Client: Civil Engineering Division of Rijkswaterstaat

Statistical inference and hypothesis testing for Markov chains with Interval Censoring

Application to bridge condition data in the Netherlands

Author: Monika Skuriat-Olechnowska
Delft University of Technology

STATISTICAL INFERENCE AND HYPOTHESIS TESTING FOR MARKOV CHAINS WITH INTERVAL CENSORING

Application to bridge condition data in the Netherlands

Monika Skuriat-Olechnowska

A thesis submitted to the Delft University of Technology in conformity with the requirements for the degree of Master of Science

Delft, The Netherlands

Members of The Committee:

Chairperson of Graduate Committee: Prof. Dr. Ir. Jan M. Van Noortwijk

Graduate Committee:

Prof. Dr. Ir. Jan M. Van Noortwijk	Delft University of Technology, Faculty of Electrical Engineering, Mathematics and Computer Science
Dr. Hans van der Weide	Delft University of Technology, Faculty of Electrical Engineering, Mathematics and Computer Science
Ir. Maarten-Jan Kallen	HKV Consultants
Ir. Leo Klatter	Civil Engineering Division of Rijkswaterstaat

ACKNOWLEDGMENTS

I would like to express my sincerest appreciation to the Delft University of Technology, especially to Rooger Cooke and Dorota Kurowicka, where I have had opportunity to spend wonderful two years in the Risk and Environmental Modelling program.

Thanks to HKV Consultants for giving me opportunity to work within their organization, specially to Maarten-Jan Kallen for introducing the subject of this work, providing data and helpful explanations.

Many thanks go out to Jan M. Van Noortwijk for his generous supervision of my work for this thesis. I would also like to acknowledge Beata, Weronika, Agnieszka, Marcin, Slawek, Sebastian … for being good friends during these two years of studying in Delft.

I am grateful to my family for their love and support that they gave me and which I could feel despite of the large distance. Last but not the least, I would like to thank my husband Maciej for his belief in me, continuous support and care.

Contents

List of tables

List of figures

Samenvatting

Dit rapport is het resultaat van het afstudeerproject van Monika Skuriat-Olechnowska, studente Technische Wiskunde aan de faculteit Elektrotechniek, Wiskunde en Informatica (EWI) van de Technische Universiteit Delft. Het project is uitgevoerd in de periode van januari tot juli 2005 bij HKV LIJN IN WATER te Lelystad onder begeleiding van ir. M.J.Kallen en prof. dr. ir. J.M. van Noortwijk.

Introductie

Bruggen en viaducten die onderdeel uitmaken van de rijkswegen in Nederland worden beheerd door de Bouwdienst (Rijkswaterstaat, Ministerie van Verkeer en Waterstaat). Om de kwaliteit van deze belangrijke objecten te waarborgen, worden ze periodiek geïnspecteerd. Dit zijn visuele inspecties die op een doorlopende basis over het hele netwerk van bruggen en viaducten worden uitgevoerd. Tijdens de inspecties worden verschillende onderdelen van een brug nauwkeurig bekeken en kent de inspecteur aan elk onderdeel een toestandsindicator toe. Er zijn zeven discrete toestanden gedefinieerd en deze zijn weergegeven in Tabel 0-1.

Indicator	Staat van onderhoud van kunstwerkdeel
0	in prima staat
1	in zeer goede staat
2	in goede staat
3	in redelijke staat
4	in matige staat
5	in slechte staat
6	in zeer slechte staat

Tabel 0-1: toestandsindicatoren in DISK

De gegevens van elke inspectie worden geregistreerd in het Data Informatie Systeem Kunstwerken (DISK). Dit systeem is al sinds december 1985 in gebruik en bevat derhalve bijna 20 jaar aan gegevens.

