122.24	0.24448	125.34	0.25068
128.74	0.25748	129.37	0.25874
125.49	0.25098	132.00	0.26400
118.77	0.23754	123.65	0.24730
129.03	0.25806	119.69	0.23938
121.41	0.24282	120.37	0.24074
131.81	0.26362	119.44	0.23888
139.72	0.27944	124.78	0.24956
129.08	0.25816	114.76	0.22952
125.18	0.25036	133.20	0.26640

Average Albedo = 0.250904

14
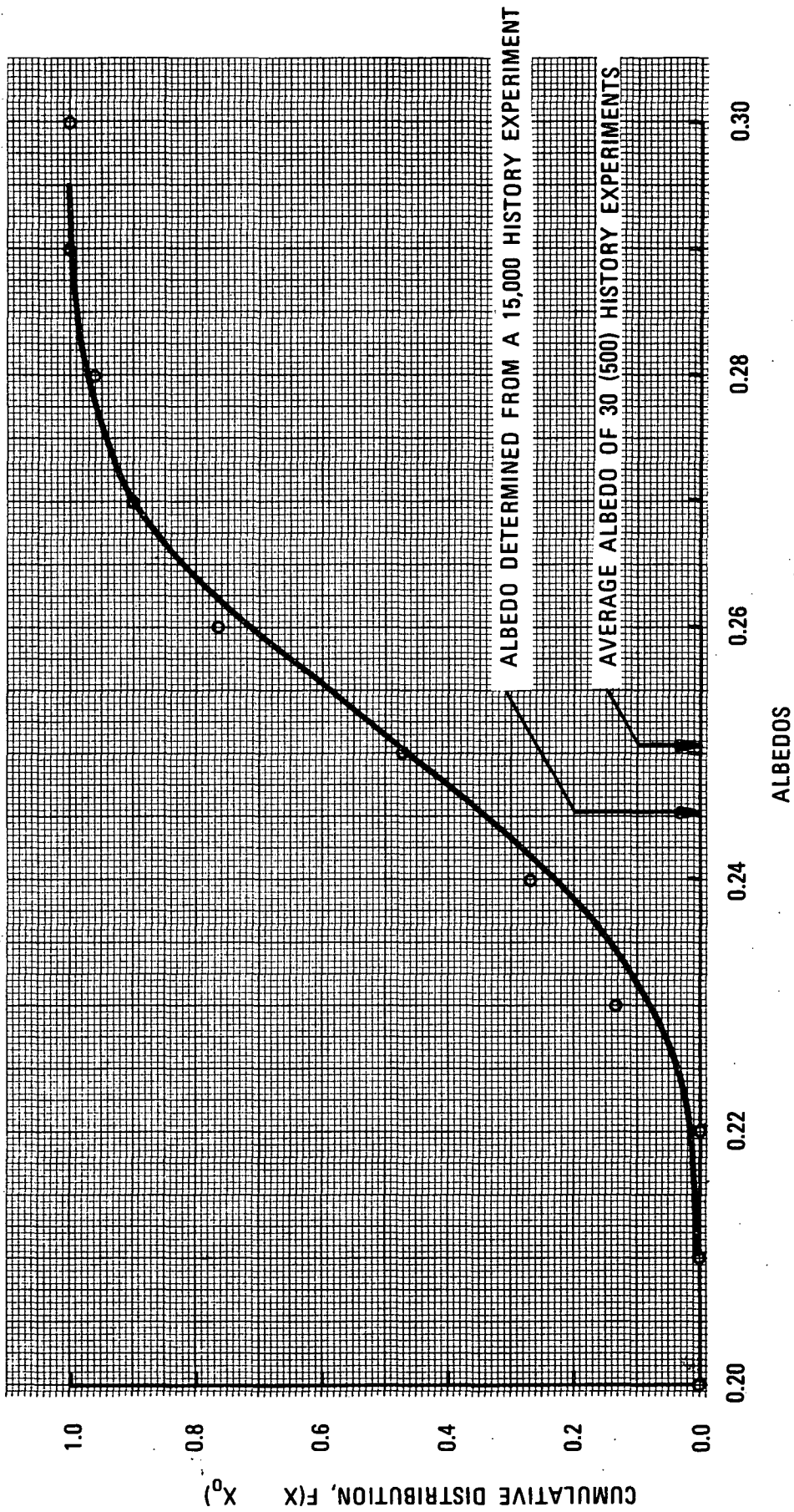
FIGURE 1

FREQUENCY DISTRIBUTION OF MONTE CARLO ALBEDOS FOR 0.662 MeV
PHOTONS INCIDENT ON A 10CM. X 10CM. CYLINDRICAL ALUMINUM
TRANSPORT MEDIUM

