

Modelling Non-homogeneous Markov Processes via Time Transformation

Structure of the Talk

- Recap and motivation
- Estimation procedure
- Simulation studies
- Application to real data

Motivation - Scientific Objectives

Scientific Goal: Describe a disease process in terms of its transitions through discrete states.

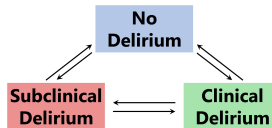
- Progressive disease: subjects traverse disease states in one direction.

- ▣ Stages of HIV infection



- Non-progressive disease: subjects may visit some or all states repeatedly, just once, or not at all.

- ▣ Delirium



Motivation - Methodological Objectives

Goal: Obtain estimates and covariance matrices for transition intensity parameters in non-homogeneous Markov process models.

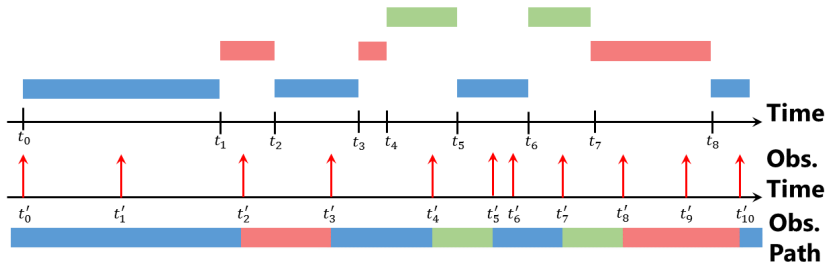
Problem #1 - Panel Data: Subjects are observed at a sequence of discrete times, observations consist of the states occupied by the subjects at those times.

- The exact transition times are not observed.
- The complete sequence of states visited by a subject may not be known.

Motivation - Methodological Objectives

Delirium Process Sample Path

■ No Delirium ■ Subclinical Delirium ■ Clinical Delirium



Motivation - Methodological Objectives

Problem #1 - Panel Data: Solution by Kalbfleisch & Lawless

- Estimation of transition parameters for *homogeneous* Markov processes via maximum likelihood.
 - A process is *Markov* if the future state of the process depends only on its current state. i.e.

$$P(X(t+s) = j | X(t) = i, X(u) = x(u), 0 \leq u < s) = P(X(t+s) = j | X(t) = i) = p_{ij}(t, t+s)$$

- A process is *homogeneous* if the transition probabilities do not depend on the chronological time t . i.e.

$$p_{ij}(t, t+s) = p_{ij}(s)$$

- Fisher scoring algorithm based on expectations of first order derivatives of the log-likelihood provides parameter and variance estimates.

Motivation - Methodological Objectives

Problem #2 -Non-homogeneity in the Markov Process

- If the process is non-homogeneous, $p_{ij}(t, t + s) \neq p_{ij}(s)$, so we must estimate a new transition probability matrix for every t .
- **Contribution of this paper:** If all of the non-homogeneity in the process is purely due to a time-varying multiplicative change in the transition intensities, we may use the results by Kalbfleisch and Lawless with minor adjustments.

Key Proposition: Time Transformation

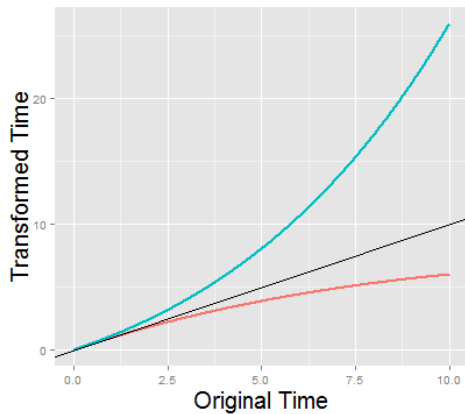
Let u denote the original time scale of the observations. If there exists an invertible transformation of the time scale such that the process is homogeneous on $t=h(u)$, with transition intensity matrix Q_0 , then

$$\begin{aligned} P(u_1, u_2) &= P(h(u_1), h(u_2)) = P(t_2 - t_1) \\ &= \exp[Q_0(t_2 - t_1)] = \exp[Q_0(h(u_2) - h(u_1))] \end{aligned}$$

