A SIMULATION STUDY OF CONFIDENCE INTERVALS FOR THE TRANSITION MATRIX OF A REVERSIBLE MARKOV CHAIN

by

XIAOJING ZHANG

Ph.D., Kansas State University, 2012

A REPORT

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

Approved by:

Major Professor
Dr. James W Neill

Abstract

Let $\{X_i; i = 0,1,2,...,\}$ be an irreducible, aperiodic, positive recurrent Markov chain with state space a subset of $\{0,1,...\}$, time-homogeneous transition matrix **P**= $\{p_{ij} = P(X_1 = j|X_0 = i)\}$ and limiting distribution $\{\pi_i > 0\}$. Based on observing $\{X_i = x_i; i = 0,1,2,\dots,t\}$, we study and compare two estimators of the transition probabilities $\{p_{ij}\}$:

(i) The maximum likelihood estimators $\{\hat{p}_{ij}\}$ of $\{p_{ij}\}$:

$$\hat{p}_{ij}=[N_{ij}(t)/N_i(t)]I(N_i(t)>0),$$

where

$$N_{ij}(t) = \sum_{l=0}^{t-1} I(X_l = i, X_{l+1} = j),\ N_i(t) = \sum_{l=0}^{t} I(X_l = i).$$

(ii) The symmetrized estimators $\{\hat{p}_{ij}^{(R)}\}$:

$$\hat{p}_{ij}^{(R)} = \left[(N_{ij}(t) + N_{ji}(t))/2N_i(t)\right]I(N_i(t) > 0).$$

It is well known that as $t \to \infty$, in distribution,

$$\sqrt{t}(\hat{p}_{ij}-p_{ij}) \to N(0, p_{ij}(1-p_{ij})/\pi_i). \qquad \text{(I)}$$

It was shown in Annis et. al. (2010) that *if the chain is reversible*, in distribution,

$$\sqrt{t}(\hat{p}_{ij}^{R}-p_{ij}) \to N(0, \sigma_{ij}^2(R)), \qquad \text{(II)}$$

with σ_{ij}^2(R) / $[p_{ij}(1-p_{ij})/\pi_i] \in [1/2, 1]$, implying that for a reversible chain $\hat{p}_{ij}^{(R)}$ is asymptotically as least as good as $\hat{p}_{ij}$ for reversible chains.

We designed and carried out a simulation study, using representative choices of n and **P**, to compare the performance, in terms of coverage rate and mean width, of nominal 0.95 confidence intervals for the elements of the transition matrix **P** constructed using (I) and (II), where the limiting variances are replaced by appropriate sample estimates. When the chain is reversible, both intervals are asymptotically correct and the intervals based on II are asymptotically no wider than those based on I. However, our simulations indicate that for the finite sample sizes and models used here, the estimated coverage rates based on II appear to be considerably below their nominal values in some cases. Since the coverage rates for the intervals based on (I) appear to approach their nominal levels as sample sizes increase, we recommend using them rather than the more complicated intervals based on II.

KEY WORDS: Stationary distribution; Reversibility; Transition probability estimation.

Table of Contents

List of Figures

List of Tables

Acknowledgements

Though only two names appear on the cover of this report, a great many people have contributed to its production. I owe my gratitude to all those people who have made this report possible.

My deepest gratitude is to my advisor, Dr. Paul I. Nelson. I have been very fortunate to have an advisor who is so dedicated to statistics research and has been very patient and supportive to my study even after his retirement. His profound knowledge and valuable suggestions have helped me overcome many crisis situations. I am very thankful to him for carefully reading and commenting on countless revisions of this manuscript. Without him, I would not be able to finish this report.

I am deeply grateful to Dr. James W Neill, who has been my major professor since Dr. Nelson's retirement. He has been always there to encourage and give advice. I am also thankful to him for reading my reports, commenting on my views and enrich my ideas.

I would like to thank my committee members Dr. Pietro Poggi-Corradini and Dr. Weixing Song, who were more than generous with their expertise and precious time. Thank you for agreeing to serve on my committee and for being so friendly, supportive and helpful throughout my graduate study.

Finally, I would like to thank my family. None of this would have been possible without the love and patience of my family. My husband Hongzhou Huang, my son Stephen and my parents Libing Zhu and Shangwei Zhang have been constant sources of love, concern, support and strength all these years. I would like to express my heart-felt gratitude to my whole family.

Dedication

This report is dedicated to:

My advisor Dr. Paul I. Nelson.

My son David S. Huang, whom I will meet in heaven.

Chapter 1 - Introduction

We consider a stochastic process $\{X_t, t = 0,1,2, ...\}$ that takes on a finite or countably infinite number of possible values, called states, denoted by $S \subset \{0,1,2, ...\}$. If $X_t = i$, the process is said to be in state *i* at time *t*. We suppose that whenever the process is in state *i,* the probability p_{ij} that it will next be in state *j* is the same, regardless of the path it took before reaching state *i* and the time *t* at which the transition takes place. That is, we suppose that

$$P(X_{t+1}=j|X_t=i, X_{t-1}=i_{t-1},...,X_1=i_1, X_0=i_0)$$

$$= P(\mathbf{X_{t+1}}=j|\mathbf{X_t}=i)=P(\boldsymbol{X_1}=j|\boldsymbol{X_0}=i)=\boldsymbol{p_{ij}} \qquad \textbf{(1.1)}$$

for all states $i_0, i_1,...,i_{-1}$, i, j and all $t \geq 0$. Such a stochastic process is known as a ***Markov chain***. Since probabilities are nonnegative and since the process must make a transition into a different state or stay where it is, we have that

$$p_{ij}\geq 0,\ i, j\geq 0;\ \sum_{j=0}^{\infty} p_{ij}=1,\ i=0,1,\ldots.$$

In applications, the transition probabilities $\{p_{ij}\}$ are typically unknown and have to be estimated from an observed trajectory.

It may be shown that Equation (1.1) is equivalent to stating that, for a Markov chain, conditional on any present state X_t, any finite collection of future states $\{X_{t+j}; j=1,2,\ldots, m\}$ is independent of the past states $\{X_0, X_1,\ldots, X_{t-1}\}$. This formulation of the Markov Property makes no distinction between the 'past' and the 'present.' Hence, for any positive integer n, $\{Y_i^{(n)}= X_{n-i}$; i=0, 1, 2,…, n$\}$ is a Markov chain, a type of ***reversibility*** that motivated this report. A precise definition of reversibility is given below.

We now define the *n*-step transition probabilities p_{ij}^n to be the probability that a process in state *i* will be in state *j* after *n* transitions. That is,

$$p_{ij}^n = P\{X_{n+k} = j|X_k = i\}, n \geq 0, i,j \geq 0, k \geq 0.$$

Of course, $p_{ij}^1 = p_{ij}$. Let $\mathbf{P} = \{p_{ij}\}$ denote the matrix of time-homogeneous transition probabilities and let $\mathbf{P}^{(n)} = \{p_{ij}^n\}$ denote the matrix of *n*-step transition probabilities. By induction it can be shown that $\mathbf{P}^{(n)} = \mathbf{P}^n$, where $P^k= PP^{k-1}$denotes the transition matrix P raised to the power k, k = 1,2, … and $\mathbf{P}^0$ is the identity matrix. That is, the *n*-step transition matrix may be obtained by multiplying the matrix **P** by itself *n* times.

State *j* is said to be *accessible* from state *i* if $p_{ij}^n > 0$ for some $n \geq 0$. This implies that state *j* is accessible from state *i* if and only if, starting in *i,* there is positive probability that the process will at some finite future time enter state *j.* Two states *i* and *j* that are accessible to each other are said to *communicate*, and a Markov Chain is said to be ***irreducible*** if all states communicate with each other.

State *i* is said to have *period d* if $p_{ii}^n = 0$ whenever *n* is not divisible by *d*, and *d* is the largest integer with this property. A state with period 1 is said to be ***aperiodic***.

For any state *i* we let f_i denote the probability that, starting in state *i,* the process will re-enter state *i*. State *i* is said to be *recurrent* if $f_i = 1$ and *transient* if $f_i < 1$. If state *i* is recurrent, then it is said to be ***positive recurrent*** if, starting in *i,* the expected time until the process returns to state *i* is finite.

From now on, we suppose that $\{X_i$; i=0,1,2,…$\}$ is irreducible, aperiodic, and positive recurrent. Then, from Ross (2003), $Lim(P^n) = \Pi$, a matrix consisting of identical copies of a probability vector $\boldsymbol{\pi} = (\pi_1, \pi_2, ...)$, so that for all possible states *i* and *j*

$$\pi_j = \lim_{n\to\infty} \left(P(X_n = j | X_0 = i) \right).$$

Under our assumptions, $\{\pi_j\}$ are the unique nonnegative solutions of

$$\pi_j = \sum_{i=0}^{\infty} \pi_i p_{ij},\ j \geq 0,$$

$$\sum_{j=0}^{\infty} \pi_j = 1.$$

The vector $\boldsymbol{\pi} = \{\pi_j > 0,\ j=0,1,2,\ldots\}$ is called the **stationary distribution** of $\{X_i$; i=0,1,2,…$\}$. It follows directly from this definition that for such a chain having a finite number of states and a symmetric transition matrix, the stationary distribution assigns equal probability to each of its states. Another simple example is given by the 3×3 transition matrix

$$\mathrm{P} = \begin{pmatrix} .48 & .26 & .26 \\ .3 & .4 & .3 \\ .17 & .17 & .66 \end{pmatrix}$$

The stationary distribution, computed using (Appendix)[1], here is given approximately by

$$\boldsymbol{\pi} \approx (0.2944573,\ 0.2551963,\ 0.4503464).$$

There are many applications of Markov chains to a wide range of disciplines, such as physics, chemistry, medicine, music, game theory and sports. A famous Markov chain is the "drunkard's walk", a random walk on the integers, where, at each step, the position may change by +1 or -1 with equal probability. From any position, there are two possible transitions, to the next or previous integer. The transition probabilities depend only on the current position, not on the manner in which the position was reached. More generally, a game which is played repeatedly and whose changes of state are independently determined by the same stochastic mechanism at each time period is a Markov chain.

In the following chapters, based on having observed $\{X_i, i = 0,1,2,...,t\}$,we will design and carry out simulation studies, using representative choices of t and **P**, to compare the performance of the classical maximum likelihood estimator (mle) and the reversible estimator, in terms of coverage rate and mean width, of nominal 0.95 confidence intervals for the elements of the transition matrix **P**. We will replace the limiting variances by appropriate sample estimates.

