

Some aspects of stochastically approximative analysis

Kitagawa, Toshio
Kyushu University

<https://doi.org/10.5109/12971>

出版情報 : 統計数理研究. 6 (3/4), pp.109-129, 1956-03. Research Association of Statistical Sciences

バージョン :

権利関係 :



SOME ASPECTS OF STOCHASTICALLY APPROXIMATIVE ANALYSIS

By

Tosio KITAGAWA

(Received March 20, 1956)

Introduction.

§ 1. **The objects of Introduction.** The main objects of this Introduction are to make clear as far as possible some of the methodological aspects of what I am now giving a tentative name of stochastically approximative analysis or more simply stochastic analysis. Let us illustrate first the backgrounds and needs for and secondly the methodological characteristic aspects of what we shall call stochastically approximative analysis.

§ 2. Backgrounds and needs for the stochastically approximative analysis.

[1] *Numerical analysis.* In 1951 the author of the present paper [3] introduced the notion of random integration by which to make comparisons of efficiencies among various mechanical quadratures such as we owe to Euler-Maclaurin, Tschebycheff, Gauss and so on. It is our standpoint that their relative efficiencies should be referred to certain families of functions to which their respective measure of probability should be defined in view of the corresponding situation of applications. It is hardly necessary to remark that the advent of high-speed sequenced computing machinery is not only revolutionizing the art of computation, but also our theoretical formulations by which some of theoretical aspects of numerical computations to be described. In fact the needs for stochastic formulations to numerical computations seem to us to become the more important, the more speedy machinery becomes available to us which can yield us numerical computation in a scale of mass productions.

[2] *Analytical studies in sample surveys and in designed experiments.* In some kinds of sample surveys and specially in a sequence of sample surveys an accumulation of our information will make it possible and our purposes of surveys will make it necessary to enter into discussions of more detailed aspects of our populations sampled. The three papers of the author [13], [14] and [15] were devoted to such problems as belonging to this category of the problems. For instance, mapping problems and interpenetrating procedures discussed in these papers will suggest us some correspondences to classical problems treated in the realm of calculus such as those of finding out the extremal values of the response surfaces. The main

differences lie in the fact that we should treat these problems not from deterministic formulations as in the classical calculus but from the stochastically approximative analysis. Consequently some preparations should be performed for fundamental formulations before we shall be able to enter into any adequate discussion manifesting the characteristic aspects of such analytical studies in these fields of statistical activities.

[3] *Statistical controls.* In 1952 the author wrote a paper on successive processes of statistical controls [8] by which to formulate some cybernetical theory of statistical controls. One of our emphasis is to suggest the possibilities and needs for giving a risk function formulation which should include not only decisions on statistical inferences but also controls on our objective worlds themselves. In our formulations controls mean transformations of the value of our population parameter to the other value, and our process of controlling scheme will sometimes yield us some functional equations concerning these transformations.

§ 3, Methodological characteristic aspects of stochastically approximative analysis. The main idea in giving a frame for developing our stochastically approximative analysis is to prepare with ourselves the standpoints with regards to the three aspects of logics according to which to coordinate our detailed investigations, i.e., those of objectivity, practice and subjectivity. In our schemes each of these three aspects has its subordinate principles on which to be based; namely, (i) the objectivity on the principles of factorization and of probabilization, (ii) the practices on the principles of efficiency and of successiveness, and (iii) the subjectivity on the principles of valuations and of strategy. Let us illustrate our viewpoints in what follows.

It is to be noted that the frame of logical aspects for developing our stochastically approximative analysis is subject to certain specializations of our general frame applicable in formulating our unified viewpoint including not only statistical inferences, statistical decision functions, but also stochastic analysis, statistical controls and cybernetics as its subordinate fields.

However we shall be here content with rather specialized frame suited for our stochastic analysis, not entering into the general frame as a whole.

In the first place, the two principles belonging to the objectivity aspect concern themselves with a formulation by which to describe our objective worlds in our concerns. As in almost all schemes in which any stochastic approaches may be found to be both necessary and adequate, it is almost evident that these two principles are indispensable in formulating our scheme of stochastic analysis. It is to be mentioned here that broader concepts of stochastic schemes than those which have been actually used in the statistical inference theories so far developed in most of current literatures should be prepared in order to give us mathematical formulations which may be re-

cognized to be better suited for approaching our problems in this field of stochastically approximative analysis. For instance a new formulation of analysis variance scheme in reference to the function space was given by the author [5] in 1951 in which the mathematical model appealing to the Wiener space of continuous functions was adopted as in CAMERON-MARTIN [1]. In the same idea the classical problems allied to the procedure how to define empirical functions were discussed in two papers of the author [11] and [13] in order to give better descriptions of error terms originating from various kinds of samplings and of observations.

Secondly let us turn to the aspects of practices. It can be readily observed that a notion of efficiency in some sense will be one of dominating principles ruling our procedures of stochastically approximative analysis. This is in the first place because of the real situations to be met in our practices in which our resources to be used for our investigations are almost always limited within certain limits, and secondly because of our allowable tolerances within which it suffices us to attain our aims. This is indeed the essential reason why we call our analysis as approximative in its adequate sense. Now the second subordinate principle, which we have called the principle of successiveness, also reflects our real situations of practices where multistage approaches may be applicable. In its special formulation some aspects of feedback principle will be found to be useful as in sequential or successive processes of statistical inferences and controls.

Lastly let us discuss with the logical aspects of subjectivity in our stochastically approximative analysis. We are placed in the circumstance where we have to start with an incomplete information as to our objective worlds which will be described by two principles of objectivity. An accumulation of our knowledges having been provided to us by each step of our experiments will gradually manifest the situations of our objective world and therefore will become more probable to have better recognition about our objective world at the expenses of costs of experimentations. In views of these considerations as well as those enunciated with regards to the principles of practices it seems to us quite promising to appeal to some decision function formulation as our frame work in developing stochastically approximative analysis.

§ 4. Decision theoretical approaches and cybernetical formulations of successive process of inferences and controls. It is to be noted here that, unless our procedures of solving various problems in our stochastically approximative analysis are categorized in the objective way, it would be impossible for us to establish any scientific approach. For instance it may be very practical as well as fairly efficient in some circumstances to fit a curve by free-hand drawing to observed data. However any decision theoretical approach may be impossible for us to develop, because our pro-

cedures of drawing free-hand curves will not be defined objectively in its strict sense. Thus the possibility of decision theoretical approaches means certain restrictions upon our research attitudes how to evaluate the scopes of attainments and how to interpret our observations. It is to be compared with purely scientific approaches where it should be completely free for research workers how to interpret the data and how to make questions in further investigations without any considerations upon our resources for investigations.

The distinction between decision theoretical approaches and scientific experimentations has been clearly described by a recent paper of R. A. FISHER [1]. It is to be noted however that our stochastically approximative analysis can be better formulated at least in its starting simpler schemes through the frame of decision theoretical approaches. This comes from two reasons. In the first place we want to have gradually progressive procedure of formulating our more extensive problems, starting therefore with the simpler and more clear ones. Secondly even some of the quite common problems such as those which will be formulated in the Parts I, II and III of this paper have not been well formulated in reference to our frame of stochastically approximative analysis in the sense enunciated in the previous paragraph.

