University of Tartu
Faculty of Science and Technology
Institute of Mathematics and Statistics

Zhen Xu

Parameter Estimation in Progressive Multistate Models

Actuarial and Financial Engineering

Master's Thesis (30 ECTS)

Supervisor: Prof. Jüri Lember

Tartu 2025

PARAMETER ESTIMATION IN PROGRESSIVE MULTISTATE MODELS

Master thesis

Zhen Xu

Abstract

Multistate models are widely used to analyze systems or entities that transition between different states over time. These models have applications in various fields, including survival analysis, reliability engineering, and financial mathematics. This thesis will focus on parameter estimation for time-homogeneous progressive Markov chains, where transitions are limited to higher-numbered states, and the final state is absorbing. The thesis can be divided into two parts, the first part focusing on normal multistate models and the other part focusing on hidden multistate models.

In the first part, we study progressive, time-homogeneous multistate Markov chains. During this part, maximum likelihood estimators (MLEs) for transition probabilities are derived for uncensored, fixed-censored, and random-censored scenarios. Consistency proofs for these estimators are provided using theoretical tools such as the Strong Law of Large Numbers and the Continuous Mapping Theorem.

The second part focuses on hidden multistate models, specifically the two-state hidden Markov chains. In this case, the underlying states are not fully observable: In each observation, each position of the chain has a constant probability of being observed. We study the block structure of the observations and derive statistical properties of block lengths under different censoring mechanisms. Here we only have one transition probability to estimate, moment-based estimators and maximum likelihood estimators are developed for the transition probability, and their consistency is validated through simulations.

Simulation experiments are also conducted for both parts to evaluate the performance of the derived estimators under various parameter configurations and censoring mechanisms. The results confirm the consistency of all proposed estimators.
CERCS research specialisation: P160 Statistics, operations research, programming, financial and actuarial mathematics.
Key Words: Progressive Markov chains, hidden Markov chains, parameter estimation, maximum likelihood estimation, method of moments estimators, consistency, simulation studies

PARAMEETRITE HINDAMINE ÜHESUUNALISTES MUDELITES

Magistritöö

Zhen Xu

Lühikokkuvõte

Magistritöös vaadeldud ühesuunalisi Markovi mudeleid kasutatakse väga erinevates valdkondades – medistsiinistatistikas (elukestvusanalüüs), signaalitöötluses ning ka finantsmatemaatikas. Käesolev magisritöö keskendub mudeli parameetrite hindamisele. Töö koosneb kahest osasat, esimeses käsitletakse üldisi ühesuunalisi mudeleid, teine osa fokuseerub ühele lihtsale kuid osaliselt varjatud mudelile.

Magistritöö esimeses osas leitakse üldise (mittevarjatud) ühesuunalise Markovi ahela parameetrite suurima tõepära hinnangud nii tsenseerimata kui ka tsenseeritud andmete korral, samuti tõestatakse hinnangute mõjusus. Selleks kasutataks klassikalisi tõenäosusteooria tööriistu nagu suurte arvude seadused ning geomeetrilise jaotuse omadused. Töö teises osas käsitletakse lihtsat kahesisundilist ühesuunalist Markovi mudelit, mis aga on osaliselt varjatud – iga ajahetkel on positiivne tõenäosus sisundit mitte registreerida. Seega on käsitletav mudel väga spetsiifiline varjatud Markovi ahel. Mudeli spetsiifikast lähtudes jagunevad sel moel saadud vaatlused kaheks plokiks, iga ploki jaotus ja keskväärtus on töö kesksed uurimisobjektid. Sarnaselt töö esimese osaga käsitletakse ka teises osas nii tsenseeritud kui ka tsenseerimata vaatlusi, hinnatav parameeter on üks üleminekutõenäosus. Töös leitakse parameetrile hinnangud nii täieliku kui osalise informatsiooni põhjal ning tõestatakse nende hinnangute mõjusus.

Toetamaks nii esimese kui teise osa teooriat on töös läbi viidud simulatsioonid, mis kinnitavad teooria paikapidavust. Muuhulgas näitavad simulatsioonid, et saadud hinnangud on tõepoolest möjusad.
CERCS teaduseriala: P160 Statistika, operatsioonianalüüs, programmeerimine, finants- ja kindlustusmatemaatika.
Märksõnad: Ühesuunalised Markovi mudelid, varjatud Markovi mudelid, parameetrite hindamine, momentide meetod, suurima tõepära hinnang, mõjusus, simulatsioonid.

Contents

Introduction

Multistate models are mathematical tools used to describe the transitions of entities between discrete states over time. These models have wide applications in fields such as survival analysis, reliability engineering, and financial mathematics. This thesis focuses on two types of progressive multistate models: normal multistate models, where the states are fully observable, and progressive hidden multistate models, where the underlying states are only partially observed. In these models, transitions are restricted to higher-numbered states, and the final state is absorbing.

The first part of this thesis deals with progressive, time-homogeneous multistate Markov models. A key challenge in parameter estimation arises due to censoring, where observations may be incomplete because of fixed or random stopping times. This part derives general maximum likelihood estimator (MLE) for transition probabilities under various censoring scenarios and provides rigorous proofs of the consistency.

The second part of this thesis focuses on two-state progressive Hidden Markov Chains (HMC). In these models, there is only one transition probability needs to be estimated, and the underlying states are not directly observable; instead, each state has a constant probability of being observed. This part derives the statistical properties of block structures in the observations, such as sequences of ones and zeros, and provides moment-based and maximum likelihood estimators for the transition probability. Consistency is validated through simulation experiments, and the influence of censoring and noise levels on the estimators is analyzed.

The primary contributions of this thesis are:

- Derivation of MLEs for normal progressive multistate models under uncensored, fixed-censored, and random-censored scenarios, along with consistency proofs.
- Derivation of estimators for the two-state progressive HMC, using observed

block structures and statistical properties of block lengths, including MLEs and moment estimators.

- Validation of the proposed estimators through simulation studies, highlighting their consistency under various parameter configurations and censoring mechanisms.

1 Background

1.1 Geometric Distributions

The geometric distribution is a fundamental discrete probability distribution that models the number of trials needed to achieve the first success in a sequence of independent Bernoulli trials. In this thesis, we encounter two common parameterizations of this distribution, which we distinguish as follows:

1.1.1 Two Definitions

Definition (Geometric Distribution G_1)**.** *A random variable* $X \sim G_1(p)$ *follows the geometric distribution with parameter* $p \in (0,1)$ *if:*

$$P(X = k) = (1-p)^{k-1}p, \quad \textit{for } k = 1, 2, 3, \ldots$$

This distribution represents the number of trials until the first success, including the success trial.

Definition (Geometric Distribution G_0)**.** *A random variable* $Y \sim G_0(p)$ *follows the geometric distribution with parameter* $p \in (0,1)$ *if:*

$$P(Y = k) = (1-p)^{k}p, \quad \textit{for } k = 0, 1, 2, \ldots$$

This distribution represents the number of failures before the first success.

The relationship between these two distributions is straightforward: if $X \sim G_1(p)$, then $X - 1 \sim G_0(p)$, and conversely, if $Y \sim G_0(p)$, then $Y + 1 \sim G_1(p)$.

1.1.2 Expectations

If $X \sim G_1(p)$, then $E(X) = \frac{1}{p}$.

Proof.

$$\begin{aligned} E(X) &= \sum_{k=1}^{\infty} k \cdot P(X = k) \\ &= \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} p \\ &= p \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} \end{aligned}$$

Let $S = \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1}$. We can compute this sum using a well-known trick:

$$\begin{aligned} (1-p) \cdot S &= (1-p) \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} \\ &= \sum_{k=1}^{\infty} k \cdot (1-p)^{k} \end{aligned}$$

Since:

$$\begin{aligned} S &= \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} \\ &= \sum_{k=0}^{\infty} (k+1) \cdot (1-p)^{k} \\ &= 1 + \sum_{k=1}^{\infty} (k+1) \cdot (1-p)^{k} \end{aligned}$$

Thus

$$\begin{aligned}
E(X) = pS = S - (1-p)S \\
&= 1 + \sum_{k=1}^{\infty}(k+1)\cdot(1-p)^k - \sum_{k=1}^{\infty} k\cdot(1-p)^k \\
&= 1 + \sum_{k=1}^{\infty}(1-p)^k \\
&= 1 + \frac{(1-p)^1 - (1-p)^\infty}{1-(1-p)} \\
&= 1 + \frac{1-p}{p} \\
&= \frac{1}{p}
\end{aligned}$$

□

If $Y \sim G_0(p)$, then $E(Y) = E(X) - 1 = \frac{1-p}{p}$.

Throughout this thesis, we will consistently use G_0 to denote the geometric distribution starting at 0 and G_1 for the distribution starting at 1. This distinction is crucial for the correct formulation and analysis of the partially hidden Markov models that form the core of our study.

1.2 Markov Chains

A stochastic process $\{X_t\}_{t\geq 1}$ taking values in a finite state space $S = \{1, ..., r+1\}$ is called a Markov chain (Learn more from Norris, 1998) if it satisfies the Markov property: the future state depends only on the present state, not on the sequence of states that preceded it. Mathematically, for all states $i, j \in S$ and all $t \geq 1$:

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

In this thesis, we focus on time-homogeneous Markov chains, where transition probabilities remain constant over time. These probabilities are denoted as:

$$p_{ij} = P(X_{t+1} = j | X_t = i)$$

1.3 Progressive Multistate Models

Markov chains provide a fundamental mathematical framework for modeling stochastic processes with state transitions. In a Markov chain, the future state depends solely on the current state, independent of the past trajectory. Multistate models extend this concept by focusing on how entities move between a finite set of states over time.

Progressive multistate models represent a specialized class of Markov chains with additional structural constraints. While general multistate models allow transitions between any states, progressive multistate models restrict transitions to only higher-numbered states, creating a unidirectional progression. The transition matrix for such models has an upper triangular structure, with:

$$p_{ij} = 0 \quad \text{for } j < i$$

The final state r+1 is absorbing, meaning once entered, it cannot be left:

$$p_{r+1,r+1} = 1$$

1.4 Censoring Mechanisms

In practice, complete observation of the process until absorption is often impossible. We consider three types of censoring:

- No censoring: The process is observed until absorption
- Fixed censoring: Observation stops at a predetermined time m
- Random censoring: Observation stops at a random time M independent of the process

Note that fixed censoring is a special case of random censoring.

1.5 Maximum Likelihood Estimation

Maximum Likelihood Estimation (MLE) is a widely used method for estimating the parameters of statistical models (Framework established by Fisher, 1922). The central idea is to determine the parameter values that maximize the likelihood function, which represents the probability of observing the given data under the model.

Definition of the Likelihood Function

Let $\boldsymbol{\theta}$ represent the vector of parameters to be estimated in the model. Suppose we observe data $\mathcal{D} = \{x_1, x_2, \ldots, x_n\}$, where x_i represents the observed transitions or outputs for the i-th entity. The likelihood function is defined as:

$$L(\boldsymbol{\theta}; \mathcal{D}) = P(\mathcal{D} \mid \boldsymbol{\theta}),$$

which expresses the probability of observing the data $\mathcal{D}$ given the parameter vector $\boldsymbol{\theta}$. In this thesis, for simplicity we sometimes just use notation L when context is clear. In practice, it is often more convenient to work with the log-likelihood function:

$$\ell(\boldsymbol{\theta}; \mathcal{D}) = \log L(\boldsymbol{\theta}; \mathcal{D}).$$

Maximizing $\ell(\boldsymbol{\theta}; \mathcal{D})$ with respect to $\boldsymbol{\theta}$ yields the Maximum Likelihood Estimator:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}; \mathcal{D}).$$

1.6 Stopping Times

A key theoretical tool in our analysis is the concept of stopping times, particularly first hitting times (Learn more from Capasso and Bakstein, 2012):

$$T_i = min\{t \geq 1 : X_t = i\}$$

These help decompose the complex problem of proving consistency into more manageable pieces by conditioning on the first time each state is reached.

1.7 Strong Law of Large Numbers

The consistency proofs rely heavily on the Strong Law of Large Numbers (SLLN) (Learn more from Billingsley, 2012). For a sequence of independent, identically distributed random variables Y_n with $E|Y_1| < \infty$:

$$\frac{1}{n}\sum_{k=1}^{n} Y_k \xrightarrow{a.s.} E[Y_1]$$

It indicates that through enough trials, the average outcome of all observations of a random variable will be the expectation.

This theorem provides the theoretical foundation for proving that our estimators converge almost surely to the true parameters.

1.8 Continuous Mapping Theorem

The Continuous Mapping Theorem (CMT) is a foundational result in probability theory that ensures the preservation of convergence under continuous transformations (Also introduced by Pollard, 1984). It plays a vital role in proving the consistency of estimators derived from convergent sequences of random variables.

Let Z_n be a sequence of random variables and Z a random variable such that $Z_n \xrightarrow{a.s.} Z$. Let $g : \mathbb{R}^d \to \mathbb{R}^k$ be a function that is continuous almost everywhere on the support of Z. Then:

$$g(Z_n) \xrightarrow{a.s.} g(Z)$$

2 Maximum Likelihood Estimation for Normal Multistate Models

2.1 Model Setup

2.1.1 Model Setup for a Single Realization

Let us first consider a single realization of a progressive Markov chain $\{X_t\}_{t\geq 1}$, with state space $S = \{1, ..., r+1\}$ and transition matrix:

$$\begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} & \cdots & p_{1(r+1)} \\ 0 & p_{22} & p_{23} & p_{24} & \cdots & p_{2(r+1)} \\ 0 & 0 & p_{33} & p_{34} & \cdots & p_{3(r+1)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & p_{rr} & p_{r(r+1)} \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

where $p_{ij} \geq 0$ and $\sum_{j=i}^{r+1} p_{ij} = 1$ for each $i = 1, ..., r$.

The realization starts with $1(P(X_1 = 1) = 1)$, and is observed until a censoring time M, which may be:

- Constant: $M = m$
- Random(Independent Censoring): M is a random variable that is independent of $\{X_t\}$

Note that constant censoring is just a special case of independent censoring with distribution:

$$P(M = k) = \begin{cases} 1 & k = m \\ 0 & k \neq m \end{cases}$$

We also consider ideally uncensored case: $M = \infty$. In this case, a transition from a non-absorption state to another state will happen for sure, the realization will end up with state $r + 1$. Note that with censoring or not, a transition from each state to another state can happen at most once(progressive property).

Obviously M is a positive discrete random variable(taking values only of integers), denote $q_m = P(M = m)$, note that for constant censoring m we have $P(M = m) = 1$, for non-censoring we have $P(M = \infty) = 1$, so for these two special cases we take $q_M = 1$ during likelihood derivations.

Let us define the key counting variables. For each state i, let:

$$
\begin{aligned}
&I_i(X_t) \text{ denote indicator function :} \\
&I_i(X_t) = \begin{cases} 0 & X_t \neq i \\ 1 & X_t = i \end{cases} \\
K_i &:= \sum_{t=1}^{\infty} I_i(X_t) \quad \text{(total occurrences of state i)} \\
K_i^M &:= \sum_{t=1}^{M} I_i(X_t) \quad \text{(occurrences of state i up to time M)} \\
I_{ij} &:= \sum_{t=1}^{\infty} I_i(X_t) I_j(X_{t+1}) \quad \text{(total transitions from i to j)} \\
I_{ij}^M &:= \sum_{t=1}^{M-1} I_i(X_t) I_j(X_{t+1}) \quad \text{(transitions from i to j up to time M)}
\end{aligned}
$$

Note that for indicator functions, $E[I_i(X_t)] = P(X_t = i)$. There will also be another way of defining an indicator function:

$$
I(X \in \mathcal{X}) = \begin{cases} 0 & X \notin \mathcal{X} \\ 1 & X \in \mathcal{X} \end{cases}
$$

2.1.2 The Expectation of K_1^m

Let X be a discrete random variable with distribution $G_1(1-p)$, i.e.:

$$P(X=k) = (1-p)p^{k-1}, \qquad k > 0$$

Denote $X^m = \min\{X, m\}$, then we have:

$$P(X^m = k) = \begin{cases} p^{k-1}(1-p) & k < m; \\ \sum_{k=m}^{\infty} p^{k-1}(1-p) = p^{m-1} & k = m. \end{cases}$$

Claim 2.1. $E(X^m) = \frac{1-p^m}{1-p}$

Proof. First, we have:

$$EX^m = \sum_{k=1}^{m-1} kp^{k-1}(1-p) + mp^{m-1}$$

Denote:

$$S := \sum_{k=1}^{m-1} kp^{k-1} = 1 + \sum_{k=2}^{m-1} kp^{k-1}$$

Thus, we have:

$$\begin{aligned} pS &= \sum_{k=1}^{m-1} kp^k \\ &= \sum_{k=2}^{m} (k-1)p^{k-1} \\ &= \sum_{k=2}^{m-1} (k-1)p^{k-1} + (m-1)p^{m-1} \end{aligned}$$

and:

$$\begin{aligned} S - pS &= 1 + \sum_{k=2}^{m-1} p^{k-1} - (m-1)p^{m-1} \\ &= \sum_{k=1}^{m} p^{k-1} - mp^{m-1} \\ &= \frac{1-p^m}{1-p} - mp^{m-1} \end{aligned}$$

So:

$$EX^m = (1-p)S + -mp^{m-1} = \frac{1-p^m}{1-p}$$

□

Obviously $K_1 \sim G_1(1-p_{11})$ and $K_1^M = \min\{K_1, M\}$.

2.1.3 Observed Data from Multiple Realizations

Now consider n independent realizations of a progressive Markov chain $\{X_t^k\}_{t\geq 1}$, k=1,2,...,n, with state space $S = \{1, ..., r+1\}$ and the same transition matrix.

Each observation k always starts with 1, and is observed until a censoring time M_k, which may be:

- Constant: $M_k = m$
- Random(Independent Censoring): All M_k are independent and identically distributed (i.i.d.) random variables that are also independent of all $\{X_t^k\}$

2.2 Maximum Likelihood Estimator

Our objective is to estimate the transition probabilities $\{p_{ij}\}$ of the progressive Markov chain using maximum likelihood estimation. To derive the MLE(Maximum

Likelihood Estimator), we need to find the p_{ij} that maximize the likelihood function. For one or several observations, likelihood function means the probability(in discrete cases) of obtaining such observations.

Let n_{ij} denote the number of observed transitions from state i to state j over all n observations:

$$n_{ij} = \sum_{k=1}^{n} \sum_{t=1}^{M_k - 1} I_i(X_t) I_j(X_{t+1}) = \sum_{k=1}^{n} I_{ij}^{M_k}$$

If $M_k = \infty$(uncensored),

$$n_{ij} = \sum_{k=1}^{n} I_{ij}$$

All the observations and censorings are independent of each other, and in each observation each transition is also independent. In a certain observation k, transition from i to j has the probability of p_{ij} and occurrence of $I_{ij}^{M_k}$. So for observation k, the likelihood function should be:

$$q_{M_k} \prod_{i=1}^{r} \prod_{j=i}^{r+1} p_{ij}^{I_{ij}^{M_k}}$$

Thus the likelihood function for all n observations is:

$$L = \prod_{k=1}^{n} \left(q_{M_k} \prod_{i=1}^{r} \prod_{j=i}^{r+1} p_{ij}^{I_{ij}^{M_k}} \right) = \prod_{k=1}^{n} q_{M_k} \prod_{i=1}^{r} \prod_{j=i}^{r+1} p_{ij}^{n_{ij}}$$

Taking the logarithm:

$$\ell_n(p) = \log L = \sum_{k=1}^{n} \log q_{M_k} + \sum_{i=1}^{r} \sum_{j=i}^{r+1} n_{ij} \log p_{ij}$$

Note that $\sum_{k=1}^{n} \log q_{M_k}$ is constant with respect to p.

To find the MLE, we maximize this subject to the constraints:

$$\sum_{j=i}^{r+1} p_{ij} = 1 \text{ for each } i = 1, ..., r$$
$$p_{ij} \geq 0, \quad j \geq i$$
$$p_{ij} = 0, \quad j < i$$

Using Lagrange multipliers (Shown by Ito and Kunisch, 2008) λ_i for each constraint:

$$\mathcal{L} = \sum_{k=1}^{n} \log q_{M_k} + \sum_{i=1}^{r}\sum_{j=i}^{r+1} n_{ij} \log p_{ij} - \sum_{i=1}^{r} \lambda_i (\sum_{j=i}^{r+1} p_{ij} - 1)$$

Taking partial derivatives and setting to zero:

$$\frac{\partial \mathcal{L}}{\partial p_{ij}} = \frac{n_{ij}}{p_{ij}} - \lambda_i = 0 \quad \text{ for each } i = 1, ..., r, \forall j \geq i$$

This means for each fixed i, $\frac{n_{ij}}{p_{ij}}$ is constant for every possible j.

So for specific $i, j : j \geq i$:

$$p_{ik} = \frac{n_{ik}}{n_{ij}} p_{ij} \text{ for each } k = i, ..., r+1. \text{ Thus :}$$
$$\sum_{k=i}^{r+1} p_{ik} = \sum_{k=i}^{r+1} \frac{n_{ik}}{n_{ij}} p_{ij} = 1$$

which yields:

$$\hat{p}_{ij,n} = \frac{n_{ij}}{\sum_{k=i}^{r+1} n_{ik}} \quad \text{ for each } i = 1, ..., r, \forall j \geq i$$

The MLE is defined only when the denominator $\sum_{k=i}^{r+1} n_{ik}$ is non-zero. When no transitions from state i are observed (which can happen when i exceeds the maximum censoring time), the MLE is undefined.

