POLSKA AKADEMIA NAUK INSTYTUT BADAŃ SYSTEMOWYCH

PROCEEDINGS OF THE 3rd ITALIAN-POLISH CONFERENCE ON APPLICATIONS OF SYSTEMS THEORY TO ECONOMY, MANAGEMENT AND TECHNOLOGY

WARSZAWA 1977

Redaktor techniczny Iwona Dobrzyńska

1. i

...

Korekta Halina Wołyniec

POLSKA AKADEMIA NAUK INSTYTUT BADAŃ SYSTEMOWYCH

PROCEEDINGS OF THE 3rd ITALIAN-POLISH CONFERENCE ON APPLICATIONS OF SYSTEMS THEORY TO ECONOMY, MANAGEMENT AND TECHNOLOGY

BIAŁOWIEŻA, POLAND MAY 26-31, 1976

EDITED BY J. GUTENBAUM

WARSZAWA 1977

Redaktor techniczny Iwona Dobrzyńska

Korekta Halina Wołyniec



Ir inw. IBS PAN 32959

NAME: STATE



While the first two parts are or tather homogeneous observate, the ment pair contains the capera concerning the different types of models — for the economic fushinological, management and data processing sympton.

J. Gren

Central School of Planning and Statistics, Warsaw

ON CERTAIN APPLICATION OF STATISTICAL GAMES THEORY TO QUALITY CONTROL

1. INTRODUCTION

It is very well known that the whole set of methods which we call Statistical Quality Control, may be divided into two main groups. Quality control during production, and quality control after production, i.e. during collecting the completed products. In this paper one method of the last group will be presented. Very often, the main aim of quality control after production is to classify the controlled goods into some sorts of quality. When many important, measurable characters of a particular manufactured product are under control, the direct decision which should be taken for each unit of product during statistical quality inspection, may be difficult and too expensive. Some kind of linear index of quality, which takes values in one-dimensional space can be used instead.

The Bayes method for finding such linear index of quality is proposed and discussed in details in this paper. Generally speaking, this method is based on solution of some statistical game, which can be built for the Statistical Quality Inspection problem. By using the Bayes strategy in this statistical game, we obtain weights for linear index of quality, which may be a very useful tool in the multi-dimensional Quality Control.

2. SPECIFICATION OF THE MULTI-DIMENSIONAL QUALITY CONTROL PROBLEM

We are dealing with rather complicated product, which is produced in a great number of items using the same technology. A special staff is employed in the factory to control the quality of each finished unit. Many specific characters of our product are investigated and an adequate decision, namely, the classification of controlled items of product into some sorts of quality should be taken. Because of complexity of product, its quality is now expressed by a fixed set of value of investigated, measurable variables. Each variable

306

represents one specific feature of our article, and may take its value from a known finite interval. Assuming we have "r" such variables: let $P_j(j=1, 2, ..., r)$ denote the value of variable "j", and $P_j \in (a_j, b_j)$, where a_j and b_j are known numbers given by technical experts. Then, the vector $Q = (P_1, P_2 ... P_r)$ can express the quality of our product.

It will be convenient to normalize each value P_j in order to obtain a unit interval.

So, if we make the normalization:

$$x_j = \frac{P_j - a_j}{b_j - a_j}$$
 or $x_j = \frac{b_j - P_j}{b_j - a_j}$

where having any of these formulas depends on which of these two values X_j has increasing tendency to 1, which indicates the highest quality, then we have

$$X_i \in [0, 1]$$
 for $j = 1, 2, ..., r$

Hence, during performing control, we can measure value P_j for all r characters of the controlled items, and know at the same time, values $X_j(j=1,2,...,r)$ for this item.

Now, we may introduce the important following definition:

Definition

The normalized vector $\Theta = (X_1, X_2, ..., X_r)$ is called the level of quality of controlled items, if each X_j for j = 1, 2, ..., r is obtained (from an adequate P_j) during Quality Inspection of this item of product.

Let us make now the following basic assumptions:

Set of basic assumptions

1. We assume that Θ is a random vector, which takes values in *r*-dimensional space, denoted by Ω . So, $\Theta \in \Omega$ and Ω is, of course, the *r*-dimensional unit cube.

2. We assume that each random variable X_j with values in the unit interval [0, 1], has some continuous probability distribution, generated by the homogeneous technological process in which our article is produced.

