
Charles J. Mode
Department of Mathematics and Computer Science
Drexel University
Philadelphia, PA 19104

Overview of Topics.

I. Examples of State Spaces.

A first step in developing a model based on semi-Markovian and related ideas is to consider an appropriate state space for the situation under investigation.

II. Modeling Transition Density Matrices.

After the state space has been chosen, the next step to take is that modelling the transition densities. For example, shall the sojourn time distributions in states be exponentials or some other parametric distribution. Will covariates need to be taken into account?

III. Computational and Computer Intensive Methods.

After choosing functional forms for the sojourn time distributions, some decisions will need to be made as to which computational and computer intensive methods will be appropriate.
I. Examples of State Spaces.

1. Illness - Death Process – Fix and Neyman

Imagine a set of patients under observation at some clinic and consider their evolution in time with respect to some disease. One of the simplest state spaces for such a situation is as follows.

- **Absorbing States** - Observations Terminate.
  
  $E_1$ - Patient Died.
  $E_2$ - Patient Lost to Follow Up.

- **Transient States**.
  
  $E_3$ - Patient Ill and Under Treatment.
  $E_4$ - Patient Well, For the case of cancer or arthritis in remission.
2. Patients with HIV/AIDS.

Imagine a set of individuals who have been infected with HIV, the virus that causes AIDS. Then, the following type of state space has not only been used in clinical studies but also in stochastic models who purpose is to provide frameworks for studying HIV/AIDS epidemics at a population level.

• Absorbing States.

$E_{11}$ - Death From a Cause Other Than AIDS.
$E_{12}$ - Death Due to an AIDS Defining Disease.
$E_{13}$ - Patient Cleared of HIV by Treatment with Anti-Viral Drugs.

• Transient States.

$E_{2j}$ - Classes $j = 1, 2, \cdots, n$ of CD4$^+$ Counts of T-Lymphocytes Used to Define Clinical Stages of HIV Disease.
3. Actual Clinical Trial on Drugs to Treat Rheumatoid Arthritis (D. DiPrimeo) First Attempt.

American College of Rheumatology (ACR) has established some standards for measuring response to drugs. For example, a ACR-20 responder is a patient who experiences at least a 20% reduction in both painful and swollen joints, and a 20% reduction in at least 3 of 5 efficacy parameters: patient opinion, investigator opinion, disability function, and sedimentation rate. Data on 1,400 to 1,500 patients are available for study.

- Absorbing States.

\[ E_{11} \] - Withdrawn From Study for an Adverse Reaction to Drug (ADR).
\[ E_{12} \] - Withdrawn From Study due to Lack of Efficacy (LOE).
\[ E_{13} \] - Withdrawn From Study for Other Reasons.
\[ E_{14} \] - Completed Study.

- Transient States.

\[ E_{21} \] - Patients Who are not ACR 20 responders. By definition and trial design, all patients enter study in this state.
\[ E_{22} \] - Patients Who are ACR 20 Responders but not ACR 50 Responders.
\[ E_{23} \] - Patients Who are ACR 50 Responders but not ACR 70 Responders.
\[ E_{24} \] - Patients Who are ACR 70 Responders.
• Covariates.

1. Age  
2. Sex  
3. Race  
4. Weight  
5. Duration of Disease  
6. Concomitant Steroid Use  
7. Prior Anti-Rheumatic Drug  
8. Rheumatoid Factor + or -  
9. Sedimentation Rate.  
10. Treatment Group  

Some covariates, such as sedimentation rate, may change in time for each patient, which gives rise to the problem of dealing with time varying covariates. Clinicians often group sedimentation rates into discrete classes such as low, medium, and high. One approach to handling discrete time varying covariates is to suppose the transient states may be described by a pair \((k_1, k_2)\) of indicators, where \(k_1\) is the level ACR Response and \(k_2\) is the sedimentation rate class. If this approach were used, the set of transient states would contain 12 elements.
4. Second Attempt - Simplified State Space Based on Paulus Criteria.

- Absorbing States.

\[ E_{11} \] - Withdrawn for Lack of Efficacy
\[ E_{12} \] - Withdrawn for Adverse Reaction
\[ E_{13} \] - Withdrawn for Other Reasons.

- Transient States.

\[ E_{21} \] - Non-Response
\[ E_{22} \] - Response
II. Modeling Transition Density Matrices.

• Overview of Problem.