Key Proposition: Time Transformation

- Kalbfleisch and Lawless suggested that nonhomogeneity arising from a transformation of the time scale could be accounted for using $\exp(Q_0 \int_{u_1}^{u_2} g(s) ds)$. An advantage of the method in my paper is that it does not require that the time transformation be integrable.
- $t = h(u)$ is a time scale, so we require $h(u) \geq 0$ and $\frac{\partial h(u)}{\partial u} \geq 0$.
- Examples of time transformations
 - ▣ Exponential: $t = h(u; \theta) = u\theta^u$
 - ▣ Nonparametric: Knots at u_k , $k = 1, \dots, d$

$$\begin{aligned} t = h(u) &= u\theta(u) \\ \theta(u) &= \sum_{k=1}^d c(u)\theta_k \left\{ \frac{1}{\gamma} K \left(\frac{u - u_k}{\gamma} \right) \right\} \\ c(u) &= \left(\sum_{k=1}^d \frac{1}{\gamma} K \left(\frac{u - u_k}{\gamma} \right) \right)^{-1} \end{aligned}$$



Transformation

- Exponential ($\theta=0.95$)
- Exponential ($\theta=1.1$)

Maximum Likelihood Estimation

Maximum likelihood estimation for a nonhomogeneous Markov process via time transformation proceeds exactly as in Kalbfleisch and Lawless.

The likelihood for one subject is:

$$L(Q_0) = P(X(u_1) = x_1) \prod_{i=2}^n p_{x_{i-1}, x_i}(u_{i-1}, u_i)$$

Maximum Likelihood Estimation

Maximum likelihood estimation for a nonhomogeneous Markov process via time transformation proceeds exactly as in Kalbfleisch and Lawless.

The likelihood for one subject is:

$$L(Q_0) = P(X(u_1) = x_1) \prod_{i=2}^n p_{x_{i-1}, x_i}(u_{i-1}, u_i)$$

† No absorbing states $\Rightarrow$ Asymptotics kick in as one subject is observed for increasing period of time, or as many independent subjects are each observed for finite period.

Maximum Likelihood Estimation

Maximum likelihood estimation for a nonhomogeneous Markov process via time transformation proceeds exactly as in Kalbfleisch and Lawless.

The likelihood for one subject is:

$$L(Q_0) = P(X(u_1) = x_1) \prod_{i=2}^n p_{x_{i-1}, x_i}(u_{i-1}, u_i)$$

↑ No absorbing states $\Rightarrow$ Asymptotics kick in as one subject is observed for increasing period of time, or as many independent subjects are each observed for finite period.

↑ Absorbing states $\Rightarrow$ Asymptotics rely on increasing number of subjects under observation and independence across subjects.

Maximum Likelihood Estimation

In our simulations and application, we have an absorbing state.
Therefore, the likelihood for the sample is:

$$L(Q_0) = \prod_{j=1}^m P(X(u_{j,1}) = x_{j,1}) \prod_{i=2}^n p_{x_{j,i-1}, x_{j,i}}(u_{j,i-1}, u_{j,i})$$

Maximum Likelihood Estimation

Since the time-transformed process is homogeneous, we have

$$\begin{aligned} L(Q_0, \theta) &= \prod_{j=1}^m P(X(u_{j,1}) = x_{j,1}) \prod_{i=2}^n p_{x_{j,i-1}, x_{j,i}}(h(u_{j,i}; \theta) - h(u_{j,i-1}; \theta)) \\ &= \prod_{j=1}^m P(X(u_{j,1}) = x_{j,1}) \prod_{i=2}^n \left[e^{Q_0(h(u_{j,i}; \theta) - h(u_{j,i-1}; \theta))} \right]_{x_{j,i-1}, x_{j,i}} \end{aligned}$$

