Published in final edited form as:

J Appl Stat. 2015 ; 42(12): 2531–2538. doi:10.1080/02664763.2015.1043865.

Joint coverage probability in a simulation study on Continuous-Time Markov Chain parameter estimation

Julia S. Benoit, PhD^{a,b}, Wenyaw Chan, PhD^b, and Rachelle S. Doody, MD PhD^c

^aDepartment of Vision Sciences/Texas Institute of Measurement, Evaluation, & Statistics (TIMES), University of Houston, Houston, TX, USA

^bDepartment of Biostatistics, University of Texas School of Public Health, Houston, TX, USA

^cDepartment of Neurology, Baylor College of Medicine, Houston, TX, USA

Abstract

Parameter dependency within data sets in simulation studies is common, especially in models such as Continuous-Time Markov Chains (CTMC). Additionally, the literature lacks a comprehensive examination of estimation performance for the likelihood-based general multi-state CTMC. Among studies attempting to assess the estimation, none have accounted for dependency among parameter estimates. The purpose of this research is twofold: 1) to develop a multivariate approach for assessing accuracy and precision for simulation studies 2) to add to the literature a comprehensive examination of the estimation of a general 3-state CTMC model. Simulation studies are conducted to analyze longitudinal data with a trinomial outcome using a CTMC with and without covariates. Measures of performance including bias, component-wise coverage probabilities, and joint coverage probabilities are calculated. An application is presented using Alzheimer's disease caregiver stress levels. Comparisons of joint and component-wise parameter estimates yield conflicting inferential results in simulations from models with and without covariates. In conclusion, caution should be taken when conducting simulation studies aiming to assess performance and choice of inference should properly reflect the purpose of the simulation.

Keywords

Alzheimer's disease; estimation: joint coverage probability; continuous-time Markov chain; longitudinal study

1.0 Introduction

Progressive scientific research requires such complex statistical modeling that statisticians rely on effective simulation studies to develop and establish novel methodology that may lack an analytical form. Mean estimates, standard errors, bias, and coverage probabilities, among others, are useful to evaluate the performance of the results obtained from the estimating procedure [1]. Multivariate and univariate approaches have been discussed in the

literature for statistical testing and obtaining Confidence Intervals but not with regards to coverage probability.

1.1 Coverage Probability

Coverage, probability (CP), or simulated coverage rate has been defined in the literature and is reported as the proportion of time the $100(1-\alpha)\%$ confidence interval (CI) covers the true parameter constructed from the estimates of each randomly generated data set all with the same known truth [1]. Thus, we are provided with a measure of how often in practice the model is likely to produce effective results and CP can be formulated as the proportion of

times $\hat{\theta} \pm Z_{1-\alpha/2} SE(\hat{\theta})$ includes θ for $i=1, \dots, S$ datasets. In this research we present a multivariate approach for reporting simulated coverage when interest lies in making inference on several parameters. We discuss the current literature on reported accuracy of simulated results, compare and lay the groundwork for univariate and multivariate confidence regions (intervals), and revisit the Continuous-Time Markov Chain (CTMC) in previously published simulation studies. This paper focuses on the comparison of coverage probability calculated using joint confidence regions versus simultaneous confidence intervals with one at-a-time (or component-wise) intervals to evaluate the precision of an estimation procedure in simulation studies.

Let S denote the number of independent data sets generated under the assumptions of a specified statistical model that depends on an unknown parameter set, $\theta = \{\theta_1, \theta_2, \dots, \theta_p\}$. Let $T^{(\nu)}$ be the numerical values for the estimators for parameter θ_ν , $\nu = 1$ to p . Then for S

data replicates $T^{(\nu)} = (T_1^{(\nu)}, T_2^{(\nu)}, \dots, T_s^{(\nu)})'$. Hence, $\bar{T}^{(\nu)} = S^{-1} \sum_{k=1}^S T_k^{(\nu)} = \widehat{\text{mean}}$,

$\widehat{\text{bias}} = \bar{T}^{(\nu)} - \theta_\nu$, and $\widehat{\text{SD}} = \sqrt{(S-1)^{-1} \sum_{k=1}^S (T_k^{(\nu)} - \bar{T}^{(\nu)})^2}$. Furthermore, the simulated coverage rate, expected to achieve nominal 95% coverage [9], can be calculated as

$CP_\nu = \frac{1}{s} \sum_{j=1}^S I_j$, where $I_j = 1$ indicates that the 95% CI for sample j captures the true parameter θ_ν and $I_j = 0$ otherwise.

