Memoirs of the Faculty of Science, Kyūsyū University
Ser. A, Vol. 8, No. 1, 1953

SUCCESSIVE PROCESS OF STATISTICAL INFERENCES (6)

By

Tosio KITAGAWA

(Received May 20, 1953)

Contents

Introduction

In Part XII of our fifth paper of our series we have introduced a new formulation of appealing to previous knowledges which may be recognised to have some connections with the operational points of view. Now in Part XIII of the present paper our chief concerns are with such operational interpretations similar to that of Part XII, which will hold true for the problems of prognosis. The problems of prognosis will involve always so-called halo-effects, and we shall define halo-effect transformations, which can be suitably observed from the standpoints of the loss functions. § 1 in Part XIII will provide some commentary observations to our generalised formulations of successive decision functions enunciated in Part VIII, while

§ 2 gives moreover some contribution to Part IV. Although § 2 deals with a special problem, what we want to show is concerned with a rather general strategic principle, that is to say, the one which is taught in a Japanese proverb, " Don't use a batcher's knife for cooking chicken," and in an Indian proverb, " Don't shoot a fly with a gun." In § 3 we shall compare with our predictive behaviours with some fundamental principles which we may observe, as E. S. PEARSON pointed out, in the pioneer works of STUDENT and in the brilliant theories of R. A. FISHER. To make clear our real situations which we formulated in Parts II, IV, V, VII and VIII and specially to make more comprehensible our generalised formulation in Part VIII, we shall illustrate our points of view by a simple example, and then by discussing analysis of variance from our standpoint.

Part XIV in this paper is a continuation to Part XII in our fifth paper, and is a partial preparation to our successive design of experiment. It will chiefly concern itself with the problem how to determine the numbers of levels of factors in our design of experiments. In some engineering applications the numbers of levels of these factors are not so fixed ones as those of varieties in agricultural experimentations. The levels of temperature and those of pressures in some chemical experimentations may be chosen as great as we please. Indeed in such situations what we require to establish is to determine some functional relations which will hold among continuous variables. Consequently our problems must be reformed from the standpoint of functional analysis. We have already introduced some formulations of analysis of variance in function space in KITAGAWA [6]. In § 1 after certain generalisations of our previous paper we shall define a sequence of resolutions of identity and introduce main-effect functions and interaction functions. Here the expansion theory appealing to some orthogonal function systems will play its important rôles. To proceed our formulation in more details, we must give some theory of empirical function from stochastic point of view. There may be various formulations of empirical functions. We have already given some of them in KITAGAWA [7], which, in combination of the present Part XIV, will constitute a partial preparation for discussing successive design of experiments so far as our concerns are with adjustments of numbers of levels of factors previously determined.

In § 2 we shall make use of orthogonal expansions of effect function in analysis of variance, and we shall regard any design of experiments as a method of estimating coefficients of these orthogonal expansions. We shall observe some of fundamental factorial designs such as 2×2 and 3×3, and give some observations upon general cases. In spite of our several preparations, there still remain various considerations which should be done before we may be able to proceed to some theory of successive designs of experiments. In §3 we shall enunciate some classes of successive designs of experiments concerning levels of factors.

Part XIII. Predictive behaviours and risk functions

§ **1. Risk functions and prognosis.** The two sample formulation of the problems of prognosis developed in Part II of our paper [1] will show us that the solutions of these problems involve somewhat greater confidence intervals, larger acceptance regions than those in the inferences from a sample to its population. For example, we have enunciated in Theorem 2.1 in Part II, p. 153, the formula which is equivalent to

$$(1.01) \qquad Pr.\left\{|\bar{x}_2 - \bar{x}_1| \leq \frac{t_{n_1-1}(\alpha)}{\sqrt{n_1 - 1}} s_1 \sqrt{\frac{n_1 + n_2}{n_2}}\right\} = \alpha ,$$

which involves the multiplying factor $(n_1 + n_2)^{1/2} n_2^{-1/2}$ greater than 1, compared in addition to the case when $n_2 = \infty$, that is, when we infer from a sample to its parent population itself. The formulae similar to this particular one are valid, as may be seen from Theorems 2.2–2.10. Furthermore the relative efficiency theory developed in Part IV of our paper [2] may be observed to imply the similar results.

On the other hand we have introduced operational interpretation of previous knowledges in Part XII of our paper [5]. We point out here that such *operational interpretation* will hold true also for the problem of prognosis.

In an intuitive expression the problems of prognosis will involve nearly always a halo of vagueness around our moon, that is, the population parameter in our concern. In our two and three sample formulations, these *halo-effects* have been described by stochastic variable distributed around the population parameter. Now from our second standpoint of operationalism, the halo-effects may be recognised as certain operations which transform our estimates into other ones which are more vague and more vast than the formers.

Such halo-effect will happen when (a) there are possibilities that population parameter will change into another value and/or when (b) our real objects of prognosis are not population parameters themselves but rather a second sample drawn from the population. The first condition (a) will be reasonably treated within the scopes of statistical inferences under certain formulation of stochastic processes, while the second condition (b) belongs to the realm of two and three sample formulations. These two conditions will be sometimes amalgamated into one halo-operation.

Example 13. 1. Let $\hat{\theta}$ be an estimate of the population parameter θ with its density function $f(\hat{\theta})$ of sampling distribution. Our object of prognosis $\varepsilon\{z\}$ is not θ itself, but $\varepsilon\{z\}$ is assumed to be uniformly distributed in an interval $(\theta-a, \theta+a)$ when θ is true. Under this circumstance the halo-effect will be described by the transformation T which transforms a point estimate $\hat{\theta}$ into an interval predictor $(\hat{\theta}-a, \hat{\theta}+a)$, since we have now

$$(1.02)\qquad Pr.\{x<\hat{\theta}+\varepsilon\{z\}<x+dx\}=\frac{dx}{2a}\int_{-a}^{a} f(x+t)\,dt\,.$$

Example 13. 2. Let $\hat{\varphi}$ be an estimate of the population parameter $\varphi \geqq 0$ with its density function $g(\hat{\varphi}), 0 \leqq \hat{\varphi} < \infty$, of the sampling distribution. Now let the object of our prognosis $\varepsilon\{z\}$ be distributed uniformly in the interval $(a\varphi, b\varphi)$ when φ is true. Then the halo-effect will be described by the transformation T_1, which transforms a point estimate $\hat{\varphi}$ into an interval predictor $(a\hat{\varphi}, b\hat{\varphi})$.

Example 13. 3. More generally a halo-effect can be defined by a transformation of an estimate $\hat{\theta}$ of the population parameter θ into $T\hat{\theta}$ having its density function $h(x)$ of sampling distribution

$$(1.03)\qquad h(x\,;\,\theta)=\int_{-\infty}^{\infty} f(t\,;\,\theta)\,K(x,t)\,dt$$

with certain determined functlon $K(x, t)$.

The essential rôle of these halo-effect transformations can be observed from the standpoints of the loss functions. In view of the notations and terminolgies introduced in Part VII in our paper [3], we may enunciate the following

Theorem 13.1. *In addition to the formulation in Part* VIII, § 4, *let us assume furthermore the following* (i) – (iv):

(i) *The object of our prognosis* $\varepsilon(z)$ *reduces to a stochastic variable* z *itself such that* $\varepsilon(z)=z=\theta+u$, *where* θ *is a population parameter and* u *is a stochastic variable with the mean value* 0.

(ii) *The simultaneous distribution function* $K(x, y, z)$ *of the stochastic variables* X, Y *and* Z *will be decomposed into two mutually independent components to the effect that*

$$dK(x, y, z) = dL(z)\, dH(x, y)\,. \tag{1.04}$$

(iii) *The a priori distribution function of* θ *is assumed to be given by* $\xi(\theta)$ *in the parameter space* Ω_θ.

(iv) *Let us define* $r_1(\xi, \varepsilon\,\delta\mu)$ *by*

$$r_1(\xi, \varepsilon\,\delta\mu) \equiv \int_{\Omega_\theta} r_1(K, \varepsilon\,\delta\mu)\, d\xi(\theta)\,, \tag{1.05}$$

where $r_1(K, \varepsilon\,\delta\mu)$ *is defined as in* (4.01) *in Part* VIII.

Then we have

$$r_1(\xi, \varepsilon\,\delta\mu) = \int_{\Omega_\varphi}\int_{R_1(y,x)} W(\varphi, \delta\mu)\, dH(x, y)\, d\xi_1(\varphi)\,, \tag{1.06}$$

where Ω_φ *denotes the whole parameter space,* $R_1(y, x)$ *the whole sample space associated with* (Y, X), *and*

$$\xi_1(\varphi) = \int_{R(u)} \xi(\varphi - u)\, dL(u)\,, \tag{1.07}$$

$R(u)$ *being the whole sample space of* u.

