Article 1.3

Ordinal Discrete Time Markov Transition Models

Guillermo Marshall and Wensheng Guo

Pontificia Universidad Católica de Chile and
University of Pennsylvania School of Medicine, Philadelphia, USA

Abstract. Markov transition models in equally spaced settings have been widely used in many applications. In equally spaced settings, the dependency from the previous observations are modeled explicitly by including the previous history as covariates. The estimation and inference are straightforward. However, in real practice such as data collected from a clinic, the observations are usually unequally spaced and missing data are common. The equally spaced Markov transition models break down in these general settings, because the transition probability not only depends on the previous stages, but also depends on the time interval between the observations. For rapid transition diseases, multiple transitions can happen if the interval between two observations are long enough. Another long existing problem for multi-stage Markov transition models is that the number of parameters expands exponentially with the increase of the number of states and covariates, which limits the application of such models to relatively simple cases. In this paper, we propose to model the transition rates as functions of covariates instead of the transition probabilities. We focus on the cases with ordinal outcomes, which can include absorbing stages such as dropout or death as special cases. We first propose general methodology and then propose a reduced common slope model that significantly reduces the number of parameters and increases the interpretability of the model. For computational convenience, we approximate the transition density matrix in the continuous time setting by the one-step transition probability matrix in a discrete time scale. This leads to an efficiently parametrized model and a computationally efficient estimation procedure. The model is illustrated using a four-state model for diabetic retinopathy in young subjects with insulin-dependent diabetes mellitus.

Keywords: Diabetic Retinopathy; Markov transition Model; Multi-State Model; Ordinal Response Models; Time-dependent Covariables.

Reference: Partially published in Guo W and Marshall G. ORDMKV: A Computer Program Fitting Proportional Odds Model for Multi-State Markov Models. *Computers Methods and Programs in Biomedicine* 46: 257-263.

1 Introduction

Markov transition models have been widely used to model the progressions of chronic diseases such as cancer, AIDS, and diabetes. When patients are observed at equally spaced intervals, the dependency of the history can be modeled explicitly by including the previous observations as covariates. By doing so, one assumes that the exact transition time is observed. This has been the common practice of most of the Markov transition model.

Unfortunately, in real practice patients visit a clinic at irregular time intervals and there are usually missing data even in the best controlled clinical trials. In such settings, modeling the transition probability by including the previous observations as covariate can be misleading, since (1) the longer the interval is, the more likely a transition may have occurred; (2) we seldom observe the exact transition time; (3) the patient may have been though several intermediate transitions between two consecutive observations. Therefore, we need to model the disease progression through the transition intensities instead of the transition probabilities.

In continuous time settings, the Markov transition model is completely specified by the transition intensity matrix. Kay [1] introduces a k-state Markov transition model with a special structure, in which a patient can progress to the adjacent stage or transfer to an absorbing stage. The exact transition times are not observed except in certain circumstances such as death. The model assumes that the underlying biological process can be characterized as a continuous-time Markov process and the covariates affect the disease progression through the transition densities.

Longini et al. [2] used this model to describe the distribution of the incubation period of HIV on patients with AIDS. Garg, Marshall, Chase et al. [6] used this model to describe the natural course of diabetic retinopathy. Marshall and Jones [3, 4] and Marshall, Guo and Jones [5] extended the k-state Markov transition model by proposing an exact likelihood function when the exact arrival time is observed and by introducing time-invariant and time-dependent covariables in the model. Marshall, Garg, Jackson et al. [7] used this regression model to identify various factors influencing the progression and regression of diabetic retinopathy in young subjects with type I diabetes mellitus. The limitations of these models are: (1) the model structure is specific; (2) the number of parameters increases exponentially as the number of stages and covariates increases; (3) it is computationally expensive to evaluate the likelihood in the continuous time setting; (4) it is difficult to interpret the parameters in terms of the overall progression, since each element in the transition intensity matrix has its own set of parameters.

