No. 11

A Formal Approach to the Evaluation of Forecasts

April 1977

Hajime Eto
University of Tsukuba

Abstract

In a forecasting practice there often arises a problem of evaluating the alternative forecastings and determining the best among them on some criterion. A forecasting practice is often sequential in that one forecast is based on the result of another foregoing one. In order to formulate this process of forecasting practice this paper introduces distinctly the notion of a forecasting function as well as that of a resultant forecasting function, the letter being a function (or a functional) of the former. Various operations are analysed to construct the resultant forecasting function for the purpose of the evaluation and determination.

1. INTRODUCTION

The flood of new things raises a problem of evaluating the alternatives and selecting the "best" one among them. A forecasting practice makes no exception in this context. This is extremely important in an organization wherein each member has his own forecast. In operations research a lot of efforts have been devoted to precise calculation techniques of forecastings while in many realistic cases priority is on the organizational design of the process of forecasting practice rather than the calculational precision. As the apparently first attempt in forecasting research context this paper is devoted, if not to design yet, to formulation and partly to systematization of the process of forecasting practice.

Consider a situation in which A makes his decision based on his forecast of the behavior of B. In a realistic case this forecasting is not once for all but is composite in sequence. That is, A first forecasts a basic event B_1 (e.g., level of economic activities, event in 1980 or policy of B) and then its resultant event B_2 (e.g., level of transportation demand, event in 1990 or effect of the policy taken) upon the forecasted information of B_1 as its cause. A forecast is good in a simple case if it fits the reality, but this criterion is of no use when the long future is forecasted and makes no sense when the future may be changed by policy of B. Consequently some operational criterion is needed on which some operation determines the best forecast, say via an optimization.

The above situation may procedurely be summarized as follows.

(i) The main factor or cause a_i which is uncontrollable at least to analyst A is taken by him as a prameter varying in his properly assumed or forecasted domain which may be a range of previous forecast.

(ii) The value of a_i is appropriately forecasted or assumed by a proper method in the domain.

(iii) A certain function or functional is formed from the forecasting function for the forecasted or assumed value of parameter. This is termed a resultant function.

(iv) Criterion K is applied to assessing the forecasts to determine the "best" one.

Each of the above steps is separately not new in forecasting research but the whole process is apparently unexplored yet from an organizational point of view.

Part I discusses a case where the disturbance is negligibly small for the main factor and Part II will discuss a stochastic case where the disturbance is innegligible. The sequel of Part I consists of two sections. §2 introduces the notion of a forecasting function in respect to a given parameter and that of a resultant forecasting function. §3 discusses how to construct resultant forecasting functions upon various criteria via associated operations wherein a criterion is chosen according to the type of parameter.

2. BASIC NOTIONS

Let B be a forecasting object under investigation described by vector $\bar{X}(t) = (x_1(t), x_2(t), \ldots, x_\ell(t))$, where $x_i(t) = s_i(t\ ;\ a_i) + n_i(t)$, i-th characteristic describing a forecasting object under analysis B ($i = 1, 2, \ldots, \ell$), treated further on as a stochastic process. Component $s_i(t\ ;\ a_i)$ of the process $X_i(t)$ stands for main forecasting information depending on the parameter a_i, and $n_i(t)$ is a component of disturbance or noise.

Let $y_i(t\ ;\ a_i)$ be a function forecasting the value of i-th coordinate of vector $\bar{X}(t)$ at the time t in respect to parameter a_i. Let A_i be the domain of a_i, or the set of its estimated values d_i. That is, $d_i \in A_i$. A_i is the range of some forecasting function whose specification is not needed to our investigation. Let us assume in Part I that disturbance n_i is negligible.

Definition 1. A function $y_i(t\ ;\ d_i)$ which forecasts the value of i-th coordinate of vector $\bar{X}(t)$ at the moment t in respect to parameter a_i is determined by formula $y_i(t\ ;\ d_i) \triangleq s_i(t\ ;\ d_i)$.

By forecast of the value of i-th coordinate of vector $\bar{X}(t)$ at the moment t_o with advance T in respect to parameter a_i we understand an estimation of the forecasting function $y_i(t\ ;\ d_i)$ at the moment $t = t_o + T$, that is, $y_i(t_o + T\ ;\ d_i) = s_i(t_o + T\ ;\ d_i)$. It is easy to notice that different values of forecasting function, therefore, different forecasts of a fixed coordinate of vector $\bar{X}(t)$ are obtained for different values of parameter a_i. Hence there arises a problem of determining the best one and this may be solved by building a resultant forecast.