APPENDIX 1

THE BERNOULLI SCHEME, BINOMIAL DISTRIBUTIONS

AND BERNOULLI'S LAW OF LARGE NUMBERS

Appendix 1: The Bernoulli Scheme, Binomial Distributions and
Bernoulli's Law of Large Numbers.

Most discussions on the accuracy of the Monte Carlo method start
from a statement of the binomial distribution:

$$P^n_m = C^n_m \, p^m \, (1 - p)^{n-m} \quad ,$$

which gives the probability of "exactly" m successful outcomes in
n random experiments. Very little discussion is given to the under-
lying principles of the binomial distribution, which, more often
than not, must be weened from texts on probability theory and
mathematical statistics. For this reason it is felt that many readers
might benefit from a somewhat pedagogical, but unified, treatment
of the subject starting from the concept of the Bernoulli scheme of
sampling and carrying the development through to the accuracy of
the Monte Carlo method and its principal features.

The simplest case of a random event is one for which there are only
two possible outcomes, the event A and the complement of A. This
situation might be abstracted as a sample space consisting of the
events "zero", which we define as an unsuccessful event, and
"one", which we define as a successful event. This "zero-one"
sample space has numerous realizations, some examples of which are
the toss of a coin ("heads" or "tails"), the position of a switch ("on"
or "off"), the random selection of a binary digit ("0" or "1"), Russian
roulette ("loaded" or "unloaded" chamber), etc. For purposes of this
presentation, it will be assumed that the outcome of any given ex-
periment does not depend on the previous results obtained so that the

events are mutually exclusive (the occurrence of A precludes the occurrence
of the complement of A) and the random variable $X(X = 0, 1)$ is independent.
The concepts of mutual exclusiveness and independence implies that associated
with the successful event $X = 1$ is a probability p and with the unsuccessful
event $X = 0$ a probability $1 - p$:

$$
\left.
\begin{array}{l}
P\ (X = 1) = p \\
P\ (X = 0) = 1\text{-}p
\end{array}
\right\}
\qquad (1.2)
$$

Performance of experiments from a "zero-one" sample space, whose
outcome is represented by an independent random variable X, forms
the basis for the Bernoulli scheme.

If each outcome of an experiment is a realization of the "zero-one"
distribution, then it is possible to define an independent random
variable X as the linear superposition of the outcome of n random
experiments performed in a "zero-one" sample space:

$$
X = X_1 + X_2 + \ldots + X_n = \Sigma X_r \qquad (1.3)
$$

where $X_r = 0, 1$ and $r = 1, 2 \ldots, n$. Each of the random variables X_r
has a "zero-one" outcome so that the random variable X can take on
the values $X = k$ $(k = 0, 1, 2, \ldots, n)$. The outcome $X = k$ can occur if,
and only if, exactly k of the experiments have the outcome $X = 1$
and n-k have the outcome $X = 0$. To facilitate discussions suppose that
the successful event A contains ϵ_1 elements and the unsuccessful event
$\tilde{A}$ (complement of A) contains ϵ_2 elements, then the sample space $A \cup \tilde{A}$

contains $\epsilon = \epsilon_1 + \epsilon_2$ elements*. If each of these elements are equally likely to occur (a true die), then the à priori probability of the event A and $\tilde{A}$ would be $p = \epsilon_1/\epsilon$ and $1-p = \epsilon_2/\epsilon$, respectively. If the experiment is performed n times, then the number of ways to arrive at the event A exactly k times is ϵ_1^k and at the event $\tilde{A}$ exactly n-k times is ϵ_2^{n-k}. Consequently, the number of ways one can obtain exactly k successes and n-k failures is $\epsilon_1^k \epsilon_2^{n-k}$. Obviously, the number of possible outcomes in n random experiments is ϵ^n. As yet nothing has been said about the ordering of successful or unsuccessful outcomes. If the successful outcomes could be distinguished; that is one could somehow "tag" the $\{l's\}$ (the set l_j , $j = 1,2...,k$) in order to differentiate between $X = l_i$ and $X = l_j$, then the number of different ways of arriving at exactly k successes and n-k failures is:

$$n(n-1)...(n-k+1) = \frac{n!}{(n-k)!}$$

The number of possible ways of arranging the set l_j is given by $k(k-1)...1 = k!$, and, therefore, the number of indistinguishable ways of arranging exactly k successes and n-k failures in n random experiments is:

$$C_k^n \equiv \frac{n!}{k!\,(n-k)!} \equiv \binom{n}{k} \qquad (1.4)$$

---

*A realization of such a sample space would be the possible outcomes from the throw of a die. The six faces of a die are numbered 1 through 6. If the outcome 1 or 2 is considered a successful outcome (the event A) and 3,4,5 and 6 is considered an unsuccessful outcome (the event $\tilde{A}$), then $\epsilon_1 = 2$; $\epsilon_2 = 4$ and $\epsilon = \epsilon_1 + \epsilon_2 = 6$. The random variable $X = 1$ is assigned when the die shows a 1 or 2, and $X = 0$ is assigned when the die shows a 3,4,5 or 6.

The product of the number of ways to arrange the k successes in n experiments and the number of ways arriving at the k successes (and n-k failures) gives the total number of indistinguishable ways of obtaining exactly k successes and n-k failures in n random experiments: $C_k^n \epsilon_1^k \epsilon_2^{n-k}$. This quantity normalized to the total number of all possible outcomes for the n experiments is defined as the "probability of exactly k successes":

$$P_k^n = C_k^n \epsilon_1^k \epsilon_2^{n-k} = C_k^n \left(\frac{\epsilon_1}{\epsilon}\right)^k \left(\frac{\epsilon_2}{\epsilon}\right)^{n-k} \qquad ,$$

or

$$P_k^n = C_k^n p^k (1-p)^{n-k} \qquad , \qquad (1.5)$$

which is recognized as the binomial distribution presented in equation (1.1). Thus, the binomial distribution is seen to arise from the "zero-one" distribution by performing the experiment n times in accordance with the Bernoulli scheme. It is readily shown that:

$$\sum_{k=0}^{n} P_k^n = \sum_{k=0}^{n} C_k^n p^k (1-p)^{n-k} = 1 \qquad (1.6)$$

by expanding $(x + y)^n$, substituting $x = \epsilon_1/\epsilon$ and $y = \epsilon_2/\epsilon$, and nothing that $x + y = 1$.

The expectation value of the random variable X, denoted by E(X), is defined by:

$$E(X) \equiv \sum_{k=0}^{n} k P_k^n = \sum_{k=0}^{n} k C_k^n p^k (1-p)^{n-k} \qquad (1.7)$$

Since:

$$\sum_{k=0}^{n} k \, C_k^n \, p^k \, (1-p)^{n-k} = \sum_{k=0}^{n} \frac{k \, (n!)}{k! \, (n-k)!} \, p^k \, (1-k)^{n-k} \quad ,$$

$$= \sum_{k=0}^{n} \frac{n!}{(k-1)! \, (n-k)!} \, p^k \, (1-p)^{n-k} \quad ,$$

$$= np \sum_{\ell=0}^{m} \frac{m!}{\ell! \, (m-\ell)!} \, p^\ell \, (\ell-p)^{m-\ell} \quad ,$$

then equations (1.6) and (1.7) give:

$$E(X) = np \qquad . \quad (1.8)$$

Similarly, the expectation value of the random variable X , denoted by $E(X^2)$, is defined by:

$$E(X^2) = \sum_{k=0}^{n} k^2 \, C_k^n \, p^k \, (1-p)^{n-k} \qquad , \quad (1.9)$$

which can be shown to reduce to:

$$E(X^2) = np + n(n-1) \, p^2 \qquad , \quad (1.10)$$

following the procedure outlined above. The second central moment $D(X^2)$ is obtained from equations (1.8) and (1.10):

$$D(X^2) = E(X^2) - E^2(X) = np \, (1-p) \qquad . \quad (1.11)$$

If instead of the random variable X one defines a new random variable Y:

$$Y = \frac{X}{n} \qquad ,$$

where Y can take on the values

$$0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1 \qquad ,$$

then the probability of $Y = k/n$ is equal to the probability of exactly k successes in n random experiments:

$$P\left(Y - \frac{k}{n}\right) = P_k^n = C_k^n \, p^k \, (1-p)^{n-k} \qquad . \qquad (1.12)$$