Proof: In virtue of (i) and (ii), the right-hand side of (4.01) in Part VIII will reduce to

$$\begin{aligned} r_1(K, \varepsilon\,\delta\mu) &= \int_{D^t}\iiint_{R(z,y,x)} W(z, d^t)\, dP(D^t, \delta_y\,\mu_x)\, dK(z, y, x) \\ &= \int_{R(u)} dL(u)\int_{D^t}\int_{R_1(y,x)} W(\theta+u, d^t)\, dP(D^t, \delta_y\mu_x)\, d\,H(y, x)\,. \end{aligned} \tag{1.08}$$

The combination of (1.08) with (iii) and (iv) yields us (1.06) and (1.07), as we were to prove.

Corrollary 13.1. Specially when $W(z, d) \equiv (z - d^t)^2$ in the Assumptions to Theorem 13.1, then we have

$$(1.09)\quad r_1(\xi, \varepsilon\,\delta\mu) = \int_{D^t}\int_{\Omega_\theta}\int_{R(y,x)} W(\theta, d^t)\, dP(D^t, \delta_y\mu_x)\, dH(y, x)\, d\xi(\theta) + \int_{R(u)} u^2\, dL(u)\,.$$

Corrollary 13.1 is remarkable in the sense that for any strategy on sequential designs which we may employ the average loss is always greater than a certain constant:

$$(1.10)\qquad r_1(\xi, \varepsilon\,\delta\mu) \geqq \sigma_u^2 = \int_{R(u)} u^2\, dL(u)\,.$$

In a more general circumstance, the characteristic features of the problems of prognosis will be observed from an inequality

$$(1.11)\qquad r(\xi, \varepsilon\,\delta\mu) \geqq c\,,$$

c being a prescribed constant, which may be compared with the use of previous experiences in reaching decision functions in which a restricted Bayes solution has been introduced by HODGES, J. L. and LEHMANN, E. L. [1]: denoting by $R_\delta(\theta)$ the risk function of the decision procedure δ, a procedure δ_0 is said to be a restricted Bayes solution with respect to the a priori distribution λ and subject to the restriction

$$(1.12)\qquad R_\delta(\theta) \leqq C_0 \qquad \text{for all } \theta\,,$$

if it minimizes $\int R_\delta(\theta)\, d\lambda(\theta)$ among all procedures satisfying (1.13). Indeed (1.11) and (1.12) are in striking constrast with each other.

§ **2. Predictor in sequential analysis.** In this paragraph we shall show that real significance of prognosis problems can be observed from suitably defined risk functions. For the present purpose we shall discuss the following special problem of determining a predictor in sequential analysis which minimizes the maximum risk. The whole problem belongs to the realm of successive decision functions. In this paragraph we adopt the following Assumptions (1°) ~ (6°).

Assumption (1°). The object of prognosis $\varepsilon\{Z\}$ is to give a predicted value for a stochastic variable Z itself, where Z is distributed according to a normal distribution $N(\theta, \sigma^2/m)$.

Assumption (2°). Two sample formulation is adopted in the sense that there is no previous knowledge which secures a stochastic process X, and therefore, the simultaneous distribution $H(y, x)$ function of (Y, X) reduces to a distribution function $H(y)$ of Y alone, which is assumed to be given by

$$(2.01) \qquad H(y) = Pr.\{Y_i < y_i\,;\ i = 1, 2, 3, \cdots\}$$
$$= \prod_{i=1}^{\infty} \left\{(2\pi)^{-1/2}\sigma^{-1}\int_{-\infty}^{y_i} \exp\{-(v-\theta)^2/2\sigma^2\}\,dv\right\}.$$

Assumption (3°). The a priori distribution function $\xi_{\sigma_1}(\theta)$ of the population parameter θ is given by

$$(2.02) \qquad \xi_{\sigma_1}(\theta) = (2\pi)^{-1/2}\sigma_1^{-1}\int_{-\infty}^{\infty} \exp\{-\theta^2/2\sigma_1^2\}\,d\theta.$$

Assumption (4°). Our predictor is denoted by $d(y)$, and the loss function is defined as

$$(2.03) \quad W(\varepsilon(Z), \varepsilon\,\delta_y\gamma_x) = W(Z, d(y)) = w(Z - d(y))^2\sigma^{-2},$$

where w means a constant independent of Z and y.

Assumption (5°). The costs for making uses of storages of memories, and for formulating the statistics to be used in determining our successive decision function become zero, and the cost of making observations on $Y = \{Y_i\}$ is proportional to the number of observation n, that is,

$$(2.04) \qquad c(s_1, s_2, \cdots, s_k, d^t(\varepsilon_k)/\varepsilon_Z\,\delta_y\mu_x)$$
$$= c_2(y\,; s_1, s_2, \cdots, s_k) = c\,n,$$

where n denotes all the numbers of elements in the set of integers $(s_1, s_2, \cdots, s_k)$, c being a constant independent of n.

Assumption (6°). Each stage of experimentation consists exactly of a single observation on Y_i $(i = 1, 2, 3, \cdots)$.

We shall now observe by direct calculation

Lemma 13.1. *Let $d_n(y)$ be a decision function which depends only on the first n observations of Y, and let us define*

$$(2.05)\quad r(\xi, d_n(y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{cn + W(z, d_n(y))\}\, dL(z)\, d\xi(\theta).$$

Let $d_n^0(y)$ be the decision function which minimizes $r(\xi, d_n(y))$ with respect to d_n.

Then we have

$$(2.06)\quad d_n^0(y) = n\sigma_1^2(n\sigma_1^2 + \sigma^2)^{-1}\bar{y},$$

where $\bar{y} = (y_1 + y_2 + \cdots + y_n)/n$.

Furthermore we have

$$(2.07)\quad r(\xi_\sigma, d_n^0(y)) = r(n)\, g_n(y),$$

where we put

$$(2.08)\quad r(n) = cn + w\{m^{-1} + \sigma_1^2(n\sigma_1^2 + \sigma^2)^{-1}\}$$

$$(2.09)\quad g_n(y) = k_n\, exp\{-(\sum_{i=1}^{n} y_i^2 - n^2\sigma_1^2\bar{y}^2(n\sigma_1^2 + \sigma^2)^{-1})(2\sigma^2)^{-1}\},$$

with

$$(2.10)\quad k_n = \{(2\pi)^{1/2}\sigma\}^{-n}(2\pi)^{-1/2}(n\sigma_1^2 + \sigma^2)^{-1/2}\sigma.$$

Furthermore the mean value of $r_1(\xi_\sigma, d_n^0(y))$ under the condition that the set of values $(y_1, y_2, \cdots, y_{n-1})$ is assigned is

$$(2.11)\quad E_{n-1}\{r(\xi_\sigma, d_n^0(y))\}$$
$$= (2\pi)^{-1/2}\sigma^{-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{cn + W(z, d_n(y))\}$$
$$\cdot \exp\{-(y_n - \theta)^2/2\sigma^2\}\, dy\, dL(z)\, d\xi(\theta)$$
$$= r(n)\, g_{n-1}(y).$$

Here we shall make use of the following two Theorems which are valid under fairly general conditions without Assumptions (1°) ∼ (6°).

Theorem A. (MIYASAWA [1]). Let us define $r_j(\xi, y)$ and $\{\alpha_{jn}(\xi, y)\}$ $(j = 0, 1, 2, \cdots, n)$ in the recurrent manner:

(1°) $r_j(\xi, y) = r(\xi, d_j^0(y))$

(2°) $\alpha_{n,n}(\xi, y) = r_n(\xi, y)$

(3°) For $j < n$, we define

(2.12) $$\alpha_{jn}(\xi, y) = \min[r_j(\xi, y),\ E_j\{\alpha_{j+1,n}(\xi, y)\}],$$

where

(2.13) $$E_j\{\alpha_{j+1,n}(\xi, y)\} = \int_{-\infty}^{\infty} \alpha_{j+1,n}(\xi, y)\, dg(y_{n+1}),$$

g being the probability distribution function of Y_{n+1}.

Then we have the following assertions:

(a) $\alpha_{jn}(\xi, y)$ is a monotonically decreasing function of n. If we denote by $\alpha_j(\xi, y)$ the limit of $\alpha_{jn}(\xi, y)$ as $n\rightarrow\infty$, then we have

(2.14) $$\alpha_j(\xi, y) = \min.[r_j(\xi, y),\ E_j\{\alpha_{j+1}(\xi, y)\}].$$

(b) Let us define a sequence of sets $\{S_j\}$ in the sample space $y = (y_1, y_2, \cdots, y_n, \cdots)$ by

(2.15) $$S_j = \{y;\ r_i(\xi, y) > \alpha_i(\xi, y) \text{ for } i < j, \text{ and } r_j(\xi, y) = d_j(\xi, y)\}$$

for all $j = 1, 2, 3, \cdots$. Let d_ξ be the decision function which proceeds such that if y belongs to S_j, we shall stop the experimentation at the j-th observation and determine the terminal decision $d_j^0(y)$. Then d_ξ is the BAYES solution relative to ξ.