In het kader van huidig onderzoek naar onderhoudsoptimalisatie bij de Bouwdienst, is men geïnteresseerd om de historie van toestandsindicatoren in DISK te gebruiken voor het voorspellen van de veroudering. Deze veroudering is onzeker in de tijd en mogelijk afhankelijk van één of meerdere eigenschappen van een brug. Een logische keuze voor het modelleren van veroudering op een discrete toestandsschaal, zijn de zogenaamde Markovprocessen. Een Markovproces is een stochastisch proces als functie van de tijd waarbij de toekomstige toestand slechts (conditioneel) afhankelijk is van de huidige toestand. Deze aanname vereenvoudigt de wiskundige modellering van de onzekere veroudering. Het gebruik van Markovprocessen in Bridge Management Systemen (BMS) is niet nieuw. Het Amerikaanse PONTIS (en ook het minder gebruikte BRIDGIT) maakt gebruik van Markovketens. Bij deze ketens neemt men aan dat transities tussen de verschillende toestanden slechts jaarlijks voorkomen. Met andere woorden: Markovketens zijn Markovprocessen gedefinieerd op een discrete tijdsas.

Men kan zich nu afvragen in hoeverre een model zoals door PONTIS wordt gebruikt, toepasbaar is op de situatie in Nederland. Zo niet, kan een vergelijkbaar model opgesteld worden die wel voor de Nederlandse situatie geschikt is? In ieder geval moet nagegaan worden of de aanname van de Markoveigenschap wel terecht is. Op basis hiervan zijn de volgende onderzoeksvragen gedefinieerd:

1. kunnen de gegevens in de DISK database getoetst worden op de Markoveigenschap en zo ja, wat is het resultaat van deze toets?
2. hoe moeten de parameters van de Markovketen geschat worden zodanig dat rekening gehouden wordt met de bijzondere vorm van censurering? (we kennen niet het exacte moment van een transitie tussen de toestandsindicatoren, alleen dat deze tussen twee tijdsstippen heeft plaatsgevonden)
3. zijn de transitiekansen verschillend per toestand? M.a.w. blijft een object in sommige toestand significant langer of korter dan ik andere toestanden?
4. welke karakteristieken van een brug of viaduct hebben invloed op de transitiekansen?
5. kunnen we de subjectiviteit van de visuele inspecties meenemen in het verouderingsmodel en zo ja, hoe?

Analyse

In het verslag wordt uitgegaan van twee Markovmodellen. In hoofdstuk 4 wordt gebruik gemaakt van een model waarbij transities over meerdere condities tegelijk kunnen voorkomen. Dit model wordt ook door PONTIS gebruikt en bestaat uit een discrete-tijd Markovketen met een volledig gevulde transitie kansmatrix van de vorm:

$$P = \begin{bmatrix} P_{00} & P_{01} & \cdots & P_{0n} \\ P_{10} & P_{11} & \cdots & P_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n0} & P_{n1} & \cdots & P_{nn} \end{bmatrix}.$$

Vanaf hoofdstuk 5 wordt de aanname gemaakt dat een transitie altijd naar de eerstvolgende toestandsindicator geschiedt. Deze vorm komt veelvuldig voor in de wetenschappelijke literatuur en lijkt een goede benadering van veroudering bij een kleine tijdseenheid tussen de transities. De transitiematrix is in dit model van de vorm:

$$P = \begin{bmatrix} 1-P_{01} & P_{01} & 0 & 0 \\ 0 & 1-P_{12} & P_{12} & 0 \\ 0 & 0 & 1-P_{23} & P_{23} \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

waarbij dit een voorbeeld is met slechts vier toestanden.