This is a very intuitive result, also the most reasonable estimator one can come up with obtained information. The denominator is the occurrences of all transitions start with i during all n observations. And it also shows that the MLE remains the same for constant, independent and no censoring.

NB! The likelihood function above does not hold when M depends on realizations, which means M_k is determined by each observation k. Take a two-state model for example(In this case, the number of observed 1-s(K_1^M) contains all the information):

We censor the realization always before last two 1-s, which means, when we observe 111($K_1^M = 3$), it means there are 2 another 1-s before it enters state 2(i.e. $K_1 = 5$, the actual sequence is 1111122...). So in a certain observation:

$$P(K_1^M = 0) = P(K_1 = 1) + P(K_1 = 2) = p_{12}(1 + p_{11})$$
$$P(K_1^M = l, l \in N^+) = P(K_1 = l + 2) = p_{12}p_{11}^{l+2}$$

So for each observation, the likelihood function is a pairwise function, the overall likelihood function will be much more complicated, let alone we have unlimited ways to do dependent censoring.

2.3 Consistency of Estimators

In statistical inference, consistency (Also introduced by Fisher, 1922) represents a fundamental asymptotic property that characterizes the behavior of an estimator as the sample size increases indefinitely. Formally, an estimator $\hat{\theta}_n$ is consistent for parameter θ if it converges to the true parameter value almost surely as $n \to \infty$:

$$\hat{\theta}_n \xrightarrow{a.s.} \theta \quad \text{as} \quad n \to \infty$$

where $\xrightarrow{a.s.}$ denotes almost sure convergence, meaning that the probability of the estimator converging to the true parameter equals one, i.e. $\lim_{n\to\infty} P(\hat{\theta}_n = \theta) = 1$. Note that an estimator is actually a random variable that usually depends on n.

Important Statistical Properties:

The convergence of consistent estimators is characterized by several key properties:

- **Asymptotic Unbiasedness**: While an estimator may be biased for finite samples, consistency ensures that $\lim_{n\to\infty} E[\hat{\theta}_n] = \theta$, provided the expectation exists.
- **Variance Reduction**: For a consistent estimator, the variance typically decreases to 0 as $n \to \infty$. This variance reduction is crucial in ensuring that the estimator concentrates increasingly around the true parameter value.

In what follows, we rigorously establish the consistency of our MLE under independent censoring, demonstrating that our estimation approach remains valid asymptotically even when observations are incompletely observed.

2.4 Consistency Under Independent Censoring

To prove the consistency, we need to show that $\hat{p}_{ij,n} \xrightarrow{a.s.} p_{ij}$ as $n \to \infty$.

Lemma 2.1 (Ratio of Conditional Expectations)**.** *Let X, Y, M be discrete random variables such that for every possible value m of M,*

$$\frac{E[X|M=m]}{E[Y|M=m]} = p$$

where we define $0/0 = p$. Then

$$\frac{E[X]}{E[Y]} = p$$

Proof. Let M take values in some countable set S. Then

$$\begin{aligned} E[X] = E[E[X|M]] &= \sum_{m \in S} E[X|M = m]P(M = m) \\ &= \sum_{m \in S} pE[Y|M = m]P(M = m) \\ &= p \sum_{m \in S} E[Y|M = m]P(M = m) \\ &= pE[Y] \end{aligned}$$

Therefore,

$$\frac{E[X]}{E[Y]} = p$$

□

This lemma is crucial for our consistency proof as it allows us to move from conditional expectations to unconditional expectations while preserving their ratio. It is particularly useful when dealing with censored data where we condition on the censoring time M.

Lemma 2.2. *For independent censoring M, the following equations hold:*

$$\frac{E(K_i^M) - P(K_i^M > 0)}{E(K_i^M) - P(K_i^M > 0) + P(\sum_{j>i} I_{ij}^M = 1)} = p_{ii} \tag{2.4.1}$$

$$\frac{E(I_{ij}^M)}{E(K_i^M) - P(K_i^M > 0) + P(\sum_{j>i} I_{ij}^M = 1)} = p_{ij} \tag{2.4.2}$$

And in the uncensored case:

$$\frac{E(K_i) - P(K_i > 0)}{E(K_i)} = p_{ii} \tag{2.4.3a}$$

$$\frac{E(I_{ij})}{E(K_i)} = p_{ij} \tag{2.4.3b}$$

Here in I_{ij}^M and I_{ij}, $j > i$.

Proof. Let us first prove these equations hold for fixed censoring time m. Define the stopping time:

$$T_i := \min\{t \geq 1 : X_t = i\}$$

T_i refers to the first time state i is observed, obviously $T_1 = 1$.

First, let's prove that for any $l = 1, ..., m$, we have:

$$E[K_i^m | T_i = l] = \frac{1 - p_{ii}^{m-l+1}}{1 - p_{ii}}$$

$$P[\sum_{j>i} I_{ij}^m = 1 | T_i = l] = 1 - p_{ii}^{m-l}$$

$$E[I_{ij}^m | T_i = l] = \frac{p_{ij}}{1 - p_{ii}}(1 - p_{ii}^{m-l})$$

Recall that the random variables involved are under single observation.

For the first one, $E[K_i^m | T_i = l]$ means the expected occurrences of state i under constant censoring time m while at time l X_t enters state i for the first time. It's equivalent to $E[K_i^{m-l+1} | X_1 = i]$(Cut away the first $l - 1$ time points, consider it starts with state i), so the distribution of this random variable should be:

$$P(K_i^{m-l+1} = k | X_1 = i) = \begin{cases} p_{ii}^{k-1}(1 - p_{ii}) & k < m - l + 1; \\ 1 - \sum_{k=1}^{m-l} p_{ii}^{k-1}(1 - p_{ii}) & k = m - l + 1. \end{cases}$$

By claim 2.1 we have: $E[K_i^m | T_i = l] = \frac{1 - p_{ii}^{m-l+1}}{1 - p_{ii}}$.

For the second one, $P(\sum_{j>i} I_{ij}^m = 1 | T_i = l)$ means the probability of that, the number of transitions from state i to any other state under constant censoring time m is equal to 1, while at time l X_t enters state i for the first time. Similarly we can also cut away the first $l - 1$ time points, it will be $P(\sum_{j>i} I_{ij}^{m-l+1} = 1 | X_1 = i)$. We explained earlier that the sum can take only two values: 0 and 1, when it takes 0, then all $m - l$ transitions are from i to i.

Thus:

$$\begin{aligned}
P[\sum_{j>i} I_{ij}^m = 1|T_i = l] &= 1 - P[\sum_{j>i} I_{ij}^m = 0|T_i = l] \\
&= 1 - P[\sum_{j>i} I_{ij}^{m-l+1} = 0|X_1 = i] \\
&= 1 - p_{ii}^{m-l}
\end{aligned}$$

For the last one, similarly $E[I_{ij}^m|T_i = l]$ is equivalent to $E[I_{ij}^{m-l+1}|X_1 = i]$, which means the expected number of transitions from i to $j(j > i)$under constant censoring time $m - l + 1$ with $X_1 = i$. It can take two values: 0 and 1, note that we have $m - l$ places to choose from:

$$\begin{aligned}
E(I_{ij}^m|T_i = l) &= E(I_{ij}^{m-l+1}|X_1 = i) \\
&= 1 \cdot P(I_{ij}^{m-l+1} = 1|X_1 = i) + 0 \cdot P(I_{ij}^{m-l+1} = 0|X_1 = i) \\
&= P(I_{ij}^{m-l+1} = 1|X_1 = i) \\
&= \sum_{k=1}^{m-l} p_{ij} p_{ii}^{k-1} \\
&= p_{ij} \frac{1 - p_{ii}^{m-l}}{1 - p_{ii}}
\end{aligned}$$

Note that when $T_i = 0$ (which means state i is never reached), we have $K_i^m = 0$, $P(K_i^M > 0) = 0$, and $\sum_{j>i} I_{ij}^M = 0$ by definition, as no transitions involving state i can occur if the state is never reached, thus all these 3 quantities will be zero.

Using these conditional expectations:

Obviously $P[K_i^m > 0|T_i = l] = 1, l > 0$, For (2.4.1):

$$\begin{aligned}
&\frac{E[K_i^m|T_i=l]-P[K_i^m>0|T_i=l]}{E[K_i^m|T_i=l]-P[K_i^m>0|T_i=l]+P[\sum_{j>i}I_{ij}^m=1|T_i=l]}\\
&=\frac{\frac{1-p_{ii}^{m-l+1}}{1-p_{ii}}-1}{\frac{1-p_{ii}^{m-l+1}}{1-p_{ii}}-1+1-p_{ii}^{m-l}}\\
&=\frac{1-p_{ii}^{m-l+1}-1+p_{ii}}{1-p_{ii}^{m-l+1}-p_{ii}^{m-l}(1-p_{ii})}\\
&=\frac{p_{ii}(1-p_{ii}^{m-l})}{1-p_{ii}^{m-l}}\\
&=p_{ii}
\end{aligned}$$

For (2.4.2):

$$\begin{aligned}
&\frac{E[I_{ij}^m|T_i=l]}{E[K_i^m|T_i=l]-P[K_i^m>0|T_i=l]+P[\sum_{j>i}I_{ij}^m=1|T_i=l]}\\
&=\frac{\frac{p_{ij}}{1-p_{ii}}(1-p_{ii}^{m-l})}{\frac{1-p_{ii}^{m-l+1}}{1-p_{ii}}-1+1-p_{ii}^{m-l}}\\
&=\frac{p_{ij}(1-p_{ii}^{m-l})}{1-p_{ii}^{m-l}}\\
&=p_{ij}
\end{aligned}$$

According to lemma 2.1, conditioning on T we have:

$$\frac{E(K_i^m)-P(K_i^m>0)}{E(K_i^m)-P(K_i^m>0)+P(\sum_{j>i}I_{ij}^m=1)}=p_{ii}$$
$$\frac{E(I_{ij}^m)}{E(K_i^m)-P(K_i^m>0)+P(\sum_{j>i}I_{ij}^m=1)}=p_{ij}$$

For independent censoring M, we can apply the lemma 2.1 again by conditioning on M to obtain equations (2.4.1) and (2.4.2). It also applies to dependent censoring, since we are interested in proving the consistency of MLE, so we will focus on the independent censoring.

The uncensored case (2.4.3) follows as a special case by taking $m \to \infty$:

For (2.4.3a):

$$\frac{E[K_i|T_i = l] - P[K_i > 0|T_i = l]}{E[K_i|T_i = l]} = \frac{\frac{1}{1-p_{ii}} - 1}{\frac{1}{1-p_{ii}}}$$

$$= 1 - 1 + p_{ii}$$

$$= p_{ii}$$

For (2.4.3b):

$$\frac{E[I_{ij}|T_i = l]}{E[K_i|T_i = l]} = \frac{\frac{p_{ij}}{1-p_{ii}}}{\frac{1}{1-p_{ii}}}$$

$$= p_{ij}$$

Applying lemma 2.1 with conditioning on T we have the equations (2.4.3). □

Note that constant censoring is included in the independent censoring.

Next, we need to rewrite $\hat{p}_{ij,n}$ in a random variable form using K, K^M, I_{ij} and I_{ij}^M, as we pointed out before, estimators are actually random variables depending on n, and during practical estimations, we use their realizations. Consistency makes sure that as sample size n grows, the estimator's variance goes to 0, thus the realizations should converge to a constant value, which should be the true value of the parameter of our interest, almost surely.

Note that the index k after the comma in $K_{i,k}$ means k-th observation, we don't need this index for notations that contain M_k since it indicates which observation already:

For independent censoring, for each $i = 1, 2, ..., r$:

$$n_{ii} = \sum_{k=1}^{n}(K_i^{M_k} - I(K_i^{M_k} > 0))$$

(Total occurrences of state i subtracted by how many observations it occurs in)

$$n_{ij} = \sum_{k=1}^{n} I_{ij}^{M_k}, j > i$$

$$\sum_{j=i+1}^{r+1} n_{ij} = \sum_{j=i+1}^{r+1}\sum_{k=1}^{n} I_{ij}^{M_k}$$

$$= \sum_{k=1}^{n}\sum_{j=i+1}^{r+1} I_{ij}^{M_k}$$

Thus:

$$\hat{p}_{ii,n} = \frac{\sum_{k=1}^{n}(K_i^{M_k} - I(K_i^{M_k} > 0))}{\sum_{k=1}^{n}(K_i^{M_k} - I(K_i^{M_k} > 0) + \sum_{j=i+1}^{r+1} I_{ij}^{M_k})}$$

$$\hat{p}_{ij,n} = \frac{\sum_{k=1}^{n} I_{ij}^{M_k}}{\sum_{k=1}^{n}(K_i^{M_k} - I(K_i^{M_k} > 0) + \sum_{j=i+1}^{r+1} I_{ij}^{M_k})}, j > i$$

For no censoring, for each $i = 1, 2, ..., r$:

$$n_{ii} = \sum_{k=1}^{n}(K_{i,k} - I(K_{i,k} > 0))$$

(Total occurrences of state i subtracted by how many observations it occurs in)

$$n_{ij} = \sum_{k=1}^{n} I_{ij,k} \quad j > i$$

$$I(K_{i,k} > 0) = \sum_{j=i+1}^{r+1} I_{ij,k}$$

(If i appears, then transition from i must happen only once;

if not, the transition doesn't happen either)

So:

$$\begin{aligned} n_{ii} &= \sum_{k=1}^{n} K_{i,k} - \sum_{k=1}^{n} I(K_{i,k} > 0) \\ &= \sum_{k=1}^{n} K_{i,k} - \sum_{k=1}^{n} \sum_{j=i+1}^{r+1} I_{ij,k} \\ &= \sum_{k=1}^{n} K_{i,k} - \sum_{j=i+1}^{r+1} \sum_{k=1}^{n} I_{ij,k} \\ &= \sum_{k=1}^{n} K_{i,k} - \sum_{j=i+1}^{r+1} n_{ij} \end{aligned}$$

Therefore:

$$\sum_{j=i}^{r+1} n_{ij} = \sum_{k=1}^{n} K_{i,k}$$

Thus:

$$\begin{aligned} \hat{p}_{ii,n} &= \frac{\sum_{k=1}^{n}(K_{i,k} - I(K_{i,k} > 0))}{\sum_{k=1}^{n} K_{i,k}} \\ \hat{p}_{ij,n} &= \frac{\sum_{k=1}^{n} I_{ij,k}}{\sum_{k=1}^{n} K_{i,k}}, \quad j > i \end{aligned}$$

Now we can prove the consistency of the MLE:

Theorem 1. *Under independent censoring, the MLE $\hat{p}_{ij,n}$ is consistent for p_{ij}, if the following condition holds:*

For transition $i \to j$, there exists a non-zero probability of observing $i \to j$ under the censoring mechanism.

The condition makes sure that at least two kinds of transition($i \to i$, $i \to j$) have a chance to be observed in the data, thus the denominator of MLE($\sum_{k=i}^{r+1} n_{ik}$) could be non-zero, and $\hat{p}_{ij,n}$ does not converge to 0 if p_{ij} is not 0. The MLE is undefined if the observed denominator is 0.

Proof. The proof consists of two main parts:

1) First, we recall equations (2.4.1) and (2.4.2) :

$$\frac{E(K_i^M) - P(K_i^M > 0)}{E(K_i^M) - P(K_i^M > 0) + P(\sum_{j>i} I_{ij}^M = 1)} = p_{ii}$$

$$\frac{E(I_{ij}^M)}{E(K_i^M) - P(K_i^M > 0) + P(\sum_{j>i} I_{ij}^M = 1)} = p_{ij}$$

And also we have obtained before:

$$\hat{p}_{ii,n} = \frac{\sum_{k=1}^{n}(K_i^{M_k} - I(K_i^{M_k} > 0))}{\sum_{k=1}^{n}(K_i^{M_k} - I(K_i^{M_k} > 0) + \sum_{j=i+1}^{r+1} I_{ij}^{M_k})}$$

$$\hat{p}_{ij,n} = \frac{\sum_{k=1}^{n} I_{ij}^{M_k}}{\sum_{k=1}^{n}(K_i^{M_k} - I(K_i^{M_k} > 0) + \sum_{j=i+1}^{r+1} I_{ij}^{M_k})}, j > i$$

Since M is independent of K_i, and each observation is also independent from each other, so $K_i^{M_k}$ are a sequence of independent and identically distributed random variables with expectation of $E(K_i^M)$, same for $I_{ij}^{M_k}$.

2) By the Strong Law of Large Numbers, when $n \to \infty$:

$$\frac{1}{n}\sum_{k=1}^{n}(K_i^{M_k} - I(K_i^{M_k} > 0)) \xrightarrow{a.s.} E[K_i^M - I(K_i^M > 0)]$$

$$\frac{1}{n}\sum_{k=1}^{n} I_{ij}^{M_k} \xrightarrow{a.s.} E[I_{ij}^M]$$

$$\frac{1}{n}\sum_{k=1}^{n}\left(K_i^{M_k} - I(K_i^{M_k} > 0) + \sum_{j=i+1}^{r+1} I_{ij}^{M_k}\right) \xrightarrow{a.s.} E[K_i^M - I(K_i^M > 0) + \sum_{j=i+1}^{r+1} I_{ij}^M]$$

$$Note\ that : E(\sum_{j=i+1}^{r+1} I_{ij}^M) = P(\sum_{j=i+1}^{r+1} I_{ij}^M = 1).$$

Therefore by Continuous Mapping Theorem:

$$\hat{p}_{ii,n} = \frac{\sum_{k=1}^{n}(K_i^{M_k} - I(K_i^{M_k} > 0))}{\sum_{k=1}^{n}(K_i^{M_k} - I(K_i^{M_k} > 0) + \sum_{j=i+1}^{r+1} I_{ij}^{M_k})} \xrightarrow{a.s.} \frac{E(K_i^M) - P(K_i^M > 0)}{E(K_i^M) - P(K_i^M > 0) + P(\sum_{j>i} I_{ij}^M = 1)} = p_{ii}$$

$$\hat{p}_{ij,n} = \frac{\sum_{k=1}^{n} I_{ij}^{M_k}}{\sum_{k=1}^{n}(K_i^{M_k} - I(K_i^{M_k} > 0) + \sum_{j=i+1}^{r+1} I_{ij}^{M_k})} \xrightarrow{a.s.} \frac{E(I_{ij}^M)}{E(K_i^M) - P(K_i^M > 0) + P(\sum_{j>i} I_{ij}^M = 1)} = p_{ij}$$

□

Note that constant censoring is just a special case of independent censoring.

The consistency for the uncensored case can also be proved similarly, recall what we obtained for uncensored case before:

$$\hat{p}_{ii,n} = \frac{\sum_{k=1}^{n}(K_{i,k} - I(K_{i,k} > 0))}{\sum_{k=1}^{n} K_{i,k}} \xrightarrow{a.s.} \frac{E[K_i] - P[K_i > 0]}{E[K_i]} = p_{ii}$$
$$\hat{p}_{ij,n} = \frac{\sum_{k=1}^{n} I_{ij,k}}{\sum_{k=1}^{n} K_{i,k}} \xrightarrow{a.s.} \frac{E[I_{ij}]}{E[K_i]} = p_{ij} \quad j > i$$

2.5 Simulation Examples

To empirically validate the consistency results established in Theorem 1, we conducted a series of simulation studies examining the behavior of the maximum likelihood estimator under various conditions.

2.5.1 Simulation Setup

We simulated a progressive Markov chain with $r = 3$ states, using the following transition matrix:

$$P = \begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0 & 0.6 & 0.4 \\ 0 & 0 & 1 \end{pmatrix}$$

For each simulation setting, we generated n independent realizations of the Markov chain under no censoring, constant censoring and random censoring. The random censoring times followed a shifted Poisson distribution (ensuring a minimum censoring time of 2).

2.5.2 Results

For no censoring:

Figure 1 shows us that without censoring, all estimations converge quickly to the true value.

For constant censoring:

Figure 2 shows us simulations results under constant censoring of 12. We can see clearly that estimations for p_{22} and p_{23} converge slower than when without censoring. It's mainly because p_{11} is relatively large so we don't have many observations of state 2 with small sample size.

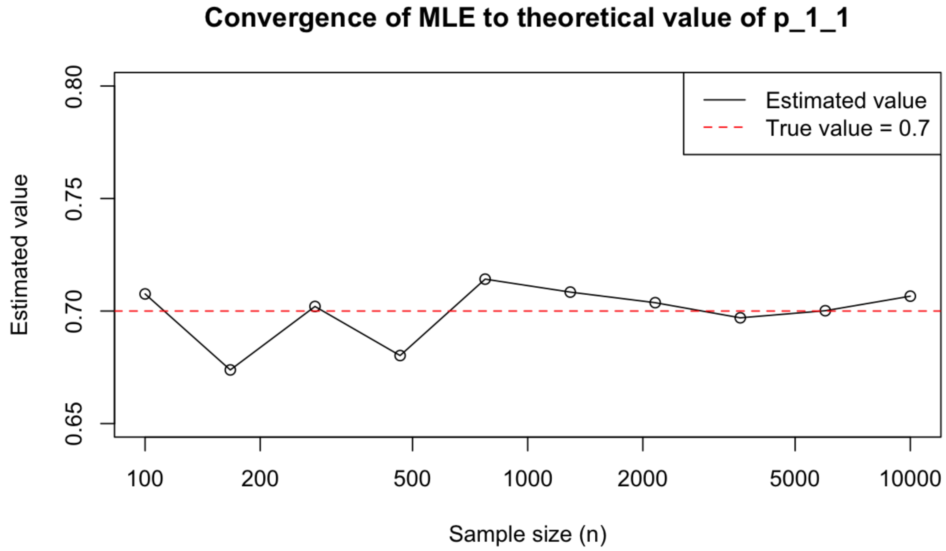
For random censoring

Figure 3 shows us simulations results under random censoring, $M \sim Poi(12) + 2$. We can observe the same problem as in constant censoring, but eventually they will converge to the true value.