Let $\mathcal{S}_1$ denote a set of $m_1 \geq 1$ absorbing states, let $\mathcal{S}_2$ denote a set of $m_2 \geq 1$ transient states, let $m = m_1 + m_2$ be the total number of states, and let $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2$ denote the state space. In most applications, the set $\mathcal{S}_2$ of transient states is an irreducible class of communicating states. The problem is to construct a $m \times m$ matrix of density functions

$$a(t) = (a_{ij}(t))$$

such that if at time $t = 0$ the process is in state $i \in \mathcal{S}$, then

$$A_{ij}(t) = \int_0^t a_{ij}(s)ds$$

is the conditional probability of a jump to state $j \in \mathcal{S}$ during the time interval $(0, t], \ t > 0$. The time set $\mathbb{T}$ may be continuous

$$\mathbb{T} = [0, \infty) = [t \in \mathbb{R} \mid t \geq 0]$$

or a discrete lattice

$$\mathbb{T} = [th \mid t = 0, 1, 2, \ldots] ,$$

where $h > 0$.

• Because no exits from an absorbing state are possible, the $m \times m$ density matrix $a(t)$ may be represented in the partitioned form

$$a(t) = \begin{bmatrix} 0_{m_1m_1} & 0_{m_1m_2} \\ r_{m_1m_2}(t) & q_{m_2m_2}(t) \end{bmatrix} \text{ for } t \in \mathbb{T}.$$
• Competing Risks in Continuous Time.

When using the classical theory of competing risks to derive explicit formulas for the elements of the density matrix \( \mathbf{a}(t) \), the starting point may be the specification of a \( m_2 \times m \) matrix

\[
\Theta(t) = (\theta_{ij}(t))
\]

of non-negative latent risks, where \( t > 0 \) and \( \theta_{ii}(t) = 0 \) for all \( t > 0 \). Then, in the ”absence” of other competing risks, the conditional probability that if the process is in transient state \( i \in \mathcal{S}_2 \) at time \( t = 0 \), then the conditional probability of a jump to state \( j \in \mathcal{S} \) during the time interval \((0, t]\) is latent distribution function

\[
F_{ij}(t) = 1 - \exp \left[ - \int_0^t \theta_{ij}(s) ds \right]
\]

of a latent random variable \( T_{ij} \). For each latent random variable there is a latent survival function

\[
S_{ij}(t) = 1 - F_{ij}(t) = \exp \left[ - \int_0^t \theta_{ij}(s) ds \right].
\]

Moreover, if it is assumed that for each transient state \( i \in \mathcal{S}_2 \) the latent random variables \( \{T_{ij} \mid j \neq i\} \) are independent, then the survival function

\[
S_i(t) = \prod_{j \neq i} S_{ij}(t)
\]

is the conditional probability that the process is still in transient state \( i \in \mathcal{S}_2 \) at time \( t > 0 \), given that is was in state \( i \) at time \( t = 0 \).
If the process is in transient state $i$ at time $t$, then during a small time interval $(t, t + dt]$, $\theta_{ij}(t)dt$ is the conditional probability of a jump to state $j \neq i$. Hence, by integrating it can be seen that

$$A_{ij}(t) = \int_0^t S_i(u)\theta_{ij}(u)du$$

is the conditional probability of a jump to state $j \in \mathcal{S}$ during $(0, t]$, given that the process was in state $i \in \mathcal{S}_2$ at time $t = 0$. Thus, the non-zero elements of the density matrix $a(t)$ have the general form

$$a_{ij}(t) = S_i(t)\theta_{ij}(t).$$

Equivalently, if, by definition,

$$\theta_i(t) = \sum_{j \neq i} \theta_{ij}(t)$$

is the total risk function for transient state $i \in \mathcal{S}_2$, then an convenient alternative form of a density is

$$a_{ij}(t) = \exp \left[ - \int_0^t \theta_i(s)ds \right] \theta_{ij}(t).$$

Let $P = (p_{ij})$ denote the transition matrix for the Markov chain embedded in the semi-Markov process. Then for $i \in \mathcal{S}_2$ and $j \in \mathcal{S}$ the element $p_{ij}$ is given by

$$p_{ij} = \lim_{t \to \infty} A_{ij}(t) = \lim_{t \to \infty} \int_0^t S_i(u)\theta_{ij}(u)du.$$

