On the Quantification of Qualitative Data from the Mathematico-Statistical Point of View

(An Approach for Applying this Method to the Parole Prediction)

By Chikio HAYASHI

(Received June 5, 1950)

I. Introduction

There have been some approaches made its aim being to grasp scientifically the social and psychological phenomena, by quantifying the qualitative data, but there are few that rest on the theoretical foundations reliable from the methodological point of view. Moreover, the quantities given to qualitative data by these approaches are quite optional.

Quantification should be done only for the purpose of solving the concrete problems. In other words, quantification should be made from the best point of view and by the most reasonable means that may answer our purpose, as we wish either to acquire some reasonable knowledge on something or to make reasonable, effective and positive criteria how we have to act or behave ourselves in managing some affairs. Moreover it also must be proved from a scientific standpoint that quantifications made can give useful conclusions. Thus it must be said that quantification has no absolute meaning, but only relative and functional meaning to our purpose.

Remarkable results concerning quantification problems have been obtained by Mr. Guttman and Mr. von Neumann; scale analysis, intensity analysis, paired comparison, and theory of games. In Japan, there have appeared a few interesting researches concerning the quantification problems. In this paper, the quantification problem is considered from the point of maximizing the precision of prediction of social phenomena in the sense of the theory of probability. The prediction of the social prognoses of criminals (the prediction of parole outcome) has been taken up as an example. The general and theoretical considerations and the practical method of quantification will be dis cussed below.

II. Parole Prediction

There is a regime that a criminal who has finished more than one-third of his term may be paroled when he is fairly educated and is recognized, as the result of investigation of his career and behaviour in the prison, not to commit a crime again in society. The method currently adopted in reasoning involves many questionable points.

We proceed in maximizing the precision of prediction, that is, the success rate of prediction, in other words, the rate the prediction "he will be good or bad in prognosis" turns out to be true, by quantifying the qualitative results of investigations. It is here the problem of quantifying the qualitative data arises.

What are the contents of current investigations? The contents consist of the items concerning the factors that are considered to be useful from various analyses of past data in prediction of a criminal's prognosis in the society. For example, physical and hereditary evidences, character, family life, occupation, circumstances, economical life, etc. But human relations are much too complicated to be judged merely by these rather superficial factors. They must be contemplated from a higher point of view and we have adopted the factors emerging from these considerations. For example, the growing pattern of a criminal, to the present from childhood, his behaviour in prison, his psychological state in the offence and concerning the victims, feelings in prison, desires, interest, attitude towards the society etc.

As the results of the investigations on the factors we obtain the reactions of a criminal in every item. In quantifying these reactions, we predict a criminal's prognosis. The problem now is to quantify these reactions, and to synthetize the quantities obtained, in order to maximizing the success rate of prediction.

We repeat again that our aim is to maximize the success ratio of parole prediction by representing it in terms of probability from the collective theory.

III. Practical Method of Quantifications

Let group A, B be respectively the criminals who are good in prognoses and bad. We investigate every one of groups A, B. After this we quantify all the reactions in every item.

Now suppose that every reaction has been quantified. Every one in A, B has a quantity which, as it were, synthetizes his reactions in all items of investigation. This quantity is a weighted total value of the quantities given

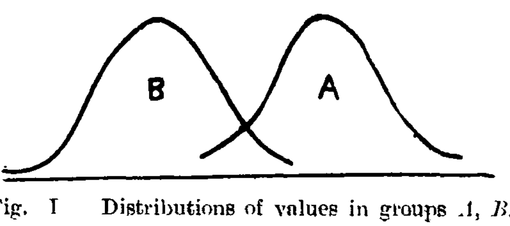


Fig. I Distributions of values in groups A, B.

to his reaction in each item. For convenience, we shall call this quantity a "value." Every one has only one value. Then the distribution of these values in A and B respectively are observed. (see Fig. I)
From these distributions the criterion will be made whether we can parole a criminal or not.

The method employed cannot be said the best. But it can be recognized as the first step of succesive approximation towards the goal in formulating a scheme necessary to solve problems involving complicated phenomena and attempting to construct practically the most effective prediction formulae and the most reasonable criteria of judgment.

Now suppose that the distribution of values in group A and B is obtained. We interpret the distribution of values in group A or B as the distribution function, strictly speaking, the probability density function in population A or B, constructed by investing every one from group A or B an equal sampling probability.