Chapter 2 - Reversible Markov Chain

A Markov Chain having the properties described above is said to be ***reversible*** if

$$\boldsymbol{\pi_i p_{ij} = \pi_j p_{ji}} \tag{2.1}$$

for each pair of states *i* and *j*. Reversibility implies that the long-term flow rate from state *i* to state *j* equals the long time flow rate from state *j* to *i*. Specifically, suppose that a reversible chain is started with its stationary distribution, so that for all states

$$P(X_l = i) = \pi_i,\ l=0,1,\ldots,t.$$

Then, as shown in Section 6.1 of Bremaud (1998), $Q = (q_{ij}) = (\pi_j p_{ji}/\pi_i)$ is the transition matrix of the chain run backward, which is also a Markov Chain, and $q_{ij} = P(X_t=j|X_{t+1}=i) = p_{ij}$ for all *t* and pairs of states (i, j) and

$$\lim_{t\to\infty} \boldsymbol{P(X_0 = j|X_t = i)} = \lim_{t\to\infty} \boldsymbol{p_{ij}^{(t)} = \pi_j} \tag{2.2}$$

It is well known, see Jiang (2009) for example, that a positive recurrent, irreducible, homogeneous Markov chain having a finite number of states and a symmetric transition matrix is reversible. A slight generalization of (2.1) occurs when such a chain is symmetric but only recurrent and not *positive* recurrent, which can only happen, as shown in Theorem 3.3 of Bremaud (1998), when the state space is not finite. Specifically, from Theorem 3.4 of Karlin (1966), for an irreducible, homogeneous Markov Chain that is *recurrent* but not necessarily *positive* recurrent, the system of equations

$$V_j=\sum_{i=0}^{\infty} V_i p_{ij}\ ,\ V_0 = 1, j = 1,2,\ldots$$

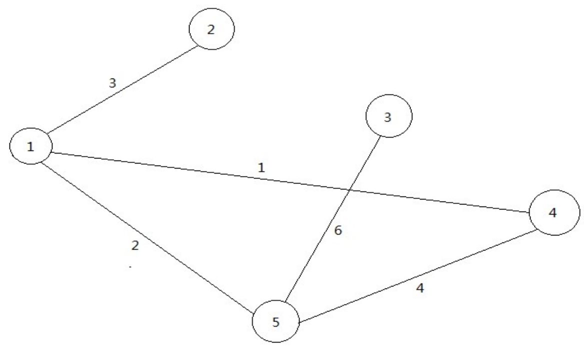
has a unique solution. Further, when $\{V_i > 0; i = 1,2,\ldots\}$, $Q = (q_{ij}) = (V_j p_{ij}/V_i)$ is the transition matrix of the chain run backwards. This holds, for example, if the transition matrix of such a chain is symmetric, whence $V_j = 1$ for all states *j* and the chain is called *quasi- reversible* since $q_{ij} = p_{ij}$ for all pairs of states. But, the limiting distribution (2.2) with π_j replaced by V_j need not hold.

Several broad classes of Markov chains, including those with symmetric transition matrices, random walks on graphs, and birth and death processes, could be reversible. For one specific example, consider an arbitrary connected graph having a positive number w_{ij} associated with arc (i, j). It may be helpful to think of various U.S. cities as the states in the chain, with an arc existing between cities *i* and *j* when it is possible to fly directly from city *i* to *j* , where the cost of traveling directly from city i to j is w_{ij}. Symmetry is assumed in that one can fly directly from j to i if it is possible to fly directly from i to j. We also take $w_{ij} = w_{ji}$ and the probability, regardless of previous states visited, of undergoing a transition from i to j is taken to be proportional to its cost, so that

$$p_{ij} = \frac{w_{ij}}{\sum_{j \in S} w_{ij}}.$$

For instance, for the graph of Figure 2-1, where weights are given along the lines connecting nodes, P_{12}=3/(3+1+2)=1/2.

Figure 2-1



This Markov chain is reversible with $\pi_i = \frac{\sum_j w_{ij}}{\sum_i \sum_j w_{ij}}$ and $q_{ij} = \frac{w_{ji}}{\sum_j w_{ij}}$.

A second example is a hypothetical stock market exhibiting a bull market, bear market, or stagnant market trend during a given week. Suppose a bull week is followed by another bull week 90% of the time, a bear week 7.5% of the time, and a stagnant week the other 2.5% of the time. Labelling the state space {1 = bull, 2 = bear, 3 = stagnant} the transition matrix for this example is

$$\mathbf{P} = \begin{pmatrix} .9 & .075 & .025 \\ .15 & .8 & .05 \\ .25 & .25 & .5 \end{pmatrix},$$

where, respectively, the second and third rows represent transition probabilities from a bear market and from a stagnant market. This Markov chain is reversible with

$$\boldsymbol{\pi} = (0.625, 0.3125, 0.0625).$$

Chapter 3 - Estimation in Reversible Markov Chains

According to Annis et. al. (2010), suppose we observe the data consisting of realized values of $X_0, \dots, X_t$ and we wish to estimate the one-step-ahead transition probabilities p_{ij} for all pairs of states *i, j* ∊ S. The well-known classical maximum likelihood estimator (mle) of p_{ij} is

$$\boldsymbol{\hat{p}_{ij}(t)} = \frac{\boldsymbol{N_{ij}(t)}}{\boldsymbol{N_i(t)}} \mathbf{1}_{[\boldsymbol{N_i(t)>0}]} \tag{3.1}$$

where $1_{[A]}$ is an indicator that is 1 when the event *A* occurs and zero otherwise, $N_{ij}(t)$ is the number of one-step-ahead transitions from *i* to *j*, and $N_i(t)$ is the number of times state *i* is visited up to time *t*. The indicator $1_{[N_i(t)>0]}$ in (3.1) is introduced to avoid division by zero. The counts $N_{ij}(t)$ and $N_i(t)$ are explicitly defined by

$$N_{ij}(t) = \sum_{l=0}^{t-1} 1_{[X_l = i \cap X_{l+1} = j]} \text{ and}$$

$$\boldsymbol{N_i(t)} = \sum_{\boldsymbol{l=0}}^{\boldsymbol{t}} \mathbf{1}_{[\boldsymbol{X_l = i}]}.$$

For reversible chains, following questions may arise—is the mle the best asymptotic estimator, does priori knowledge of a chain's reversibility aid transition probability estimation? These questions were beautifully answered by Greenwood and Wefelmeyer (1999) and Greenwood, Schick, and Wefelmeyer (2001) who showed that the symmetrized (reversible) estimator

$$\boldsymbol{\hat{p}_{ij}^{(R)}(t)} = \frac{\boldsymbol{N_{ij}(t)+N_{ji}(t)}}{\boldsymbol{2N_i(t)}} \mathbf{1}_{[\boldsymbol{N_i(t)>0}]}. \tag{3.2}$$

is not only preferable for $i \neq j$, but also asymptotically most efficient.

A heuristic motivation for (3.2) follows from the law of large numbers applied to irreducible, positive recurrent Markov chains. Specifically, for large *t*, we have that $N_{ij}/N_i \approx p_{ij}$, $N_{ji}/N_j \approx p_{ji}$, $N_j/t \approx \pi_j$ and $N_i/t \approx \pi_i$. Then, for large *t* and $N_i N_j > 0$, using the definition of reversibility given in (2.1),

$$\hat{p}_{ij}^{(R)} = \frac{(N_{ij}/N_i)N_i + (N_{ji}/N_j)N_j}{2N_i}$$

$$\approx [p_{ij}N_i + p_{ji}N_j]/2\,N_i$$
$$\approx [p_{ij}(N_i/t)t + p_{ji}(N_j/t)t]/[(2\,N_i/t)t]$$
$$\approx [p_{ij}\pi_i t + p_{ji}\pi_j t]/2\,\pi_i t$$
$$=2p_{ij}\pi_i/2\,\pi_i=p_{ij}.$$

Clearly, the maximum likelihood and reverse estimators are identical when $i = j$. The estimator in (3.2) can be viewed as merely averaging forward and backward versions of (3.1). Annis et. al. (2010) show that the asymptotic variance of the reversible estimator is never lager than that of the classical estimator, specifically that

$$\lim_{t\to\infty}\frac{Var(\hat{p}_{ij}^{(R)}(t))}{Var(\hat{p}_{ij}(t))} \in \left[\frac{1}{2}, 1\right].$$

They show that for $i \neq j$ as $t \to \infty$, we have the following distributional convergence.

$$\sqrt{t}\left(\hat{p}_{ij}^{(R)}(t) - p_{ij}\right) \xrightarrow{D} N\left(0, \frac{\left(p_{ij}-p_{ij}^2\right)+\left(p_{ij}\sum_{k=0}^{\infty} ip_{ij}^{(k)}p_{ji}-p_{ij}^2\right)}{2\pi_i}\right). \qquad (3.3)$$

where $ip_{ij}^{(k)}$ is the "taboo probability" that starting from state i, the chain is in state j at time t and the first return time to state i is greater than k. Here, the adjective "taboo" indicates that state i must be avoided during the interior times in the cycle. Mathematically, when k=1, $ip_{ij}^{(k)}$=p_{ij}, and for k≥2,

$$ip_{ij}^{(k)}=\sum_{l_1,\dots,l_{k-1}\neq i} p_{i,l_1}p_{l_1,l_2}\dots p_{l_{k-1}j}=(p_{it})_{t\neq i}(p_{ts})_{t,s\neq i}^{k-2}(p_{sj})_{s\neq i}. \qquad (3.4)$$

To use (3.3), we must approximate the infinite sum $\sum_{k=0}^{\infty} ip_{ij}^{(k)}p_{ji}$, by truncating it. We found, using the package R (Appendix A[3]), that partial sums $\sum_{k=0}^{m} ip_{ij}^{(k)}p_{ji}$ appeared to converge very rapidly for $m \geq 75$. Consequently, in our estimation of the standard errors of reversible estimators, we approximated $\sum_{k=0}^{\infty} ip_{ij}^{(k)}p_{ji}$ by $\sum_{k=0}^{75} ip_{ij}^{(k)}p_{ji}$.

In addition, it is well known, as shown in Annis et. al. (2010), that for all pairs (i, j) $\in S$, in distribution as $t \to \infty$,

$$\sqrt{t}\left(\hat{p}_{ij}(t) - p_{ij}\right) \xrightarrow{D} N\left(0, \frac{p_{ij}-p_{ij}^2}{\pi_i}\right). \qquad (3.5)$$

Note that since $\hat{p}_{ii} = \hat{p}_{ii}^{(R)}$, (3.5) is the appropriate standard error for both estimators when estimating transitions from a state back to itself.

.