3. We assume that the random variables X_j for j = 1, 2, ..., r are independent. It means that the set of r different characters of controlled product, should be choosen as a set of independent variables.

4. We assume that the main aim of Quality Control is to classify controlled items into 3 sorts: the best one — I sort (the highest quality), II sort and III sort (the lowest quality of product). Such situation is typical for example in the light and food industry, however if we need, we may use only 2 or 4 and more different sorts.

5. We assume that we have fixed definition for each sort of quality, as some combinations of values P_j for r characters of product according to consumers' demand.

It means, that we have some rules or regulations to partition the r-dimensional unit cube Ω into 3 separate subsets, i.e.:

 $\Omega = \Omega_1 \cup \Omega_2 \cup \Omega_3$, where $\Omega_k \cap \Omega_l \neq \emptyset$, for $k \neq l = 1, 2, 3$.

If $\Theta \in \Omega_1$ we say, that our product is in the I sort of quality. If $\Theta \in \Omega_2$ we say, that our product is in the II sort of quality. If $\Theta \in \Omega_3$ we say, that our product is in the III sort of quality.

6. We assume, that different sorts of quality have different prices (it means higher profit of producer for higher quality of product), and error in the right classification gives some kind of loss for the producer.

It is obvious, that only for very small number of variables X_j (elements of the vector Θ) the direct method of classification in the *r*-dimensional space can be used. For each controlled item, the direct method of classification which utilizes definitions of each sort of quality, need searching in which subset Ω_1 , Ω_2 or Ω_3 the vector Θ — level of quality lies. However, if the vector Θ contains a great number of variables, for instance 5 or more, such direct method of classification of controlled items of product is going to be slow, difficult and then too expensive at 100 percent Quality Control and Statistical Control as well. Indeed, in multi-dimensional Quality Control, the controller, who has to classify each item of product into some sort of quality, should have another indirect but more easy tool. Such tool (for instance a physical or mathematical model) should still utilize all values X_j , but in order to make control quicker and easier, it should reduce (or transform) the *r*-dimensional set Ω to another simple set.

We propose, as such tool, a Linear Index of Quality (LIQ) which will be denoted by J, and is in the form:

$$J = \sum_{j=1}^{r} \alpha_j x_j, \quad \text{where} \quad \alpha_j > 0, \ \sum_{j=1}^{r} \alpha_j = 1$$

The Linear Index of Quality -J is, in fact, convex linear combination of all elements of the normalized vector Θ , and takes value in the one-dimensional unit interval [0, 1], which is easy to search in. The main goal is now, how the set of weights $\alpha_1, \alpha_2, ..., \alpha_r$ should be choosen in order to utilize LIQ-J instead of direct classification of controlled items of our product.

Among possible criteria which can be considered, in order to obtain the set of weights for LIQ-J, the economic consequencies of errors in classification during quality control, seems to be very important.

Considering economic viewpoint of decision problem (classification) in the quality control of our product, and taking into account the stochastic character of LIQ-(J is random variable, because it is a function of random variables X_j), we conclude that it is possible to adopt the statistical games theory in order to obtain the optimal set of weights for LIQ-J.

Before we do it, let us review the most important definitions in statistical games theory and introduce notation, which we shall use (see [1], [2]).

3. MAIN DEFINITIONS AND NOTATIONS IN STATISTICAL GAMES

We consider the game with two players:

I — Nature, who chooses the state Θ , from the known states space Ω . II — Statistician, who chooses the decision a, from the known decisions space A.

Statistician can measure all consequences of his decision a, when the true state of nature is Θ , using a loss function $L(\Theta, a)$, which is real valued function defined on the Cartesian product $\Omega \times A$.

The triple (Ω, A, L) is called a primary strategic game in each statistical decision problem.

However, before taking his decision, statistician may perform some statistical experiment, that allows him to observe the realization X of such random variable, which has known conditional distribution $F(x|\Theta)$. That realization x, gives him some indirect information about the state Θ , which had been choosen actually by the Nature.

Let us denote by X, the sample space in the experiment performed by statistician.

A function $d: X \to A$, is called the stististical decision function. Now, this function d(x) = a is the strategy for statistician, who chooses it from the set D of all possible such functions d.