In many cases, a numerical integration procedure will be needed to evaluate this limit.
Before proceeding it is appropriate to pause and give a simple example. For the case of the illness-death process of Fix and Neyman, there are \( m_1 = 2 \) absorbing states, \( m_2 = 2 \) transient states so that the \( 2 \times 4 \) matrix \( \Theta(t) \) of latent risks has the form

\[
\Theta(t) = \begin{bmatrix}
\theta_{31}(t) & \theta_{32}(t) & 0 & \theta_{34}(t) \\
\theta_{41}(t) & \theta_{42}(t) & \theta_{43}(t) & 0
\end{bmatrix}.
\]

Among the choices of functional form for these risk functions is that for a Weibull type distribution

\[
\theta_{ij}(t) = \alpha_{ij} \beta_{ij} t^{\alpha_{ij}-1},
\]

where \( \alpha_{ij} \) and \( \beta_{ij} \) are positive parameters and \( t > 0 \). If all these parameters are distinct, then a 12 parameter model would arise. So there is frequently a need to reduce the number of parameters. One approach is to suppose that the scale type parameters \( \beta_{ij} = \beta \), a constant for all \( i \) and \( j \). Under this assumption, the model would depend on 7 parameters.
The Case of Constant Latent Risks.

An interesting case arises when there are constants $\theta_{ij}$ such that

$$\theta_{ij}(t) = \theta_{ij}$$

for all $t > 0$, $i \in \mathcal{S}_2$ and $j \in \mathcal{S}$. For in this case, the total risk functions $\theta_i(t) = \theta_i$ are also constant for all transient states $i \in \mathcal{S}_2$ and

$$A_{ij}(t) = \int_0^t e^{-\theta_i s} \theta_{ij} ds = \left(1 - e^{-\theta_i t}\right) \frac{\theta_{ij}}{\theta_i}.$$

By construction, $\theta_i > 0$ for all transient states $i \in \mathcal{S}_2$. Therefore, the distribution function of a sojourn time in state $i$ has the exponential form

$$A_i(t) = \sum_{j \neq i} A_{ij}(t) = 1 - e^{-\theta_i t}$$

for $t \geq 0$. Hence, a semi-Markov process with constant latent risks is equivalent to a Markov jump process in continuous time. It is also easy to see that for $i \in \mathcal{S}_2$ and $j \in \mathcal{S}$ the transition probabilities of the embedded Markov chain have the simple form

$$p_{ij} = \lim_{t \to \infty} A_{ij}(t) = \frac{\theta_{ij}}{\theta_i},$$

which may be easily calculated.
Mixtures of Distributions as Alternatives to Competing Risks.

For $t$ large, many expectations and variances of random variables that arise in the analysis of semi-Markov process may be computed in terms of the transition probabilities of the embedded Markov chain. It is, therefore, of interest to consider some alternatives to competing risks in modeling density functions so as to avoid problems that may arise in computing these probabilities by numerical integration. If $\theta_{ij}$ is a constant positive latent risk, then it can also be interpreted as a parameter in an exponential distribution such if $T_{ij}$ is a random variable with this distribution, then its expectation is

$$E\left[T_{ij}\right] = \mu_{ij} = \frac{1}{\theta_{ij}}.$$ 

Thus, an alternative way of writing the above formula for a transition probability is

$$p_{ij} = c_i \frac{1}{\mu_{ij}},$$

where $c_i$ is the normalizing constant

$$c_i = \sum_{j \neq i} \frac{1}{\mu_{ij}}.$$

Now suppose at time $t = 0$ the process is in transient state $i$ and let $F_{ij}(t)$ denote the conditional "latent" distribution function of the waiting time to a jump to state $j \neq i$, given that this jump occurs eventually. Let $\mu_{ij}$ the conditional expectation of this waiting time, and let $p_{ij}$ be the conditional probability of an eventual jump to state $j$. It seems plausible to assume that $p_{ij}$ is inversely proportional to $\mu_{ij}$. Hence, the above formula could be used to compute $p_{ij}$ and the function $A_{ij}(t)$ could be modeled as

$$A_{ij}(t) = F_{ij}(t)p_{ij}.$$
If \( f_{ij}(t) \) is the probability density function of distribution function \( F_{ij}(t) \), then

\[
a_{ij}(t) = f_{ij}(t)p_{ij},
\]

and the density of sojourn times in transient state \( i \) is the mixture

\[
a_i(t) = \sum_{j \neq i} f_{ij}(t)p_{ij}.
\]