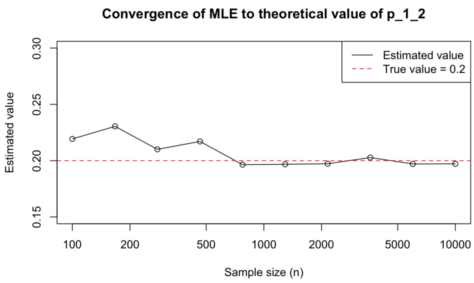
The simulation results clearly demonstrate the consistency of the maximum likelihood estimator. As predicted by Theorem 1, we observe:

1. The estimates converge to the true parameter values as the sample size increases
2. The convergence occurs even under random censoring, confirming the theoretical result that independent censoring preserves consistency

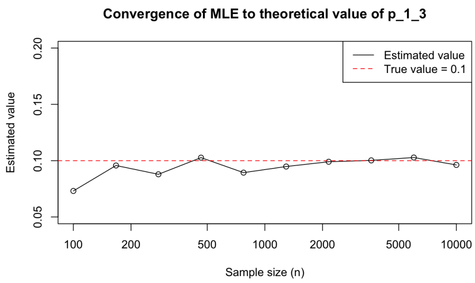
For transition probabilities associated with rarely visited states (such as p_{22} and



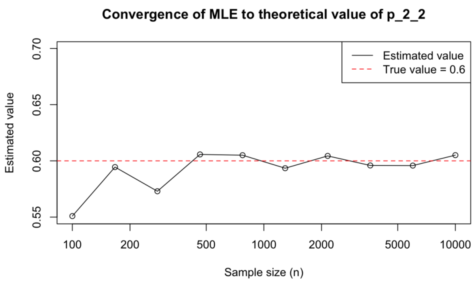
(a) Convergence of p_{11}



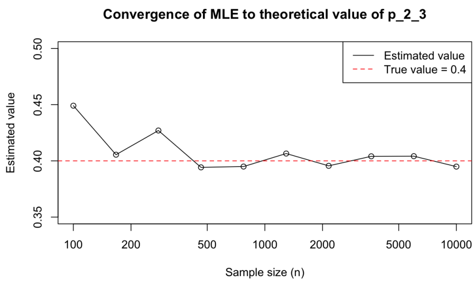
(b) Convergence of p_{12}



(c) Convergence of p_{13}

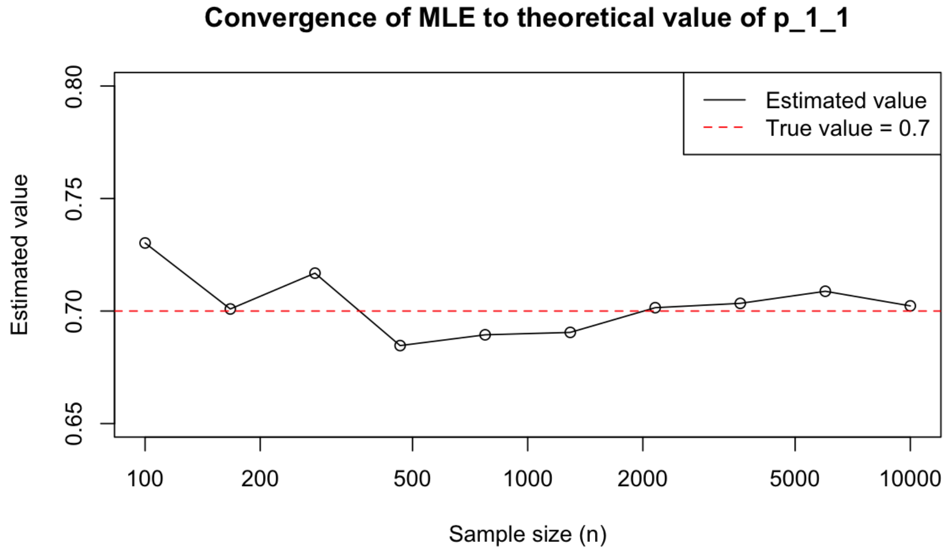


(d) Convergence of p_{22}

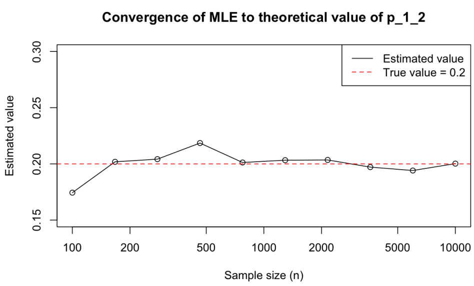


(e) Convergence of p_{23}

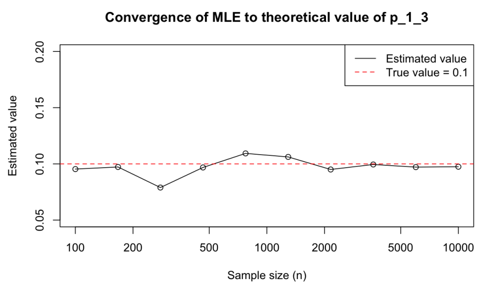
Figure 1: Convergence under no censoring



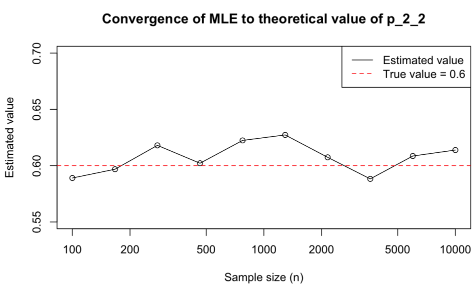
(a) Convergence of p_{11}



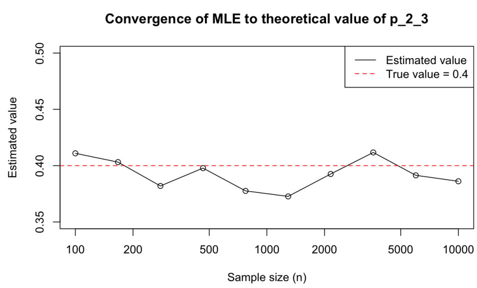
(b) Convergence of p_{12}



(c) Convergence of p_{13}

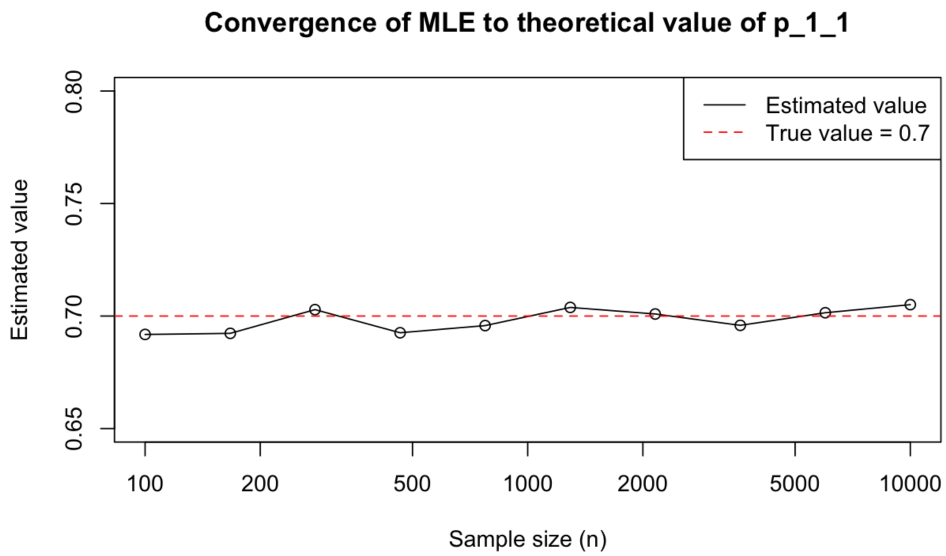


(d) Convergence of p_{22}

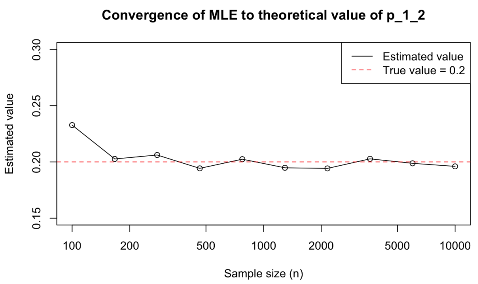


(e) Convergence of p_{23}

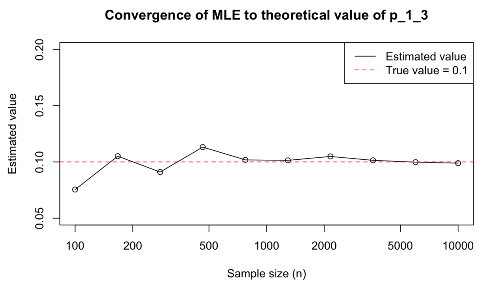
Figure 2: Convergence under fixed censoring



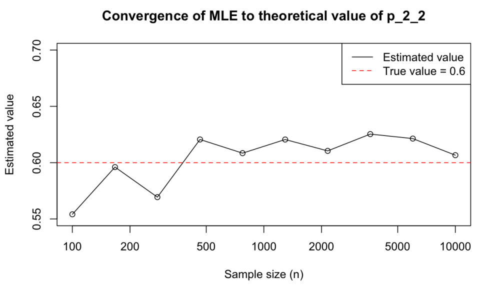
(a) Convergence of p_{11}



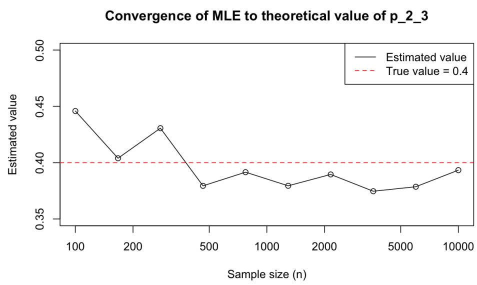
(b) Convergence of p_{12}



(c) Convergence of p_{13}



(d) Convergence of p_{22}



(e) Convergence of p_{23}

Figure 3: Convergence under random censoring

p_{23} when p_{12} is small), larger sample sizes are required to achieve the same level of precision as more frequently visited states.

3 Two-State Hidden Markov Chain

3.1 Model Definition

Let $\{X_t\}_{t\geq 1}$ be a Markov chain with state space $S = \{1, 2\}$ and transition matrix

$$\begin{pmatrix} p & 1-p \\ 0 & 1 \end{pmatrix}$$

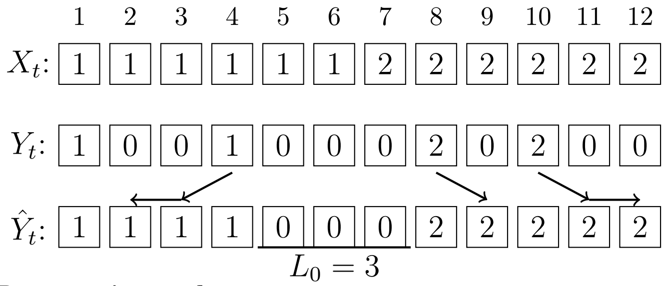
where $X_1 = 1$. This represents a simple progressive Markov chain with an absorbing state.

The observation process $\{Y_t\}_{t\geq 1}$ is defined as:

$$Y_t = \begin{cases} X_t, & \text{with probability } 1-q \\ 0, & \text{with probability } q \end{cases}$$

Due to the progressive structure of the chain, we can reconstruct some missing observations. Let $\{\hat{Y}_t\}_{t\geq 1}$ be the reconstructed process where:

$$\hat{Y}_t = \begin{cases} 1, & \text{if we can observe or recover } X_t = 1 \\ 2, & \text{if we can observe or recover } X_t = 2 \\ 0, & \text{otherwise} \end{cases}$$



Recovering rules:

1. Backward: If $X_t = 1$ observed, all previous unobserved must be 1

2. Forward: If $X_t = 2$ observed, all subsequent unobserved must be 2

3. Zero-block (L_0): Area where state cannot be inferred

Figure 4: Reconstruction of censored observations in a two-state hidden Markov model

Note that $\hat{Y}_1 = 1$.

3.2 Block Structure in Observations

The process $\{\hat{Y}_t\}$ typically contains three distinct segments:

- An initial block of 1s (observed state 1, with length of at least 1)
- A middle block of 0s (unobserved states, length of which could be 0)
- A final block of 2s (observed state 2, infinite long if it's uncensored)

We define the following block length random variables:

$$L_1 = \sum_{t=1}^{\infty} I_1(\hat{Y}_t) \qquad \text{(length of observed 1-block)}$$
$$L_0 = \sum_{t=1}^{\infty} I_0(\hat{Y}_t) \qquad \text{(length of observed 0-block)}$$

3.3 Distributions of Block Lengths

3.3.1 Distribution of L_1

To analyze the distributions of L_1, we introduce two fundamental random variables:

- $X \sim G_0(1-p)$: The number of additional states 1 after the first state (excluding the first state). X follows a geometric distribution with parameter $(1-p)$:

$$P(X = k) = p^k(1-p), \quad k \geq 0$$

- $Y \sim G_0(1-q)$: The number of consecutive zeros counting backward from the position of the last state 1. Y follows a geometric distribution with parameter $(1-q)$:

$$P(Y = k) = q^k(1-q), \quad k \geq 0$$

The observed length L_1 can be expressed in terms of these random variables as:

$$L_1 = X - (X \wedge Y) + 1$$

where $X \wedge Y = \min\{X, Y\}$ represents the number of state 1s that are hidden (not observed) in the zero-block, and "+1" accounts for the first observed state 1.

To explain the distribution of $X \wedge Y$, let's take a look at X first. We know that $P(X = k) = p^k(1-p), \quad k \geq 0$, which can be considered as a number starts with 0, with probability p to increase by 1, with probability $1-p$ to stop the process. Then it's not hard to observe that, for $X \wedge Y$, it has the probability pq to increase by 1(it can increase only when both X and Y increase), and $1-pq$ to stop the

Time points:										
1	2	3	4	5	6	7	8	9	10	11
X_t (underlying states):										
1	1	1	1	1	1	2	2	2	2	2
$\hat{Y}_t$ (observed process):										
1	1	1	0	0	0	0	2	2	2	2

Variable	Description	Value
L_1	Length of observed 1-block	3
L_0	Length of observed 0-block	4
X	Number of additional 1s after first position	5
Y	Number of 0s counting backward from last state 1	3
$X \wedge Y$	Number of hidden state 1s	3

Figure 5: Illustration of block structure and random variables. Note that Y counts backwards from the last state 1, explaining why $X \wedge Y$ represents the number of hidden 1s.

process(it stops when one of them stops). Thus $X \wedge Y \sim G_0(1 - pq)$:

$$P(X \wedge Y = k) = (pq)^k (1 - pq), \quad k \geq 0$$

To derive the probability mass function of L_1, we need to analyze how L_1 relates to the events involving X and Y. First, consider the special case when $L_1 = 1$:

$$\begin{aligned} P(L_1 = 1) &= P(X - (X \wedge Y) + 1 = 1) \\ &= P(X - (X \wedge Y) = 0) \\ &= P(X = X \wedge Y) \\ &= P(X \leq Y) \end{aligned}$$

For two independent geometric random variables, we can compute:

$$\begin{aligned}
P(X \leq Y) &= \sum_{k=0}^{\infty} P(X = k) P(Y \geq k) \\
&= \sum_{k=0}^{\infty} p^k (1-p) \cdot \sum_{j=k}^{\infty} q^j (1-q) \\
&= \sum_{k=0}^{\infty} p^k (1-p) \cdot q^k \\
&= (1-p) \sum_{k=0}^{\infty} (pq)^k \\
&= \frac{1-p}{1-pq}
\end{aligned}$$

Now for $L_1 = l > 1$, we have:

$$\begin{aligned}
P(L_1 = l) &= P(X - (X \wedge Y) + 1 = l) \\
&= P(X - (X \wedge Y) = l - 1)
\end{aligned}$$

This occurs when $X > Y$ and $X - Y = l - 1$. We can calculate:

$$\begin{aligned}
P(L_1 = l) &= \sum_{k=l-1}^{\infty} P(X = k, Y = k - (l-1)) \\
&= \sum_{k=l-1}^{\infty} P(X = k)P(Y = k - (l-1)) \\
&= \sum_{k=l-1}^{\infty} p^k(1-p) \cdot q^{k-(l-1)}(1-q) \\
&= (1-p)(1-q) \cdot \sum_{k=l-1}^{\infty} p^k \cdot q^{k-(l-1)} \\
&= (1-p)(1-q) \cdot q^{-(l-1)} \cdot \sum_{k=l-1}^{\infty} (pq)^k \\
&= (1-p)(1-q) \cdot q^{-(l-1)} \cdot \frac{(pq)^{l-1}}{1-pq} \\
&= \frac{(1-p)(1-q)p^{l-1}}{1-pq}, \quad l > 1
\end{aligned}$$

Therefore, the complete probability mass function of L_1 is:

$$\begin{aligned}
P(L_1 = 1) &= \frac{1-p}{1-pq} \\
P(L_1 = l) &= \frac{(1-p)(1-q)p^{l-1}}{1-pq}, \quad l > 1
\end{aligned}$$

3.3.2 Distribution of L_0

For the 0-blocks, we can decompose L_0 into two independent components:

$$L_0 = Z_1 + Z_2$$

Where:

- Z_1 represents the number of zeros that occur after the last unobserved state 1 (in the "1-side" of the hidden block)

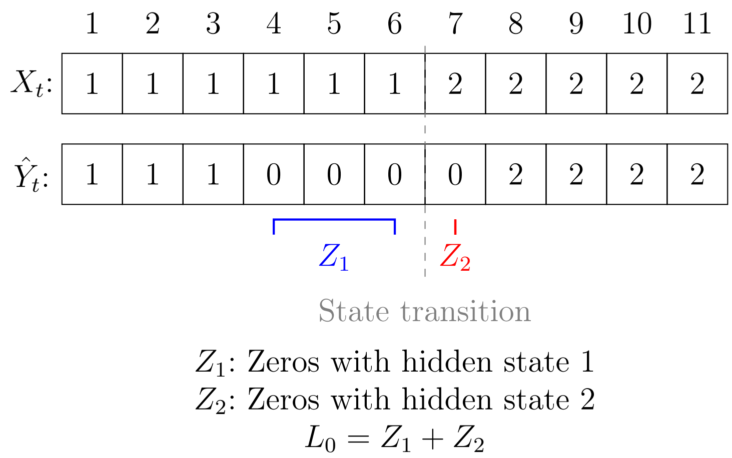


Figure 6: Decomposition of 0-block into Z_1 and Z_2 components. Z_1 represents zeros that correspond to hidden state 1, while Z_2 represents zeros that correspond to hidden state 2.

- Z_2 represents the number of zeros that occur after transitioning to state 2, before the first observed 2 (in the "2-side")

Analyzing their distributions:

For Z_1, note that it is actually the same as $X \wedge Y$ introduced in previous subsection. Thus:

$$Z_1 \sim G_0(1 - pq)$$

with probability mass function:

$$P(Z_1 = k) = (pq)^k (1 - pq), \quad k \geq 0$$

For Z_2, it follows a shifted geometric distribution based on observing zeros after transitioning to state 2:

$$Z_2 \sim G_0(1 - q)$$

with probability mass function:

$$P(Z_2 = k) = q^k(1-q), \quad k \geq 0$$

Since $L_0 = Z_1 + Z_2$ and Z_1 and Z_2 are independent, the probability mass function of L_0 is the convolution of the distributions of Z_1 and Z_2:

$$\begin{aligned}
P(L_0 = l) &= \sum_{j=0}^{l} P(Z_1 = j) \cdot P(Z_2 = l - j) \\
&= \sum_{j=0}^{l} (pq)^j (1 - pq) \cdot q^{l-j} (1 - q) \\
&= (1 - pq)(1 - q) \cdot \sum_{j=0}^{l} (pq)^j \cdot q^{l-j} \\
&= (1 - pq)(1 - q) \cdot q^l \cdot \sum_{j=0}^{l} p^j \\
&= \frac{(1 - pq)(1 - q)}{1 - p} \cdot q^l \cdot (1 - p^{l+1}), \quad l \geq 0
\end{aligned}$$

3.4 Expected Length of 1-Blocks

3.4.1 Using X and Y

Using the representation $L_1 = X - (X \wedge Y) + 1$, and since X and Y are independent, we have:

$$\begin{aligned}
E(L_1) &= E(X) - E(X \wedge Y) + 1 \\
&= \frac{p}{1-p} - \frac{pq}{1-pq} + 1
\end{aligned}$$

Note that $\frac{p}{1-p} = \frac{1}{1-p} - 1$, so it also can be written as:

$$E(L_1) = \frac{1}{1-p} - \frac{1}{1-pq} + 1$$

Simplifying:

$$\begin{aligned}
E(L_1) &= \frac{1}{1-p} - \frac{1}{1-pq} + 1 \\
&= \frac{1}{1-p} - \frac{pq}{1-pq} \\
&= \frac{1-pq-(1-p)pq}{(1-p)(1-pq)} \\
&= \frac{1-2pq+p^2q}{(1-p)(1-pq)}
\end{aligned}$$

3.4.2 Using distribution

We can also compute $E(L_1)$ directly from its probability mass function:

$$\begin{aligned}
E(L_1) &= \sum_{l=1}^{\infty} l \cdot P(L_1 = l) \\
&= 1 \cdot P(L_1 = 1) + \sum_{l=2}^{\infty} l \cdot P(L_1 = l) \\
&= 1 \cdot \frac{1-p}{1-pq} + \sum_{l=2}^{\infty} l \cdot \frac{(1-p)(1-q)p^{l-1}}{1-pq} \\
&= \frac{1-p}{1-pq} + \frac{(1-p)(1-q)}{1-pq} \cdot \sum_{l=2}^{\infty} l \cdot p^{l-1}
\end{aligned}$$

Using the identity $\sum_{l=1}^{\infty} l \cdot p^{l-1} = \frac{1}{(1-p)^2}$ for $|p| < 1$, we have:

$$\begin{aligned}
\sum_{l=2}^{\infty} l \cdot p^{l-1} &= \sum_{l=1}^{\infty} l \cdot p^{l-1} - 1 \cdot p^{1-1} \\
&= \frac{1}{(1-p)^2} - 1
\end{aligned}$$

Substituting:

$$\begin{aligned}
E(L_1) &= \frac{1-p}{1-pq} + \frac{(1-p)(1-q)}{1-pq} \cdot \left(\frac{1}{(1-p)^2} - 1 \right) \\
&= \frac{1-p}{1-pq} + \frac{1-q}{1-pq} \cdot \left(\frac{1}{1-p} - (1-p) \right) \\
&= \frac{1-p}{1-pq} + \frac{1-q}{1-pq} \cdot \frac{1-(1-p)^2}{1-p} \\
&= \frac{1-p}{1-pq} + \frac{1-q}{1-pq} \cdot \frac{1-1+2p-p^2}{1-p} \\
&= \frac{1-p}{1-pq} + \frac{1-q}{1-pq} \cdot \frac{p(2-p)}{1-p} \\
&= \frac{(1-p)^2 + p(1-q)(2-p)}{(1-pq)(1-p)} \\
&= \frac{1-2p+p^2+2p-2pq-p^2+p^2q}{(1-pq)(1-p)} \\
&= \frac{1-2pq+p^2q}{(1-pq)(1-p)}
\end{aligned}$$

which matches the result from the first approach.