Then we use the theory of Richard von Mises. Let x be a label of population represented by a real number. Consider n number of populations, the probability density function being $p_1(x), p_2(x), \cdots, p_n(x)$, respectively. We treat the problem of determining the population the observed value x belonging to. For example, we consider three populations A, B and C. Let the probability density functions of the populations A, B and C be all Gaussian with the means 75, 50 and 25, and with the standard deviations $4\sqrt{2}$, $8\sqrt{2}$ and $12\sqrt{2}$ respectively. Now if we obtain $x = 45$, which population A, B or C should we consider this to belong to? We have e% of confidence, (the central point of the v. Mises' thought is to maximize this percentage of confidence) if we set the criterion:

x bigger than u belongs to A
x smaller than v belongs to C
x between them belongs to B

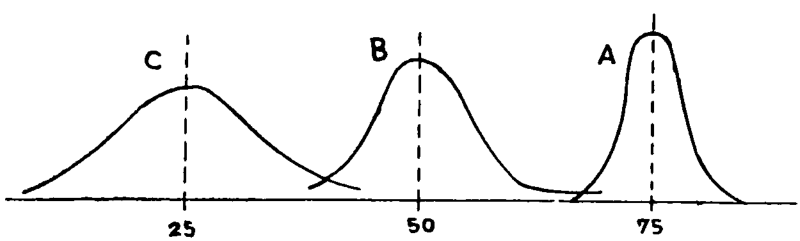


Fig. II

We consider our parole prediction as follows. Let A be the group of the ones with whom the parole succeeds, and the label be the value, the weighted total score of all the factors mentioned. Each individual is considered to be drawn at random from the group A. The sequence of individuals drawn is considered to form a "collective" with a distribution $p(x)$. A is considered a population in this standpoint. That is to say, the population A is considered to have a density function $p(x)$. Similarly, B is a population of the failures group in parole with the distribution of the total score of factors $q(x)$. That is, the total score of factors of the failures has a density function $q(x)$. Now, our problem is to predict which class, A or B, a criminal having the total score x will belong to, and to take measure of its confidence.

We give only the result of this theory.

Definition : The success rate is the probability to obtain the right result when we give a definite proposition.—the probability to judge what belongs to A to be A.

Notation : Let

$$P_\nu = \int_{R_\nu} p(x)\,dx \qquad \nu = 1, 2, 3, \cdots n,$$

where R_ν is an interval of variable x, and does not overlap and $\sum_{\nu=1}^{n} R_\nu =$ whole interval. P_ν is the probability that the variable of the population ν belongs to R_ν.

Proposition ; Divide the whole space into n intervals so that P_ν is all the same, P. Then we obtain an optimum result, and consequently a high success rate. The success rate is then equal to P.

Omitting the proof, here we consider the above example (class A, B and C). We have only to determine u, v so that

$$P = \int_u^\infty \frac{1}{\sqrt{2\pi}\,\sigma_A} e^{-\frac{(x-m_A)^2}{2\sigma_A^2}}\,dx = \int_v^u \frac{1}{\sqrt{2\pi}\,\sigma_B} e^{-\frac{(x-m_B)^2}{2\sigma_B^2}}\,dx$$

$$= \int_{-\infty}^{v} \frac{1}{\sqrt{2\pi}\,\sigma_C} e^{-\frac{(x-m_C)^2}{2\sigma_C^2}}\,dx$$

where

$$m_A = 75, \qquad \sigma_A = 4\sqrt{2}$$

$$m_B = 50, \qquad \sigma_B = 8\sqrt{2}$$

$$m_C = 25, \qquad \sigma_C = 12\sqrt{2}.$$

In this way we get $u = 70$, $v = 40$. In this case the success rate P is 96.1%.

Similarly in the parole prediction, let A have a density function

$\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$ and B a density $\frac{1}{\sqrt{2\pi}\,\sigma'} e^{-\frac{(x-m')^2}{2\sigma'^2}}$. The optimum division is to determine x_0 so that

$$\frac{1}{\sqrt{2\pi}\,\sigma}\int_{x_0}^{\infty} e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx = \frac{1}{\sqrt{2\pi}\,\sigma'}\int_{-\infty}^{x_0} e^{-\frac{(x-m')^2}{2\sigma'^2}}\,dx$$

The value of x_0 is given in the form

$$\frac{x_0 - m}{\sigma} = -\frac{x_0 - m'}{\sigma'}$$

$$x_0 = \frac{m\sigma' + m'\sigma}{\sigma + \sigma'}.$$

We judge that a criminal having a "value" higher than x_0 belongs to A—the parole will succeed with him—lower than x_0 belongs to B—the parole will fail with him.