One of the advantages of this formulation is that it holds for both continuous and discrete time. For example, the densities \( f_{ij}(t) \) could belong to a well known family of discrete type distributions such as the Poisson or negative binomial. In discrete time formulations, it is reasonable to suppose that no jumps can occur instantaneously so that \( f_{ij}(0) = 0 \) and the support of the "latent" random variable \( T_{ij} \) is \( t = 1, 2, 3, \ldots \). For example, suppose the density \( f_{ij}(t) \) is chosen as a Poisson

\[
f_{ij}(t) = \frac{e^{-\lambda_{ij}t} (\lambda_{ij}t)^{t-1}}{(t-1)!},
\]

where \( t = 1, 2, \ldots \). Then, the expectations

\[
E[T_{ij}] = \mu_{ij} = 1 + \lambda_{ij}
\]

could be used to determine the transition probabilities \( p_{ij} \) of the embedded Markov chain according to the above formula.
• Multiple Decrement Life Table Algorithms - Discrete Time.

These ideas have been used in demographic studies and occasionally in biostatistics. Given that the process is in state \( i \in \mathcal{S}_2 \) at time \( t \), let \( q_{ij}(t) \) be the conditional probability that there is a jump to state \( j \in \mathcal{S} \) during the time interval \((t, t + 1]\). Then,

\[
q_i(t) = \sum_{j \neq i} q_{ij}(t)
\]

is the conditional probability there is a jump from \( i \) during the time interval and \( p_i(t) = 1 - q_i(t) \) is the conditional probability the process remains in state \( i \) during \((t, t + 1]\). Be definition, \( p_i(0) = 1 \).

Given that the process is in state \( i \) at time \( t = 0 \),

\[
S_i(t) = \prod_{\nu=1}^{t} p_i(\nu)
\]

is the conditional probability the process is still in state \( i \) at time \( t > 0 \). In this discrete time approach to modeling the transition density \( a_{ij}(t) \), for \( j \neq i \), has the form

\[
a_{ij}(t) = S_i(t - 1)q_{ij}(t).
\]

It is the conditional probability that the process jumps to state \( j \) during the time interval \((0, t]\) for \( t = 1, 2, \cdots \), given that the process was in state \( i \) at time \( t = 0 \). Some remarks on modelling the conditional probabilities \( q_{ij}(t) \) will be made later. It should be mentioned that in some applications, it will be possible to obtain non-parametric estimates of the risk functions \( q_{ij}(t) \).
III. Incorporating Covariates Into the Model.

• The Case of Constant Latent Risks.

In this case, it seems reasonable to borrow some ideas from loglinear models in statistics. For example, suppose there are \( k \geq 1 \) covariates under consideration \( x_1, x_2, \ldots, x_k \). Then the latent risk \( \theta_{ij} \) could be chosen as the function

\[
\theta_{ij}(x) = \exp \left[ \beta_{ij0} + \sum_{\nu=1}^{k} \beta_{ij\nu} x_{\nu} \right],
\]

where the parameters \( \{ \beta_{ij\nu} \mid \nu = 0, 1, 2, \ldots, k \} \) are to be estimated from data and

\[
x = (1, x_1, \ldots, x_k),
\]

is a vector of covariates. These ideas would probably work well only for relatively small state spaces and a few covariates. A preliminary analysis would be required to determine which covariates are significant. This procedure would also work for the case of discrete time mixture models with Poisson type densities.

• Polytomous Logits.

Polytomous logits are multivariate generalizations of the logit function that is used in statistics under the rubric, logistic regression. Suppose, for example, an investigator wishes to express the transition probabilities \( p_{ij} \) of a Markov chain as functions of covariates. For each vector of regression parameters

\[
\beta_{ij} = (\beta_{ij\nu} \mid \nu = 0, 1, 2, \ldots, k)
\]

and vector \( x \) covariates, let

\[
\eta_{ij}(\beta_{ij}, x) = \exp \left[ \beta_{ij0} + \sum_{\nu=1}^{k} \beta_{ij\nu} x_{\nu} \right].
\]
Then, for every transient state $i \in \mathcal{S}_2$ and state $j \in \mathcal{S}$,$$
p_{ij} = \frac{\eta_{ij}(\beta_{ij}, x)}{\sum_{\nu=1}^{m} \eta_{i\nu}(\beta_{i\nu}, x)}.
$$Such software packages as SAS have procedures for estimating the regression parameters. These procedures work relatively well using maximum likelihood methodologies, because log-likelihood functions, expressed in terms of polytomous logits, are concave.

