MARTIN MARIETTA ENERGY SYSTEMS LIBRARIES
3 4456 0250281 5

CENTRAL RESEARCH LIBRARY
DOCUMENT COLLECTION

ORNL-1508
Physics
4A

A GENERAL THEORY OF STOCHASTIC ESTIMATES OF THE NEUMANN SERIES FOR THE SOLUTIONS OF CERTAIN FREDHOLM INTEGRAL EQUATIONS AND RELATED SERIES

G. E. Albert

CENTRAL RESEARCH LIBRARY
DOCUMENT COLLECTION

LIBRARY LOAN COPY

DO NOT TRANSFER TO ANOTHER PERSON

If you wish someone else to see this document, send in name with document and the library will arrange a loan.

OAK RIDGE NATIONAL LABORATORY
OPERATED BY
CARBIDE AND CARBON CHEMICALS COMPANY
A DIVISION OF UNION CARBIDE AND CARBON CORPORATION
UCC
POST OFFICE BOX P
OAK RIDGE, TENNESSEE

ORNL-1508

This document consists of 48 pages
Copy 4 of 363 copies Series A.

Contract No W-7405-eng-26

MATHEMATICS PANEL

A GENERAL THEORY OF STOCHASTIC ESTIMATES OF THE NEUMANN SERIES FOR THE SOLUTIONS OF CERTAIN FREDHOLM INTEGRAL EQUATIONS AND RELATED SERIES

G. E. Albert

Date Issued

AUG 13 1953

OAK RIDGE NATIONAL LABORATORY
Operated by
CARBIDE AND CARBON CHEMICALS COMPANY
A Division of Union Carbide and Carbon Corporation
Post Office Box P
Oak Ridge, Tennessee

MARTIN MARIETTA ENERGY SYSTEMS LIBRARIES
3 4456 0250281 5

INTERNAL DISTRIBUTION

1. C. E. Center
2. Biology Library
3. Health Physics Library
4-5. Central Research Library
6. Reactor Experimental Engineering Library
7-25. Laboratory Records Dept.
26. Laboratory Records, ORNL R.C.
27. C. E. Larson
28. W. B. Humes (K-25)
29. L. B. Emlet (Y-12)
30. A. M. Weinberg
31. E. H. Taylor
32. E. D. Shipley
33-82. A. S. Householder
83. F. C. Vonderlage
84. C. P. Keim
85. R. C. Briant
86. J. A. Swartout
87. S. C. Lind
88. F. L. Culler
89. R. S. Livingston
90. A. H. Snell
91. A. Hollaender
92. M. T. Kelley
93. G. H. Clewett
94. J. H. Frye, Jr.
95. K. Z. Morgan
96. C. S. Harrill
97. C. E. Winters
98. D. W. Cardwell
99. E. M. King
100. Lewis Nelson
101. D. D. Cowen
102. P. M. Reyling
103. J. Moshman
104. N. M. Dismuke
105. W. S. Snyder (consultant)
106. M. J. Skinner
107-112. G. E. Albert

EXTERNAL DISTRIBUTION

113. D. B. Benedict, C&CCC, South Charleston
114. R. F. Bacher, California Institute of Technology
115. C. L. Perry, U. S. Naval Postgraduate School
116. G. H. Goertzel, New York University
117. M. Kac, Cornell University
118. M. R. Hestenes, University of California
119. G. A. Forsythe, University of California
120. D. Teichroew, University of California
121. P. Hammer, University of Wisconsin
122. R. L. Anderson, North Carolina State College
123. H. A. Myers, University of Florida
124. C. L. Seebeck, University of Alabama
125. J. Neyman, University of California
126. M. A. Girshick, Stanford University
127-363. Given distribution as shown in TID-4500 under Physics category

DISTRIBUTION PAGE TO BE REMOVED IF REPORT IS GIVEN PUBLIC DISTRIBUTION

CONTENTS

PAGE

CHAPTER 1. INTRODUCTION 1
1.1. Background 1
1.2. Notations and Assumptions 1
1.3. Related Problems 1

CHAPTER 2. STOCHASTIC PROCESSES 2
2.1. Summary and Remarks 2
2.2. The Basic Process 2
2.3. More Elaborate Processes 8

CHAPTER 3. ESTIMATION AT A FIXED POINT 18
3.1. Summary and Remarks 18
3.2. The Basic Estimating Variable 18
3.3. Zero Variance Processes 19
3.4. Representative Sampling from a Stratified Process 21

CHAPTER 4. WEIGHTED AVERAGES OF $\phi(x)$: SOLUTION IN THE LARGE 27
4.1. Estimation of a Weighted Average 27
4.2. Zero Variance Estimates of I 29
4.3. A Much Used Estimator of $\phi(x)$ 29
4.4. Estimation of $\phi(x)$ in the Large 32

APPENDIX A. AN ILLUSTRATION 35

APPENDIX B. THE ESTIMATION OF DEFINITE INTEGRALS OF FIXED MULTIPLICITY 37

REFERENCES 41

CHAPTER 1. INTRODUCTION

1.1. BACKGROUND

Considerable interest has been focused in recent years on the statistical (Monte Carlo) estimation of the solutions of integral equations of the type

$$(1.1)\quad \phi(x) = g(x) + \lambda \int_R K(x,y)\, \phi(y)\, dy\ , \qquad x \text{ in } R\ ,$$

R being a Euclidean space, and on related estimation problems. Some research on these problems has been published,[1,2,3,4] and much more of the research has been done via intra- and interlaboratory communications, either oral or written. The present report leans heavily upon such research, with little or no indication of such dependence.

Fundamentally, the aim of this report is to present (1) the application of rigorous stochastic process theory to the generation of a unified theory of estimation processes for the solution $\phi(x)$ of Eq. 1.1 at a single point $x = x_0$, (2) the application of elaborate sampling schemes based upon simple distribution functions in an attempt to realize an over-all computing efficiency by appropriate balance between statistical efficiency and machine efficiency, and (3) the estimation of the entire function $\phi(x)$ by a weighted interpolation of many single-point estimates of ϕ. Related estimation problems will be mentioned only briefly.

1.2. NOTATIONS AND ASSUMPTIONS

Define the sequence of functions

$$(1.2)\quad G_N(x,\ u_1, u_2,\ \ \ , u_N) \equiv \lambda^N\, K(x,u_1)\, K(u_1,u_2) \ldots K(u_{N-1},u_N)\, g(u_N)\ , \qquad N = 1,\ 2,\ 3,\ \ldots\ .$$

The Neumann series for Eq 1.1 has the form

$$(1.3)\quad \phi(x) = g(x) + \sum_{N=1}^{\infty} I_N(x)\ ,$$

in which

$$(1.4)\quad I_N(x) \equiv \int_R \ldots \int_R G_N(x,\ u_1, u_2, \ldots, u_N)\, du_1 \ldots du_N\ .$$

It will be assumed throughout the report that series 1.3 for Eq. 1.1 converges for x in R. In much of the report the main interest will be centered about Eqs. 1.1 for which the source term $g(x)$ is nonnegative and bounded on R and $\lambda\ K(x,y)$ is nonnegative and bounded for both variables in R. These assumptions will not be mentioned explicitly except in the sections in which the stochastic processes that depend critically upon them are first introduced.

1.3. RELATED PROBLEMS

The Basic Estimation Idea. Let A be an $r \times r$ nonsingular matrix of real numbers and let ψ and h denote column vectors of length r, where h is known and ψ is unknown. The linear system of algebraic equations $A \cdot \psi + h = 0$ may be written as

$$(1.5)\quad \psi = h + (A + I) \cdot \psi\ .$$

A formal similarity between Eqs. 1.5 and 1 1 is evident. By iteration of Eq. 1 5, an analogue to Eq. 1.3 is obtained

$$(1.6)\quad \psi = h + \sum_{N=1}^{\infty} (A + I)^N \cdot h\ .$$

If series 1.6 is convergent, stochastic methods for estimating series 1.3 may be adapted to series 1.6 by replacing the continuous distribution functions by discrete distributions

The similarity between Eqs. 1.1 and 1.5 is important in estimating the solutions of certain equations of type 1.1 Most of the estimation methods to be discussed for Eq 1.1 are inapplicable if the kernel, $K(x,y)$, is unbounded in R^2. Nevertheless, if the integral

$$\int_R K(x,y)\ dy$$

is a bounded function of x in R, Eq. 1.1 may be approximated by Eq. 1.5 and the solution estimated by a stochastic process for series 1.6. To see this, partition R into nonoverlapping sets, $\Delta_1, \Delta_2, \ldots, \Delta_r$, and choose points x_i, $i = 1, 2, \ldots, r$, in Δ_i. Let ψ denote the column vector whose elements are the values $\phi(x_i)$ of the solution to Eq. 1.1, and let h be the vector whose elements are $g(x_i)$. Define the matrix A of elements $a_{\alpha\beta}$ so that

$$a_{\alpha\beta} + \delta_{\alpha\beta} = \lambda\, K(x_\alpha, x_\beta)\, m(\Delta_\beta) ,$$

where

$\delta_{\alpha\beta}$ = the Kronecker delta,

$m(\Delta_\beta)$ = the measure $\int_{\Delta_\beta} dy$ of Δ_β .

If $\phi(x)$ and $g(x)$ do not vary much over each of the sets Δ_i, the solution ψ of system 1.5 approximates the set of values $\phi(x_i)$, $i = 1, 2, \ldots, r$, and a stochastic estimate of ψ gives a stochastic estimate of $\phi(x_i)$.

The general methods to be discussed in these pages estimate series 1.3 or 1.6. The basic idea of the methods is to define a class of stochastic processes and random variables thereon whose expectation is the series in question. Computational efficiency would then be achieved by choosing the process and the random variable that give the best balance between statistical efficiency (minimum dispersion) and speed of computing. This optimal choice may be difficult and has not been found, in general. It is hoped that sufficient variety has been included here to allow good choices in many problems.

Physical problems that give rise to Eqs. 1.1 and 1.5 are sometimes called steady-state problems.[1] The discerning reader will easily see that, in general, the estimation processes to be discussed in chapter 3 of this report are not limited in their applicability to only Eqs. 1.1 and 1.5. The same processes could be used to estimate series like 1.3 or 1.6 in which the integrands, G_N, or matrices, $(A + I)^N$, are of more general form. For example, the functions G_N need not have the explicit form given in Eq. 1.2. The factors K might vary in functional form with the indices on their arguments. Similarly, the matrices $(A + I)^N$ in the series 1.6 might be replaced by the more general forms $\prod_{i=1}^{N} B_i$

CHAPTER 2. STOCHASTIC PROCESSES

2.1. SUMMARY AND REMARKS

This chapter presents the formal description of certain stochastic processes that will be applied in Chap. 3 to the statistical estimation of solutions of equations of types 1.1 and 1.5. The uninitiated reader might study sections 2.2, 3.2, and 3.3 before attempting to follow the elaborate descriptions given in sections 2 3 through 2.3.3.

Any statistical theory should begin with a description of a sampling space and a probability measure defined thereon. Doob[5] has given a rigorous theory for the type of stochastic processes needed here. The definitions to be given below conform with his definitions and theorems. Section 2.2 contains a description of a process that will be used in Sections 3.2 and 3.3 to define the basic estimation procedure to be used in this report. Most of the estimation methods described in refs. 1, 2, 3, and 4 are special cases of that basic procedure. Sections 2.3 through 2.3.3 define more complicated stochastic processes for use in the elaborate sampling schemes involving stratified and multiple sampling described in section 3.4.

2.2 THE BASIC PROCESS

For convenience and simplicity, the space R of the integral in Eq. 1.1 will be assumed henceforth to be an interval $a \leq x \leq b$ of the real number

axis. Generalization to multidimensional intervals R offers no difficulty in theory. For unity in notation, in dealing with the system of Eqs. 1 5, the set of integers (1, 2, 3, ..., r) will be denoted also by R.

The following abbreviated notations will simplify the formulas and will be used freely without comment

$$
\begin{aligned}
\gamma_N &= (x_1, x_2, x_3, \ldots, x_N) \\
&\qquad = \text{point in } R^N, \\
\gamma_N^* &= (x_0, x_1, x_2, \ldots, x_N) = \\
&\qquad (x_0, \gamma_N) = \text{point in } R^{N+1}, \\
d\gamma_N &= dx_1\, dx_2 \ldots dx_N, \\
d\gamma_N^* &= dx_0\, dx_1 \ldots dx_N \\
&\qquad = dx_0\, d\gamma_N, \\
f(\gamma_N) &= f(x_1, x_2, \ldots, x_N), \\
f(\gamma_N^*) &= f(x_0, x_1, \ldots, x_N).
\end{aligned}
\tag{2.1}
$$

Definition 1. *Define a sequence $F_N(\gamma_N^*)$, $N = 0, 1, 2, \ldots$, of cumulative distribution functions on the spaces R^{N+1} with the consistency properties $F_N(x_0, x_1, \ldots x_p, \infty, \ldots, \infty) \equiv F_p(\gamma_p^*)$ for each $N \geq 1$ and $p < N$.* Also, define a sequence of functions $p_N(\gamma_N^*)$, $N = 0, 1, 2, \ldots$, on R^{N+1} such that for each N and each γ_N^*, $0 \leq p_n(\gamma_N^*) \leq 1$. Let $q_N(\gamma_N^*) \equiv 1 - p_N(\gamma_N^*)$.*

Assumption 1. *(a) If the solution $\phi(x)$ of Eq 1 1 is to be estimated, it will be assumed that a piecewise continuous-density function $f_N(\gamma_N^*)$ exists such that*

$$
F_N(\gamma_N^*) = \int_a^{x_0} \cdot \int_a^{x_N} f_N(\gamma_N^{*\prime})\, d\gamma_N^{*\prime}, \qquad (= 1 \text{ if } \gamma_N^* = b, b, \ldots, b)
$$

(b) If the vector ψ of Eqs. 1.5 is to be estimated, it will be assumed that $F_N(\gamma_N^*)$ is a discontinuous distribution with all the probability concentrated at points γ_N^*, for which each component x_i, $0 \leq i \leq N$, is one of the integers 1, 2, ..., r. For uniformity in notation, $f_N(\gamma_N^*)$ will denote the probability concentrated at γ_N^*

It follows from Assumption 1 that $F_N(\gamma_N^*)$ vanishes if one or more of the components x_i of γ_N^* is less than a and is unity if all the x_i of γ_N^* exceed b.

The following notations will be useful

$$
P_N(\gamma_N^*) = p_N(\gamma_N^*) \prod_{i=0}^{N-1} q_i(\gamma_i^*), \tag{2 2}
$$

$$
\begin{aligned}
f_N(\gamma_N | \gamma_i^*) &\equiv f_N(x_{i+1}, x_{i+2}, \ldots, x_N | \gamma_i^*) \\
&\equiv \frac{f_N(\gamma_N^*)}{f_i(\gamma_i^*)}, \quad i < N,
\end{aligned}
\tag{2.3}
$$

$$
H_{N+1}(\gamma_{N+1} | \gamma_N^*) \equiv f_{N+1}(\gamma_{N+1} | \gamma_N^*) \prod_{i=0}^{N} q_i(\gamma_i^*). \tag{2.4}
$$

The function 2.3 denotes the conditional probability (density) of x_{i+1}, x_{i+2}, ..., x_N, if γ_i^* is given.

The sampling space Ω to be envisaged may be thought of as a space of infinitely many dimensions, provided "dimension" is not interpreted strictly. The definition of the sampling space follows Assign two distinct states, S and T, for which intuitive meanings will be given later. Let X denote the collection of all elements a that are pairs of the forms (x, S) and (x, T), where x is in R. For any fixed x in R, the two pairs (x, S) and (x, T) are to be regarded as distinct elements a of X, as are (x_1, S) and (x_2, S) or (x_1, T) and (x_2, T) for distinct x_1, x_2 in R. Subsets of X will be denoted by A and B, with, possibly, subscripts or superscripts or both. The symbols $A^{(S)}$ and $B^{(S)}$ will denote subsets of X that are composed entirely of elements of the form (x, S). Similarly, $A^{(T)}$ and $B^{(T)}$ contain only (x, T) for various x. The symbols $[A^{(S)}]^{-1}$, $[A^{(T)}]^{-1}$, etc. will denote the complete subsets of R for which (x, S) is in $A^{(S)}$, $A^{(T)}$, etc., respectively. Let ω denote an

*See reference 6, p 19, for further properties.

arbitrary, denumerably-infinite sequence $\{\alpha_n\}_0^\infty$ of elements $\alpha_0, \alpha_1, \alpha_2, \ldots \alpha_n, \ldots$ from X. The sampling space Ω will be composed of the collection of all such sequences ω.

The class F_x of Borel subsets of R induces a Borel field F_α of subsets of X in the following way. Any A in X may be decomposed into a sum $A = A^{(S)} + A^{(T)}$ of disjoint sets of the characters described above. The set A will belong to F_α if the complete inverse sets $[A^{(S)}]^{-1}$ and $[A^{(T)}]^{-1}$ both belong to F_x.

Let $n_1 < n_2 < \quad < n_k$ be k integers. The symbol $\Lambda_{n_1 n_2 \ldots n_k}$ will be generic for a Borel cylinder subset of Ω over $\alpha_{n_1}, \alpha_{n_2}, \ldots, \alpha_{n_k}$. Such sets are composed in the following way. Let A_i, $i = 1, 2, \ldots, k$, be k elements of F_α (not necessarily distinct). The set of all ω in Ω for which α_{n_i} is in A_i, $i = 1, 2, \ldots, k$, is a Borel cylinder set over $\alpha_{n_1}, \alpha_{n_2}, \ldots, \alpha_{n_k}$, briefly, a B-cylinder set over $\alpha_{n_1}, \alpha_{n_2}, \ldots, \alpha_{n_k}$ Doob* has shown how to define a probability measure on a Borel field of subsets of Ω. In particular, he has given a formula for the calculation of the value of that measure for any B-cylinder set over a finite number of α's. His results will be specialized suitably in the following for later applications.