For component-wise parameter interest when testing the single hypothesis, $H_0: \theta_\nu = \theta_{\nu 0}$

then $z = (\hat{\theta}_\nu - \theta_{\nu 0}) / SE \xrightarrow{d} N(0, 1)$ under the null hypothesis in large samples. This Wald

statistic is used to calculate confidence intervals for each $\theta_{\nu 0}$ as $\hat{\theta}_\nu \pm z_{\alpha/2} \sqrt{SE(\hat{\theta}_\nu)}$ [2].

In the multi-dimensional parameter space, if multivariate normality (MVN) is assumed for

large samples, then $Y = V^{-1/2} (\hat{\theta} - \theta_0) \xrightarrow{d} N(0, I_r)$, thus the quadratic form

$Y'Y = (\hat{\theta} - \theta_0)' V^{-1} (\hat{\theta} - \theta_0) \sim X_r^2$, where $\hat{\theta}$ is the vector of maximum likelihood estimates, $\theta_0 = (\theta_{01}, \dots, \theta_{0p})'$, V is the variance-covariance matrix for $\hat{\theta}$, and r is the rank(V). Therefore, a joint confidence region can be constructed as:

$$X_{\frac{\alpha}{2},r}^2 < (\theta - \theta_0)' V^{-1} (\theta - \theta_0) < X_{1-\frac{\alpha}{2},r}^2 \quad (1) \quad [6]$$

This study focuses on examining the joint and component-wise coverage probabilities for simulation studies aiming at estimation of the transition rates of the CTMC and an application on self-reported stress levels from caregivers of Alzheimer's disease patients are presented.

2.0 Application to the CTMC Model

A CTMC consists of a family of random variables which can describe the state of a system at a given time. In Markov chains, the dynamic behavior of the future process (including the probability of being at a particular state in the future) depends only on the current state, not on the past history of the process. Consider a study where $Y_k(t_{k,1}), Y_k(t_{k,2}), \dots, Y_k(t_{k,n_k})$ is a sequence of observed ternary outcomes recorded as 1, 2, or 3 and measured longitudinally at times $t_{k,1}, t_{k,2}, \dots, t_{k,n_k}$ on subject k , and n_k is the number of observations on subject k . Assume $Y_k(\cdot)$ follows a three-state CTMC that can be fully described by the instantaneous transition rates, q_{ij} , the rate at which the process transitions from state ' i ' to state ' j ', where $i, j=1, 2, 3$ and $i \neq j$. Then the infinitesimal matrix R is formed by these parameters:

$$R = \begin{bmatrix} -(q_{12}+q_{13}) & q_{12} & q_{13} \\ q_{21} & -(q_{21}+q_{23}) & q_{23} \\ q_{31} & q_{32} & -(q_{31}+q_{32}) \end{bmatrix} \quad (2)$$

From the property of a continuous-time Markov chain, the sojourn time or amount of time a process stays in category ' i ' before exiting follows an exponential distribution with mean

$(\sum_{l=1, l \neq i}^3 q_{il})^{-1}$ and is generally unobservable. At transition time, the probability of transitioning into state ' j ' given that the process is currently in state ' i ' is calculated as

$Q_{ij} = q_{ij} (\sum_{l=1, l \neq i}^3 q_{il})^{-1}$. The transition rates make up the probability mechanism used to derive the probability of transition over a specific interval of time t , $P_{ij}(t) = P(Y(s+t) = j | Y(s) = i)$, $s > 0$. In practice each CTMC and its associated rate can also depend on covariates.