3.5 Expected Length of 0-Blocks

3.5.1 Using Z_1 and Z_2

Using the representation $L_0 = Z_1 + Z_2$, and since Z_1 and Z_2 are independent, we have:

$$E(L_0) = E(Z_1) + E(Z_2)$$

For a geometric random variable $G_0(1-r)$, the expected value is $\frac{r}{1-r}$. Since $Z_1 \sim$

$G_0(1-pq)-1$ and $Z_2 \sim G_0(1-q)-1$, we have:

$$
\begin{aligned}
E(Z_1) &= \frac{pq}{1-pq} \\
E(Z_2) &= \frac{q}{1-q}
\end{aligned}
$$

Therefore:

$$
\begin{aligned}
E(L_0) &= \frac{pq}{1-pq} + \frac{q}{1-q} \\
&= \frac{pq(1-q) + q(1-pq)}{(1-pq)(1-q)} \\
&= \frac{pq - pq^2 + q - pq^2}{(1-pq)(1-q)} \\
&= \frac{q(1+p-2pq)}{(1-pq)(1-q)}
\end{aligned}
$$

The advantage of this approach is that it can be easily interpreted, and can handle the scenarios simply when we have different probabilities of observing two states. Let's say we have chance of $1-q_1$ to observe 1 and $1-q_2$ for state 2, then:

$$
E(L_0) = \frac{pq_1}{1-pq_1} + \frac{q_2}{1-q_2}
$$

In future steps when we want to estimate p, it also provides simpler calculations.

NB! The case that two states have different probabilities of being observed is not addressed in this thesis.

3.5.2 Using distribution

We can compute $E(L_0)$ directly from its probability mass function:

$$\begin{aligned} E(L_0) &= \sum_{l=0}^{\infty} l \cdot P(L_0 = l) \\ &= \sum_{l=0}^{\infty} l \cdot \frac{(1-pq)(1-q)}{1-p} \cdot q^l \cdot (1-p^{l+1}) \end{aligned}$$

This can be expanded as:

$$E(L_0) = \frac{(1-pq)(1-q)}{1-p} \cdot \left[\sum_{l=0}^{\infty} l \cdot q^l - \sum_{l=0}^{\infty} l \cdot q^l \cdot p^{l+1} \right]$$

Using the identities $\sum_{l=0}^{\infty} l \cdot q^l = \frac{q}{(1-q)^2}$ and $\sum_{l=0}^{\infty} l \cdot (pq)^l = \frac{pq}{(1-pq)^2}$ for $|q|, |pq| < 1$, we get:

$$
\begin{aligned}
E(L_0) &= \frac{(1-pq)(1-q)}{1-p} \cdot \left[\frac{q}{(1-q)^2} - p \cdot \sum_{l=0}^{\infty} l \cdot (pq)^l \right] \\
&= \frac{(1-pq)(1-q)}{1-p} \cdot \left[\frac{q}{(1-q)^2} - p \cdot \frac{pq}{(1-pq)^2} \right] \\
&= \frac{q(1-pq)(1-q)}{1-p} \cdot \left[\frac{1}{(1-q)^2} - \frac{p^2}{(1-pq)^2} \right] \\
&= \frac{q(1-pq)(1-q)}{1-p} \cdot \left[\frac{(1-pq)^2}{(1-q)^2(1-pq)^2} - \frac{p^2(1-q)^2}{(1-q)^2(1-pq)^2} \right] \\
&= \frac{q(1-pq)(1-q)}{1-p} \cdot \frac{(1-pq)^2 - p^2(1-q)^2}{(1-q)^2(1-pq)^2} \\
&= \frac{q}{1-p} \cdot \frac{1-2pq+p^2q^2-p^2(1-2q+q^2)}{(1-q)(1-pq)} \\
&= \frac{q}{1-p} \cdot \frac{1-2pq-p^2+2p^2q}{(1-q)(1-pq)} \\
&= \frac{q}{1-p} \cdot \frac{1-p^2-2pq+2p^2q}{(1-q)(1-pq)} \\
&= \frac{q}{1-p} \cdot \frac{(1+p)(1-p)-2pq(1-p)}{(1-q)(1-pq)} \\
&= \frac{q(1+p-2pq)}{(1-pq)(1-q)}
\end{aligned}
$$

which matches our first approach.

3.5.3 Using X and Y

We can derive $E(L_0)$ using a direct approach based on conditional expectation. In this context, we define:

- $X \sim G_0(1-p)$: The number of hidden state 1s after the last observed state 1
- $Y \sim G_0(1-q)$: The number of consecutive zeros counting forward from the last observed state 1

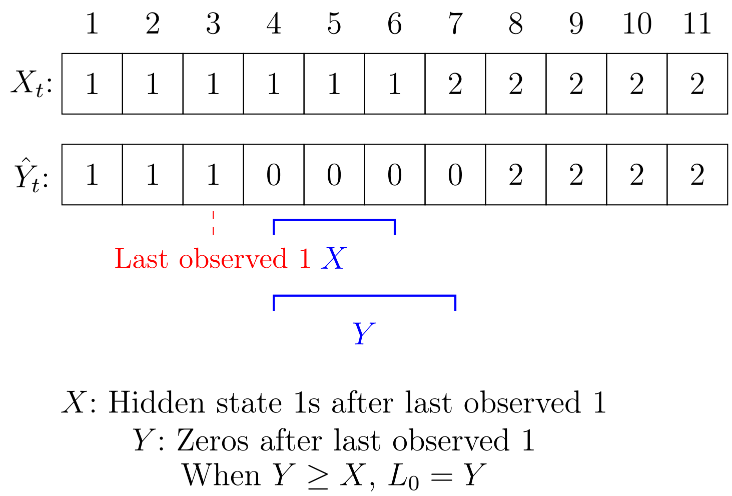


Figure 7: Illustration of X and Y in the conditional expectation approach.

Note that unlike the case in L_1, X and Y here count different things, but still follow same corresponding distributions.

Only when $Y \geq X$, the zero blocks can survive after the modification, and the total length of the 0-block equals Y, the number of consecutive zeros after the last observed state 1. Therefore:

$$L_0 = Y \quad \text{when } Y \geq X$$

This gives us:

$$E(L_0) = E(Y|Y \geq X)$$

To calculate $E(Y|Y \geq X)$, we need:

$$E(Y|Y \geq X) = \frac{E(Y \cdot I(Y \geq X))}{P(Y \geq X)}$$

First, we know $P(Y \geq X) = \frac{1-p}{1-pq}$.

Next, we compute $E(Y \cdot I(Y \geq X))$:

$$
\begin{aligned}
E(Y \cdot I(Y \geq X)) &= \sum_{k=0}^{\infty} \sum_{j=k}^{\infty} j \cdot P(X = k, Y = j) \\
&= \sum_{k=0}^{\infty} \sum_{j=k}^{\infty} j \cdot P(X = k) \cdot P(Y = j) \\
&= \sum_{k=0}^{\infty} \sum_{j=k}^{\infty} j \cdot p^k (1-p) \cdot q^j (1-q) \\
&= (1-p)(1-q) \sum_{k=0}^{\infty} p^k \sum_{j=k}^{\infty} j \cdot q^j
\end{aligned}
$$

For the inner sum, we have:

$$
\begin{aligned}
\sum_{j=k}^{\infty} j \cdot q^j &= q^k \sum_{i=0}^{\infty} (k+i) \cdot q^i \quad \text{(with substitution } j = k+i\text{)} \\
&= q^k \left[k \sum_{i=0}^{\infty} q^i + \sum_{i=0}^{\infty} i \cdot q^i \right] \\
&= q^k \left[\frac{k}{1-q} + \frac{q}{(1-q)^2} \right] \\
&= q^k \frac{k(1-q) + q}{(1-q)^2}
\end{aligned}
$$

Substituting back:

$$
\begin{aligned}
E(Y \cdot I(Y \geq X)) &= (1-p)(1-q) \sum_{k=0}^{\infty} p^k \cdot q^k \cdot \frac{k(1-q)+q}{(1-q)^2} \\
&= \frac{1-p}{1-q} \sum_{k=0}^{\infty} (pq)^k \cdot [k(1-q) + q] \\
&= \frac{1-p}{1-q} \left[(1-q) \sum_{k=0}^{\infty} k \cdot (pq)^k + q \sum_{k=0}^{\infty} (pq)^k \right]
\end{aligned}
$$

Using the identities $\sum_{k=0}^{\infty} k \cdot r^k = \frac{r}{(1-r)^2}$ and $\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$ for $|r| < 1$:

$$
\begin{aligned}
E(Y \cdot I(Y \geq X)) &= \frac{1-p}{1-q}\left[(1-q)\cdot\frac{pq}{(1-pq)^2} + q\cdot\frac{1}{1-pq}\right] \\
&= \frac{1-p}{1-q}\cdot\frac{pq(1-q)+q(1-pq)}{(1-pq)^2} \\
&= \frac{(1-p)[pq(1-q)+q(1-pq)]}{(1-q)(1-pq)^2} \\
&= \frac{(1-p)[pq-pq^2+q-pq^2]}{(1-q)(1-pq)^2} \\
&= \frac{(1-p)q[1+p-2pq]}{(1-q)(1-pq)^2}
\end{aligned}
$$

Finally:

$$
\begin{aligned}
E(Y|Y \geq X) &= \frac{E(Y \cdot I(Y \geq X))}{P(Y \geq X)} \\
&= \frac{\frac{(1-p)q[1+p-2pq]}{(1-q)(1-pq)^2}}{\frac{1-p}{1-pq}} \\
&= \frac{q[1+p-2pq]}{(1-q)(1-pq)}
\end{aligned}
$$

Therefore:

$$
E(L_0) = E(Y|Y \geq X) = \frac{q(1+p-2pq)}{(1-pq)(1-q)}
$$

3.6 Parameter Estimation

The observed block lengths provide valuable information for estimating the parameter p of the hidden Markov chain. We will explore several approaches to parameter estimation, let's start with some important notations first:

To estimate a parameter, we need to establish the relationship between it and real observations. Assume we have n observations, each observation contains a 1-block and 0-block.

Let's consider 1-block, the 0-block is similar:

For i-th observation, the observed length of 1-block is denoted as l_1^i, and the length of 1-block in i-th trial is a random variable denoted as L_1^i. So l_1^i the is realization of L_1^i, and $L_1^i,\ i = 1, 2, ..., n$ are n i.i.d. random variables.

We also denote:

$$L_1(n) := \sum_{i=1}^{n} L_1^i$$

and:

$$l_1(n) := \sum_{i=1}^{n} l_1^i$$

We will also consider cases when we don't know about q, in this case we can estimate it using number of zeros and non-zeros before we obtain $\hat{Y}_t$, i.e. using Y_t.

Denote $z(n)$ as the number of zeros in all n observations before modification and $\tilde{z}(n)$ as the counterpart, we can have a consistent estimator for q:

$$\hat{q} = \frac{z(n)}{z(n) + \tilde{z}(n)}$$

3.6.1 Moments Estimator Based on L_1

We already know that $E(L_1)$ is a function of p, if we denote that $E(L_1) = h_1(p)$, then we can define $f_1 = h_1^{-1}$, so that $p = f_1(E(L_1))$.

First we need to show that the inverse function f_1 exists:

Let's define $h_1(p) = E(L_1)$. We need to show that h_1 is strictly monotonic (to ensure f_1 is well-defined) and continuous.

Taking the derivative:

$$\begin{aligned}\frac{dh_1}{dp} &= \frac{\partial}{\partial p}\left(\frac{1}{1-p} - \frac{1}{1-pq} + 1\right) \\ &= \frac{1}{(1-p)^2} - \frac{q}{(1-pq)^2} \\ &= \frac{(1-pq)^2 - (1-p)^2 q}{(1-p)^2(1-pq)^2} \\ &= \frac{1 - 2pq + p^2q^2 - (1 - 2p + p^2)q}{(1-p)^2(1-pq)^2} \\ &= \frac{1 - q + p^2q^2 - p^2q}{(1-p)^2(1-pq)^2} \\ &= \frac{(1-q)(1-p^2q)}{(1-p)^2(1-pq)^2}\end{aligned}$$

Thus, $\frac{dh_1}{dp} > 0$ for all $p, q \in (0, 1)$, confirming h_1 is strictly increasing.

Since h_1 is strictly increasing, $f_1 = h_1^{-1}$ exists and is well-defined. Also, since h_1 is continuous, f_1 is continuous.

Note that in case q is unknown, it can also be written as $p = f_1(EL_1; q)$. Thus we can define estimator for p:

$$\hat{p}_{L_1,n} = f_1(\frac{L_1(n)}{n})$$

For unknown q:

$$\hat{p}_{L_1,n} = f_1(\frac{L_1(n)}{n}; \hat{q})$$

Note that the second form is more general, the first form can be considered as a special case of it. So during the following proofs, we will focus on proving the second form.

Let's find f_1 first:

$$\begin{aligned}
&EL_1 = \frac{1 - 2pq + p^2 q}{(1-p)(1-pq)} \\
&EL_1(1-p)(1-pq) = 1 - 2pq + p^2 q \\
&EL_1(1 - p - pq + p^2 q) + 2pq - p^2 q - 1 = 0 \\
&(EL_1 q - q)p^2 + (2q - EL_1 - EL_1 q)p + EL_1 - 1 = 0
\end{aligned}$$

if $\Delta = (2q - EL_1 - EL_1 q)^2 - 4(EL_1 q - q)(EL_1 - 1) = (EL_1)^2(1-q)^2 + 4q(1-q)(EL_1 - 1) \geq 0$,

then:

$$p = \frac{EL_1 + EL_1 q - 2q \pm \sqrt{\Delta}}{2(EL_1 q - q)}$$

We know only one form is correct, and if take $EL_1 \to \infty$, we should have $p \to 1$:

$$\begin{aligned}
&\lim_{EL_1 \to \infty} \frac{EL_1 + EL_1 q - 2q \pm \sqrt{(EL_1)^2(1-q)^2 + 4q(1-q)(EL_1 - 1)}}{2(EL_1 q - q)} \\
&= \lim_{EL_1 \to \infty} \frac{EL_1(1+q) \pm \sqrt{(EL_1)^2(1-q)^2}}{2qEL_1} \\
&= \frac{(1+q) \pm (1-q)}{2q}
\end{aligned}$$

If we take plus, then it becomes $\frac{1}{q} > 1$, which is obvious not our function;

if we take minus, it will be 1, which is what we expected, thus:

$$f_1(EL_1; q) = \frac{EL_1 + EL_1 q - 2q \pm \sqrt{\Delta}}{2(EL_1 q - q)}$$

Therefore:

$$\hat{p}_{L_1,n} = f_1(\frac{L_1(n)}{n}; \hat{q}) = \frac{(1+\hat{q})\frac{L_1(n)}{n} - 2\hat{q} - \sqrt{\Delta_n}}{2(\hat{q}\frac{L_1(n)}{n} - \hat{q})}$$

Here $\Delta_n = (\frac{L_1(n)}{n})^2(1-\hat{q})^2 + 4\hat{q}(1-\hat{q})(\frac{L_1(n)}{n} - 1)$.

Observe that $f_1(EL_1; q)$ is also continuous with respect to q.

Theorem 2. *The estimator $\hat{p}_{L_1,n}$ for p is consistent.*

Proof. By the Continuous Mapping Theorem, since f_1 is continuous with respect to both arguments, $\frac{L_1(n)}{n} \xrightarrow{a.s.} E(L_1)$ and $\hat{q} \xrightarrow{a.s.} q$, we have:

$$f_1\left(\frac{L_1(n)}{n}; \hat{q}\right) \xrightarrow{a.s.} f_1(E(L_1); q) = p$$

□

3.6.2 Moments Estimator Based on L_0

Similarly, we show that f_0 exists, where $f_0 : E(L_0) \mapsto p$ is the inverse of $h_0 : p \mapsto E(L_0)$.:

Let's define $h_0(p) = E(L_0)$. We need to show that h_0 is strictly monotonic (to ensure f_0 is well-defined) and continuous.

Taking the derivative:

$$\begin{aligned}\frac{dh_0}{dp} &= \frac{\partial}{\partial p}\left(\frac{pq}{1-pq} + \frac{q}{1-q}\right) \\ &= \frac{\partial}{\partial p}\left(\frac{1}{1-pq} + \frac{q}{1-q} - 1\right) \\ &= \frac{q}{(1-pq)^2}\end{aligned}$$

Thus, $\frac{dh_0}{dp} > 0$ for all $p, q \in (0, 1)$, confirming h_0 is strictly increasing.

Since h_0 is strictly increasing, $f_0 = h_0^{-1}$ exists and is well-defined. Also, since h_0 is continuous, f_0 is continuous.

Thus we can define another moment estimator:

$$\hat{p}_{L_0,n} = f_0(\frac{L_0(n)}{n}; \hat{q})$$

Let's find f_0:

$$\begin{aligned} EL_0 &= \frac{q + pq - 2pq^2}{(1-q)(1-pq)} \\ EL_0(1-q)(1-pq) &= q + pq - 2pq^2 \\ -EL_0(1-q)qp - pq + 2pq^2 &= -EL_0(1-q) + q \\ p &= \frac{q - EL_0(1-q)}{2q^2 - q - EL_0(1-q)q} = f_0(EL_0; q) \end{aligned}$$

Thus:

$$\hat{p}_{L_0,n} = \frac{\hat{q} - (1-\hat{q})\frac{L_0(n)}{n}}{2\hat{q}^2 - \hat{q} - (1-\hat{q})\hat{q}\frac{L_0(n)}{n}}$$

Observe that $f_0(EL_0; q)$ is also continuous with respect to q.

Theorem 3. *The estimator $\hat{p}_{L_0,n}$ for p is consistent.*

Proof. By the Continuous Mapping Theorem, since f_0 is continuous with respect to both arguments, $\frac{L_0(n)}{n} \xrightarrow{a.s.} E(L_0)$ and $\hat{q} \xrightarrow{a.s.} q$, we have:

$$f_0\left(\frac{L_0(n)}{n}; \hat{q}\right) \xrightarrow{a.s.} f_0(E(L_0); q) = p$$

□

There is another way to choose function $E(L_0) \mapsto p$:

Using the representation $L_0 = Z_1 + Z_2$, with sample size of n, we have n i.i.d. random variables Z_1^i following the same distribution of Z_1. Denote $Z_1(n) = \sum_{i=1}^{n} Z_1^i$, we can simplify to:

$$\begin{aligned} \hat{p}_{L_0,n} &= g_0\left(L_0(n)/n - \frac{q}{1-q}\right) \\ &= g_0\left(\frac{Z_1(n)}{n}\right) \end{aligned}$$

where g_0 is the inverse of $p \mapsto E(Z_1) = \frac{pq}{1-pq}$.