Let Λ_0 denote any B-cylinder set over α_0 as specified by α_0 is in $A_0^{(S)} + A_0^{(T)}$. Define

$$(2.5)\quad P_0(\Lambda_0) \equiv \int_{[A_0^{(S)}]^{-1}} f_0(x_0)\, p_0(x_0)\, dx_0 + \int_{[A_0^{(T)}]^{-1}} f_0(x_0)\, q_0(x_0)\, dx_0 \;.^{**}$$

Let Λ_{n+1} be any B-cylinder set over α_{n+1}, $n = 0, 1, 2, 3, \ldots$. Clearly, Λ_{n+1} is defined by α_{n+1} in $A_{n+1}^{(S)} + A_{n+1}^{(T)}$. Define

$$(2.6)\quad Q_{n+1}(\Lambda_{n+1}|\gamma_n^*) \equiv \left[\int_{[A_{n+1}^{(S)}]^{-1}} H_{n+1}(\gamma_{n+1}|\gamma_n^*)\, p_{n+1}(\gamma_{n+1}^*)\, dx_{n+1} + \int_{[A_{n+1}^{(T)}]^{-1}} H_{n+1}(\gamma_{n+1}|\gamma_n^*)\, q_{n+1}(\gamma_{n+1}^*)\, dx_{n+1}\right] \left[\int_R H_{n+1}(\gamma_{n+1}|\gamma_n^*)\, dx_{n+1}\right]^{-1} .$$

In terms of Eq. 2.6, define

$$(2.7)\quad P_{n+1}\left(\Lambda_{n+1}|\alpha_0, \alpha_1, \ldots, \alpha_n\right) \equiv \begin{cases} Q_{n+1}(\Lambda_{n+1}|\gamma_n^*), & \text{if } \alpha_n = (x_n, T), \\ 1, & \text{if } \alpha_n = (x_n, S) \text{ and } x_n \text{ is in } \left[A_{n+1}^{(S)}\right]^{-1}, \\ 0, & \text{if } \alpha_n = (x_n, S) \text{ and } x_n \text{ is not in } \left[A_{n+1}^{(S)}\right]^{-1}. \end{cases}$$

Theorem 1 *Let $\mathcal{B}$ denote a Borel field of subsets of Ω that includes all the B-cylinder sets defined above A probability measure $P(\Lambda)$ exists on the elements Λ of $\mathcal{B}$. The functions 2 7 are the conditional probabilities $P(\Lambda|\alpha_0, \alpha_1, \quad, \alpha_n)$, if $\Lambda = \Lambda_{n+1}$ is a cylinder set over α_{n+1} and $\alpha_0, \alpha_1, \ldots, \alpha_n$ are given If $\Lambda = \Lambda_{n_1 n_2 \ldots n_k}$ is any B-cylinder set over $\alpha_{n_1}, \alpha_{n_2}, \ldots, \alpha_{n_k}$ and $\delta(\Lambda)$ denotes its characteristic function, the probability $P(\Lambda)$ is given by the formula*

*See reference 5, p. 102-3.

**Integrals are to be replaced by sums for case b of Assumption 1

$$(2.8)\qquad P(\Lambda) = \int P_0(de_0) \int P(de_1|\alpha_0) \ldots \int \delta(\Lambda)\, P\left(de_{n_k}|\alpha_0,\alpha_1,\ldots,\alpha_{n_k-1}\right) .$$

Theorem 1 is a special case of one given by Doob in the last citation. The interpretation of the symbolic iterated integral of Eq. 2.8 should be made clear by the calculations that follow.

The stochastic process given by Theorem 1 will be denoted by $[\Omega,P(\Lambda)]$. The process is a generalization of the Markoff chain process, with absorption, that is familiar in particle physics. Indeed, $[\Omega,P(\Delta)]$ reduces to such a process if the conditional probabilities 2.7 reduce to the form $P(\Lambda_{n+1}|\alpha_n)$ for each $n \geq 0$. Note that the index $n + 1$ on P has been dropped. Intuitively, the general process may be thought of in terms of the following random walk. A particle moving through a space R in a random manner suffers collisions at successive points x_0, x_1, x_2, . in R. It may be absorbed at any one of these points x_n, with probability $p_n(\gamma_n^*)$, or go on to x_{n+1}, with probability $q_n(\gamma_n^*)$. The distribution (density) of x_0 is $f_0(x_0)$ and, for any $n \geq 0$, if the particle is not absorbed at x_n, the distribution of x_{n+1} is $f_{n+1}(\gamma_{n+1}|\gamma_n^*)$. The choice $\alpha = (x,S)$ indicates that the particle is absorbed at x, and $\alpha = (x,T)$ indicates that the particle is not absorbed at x. Note that the notations allow the absorption and transition probabilities at any x_n to depend upon *all* previous collision positions, as does the distribution of the position of any collision. Moreover, the probability functions may change from one collision to the next.

The special values 1 and 0 stipulated in probabilities 2.7 for the case $\alpha_n = (x_n,S)$, guarantee that absorption (state S) at x_n will be followed by $\alpha_{n+1} = (x_n,S)$, with probability 1. The definition may seem artificial in that no attention is paid to the forms of α_0, α_1, ..., α_{n-1}. The (1,0) cases for smaller index take care of this omission in such a way that Eq. 2.8 assigns zero probability to any set Λ of sequences ω in which a state S is followed by a state T or even by a state S at a different point of R.

The following definition continues the description of a stochastic process for estimating the series 1 3 or 1.6.

Definition 2. *Let $n(\omega)$ be defined, for each ω in the space Ω, as equal to the smallest integer k such that the element α_k of ω has the form (x,S). Let Σ_k denote the set of all ω in Ω on which $n(\omega) = k$.*

The set Σ_k has the following form

$$(2.9)\qquad \begin{cases} \Sigma_k = \prod_{\iota=0}^{k} \Lambda_\iota , \\ \Lambda_\iota = \{\text{set of all } \omega \text{ such that } \alpha_\iota = (x,T),\ x \text{ in } R\}, \text{ if } \iota = 0, 1, \ldots, k-1, \\ \Lambda_k = \{\text{set of all } \omega \text{ such that } \alpha_k = (x,S),\ x \text{ in } R\} . \end{cases}$$

The definitions 2.5 and 2.7 apply to the Λ_ι in set 2 9 so that

$$(2\ 10)\qquad \begin{cases} P_0(\Lambda_0) = \int_a^b f_0(x_0)\, q_0(x_0)\, dx_0 , \\ P_\iota(\Lambda_\iota|\alpha_0,\alpha_1,\ldots,\alpha_{\iota-1}) = \dfrac{\int_a^b H_\iota(\gamma_\iota|\gamma_{\iota-1}^*)\, q_\iota(\gamma_\iota^*)\, dx_\iota}{\int_a^b H_\iota(\gamma_\iota|\gamma_{\iota-1}^*)\, dx_\iota}, \qquad 1 \leq \iota \leq k-1 , \\ P_k(\Lambda_k|\alpha_0,\alpha_1,\ldots,\alpha_{k-1}) = \dfrac{\int_a^b H_k(\gamma_k|\gamma_{k-1}^*)\, p_k(\gamma_k^*)\, dx_k}{\int_a^b H_k(\gamma_k|\gamma_{k-1}^*)\, dx_k} \end{cases}$$

The integration elements needed in Eq. 2 8 to compute $P(\Sigma_k)$ are obtained from Eqs. 2.10 by replacing the integration ranges (a,b) of the numerators by the ranges $(x_i, x_i + dx_i)$, $i = 0, 1, 2, \ldots, k$, in the respective cases. Then $P(\Sigma_k)$ is given by the iterated integral

$$P[n(\omega) = k] \equiv P(\Sigma_k)$$

$$= \int_a^b f_0(x_0)\, q_0(x_0)\, dx_0 \left[\prod_{i=1}^{k-1} \frac{\int_a^b H_i(\gamma_i|\gamma_{i-1}^*)\, q_i(\gamma_i^*)\, dx_i}{\int_a^b H_i(\gamma_i|\gamma_{i-1}^*)\, dx_i} \right] \times$$

$$\frac{\int_a^b H_k(\gamma_k|\gamma_{k-1}^*)\, p_k(\gamma_k^*)\, dx_k}{\int_a^b H_k(\gamma_k|\gamma_{k-1}^*)\, dx_k} \tag{2.11}$$

$$= \int_a^b \cdots \int_a^b f_k(\gamma_k^*)\, P_k(\gamma_k^*)\, d\gamma_k^* , \qquad k = 0, 1, 2, \ldots .$$

The symbol $P_k(\gamma_k^*)$ in the integrand here refers to Eq. 2.2 (see also Eq. 2.1 for notations).

It will be important that the random variable $n(\omega)$ be finite with probability 1. That is, it is required that

$$\lim_{N\to\infty} \sum_{k=0}^{N} P(\Sigma_k) = 1 . \tag{2.12}$$

Since

$$\Omega = \sum_{k=0}^{\infty} \Sigma_k$$

and $P(\Omega) = 1$, to establish the limit 2.12 it is sufficient to show that

$$\lim_{N\to\infty} \sum_{k=N}^{\infty} P(\Sigma_k) = 0 , \tag{2.13}$$

which will be established as a consequence of the following assumption.

Assumption 2. *There is a constant c, $0 \le c < 1$, and an integer M_c such that for an arbitrary integer j and all integers $m \ge M_c + j$, the inequality*

$$\int_a^b \cdots \int_a^b f_m(\gamma_m|\gamma_j^*)\, q_{j+1}(\gamma_{j+1}^*) \cdots q_m(\gamma_m^*)\, dx_{j+1} \cdots dx_m \le c \tag{2.14}$$

holds for all γ_j^* in R^{j+1} uniformly in j.

Theorem 2. *Under Assumption 2, the limit 2 13 holds. Moreover, there is a positive t_0 such that the moment generating function*

$$\sum_{k=0}^{\infty} P(\Sigma_k)\, e^{kt}$$

of the distribution of the duration $n(\omega)$ exists for all $t < t_0$. All the moments of this distribution exist

Proof. Consider the sum (for arbitrary j)

$$V_{m,j}(\gamma_j^*) = p_j(\gamma_j^*) + \sum_{N=j+1}^{m} \int_a^b \cdots \int_a^b f_N(\gamma_N|\gamma_j^*)\, q_j(\gamma_j^*) \cdots q_{N-1}(\gamma_{N-1}^*)\, p_N(\gamma_N^*)\, dx_{j+1} \cdots dx_N .$$

Since $p_N(\gamma_N^*) = 1 - q_N(\gamma_N^*)$, this sum telescopes into

$$V_{m,j}(\gamma_j^*) = 1 - \int_a^b \cdots \int_a^b f_m(\gamma_m|\gamma_j^*)\, q_j(\gamma_j^*) \cdots q_m(\gamma_m^*)\, dx_{j+1} \cdots dx_m . \tag{2.15}$$

Assumption 2 guarantees that, for any integer k, if $m \geq kM_c + j$, the integral on the left in Eq. 2 14 or on the right in Eq. 2.15 will be dominated by c^k. Thus

$$1 - c^k \leq V_{m,j}(\gamma_j^*) \leq 1, \quad \text{if } m \geq kM_c + j,$$

is satisfied uniformly in γ_j^*. This establishes

$$\lim_{m\to\infty} V_{m,j}(\gamma_j^*) = 1$$

uniformly in γ_j^* over R^{j+1} for arbitrary j Then

$$(2.16) \qquad \sum_{k=N}^{\infty} P(\Sigma_k) = \int_a^b \int_a^b f_N(\gamma_N^*)\, q_0(\gamma_0^*) \ldots q_{N-1}(\gamma_{N-1}^*) \times \left[\lim_{m\to\infty} V_{m,N}(\gamma_N^*)\right] d\gamma_N^* \leq c^r,$$

if $N \geq rM_c$, as before. This establishes limit 2.13.

Consider the sum

$$\sum_{k=M_c}^{\infty} P(\Sigma_k)\, e^{kt} = \sum_{j=1}^{\infty} \sum_{k=jM_c}^{(j+1)M_c-1} P(\Sigma_k)\, e^{kt}$$

From Eq. 2.16, it is seen that $jM_c \leq k \leq (j+1)M_c - 1$ implies that $P(\Sigma_k) \leq c^j$ so that

$$\sum_{k=M_c}^{\infty} P(\Sigma_k)\, e^{kt} \leq \sum_{j=1}^{\infty} c^j\, e^{jM_c t} \sum_{k=0}^{M_c-1} e^{kt} \leq \left(\sum_{k=0}^{M_c-1} e^{kt}\right) \sum_{j=1}^{\infty} \left(c\, e^{M_c t}\right)^j$$

The infinite sum on the right is convergent if $ce^{M_c t} < 1$, that is, if

$$t < t_0 = \frac{1}{M_c} \ln \frac{1}{c}.$$

This completes the proof of Theorem 2.

The sequence of numbers $P(\Sigma_k)$, $k = 0, 1, 2, 3, \ldots$, forms a probability distribution on the nonnegative integers, it is the distribution of the duration $n(\omega)$ of the random walk. For any fixed nonnegative integer k, the function

$$(2\ 17) \quad g_k(\gamma_k^* | n(\omega) = k) \equiv \frac{f_k(\gamma_k^*)\, P_k(\gamma_k^*)}{P(\Sigma_k)}$$

represents the conditional probability (density) of the finite chain $\gamma_k^* = (x_0, x_1, \ldots, x_k)$ over R^{k+1} under the condition $n(\omega) = k$ Since a process defined by Theorem 1 ignores sequences ω for which $\alpha_m \neq \alpha_{n(\omega)}$ when $m > n(\omega)$, function 2 17 may also be regarded as the conditional distribution of ω under the condition $n(\omega) = k$

It is possible now to envisage a different intuitive concept of experimental sampling for chains ω from the process $[\Omega, P(\Lambda)]$ that is perhaps simpler than the random walk concept. First, draw a value $n(\omega) = k'$ at random from the population whose distribution is given by Eq. 2.11, and then draw a finite chain

$$\gamma_{k'}^* = \left(x_0', x_1', \ldots, x_{k'}'\right)$$

from the population with distribution function 2.17 for $k = k'$. The sequence ω is then defined by $\alpha_l = (x_l', T)$ for $0 \leq l < k'$ and $\alpha_l = (x_{k'}', S)$ for $l \geq k'$. Considerable attention will be given in the sequel to special processes $[\Omega, P(\Lambda)]$ for which this sampling concept is practicable.

Remark In most of the applications of processes $[\Omega, P(\Lambda)]$ to the stochastic estimation of the solutions of Eqs. 1.1 or 1.5, the first chain point α_0 will be preassigned as (x_0, T), where x_0 is a fixed point in R Thus the entire process must be conditioned for the choice $\alpha_0 = (x_0, T)$. To effect this conditioning, replace $f_k(\gamma_k^*)$ in Eqs. 2.11 and 2.17 by $f_k(\gamma_k | x_0)$, choose $p_0(x_0) \equiv 0$, and drop integration with respect to x_0. The notations $P[n(\omega) = k | x_0]$ for the modified form of Eq.

2.11 and $g_k(\gamma_k | n(\omega) = k, x_0)$ for the modified form of Eq. 2.17 will place this conditioning in evidence.

2.2 1. Important Special Cases. (*a*) Most of the existing literature on stochastic processes of the type studied in Theorems 1 and 2 has been further specialized to Markoff chains, that is, the transition probabilities 2.4 actually depend only upon x_N, not upon all of γ_N^*. Some attention will be given to this type of process in sections 3.3 and 4.2. (*b*) The following specialization is in some ways simpler than a Markoff chain and should be very useful for high-speed sampling on an automatic-sequence computer.

Suppose that for each $k = 0, 1, 2, 3, \ldots$, the functions $p_k(\gamma_k^*)$ defined in Definition 1 depend at most upon k and x_0. In this case $P_N(\gamma_N^*)$ of Eq. 2.2 reduces to a form $P_N(x_0)$ and Eq. 2.11, conditioned by x_0, reduces to

$$P[n(\omega) = k | x_0] = P_k(x_0) , \quad k = 0, 1, 2, \ldots . \tag{2.18}$$

In this case the nonsequential (nonrandom walk) type of sampling described at the end of section 2.2 becomes practicable. Indeed, such sampling would be practicable in any process for which the integrals of Eq. 2.11 are amenable to explicit computation.

A further simplification of the process $[\Omega, P(\Lambda)]$ occurs if the probability (density) functions $f_k(\gamma_k | x_0)$ are chosen so that they are independent of x_0 and the $x_1, x_2, \ldots, x_k$ of γ_k are stochastically independent, that is, let $f(x)$ be a probability (density) function over R and let

$$f_k(\gamma_k | x_0) \equiv \prod_{i=1}^{k} f(x_i)$$

for

$$k = 1, 2, 3, \ldots .$$

If $f(x)$ is such that machine sampling is easy for values x_i randomly drawn from the population whose probability distribution is specified by $f(x)$, high-speed sampling for the chains $a_1, a_2, \ldots, a_{n(\omega)}$ should be achievable.

2.3. MORE ELABORATE PROCESSES

The three parts of this section will be given to the formal description of fairly elaborate processes derived from the basic process defined by Theorem 1. They will be used later in various rather complicated sampling schemes. The processes will be described in considerable generality, but the author wishes to emphasize that the general descriptions are likely to be of theoretical interest only. Practicable special cases are given by using a simple basic process of the type described in example *b* of section 2.2.1. In such cases the elaborate formal descriptions given below simplify to such an extent that the resulting processes should be very easy to construct and use for experimental sampling on high-speed computing machinery.

2.3.1. Stratification of a Process $[\Omega, P(\Lambda)]$. Let the collection Ω be partitioned into a finite number S of disjoint Borel sets $\Delta_1, \Delta_2, \ldots, \Delta_S$ and let $[\Omega, P(\Lambda)]$ and $[\Omega, P^{(j)}(\Lambda)]$, $j = 1, 2, \ldots, S$, be $S + 1$ stochastic processes, say of the kind specified by Theorem 1 and of the same type, either all continuous or all discrete.* Assume that $P(\Delta_j) > 0$ for each $j = 1, 2, \ldots, S$.

Definition 3. *The process* $[\Omega, P(\Lambda)]$ *will be said to be stratified into the* S *strata* $[\Omega, P^{(j)}(\Lambda)]$, $j = 1, 2, \ldots, S$, *if for each* j *and for each Borel set* Λ

$$P^{(j)}(\Delta_j) = 1 ,$$

$$P(\Lambda\Delta_j) = P(\Delta_j)\, P^{(j)}(\Lambda) , \tag{2.19}$$

that is, the probability measure $P^{(j)}(\Lambda)$ *is the conditional probability of* Λ *in* $[\Omega, P(\Lambda)]$, *subject to the condition* ω *in* Δ_j.

