Estimation of Transition Probabilities in Ischemic Heart Disease by Markov Model

Go TOMONAGA*, Mitsuyasu KAGIYAMA**,
Yasuo OGASAWARA*, Masami GOTO***,
Sinichiro TADAOKA*, Masanobu NAKAI*,
Osamu HIRAMATSU*, Keiichiro MITO****,
Noritake HOKI* and Fumihiko KAJIYA*

*Department of Medical Engineering and Systems Cardiology,
  Kawasaki Medical School, Kurashiki 701–01, Japan
**Computer Center, Kawasaki Medical School
***Department of Medicine, Kawasaki Hospital,
  Kawasaki Medical School, Okayama 700, Japan
****Kawasaki Para–Medical College, Kurashiki 701–01, Japan
Accepted for Publication on July 13, 1982

ABSTRACT. The purpose of this study is to develop a method of estimating transition probabilities among the disease "states" indicating its severity for the chronic diseases. This method was applied to the analysis of the natural history of patients with coronary heart diseases. The transition process among "states" is assumed to be expressed by a time discrete simple Markov process. The severity of the disease was classified into 3 "states", i. e., S_1 : single vessel disease, S_2 : double vessel disease, S_3 : triple vessel disease. Estimation of the transition probabilities was made by the maximum likelihood method, using the follow–up data of the numbers of the survival. The accuracies of the estimated values are evaluated by the asymptotic variances. From the present study the followings were observed : (1) the accuracy of the curve fiitting for the follow–up data was satisfactory, (2) the catenary model was the most prominent in the sense that an information criterion AIC is minimum, and (3) there may exist the reversible transitions among some disease states.

1. Introduction

In medical care of the chronic disease, it is necessary to know the stage of the disease relating to its severity, since the observation of transitions among stages provides an important information on progression or regression of the disease. For the evaluation of transitions, it is essential to monitor changes of the stages for patient groups repeatedly at regular time intervals. These monitoring, however, is not always an easy task, especially in such cases that the observation is invasive to patients and/or the method has technical difficulties. An example is the monitoring of vascular lesions in ischemic heart disease (IHD) ; IHD is frequently classified into single, double and triple vessel diseases by

友永　轟，鍵山光庸，小笠原康夫，後藤真己，忠岡信一郎，中井正信，平松　修，三戸恵一郎，伯耆徳武，梶谷文彦

selective coronary angiography. However, invasiveness of coronary angiography prevents its repetitive use at regular time intervals.

In the present study, we proposed a method of estimating the transition probabilities among stages of the disease by a compartmental model. An application was made to analyze the prognosis of IHD.

In the use of a compartmental model in clinical care, compartments correspond to the stages of the disease and the amount of a tracer in a compartment to the number of patients belonging to a stage. Transitions of tracers (patients) among compartments (stages) occur stochastically. In this sense, the compartmental model for evaluating the prognosis of the chronic disease can be considered as a common framework to the Markovian process. Several reports have indicated that the natural history of chronic disease could be well modeled by the Markovian process in most cases[1-4]. In their studies, long-term prognosis was estimated by the transition probabilities. However, transition probabilities are not always easy to obtain as already mentioned.
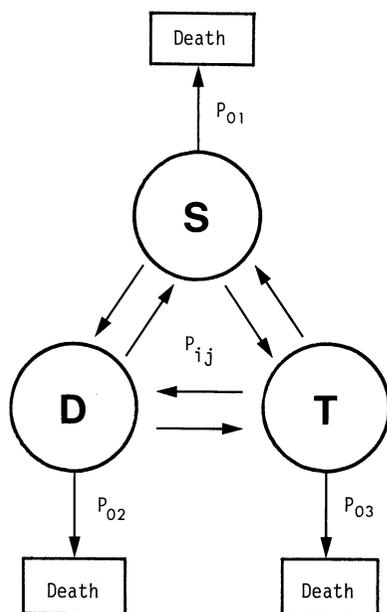

Fig. 1. Reversible 3 compartment model with an absorbing point (death). Compartments correspond to "disease states" and the amount of tracer to "the number of patients" belonging to each state. This kind of compartmental model can be considered as a common framework to Markovian process.