Following the procedures outlined above, it can be shown that the expectation value of Y is:

$$E(Y) = \frac{1}{n} E(X) = p \qquad , \qquad (1.13)$$

and the second central moment (or variance) of Y is:

$$\sigma_n^2 = \frac{1}{n^2} \left[E(X^2) - E^2(X) \right] = \frac{p(1-p)}{n} \qquad (1.14)$$

Before deriving Bernoulli's Law of Large Numbers, it is necessary to develop first the so-called Chebychev inequality. To this end, define the random variable:

$$\psi = [Y - E(Y)]^2 = \frac{1}{n^2} [X - E(X)]^2 \qquad , \qquad (1.15)$$

then

$$E(\psi) = \frac{1}{n^2} E\left\{ [X - E(X)]^2 \right\} = \frac{1}{n^2} \left[E(X^2) - 2XE(X) + E^2(X) \right] \qquad ,$$

and from equation (1.14):

$$E(\psi) = \frac{1}{n^2} \left\{ E(X^2) - E^2(X) \right\} = \sigma^2 \qquad . \qquad (1.16)$$

If a random variable can take on only non-negative values, then

$$E(X) = \sum_{\substack{k=m \\ m \neq 0}}^{n} k P_k^n \geq m \, P \, (k \geq m)$$

or

$$P(k \geq m) \leq \frac{E(X)}{m} \qquad (1.17)$$

where

$$P(k \geq m) \equiv \sum_{\substack{k=m \\ m > 0}}^{n} P_k^n$$

and represents the probability of m or more successes in n random experiments. Since the random variable ψ, defined by equation (1.15), satisfies the requirements leading to equation (1.17) we can immediately write

$$P(\psi \geq m^2) \leq \frac{E(\psi)}{m^2} = \frac{\sigma^2}{m^2}$$

defining $m^2 = \alpha^2 \sigma^2$:

$$P(\psi \geq \alpha^2 m^2) \leq \frac{1}{\alpha^2} \qquad .$$

or, recognizing that $\psi \geq \alpha^2 \sigma^2$ is equivalent to $|\psi| \geq \alpha \sigma$, then

$$P(\, |\psi| \geq \alpha \sigma) \leq \frac{1}{\alpha^2} \qquad (1.18)$$

where

$$P(|\psi| \geq \alpha\sigma) = P(\sqrt{\psi} \leq \alpha\sigma) + P(\sqrt{\psi} \geq \alpha\sigma)$$

Equation (1.18) is the Chebychev inequality.

Making a slight change in nomenclature, the random variable ψ_n is defined by

$$\psi_n \equiv [Y_n - E(Y_n)]$$

where the subscript n is used to denote the results obtained from n random experiments. Consequently, the Chebychev inequality gives:

$$P(|\psi_n| \geq \alpha\sigma_n) \leq \frac{1}{\alpha^2}$$

From equation (1.14):

$$\sigma_n = \sqrt{\frac{p(1-p)}{n}}$$

and letting

$$\alpha = \sqrt{\frac{\epsilon}{p(1-p)}}$$

where $\epsilon < 0$, then

$$P(|\psi_n| \geq \epsilon) \leq \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{n\epsilon^2} \qquad (1.19)*$$

---

* Since $0 \leq p \leq 1$, then the expression p(1-p) has a maximum value at p = 1/2; that is
$$p(1-p) \leq 1/4 \leq 1$$
which is used in equation (1.19)

Equation (1.19) is the mathematical statement of the "Bernoulli Law of Large Numbers." In effect, it requires that for every $\epsilon > 0$

$$\lim_{n \to \infty} P(|\psi_n| \geq \epsilon) = 0 \qquad , \qquad (1.20)*$$

which is the condition for the sequence $\{\psi_n\}$ to be "stochastically convergent to zero." The concept of stochastic convergence to zero implies that the probability of the event $|\psi_n| \geq \epsilon$ tends to zero as $n \to \infty$.