Theorem B. (LEHMANN [1]). Let $d_{\xi k}$ be the BAYES solution relative to the a priori distribution of the population parameter θ, $k = 1, 2, 3, \cdots$. Let d^* be a decision function for which

(2.16) $$\sup r(\theta, d^*) \leqq \lim_{k\rightarrow\infty} \sup r(\xi_k, d_k).$$

Then the decision function d^* is a minimax solution.

In virtue of these two Theorems, we shall observe

Theorem 13.2. *Under the Assumptions* (1°) ~ (6°) *in this paragraph, we have the following two assertions:*

(1°) *Let* $n = n_{c_1\sigma_1}{}^*$ *be the greatest integer which is smaller than*

(2.17) $$2^{-1}\{1 + (1 + 4c^{*-1})^{1/2}\} - \sigma^{*-2},$$

where we have put

(2.18) $$c^* = c\,w^{-1}, \quad \sigma^{*2} = \sigma_1{}^2\sigma^{-2}.$$

Then the dicision function d_{σ_1} *which indicates to take n observation* $y_1, y_2, \cdots, y_n$ *and there stoppes the experimentation and which predicts* $\varepsilon\{Z\} = Z$ *by* $d_n^0(y)$ *defined in Lemma* 13.1 *is the* BAYES *solution relative to* ξ_{σ_1} *when* σ_1 *is sufficiently large.*

(2°) *Let d^* be the decision function which indicates to take these $n = n_{c_1 \infty}$ observations $x_1, x_2, \cdots, x_n$ and there terminate the experimentation and to predict Z by the sample mean $\bar{y} = (y_1 + y_2 + \cdots + y_n)n^{-1}$. Then d^* is the minimax solution of our prognosis problem.*

The proofs of (1°) and (2°) follow immediately from Theorems A, B and Lemma 13.1, as quite similarly to those of Theorems 3.1 and 3.2 in MIYASAWA [1]. Indeed it suffices us to notice that $n = n_{c_1 \sigma_1}$ is determined as the smaller root of the equation $\varphi(n) = 0$, where

$$\begin{aligned}\varphi(n) &= E_{n-1}\{r_n(\xi_{\sigma_1}, y)\} - r_{n-1}(\xi_{\sigma_1}, y) \\ &= E_{n-1}\{r(\xi_{\sigma_1}, d_n^0(y)\} - r(\xi_{\sigma_1}, d_{n-1}^0(y)) \\ &= \{r(n) - r(n-1)\} g_{n-1}(y),\end{aligned} \tag{2.19}$$

that is equivalent to the equation $r(n) - r(n-1) = 0$.

In view of Theorem 13.2 both of the BAYES solution relative to ξ_σ and minimax solution are shown to be independent of the precision of our object of prognosis, requiring experimenters to make their bests independently from the latter. This is somewhat curious from our practical considerations. Why should they trouble with reducing the average risk under their controls into the amount smaller than ten dollars, when there are certain possibilities of suffering from dangers amounting one million dollars which will derive from uncontrollable causes? The discrepancies of our results in Theorem 13.2 with our practical considerations may be remediable to some extent by introducing a new definition of the loss function. For example we may define

$$W(Z, d(y)) = w(Z - d(y))^2/\sigma^2(Z) \tag{2.20}$$

w being a constant, in stead of (2.03). We shall observe now

Theorem 13.3. *Under the same hypothesis to Theorem* 13.2, *except the definition of the loss function, which is now replaced by* (2.20) *in stead of* (2.03), *the results similar to* (1°) *and* (2°) *hold true, except that $n_{c^*_1 \sigma^*_1}$, is now defined by another constant c^{**} such that $n_{c^{**}, \sigma^{**}_1}$ is the greatest integer which is smaller than*

$$2^{-1}\{1 + (1 + 4c^{**-1})^{1/2}\} - \sigma^{**-2}, \tag{2.21}$$

where

$$(2.22) \qquad c^{**} = c(mw)^{-1}, \quad \sigma^{**2} = \sigma_1^2 \sigma^{-2} = \sigma^{*2}$$

and similarly for $n_{C^{**},\infty}$, *the limit of* $n_{C^{**},\sigma^{**}}$, *as* σ^{**} *tends to* ∞.

This Theorem shows among others that (1°) the larger m becomes, that is, the higher the precision of the object of our prognosis becomes, then we should necessarily make the larger observations and that (2°) the smaller m becomes, that is, the lower the precision of the object of our prognosis becomes, then we may dispense with the smaller observations.

Such characteristic procedures for predictions may be generally introduced into our general formulation of successive decision functions, provided that we may be able to define the risk intrinsically associated with the object of prognosis, say $r\{\varepsilon\}$. There are indeed two ways: the first is similar to that discussed in this paragraph, that is, for example, we may define

$$(2.23) \qquad W(\varepsilon_k(Z), d^t(\varepsilon_k)) = (\varepsilon_k(Z) - d^t((\varepsilon_k))^2/r\{\varepsilon_k\},$$

while the second will be defined along the idea introduced in the previous paragraph. For example it may be of some interest to introduce

Definition 13.1. *A procedure* δ_0 *is said to be a Bayes solution of prognosis of* $\varepsilon\{Z\}$ *with respect to the a priori distribution and subject to the condition*

$$(2.24) \qquad r(K, \varepsilon\,\delta\eta) \geqq \alpha\, r\{\varepsilon\} \quad \textit{for all } K,$$

α *being a prescribed positive constant, if* δ_0 *minimizes* $\int r(K, \varepsilon\,\delta\eta)\, d\xi(K)$ *among all the procedures satisfying* (2.24).

The detailed discussion of Definition 13.1 is postponed to another occasion.

§ 3. Predictive behaviours and analysis of loss into its components. In his article in the memory of WILLIAM SEALY GOSSET, said E. S. PEARSON [1], p. 211–212 "There is one very simple and illuminating theme which will be found to run as a keynote through much of his work, and may be expressed in the two formulae:

$$\sigma_{x+y}^{2} = \sigma_x^2 + \sigma_y^2 + 2\rho\sigma_x\sigma_y \cdots (1)$$
$$\sigma_{x-y}^{2} = \sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y \cdots (2).$$

Perhaps we may count as one of his achievements the demonstration in many fields of the meaning of that short equation (2); as he wrote in 1923 (11, p. 273, but with modified notation):

The art of designing all experiments lies even more in arranging matters so that ρ is as large as possible than in reducing σ_x^2 and σ_y^2.

It is a simple idea, certainly, but I cannot doubt that its emphasis and amplification helped to open the way to all the modern development of analysis of variance, and …." R. A. FISHER [1] illustrates the fundamental ideas on design of experiments in Chaper III, specially in § 16.

In these connections we shall point out the needs of emphasis upon our predictive behaviours which will make clear our real situations.

Let us illustrate our points of view by a simple example. Let θ be a population parameter, and let $d^t = y$ be an estimater for θ with probability density function $f(y)$ such that the mean and the variance of y are given by

$$E\{y\} = \int_{-\infty}^{\infty} yf(y)\,dy = \theta' \tag{3.01}$$

$$\sigma^2\{y\} = \sigma^2. \tag{3.02}$$

Let $W(\theta, y) = (\theta - y)^2$ be the loss which we suffer when θ is true and y is its estimated value. The average loss when θ is true and d^t is adopted as estimator is given then

$$r_1(\theta, d^t) = \int_{-\infty}^{\infty} (\theta - y)^2 f(y)\,dy = (\theta - \theta')^2 + \sigma^2. \tag{3.03}$$

For the sake of simplicity and emphasis on essential points, let us assume that σ^2 is a constant independent of θ and θ'.

Under these formulations we can make the following four observations.

(1°) In the case when $\theta' = \theta$ the average risk $r_1(\theta, d^t)$ becomes constant independent of θ and θ'. There is no need of an a priori distribution for θ.

(2°) In the case when θ' does not coincide with θ, there leaves some discussion or some presumption for possible differences between

θ and θ'. The classical BAYES approach is to presume an a priori distribution $\xi(\theta-\theta')$ and to introduce

$$(3.04)\qquad r(\xi, d) = \int r(\theta, d^t)\, d\xi(\theta-\theta') = \int (\theta-\theta')^2 d\xi(\theta-\theta') + \sigma^2 .$$

On the other hand the minimax principle approach due to WALD [1] concerns itself how to minimize the maximum risk, that is, $\max_\theta r(\theta, d^t)$. In spite of the essential differences between these approaches there is an intimate relation between minimax solution and BAYES one to the effect that under certain rather weak conditions any minimax solution is also a BAYES solution relative to a least favourable a priori distribution.

The real duties of practical statisticians in designing his experiments cannot be completely fulfilled by seeking merely minimax strategy for rather broad class of possible a priori distributions. A practical statistician should endevour at first to reduce a broad class of possible a priori distributions into narrow one.