But the loss function $L(\Theta, a)$, is now a random variable because through the decision function d, decision "a" is taken while observing the random variable X.

Therefore, we have to define the risk function as

$$R(\Theta, d) = E_{\Theta}L(\Theta, a) = \int_{\Theta} L(\Theta, a) dF(x|\Theta)$$

Now, we have the fundamental definition of statistical game. The triple (Ω, D, R) is called the statistical game, or the game against Nature with utilization of additional statistical information.

Of course, statisticians want to find and use such optimal strategy $d \in D$ which minimizes the risk function $R(\Theta, d)$ in the whole states space Ω , but such strategy very often does not exist.

For this reason, two main rules (Bayes and minimax) are suggested, in order to obtain sub-optimal solution in the statistical game (Ω, D, R) . When the statistical decision problem, described as a statistical game, is going to be repeated many times, the best solution for statistical game is obtained by means of Bayes rule.

It can be proved, that the Bayes solution does not need randomization in the set D, and that it is admissible. In order to find Bayes strategy, we shall assume, that the state of Nature Θ is a random variable, and has the so-called prior distribution $G(\Theta)$. That prior distribution represents our knowledge about state of Nature, before we will obtain a new information from the sample, i.e. observation x. Using prior distribution, we can take the expectation of the risk function $R(\Theta, d)$ and define the Bayes risk as:

$$r(G, d) = E_d R(\Theta, d) = \int_{\Omega} R(\Theta, d) dG$$

Finally, we define the Bayes decision function, (or strategy), say $d_0 \in D$ for the statistical game (Ω, D, R) , with respect to the prior distribution $G(\Theta)$, which minimizes the Bayes risk r(G, d), that is

$$r(G, d_0) = \min_{d \in D} r(G, d)$$

Sometimes it may be more convenient to find the Bayes strategy, using the posterior distribution $G(\Theta|x)$, which may be calculated from the prior distribution $G(\Theta)$ and conditional distribution $F(x|\Theta)$, but in our case, the definition of Bayes decision function is quite useful to operate it. Let us now return to our problem of quality, controlled items of some product.

4. SOLUTION OF THE PROBLEM

In order to obtain the optimal system of weights $\alpha_j (j=1, 2, ..., r)$ for the linear Index of Quality: $J = \sum_{j=1}^{r} \alpha_j x_j$, we shall build the adequate statistical game concerning the problem of quality control and find the optimal (bayes) strategy in this game. First of all, let us define the structure of suitable primary strategic two-person game (Ω, A, L) .

Let the first player in our game will be a Nature, which we may interpret as the whole technological process of production of some particular product. The simple strategy of Nature is just the level of quality of each item of product, namely the vector $\Theta = (x_1, x_2, ..., x_r)$, where each $X_j \in [0, 1]$ is obtained from an adequate p_j , the value of controlled variable *j*. Let $\Theta \in \Omega$, which is *r*-dimensional unit cube, containing all possible levels of quality Θ . Ω is the states space in our game. This space is divided into 3 separate subsets, according to the basic assumption no. 5 on definitions of sort I, sort II and sort III of quality ($\Theta \in \Omega_1$ or $\Theta \in \Omega_2$ or $\Theta \in \Omega_3$). Let the second player in our game will be a Statistician (the controller in quality control of our product) who wants to classify each item of our product into one of the three sort of quality. The set of possible decisions of our controller is $A = \{a_1, a_2, a_3\}$, where a_1 is decision to quality the item as the I sort, a_2 is decision to quality.

So we have got 2 of 3 main elements of our strategic game, and now we shall construct the loss function $L(\Theta, a)$, which is the pay-off function, i.e. the last element in the game (Ω, A, L) .

Of course, we want to have a consistency between decision given by controller and the real sort of quality of controlled items.

According to the basic assumption no. 6, we may take into account all economic consequences of errors in the sort-classification, and built the loss

function, based on the accounting records in the enterprise where our article is produced.

decision			
actual sort	<i>a</i> 1	<i>a</i> ₂	a ₃
$\boldsymbol{\varTheta}\in\boldsymbol{\varOmega}_1$	0	l_1	<i>l</i> ₂
$\varTheta\in \varOmega_2$	k_1	0	<i>l</i> ₃ .
$\varTheta\in \varOmega_3$	<i>k</i> ₂	<i>k</i> ₃	0

The values of such loss function $L(\Theta, a)$ are in the table

The numbers l_1 , l_2 and l_3 are just positive losses for the enterprise, which produce the item of higher sort of quality, but because of error in classification during control, the item of article can be sold at a lower price, as the lower sort of quality. For example, l_1 — is the difference between profit from selling 1 item of I sort, which might be obtained by the enterprise, and the actual profit from selling this item at the lower price for II sort, and so on.