In this paper we propose a general discrete time ordinal Markov transition model, because the stages of a disease are usually ordinal. In the discrete time setting, the progression of the disease is completely specified by the one-step transition probability matrix. When the time scale is chosen fine enough, it can be viewed as an approximation to the transition intensity matrix in the continuous time. The structure of the one-step transition probability matrix defines the possible transition paths of the disease and need to be specified according to the knowledge of the disease or the experimental design. When a path in the one-step transition is not allowed, its associated one-step transition probability is defined as a structural zero. A structural zero in the one-step transition probability matrix does not imply the same structural zero in the observed transition probability matrix. For example, we can restrict the disease to progress one stage at a time and therefore a direct progression from stage 1 to stage 3 is prohibited in the one-step transition. This translates into a structural zero probability for a transition from stage 1 to stage 3 in the one-step transition matrix. However, when the observation interval is longer than the underlying discrete time scale, it is possible to observed a transition from stage 1 to stage 3, simply because we do not observe the transitions from stage 1 to stage 2 and from stage 2 to stage 3. We then model each row of the one-step transition probability matrix as a proportional odds model, which reduces the number of parameters needed to characterize the disease progression significantly and increase the interpretability of these parameters. This is further extended to restrict different rows of the one-step transition probability matrix to share the same set of covariates and with same slops. This further reduces the number of parameters and makes the interpretation of the parameters even clearer. Further model reductions by putting constraints on the intercepts are discussed in section 2. These model assumptions can be easily checked by likelihood ratio tests because they are nested models.

A special case of particular interest is when there are absorbing stage(s) such as death or dropout. In many cases, death may be the worst stage. We are not only interested in the time to death, but also interested in the progression of the disease. This is particularly useful for quality of life data. In the cases with informative dropouts, a patient may drop out because the disease gets worse or because of remedy. In these cases, it is possible to include dropout as the worst or the best stage. An absorbing stage is defined by having a probability one to stay in its current stage and all the transition probabilities to other stages being zero in the one-step transition probability matrix. The difference between the dropout and death in terms our model is that we usually observe the exact transition time in the cases with death, but not with dropouts. They contribute differently to the likelihood, which will be further explained in section 4.

In the following of this paper, we describe the general model in section 2 and its estimation procedure in section 3. We describe some special cases with absorbing stage(s) in section 4. In section 5, we illustrate the model by an application to the diabetic retinopathy in young subjects with insulin-dependent diabetes mellitus. Discussion about limitations and future works are in section 6.

2 The Multi-State Markov Transition Model

For a biological process with k ordinal disease states, the progression or regression of the disease within a small interval (one-step transition) depends on the m previous stages as well as covariates. Such biological process can be modeled by a mth order Markov process. In the continuous time setting, the progression is completely specified by the transition intensities [8]. Marshall and Jones [4] modeled these transition densities as functions of covariates. In the discrete time setting, the process is determined by the one-step transition matrix. We can then model the one-step transition probabilities as functions of covariates. For simplicity, we focus on first order Markov transition model in the current paper. Our method can be directly extended to incorporate high order Markov transition models by extending the dimension of the one-step transition matrix. When the one-step interval is chosen fine enough, the discrete time Markov process is a good approximation to the continuous time Markov process. However the finer the time grid, the more computationally expensive it is to estimate the parameters. In real practice, the grid is chosen large enough to ease the computational demand and small enough so that no more than one transition can happen within that interval.

Figure 1. The Multi-state Disease Process with associated one-step transition probabilities

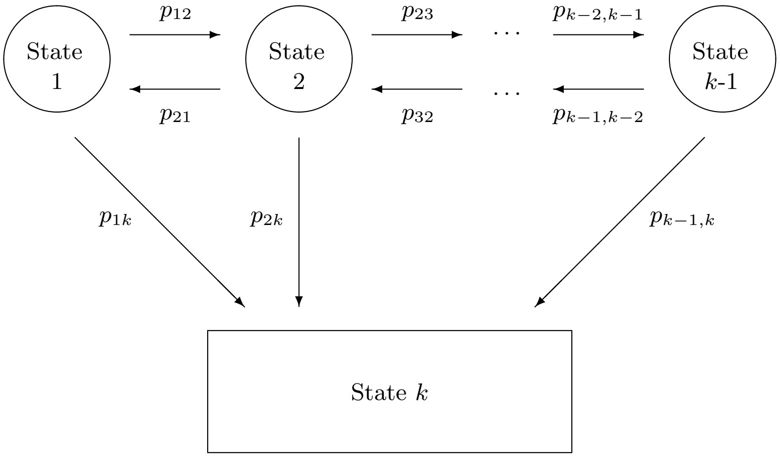

Figure 1 shows Markov process with k stages. The paths represent possible one-step transition among different stages. Some of the paths can be forbidden due to biological or physical beliefs. An absorbing stage is a stage whose outgoing paths to other stages are all forbidden. The transition paths can be represented by the one-step transition probability matrix as