Solving for p in terms of $E(Z_1) = e$:

$$\begin{aligned}\frac{pq}{1-pq} &= e \\ pq &= e(1-pq) \\ pq &= e - epq \\ pq(1+e) &= e \\ p &= \frac{e}{q(1+e)}\end{aligned}$$

Therefore:

$$g_0(e) = \frac{1}{q} \cdot \frac{e}{1+e}$$

And our estimator becomes:

$$\hat{p}_{L_0,n} = \frac{1}{q} \cdot \frac{Z_1(n)/n}{1+Z_1(n)/n}$$

It's not hard to see that g_0 is just a shifted function of f_0:

$$\begin{aligned}f_0(EL_0) &= \frac{q - EL_0(1-q)}{2q^2 - q - EL_0(1-q)q} \\ &= \frac{q - (EZ_1 + \frac{q}{1-q})(1-q)}{2q^2 - q - (EZ_1 + \frac{q}{1-q})(1-q)q} \\ &= \frac{-EZ_1(1-q)}{q^2 - q - EZ_1(1-q)q} \\ &= \frac{-EZ_1}{-q - EZ_1 q} \\ &= \frac{EZ_1}{q(1+EZ_1)} \\ &= g_0(EZ_1)\end{aligned}$$

3.6.3 Combined Estimator

When both L_1 and L_0 block information is available, a weighted combination can be used:

$$\hat{p}_{c,n} = w_1 \hat{p}_{L_1,n} + w_0 \hat{p}_{L_0,n}$$

The weights might be chosen inversely proportional to the variances of the individual estimators, or based on the relative observed block lengths:

$$w_1 = \frac{l_1(n)}{l_1(n) + l_0(n)}, \quad w_0 = \frac{l_0(n)}{l_1(n) + l_0(n)}$$

3.6.4 MLE Based on both blocks

Given a single realization with observed block lengths l_1 and l_0, we know that $X_{l_1} = 1$ and $X_{l_0+l_1+1} = 2$. The probability of this event is:

$$\begin{aligned} P(X_{l_1} = 1, X_{l_0+l_1+1} = 2) &= P(X_{l_1} = 1)P(X_{l_0+l_1+1} = 2|X_{l_1} = 1)f(q, l_0) \\ &= p^{l_1-1}(1 - P(X_{l_0+l_1+1} = 1|X_{l_1} = 1))f(q, l_0) \\ &= p^{l_1-1}(1 - p^{l_0+1})f(q, l_0) \end{aligned}$$

where $f(q, l_0) = q^{l_0}(1-q)^2$ is the probability of having this observation. Since it's free from p, so in MLE derivation it will be omitted(after taking logarithm and taking derivative with respect to p, it will become 0).

Note that $1 - p^{l_0+1}$ can be interpreted as the probability of length of hidden 1-s is less or equal than l_0. This will come in handy when we are trying to derive the probability of same event under constant censoring.

The maximum likelihood estimator for p is:

$$\hat{p} = \arg\max_p p^{l_1-1}(1 - p^{l_0+1})$$

Taking logarithm of the likelihood:

$$\ell_1(p) = (l_1 - 1)\ln p + \ln(1 - p^{l_0+1})$$

Taking the derivative with respect to p:

$$\frac{d\ell_1(p)}{dp} = \frac{l_1 - 1}{p} - \frac{(l_0 + 1)p^{l_0}}{1 - p^{l_0+1}}$$

And setting it to zero:

$$\begin{aligned}
\frac{l_1 - 1}{p} &= \frac{(l_0 + 1)p^{l_0}}{1 - p^{l_0+1}} \\
l_1 - 1 &= \frac{(l_0 + 1)p^{l_0+1}}{1 - p^{l_0+1}} \\
\frac{l_1 - 1}{l_0 + 1} &= \frac{p^{l_0+1}}{1 - p^{l_0+1}} \\
p^{-(l_0+1)} - 1 &= \frac{l_0 + 1}{l_1 - 1} \\
p^{-(l_0+1)} &= \frac{l_0 + l_1}{l_1 - 1} \\
\hat{p} &= \left(\frac{l_1 - 1}{l_0 + l_1}\right)^{\frac{1}{l_0+1}}
\end{aligned}$$

With a sample size n, let $l_1^1, ..., l_1^n$ and $l_0^1, ..., l_0^n$ be the observed block lengths. The MLE becomes:

$$\begin{aligned}
\hat{p}_n &= \arg\max_p \prod_{i=1}^{n} p^{l_1^i-1}(1 - p^{l_0^i+1}) \\
&= \arg\max_p \left\{ p^{\sum_{i=1}^n l_1^i - n} \prod_{i=1}^{n}(1 - p^{l_0^i+1}) \right\}
\end{aligned}$$

Taking the logarithm and differentiating:

$$\frac{d}{dp}\left\{\left(\sum_{i=1}^{n} l_1^i - n\right)\ln p + \sum_{i=1}^{n}\ln(1-p^{l_0^i+1})\right\} = 0$$
$$\frac{\sum_{i=1}^{n} l_1^i - n}{p} - \sum_{i=1}^{n}\frac{(l_0^i+1)p^{l_0^i}}{1-p^{l_0^i+1}} = 0$$

This leads to:

$$\sum_{i=1}^{n}\frac{(l_0^i+1)p^{l_0^i+1}}{1-p^{l_0^i+1}} = l_1(n) - n$$

The MLE $\hat{p}_n$ is the solution of this equation.

Note that this estimator requires no knowledge of q.

3.6.5 MLE Based on L_1

When we focus solely on the 1-block lengths in the uncensored case, we can develop a specialized maximum likelihood estimator. The distribution of L1 is given by:

$$P(L_1 = 1) = \frac{1-p}{1-pq}$$
$$P(L_1 = l) = \frac{(1-p)(1-q)p^{l-1}}{1-pq}, \text{ for } l > 1$$

Likelihood function based on L_1 blocks

$$
\begin{aligned}
L(l_1^1, \ldots, l_1^n; p) &= \prod_{i=1}^{n} P(L_1 = l_1^i) \\
&= \prod_{i:l_1^i=1} \frac{1-p}{1-pq} \cdot \prod_{i:l_1^i>1} \frac{(1-p)(1-q)p^{l_1^i-1}}{1-pq} \\
&= \left(\frac{1-p}{1-pq}\right)^{n_1} \cdot \left(\frac{(1-p)(1-q)}{1-pq}\right)^{n-n_1} \cdot \prod_{i:l_1^i>1} p^{l_1^i-1} \\
&= \left(\frac{1-p}{1-pq}\right)^{n_1} \cdot \left(\frac{(1-p)(1-q)}{1-pq}\right)^{n-n_1} \cdot p^{\sum_{i:l_1^i>1}(l_1^i-1)} \\
&= \left(\frac{1-p}{1-pq}\right)^{n} \cdot (1-q)^{n-n_1} \cdot p^{\sum_{i:l_1^i>1}(l_1^i-1)}
\end{aligned}
$$

Where n_1 is the number of L_1 blocks of length 1.

Log-likelihood function:

$$
\begin{aligned}
\ell_n(p) &= n \ln\left(\frac{1-p}{1-pq}\right) + (n-n_1)\ln(1-q) + \left(\sum_{i:l_1^i>1} (l_1^i - 1)\right) \ln p \\
&= n\ln(1-p) - n\ln(1-pq) + (n-n_1)\ln(1-q) + \left(\sum_{i:l_1^i>1} (l_1^i - 1)\right) \ln p
\end{aligned}
$$

Derivative of log-likelihood:

$$
\frac{d\ell_n(p)}{dp} = \frac{-n}{1-p} + \frac{nq}{1-pq} + \frac{\sum_{i:l_1^i>1}(l_1^i - 1)}{p}
$$

Set derivative equal to zero:

$$\begin{aligned}
\frac{-n}{1-p} + \frac{nq}{1-pq} + \frac{\sum_{i:l_1^i>1}(l_1^i - 1)}{p} &= 0 \\
\frac{-n}{1-p} + \frac{nq}{1-pq} &= -\frac{\sum_{i:l_1^i>1}(l_1^i - 1)}{p} \\
\frac{-n(1-pq) + nq(1-p)}{(1-p)(1-pq)} &= -\frac{\sum_{i:l_1^i>1}(l_1^i - 1)}{p} \\
\frac{-n + npq + nq - npq}{(1-p)(1-pq)} &= -\frac{\sum_{i:l_1^i>1}(l_1^i - 1)}{p} \\
\frac{-n + nq}{(1-p)(1-pq)} &= -\frac{\sum_{i:l_1^i>1}(l_1^i - 1)}{p} \\
\frac{n(1-q)}{(1-p)(1-pq)} &= \frac{\sum_{i:l_1^i>1}(l_1^i - 1)}{p}
\end{aligned}$$

Multiply both sides by $p(1-p)(1-pq)$

$$\begin{aligned}
np(1-q) &= (1-p)(1-pq) \sum_{i:l_1^i>1} (l_1^i - 1) \\
np(1-q) &= (1-p-pq+p^2q) \sum_{i:l_1^i>1} (l_1^i - 1) \\
np(1-q) &= (1-p-pq+p^2q)S \quad \text{where } S = \sum_{i:l_1^i>1} (l_1^i - 1)
\end{aligned}$$

Rearranging to form a quadratic equation in p

$$\begin{aligned}
np(1-q) &= (1-p-pq+p^2q)S \\
np(1-q) &= S - Sp - Spq + Sp^2q \\
Sp^2q - Spq - Sp - np(1-q) + S &= 0 \\
Sqp^2 - (Sq + S + n(1-q))p + S &= 0
\end{aligned}$$

This is a quadratic equation in p:

$$Sqp^2 - (S(1+q) + n(1-q))p + S = 0$$

Using the quadratic formula:

$$p = \frac{S(1+q) + n(1-q) \pm \sqrt{(S(1+q) + n(1-q))^2 - 4S^2q}}{2Sq}$$

Still, let take $S \to \infty$:

$$\begin{aligned}
&\lim_{S\to\infty} \frac{S(1+q) + n(1-q) \pm \sqrt{(S(1+q) + n(1-q))^2 - 4S^2q}}{2Sq} \\
&= \lim_{S\to\infty} \frac{S(1+q) \pm \sqrt{(S(1+q))^2 - 4S^2q}}{2Sq} \\
&= \lim_{S\to\infty} \frac{S(1+q) \pm \sqrt{(S(1-q))^2}}{2Sq} \\
&= \lim_{S\to\infty} \frac{S(1+q) \pm \sqrt{(S(1-q))^2}}{2Sq} \\
&= \lim_{S\to\infty} \frac{(1+q) \pm (1-q)}{2q}
\end{aligned}$$

Like the argument in moment estimation based on L_1, we should take minus:

$$\hat{p}_{1,n} = \frac{S_n(1+q) + n(1-q) - \sqrt{(S_n(1+q) + n(1-q))^2 - 4S_n^2q}}{2S_nq}$$

where $S_n = \sum_{i=1}^{n} I(L_1^i > 1) \cdot (L_1^i - 1)$.

In practice, numerical optimization is employed to find the MLE, ensuring that the constraint $p \in (0,1)$ is satisfied.

3.6.6 MLE Based on L_0

We can also develop a maximum likelihood estimator based solely on the 0-block lengths. The distribution of L_0 in the uncensored case is:

$$P(L_0 = l) = \frac{(1-q)(1-qp)}{(1-p)} \left(q^l - p(qp)^l\right)$$

Given observed 0-block lengths $l_0^1, \ldots, l_0^n$, the likelihood function is:

$$\begin{aligned} L(l_0^1, \ldots, l_0^n; p) &= \prod_{i=1}^{n} P(L_0 = l_0^i) \\ &= \left[\frac{(1-q)(1-qp)}{(1-p)}\right]^n \prod_{i=1}^{n} \left(q^{l_0^i} - p(qp)^{l_0^i}\right) \end{aligned}$$

Taking the logarithm:

$$\begin{aligned} \ell_n(p) &= n \ln\left[\frac{(1-q)(1-qp)}{(1-p)}\right] + \sum_{i=1}^{n} \ln\left(q^{l_0^i} - p(qp)^{l_0^i}\right) \\ &= n\ln(1-q) + n\ln(1-qp) - n\ln(1-p) + \sum_{i=1}^{n} \ln\left(q^{l_0^i} - p(qp)^{l_0^i}\right) \\ &= n\ln(1-q) + n\ln(1-qp) - n\ln(1-p) + \sum_{i=1}^{n} \ln\left(q^{l_0^i}(1 - p^{l_0^i+1})\right) \\ &= n\ln(1-q) + n\ln(1-qp) - n\ln(1-p) + \sum_{i=1}^{n} l_0^i \ln q + \sum_{i=1}^{n} \ln(1 - p^{l_0^i+1}) \end{aligned}$$

Taking the derivative with respect to p:

$$\frac{d\ell_n(p)}{dp} = \frac{-nq}{1-qp} + \frac{n}{1-p} - \sum_{i=1}^{n} \frac{(l_0^i+1)p^{l_0^i}}{1-p^{l_0^i+1}}$$

Setting this equal to zero:

$$\frac{-nq}{1-qp} + \frac{n}{1-p} - \sum_{i=1}^{n} \frac{(l_0^i+1)p^{l_0^i}}{1-p^{l_0^i+1}} = 0$$

Rearranging:

$$\begin{aligned}
\frac{n}{1-p} - \frac{nq}{1-qp} &= \sum_{i=1}^{n} \frac{(l_0^i+1)p^{l_0^i}}{1-p^{l_0^i+1}} \\
\frac{n(1-qp) - nq(1-p)}{(1-p)(1-qp)} &= \sum_{i=1}^{n} \frac{(l_0^i+1)p^{l_0^i}}{1-p^{l_0^i+1}} \\
\frac{n - nqp - nq + nqp}{(1-p)(1-qp)} &= \sum_{i=1}^{n} \frac{(l_0^i+1)p^{l_0^i}}{1-p^{l_0^i+1}} \\
\frac{n(1-q)}{(1-p)(1-qp)} &= \sum_{i=1}^{n} \frac{(l_0^i+1)p^{l_0^i}}{1-p^{l_0^i+1}}
\end{aligned}$$

This equation defines the MLE $\hat{p}_{0,n}$ but does not yield a closed-form solution. The complexity arises from the sum on the right-hand side, which involves rational functions of p raised to different powers. Still it can be obtained by numeric method.

4 Censoring in Two-State Hidden Markov Chain

4.1 Distributions Under Fixed Censoring

In practical applications, observations are often limited to a fixed maximum length. We define this as *fixed censoring* with censoring parameter m, where we observe the process only up to time m. Under this constraint, we define L_1^m as the observed length of 1-blocks and L_0^m as the observed length of 0-blocks.

Note that $m \geq 2$, otherwise it's meaningless.

It is important to note that L_1^m is not simply the minimum of L_1 and m (i.e., $L_1 \wedge m$) since censoring occurs prior to state recovery(modification) in our model.

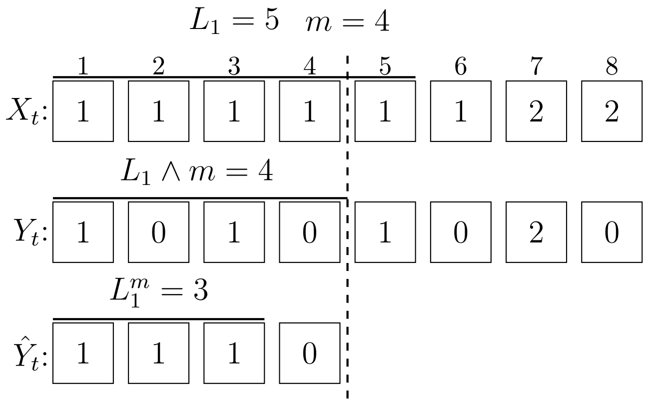


Figure 8: Example showing why $L_1^m \neq L_1 \wedge m$ with censoring limit $m = 4$. Without censoring the recovered 1-block has length $L_1 = 5$, but in reality we can only recover 3 out of 5. Thus L_1^m is totally different from $L_1 \wedge m$.

4.1.1 Distribution of L_1^m

Use same X and Y from the discussion of L_1, recall that $X \sim G_0(1-p), Y \sim G_0(1-q)$. The distribution of L_1^m can be derived as follows:

$$\begin{aligned}
P(L_1^m = 1) &= P(X \wedge (m-1) - X \wedge Y \wedge (m-1) = 0) \\
&= P(X \wedge (m-1) = X \wedge Y \wedge (m-1)) \\
&= P(Y \geq X \wedge (m-1)) \\
&= \sum_{k=0}^{m-1} P(Y \geq k) P(X \wedge (m-1) = k) \\
&= \sum_{k=0}^{m-2} q^k P(X = k) + q^{m-1} P(X \geq m-1) \\
&= \sum_{k=0}^{m-2} q^k p^k (1-p) + (pq)^{m-1} \\
&= \frac{1-p}{1-pq}(1-(pq)^{m-1}) + (pq)^{m-1}
\end{aligned}$$

For $m \geq l > 1$, we have:

$$\begin{aligned}
P(L_1^m = l) &= P(X \wedge (m-1) - X \wedge Y \wedge (m-1) = l-1) \\
&= P(X - Y = l-1, X < m-1) + P(X - Y = l-1 | X = m-1) P(X \geq m-1) \\
&= \sum_{k=l-1}^{m-2} P(Y = k-l+1) P(X = k) + P(Y = m-l) P(X \geq m-1) \\
&= \sum_{k=l-1}^{m-2} q^{k-l+1}(1-q) p^k (1-p) + q^{m-l}(1-q) p^{m-1} \\
&= (1-q)(1-p) q^{-l+1} \sum_{k=l-1}^{m-2} q^k p^k + q^{m-l}(1-q) p^{m-1} \\
&= (1-q)(1-p) q^{-l+1} \frac{(pq)^{l-1} - (pq)^{m-1}}{1-pq} + q^{m-l}(1-q) p^{m-1} \\
&= (1-q) q^{-l} p^{-1} \left[(1-p) \frac{(pq)^l - (pq)^m}{1-qp} + (pq)^m \right]
\end{aligned}$$

As $m \to \infty$, this distribution converges to the uncensored case:

$$\lim_{m\to\infty} P(L_1^m = 1) = \frac{1-p}{1-pq}$$
$$\lim_{m\to\infty} P(L_1^m = l) = \frac{(1-p)(1-q)p^{l-1}}{1-pq}, \quad l > 1$$

4.1.2 Distribution of L_0^m

For the distribution of L_0^m, we define the following events:

$$A_{t,k} := \{Y_t = 1, Y_{t+1} = \ldots = Y_{t+k} = 0, Y_{t+k+1} = 2\},$$
$$t = 1, \ldots, m-k-1,\ k = 0, \ldots, m-2.$$
$$B_k := \{Y_{m-k} = 1, Y_{m-k+1} = \ldots = Y_m = 0\}, \quad k = 0, \ldots, m-1$$

Here k is the length of 0-block, $A_{t,k}$ stands for the cases when state 2 is observed under the fixed censoring, while B_k stands for the cases when state 2 is not observed. Then, the probability of observing a 0-block of length k is:

$$P(L_0^m = k) = \begin{cases} \sum_{t=1}^{m-k-1} P(A_{t,k}) + P(B_k) & k = 0, 1, ..., m-2 \\ P(B_k) & k = m-1 \end{cases}$$

We compute these probabilities as follows:

$$P(A_{t,k}) = \begin{cases} r_k, & \text{when } t = 1 \\ p^{t-1}(1-q)r_k, & \text{when } t = 2, \ldots, m-k-1 \end{cases}$$
$$P(B_k) = \begin{cases} q^{m-1}, & \text{when } k = m-1 \\ p^{m-k-1}(1-q)q^k, & \text{when } k = 0, \ldots, m-2 \end{cases}$$

where $r_k = P(Y_{t+1} = \ldots = Y_{t+k} = 0, Y_{t+k+1} = 2 | Y_t = 1) = (1-p^{k+1})q^k(1-q)$.

First, let's evaluate $\sum_{t=1}^{m-k-1} P(A_{t,k})$:

$$\begin{aligned}
\sum_{t=1}^{m-k-1} P(A_{t,k}) &= r_k + \sum_{t=2}^{m-k-1} p^{t-1}(1-q)r_k \\
&= r_k + r_k(1-q)\sum_{t=2}^{m-k-1} p^{t-1} \\
&= r_k + r_k(1-q)\cdot p \cdot \sum_{j=0}^{m-k-3} p^j \quad \text{(where } j = t-2) \\
&= r_k + r_k(1-q)\cdot p \cdot \frac{1-p^{m-k-2}}{1-p} \\
&= r_k \left(1 + (1-q)\cdot p \cdot \frac{1-p^{m-k-2}}{1-p}\right)
\end{aligned}$$

Thus:

$$P(L_0^m = k) = \begin{cases} q^{m-1}, & \text{when } k = m-1 \\ r_k \left(1 + (1-q)\cdot p \cdot \frac{1-p^{m-k-2}}{1-p}\right) + p^{m-k-1}(1-q)q^k, & \text{when } k = 0, \ldots, m-2 \end{cases}$$

As $m \to \infty$, these probabilities converge to those of the uncensored case:

$$\lim_{m\to\infty} P(L_0^m = k) = \frac{(q^k - p^{k+1}q^k)(1-q)(1-qp)}{1-p}$$

4.2 Expected Values Under Fixed Censoring

Generally, we can use probability distribution to obtain expectations. Given the complexity of probabilities of L_1^m and L_0^m, here we will only use decomposition approaches.

4.2.1 Calculation of EL_1^m Using Decomposition Approach

In section 2, for a random variable $K \sim G_1(1-p)$, we have $E(K \wedge m) = \frac{1-p^m}{1-p}$, similarly for random variable $X \sim G_0(1-p)$:

$$E(X \wedge m) = 0 \cdot P(X=0) + pE(K \wedge m) = \frac{p - p^{m+1}}{1-p}$$

Hence, the expected value of L_1^m can be expressed as:

$$\begin{aligned} E(L_1^m) &= E(X \wedge (m-1)) - E(X \wedge Y \wedge (m-1) + 1 \\ &= \frac{p - p^m}{1-p} - \frac{pq - (pq)^m}{1-pq} + 1 \\ &= \frac{1 - p^m}{1-p} - \frac{1 - (pq)^m}{1-pq} + 1 \end{aligned}$$

This can be related to the uncensored expectation:

$$E(L_1^m) = E(L_1) - \frac{p^m}{1-p} + \frac{(pq)^m}{1-pq}$$

4.2.2 Calculation of EL_0^m Using Decomposition Approach

For the expected length of the zero-block under censoring (L_0^m), we use the decomposition approach where L_0^m is viewed as the sum of two components: Z_1 (zeros in the state 1 region) and Z_2 (zeros in the state 2 region).