By resultant function forecasting the value of i-th coordinate of vector $\bar{X}(t)$ at the moment t we understand a forecasting function $y_i(t\ ;\ d_i)$ in respect to parameter a_i at a fixed value $a_i = d_i^*$. Value $a_i = d_i^*$ defining a resultant forecasting function $y_i\ (t\ ;\ d_i^*)$ is determined on some fixed criterion.

In order to specify the notion of resultant forecasting function, let us introduce the following symbols:

$\mathcal{A}$: family of sets in some space Y. That is, $\mathcal{A} = \{A : A \sqsubset Y\}$

K_r: $\mathcal{A} \rightarrow W$; univocal operation (fixed criterion),

where $W = \{K_r(A) : A \in \mathcal{A}\} \sqsubset Y$, $r = 1, 2, \ldots, q$. Operation K_r, space

Y and family $\mathcal{A}$ will be specified later.

Definition 2. A resultant function forecasting the value of i-th coordinate of vector $\bar{X}(t)$ at the moment t at a given operation K_r is defined by formula

$$y_i(t\ ;\ d_i^*) \triangleq s_i(t\ ;\ d_i^*)$$

where

$$d_i^* = K_r(A), \quad A \in \mathcal{A}, \quad d_i^* \in W, \quad r = 1, 2, \ldots, q.$$

Value of the resultant forecasting function at the moment $t = t_o + T$, that is, $y_i(t_o + T\ ;\ d_i^*) = s_i(t_o + T\ ;\ d_i^*)$ is called the resultant forecast of the value of i-th coordinate of vector $\bar{X}(t)$ at the moment t with advance T. As we can see, the problem of resultant forecasts reduces itself to determination of some operation K_r (some criterion) defining the value of parameter a_i in a univocal way.

3. SEQUENTIAL COMBINATION OF FORECASTS

3.1 Construction of Resultant Forecasting Function on Criterion of Discrimination

Consider a situation wherein the factor a_i is bi-valued. For example, a policy i is taken or not, a pollution level exceeds a certain limit or not, a conflict exceeds a certain level upto a hostility or not, etc. which are pertinent to strategic decision.

In order to determine a resultant function forecasting the value of i-th coordinate of vector $\bar{X}(t)$ at the moment t, let us denote by:

X_i the observation space of i-th characteristic $X_i(t)$ describing a forecasting object B, composed of vectors $\bar{x}_i = (x_i^{(1)},$

$x_i^{(2)}, \ldots, x_i^{(m)})$, where $x_i^{(1)} = X_i(t_1)$, $x_i^{(2)} = X_i(t_2)$, ..., $x_i^{(m)} = X_i(t_m)$,

$f_i(x_1, x_2, \ldots, x_m \mid a_i)$ the conditional density of a random vector $(X_i(t_1), X_i(t_2), \ldots, X_i(t_m))$ under the condition a_i,

$d_i^{(1)} \varepsilon R$ and $d_i^{(2)} \varepsilon R$ the values of parameter a_i which it can assume respectively with probabilities $p_i^{(1)}$ and $p_i^{(2)}$.

Specifying the criterion K, let $Y \subset R^2$ and $\mathcal{A}$ be a family of sets in Y so that $A_i = \{d_i^{(1)}, d_i^{(2)}\} \varepsilon \mathcal{A}$. In sequel the index r is omitted from K.

Criterion K and the associated operation necessary to obtain a resultant forecast can be determined by means of a univocal division of the observation space X_i into the union of two disjoint sets $A_i^{(1)}$ and $A_i^{(2)}$. This division determines a resultant forecasting function $s_i(t \, ; \, d_i^*)$ in the following way:

$$s_i(t \, ; \, d_i^*) = \begin{cases} s_i(t \, ; \, d_i^{(1)}) & \text{if} \quad \bar{x}_i \varepsilon A_i^{(1)}, \\ s_i(t \, ; \, d_i^{(2)}) & \text{if} \quad \bar{x}_i \varepsilon A_i^{(2)}. \end{cases}$$