Now the role of decision function approaches in our stochastically approximative analysis can be observed from the fact that there are three fundamental attitudes in adopting our strategies in dealing with our problems, namely

(i) the strategy appealing to our apriori informations and/or strategic variables possibly after obtaining some functional relationships among inter-related variables

(ii) the strategy appealing to randomization principle in dealing with unknown situations and unknown effects of various uncontrolled factors

(iii) the strategy appealing to game theoretical approach in formulating unknown situations and overcoming unknown effects.

Indeed the first strategy belongs to current mathematical analysis, as we can observe specially in some of functional analysis. The second strategy covers the standpoint of stochastic approximations developed by ROBBINS and MONRO [1], WOLFOWITZ [1] and others, as well as that of random integration developed by the auther [3]. On the other hand the principal aspect of the uses of the third strategy, so far as stochastically approximative analysis is concerned with, has not been well pointed out by any other author.

§ 5. Summary of this paper. The first purpose of this paper is to enunciate our general viewpoints how to formulate our stochastically approximative analysis, as we have given in this Introduction. The second

purpose of this paper is to give each formulation to each of some fundamental problems associated with numerical analysis from the view points of the decision function approaches to stochastically approximative analysis. In Part I we give two theorems concerning linear estimation formula on the basis of the non-sequential minimax procedures, which may be recognized to be preparatory to stochastically approximative analysis of more advanced and more elaborate types.

In Part II we are concerned with the discussions about the uses of difference-tables in interpolatory analysis from the standpoints of stochastically approximative analysis, appealing to the frame of the WIENER space and also to successive process of statistical inferences. In Part III our emphasis lies also in the formulation of our current procedures in developing our concurrence-function in the series of given orthogonal functions and in stopping the calculation of the expansion until some adequate terms. We shall point out here as well as elsewhere the situation that some of the current procedures of numerical analysis can be formulated from the standpoints of successive process of statistical inferences after giving at least substantially their respective suitable formulation concerning the three fundamental logical schemes. Some of Part II and most of Part III are merely giving certain formulations to our problem, and it is open to further development to solve various detailed problems suggested by our formulation.

Part I. Preparatory considerations upon risk function approaches to linear estimations.

§ 1. **Introductory.** Several types of stochastically approximative integrations have been introduced by our previous paper [3] in the name of random integrations. This will naturally suggest the possibility and the needs for introducing stochastically approximative differentiations. However it will be readily observed that some adequate formulations of stochastic approximations should have been introduced before we shall be able to discuss them. Moreover it seems to the author that some more fundamental reconsideration should have been performed before we shall be able to obtain some adequate formulation for stochastic approximations in general. This means that, although stochastic approximations should be concerned with limiting process of usually infinite sequence, some preparatory formulations of biased estimations concerning finite sequence should be developed in some scheme appealing to some sorts of inequalities giving restrictions to our realm of possible estimations.

We want to enunciate our idea very briefly rather confining ourselves within the simplest linear estimations. Although the two Theorems can be enunciated in a fairly general way, the uses of these two Theorems in discussing stochastic approximation can be observed from what follows.

Let us define a sequence of stochastic variables $Y_i = f(T_i)$ for $i = 1, 2, 3, \dots$, whose realization will be $y_i = f(t_i)$ for our partial realization $t = (t_1, t_2, \dots, t_n)$ of $T = (T_1, T_2, \dots, T_n, \dots)$.

In order to discuss the stochastic behaviours of the sequence of $\{f(T_i)\}$, we have to prepare ourselves with two fundamental notions. The first thing to do is how to define the meanings of convergence to the limit value $f(t_0)$. There are various kinds of convergences such as (i) convergence in law, (ii) mean convergence (iii) convergence in probability and (iv) almost everywhere convergence. The second thing to be considered is how to choose our estimator using our observations $\{f(t_i)\}$ in order to approximate our limit $f(t_0) \equiv m_0$ in its finite step. It is not altogether apparent whether or not there is a linear summation formula defined by the sequence

$$(1.01) \quad F_{c,n} \equiv \sum_{i=1}^n c_{n,i} f(t_i),$$

which is a better estimator of $f(t_0)$ in some sense of efficiency with reference to some notion of convergence than the sequence $\{f(t_i)\}$ ($i = 1, 2, \dots$). The situation can be recognized by the following fundamental relations to the effect that under the assumptions of existences of means and variances of the stochastic variables Y_i we have

$$(1.02) \quad E \left\{ \sum_{i=1}^n c_{n,i} Y_i \right\} = \sum_{i=1}^n c_{n,i} E \{ Y_i \} \equiv \sum_{i=1}^n c_{n,i} m_i$$

$$(1.03) \quad \sigma^2 \left\{ \sum_{i=1}^n c_{n,i} Y_i \right\} = \sum_{i=1}^n c_{n,i}^2 \sigma^2 \{ Y_i \} \equiv \sum_{i=1}^n c_{n,i}^2 \sigma_i^2.$$

Hence there may be some possibility of obtaining the means and the variances such that

$$(1.04) \quad \left| \sum_{i=1}^n c_{n,i} m_i - m_0 \right| \leq \left| m_n - m_0 \right|$$

$$(1.05) \quad \sum_{i=1}^n c_{n,i}^2 \sigma_i^2 \leq \sigma_n^2$$

by a suitable choice of the sequence $\{c_{n,i}\}$. However, since we are not acquainted with the values of these $\{m_i\}$ and $\{\sigma_i^2\}$, we are placed to be in the situation of game player whose strategy is to choose $\{c_{n,i}\}$ under the incomplete informations about the sets of values $\{m_i\}$ and $\{\sigma_i^2\}$.

§ 2. Minimax approaches.

[1] Let us assume to adopt the mean square error in order to evaluate the degree of approximation to an unknown value m_0 . Here let us confine ourselves with non-sequential approach, and let us denote $c_{n,i} = c_i$, $m_i - m_0 = A_i$, $\sigma_i^2 = B_i^2$ ($i = 1, 2, \dots, n$) for a fixed number n . Then the risk will be denoted by

$$(2.01) \quad R((A, B); c) \equiv \left(\sum_{i=1}^n c_i A_i \right)^2 + \sum_{i=1}^n c_i^2 B_i^2,$$

where $A = (A_1, A_2, \dots, A_n)$, $B = (B_1, B_2, \dots, B_n)$, $c = (c_1, c_2, \dots, c_n)$ and (A, B) denote vectorial notations.

Let us assume certain domain \mathcal{Q} of (A, B) in the $2n$ -dimensional euclidian space, and the range C of c in the n -dimensional space satisfying the condition $\sum c_i = 1$. We shall be concerned with the fact that the relations hold true that

$$(2.02) \quad \text{Max}_{\mathcal{Q}} \text{Min}_c R((A, B); c) = \text{Min}_c \text{Max}_{\mathcal{Q}} R((A, B); c)$$

for suitable choices of the domain \mathcal{Q} . The relation (2.02) shows us the existence of minimax solution for their respective domain \mathcal{Q} . For the sake of simplicity let us introduce

$$(2.03) \quad \Theta \equiv \sum_{i=1}^n \frac{A_i^2}{B_i^2}, \quad \Phi \equiv \sum_{i=1}^n \frac{A_i}{B_i^2}, \quad \Psi \equiv \sum_{i=1}^n \frac{1}{B_i^2}.$$

The direct calculations yield us the following two Lemmas.