In contrast to the applications of the Markovian model so far, this paper presents a method of estimating transition probabilities among stages of the disease from the survival data (data of death). According to the terms employed

in the Markovian model, we use "disease states" instead of "stages of disease" hereafter. The "disease states" are classified into "single", "double" and "triple" vessel diseases and "death". Three different types of models (one way, catenary and reversible) were used for the analysis (see Fig. 2) and the transition probabilities among the states were estimated by the maximum likelihood method. The accuracy of identifications was evaluated by asymptotic variances and an information criterion (AIC)[7-8].
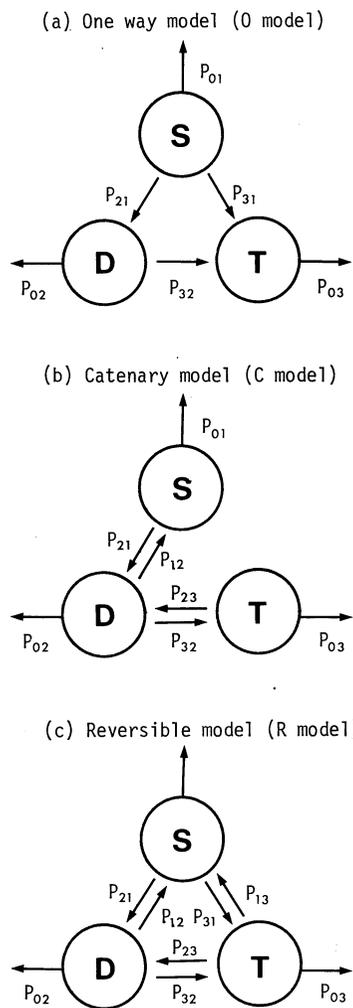

Fig. 2. Three different compartment (Markovian) models used in the present study. The compartments (states) S, D and T denote single, double and triple vessel disease, respectively. The paths to the outside of the system imply death.

2. Formulation of Problem

Suppose that patients with a chronic disease are initially classified into several patients groups —disease state— by the first diagnostic procedure and that the survival data for each patient group are obtained every year for many years. Classification of the disease state cannot performed repeatedly. The problem is to estimate transition probabilities P_{ij} among disease states from the survival data. Here $P_{ij}(i\neq0)$ indicates the transition probability from state j to state i for a year and P_{oj} is the death probability from state j. The matrix (P_{ij}) is assumed to be time–invariant.

Consider that there are initially $N_i(0)$ patients in state i, $i=1, \cdots\cdots n$, and that $N_i(j)$, $j=1, \cdots\cdots, t$, is the number of survival patients after j years who initially belonged to state i. In other words, $N_i(k)-N_i(k+1)$, $k=1,\cdots\cdots, t-1$, represents the number of death during the year $k\sim k+1$, and $N_i(t)$ the number of survival patients in tth year.

Our problem is the estimation of the transition probabilities $\{P_{ij}\}$ among disease states from the observation of $N_i(j)$, $i=1, \cdots\cdots, n$, $j=0, \cdots\cdots, t$.

3. Likelihood Function for Survial Data

In this section, we derive the expression for the likelihood function to estimate P_{ij} when the survival data are given by the form described in the previous section.

First, we derive the recursion formula to get the probability of death $Q_i(s)$ of a patient initially been in state i, during the period from $(s-1)$th year. to sth year. Since $Q_i(1)$ implies the probability of death for the first one year, it is simply written as,

$$Q_i(1)=P_{oi}, \quad i=1, 2, \cdots\cdots, n \tag{1}$$

Since $Q_i(2)$ is the probability of death by way of any one of the states 1, 2, $\cdots\cdots$, and n from the initial state i, we have,

$$Q_i(2)=\sum_{k=1}^{n} Q_k(1) P_{ki}$$

In the same way, the recursion formula is given by,

$$Q_i(s)=\sum_{k=1}^{n} Q_k(s-1) P_{ki} \tag{2}$$

The probability of a patient being alive for t years is written as

$$\overline{Q}_i(t)=1-\sum_{s=1}^{n} Q_i(s) \tag{3}$$

Using Eq's (1)–(3) and considering the independent nature of the Markovian process, the likelihood function for the survival data of patients initially belonged to state i is expressed by

$$L_i = \binom{N_i(0)}{N_i(0) - N_i(1)} \binom{N_i(1)}{N_i(1) - N_i(2)} \cdots \binom{N_i(t-1)}{N_i(t-1) - N_i(t)}$$