In hoofdstuk 4 worden drie statistische toetsen uitgevoerd. De eerste twee zijn voor het toetsen van de Markoveigenschap. Allereerst wordt een kruistabel (contingency table) opgesteld met een selectie van transitiefrequenties. Deze frequenties zijn de geobserveerde aantallen transities en deze worden vergeleken met de aantallen die men onder de Markoveigenschap zou verwachten. Indien de verdeling van de geobserveerde aantallen significant verschilt van de verwachte aantallen, dan wordt de hypothese van de Markoveigenschap verworpen. Hetzelfde wordt gedaan in de tweede toets, maar deze keer worden de frequenties van tweede orde transities vergeleken met eerste orde transities. Een Markovketen van tweede orde is niet alleen afhankelijk van de huidige toestand, maar ook van één toestand in het verleden. Indien de verdeling van tweede orde transitiekansen significant anders is dan die van een gewone eerste orde transitiekansen, dan wordt de aanname van de Markoveigenschap verworpen. De derde toets is voor het toetsen van de aanname van stationariteit. Een Markovproces is stationair als de transitiekansen gelijk zijn gedurende de leeftijd van het proces. D.w.z. dat de kans op een bepaalde transitie gelijk is op elk moment van het proces.

Omdat er sprake is van een soort intervalcensurering, wordt in hoofdstuk 5 een schattingsmethodiek gepresenteerd, die voor deze situatie geschikt is. De maximum-likelihoodschatter wordt bepaald door het maximaliseren van de aannemelijkheid van de observaties als functie van de parameters $P_{01}, P_{12}, \ldots, P_{n-1,n}$. Elke observatie bestaat uit drie gegevens: de toestanden ten tijde van twee opeenvolgende inspecties en de tijd die verstreken is tussen de twee inspecties. Het numerieke algoritme dat in hoofdstuk 5 gepresenteerd wordt, maakt het mogelijk om de standaarddeviatie van de schatter te bepalen. Hiermee kan vervolgens een betrouwbaarheidsband voor de schatter uitgerekend worden. Als laatste wordt in dit hoofdstuk een likelihood-ratiotoets gebruikt om na te gaan of de parameters $P_{01}, P_{12}, \ldots, P_{n-1,n}$ significant van elkaar verschillen. De nulhypothese in deze toets is daarmee $H_0 : P_{01} = P_{12} = \ldots = P_{n-1,n}$ en er wordt getoetst of deze hypothese beter bij de gegevens past dan de andere aanname, namelijk dat de transitiekansen verschillend zijn. Als de kansen verschillend zijn, dan worden ze in dit rapport toestandsafhankelijk genoemd.

Om de invloed van bepaalde karakteristieken zoals locatie, leeftijd, gebruiksintensiteit, enz. na te gaan, wordt in hoofdstuk 6 logistische regressie toegepast. In deze techniek worden de transitiekansen geformuleerd als functie van de verschillende hulpvariabelen (covariates). Deze hulpvariabelen kunnen twee waarden aannemen: 0 of 1. De 0 staat dan bijv. voor ‘kustprovincies’, ‘gebouwd vóór 1976’ en ‘licht/weinig verkeer’ en de 1 staat voor het tegenovergestelde hiervan: ‘niet-kustprovincies’, ‘gebouwd na 1976’ en ‘zwaar/veel verkeer’. Het doel van deze analyse is om na te gaan of deze onderverdelingen wel of geen statistisch significante invloed hebben of de parameters van het model.

In het laatste hoofdstuk, worden formules afgeleid die het mogelijk maken om de subjectiviteit van visuele inspecties mee te nemen in het Markovproces. Er wordt hierbij vanuit gegaan dat er een kans is dat de inspecteur de toestand beter of slechter inschat dan deze in werkelijkheid is. Er wordt dan gesproken van een geobserveerd proces en een verborgen proces. Omdat we vooral geïnteresseerd zijn in de daadwerkelijke snelheid van de veroudering in bruggen, willen we de transitiekansen van het verborgen proces schatten.

Resultaten en aanbevelingen

De resultaten voor de eerdergenoemde vragen zijn als volgt:

1. **kunnen de gegevens in de DISK database getoetst worden op de Markoveigenschap en zo ja, wat is het resultaat van deze toets?**
   Ja, de Markoveigenschap kan getoetst worden en de beste toets hiervoor blijkt de toets van de orde van het Markovproces. De eerste toets, op basis van een kruistabel blijkt erg gevoelig voor hoe men deze tabellen opstelt. Het is namelijk makkelijk om alleen die waarden te gebruiken die resulteren in een gunstig resultaat. Bij de toets van de

orde is dit moeilijker en hierdoor is deze toets minder gevoelig voor manipulatie. Het resultaat van beide toetsen is dat voor de DISK gegevens de Markoveigenschap niet verworpen kan worden. Dit wil zeggen dat er geen statistisch significant bewijs is voor het tegengestelde.