Suppose $[\Omega, P(\Lambda)]$ conditioned for fixed x_0 and $p_0(x_0) \equiv 0$. The sets $\Sigma_{kj} = \Sigma_k \Delta_j$, $k = 0, 1, 2, \ldots, j = 1,$

*More general processes than those specified by Theorem 1 could be stratified

2, ..., S, are those subsets of Δ_j on which $n(\omega) = k$, and

$$\Delta_j = \sum_{k=0}^{\infty} \Sigma_{kj} .$$

Define

$$(2.20) \quad \nu_j = P(\Delta_j) = \sum_{k=1}^{\infty} P(\Sigma_{kj}),^{*} \quad j = 1, 2, \ldots, S .$$

From Eqs. 2.19, the probability measure $P^{(j)}(\Lambda)$ of the jth stratum is given in terms of the original measure $P(\Lambda)$ by

$$(2.21) \quad P^{(j)}(\Lambda) = \nu_j^{-1} \sum_{k=1}^{\infty} P(\Lambda\Sigma_{kj}) .$$

In particular, if $\Lambda = \Sigma_{lj}$, Eq. 2.21 gives the probability that $n(\omega) = l$ under the conditions x_0 and ω in Δ_j. For each $j = 1, 2, \ldots, S$,

$$(2.22) \quad P^{(j)}[n(\omega) = l \mid x_0] = \begin{cases} 0, \text{ if } l = 0 \\ \nu_j^{-1} P[\Sigma_l \Delta_j], \text{ if } l = 1, 2, 3, \ldots . \end{cases}$$

*The assumption $p_0(x_0) = 0$ has the effect $P(\Sigma_{0j}) = 0$ for each j

The analogue of distribution 2.17 is obtained in the following manner. For any ω in which the x's of α_0, α_1, ..., α_l are specified by $\gamma_l^* = (x_0, x_1, \ldots, x_l)$, let

$$\Lambda_i = \left\{ \text{set of all } \omega \text{ for which } \alpha_i = (x_i', T),\ x_i' \text{ in } (x_i,\ x_i + dx_i) \right\} \quad \text{for } i = 1, 2, \ldots, l-1 ,$$

$$\Lambda_l = \left\{ \text{set of all } \omega \text{ for which } \alpha_l = (x_l', S),\ x_l' \text{ in } (x_l,\ x_l + dx_l) \right\} .$$

From the law of the mean for integrals and the formulas above, it is found that the probability element

$$g^{(j)}(\gamma_l \mid n(\omega) = l, x_0)\ d\gamma_l = \frac{P^{(j)}\left[\Lambda_l \prod_{i=1}^{l-1} \Lambda_i\right]}{P^{(j)}[n(\omega) = l \mid x_0]}$$

$$= \frac{\nu_j^{-1} P\left[\Delta_j \Lambda_l \prod_{i=1}^{l-1} \Lambda_i\right]}{\nu_j^{-1} P[\Delta_j \cdot \Sigma_l]} = \frac{\sum_{k=1}^{\infty} P\left[\Sigma_{kj} \Lambda_l \prod_{i=1}^{l-1} \Lambda_i\right]}{P(\Sigma_{lj})}$$

$$= \frac{P\left[\Sigma_{lj} \Lambda_l \prod_{i=1}^{l-1} \Lambda_i\right]}{P(\Sigma_{lj})} = \begin{cases} 0, \text{ if } \omega \text{ not in } \Delta_j , \\ \dfrac{f_l(\gamma_l \mid x_0)\ P_l(\gamma_l^*)\ d\gamma_l}{P(\Sigma_{lj})}, \text{ if } \omega \text{ in } \Delta_j . \end{cases}$$

From this,

$$(2.23)\quad g_l^{(j)}(\gamma_l \mid n(\omega) = l, x_0) = \begin{cases} 0, \text{ if } \omega \text{ not in } \Delta_j , \\ g_l(\gamma_l \mid n(\omega) = l, x_0) \dfrac{P[n(\omega) = l \mid x_0]}{\nu_j P^{(j)}[n(\omega) = l \mid x_0]} , \\ \qquad\qquad \text{if } \omega \text{ in } \Delta_j . \end{cases}$$

Each of the strata $[\Omega, P^{(j)}(\Lambda)]$ is specifiable in terms of Definition 1 and the developments leading up to Theorem 1. This will not be shown in complete detail for the general case. It is important for sampling purposes to know the functions

$$f^{(j)}(\gamma_l \mid x_0)\ P_l^{(j)}(\gamma_l^*) .$$

From Eq. 2.11, it is seen that

$$(2.11a)\qquad P^{(j)}[n(\omega) = l \mid x_0] = \int_a^b \cdots \int_a^b f_l^{(j)}(\gamma_l \mid x_0)\ P_l^{(j)}(\gamma_l^*)\ d\gamma_l .$$

On comparing this with Eq. 2.22, it is evident that the definitions

$$(2.24)\qquad f_l^{(j)}(\gamma_l \mid x_0)\ P_l^{(j)}(\gamma_l^*) = \begin{cases} 0, \text{ if } \omega \text{ not in } \Delta_j , \\ \dfrac{f_l(\gamma_l \mid x_0)\ P_l(\gamma_l^*)}{\nu_j}, \text{ if } \omega \text{ in } \Delta_j , \end{cases}$$

will serve for each $l \geq 1$ and each $j = 1, 2, \ldots, S$. It is now evident that Eq. 2.23 may be written in the alternative form

$$(2.17a)\quad g_l^{(j)}(\gamma_l \mid n(\omega) = l, x_0) = \frac{\nu_j\ f_l^{(j)}(\gamma_l \mid x_0)\ P_l^{(j)}(\gamma_l^*)}{P(\Sigma_{lj})} = \frac{f_l^{(j)}(\gamma_l \mid x_0)\ P_l^{(j)}(\gamma_l^*)}{P^{(j)}(\Sigma_{lj})}$$

and that this is the analogue of Eq. 2.17.

Example a. The following particularly simple stratification may be of considerable value in estimating series 1.3 and 1.6, especially when it is based upon a process $[\Omega, P(\Lambda)]$ of the simple character of example *b* of section 2.2.1. The discussion will be given for a general process $[\Omega, P(\Lambda)]$, subject only to the conditions x_0 fixed and $p_0(x_0) \equiv 0$. For practical sampling purposes, the results are of little value unless $[\Omega, P(\Lambda)]$ is simple enough for easy computation of the integrals of Eq. 2.11, say as in the quoted example. The derivations for the general case will serve as further illustrations of the use of Theorem 1.

Assign S integers, $0 = N_1, N_2, N_3, \ldots, N_S$, in increasing order, and denote ∞ by N_{S+1} for convenience. The partition Δ_j, $j = 1, 2, \ldots, S$, of Ω will be defined by ω in Δ_j if $N_j \leq n(\omega) < N_{j+1}$. In this case $\Sigma_{kj} = \Sigma_k$ if $N_j \leq k < N_{j+1}$ and is empty if this condition on k is not met. The formulas 2.20, 2.21, and 2.22 specialize to

$$(2.25)\quad \begin{cases} \nu_j = \sum_{k=N_j}^{N_{j+1}-1} P[n(\omega) = k \mid x_0] , \\ P^{(j)}(\Lambda) = \nu_j^{-1} \sum_{k=N_j}^{N_{j+1}-1} P(\Lambda\Sigma_k) , \\ P^{(j)}[n(\omega) = l \mid x_0] = \begin{cases} 0, \text{ if } l < N_j \text{ or } l \geq N_{j+1} , \\ \nu_j^{-1} P[n(\omega) = l \mid x_0], \text{ if } N_j \leq l < N_{j+1} . \end{cases} \end{cases}$$

In connection with Eqs. 2 25, it should be recalled that $P[\Sigma_0] = 0$. The third of Eqs. 2.25 reduces Eq. 2.23 to

$$(2.17b)\qquad g_l^{(j)}(\gamma_l \mid n(\omega) = l, x_0) = \begin{cases} 0, \text{ if } l < N_j \text{ or } l \geq N_{j+1} , \\ g_l(\gamma_l \mid n(\omega) = l, x_0), \text{ if } N_j \leq l < N_{j+1} . \end{cases}$$

Formula 2.24 needs no specialization other than the interpretation of Δ_j as $N_j \leq n(\omega) < N_{j+1}$.

As will be proposed later, actual experimental sampling would be done by using the distributions specified in Eqs. 2 25 and 2.17*b*. However, in this example of stratification, it is possible to exhibit specific functions $f_l^{(j)}(\gamma_l \mid x_0)$ and $p_l^{(j)}(\gamma_l^*)$ for the construction of the processes $[\Omega, P^{(j)}(\Lambda)]$ from Theorem 1. This will be done next for its mathematical interest.

Careful consideration of the construction leading to Theorem 1 shows that if the index l is equal to one of the partitioning integers N_j, $f_{N_j}^{(j)}(\gamma_{N_j} \mid x_0)$ is the marginal probability (density) of γ_{N_j} computed from the conditional distribution of ω, subject to the condition $N_j \leq n(\omega) < N_{j+1}$. Define the B-cylinder sets

$$\Lambda_i^{(T)} = \left\{ \text{all } \omega \text{ such that } \alpha_i = (x_i', T),\ x_i' \text{ in } (x_i,\ x_i + dx_i) \right\} ,$$

for $i = 1, 2, \ldots, N_j$,

$$\Lambda_{N_j}^{(S)} = \left\{ \text{all } \omega \text{ such that } \alpha_{N_j} = (x_{N_j}', S),\ x_{N_j}' \text{ in } (x_{N_j},\ x_{N_j} + dx_{N_j}) \right\}$$

and

$$(2.26)\quad \Lambda = \prod_{i=1}^{N_j} \Lambda_i^{(T)} + \Lambda_{N_j}^{(S)} \prod_{i=1}^{N_j-1} \Lambda_i^{(T)} .$$

Then

$$(2.27)\quad f_{N_j}^{(j)}(\gamma_{N_j} \mid x_0)\, d\gamma_{N_j} = \nu_j^{-1} \sum_{k=N_j}^{N_{j+1}-1} P(\Lambda\Sigma_k) .$$

Since the second set product on the right in Eq. 2 26 is a subset of Σ_{N_j}, its contribution to Eq 2.27 is

$$\nu_j^{-1} P(\Lambda\Sigma_{N_j}) = \nu_j^{-1} f_{N_j}(\gamma_{N_j} \mid x_0)\, P_{N_j}(\gamma_{N_j}^*)\, d\gamma_{N_j}$$

The contribution of the first term on the right in Eq. 2.26 is

$$\left\{ \nu_j^{-1} \sum_{k=N_j+1}^{N_{j+1}-1} \int_a^b \cdots \int_a^b f_k(\gamma_k | x_0)\, P_k(\gamma_k^*)\, dx_{N_j+1} \cdots dx_k \right\} d\gamma_{N_j} \ .$$

On defining the functions

$$(2.28) \quad M_l^{(j)}(\gamma_l^*) \equiv P_l(\gamma_l^*) + \sum_{k=l+1}^{N_{j+1}-1} \int_a^b \cdots \int_a^b f_k(\gamma_k | \gamma_l^*)\, P_k(\gamma_k^*)\, dx_{l+1} \cdots dx_k \ , \quad \text{for } l \le N_{j+1} - 2 \ ,$$

$$\equiv P_{N_{j+1}-1}(\gamma^*_{N_{j+1}-1}) \ , \quad \text{for } l = N_{j+1} - 1 \ ,$$

it is seen that Eqs. 2.25 and 2.27 give

$$\nu_j = \int_a^b \cdots \int_a^b f_{N_j}(\gamma_{N_j} | x_0)\, M_{N_j}^{(j)}(\gamma^*_{N_j})\, d\gamma_{N_j} \ ,$$

(2.29)

$$f_{N_j}^{(j)}(\gamma_{N_j} | x_0) = \nu_j^{-1} f_{N_j}(\gamma_{N_j} | x_0)\, M_{N_j}^{(j)}(\gamma^*_{N_j}) \ .$$

Similarly, for $N_j \le l \le N_{j+1} - 2$, the conditional (density) functions $f_{l+1}^{(j)}(\gamma_{l+1} | \gamma_l^*)$ are the marginal (density) functions of x_{l+1} under the condition that $l + 1 \le n(\omega) < N_{j+1}$ and that $\alpha_0, \alpha_1, \ldots, \alpha_l$ is of form (x, T), with the position vector γ_l^* in R^{l+1}, that is, let $\Lambda_{l+1}^{(T)}$ and $\Lambda_{l+1}^{(S)}$ be defined as above and write Λ for their sum. Then, using $\alpha_0, \alpha_1, \ldots, \alpha_l$ as specified,

$$f_{l+1}^{(j)}(\gamma_{l+1} | \gamma_l^*)\, dx_{l+1} \equiv \frac{\sum_{k=l+1}^{N_{j+1}-1} P(\Lambda\Sigma_k | \alpha_0, \alpha_1, \ldots, \alpha_l)}{\sum_{k=l+1}^{N_{j+1}-1} P(\Sigma_k | \alpha_0, \alpha_1, \ldots, \alpha_l)} \ .$$

A calculation much like that leading to Eqs. 2.29, but using the conditional form

$$P(\Lambda | \alpha_0, \alpha_1, \ldots, \alpha_l) = \int P(de_{l+1} | \alpha_0, \ldots, \alpha_l) \cdots \int \phi(\Lambda)\, P(de_{N_{j+1}-1} | \alpha_0, \ldots, \alpha_l, \alpha_{l+1}, \ldots, \alpha_{N_{j+1}-2})$$

of Eq. 2.8, leads to

$$(2.30) \quad f_{l+1}^{(j)}(\gamma_{l+1} | \gamma_l^*) = \frac{f_{l+1}(\gamma_{l+1} | \gamma_l^*)\, M_{l+1}^{(j)}(\gamma^*_{l+1})}{\int_a^b f_{l+1}(\gamma_{l+1} | \gamma_l^*)\, M_{l+1}^{(j)}(\gamma^*_{l+1})\, dx_{l+1}} \ , \quad N_j \le l \le N_{j+1} - 2 \ .$$

For $l < N_j$, the distributions $f_l^{(j)}(\gamma_l | x_0)$ are defined from Eqs. 2 29 by integrating with respect to the variables $x_{l+1}, \ldots, x_{N_j}$ over (a,b). Definitions for $l \geq N_{j+1}$ are unnecessary since $P^{(j)}[n(\omega) \leq N_{j+1} - 1] = 1$ in the process $[\Omega, P^{(j)}(\Lambda)]$.

Define $p_l^{(j)}(\gamma_l^*) \equiv 0$ for $l < N_j$. The functions $p_l^{(j)}(\gamma_l^*)$, $N_j \leq l < N_{j+1}$ are now easily defined in succession by using the form of Eqs. 2.2 and 2.24 and the results of Eqs. 2.29 and 2.30. It is convenient, first, to note the recurrence relation

$$M_l^{(j)}(\gamma_l^*) = P_l(\gamma_l^*) + \int_R f_{l+1}(\gamma_{l+1} | \gamma_l^*)\, M_l^{(j)}(\gamma_{l+1}^*)\, dx_{l+1} \tag{2.31}$$

for the function defined in Eq. 2.28. This leads to

$$p_l^{(j)}(\gamma_l^*) = \frac{P_l(\gamma_l^*)}{M_l^{(j)}(\gamma_l^*)}, \qquad N_j \leq l \leq N_{j+1} - 1 . \tag{2.32}$$

Evidently,

$$p_{N_{j+1}-1}^{(j)}(\gamma_{N_{j+1}-1}^*) \equiv 1 ,$$

in agreement with

$$P^{(j)}[n(\omega) < N_{j+1}] = 1 .$$

As already mentioned, sampling from the process $[\Omega, P^{(j)}(\Lambda)]$ will be performed by using Eqs. 2.25 and 2.17*b*. First, draw a value of $n(\omega)$ by using the third formula in Eqs. 2.25 for the distribution of $n(\omega)$, and then draw $x_1, x_2, \ldots, x_{n(\omega)}$ from Eq. 2.17*b* to obtain $\gamma_{n(\omega)}$. For such an order of procedure to be practicable, explicit calculation of Eq. 2.25 is required. For a general process $[\Omega, P(\Lambda)]$ this might be very difficult but, if the basic process $[\Omega, P(\Lambda)]$ used in the construction described above is of the simple type described in example *b* of 2.2.1, the necessary calculations could be performed easily. Moreover, sampling from Eq. 2.17*b* would be simple. Indeed, in this case, the probability functions 2.25, 2.17*b*, 2.29, and 2.30 simplify to such an extent that sampling from the individual strata could be performed either as indicated above or in the sequential manner of a random walk. To see this last remark, first recall that $P_l(\gamma_l^*)$ has the simple form $P_l(x_0)$. Then Eq. 2.28 reduces to

$$M_l^{(j)}(x_0) = \sum_{k=l}^{N_{j+1}-1} P_k(x_0) .$$

On using this, Eqs 2.29, 2.30, and 2.32 reduce to

$$\begin{cases} f_{N_j}^{(j)}(\gamma_{N_j} | x_0) \equiv f_{N_j}(\gamma_{N_j} | x_0) , \\ f_{l+1}^{(j)}(\gamma_{l+1} | \gamma_l^*) \equiv f_{l+1}(\gamma_{l+1} | \gamma_l^*) , \quad \text{if } N_j \leq l \leq N_{j+1} - 2 , \\ p_l^{(j)}(x_0) = \dfrac{P[n(\omega) = l | x_0]}{\sum_{k=l}^{N_{j+1}-1} P[n(\omega) = k | x_0]} , \quad \text{if } N_j \leq l \leq N_{j+1} - 1 . \end{cases} \tag{2.33}$$

In the last of these three formulas, the terms $P[n(\omega) = k | x_0]$ are given by Eq. 2.18.

Example b This example is quite complicated, perhaps too much so to be practicable. It is constructed from

example a above, that is, the simple case given in Eqs 2.33

Select a sequence of integers r_1, r_2, .. , r_S and, for each $j = 1, 2, \ldots, S$ of example a, partition the set Δ_j of all ω for which $N_j \leq n(\omega) < N_{j+1}$ into r_j Borel subsets Δ_{j1}, Δ_{j2} ..., Δ_{jr_j} that are nonoverlapping in pairs. First, partition the space R^{N_j} of $\gamma_{N_j} = (x_1, x_2, \ldots, x_{N_j})$ into r_j disjoint Borel sets Γ_{j1}, Γ_{j2}, ..., Γ_{jr_j}. Second, define Δ_{ji}, $i = 1, 2, \ldots, r_j$, as the set of all sequences ω contained in Δ_j for which γ_{N_j} is in Γ_{ji}. This Ω is partitioned into the $\sum_{j=1}^{S} r_j$ subsets Δ_{ji}, $j = 1\ 2, \ldots, S$, $i = 1, 2, \ldots, r_j$. Next, stratify $[\Omega, P(\Lambda)]$ into $\sum_{j=1}^{S} r_j$ strata by as many processes $[\Omega, P^{(j,i)}(\Lambda)]$ for which

$$P^{(j,i)}(\Delta_{ji}) = 1 , \qquad P^{(j,i)}(\Lambda) = \frac{P(\Lambda\Delta_{ji})}{P(\Delta_{ji})} , \qquad j = 1, 2, \ldots, S, \quad i = 1, 2, \ldots, r_j . \tag{2.34}$$

Since, for each $i = 1, 2, \ldots, r_j$, the Δ_{ji} are in Δ_j, the second of Eqs. 2.34 may be written in terms of the probability functions $P^{(j)}(\Lambda)$ of the stratifications over the Δ_j. Thus, since

$$P^{(j)}(\Lambda\Delta_{ji}) = \frac{P(\Lambda\Delta_{ji}\Delta_j)}{P(\Delta_j)} = \frac{P(\Lambda\Delta_{ji})}{P(\Delta_j)}$$

and

$$P^{(j)}(\Delta_{ji}) = \frac{P(\Delta_{ji}\Delta_j)}{P(\Delta_j)} = \frac{P(\Delta_{ji})}{P(\Delta_j)} ,$$

Eqs. 2.34 become

$$P^{(j,i)}(\Delta_{ji}) = 1 , \qquad P^{(j,i)}(\Lambda) = \frac{P^{(j)}(\Lambda\Delta_{ji})}{P^{(j)}(\Delta_{ji})} . \tag{2.34a}$$

Thus Eqs. 2.34 amount to substratification of the individual strata $[\Omega, P^{(j)}(\Lambda)]$ over the sets Δ_j, $j = 1, 2, \ldots, S$.