In the rheumatoid arthritis study, polytomous logits are being used in preliminary screening procedures to determine what covariates are important. The preliminary model is a Markov chain, and the discrete time points of the Markov chain are successive visits to a clinic by patients.

For the case of a multiple decrement life table formulation, when the conditional probabilities $q_{ij}(t) = q_{ij}$ do not depend on $t$, the formula

$$q_{ij} = \frac{\eta_{ij}(\beta_{ij}, x)}{\sum_{\nu=1}^{m} \eta_{i\nu}(\beta_{i\nu}, x)}$$

to incorporate covariates into the model. Relatively little attention has been given to problem of modelling the discrete risk functions $q_{ij}(t)$ when they depend on time. But, this is a problem worthy of attention.
• Cox-Type Models of Competing Risks.

When latent risks depend on time, then it may be feasible to use a Cox-Type model with covariates. For example, suppose for each pair of states $i \in \mathcal{S}_2$ and $j \in \mathcal{S}$, some risk function $\eta_{ij}(t)$ has been specified. Then, the latent risks in a classical model of competing risks could be modeled as

$$\theta_{ij}(t; \beta, x) = \eta_{ij}(t) \exp \left[ \sum_{\nu=1}^{k} \beta_{ij\nu} x_{\nu} \right].$$

In principle, one can express a transition density as a function of unknown $\beta$-parameters and a vector $x$ of covariates, using these formulas for latent risks and the classical theory of competing risks. Thus,

$$a_{ij}(t; \beta, x) = S_i(t; \beta, x) \theta_{ij}(t; \beta, x)$$

for $i \in \mathcal{S}_2$ and $j \in \mathcal{S}$. For any transient state $i \in \mathcal{S}_2$, results in the literature on estimating the unknown $\beta$-parameters by maximum likelihood procedures may be used. But, in any practical application of semi-Markov processes, extensive numerical integrations would need to be carried out in implementing this approach to including covariates in the model.
IV. Computational and Computer Intensive Methods.

- Objectives and Applications.

1. Estimation of Unknown Parameters Given a Data Set.

   When using the method of maximum likelihood to estimate unknown parameters, for example, quite general procedures for finding a maximum of a log-likelihood function are available in such well known software packages as SAS, S-PLUS, and MATLAB to name only three.

2. Plots of Current State Probabilities.

   Given that the process starts in some transient state \( i \in S_2 \), let \( P_{ij}(t; x) \) denote the conditional probability that the process is in some state \( j \in S \) as a function of a vector \( x \) of covariates that do not depend on time. For every \( i \in S_2 \),

   \[
   \sum_{j \in S} P_{ij}(t; x) = 1
   \]

   for all \( t \geq 0 \) and \( x \). For a fixed initial transient state \( i \in S_2 \), it would be informative for clinicians to have plots of the collection of probabilities

   \[
   \{ P_{ij}(t; x) \mid j \in S \}
   \]

   as functions of \( t \geq 0 \) for selected vectors of covariates \( x \). Different values of the vector \( x \) of covariates would correspond to various risk groups in a heterogeneous population. For example, for the case of a clinical trial to test the efficacy of drugs to treat rheumatoid arthritis, different values of \( x \) may represent age groups or those who were positive or negative for the rheumatoid factor.

Because the current state probabilities will need to be computed for several values of a vector of covariates $\mathbf{x}$, efficient methods for computing current state probabilities will be required. There are two cases to consider.

3.1. Latent Risk Functions Do Not Depend on Time.

For this case, the exponential matrix may be used to compute the current state probabilities as functions of time and a vector of covariates. An example will be given.

3.2. Latent Risk Functions Do Depend on Time.

In this case, the current state probabilities satisfy systems of renewal type integral equations that will need to be solved numerically. Some numerical methods will be discussed briefly.
The Case Latent Risk Functions Do Not Depend on Time.