$$\times Q_i(1)^{N_i(0) - N_i(1)} \; Q_i(2)^{N_i(1) - N_i(2)}$$

$$\cdots Q_i(1)^{N_i(t-1) - N_i(t)} \; \overline{Q_i}(t)^{N_i(t)} \tag{4}$$

As the joint likelihood function L is the product of L_1, L_2, $\cdots$, and L_n, L is given by

$$L = L_1 \; L_2 \cdots \cdots L_n \tag{5}$$

4. Estimation of Parameters

Now our problem is to estimate P_{ij} so as to maximize L in Eq. 5. However, P_{ij} has the constraints based on the statistical nature of probabilities, i. e.,

$$\sum_{k=1}^{n} P_{kj} = 1.$$

$$0 \leqq P_{ij} \leqq 1, \quad i = 0, 1, \cdots, n, \quad j = 1, \cdots, n \tag{6}$$

In order to remove the constraints in Eq. 6, transformation of parameters from P_{ij} to α_{ij} was made. It follows,

$$P_{ij} = \alpha_{kj}^2 \Big/ \sum_{k=0}^{n} \alpha_{kj}^2, \quad i = 0, 1, \cdots, n, \quad j = 1, 2, \cdots, n \tag{7}$$

where

$$\alpha_{0j} = 1, \quad j = 1, 2, \cdots, n$$

Then estimate of parameter P_{ij} are obtained by solving an ordinary non–linear optimization problem[6]. For this problem, we adopted the simplex method of Nelder and Mead[5], because it can create its own scaling factors and converge for a wide range of starting points.

5. Evaluation of the Accuracy of Estimates

In the estimation of parameters, it is important to evaluate the accuracy of estimates objectively. For this evaluation, we used the asymptotic variance which was obtained in the following manner.

Consider a generalized Fisher's information matrix I $(I_{kl})_{p \times p}$, where

$$I_{kl} = -E\left(\frac{\partial^2}{\partial \theta_k \partial \theta_l} \ln L\right) \tag{8}$$

and L is the likelihood function given in Eq. 5. As a generalization of the Cramér-Rao inequality, it can be shown that the following holds : for any unbiased estimator $(\tilde{\theta}_1, \cdots, \tilde{\theta}_p)$ of $(\theta_1, \cdots, \theta_p)$ which are equivalent to $(P_{01}, P_{11}, \cdots)$ in our case, let V be the variance-covariance matrix. V must

be larger than or equal to the matrix I^{-1} in the sense that $V-I^{-1}$ is non-negative definite. For large samples the diagonal of I^{-1} provides an estimator of variances of the maximum likelihood estimators of the parameters.

Substituting Eq. 4 into Eq. 5 and taking the logarithm, it follows,

$$\log L = \sum_{k=1}^{n} \left[\sum_{s=1}^{t} (N_k(s-1) - N_k(s)) \log Q_k(s) + N_k(t) \log \overline{Q}(t) \right]$$
$$+ \text{(the term independent of } \Theta\text{'s)} \tag{9}$$

Partially differentiating Eq. 9 with respect to Θ_j, we have

$$\frac{\partial \log L}{\partial \theta_j} = \sum_{k=1}^{n} \left[\sum_{s=1}^{t} \frac{(N_k(s-1) - N_k(s))}{Q_k(s)} \frac{\partial Q_k(s)}{\partial \theta_j} \right.$$
$$\left. - \frac{N_k(t)}{\overline{Q}_k(t)} \sum_{s=1}^{t} \frac{\partial Q_k(s)}{\partial \theta_j} \right] \tag{10}$$

Partial differentiation of Eq. 10 with respect Θ_i gives

$$\frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j} = \sum_{k=1}^{n} \left[\sum_{s=1}^{s} (N_k(s-1) - N_k(s)) \left\{ \frac{1}{Q_k(s)} \frac{\partial^2 Q_k(s)}{\partial \theta_i \partial \theta_j} \right. \right.$$
$$\left. - \frac{1}{Q_k(s)^2} \frac{\partial Q_k(s)}{\partial \theta_i} \frac{\partial Q_k(s)}{\partial Q_j} \right\} - N_k(t) \left\{ \frac{1}{\overline{Q}_k(t)} \sum_{s=1}^{t} \frac{\partial^2 Q_k(s)}{\partial \theta_i \partial \theta_j} \right.$$
$$\left. \left. + \frac{1}{\overline{Q}_k(t)^2} \sum_{s=1}^{t} \frac{\partial Q_k(s)}{\partial \theta_i} \sum_{s=1}^{t} \frac{\partial Q_k(s)}{\partial \theta_j} \right\} \right] \tag{11}$$