Li [7] used a CTMC model to develop a likelihood-based iterative method to analyze longitudinal data with a trinomial outcome variable. For each iteration of the estimation procedure the likelihood function was updated using one of three forms of algebraically complex explicit expressions derived for the transition probabilities. Each of these expressions was a function of instantaneous transition rates. The log-likelihood function was maximized using the Nelder-Mead simplex technique. The primary purpose of this research method was to compare transition rates between two groups. Although the method

performed well, neither standard error nor a measure of convergence was reported in their simulation study. The model did not include fixed or time-varying covariates and assumed the probability of the initial state was equally likely.

As an extension to the aforementioned work three covariates (two binary and one continuous) were included and a log link function was applied to link the transition rates with the linear combination of covariates in Mhoon et al. [8]. Convergence rates were poor and low bias was reported. Standard errors were reported but intervals were not used in inference. This paper focuses on examining coverage probability of the parameter estimates obtained via simulation studies of the three-state continuous time Markov chain model without covariates and with two covariates.

2.1 Model I-Simple Tertiary CTMC

For this simulation study, a continuous-time Markov process with $K=3$ states and model parameters q_{ij} were generated for $m=1000$ subjects, beginning with the generation of the initial state modeled as $\gamma_i = P(Y(t_1) = i)$, where ' i ' can take on values 1, 2, or 3. For each subject, data were generated for up to 10 time units and outcomes at times $t = 0, \dots, 10$ were observed. This simulation procedure was repeated 1000 times and the transition rates were estimated for each dataset. The Quasi Newton optimization [4] technique was used to maximize the log-likelihood function:

$$L(\Theta) = \sum_{k=1}^m \left[\sum_{i=1}^3 I_{\{i\}}(Y(t_{k,1})) \log(\gamma_i) + \sum_{j=1}^{n_k-1} \sum_{\text{all } j's} \log \left(P_{Y(t_{k,j})Y(t_{k,j+1})}(t_{k,j+1} - t_{k,j}) \right) \right], \quad (3)$$

where the indicator function $I_{\{i\}}(Y(t_{k,1}))$ is defined as $I_{\{i\}}(Y(t_{k,1})) = 1$ if $Y(t_1) = i$; 0 otherwise and all $\gamma_i = P(Y(t_1) = i)$ were treated as nuisance parameters since each γ_i does not depend on

the infinitesimal transition matrix. $\left(P_{Y(t_{k,j})Y(t_{k,j+1})}(t_{k,j+1} - t_{k,j}) \right)$ is the transition probability as defined in section 2.0. Estimates were compared to the true parameters using a measure of bias and coverage probability. Following the methodology from section two and modifying equality (1) the joint confidence region for the model parameters can be expressed as:

$$X_{\frac{\alpha}{2}, r}^2 < (\hat{q} - q)' V^{-1} (\hat{q} - q) < X_{1-\frac{\alpha}{2}, r}^2 \quad (4)$$

The coverage probability was calculated as the proportion of the data sets that have an estimated joint confidence region capturing the true q vector, where $q = (q_{12}, q_{13}, q_{21}, q_{23}, q_{31}, q_{32})'$. Note that V is the variance covariance matrix of $\hat{q}$, the MLEs of the parameter vector q .

The simulation study was conducted in statistical package SAS version 9.3 (SAS Institute) and implemented the NLMIXED procedure for nonlinear models.

Of the 1000 samples that were simulated, the overall joint coverage probability for the parameter set was 94.1% for the 998 replicates where the variance-covariance matrix was produced. Marginal coverage probabilities, standard errors, and bias of component transition intensities analyzed one-at-a-time are reported (Table 1).