Consider a realization of length m where the transition from state 1 to state 2 occurs at position $m-k$ (i.e., $X_{m-k} = 1$ and $X_{m-k+1} = 2$):

$$1 \ \underbrace{1 \ldots 1}_{X=m-k-1} \ \underbrace{2 \ldots 2}_{k}$$

In this censoring context, k refers to the length of real 2-block. There are two separate cases for $k : k > 0$ and $k = 0$, indicating whether Z_2 exists or not. Recall that $X \sim G_0(1-p)$ and $Y \sim G_0(1-q)$, and $EX \wedge m = \frac{p-p^{m+1}}{1-p}$, similar for EY.

Let consider $k > 0$ first, :

$$\begin{aligned} E[Z_1|X = m-k-1] &= E[Y \wedge (m-k-1)] = \frac{q-q^{m-k}}{1-q} \\ E[Z_2|X = m-k-1] &= E[Y \wedge k] = \frac{q-q^{k+1}}{1-q} \end{aligned}$$

Thus, the conditional expectation of L_0^m given $X = m-k-1$ is:

$$\begin{aligned} E[L_0^m|X = m-k-1] &= E[Z_1|X = m-k-1] + E[Z_2|X = m-k-1] \\ &= \frac{2q - q^{k+1} - q^{m-k}}{1-q}, \quad k = 1, \ldots, m-1 \end{aligned}$$

When $X \geq m-1$ (i.e. $k = 0$, the transition doesn't occur within the first m positions):

$$E[L_0^m|X \geq m-1] = E[Z_1|X \geq m-1] = E[Y \wedge m-1] = \frac{q-q^m}{1-q}$$

The unconditional expectation is:

$$\begin{aligned} E(L_0^m) &= \sum_{k=1}^{m-1} E[L_0^m|X = m-k-1]P(X = m-k-1) + E[L_0^m|X \geq m-1]P(X \geq m-1) \\ &= \sum_{k=1}^{m-1} \frac{2q - q^{k+1} - q^{m-k}}{1-q}(1-p)p^{m-k-1} + \frac{q-q^m}{1-q}p^{m-1} \end{aligned}$$

Simplifying the sums:

$$\begin{aligned}
\frac{1-p}{1-q}\sum_{k=1}^{m-1}(2q)p^{m-k-1} &= \frac{1-p}{1-q}(2q)\sum_{k=0}^{m-2}p^k = \frac{2q(1-p^{m-1})}{1-q} \\
\frac{1-p}{1-q}\sum_{k=1}^{m-1}q^{k+1}p^{m-(k+1)} &= \frac{1-p}{1-q}\cdot p^m\sum_{k=1}^{m-1}\left(\frac{q}{p}\right)^{k+1} \\
&= \frac{1-p}{1-q}\cdot\frac{p^m((\frac{q}{p})^2-(\frac{q}{p})^m)}{1-\frac{q}{p}} \\
&= \frac{1-p}{1-q}\cdot\frac{p^m(q^2-\frac{q^m}{p^{m-2}})}{p(p-q)} \\
&= \frac{1-p}{1-q}\cdot\frac{p^2q^2(p^{m-2}-q^{m-2})}{p(p-q)} \\
&= \frac{1-p}{1-q}\cdot\frac{pq^2(p^{m-2}-q^{m-2})}{p-q}
\end{aligned}$$

Similarly:

$$\begin{aligned}
\frac{1-p}{1-q}\sum_{k=1}^{m-1}q^{m-k}p^{m-k-1} &= \frac{1-p}{1-q}p^{-1}\sum_{k=1}^{m-1}(pq)^{m-k} \\
&= \frac{1-p}{1-q}p^{-1}\sum_{k=1}^{m-1}(pq)^k \\
&= \frac{1-p}{1-q}\cdot\frac{q(1-(pq)^{m-1})}{1-pq}
\end{aligned}$$

Combining all terms:

$$
\begin{aligned}
E(L_0^m) &= \frac{2q(1-p^{m-1})}{1-q} - \frac{1-p}{1-q} \cdot \frac{pq^2(p^{m-2}-q^{m-2})}{p-q} \\
&\quad - \frac{1-p}{1-q} \cdot \frac{q(1-(pq)^{m-1})}{1-pq} + \frac{q-q^m}{1-q} p^{m-1} \\
&= \frac{2q(1-p^{m-1}) + qp^{m-1} - q^m p^{m-1}}{1-q} \\
&\quad - \frac{1-p}{1-q} \left[\frac{pq^2(p^{m-2}-q^{m-2})}{p-q} + \frac{q(1-(pq)^{m-1})}{1-pq} \right] \\
&= q\frac{2-p^{m-1}-(pq)^{m-1}}{1-q} \\
&\quad - \frac{q-pq}{1-q} \left[\frac{pq(p^{m-2}-q^{m-2})}{p-q} + \frac{1-(pq)^{m-1}}{1-pq} \right]
\end{aligned}
$$

When $m \to \infty$, we recover the uncensored result:

$$
\begin{aligned}
\lim_{m\to\infty} E(L_0^m) &= \frac{2q}{1-q} - \frac{q-pq}{(1-q)(1-pq)} \\
&= \frac{q}{1-q} + \frac{q(1-pq)}{(1-q)(1-pq)} - \frac{q-pq}{(1-q)(1-pq)} \\
&= \frac{q}{1-q} + \frac{q-pq^2-q+pq}{(1-q)(1-pq)} \\
&= \frac{q}{1-q} + \frac{pq}{1-pq} = E(L_0)
\end{aligned}
$$

4.3 Method of Moments Estimators Under Fixed Censoring

4.3.1 Estimator Based on L_1^m

To develop a method of moments estimator based on L_1^m, we first establish that the function $p \mapsto e_1^m(p) = E(L_1^m)$ is strictly increasing for any $m > 1$ and $q \in (0, 1)$.

Taking the derivative:

$$\begin{aligned}\frac{d}{dp}e_1^m(p) &= \frac{d}{dp}\left(\frac{1-p^m}{1-p} - \frac{1-(pq)^m}{1-pq} + 1\right) \\ &= \frac{d}{dp}\left(\frac{1-p^m}{1-p} - \frac{1-(pq)^m}{1-pq}\right) \\ &= \frac{d}{dp}\left(\sum_{k=0}^{m} p^k - \sum_{k=0}^{m}(pq)^k\right) \\ &= \sum_{k=1}^{m} kp^{k-1} - \sum_{k=1}^{m} kq^k p^{k-1} \\ &= \sum_{k=1}^{m} kp^{k-1}(1-q^k)\end{aligned}$$

This derivative is positive for all $p \in (0,1)$, confirming that $e_1^m(p)$ is strictly increasing. Therefore, the inverse function f_1^m exists and is continuous with respect to both arguments. Like what we did in uncensored case, we can define our consistent estimator as:

$$\hat{p}_{L_1^m,n} = f_1^m\left(\frac{L_1^m(n)}{n};\hat{q}\right)$$

where $L_1^m(n)$ is the sum of n i.i.d. random variables: $L_1^m(n) = \sum_{i=1}^n L_{1,i}^m$, $L_{1,i}^m$ stands for the length of 1-block under constant censoring m in the i-th trial. The notations for realizations can stay the same.

For general m, numerical methods are typically required to compute f_1^m.

4.3.2 Estimator Based on L_0^m

Due to the complicated form of EL_0^m, the estimator is defined via numeric method, define $EL_0^m = g(p)$, then with obtained observations, $\hat{p}_{L_0^m,n}$ is the root of:

$$g(p) - \frac{l_0(n)}{n}$$

The consistency is also validated through simulations.

4.3.3 Combined Estimator

When both L_1^m and L_0^m block information is available, a weighted combination can be used:

$$\hat{p}_{c,n} = w_1 \hat{p}_{L_1^m,n} + w_0 \hat{p}_{L_0^m,n}$$

The weights might be chosen inversely proportional to the variances of the individual estimators, or based on the relative observed block lengths:

$$w_1 = \frac{l_1(n)}{l_1(n) + l_0(n)}, \quad w_0 = \frac{l_0(n)}{l_1(n) + l_0(n)}$$

4.4 MLE Under Fixed Censoring

Given a single realization with observed block lengths l_1 and l_0, we know that $X_{l_1} = 1$ and $X_{l_0+l_1+1} = 2$. Unlike the case when it's uncensored, when $l_1 + l_0 = m$ we have no idea how many hidden 1-s are there, so the probability of this event is:

$$P(l_1, l_0) = \begin{cases} p^{l_1-1}(1 - p^{l_0+1}) f(q, l_0) & if\ l_1 + l_0 < m \\ p^{l_1-1} g(q, l_0) & if\ l_1 + l_0 = m \end{cases}$$

where $f(q, l_0) = q^{l_0}(1 - q)^2$ and $g(q, l_0) = q^{l_0}(1 - q)$ are the probabilities of having this certain observation. Since they are free from p, so in MLE derivation they will be omitted.

With a sample size n, and $l_1^1, ..., l_1^n$ and $l_0^1, ..., l_0^n$ be the observed block lengths. The

MLE becomes:

$$\begin{aligned}\hat{p}_n &= \arg\max_p \prod_{i:l_1+l_0<m}^{n} p^{l_1^i-1}(1-p^{l_0^i+1}) \cdot \prod_{i:l_1+l_0=m}^{n} p^{l_1^i-1} \\ &= \arg\max_p \left\{ p^{\sum_{i=1}^n l_1^i - n} \prod_{i:l_1+l_0<m}^{n} (1-p^{l_0^i+1}) \right\}\end{aligned}$$

Taking the logarithm and differentiating:

$$\begin{aligned}\frac{d}{dp}\left\{ \left(\sum_{i=1}^{n} l_1^i - n\right)\ln p + \sum_{i:l_1+l_0<m}^{n} \ln(1-p^{l_0^i+1}) \right\} &= 0 \\ \frac{\sum_{i=1}^n l_1^i - n}{p} - \sum_{i:l_1+l_0<m}^{n} \frac{(l_0^i+1)p^{l_0^i}}{1-p^{l_0^i+1}} &= 0\end{aligned}$$

This leads to:

$$\sum_{i:l_1+l_0<m}^{n} \frac{(l_0^i+1)p^{l_0^i+1}}{1-p^{l_0^i+1}} = l_1(n) - n$$

The MLE $\hat{p}_n$ is the solution to this equation.

4.5 Random Censoring

Once we obtain the expectations for those two blocks under constant censoring, we can also obtain the expectations under random censoring.

Let $M \sim Q$, Q is some specific positive discrete probability measure, then we have formula:

$$EL^M = \sum_{m=2}^{\infty} EL^m \cdot Q(m)$$

here L could be either L_0 or L_1.

Note that m starts with 2.

For better understanding, let's use Poisson distribution as an example, which is also used in simulations. Since standard Poisson distribution starts with 0, so I shifted it by 2, thus $M \sim Poi(\lambda) + 2$:

$$P(M = m) = \frac{\lambda^{m-2}}{(m-2)!} e^{-\lambda}, m \geq 2$$

hence:

$$\begin{aligned}
EL_1^M &= \sum_{m=2}^{\infty} \left(E(L_1) - \frac{p^m}{1-p} + \frac{(pq)^m}{1-pq} \right) \frac{\lambda^{m-2}}{(m-2)!} e^{-\lambda} \\
&= \sum_{m=0}^{\infty} \left(\frac{-p^{m+2}}{1-p} + \frac{(pq)^{m+2}}{1-pq} \right) \frac{\lambda^m}{m!} e^{-\lambda} + \sum_{m=0}^{\infty} E(L_1) \frac{\lambda^m}{m!} e^{-\lambda} \\
&= \sum_{m=0}^{\infty} \left(\frac{-p^{m+2}}{1-p} \right) \frac{\lambda^m}{m!} e^{-\lambda} + \sum_{m=0}^{\infty} \left(\frac{(pq)^{m+2}}{1-pq} + 1 \right) \frac{\lambda^m}{m!} e^{-\lambda} + E(L_1) \\
&= \sum_{m=0}^{\infty} \left(\frac{-p^2}{1-p} \right) \frac{(p\lambda)^m}{m!} e^{-p\lambda} e^{-\lambda + p\lambda} + \sum_{m=0}^{\infty} \left(\frac{(pq)^{m+2}}{1-pq} + 1 \right) \frac{\lambda^m}{m!} e^{-\lambda} + E(L_1) \\
&= \frac{-p^2}{1-p} e^{-\lambda + p\lambda} + \frac{(pq)^2}{1-pq} e^{-\lambda + pq\lambda} + E(L_1)
\end{aligned}$$

Different combinations of Q and EL^m could create rather complicated results that is impossible to simplify, so generally numeric method is applied. To handle infinite sum, we need to keep adding terms until the absolute value of the next term is small enough.

Now with the expectation under random censoring, and observations of lengths of those blocks, we can use numeric method again to estimate p. The consistency is again validated through simulations.

The estimator for p in this case can be defined as:

$\hat{p}_{L^M,n}$ is the root of

$$EL^M - \frac{l(n)}{n}$$

where $l(n)$ can be $l_1(n)$ or $l_0(0)$ according to L. Note that EL^M is a function of p.

5 Simulation Results for Two-State Hidden Markov Chain

5.1 Simulation Setup

To validate our theoretical results and examine the performance of the proposed estimators, we conducted extensive simulation studies using different parameter configurations:

- **For expectation validation:** Fixed $p = 0.7$ with $q = 0.3$ are chosen to verify our formulae
- **For consistency validation:** 4 pairs of p, q are used to evaluate estimator performance under various censoring scenarios, $(p, q) \in \{(0.3, 0.3), (0.3, 0.7), (0.7, 0.3), (0.7, 0.7)\}$

We implemented the estimators for p we have obtained:

- **Moment estimator based on** L_1 ($\hat{p}_{L_1,n}$): Utilizes only the observed 1-block lengths
- **Moment estimator based on** L_0 ($\hat{p}_{L_0,n}$): Utilizes only the observed 0-block lengths
- **Combined moment estimator** ($\hat{p}_{c,n}$): Weighted combination of L_1 and L_0 estimators, with weights proportional to observed block lengths
- **Maximum likelihood estimators** (MLE): Based on the distribution of each or both block types

For the censoring mechanisms, we examined:

- **No censoring:** Complete observation until absorption (all estimators implemented)

- **Fixed censoring:** Constant censoring time $m = 10$ (implemented $\hat{p}_{L_1^m,n}$, $\hat{p}_{L_0^m,n}$, $\hat{p}_{c,n}$, and MLE based on both blocks)

- **Random censoring:** Censoring time following a shifted Poisson distribution $M \sim \text{Pois}(\lambda = 10) + 2$ (implemented $\hat{p}_{L_1^M,n}$, $\hat{p}_{L_0^M,n}$ and $\hat{p}_{c,n}$)

We also investigated scenarios where q is unknown where it must be estimated from the data, using the proportion of zeros in the original observation process before reconstruction.

5.2 Results

5.2.1 Expectation Validation

Figure 9, Figure 10 and Figure 11 shows us clearly that with or without censoring, the observed sample averages converge to the expectations calculated by our formulas for both L_1 and L_0, and the absolute difference between observed and theoretical values is decreasing as samples size grows.

5.2.2 Consistency Validation

To evaluate the consistency of the proposed estimators for p, we conducted simulations using four different pairs of (p, q) values. These combinations were chosen to investigate the influences of both small and large values of p and q on the performance of the estimators.

The expected observations for each (p, q) pair are as follows:

- For small p and q (e.g., $(0.3, 0.3)$), the 1-blocks are rare, and fewer zeros occur.

- When p is small and q is large (e.g., $(0.3, 0.7)$), the 1-blocks are rare because transitions to state 2 occur quickly, and more zeros are observed.

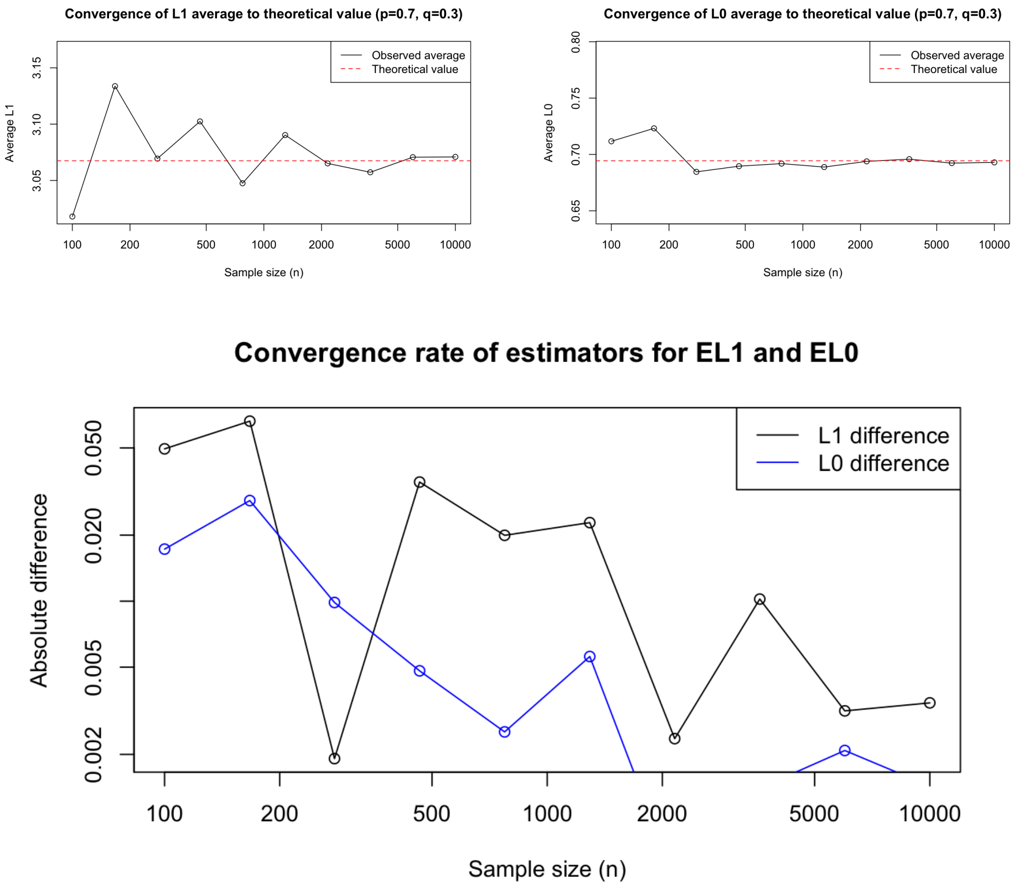


Figure 9: Expectation under no censoring

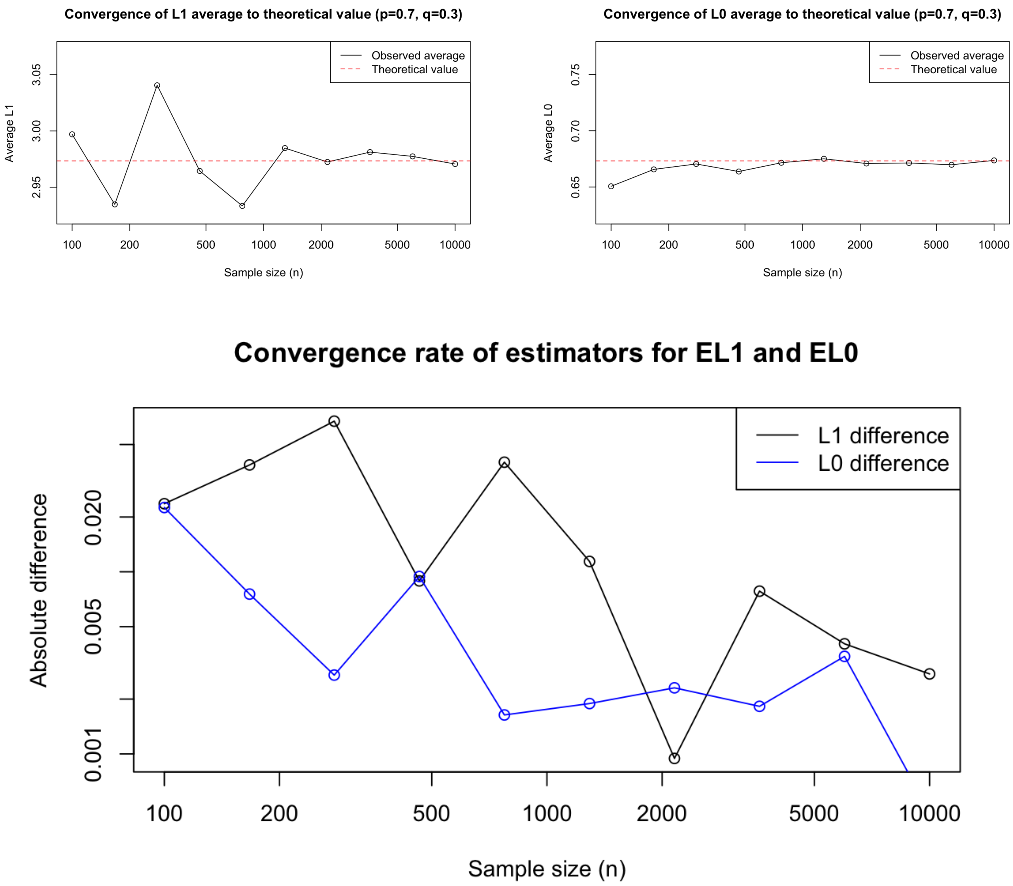


Figure 10: Expectation under fixed censoring

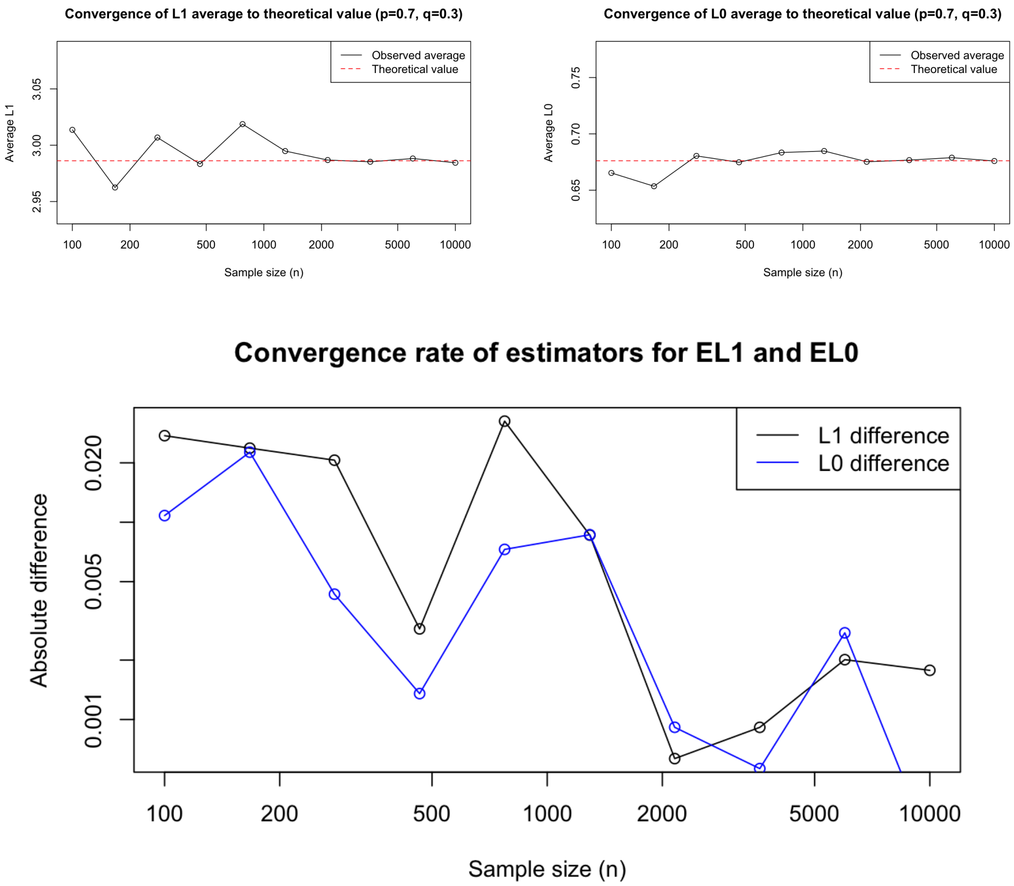


Figure 11: Expectation under random censoring

- When p is large and q is small (e.g., $(0.7, 0.3)$), the 1-blocks are frequent because the chain stays in state 1 longer, and fewer zeros are observed since most states are directly observed.
- For large p and q (e.g., $(0.7, 0.7)$), the 1-blocks are frequent, but more zeros are observed.