Taking the expected value of Eq. 4 and making use of the following relations,

$$\langle N_k(s-1) \rangle - \langle N_k(s) \rangle = N_k(0) Q_k(s), \quad s = 1, 2, \ldots, t-1$$

$$\langle N_k(t) \rangle = N_k(0) \overline{Q}_k(t) \tag{12}$$

we have,

$$I_{ij} = \sum_{k=1}^{n} N_k(0) \left\{ \sum_{s=1}^{t} \frac{1}{Q_k(s)} \frac{\partial^2 Q_k(s)}{\partial \theta_i \partial \theta_j} + \frac{1}{\overline{Q}_k(t)} \sum_{s=1}^{t} \frac{\partial Q_k(s)}{\partial \theta_i} \sum_{s=1}^{t} \frac{\partial Q_k(s)}{\partial \theta_j} \right\} \tag{13}$$

Since Θ_1, Θ_2, $\ldots$ correspond to transition probabilities in our problem, the partial derivatives $\partial Q_i(s)/\partial \theta_j$, $\ldots$ can be calculated by the following recursion formulae, which are obtained by partial differentiation of Eqs. 1 and 2 with respect to Θ_i

$$\frac{\partial Q_i(1)}{\partial P_{oi}} = 1, \quad \frac{\partial Q_i(1)}{\partial P_{hk}} = 0, \quad (h, k) \neq (o, i)$$

$$\frac{\partial Q_i(t)}{\partial \theta_j} = \sum_{k=1}^{n} \frac{\partial P_{ik}}{\partial \theta_j} Q_k(s-1) + \sum_{k=1}^{n} P_{ik} \frac{\partial Q_k(s-1)}{\partial \theta_j} \tag{14}$$

$$t = 2, 3, \ldots, s, \quad i, k = 1, 2, \ldots, n$$

6. Comparison of Models

In this paper we proposed three types of models for fitting the survival data by the Markovian process. For a measure of fittness of a model, we considered an application of criterion called an information criterion (AIC)[7,8] which was introduced by Akaike and which has been successfully applied for the identification of a statistical model.

Let N_1, N_2 ······, N_t be the number of t independent observations of a random variable with probabilility density function f (N/θ). AIC is defined as

$$AIC(\theta) = -2\sum_{i=1}^{t} \ln f(N_i/\widehat{\theta}) + 2k \tag{15}$$
$$= -2\ln(\text{maximum likelihood}) + 2k$$

Where k is the number of independently adjustable parameters within the model and $\widehat{\theta}$ is the vector of the maximum likelihood estimates of θ.

The model corresponding to the minimum AIC is selected as the optimal one.

7. Survival Data of Ischemic Heart Disease

An analysis of the patient survival data in a chronic disease is very important for the estimation of the natural course of the disease of the comparison of effects of several treatments. This is also true for ischemic heart disease (IHD) and many follow-up studies for IHD have been made.

In 1973 Bruchke et al.[9] published a report on consecutive nonsurgical cases of IHD followed for 5-9 years. In their study, 553 patients were classified initially into three categories, by finding of selective coronary angiography: single vessel diease (S), double vessel disease (D) and triple vessel disease (T). The number of patients survived was followed in each patient group for 5-9 years.

In this study, we used Bruschke et al.'s data for estimating P_{ij} in the Markovian model, since the number of patients followed was larger than others' and their data are appropriate for an analysis of natural history of IHD.

8. Results
8. 1 Estimation of Transition Probabilities
(1) Reversible model

The annual transition probabilities among S, D, T and death calculated in the reversible model are shown in Fig. 3. Annual transition probabilities from the S, D, or T state to the death were 0.026, 0.090 and 0.220, respectively. These values indicated that the mortality in ischemic heart disease rises steeply with an increase in the number of affected coronary arteries. This model suggested that transitions from D to S or to T occur with relatively high probabilities, 0.48 and 0.43, respectively, while the transition from S to D or to T occurs with low probabilities.