Seminal Paper: *Kalbfleish & Lawless (1985)*

- The scoring method proposed in Kalbfleish and Lawless, and implemented in Hubbard et al., uses the expected information matrix. This simplifies the algorithm by only requiring use of the first-order derivatives of the likelihood.
- Assume independence across subjects. Letting ψ be the vector of functionally independent elements of Q_0 and θ , we have that the MLEs are the solutions to $\frac{\partial}{\partial \psi} \log L_m(\psi) = 0$ and

$$\hat{\psi} \sim \mathcal{N}(0, \mathcal{I}_m^{-1}(\psi_0)), \text{ where } \mathcal{I}_m = E \left[\left(\frac{\partial}{\partial \psi} \log L_m(\psi) \right) \left(\frac{\partial}{\partial \psi} \log L_m(\psi) \right)^T \right]$$

- Given initial estimates $\psi^{(0)}$, the $(k+1)^{st}$ step for the MLEs is

$$\hat{\psi}^{(k+1)} = \hat{\psi}^{(k)} + \hat{\mathcal{I}}_m(\hat{\psi}^{(k)})^{-1} \frac{\partial}{\partial \psi} \log L_m(\psi)$$

- Last time, we derived in (excruciating) detail expressions for the elements of the score vector and the Fisher information matrix.
- Large sample properties of the estimates can be derived using standard asymptotic theory (Billingsley, 1961; Albert, 1962; Bladt and Sorensen, 2005).
- Were they relevant, we could derive other quantities using estimates of the transition intensity parameters and time transformation parameters.
 - Mean sojourn times in each state and their variances
 - Steady state distribution (for an ergodic process)

Data Simulation Procedure - Overview

We use *Gillespie's direct method* to simulate data (Gillespie, 1977).

- 1 We recognize the off diagonal elements of Q as intensity parameters of independent Poisson processes from state i to state j . Diagonal elements are the negated intensities of arrival to any non-diagonal state.
- 2 We can simulate the path for homogeneous process from a given transition intensity matrix by generating transition times, then randomly selecting the new state conditional on a move.
- 3 Incorporate non-homogeneity and the panel structure of observations is trivial once the path has been simulated.

Data Simulation Procedure - Observation Times

We simulate data that is homogeneous on a transformed time scale as follows:

- 1 On the original time scale, simulate i.i.d. $Unif(0.2, 1)$ inter-observation times. Cumulative summation gives the raw observation times.
- 2 Transform the observation times. This gives observation times in the operational time scale on which the process is homogeneous.
- 3 Simulate a path from the transition intensity matrix. Record sequence of states and transition times.

Data Simulation Procedure - Simulating a Sample Path

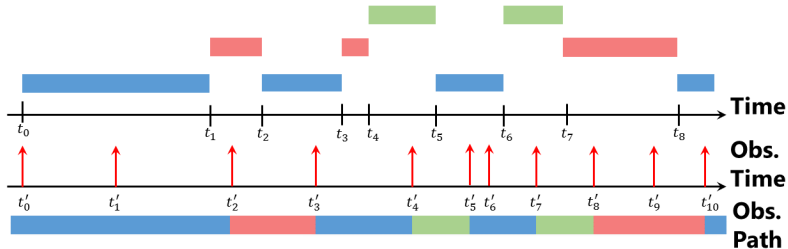
Simulating a path proceeds as follows:

- 1 Given state i at time $t = 0$, draw $u \sim Unif(0, 1)$. The sojourn time in state i is given by $\Delta t = \frac{\log(u)}{q_{ii}} \sim exp("mean" = -\frac{1}{q_{ii}})$
- 2 Conditioned on exiting state i , the probability of moving to state j is $p_{ij} = \frac{q_{ij}}{\sum_{j \neq i} q_{ij}}$. We then partition the interval $[0, 1]$ into the lengths p_{ij} for $j \neq i$, and draw $v \sim Unif(0, 1)$. The interval into which v falls gives it's next state.
- 3 If the new state is an absorbing state (death), or if the transition data is greater than the termination date of the period of observation, the path is terminated. Otherwise, the state is recorded under the next time in the panel observation, we increment the time, then proceed.