According to (3.3) and (3.5), we can estimate the standard error of the mle estimator by

$$\mathrm{se}(\hat{\boldsymbol{p}}_{ij})=\sqrt{\frac{\hat{\boldsymbol{p}}_{ij}-\hat{\boldsymbol{p}}_{ij}^{2}}{\boldsymbol{N}_i}} \qquad \textbf{(3.6)}$$

and for $i \neq j$ the standard error of the reversible estimator by

$$\mathrm{se}(\hat{\boldsymbol{p}}_{ij}^{(R)})=\sqrt{\frac{\left(\hat{\boldsymbol{p}}_{ij}^{(R)}-\hat{\boldsymbol{p}}_{ij}^{(R)^2}\right)-\left(\hat{\boldsymbol{p}}_{ij}^{(R)}\sum_{k=0}^{75} i\hat{\boldsymbol{p}}_{ij}^{(R)^{(k)}}\hat{\boldsymbol{p}}_{ji}^{(R)}-\hat{\boldsymbol{p}}_{ij}^{(R)^2}\right)}{2\boldsymbol{N}_i}}\,. \qquad \textbf{(3.7)}$$

In simulations, since the finite approximation $\sum_{k=0}^{75} i\hat{p}_{ij}^{(R)^{(k)}}\hat{p}_{ji}^{(R)}$ in (3.7) may yield a negative numerator inside the square root, we set the standard error of the reversible estimator equal to the standard error of the mle (which is asymptotically never smaller) in such cases and we record the proportion of times that this happens. For $i = j$, $se(\hat{p}_{ii}^{(R)}) = se(\hat{p}_{ii})$. Since estimating taboo probabilities adds considerable variability to estimating the standard errors of the reverse estimators, as shown below, we were not surprised to find that our simulation study failed to find a systematic finite sample size advantage for this asymptotically better procedure.

In next chapter, we will set up a four-step estimation algorithm and compare 95% confidence intervals of the mle and reversible estimators. Explicitly, an approximate 1- α confidence interval for the mle will be

$$\hat{p}_{ij} \pm z_{\alpha/2} se(\hat{p}_{ij})$$

and an approximate 1- α confidence interval for the reversible estimator will be

$$\hat{p}_{ij}^{(R)} \pm z_{\alpha/2} se(\hat{p}_{ij}^{(R)}).$$

Chapter 4 - Simulations

4.1 A Four-step Estimation Algorithm

In this Section we describe and summarize the results of our simulation study of the performance, in terms of coverage rates and interval widths of 1- α = .95 confidence intervals for the transition probabilities of a Markov Chain assumed to be reversible. Since, as noted above, the reversible and maximum likelihood estimators are identical for $i = j$, the diagonal entries in the summary matrices below are likewise identical. We start by illustrating our simulation using the symmetric, and hence reversible, 3-state Markov Chain whose symmetric transition matrix is given by:

$$\mathbf{P}=\begin{pmatrix} .5000000 & .2500000 & .2500000 \\ .2500000 & .3333333 & .4166667 \\ .2500000 & .4166667 & .3333333 \end{pmatrix}.$$

Since $\mathbf{P}$ is symmetric, its stationary distribution is the vector

$$\boldsymbol{\pi} = (0.3333333, 0.3333333, 0.3333333).$$

The four steps for generating a trajectory consisting of n observed transitions are as follows.

1. Generate a trajectory $X_0,\ldots, X_n$ with X_0 taken from the stationary distribution. For example, we generated a 100 steps, resulting in the states visited sequentially as:

   3 3 3 3 3 3 1 3 1 3 2 1 2 3 1 2 2 1 1 2 3 1 1 1 2 2 2 2 2 1 3 2 3 1 1 1 2
   3 3 3 2 1 3 2 3 3 3 2 1 3 2 3 2 3 2 1 1 3 2 3 1 2 3 1 2 2 2 1 1 1 1 1 1 2
   1 1 1 1 2 3 1 2 2 3 2 1 1 3 2 3 2 3 1 1 2 3 2 3 1 1 3

2. Estimate the maximum likelihood estimators $\{\hat{p}_{ij}\}$ and the symmetrized estimators $\{\hat{p}_{ij}^{(R)}\}$ of $\{p_{ij}\}$. For instance, based on the simulation example from step 1, we estimate the maximum likelihood estimators $\{\hat{p}_{ij}\}$ by (3.1) and the symmetrized estimators $\{\hat{p}_{ij}^{(R)}\}$ by (3.2) and get:

$$\widehat{\mathbf{P}} = \{\hat{p}_{ij}\} = \begin{pmatrix} 0.4444444 & 0.2592593 & 0.2962963 \\ 0.1764706 & 0.3823529 & 0.4411765 \\ 0.2250000 & 0.3500000 & 0.4000000 \end{pmatrix}$$

and

$$\widehat{\mathbf{P}}^{(R)} = \left\{\hat{p}_{ij}^{(R)}\right\} = \begin{pmatrix} 0.4444444 & 0.2407407 & 0.3148148 \\ 0.1911765 & 0.3823529 & 0.4264706 \\ 0.2125000 & 0.3625000 & 0.4000000 \end{pmatrix}$$

respectively.

3. Compute se($\hat{p}_{ij}$) and se($\hat{p}_{ij}^{(R)}$) . According formulas (3.6) and (3.7) we get

$$\text{se}(\widehat{\mathbf{P}}) = \begin{pmatrix} 0.09562922 & 0.08433704 & 0.08787719 \\ 0.06537870 & 0.08334181 & 0.08515380 \\ 0.06602556 & 0.07541552 & 0.07745967 \end{pmatrix}$$

and

$$\text{se}(\widehat{\mathbf{P}}^{(R)}) = \begin{pmatrix} 0.09562922 & 0.05772281 & 0.06215590 \\ 0.04751910 & 0.08334181 & 0.05880081 \\ 0.04573617 & 0.05374637 & 0.07745967 \end{pmatrix}.$$

4. Construct 95% C.I.'s for $\{\hat{p}_{ij}\}$ and $\{\hat{p}_{ij}^{(R)}\}$. The matrix of 95% C.I.'s for $\{\hat{p}_{ij}\}$ is:

$$\begin{pmatrix} (0.25701118, 0.6318777) & (0.09395865, 0.4245599) & (0.1240570, 0.4685356) \\ (0.04832834, 0.3046128) & (0.21900299, 0.5457029) & (0.2742750, 0.6080779) \\ (0.09558990, 0.3544101) & (0.20218559, 0.4978144) & (0.2481791, 0.5518209) \end{pmatrix}$$

and the matrix of 95% C.I. for $\{\hat{p}_{ij}^{(R)}\}$ is:

$$\begin{pmatrix} (0.25701118, 0.6318777) & (0.1276040, 0.3538774) & (0.1929892, 0.4366404) \\ (0.09803904, 0.2843139) & (0.2190030, 0.5457029) & (0.3112210, 0.5417202) \\ (0.12285711, 0.3021429) & (0.2571571, 0.4678429) & (0.2481791, 0.5518209) \end{pmatrix}$$

From these results, for this particular chain and this particular trajectory, we see that for $i \neq j$, as expected, using the reversible estimator we get smaller standard errors and hence narrower confidence intervals. And for this particular chain and trajectory, for both the mle and reversible estimator, every interval successfully contains the target p_{ij}. However, to draw more reliable conclusions, we need to check and compare the coverage rates and mean widths of both estimators obtained from a suitably large number of simulations.

4.2 Simulation of Reversible 3-state Markov Chain

Now, independently repeat steps 1-4 in Section 4.1 one thousand times and calculate the attained coverage rates and mean lengths. Using t to denote the number of observed transitions, we constructed tables and summarized the results below:

Table 4-1 Comparing Estimated Coverage Rates of Nominal 95% Confidence Intervals

t	Coverage Rate For $\{\hat{p}_{ij}\}$	Coverage Rate For $\{\hat{p}_{ij}^{(R)}\}$	Proportion of Negative Numerator For $\{\hat{p}_{ij}^{(R)}\}$
25	$\begin{pmatrix} 0.874 & 0.859 & 0.846 \\ 0.853 & 0.854 & 0.886 \\ 0.874 & 0.864 & 0.886 \end{pmatrix}$	$\begin{pmatrix} 0.874 & 0.834 & 0.846 \\ 0.838 & 0.854 & 0.818 \\ 0.826 & 0.798 & 0.886 \end{pmatrix}$	$\begin{pmatrix} 0.000 & 0.001 & 0.002 \\ 0.003 & 0.000 & 0.007 \\ 0.000 & 0.004 & 0.000 \end{pmatrix}$
50	$\begin{pmatrix} 0.920 & 0.911 & 0.910 \\ 0.907 & 0.904 & 0.908 \\ 0.901 & 0.934 & 0.909 \end{pmatrix}$	$\begin{pmatrix} 0.920 & 0.852 & 0.843 \\ 0.862 & 0.904 & 0.841 \\ 0.865 & 0.835 & 0.909 \end{pmatrix}$	
75	$\begin{pmatrix} 0.949 & 0.926 & 0.921 \\ 0.916 & 0.927 & 0.924 \\ 0.929 & 0.944 & 0.909 \end{pmatrix}$	$\begin{pmatrix} 0.949 & 0.877 & 0.895 \\ 0.872 & 0.927 & 0.844 \\ 0.875 & 0.868 & 0.909 \end{pmatrix}$	
100	$\begin{pmatrix} 0.946 & 0.937 & 0.918 \\ 0.953 & 0.921 & 0.925 \\ 0.925 & 0.933 & 0.927 \end{pmatrix}$	$\begin{pmatrix} 0.946 & 0.881 & 0.892 \\ 0.869 & 0.921 & 0.862 \\ 0.876 & 0.859 & 0.927 \end{pmatrix}$	
125	$\begin{pmatrix} 0.935 & 0.937 & 0.925 \\ 0.939 & 0.926 & 0.944 \\ 0.928 & 0.923 & 0.926 \end{pmatrix}$	$\begin{pmatrix} 0.935 & 0.878 & 0.882 \\ 0.876 & 0.926 & 0.873 \\ 0.887 & 0.882 & 0.926 \end{pmatrix}$	
150	$\begin{pmatrix} 0.937 & 0.932 & 0.935 \\ 0.944 & 0.923 & 0.945 \\ 0.908 & 0.935 & 0.914 \end{pmatrix}$	$\begin{pmatrix} 0.937 & 0.883 & 0.883 \\ 0.881 & 0.923 & 0.880 \\ 0.894 & 0.861 & 0.914 \end{pmatrix}$	
175	$\begin{pmatrix} 0.934 & 0.933 & 0.941 \\ 0.945 & 0.944 & 0.942 \\ 0.939 & 0.938 & 0.939 \end{pmatrix}$	$\begin{pmatrix} 0.934 & 0.885 & 0.894 \\ 0.873 & 0.944 & 0.878 \\ 0.888 & 0.868 & 0.939 \end{pmatrix}$	
200	$\begin{pmatrix} 0.941 & 0.938 & 0.946 \\ 0.926 & 0.939 & 0.942 \\ 0.951 & 0.949 & 0.943 \end{pmatrix}$	$\begin{pmatrix} 0.941 & 0.873 & 0.893 \\ 0.873 & 0.939 & 0.858 \\ 0.877 & 0.850 & 0.943 \end{pmatrix}$	