$$\mathbf{P}(\mathbf{z}) = \begin{bmatrix} p_{11}(\mathbf{z}) & p_{12}(\mathbf{z}) & \cdots & p_{1k}(\mathbf{z}) \\ p_{21}(\mathbf{z}) & p_{22}(\mathbf{z}) & \cdots & p_{2k}(\mathbf{z}) \\ \vdots & \vdots & \ddots & \vdots \\ p_{k1}(\mathbf{z}) & p_{k2}(\mathbf{z}) & \cdots & p_{kk}(\mathbf{z}) \end{bmatrix}, \tag{1}$$

where $\mathbf{z}$ is a set of covariates at the current time and each row of the transition probability matrix satisfies the basic property $p_{i1}(\mathbf{z}) + p_{i2}(\mathbf{z}) + \ldots + p_{ik}(\mathbf{z}) = 1$ $(i = 1, \ldots, k)$. The n-step transition probability for a subject with vector of covariates $\mathbf{z}$, denoted by $p_{ij}^{(n)}(\mathbf{z})$, represents the probability that the process is in the state j after n-steps given that the process started on state i at time t, that is

$$p_{ij}^{(n)}(\mathbf{z}) = Pr\{X(t+n) = j | X(t) = i, \mathbf{z}\}. \tag{2}$$

This can be evaluated by computing the nth power of the transition probability matrix $\mathbf{P}(\mathbf{z})$. The transition probability $p_{ij}^{(n)}(\mathbf{z})$ is the element (i, j) of the product matrix $\mathbf{P}^n(\mathbf{z})$. The evaluation of $\mathbf{P}^n(\mathbf{z})$ can be obtained as

$$\mathbf{P}^n(\mathbf{z}) = \mathbf{A}(\mathbf{z}) \begin{bmatrix} \rho_1(\mathbf{z})^n & 0 & \cdots & 0 \\ 0 & \rho_1(\mathbf{z})^n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \rho_k(\mathbf{z})^n \end{bmatrix} \mathbf{A}^{-1}(\mathbf{z}), \tag{3}$$

where $\mathbf{A}(\mathbf{z})$ is the square matrix containing, in each column, the eigenvector associated with each of the eigenvalues, $\rho_i(\mathbf{z})$, of the transition probability matrix $\mathbf{P}(\mathbf{z})$.

From the formulation of the discrete-time Markov transition model, we can see that the model is characterized by the one-step transition probability matrix $P(\mathbf{z})$, which is indexed by the covariates $\mathbf{z}$. As mentioned in the introduction, the long-existing challenge for Markov transition models is to define an efficient parameterization so that the parameters can be stably estimated and have clear interpretations, because in the traditional parameterization there are k^2 elements in the one-step transition probability matrix $P(\mathbf{z})$ and each has its own set of parameters. In this section, we propose to model each row of $P(\mathbf{z})$ by a proportional odds model, and further reduce the model for the whole one-step transition probability matrix $P(\mathbf{z})$ to a "common slope model", which enforces different rows to share same slope parameters. This approach reduces significantly the number of parameters needed to characterize the progression and provides clear interpretations for these parameters in terms of the effects of the covariates on the disease progression. Further model reduction by putting constraint on the intercept parameters will also be discussed.

We define the cumulative one-step transition probability as

$$\gamma_{ij}(\mathbf{z}) = \Pr\left\{X(t+1) \leq j | X(t) = i \mid z\right\}.$$

Using a proportional odds model, the cumulative one-step transition probabilities can be written as:

$$\gamma_{ij}(\mathbf{z}) = \frac{\exp\{\theta_{ij} - \beta_i'\mathbf{z}\}}{1 + \exp\{\theta_{ij} - \beta_i'\mathbf{z}\}}, \tag{4}$$

for $j = 1, \ldots, k-1$, and $\gamma_{ik}(\mathbf{z}) = 1$. The intercept parameters θ_{ij} are restricted to be ordered as $\theta_{i1} < \theta_{i2} < \cdots < \theta_{i,k-1}$, to preserve the ordinal structure of the model and the definition of a probability measure. The one-step transition probabilities can be calculated from the cumulative one-step transition probabilities as:

$$p_{ij}(\mathbf{z}) = \begin{cases} \gamma_{ij}(\mathbf{z}) & \text{if } j = 1 \\ \gamma_{ij}(\mathbf{z}) - \gamma_{i,j-1}(\mathbf{z}) & \text{if } j \neq 1 \end{cases}$$