Data Simulation Procedure

Delirium Process Sample Path

■ No Delirium ■ Subclinical Delirium ■ Clinical Delirium



Simulation Setup

- Transition intensity matrix $Q_0 = \begin{pmatrix} -0.3 & 0.2 & 0.1 \\ 0.2 & -0.3 & 0.1 \\ 0 & 0 & 0 \end{pmatrix}$
- 1,000 datasets were generated for each combination of the following parameters:
 - Number of subjects: 100, 250, 500
 - Observations per subject: 12, 24
 - Exponential transformation parameter θ : 0.95, 1, 1.1
- Initial values based on estimates assuming exact knowledge of path and transition times:
 - $q_{ij}^{(0)} = \frac{\# \text{ Transitions from } r \text{ to } s}{\text{Total time in state } r}$
 - $\hat{\theta}^{(0)} = 1$
- Parameters for each dataset estimated for time transformation (TT) method and with standard homogeneous model (HM).

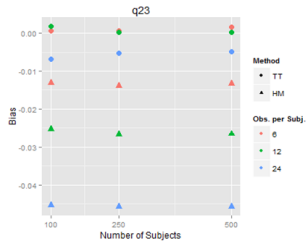
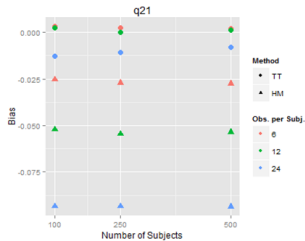
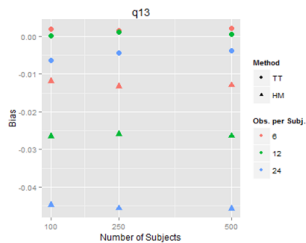
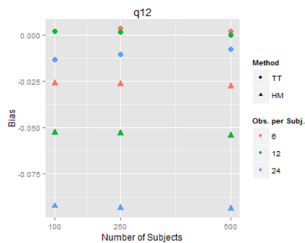
Simulation Results: $\theta = 0.95$

$\theta = 0.95$ with 24 measurements per subject was the only simulation scenario with severe convergence issues.

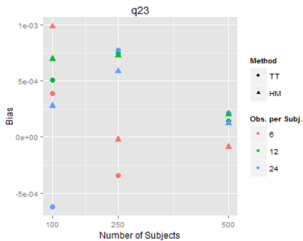
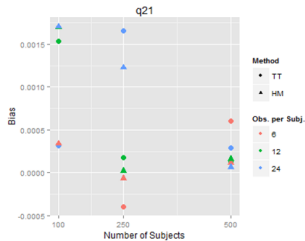
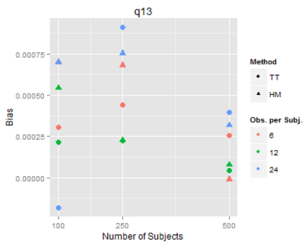
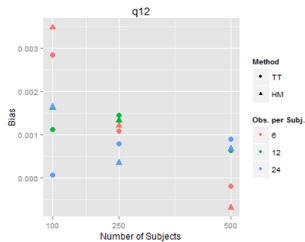
- $m = 100$: Model failed to converge for 539 of 1000 datasets.
- $m = 250$: Failed to converge for 680 datasets.
- $m = 500$: Failed to converge for 742 datasets.

Explanation: $\theta < 1$ corresponds to a process whose rate of evolution is decreasing, meaning that we will artificially observe more self-transitions, particularly as the observation period becomes longer.