2.2 Basic tertiary CTMC with covariates

For this study we modified the model from section 2.1 to include two covariates (X_1 and X_2) to reflect the real data example and the transition rates were linked to the transition rates q_{ij} using a log-link function denoted as:

$$\log(q_{ij}) = \alpha_{ij} + \beta_p X_p, \text{ where } i, j = 1, 2, 3; i \neq j; p = 1, 2, \quad (5)$$

where α_{ij} is the unique intercept for each q_{ij} and β_p is the shared coefficients for covariates X_p . Using SAS version 9.3, 2000 replicate data sets were generated with the goal of achieving minimal bias, high convergence rates, and nominal coverage. The two covariates $X_1 \sim \text{Bin}(1, .6)$ and $X_2 \sim N(9.4, 6.1)$ were simulated for $m=1000$ subjects. For each simulated continuous-time process, data were observed at times $t = 0, \dots, 10$. To obtain initial values for α_{ij} when implementing the NLMIXED procedure in SAS first X_2 was stratified into deciles and the mean value of each decile was imputed with one of 10 distinct values, thus resulting in $X_1 \times X_2 = 10 \times 2$ possible combinations of individual information. Then within each group, the maximum likelihood estimates for q_{ij} were calculated as the $\{\text{\#transitions from } i \text{ to } j\} / \{\text{duration of time spent in } i\}$ and used as the dependent variable in Equation (2.3) to obtain regression coefficients for α_{ij} . The average $\hat{\beta}_1, \hat{\beta}_2$ and $\hat{\alpha}_{ij}$ from the six models were used as initial starting parameters for estimation. The Quasi Newton optimization technique was used to find the MLE of the parameter set of the log-likelihood function and then compared to the true parameters using bias and coverage probability measures. Since the parameter estimates are dependent, univariate confidence intervals and a joint confidence region was calculated.

Of 2000 sample datasets simulated, 10.8% either did not converge or the variance-covariance matrix was not invertible. The overall joint coverage probability was 92.9% ($n=1,659$) for the 1,785 useable replicate datasets. Component-wise coverage probabilities for each parameter are also calculated and displayed (Table 2).

Additionally, simulation studies were repeated for $m=100, 200$, and 500 subjects and 90%, 95%, and 99% joint coverage rates are presented in Figure 1. As expected, larger sample sizes yield better coverage, however, even with sample sizes of $m=100$, coverage is stable.

Often in real data it may be unrealistic for certain recurrent transitions to occur. For example, it is highly improbable for an individual who is in state 3 of an irreversible disease to transition back to state 1. Thus, we assessed the performance of the special case of the 3-state CTMC parameter estimation with $q_{31} = 0$ via simulation with 1,000 subjects and 2,000

replicates of the data and the results are presented in Table 3. The overall joint coverage probability was 93.9% ($n=1,869$) for the 1,999 useable replicate datasets.

3.0 Data Analysis: An Example

Longitudinal data collected from the Alzheimer's Disease and Memory Disorders Center (ADMDC) at Baylor College of Medicine is used for illustration [5]. The database consists of probable AD patients followed over 20 years, many of whom remain in active follow-up. Intervals of time between visits varied among individuals as did the number of visits themselves. Patient socio-demographic information such as age, sex and years of education, medical history, estimated symptom duration, neuropsychological tests scores, and caregiver self-reported stress levels were collected. In this study, longitudinal self-rated stress levels of each caregiver were modeled as a CTMC with three categories (mild, moderate and extreme) and patient baseline demographic characteristics, age and gender, were examined to better understand the movement between stages of AD patient care giver stress levels. Subjects whose caregivers did not self-rate stress levels at least once were excluded from this analysis. This model's ability to measure transitions over time while accounting for uneven intervals and number of observations allows the inclusion of patients with intermittent and monotone missing data under the assumption of missing completely at random (MCAR).

A total of 952 (1,123) patients had at least one caregiver self-rated stress level and were included in the analysis. For these subjects, the median of the possible visits was 3 (range 1–11). The average age of the patient was 74, ranging from 44–93, and the majority (68%) was female. The total number of visits was 3.4 and baseline stress levels were distributed as 8.7% mild, 25.7% moderate, and 65.51% extreme levels of stress.

The mean time to caregiver stress level change for female patients is longer than the male group (Table 4). Specifically, the mean time to caregiver stress transition for the females is 1.17 times that of males and, on average, caregivers of subjects who are one year older have their mean time to change stage increased by a multiplicative factor of 1.17.