Table 4-2 Comparing Mean Lengths of a 3-state Markov Chain

t	Mean Length for $\{\hat{p}_{ij}\}$	Mean Length for $\{\hat{p}_{ij}^{(R)}\}$
25	$\begin{pmatrix} 0.6091309 & 0.5158375 & 0.5331273 \\ 0.5308894 & 0.5453941 & 0.6134530 \\ 0.5187765 & 0.6188529 & 0.5377003 \end{pmatrix}$	$\begin{pmatrix} 0.6091309 & 0.3910093 & 0.3904539 \\ 0.3841363 & 0.5453941 & 0.4263079 \\ 0.3833964 & 0.4251637 & 0.5377003 \end{pmatrix}$
50	$\begin{pmatrix} 0.4656950 & 0.4123620 & 0.4068140 \\ 0.4054860 & 0.4249900 & 0.4601700 \\ 0.3960310 & 0.4584800 & 0.4218490 \end{pmatrix}$	$\begin{pmatrix} 0.4656950 & 0.2885180 & 0.2910710 \\ 0.2812810 & 0.4249900 & 0.3209130 \\ 0.2846890 & 0.3225200 & 0.4218490 \end{pmatrix}$
75	$\begin{pmatrix} 0.3876573 & 0.3381155 & 0.3395952 \\ 0.3306596 & 0.3585567 & 0.3805194 \\ 0.3344476 & 0.3788799 & 0.3576499 \end{pmatrix}$	$\begin{pmatrix} 0.3876573 & 0.2385560 & 0.2364456 \\ 0.2364594 & 0.3585567 & 0.2659568 \\ 0.2374712 & 0.2665049 & 0.3576499 \end{pmatrix}$
100	$\begin{pmatrix} 0.3368830 & 0.2912000 & 0.2947190 \\ 0.2903310 & 0.3113930 & 0.3308150 \\ 0.2917050 & 0.3317860 & 0.3125170 \end{pmatrix}$	$\begin{pmatrix} 0.3368830 & 0.2056140 & 0.2067460 \\ 0.2054170 & 0.3113930 & 0.2304730 \\ 0.2041480 & 0.2314550 & 0.3125170 \end{pmatrix}$
125	$\begin{pmatrix} 0.3023364 & 0.2598999 & 0.2607389 \\ 0.2577371 & 0.2808252 & 0.2952987 \\ 0.2591458 & 0.2965395 & 0.2810984 \end{pmatrix}$	$\begin{pmatrix} 0.3023364 & 0.1845738 & 0.1854798 \\ 0.1839777 & 0.2808252 & 0.2079565 \\ 0.1843693 & 0.2079322 & 0.2810984 \end{pmatrix}$
150	$\begin{pmatrix} 0.2756090 & 0.2364420 & 0.2379820 \\ 0.2378760 & 0.2570920 & 0.2710830 \\ 0.2382120 & 0.2714630 & 0.2570150 \end{pmatrix}$	$\begin{pmatrix} 0.2756090 & 0.1681980 & 0.1695130 \\ 0.1674840 & 0.2570920 & 0.1907330 \\ 0.1687680 & 0.1913110 & 0.2570150 \end{pmatrix}$
175	$\begin{pmatrix} 0.2549996 & 0.2216255 & 0.2212562 \\ 0.2211639 & 0.2391926 & 0.2509283 \\ 0.2218496 & 0.2507953 & 0.2387933 \end{pmatrix}$	$\begin{pmatrix} 0.2549996 & 0.1564232 & 0.1566066 \\ 0.1563496 & 0.2391926 & 0.1768742 \\ 0.1550236 & 0.1771447 & 0.2387933 \end{pmatrix}$
200	$\begin{pmatrix} 0.2387940 & 0.2074410 & 0.2080130 \\ 0.2071640 & 0.2242150 & 0.2357600 \\ 0.2065900 & 0.2359420 & 0.2235820 \end{pmatrix}$	$\begin{pmatrix} 0.2387940 & 0.1468570 & 0.1463120 \\ 0.1456700 & 0.2242150 & 0.1652270 \\ 0.1459990 & 0.1651290 & 0.2235820 \end{pmatrix}$

The results of Table 4-1 and Table 4-2 indicate that the mean lengths based on reversible estimators tend to be smaller than mean lengths of 95% C.I.'s based on maximum likelihood estimators. However, the estimated coverage rates of intervals based on the reverse estimators are unacceptably below the target 95% rate for the small values of *t* used in this simulation.

To graphically display these tables, we used scatterplots and boxplots to further compare coverage rates and mean lengths confidence intervals for all nine $\{p_{ij}\}$ constructed using the mle and the reversible estimators:

Figure 4-1 Scatterplots of Coverage Rate for 3-state Chain

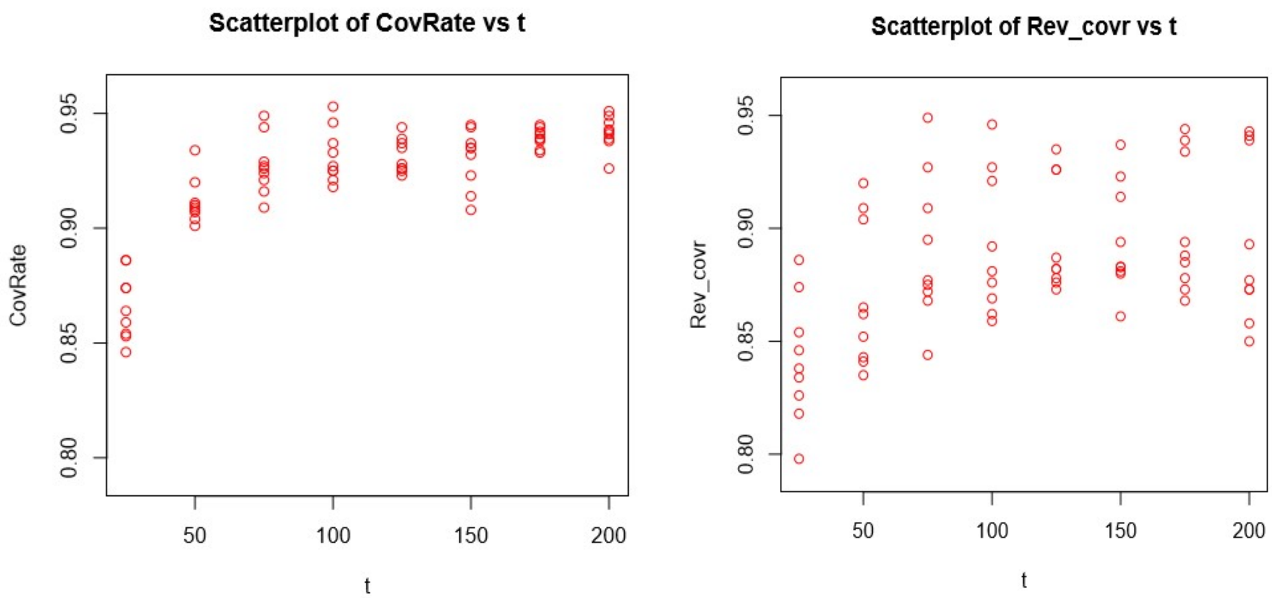
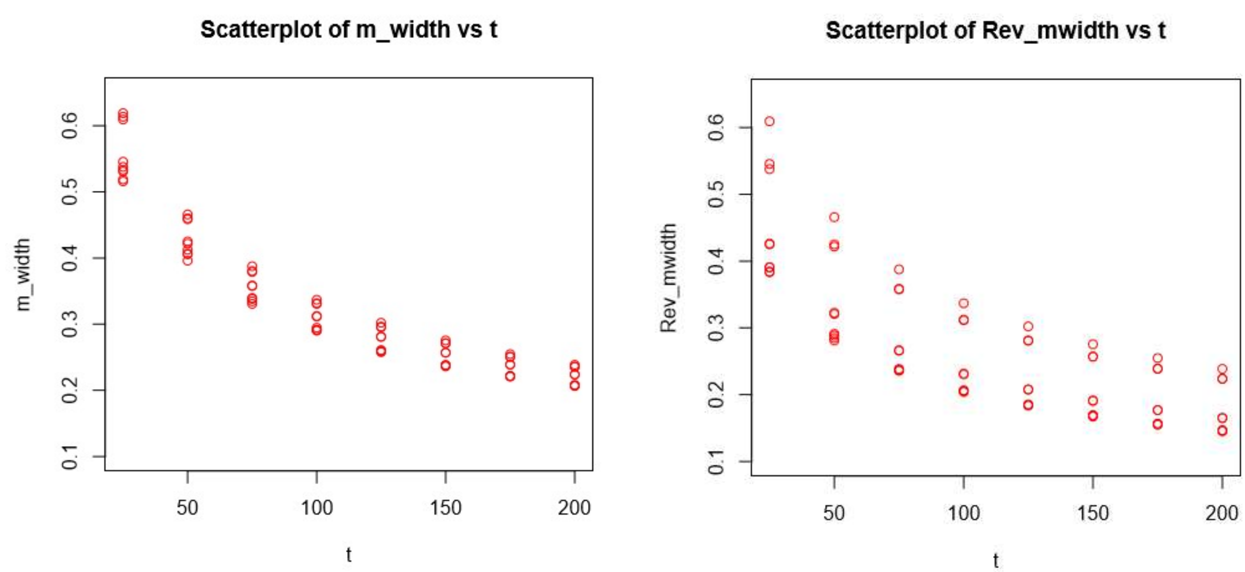


Figure 4-2 Scatterplots of Mean Width for 3-state Chain



From these results, as expected, we see that the widths of both types of confidence intervals tend to decrease with increasing t and that estimated mean lengths based on reversible estimator tend to be smaller than mean lengths of 95% C.I.'s based on maximum likelihood estimator. Unfortunately, as noted above, this advantage is outweighed by coverage rates of intervals based on the reverse estimators appearing to be unacceptably below their nominal .95 values. To better visualize this relationship, we picked one pair of states (*i*=2, *j*=1) as an example and present below side by side box plots of the estimated coverage rates aggregated over the eight values of *t* and widths of the two types of confidence intervals for selected values of t.