In our two sample points of views our predictive behaviours associated with the analysis of variance seems to play their essential roles for this purpose, as will be illustrated in what follows.

(3°) It may be frequently observed that "what we need in practice is not the inference about the population parameter, but rather the inferences about its another sample which will be drawn from this population on another occasion." (KITAGAWA [1] p. 141–142).

Let us illustrate our points of views by one factor formulation in which the whole population will be divided into m strata with stratum means $\mu_i + Y_i$ $(i = 1, 2, \cdots, m)$, where μ_i are the stratum parameters while $\{Y_i\}$ are mutually independent stochastic variables distributed in a normal distribution $N(0, \sigma_B^2)$. Any individual observation will be assumed to be recognised as a realisation of stochastic variable $\mu + \mu_i + Y_i + W_i$, where μ is a grand population mean, and $\{W_i\}$ $(i = 1, 2, \cdots, m)$ are mutually independently distributed with each others and with $\{Y_i\}$ according to the normal distribution $N(0, \sigma_W^2)$. For example let us denote by W_i the sampling fluctuation due to random sampling from the i-th stratum, while

Y_i denotes the material fluctuation of the i-th stratum mean itself. Let it be assumed that $\sum_{i=1}^{m} \mu_i = 0$. From the standpoints of predictive behaviours the real objects of our inferences may be sometimes recognised not itself, but rather $\bar{x}^{(2)}_{mn_2}$ which will correspond to the stochastic variable

$$(3.05) \qquad \bar{X}^{(2)}_{mn_2} = (mn_2)^{-1} \sum_{i=1}^{m} \sum_{k=1}^{n_2} X^{(2)}_{ik} = (mn_2)^{-1} \sum_{i=1}^{m} \sum_{k=1}^{n_2} (\mu + \mu_i + Y_i + W^{(2)}_{ik}),$$

when we have obtained the sample mean $\bar{x}_{ln_1}{}^{(1)}$ which corresponds to the stochastic variable

$$(3.06) \qquad \bar{X}^{(1)}_{ln_1} = (ln_1)^{-1} \sum_{h=1}^{l} \sum_{k=1}^{n_1} X^{(2)}_{i_h k} = (ln_1)^{-1} \sum_{h=1}^{l} \sum_{k=1}^{n_1} (\mu + \mu_{i_h} + Y_{i_h} + W^{(1)}_{i_h k}).$$

In the case when $l = m$ and $i_h = h\ (h = 1, 2, \cdots, m)$, $\bar{X}^{(1)}_{mn_1}$ is an unbiassed predictor for $\bar{X}^{(2)}_{mn_2}$, and we have

$$(3.07) \qquad W\{\bar{X}^{(2)}_{mn_2}, \bar{X}^{(1)}_{mn_1}\} = \sigma^2\{(\bar{X}^{(2)}_{mn_2} - \bar{X}^{(1)}_{mn_1})\} = \{(mn_2)^{-1} + (mn_1)^{-1}\}\sigma_W{}^2.$$

If $\sigma_W{}^2$ is independent of other population parameter, then there is no need of assuming any a priori distribution ξ for the grand population mean μ. On the other hand otherwise $\bar{X}_{ln_1}{}^{(1)}$ may not be an unbiassed predictor for $\bar{X}^{(2)}_{mn_2}$. For example let $l = m$ and $i_h = h\ (h = 1, 2, \cdots, l)$. Then we have

$$(3.08) \qquad E\{\bar{X}^{(2)}_{mn_2} - \bar{X}^{(1)}_{ln_1}\} = \sum_{i=l+1}^{m} \mu_i,$$

which is not necessarily zero, and

$$(3.09) \qquad W\{\bar{X}^{(2)}_{mn_2}, \bar{X}^{(1)}_{ln_1}\} = (\sum_{i=l+1}^{m} \mu_i)^2 + (m-l)(ml)^{-1}\sigma_Y^2 + \{(mn_2)^{-1} + (ln_1)^{-1}\}\sigma_W^2.$$

On this circumstance, not only the loss becomes greater than (3.09), but also there is some needs for previous knowledges upon $(\mu_1, \mu_2, \cdots, \mu_m)$ or otherwise for appealing to some strategy such as minimax principle. If there is an a priori distribution $\xi(\mu_{l+1}, \cdots, \mu_m)$, then the average loss will be given by

$$(3.10)\quad r_1(\xi, d) = \int \cdots \int (\sum_{i=l+1}^{m} \mu_i)^2 \, d\xi(\mu_{l+1}, \cdots, \mu_m) + (m-l)(m\,l)^{-1}\sigma_B^2 + \{(m\,n_2)^{-1} + (l\,n_1)^{-1}\}\,\sigma_W^2.$$

It is also to be noted that the correlation coefficient between $\bar{X}_{m n_1}$ and $\bar{X}_{l n_2}$ is given by

$$(3.11)\quad \rho(\bar{X}_{m n_1}, \bar{X}_{l n_2}) = \left(\frac{l}{m}\right)^{1/2} (1 + \sigma_W^{*2} n_1^{-1})^{-1/2} (1 + \sigma_W^{*2} n_2^{-1})^{-1/2},$$

where we put $\sigma_W^{*2} = \sigma_W^2 \sigma_B^{-2}$. This correlation coefficient becomes greater as l becomes nearer to m.

(4°) If we would treat our problem merely as estimation problem, then the a priori distribution of μ would be suitably observed to be

$$(3.12)\quad \xi^*(\mu) = \int_{-\infty}^{\infty} \xi_1(\mu - v)\, g(v)\, dv,$$

where we define ξ_1 by

$$(3.13)\quad \xi_1(z) = \int \cdots \iint_{(\mu_1^2+\mu_2^2+\cdots+\mu_m^2)^{1/2}<z} d\xi(\mu_1, \mu_2, \cdots, \mu_m),$$

while g is the density function of the normal distribution $N(0, \sigma_B^2 m^{-1})$.

It is to be noted that $\xi^*(\mu)$ will surely have larger variance than each of that due to its components ξ_1 and g.

In conclusion of this paragraph we emphasize what we want to point out: (a) The predictive behaviour is indispensable for applying the Student keynote to our statistical inferences; (b) The rôle of predictive behaviour is to recognise the valiability of the object of prognosis, but at the same time to design our experiment in which our observations will be recognised as stochastic variables with intimate correlation with those corresponding to the object of prognosis; (c) The broad class of a priori distribution function can be reduced to a narrower one by our procedure of analysis of variances in our design of experiments.

Part XIV. Uses of orthogonal functions on successive designs of experiments concerning levels of factors

§ 1. Successive analysis of variance with adjustments of the numbers of levels of each factor. Let us consider a factorial experiment $A \times B$ involving two factors A and B. In an ordinary analysis of variance, these two factors are considered at m levels and n levels respectively, and hence $m \times n$ design will be used with certain number of replications r. It is however to be noted that there occur frequently cases when setting of any large numbers of levels may be possible. Thus temperatures and pressures in some of chemical experiments may be chosen to any prescribed values ranging almost continuously in certain intervals. In such situations it may be said to be routine procedures to appeal to successive process of designing the experiments by adjusting the number of levels of the factors. In such situations our objects of statistical inferences may change from stage to stage. Sometimes the first object of experiment may be to test the null hypothesis that there is no main effect due to the factor A. After the significance of this null hypothesis has been verified, then at a second stage we may sometimes aim to estimate the relationship between the characteristic in which we are interested and the factors in our concerns, such as to estimate regression curves, or sometimes to seek the maximum point of the characteric when each factor is ranging in its domain respectively. In this preliminary § 1, we shall begin with the sequential processes of testing the null hypothesis which will adopt the adjustments of numbers of levels of each factor. Even within this restricted object it seems to us to be necessary to adopt the notion of *analysis of variance in certain function space* C_2 consisted of all functions which are defined in the two-dimensional intervals: $0 \leqq t,\ \tau \leqq 1$, and are everywhere continuous in a certain sense vanishing on the sides $t=0$ and $\tau=0$. The generalised main effects due to the factors A and B may be expressed in terms of some functions $F(t)$ and $G(\tau)$, while the generalised interactions between A and B in terms of some functions of two variables $H(t, \tau)$, when A and B are assigned to take the values t and τ respectively.

We have already discussed the analysis of variance applied to function spaces in KITAGAWA [6]. The following slight generalisations will be necessary in the present paper.