Lemma 1.1. *For each assigned vector (A, B) , we have*

$$(2.04) \quad \text{Min}_c R((A, B); c) = \frac{1 + \Theta}{\Psi(1 + \Theta) - \Phi^2}$$

and this minimum is attained by $c^0 = (c_1^0, c_2^0, \dots, c_n^0)$ defined by

$$(2.05) \quad c_i^0 = \frac{1}{B_i^2} \cdot \frac{(1 + \Theta) - A_i \Phi}{\Psi(1 + \Theta) - \Phi^2}, \quad (i = 1, 2, \dots, n).$$

Lemma 1.2. *We have*

$$(2.06) \quad R((A, B); c) \geq \sum_{i=1}^n c_i^2 B_i^2 (1 + \Theta).$$

Theorem 1.1. *Let \mathcal{Q}_1 be the domain defined in the $2n$ -dimensional euclidian space by the relation that*

$$(2.07) \quad \Theta = \Theta_0, \quad B_i^2 \leq B_0^2 \varphi(i)^{-1} \quad (i = 1, 2, \dots, n)$$

for an assigned set of values Θ_0 and B_0^2 , where $\varphi(i)$ is a given positive valued function of i .

Then the minimax solution for the non-sequential stochastic approximation within the class of all the linear combinations of $\{f(t_i)\}$, i.e. $y_n = \sum_{i=1}^n c_i f(t_i)$ with $\sum c_i = 1$, with respect to the risk function (2.01) and \mathcal{Q}_1 , is given by

$$(2.08) \quad c_i = \varphi(i) \left\{ \sum_{i=1}^n \varphi(i) \right\}^{-1}, \quad (i = 1, 2, \dots, n).$$

Corollary 1.1. *Specially when $\varphi(i) \equiv 1$ for $i = 1, 2, \dots, n$, then we have*

$$(2.09) \quad c_1 = c_2 = \dots = c_n = n^{-1}.$$

Proof of Theorem 1.1.: The Schwarz inequality gives us $\psi\theta \geq \phi^2$, where the equality holds when and only when $A_i^2 B_i^2 B_i^{-2} = A_i^2$ are equal to a constant independent of i . Consequently we can observe

$$(2.10) \quad \begin{aligned} \text{Max}_{\Omega_1} \text{Min}_c R((A, B); c) &= \text{Max}_{\Omega_1} \frac{1 + \theta}{\psi(1 + \theta) - \phi^2} \\ &= \frac{1 + \theta_0}{\text{Min}_{\Omega_1} \psi} = \frac{B_0^2(1 + \theta_0)}{\sum_{i=1}^n \varphi(i)}, \end{aligned}$$

which is attained by the point (A, B) such that

$$(2.111) \quad B_i^2 = B_0^2 \varphi(i)^{-1} \quad (i = 1, 2, \dots, n)$$

$$(2.112) \quad A_1^2 = A_2^2 = \dots = A_n^2 = \theta_0 B_0^2 \left[\sum_{i=1}^n \varphi(i) \right]^{-1},$$

for which the corresponding values of the minimizing set of $\{c_i^0\}$ are given by (2.08) as immediately seen from (2.05).

On the other hand Lemma 1.2 gives us that

$$(2.12) \quad \text{Max}_{\Omega_1} R((A, B); c) = (1 + \theta_0) B_0^2 \sum_{i=1}^n c_i^2 \varphi(i)^{-1}$$

and hence that

$$(2.13) \quad \text{Min}_c \text{Max}_{\Omega_1} R((A, B); c) = \frac{(1 + \theta_0) B_0^2}{\sum_{i=1}^n \varphi(i)},$$

which is attained just by the same set of the values $\{A_i\}$, $\{B_i\}$ and $\{c_i^0\}$ as those defined by (2.111) \sim (2.112).

[2] Another type of minimax approaches will be given by introducing a mean risk function under the circumstances that there is a class of a priori distribution functions $G(A)$ such that

$$(2.14) \quad \int A_i^2 dG(A) = D_i^2, \quad \int A_i A_j dG(A) = 0 \quad (i \neq j)$$

for $i, j = 1, 2, \dots, n$. We have then

$$(2.15) \quad \int R((A, B); c) dG(A) = \sum_{i=1}^n c_i^2 (D_i^2 + B_i^2).$$

It will be our common situations that we do not have any exact information about the values of $\{D_i^2 + B_i^2\}$ ($i = 1, 2, \dots, n$) until that we can expect that $B_i^2 + D_i^2 = O(\varphi(i)^{-1})$, where $\varphi(i)^{-1}$ tends to zero as i tends to infinity. In view of our analytical tool to be used in such an approach, let us consider the domain $\mathcal{Q}_\rho(H_0)$ in the space $(B_1^2, B_2^2, \dots, B_n^2, D_1^2, D_2^2, \dots, D_n^2)$ defined such as

$$(2.16) \quad B_i^2 + D_i^2 = H_i \varphi(i)^{-1} \quad (i = 1, 2, \dots, n),$$

where

$$(2.17) \quad \sum_{i=1}^n H_i^\rho = H_0,$$

with assigned positive value of ρ and H_0 . Let us denote the weighted risk (2.15) by $K_1(H, c)$.

Theorem 1.2. *The minimax solution of the stochastically approximative limit within the class of all linear combinations of $\{f(t_i)\}$ with respect to the risk function (2.15) and $\Omega_\rho(H_0)$ is given by*

$$(2.18) \quad c_i^0 = [\varphi(i)]^{\rho/(1+\rho)} / \sum_{j=1}^n [\varphi(j)]^{\rho/(1+\rho)} \quad (i = 1, 2, \dots, n).$$

Proof: The direct calculation of the value gives us

$$(2.19) \quad \max_{\Omega_\rho(H_0)} \min_c K_1(H; c) = \frac{1}{\min_{\Omega_\rho(H_0)} \left\{ \sum_{i=1}^n \frac{\varphi(i)}{H_i} \right\}} = \frac{1}{H_0^{1/\rho}} \left(\sum_{i=1}^n [\varphi(i)]^{\rho/(1+\rho)} \right)^{(1+\rho)/\rho},$$

which is attained by the set of the values of $\{H_i\}$ and $\{c_i^0\}$ defined as follows

$$(2.201) \quad H_i = H_0^{1/\rho} [\varphi(i)]^{1/(1+\rho)} / \left(\sum_{i=1}^n [\varphi(i)]^{\rho/(1+\rho)} \right)^{1/\rho}$$

$$(2.202) \quad c_i^0 = [\varphi(i)]^{\rho/(1+\rho)} / \sum_{i=1}^n [\varphi(i)]^{\rho/(1+\rho)}.$$

On the other hand we can calculate directly

$$(2.21) \quad \min_c \max_{\Omega_\rho(H_0)} K_1(H; c),$$

which turns to be the same as (2.19), being attained by the set of the values defined (2.201) and (2.202).

Part II. Interpolations viewed from the standpoints of stochastically approximative analysis.

§ 1. Stochastic approaches to interpolation analysis. It is rather a common procedure to be observed in most of our applications that we shall first look at the difference-table before we shall have determined any choice of interpolation formula. In such a situation some informations are to be sought for which will lead us to decide how to answer to various problems. The first use of the difference table is to be served in deciding what is the suitable degree of polynomial to be used to fit for our data, and the second use in detecting any gross error which may be sometimes involved. In many cases of practical applications there will be various sources of errors which should be treated in probabilistic formulations such as those derived from random errors of observations. The relative efficiencies and roles of various procedures of interpolations can be discussed from the

standpoints of stochastically approximative analysis where we are not necessarily interested in each individual function but in a family of functions constituting a suitable probability field.