If the latent risk functions depend only on a vector $\mathbf{x}$ of covariates but not on time, then the distribution of the sojourn time in a transient state $i \in \mathcal{S}_1$ is exponential with parameter $\theta_i(\mathbf{x})$, the total risk for this state as a function of $\mathbf{x}$. Then for each vector $\mathbf{x}$ the semi-Markov process is equivalent to a continuous time Markov jump process. Let $\mathbf{P}(t; \mathbf{x}) = (P_{ij}(t; \mathbf{x}))$ denote the $m \times m$ matrix of current state probabilities for this continuous time Markov jump process. Then, there is a $m \times m$ matrix $\mathbf{Q}(\mathbf{x})$, depending on $\mathbf{x}$ such that the matrix $\mathbf{P}(t; \mathbf{x})$ satisfies the Kolmogorov differential equations

$$\frac{d\mathbf{P}(t; \mathbf{x})}{dt} = \mathbf{P}(t; \mathbf{x})\mathbf{Q}(\mathbf{x}) = \mathbf{Q}(\mathbf{x})\mathbf{P}(t; \mathbf{x})$$

for $t \geq 0$ with the initial condition

$$\mathbf{P}(0) = \mathbf{I}_m ,$$

where $\mathbf{I}_m$ is a $m \times m$ identity matrix. As is well known, the unique to solution to this finite system of equations is the well-known exponential matrix

$$\mathbf{P}(t; \mathbf{x}) = \exp \left[ \mathbf{Q}(\mathbf{x})t \right] ,$$

which needs to be computed for selected values of the vector $\mathbf{x}$. 

• Computation of Exponential Matrix.

To carry out these computations, the matrix $Q(x)$ needs to be expressed in terms of the $m_2 \times m$ matrix $\Theta(x)$ of latent risks for the semi-Markov process. To this end, define a $m \times m$ matrix $\Theta^*(x)$ by the partitioned form

$$\Theta^*(x) = \begin{bmatrix} 0_{m_1m} \\ \Theta(x) \end{bmatrix} = (\theta_{ij}^*(x)),$$

where $0_{m_1m}$ is a zero matrix, and let

$$D(x) = \text{diag} (\theta_1^*(x), \cdots, \theta_m^*(x))$$

be a $m \times m$ diagonal matrix with the diagonal elements

$$\theta_i^*(x) = \sum_{j \neq i} \theta_{ij}^*(x)$$

for $i = 1, 2, \cdots, m$. Then,

$$Q(x) = \Theta^*(x) - D(x).$$

With the matrix $Q(x)$ so defined, the exponential matrix has the partitioned form

$$P(t; x) = \exp [Q(x)t] = \begin{bmatrix} I_{m_1} & 0_{m_1m_2} \\ P_{21}(t) & P_{22}(t) \end{bmatrix}.$$

SAS and other software packages such as MATLAB may be used to compute the exponential matrix for selected values of $t$. 

21
• A Simple Numerical Example of the Illness-Death Process.

If matrix $\Theta^*(x)$ has the form

$$\Theta^*(x) = \begin{bmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
\frac{1}{10} & \frac{2}{10} & 0 & \frac{3}{10} \\
\frac{1}{20} & \frac{2}{20} & \frac{3}{20} & 0
\end{bmatrix},$$

then

$$Q(x) = \begin{bmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
\frac{1}{10} & \frac{2}{10} & -\frac{6}{10} & \frac{3}{10} \\
\frac{1}{20} & \frac{2}{20} & \frac{3}{20} & -\frac{6}{20}
\end{bmatrix}.$$

And

$$P(10) = \begin{bmatrix}
1.0 & 0 & 0 & 0 & 0 \\
0 & 1.0 & 0 & 0 & 0 \\
0.2940 & 0.5880 & 3.2198 \times 10^{-2} & 8.5710 \times 10^{-2} & 0 \\
0.2797 & 0.5594 & 4.2855 \times 10^{-2} & 0.11790 & 0
\end{bmatrix}.$$ 

This example was done by MAPLE, a computer algebra package, which is linked to the word processor used to prepare this presentation.
Maximum Likelihood Estimation Based on Exponential Matrix.

Whenever software is available to compute the exponential matrix efficiently and the model being used is a Markov jump process in continuous time, maximum likelihood procedures of estimation become feasible. For example, suppose \( n \) patients are under study in a clinical trial and the \( i \)-th patient has the covariate vector \( \mathbf{x}_i \). Let \( t_{i0} < t_{i1} < t_{i2} < \cdots < t_{in_i} \) be the times the \( i \)-th patient visits the clinic and suppose at the \( \nu \)-th visit the patient is observed in state \( k_{i\nu} \). Then, by using the Markov property, the likelihood function for the \( i \)-th patient is

\[
L_i(\beta, \mathbf{x}_i) = \prod_{\nu=1}^{n_i} P_{k_{\nu-1}k_{\nu}}(t_{i\nu} - t_{i(\nu-1)}, \mathbf{x}_i)
\]