Figure 4-3 Boxplots Comparing Coverage Rates of 1000 Simulated Confidence Intervals for p_{21}

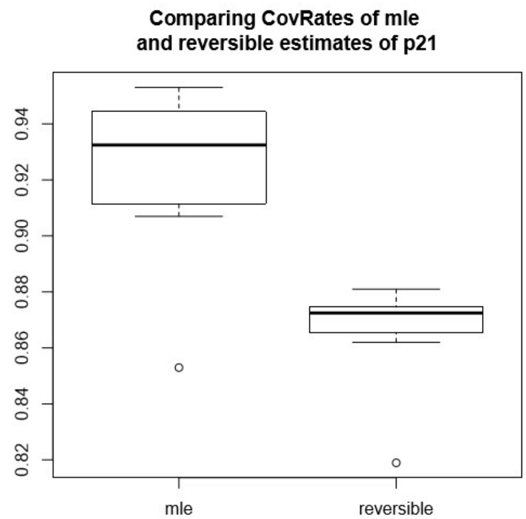
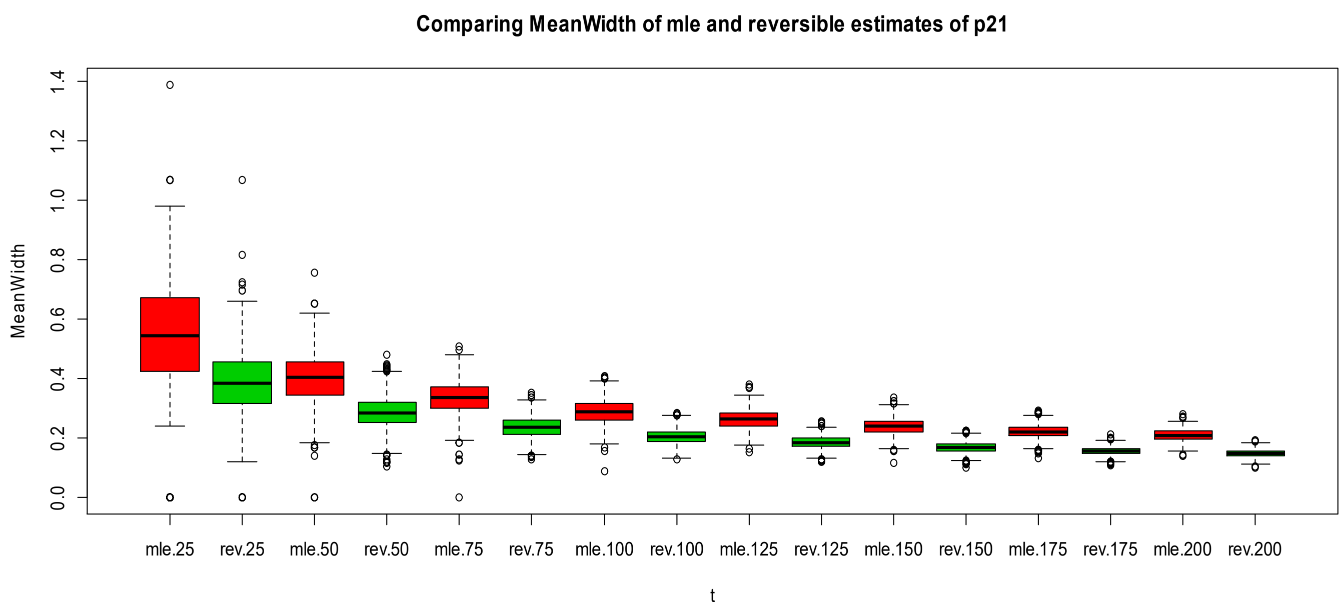


Figure 4-4 Boxplots Comparing Widths of 1000 Simulated Confidence Intervals for p_{21}



From these boxplots we can see that as the number of transitions t increases, the boxes become very narrow and the widths of the one thousand 95% confidence intervals based on the mle appear to approach 0.2 and those based on the reversible estimator appear to approach 0.15. But, again as noted above, this advantage of intervals based on the reverse estimators is undercut the by the coverage rates of the mle based confidence intervals being much closer to their nominal .95 target than the coverage rates based on the reversible estimator, which tend to be unacceptably low.

4.3 Simulation of Reversible 4-state Markov Chain

We used the 4-state chain whose transition matrix is given by

$$\mathbf{P}_4 = \begin{pmatrix} 0.2500000 & 0.2500000 & 0.3333333 & 0.1666667 \\ 0.1875000 & 0.3375000 & 0.1000000 & 0.3750000 \\ 0.5000000 & 0.2000000 & 0.1500000 & 0.1500000 \\ 0.1666667 & 0.5000000 & 0.1000000 & 0.2333333 \end{pmatrix}.$$

Its stationary distribution is, approximately,

$$\boldsymbol{\pi} = (0.2500000,\ 0.3333333,\ 0.1666667,\ 0.2500000).$$

Independently repeating steps 1-4 in section 4.1 one thousand times, we obtained:

Table 4-3 Comparing Coverage Rates for a 4-state Markov Chain

n	Coverage Rate For $\{\hat{p}_{ij}\}$	Coverage Rate For $\{\hat{p}_{ij}^{(R)}\}$	Proportion of Negative Numerator For $\{\hat{p}_{ij}^{(R)}\}$
25	$\begin{pmatrix} 0.709 & 0.797 & 0.846 & 0.677 \\ 0.810 & 0.871 & 0.571 & 0.885 \\ 0.758 & 0.595 & 0.473 & 0.472 \\ 0.666 & 0.854 & 0.478 & 0.721 \end{pmatrix}$	$\begin{pmatrix} 0.709 & 0.840 & 0.807 & 0.788 \\ 0.798 & 0.871 & 0.798 & 0.810 \\ 0.772 & 0.783 & 0.473 & 0.695 \\ 0.786 & 0.801 & 0.698 & 0.721 \end{pmatrix}$	$\begin{pmatrix} 0.000 & 0.004 & 0.003 & 0.002 \\ 0.000 & 0.000 & 0.000 & 0.001 \\ 0.047 & 0.009 & 0.000 & 0.008 \\ 0.001 & 0.022 & 0.000 & 0.000 \end{pmatrix}$
50	$\begin{pmatrix} 0.867 & 0.891 & 0.871 & 0.874 \\ 0.884 & 0.889 & 0.832 & 0.908 \\ 0.881 & 0.821 & 0.673 & 0.731 \\ 0.878 & 0.884 & 0.880 & 0.737 \end{pmatrix}$	$\begin{pmatrix} 0.867 & 0.868 & 0.847 & 0.865 \\ 0.870 & 0.889 & 0.846 & 0.846 \\ 0.848 & 0.849 & 0.673 & 0.839 \\ 0.874 & 0.821 & 0.850 & 0.737 \end{pmatrix}$	
75	$\begin{pmatrix} 0.898 & 0907 & 0.910 & 0.894 \\ 0.901 & 0.928 & 0.905 & 0.922 \\ 0.894 & 0.890 & 0.785 & 0.844 \\ 0.895 & 0.917 & 0.864 & 0.901 \end{pmatrix}$	$\begin{pmatrix} 0.898 & 0.884 & 0.878 & 0.880 \\ 0.864 & 0.928 & 0.893 & 0.830 \\ 0.859 & 0.868 & 0.785 & 0.861 \\ 0.888 & 0.853 & 0.859 & 0.901 \end{pmatrix}$	

100	$\begin{pmatrix} 0.906 & 0.930 & 0.932 & 0.907 \\ 0.907 & 0.937 & 0.878 & 0.932 \\ 0.923 & 0.896 & 0.847 & 0.897 \\ 0.917 & 0.947 & 0.883 & 0.914 \end{pmatrix}$	$\begin{pmatrix} 0.906 & 0.877 & 0.871 & 0.904 \\ 0.887 & 0.937 & 0.873 & 0.878 \\ 0.864 & 0.907 & 0.847 & 0.877 \\ 0.888 & 0.875 & 0.892 & 0.914 \end{pmatrix}$
125	$\begin{pmatrix} 0.927 & 0.939 & 0.921 & 0.921 \\ 0.937 & 0.937 & 0.902 & 0.945 \\ 0.937 & 0.909 & 0.881 & 0.885 \\ 0.919 & 0.941 & 0.870 & 0.920 \end{pmatrix}$	$\begin{pmatrix} 0.927 & 0.906 & 0.876 & 0.884 \\ 0.888 & 0.937 & 0.900 & 0.885 \\ 0.892 & 0.909 & 0.881 & 0.916 \\ 0.904 & 0.864 & 0.900 & 0.920 \end{pmatrix}$
150	$\begin{pmatrix} 0.915 & 0.946 & 0.927 & 0.922 \\ 0.921 & 0.940 & 0.912 & 0.938 \\ 0.930 & 0.927 & 0.889 & 0.892 \\ 0.919 & 0.948 & 0.897 & 0.918 \end{pmatrix}$	$\begin{pmatrix} 0.915 & 0.916 & 0.890 & 0.885 \\ 0.884 & 0.940 & 0.901 & 0.868 \\ 0.883 & 0.909 & 0.889 & 0.883 \\ 0.894 & 0.877 & 0.883 & 0.918 \end{pmatrix}$
175	$\begin{pmatrix} 0.926 & 0.944 & 0.942 & 0.926 \\ 0.935 & 0.935 & 0.927 & 0.948 \\ 0.926 & 0.914 & 0.895 & 0.906 \\ 0.925 & 0.949 & 0.922 & 0.917 \end{pmatrix}$	$\begin{pmatrix} 0.926 & 0.895 & 0.895 & 0.906 \\ 0.887 & 0.935 & 0.893 & 0.880 \\ 0.865 & 0.906 & 0.895 & 0.914 \\ 0.897 & 0.904 & 0.892 & 0.917 \end{pmatrix}$
200	$\begin{pmatrix} 0.929 & 0.936 & 0.923 & 0.946 \\ 0.936 & 0.931 & 0.906 & 0.936 \\ 0.944 & 0.922 & 0.889 & 0.929 \\ 0.939 & 0.955 & 0.920 & 0.926 \end{pmatrix}$	$\begin{pmatrix} 0.929 & 0.916 & 0.857 & 0.893 \\ 0.887 & 0.931 & 0.886 & 0.869 \\ 0.874 & 0.911 & 0.889 & 0.915 \\ 0.907 & 0.889 & 0.913 & 0.926 \end{pmatrix}$