An interpretation of the Bernoulli Law of Large Numbers is important in understanding the statistics of the Monte Carlo Method, since it relates the accuracy of the calculation to the number of histories used in the calculations. From equations (1.13) and (1.15), the Bernoulli Law of Large Numbers can be expressed in the form:

$$P(|Y_n - p| \geq \epsilon) \leq \frac{1}{n\epsilon^2} \qquad . \qquad (1.21)$$

The sequence $\{Y_m\}$ represents the normalized values of the random variable X_n possible in n random experiments:

$$\left\{ Y_m = \frac{X_n}{m} \right\} : 0, \frac{1}{m}, \frac{2}{m}, \dots, \frac{n-1}{n}, 1 \qquad ,$$

when the n experiments are conducted according to the Bernoulli scheme (ie, the possible outcomes of n Bernoulli trials). The sequence $\{Y_n - p\}$ merely centers the sequence $\{Y_n\}$ relative to the expectation value of Y_n; ie., $E(Y_n) = p$. Thus, equation (1.19) relates the frequency of observing values of $|Y_n - p| \geq \epsilon$ as the number of Bernoulli trials is increased. For example, taking $\epsilon = 0.1$ then equation (1.19) becomes

---

*Equation (1.20) is obtained in the straightforward manner from equation (1.19). Since $P(|\psi_n| \geq \epsilon) \geq 0$, then

$$\lim_{n \to \infty} P(|\psi_n| \geq \epsilon) \leq \lim_{n \to \infty} \frac{1}{n\epsilon^2} = 0$$

which can be satisfied if, and only if $P(|\psi_n| \geq \epsilon) = 0$ as $n \to \infty$ provided the limit exists

25

$$P\ (|Y_n - p|\ > 0.1)\ \leq \frac{10^2}{n}$$

so that in 10^2 experiments, $P(|Y_n-p| \geq 0.1) \leq 1.0$; in 10^3 experiments, $P(|Y_n-p| \geq 0.1) \leq 0.1$; and in 10^5 experiments, $P(|Y_n-p| \geq 0.1) \leq 0.001$. Consequently, as the number of Bernoulli trials is increased (with constant ϵ), the frequency of observing $|Y_n-p| \geq \epsilon$ decreases toward zero (stochastically converges to zero). The value of ϵ can be interpreted as the error associated with the determination of expectation value of $\{Y_n\}$ from n Bernoulli trials. In the above example, the Bernoulli Law of Large Numbers suggests a large uncertainty in the determination of p from only 100 experiments while for 10^5 experiments the frequency of observing experimental values for $\{Y_n\}$ in the range $p \pm 0.1$ is better than 0.999.

If, instead of holding ϵ constant, one allows ϵ to approach zero as n increases without bound, then the probability that the observed frequency of the (successful) event A differs little from p is close to unity. The rate of convergency is slow since ϵ cannot decrease any faster than $n^{-1/2}$ for stochastic convergency to zero; ie., one must increase the number of Bernoulli trials by more than two-orders-of-magnitude to achieve a reduction in the error by one-order-of-magnitude. This imposes a practical limitation on the Monte Carlo method, because the computational time varies directly as the number of Bernoulli trials (histories or trajectories) used in the calculation: $t \propto n$.

APPENDIX 2

DeMOIVRE – LAPLACE LIMIT THEOREM

Appendix 2: <u>DeMoivre - Laplace Limit Theorem</u>

In the previous discussion of the Bernoulli Law of Large Numbers [Appendix 1, equation (1.21)] it was implied that stochastic convergence existed for all values of the argument:

$$\underset{n \to \infty}{\text{Lim}} \ P(\left| Y_n - p \right| \geq \epsilon) \ = \ 0$$

for each $\epsilon > 0$. However, if $\psi_n \to 0$ as $n \to \infty$ is then one is presented with a difficulty. Since

$$\psi_n \ = \ Y_n - p \ = \ \frac{X_n - np}{n}$$

then one is interested in those cases where $n \to \infty$ and $X_n(=k) \to \infty$ in such a manner that

$$\frac{\mu_k}{n} \equiv \frac{k - np}{n} \longrightarrow 0. \tag{2.1}$$

For reasons to become obvious below, we will require that $n \ \infty$ and $k \ \infty$ in such a manner that

$$\frac{\mu_k^2}{n} \longrightarrow 0 \tag{2.2}$$

which also contains the condition of equation (2.1);