Model (4) can be further reduced to a common slope model:

$$\gamma_{ij}(\mathbf{z}) = \frac{\exp\{\theta_{ij} - \beta'\mathbf{z}\}}{1 + \exp\{\theta_{ij} - \beta'\mathbf{z}\}}, \tag{5}$$

which implies the covariates have the same impact on the progression regardless of the patient's current stage. This assumption can be checked by a likelihood ratio test comparing models (4) and (5). A further reduction of the number of parameters can be introduced by modeling the intercepts θ_{ij} as $\theta_{ij} = \theta_i \alpha_i^{(i-j)}$ where $0 < \alpha_i < 1$, which can also be tested by a likelihood ratio test.

3 Parameter Estimation and the Likelihood Function

First we denote θ as the whole collection of all unknown parameters, which include the slope and intercept parameters defined above. In this section we describe how to estimate these parameters via maximum likelihood.

Suppose we observe ith $(i = 1, \ldots, m)$ at time t_{ij} with the stage s_{ij} and covariates $\mathbf{z}_{ij}$ $(j = 1, \ldots, L_i)$. The interval between two consecutive observations is calculated as $n_{ij} = t_{ij} - t_{i,j-1}$. In the absence of absorbing states, the log-likelihood can then be calculated as

$$l(\theta) = \sum_{i=1}^{m} \sum_{j=1}^{L_i} log\{p^{n_{ij}}_{s_{i,j-1}, s_{i,j}}(\mathbf{z}_{ij})\}, \tag{6}$$

where the transition probability $p^{n_{ij}}_{s_{i,j-1}, s_{i,j}}(\mathbf{z}_{ij})$ takes into account that multiple transitions can have happened within the interval $[s_{i,j-1}, s_{i,j}]$ and we do not observe the exact transition time.

In the presence of absorbing stages such as death, we may observe the exact transition time. The contribution to the likelihood needs to take into account this additional information explicitly. Suppose that we observe the exact transition time of *ith* subject to an absorbing stage s_{ij}. Using a similar formulation by Kay (1986), we can write the contribution to the likelihood function as

$$\bar{p}^{(n_{i,j})}_{s_{i,j-1},s_{ij}}(\mathbf{z}_{ij}) = \sum_{l \neq s_{ij}} p^{(n_{ij}-1)}_{s_{i,j-1},l}(\mathbf{z}_{ij}) p_{l,s_{ij}}(\mathbf{z}_{ij}). \tag{7}$$

Note that this contribution term explicitly forces the transition to the absorbing stage to happen within the last unit of time. We will further discuss some applications with absorbing stages in section 4.

The following outlines the steps in calculating the log-likelihood:

1) For a given value of θ, use model (4) or (5) to calculate cumulative one-step transition probabilities $\gamma_{ij}(\mathbf{z}_{ij}, \theta)$.

2) Calculate one-step transition probabilities $p_{ij}(\mathbf{z}_{ij}, \theta)$ from $\gamma_{ij}(\mathbf{z}_{ij}, \theta)$.

3) If we do not observe the exact transition time, calculate $\mathbf{P}^{n_{ij}}(\mathbf{z}_{ij}, \theta)$ using equation (3). The $(s_{i,j-1}, s_{ij})$ element of the $\mathbf{P}^{n_{ij}}(\mathbf{z}_{ij}, \theta)$ is the contribution to the likelihood from this transition.

4) If we observe the exact transition time, calculate $\mathbf{P}^{(n_{ij}-1)}(\mathbf{z}_{ij}, \theta)$ or using equation (3), and use equation (7) to calculate $\bar{p}_{s_{i,j-1},s_{ij}}$, which is the contribution to the likelihood from this transition.

5) Repeat step (2)-(4) for all transitions and sum over all the log transition probabilities using equation (6).

The log-likelihood can then be numerically maximized using a Quasi-Newton algorithm. Initial values can be critical in the process of finding the maximum likelihood estimates of the parameters. In the context of a continuous-time Markov model, a way to obtain the initial values was proposed by [1] to use the estimates of the parameters found by assuming that the observed times are the actual transition time of the process. We adopt the similar idea here. The details on how to obtain the initial values is given in the appendix.