2. **hoe moeten de parameters van de Markovketen geschat worden zodanig dat rekening gehouden wordt met de bijzondere vorm van censurering? (we kennen niet het exacte moment van een transitie tussen de toestandsindicatoren, alleen dat deze tussen twee tijdsstippen heeft plaatsgevonden)**
   In hoofdstuk 5 wordt een algoritme gepresenteerd die correct rekening houdt met de intervalcensurering. De schatter is een zogenaamde maximum-likelihoodschatter en het algoritme laat ons toe om het betrouwbaarheidsinterval van de schatter uit te rekenen.
3. **zijn de transitiekansen afhankelijk van de toestand waarin een object zich bevindt?**
   Ja, de likelihood-ratiotoets in hoofdstuk 5 geeft overduidelijk aan dat de transitiekansen verschillend zijn voor elke toestand. De objecten blijken sneller door toestanden 0 t/m 2 te gaan en vervolgens langzamer door toestanden 3 en 4.
4. **welke karakteristieken van een brug of viaduct hebben invloed op de transitiekansen?**
   In hoofdstuk 6 worden de volgende karakteristieken onderscheiden: jaar van constructie (voor of na 1976), locatie in weg (in of over de rijksweg), type brug (brug of viaduct), type verkeer (alleen auto's en vrachtwagens of ook fietsers en voetgangers) en provincie (twee groeperingen: kustprovincie of niet en dicht- of lichtbevolkte provincie). De invloed van elk van deze karakteristieken is per toestand bekeken. Het blijkt dat de groepering van de provincies volgens bevolkingsdichtheid statistisch significante invloed heeft op de transitiekansen tussen alle toestanden. Deze invloed is echter niet erg groot. Opmerkelijk is dat een hogere bevolkingsdichtheid alleen een hogere transitiekans heeft in toestanden 0 en 4, in de overige toestanden is er een (licht) lagere kans om naar de volgende toestand te gaan. Bruggen die vóór 1976 gebouwd zijn hebben in toestanden 0, 1 en 2 een duidelijk hogere transitiekans dan bruggen gebouwd in de periode na 1976. In toestand 3 is er geen significant verschil en in toestand 4 hebben de pre-1976 bruggen een lagere transitiekans. Dit wil zeggen dat pre-1976 bruggen korter in de toestanden 0 t/m 2 verblijven en langer in toestand 4 vergeleken met de post-1976 bruggen. Verder blijken bruggen sneller door de toestanden 0 en 1 te gaan dan viaducten.
5. **kunnen we de subjectiviteit van de visuele inspecties meenemen in het verouderingsmodel en zo ja, hoe?**
   Ja, maar het berekenen van de formule voor de aannemelijkheid van de subjectieve observaties in hoofdstuk 7, vergt een numerieke aanpak. Voor de implementatie hiervan was onvoldoende tijd tijdens het afstudeerproject en daarom zijn er geen resultaten beschikbaar.

Als aanbevolen acties worden vermeld:

- Implementeren van de relevante formules in hoofdstuk 7 voor het berekenen van de aannemelijkheid van subjectieve visuele observaties.
- De toets voor stationariteit in hoofdstuk 4 maakt gebruik van leeftijdsgroepen met een lengte van 5 jaar. Aanbevolen wordt om een andere leeftijdsindeling te maken of om de analyse zelfs op jaarbasis te doen. Hiervoor is echter een andere aanpak nodig, omdat de gegevens niet geschikt zijn.