It should be evident to the reader that if $[\Omega, P^{(j)}(\Lambda)]$ is specified by Theorem 1 in terms of functions 2.33, then $[\Omega, P^{(j,i)}(\Lambda)]$ is specified by Theorem 1 and the functions

$$\begin{cases} f_{N_j}^{(j,i)}(\gamma_{N_j}|x_0) = \begin{cases} \dfrac{f_{N_j}(\gamma_{N_j}|x_0)}{\int \cdot \cdot \int_{\Gamma_{ji}} f_{N_j}(\gamma_{N_j}|x_0)\, d\gamma_{N_j}} , & \text{for } \gamma_{N_j} \text{ in } \Gamma_{ji} , \\ 0, \text{ otherwise,} \end{cases} \\ f_{l+1}^{(j,i)}(\gamma_{l+1}|\gamma_l^*) = f_{l+1}(\gamma_{l+1}|\gamma_l^*) , \quad \text{for } N_j \leq l \leq N_{j+1} - 2 , \\ p_l^{(j,i)}(x_0) = p_l^{(j)}(x_0) , \quad \text{for } N_j \leq l \leq N_{j+1} - 1 . \end{cases} \tag{2.35}$$

2.3.2. Processes for Repeated Independent Sampling of Chains of a Given Length $n(\omega)$. It has been remarked, in connection with example *b* of 2.2.1 and the stratifications of it defined by Eqs. 2.33 or 2.35 and Theorem 1, that experimental sampling could proceed as follows. For some integer M, let $n_i(\omega)$, $i = 1, 2, \ldots, M$, be M stochastically independent values drawn from the distribution of $n(\omega)$. These having been obtained, the associated position vectors $\gamma_{n_i(\omega)}$ would be drawn independently from the appropriate form of Eq. 2.17. Such a procedure would probably not utilize the computing efficiency possible in high-speed automatic equipment. Once a value of $n(\omega)$ had been chosen and the sampling process for $\gamma_{n(\omega)}$ set up in a machine, many independent chains $\gamma_{n(\omega)}$ might be run in a very short time. A process design is needed that will allow repeated sampling of $\gamma_{n(\omega)}$ for each $n(\omega)$ that is obtained.

A new sample space must be defined. The explanation will be given in terms of a general process $[\Omega, P(\Lambda)]$, supposed already defined, but any practical case would likely be based upon a simple $[\Omega, P(\Lambda)]$ of the type described in example *b* of section 2 2.1. It will be assumed, without further mention, that the given process $[\Omega, P(\Lambda)]$ is conditioned for some specific choice $\alpha_0 = (x_0, T)$, all finite chains γ_n will begin at x_0 and $p_0(x_0) = 0$ is assumed. Let M be an assigned integer and denote M-tuples $(x^{(1)}, x^{(2)}, \ldots, x^{(M)})$, of points $x^{(i)}$ in R by the single letter z. Let Z denote the collection of all elements of the forms (z, S) and (z, T) That is, Z is a collection similar to X of section 2.2, except that the single point x in the elements (x, S) and (x, T) of X is replaced by an M-tuple z of such points Denote elements of Z by β. Let Ω_M denote the collection of all denumerably infinite sequences $\xi = (\beta_0, \beta_1, \beta_2, \ldots)$ of elements β_i chosen from Z Clearly, Ω_1 is the original collection Ω. A finite chain $(z_0, z_1, \ldots, z_n)$ denotes the position components of the elements $\beta_0, \beta_1, \ldots, \beta_n$ of ξ. Such a chain consists of n chains $\gamma_n^{*(i)} = (x_0^{(i)}, x_1^{(i)}, \ldots, x_n^{(i)})$, $i = 1, 2, \ldots, M$, of the type already used, and this notation will be useful on occasion.

Let l and k be any pair of integers, with $l \leq k$. By using the probability functions 2.2, 2 3, and $P[n(\omega) = k | x_0]$ of the given process $[\Omega, P(\Lambda)]$, define the collection of functions

$$(2.36) \quad J_k(z_0, z_1, \ldots z_l) = \begin{cases} \dfrac{\prod_{i=1}^{M} \int_R \cdots \int_R f_k(\gamma_k^{(i)} | \gamma_l^{*(i)})\, P_k(\gamma_k^{*(i)})\, dx_{l+1}^{(i)} \cdots dx_k^{(i)}}{\{P[n(\omega) = k | x_0]\}^M}, & \text{if } l < k, \\[2ex] \dfrac{\prod_{i=1}^{M} P_k\left(\gamma_k^{*(i)}\right)}{\{P[n(\omega) = k | x_0]\}^M}, & \text{if } l = k. \end{cases}$$

Probability functions for the z's that are analogous to Eqs. 2.2 and 2.3 are defined by

$$
(2.37)\quad
\begin{aligned}
f_{1M}(z_1|z_0) &\equiv \left\{\prod_{i=1}^{M} f_1(x_1^{(i)}|x_0)\right\} \sum_{k=1}^{\infty} P[n(\omega)=k|x_0]\ J_k(z_0,z_1)\ ,\\
f_{NM}(z_N|z_0,z_1,\ldots,z_{N-1})&\\
&\equiv \left\{\prod_{i=1}^{M} f_N(x_N^{(i)}|\gamma_{N-1}^{*(i)})\right\} \frac{\sum_{k=N}^{\infty} P[n(\omega)=k|x_0]\ J_k(z_0,z_1,\ldots,z_N)}{\sum_{k=N}^{\infty} P[n(\omega)=k|x_0]\ J_k(z_0,z_1,\ldots,z_{N-1})}\ ,
\end{aligned}
$$

and

$$
(2.38)\quad
\begin{cases}
p_{0M}(z_0)=0,\ p_{NM}(z_0,z_1,\ .\ ,z_N)\\
\qquad \equiv \dfrac{P[n(\omega)=N|x_0]\ J_N(z_0,z_1,\ldots,z_N)}{\sum_{k=N}^{\infty} P[n(\omega)=k|x_0]\ J_k(z_0,z_1,\ldots,z_N)},\ N=1,\ 2,\ 3,\ \ldots,\\
q_{NM}\equiv 1-p_{NM}\ ,\\
P_{NM}(z_0,z_1,\ldots,z_N)\equiv q_{1M}(z_0,z_1)\ldots q_{N-1,M}(z_0,\ldots,z_{N-1})\ p_{NM}(z_0,\ldots,z_N)\ .
\end{cases}
$$

Theorem 1 may be revamped by using Eqs. 2.37 and 2.38 in place of their analogues to generate a stochastic process on Ω_M, to be denoted by $[\Omega_M, P_M(\Lambda)]$, which will be conditioned for $\beta_0 = (z_0, T)$, where $z_0 = (x_0, x_0, \ldots, x_0)$ to M components. The random variable $n(\xi)$ on the elements ξ of Ω_M is defined as the smallest integer for which the component $\beta_{n(\xi)}$ of ξ has the form $(z_{n(\xi)}, S)$. It is not difficult to see that the process $[\Omega_M, P_M(\Lambda)]$ is such that the probability (density) of $z_1, z_2, \ldots, z_{n(\xi)}$ and $n(\xi)$ is given by (for $N = 1, 2, 3, \ldots$)

$$
(2.39)\quad
\begin{aligned}
&f_{NM}(z_1,\ldots,z_N|z_0)\ P_{NM}(z_0,\ldots,z_N)\\
&\equiv f_{1M}(z_1|z_0)\ f_{2M}(z_2|z_0,z_1)\ldots f_{NM}(z_N|z_0,\ldots,z_{N-1})\ P_{NM}(z_0,\ldots,z_N)\\
&\equiv P[n(\omega)=N|x_0]\ \prod_{i=1}^{M}\left\{\frac{f_N(\gamma_N^{(i)}|x_0)\ P_N(\gamma_N^{*(i)})}{P[n(\omega)=N|x_0]}\right\}\\
&\equiv P[n(\omega)=N|x_0]\ \prod_{i=1}^{M}\{g_N(\gamma_N^{(i)}|x_0,\ n(\omega)=N\}\ .
\end{aligned}
$$

This shows that sampling may be conceived entirely in terms of Ω and its probability functions. The probability $P[n(\xi) = N|x_0]$ will be obtained by integrating Eq. 2.39 with respect to all of the x's involved except x_0. Clearly, the result is

$$(2.40) \quad P[n(\xi) = N|x_0] = P[n(\omega) = N|x_0] \ .$$

Experimental sampling for the relevant part of elements ξ is performed in the following manner. First, draw $n(\xi) = n(\omega)$ at random from the distribution of $n(\omega)$ in $[\Omega, P(\Lambda)]$. Second, draw M independent chains $\gamma^{(i)}_{n(\omega)}$, $i = 1, 2, \ldots, M$, at random by using the conditional distribution $g_{n(\omega)}[\gamma_{n(\omega)}|x_0, n(\omega)]$. Then form $\beta_0, \beta_1, \ldots, \beta_{n(\xi)}$ in the manner indicated above in the description of the sampling space Ω_M. The elements β_k, $k > n(\xi)$, of ξ are irrelevant.

As a consequence of the reduction obtained in Eqs. 2 39 and 2 40, it will be possible to combine multiple sampling and stratification. This will be done in section 3.4, and a discussion of the process will be given.

2.3.3. Product Spaces and Measures. Let $[\Omega, P(\Lambda)]$ be a stochastic process given by Theorem 1. For some positive integer M, let $\Omega^{(M)} \equiv \Omega \times \Omega \times \ldots \times \Omega$ be the familiar product space of M factors Ω That is, $\Omega^{(M)}$ is the collection of all M-tuples $(\omega_1, \omega_2, \ldots, \omega_M)$ with ω_i in Ω for $i = 1, 2, \ldots, M$. A probability measure may be defined on a Borel field of subsets of $\Omega^{(M)}$, as follows. Let A be such a set and let A_{ik}, $i = 1, 2, \ldots, M$, $k = 1, 2, \ldots, l$, be Borel subsets of Ω such that $\sum_{k=1}^{l} \prod_{i=1}^{M} A_{ik}$ covers A. Define the probability of the set A as

$$(2.41) \quad P(A) \equiv \text{g.l.b.} \sum_{k=1}^{l} \prod_{i=1}^{M} P(A_{ik}) \ ,$$

where the greatest lower bound is taken relative to all possible coverings of A of the type described. The measure defined by Eq. 2.41 is termed a product measure, and the combination $[\Omega^{(M)}, P(A)]$ is a stochastic process on the space $\Omega^{(M)}$ If A is a Borel product set, $A = \prod_{i=1}^{M} A_i$, where A_i is a Borel subset of Ω for each i, from Eq. 2.41 it is seen that $P(A) = \prod_{i=1}^{M} P(A_i)$. The sets A_i are called stochastically independent. If M chains $\omega_1, \omega_2, \ldots, \omega_M$ are subject to ω_i in A_i for each i, the chains are said to be stochastically independent.

A modification of the product measure given by Eq. 2.41 will be used in section 3.4 in connection with a stratified process such as is given by Definition 3. Let $M_1, M_2, \ldots, M_S$ be positive integers and let $\Omega^{(M)}$, $M = \sum_{j=1}^{S} M_j$, again be a product space of factors Ω. In place of the definition 2.41, a probability measure will be defined by

$$(2\ 42) \ P(A) \equiv \text{g.l.b.} \sum_{k=1}^{l} \prod_{j=1}^{S} \prod_{i=M_{j-1}+1}^{M_j} P^{(j)}(A_{ik}) \ ,$$

where $M_0 = 1$ and the symbol $P^{(j)}$ denotes the measure for the jth stratum. The essential difference between definitions 2.41 and 2.42 is that different probability measures have been used on the right in definition 2.42 on groups of factors M_j of $\Omega^{(M)}$. Again, M chains are said to be stochastically independent if A is a product set. The intuitive concept of M independent chains is the following. The M_1 chains $\omega_1, \omega_2, \ldots, \omega_{M_1}$ are drawn independently from $[\Omega^{M_1}, P^{(1)}(A)]$ by using a measure defined by Eq. 2.41, the M_2 chains $\omega_{M_1+1}, \omega_{M_1+2}, \ldots, \omega_{M_1+M_2}$ are drawn independently from $[\Omega^{M_2}, P^{(2)}(A)]$ and so on. Note that, apart from sets of probability zero, in considering Eq. 2.42, the product space $\Omega^{(M)}$ may be replaced by the product set $\Delta_1^{(M_1)} \times \Delta_2^{(M_2)} \times \ldots \times \Delta_S^{(M_S)}$, where $\Delta_j^{(M_j)} \equiv \Delta_j \times \Delta_j \times \ldots \times \Delta_j$, to M_j factors.

CHAPTER 3. ESTIMATION AT A FIXED POINT

3.1. SUMMARY AND REMARKS

Stochastic estimation of the solution $\phi(x)$ of Eq. 1.1 at a single point $x = x_0$ (or of a single component of the vector solution ψ of Eq. 1 5 is based upon the definition of a random variable $\hat{\phi}(\omega, x_0)$ over Ω, which has some standard estimation character for $\phi(x_0)$ in a statistical sense, x_0 is a parameter. In this report it will be required that the expectation $E\{\hat{\phi}(\omega, x_0) | x_0\}$ in the conditioned process $[\Omega, P(\Lambda)]$ be equal to $\phi(x_0)$. That is, $\hat{\phi}(\omega, x_0)$ is an unbiased estimate of $\phi(x_0)$. The average of, say M, stochastically independent values of $\hat{\phi}(\omega, x_0)$ is also an unbiased estimate of $\phi(x_0)$. Variations of this simple idea are used in connection with the stochastic processes already defined for purposes of constructing estimation procedures that may balance statistical efficiency and computing efficiency in some sensible manner.

Section 3.2 contains a discussion of the important sampling characteristics of the particular random variable that has been most discussed in the literature for this estimation problem and which is the basic variable to be used in this report. Section 3.3 contains a discussion of a special Markoff chain process for which $\hat{\phi}(\omega, x_0)$ has zero variance. It is definable only if $g(x)$ and $\lambda K(x,y)$ of Eq. 1.1 are nonnegative and bounded. There has been much speculation as to the possibilities for approximation of such zero variance processes. The author, being skeptical of the feasibility of such approximations, proposes to use, as a substitute, representative sampling from stratified simple processes in conjunction with high-speed sampling. Indeed, it will be pointed out that such a substitution actually amounts to a new method for approximating zero variance processes.

3.2. THE BASIC ESTIMATING VARIABLE

Suppose that a particular equation of the type of Eq. 1.1 is to be solved at $x = x_0$ and that a process $[\Omega, P(\Lambda)]$ of the type described in section 2.2 (conditioned for $\alpha_0 = (x_0, T)$) has been selected for the purpose. Define the random variable

$$(3.1) \qquad \hat{\phi}(\omega, x_0) = g(x_0) + \frac{G_{n(\omega)}(x_0, \gamma_{n(\omega)})}{f_{n(\omega)}(\gamma_{n(\omega)} | x_0)\, P_{n(\omega)}(x_0, \gamma_{n(\omega)})}$$

for each ω in Ω by using Eqs. 1.2, 2.2, and 2.3 on the right.

The expectation of the random variable 3.1 in $[\Omega, P(\Lambda)]$ is

$$E\,\hat{\phi}(\omega, x_0) = g(x_0) + \sum_{k=1}^{\infty} P[n(\omega) = k | x_0]\; E\left\{ \frac{G_k(x_0, \gamma_k)}{f_k(\gamma_k | x_0)\, P_k(x_0, \gamma_k)} \;\middle|\; x_0,\ n(\omega) = k \right\}$$

$$= g(x_0) + \sum_{k=1}^{\infty} P[n(\omega) = k | x_0] \int_R \cdots \int_R \frac{G_k(x_0, \gamma_k)}{f_k(\gamma_k | x_0)\, P_k(x_0, \gamma_k)}\, g_k[\gamma_k | x_0,\ n(\omega) = k]\, d\gamma_k$$

$$= g(x_0) + \sum_{k=1}^{\infty} I_k(x_0) .$$

Since this is the series 1.3, it is seen that $\hat{\phi}(\omega, x_0)$ is unbiased. Its variance is given by

$$(3.2) \qquad \operatorname{var} \hat{\phi}(\omega,x_0) = \operatorname{var} [\hat{\phi}(\omega,x_0) - g(x_0)]$$
$$= E\{[\hat{\phi}(\omega,x_0) - g(x_0)]^2 | x_0\} - [\phi(x_0) - g(x_0)]^2 .$$

Now

$$(3.3) \qquad E\{[\hat{\phi}(\omega,x_0) - g(x_0)]^2 | x_0\}$$
$$= \sum_{k=1}^{\infty} P[n(\omega) = k | x_0] \int_R \cdots \int_R \left\{ \frac{G_k(x_0, \gamma_k)}{f_k(\gamma_k | x_0)\, P_k(x_0, \gamma_k)} \right\}^2 g_k[\gamma_k | x_0, n(\omega) = k]\, d\gamma_k$$
$$= \sum_{k=1}^{\infty} \int_R \cdots \int_R \frac{G_k^2(x_0, \gamma_k)}{f_k(\gamma_k | x_0)\, P_k(x_0, \gamma_k)}\, d\gamma_k .$$

In general, the variance formulas 3.2 and 3.3 are likely to be completely intractable. However, in a practical experiment that is to make use of, say M, stochastically independent values of $\hat{\phi}(\omega,x_0)$ from the process $[\Omega^{(M)}, P(A)]$ of Eq. 2.41, the variance of the average may be estimated by familiar statistical methods.

The following comment may aid the reader in an intuitive manner. The random variable

$$(3.4) \quad P[n(\omega) = k | x_0]\, [\hat{\phi}_k(\omega,x_0) - g(x_0)] , \qquad k \text{ fixed},$$

is a conditional unbiased estimate of the term $I_k(x_0)$ of series 1.3 under the condition that $n(\omega) = k$, that is, variable 3.4 is an unbiased estimate of $I_k(x_0)$ relative to the distribution $g_k[\gamma_k | x_0, n(\omega) = k]$ of γ_k. Thus it appears that the method for estimating $\phi(x_0)$ consists of three steps: (*a*) choose a term I_k of series 1.3 at random by using the distribution of $n(\omega)$, (*b*) estimate I_k by the random variable 3.4, (*c*) adjust the estimate of $I_k(x_0)$ by division by $P[n(\omega) = k | x_0]$ to obtain an estimate of the entire series.

Evidently, the series of integrals being estimated need not have the particular form of the Neumann series 1.3. Any series of integrals of steadily increasing multiplicities could be estimated in this manner.

Modification of the estimating process described above for the estimation of a component of the vector ψ given by Eq. 1.6 should be evident. The space R is discrete (the integers 1, 2, ..., r), and the functions $f_k(\gamma_k | x_0)\, P_k(x_0, \gamma_k)$ are correspondingly defined. To estimate the ith component of ψ, the analogue of variable 3.1 will be obtained by replacing $g(x_0)$ by the ith component of h and replacing the numerator $G_{n(\omega)}(x_0, n(\omega))$ by the ith component of the vector $(A + I)^{n(\omega)} h$.

3.3. ZERO VARIANCE PROCESSES

Suppose that $g(x)$ and $\lambda K(x,y)$ of Eq. 1.1 are nonnegative and bounded on the ranges of their variables. It will be shown that there exists a process $[\Omega, P(\Lambda)]$ for which the variance 3.2 of 3.1 is zero. This process is of no practical value because $\phi(x)$ must be known for its construction.

Approximation methods that may be practical will be evident from the discussion.