4 Survival Curves

The survival function defined as the probability to stay in one of the transient state at time t,

$$S_i(t; z) = \Pr\left\{X(t) < k | X(0) = i \mid \mathbf{z}\right\},$$

can be defined in the context of this discrete Markov model as $S_i(t;z) = 1 - p_{ik}^{(t)}(\mathbf{z})$. The overall survival function can be defined as

$$S(t;\mathbf{z}) = \sum_{i=1}^{k-1} \pi_i S_i(t;\mathbf{z}),$$

where π_i represents the probability to be at state i at any arbitrary time $t = 0$

5 Diabetic Retinopathy: An Example.

Two hundred seventy-seven subjects who had Type 1 diabetes for at least five years when initially seen in the Eye-Kidney Clinic at Barbara Davis Center for Childhood Diabetes were included in the analysis described by Marshall, Garg, Jackson et al. [7] using a continuous-time Markov model. In this study all subjects were seen at least twice with visits one or more years apart for a mean follow-up of almost 3 years. A grading of retinal findings for each visit was perform by a retinal specialist. The grades were based on the modified Airlie House classification in which grade I indicates no retinopathy; grade II indicates micro aneurysms only; grade III and IV indicate intermediate stages of background retinopathy, and grades V and VI indicate preproliferative and proliferative retinopathy, respectively. The worst eye grade was used for defining the state of the process in the Markov model. The Markov model was define regrouping the stages as follow:

$$\text{Grades I} \Leftrightarrow \text{Grades II-III} \Leftrightarrow \text{Grades IV-V} \Rightarrow \text{Grade VI.}$$

Different clinical factors related to the physiology and history of the disease were evaluated with respect to the influence on the onset, progression and regression of diabetic retinopathy. The most important factors related to changes in retinopathy were duration of diabetes, historical values of glycohemoglobin (HbA1) and distolic blood pressure.

A single-covariate, discrete-time Markov model was used to assess the individual effects of factors associated with diabetic retinopathy using a custom-designed computer program . The model with three regression coefficients (full model) and the model assuming common regression coefficients (restricted model) were fitted to each of the 10 factors considered in this study. Table 1 shows the fit of each model in terms of the -2 log likelihood function, the likelihood ratio test versus the null model, and the significance of the effect of each factor for both the full and restricted models. If a factor was found significantly associated with the progression or regression of diabetic retinopathy based on each of these models, the parsimonious representation of this factor was later used for multiple regression analysis.

The duration of diabetes, the age of the subject, and the mean HbA1 levels were factors most associated with transitions of diabetes retinopathy. Diastolic and systolic blood pressure, values of HbA1 at visit times, and smoking were also associated with changes in diabetic retinopathy. All factors except duration of diabetes and age are better represented by using the restrictive model that uses a common regression coefficient for all different transient states. More importantly, systolic blood pressure and smoking were found to be significantly associated with the disease process only when the restricted model was used, showing that the full model overfit the data and reduces the effect of these two factors by unnecessarily adding extra parameters.

The estimates of the regression coefficients of duration of diabetes are

$$(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) = (0.0539, 0.1152, 0.2165),$$

showing that the effect of duration increases over the course of the disease, in fact it doubles the effect from state 1 to state 2 and from state 2 to state 3.

Table 1. The -2 log-Likelihood function, the likelihood ratio test, and the P-value of single-covariable discrete-time Markov models for various factors associated with diabetic retinopathy using the full model (4), and the restrictive model $\boldsymbol{\beta}_i = \boldsymbol{\beta}$

.

	Full Model			Restrictive Model		
Factor	-2$l(\theta)$	$\chi^2(3)$	p-value	-2$l(\theta)$	$\chi^2(1)$	p-value
Duration of Diabetes	886.1	54.0	<0.001	892.5	47.6	<0.001
Age	912.4	27.8	<0.001	920.3	19.8	<0.001
Mean HbA$_{1c}$	916.6	23.6	<0.001	917.8	22.3	<0.001
Diastolic Blood Pressure	929.3	10.8	0.013	929.4	10.7	0.001
HbA$_{1c}$ at the Visit	929.4	10.7	0.013	931.0	9.1	0.003
Systolic Blood Pressure	933.3	6.8	0.079	933.8	6.3	0.012
Gender	934.6	5.6	0.136	938.5	1.6	0.203
Cholesterol	935.5	4.6	0.204	936.9	3.2	0.074
Smoking	936.0	4.1	0.251	936.2	3.9	0.048
Family Hx Hypertension	938.3	1.8	0.611	940.0	0.1	0.752

Table 2 gives the estimates and the standard errors of the estimates for the parameters of the multiple regression model. Duration of diabetes remained the most important factor for explaining changes in diabetic retinopathy. As expected, cumulative mean values of HbA1 were the second most important clinical variable associated with transitions in retinopathy. Finally, diastolic blood pressure also remained in the model after adjusting for duration and HbA1 values, showing that it is an independent factor associated with the progression and regression of diabetic retinopathy.