§ 2. Lagrange formula as a functional in the generalized Wiener space.

Let $f(t)$ be a function belonging to the function space (C_p) of all the continuous functions which are continuously differentiable to a certain p -th order and which can be written as

$$(2.01) \quad f(t) = a_0 + a_1 t + a_2 t^2 + \dots + a_{p-1} t^{p-1} + \int_0^t \frac{(t-s)^{p-1}}{(p-1)!} y(s) ds,$$

having the probability field introduced by the following conditions:

Condition (1°). $a_0, a_1, a_2, \dots, a_{p-1}$ are mutually independent stochastic variables with means 0 and finite variances $\sigma_0^2, \sigma_1^2, \dots, \sigma_{p-1}^2$ respectively.

Condition (2°) $y(\cdot)$ belongs to the function space (C) with Wiener measure.

Condition (3°) $y(\cdot)$ runs through (C) independently with $(a_0, a_1, \dots, a_{p-1})$.

Now let us assume that the values of $f(t)$ for a set of values $t = t_i$ ($i = 1, 2, \dots, n$) are observed and exactly known to us, which yields us the Lagrange interpolation formula

$$(2.02) \quad P(t) = \sum_{i=1}^n f(t_i) L_i(t),$$

with

$$(2.03) \quad L_i(t) = \frac{(t-t_1)(t-t_2)\dots(t-t_{i-1})(t-t_{i+1})\dots(t-t_n)}{(t_i-t_1)(t_i-t_2)\dots(t_i-t_{i-1})(t_i-t_{i+1})\dots(t_i-t_n)}.$$

[1] Let it be assumed that $p \leq n$. Then we have

$$(2.04) \quad R(t) \equiv f(t) - P(t) = \sum_{i=1}^n L_i(t) \int_{t_i}^t \frac{(t_i-u)^{p-1}}{(p-1)!} y(u) du,$$

which yields us

$$(2.05) \quad \int_0^1 [R(t)]^2 dt = \sum_{i=1}^n \sum_{j=1}^n \int_0^1 L_i(t) L_j(t) \int_{t_i}^t \frac{(t_i-u)^{p-1}}{(p-1)!} y(u) du \cdot \int_{t_j}^t \frac{(t_j-v)^{p-1}}{(p-1)!} y(v) dv.$$

In order to simplify our notations, let us introduce

$$(2.06) \quad \varphi_i(t, u) \equiv \begin{cases} L_i(t) \frac{(t_i-u)^{p-1}}{(p-1)!} & t_i \leq u < t \\ 0 & \text{elsewhere} \end{cases}$$

for $i = 1, 2, \dots, n$ in the two-dimensional square $0 \leq t, u \leq 1$.

Then we can rewrite (2.04) by

$$(2.07) \quad R(t) = \int_0^1 \sum_{i=1}^n \varphi_i(t, u) y(u) du.$$

For each fixed value of t in $0 \leq t \leq 1$, $R(t)$ is a functional defined over the Wiener space (C) . It will be more convenient to write (2.07) as

$$(2.08) \quad R(t) = - \int_0^1 y(u) d_u \psi(t, u),$$

with

$$(2.09) \quad \psi(t, u) \equiv \int_u^1 \sum_{i=1}^n \varphi_i(t, v) dv.$$

In view of the general theory of the Wiener space by PALEY-WIENER [1] and CAMERON-MARTIN [1], the distribution of the functional $R(t)$ is normal with the mean zero and the variance $\sigma_{R(t)}^2$, where

$$(2.10) \quad \sigma_{R(t)}^2 = \frac{1}{2} \int_0^1 [\psi(t, u)]^2 du.$$

[2] Now let it be assumed that $p \geq n+1$. Then we have

$$(2.11) \quad \begin{aligned} R(t) \equiv f(t) - P(t) &= \sum_{j=n+1}^p a_j [t^j - \sum_{i=1}^n t_i^j L_i(t)] \\ &+ \sum_{i=1}^n L_i(t) \int_{t_i}^t \frac{(t_i - u)^{p-1}}{(p-1)!} y(u) du. \end{aligned}$$

In this case the distribution of the functional $R(t)$ in the generalized WIENER space (C_p) is normal with the mean zero and the variance

$$(2.12) \quad \sigma_{R(t)}^2 = \sum_{j=n+1}^p \sum_{k=n+1}^p \sigma_j^2 \sigma_k^2 p_j(t) p_k(t) + \frac{1}{2} \int_0^1 [\psi(t, u)]^2 du,$$

where $\psi(t, u)$ is that defined in (2.09) and

$$(2.13) \quad p_j(t) \equiv t^j - \sum_{i=1}^n t_i^j L_i(t) \quad (j = n+1, \dots, p).$$

Now let it be assumed that our observations involve errors ε_i giving us $f(t_i) + \varepsilon_i$ instead of $f(t_i)$ at each point t_i , and let it be assumed that $\{\varepsilon_i\}$ ($i=1, 2, \dots, n$) is a realization of n independent stochastic variables with the mean 0 and the common variance σ^2 , satisfying the condition similar to the condition (3°). It can be now readily observed

Theorem 2.1. *Under the general formulation of the concurrence-functions and our observation errors enunciated in this paragraph, the remainder*

term $R(t)$ for each assigned t can be considered as a functional defined over the generalized Wiener space (C_p) with the addition of observation errors which has the mean value 0 and the variance

$$(2.14) \quad \frac{\sigma_{R(t)}^2}{2} + \sigma^2 \sum_{i=1}^n [L_i(t)]^2.$$

This Theorem implies in particular that, so far as $p \leq n$ holds true, the Lagrange interpolation is an unbiased estimator with the variance (2.10), for any sample value of a_0, a_1, \dots, a_{p-2} and a_{p-1} . For each assigned set of $\{t_i\}$ we can evaluate directly the value (2.14) for any t , and furthermore we can discuss the problem how to choose a set of observation points $\{t_i\}$ in order to minimize the averaged variance such as

$$(2.15) \quad \frac{1}{2} \int_0^1 \sigma_{R(t)}^2 dt + \sigma^2 \sum_{i=1}^n \int_0^1 [L_i(t)]^2 dt.$$

[3] Risk function approach. The following formulation in dealing with our problem may be of some interests. Let us consider the situation where our concurrence-function $f(t)$ can be assumed to be a polynomial of at most p -th degree, where $p \geq n+1, \{a_j\}$ being here unknown constants. Then under our hypothesis (2°) to this paragraph, we have, in view of $E_c[y(u)] = 0$,

$$(2.16) \quad R(t)^2 = (f(t) - P(t))^2 = \left\{ \sum_{j=n}^{p-1} a_j (t^j - \sum_{i=1}^n t_i^j L_i(t)) \right\}^2 \\ + \sum_{i=1}^n \sum_{j=1}^n L_i(t) L_j(t) \int_{t_i}^t \frac{(t_i - u)^{p-1}}{(p-1)!} y(u) du \int_{t_j}^t \frac{(t_j - v)^{p-1}}{(p-1)!} y(v) dv, \\ + 2 \sum_{j=n}^{p-1} a_j \left(t_j - \sum_{i=1}^n t_i^j L_i(t) \right) \cdot \sum_{i=1}^n L_i(t) \int_{t_i}^t \frac{(t_i - u)^{p-1}}{(p-1)!} y(u) du$$

$$(2.17) \quad E_c \left[\int_0^1 [R(t)]^2 dt \right] = \sum_{j=n}^{p-1} \sum_{l=n}^{p-1} a_j a_l B_{jl}^{(n)} \\ + \sum_{i=1}^n \sum_{j=1}^n \int_0^1 L_i(t) L_j(t) \int_{t_i}^t \frac{(t_i - u)^{p-1}}{(p-1)!} \int_{t_j}^t \frac{(t_j - v)^{p-1}}{(p-1)!} \\ \cdot E_c[y(u) y(v)] du dv,$$

where

$$(2.18) \quad B_{j,l}^{(n)} \equiv \int_0^1 \left(t^j - \sum_{i=1}^n t_i^j L_i(t) \right) \left(t^l - \sum_{i=1}^n t_i^l L_i(t) \right) dt$$