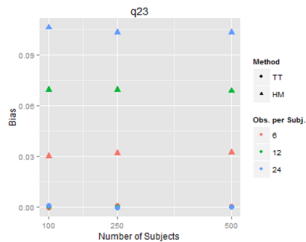
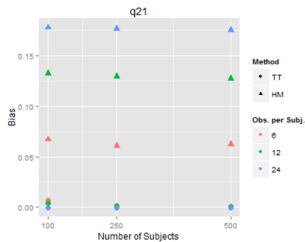
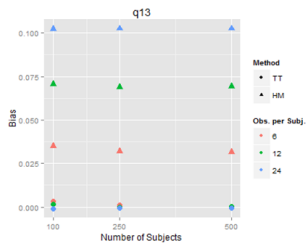
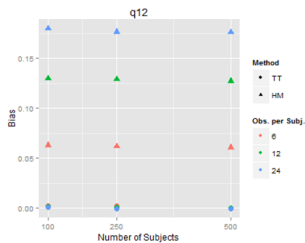
Simulation Results - Bias: $\theta = 0.95$



Simulation Results - Bias: $\theta = 1$



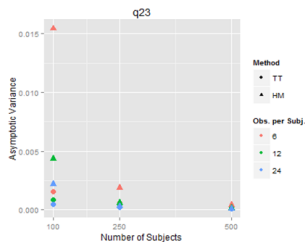
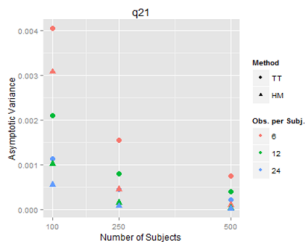
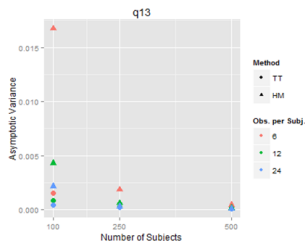
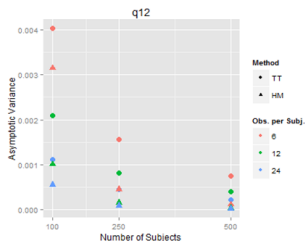
Simulation Results - Bias: $\theta = 1.1$



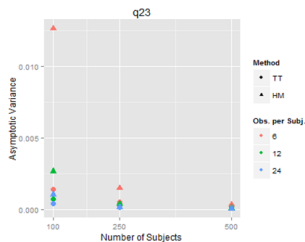
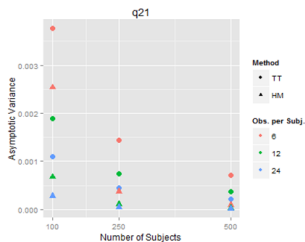
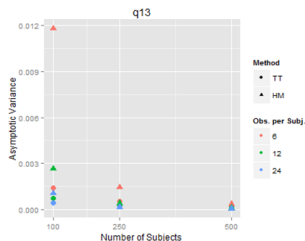
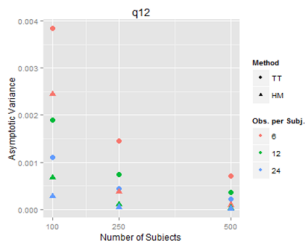
Simulation Results - Bias

- When the data are generated from a model where the rate of evolution is *decreasing*, transition intensity estimates from a model that assumes time homogeneity will be biased *downward*. i.e. the model will think the process transitions less often than is the case.
- Downward bias becomes more severe with a longer period of observation.
- When the data are generated from a model where the rate of evolution is *increasing*, transition intensity estimates will be biased *upward*. i.e. the model will think the process transitions more often than is the case.
- Upward bias becomes more severe with a longer period of observation.
- Bias does not change much with the number of subjects.