Let $\alpha(x)$ be any nonnegative, bounded function on R and define

$$(3.5)\quad \beta(x) \equiv g(x) + \lambda \int_R K(x,y)\,\alpha(y)\,dy\ .$$

In terms of these functions, define a process $[\Omega, P(\Lambda)]$ from

$$(3.6)\quad \begin{cases} h(y|x) \equiv \dfrac{\lambda\, K(x,y)\,\alpha(y)}{\beta(x) - g(x)}, \\ f_k(\gamma_k|x_0) \equiv \prod_{i=1}^{k} h(x_i|x_{i-1})\ , \\ p_0(x_0) \equiv 0 \\ p_k(\gamma_k^*) \equiv \dfrac{g(x_k)}{\beta(x_k)}\ , \qquad k \geq 1\ . \end{cases}$$

By using Eqs. 3.6, 3.1 reduces to

$$(3.7)\quad \hat{\phi}_\alpha(\omega, x_0) = g(x_0) + [\beta(x_0) - g(x_0)] \prod_{i=1}^{n(\omega)} \frac{\beta(x_i)}{\alpha(x_i)}\ .$$

Since, in this case,

$$\frac{G_k^2(x_0,\gamma_k)}{f_k(\gamma_k|x_0)\,P_k(x_0,\gamma_k)} \equiv [\beta(x_0) - g(x_0)]\ g(x_k) \prod_{i=1}^{k} \frac{\lambda\, K(x_{i-1},x_i)\,\beta(x_i)}{\alpha(x_i)}\ ,$$

the variance 3.3 reduces to

$$(3.8)\quad E\{[\hat{\phi}_\alpha(\omega,x_0) - g(x_0)]^2 | x_0\} = [\beta(x_0) - g(x_0)] \sum_{k=1}^{\infty} \int_R \cdots \int_R g(x_k) \prod_{i=1}^{k} \left[\frac{\lambda\, K(x_{i-1},x_i)\,\beta(x_i)}{\alpha(x_i)}\right] d\gamma_k\ .$$

The summation on the right in Eq. 3.8 is the formal Neumann series for the quantity $\psi(x_0) - g(x_0)$ defined by the integral equation

$$(3\ 9)\quad \psi(x) = g(x) + \lambda \int_R K(x,y)\,\frac{\beta(y)}{\alpha(y)}\,\psi(y)\,dy\ .$$

Thus, in this special case,

$$(3.10)\quad \operatorname{var} \hat{\phi}_\alpha(\omega,x_0) = [\beta(x_0) - g(x_0)]\,[\psi(x_0) - g(x_0)] - [\phi(x_0) - g(x_0)]^2\ .$$

It is evident that $\alpha(x) \equiv \phi(x)$ implies $\beta(x) \equiv \alpha(x) \equiv \psi(x) \equiv \phi(x)$ and variance 3.10 has the value zero.

Approximation to the zero variance case seems evident by choosing as $\alpha(x)$ some approximation to $\phi(x)$. However, such an approximation may be very hard to find. Indeed, the purpose of a Monte Carlo experiment is to obtain an approximation to $\phi(x)$. The circular situation is self-evident. One immediately questions the value of a Monte Carlo experiment if, in order to design it, a good approximation to $\phi(x)$ is already necessary. Certainly one would insist in such a procedure that the stochastic process provide a useful method for improving the accuracy of the approximation used in the construction of that process. Such a demand might place exceedingly stringent requirements on the statistical accuracy of the stochastic estimation method. Moreover, practicability of such a composite approximation method would imply that the stochastic process furnishes a better method (in some sense) of improving the approximation already at hand than would some other available method. Explicitly, let $0 < K < 1$, and suppose it is

desired that the average of M values of variable 3.7 be a more accurate estimate of $\phi(x_0)$ than is $\beta(x_0)$ with large probability.* This requirement might be expressed by requiring that three standard deviations of the stochastic estimate be no larger than

$$K|\beta(x_0) - \phi(x_0)| \ .$$

That is

$$\frac{9 \text{ var } \hat{\phi}_\alpha(\omega, x_0)}{M} \leq K^2|\beta(x_0) - \phi(x_0)|^2 \ .$$

This leads to the requirement that the sample size M satisfy

$$(3.11) \qquad M \geq \frac{9\{[\beta(x_0) - g(x_0)]\ [\psi(x_0) - g(x_0)] - [\phi(x_0) - g(x_0)]^2\}}{K^2[\beta(x_0) - \phi(x_0)]^2} \ .$$

The following remarks are now pertinent. In general, there would be little point to the use of a Monte Carlo method for estimating $\phi(x_0)$ if the operator of the integral equations were such that iteration was rapidly convergent, that is, if the Neumann series 1.3 converged rapidly. In the opposite case, series 1.3 is essentially useless for computational purposes. Moreover, if $\alpha(x)$ in Eq. 3.5 is a fair approximation to $\phi(x)$, slow convergence of series 1.3 very likely implies that $\beta(x)/\alpha(x)$ differs little from unity and the solution $\psi(x)$ of Eq. 3.9 is very close to $\phi(x)$. However, the right member of 3.11 is then of the order of

$$\frac{9[\phi(x_0) - g(x_0)]}{K^2[\beta(x_0) - \phi(x_0)]} \ ,$$

and it is not difficult to see the implication of this reduction. Very large M would be required to achieve an improvement of the estimate $\beta(x_0)$ already obtained.

3.4. REPRESENTATIVE SAMPLING FROM A STRATIFIED PROCESS

Suppose that a process $[\Omega, P(\Lambda)]$ and a stratification of it into $[\Omega, P^{(j)}(\Lambda)]$ over Δ_j, $j = 1, 2, \ldots, S$, are given. It will be shown in this section that a reduction in the variance of the average of M independent values of the variable 3.1, taken over $[\Omega, P(\Lambda)]$, may be achieved by an appropriate division of the sampling over the S strata. Before proceeding, it will be convenient to derive a new formula for the variance 3.2.

Referring to the notations of section 2.3.1, the variance 3.2 may be written as

$$(3.12) \qquad \text{var } \hat{\phi}(\omega, x_0) = \sum_{j=1}^{S} \nu_j \ E\{[\hat{\phi}(\omega, x_0) - \phi(x_0)]^2 \mid \omega \text{ in } \Delta_j\} \ .^{**}$$

Define $\phi_j(x_0)$ for $j = 1, 2, \ldots, S$ by

$$(3.13) \qquad \nu_j \ \phi_j(x_0) \equiv \nu_j \ g(x_0) + \sum_{k=1}^{\infty} \int \cdots \int_{\Sigma_{kj}} G_k(x_0, \gamma_k) \ d\gamma_k \ .$$

The notation Σ_{kj} for the range of integration is symbolic for the set of all γ_k for which the ω fall in $\Sigma_{kj} = \Sigma_k \Delta_j$. The random variable 3.1 has the conditional expectation

*The construction implies that $\beta(x)$ is known.

**A process conditioned for fixed $\alpha_0 = (x_0, T)$ is to be used. Indication of this conditioning will be omitted except where confusion might arise.

$$E\{\hat{\phi}(\omega,x_0)|\omega \text{ in } \Delta_j\} = g(x_0)$$

$$+\sum_{k=1}^{\infty}\int\cdots\int_{\Sigma_{kj}} \frac{G(x_0,\gamma_k)}{f_k(\gamma_k|x_0)\,P_k(x_0,\gamma_k)}$$

$$g_k^{(j)}(\gamma_k|x_0)\,P_k^{(j)}(x_0,\gamma_k)\,d\gamma_k = \phi_j(x_0)\ ,$$

for each $j = 1, 2, \ldots, S$. It follows that the variance of the average $\overline{\hat{\phi}(\omega,x_0)}$ of M stochastically independent values of variable 3.1 (from $[\Omega^{(M)},P(A)]$, Eq. 2.41) is given by

$$\text{(3.14)}\qquad \text{var}\ \overline{\hat{\phi}(\omega,x_0)} = \frac{1}{M}\left\{\sum_{j=1}^{S}\nu_j E\{[\hat{\phi}(\omega,x_0) - \phi_j(x_0)]^2|\omega \text{ in } \Delta_j\}\right.$$

$$\left. + \sum_{j=1}^{S}\nu_j[\phi_j(x_0) - \phi(x_0)]^2\right\}.$$

In this formula,

$$\text{(3.15)}\qquad E\{[\hat{\phi}(\omega,x_0) - \phi_j(x_0)]^2|\omega \text{ in } \Delta_j\}$$

$$\equiv\sum_{k=1}^{\infty}\int\cdots\int_{\Sigma_{kj}} \frac{G_k^2(x_0,\gamma_k)\,f_k^{(j)}(\gamma_k|x_0)\,P_k^{(j)}(x_0,\gamma_k)}{[f_k(\gamma_k|x_0)\,P_k(x_0,\gamma_k)]^2}\,d\gamma_k$$

$$- [\phi_j(x_0) - g(x_0)]^2\ .$$

Suppose now that for each $j = 1, 2, \ldots, S$, the quantity $\phi_j(x_0)$ defined in Eq. 3.13 is estimated by using an analogue of variable 3.1 on the jth stratum $[\Omega,P^{(j)}(\Lambda)]$ of $[\Omega,P(\Lambda)]$. The random variable used to estimate $\phi_j(x_0)$ is

$$\text{(3.16)}\qquad \hat{\phi}_j(\omega,x_0) \equiv g(x_0) + \frac{G_{n(\omega)}(x_0,\gamma_{n(\omega)})}{\nu_j\ f_{n(\omega)}^{(j)}(\gamma_{n(\omega)}|x_0)\ P_{n(\omega)}^{(j)}(x_0,\gamma_{n(\omega)})}\ ,$$

and the sampling distributions involved are Eqs. 2.22 and 2.17*a*. By methods exactly like those used to obtain formulas 3.2 and 3.3, the mean and variance of the random variable 3.16 are given by

$$E\ \hat{\phi}_j(\omega,x_0) = \phi_j(x_0)\ ,$$

$$\text{var}\ \hat{\phi}_j(\omega,x_0) = \sum_{k=1}^{\infty}\int\cdots\int_{\Sigma_{kj}} \frac{G_k^2(x_0,\gamma_k)\,d\gamma_k}{\nu_j^2\ f_k^{(j)}(\gamma_k|x_0)\ P_k^{(j)}(x_0,\gamma_k)} - [\phi_j(x_0) - g(x_0)]^2\ .$$

Note that the variance here is given by the series 3.15.

Let the integer M be decomposed into a sum

$$M = \sum_{j=1}^{S} M_j$$

of integers M_j in some manner, and suppose that, independently for each $j = 1, 2, \ldots, S$, M_j stochastically independent values $\hat{\phi}_j(\omega_{ij}, x_0)$, $i = 1, 2, \ldots, M_j$, of variable 3.16 are drawn from $[\Omega, P^{(j)}(\Lambda)]$. Let λ_{ij}, $i = 1, 2, \ldots, M_j$, $j = 1, 2, \ldots, S$, be M constants, and form the combination

$$(3.17) \quad \phi_R(x_0) \equiv \sum_{j=1}^{S} \sum_{i=1}^{M_j} \lambda_{ij} \, \hat{\phi}_j(\omega_{ij}, x_0) .$$

The combination 3.17 is a random variable on the process $[\Omega^{(M)}, P(A)]$ with measure defined by Eq. 2.42. Moreover, the specification that the chains ω_{ij}, $i = 1, 2, \ldots, M_j$, $j = 1, 2, \ldots, S$, be stochastically independent implies that they be drawn from a product set

$$\prod_{j=1}^{S} \prod_{i=1}^{M_j} A_{ij} ,$$

ω_{ij} in A_{ij}, where A_{ij} is a Borel subset of Δ_j for $i = 1, 2, \ldots, M_j$ and each $j = 1, 2, \ldots, S$. If it is desired that the combination 3.17 always furnish an unbiased estimate of the solution $\phi(x_0)$ of Eq. 1.1, then

$$E\, \phi_R(x_0) = \sum_{j=1}^{S} \sum_{i=1}^{M_j} \lambda_{ij} \, \phi_j(x_0)$$

implies the condition

$$(3.18) \quad \sum_{i=1}^{M_j} \lambda_{ij} = \nu_j , \qquad j = 1, 2, \ldots, S .$$

Consider the variance of the combination 3.17 subject to condition 3.18, and write

$$\lambda_{\cdot j} = \frac{1}{M_j} \sum_{i=1}^{M_j} \lambda_{ij}$$

for each j. Then

$$\sum_{i=1}^{M_j} \lambda_{ij}^2 = \sum_{i=1}^{M_j} (\lambda_{ij} - \lambda_{\cdot j})^2 + M_j \lambda_{\cdot j}^2$$

$$= \sum_{i=1}^{M_j} \left(\lambda_{ij} - \frac{\nu_j}{M_j}\right)^2 + M_j \left(\frac{\nu_j}{M_j}\right)^2$$

and, for the combination 3.17, the variance is

$$(3.19) \quad \operatorname{var} \phi_R(x_0) = \sum_{j=1}^{S} \left\{ \sum_{i=1}^{M_j} \left(\lambda_{ij} - \frac{\nu_j}{M_j}\right)^2 + M_j \left(\frac{\nu_j}{M_j}\right)^2 \right\} \operatorname{var} \hat{\phi}_j(\omega, x_0) ,$$

because, for a given j, all the $\hat{\phi}_j(\omega_{ij}, x_0)$ have the same variance. Clearly, for a fixed choice of the stratification and the M_j, $\operatorname{var} \phi_R(x_0)$ is minimized by the choice

$$(3.20) \quad \lambda_{1j} = \lambda_{2j} = \ldots = \lambda_{M_j, j} = \frac{\nu_j}{M_j} , \qquad j = 1, 2, \ldots, S ,$$

which is consistent with the condition 3.18. Under the conditions 3.20, the variance 3.19 becomes

$$(3.21) \quad \operatorname{var} \phi_R(x_0) = \sum_{j=1}^{S} \frac{\nu_j^2}{M_j} E\{[\hat{\phi}(\omega, x_0) - \phi_j(x_0)]^2 | \omega \text{ in } \Delta_j\} ,$$

where Eq. 3.15 is used on the right.

An important special case arises when the stratification of $[\Omega, P(\Lambda)]$ is such that the S products, $M\nu_j$, are all integers. In this case, the choices

$$(3.22) \quad M_j = M\nu_j , \qquad j = 1, 2, \ldots, S ,$$

reduce variance 3.21 to

$$(3.21a) \qquad \operatorname{var} \phi_R(x_0) = \frac{1}{M} \sum_{j=1}^{S} \nu_j \, E\{[\hat{\phi}(\omega, x_0) - \phi_j(x_0)]^2 | \omega \text{ in } \Delta_j\} .$$

In other words, the sampling method gives an estimate $\phi_R(x_0)$ whose variance is less than the variance 3.14 of $\hat{\phi}(\omega, x_0)$ by the term

$$\frac{1}{M} \sum_{j=1}^{S} \nu_j [\phi_j(x_0) - \phi(x_0)]^2 .$$

Sampling subject to 3.22 is known as representative sampling from the S processes $[\Omega, P^{(j)}(\Lambda)]$, $j = 1, 2, \ldots, S$.

Note that 3.16 is formally the same as 3.1. There is a sampling difference between the two estimates, however. In using 3.16, ω is forced into Δ_j. Also, the conditions 3.20 and 3.22 reduce the combination 3.17 to the average of M values of variable 3.16. Thus, in this case, $\phi_R(x_0)$ and the average

$$\overline{\hat{\phi}(\omega, x_0)}$$

are formally the same. Only the sampling methods differ between the two estimates.

If the products $M\nu_j$ are not all integers, one could approximate the choice 3.22. The choice 3.20 makes $\phi_R(x_0)$ an unbiased estimate of $\phi(x_0)$, but the precise reduction of variance discussed above would not be achieved. Nevertheless, the discussion above would be expected to apply in some approximate sense.

If the conditions 3.22 hold, the formula 3.21, or 3.21a, lends itself to easy estimation from an experimental sample, as follows In the combination 3.17, for each $j = 1, 2, \ldots, S$, the terms $\hat{\phi}_j(\omega_{ij}, x_0)$ with $i = 1, 2, \ldots, M_j$ are independent estimates of $\phi_j(x_0)$. They may be averaged to obtain

$$\overline{\hat{\phi}_j(\omega_{\ j}, x_0)} = M_j^{-1} \sum_{i=1}^{M_j} \hat{\phi}(\omega_{ij}, x_0)$$

and their sample variance

$$s_j^2 = \frac{1}{M_j - 1} \sum_{i=1}^{M_j} [\hat{\phi}_j(\omega_{ij}, x_0) - \overline{\phi_j(\omega_{\ j}, x_0)}]^2$$

may be computed for each j. This is an estimate of the variance

$$E\{[\hat{\phi}(\omega, x_0) - \phi_j(x_0)]^2 | \omega \text{ in } \Delta_j\}$$

that occurs in 3.21. An estimate of $\operatorname{var} \phi_R(x_0)$ is obtained by inserting the estimate s_j^2 in place of the true variance in formula 3.21, or 3.21a.

If the processes used are such that each of the conditional variances 3.15 is zero, variance 3.21a is zero. Let $[\Delta_j, P^{(j)}(\Lambda)]$ denote a process for which 3.16 is a zero variance estimator of $\phi_j(x_0)$. Assign S nonnegative numbers $\nu_1, \nu_2, \ldots, \nu_S$ and construct a process $[\Omega, P(\Lambda)]$ for which $P(\Delta_j) = \nu_j$ and the $[\Delta_j, P^{(j)}(\Lambda)]$ are the strata over the Δ_j. Under the conditions 3.20 and 3.22, $\phi_R(x_0)$ will be an unbiased estimate of $\phi(x_0)$ with zero variance.

This new way of describing zero variance processes might lend itself to approximation according to the following scheme. Suppose that for some preassigned small positive number ϵ, the partial sum

$$g(x_0) + \sum_{k=1}^{N_\epsilon} I_k(x_0)$$

of Eq. 1.3 is within ϵ of $\phi(x_0)$. As in example b of section 2.3.1, let the partitioning integers $N_2, N_3, \ldots, N_S$ be the consecutive integers from 1 to N_ϵ, and choose the partitioning of each R^k, $k = 1, 2, \ldots, N_\epsilon$, and the individual processes on the elements of the partitions in such a way that approximate zero variance is achieved for each estimator used (cf., Appendix B for stratified sampling in the estimation of multiple integrals). Approximations over coarser stratifications should suggest themselves to the reader. Undoubtedly, quite a bit of ingenuity might be required in the

design of a sampling experiment for which the variance of $\phi_R(x_0)$ is small and the experimental sampling is adaptable for high-speed machine work.

It should be evident that representative sampling could be combined with repeated sampling for single integrals. Given the process $[\Omega, P(\Lambda)]$ and a stratification into $[\Omega, P^{(j)}(\Lambda)]$, $j = 1, 2, \ldots, S$, assign an integer r_j to each stratum and estimate $\phi(x_0)$, as follows. As in forming $\phi_R(x_0)$, independently for each $j = 1, 2, \ldots, S$, draw M_j independent chain lengths $n(\omega_{ij})$, $i = 1, 2, \ldots, M_j$, from the distribution of $n(\omega)$ for the jth stratum. For each pair (i, j), draw r_j chains ω_{ijk}, $k = 1, 2, \ldots, r_j$, of length $n(\omega_{ij})$, independently, from the distribution 2.23 with $l = n(\omega_{ij})$. Average the r_j values $\hat{\phi}(\omega_{ijk}, x_0)$ of the random variable 3.16 on those chains to obtain the unbiased estimate

$$(3.23) \quad \overline{\hat{\phi}_j(\omega_{ij}, x_0)} \equiv \frac{1}{r_j} \sum_{k=1}^{r_j} \hat{\phi}(\omega_{ijk}, x_0)$$

of $\phi_j(x_0)$ for each $i = 1, 2, \ldots, M_j$, $j = 1, 2, \ldots, S$. Define the estimate

$$(3.24) \quad \phi_R^*(x_0) \equiv \sum_{j=1}^{S} \sum_{i=1}^{M_j} \lambda_{ij} \overline{\hat{\phi}_j(\omega_{ij}, x_0)}$$

of $\phi(x_0)$ by using 3.23 on the right and coefficients λ_{ij} subject to conditions 3.18 and 3.20 and, if possible, 3.22. In other words, in 3.17 replace the $\hat{\phi}_j(\omega_{ij}, x_0)$ by the averages 3.23. A discussion of the overall stochastic process on which $\phi_R^*(x_0)$ is defined will be found at the end of this section.