And the likelihood function for all patients is

\[
L = \prod_{i=1}^{n} L_i(\beta, \mathbf{x}_i)
\]

When it is feasible to compute the collection of probabilities

\[
\{ P_{k_{\nu-1}k_{\nu}}(t_{i\nu} - t_{i(\nu-1)}, \mathbf{x}_i) \mid \nu = 1, 2, \cdots, n_i; i = 1, 2, \cdots, n \}
\]

efficiently, then iterative search methods may be used to find a maximum likelihood estimate of a vector \( \beta \) of parameters.

Observe that the Markov property hold in time is essential for the validity of this procedure of estimation. Moreover, the assumption that sojourns times in every transient states follow exponential distribution must also be reasonable for situation under study. This may be the case for clinical trials with a duration of one to three years, where the results may be robust to distributional assumptions.

Ideally, to use the principle of maximum estimation when a semi-Markov process is being considered as a model, one needs complete information on the sample path of each patient (individual) under study. Suppose the following set of pairs of information

\[ \{(k_{i\nu}, t_{i\nu}) \mid \nu = 0, 1, 2, \ldots, n_i; i = 1, 2, \ldots, n\} \]

is available on a sample of \( n \) patients, where \( t_{i\nu} \) is the time state \( k_{i\nu} \) is entered for the \( i \)-th patient. Thus, for the \( i \)-th patient, the sojourn time in state \( k_{i\nu} \) is \( t_{i(\nu+1)} - t_{i\nu} \). Then, for the \( i \)th patient with vector \( x_i \) of covariates, the likelihood function based on the information for this patient is

\[
L_i(\beta, x_i) = \prod_{\nu=1}^{n_i} a_{k_{i(\nu-1)}k_{i\nu}}(t_{i\nu} - t_{i(\nu-1)}, x_i).
\]

Unfortunately, in many data sets complete information on sample paths of individuals is not available. What is usually available is the pair \((k_{i\nu}, t_{i\nu})\), where \( t_{i\nu} \) is the time of the \( \nu \)-th visit of the \( i \)-th patient and \( k_{i\nu} \in S \) is the state of the patient at that time. Hence, the method of maximum would not strictly apply in this case so that an investigator must resort to other methods of estimating unknown parameters. Among the ways to circumvent this difficulty that have been suggested are the following.

(a) If transition occur among states rather slowly, then take observations more frequently and view the pairs \((k_{i\nu}, t_{i\nu})\) as complete information on the sample paths. This is the view that D. DiPrimeo is taking in his research on drugs to treat rheumatoid arthritis. He also uses the assumption of latent exponential random variables and this makes the problem of estimating parameters very tractable.
(b) Suppose one has faith in some stochastic process as a model for a data set, but it is very difficult to write down and compute a likelihood function for the data. Furthermore, suppose that given some trial values of the unknown parameters, it is easy to develop numerical schemes for computing Monte Carlo realizations of the process. Then, one could compute a set of Monte Carlo realizations for various points in the parameter space and finding a point that minimizes a goodness of fit criterion to the data such as Chi-Square.

(c) In any case, it may be informative to carry out a number of numerical experiments to gain insights based on informed guesses for values of parameters.
• Systems of Renewal Type Integral Equations.

Because procedures for computing numerical solutions of a system of renewal types integral equations will always entail some discretization process, only the case of discrete time will be considered and time set will be taken as

\[ T = \{ t \mid t = 0, 1, 2, \ldots \} . \]

Given that the process is in transient state \( i \in \mathcal{S}_2 \) at time \( t = 0 \), let \( P_{ij}(t; \mathbf{x}) \) denote the conditional probability the process is in state \( j \in \mathcal{S} \) at time \( t > 0 \) as a function of a covariate vector \( \mathbf{x} \).