Note that:

to different operations (criteria) correspond different methods of division of the observation space X_i into two subsets,

two kinds of errors are connected with each operation:

(k) the error of taking that $s_i(t \, ; \, d_i^*) = s_i(t \, ; \, d_i^{(2)})$ in resultant function forecasting the value of i-th characteristic $X_i(t)$, while in reality $s_i(t \, ; \, d_i^*) = s_i(t \, ; \, d_i^{(1)})$,

(kk) the error of taking that $s_i(t \, ; \, d_i^*) = s_i(t \, ; \, d_i^{(1)})$ in resultant function forecasting the value of i-th characteristic $X_i(t)$, while in reality $s_i(t \, ; \, d_i^*) = s_i(t \, ; \, d_i^{(2)})$.

Let α_i and β_i be the probability of error indicated in (k) and (kk) respectively.

Then

$$\alpha_i = \int_{A_i^{(2)}} f_i\,(x_1, x_2, \ldots, x_m \mid d_i^{(1)})\, dx_1\, dx_2 \ldots dx_m$$

$$\beta_i = \int_{A_i^{(1)}} f_i\,(x_1, x_2, \ldots, x_m \mid d_i^{(2)})\, dx_1\, dx_2 \ldots dx_m$$

where $A_i^{(1)} \subset X_i,\ A_i^{(2)} \subset X_i,\ X_i = A_i^{(1)} \cup A_i^{(2)},\ A_i^{(1)} \cap A_i^{(2)} = \emptyset$

General probability of error can be expressed by formula

$$b_i = p_i^{(1)}\, \alpha_i + p_i^{(2)}\, \beta_i\ .$$

In order to determine the operation K the observation space X_i is divided into the sets $A_i^{(1)}$, $A_i^{(2)}$ using for example Kotelnikov criterion*). Hence, minimizing the general probability of error

$$b_i = p_i^{(1)} \int_{A_i^{(2)}} f_i\,(x_1, x_2, \ldots, x_m \mid d_i^{(1)})\, dx_1\, dx_2 \ldots dx_m$$

$$+\ p_i^{(2)} \int_{A_i^{(1)}} f_1\,(x_1, x_2, \ldots, x_m \mid d_i^{(2)})\, dx_1\, dx_2 \ldots dx_m,$$

we obtain

(1)
$$A_i^{(2)} = \{ (x_1, x_2, \ldots, x_m):\ p_i^{(2)} f_i\,(x_1, x_2, \ldots, x_m \mid d_i^{(2)}) - p_i^{(1)} f_i\,(x_1, x_2, \ldots, x_m \mid d_i^{(1)}) > 0 \}$$

$$A_i^{(1)} = X_i - A_i^{(2)}$$

---

*) See e.g. [1, 4].

A resultant forecasting function can also be determined using other criteria defining the operation K. Those criteria can be for example maximum of likelihood, maximum of probability a posteriori criteria as well as the Neyman-Pearson criterion and its variants.

Such a division of the observation space X_i into sets $A_i^{(1)}$ and $A_i^{(2)}$ suffices to determine the operation K, which in turn defines in a univocal way the value of parameter a_i. And indeed, according to formula (1), if the observation $\bar{x} = (x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(m)})$ being a point in space X_i satisfies the condition $p_i^{(2)} f_i (x_1, x_2, \ldots, x_m \mid d_i^{(2)}) + p_i^{(1)} f_i (x_1, x_2, \ldots, x_m \mid d_i^{(1)}) > 0$, then $\bar{x}_i \in A_i^{(2)}$, hence a resultant forecasting function has a form $s_i(t ; d_i^*) = s_i(t ; d_i^{(2)})$. In the opposite case, that is, if $\bar{x}_i \in A_i^{(1)}$, a resultant forecasting function has a form $s_i(t ; d_i^*) = s_i(t ; d_i^{(1)})$.

A resultant forecasting function determined in this way immediately determines a resultant forecast.