4.0 Discussion

The likelihood-based continuous time Markov model is useful when the focus is on characterizing movement of individuals in discrete states. The purpose of this study was to perform a comprehensive examination, using more than one performance measure, of the recently developed general three-state CTMC MLEs with and without covariates. The utilization of additional performance measures other than bias is advised in evaluating statistical methods [3]. Although, low bias was reported in [7], no standard errors nor other means of evaluating the variability of the parameter estimates were reported and researchers may be hesitant to implement this method without this reassurance. We modified the model and reported low measures of bias and assessments of coverage. From 1000 independently sampled replicates in the simulation study without covariates, the expected simulation coverage rate should fall within 93.6–96.4 [10] for 95% confidence intervals. This range was constructed as $.95 \pm 1.95 * SE(.95)$ and $SE(.95) = \sqrt{.95(1 - .95)/1000}$. The rate of simulated coverage for the 1000 simulated samples ranged from 92.8–94.9 indicating that two of the

estimated transition intensities were slightly under covered. For the second simulation study that included covariates with 2,000 replicates, simulated coverage rate for component transition intensities ranged from 93.9–95.4 and low bias was reported. Although the empirical standard error was reported in [8] no assessment of coverage relative to the bias was reported. When we decreased the sample size to $m=100$, 200, and 500, coverage rates decreased. We also showed that the biases and the coverage probabilities of the parameter estimates performed well under the special case scenario.

An additional aim of the study was to introduce multivariate approaches to make inference on the model's performance. The joint coverage probability is more accurate in terms of investigating the simulation performance of the method. In this study we focused on making inference on the parameter set. Specifically, we used multivariate techniques to jointly analyze the correlated estimated transition rate vector and compared inference made using one at a time intervals. Comparisons of joint and component-wise parameter estimates yielded conflicting inferential results in simulations from both models. For example, in model I., two of the parameter's estimates were under covered yet the joint coverage rate fell within the acceptable range. On the other hand, in model II, component-wise coverage rates were as expected and the joint confidence region suggested slight under coverage.

For estimation of the CTMC, it is not appropriate to make inference on one parameter estimate or in this example the intercepts that uniquely identify each transition rate because all of the components of the instantaneous transition rate matrix are treated as one entity. We point this out because interpretation and inference of multi-state transition rates is primitive and currently the status quo is to report standard normal-based 95% C.I. as though the parameter estimates are independent of one another. We argue not that all of the multivariate techniques presented in this paper are novel but that the multivariate nature be rigorously enforced in scenarios modeled as multi-state CTMC. From the results in our data example in section 3.0, we compared the lengths of the 95% C.I. using standard normal-based versus simultaneous chi-square based or shadows of the p -dimensional ellipsoid using the standard error and point estimates in table 3. The median increase in length of the chi-square based 95% C.I. compared to the standard-normal based was 1.8 and inference was contradicting for the three parameters whose p -values were less than .05.

We have conducted an extensive investigation of the parameter estimation of the 3-state CTMC (general and special case) via simulation studies with various samples sizes and have presented results of joint coverage probabilities for ranging theoretical levels. We also examined the behavior of the individual parameters of the model. It is convenient to study parameters of a statistical model marginally but with awareness of the correlation within datasets using an adjustment when necessary. Caution should be taken when conducting simulation studies aiming to assess performance when the data does in fact have such parameter estimator dependency as that of the CTMC and choice of inference should be stated a priori to reflect the purpose of the simulation.