- Naast de karakteristieken die in dit verslag aan bod gekomen zijn, is het misschien mogelijk om nog andere groeperingen te bedenken die van invloed kunnen zijn op de transitiekansen.
- Alleen het eerste model van hoofdstuk 4 is getoetst op de Markoveigenschap. Voor het model dat vanaf hoofdstuk 5 gebruikt wordt, zou deze eigenschap ook getoetst kunnen worden.
- Alle modellen in dit verslag gaan ervan uit dat er geen onderhoud in het spel is. Het zou interessant zijn om dit wel mee te nemen in de modellering.

Over het algemeen kan geconcludeerd worden dat Markovketens bruikbaar zijn voor de modellering van de onzeker veroudering in de tijd. De resultaten in dit rapport geven aan dat er veel onzekerheid zit in the levensduurverwachting van bruggen en hun structurele elementen. In het rapport zijn de gegevens aan grondige toetsen onderworpen om na te gaan dat alle aannames correct zijn. Deze toetsen maken het mogelijk om ook bij andere toepassingen met andere gegevens te beslissen welke vorm van Markovketens het meest geschikt is. Tegelijkertijd erkennen we ook de beperkingen van deze aanpak: de modellering is puur statistisch en er zijn slechts gegevens over een beperkte periode beschikbaar. De relatie met het fysieke proces van veroudering is nog niet gelegd en de kwaliteit van de resultaten is dus sterk afhankelijk van de kwaliteit van de gegevens.

Summary

The Civil Engineering Division ('Bouwdienst') of Rijkswaterstaat, which is part of the Netherlands Ministry of Transport, Public works and Water Management, has employed people to work on maintenance optimization. This division is responsible for the development of inspection and maintenance strategies for various infrastructures. Amongst others, they are responsible for the maintenance of the national highways and bridge network. In recent years, a lot of research has been done to better understand the physical process of deterioration of bridges, especially to understand the degradation in the quality of the concrete.

Since 1985, the Bouwdienst keeps all the inspection results for bridges in a relational database called "DISK". The condition of bridges is represented by a number from 0 to 6, where 0 is perfect and 6 is extremely bad. From the database we extracted relevant information to perform this analysis. Most importantly, we have information about the transitions from one condition to another and the duration between the observations, as well as external factors, which could be important to the analysis. This information can help us to determine (approximately) how fast a bridge moves from a perfect condition to a condition where maintenance is required.

Because bridge-inspection data concern visual ratings on a discrete condition scale varying (in seven steps) from "perfect" to "very bad", Markov chains are a suitable tool for the purpose of bridge deterioration modelling. The Markov chain is a stochastic process that takes on a finite number of possible states and the probability of moving toward a future state only depend on the current state. The parameters that have to be estimated are the probabilities of transition of the current state to the next state (transition probabilities). For this purpose, we propose statistical estimation techniques, which take into account the fact that the times of transitions, are not always known (called interval censoring). Also, we would like to have statistical tests at our disposal in order to assess the validity and relative performance of different types of Markov chains.

Interval censoring means that we do not know the exact time of an event. In our context, this means that we do not know the time when the bridge moves from one state to another state. For example, during an inspection we see that a given bridge is in state 2. After 2 years we perform another inspection in which we observe the bridge in state 3. We see that the bridge moved from state 2 to state 3 in 2 years, but we do not know exactly when this happened. These kinds of observations are called "interval censoring", because we know the interval, but we do not know the exact transition time.

In this thesis we will consider the following questions:

1. Does the Markov property hold for the data from the DISK database?
2. How to estimate a Markov chain based on the DISK data?
3. How to determine and compare the expected deterioration?
4. How to separate bridges into sensible groups and to determine the effect of this grouping on the parameters of the Markov model?
5. How to take into account the subjectivity of the inspectors into condition rating?

In this research, we consider and compare two different Markov models based on different approaches to model the transition probabilities. The first model assumes that transition probabilities do not depend on the state in which is a given bridge. The second model assumes dependence on the state. The results are expressed as one-year transition probability matrices.