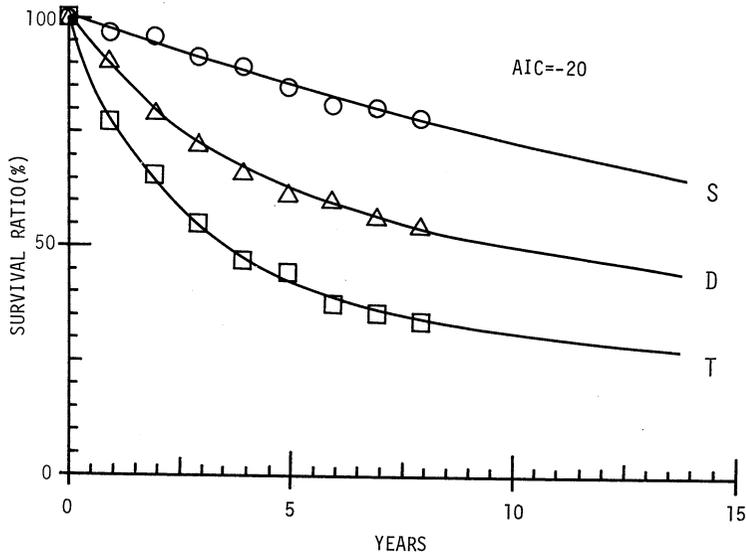Results of curve fitting for the survival data by the reversible model were

Fig. 3. Comparison of the Bruschke et al.'s survival data with those estimated by applying the reversible model. The solid curves represent estimation results and the symbols ○, △ and □ indicate Bruschke et al.'s data. The letters S, D and T denote the patients groups who initially belonged to single, double and triple vessel diseases, respectively.
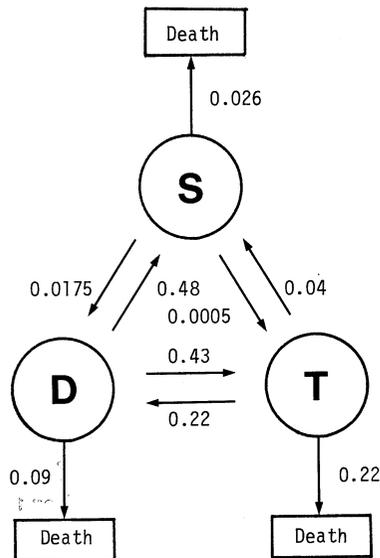

Fig. 4. Estimated values of transition probabilities in the reversible model.

shown in Fig. 4. Three curves indicate the mean survival values estimated for patient groups belonged initially to states S, D and T, respectively. These curves were in good agreements with the data by Bruschke et al.[9] For example, the 5 years survival rate in Bruschke et al.'s data were 84.8% and 44.2% for the initial disease states S, D and T, respectively. while the estimated values were 86.4%, 63.2% and 42.5%, respectively.

(2)   Catenary model

The transition probabilities obtained in the catenary model were shown in Fig. 5. Although the path between S and T is neglected in the catenary model, the transition probabilities were almost compatible with those in the reversible model. The accuracy of survival curve fitting was also satisfactory.
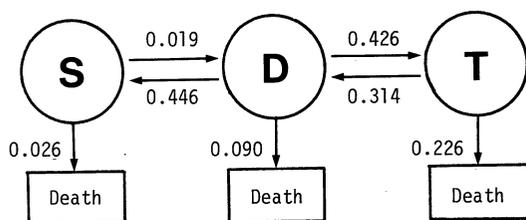

Fig. 5.   Estimated values of transition probabilities in the catenary model.

(3)   One way model

The one way model does not include the reversible–recovery–paths of transitions, i. e., T→D, T→S and D→S. Transition probabilities calculated in this
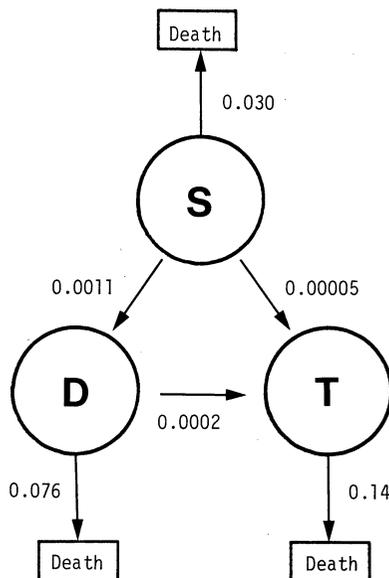

Fig. 6.   Estimated values of transition probabilities in the one way model.

model are shown in Fig. 6. The transition probabilities among S, D and T showed lower values comparing with those in other two models, while probabilities from S, D and T to death were almost the same as the results in other models. The curve fitting of survival data by the model was less satisfactory.