Notice that

$$(2.19) \quad E_c[y(u) y(v)] = \sigma^2 \text{Min}(u, v)$$

and put

$$(2.20) \quad A_{ij}^{(n)} \equiv \int_0^1 L_i(t) L_j(t) \int_{t_i}^t \frac{(t-u)^{p-1}}{(p-1)!} \int_{t_j}^t \frac{(t-v)^{p-1}}{(p-1)!} \text{Min}(u, v) du dv.$$

Then we have

$$(2.21) \quad E_c \left[\int_0^1 [R(t)]^2 dt \right] = \sum_{j=n}^{p-1} \sum_{l=n}^{p-1} a_j a_l B_{j,l}^{(n)} + \sigma^2 \sum_{i=1}^n \sum_{j=1}^n A_{i,j}^{(n)}.$$

The left-hand side is the average loss of using $p(t)$ as an estimate for $f(t)$, and is a functional dependent upon the choice of the set of division-points (t_1, t_2, \dots, t_n) and also naturally upon the concurrence-function of $f(t)$ since it is a quadratic form of the coefficients $(a_n, a_{n+1}, \dots, a_{p-1})$. In view of the right-hand side it is possible to make a risk function approach if we shall restrict ourselves with some partial problem such as to determine the number of division-points provided that the sequence $(t_1^{(n)}, t_2^{(n)}, \dots, t_n^{(n)})$ ($n = 1, 2, 3, \dots$) is assigned.

§ 3. The variate-difference procedure in interpolation analysis. One of our main concerns in applying a Lagrange interpolation formula is how to find out the degree $p-1$ of the polynomial. The use of the variate-difference procedure lies among others in contributing some informations in answering to this question. It is usually recommended to stop a continuation of constructing higher-difference just when we have reached the differences with some degrees which can be recognized to be a random sequence. This will naturally serve to decide the degree of an interpolation polynomial.

From the standpoint of our stochastically approximative analysis we may formulate accurately and precisely in the mathematical expressions this procedure of fitting a polynomial in view of sequences of differences.

Then it will turn out that this sort of procedure can be recognized as one of successive process of statistical inferences discussed specially in Parts I and III of the paper of the author [1] on this topics.

§ 4. Stochastic aspects associated with difference-variate analysis. Let our concurrence-function $f(t)$ be of the form

$$(4.01) \quad f(t) = \sum_{i=0}^{\infty} a_i t^i,$$

which is regular and analytic in the interval $0 \leq t \leq L$. Let our observation at each point t_j with equal intervals be

$$(4.02) \quad \hat{f}(t_j) = \sum_{i=0}^{\infty} a_i t_j^i + \varepsilon_j, \quad (j = 1, 2, \dots, n)$$

where $\{\varepsilon_j\}$ is a realization of the n independent stochastic variables distributed in $NID(0, \sigma^2)$, σ^2 being unknown to us.

Now let the differences of the ν -th order be denoted by

$$(4.03) \quad x_{\nu j} = \Delta^\nu \hat{f}(t_j) = \sum_{i=\nu}^{\infty} a_i \Delta^\nu t_j^i + \Delta^\nu \varepsilon_j \equiv m_{\nu j} + \Delta^\nu \varepsilon_j,$$

say ($j = 1, 2, \dots, n - \nu$).

In view of our difference formula we have

$$(4.04) \quad E(\Delta^\nu \varepsilon_j \Delta^\nu \varepsilon_k) = \sigma^2 \rho_{jk}^{(\nu)}.$$

Consequently the simultaneous probability density function of $y_\nu = (x_{\nu 1}, x_{\nu 2}, \dots, x_{\nu, n-\nu})$ is now given by

$$(4.05) \quad \frac{1}{(2\pi\sigma^2)^{(n-\nu)/2} |V_{n-\nu}|^{1/2}} \exp\left\{-\frac{1}{2\sigma^2} (y_\nu - m_\nu)' V_\nu^{-1} (y_\nu - m_\nu)\right\},$$

where the matrix $(\rho_{ik}^{(\nu)})$ is known to us, while the vector $m_\nu = (m_{\nu 1}, m_{\nu 2}, \dots, m_{\nu, n-\nu})$ and the constant σ^2 are unknown to us.

The general aspects of our situations to be considered in deciding a procedure how to choose a suitable order ν of the differences can be illustrated as follows:

(1°) For the concurrence-function $f(t)$ which reduces to a polynomial whose degree is less than ν we have $m_{\nu j} = 0$ for $j = 1, 2, \dots, n - \nu$. Broadly speaking there is a general tendency that the magnitude of m_ν becomes smaller as ν becomes greater.

(2°) The sequence of the variance matrices $\sigma^2 V_\nu^{-1}$ has a general tendency that $1' V_\nu^{-1} 1$ becomes greater as ν becomes greater.

(3°) Let us consider the null hypothesis $H_0^{(\nu)}: m_{\nu 1} = m_{\nu 2} = \dots = m_{\nu, n-\nu} = 0$. A choice of the degree ν should take into considerations the two fundamental aspects (1°) and (2°). One of the simplest procedures will be to introduce the generalised mean square error defined as

$$(4.06) \quad m_\nu' V_\nu^{-1} m_\nu + \sigma^2 1' V_\nu^{-1} 1,$$

where the first term corresponds to the square of the bias, while the second one to the variance. The quantity (4.06) is a function of ν and $a = (a_0, a_1, a_2, \dots)$, and consequently will be denoted by $W(\nu, a)$. We have, putting $V_\nu^{-1} = (b_{jk}^{(\nu)})$, ($j, k = 1, 2, \dots, n - \nu$),

$$(4.07) \quad \begin{aligned} W(\nu, a) &\equiv \sum_{j=1}^{n-\nu} \sum_{k=1}^{n-\nu} \left(\sum_{i=\nu}^{\infty} a_i \Delta^\nu t_j^i \right) b_{j,k}^{(\nu)} \left(\sum_{l=\nu}^{\infty} a_l \Delta^\nu t_k^l \right) + \sigma^2 1' V_\nu^{-1} 1 \\ &= \sum_{i=\nu}^{\infty} \sum_{l=\nu}^{\infty} a_i a_l \sum_{j=1}^{n-\nu} \sum_{k=1}^{n-\nu} b_{jk}^{(\nu)} \Delta^\nu t_j^i \Delta^\nu t_k^l + \sigma^2 1' V_\nu^{-1} 1 \\ &= \sum_{i=\nu}^{\infty} \sum_{l=\nu}^{\infty} a_i a_l B_\nu(i, l) + \sigma^2 \sum_{j=1}^{n-\nu} \sum_{k=1}^{n-\nu} b_{jk}^{(\nu)}. \end{aligned}$$

(4°) It is our common principle in choosing a suitable degree ν of the interpolation polynomial to find out the least order ν for which $H_0^{(\nu)}$ cannot be rejected. In some cases it can be performed in sequential procedure.