Also, we would like to incorporate the effect of grouping bridges into sensible groups and take into account the inspectors' subjectivity into the condition rating. Another goal is to determine the essential factors which influence the deterioration process. This was done by dividing data with respect to essential factors and compared them with transition probabilities estimated for all data.

1 Introduction

Most of the bridges in the Netherlands are constructed of concrete, and more than half of them are more than 30 years old. As bridges deteriorate at an accelerating rate from corrosion, concrete degradation and vehicle damage, the Civil Engineering Division ('Bouwdienst') of Rijkswaterstaat, which is part of the Ministry of Transport, Public works and Water Management, must repair them and, where possible, prevent further deterioration.

The ministry of Public Works and Water Management is the principal of about 3500 bridges in the Netherlands. As a principal, the ministry likes to know the remaining service lifetime of its structures. During their lifetime the structures will need repair. At this moment a repair strategy that is based on inspections is used to determine when repair will be done. This repair strategy for concrete bridges in the Netherlands results in repair of bridges every 25 till 35 years. The repair will normally be 0.5 till 1.5% of the area of the structure (van Beek *et al*., 2003).

PONTIS and BRIDGIT are two of the most common Bridge Management Systems (BMS) currently available (Golabi and Shepard, 1997; Thompson *et al*., 1998). Both have their origins in the Arizona Pavement Management System developed in the late 1970's and are almost solely used in the United States. All these models use Markov chains to model uncertain deterioration of bridges over time. Unfortunately, the aspect of fitting model parameters has been largely neglected and objections against the Markov assumption have risen in recent years.

In the Netherlands, results from bridge inspections are registered in a database, which is primarily used for record keeping. This database is a very rich source of information -it contains data collected over a period of almost 20 years- and the purpose of current research is to use this data for estimating the rate of deterioration.

Acknowledging that Markov chains are a suitable tool for the purpose of bridge deterioration modelling, we would like to propose fitting techniques which take into account the special type of censoring involved with bridge inspections. Also, we would like to have statistical tests at our disposal in order to assess the validity and relative performance of different types of Markov chains. Essentially, we are interested in obtaining the functionality of a BMS like PONTIS, while taking special care of the validity of our assumptions and the resulting models with respect to the situation in the Netherlands.

1.1 The goal of research

This thesis supports the PhD project of Maarten-Jan Kallen whose aim it is to estimate the probabilities of degradation of bridges in the Netherlands.

The goal of this thesis is to create a deterioration model for use in bridge condition assessment.

The probabilistic method which is used to predict deterioration is the Markov chain. This model is condition-based and fits well, because all bridges are inspected visually and results are collected in a discrete condition scale. The heart of Markov model is a transition probability matrix, which is based on transition probabilities between all condition states. Hence, the main idea is to determine a matrix which describes the deterioration process during one year, the one-year transition probability matrix. The initial form of the matrix can be simplified since it is accepted that deterioration is a one-way process.

To start working with the Markov chain, we need to verify the Markov property for the data set. This property entails that the probability of deteriorating to another state doesn't depend on the history of the process, but only depends on the last condition. We would like to incorporate the effect of grouping bridges into sensible groups and to take into account the subjectivity of the inspectors into condition rating. One of the goals is to mark the essential factors which influence the deterioration process.