3.2 Construction of Resultant Forecasting Function on Criterion of Minimal Risk

Let now

$$X_i(t) = s_i(t ; a_i) + n_i(t), \text{ where } a_i = [a_{i1}, a_{i2}, \ldots, a_{im}]^T$$

be a vector of parameters describing the main forecasting information s_i with a_{ij} being an unknown fixed number $(j = 1, 2, \ldots, m)$, that is, $a_{ij} \in R$,

$d_i^{(k)} = [d_{i1}^{(k)}, d_{i2}^{(k)}, \ldots, d_{im}^{(k)}]^T$ $(k = 1, 2)$ be a m-vector of random variables denoting the estimations of parameter a_i,

Y be a set of random variables of 2m dimensions,

$\mathcal{A}$ be a set of pairs of random variables,

W be a set of random variables.

Note that $A_i = \{ d_i^{(1)}, d_i^{(2)} \} \in \mathcal{A} \subset Y$

Let us denote by $C_i^{(k)} = d_i^{(k)} - a_i = [c_{i1}^{(k)}, c_{i2}^{(k)}, c_{im}^{(k)}]^T$ error of estimation, and by

$$\Lambda_i^{(k)} = E\{C_i^{(k)} (C_i^{(k)})^T\}$$

$$= \begin{bmatrix} E\{c_{i1}^{(k)} \cdot c_{i1}^{(k)}\} & E\{c_{i1}^{(k)} \cdot c_{i2}^{(k)}\} & \cdots & E\{c_{i1}^{(k)} \cdot c_{im}^{(k)}\} \\ E\{c_{i2}^{(k)} \cdot c_{i1}^{(k)}\} & E\{c_{i2}^{(k)} \cdot c_{i2}^{(k)}\} & \cdots & E\{c_{i2}^{(k)} \cdot c_{im}^{(k)}\} \\ \cdots & \cdots & \cdots & \cdots \\ E\{c_{im}^{(k)} \cdot c_{i1}^{(k)}\} & E\{c_{im}^{(k)} \cdot c_{i2}^{(k)}\} & \cdots & E\{c_{im}^{(k)} \cdot c_{im}^{(k)}\} \end{bmatrix}$$

matrix of covariance of the error of k-th estimation for i-th characteristic describing a forecasting object under investigation B.

Note that, according to Definition 1, to different estimations $d_i^{(k)}$ $(k = 1, 2)$ of the vector of parameters a_i correspond different forecasting functions $s_i(t\,;\,d_i^{(1)})$, $s_i(t\,;\,d_i^{(2)})$, therefore, different forecasts $s_i(t_o + T, d_i^{(1)})$, $s_i(t_o + T, d_i^{(2)})$ of i-th characteristic s_i at the moment t_o with advance T.

By resultant function forecasting the value of i-th coordinate of vector $\bar{X}(t)$ at the moment t with a given operation K (see Definition 2) we understand a function $s_i(t\,;\,d_i^*)$, where d_i^* is the best (in the sense of operation K) linear combination in the case of two independent estimations $d_i^{(1)}$ and $d_i^{(2)}$.

Let us proceed now to the construction on a criterion of a risk function, considered in the theory of recursive estimators, in the following form (see [2, 3])

(2)
$$R(a_i) = (d_i^{(1)} - a_i)^T (\Lambda_i^{(1)})^{-1} (d_i^{(1)} - a_i) + (d_i^{(2)} - a_i)^T (\Lambda_i^{(2)})^{-1} (d_i^{(2)} - a_i).$$

Vector d_i^* minimizing a risk function (2) has a form

$$(3) \qquad d_i^* = [(\Lambda_i^{(1)})^{-1} + (\Lambda_i^{(2)})^{-1}]^{-1} [(\Lambda_i^{(1)})^{-1} d_i^{(1)} + (\Lambda_i^{(2)})^{-1} d_i^{(2)}] .$$

And indeed, acting with a matrix operator of partial differential coefficients defined by formula

$$\nabla_{a_i}(\bar{y}) = \begin{bmatrix} \frac{\partial}{\partial a_{i1}} \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial}{\partial a_{im}} \end{bmatrix} ([y_1, y_2, \dots, y_m]) \triangleq \begin{bmatrix} \frac{\partial y_1}{\partial a_{i1}} & \frac{\partial y_2}{\partial a_{i1}} & \dots & \frac{\partial y_m}{\partial a_{i1}} \\ \frac{\partial y_1}{\partial a_{i2}} & \frac{\partial y_2}{\partial a_{i2}} & \dots & \frac{\partial y_m}{\partial a_{i2}} \\ \cdot & \cdot & \cdot & \cdot \\ \frac{\partial y_1}{\partial a_{im}} & \frac{\partial y_2}{\partial a_{im}} & \dots & \frac{\partial y_m}{\partial a_{im}} \end{bmatrix}$$