Definition 14.1. *Let $F(t)$ and $G(\tau)$ be continuous and of bounded variations in $0 \leqq t \leqq 1$ and $0 \leqq \tau \leqq 1$ respectively, and let $H(t, \tau)$ be continuous and of bounded variation in (t, τ) in the two-dimensional interval $0 \leqq t \leqq 1$, $0 \leqq \tau \leqq 1$. The functional of f and g, which we denote by $A(f, g)$, is defined by*

$$(1.01)\quad A(f, g) \equiv \int_0^1 f(t)\,dF(t) + \int_0^1 g(\tau)\,dG(\tau) + \int_0^1\int_0^1 f(t)\,g(\tau)\,dH(t, \tau) + \int_0^1\int_0^1 f(t)\,g(\tau)\,d_t d_\tau x(t, \tau),$$

where $x(t, \tau)$ *is a member that belongs to the function-space* C_2 *introduced in* KITAGAWA [6].

Definition 14.2. *A family of infinite sequence of fnnctions* $\{f_i^{(m)}(t)\}$ $(i = 1, 2, \cdots, m;\ m = 1, 2, 3, \cdots)$ *is said to be a complete family of resolutions of unit function when the following conditions are satisfied for every* m:

$$(1^\circ)\quad f_1^{(m)}(t) + f_2^{(m)}(t) + \cdots + f_m^{(m)}(t) = m, \quad (0 \leqq t \leqq 1)$$

$$(2^\circ)\quad \int_0^1 f_i^{(m)}(t)\,dt = 1 \qquad (i = 1, 2, \cdots, m)$$

$$(3^\circ)\quad \int_0^1 f_i^{(m)}(t)\,f_j^{(m)}(t)\,dt = m\,\delta_{ij} \quad (i, j = 1, 2, \cdots, m).$$

Definition 14.3. *Let* $\{f_i^{(m)}(t)\}$ $(i = 1, 2, \cdots, m;\ m = 1, 2, 3, \cdots)$ *and* $\{g_j^{(n)}(\tau)\}$ $(j = 1, 2, \cdots, n;\ n = 1, 2, 3, \cdots)$ *be two complete families of resolutions of unit functions. Then under the assumption to Definition* 14.1, *the infinite sequences of constants* (i) $\{a_i^{(m)}\}$, (ii) $\{b_j^{(n)}\}$ *and* (iii) $\{(a\,b)_{ij}^{(mn)}\}$ *are called the complete families of* (i) *the main effects due to the factor A, of* (ii) *main effects due to the factor B, and of* (iii) *the two factor interactions due to the factors A and B respectively*:

$$(1.02)\qquad a_i^{(m)} \equiv \int_0^1 f_i^{(m)}(t)\,dF(t)$$

$$(1.03)\qquad b_j^{(n)} \equiv \int_0^1 g_j^{(n)}(\tau)\,dG(\tau)$$

$$(1.04)\qquad (a\,b)_{ij}^{(mn)} \equiv \int_0^1\int_0^1 f_i^{(m)}(t)\,g_j^{(n)}(\tau)\,dH(t, \tau),$$

for $i = 1, 2, \cdots, m;\ m = 1, 2, 3, \cdots$ and $j = 1, 2, \cdots, n;\ n = 1, 2, 3, \cdots$.

From the practical standpoint the null hypothesis to be tested concerning the main effects due to the factor A should be nothing but $F(t) \equiv 0$, while at each stage of our experiments we shall be concerned with the statistical hypothesis

(1.05) $$H_0^{(m)}(A):\ a_1^{(m)} = a_2^{(m)} = \cdots = a_m^{(m)} = 0.$$

These gaps between the real objects and our ordinary mathematical formulations can be narrowed only by the successive processes of testing the sequence of these null hypotheses $\{H_0^{(m)}(A)\}$ $(m = 2, 3, \cdots)$. The alternative hypothesis should be given by assigning a function $F(t)$ defined in $0 \leqq t \leqq 1$, which does not identically vanish there, and for which the sequence of the statistical hypotheses at each stage should be given as

(1.06) $$H_1^{(m)}(A):\ a_i^{(m)} = \theta_i^{(m)},\ \sum_{i=1}^{m} \theta_i^{(m)} = 0,\ \sum_{i=1}^{m} \theta_i^{(m)2} > 0.$$

Similarly for the main effects due to the factor B the null hypothesis and its alternative one should be given as follows:

(1.07) $$H_0^{(n)}(B):\ b_1^{(n)} = b_2^{(n)} = \cdots = b_n^{(n)} = 0\ \ (n = 1, 2, 3, \cdots).$$

(1.08) $$H_1^{(n)}(B):\ b_j^{(n)} = \varphi_j^{(n)},\ \sum_{j=1}^{n} \varphi_j^{(n)} = 0,\ \sum_{j=1}^{n} (\varphi_j^{(n)})^2 > 0.$$

And also for the two-factor interaction between the factors A and B the null hypothesis and its alternative one will be given in a double series:

(1.09) $$H_0^{(mn)}(B):\ (ab)_{11}^{(mn)} = (ab)_{12}^{(mn)} = \cdots = (ab)_{mn}^{(mn)} = 0 \quad (m, n = 1, 2, 3, \cdots)$$

(1.10) $$H_1^{(mn)}(B):\ (ab)_{ij}^{(mn)} = \varphi_{ij}^{(mn)},\ \sum_{j=1}^{n} \varphi_{ij}^{(mn)} = \sum_{i=1}^{m} \varphi_{ij}^{(mn)} = 0$$
$$\sum_{i=1}^{m} \sum_{j=1}^{n} [\varphi_{ij}^{(mn)}]^2 > 0.$$

Now we shall remind the probability distribution functions of the non-central F-distribution which are due to TANG [1], and which we shall enunciate for our present purpose in the following

Theorem (TANG [1]). *Under the hypothesis $H_1^{(m)}(A)$, the probability density function of the stochastic variable $G = S_R/(S_R + S_E)$ is*

(1.11) $$h_{f_1, f_2}(G/\lambda_A^{(m)})$$
$$= \sum_{i=0}^{\infty} \frac{\lambda_A^{(m)i} \exp\{-\lambda_A^{(m)}\}}{\Gamma(i+1)\, B(f_1/2 + i, f_2/2)} G^{(f_1/2)+i-1} (1 - G)^{f_2/2-1},$$

where we put

(1.12) $f_1 = m - 1,\ f_2 = mn(r-1)$

(1.13) $\lambda_A^{(m)} = \dfrac{nr}{2\sigma_E^2} \cdot \dfrac{1}{m} \sum_{i=1}^{m} \left\{ \int_0^1 f_i^m(t)\, dF(t) \right\}^2 = \dfrac{nr}{2\sigma_E^2} \sigma_{A,m}^2,$ *say.*

In what follows in this paragraph we denote by $G(f_1, f_2;\ \alpha/0)$ or simply by $G_{f_2}^{f_1}(\alpha)$ the α-point of the distribution $h_{f_1,f_2}(G_2/0)$ such that

(1.14) $$\int_{G(f_1, f_2;\ \alpha/0)}^{1} h_{f_1,f_2}(G/0)\, dG = \alpha .$$

It is also to be noted that (i) $G = f_1 F/(f_1 F + f_2)$. And hence $G/(1-G) = f_1 F/f_2$, where $F = S_R/S_E$; (ii) let $F_{f_2}^{f_1}(\alpha)$ be the α-point of F when $\lambda_A = 0$, then $G(f_1, f_2;\ \alpha/0) \equiv G_{f_2}^{f_1}(\alpha) = f_1 F_{f_2}^{f_1}(\alpha) / (f_1 F_{f_2}^{f_1}(\alpha) + f_2)$; (iii) as f_2 tends to infinity, $G(f_1, f_2;\ \alpha/0)$ tends to zero.

On the other hand we shall make further preparation by introducing the following

Assumption 14.1. *The following limit exists:*

(1.15) $\lim_{m\to\infty} m^{-1} \sum_{i=1}^{m} \left\{ \int_0^1 f_i^m(t)\, dF(t) - \int_0^1 dF(t) \right\}^2 \equiv \sigma_A^2,$ *say.*

This Assumption will be certainly verified in various fields of applications for designs of experiments.