Its success rate is

$$P = \frac{1}{\sqrt{2\pi}}\int_{\frac{m'-m}{\sigma+\sigma'}}^{\infty} e^{-\frac{t^2}{2}}\,dt.$$

Now we consider how to determine $p(x)$, $q(x)$ of the population A and B respectively. In order to solve this problem, we set the Central Limitting Theorem.

Theorem: $X_1, X_2, \cdots, X_n$ be independent random variables (collectives) with the means $m_1, m_2, \cdots, m_n$ and with the standard deviations $\sigma_1, \sigma_2, \cdots, \sigma_n$. We consider

$$X = X_1 + X_2 + X_3 + \cdots + X_n$$

$$S_n^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2$$

(i) S_n^2 is the variance of X,

(ii) Let each X_i be bounded $|X_i| \leq A$.

The distribution of $\dfrac{X - \sum_{i=1}^{n} m_i}{S_n}$ tends to Gaussian distribution as n increases and $\dfrac{A}{S_n}$ decreases, i.e. the distribution of X tends to the Gaussian distribution with the mean $(m_1 + m_2 + \cdots + m_n)$ and with the variance S_n^2.

In our case, $X_1, X_2, \cdots, X_n$ are given from each factor. Now consider the population A (in which the parole has succeeded and will succeed with). The only knowledge about A is that about the palores whose prognoses were good in the past. Consider A here not as the object of the survey but as a theoretical model. Later on we will treat the problem of practical surveying procedure.

We give numerical values (quantities) to each category of a factor (an item)—for example, if a factor is about parents, the categories will be with parents, with only father, with only mother, without parents respectively. In some cases, X has characteristics $0, 1, 2, 3$ and its distributions are a, b, c, d respectively. In this way, we give a random variable attached to each factor. In fact, a, b, c, d are unkown, but can be estimated from the sample value. Thus $X = X_1 + X_2 + \cdots + X_n$ is considered as a random variable respecting the total score of the factors.

For convenience, we suppose that $X_1, X_2, \cdots, X_n$ are independent. The dependent case can be treated in the same way. In our case, X_i is bounded clearly from the quantification, σ_i^2 is positive and their sum increases as n increases. Then the distribution of X tends to Gaussian distribution with the mean $(m_1 + m_2 + \cdots + m_n)$ and the variance S_n^2 from the above theorem. Thus, the distribution of A is approximately considered as a Gaussian distribution.

Similarly, the total scores of B have a Gaussian distribution.

The discussion given above is limited to a model population. Now we shall discuss the application of this model to practical cases. Our knowledge concerning A and B is confined to the past (sample). A and B are the populations in which the parole succeeded and will succeed, and failed and will fail, respectively. The problem of prediction is based on this assumption, that is the assumption that the analyses of the past can be used for prediction of the future. If the past and the future is different from each other (does not belong to the same population), prediction is impossible. In our case, the prediction of the future is based on the assumption that the past and the future has the same distribution. The plausibility of this assumption is provided from past experience. Of course, we do not make prediction today from the data obtained 20 years ago.

We treat this problem supposing that the data obtained in the past about A and B are random samples from the populations A and B respectively. After obtaining the numerical values mentioned above of the random samples, we must consider its confidence. The number of available data then becomes important. In other words, the total amount of the past data must be large enough to be reliable in the light of sampling theory. of sampling theory. Here we do not discuss about this point.

Then we must examine the distribution of the total score of the factors mentioned in the past data. For this purpose, we adopt the idea of the testing hypothesis. Supposing that the population distribution has Gaussian type (theoretically gained in this case), χ^2-test is used. If this hypothesis is not rejected, we may well consider the Gaussian distribution as the population distributions.

If there is no difference between two populations A and B, we can not predict anything. Then we must take up the effective factors for the prediction, and quantify them and consider how to connect them to differ from each other as much as possible in the test.

Firstly, apart from quantification of factors we have considered the method of judging the population, the sample point belongs to A or B and how to maximize the success rate mentioned above. Next we consider how to quantify the factors to obtain the success rate.

1-st step: To quantify the i-th factor, let the quantities of categories of this factor be, for example, α, β, γ, δ, population A has then a distribution $p_\alpha, p_\beta, p_\gamma, p_\delta$, (%), population B has a distribution $q_\alpha, q_\beta, q_\gamma, q_\delta$ (%). From this we may obtain the mean and variance of the factor i.