Table 4-4 Comparing Mean Lengths for a 4-state Markov Chain

n	Mean Length for $\{\hat{p}_{ij}\}$	Mean Length for $\{\hat{p}_{ij}^{(R)}\}$
25	$\begin{pmatrix} 0.4965133 & 0.5637503 & 0.6117802 & 0.4121320 \\ 0.4559840 & 0.5454349 & 0.3009641 & 0.6005126 \\ 0.7156301 & 0.4819536 & 0.3302188 & 0.3952466 \\ 0.4358928 & 0.7001721 & 0.2888476 & 0.4672463 \end{pmatrix}$	$\begin{pmatrix} 0.4965133 & 0.4395931 & 0.4630845 & 0.3617960 \\ 0.3327371 & 0.5454349 & 0.2541011 & 0.4238571 \\ 0.5396530 & 0.4381346 & 0.3302188 & 0.3726498 \\ 0.3578734 & 0.4814722 & 0.2620159 & 0.4672463 \end{pmatrix}$
50	$\begin{pmatrix} 0.4247208 & 0.4576356 & 0.5031072 & 0.3806708 \\ 0.3565417 & 0.4249401 & 0.2540216 & 0.4538301 \\ 0.6381234 & 0.4772112 & 0.3481232 & 0.3882131 \\ 0.3611441 & 0.5340682 & 0.2665987 & 0.4089051 \end{pmatrix}$	$\begin{pmatrix} 0.4247208 & 0.3239083 & 0.3564685 & 0.2830731 \\ 0.2596190 & 0.4249401 & 0.1919330 & 0.3168055 \\ 0.4401312 & 0.3602182 & 0.3481232 & 0.3134373 \\ 0.2807591 & 0.3666626 & 0.2147910 & 0.4089051 \end{pmatrix}$
75	$\begin{pmatrix} 0.3636156 & 0.3783744 & 0.4158473 & 0.3203648 \\ 0.2987004 & 0.3586700 & 0.2232104 & 0.3736334 \\ 0.5417038 & 0.4099529 & 0.3195284 & 0.3531060 \\ 0.3209954 & 0.4414457 & 0.2405822 & 0.3586932 \end{pmatrix}$	$\begin{pmatrix} 0.3636156 & 0.2729880 & 0.2924387 & 0.2309972 \\ 0.2126301 & 0.3586700 & 0.1623520 & 0.2592844 \\ 0.3698469 & 0.3068654 & 0.3195284 & 0.2642671 \\ 0.2320489 & 0.3085292 & 0.1853728 & 0.3586932 \end{pmatrix}$
100	$\begin{pmatrix} 0.3240062 & 0.3298386 & 0.3651459 & 0.2814215 \\ 0.2588588 & 0.3145792 & 0.1951679 & 0.3240941 \\ 0.4676070 & 0.3638239 & 0.2985629 & 0.3195954 \\ 0.2824075 & 0.3878252 & 0.2187587 & 0.3172581 \end{pmatrix}$	$\begin{pmatrix} 0.3240062 & 0.2368710 & 0.2560396 & 0.2047514 \\ 0.1847045 & 0.3145792 & 0.1399821 & 0.2276834 \\ 0.3276854 & 0.2667801 & 0.2985629 & 0.2393920 \\ 0.2023500 & 0.2688822 & 0.1594219 & 0.3172581 \end{pmatrix}$
125	$\begin{pmatrix} 0.2938077 & 0.3012533 & 0.3267188 & 0.2557401 \\ 0.2337163 & 0.2812173 & 0.1765533 & 0.2904689 \\ 0.4222152 & 0.3326345 & 0.2743884 & 0.2902840 \\ 0.2552454 & 0.3444653 & 0.2007840 & 0.2860685 \end{pmatrix}$	$\begin{pmatrix} 0.2938077 & 0.2132665 & 0.2276743 & 0.1794096 \\ 0.1661011 & 0.2812173 & 0.1259528 & 0.2040227 \\ 0.2940371 & 0.2403100 & 0.2743884 & 0.2130262 \\ 0.1829629 & 0.2430123 & 0.1452807 & 0.2860685 \end{pmatrix}$

150	$\begin{pmatrix} 0.2685994 & 0.2735344 & 0.2984904 & 0.2342888 \\ 0.2138766 & 0.2572087 & 0.1630203 & 0.2652182 \\ 0.3880024 & 0.3073964 & 0.2569112 & 0.2663075 \\ 0.2373027 & 0.3177850 & 0.1839005 & 0.2637705 \end{pmatrix}$	$\begin{pmatrix} 0.2685994 & 0.1940844 & 0.2095983 & 0.1683779 \\ 0.1520738 & 0.2572087 & 0.1159364 & 0.1871348 \\ 0.2736196 & 0.2184046 & 0.2569112 & 0.1955149 \\ 0.1651030 & 0.2219857 & 0.1327118 & 0.2637705 \end{pmatrix}$
175	$\begin{pmatrix} 0.2508827 & 0.2567217 & 0.2769789 & 0.2179744 \\ 0.1990233 & 0.2389428 & 0.1518783 & 0.2463337 \\ 0.3605493 & 0.2851649 & 0.2468397 & 0.2494727 \\ 0.2166389 & 0.2931182 & 0.1725566 & 0.2445651 \end{pmatrix}$	$\begin{pmatrix} 0.2508827 & 0.1810289 & 0.1949597 & 0.1556062 \\ 0.1405978 & 0.2389428 & 0.1081832 & 0.1732793 \\ 0.2516325 & 0.2022469 & 0.2468397 & 0.1789566 \\ 0.1542924 & 0.2064010 & 0.1240249 & 0.2445651 \end{pmatrix}$
200	$\begin{pmatrix} 0.2350053 & 0.2390689 & 0.2588222 & 0.2047194 \\ 0.1862504 & 0.2244578 & 0.1424504 & 0.2309688 \\ 0.3369043 & 0.2698795 & 0.2289835 & 0.2372560 \\ 0.2018249 & 0.2751404 & 0.1615171 & 0.2304976 \end{pmatrix}$	$\begin{pmatrix} 0.2350053 & 0.1687217 & 0.1826623 & 0.1434511 \\ 0.1314895 & 0.2244578 & 0.1011995 & 0.1631052 \\ 0.2369804 & 0.1901092 & 0.2289835 & 0.1705242 \\ 0.1447819 & 0.1931570 & 0.1146795 & 0.2304976 \end{pmatrix}$

Again, let's draw scatterplots to graphically display these tables:

Figure 4-5 Scatterplots of Coverage Rate for 4-state Reversible Markov Chain

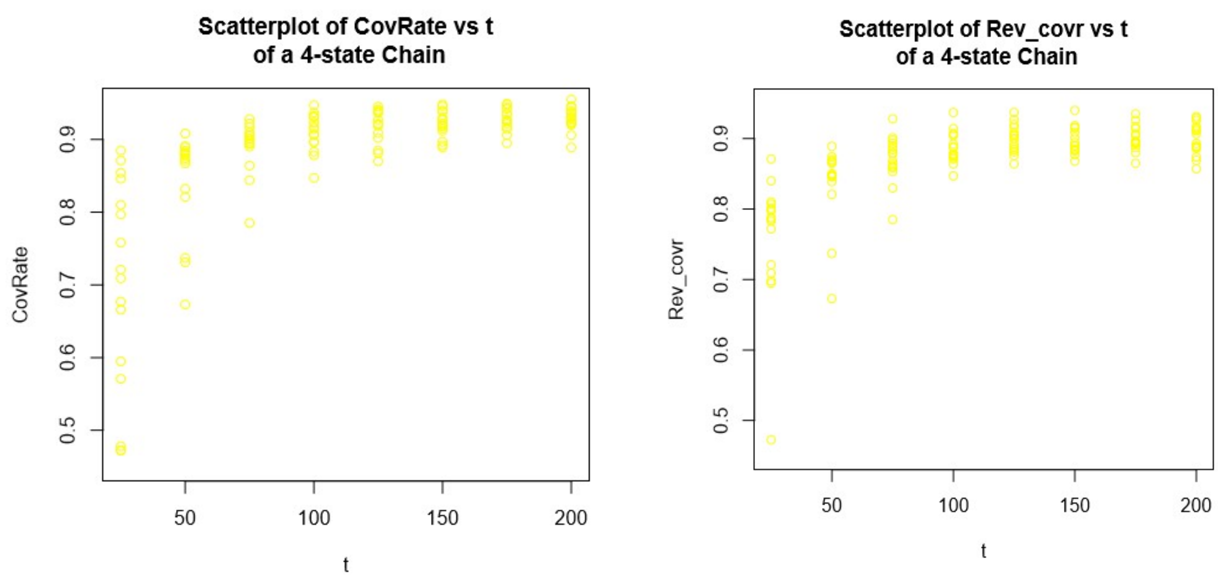
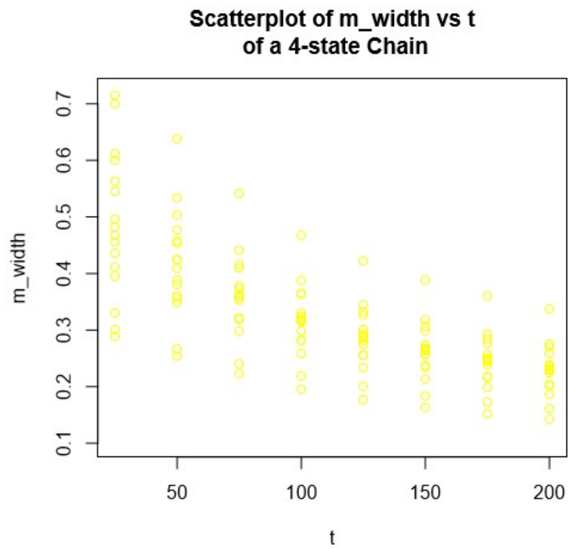
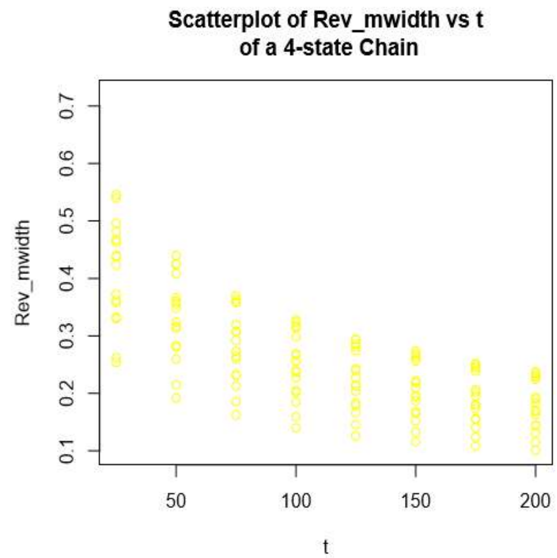


Figure 4-6 Scatterplots of Mean Width for 4-state Chain





From these results we see that with the reversible estimator we obtain slightly smaller mean lengths of 95% C.I.'s. But, unlike the simulation results reported above for a three state chain, the estimated coverage rates for this four state chain of intervals based on the reverse estimators are almost as close to their nominal values as intervals based on the on the mle. One speculative explanation for this conclusion may be that for fixed time t and state i, there are four possible transitions here and only three for a three state chain. Hence, some of the numbers of observed transitions $\{ N_{ij} \}$ between states tend to be less for a four state chain than for a three state chain. Since both estimators improve as the amount of data increases, the reverse procedure may overcome this small data problem in this example by using both N_{ij} and N_{ji} to estimate p_{ij}. This issue warrants further study.