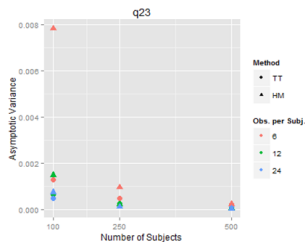
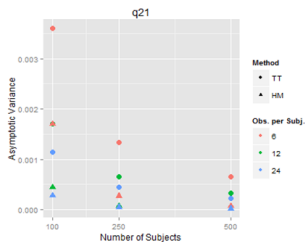
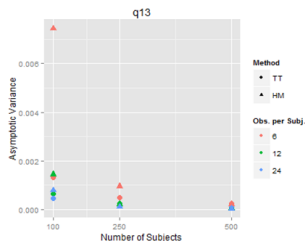
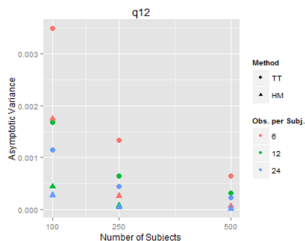
Simulation Results - Asymptotic Variance: $\theta = 0.95$



Simulation Results - Asymptotic Variance: $\theta = 1$



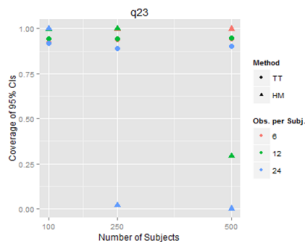
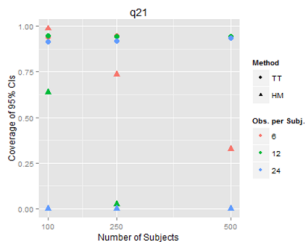
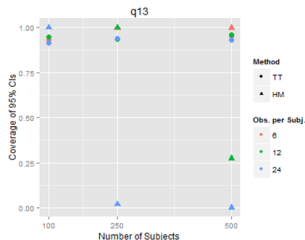
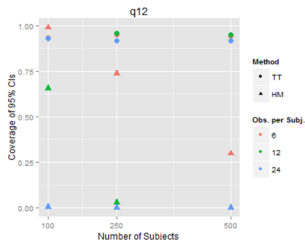
Simulation Results - Asymptotic Variance: $\theta = 1.1$



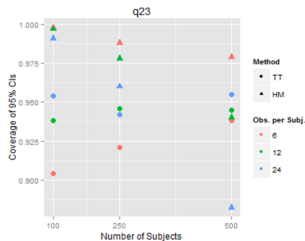
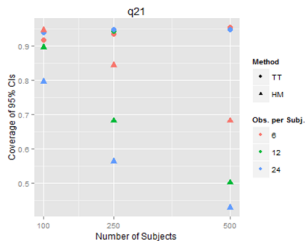
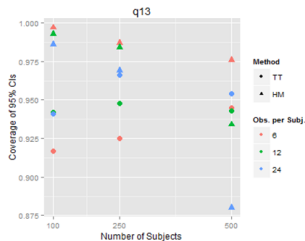
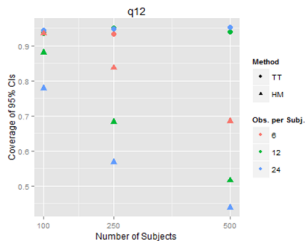
Simulation Results - Asymptotic Variance

- Asymptotic variance decreases as both the number of subjects increases and as the observation period increases.
- There is not a clear pattern for how properly accounting for the changing rate of evolution of the process using a time transformation affects the asymptotic variance of the transition intensity estimates.

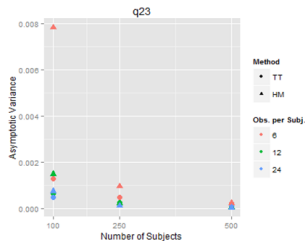
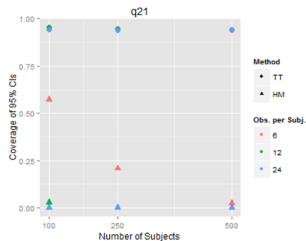
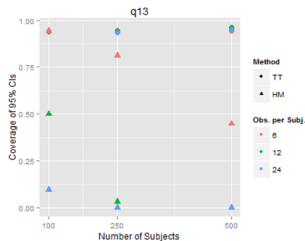
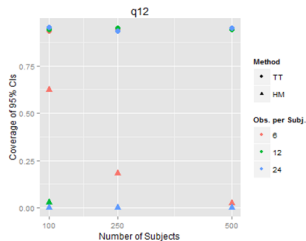
Simulation Results - Coverage of 95% CIs: $\theta = 0.95$



Simulation Results - Coverage of 95% CIs: $\theta = 1$



Simulation Results - Coverage of 95% CIs: $\theta = 1.1$



Simulation Results - Coverage of 95% CIs

- Coverage of confidence intervals is strongly affected by the bias in the estimates.
- **Moral:** If there is non-homogeneity in the process arising from a time varying multiplicative change in the transition intensities, don't trust your confidence intervals.