§ 5. Game theoretic approaches in deciding the degree of the interpolation formula. The considerations given in (3°) and (4°) in § 4 will now suggest us to formulate the following stochastic analysis.

In view of the formula (4.07) our risk function $R(\nu, a)$ can be defined as the sum of the loss $W(\nu, a)$ and the cost $c(\nu)$, which is here assumed simply to be a function of ν

$$(5.01) \quad R(\nu, a) \equiv W(\nu, a) + c(\nu).$$

The nature will choose his strategy a , while the statistician his strategy ν . In this case we shall be concerned with mixed strategies: let $\xi(\nu)$ and $\eta(a)$ be the probability distribution functions of ν and a respectively, and let us consider the problem how to find out the minimax strategy. That is to say, defining the weighted risk

$$(5.02) \quad R(\xi, \eta) \equiv \int \int R(\nu, a) d\xi(\nu) d\eta(a),$$

we want to find out the solution (ξ, η) satisfying

$$(5.03) \quad \inf_{\xi} \sup_{\eta} R(\xi, \eta) = \sup_{\eta} \inf_{\xi} R(\xi, \eta).$$

§ 6. Difference variate method and successive process of statistical inferences. Let s^2 be an independent estimate of σ^2 such that rs^2/σ^2 is distributed inchi-square distribution with the d.f.r. and let $F_r^{n-\nu}(\alpha_\nu)$ be the level of the significance of the statistic F associated to the probability α_ν of the error of the first kind. In what follows let us assume $\alpha = (\alpha_1, \dots, \alpha_{n-1})$ be an assigned sequence of the values such that $0 < \alpha_i < 1$ for $i = 1, 2, \dots, n-1$.

Our statistical procedure is defined to proceed in the following manner:

(1°) Accept the null hypothesis $H_0^{(1)}$ if

$$(6.01) \quad y_1' V_1^{-1} y_1 \leq F_r^{n-1}(\alpha_1) s^2$$

and decide to use the interpolation polynomial of the degree 0, that is, a constant.

(2°) Accept the null hypothesis $H_0^{(1)}$ if

$$(6.021) \quad y_1' V_1^{-1} y_1 > F_r^{n-1}(\alpha_1) s^2$$

$$(6.022) \quad y_2' V_2^{-1} y_2 \leq F_r^{n-2}(\alpha_2) s^2$$

and decide to use the interpolation polynomial of the degree 1.

(ν°) Accept the null hypothesis $H_0^{(\nu)}$ if

$$(6.031) \quad y_i' V_i^{-1} y_i > F_r^{n-i}(\alpha_i) s^2 \quad (i = 1, 2, \dots, \nu-1)$$

$$(6.032) \quad y_\nu' V_\nu^{-1} y_\nu \leq F_r^{n-\nu}(\alpha_\nu) s^2$$

and decide to use on interpolation polynomial of the degree $\nu-1$, for $\nu = 1, 2, 3, \dots, n-1$.

(n°) Decide to use the interpolation polynomial of the degree $n-1$, if

$$(6.04) \quad y_i' V_i^{-1} y_i > F_r^{n-i}(\alpha_i) s^2, \quad (i=1, 2, \dots, n-1).$$

It is not easy to give numerical calculations yielding us the power function of this statistical procedure. However it is to be noted that, under our formulation given in the previous paragraph, the theoretical aspects of this procedure are rather simple. Indeed we can enunciate the probability function associated with this procedure in the following way. Let us write

$$(6.05) \quad E(\Delta^p \varepsilon_i \Delta^q \varepsilon_j) = \sigma^2 \rho_{ij}^{(p, q)}, \quad (i=1, 2, \dots, n-p; j=1, 2, \dots, n-q)$$

$$(6.06) \quad V_{pq} = (\rho_{ij}^{(p, q)}),$$

which are $(n-p) \times (n-q)$ rectangular matrices in general and coincide with V_p when $p=q$.

Then the probability that we shall have the case $((\nu+1)^\circ)$ is equal to

$$(6.07) \quad p(\nu; a) \equiv \frac{1}{(2\pi\sigma^2)^{n(n+1)/2} |M^{1/2}|} \int \dots \int_{\mathfrak{D}_\nu(a)} \exp \left\{ -\frac{\sum_{i=1}^n (x_i - m_i)^2}{2\sigma^2} \right\} dx_1 \dots dx_n \\ \cdot g_r(s^2; \sigma^2) ds^2$$

where the domain of the integration is defined by

$$(6.8) \quad \begin{cases} y_i' V_i^{-1} y_i > F_r^{n-i}(\alpha_i) s^2 \\ y_i' V_\nu^{-1} y_\nu \leq F_r^{n-\nu}(\alpha_\nu) s^2. \end{cases}$$

and $g_r(s^2; \sigma^2)$ is the probability density function of the statistic s^2 where $r s^2 / \sigma^2$ is distributed in the chi-square distribution with the r degrees of freedom.

Part III. Successive process of fitting a sequence of orthogonal functions.

§ 1. Introductory. The use of orthogonal function systems in fitting empirical functions is particularly recommended in virtue of their property such that the coefficients of the fitted orthogonal expansions are determined irrespective of the degrees of the approximate formula. Therefore there have been suggested some sort of successive fittings of sections of orthogonal expansions. However any rigorous formulation of statistical procedure and any evaluation of their merits have not yet been well discussed in any literature. It is our purpose in this Part III to formulate the problems from the standpoints of the decision function approach under the general schemes of stochastically approximative analysis.

§ 2. General formulation of a successive procedure in fitting orthogonal functions. Let $\{\varphi_\nu(t; n)\}$ ($\nu = 0, 1, 2, \dots, n-1$) be a system of the normalized orthogonal functions associated with an assigned system of points where m independent observations on our unknown function $\tilde{g}(t)$ were performed which yielded us the sets of observed values $O_i: (y_{i1}, y_{i2}, \dots, y_{im})$ for $i = 1, 2, \dots, n$. Let us define

$$(2.01) \quad \hat{c}_\nu = \sum_{i=1}^n y_{i\cdot} \varphi_\nu(t_i; n) \quad (\nu = 0, 1, 2, \dots, n-1)$$

by

$$(2.92) \quad y_{i\cdot} = (y_{i1} + y_{i2} + \dots + y_{im})/m \quad (i = 1, 2, \dots, n).$$

Let us also define

$$(2.03) \quad \hat{g}_k(t) = \sum_{\nu=0}^k \hat{c}_\nu \varphi_\nu(t; n) \quad (k = 0, 1, 2, \dots, n-1)$$

$$(2.04) \quad s^2 = \sum_{i=1}^n \sum_{j=1}^m (y_{ij} - \bar{y}_{i\cdot})^2 / n(m-1)$$

$$(2.05) \quad \begin{aligned} r_k^2 = \|r_k(t)\|^2 &= \|\hat{g}_{n-1}(t) - \hat{g}_k(t)\|^2 \\ &= \sum_{i=1}^n \left[\sum_{\nu=k+1}^{n-1} \hat{c}_\nu \varphi_\nu(t_i; n) \right]^2 = \sum_{\nu=k+1}^{n-1} [\hat{c}_\nu]^2. \end{aligned}$$

The current procedures recommended in some literatures seem to proceed in the following way, if it be formulated more accurately. Let a sequence of fractions $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_{n-1})$ with the conditions $0 < \alpha_i < 1$ ($i = 0, 1, 2, \dots, n-1$) be assigned.