If \( j \in \mathcal{S}_1 \), an absorbing state, the by a renewal argument, it can be seen that the \( m_2 \times m_1 \) matrix

\[ \mathbf{P}_A(t; \mathbf{x}) = (P_{ij}(t; \mathbf{x}) \mid i \in \mathcal{S}_2, j \in \mathcal{S}_1) \]

of current state probabilities satisfies the system

\[ P_{ij}(t; \mathbf{x}) = A_{ij}(t; \mathbf{x}) + \sum_{k \in \mathcal{S}_2} \sum_{s=0}^{t} a_{ik}(s; \mathbf{x}) P_{kj}(t - s; \mathbf{x}) \]

of renewal type equations. To cast this system in a succinct matrix notation, let

\[ \mathbf{R}(t; \mathbf{x}) = (A_{ij}(t; \mathbf{x}) \mid i \in \mathcal{S}_2, j \in \mathcal{S}_1) \]

denote a \( m_2 \times m_1 \) matrix and

\[ \mathbf{q}(t; \mathbf{x}) = (a_{ij}(t; \mathbf{x}) \mid i \in \mathcal{S}_2, j \in \mathcal{S}_2) \]

a \( m_2 \times m_2 \) matrix of densities for transitions among transient states. Then, this system may be written as

\[ \mathbf{P}_A(t; \mathbf{x}) = \mathbf{R}(t; \mathbf{x}) + \sum_{s=0}^{t} \mathbf{q}(s; \mathbf{x}) \mathbf{P}_A(t - s; \mathbf{x}) \]

26
But, if $j \in \mathcal{S}_2$, a transient state, then the $m_2 \times m_2$ matrix of current state probabilities

$$P_T(t; x) = (P_{ij}(t; x) | i \in \mathcal{S}_2, j \in \mathcal{S}_2)$$

satisfy the system of renewal type equations

$$P_T(t; x) = D(t; x) + \sum_{s=0}^{t} q(s; x) P_T(t-s; x),$$

where

$$D(t; x) = (\delta_{ij} (1 - A_i(t; x) | i \in \mathcal{S}_2, j \in \mathcal{S}_2))$$

is a diagonal matrix.

Recursive procedures may be used to solve these equations. For example, suppose it takes at least one time unit to record a change of state. Then, the matrix of transition densities satisfies the condition $a(0) = 0_{mm}$, $m \times m$ zero matrix. Therefore,

$$P_A(1; x) = R(1; x)$$

and for $t \geq 1$,

$$P_A(t; x) = R(t; x) + \sum_{s=0}^{t-1} q(s; x) P_A(t-s; x),$$

because $P_A(0; x) = 0_{m_2m_1}$. Similar remarks hold for the matrix $P_T(t; x)$. 

27
Recursive Computation of Markov Renewal Density.

Let \( \ast \) stand for the operation of matrix convolution. Then, by dropping the argument \( t \), the renewal equation for the matrix of absorption probabilities may be written in the succinct form

\[
P_A(x) = R(x) + q(x) \ast P_A(x)
\]

with the formal solution

\[
P_A(x) = \left( q^{(0)} - q(x) \right)^{(-1)} \ast R(x).
\]

In this formalism, \( q^{(0)}(t) \) is the identity function for the operation of matrix convolution and \( (q^{(0)} - q(x))^{(-1)} \) stands for convolution inverse of the function \( q^{(0)}(t) - q(t;x) \) and may be identified as \( m(t;x) \), the renewal density. The identity function in question is defined as

\[
q^{(0)}(0) = I_{m_2}
\]

and

\[
q^{(0)}(t) = 0_{m_2m_2}
\]

for \( t \geq 1 \). Thus, the renewal density must satisfy the equation

\[
\sum_{s=0}^{t} m(s;x) \left( q^{(0)}(t - s) - q(t - s;x) \right) = q^{(0)}(t)
\]

for all \( t = 0, 1, 2, \ldots \). Equivalently,

\[
m(t;x) = q^{(0)}(t) + \sum_{s=0}^{t} m(s;x)q(t - s;x).
\]

Therefore,

\[
m(0) = I_{m_2}
\]
and for \( t \geq 1 \)

\[
m(t; x) = \sum_{s=0}^{t-1} m(s; x)q(t - s; x).
\]

And the solution of the renewal equation is

\[
P_A(t, x) = \sum_{s=0}^{t} m(s, x)R(t - s, x)
\]

for \( t = 0, 1, 2, \cdots \). Lastly, the Markov renewal function is

\[
M(t; x) = (M_{ij}(t; x)) = \sum_{s=0}^{t} M(s; x),
\]

and \( M_{ij}(t; x) \) is the conditional expectation of the number of visits to transient state \( j \) during the time interval \( (0, t] \) prior to absorption, given that the process was in state \( i \in \mathcal{S}_2 \) at time \( t = 0 \).