For instances this Assumption holds true when the following two conditions are satisfied: (1°) for each i, $1 \leqq i \leqq m$, $f_i^m(t) = m$ in the subinterval $(i-1)m^{-1} \leqq t < i m^{-1}$, and $f_i^m(t) = 0$ elsewhere; (2°) $F(t)$ is continuously differentiable function defined in $0 \leqq t \leqq 1$. Indeed we shall have then

(1.16) $$\lim_{m\to\infty} m^{-1} \sum_{i=1}^{m} \{F(i/m) - F((i-1)/m)\}^2 - \left(\int_0^1 dF(t) \right)^2$$
$$= \int_0^1 \{F'(t)\}^2 dt - \left(\int_0^1 F'(t)\, dt \right)^2 .$$

The goals to which the successive designs of experiments will aim to attain may be testings of statistical hypotheses as enunciated in the following table:

Inferences concerned with	Null hypothesis	alternative hypotheses
A	$H_0(A)$	$H_1(A)$
B	$H_0(B)$	$H_1(B)$
AB	$H_0(AB)$	$H_1(AB)$
A and B	$H_0(A) \cap H_0(B)$	$H_i(A) \cap H_j(B) \quad (i+j \geqq 1)$
A and AB	$H_0(A) \cap H_0(AB)$	$H_i(A) \cap H_j(AB) \quad (i+j \geqq 1)$
B and AB	$H_0(B) \cap H_0(AB)$	$H_i(B) \cap H_j(AB) \quad (i+j \geqq 1)$
A, B and AB	$H_0(A) \cap H_0(B) \cap H_0(AB)$	$H_i(A) \cap H_j(AB) \cap H_k(AB) \quad (i+j+k \geqq 1)$

Nevertheless, from the view point of successive design of experiments, what we really concern ourselves at each stage are some of the hypotheses $H_i^m(A)$, $H_j^n(B)$ and $H_{ij}^{mn}(AB)$ $(i, j = 0, 1)$ and their products, that is, simultaneous validities, and consequently all the sequences of possible courses of successive designs constitute a variety of abundancy. Moreover the types of designs of experiments, even in our restriction within the scope of $m \times n \times (r)$ designs, also constitute a three dimensional multiplicity whose element will be denoted by the coordinate (m, n, r). Thus the adjustments now in our consideration may be characterised as the transition from (m, n, r) into (m', n', r'). There are also many possibilities and their combinations amount to numerous one. The transitions of designs of experiments from (m, n, r) into (m', n', r') are called to be (i) r-type, when $m = m'$, $n = n'$; (ii) m-type, when $n = n'$, $r = r'$; (iii) n-type when $m = m'$, $r = r'$; (iv) mr-type when $n = n'$; (v) nr-type when $m = m'$; (vi) mn-type when $r = r'$.

For the detailed discussions, however, it seems to be adequate to introduce effect functions and their expansions into orthogonal functions, to which we shall devote the following two paragraphes.

§ 2. **Uses of orthogonal expansions of effect functions in analysis of variances.** In this preparatory pagraph, we shall restrict ourselves within a formal description of effect functions, which, for a sake of brevity, are assumed to be functions of two variables (x, y) in the quadrat $0 \leqq x, y \leqq 1$. For practical applications, the numbers of levels should be introduced as shown in the next paragraph. Our object in this paragraph is to show that uses of orthogonal expansions of the functions $f(x, y)$ will make clear the formal aspects of main effects and interactions adopted in analysis of variance.

Let $\{\varphi_m(x)\}$ $(m=0, 1, 2, 3, \cdots)$ be a complete normalised orthogonal system (CONS) in the function space $L^2(0, 1)$ of squarely integrable functions in the sense of LEBESGUE integration in $0 \leqq x \leqq 1$, and further let $\{\varphi_m(x)\,\varphi_n(y)\}$ $(m, n=0, 1, 2, 3, \cdots)$ be also CONS in the function spaces $L^2(Q)$ where Q is the domain $0 \leqq x, y \leqq 1$. Under certain restrictions upon these functions which secure convergence of the following series expansion

$$(2.01) \qquad f(x, y) \sim \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{m,n}\, \varphi_m(x)\, \varphi_n(y),$$

with

$$(2.02) \qquad a_{m,n} = \int_0^1 \int_0^1 f(x, y)\, \varphi_m(x)\, \varphi_n(y)\, dx\, dy, \quad (m, n = 0, 1, 2, \cdots)$$

the formal considerations will lead us to the correspondence that the main effects with respect to X-factor and Y-factor are given by

$$(2.03) \qquad a(x) = \int_0^1 f(x, y)\, dy - \int_0^1 \int_0^1 f(x, y)\, dx\, dy$$

$$(2.04) \qquad b(y) = \int_0^1 f(x, y)\, dx - \int_0^1 \int_0^1 f(x, y)\, dx\, dy$$

respectively, while the interaction between X-and Y-factors by

$$(2.05) \qquad (ab)(x, y) = f(x, y) - \int_0^1 f(x, y)\, dy - \int_0^1 f(x, y)\, dx + \int_0^1 \int_0^1 f(x, y)\, dx\, dy.$$

Now let us consider the special case when expansion is due to a system of orthogonal polynomials, each $\varphi_m(x)$ being a polynomial of m-th degree, and let us restrict ourselves for a moment with the case when

$$(2.06) \qquad f(x, y) = c_{00} + c_{10}\,\varphi_1(x) + c_{01}\,\varphi_1(y) + c_{20}\,\varphi_2(x) + c_{11}\,\varphi_1(x)\,\varphi_1(y) + c_{02}\,\varphi_2(y),$$

where we have put $\varphi_0(x) \equiv \varphi_0(y) \equiv 1$. Then we can write

$$(2.07) \qquad a(x) = c_{10}\,\varphi_1(x) + c_{20}\,\varphi_2(x)$$

(2.08) $$b(y) = c_{01}\varphi_1(y) + c_{02}\varphi_2(y)$$

(2.09) $$(a b)(x, y) = c_{11}\varphi_1(x)\varphi_1(y) .$$

The problem how to estimate effect function $f(x, y)$ will be consequently reduced to that of estimations of these five constants c_{ij}. The skills and powers of the ordinary factorial designs of experiments $m \times n$ should be judged from the efficiencies of estimations of these parameters. In this standpoint the numbers of levels m and/or n are sufficient enough to yield us these estimations. To investigate such problems concerning level numbers it seems almost necessary to appeal to orthogonal expansions of functions defined for a discrete set of x and y arguments, that is, essentially the resolution of unit functions as defined in KITAGAWA [6] and also in the previous pagraph. For the sake of brevity, it will suffice us in the present case to appeal to systems of orthogonal polynomials due to TSCHEBYCHEFF [1], defined for the interval $(-1, 1)$.

[1] 2×2 *designs and estimations of* c_{00}, c_{10}, c_{01} *and* c_{11}. The use of TSCHEBYCHEFF orthogonal polynomials give us the estimations of the four constants c_{00}, c_{10}, c_{01} and c_{11} by means of the 2×2 designs which yield us the observations on $f(-1/2, -1/2)$, $f(1/2, -1/2)$, $f(-1/2, 1/2)$ and $f(-1/2, -1/2)$ for which the relations hold:

(2.10) $$\begin{cases} (1/2)(c_{00} - c_{01} - c_{10} + c_{11}) = f(-1/2, -1/2) \\ (1/2)(c_{00} - c_{01} + c_{10} - c_{11}) = f(1/2, -1/2) \\ (1/2)(c_{00} + c_{01} - c_{10} - c_{11}) = f(-1/2, 1/2) \\ (1/2)(c_{00} + c_{01} + c_{10} + c_{11}) = f(1/2, 1/2) . \end{cases}$$

The coefficients of the linear equations in the left-hand side can be easily verified by the following Table 14.1.

[2] 3×3 *designs and estimations of* c_{00}, c_{01}, c_{02}, c_{10}, c_{11}, c_{12}, c_{20}, c_{21} *and* c_{22}. Let us put for a moment

(2.11) $$\begin{cases} Z_1 = f(-1, -1), & Z_2 = f(-1, 0), & Z_3 = f(-1, 1) \\ Z_4 = f(0, -1), & Z_5 = f(0, 0), & Z_6 = f(0, 1) \\ Z_7 = f(1, -1), & Z_8 = f(1, 0), & Z_9 = f(1, 1) \end{cases}$$

(2.12) $$Z = (Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, Z_7, Z_8, Z_9)$$

and

(2.13) $$c = (c_{00}, c_{01}, c_{02}, c_{10}, c_{11}, c_{12}, c_{20}, c_{21}, c_{22}) .$$

Table 14.1. The matrix A in 2×2

			φ_0		φ_1	
			$-1/2$	$1/2$	$-1/2$	$1/2$
			$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$	$\frac{-1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$
φ_0	$-1/2$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{-1}{2}$	$\frac{1}{2}$
	$1/2$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{-1}{2}$	$\frac{1}{2}$
φ_1	$-1/2$	$\frac{-1}{\sqrt{2}}$	$\frac{-1}{2}$	$\frac{-1}{2}$	$\frac{1}{2}$	$\frac{-1}{2}$
	$1/2$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{-1}{2}$	$\frac{1}{2}$