on the square form (2) and using the property

$$\nabla_{a_i}(\bar{x} \cdot \bar{z}) = \nabla_{a_i} \left([x_1, \dots, x_m] \begin{bmatrix} z_1 \\ z_2 \\ \cdot \\ \cdot \\ \cdot \\ z_m \end{bmatrix} \right) = \{\nabla_{a_i}(\bar{x})\} \bar{z} + \{\nabla_{a_i}(\bar{z}^T)\} \bar{x}^T$$

and the fact that matrix $(\Lambda_i^{(k)})^{-1}$ $(k = 1, 2)$ is a positive semi-definite symmetric matrix, we obtain

$$\nabla_{a_i}(R(a_i)) = 2\{\nabla_{a_i} (d_i^{(1)} - a_i)^T\} (\Lambda_i^{(1)})^{-1} (d_i^{(1)} - a_i)$$

$$+ 2\{\nabla_{a_i} (d_i^{(2)} - a_i)^T\} (\Lambda_i^{(2)})^{-1} (d_i^{(2)} - a_i)$$

$$= 2(-1)^m [(\Lambda_i^{(1)})^{-1} (d_i^{(1)} - a_i) + (\Lambda_i^{(2)})^{-1} (d_i^{(2)} - a_i)].$$

Hence it is easy to notice that vector d_i^* minimizing a risk function (2) is a solution of equation $\nabla_{a_i} (R(d_i^*)) = 0$. In this way the operation K (criterion) permitting to define a resultant forecasting function $s_i(t\ ;\ d_i^*)$, and therefore a resultant forecast $s_i(t_o + T\ ;\ d_i^*)$ at the moment t_o with advance T of i-th characteristic s_i has been determined.

Finally, let us draw the attention to the fact that, in order to determine vector d_i^* defining a resultant forecasting function, it is necessary to make three operations of matrix inversion. This process is labour-consuming and may make the calculations difficult, but it can be avoided if we reduce formula (3) to the form

$$d_i^* = L_i^{(1)} d_i^{(1)} + L_i^{(2)} d_i^{(2)} - (L_i^{(1)} M_i - L_i^{(2)} N_i)(M_i + N_i)^{-1} (d_i^{(1)} - d_i^{(2)}),$$

where $L_i^{(1)}$ and $L_i^{(2)}$ are any matrices such that $L_i^{(1)} + L_i^{(2)} = I$, (I: identity matrix), and for any non-singular matrix G_i, matrices M_i and N_i can be determined from relation $\Lambda_i^{(1)} = M_i G_i^{-1}$, $\Lambda_i^{(2)} = N_i G_i^{-1}$, requiring only one matrix inversion (see [2, 3]).

As is seen above, our logic in the process of forecasting practice somewhat resembles that in statistical inference and decision theory. The latters as an exact science specify a criterion for evaluation of inference or decision in seeking for rigidity while a forecasting practice specifies a criterion in seeking for consensus in organization. Statistical inference or decision theory seeks for its systematization for mathematical clearness while a forecasting practice in organization needs its systematization for an organization effectiveness. This difference in purpose demands a newly unified notion of resultant forecast so that the hierarchy of a forecast and a resultant forecast corresponds to the organizational hierarchy of decision levels. In this sense this notion is quite organization oriented.

4. CONCLUSION

A new notion of a resultant forecast is formulated and its construction is discussed in an operational manner from an organizational science point of view. Part II to follow will discuss a stochastic case.

REFERENCES

[1] Anderson, R., INTRODUCTION TO MULTIVARIATE STATISTICAL ANALYSIS, John Wiley, 1958.

[2] Blackman, R. B., Methods of Orbit Refinement, Bell System Tech. Journ. 43, 1964, 885-909.

[3] Claus, A. I., Blackman, R. B., Halline, E. G., and Ridgeway, W. C., III, Orbit Determination and Prediction and Computer Programs, Bell System Tech. Journ. 42, 1963, 1357-1382.

[4] Feldbaum, A. A., OSNOVY TEORII OPTIMALNYKH SISTEM, Moscow, 1966.