4.4 Simulation of Non-Reversible Markov Chain

To check how the confidence intervals compare when the chain is not reversible, we also simulated observations from a non-reversible matrix 3-state Markov Chain whose transition matrix is given by

$$P_3=\begin{pmatrix} .40 & .35 & .25 \\ .16 & .50 & .34 \\ .23 & .30 & .47 \end{pmatrix}.$$

The stationary distribution of the chain is approximately

$\boldsymbol{\pi}$ = (0.2441948, 0.3902622, 0.3655431).

--As we can simply calculate, $\pi_1 p_{12} = .35 \times .2441948 = .08546818$ while $\pi_2 p_{21} = .16 \times .3902622 = .062441952$, $\pi_1 p_{12} \neq \pi_2 p_{21}$. So by definition, $\mathbf{P_3}$ is not a reversible chain.

Again, independently repeating the four-steps of our simulation algorithm 1000 times and calculating the coverage rates and mean lengths, we obtained:

Table 4-5 Comparing Coverage Rates for Non-Reversible Chain

n	Coverage Rate For $\{\hat{p}_{ij}\}$	Coverage Rate For $\{\hat{p}_{ij}^{(R)}\}$	Proportion of Negative Numerator For $\{\hat{p}_{ij}^{(R)}\}$
25	$\begin{pmatrix} 0.807 & 0.831 & 0.791 \\ 0.806 & 0.888 & 0.889 \\ 0.869 & 0.853 & 0.884 \end{pmatrix}$	$\begin{pmatrix} 0.807 & 0.764 & 0.826 \\ 0.850 & 0.888 & 0.795 \\ 0.782 & 0.820 & 0.884 \end{pmatrix}$	$\begin{pmatrix} 0.000 & 0.013 & 0.001 \\ 0.000 & 0.000 & 0.004 \\ 0.000 & 0.001 & 0.000 \end{pmatrix}$
50	$\begin{pmatrix} 0.918 & 0.901 & 0.910 \\ 0.888 & 0.922 & 0.923 \\ 0.916 & 0.902 & 0.909 \end{pmatrix}$	$\begin{pmatrix} 0.918 & 0.823 & 0.855 \\ 0.883 & 0.922 & 0.795 \\ 0.781 & 0.847 & 0.909 \end{pmatrix}$	
75	$\begin{pmatrix} 0.928 & 0.921 & 0.915 \\ 0.919 & 0.930 & 0.929 \\ 0.927 & 0.935 & 0.928 \end{pmatrix}$	$\begin{pmatrix} 0.928 & 0.789 & 0.863 \\ 0.866 & 0.930 & 0.795 \\ 0.781 & 0.836 & 0.928 \end{pmatrix}$	
100	$\begin{pmatrix} 0.916 & 0.921 & 0.916 \\ 0.914 & 0.936 & 0.945 \\ 0.926 & 0.920 & 0.932 \end{pmatrix}$	$\begin{pmatrix} 0.916 & 0.769 & 0.861 \\ 0.868 & 0.936 & 0.788 \\ 0.798 & 0.845 & 0.932 \end{pmatrix}$	
125	$\begin{pmatrix} 0.923 & 0.931 & 0.929 \\ 0.931 & 0.949 & 0.953 \\ 0.940 & 0.939 & 0.938 \end{pmatrix}$	$\begin{pmatrix} 0.923 & 0.789 & 0.849 \\ 0.866 & 0.949 & 0.820 \\ 0.796 & 0.849 & 0.938 \end{pmatrix}$	
150	$\begin{pmatrix} 0.928 & 0.926 & 0.927 \\ 0.924 & 0.931 & 0.947 \\ 0.946 & 0.938 & 0.936 \end{pmatrix}$	$\begin{pmatrix} 0.928 & 0.769 & 0.845 \\ 0.835 & 0.931 & 0.783 \\ 0.765 & 0.836 & 0.936 \end{pmatrix}$	
175	$\begin{pmatrix} 0.941 & 0.951 & 0.955 \\ 0.926 & 0.932 & 0.936 \\ 0.925 & 0.952 & 0.949 \end{pmatrix}$	$\begin{pmatrix} 0.941 & 0.751 & 0.820 \\ 0.830 & 0.932 & 0.766 \\ 0.744 & 0.818 & 0.949 \end{pmatrix}$	
200	$\begin{pmatrix} 0.941 & 0.937 & 0.929 \\ 0.936 & 0.941 & 0.939 \\ 0.935 & 0.939 & 0.946 \end{pmatrix}$	$\begin{pmatrix} 0.941 & 0.755 & 0.798 \\ 0.797 & 0.941 & 0.787 \\ 0.743 & 0.787 & 0.946 \end{pmatrix}$	

Table 4-6 Comparing Mean Lengths for Non-Reversible Chain

n	Mean Length for $\{\hat{p}_{ij}\}$	Mean Length for $\{\hat{p}_{ij}^{(R)}\}$

25	$\begin{pmatrix} 0.5997813 & 0.6567488 & 0.5698674 \\ 0.3982364 & 0.5837112 & 0.5633525 \\ 0.5037293 & 0.5529773 & 0.5865113 \end{pmatrix}$	$\begin{pmatrix} 0.5997813 & 0.4585395 & 0.4606900 \\ 0.3240842 & 0.5837112 & 0.3859647 \\ 0.3411504 & 0.4026619 & 0.5865113 \end{pmatrix}$
50	$\begin{pmatrix} 0.5088743 & 0.5212417 & 0.4644741 \\ 0.3092303 & 0.4339715 & 0.4156602 \\ 0.3726648 & 0.4035939 & 0.4443248 \end{pmatrix}$	$\begin{pmatrix} 0.5088743 & 0.3529591 & 0.3534758 \\ 0.2372575 & 0.4339715 & 0.2852265 \\ 0.2543471 & 0.2982353 & 0.4443248 \end{pmatrix}$
75	$\begin{pmatrix} 0.4327030 & 0.4320070 & 0.3880221 \\ 0.2574282 & 0.3565660 & 0.3403778 \\ 0.3120188 & 0.3381066 & 0.3674832 \end{pmatrix}$	$\begin{pmatrix} 0.4327030 & 0.2915338 & 0.2930284 \\ 0.1973266 & 0.3565660 & 0.2319027 \\ 0.2073922 & 0.2447498 & 0.3674832 \end{pmatrix}$
100	$\begin{pmatrix} 0.3806798 & 0.3756266 & 0.3375080 \\ 0.2278844 & 0.3100004 & 0.2938761 \\ 0.2697748 & 0.2934244 & 0.3194290 \end{pmatrix}$	$\begin{pmatrix} 0.3806798 & 0.2551343 & 0.2546408 \\ 0.1703660 & 0.3100004 & 0.2040387 \\ 0.1814991 & 0.2121725 & 0.3194290 \end{pmatrix}$
125	$\begin{pmatrix} 0.3429184 & 0.3366396 & 0.3058250 \\ 0.2033516 & 0.2789631 & 0.2641808 \\ 0.2403169 & 0.2633069 & 0.2871648 \end{pmatrix}$	$\begin{pmatrix} 0.3429184 & 0.2292789 & 0.2268233 \\ 0.1552681 & 0.2789631 & 0.1816863 \\ 0.1623354 & 0.1906648 & 0.2871648 \end{pmatrix}$
150	$\begin{pmatrix} 0.3128851 & 0.3080420 & 0.2770469 \\ 0.1864277 & 0.2550182 & 0.2414901 \\ 0.2214468 & 0.2402122 & 0.2615401 \end{pmatrix}$	$\begin{pmatrix} 0.3128851 & 0.2085113 & 0.2095178 \\ 0.1407088 & 0.2550182 & 0.1660126 \\ 0.1479519 & 0.1753484 & 0.2615401 \end{pmatrix}$
175	$\begin{pmatrix} 0.2906618 & 0.2849342 & 0.2582575 \\ 0.1727582 & 0.2354117 & 0.2239843 \\ 0.2047484 & 0.2233467 & 0.2439396 \end{pmatrix}$	$\begin{pmatrix} 0.2906618 & 0.1939851 & 0.1924163 \\ 0.1307894 & 0.2354117 & 0.1539698 \\ 0.1373305 & 0.1619433 & 0.2439396 \end{pmatrix}$
200	$\begin{pmatrix} 0.2708762 & 0.2676725 & 0.2417454 \\ 0.1620023 & 0.2207693 & 0.2093134 \\ 0.1904819 & 0.2089171 & 0.2269583 \end{pmatrix}$	$\begin{pmatrix} 0.2708762 & 0.1821459 & 0.1820099 \\ 0.1217842 & 0.2207693 & 0.1445498 \\ 0.1294596 & 0.1522792 & 0.2269583 \end{pmatrix}$