Figures 12, 13, and 14 show the convergence behavior of the estimators under no censoring, fixed censoring, and random censoring, respectively.

For each scenario, we observe that:

- Almost all estimators converge to the true value of p as the sample size n increases, except for fixed censoring with pair of (0.3,0.7).
- The convergence rate is influenced by the values of p and q. When p is small and q is large, fewer transitions are observed, leading to slower convergence due to reduced information in the data. This aligns with the graphs, where larger variances are observed for these parameter combinations, particularly under censoring.
- The moment estimator based on L_1 ($\hat{p}_{L_1,n}$) and the combined estimator ($\hat{p}_{c,n}$) consistently show faster convergence compared to the estimator based on L_0 ($\hat{p}_{L_0,n}$), especially when q is small.

For no censoring (Figure 12), all estimators perform well, with rapid convergence to the true parameter values. Fixed censoring (Figure 13) introduces some bias for smaller sample sizes, particularly in cases where p is small and q is large, we see it did not converge under this sample size, but it converged under random censoring. Random censoring (Figure 14) exhibits similar behavior to fixed censoring, with slightly slower convergence due to the additional randomness in the censoring mechanism.

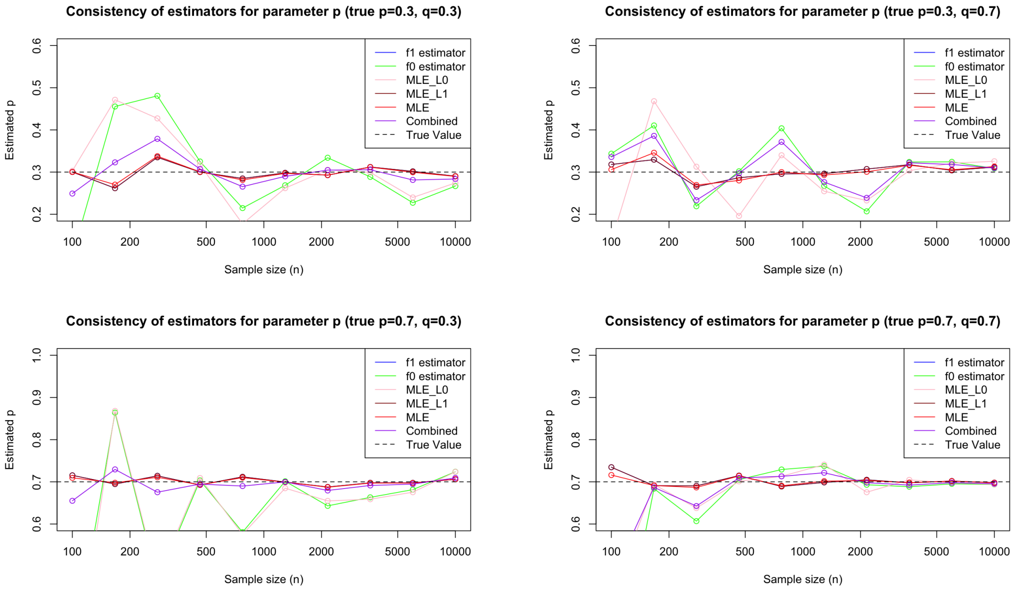


Figure 12: Consistency under no censoring

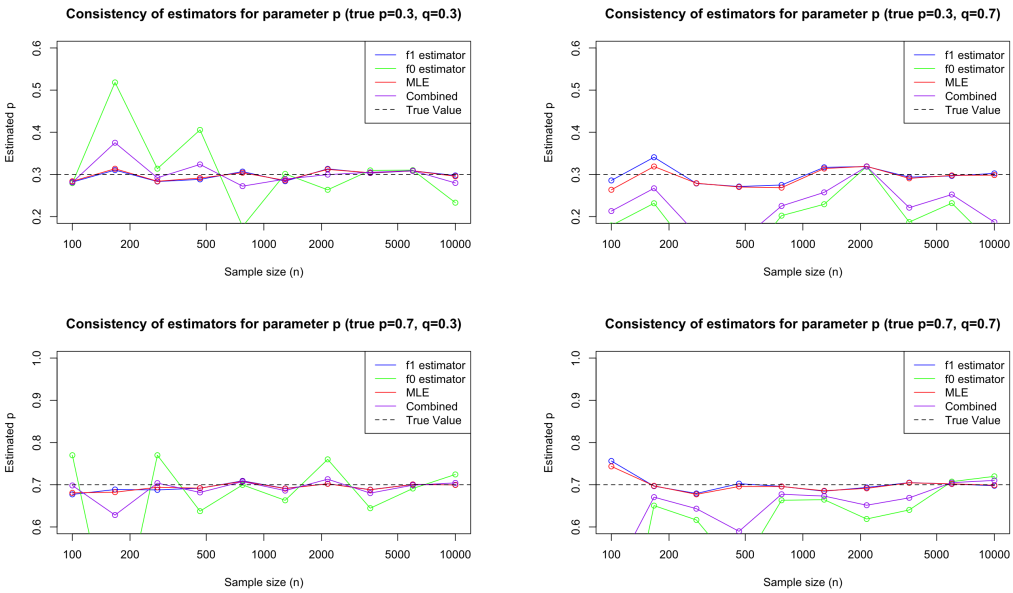


Figure 13: Consistency under fixed censoring

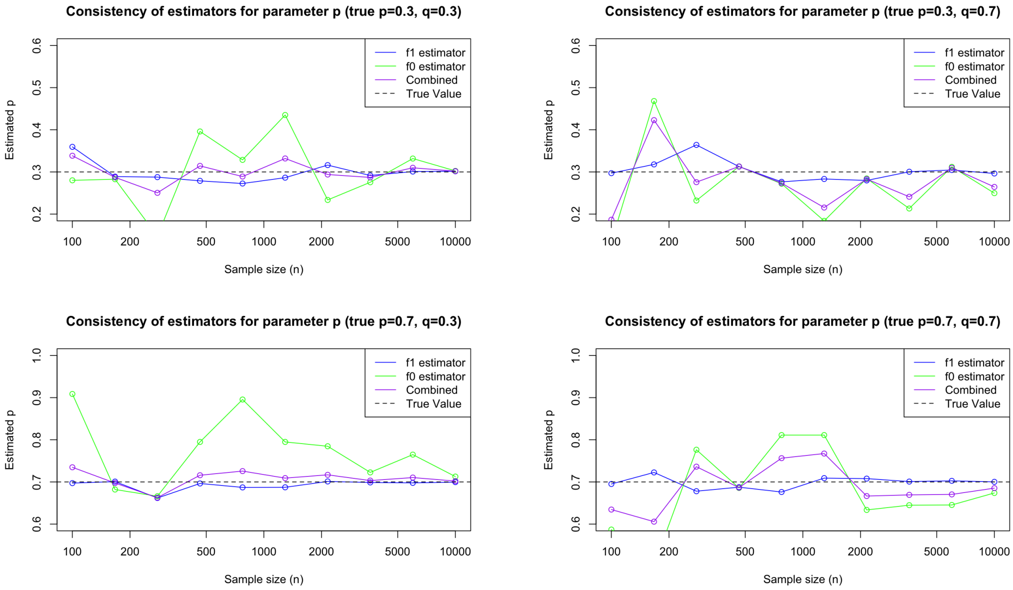


Figure 14: Consistency under random censoring

5.2.3 Consistency Validation with Estimated q

When the true value of q is unknown, we estimate it using the proportion of zeros in the original observation process before reconstruction. This introduces additional variability into the estimators, which we evaluated using the same four pairs of (p, q) values.

The results(Figure 15, Figure 16 and Figure 17) do not show substantial differences from the scenarios where q is known, confirming that the proposed estimators are robust to the estimation of q, maintaining their consistency under various censoring scenarios and parameter configurations. The graphs align with the theoretical expectations, showing slower convergence and larger variances for challenging parameter combinations such as small p and large q.

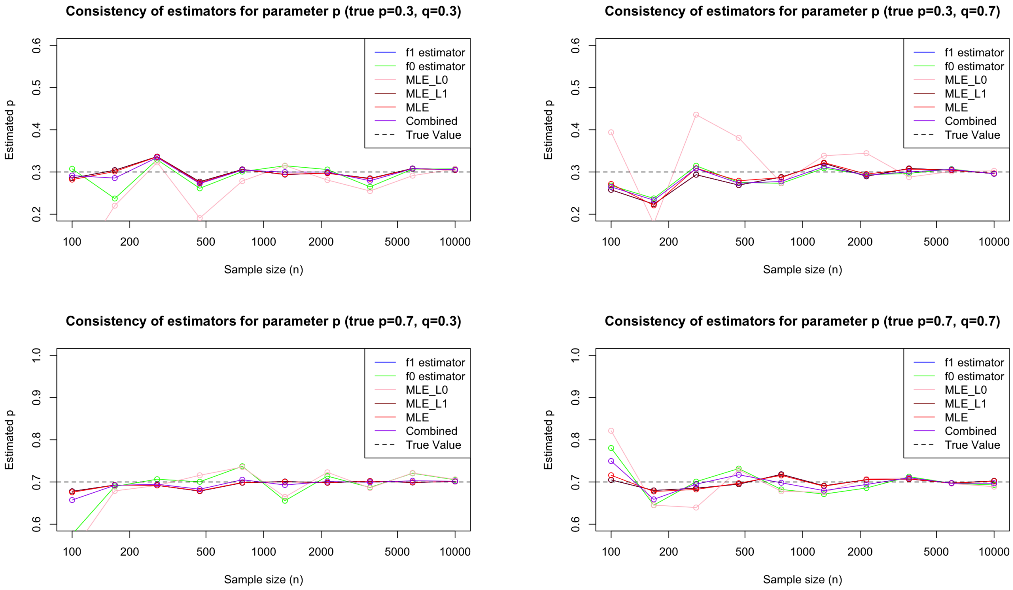


Figure 15: Consistency under no censoring with estimated q

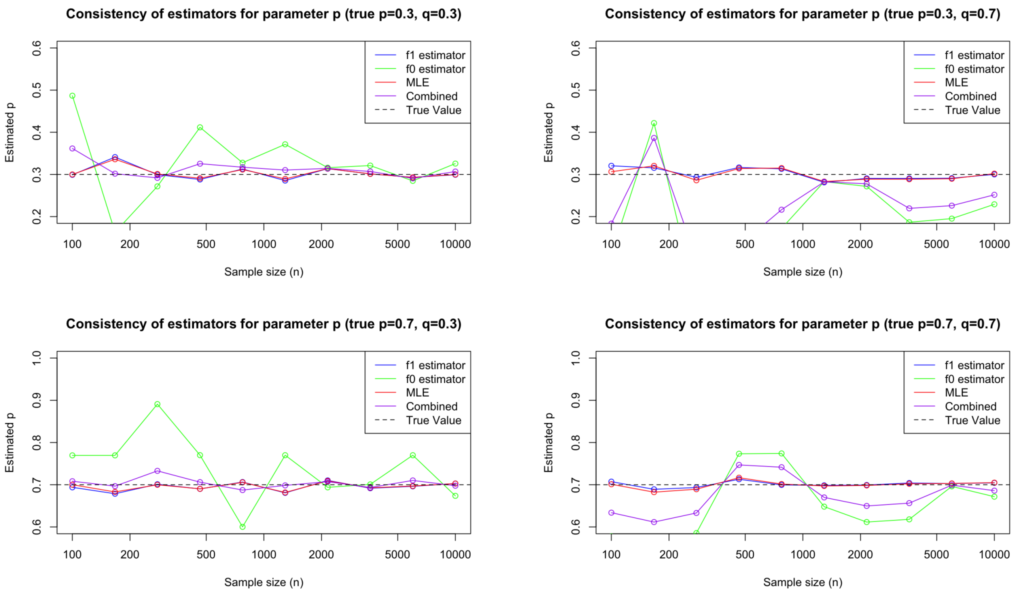


Figure 16: Consistency under fixed censoring with estimated q

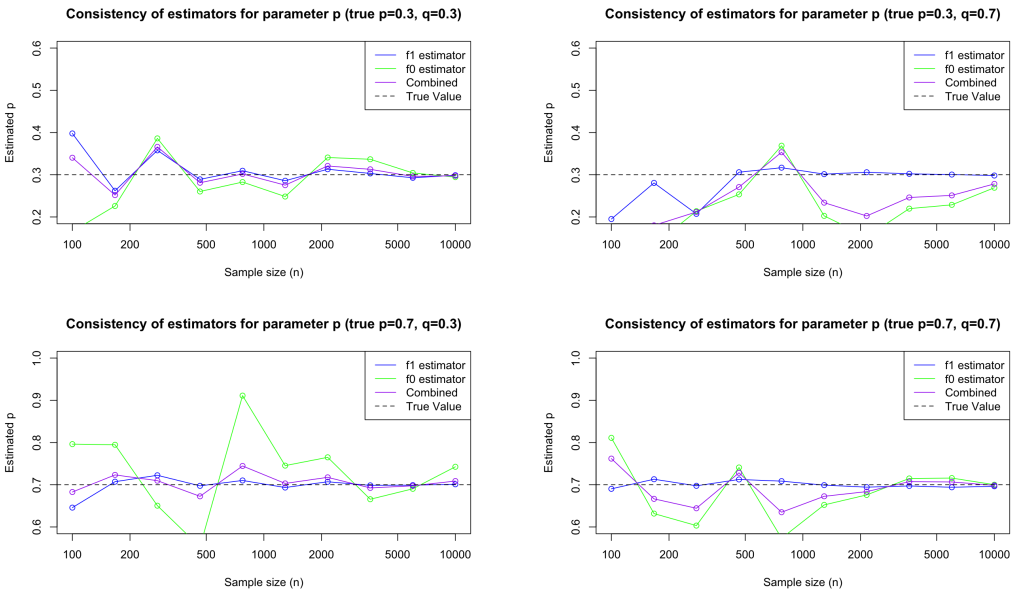


Figure 17: Consistency under random censoring with estimated q

Conclusions

This thesis has addressed parameter estimation in two types of progressive multistate models: normal progressive multistate models with fully observable states and progressive hidden multistate models where the underlying states are only partially observed. The main findings are summarized as follows:

Normal Progressive Multistate Models

In the first part of the thesis, we studied progressive, time-homogeneous Markov models, which are widely used to analyze systems where transitions occur only to higher-numbered states, and the final state is absorbing. The key contributions include:

- Derivation of maximum likelihood estimator (MLE) for all transition probabilities under uncensored, fixed-censored, and random-censored scenarios.
- Proof of consistency for the MLE, ensuring that they converge to the true parameter values as the sample size increases.
- Validation of the proposed MLE through simulation studies, which confirmed the consistency under various censoring mechanisms and parameter configurations.

Two-state Progressive Hidden Multistate Models

The second part of the thesis focused on the two-state progressive Hidden Markov Chains (HMC). The purpose is to estimate the only transition probability under all kinds of censoring mechanism. The main findings are:

- Distributions of the different block lengths without censoring and with fixed censoring scenarios.

- Derivation of moment-based estimators under all kinds of censoring mechanism.
- Derivation of MLEs based on both two blocks under under no censoring and fixed censoring scenarios.
- Derivation of MLEs based on only one of the blocks under no censoring scenarios
- Validation of the consistency of the proposed estimators through simulation studies, even under challenging scenarios such as high noise levels or incomplete observations.
- Demonstration of the robustness of the estimators when the noise probability q is unknown and must be estimated from the data.

For HMCs, we only studied the 2-state scenario, the estimations in general cases remain to be discovered. Also the MLEs based on both two blocks under random censoring and MLEs based on one of the blocks under any censoring mechanism are not given in this paper.

A Appendix: R Code Used for Simulations

A.1 Simulation for Normal Markov Chain

The following script contains 3 functions:

The first one generates multistate Markov chain with transition matrix P and censoring time M, if M is 0, then it will generate uncensored observations (It stops when last state is present).

The second one counts all the transitions in a chain, returns a matrix (Same dimensions as P).

The last one calculates corresponding MLEs using the counting matrix, also returns a matrix (Same dimensions as P).

Listing 1: Simulation of Normal Markov Chain

```
# Simulation to check consistency of MLE for transition probabilities
# in a progressive Markov chain with independent censoring

# Function to generate a single realization of a progressive Markov
   chain
# with transition matrix P and censoring time M
generate_progressive_chain <- function(P, M) {
  num_states <- nrow(P)
  states <- numeric(M)
  states[1] <- 1  # Always start in state 1

  # Uncensored
  if(M==0){
    t=1
    current_state <- states[t]
    while(current_state != num_states){
      # Generate next state based on transition probabilities
      probs <- P[current_state,]
      states[t+1] <- sample(1:num_states, 1, prob = probs)
      t = t+1
      current_state <- states[t]
    }
    return(states)
  }

  for (t in 1:(M-1)) {
    current_state <- states[t]
    if (current_state == num_states) break  # Absorbing state reached

    # Generate next state based on transition probabilities
    probs <- P[current_state,]
    states[t+1] <- sample(1:num_states, 1, prob = probs)
  }

  # Fill remaining positions with the last state
```

```
    if (t < M) {
      states[(t+1):M] <- states[t]
    }

    return(states)
  }

  # Function to count transitions in a sequence
  count_transitions <- function(states, num_states) {
    # Initialize transition counts matrix
    n_ij <- matrix(0, nrow = num_states, ncol = num_states)

    # Count transitions
    for (t in 1:(length(states)-1)) {
      i <- states[t]
      j <- states[t+1]
      n_ij[i,j] <- n_ij[i,j] + 1
    }

    return(n_ij)
  }

  # Function to calculate MLE for all transition probabilities
  calculate_mle <- function(counts) {
    num_states <- nrow(counts)
    P_hat <- matrix(0, nrow = num_states, ncol = num_states)

    for (i in 1:(num_states-1)) {  # No transitions from absorbing state
      row_sum <- sum(counts[i,])
      if (row_sum > 0) {
        P_hat[i,] <- counts[i,] / row_sum
      } else {
        P_hat[i,] <- NA  # No transitions from state i observed
      }
    }

    # Set absorbing state probability
    P_hat[num_states, num_states] <- 1

    return(P_hat)
  }
```

The following script is the function that runs simulations with provided P and m, and the maximum sample size.