But there exist also losses (denoted by k in the above table) which appear when the error of classification has been done in the opposit direction. In order to protect consumers from purchasing article which has been qualified as better or the best sort, but in fact, its quality is worse, there exist some regulations in the law, which force the producer to pay all cost connected with guarantees, reparations or replacement by another item, if the consumer claims it. So, the positive numbers k_1 , k_2 and k_3 , as values of above loss function, can be evaluated (from the accounting records) as such kind of cost for enterprise.

Construction of the loss function $L(\Theta, a)$ in such manner completed the whole structure of strategic game (Ω, A, L) and we may now transform it into statistical game (Ω, D, R) using additional stratistical information about state of Nature, i.e. about vector Θ — the real level of quality.

Indeed, we have got some information about state of Nature Θ , transforming P_j into $X_j \in [0, 1]$, for controlled item of product. We do not use the direct method of classification of each item into one sort of quality, because it is difficult and costly for the great r. Having information about all elements X_j of the vector Θ , we still do not know, in which part (subset Ω_1, Ω_2 or Ω_3) of the unit cube, the vector Θ excatly lies. But now, the controller can use some kind of simple tool, which help him to take decision about the real sort of quality of controlled item. This tool will be — from statistical decision point of view — a statistical decision function d, which transform observations X_j (j=1, 2, ..., r) into decision $a \in A$.

As such statistical decision function, we propose the Linear Index of Quality J, defined in section II of this paper, as

$$J = \sum_{j=1}^{r} \alpha_j x_j, \quad \text{where} \quad \alpha_j > 0, \ \sum_{j=1}^{r} \alpha_j = 1$$

LIQ-J takes values from the unit interval [0, 1].

Let us denote A = [0, 1], and we split this unit interval A into 3 separate intervals $A_1, A_2, A_3(A = A_1 + A_2 + A_3)$. This division (corresponding to Ω_1, Ω_2 and $\Omega_3 - r$ -dimensional subsets of Ω) can be made arbitrarily, but we will see at the end, how we may choose the best split of unit interval A into A_1, A_2, A_3 .

So a proposed statistical decision function, based on LIQ-3, is defined as follows:

If the value $J \in A_1$, then controller takes the decision a_1 (qualifying the item to the I sort)

If $J \in A_2$, the decision a_2 is taken (qualifying the items to the II sort)

If $J \in A_3$, then the decision a_3 is taken (qualifying the item to the III sort).

Of course, LIQ-3 as a decision function, differs one from another, when we change the set of weights $(\alpha_1, \alpha_2, ..., \alpha_r)$. For this reason, we can write one decision function, as $d = (\alpha_1, \alpha_2, ..., \alpha_r)$.

Let D denote the set of all possible such decision functions d.

Therefore, the choice of optimal decision function d from the set D, is in fact the choice of optimal set of weights α_j for the LIQ-3, used in multidimensional quality control of our product.

Now- to complete the structure of statistical game (Ω, D, R) for our problem, we have to calculate the risk function $R(\Theta, d) = R(\Theta, J)$.

In order to obtain the risk function $R(\Theta, J)$ according to the definition in section 3 of this paper, we must know the conditional probability distributions of random variable J, given $\Theta \in \Omega_1$, $\Theta \in \Omega_2$ and $\Theta \in \Omega_3$.

Since J is a function of random variables X_j , these conditional distributions can be obtained from distributions of random variables X_j .

So, we need for each continuous (see assumption 2) random variable X_j its conditional probability density function

 $f_j^i(\mathbf{x}_i | \Theta \in \Omega_i)$ for j = 1, 2, ..., r and i = 1, 2, 3.

There are 3. r of such density functions. We can estimate them in the following way.