Now, consider the problem of determining α, β, γ, δ maximizing the difference of means under the condition that the variances (of the population A and B) are fixed, for example, equal to 1. We have to determine α, β, γ, δ unequaly assuming the sum of two means, or the mean of A or B to be zero. Thus we quantify the categories of the factors by the above method, without giving $0, 1, \cdots$ a priori. The obtained random variables of the populations A and B can be then expressed by X_i, Y_i and their means by m_i, m_i'. In practice, $p_\alpha, p_\beta, p_\gamma, p_\delta$; $q_\alpha, q_\beta, q_\gamma, q_\delta$; m_i, m_i' can be estimated from the value of the past.

2-nd step: We consider not merely total score, but the score

$$X = a_1X_1 + a_2X_2 + \cdots + a_nX_n$$
$$Y = a_1Y_1 + a_2Y_2 + \cdots + a_nY_n,$$

where $\sum a_i = n$ and a_i are weights.

In this case we obtain

the mean of $X = a_1m_1 + a_2m_2 + \cdots + a_nm_n$

the mean of $Y = a_1m_1' + a_2m_2' + \cdots + a_nm_n$

the variance of $X = a_1^2 + a_2^2 + \cdots + a_n^2$

the variance of $Y = a_1^2 + a_2^2 + \cdots + a_n^2$

(for the variances of X_i and Y_i are 1 by the assumption).

The distribution of X and Y tends to the Gaussian distribution as n increases.

It is reasonable that we determine the weights $(a_1, a_2, \cdots a_n)$ to maximize the success rate of v. Mises (correct judgment rate), $P=\frac{1}{\sqrt{2\pi}}\int_{\frac{m'-m}{\sigma+\sigma'}}^{\infty} e^{-\frac{t^2}{2}}dt$, that is to minimize $\frac{m'-m}{\sigma+\sigma'}$, that is, to maximize $\frac{m-m'}{\sigma+\sigma'}$ can be expressed in the form

$$\frac{a_1(m_1-m_1')+\cdots+(m_n-m_n')}{2\sqrt{a_1^2+\cdots+a_n^2}}$$

Let

$$m_i-m_i'=l_i,$$

we have only to maximize

$$f=\frac{a_1l_1+\cdots+a_nl_n}{\sqrt{a_1^2+\cdots+a_n^2}}.$$

For this purpose, under the condition of

$$\sum a_i=n,$$

we set

$$\frac{\partial f}{\partial a_i}=0 \qquad (l_i=1,\cdots n)$$

so we get the next relation,

$$a_i=\frac{l_i}{l_1+\cdots+l_n}\cdot n$$

or

$$a_i=\frac{l_in}{\sum l_i}$$

Satisfying this relation, we obtain the maximum success rate. This is the second step of quantification. Using this we quantify the factors.

Summarizing the first process, the second process and the last proposition, we can say as follows.

(1) At first we calculate the score of a criminal z from the next formula.

$$z=a_1z_1+a_2z_2+\cdots+a_nz_n.$$

(2) Then we determine to which population, A or B, this z belongs to. In order to judge from which population z is drawn, we proceed as follows. Comparing this z with x_0, we see which is greater than the other, where

$$x_0=(m\sigma'+m'\sigma)(\sigma+\sigma')^{-1}$$

$$m=a_1m_1+a_2m_2+\cdots+a_nm_n,\quad m'=a_1m_1'+a_2m_2'+\cdots+a_nm_n',$$

$$\sigma^2=\sigma'^2=a_1^2+a_2^2+\cdots+a_n^2.$$

As z is greater than x_0, z belongs to A, and vice versa.
(3) Then the success-rate is given as

$$P = \frac{1}{\sqrt{2\pi}} \int_{\frac{m'-m}{\sigma+\sigma'}}^{\infty} e^{-\frac{1}{2}t^2} dt$$

where

$$\frac{m'-m}{\sigma+\sigma'} = -\frac{1}{2}\sqrt{\sum l_i^2}.$$

Thus, the quantification is not done abstractly and generally, but should be done with the propositon required in each case in mind, and we may get profitable quantifications for the prediction.

In the practical use of these processes, the sampling method may be applied, and the population values of this quantification and determination of criterion is estimated from samples, in which confidence degree of estimation has been considered, but this problem of error is so complicated, that now we shall not refer to it here.