Since the terms on the right in 3.23 all have the same determination (random) of chain length $n(\omega_{ij})$, they are correlated. Thus, to compute the variance of 3.24, it is not sufficient merely to divide the expectations on the right in Eq. 3.21 by the integers r_j. Instead, for each $i = 1, 2, \ldots, M_j$,

$$(3.25) \quad E\left\{\overline{[\hat{\phi}_j(\omega_{ij}, x_0)} - \phi_j(x_0)]^2 \,|\, \omega_{ij} \text{ in } \Delta_j\right\}$$

$$= \sum_{l=1}^{\infty} P^{(j)}[n(\omega) = l \,|\, x_0] \; E\left\{\overline{[\hat{\phi}_j(\omega_{ij}, x_0)} - \phi_j(x_0)]^2 \,|\, n(\omega_{ij}) = l\right\}.$$

The expectations on the right in Eq. 3.25 are relative to the conditional process $[\Omega, P^{(j)}(\Lambda)]$. Under the condition $n(\omega_{ij}) = l$, the terms on the right in 3.23 do not have $\phi_j(x_0)$ as their mean. Instead,

$$E[\hat{\phi}_j(\omega_{ijk}, x_0) \,|\, n(\omega_{ijk}) = l]$$

$$= g(x_0) + [\nu_j P^{(j)}\{n(\omega) = l \,|\, x_0\}]^{-1} \int \cdots \int_{\Sigma_{lj}} G_l(x_0, \gamma_l)\, d\gamma_l .$$

If the multiple integral on the right is symbolized by $I_{lj}(x_0)$, then

$$E\{[\hat{\phi}_j(\omega_{ij}, x_0) - \phi_j(x_0)]^2 | n(\omega_{ij}) = l\} \; E\left\{\left[\hat{\phi}_j(\omega_{ij}, x_0) - g(x_0) - \frac{I_{lj}(x_0)}{\nu_j \, P^{(j)}\{n(\omega) = l | x_0\}} + g(x_0) + \frac{I_{lj}(x_0)}{\nu_j \, P^{(j)}\{n(\omega) = l | x_0\}} - \phi_j(x_0)\right]^2 | n(\omega_{ij}) = l\right\} = \left[\phi_j(x_0) - g(x_0) - \frac{I_{lj}(x_0)}{\nu_j \, P^{(j)}\{n(\omega) = l | x_0\}}\right]^2 + r_j^{-1} E\left\{\left[\hat{\phi}_j(\omega, x_0) - g(x_0) - \frac{I_{lj}(x_0)}{\nu_j \, P^{(j)}\{n(\omega) = l | x_0\}}\right]^2 | n(\omega) = l\right\} ,$$

in which all expectations are relative to $[\Omega, P^{(j)}(\Lambda)]$. By using this result in Eq. 3.25 and recalling Eq. 2.25, it is seen that

$$\text{(3.26)} \quad \operatorname{var} \phi_R^*(x_0) = \sum_{j=1}^{S} \frac{\nu_j^2}{M_j} \sum_{l=1}^{\infty} P^{(j)}\{n(\omega) = l | x_0\} \left\{\left[\phi_j(x_0) - g(x_0) - \frac{I_{lj}(x_0)}{P\{n(\omega) = l | x_0\}}\right]^2 + r_j^{-1} E\left(\left[\hat{\phi}_j(\omega, x_0) - g(x_0) - \frac{I_{lj}(x_0)}{P\{n(\omega) = l | x_0\}}\right]^2 | n(\omega) = l\right)\right\} .$$

For $r_1 = r_2 = \cdot \quad = r_S = 1$, Eq. 3.26 constitutes a breakdown of Eq. 3.21. The effect of the averaging in 3.23 on the variance is easily seen.

It should be clear to the reader that the sampling procedure described above *requires* that the stochastic processes involved be simple enough to make practicable the order of sampling first, chain lengths, second, chain points. This is not necessarily required if $r_1 = r_2 = \quad = r_S = 1$.

Some insight may be gained from Eq. 3 26 into the problem of choosing a good stratification for the representative sampling. The terms involving the r_j^{-1} as factors may be made small by making the r_j large. This involves repeated sampling. Repetitive calculations can be made very rapidly on high-speed calculators. It appears then that in choosing the stratification, attention should be paid to making the terms

$$\nu_j \, P^{(j)}\{n(\omega) = l | x_0\} \left[\phi_j(x_0) - g(x_0) - \frac{I_{lj}(x_0)}{P\{n(\omega) = l | x_0\}}\right]^2$$

small.

Estimation of the variance 3.26 from an experimental sample appears to be complicated. The difficulty is in finding an estimate of the variance of 3.23 for any fixed index j. The r_j terms on the right in 3.23 all have the same random determination of chain length and so are correlated in pairs, each pair having the same correlation. Thus a method is needed for estimating the mean of several random variables in the presence of intraclass correlation. As far as the author knows, no such estimate is available in the statistical literature. It would be desirable to have an estimate of the variance 3.26 to use in connection with approximate normality of the estimate 3 24. The author conjectures that the estimate 3.24 should be more nearly normally distributed than any of the other estimates

of the series 1.3 that have been suggested.

Consider the definition of a stochastic process for the estimate 3.24. The notation there was used for ease in describing the sampling, but it is misleading for the present purpose. On recalling section 2.3.2, it is seen that the sampling space is actually the product space

$$\Omega_{r_1}^{(M_1)} \times \Omega_{r_2}^{(M_2)} \times \ldots \times \Omega_{r_S}^{(M_S)} .$$

On the component Ω_{r_j}, the probability measure $P_{r_j}^{(j)}$ is to be derived from $[\Omega, P^{(j)}(\Lambda)]$, as was P_M from $[\Omega, P(\Lambda)]$ in section 2.3.2, that is, in place of the functions 2.2 and 2.3 for $[\Omega, P(\Lambda)]$ used in section 2.3.2, the functions of the type 2.2 and 2.3 appropriate to $[\Omega, P^{(j)}(\Lambda)]$ in the construction 2.36 through 2.39 are used. This will generate the probability density functions $f_{N,r_j}(z_N | z_0, z_1, \ldots, z_{N-1})$ and the stopping probabilities $p_{N,r_j}(z_0\ z_1, \ldots, z_N)$ for $N \geq 1$. Theorem 1, revamped for Ω_{r_j}, then gives the process $[\Omega_{r_j}, P_{r_j}^{(j)}(\Lambda)]$. The measure on the product space

$$\prod_{j=1}^{S} \Omega_{r_j}^{(M_j)}$$

is then defined as a product measure by using M_j factors $P_{r_j}^{(j)}(\Lambda)$ for each $j = 1, 2, \ldots, S$.

It should be remarked that the probability functions of the type 2.2 and 2.3 that are needed in the construction indicated above have been described only for the special cases of examples *a* and *b* of section 2.3.1. The general process description is then dependent on the possibility of defining the needed functions. On the other hand, the sampling description for 3.23 and 3.24 is given in terms of the generally defined sampling functions 2.11*a* and 2.17*a*. The author's opinion is that a structure theorem for stochastic processes of the type needed for this work, based directly upon the sampling functions of the types 2.11 and 2.17, might be more convenient than the structure theorem by Doob upon which Theorem 1 is based.

CHAPTER 4. WEIGHTED AVERAGES OF $\phi(x)$: SOLUTION IN THE LARGE

4.1. ESTIMATION OF A WEIGHTED AVERAGE

The discussion in chapter 3 was devoted entirely to the problem of estimating the solution $\phi(x)$ of Eq. 1.1 at a single point $x = x_0$ or a single component of the vector ψ of Eq. 1.5. This chapter will present stochastic estimates of integrals of the form $\int_R w(x)\ \phi(x)\ dx$, where $w(x)$ is a specified function, and the estimation of the entire function $\phi(x)$ over R

Consider first the integral

$$(4.1) \qquad I = \int_R w(x)\ \phi(x)\ dx ,$$

where $w(x)$ is a known function and $\phi(x)$ is the solution of Eq. 1.1. A well-known stochastic technique for estimating I proceeds as follows (cf., Appendix B). Let $f_0(x)$ be an arbitrary probability density function defined and positive on R. For some integer r, draw r stochastically independent values $x^{(1)}, x^{(2)}, \ldots, x^{(r)}$ at random from the population whose distribution is $f_0(x)$. An unbiased estimate of I is given by

$$(4.2) \quad (\text{est } I) = \frac{1}{r} \sum_{i=1}^{r} \frac{w(x^{(i)})\ \phi(x^{(i)})}{f_0(x^{(i)})},$$

which involves a knowledge of the values $\phi(x^{(i)})$, $i = 1, 2, \ldots, r$, of the solution of Eq. 1.1 at the points $x^{(i)}$. Estimates of these values could be obtained by any one of the processes described in chapter 3, with the r values $x^{(i)}$ used successively as the fixed point x_0. Let the

estimate of I so obtained be denoted by $\hat{I}$, that is,

$$(4.3) \qquad \hat{I}_r = \frac{1}{r} \sum_{i=1}^{r} \frac{w(x^{(i)})\ \hat{\phi}(x^{(i)})}{f_0(x^{(i)})},$$

where the $x^{(i)}$ are randomly drawn from $f_0(x)$ and $\hat{\phi}(x^{(i)})$ denotes an estimate of $\phi(x)$ at the (fixed) point $x^{(i)}$. Thus

$$E\hat{I}_r = \frac{1}{r} \sum_{i=1}^{r} \int_R \frac{w(x^{(i)})}{f_0(x^{(i)})} E\{\hat{\phi}(x_0) \mid x_0 = x^{(i)}\}\ f_0(x^{(i)})\ dx^{(i)}$$

$$= \frac{1}{r} \sum_{i=1}^{r} \int_R w(x^{(i)})\ \phi(x^{(i)})\ dx^{(i)} = I\ ,$$

according to the unbiased character of each $\hat{\phi}(x^{(i)})$. The variance of $\hat{I}_r$ is given by the formula

$$(4.4) \qquad I^2 + r \operatorname{var} \hat{I}_r = E\left\{\frac{w(x)\ \hat{\phi}(x)}{f_0(x)}\right\}^2$$

$$= \int_R \left[\frac{w(x)}{f_0(x)}\right]^2 E\{[\hat{\phi}(x_0)]^2 \mid x_0 = x\}\ f_0(x)\ dx$$

$$= \int_R \frac{w^2(x)}{f_0(x)} \{\operatorname{var}[\hat{\phi}(x_0) \mid x_0 = x] + \phi^2(x)\}\ dx\ .$$

It is clear at once that, even though zero variance processes are used to obtain the $\hat{\phi}(x^{(i)})$, the variance of 4.3 will not be zero. Does a zero variance estimator of I exist?

The answer in the affirmative seems to have been given first by Goertzel.[7] His result and an alternative one will be given in the next section. First, the integral I of Eq. 4.1 will be put into a different form and an alternative estimate to 4.3 will be introduced.

Assuming uniform convergence of the Neumann series 1.3 over R, one has from Eqs. 4.1 and 1 3

$$I = \int_R w(x)\ g(x)\ dx + \sum_{N=1}^{\infty} \int_R w(x)\ I_N(x)\ dx\ .$$

By using 1.4 and the change of variables $x = v_N$, $u_1 = v_{N-1}$, ..., $u_{N-1} = v_1$, $u_N = y$ for $N \geq 1$, it is seen that

$$\int_R w(x)\ I_N(x)\ dx =$$

$$\lambda^N \int_R g(y)\ dy \int_R \ldots \int_R K(v_1, y)\ K(v_2, v_1)\ \ldots\ K(v_N, v_{N-1})\ w(v_N)\ dv_1\ \ldots\ dv_N\ .$$

Thus

$$(4.5) \quad I = \int_R g(y) \left\{ w(y) + \sum_{N=1}^{\infty} \lambda^N \int_R \ldots \int_R K(v_1, y)\ K(v_2, v_1)\ \ldots\ K(v_N, v_{N-1})\ w(v_N)\ dv_1\ \ldots\ dv_N \right\} dy\ .$$

Consider the integral equation

$$(4.6)\quad Q(y) = w(y) + \lambda \int_R K(x,y)\, Q(x)\, dx .$$

The series in the bracket in Eq. 4.5 is the Neumann series for Eq. 4.6. Thus I may be written as

$$(4.7)\qquad I = \int_R g(y)\, Q(y)\, dy .$$

Upon applying the method of Eq. 4.2 to Eq. 4.7, an estimate of I is obtained that is an alternate to estimate 4.3, it is,

$$(4.8)\qquad \tilde{I}_r = \frac{1}{r} \sum_{i=1}^{r} \frac{g(x^{(i)})\, \hat{Q}(x^{(i)})}{f_0(x^{(i)})} ,$$

where $\hat{Q}$ denotes a point estimate of the function Q.

This reformulation of the estimation procedure leads to two very important items (1) a generalization of a much used method for estimating $\phi(x)$ over a fixed net of points $x_1, x_2, \ldots, x_n$, (2) a stochastic method for estimation of the entire function $\phi(x)$ over R. These problems will be discussed in sections 4.3 and 4.4.

4.2. ZERO VARIANCE ESTIMATES OF I

Consider the variance formula 4.4. Suppose that the integral equation 1.1 satisfies the conditions (cf., section 3.3) for the existence of a zero variance estimate of $\phi(x_0)$ for each x_0 in R. Then $\phi(x)$ is nonnegative on R. If, in addition, $w(x)$ is nonnegative on R, the choice

$$f_0(x) \equiv \frac{w(x)\, \phi(x)}{I}$$

may be used. Suppose that the zero variance process for $\phi(x_0)$ is used in obtaining each of the $\hat{\phi}(x^{(i)})$ in 4.3. The right member of Eq. 4.4 easily reduces to I^2 from which var $\hat{I}_r = 0$ follows.

Clearly, the above construction may be used in connection with Eqs. 4.7 and 4.8 and a zero variance process for Eq. 4.6. In this case the additional requirement is that $g(x)$ be nonnegative and the choice $f_0(x) = g(x)\, Q(x)/I$ must be made. This latter procedure is the one introduced by Goertzel.[7]

4.3. A MUCH USED ESTIMATOR OF $\phi(x)$

The fundamental basis for all the estimation processes discussed so far has been the random variable 3.1. Many Monte Carlo experiments that have been actually performed in computing laboratories have not been based upon Eq. 3.1 but rather upon the physical analogue scheme to be described and generalized in this section.

Suppose that Eq. 1.1 arises in a particle diffusion problem. A particle moves about in a phase space R and collides with other particles in a random manner and is subjected to possible absorption on any collision. Let the rules of this process be the following (1) $g(x)$ is the probability density of the point at which the first collision occurs, (2) for each fixed y in R, $\lambda\, K(x,y)$ is the probability density of the position x of the collision immediately following a collision at y, (3) $\rho(x) = 1 - \lambda \int_R K(x,y)\, dy$ is the probability of absorption on collision at x. It is not difficult to see that the term $I_N(x)$ of the series 1.3 is the density function for the $(N+1)$ collision, that is, $I_N(x)\, dx$ is the probability that the particle gets as far as the $(N+1)$ collision and that the collision will occur in the interval $(x, x+dx)$. For any subset A of R, the integrals

$$J_1 = \int_A \phi(x)\, dx$$

and

$$J_2 = \int_A \rho(x)\, \phi(x)\, dx$$

represent, respectively, the probability of at least one collision in A and the probability that the particle is absorbed on some collision in A.

Let $\delta(x,A)$ denote the characteristic function of the set A and define the two functions $w_1(x,A) \equiv \delta(x,A)$ and $w_2(x,A) \equiv \rho(x)\, \delta(x,A)$. Then the integrals J_1 and J_2 take the forms

$$(4.9)\quad \begin{cases} J_1 = \int_R w_1(x, A)\, \phi(x)\, dx\,, \\ J_2 = \int_R w_2(x, A)\, \phi(x)\, dx\,, \end{cases}$$

that is, they are special cases of Eq. 4.1.

Define the integral equations

$$(4.10)\qquad Q_i(x,A) = w_i(x,A) + \lambda \int_R Q_i(y,A)\, K(y,x)\, dy\,, \quad i = 1,\ 2\,.$$

These are of the form of Eq. 4.6, and the integrals 4.9 may be written as

$$(4.11)\quad J_i = \int_R g(x)\, Q_i(x,A)\, dx\,, \qquad i = 1,\ 2\,.$$

Either of the integrals J_i can then be estimated by the method of Eq. 4.8. This method requires an unbiased estimate of $Q_i(x,A)$. The appropriate specialization of Eq. 3 1 could be used. As already mentioned, an alternative to Eq. 3.1 is of interest.

Consider the following procedure for estimating $Q_i(x_0, A)$ for an *arbitrary choice of* x_0 in R. In following the description of the particle process above, by some random device decide that the particle is absorbed at x_0 with probability $\rho(x_0)$ or not absorbed with probability $1 - \rho(x_0)$. If it is decided that the particle is not absorbed at x_0, by an appropriate device select a point $x = x_1$ in R by using the distribution $\lambda\, K(x, x_0)$. Again decide upon absorption or no absorption at x_1, with probabilities $\rho(x_1)$ and $1 - \rho(x_1)$, respectively. If nonabsorption at x_1 is decided, select $x = x_2$ at random by using the distribution $\lambda\, K(x, x_1)$ and so on until the particle is absorbed at some point x_n.

To estimate $Q_1(x_0, A)$, simply count those of the points $x_0, x_1, x_2, \ldots, x_n$ that fall in the set A. This count is the estimate of $Q_1(x_0, A)$. To estimate $Q_2(x_0, A)$, use the sum of all of the values of $\rho(x_i)$ for all of those collision points x_i that fall in A. In formulas,

$$(4.12)\qquad \text{est } Q_j(x_0, A) = \sum_{i=0}^{n} w_j(x_i, A)\,, \qquad j = 1,\ 2\,.$$

The estimate 4.12 of $Q_1(x_0, A)$, used in conjunction with 4.8 to estimate 4.11 for $i = 1$, provides a new method for estimating $\phi(\xi)$ at any ξ in R, as follows. Let A be a set containing ξ for which the measure $m(A) = \int_A dx$ is small and over which the variation of $\phi(x)$ is small. Then $\phi(\xi) \simeq J_1/m(A)$, according to the law of the mean for integrals. Consequently, an estimate of J_1 gives an estimate of $\phi(\xi)$. Clearly, this estimate of $\phi(\xi)$ is simply the total number of collisions that occur in A divided by the measure $m(A)$. Such an estimating procedure could be used to obtain simultaneous estimates of ϕ on a partition of R into small sets $A_1, A_2, \ldots, A_n$.