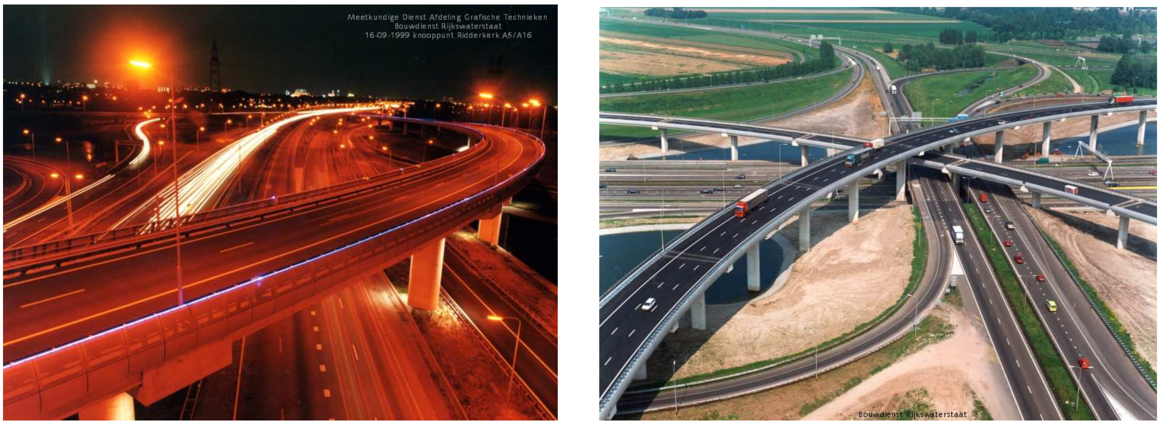

Figure 1-1 Bridges in the Netherlands (taken from the website www.bouwdienst.nl)

The present research work is performed under general supervision of HKV Consultants in cooperation with The Civil Engineering Division of Rijkswaterstaat, which provided data and essential instructions for performing this work.

HKV Consultants is an independent consulting company providing consultancy and research services in the field of water management and water-management related problems. The activities of HKV research include desk studies, software engineering, project management and turn-key system implementation.

1.2 Approach and assumptions

In this research, a deterioration model for bridges in the Netherlands is developed. This model is based on a Markov chain with a discrete condition scale. The transitions in the Markov chain occur at discrete units of time, usually every year.

The condition of bridges is represented by a number from 0 to 6, where 0 is perfect and 6 is extremely bad. From the database we extracted information to perform this analysis. Most importantly, we have information about the transitions from one condition to another and the duration between the observations, as well as external factors which could be important in the analysis. This information can help us to determine (approximately) how fast a bridge moves from a perfect condition to a condition where maintenance is required.

This MSc project has identified the following five essential activities:

- To verify the Markov property

- To estimate a Markov chain
- To determine and compare the expected deterioration process.
- To separate bridges into sensible groups and to determine the effect of this grouping on the parameters of the Markov model
- To take into account the subjectivity of the inspectors into condition rating

For maintenance activities for any bridge, the information on the current state and potential future bridge condition is essential. Decisions about a structure's safety and maintenance needs are made based on the mixed information obtained in the field. By collecting data on a regular basis, it is possible to identify trends and to predict future condition states.

Currently, bridge inspection data is primarily qualitative and prediction models are obtained empirically based on statistical trends. Previous inspections can provide global information for the bridge. However we find some of the limitations in the data collected which are:

- Lack of transition times, the data contains information about states randomly-timed, inspections were performed randomly.
- The prediction models are based on statistical relationships of visual condition ratings and do not consider the cause of deterioration or damage as governing parameters.
- Data does not come from quantitative measurements but from visual inspections. The condition of a bridge is rated with respect to a general rule described in an inspection manual, but it can contain inspector’s subjectivity.

In this research we consider two different Markov models based on different approaches to model the transition probabilities. The first model assumes that transition probabilities do not depend on the state in which a given bridge is. The second model assumes dependence on the state. The results are expressed as one-year transition probability matrices which form the heart of the Markov chain. A comparison of the two different models is made.

1.3 Outline of the thesis

Firstly, Chapter 2 presents information about Markov chains, which is the basis of this research. The general idea of the method, the terms used, and description of the data are briefly presented.

Chapter 3 presents an overview of the statistical tests of hypotheses, which are used to test statistical significance of the Markov property and any other factors in later analysis.

The results of testing the Markov property are shortly presented in Chapter 4. This chapter contains only the main results and descriptions of three different methods. The full analysis is presented in the internship report of Skuriat-Olechnowska (2005).