Figure 4-7 Scatterplots of Coverage Rate for Non-Reversible Chain

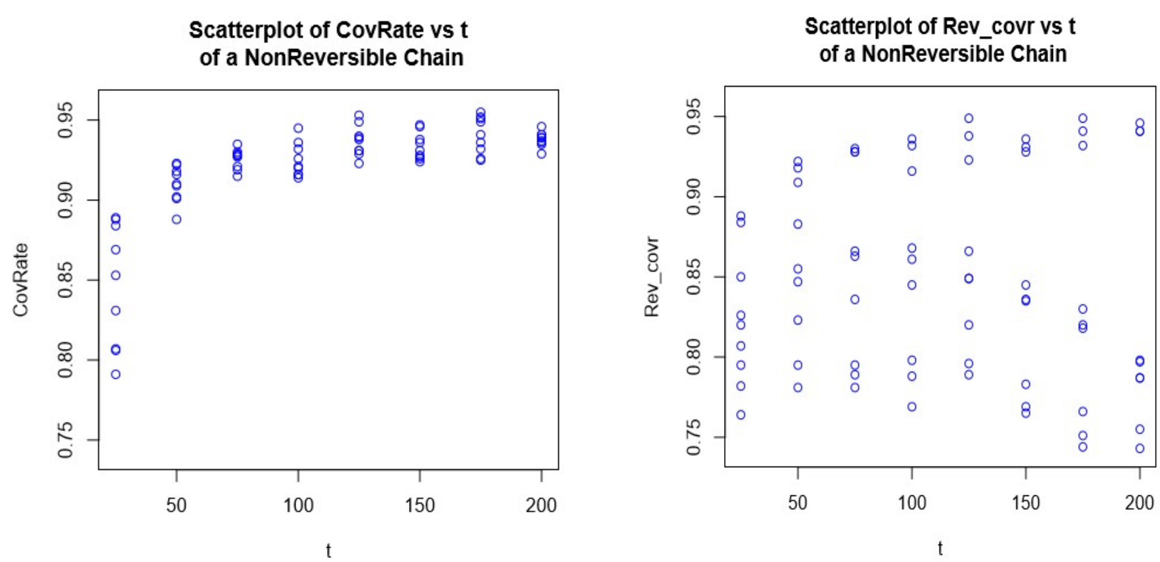
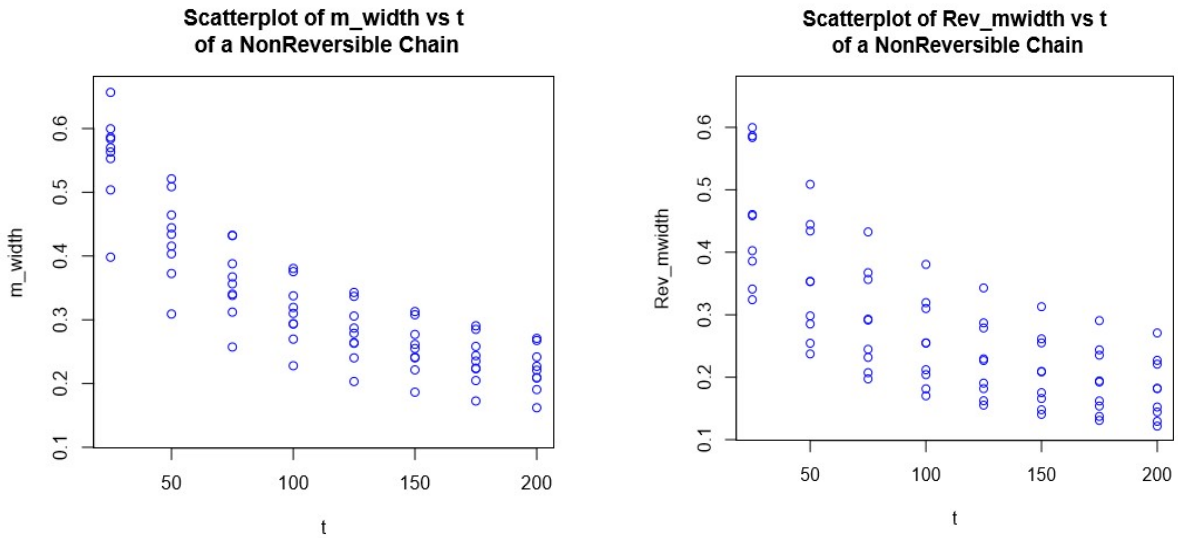


Figure 4-8 Scatterplots of Mean Width for Non-Reversible Chain



From these results we see that for a non-reversible Markov Chain, the reversible estimator performs poorly. It does not provide smaller mean lengths of 95% C.I.'s, while the coverage rates even get worse as the number of observed transitions *t* increases.

Chapter 5 - Conclusion

Based on the observed results from Sections 4.2-4.4, we conclude that when the chain is reversible, the nominal 0.95 confidence intervals constructed by both types of estimators are asymptotically correct and the intervals based on the reversible estimator are asymptotically no wider than those based on the mle. However, our simulations indicate that for the finite sample sizes and models used here, the estimated coverage rates based on the reversible estimator appear to be considerably below their nominal values for the three-state chain used in this study. But for the four-state chain used in the study, the reversible estimator performed more reasonably. This issue warrants further study. Nevertheless, since the coverage rates for the intervals based on the mle appear to approach their nominal levels as sample sizes increase, we recommend using them rather than the more complicated intervals based on the reversible estimator.

References

Annis, D.H., Kiessler, P.C., Lund, R. and Steuber, T.L. (2010). Estimation in reversible Markov chains. *American Statistician*, 64, 116-120.

Bremaud, P. (1998). Markov Chains Gibbs Fields, Monte Carlo Simulation and Queues. Springer-Verlag. New York.

Greenwood, P.E. and Wefelmeyer, W. (1999). Reversible Markov chains and optimality of symmetrized empirical estimators. *Bernoulli*, 5, 109-123.

Greenwood, P.E., Schick, A. and Wefelmeyer, W. (2001). Comments on "Inference for semiparametric models: Some questions and answers," by P.J. Bickel and J. Kwon., Statistica Sinica, 11, 892-906.

Karlin, S. (1966). A First Course in Stochastic Processes. Academic Press, New York.

Jiang, Q. (2009). Construction of Transition Matrices of Reversible Markov Chains. Ph.D. thesis, University of Windsor.

Ross, S. M. (2003). Introduction to Probability Models. Academic Press, San Diego, CA

Appendix A - R codes

[1] Compute the stationary distribution of matrix P

```
• Pi=function(P){
  y=P
  n=ncol(P)
  Q=matrix(NA,ncol=n,byrow=TRUE)
  for (i in 1:25){
  y=y%*%P}
  for (i in 1:n){
  Q[1,i]=y[1,i]}
  return(Q)
  }
```

[2] Generate a trajectory from the stationary distribution of matrix P

```
• sim=function(n,P){
 Q=Pi(P)
 sim<-as.numeric(n+1)
 sim[1]<-sample(1:ncol(P),1,prob=Q)
 for (i in 2:(n+1)) {
      newstate<-sample(1:ncol(P),1,prob=P[sim[i-1],])
      sim[i]<-newstate
   }
 sim
 }
```

[3] Estimate the mle's and reversible mle's

```
• mle= function(P,x){
```

```
    n<-length(x)
    t=ncol(P)
    N=array(NA,t)
    simP<-matrix(NA,t,t,byrow=TRUE)
    for (i in 1:t){
     N[i]<-sum(x[1:n]==i)}
    M=matrix(NA,t,t,byrow=TRUE)
    for (i in 1:t){
     for (j in 1:t){
       if (N[i]==0) {simP[i,j]=0}
       else {M[i,j]<-sum(x[-n]==i & x[-1]==j)
             simP[i,j]=M[i,j]/N[i]}
     }
    }
    return(simP)
}
```

- rmle= function(P,x){

```
    n<-length(x)
    t=ncol(P)
    N=array(NA,t)
    revP<-matrix(NA,t,t,byrow=TRUE)
    for (i in 1:t){
     N[i]<-sum(x[1:n]==i)}
    M=matrix(NA,t,t,byrow=TRUE)
    for (i in 1:t){
     for (j in 1:t){
       M[i,j]<-sum(x[-n]==i & x[-1]==j)
     }}
    for (i in 1:t){
     for (j in 1:t){
```

```
    if (N[i]==0) {revP[i,j]=0}
    else {revP[i,j]=(M[i,j]+M[j,i])/(2*N[i])}
   }}
  return(revP)
}
```

[4] Compute the standard errors

- se= function(P,x){

```
  Q=mle(P,x)
  n<-nrow(Q)
  se=matrix(NA,n,n,byrow=TRUE)
  for (i in 1:n){
   for (j in 1:n){
     se[i,j]=sqrt((Q[i,j]-Q[i,j]^2)/(sum(x==i)))
   }
  }
  return(se)
}
```

- rse= function(i,j,m,P,x){

```
  Q=rmle(P,x)
  n<-nrow(Q)
  for (k in 1:m){
  if (k==1){T[k]<-Q[i,j]}
  else{
  T[k]<-array(Q[i,(1:n)[-i]])%*%(array(Q[(1:n)[-i],(1:n)[-i]],c(n-1,n-1))%^%(k-
2))%*%array(Q[(1:n)[-i],j],c(n-1,1))}
  }
  A=(Q[i,j]-Q[i,j]^2)+(Q[i,j]*sum(T[1:m])*Q[j,i]-Q[i,j]^2)
  if (A<0) {
  rse=se2(P,x)[i,j]
```

```
    }
    else{rse<-sqrt(((Q[i,j]-Q[i,j]^2)+(Q[i,j]*sum(T[1:m])*Q[j,i]-Q[i,j]^2))/(2*(sum(x==i))))}
    return(rse)
}
```

- Rse= function(m,P,x){

```
     n=ncol(P)
     Rse=matrix(NA,n,n,byrow=TRUE)
     for (i in 1:n){
       for (j in 1:n){
        Rse[i,j]=rse(i,j,m,P,x)
       }
     }
     return(Rse)
}
```

[5] Construct 95% C.I.'s for mle and rmle

The Lower Bound for mle:

```
Lm=function(P,x){
  Lm=mle(x)-1.96*se(P,x)
  return(Lm)
  }
```

The Upper Bound for mle:

```
Um=function(P,x){
  Um=mle(x)+1.96*se(P,x)
  return(Um)}
```

The Lower Bound for rmle:

```
 Lr=function(m,P,x){
    Lr=rmle(x)-1.96*Rse(m,P,x)
    return(Lr)
```

```
}
```

The Upper Bound for rmle:

```
Ur=function(m,P,x){
  Ur=rmle(x)+1.96*Rse(m,P,x)
  return(Ur)
}
```

[6] 1000 times Simulations

```
rCWNeg=function(t,n,m,P){
  s=ncol(P)
  Q1=matrix(NA,s,s,byrow=TRUE)
  Q2=matrix(NA,s,s,byrow=TRUE)
  R=matrix(NA,s,s,byrow=TRUE)
  for (i in 1:s){
   for (j in 1:s){
     a=array(NA,m)
     b=array(NA,m)
     c=array(NA,m)
     for (k in 1:m){
       y=sim(n,P)
       a[k]=mean(P[i,j]>=Lr2(t,P,y)[i,j] & P[i,j]<=Ur2(t,P,y)[i,j])
       b[k]=Dr2(t,P,y)[i,j]
       c[k]=Nrse(i,j,t,P,y)
      }
     Q1[i,j]=mean(a)
     Q2[i,j]=mean(b)
     R[i,j]=mean(c)
    }
   }
  return(list(v1=Q1,v2=Q2,v3=R))}
```