$(0^\circ)_0$ Test the null hypothesis $H_0^{(0)}: c_1 = c_2 = \dots = c_{n-1} = 0$. For this purpose we appeal to the statistic

$$(2.06) \quad F_0 = \|r_0\|^2 / s^2,$$

which, under the null hypothesis $H_0^{(0)}$, is distributed according to the F -distribution with the pair of the degrees of the freedom $[n-1, n(m-1)]$. That is, if $F_0 > F_{n(m-1)}^{n-1}(\alpha_0)$, we say that the test for $H_0^{(0)}$ is significant, and if otherwise, we say non-significant.

$(0^\circ)_1$ If the test $H_0^{(0)}$ is non-significant in the step $(0^\circ)_0$, then we adopt

$$(2.07) \quad \hat{g}(t) = \hat{g}_0(t).$$

$(1^\circ)_0$ If the test $H_0^{(0)}$ is significant in the step $(0^\circ)_0$, then we shall now test the null hypothesis $H_0^{(1)}: c_2 = c_3 = \dots = c_{n-1} = 0$.

For this purpose we appeal to the statistic

$$(2.08) \quad F_1 = \|r_1\|^2 / s^2,$$

which, under the null hypothesis $H_0^{(1)}$, is distributed according to the F -distribution with the pair of the degrees of the freedom $[n-2, n(m-1)]$.

That is, $F_1 > F_{n(m-1)}^{\alpha-2}(\alpha_1)$, we say that the test for $H_0^{(1)}$ is significant, and if otherwise, we say non-significant.

(1°)₁ If the test for $H_0^{(1)}$ is non-significant in the step (1°)₀, then we adopt

$$(2.09) \quad \hat{g}(t) = \hat{g}_1(t).$$

(2°)₀ If the test for $H_0^{(1)}$ is significant in the step (1°)₀, then we shall proceed to test the null hypothesis $H_0^{(2)}: c_3 = c_4 = \dots = c_{n-1} = 0$.

For this purpose we shall appeal to the statistic

$$(2.10) \quad F_2 = \|r_2\|^2 / s^2,$$

which, under the null hypothesis $H_0^{(2)}$, is distributed according to the F -distribution with the pair of the degrees of the freedom $[n-3, n(m-1)]$.

(2°)₁ Similarly we shall proceed on.

This procedure seems to be very natural in some respects. In more accurate considerations some preassumptions seem to have been set out beforehand without which the procedure described above would lack their validities and their effectiveness.

For instance, let $m=10$, $n=10$ and $\sigma^2=1$, and let our concurrence function be given by

$$(2.11) \quad \tilde{g}(t) = 1 + 0.01 \varphi_8(t; 10).$$

In such a situation, the procedure will have a relatively large chance of adopting some polynomial of very low degree. This is due to the one characteristic aspect of our statistical procedure, which gives the predetermined ascending order of testing the sequence of the null hypotheses. Such preferences on the lower degrees do not necessarily hold adequate. In the case when no information about the coefficients $\{c_\nu\}$ is provided for us, a more efficient approach may be possible which will test the significance of the coefficients in the order of the descending values of the statistics $\{\hat{c}_\nu\}$ ($\nu=0, 1, 2, \dots, n-1$).

§ 3. Successive process of statistical inferences associated with a sequence of fitting orthogonal functions. In this paragraph let our concurrence-function $\tilde{g}(t)$ be denoted by

$$(3.01) \quad \tilde{g}(t) \equiv \sum_{\nu=0}^{n-1} c_\nu \varphi_\nu(t; n),$$

which can be characterized by the vector $c = (c_0, c_1, \dots, c_{n-1})$.

Under this circumstance $\|r_k\|^2/s^2$ introduced in § 2 is distributed in the non-central F -distribution

$$(3.02) \quad F(n-1-k; n(m-1); 1, \sum_{\nu=k+1}^{n-1} c_\nu^2/\sigma^2),$$

where the pair of the degrees of freedom is $[n-1-k, n(m-1)]$ and the parameter of non-centrality is $\sum_{\nu=k+1}^{n-1} c_\nu^2 / \sigma^2$.

The procedure indicated in § 2 gives us the estimate $\hat{g}(t)$ for an unknown function $\tilde{g}(t)$ which is defined as follows:

- (i) If $H_0^{(0)}$ is non-significant, $\hat{g}(t) = \hat{g}_0(t)$.
- (ii) If $H_0^{(0)}$ is significant and $H_0^{(1)}$ is non-significant, $\hat{g}(t) = \hat{g}_1(t)$.
- (ν) If each of $H_0^{(i)}$ is significant for $i = 0, 1, 2, \dots, \nu-1$ and $H_0^{(\nu)}$ is non-significant, $\hat{g}(t) = \hat{g}_\nu(t)$. ($\nu = 2, 3, \dots, n-2$).
- ($n-1$) If each of $H_0^{(i)}$ is significant for $i = 0, 1, 2, \dots, n-1$, then $\hat{g}(t) = \hat{g}_{n-1}(t)$.

Let us denote by $d(y) = (d_0(y), d_1(y), \dots, d_{n-1}(y))$ the decision function defined by our statistical procedure just enunciated, where $d_\nu(y)$ is equal to one, when and only when we decide to employ $\hat{g}_\nu(t)$ in our statistical procedure, and is equal to zero, if otherwise. In virtue of this function, we can write

$$(3.03) \quad \hat{g}(t) = \sum_{\nu=0}^{n-1} d_\nu(y) \hat{g}_\nu(t).$$

Now let us calculate

$$(3.04) \quad \begin{aligned} \|\hat{g}(t) - \tilde{g}(t)\|^2 &= \sum_{j=1}^n [\hat{g}(t_j) - \tilde{g}(t_j)]^2 \\ &= \sum_{i=1}^n \left[\sum_{\nu=0}^{n-1} d_\nu(y) \sum_{i=0}^{\nu} \hat{c}_i \varphi_i(t_j; n) - \sum_{i=0}^{n-1} c_i \varphi_i(t_j; n) \right]^2 \\ &= \sum_{j=1}^n \left[\sum_{i=0}^{n-1} \left(\sum_{\nu=i}^{n-1} d_\nu(y) \right) \varphi_i(t_j; n) - \sum_{i=0}^{n-1} c_i \varphi_i(t_j; n) \right]^2 \\ &= \sum_{i=0}^{n-1} \left[\hat{c}_i \sum_{\nu=i}^{n-1} d_\nu(y) - c_i \right]^2. \end{aligned}$$

Writing

$$(3.05) \quad e_i(y) \equiv \sum_{\nu=i}^{n-1} d_\nu(y) \quad (i = 0, 1, 2, \dots, n-1),$$

we have

$$(3.06) \quad \begin{aligned} \|\hat{g}(t) - \tilde{g}(t)\|^2 &= \sum_{i=0}^{n-1} (\hat{c}_i - c_i)^2 e_i^2(y) \\ &\quad + 2 \sum_{i=0}^{n-1} c_i (\hat{c}_i - c_i) e_i(y) (e_i(y) - 1) \\ &\quad + \sum_{i=0}^{n-1} c_i^2 (e_i(y) - 1)^2 \\ &= \sum_{i=0}^{n-1} (\hat{c}_i - c_i)^2 e_i^2(y) + \sum_{i=0}^{n-1} c_i^2 (e_i(y) - 1)^2, \end{aligned}$$

since $e_i(y)(e_i(y) - 1) = 0$ for almost all y .