The binomial distribution [Appendix 1, euqation (1.5)] can be cast into the form:

$$P_k^n \ \sim \ \left\{ \frac{n}{2\pi k(n-k)} \right\}^{1/2} \ \left(\frac{np}{k} \right)^k \ \left(\frac{nq}{n-k} \right)^{n-k} \tag{2.3}$$

28

when Stirling's formula

$$m! \sim \sqrt{2\pi}\ m^{m+\frac{1}{2}}\ e^{-m}$$

is used to approximate the factorial terms of C_k^n. In equation (2.3) $q = 1-p$ and the sign $\sim$ denotes that the ratio of the two sides tends to unity. Introducing the variable μ_k into equation (2.3)

$$-\ln P_k^n \sim 1/2 \ln\left[2\pi npq\left(1 + \frac{\mu_k}{np}\right)\left(1 - \frac{\mu_k}{nq}\right)\right] + \ln\left[\left(1 + \frac{\mu_k}{np}\right)^{np+\mu_k}\left(1 - \frac{\mu_k}{nq}\right)^{nq-\mu_k}\right]$$

and using the Taylor representation

$$\ln(1 + X) = 1 + X - \frac{X^2}{2} + \frac{X^3}{3} - \frac{X^4}{4} \pm \ldots$$

for the second term of the right-hand side gives

$$-\ln P^n \sim 1/2 \ln\left[2\pi npq\left(1 + \frac{\mu_k}{np}\right)\left(1 - \frac{\mu_k}{np}\right)\right] + \frac{\mu_k^2}{2npq}\left[1 + \frac{p-q}{3pq}\left(\frac{\mu_k}{n}\right) + \ldots\right] \quad (2.4)$$

Subject to the condition of equation (2.2), in the limit of large n, equation (2.4) takes on the asymptotic form:

$$P_k^n \sim \frac{1}{\sqrt{2\pi npq}}\ \exp\left\{-1/2\left(\frac{d - np}{\sqrt{npq}}\right)^2\right\} \quad (2.5)$$

which is recognized as the "normal" distribution function. Defining the standardized random number:

$$\psi_k \equiv \frac{k - np}{\sqrt{npq}}$$

then the equation (2.5) can be cast in the form:

$$P_k^n \sim h\ \phi(\psi_k) \quad (2.6)$$

where $h \equiv (mpq)^{-1/2}$. The conditions expressed by equations (2.1) and (2.2) can be expressed in the form

$$\frac{\mu_k^3}{n^2} = h\psi_k^3 \to 0$$

as $n \to \infty$. The probability of $\alpha \leq X_n \leq \beta$ is:

$$P(\alpha \leq X_n \leq \beta) = P_\alpha^n + P_{\alpha+1}^n + \ldots + P_{\beta-1}^m + P_\beta^n \tag{2.7}$$

so that if equation (2.6) is satisfied for all k ($k = \alpha, \alpha + 1, \ldots, \beta$) in the interval, equations (2.6) and (2.7) give

$$P(\alpha \leq X_n \leq \beta) \sim h\left[\phi(X_\alpha) + \phi(X_{\alpha+1}) + \ldots + \phi(X_\beta)\right]. \tag{2.8}$$

To digress for a moment consider the area under the curve $y = f(X)$ in the interval $x_k - 1/2h \leq x \leq x_k + 1/2h$, then from the first mean value theorem:

$$\int_L f(X)\, dX = hf(\xi_k); \quad x_k - 1/2h < \xi_k < x_k + 1/2h$$

From this it is clear that the right hand member of equation (2.6) can be represented as an integral of the function:

$$d\Phi(X) = \phi(X)\, dX$$

over the interval $x_{k-1/2} \leq x_{k+1/2}$*. Further, if $h X_k^3 \to 0$ ($h = \alpha, \alpha + 1, \ldots, \beta$) for all $\alpha \leq X_n \leq \beta$, then equation (2.8) can be represented to some order of approximation, by the integral over the interval $\alpha - 1/2 \leq x \leq \beta + 1/2$. The questing remains as to the "goodness" of this approximation as a fit to $h\phi(x_k)$.