Let us assume, that before we try to find an easy method for classification of our product, in the enterprise had been performed, for some past period of time, the direct method of classification for a multidimensional searching (it means using definition of each kind of sort of quality in multidimensional unit cube Ω). So, we may have records of results of this direct qualification of our product and from those previous statistical data we can estimate not only 3.r conditional density functions $f_i^i(x_j | \Theta \in \Omega_i)$, but also prior distribution for Θ , using the division Ω into subsets $\Omega_1, \Omega_2, \Omega_3$, i.e. prior probabilities $P(\Theta \in \Omega_1) = p_1, P(\Theta \in \Omega_2) = p_2$ and $P(\Theta \in \Omega_3) = p_3$. We may expect, that the conditional distributions of X_j , with p.d.f. denoted by $f_j^i(x_j | \Theta \in \Omega_i)$, will be mainly from a beta distribution system, with p.d.f. defined as:

$$f(x) = \frac{1}{B(p, q)} x^{p-1} (1-x)^{q-1}$$

where $0 \leq X \leq 1$, p, q > 0.

These conditional distributions will have, of course, different values of parameters p and q for each condition $\Theta \in \Omega_1$ and for each random variable X_j , but the typical situation will be probably like these graphs: $(f^1, f^2 \text{ and } f_3 \text{ denote p.d.f. of } X_j \text{ given } \Theta \in \Omega_1, \Theta \in \Omega_2 \text{ and } \Theta \in \Omega_3 \text{ respectively})$



Since we have restricted the type of these distribution to the beta type, we can estimate its parameters p and q rather easily.

It is well know, that all important moments of the beta distriution are just simple functions of these two parameters p and q.

For the expected value we have $E(x) \varnothing = \frac{p}{p+q}$, and for the variance of the

beta distribution we have $D^2(x) = \frac{pq}{(p+q)^2(p+q+1)}$

From the sample statistical data which are results of the previous direct classification of controlled items of our product, we may estimate (by method of moments) the expected values and variance as $E(X) \approx \bar{x}$, $D^2(X) \approx S^2$ and solving the above two equations, we can obtain estimates for parameters p and q, which gives us the complete form of p.d.f. for the beta distribution.

Now, we come to another important problem. In order to calculate the risk function $R(\Theta, J)$ for our statistical game, we must find 3 conditional probability density functions of the random varibale $J = \sum_{j=1}^{r} \alpha_j x_j$, knowing 3.r p.d.f's $f_i^i(x_j | \Theta \in \Omega_i)$.

Let us denote those p.d.f's by $f^i(J|\Theta \in \Omega_i)$ for i = 1, 2, 3. To obtain such 3 p.d.f's of the random variable J, which is simple linear function of r independent random variables X_j , we may utilize the theorem for the direct transformation of set random variables(with the Jacobian) or we may, alternatively utilize the tool of characteristic functions $\varphi_x(t) = Ee^{itx}$

Considering the linearity of J, its characteristic function will be:

 $\varphi_J(t) = \prod_{j=1} \varphi_{x_j}(\alpha_j t)$, where φ_{x_j} is the characteristic function of the

above mentioned known conditional distribution of the random variable X_j . Now, by utilization of the inversion theorem for p.d.f. from the characteristic function, we obtain $f^i(J|\Theta \in \Omega_i)$ for i=1, 2, 3. These functions will depend, of course, on the unknown parameters α_j , the weights in LIQ-J, which we are looking for.

Finally, we can calculate the risk function $R(\Theta, J)$ for our problem, as follows:

for
$$\Theta \in \Omega_1$$
: $R^1(\Theta, J) = l_1 \cdot P \{J \in A_2 | \Theta \in \Omega_1\} + l_2 \cdot P \{J \in A_3 | \Theta \in \Omega_1\}$
for $\Theta \in \Omega_2$: $R^2(\Theta, J) = k_1 \cdot P \{J \in A_1 | \Theta \in \Omega_2\} + l_3 \cdot P \{J \in A_3 | \Theta \in \Omega_2\}$
for $\Theta \in \Omega_3$: $R^3(\Theta, J) = k_2 \cdot P \{J \in A_1 | \Theta \in \Omega_3\} + k_3 \cdot P \{J \in A_2 | \Theta \in \Omega_3\}$

The probabilities $P\{J \in A_i | \Theta \in \Omega_k\}$ for $i \neq k = 1, 2, 3$ in these risk functions can be obtained by the integration of respective p.d.f. $f^i(J | \Theta \in \Omega_i)$.