References

1. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. *Stat Med.* 2006; 25:4279–4292. [PubMed: 16947139]

2. Casella, G., Berger, RL. Statistical inference. Duxbury Press; Belmont, CA: 1990.
3. Collins LM, Schafer JL, Kam C. A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychol Methods*. 2001; 6:330–351. [PubMed: 11778676]
4. Dennis J, John E, Moré JJ. Quasi-Newton methods, motivation and theory. *SIAM Rev*. 1977; 19:46–89.
5. Doody R, Pavlik V, Massman P, Kenan M, Yeh S, Powell S, Cooke N, Dyer C, Demirovic J, Waring S. Changing patient characteristics and survival experience in an Alzheimer's center patient cohort. *Dement Geriatr Cogn Disord*. 2005; 20:198–208. [PubMed: 16088145]
6. Johnson, RA., Wichern, DW. Applied multivariate statistical analysis. Prentice Hall; Upper Saddle River, NJ: 2002.
7. Li Y, Chan W. Analysis of longitudinal multinomial outcome data. *Biometrical Journal*. 2006; 48:319–326. [PubMed: 16708781]
8. Mhoon KB, Chan W, Del Junco DJ, Vernon SW. A continuous-time Markov chain approach analyzing the stages of change construct from a health promotion intervention. *JP J Biostat*. 2010; 4(3):213–226. [PubMed: 23504410]
9. Natrella MG. The relation between confidence intervals and tests of significance: A teaching aid. *The American Statistician*. 1960; 14:20–22.
10. Tang L, Song J, Belin TR, Unützer J. A comparison of imputation methods in a longitudinal randomized clinical trial. *Stat Med*. 2005; 24:2111–2128. [PubMed: 15889392]

Figure 1.

Comparison of empirical and theoretical coverage probability for simulation studies on CTMC with covariates for various sample sizes of $N=100, 200, 500$, and 1000 .

Table 1

Bias and Coverage Probability of component transition intensities analyzed one-at-a-time from simulation study.

Parameter	TRUE	BIAS	CP (%)	Standard Error
q_{12}	0.6	-.007	93.7	.033
q_{13}	0.5	-.001	94.5	.027
q_{21}	0.4	-0.05	92.8	.022
q_{23}	0.3	0.002	92.9	.017
q_{31}	0.2	-0.0004	94.1	.011
q_{32}	0.1	.001	94.9	.008

Table 2

Bias and Coverage Probability of component means analyzed one-at-a-time from simulation study.

Parameter	TRUE	BIAS	%BIAS	CP (%)
α_{12}	-4.32	.013	.29	94.6
α_{13}	-4.50	.024	.53	93.9
α_{21}	-4.73	.016	.34	95.1
α_{23}	-5.01	.012	.24	94.6
α_{31}	-5.42	.020	.37	94.5
α_{32}	-6.11	.026	.42	95.4
β_1	1.0	-.001	.09	94.1
β_2	0.5	-0.003	.44	94.3

Table 3

Bias and Coverage Probability of component means analyzed one-at-a-time from simulation study: a special case where $q_{31}=0$.

Parameter	TRUE	BIAS	%BIAS	CP (%)
α_{12}	-4.32	.015	.29	95.1
α_{13}	-4.50	.014	.34	95.3
α_{21}	-4.73	.013	.27	94.7
α_{23}	-5.01	.025	.50	94.6
α_{31}	—	—	—	—
α_{32}	-6.11	.018	.30	94.8
β_1	1.0	-.004	.48	94.1
β_2	0.5	-0.002	.44	94.5

Table 4

Parameter Estimates AD Patient Caregiver Stress Levels.

Parameter Estimate	$\hat{\alpha}_{12}$	$\hat{\alpha}_{13}$	$\hat{\alpha}_{21}$	$\hat{\alpha}_{23}$	$\hat{\alpha}_{31}$	$\hat{\alpha}_{32}$	Gender $\left(\hat{\beta}_1\right)$	Age $\left(\hat{\beta}_2\right)$
Proposed Estimate (Standard Error)	-0.1975 (0.4638)	-5.9411 (8.8535)	-1.5215 (0.4869)	-0.4882 (0.4614)	-3.9734 (0.6528)	-1.7100 (0.4669)	0.1533 (0.1148)	0.1587 (0.6254)
$e^{\gamma_{ij}}$ or $e^{-\beta_k}$	0.821	0.003	0.218	0.632	0.019	0.181	1.1657	1.172
(P-value)	0.67	0.52	<0.005	0.29	<.0001	<0.0005	0.18	0.78