In the above method, we have set $X_1, X_2, \cdots, X_n$ mutually independent, for the sake of convenience. If they are dependent, we may think as follows. Let the correlation-coefficients between X_i and X_j, Y_i and Y_j be equal to ρ_{ij}. And let

$$X = a_1X_1 + a_2X_2 + \cdots + a_nX_n$$
$$Y = a_1Y_1 + a_2Y_2 + \cdots + a_nY_n$$

In this case the Central Limit Theorem stands similarly under certain conditions. The variances σ_X^2, σ_Y^2 of X and Y are then

$$a_1^2 + a_2^2 + \cdots + a_n^2 + \sum_{i \neq j} a_i a_j \rho_{ij}$$

therefore

$$\frac{m-m'}{\sigma+\sigma'} = \frac{\sum a_i l_i}{2\sqrt{\sum a_i^2 + \sum a_i a_j \rho_{ij}}}$$

Then we obtain the values of $a_i (i = 1, 2, \cdots, n)$ by maximizing the above formula under the condition $\Sigma a_i = n$. In practice the case is usually dependent.

It may be sometimes very effective for parole judgement to make up singular complex factors combining two or more factors.

In the above process we mentioned the idea of "v. Mises' Optimum." We shall now consider this "Optimum." The above mentioned

$$P_\nu = \int_{R_\nu} p_\nu(x)\, dx$$

is the probability of the ν-th class appearing in R_ν.

Suppose that we have, in the first N trials of the infinite sequence, N_ν ones belong to the ν-th class. If only ones having label x falls in R_ν are judged to belong to the ν-th class, in these N_ν trials we have $N_\nu(P_\nu + \varepsilon_\nu)$ trials belonging to the ν-th class, where $\varepsilon_\nu \to 0$ as $N_\nu \to \infty$. Then in the N trials the number of right judgement from observed values is

$$N_1(P_1 + \varepsilon_1) + \cdots + N_n(P_n + \varepsilon_n)$$

and in this case

$$\frac{N_1}{N}(P_1 + \varepsilon_1) + \cdots + \frac{N_n}{N}(P_n + \varepsilon_n)$$

When $N \to \infty$, $N_\nu \to \infty$; so when $N \to \infty$, $\varepsilon \to 0$. From this, the ratio of right judgement is seen to be equal to

$$\frac{1}{N}(N_1P_1 + \cdots + N_nP_n)$$

This N_ν is unknown. Let $P_{min} = Min\ P_\nu$. To minimize the above formula, or the ratio of right judgement, it is neccessary that $N_{min} = N$ and $N_i = 0$ $(i \neq \nu)$. Then this value becomes P_{min}. Under this consideration, the proportion of right judgement must be equal to P_{min}. By maximizing this P_{min}, the optimum result is obtained. When N_ν / N is unknown, we have to employ the above process.

Next let $\lim\limits_{n\to\infty} \frac{N_\nu}{N} = Q_\nu$ be known. In this case, we assume Q_A and Q_B $(= 1 - Q_A)$ of A and B respectively, are known: in other words, the ratio of parole-failure $(Q_B = Q)$ is known. Then the right judgement ratio is given by

$$S = Q_BP_B + Q_AP_A = QP_B + (1 - Q)P_A$$

Now let

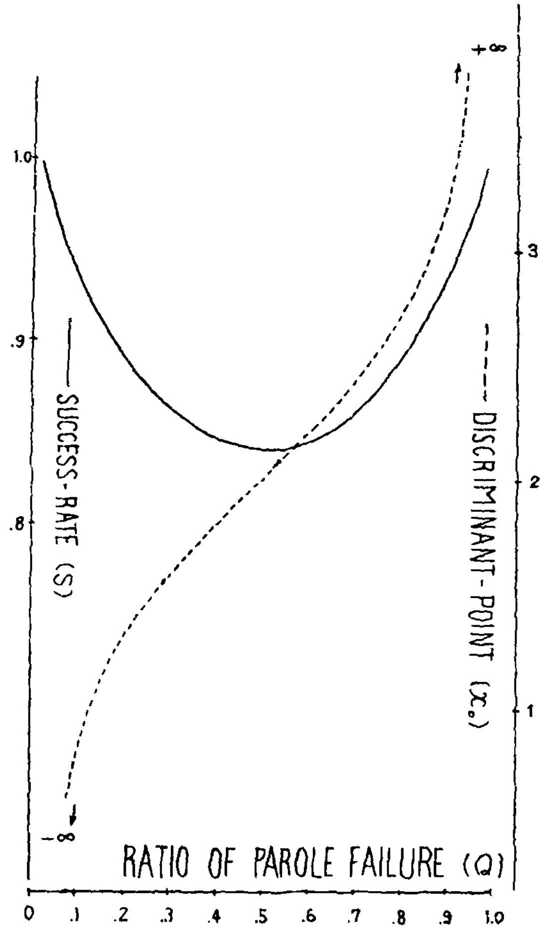
$$P_B = \frac{1}{\sqrt{2\pi}\,\sigma'}\int_{-\infty}^{x} e^{-\frac{(x'-m')^2}{2\sigma'^2}}\,dx'$$