Listing 2: Consistency of MLEs for Transition Probabilities in Normal Markov Chain

```
  show_MC_MLE = function(P, max_n=10000, m=0, steps=10){
    # If m<0, use poisson distribution with parameter -m
    sample_sizes <- round(exp(seq(log(100), log(max_n), length.out =
        steps)))
    # Storage for results
    results <- matrix(NA, nrow = length(sample_sizes), ncol = num_states
        * num_states)
    colnames(results) <- paste0("p_", rep(1:num_states, each = num_
        states), "_", rep(1:num_states, num_states))

    for (s in 1:length(sample_sizes)) {
      n <- sample_sizes[s]
```

```
    # For each sample size, generate n chains with random censoring
    total_counts <- matrix(0, nrow = num_states, ncol = num_states)

    for (i in 1:n) {
      # Generate random censoring time
      if(m<0){
        M <- rpois(1, -m) + 2  # Ensure censoring time is at least 2
      }else{
        M = m
      }
      # Generate chain
      chain <- generate_progressive_chain(P_true, M)

      # Count transitions
      counts <- count_transitions(chain, num_states)

      # Accumulate counts
      total_counts <- total_counts + counts
    }

    # Calculate MLE
    P_hat <- calculate_mle(total_counts)

    # Store results
    results[s,] <- as.vector(t(P_hat))
  }

  # Only plot relevant transition probabilities (diagonal and upper
      triangle)
  for (i in 1:(num_states-1)) {
    for (j in i:num_states) {
      col_idx <- (i-1)*num_states + j
      param_name <- paste0("p_", i, "_", j)
      p = P_true[i,j]
      plot(sample_sizes, results[, col_idx], type = "o", log = "x",
        ylim = c(p-0.05, p+0.1),
        xlab = "Sample size (n)", ylab = "Estimated value",
        main = paste0("Convergence of MLE to theoretical value of ",
            param_name))
      abline(h = p, col = "red", lty = 2)
      legend("topright", legend = c("Estimated value", paste("True
          value =", p)),
           col = c("black", "red"), lty = c(1, 2))
    }
  }
}
```

Here is the example usage:

Listing 3: Example usage

```
num_states <- 3  # Number of states in the chain
# Define true transition matrix for a progressive chain
P_true <- matrix(0, nrow = num_states, ncol = num_states)
P_true[1,] <- c(0.7, 0.2, 0.1)  # Transitions from state 1
P_true[2,] <- c(0, 0.6, 0.4)    # Transitions from state 2
P_true[3,] <- c(0, 0, 1)         # State 3 is absorbing
show_MC_MLE(P)
```

A.2 Simulation for Two-State HMC

The following R script contains a function that generates Two-State HMC:

Listing 4: Generating Two-State HMC

```
# Simulation hidden two-state Markov chain
# Parameters:
# p: probability of staying in state 1
# q: probability of not observing a state
# num_trials: number of independent realizations to simulate
# max_length: maximum length of each chain
#   - if max_length=0: generate until first state 2 is observed (stop
    at first Y=2)
#   - if max_length<0: generate each observation with random censoring
     ~ Poi(-max_length)
#   - if max_length>0: fixed maximum length (original behavior)
simulate_hidden_chain <- function(p, q, num_trials = 2000, max_length
    = 0) {
  results <- data.frame(L1 = numeric(num_trials), L0 = numeric(num_
      trials))
  zeros = 0
  nonzeros = 0
  for (i in 1:num_trials) {
    if (max_length == 0) {
      # Generate until first state 2 is OBSERVED
      Y <- c(1)  # First observation
      U <- c(1)
      is_absorbing = FALSE
      while (TRUE) {
        # Generate next state
        if(is_absorbing){
          new_X <- 2
        }else{
            if (runif(1) > p) {
            new_X <- 2  # State 2 is absorbing
            is_absorbing = TRUE
          } else {
            new_X <- 1  # Stay in state 1
          }
        }
        # Generate observation indicator
        new_U <- rbinom(1, 1, 1-q)
        U <- c(U, new_U)

        # Generate observation
        new_Y <- new_X * new_U
        Y <- c(Y, new_Y)

        # Check if we've observed state 2
        if (new_Y == 2) {
          break  # Stop when state 2 is observed
        }
      }
    } else if (max_length < 0) {
      # Generate with random length for each observation
      random_length <- rpois(1, -max_length) + 2
      X <- rep(1, random_length)  # Start with all 1s
      for (t in 2:random_length) {
        # State 1 can transition to state 2 with probability 1-p
        if (runif(1) > p) {
          X[t:random_length] <- 2  # State 2 is absorbing
          break
        }
```

```
      }

      # Generate observation indicators (U)
      U <- rbinom(random_length, 1, 1-q)
      U[1] <- 1  # We always observe the first state

      # Generate observed process (Y)
      Y <- X * U
    } else {
      # Original behavior with fixed maximum length
      X <- rep(1, max_length)  # Start with all 1s
      for (t in 2:max_length) {
        # State 1 can transition to state 2 with probability 1-p
        if (runif(1) > p) {
          X[t:max_length] <- 2  # State 2 is absorbing
          break
        }
      }

      # Generate observation indicators (U)
      U <- rbinom(max_length, 1, 1-q)
      U[1] <- 1  # We always observe the first state

      # Generate observed process (Y)
      Y <- X * U
    }

    # Create modified observations using knowledge of the model
        structure
    Y_hat <- Y

    # Fill in known values based on structure:
    # 1. If we observe X_t = 1, then all previous values must be 1
    # 2. If we observe X_t = 2, then all subsequent values must be 2
    last_1 <- max(which(Y_hat == 1), 0)
    if (last_1 > 0) {
      Y_hat[1:last_1] <- 1
    }

    first_2 <- min(which(Y == 2), length(Y) + 1)
    if (first_2 <= length(Y)) {
      Y_hat[first_2:length(Y)] <- 2
    }

    # Calculate L1 (length of observed 1-block)
    L1 <- sum(Y_hat == 1)

    # Calculate L0 (length of observed 0-block)
    # This is the number of zeros after the last 1
    L0 <- sum(Y_hat == 0)

    results$L1[i] <- L1
    results$L0[i] <- L0

    zeros = zeros + sum(U == 0)
    nonzeros = nonzeros + sum(U == 1) - 1
  }

  return(list(res = results, est_q = zeros/(zeros+nonzeros)))
}
```

The following R script contains functions that calculate expectations of lengths of two blocked under different censoring mechanism:

Listing 5: Generating Two-State HMC

```
EL1 = function(p,q){
  return((1-2*p*q+p^2*q)/((1-p)*(1-p*q)))
}
EL0 = function(p,q){
  return((q+p*q-2*p*q**2)/((1-q)*(1-p*q)))
}
EL1m = function(p,q,m){
  return(EL1(p,q)-p^m/(1-p)+(p*q)^m/(1-p*q))
}

EL0m = function(p,q,m){
  pm1=p^(m-1)
  qm1=q^(m-1)
  pm2=p^(m-2)
  qm2=q^(m-2)
  pq=p*q
  pqm1=pq^(m-1)
  term1 = q*(2-pm1-pqm1)/(1-q)
  term2 = (q-pq)*(pq*(pm2-qm2)/(p-q)+(1-pqm1)/(1-pq))/(1-q)
  return(term1-term2)
}

EL1M = function(p,q,mu){
  p1 = p*q
  return(EL1(p,q)-p^2*exp(-mu+p*mu)/(1-p)+p1^2*exp(-mu+p1*mu)/(1-p1))
}

EL0M = function(p,q,mu){
  eM = 0
  next_term = function(m) EL0m(p,q,m)*mu^(m-2)*exp(-mu)/factorial(m-2)
  m = 2
  while (abs(next_term(m))>1e-6) {
    eM = eM + next_term(m)
    m = m+1
  }
  return(eM)
}
```

The following R script contains two functions:

The first one runs simulations and return observed average of lengths of blocks and differences between theoretical values.

The second one plots the results.

Listing 6: Expectations of block length in Two-State HMC

```
# Function to examine convergence as sample size increases
validate_expectations <- function(p, q, max_n = 10000, steps = 10, m =
    0) {
  n_values <- round(exp(seq(log(100), log(max_n), length.out = steps))
    )
  if(m==0){
    EL1 = EL1(p,q)
    EL0 = EL0(p,q)
  }else if(m>0){
    EL1 = EL1m(p,q,m)
    EL0 = EL0m(p,q,m)
```

```
  }else{
    EL1 = EL1M(p,q,-m)
    EL0 = EL0M(p,q,-m)
  }
  results <- data.frame(
    n = n_values,
    avg_L1 = numeric(steps),
    EL1 = rep(EL1, steps),
    avg_L0 = numeric(steps),
    EL0 = rep(EL0, steps),
    L1_diff = numeric(steps),
    L0_diff = numeric(steps)
  )
  for(i in 1:steps) {
    n <- n_values[i]
    # Run multiple simulations and average results
    sim_runs <- 30  # Number of simulation runs to average
    L1_vals <- numeric(sim_runs)
    L0_vals <- numeric(sim_runs)

    for(j in 1:sim_runs) {
      sim_data <- simulate_hidden_chain(p, q, num_trials = n, max_
          length = m)$res
      L1_vals[j] <- mean(sim_data$L1)
      L0_vals[j] <- mean(sim_data$L0)
    }

    results$avg_L1[i] <- mean(L1_vals)
    results$avg_L0[i] <- mean(L0_vals)
    results$L1_diff[i] <- abs(results$avg_L1[i] - results$EL1[i])
    results$L0_diff[i] <- abs(results$avg_L0[i] - results$EL0[i])
  }

  return(results)
}

show_validation_results = function(p, q, max_n=10000, steps=10, m=0){
  if(m==0){
    EL1 = EL1(p,q)
    EL0 = EL0(p,q)
  }else if(m>0){
    EL1 = EL1m(p,q,m)
    EL0 = EL0m(p,q,m)
  }else{
    EL1 = EL1M(p,q,-m)
    EL0 = EL0M(p,q,-m)
  }
  # Run validation for selected (p,q) pair
  validation_results <- validate_expectations(p, q, max_n, steps, m)

  # Plot convergence of L1
  plot(validation_results$n, validation_results$avg_L1, type = "o",
      log = "x",
       ylim = c(EL1-0.05, EL1+0.1),
       xlab = "Sample size (n)", ylab = "Average L1",
       main = paste0("Convergence of L1 average to theoretical value (
           p=", p, ", q=", q, ")"))
  abline(h = validation_results$EL1[1], col = "red", lty = 2)
  legend("topright", legend = c("Observed average", "Theoretical value
      "),
         col = c("black", "red"), lty = c(1, 2))

  # Plot convergence of L0
  plot(validation_results$n, validation_results$avg_L0, type = "o",
      log = "x",
       ylim = c(EL0-0.05, EL0+0.1),
```

```
       xlab = "Sample size (n)", ylab = "Average L0",
       main = paste0("Convergence of L0 average to theoretical value (
           p=", p, ", q=", q, ")"))
  abline(h = validation_results$EL0[1], col = "red", lty = 2)
  legend("topright", legend = c("Observed average", "Theoretical value
      "),
         col = c("black", "red"), lty = c(1, 2))

  # Plot absolute differences to show convergence rate
  plot(validation_results$n, validation_results$L1_diff, type = "o",
      log = "xy",
       xlab = "Sample size (n)", ylab = "Absolute difference",
       main = "Convergence rate of estimators for EL1 and EL0")
  lines(validation_results$n, validation_results$L0_diff, type = "o",
      col = "blue")
  legend("topright", legend = c("L1 difference", "L0 difference"),
         col = c("black", "blue"), lty = 1)
}
```

And an example:

Listing 7: Example Usage

```
p_test <- 0.7
q_test <- 0.3
show_validation_results(p_test, q_test, m=0)
```

Next we validate the consistency of our estimators, first the functions that define them:

Listing 8: Estimators for p in Two-State HMC

```
f1 <- function(q,e){
  d <- e^2*(1-q)^2+4*q*(1-q)*(e-1)
  if(d<0){
    return(NA)
  }
  p <- (e+e*q-2*q-sqrt(d))/(2*(e*q-q))
  if(p > 0 & p < 1){
    return(p)
  }
  return(NA)
}

f0 <- function(q,e){
  e1 = e - q/(1-q)
  return(e1/q/(1+e1))
}

f1m <- function(q,m,e){
  f <- function(p) EL1m(p,q,m)-e
  result <- try(uniroot(f, lower = 0.001, upper = 0.999), silent =
      TRUE)

  if (inherits(result, "try-error")) {
    return(NA)
  } else {
    return(result$root)
  }
```

```
}

f0m <- function(q,m,e){
  f <- function(p) EL0m(p,q,m)-e
  abs_f <- function(p) abs(f(p))
  result <- optim(par = 0.5, fn = abs_f, method = "L-BFGS-B", lower =
      0.001, upper = 0.999)
  # result <- try(uniroot(f, lower = 0.001, upper = 0.999), silent =
      TRUE)
  if (inherits(result, "try-error")) {
    return(NA)
  } else {
    return(result$par)
  }
}

f1M <- function(q,mu,e){
  f <- function(p) EL1M(p,q,mu)-e
  result <- try(uniroot(f, lower = 0.001, upper = 0.999), silent =
      TRUE)

  if (inherits(result, "try-error")) {
    return(NA)
  } else {
    return(result$root)
  }
}

f0M <- function(q,mu,e){
  f <- function(p) EL0M(p,q,mu)-e
  abs_f <- function(p) abs(f(p))
  result <- optim(par = 0.5, fn = abs_f, method = "L-BFGS-B",lower =
      0.001, upper = 0.999)
  # result <- try(uniroot(f, lower = 0.001, upper = 0.999), silent =
      TRUE)

  if (inherits(result, "try-error")) {
    return(NA)
  } else {
    return(result$par)
  }
}

Est_comb = function(p1, p0, obs){
  l0n = sum(obs$L0)
  l1n = sum(obs$L1)
  w1 = l1n/(l1n+l0n)
  w0 = 1-w1
  return(w1*p1+w0*p0)
}

MLE <- function(obs){
  temp = obs$L0+1
  f <- function(p) sum(temp*p^temp/(1-p^temp)) - sum(obs$L1) + nrow(
      obs)
  result <- try(uniroot(f, lower = 0.001, upper = 0.999), silent =
      TRUE)

  if (inherits(result, "try-error")) {
    return(NA)
  } else {
    return(result$root)
  }
}

MLEm <- function(m,obs){
```

```
  temp = obs$L0+1
  f <- function(p) sum((obs$L0+obs$L1 < m)*temp*p^temp/(1-p^temp)) -
      sum(obs$L1) + nrow(obs)
  result <- try(uniroot(f, lower = 0.001, upper = 0.999), silent =
      TRUE)

  if (inherits(result, "try-error")) {
    return(NA)
  } else {
    return(result$root)
  }
}

# MLE based on L0
MLE0 <- function(q,obs){
  n = nrow(obs)
  temp = obs$L0+1
  f <- function(p) n*(1-q)/((1-p)*(1-p*q))-sum(temp*p^obs$L0/(1-p^temp
      ))
  result <- try(uniroot(f, lower = 0.001, upper = 0.999), silent =
      TRUE)

  if (inherits(result, "try-error")) {
    return(NA)
  } else {
    return(result$root)
  }
}

# MLE based on L1
MLE1 <- function(q,obs){
  n = nrow(obs)
  S = sum(obs$L1[obs$L1>1]-1)
  b = S*(1+q)+n*(1-q)
  a = b/(2*S*q)
  b = sqrt(b^2-4*S^2*q)/(2*S*q)
  if(a-b >= 0 & a-b <1){
    return(a-b)
  }else{
    return(NA)
  }
}
```

Then the functions that plot the results:

Listing 9: Consitency Validation for p estimators in Two-State HMC

```
# Function to examine consistency of estimators
check_consistency <- function(true_p, true_q, max_n=10000, steps=10, m
    =0, use_TQ=TRUE) {
  n_values <- round(exp(seq(log(100), log(max_n), length.out = steps))
      )
  results <- data.frame(
    sample_size = n_values,
    f1_est = numeric(steps),
    f0_est = numeric(steps),
    mle_est_0 = numeric(steps),
    mle_est_1 = numeric(steps),
    mle_est = numeric(steps),
    combined_est = numeric(steps)
  )

  for(i in 1:steps) {
    n <- n_values[i]
```

```
    # Run multiple simulations and average results
    sim_runs <- 1
    f1_vals <- numeric(sim_runs)
    f0_vals <- numeric(sim_runs)
    mle_vals_0 <- numeric(sim_runs)
    mle_vals_1 <- numeric(sim_runs)
    mle_vals <- numeric(sim_runs)
    combined_vals <- numeric(sim_runs)

    for(j in 1:sim_runs) {
      sim_data <- simulate_hidden_chain(true_p, true_q, num_trials = n
          , max_length = m)
      sim_data_res <- sim_data$res
      avg_L1 <- mean(sim_data_res$L1)
      avg_L0 <- mean(sim_data_res$L0)

      if(use_TQ){
        q <- true_q
      }else{
        q <- sim_data$est_q
      }
      if(m>0){
        f1_vals[j] <- f1m(q, m, avg_L1)
        f0_vals[j] <- f0m(q, m, avg_L0)
        mle_vals[j] <- MLEm(m,sim_data_res)
      }else if(m==0){
        f1_vals[j] <- f1(q, avg_L1)
        f0_vals[j] <- f0(q, avg_L0)
        mle_vals[j] <- MLE(sim_data_res)
        mle_vals_0[j] = MLE0(q, sim_data_res)
        mle_vals_1[j] = MLE1(q, sim_data_res)
      }else{
        f1_vals[j] <- f1M(q,-m, avg_L1)
        f0_vals[j] <- f0M(q,-m, avg_L0)
      }
      combined_vals[j] <- Est_comb(f1_vals[j],f0_vals[j], sim_data_res
          )

    }

    results$f1_est[i] <- mean(f1_vals)
    results$f0_est[i] <- mean(f0_vals)
    results$mle_est_0[i] <- mean(mle_vals_0)
    results$mle_est_1[i] <- mean(mle_vals_1)
    results$mle_est[i] <- mean(mle_vals)
    results$combined_est[i] <- mean(combined_vals)
  }

  return(results)
}

check_existence = function(x){
  return(sum(x)!=0)
}

show_consistency_results = function(p, q, max_n=10000, steps=10, m=0 ,
     use_TQ=TRUE){
  # Check consistency for selected (p,q) pair
  consistency_results <- check_consistency(p, q,max_n=max_n, steps=
      steps, m=m, use_TQ=use_TQ)
  # print(consistency_results)
  cols = c("blue", "green")
  lty = c(1, 1)
  legends = c("f1 estimator", "f0 estimator")
  attach(consistency_results)
  # Plot consistency of estimators
```

```
  plot(sample_size, f1_est, type = "o", col = "blue", log = "x",
       ylim = c(p-0.1, p+0.3),
       xlab = "Sample size (n)", ylab = "Estimated p",
       main = paste0("Consistency of estimators for parameter p (true
           p=", p, ", q=", q, ")"))
  lines(sample_size, f0_est, type = "o", col = "green")
  if(check_existence(mle_est_0)){
    lines(sample_size, mle_est_0, type = "o", col = "pink")
    cols = c(cols, "pink")
    lty = c(lty, 1)
    legends = c(legends, "MLE_L0")
  }
  if(check_existence(mle_est_1)){
    lines(sample_size, mle_est_1, type = "o", col = "darkred")
    cols = c(cols, "darkred")
    lty = c(lty, 1)
    legends = c(legends, "MLE_L1")
  }
  if(check_existence(mle_est)){
    lines(sample_size, mle_est, type = "o", col = "red")
    cols = c(cols, "red")
    lty = c(lty, 1)
    legends = c(legends, "MLE")
  }
  if(check_existence(combined_est)){
    lines(sample_size, combined_est, type = "o", col = "purple")
    cols = c(cols, "purple")
    lty = c(lty, 1)
    legends = c(legends, "Combined")
  }
  detach(consistency_results)
  abline(h = p, lty = 2)
  cols = c(cols, "black")
  lty = c(lty, 2)
  legends = c(legends, "True Value")
  legend("topright", legend = legends, col = cols, lty = lty)
}
```

Two examples:

Listing 10: Example Usage

```
# No censoring using true q
for (p_test in c(0.3,0.7)) {
  for (q_test in c(0.3,0.7)) {
    show_consistency_results(p_test, q_test, m=0)
  }
}

# Random censoring using estimated q  (M ~ Poi(10)+2)
show_consistency_results(p_test, q_test, m=-10, use_TQ = F)
```

References

Billingsley, P. (2012). *Probability and Measure*. Wiley Series in Probability and Statistics. Wiley. ISBN: 9781118341919. URL: https://books.google.ee/books?id=a3gavZbxyJcC.

Capasso, V. and D. Bakstein (2012). *An Introduction to Continuous-Time Stochastic Processes: Theory, Models, and Applications to Finance, Biology, and Medicine*. Modeling and Simulation in Science, Engineering and Technology. Birkhäuser Boston. ISBN: 9780817683467. URL: https://books.google.ee/books?id=h83lVwvgJnoC.

Fisher, R. A. (1922). "On the Mathematical Foundations of Theoretical Statistics". In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 222, pp. 309–368. ISSN: 02643952. URL: http://www.jstor.org/stable/91208 (visited on 05/20/2025).

Ito, K. and K. Kunisch (2008). *Lagrange Multiplier Approach to Variational Problems and Applications*. Advances in Design and Control. Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104). ISBN: 9780898718614. URL: https://books.google.ee/books?id=h0b8s6ASY-oC.

Norris, J.R. (1998). *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. ISBN: 9780521633963. URL: https://books.google.ee/books?id=qM65VRmOJZAC.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Clinical Perspectives in Obstetrics and Gynecology. Springer New York. ISBN: 9780387909905. URL: https://books.google.ee/books?id=B2vgGMa9vd4C.

Non-exclusive licence to reproduce thesis and make thesis public

I, Zhen Xu,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright, Parameter Estimation in Progressive Multistate Models, supervised by Jüri Lember.

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Zhen Xu
21/05/2025