Table 14.2. The matrix A in 3×3

			φ_0			φ_1			φ_2		
			-1	0	1	-1	0	1	-1	0	1
			$\frac{1}{\sqrt{3}}$	$\frac{1}{\sqrt{3}}$	$\frac{1}{\sqrt{3}}$	$\frac{-1}{\sqrt{2}}$	0	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{6}}$	$\frac{-2}{\sqrt{6}}$	$\frac{1}{\sqrt{6}}$
φ_0	-1	$\frac{1}{\sqrt{3}}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$-\frac{1}{\sqrt{6}}$	0	$\frac{1}{\sqrt{6}}$	$\frac{1}{\sqrt{18}}$	$-\frac{2}{\sqrt{18}}$	$\frac{1}{\sqrt{18}}$
	0	$\frac{1}{\sqrt{3}}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$-\frac{1}{\sqrt{6}}$	0	$\frac{1}{\sqrt{6}}$	$\frac{1}{\sqrt{18}}$	$-\frac{2}{\sqrt{18}}$	$\frac{1}{\sqrt{18}}$
	1	$\frac{1}{\sqrt{3}}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$-\frac{1}{\sqrt{6}}$	0	$\frac{1}{\sqrt{6}}$	$\frac{1}{\sqrt{18}}$	$-\frac{2}{\sqrt{18}}$	$\frac{1}{\sqrt{18}}$
φ_1	-1	$-\frac{1}{\sqrt{2}}$	$-\frac{1}{\sqrt{6}}$	$-\frac{1}{\sqrt{6}}$	$-\frac{1}{\sqrt{6}}$	$\frac{1}{2}$	0	$-\frac{1}{2}$	$-\frac{1}{\sqrt{12}}$	$\frac{2}{\sqrt{12}}$	$-\frac{1}{\sqrt{12}}$
	0	0	0	0	0	0	0	0	0	0	0
	1	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{6}}$	$\frac{1}{\sqrt{6}}$	$\frac{1}{\sqrt{6}}$	$-\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{\sqrt{12}}$	$\frac{-2}{\sqrt{12}}$	$\frac{1}{\sqrt{12}}$
φ_2	-1	$\frac{1}{\sqrt{6}}$	$\frac{1}{\sqrt{18}}$	$\frac{1}{\sqrt{18}}$	$\frac{1}{\sqrt{18}}$	$-\frac{1}{\sqrt{12}}$	0	$\frac{1}{\sqrt{12}}$	$\frac{1}{6}$	$-\frac{2}{6}$	$\frac{1}{6}$
	0	$-\frac{2}{\sqrt{6}}$	$-\frac{2}{\sqrt{18}}$	$-\frac{2}{\sqrt{18}}$	$-\frac{2}{\sqrt{18}}$	$\frac{2}{\sqrt{12}}$	0	$-\frac{2}{\sqrt{12}}$	$-\frac{2}{6}$	$\frac{4}{6}$	$-\frac{2}{6}$
	1	$\frac{1}{\sqrt{6}}$	$\frac{1}{\sqrt{18}}$	$\frac{1}{\sqrt{18}}$	$\frac{1}{\sqrt{18}}$	$-\frac{1}{\sqrt{12}}$	0	$\frac{1}{\sqrt{12}}$	$\frac{1}{6}$	$-\frac{2}{6}$	$\frac{1}{6}$

Then it can be verified that estimations of c by means of observations Z in the 3×3 designs will be obtained from the matrix relation:

$$(2.14) \qquad Z = A_{3,3}\, c,$$

where $A_{3,3}$ is an orthogonal matrix 9×9 uniquely determined from the 3×3 design as verified from Table 14.2. The following formulae obtained from (2.14) will be useful for the present purpose.

$$(2.15) \quad \begin{cases} c_{00} = \dfrac{1}{3}(Z_1 + Z_2 + Z_3 + Z_4 + Z_5 + Z_6 + Z_7 + Z_8 + Z_9) \\ c_{01} = \dfrac{1}{\sqrt{6}}(Z_3 + Z_6 + Z_9 - Z_1 - Z_4 - Z_7) \\ c_{02} = \dfrac{1}{\sqrt{18}}\{Z_1 + Z_4 + Z_7 - 2(Z_2 + Z_5 + Z_8) + Z_3 + Z_6 + Z_9\} \\ c_{10} = \dfrac{1}{\sqrt{6}}(Z_7 + Z_8 + Z_9 - Z_1 - Z_2 - Z_3) \\ c_{11} = \dfrac{1}{2}(Z_1 + Z_9 - Z_3 - Z_7) \\ c_{12} = \dfrac{1}{\sqrt{12}}(2Z_2 + Z_7 + Z_9 - Z_1 - Z_3 - 2Z_8) \\ c_{20} = \dfrac{1}{\sqrt{18}}(Z_1 + Z_2 + Z_3 - 2(Z_4 + Z_5 + Z_6) + Z_7 + Z_8 + Z_9) \\ c_{21} = \dfrac{1}{\sqrt{12}}(-Z_1 + Z_3 + 2Z_4 - 2Z_5 - Z_7 + Z_9) \\ c_{22} = \dfrac{1}{6}(Z_1 - 2Z_2 + Z_3 - 2Z_4 + 4Z_5 - 2Z_6 + Z_7 - 2Z_8 + Z_9) \end{cases}$$

[3] *General considerations.* It can be now readily observed from the results and the estimation procedures in [1] and [2] that the following general assertions will be established:

(1°) In an $m \times n \times (r)$ factorial design, we can estimate all the following mn constants $\{C_{ij}\}$ encountered in the effect-function

$$(2.16) \qquad f(x, y) = \sum_{i=0}^{m} c_{i0}\,\varphi_i(x) + \sum_{j=0}^{n} c_{0j}\,\varphi_j(y) + \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}\,\varphi_i(x)\,\varphi_j(y).$$

The similar results can be obtained for any factorial designs $m_1 \times m_2 \times \cdots \times m_k$ with k factors ($k \geqq 3$).

(2°) From the standpoint of our purposes which aim to give estimations of all constants $\{c_{ij}\}$ in effect functions one and more replications of all the combinations of treatments in factorial designs will sometimes neither be necessary nor sufficient. The weighing problems due to HOTELLING, KISHEN, BANERJEE and others may be recognised to belong to our estimation-problems as special cases when all the interactions of factorial designs become inexistent. Specially the cases well discussed in literatures are those when 2^n designs are in consideration and the constants $\{c_{i_1 i_2 \cdots i_n}^{(n)}\}$, where all i_j are zero except one of them are to be estimated.

(3°) The problem how to choose numbers of levels in a factorial design must be also investigated in connection with the problem how to locate the observation-points in the domain of factorial arguments. For instance we have used the observation-points $(-1, -1)$, $(-1, 1)$, $(1, -1)$ and $(1, 1)$ in the domain $-1 \leqq x, y \leqq 1$. In case of 2×2 design, and the 9 observation points (i, j) $(i, j = -1, 0, 1)$ in case of 3×3 design. In fact this is intimately connected with the use of orthogonal polynomial system, but there remain certain possibilities for making use of observation points of other types. Indeed it seems to be necessary and adequate to develop such general theory as giving us some principal indications for choosing observation-points.

(4°) The problem whether there may be any pecularities in choosing orthogonal polynomials rather than more general orthogonal function systems remains unsolved in this paper. It seems to the present auther that there will be necessity to make clear the notion of levels in connection with the adoptation of polynomial expansions.

§3. Successive designs of experiments concerning levels of factors. In order to discuss the problems of successive designs of experiments concerning levels of factors, we must treat at first with the problem how to introduce and to define the dimensionality of each factor. The various situations which we encounter in our statistical researches, for example, in psychological ones, will sometimes require us to adopt generalised dimension functions which have been introduced in the lattice theory and developed in continuous geometry by J. NEUMAN and others. Nevertheless in this present paper we shall restict ourselves within the classical dimension ideas and specially treat with the case of finite dimension.

Let an effect function $g(t)$ be assigned with its expansion into the orthogonal polynomials of TSCHEBYCHEFF type in the interval $(0, 1)$:

$$g(t) = \sum_{k=0}^{n-1} c_n \phi_{n,k}(t). \tag{3.01}$$

On the other hand our observations will be always finite dimensional. In our terminology introduced in 1, we shall adopt some k-th resolution $\{f_i^{(k)}(t)\}$ $(i = 1, 2, \cdots, k)$ of unit function. Then we can not treat $g(t)$ itself, but we can only investigate

$$c_i^{(k)} = \int_0^1 f_i^{(k)}(t)\, dG(t), \quad (i = 1, 2, \cdots, k) \tag{3.02}$$

where $G(t)$ is defined by

$$G(t) = \int_0^t g(t)\, dt. \tag{3.03}$$

To make the matter simpler without any fear of conclusion, however, we may and we shall hereafter assume the circumstances that our observations will be done at a set of discrete points $\tau = (t_1, t_2, \cdots, t_n)$, $0 \leqq t_1 \leqq t_2 \leqq \cdots \leqq t_n \leqq 1$, and that each observation at the point t_i can be expressed by

$$x(t_i) = \sum_{k=0}^{n-1} c_k \phi_{n,k}(t_i) + \varepsilon_{t_i}, \tag{3.04}$$

where $\{\varepsilon_{t_i}\}$ $(i = 1, 2, \cdots, n)$ are mutually independently distributed according to a normal distribution $N(0, \sigma^2)$, σ^2 being unknown to us.