Table 2. Parameter estimates and standard errors for the best multiple regression model.

Factor	Parameter	Estimate	Standard Error
Constant	θ_{11}	7.643	0.7654
Constant	θ_{21}	1.172	0.7506
Constant	θ_{22}	9.760	0.0358
Constant	θ_{32}	3.775	1.1420
Constant	θ_{33}	13.24	0.0639
Duration of Diabetes	β_{11}	0.0667	0.0368
Duration of Diabetes	β_{12}	0.1286	0.0319
Duration of Diabetes	β_{13}	0.2193	0.0660
Mean HbA$_{1c}$	$\beta_{21} = \beta_{22} = \beta_{23}$	0.2364	0.0403
Diastolic Blood Pressure	$\beta_{31} = \beta_{32} = \beta_{33}$	0.1856	0.0058

6 Discussion

The use of the ordinal structure of the disease stages by introducing a proportional odds model in a discrete-time Markov chain successfully reduced the number of parameters involved in a multi-state disease model, and particularly compared to the representation of this model by a continuous-time Markov process. The estimation process and the inferences about the effect of factors affecting the progression and regression of the disease process gained in stability and consistency as shown with the example of diabetic retinopathy.

The results of our analysis of diabetic retinopathy shows that most of the factors related to the disease process can be represented by a simple restrictive model with only one parameter, and given the ordinal nature of this disease the loss of information is minimal. The other two factors were well-represented by the full model with three parameters which is still a reduction of the number of regression coefficients when compared to the continuous-time model.

By using the restrictive model in our example we found that two factors, smoking and systolic blood pressure, were significantly associated with changes in diabetic retinopathy. This type of finding is likely to occur in other applications and in those factors that while are not strongly associated with the disease process are still significant to clinicians.

This discrete model is a step forward to provide a feasible and meaningful methodology for the analysis of transitional data when exact times are not available, particularly to applications dealing with a large number of covariates and where the purpose is to find the best multiple regression model. More research will have to be done to compare the properties of this model with other link functions and to study the properties of the proportionality test using a continuous-time model as the reference.

References

1. Kay. R. (1986). A Markov Model for Analyzing Cancer Markers and Disease States in Survival Studies. Biometrics 42: 855-865.
2. Longini Jr., I. M., Clark, W. S., Byers, R. H., et al. (1989) Statistical Analysis of the Stages of HIV Infection Using a Markov Model. Statistics in Medicine 8: 831-843.
3. Marshall G. and Jones R. H. (1993). Multi-State Survival Models. Technical Report. Department of Preventive Medicine and Biometrics, University of Colorado.
4. Marshall G. and Jones R. H. (1995). Multi-State Markov Models and Diabetes Retinopathy. Statistics in Medicine. 14: 1975-1983
5. Marshall G., Guo W., Jones R. H. (1995). Markov: A Computer Program for Multi-State Markov Models with Covariables. Computers Methods and Programs in Biomedicine. 47: 147-156
6. Garg, S.K., Marshall, G., Chase, H. P., et al. (1990). The Use of the Markov Processes in Describing the Natural Course of Diabetic Retinopathy. Arch. Ophthalmology, 108: 1245-1247.
7. Marshall, G Garg, S. K. Jackson, W. E., et al. (1993) Factors Influencing the Oncet and Progression of Diabetic Retinopathy in Subjects with Insulin-Dependent Diabetes Mellitus. Ophthalmology, 100: 1133-39
8. Chiang, Ch. L. (1968) Introduction to Stochastic Processes in Biostatistics. Wiley, & Sons, Inc., New York, London,1968.
9. Dennis Jr., J. E., Schnabel, R. B. (1983). Numerical Methods for Unconstrained Optimization and Nonlinear Equations. Prentice-Hall, Englewood Cliffs, New Jersey (1983).
10. Cox, D. R. (1972). Regression Models and Life-Tables (with discussion). J.R. Statist. Soc.,B, 34: 187-220.