Before considering the statistical properties of the estimates 4.12, it will be convenient to generalize the method. Suppose that a stochastic process $[\Omega, P(\Lambda)]$ of the type given by Theorem 1 is at hand. It will be assumed that the process is conditioned for a fixed x_0 (cf., section 2.2). Consider the random variable

$$(4.13)\quad \hat{\hat{Q}}_j(x_0, A) \equiv w_j(x_0, A) + \sum_{l=0}^{n(\omega)} \frac{\lambda^l\, K(x_1, x_0)\, K(x_2, x_1) \ldots K(x_l, x_{l-1})\, w_j(x_l, A)}{f_l(\gamma_l \mid x_0) \prod_{l=0}^{l-1} q_l(\gamma_l^*)}\,, \quad j = 1,\ 2\,,$$

where the functions in the denominators are defined in section 2.2.* It will be shown that 4.13 generalizes Eq. 4.12 in the sense that, if the process $[\Omega, P(\Lambda)]$ is the fundamental particle process described above and conditioned for fixed x_0, 4.13 reduces to 4.12. Moreover, it will be shown that

*The summation $\sum_{l=1}^{n(\omega)}$ is to be omitted if $n(\omega) = 0$

the random variable 4.13 provides an unbiased estimate of $Q_j(x_0, A)$ relative to the process $[\Omega, P(\Lambda)]$.

To show that 4.13 generalizes 4.12 construct $[\Omega, P(\Lambda)]$ from the particle process, as follows. With reference to Definition 1 and the functions 2.2 and 2.3 of section 2.2, define $p_N(\gamma_N^*) \equiv \rho(x_N)$ for each $N = 0, 1, 2, \ldots$. Then $q_N(\gamma_N^*) \equiv 1 - \rho(x_N)$. Also, define

$$f_N(\gamma_N | \gamma_i^*) \equiv \prod_{l=i+1}^{N} \frac{\lambda\, K(x_l, x_{l-1})}{1 - \rho(x_{l-1})} .$$

It can be seen immediately that the denominators in 4.13 are

$$f_i(\gamma_i | x_0) \prod_{l=0}^{i-1} q_l(\gamma_l^*) = \prod_{l=1}^{i} \frac{\lambda\, K(x_l, x_{l-1})}{q_{l-1}(\gamma_{l-1}^*)} \prod_{l=0}^{i-1} q_l(\gamma_l^*) = \prod_{l=1}^{i} \lambda\, K(x_l, x_{l-1})$$

and 4.13 reduces to 4.12.

To see that 4.13 gives an unbiased estimate of $Q_j(x_0, A)$, $j = 1, 2$, relative to the process $[\Omega, P(\Lambda)]$, consider

$$(4.14) \quad E\{\hat{\hat{Q}}_j(x_0, A) | x_0\} = P\{n(\omega) = 0 | x_0\}\, w_j(x_0, A) + \sum_{N=1}^{\infty} P\{n(\omega) = N | x_0\}\, E\{\hat{\hat{Q}}_j(x_0, A) | n(\omega) = N, x_0\} , \quad j = 1, 2 .$$

The conditional expectation under the condition $n(\omega) = N$ occurring on the right in Eq. 4.14 is

$$\begin{aligned} & E\{\hat{\hat{Q}}_j(x_0, A) | n(\omega) = N, x_0\} \\ &= [P\{n(\omega) = N | x_0\}]^{-1} \int_R \cdots \int_R \hat{\hat{Q}}_j(x_0, A)\, f_N(\gamma_N | x_0)\, P_N(\gamma_N)\, d\gamma_N \\ &= w_j(x_0, A) + \sum_{i=1}^{N} \int_R \cdots \int_R w_j(x_i, A) \left[\prod_{l=1}^{i} \frac{\lambda\, K(x_l, x_{l-1})}{q_{l-1}(\gamma_{l-1}^*)} \right] \frac{f_N(\gamma_N | x_0)\, P_N(\gamma_N^*)}{f_i(\gamma_i | x_0)}\, d\gamma_N \\ &= w_j(x_0, A) + \sum_{i=1}^{N} \int_R \cdots \int_R w_j(x_i, A) \left[\prod_{l=1}^{i} \lambda\, K(x_l, x_{l-1}) \right] f_N(\gamma_N | \gamma_i^*) \\ &\qquad \cdot\, q_i(\gamma_i^*) \cdots q_{N-1}(\gamma_{N-1}^*)\, p_N(\gamma_N^*)\, d\gamma_N . \end{aligned}$$

By using this in Eq. 4.14 and permuting the double summation $\sum_{N=1}^{\infty} \sum_{i=1}^{N}$ to the form $\sum_{i=1}^{\infty} \sum_{N=i}^{\infty}$, the following is obtained

$$(4.15) \quad E\{\hat{\hat{Q}}_j(x_0, A) | x_0\} = w_j(x_0, A) + \sum_{i=1}^{\infty} \int_R \cdots \int_R w_j(x_i, A) \left[\prod_{l=1}^{i} \lambda\, K(x_l, x_{l-1}) \right] \left[\lim_{m \to \infty} V_{m,i}(\gamma_i^*) \right] d\gamma_i .$$

Since it was shown in the proof of Theorem 2 that $\lim_{n \to \infty} V_{n,1}(\gamma_1^*) = 1$, Eq. 4.15 is the Neumann series for Eq. 4.10.

The variance of the random variable 4.13 seems to be difficult to handle. Presumably, stratification and repeated sampling devices should be useful in reducing that variance just as they were in connection with 3.1. The estimator 4.13 has much intuitive appeal in that it reduces to the standard Monte Carlo procedure of counting collisions when the estimating process $[\Omega, P(\Lambda)]$ is an analogue of the physical particle diffusion process. The author hopes to be able to give more consideration to 4.13 at some later date.

4.4. ESTIMATION OF $\phi(x)$ IN THE LARGE

Throughout chapter 3 the discussion was concerned with the estimation of the value of the solution $\phi(x)$ of Eq. 1.1, at a single point $x = x_0$ in in the space R, or of a single component of the vector solution of Eq. 1 5. The reader may have recognized that in certain special cases the estimation process actually yields an estimate of the entire function $\phi(x)$ over R. To be explicit, if the distributions $f_N(\gamma_N | x_0)$ are independent of x_0 for each $N \geq 1$ and if the stopping probabilities $p_N(\gamma_N^*)$ are constants independent of γ_N^*, the random variable 3.1 will depend only upon x_0 as a parameter. That is, x_0 appears in the formula 3.1 only in the term $g(x_0)$ and the factor $K(x_0, x_1)$ of $G_{n(\omega)}(x_0, \gamma_{n(\omega)})$, the sampling procedure for obtaining $n(\omega)$ and $\gamma_{n(\omega)}$ is independent of the value of x_0. It follows at once that $\hat{\phi}(\omega, x_0)$ is an estimate of $\phi(x_0)$ for arbitrary x_0. More general procedures for estimating the entire function $\phi(x)$ will be given next.

In section 4.1 there is nothing in the theory to prevent the weight function $w(x)$ from depending on a parameter. Consider, in fact, the integral 4.1 in which $w(x) = \lambda K(u,x)$. Then Eq. 4.6 has a solution $Q(y,u)$, and the equation becomes

$$(4.16) \quad Q(y,u) = \lambda K(u,y) + \lambda \int_R K(x,y)\, Q(x,u)\, dx .$$

The Neumann series for Eq. 4.16 is

$$Q(y,u) = \lambda K(u,y) \sum_{N=1}^{\infty} \lambda^{N+1} \int_R \cdots \int_R K(t_N, t_{N-1}) \cdots K(t_1, y)\, K(u, t_N)\, dt_1 \cdots dt_N .$$

By using this in Eq. 4.7 and writing $I(u)$ instead of $\hat{I}$,

$$(4.17) \quad \begin{aligned} I(u) &= \sum_{k=1}^{\infty} \lambda^k \int_R \cdots \int_R K(u, x_1)\, K(x_1, x_2) \cdots K(x_{k-1}, x_k)\, g(x_k)\, dx_1 \cdots dx_k \\ &= \phi(u) - g(u) . \end{aligned}$$

In obtaining Eq. 4.17, the change of variables $x_1 = t_N$, $x_2 = t_{N-1}$, ..., $x_N = t_1$ was used and followed by inversion of the order of integration.

Any unbiased stochastic estimate of $I(u)$ will be an estimate of $\phi(u) - g(u)$, according to Eq. 4.17. If the estimating process can be so chosen that the experimental determination of the chains $x_0, x_1, x_2, \ldots, x_{n(\omega)}$ is independent of the value of u in R, the corresponding statistics for estimation of $I(u)$ will be valid for all u in R.

Suppose that a stochastic process $[\Omega, P(\Lambda)]$ conditioned for arbitrary fixed x_0 and with $p_0(x_0) = 0$ has been chosen by Theorem 1. The process is defined by the sequences of probability functions

$$f_N(\gamma_N | x_0), \quad N = 1, 2, 3, \ldots,$$

$$P_0(x_0) = 0, \quad P_N(x_0, \gamma_N), \quad N = 1, 2, 3, \ldots .$$

Let $f_0(x_0)$ be an arbitrary probability (density) function on R. For any fixed integer r, let r points $x_0^{(i)}$, $i = 1, 2, \ldots, r$, be drawn independently and at random from $f_0(x_0)$, and let the r chains ω_i, beginning at $x_0^{(i)}$, $i = 1, 2, \ldots, r$, be drawn from $[\Omega, P(\Lambda)]$. These chains are determined by the values $n(\omega_i)$ and the finite chains $\gamma^*_{n(\omega_i)}$ $x_0^{(i)}, x_1^{(i)}, \ldots, x_{n(\omega_i)}^{(i)}$, $i = 1, 2, \ldots, r$. In terms of these samples, Eq. 3.1 gives the r point estimates $\hat{\phi}(\omega_i, x_0^{(i)})$ of $\phi(x_0^{(i)})$ at the r points $x_0^{(i)}$

$$(4.18)\quad \hat{\phi}(\omega_i, x_0^{(i)}) = g(x_0^{(i)}) + \frac{G_{n(\omega_i)}(x_0^{(i)}, \gamma_{n(\omega_i)})}{f_{n(\omega_i)}(\gamma_{n(\omega_i)} | x_0^{(i)})\, P_{n(\omega_i)}(x_0, \gamma_{n(\omega_i)})}, \quad i = 1, 2, \ldots, r .$$

Also, on modifying Eq. 3.1 appropriately for estimating the solution $Q(y, u)$ of Eq. 4.16, if $[\Omega, P(\Lambda)]$ is independent of u, these *same* r chains give the r estimates

$$(4.19)\quad \hat{Q}(\omega_i, u) = \lambda\, K(u, x_0^{(i)}) + \frac{\lambda^{n(\omega_i)+1} K(x_1^{(i)}, x_0^{(i)}) \ldots K(x_{n(\omega_i)}^{(i)}, x_{n(\omega_i)-1}^{(i)})\, K(u, x_{n(\omega_i)}^{(i)})}{f_{n(\omega_i)}(\gamma_{n(\omega_i)} | x_0^{(i)})\, P_{n(\omega_i)}(x_0, \gamma_{n(\omega_i)})}, \quad i = 1, 2, \ldots, r ,$$

of $Q(x_0^{(i)}, u)$. By applying 4.19 in the manner indicated above, the quantity

$$(4.20)\qquad \hat{I}_r(u) \equiv \frac{1}{r} \sum_{i=0}^{r} \frac{g(x_0^{(i)})\, \hat{Q}(\omega_i, u)}{f_0(x_0^{(i)})}$$

is an unbiased estimate of $I(u) = \phi(u) - g(u)$ that is valid for all u in R.

The estimate 4.20 may be obtained from the r point estimates 4.18 by a kind of interpolation formula. To see this, define the function

$$(4.21)\qquad S(u, x_0, \gamma_N) \equiv \frac{K(x_1, x_0)\, K(x_2, x_1) \ldots K(x_N, x_{N-1})\, \lambda\, K(u, x_N)}{K(x_0, x_1)\, K(x_1, x_2) \ldots K(x_{N-1}, x_N)\, g(x_N)}$$

for $\gamma_N = x_1, x_2, \ldots, x_N$, $N \geq 1$. In terms of 4.18 and 4.21, the random variable 4.20 becomes

$$(4.22)\quad \hat{I}_r(u) = \frac{1}{r} \sum_{i=0}^{r} \frac{g(x_0^{(i)})}{f_0(x_0^{(i)})} \{\lambda\, K(u, x_0^{(i)}) + S(u, x_0^{(i)}, \gamma_{n(\omega_i)})\, [\hat{\phi}(\omega_i, x_0^{(i)}) - g(x_0^{(i)})]\} .$$

The relationship 4.22 between estimates 4.18 and 4.20 is of little more than academic interest, since 4.20 is the simpler formula to use. Moreover, the efficiency of the process $[\Omega, P(\Lambda)]$ for estimating $\phi(u) - g(u)$ is best judged in terms of 4.20 and the appropriate modification of the variance formula 4.4. The latter takes the form·

$$[\phi(u) - g(u)]^2 + r \operatorname{var} \hat{I}_r(u) = \int_R \frac{g^2(x)}{f_0(x)} \{\operatorname{var} [\hat{Q}(\omega, u) | x_0 = x] + Q^2(x, u)\} \, dx \, .$$

It should be evident that stratification and repeated sampling techniques could be used in the estimation of $Q(x,u)$ in 4.20. The structure of the zero variance process of the type described in section 3.3 is such that, when applied to the integral of Eq. 4.16 for $Q(x,u)$, the process will depend upon a particular choice of u. Thus, zero variance for $\hat{Q}(\omega, u)$ can be achieved only at one value of u.

The formula 4.22 reduces considerably if the kernel $K(x,y)$ of the integral of Eq. 1.1 is symmetrical and $g(x)$ is a probability density function on R. In this case Eq. 4.21 reduces to $\lambda \, K(u, x_N)/g(x_N)$, and the choice $f_0(x) \equiv g(x)$ may be used. Thus 4.22 reduces to

$$(4.23) \qquad \hat{I}_r(u) = \frac{\lambda}{r} \sum_{i=1}^{r} \left\{ K(u, x_0^{(i)}) + \frac{K(u, x_{n(\omega_i)}^{(i)})}{g(x_{n(\omega_i)}^{(i)})} [\hat{\phi}(\omega_i, x_0^{(i)}) - g(x_0^{(i)})] \right\} .$$

Various modifications of the above procedure should suggest themselves to the reader. The following question poses an unsolved problem. A close examination of 4.22 shows that $\hat{I}_r(u)$ is composed of two parts, one arising from the fact that the estimate is made of $\phi(u) - g(u)$ instead of $\phi(u)$, and one arising as a weighted sum of the separate point estimates 4.18. Is the weighting-interpolation operator 4.21 of the second part of $\hat{I}_r(u)$ optimal in some sense? A negative answer would raise the problem of determining such an optimal operator.

APPENDIX A. AN ILLUSTRATION

Consider the simple integral equation

$$\text{(A.1)}\quad \phi(x) = 1 + \frac{1}{2}\int_0^1 \exp\,[-(x+y)]\,\phi(y)\,dy\,, \qquad 0 \le x \le 1\,,$$

which was chosen for ease of computation. The terms 1.4 of the Neumann series 1.3 are readily integrated for A.1 and the series is summed

$$\phi(x) = 1 + e^{-x}\sum_{N=1}^{\infty} 2^{-N}\int_0^1 \cdots \int_0^1 \left[\prod_{i=1}^{N-1} e^{-2x_i}\right] e^{-x_N}\,dx_1 \cdots dx_N = 1 + c\,e^{-x}\,,$$

$$c = \frac{2(e^2 - e)}{3e^2 + 1} \cong 0.40322\,.$$

A simple process $[\Omega, P(\Lambda)]$, of the type defined by Theorem 1, conditioned for fixed x_0 and with $p_0(x_0) \equiv 0$, is constructed in the following manner.* Choose $p_0(x_0) \equiv 0$, $q_0(x_0) \equiv 1$, and $p_N(\gamma_N^*) \equiv 1/2 = q_N(\gamma_N^*)$ for every $N \ge 1$. Let $f_N(\gamma_N | x_0)$ be the joint distribution of N variables that are independently and uniformly distributed over (0,1) for each $N \ge 1$, that is, for each $N \ge 1$, the function $f_N(\gamma_N | \gamma_{N-1}^*)$ is the uniform distribution for x_N over (0,1) and therefore is independent of γ_{N-1}^*.

The distribution of $n(\omega)$, given x_0, is given by the special case Eq. 2.18 of Eq. 2.11, and is

$$\text{(A.2)}\quad Pr[n(\omega) = k\,|\,x_0] = \begin{cases} 0, & \text{if } k = 0\,, \\ 2^{-k}, & \text{if } k \ge 1\,. \end{cases}$$

Sampling from Eq. A.2 is simple. From the Neumann series, it can be seen that the terms for $N \le 8$ should give sufficient accuracy to five decimal places. Divide the integers from 1 through 128 into classes according to Eq. A.2, as follows

CLASS	INTEGERS IN CLASS
$k = 1$	1 through 64
2	65 through 96
3	97 through 112
4	113 through 120
5	121 through 124
6	125 through 126
7	127
8	128

Draw an integer in the range 1 to 128 from a table of random numbers. The class k' to which that integer is assigned in the table above gives $n(\omega) = k'$. The sampling for the relevant part of ω continues by drawing k' values $x_1, x_2, \ldots, x_{k'}$ independently from a uniform distribution on (0,1). The basic estimate 3.1 then gives

$$\text{(A.3)}\quad \hat{\phi}(\omega, x_0) = 1 + e^{-x_0 - x_{k'}} \prod_{i=1}^{k'} e^{-2x_i}\,.$$

for an arbitrary value of x_0. This is an example of the type mentioned at the beginning of section 4.4, in which 3.1 estimates the entire function $\phi(x)$.

The sampling described above was carried out four times, in a rough manner, with the following results.

$$\text{(A.4)}\quad \begin{cases} n(\omega_1) = 3\,, & \hat{\phi}(\omega_1, x_0) = 1 + 0.0362\,e^{-x_0}\,, \\ n(\omega_2) = 1\,, & \hat{\phi}(\omega_2, x_0) = 1 + 0.5543\,e^{-x_0}\,, \\ n(\omega_3) = 2\,, & \hat{\phi}(\omega_3, x_0) = 1 + 0.0750\,e^{-x_0}\,, \\ n(\omega_4) = 1\,, & \hat{\phi}(\omega_4, x_0) = 1 + 0.8025\,e^{-x_0}\,. \end{cases}$$

*See the remark at the end of section 2 2 and the special process b of section 2.2 1

The average estimate is $\bar{\hat{\phi}} = 1 + 0.3668\, e^{-x_0}$.

Consider the application of 4.22. In A.1, the kernel $K(x,y)$ is symmetrical and $g(x) \equiv 1$ may be used as $f_0(x)$, therefore 4.23 applies. The estimate of $\phi(u)$ is

$$\text{(A.5)} \qquad \text{est } \phi(u) = 1 + \frac{1}{8} e^{-u} \sum_{i=1}^{4} e^{-x_0^{(i)}} \left[1 + c_i\, e^{-x_{n(\omega_i)}^{(i)}} \right],$$

in which the c_i and $x_{n(\omega_i)}^{(i)}$, $i = 1, 2, 3, 4$, are the coefficients of e^{-x_0} and the final chain points in the four estimates A.4 and the $x_0^{(i)}$, $i = 1, 2, 3, 4$, are to be drawn independently from a uniform distribution on (0,1). The $x_{n(\omega_i)}^{(i)}$ of A.4 were 0.78, 0.59, 0.63, 0.22, respectively, and four numbers suitable for the $x_0^{(i)}$ are 0.44, 0.30, 0.88, and 0.41. The result is

$$\text{est } \phi(u) = 1 + 0.3932\, e^{-u}.$$

This is considerably better than the average of the four estimates A.4. Such an improvement could have been anticipated on theoretical grounds, in this example.