It will be interesting to determine our decision function $d(y)$ under some adequate restrictions. However let us restrict ourselves with the calculation of the mean value of (3.06) when we appeal to the statistical procedure in § 2.

In view of the relation (3.05) the statistical procedure in § 2 gives us the following assertions:

(0°) $e_0(y)$ is always equal to 1.

(1°) $e_1(y)$ is equal to 1, provided that y belongs to the domain D_1 defined by

$$(3.07) \quad |r_0|^2/s^2 > F_{n(n-1)}^{n-1}(\alpha_0).$$

And $e_1(y)$ is equal to zero, if y belongs to elsewhere.

(2°) $e_2(y)$ is equal to 1, provided that y belongs to the domain D_2 defined by

$$(3.08) \quad \begin{aligned} |r_0|^2/s^2 &> F_{n(m-1)}^{n-1}(\alpha_0) \\ |r_1|^2/s^2 &> F_{n(m-1)}^{n-2}(\alpha_1). \end{aligned}$$

And $e_2(y)$ is equal to zero, if y belongs to elsewhere.

(ν°) In general $e_\nu(y)$ is equal to 1, provided that y belongs to the domain D_ν defined by

$$(3.09) \quad \begin{aligned} |r_0|^2/s^2 &> F_{n(m-1)}^{n-1}(\alpha_0) \\ |r_1|^2/s^2 &> F_{n(m-1)}^{n-2}(\alpha_1) \\ &\dots\dots\dots \\ |r_{\nu-1}|^2/s^2 &> F_{n(m-1)}^{n-\nu}(\alpha_{\nu-1}). \end{aligned}$$

And $e_\nu(y)$ is equal to zero, if y belongs to elsewhere ($\nu = 3, 4, \dots, n-1$).

Now the mean distance between our concurrence-function $\tilde{g}(t)$ and the estimated function $\hat{g}(t)$ introduced in this paragraph will be defined by the formula

$$(3.10) \quad \begin{aligned} E_c\{|\hat{g}(t) - \tilde{g}(t)|^2\} \\ = \int |\hat{g}(t) - \tilde{g}(t)|^2 p(y/c) dy, \end{aligned}$$

where $y = (y_{11}, y_{12}, \dots, y_{1m}, y_{21}, \dots, y_{2m}, \dots, y_{n1}, y_{n2}, \dots, y_{nm})$ and

$$(3.11) \quad p(y/c) = (2\pi\sigma)^{-(nm)/2} \exp\left\{-\sum_{i=1}^n \sum_{j=1}^m (y_{ij} - \tilde{y}_i)^2 / 2\sigma^2\right\},$$

where we denote $c = (c_0, c_1, \dots, c_{n-1})$ because of the relation

$$(3.12) \quad c_\nu = \sum_{i=1}^n \tilde{y}_i \varphi_\nu(t_i, n) \quad (\nu = 0, 1, 2, \dots, n-1).$$

In our present case we have

$$(3.13) \quad \begin{aligned} E_c\{|\hat{g}(t) - \tilde{g}(t)|^2\} \\ = \sum_{\nu=0}^{n-1} \int_{D_\nu} (\hat{c}_\nu - c_\nu)^2 p(y/c) dy + \sum_{\nu=0}^{n-1} c_\nu^2 \left(1 - \int_{D_\nu} p(y/c) dy\right). \end{aligned}$$

Although the formula (3.13) is too complicated to observe directly any consequence which will depend upon a choice of $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_{n-1})$,

it may be still worth while to give this exact formula to which any approximation may be applied.

KYUSHU UNIVERSITY.

References

- (1) BLACKWELL, D. and GIRSHICK, M. A.: Theory of games and statistical decisions, New York, 1954.
- (2) CAMERON, R. H. and MARTIN, W. T.: *Transformations of Wiener integrals under translations*, Ann. Math., **54** (1944), 386.
- (3) FISHER, R. A.: [1] *Statistical methods and scientific induction*, Jour. Roy. Stat. Soc., B, **17** (1955), 69-78.
- (4) KITAGAWA, T.: [1] *Successive process of statistical inferences*, (1), Mem. Fac. Sci. Kyushu Univ., A, **5** (1950), 139-180.
- (5) KITAGAWA, T.: [2] *Sampling from processes depending upon a continuous parameter*, Mem. Fac. Sci., Kyushu Univ., A, **5** (1950), 181-188.
- (6) KITAGAWA, T.: [3] *Random integrations*, Bull. Math. Statistics, **5** (1950), 15-21.
- (7) KITAGAWA, T.: [4] *Successive process of statistical inferences*, (2), Mem. Fac. Sci. Kyushu Univ., A, **6** (1951), 53-95.
- (8) KITAGAWA, T.: [5] *Analysis of variance applied to function spaces*, Mem. Fac. Sci. Kyushu Univ., A, **6** (1951), 41-53.
- (9) KITAGAWA, T.: [6] *Successive process of statistical inferences*, (3) Mem. Fac. Sci. Kyushu Univ., A, **6** (1952), 131-155.
- (10) KITAGAWA, T.: [7] *Successive process of statistical inferences*, (4), Bull. Math. Statistics **5** (1951), 35-50.
- (11) KITAGAWA, T.: [8] *Successive process of statistical controls* (1), Mem. Fac. Sci. Kyushu Univ., A, **7** (1952), 13-26.
- (12) KITAGAWA, T.: [9] *Successive process of statistical inferences* (5), Mem. Fac. Sci. Kyushu Univ., A, **7** (1953), 81-106.
- (13) KITAGAWA, T.: [10] *Successive process of statistical inferences* (6), Mem. Fac. Sci. Kyushu Univ., A, **8** (1953), 1-29.
- (14) KITAGAWA, T.: [11] *Some stochastic considerations upon empirical functions of various types*, Mem. Fac. Sci. Kyushu Univ., A, **5** (1953), 19-33.
- (15) KITAGAWA, T.: [12] *The t-distribution concerning random integrations*, Mem. Fac. Sci. Kyushu Univ. A, **8** (1953), 31-41.
- (16) KITAGAWA, T.: [13] *Empirical functions and interpenetrating sampling procedures*, Mem. Fac. Sci. Kyushu Univ., A, **8** (1954), 109-152.
- (17) KITAGAWA, T.: [14] *Some contributions to the design of sample surveys*, Sankhya, **14** (1955), 317-362.
- (18) KITAGAWA, T.: [15] *Some contributions to the design of sample surveys*, (2), Sankhya **16** (1956). (in press)
- (19) ROBBINS, H. and MONRO, S.: [1]: *A stochastic approximation method*, Ann. Math. Stat., **22** (1951), 400-407.
- (20) WALD, A.: [1]: Statistical decision functions, New York, 1950.
- (21) WIENER, N.: Cybernetics (1947).
- (22) WOLFOWITZ, J.: [1]: *On the stochastic approximation method of Robbins and Monro*, Ann. Math. Stat., **23** (1952), 457-461.