$$P_A = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{x}^{\infty} e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx$$

so

$$S = Q\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{x-m'}{\sigma'}} e^{-\frac{t^2}{2}}\,dt + (1 - Q)\frac{1}{\sqrt{2\pi}}\int_{\frac{x-m}{\sigma}}^{\infty} e^{-\frac{t^2}{2}}\,dt$$

From this formula let us find x maximizing S.


1.0
.9
.8
SUCCESS-RATE (S)
DISCRIMINANT-POINT (x_0)
+∞
−∞
3
2
1
RATIO OF PAROLE FAILURE (Q)
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0

$$\left(\frac{1}{\sigma'^2}-\frac{1}{\sigma^2}\right)x^2+2\left(\frac{m}{\sigma^2}-\frac{m'}{\sigma'^2}\right)x+\left(\frac{m'^2}{\sigma'^2}-\frac{m^2}{\sigma^2}\right)-2L=0$$

where

$$L=\log_e\frac{Q}{1-Q}\cdot\frac{\sigma}{\sigma'}$$

If let $\sigma'=\sigma$, x is given by

$$x=\frac{1}{2}\left\{(m+m')+\frac{2L\sigma^2}{m-m'}\right\}$$

By v. Mises' method, x is given by

$$x=\frac{1}{2}(m+m')$$

Comparing the success-rate of these two cases by the graph, we can easily see that the former is more effective than the latter. The former is thus more useful in practice.

Profitable results may be obtained by treating practical cases on the basis of this theory as follows. The quantification of the first and second process are the same as before, and we calculate

$$X=\sum a_iX_i, \qquad Y=\sum a_iY_i.$$

From this, determining the distribution-type (type of population) and calculating the mean and standard deviation and then under this type of population characterising the two groups, we determine the deviding point and success-rate. If the distribution type is regarded as Gaussian type (perhaps it will be the usual case), our discussion can be applied directly.

IV. Complements

As is seen from this process, quantification is depended upon the quality and number of adopted facters, the variety of populations, the method of treatment, and the kind of conclusion that we want: therefore the quantifications is very functional.

In this paper we have discussed a method of quantification from the predictional point of view. Many other useful quantification methods are being considered, using similar methods.

V. Complements

Moreover we can widen the field of population, and generalize our prediction stand point.

For instance, at the time t_1, let the condition of our population be E_1 and in this case the method of prediction be F_1 symbolically. Now we assume that the condition E_i is predicted with a reliability, then we may predict F_i by analysing E_1, F_1 and E_i.

In this case, the success rate will be estimated under a certain degree of confidence.

Refferences

(1) Yoshitomo Ushijima: "Judging Method of Delinquent Juveniles in Early Stage" Kaneko Co. Tokyo. 1948.

(2) R. v. Mises: "On the Classification of Observation Data in the Distinct Groups," Annals of Mathematical Statistics, 1945, vol. 16 No. 1.

(3) R. v. Mises: "Vorlesungen aus dem Gebiete der Angewandte Mathematik. I. Band. Wahrscheinlichkeitsrechnung," F. Deuticke, Leipzig und Wien, 1931.

(4) Katsuhiko Nishimura, Chikio Hayashi: "Parole Prediction," Research Series I, The Central Training Institute of Correction and Rehabilitation Personnel, 1949.

(5) Chikio Hayashi: "On the Quantification of Qualitative Data from the Mathematico-Statistical Point of View," Research Series II, the Central Training Institute of Correction and Rehabilitation Personnel, 1950.

(6) Chikio Hayashi: "On the Scientifical Stand Point of Parole Prediction," Hogakushirin, Hosei University 1950.

(7) Masatsugu D. Ishida: "Fire Prediction," Research Memoir of the Institute of Statistical Mathematics, 1949, Vol. 5.