---

* From the definition of x_k:

$$x_k \pm 1/2h = [k - np] \pm 1/2h = h[(k \pm 1/2) - np] \equiv x_k \pm 1/2.$$

In the interval $x_k - 1/2h \leq x \leq x_k + 1/2h$:

$$\Phi(x_k + 1/2) - \Phi(x_k - 1/2) = \int \phi(x)dx = h\phi(\xi_k)$$

or, from equation (2.5)

$$\exp.\left\{1/2(\xi_k^2 - \psi_k^2)\right\}\left\{\Phi(x_k + 1/2) - \Phi(x_k - 1/2)\right\} = h\,\Phi(x_k) \qquad (2.9)$$

Since $x_k - 1/2h < \xi_k < x_k + 1/2h$:

$$1/2(\xi_k^2 - x_k^2) \leq 1/2|\xi_k - x_k||\xi_k + x_k| \leq 1/2h|x_k + 1/4h||x_k|+1/\ h<\epsilon$$

and

$$1/2(\xi_k^2 - x_k^2) = 1/2(\xi_k - x_k)(\xi_k + x_k) = 1/2h(x_k - 1/4h) \geq -h\left||x_k/+1/4h\right| > -\epsilon$$

then equation (2.9) can be written

$$e^{-\epsilon}\left\{\Phi(x_k + 1/2) - \Phi(x_k - 1/2)\right\} < h\,\phi(x_k) < e^{\epsilon}\left\{\Phi(x_k + 1/2) - \Phi(x_k - 1/2)\right\}$$

or, for an arbitrary $\epsilon > 0$

$$\left|\frac{h\,\phi(x_k)}{\Phi(x_k + 1/2) - \Phi(x_k - 1/2)}\right| - 1 < \epsilon$$

since $h/x_k \to 0$ and $h \to 0$ as $n \to 0$. Consequently;

$$h\,\phi(x_k) \sim \Phi(x_k + 1/2) - \Phi(x_k - 1/2) = \int_{x_{k-1/2}}^{x_{k+1/2}} \phi(x)dx$$

and summing over $k(k = \alpha, \alpha + 1, \ldots, \beta)$ gives

$$P(\alpha \le X_n \le \beta) \sim \Phi(x_{\beta + 1/2}) = \int_{x_{\alpha - 1/2}}^{x_{\beta + 1/2}} \phi(X)dX \qquad (2.10)*$$

which is the DeMoivre-Laplace limit theorem. This limit theorem states that, to some order of approximation, the probabilities of the binomial distribution can be estimated by treating the standarized random variable as "a continuous random variable with a "normal" distribution.

The DeMoivre-Laplace limit theorem is a special case of a more general limit theorem known as the "Central Limit Theorem".

The central limit theorem is obeyed by a sequence $\{X_n\}$ if for every $\alpha < \beta$:

$$P(\alpha < Z_m < \beta) \sim \Phi(\beta) - \Phi(\alpha).$$

This statement of the central limit theorem is formally similar to the DeMoivre-Laplace limit theorem, but it need not be confined to the X_k being drawn from a common binomial distribution. The Linderberg-Levy theorem insures that the central limit theorem holds for every uniformly bounded sequence $\{X_n\}$ of mutually independent random variables, i.e., $|X_n| < A$ for all n. By this we mean that for every n there are n mutually independent random variables $X, X_2, \ldots, X_n$ with prescribed distributions such that $|X_k| < A$ for all k.

---

* Since $x_e \pm 1/2 = x_k + 1/2h$ and $h \rightarrow 0$ as $n \rightarrow \infty$, then equation (2.10) can be written:
$$P(\alpha \le X_n \le \beta) \sim \Phi(X_\beta) - \Phi(X_\alpha).$$
This form of the Demoivre-Laplace limit theorem is found in the literature, but it is more restrictive than equation (2.10).

NUS CORPORATION