Therefore, we have got now the whole structure of our statistical game (Ω, D, R) constructed for the classification problem in the discussed multidimensional quality control, and we may solve that game.

Since our statistical decision problem, i.e. classification the controlled items of our product is going to be repeated many times using the LIQ-J, the Bayes solution of our statistical game is appropriate.

Our prior distribution $G(\Theta)$ which is necessary to obtain the Bayes strategy, can be estimated from results of the previous direct qualification of our product, as probabilities $p_i = P(\Theta \in \Omega_i)$ for i = 1, 2, 3.

It allows us to calculate the Bayes risk in our statistical game, as:

$$r(G, J) = ER(\Theta, J) = R^{1}(\Theta, J) p_{1} + R^{2}(\Theta, J) p_{2} + R^{3}(\Theta, J) p_{3}$$

The full form of this Bayes risk is as follows:

$$r(G, J) = p_1 l_1 \int_{A_2} f^1(J | \Theta \in \Omega_1) dJ + p_1 l_2 \int_{A_3} f^1(J | \Theta \in \Omega_1) dJ +$$

+ $p_2 k_1 \int_{A_1} f^2(J | \Theta \in \Omega_2) dJ + p_2 l_3 \int_{A_3} f^2(J | \Theta \in \Omega_2) dJ +$
+ $p_3 k_2 \int_{A_1} f^3(J | \Theta \in \Omega_3) dJ + p_3 k_3 \int_{A_3} f^3(J | \Theta \in \Omega_3) dJ$

Bayes risk r(G, J) is a function of unknown weights α_j of LIQ-J, and to find the optimal Bayes solution of our statistical game, which will give us the

314

to optimal set of weights α_j , we shall now minimize the function r(G, J) = r(d) with respect J, i.e. with respect to the vector $d = (\alpha_1, \alpha_2, ..., \alpha_r)$. It means, that we shall find such vector d_0 , which fulfills the equations:

$$r(\boldsymbol{d}_0) = \min_{\boldsymbol{d} \in D} r(\boldsymbol{d})$$

Of course, because of the restrictions $\alpha_j > 0$ and $\sum_{j=1}^{r} \alpha_j = 1$, we have to apply here the conditional minimalization of the Bayes risk r(d) = r(G, J), By applying such a minimalization, the optimal vector $d_0 = (\alpha_1, \alpha_2, ..., \alpha_r)$, will give us the optimal LIQ-J which can be used later in multidimensional quality control of our product, to make such control easier and cheaper than the direct classification into sorts of quality.

We have solved our problem using an arbitrary division of the unit interval A into subsets A_1 , A_2 and A_3 . Now, we see, that each such division gives us different value $r(d_0)$ of the minimal Bayes risk, and using computer we can easily find the optimal division of interval A, which gives min $(\min r(d))$

The proposed method of finding the Linear Index of Quality — J, will give us such a linear function J, and the control maker can use it as a cheap and easy tool in the multidimensional quality control to classification of controlled items into some sorts of quality. This tool assures and ascertains a minimal level of expected losses caused by errors in classifications, which is important feature from the economic point of view.

However, the way of obtaining the LIQ-J is rather complicated but we think that it may be a helpful tool in many factories. The outline of flow-diagram of proposed method for obtaining LIQ-J is as follows:





REFERENCES

- J. GREN Gry statystyczne i ich zastosowanie, PWE, Warszawa 1972 (in Russian 1975, Moscow).
- [2] A. WALD Statistical Decision Functions, J. Wiley, New York 1950.

SUMMARY

The method for classification of controlled products into some sorts of quality is suggested. This method is based on certain linear indexes of quality with weights obtained by solving an appropriate statistical game using the Bayes strategy.

The method proposed may be useful in many-dimensional Quality Control problems.



and acts in the second se There are the state of the second

Instytut Badań Systemowych PAN Nakład 300 egz. Ark. wyd. 25,0. Ark. druk. 23,75. Papier druk. sat. kl. III 80 g 61×86. Oddano do składania 8 X 1976 Podpisano do druku w sierpniu 1978 r. Druk ukończono w sierpniu 1978 roku

CDW - Zakład nr 5 w Bielsku-Białej zam. 62/K/77 J-124