8.2 Comparison of Three Models

The accuracy of the transition probabilities obtained in each model was evaluated by asymptotic variances. The results are shown in Table. In general, the asymptotic variance in the catenary model showed the smallest value, comparing with the variances in other two models.

The values of AIC relative to that of the one way models were shown in the same table. The value of catenary model is the smallest.

These results indicated that the catenary model is the most approriate for the simulation of the survival data.

9. Discussion and Concluding Remarks

By using three compartment models, we analyzed the progression and

TABLE Estimated values of transition probabilities with the square roots of asymptotic variances in three different compartment (Markovian) models. The values of AIC are given at the bottom of the Table. Note that both AIC value and asymptotic variance of each parameter are the smallest in the catenary model.

	O Model		C Model		R Model	
p_{01}	0.031	± 0.011	0.026	± 0.009	0.026	± 0.011
p_{21}	0.001	± 0.65	0.019	± 0.04	0.018	± 0.31
p_{31}	0.000048	± 0.32			0.0005	± 0.18
p_{02}	0.0763	± 0.013	0.902	± 0.019	0.0904	± 0.019
p_{12}			0.446	± 0.23	0.480	± 0.28
p_{32}	0.0002	± 0.052	0.426	± 0.25	0.428	± 0.29
p_{03}	0.144	± 0.015	0.226	± 0.033	0.225	± 0.039
p_{13}					0.0419	± 0.30
p_{23}			0.314	± 0.13	0.220	± 0.68
AIC	0		−24		−20	

regression of ischemic heart disease (IHD) in terms of single, double, triple vessel diseases and death. The estimated survival data showed good agreements with Bruschke et al.'s data except the case of one way model. This suggested that the survival data of IHD could be modelled by a Markovian process.

Comparing the reversible model with the catenary, we selected the catenary model as an optimal one, since the AIC value and asymptotic variances were lower in the catenary model. The results that the estimated transition probabilities for S $\rightleftarrows$ T showed low values, will also support this selection.

On the transition probabilities in the catenary model, it is noted that the

values of the paths from D to T and from T to D were relatively high. These indicate that regression of arterial coronary lesions may occur in the chronic course of IHD. However, the regression will not necessarily imply a morphological improvement of the vascular lesion. It will also include a functional improvement. Progression of the disease will be rapid from D to T, since its transition probability was higher than that from S to D.

From the present study, it was indicated that the Markovian model provides useful information on evaluating the prognosis of chronic disease on the basis of survival data.

REFERENCES

1) Moriwaki, T., Inoue, M. and Sugiyasu, K. : Estimation of prognosis of diabetic retino-pathy using Markov process. 7th Congress IDF, 1970
2) Kitabatake, A., Inoue, M. and Sakakibara, H. : Long-range Prognosis of vascular heart disease. Saishin Igaku **33** : 120-121, 1978 (in Japanese)
3) Furukawa, T., Inada, H., Kajiya, F., et al. : Estimation of prognosis of chronic renal failure–Application of Markovian process–. Saishin Igaku **25** : 346-352, 1970 (in Japanese)
4) Kajiya, F., Inoue, M. and Furukawa, T. : Clinical course and therapy of disease from the point of view of control theory. Nihhon Rinsho **34** : 1052-1057, 1976 (in Japanese)
5) Nelder, J.A. and Mead, R. : A simplex method for function minimization. Computer Journal **7** : 308-313, 1965
6) Jacoby, S.L.S., Kowalik, J.S. and Pizzo, J.T. : Iterative methods for nonlinear optimi-zation problems. New Jersey, Prentice-Hall Inc. 1972
7) Akaike, H. : A new look at the statistical model identification, IEEE Trans. AC-16, 716-723, 1974
8) Akaike, H. : Information theory and an extension of the maximum likelihood principle. 2nd Int. Symp. Inf. Theory, pp. 267-281, 1973
9) Bruschke, A.V.G., Proudfit, W.L. and Sones, F.M. : Progress study of 590 consecutive non-surgical cases of coronary disease followed 5-9 years. 1. Arteriographic correlations. Circulation **47** : 1147-1153, 1973