The real statistical problems with which we are concerned essentially are to make inferences about the effect function $g(t)$ itself, while the observation $(t_1, t_2, \cdots, t_n)$ must be regarded to be chosen according to our strategy in making these inferences. Out strategy will be determined by our previous knowledge, by our object of prognosis, and by cost considerations, in short by risk considerations. In the following example we shall explain our problems how to increase or to maintain or to decrease the number of levels in connection with the number of observation points.

Example 14.1. Let an effect function $g(t)$ be of the third degree

$$g(t) = c_0 \varphi_{30}(t) + c_1 \varphi_{31}(t) + c_2 \varphi_{32}(t). \tag{3.05}$$

In view of the values $\varphi_{3j}(t)$ $(j=0,1,2;\ t=-1,0,1)$, we have

$$(3.06)\quad \begin{cases} g(-1)=\dfrac{c_0}{\sqrt{3}}-\dfrac{c_1}{\sqrt{2}}+\dfrac{c_2}{\sqrt{6}}\equiv l(c;\ -1) \\ g(0)=\dfrac{c_0}{\sqrt{3}}-\dfrac{2c_2}{\sqrt{6}}\equiv l(c;\ 0) \\ g(1)=\dfrac{c_0}{\sqrt{3}}+\dfrac{c_1}{\sqrt{2}}+\dfrac{c_2}{\sqrt{6}}\equiv l(c;\ 1), \quad \text{say.} \end{cases}$$

Let us consider all the possible sequences of designs of experiments in which the number of the observation points may increase in a sense of successive process from one to three, by certain rule of procedure. There are consequently 6 ! possible ways as the permutation of three points -1, 0 and 1. Due to the symmetry of the matrix $(\varphi_{ij}(t))$ $(j=0,1,2;\ t=-1,0,1)$, these 6 possible procedures of experiments will be divided into three classes:

(1°) $(0,-1,1)$ and $(0,1,-1)$; (2°) $(-1,0,1)$ and $(1,0,-1)$; (3°) $(-1,1,0)$ and $(1,-1,0)$.

(a) Testing of the null hypothesis $H_0: f(t)\equiv 0$.

(i) First let us make observations at $t=t_1$ with n_1 replications, which yield us a random sample of size n_1, $O_{n_1}^{(1)}$: $(x_{11}(t_1), x_{12}(t_1), \cdots, x_{1n_1}(t))$.

(ii) Let us now test the null hypothesis H_0 by means of the statistic

$$(3.07)\qquad F_1=(n_1-1)\,n_1\,(\bar{x}_1(t_1))^2/\{\sum_{j=1}^{n_1}(x_{1j}(t_1)-\bar{x}_1(t_1))^2\},$$

for a prescribed level significance α_1, $0<\alpha_1<1$, in view of the F-distribution with the pair of degrees of freedom $[1, n_1-1]$, where we have put

$$(3.08)\qquad \bar{x}_1(t_1)=(x_{11}(t_1)+x_{12}(t_1)+\cdots+x_{1n}(t_1))/n_1.$$

$(\text{iii})_1$ If F_1 is significant in (ii), then we shall stop the experimentation, and we reject the null hypothesis H_0.

$(\text{iii})_2$ If F_1 is non-significant in (ii), then we shall continue the experimentation, by making n_2 independent observations at the point t_2, which will yield us a random sample of size n_2, $O_{n_2}^{(2)}$: $(x_{21}(t_2), x_{22}(t_2), \cdots, x_{2n_2}(t_2))$.

(iv) Let us now test the null hypothesis H_0 by means of the statistic

$$(3.09) \qquad F_2 = (n_2 - 1)\, n_2(\bar{x}_2(t_2))^2 / \sum_{j=1}^{n_2} (x_{2j}(t_2) - \bar{x}_2(t_2))^2$$

for a prescribed level of significance α_2, $0 < \alpha_2 < 1$, in view of F-distribution with the pair of degrees of freedom $[1, n_2 - 1]$, where we have put

$$(3.10) \qquad \bar{x}_2(t_2) = (x_{21}(t_2) + x_{22}(t_2) + \cdots + x_{2n_2}(t_2))/n_2 .$$

$(\mathrm{v})_1$ If F_2 is significant in (iv), then we shall stop the experimentation, and we reject the null hypothesis H_0.

$(\mathrm{v})_2$ If F_2 is non-significant in (iv), then we shall continue the experimentation by making n_3 independent observations at the point t_3, which will yields us a random sample of size n_3, $O_{n_3}^{(3)}$: $(x_{31}(t_3), x_{32}(t_3), \cdots, x_{3n}(t_3))$.

(vi) Let us now test the null hypothesis H_0 by means of the statistic

$$(3.11) \qquad F_3 = (n_3 - 1)\, n_3(\bar{x}_3(t_3))^2 / \sum_{j=1}^{n_3} (x_{3j}(t_3) - \bar{x}_3(t_3))^2$$

for a prescribed level of significance α_3, $0 < \alpha_3 < 1$, in view of F-distribution with the pair of degrees of freedom $[1, n_3 - 1]$, where we have put

$$(3.12) \qquad \bar{x}_3(t_3) = (x_{31}(t_3) + x_{32}(t_3) + \cdots + x_{3n_2}(t_3))/n_3 .$$

$(\mathrm{vii})_1$ If F_3 is significant in (vi), then we reject the null hypothesis H_0.

$(\mathrm{vii})_2$ If F_3 is non-significant in (vi), then we shall not reject the null hypothesis H_0.

(b) Testing of the null hypothesis $H_0: f(t) \equiv 0$.

This successive process is defined quite similarly as in (a) except the statistic which will be used in F-tests. So far as the assumptions that the variances of $\{\varepsilon_t\}$ $(t = -1, 0, 1)$ are constants for all t, we may replace F_2 and F_3 by the following F_2^* and F_3^* respectively:

$$(3.13) \qquad F_2^* = \frac{\{(n_1 - 1) + (n_2 - 1)\}\{n_1 \bar{x}_1^2(t_1) + n_2 \bar{x}_2^2(t_2)\}}{2 \sum_{i=1}^{2} \sum_{j=1}^{n_i} (x_{ij}(t_i) - \bar{x}_i(t_i))^2}$$

and

(3.14) F_3^*

$$= \frac{\{(n_1-1)+(n_2-1)+(n_3-1)\}(n_1\bar{x}_1^2(t_1)+n_2\bar{x}_2^2(t_2)+n_3\bar{x}_3^2(t_3))}{3\sum_{i=1}^{3}\sum_{j=1}^{n_i}(x_{ij}(t_i)-\bar{x}_i(t_i))^2},$$

which can be tested by uses of F-distributions with the pairs of degrees of freedom $[2, n_1+n_2-2]$ and $[3, n_1+n_2+n_3-3]$ respectively.

The following two points are essentially important in discussing these procedures: (a) First these procedures manifestly appeal to successive processes of testing hypotheses, and consequently we must reconsider the errors of the first and the second kinds from the standpoints developed in Parts I ~ V, VII and VIII; (b) Differences of efficiencies of successive designs will be seen choosing course of successive processes.

Literatures

BLYTH, C. R.: [1] *On minimax statistical decision problems and their admissibility*, Ann. Math. Stat., **22** (1951), 22-42.

FISHER, R. A.: [1] *The designs of experiments*, Oliver Boyd, Edinburgh, 5th ed., (1949).

HODGES, J. and LEHMANN: [1] *Some problems in minimax point estimation*, Ann. Math. Stat., **21** (1950), 182-197.

[2] *The use of previous experience in reaching statistical decisions*, Ann. Math. Stat. **22** (1951), 573-580.

KITAGAWA, T.: [1] *Successive process of statistical inferences* (1), Mem. Fac. Sci., Kyusyu Univ., Ser. A, **5** (1950), 139-180.

[2] *Successive process of statistical inferences* (2), Mem Fac. Sci., Kyusyu Univ., Ser. A, **6** (1951), 55-95.

[3] *Successive process of statistical inferences* (3), Mem. Fac. Sci., Kyusyu Univ., Ser. A, **7** (1952), 131-155.

[4] *Successive process of statistical inferences* (4), Bull. Math. Stat. 5 (1952), 35-50.

[5] *Successive process of statistical inferences* (5), Mem. Fac. Kyusyu Univ., Ser. A, **7** (1953), 95-120.

[6] *Analysis of variance applied to function spaces*, Mem. Fac. Sci. Kyusyu Univ., Ser. A, **6** (1951), 41-53.

[7] *Some stochastic considerations concerning the empirical functions of various types*, Bull. Math. Stat. 5 (1953), 19-33.

MIYASAWA, K.: [1] *On the statistical decision functions*, (Doctor Thesis to Kyusyu Univ.), (1952).

PEARSON, E. S.: [1] *"Student" as a statistician*, Biometrika, **30** (1938), 210-250.

WALD, A.: [1] *Statistical decision functions*, John Wiley and Sons, New York (1950).