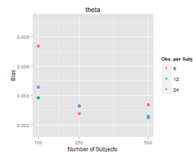
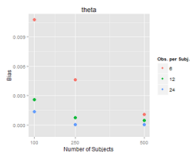
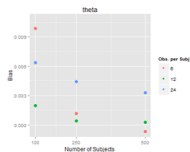
Simulation Results - Estimation of θ

$\theta = 0.95$

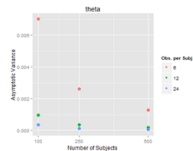
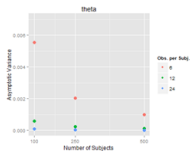
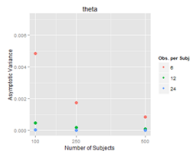
$\theta = 1$

$\theta = 1.1$

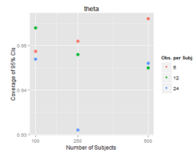
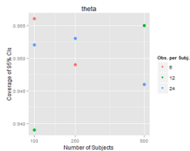
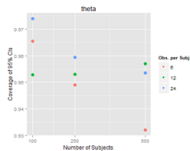
Bias



Asymptotic Variance



95% CI Coverage



Simulation Results - Estimation of θ

- Coverage is generally quite good, even at smaller sample sizes.
- Variance decreases rapidly as the length of the observation period/number of subjects increases.
- For data generating mechanisms whose rate of evolution is not constant, there is a small amount of positive bias that decreases as the number of subjects and the number of observations per subject each increase.

Application - Modeling Health Status in Older Adults

- Cardiovascular Health Study (CHS) - Longitudinal study of adults aged 65 and older recruited from Medicare eligibility rolls.
- **Note: 55 is young.**
- Interested in the progression of self-rated health.
- Follow-up observations in the original data set were scheduled annually, but actual elapsed time ranged between one month to one year.
- The data released for this project was anonymized - information on demographic characteristics and actual observation times were removed. Recorded times are given in six month increments from baseline.
- 5850 subjects, average number of follow-up visits was 7.5, including ascertainment of death.

Application - Modeling Health Status in Older Adults

- Self-reported health reported as: "excellent", "very good", "good", "fair", or "poor".
- Further dichotomized into three states
 - (1) Healthy: subject reports excellent, very good, or good health
 - (2) Unhealthy: subject reports fair or poor health
 - (3) Dead
- Observed transitions:

	Healthy (1)	Unhealthy (2)	Dead (3)
Healthy (1)	90071	15578	1267
Unhealthy (2)	13103	34960	3768

Application - Modeling Health Status in Older Adults

- Using the time transformation method with exponential transformation we estimate that $\theta = 1.017$ with estimated standard error of 0.00061, indicating that the rate of evolution of the process is accelerating.
- Parameter estimates and standard errors:

	Time Trans.	Homog. (MSM)
q_{12}	0.306 (0.0039)	0.394 (0.0084)
q_{13}	0.007 (0.0006)	0.010 (0.0758)
q_{21}	0.501 (0.0067)	0.684 (0.0091)
q_{23}	0.129 (0.0027)	0.175 (0.0168)

Application - Modeling Health Status in Older Adults

Using the time transformation method, we estimate that for a random individual, the transition probability matrix for an elapsed time of one year from baseline is:

	Healthy	Unhealthy	Dead
Healthy	0.780	0.199	0.022
Unhealthy	0.325	0.575	0.101
Dead	0	0	1