The variance of the fundamental estimate A.3 is easily computed from

$$\text{var } \hat{\phi}(\omega, x_0) = \text{var } [\hat{\phi}(\omega, x_0) - 1]$$

and

$$[\phi(x_0) - 1]^2 + \text{var } [\hat{\phi}(\omega, x_0) - 1]$$

$$= \sum_{N=1}^{\infty} 2^{-N} e^{-2x_0} \int_0^1 \cdots \int_0^1 \left[\prod_{i=1}^{N-1} e^{-4x_i} \right] e^{-2x_N}\, dx_1 \cdots dx_N,$$

with the result

$$\text{var } \hat{\phi}(\omega, x_0) = 0.0838\, e^{-2x_0}.$$

The standard deviation is

$$\sqrt{\text{var } \hat{\phi}(\omega, x_0)} = 0.2895\, e^{-x_0}.$$

Thus the standard deviation of the average of estimates A.4 is $0.1448\, e^{-x_0}$. The variance of the estimate A.5 may be found as one-fourth of the variance of a single one of the terms on the right, the calculation is similar to that above. The result is var (A.5) = $0.0050\, e^{-2u}$, or, for the standard deviation, $\cong 0.07\, e^{-u}$, which is less than one-half the standard deviation of the average of estimates A.4.

It is interesting to see the effect of stratification on the variance of the average of M estimates of the type A.4. Let the process $[\Omega, P(\Lambda)]$ used above be stratified as in example (*a*) of section 2.3.1. The integers N_1, N_2, ..., N_s will be chosen as $N_1 = 0$, $N_2 = 2$, $N_3 = 4$, $N_4 = 7$, $N_5 = \infty$. Thus, there are four strata given by all sequences ω for which $n(\omega) = 1$, $n(\omega) = 2, 3$, $n(\omega) = 4, 5, 6$, or $n(\omega) \geq 7$, for which $\nu_1 = 1/2$, $\nu_2 = 3/8$, $\nu_3 = 7/64$, and $\nu_4 = 1/64$. Assume that M is such that the sample sizes for the strata may be chosen by Eq. 3.22. The reduction in variance because of representative sampling over the strata will be given by

$$\text{(A.6)} \quad \frac{1}{M} \sum_{j=1}^{4} \nu_j [\phi_j(x_0) - \phi(x_0)]^2$$

where, from Eq. 3.13, the $\phi_j(x_0)$ are given by

$$\nu_1\, \phi_1(x_0) = \nu_1 + \frac{1}{2} e^{-x_0} \int_0^1 e^{-x_1}\, dx_1$$

$$\nu_j\, \phi_j(x_0) = \nu_j + \frac{1}{2} e^{-x_0} \sum_{k=N_j}^{N_{j+1}-1} \int_0^1 \cdots \int_0^1 e^{-x_1 - \cdots - x_k}\, dx_1 \cdots dx_k, \quad j = 2, 3, 4.$$

Formula A.6 can be easily computed to find the value $0.0558\ e^{-2x_0}/M$. Since the variance of the average of M estimates drawn from $[\Omega, P(\Lambda)]$ is $0.0838\ e^{-2x_0}/M$, the variance for the representative sample will be $0.0280\ e^{-2x_0}/M$. The effect is roughly equivalent to tripling the sample size M for the random sampling case.

APPENDIX B. THE ESTIMATION OF DEFINITE INTEGRALS OF FIXED MULTIPLICITY

It was suggested in section 3.4 that representative sampling might be advantageous in the estimation of individual integrals of the Neumann series 1.3. This matter will be considered in some detail here. It will be clear that the integral being estimated need not be a term of a Neumann series.

Consider Eq. 1.4

$$\text{(B.1)} \quad I_N(x_0) \equiv \int_R \cdots \int_R G_N(x_0, \gamma_N)\, d\gamma_N \ .$$

Let $g_N(\gamma_N | x_0)$ be a probability density function on R_N, it may be independent of x_0. In section 3.4, g_N would be given by Eq. 2.17 conditioned for x_0 and with $k = N$. If γ_N is a chain $x_1, x_2, \ldots, x_N$ whose distribution is $g_N(\gamma_N | x_0)$, then the random variable

$$\text{(B.2)} \quad \hat{I}_N(\gamma_N | x_0) \equiv \frac{G_N(x_0, \gamma_N)}{g_N(\gamma_N | x_0)}$$

is an unbiased estimate of $I_N(x_0)$. Likewise, the average

$$\text{(B.3)} \quad \hat{I}_{N,r}(x_0) \equiv \frac{1}{r} \sum_{i=1}^{r} \hat{I}_N(\gamma_N^{(i)} | x_0)$$

of r stochastically independent values of B.2 will be an unbiased estimate of B.1 relative to the probability density

$$\text{(B.4)} \quad \prod_{i=1}^{r} g_N(\gamma_N^{(i)} | x_0) \text{ on } R^{rN} \ .$$

The variance of B.3 will be computed, presently, in a form comparable to 3.14, that will be convenient for comparison with the variance of an estimate of B.1 obtained by representative sampling.

Partition the space R^N of γ_N into m nonoverlapping subsets $\Delta_{N,l}$, $l = 1, 2, \ldots, m$, for some integer m, and define the probabilities

$$\pi_{N,l} = \int \cdots \int_{\Delta_{N,l}} g_N(\gamma_N | x_0)\, d\gamma_N \ , \quad l = 1, 2, \ldots, m \ ,$$

and the conditional probability density functions

$$\text{(B.5)} \quad g_N(\gamma_N | x_0, \Delta_{N,l}) \equiv \begin{cases} \dfrac{g_N(\gamma_N | x_0)}{\pi_{N,l}} \ , & \text{if } \gamma_N \text{ in } \Delta_{N,l} \ , \\ 0, & \text{if } \gamma_N \text{ not in } \Delta_{N,l} \ , \end{cases}$$

$l = 1, 2, \ldots, m$. Also, define the quantities $I_{N,l}(x_0)$, by

$$\text{(B.6)} \quad \pi_{N,l}\, I_{N,l}(x_0) \equiv \int \cdots \int_{\Delta_{N,l}} G_N(x_0, \gamma_N)\, d\gamma_N \ , \quad l = 1, 2, \ldots, m \ .$$

The variance of the estimate B.2 may be expressed as

$$\text{var}\ [\hat{I}_N(\gamma_N|x_0)] = \sum_{l=1}^{m} \pi_{N,l}\ E\{[\hat{I}_N(\gamma_N|x_0) - I_N(x_0)]^2|\gamma_N \text{ in } \Delta_{N,l}\} .$$

As can be seen, $I_N(x_0)$ is not the mean of B.2 under the condition γ_N in $\Delta_{N,l}$. In fact,

$$\text{(B.7)} \qquad E\{\hat{I}_N(\gamma_N|x_0)|\gamma_N \text{ in } \Delta_{N,l}\} = I_{N,l}(x_0) .$$

Thus the expectations on the right in var $\hat{I}_N$ are broken down in the following manner

$$E\{[\hat{I}_N(\gamma_N|x_0) - I_{N,l}(x_0) + I_{N,l}(x_0) - I_N(x_0)]^2|\gamma_N \text{ in } \Delta_{N,l}\}$$

$$= E\{[\hat{I}_N(\gamma_N|x_0) - I_{N,l}(x_0)]^2|\gamma_N \text{ in } \Delta_{N,l}\} + [I_{N,l}(x_0) - I_N(x_0)]^2 .$$

Thus

$$\text{(B.7)} \quad \text{var}\ [\hat{I}_N(\gamma_N|x_0)] = \sum_{l=1}^{m} \pi_{N,l}\ E\{[\hat{I}_N(\gamma_N|x_0) - I_{N,l}(x_0)]^2|\gamma_N \text{ in } \Delta_{N,l}\}$$

$$+ \sum_{l=1}^{m} \pi_{N,l}[I_{N,l}(x_0) - I_N(x_0)]^2 .$$

It follows immediately that the variance of the average B.3 is obtained by dividing B.7 by r. In B.7,

$$\text{(B.8)} \quad E\{[\hat{I}_N(\gamma_N|x_0) - I_{N,l}(x_0)]^2|\gamma_N \text{ in } \Delta_{N,l}\}$$

$$= \int_R \cdots \int_R \frac{G_N^2(x_0, \gamma_N)}{g_N^2(\gamma_N|x_0)}\ g_N(\gamma_N|x_0, \Delta_{N,l})\ d\gamma_N - [I_{N,l}(x_0)]^2$$

$$= \frac{1}{\pi_{N,l}} \int \cdots \int_{\Delta_{N,l}} \frac{G_N^2(x_0, \gamma_N)}{g_N(\gamma_N|x_0)}\ d\gamma_N - [I_{N,l}(x_0)]^2 .$$

The integrals $I_{N,l}(x_0)$ defined in B.6 may be estimated by using the conditional distribution B.5 in much the same manner as Eq. B.1 was estimated. The random variable

$$\text{(B.9)} \qquad \hat{I}_{N,l}(\gamma_N|x_0) \equiv \frac{G_N(x_0, \gamma_N)}{\pi_{N,l} g_N(\gamma_N|x_0, \Delta_{N,l})} , \quad \gamma_N \text{ in } \Delta_{N,l}$$

is an unbiased estimate of $I_{N,l}(x_0)$ relative to B.5 for any $l = 1, 2, \ldots, m$. Note that B.9 is the same as B.2 except that γ_N is forced to fall in $\Delta_{N,l}$, that is, the $\pi_{N,l}$ cancels out.

Let the integer r used in B.3 be partitioned into a sum

$$r = \sum_{l=1}^{m} r_l$$

of integers r_l and let

$$\hat{I}_{N,l}(\gamma_N^{(j)}|x_0) , \quad j = 1, 2, \ldots, r_l ,$$

be stochastically independent values of B.9 for each $l = 1, 2, \ldots, m$. Clearly the variance above will be minimized by the choices

$$\mu_{lj} = \frac{\pi_{N,l}}{r_l} , \tag{B.12}$$

$j = 1, 2, \ldots, r_l$ for each $l = 1, 2, \ldots, m$. It is not hard to see that $\text{var}\,[\hat{I}_{N,l}(\gamma_N|x_0)]$ is given by Eq. B.8. Thus

$$\text{var}\,[I_{N,R}(x_0)] = \sum_{l=1}^{m} \frac{\pi_{N,l}^2}{r_l} E\{[\hat{I}_N(\gamma_N|x_0) - I_{N,l}(x_0)]^2|\gamma_N \text{ in } \Delta_{N,l}\} \tag{B.13}$$

Consider

$$I_{N,R}(x_0) \equiv \sum_{l=1}^{m} \sum_{j=1}^{r_l} \mu_{lj} \hat{I}_{N,l}(\gamma_N^{(j)}|x_0) , \tag{B.10}$$

where the μ_{lj} are constants. The random variable B.10 will be an unbiased estimate of B.1, relative to the distribution

$$\prod_{l=1}^{m} \prod_{j=1}^{r_l} g_N(\gamma_N^{(j)}|x_0, \Delta_{N,l}) ,$$

if the constants μ_{lj} are chosen to satisfy

$$\sum_{j=1}^{r} \mu_{lj} = \pi_{N,l} , \quad l = 1, 2, \ldots, m . \tag{B.11}$$

Since the terms on the right in B.10 are stochastically independent, the variance of $I_{N,R}(x_0)$ will be given by

$$\text{var}\,[I_{N,R}(x_0)] = \sum_{l=1}^{m} \sum_{j=1}^{r_l} \mu_{lj}^2 \,\text{var}\left[\hat{I}_{N,l}(\gamma_N^{(j)}|x_0)\right] = \sum_{l=1}^{m} \sum_{j=1}^{r_l} \mu_{lj}^2 \,\text{var}\left[\hat{I}_{N,l}(\gamma_N|x_0)\right] .$$

Now write

$$\frac{1}{r_l} \sum_{j=1}^{r_l} \mu_{lj} = \frac{\pi_{N,l}}{r_l}$$

if Eq. B.11 is satisfied. Then

$$\sum_{j=1}^{r_l} \mu_{lj}^2 = \sum_{j=1}^{r_l} \left(\mu_{lj} - \frac{\pi_{N,l}}{r_l}\right)^2 + \frac{\pi_{N,l}^2}{r_l} .$$

if Eq. B.12 is satisfied. Moreover, this is the minimal variance that can be achieved by this method.

In the special case in which the products $r\pi_{N,l}$, $l = 1, 2, \ldots, m$, are integers, the choices

$$r_l = r\pi_{N,l} , \quad l = 1, 2, \ldots, m , \tag{B.14}$$

make B.13 comparable to B.7 divided by r (that is, with the variance of B.3). Clearly, in this case

$$\text{var}\,[I_{N,R}(x_0)] < \text{var}\,[\hat{I}_{N,r}(x_0)]$$

by the amount

$$\frac{1}{r} \sum_{l=1}^{m} \pi_{N,l}[I_{N,l}(x_0) - I_N(x_0)]^2 . \tag{B.15}$$

If Eqs. B.12 and B.14 are satisfied, the estimates B.3 and B.10 for B.1 will be formally similar and B.10 will have the smaller variance.

That is, it is advantageous in forming B.3 to force r_l of the vectors γ_N to fall in $\Delta_{N,l}$, $l = 1, 2, \ldots, m$, by using the conditional sampling distributions B.5 for the sampling.

An important special case arises when the integrand $G_N(x_0, \gamma_N)$ of B.1 is nonnegative over the region of integration R^N. In this case, by assuming that $I_N(x_0)$ is known, for

the moment, a possible choice for the sampling distribution $g_N(\gamma_N|x_0)$ used above would be

$$\text{(B.16)} \qquad g_N(\gamma_N|x_0) \equiv \frac{G_N(x_0, \gamma_N)}{I_N(x_0)} .$$

The choice B.16 makes B.2 and all estimates derived from it exact, that is,

$$\text{var } [\hat{I}_N(\gamma_N|x_0)] = 0 .$$

It can then be seen that the individual terms of B.15 would all be zero (Eq. B.14 is assumed). The converse appears difficult, but it may be conjectured that under appropriate hypotheses on the integrand $G_N(x_0, \gamma_N)$ the following might be true. If for every partition $\Delta_{N,l}$, $l = 1, 2, \ldots, m$, for which there is an integer r such that Eq. B.14 holds, the expression B.15 vanishes and then the variance of B.2 is zero. Under conditions such that this conjecture is true, it should be possible to construct a practical approximation to the zero variance estimate indicated here. The following procedure seems clear for such a construction.

Let R be a finite interval (a,b) and suppose that $G_N(x_0, \gamma_N)$ is nonnegative for γ_N or R_N. Partition R into q subintervals for some integer q. These serve to partition the product set R^N into $m = q^N$ oriented N-dimensional intervals $\Delta_{N,l}$, $l = 1, 2, \ldots, q^N$. Choose a point z_l in each of the $\Delta_{N,l}$, and define the step function

$$\text{(B.17)} \qquad g_N(\gamma_N, x_0) \equiv \frac{H(x_0, \gamma_N)}{\int_R \ldots \int_R H(x_0, \gamma_N') \, d\gamma_N'}$$

be made. It is evident that the q^N conditional distributions B.5 defined by using B.17 will be uniform over their respective intervals $\Delta_{N,l}$, that is,

$$\text{(B.18)} \qquad g_N(\gamma_N|x_0, \Delta_{N,l}) = \begin{cases} \dfrac{1}{\pi_{N,l}}, & \text{if } \gamma_N \text{ in } \Delta_{N,l} , \\ 0, & \text{otherwise} . \end{cases}$$

It follows that, under the condition γ_N in $\Delta_{N,l}$, the components $x_1, x_2, \ldots, x_N$ will be uniformly and independently distributed over their respective subintervals of (a,b). Thus, sampling for γ_N in $\Delta_{N,l}$ by using B.18 will be very simple.

If the partitioning is done as described above and the integer r is such that the products B.14 are integers, the estimate B.10, subject to the choice B.12, reduces to

$$\text{(B.19)} \qquad I^*_{N,R}(x_0) = \frac{1}{r} \sum_{l=1}^{q^N} \sum_{j=1}^{r_l} G_N(x_0, \gamma_N^{(j)}) ,$$

and the variance B.13 becomes

$$\text{(B.20)} \qquad \text{var } [I^*_{N,R}(x_0)] = \frac{1}{r} \sum_{l=1}^{q^N} \pi_{N,l} \int \ldots \int_{\Delta_{N,l}} [G_N(x_0, \gamma_N) - I_{N,l}(x_0)]^2 \frac{d\gamma_N}{\pi_{N,l}}$$

$$= \frac{1}{r} \sum_{l=1}^{q^N} \left\{ \int \ldots \int_{\Delta_{N,l}} G_N^2(x_0, \gamma_N) \, d\gamma_N - \pi_{N,l} [I_{N,l}(x_0)]^2 \right\} .$$

$$H(x_0, \gamma_N) \equiv G_N(x_0, z_l) , \quad \text{if } \gamma_N \text{ in } \Delta_{N,l} .$$

Let the choice

In many practical examples, Eq. B.20 might be used in designing the estimation process. Let

$$U_l = LUB \; [G_N^2(x_0, \gamma_N)]$$

over $\Delta_{N,l}$ and

$$L_l = GLB\ [G_N(x_0, \gamma_N)]$$

over $\Delta_{N,l}$, $l = 1, 2, \ldots, q^N$. Then the variance B.20 is dominated by

$$\frac{1}{r} \sum_{l=1}^{q^N} \pi_{N,l} [U_l - L_l^2] \ . \tag{B.21}$$

Obviously the partitioning should be chosen to make B.21 small.

REFERENCES

1. J. H. Curtiss, "Probability Methods and Sampling Techniques," *Quarterly Report 1661 for January Through March 1952,* NBS, p. 5-9.

2. R. E. Cutkosky, "A Monte Carlo Method for Solving a Class of Integral Equations," *J. Research Natl. Bur. Standards* **47**, 113-115 (1951).

3. G. E. Forsythe and R. A. Leibler, "Matrix Inversion by a Monte Carlo Method," *Math. Tables and Other Aids to Computation* **4**, 127-129 (1950).

4. H. Kahn, "Random Sampling (Monte Carlo) Techniques in Neutron Attenuation Problems," *Nucleonics* **6**, No. 5, 27-37 (1950), **6**, No. 6, 60-65 (1950).

5. J. L. Doob, "Stochastic Processes with an Integral-Valued Parameter," *Trans. Amer. Math. Soc.* **44**, 87-150 (1938).

6. A. N. Kolmogorov, *Foundations of the Theory of Probability,* tr. by N. Morrison, Chelsea, New York, 1950.

7. G. Goertzel, *Quota Sampling and Importance Functions in Stochastic Solution of Particle Problems,* NDA memo, June 21, 1949.

8. W. Wasow, "On the Duration of Random Walks," *Annals Math. Stat.* **22**, 199-216 (1951).