Chapter 5 introduces two models that we develop to reach the goal of the research. We present here the influence of transition probabilities on the condition state. The last part of this chapter tests adjustment of these models.

In Chapter 6, data analysis is performed. This part of the project contains information about the effect of grouping bridges into sensible groups and expected conditions at given times. On these pages we estimate the most influential covariates for the transition probabilities.

Chapter 7 tries to take into account subjectivity of the inspectors into condition rating. The main formula is developed, but there is not yet an algorithm to solve it.

The conclusions of the analysis and recommendations for future research are presented in Chapter 8

2 Short introduction to Markov Chains

In this Chapter we introduce the reader to the theory of Markov chains, which is the basis of this research. The general idea of the Markov chain and description of the data are briefly presented.

2.1 Markov Chains

The Markov Chain is a discrete-time stochastic process $\{X_t, t = 0,1,2,...\}$ that takes on a finite or countable number of possible values. This set of possible values of the process will be denoted by the set of nonnegative integers $\{0,1,2,...\}$. When $X_t = i$, then the process is said to be in state i at time t. We assume that whenever the process is in certain state i, then there is a fixed probability P_{ij} that it will be in state j one time unit later. That is

$$P_{ij} = \Pr\{X_{t+1} = j \mid X_t = i, X_{t-1} = i_{t-1}, ..., X_1 = i_1, X_0 = i_0\} \tag{2.1}$$

For all states $i_0, i_1, ..., i_{n-1}, i, j$ and all $t \geq 0$. The Markov process has the very important property that the future development of a process is independent of anything that has happened in the past given its present value. In other words, this means that the conditional probability of moving to a state *j* only depends on the current state *i* and is independent of all states in the past:

$$\begin{aligned} P_{ij} &= \Pr\{X_{t+1} = j \mid X_t = i, X_{t-1} = i_{t-1}, ..., X_1 = i_1, X_0 = i_0\} \\ &= \Pr\{X_{t+1} = j \mid X_t = i\} = P_{ij}(t) \qquad t = 0,1,2,.... \end{aligned} \tag{2.2}$$

$P_{ij}(t)$ is called the transition probability from state i at time t to state j at time $t+1$.

If $P_{ij}(t)$ does not depend on *t*, we say that the Markov Chain is time homogeneous (i.e. does not depend on how long the process has been running).

$$P_{ij} = \Pr\{X_{t+1} = j \mid X_t = i\} \quad = \quad \underbrace{\Pr\{X_1 = j \mid X_0 = i\}}_{stationarity}$$

In practice, it is very difficult to estimate the parameters for a time-inhomogeneous Markov Chain; therefore most known maintenance applications like PONTIS (Golabi and Shepard, 1997; Thompson *et al.*, 1998) and Arizona Pavement Management System (Golabi *et al.*, 1982; Wang and Zaniewski, 1996), make use of time-homogeneous Markov chains. We restrict ourselves to the time-homogeneous chains unless otherwise mentioned.

More generally, for a finite Markov chain with *k* states, we need *(k+1) x (k+1) transition* probabilities. Let us we write these *(k+1) x (k+1)* transition probabilities as a matrix, called the transition probability matrix, as follows:

$$\mathrm{P} = \begin{bmatrix} p_{00} & p_{01} & \cdots & p_{0k} \\ p_{10} & p_{11} & \cdots & p_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k0} & p_{k1} & \cdots & p_{kk} \end{bmatrix} \tag{2.3}$$

This is called a one-step transition probability matrix, because it gives the transition probabilities for one time unit. Obviously, the probability of moving from one state to any other state (including itself) should be 1. In other words, the sum of each row in the matrix must be one:

$$\sum_{j=0}^{k} p_{ij} = 1 \text{ for } i = 0,1,...,k \quad \text{and} \quad p_{ij} \le 1 \,\&\, p_{ij} \ge 0$$

2.1.1 Chapman-Kolmogorov equation

In the previous section we defined the one-step probability matrix P. Now we would like to define the *m*-step probability matrix $\mathrm{P}^{(m)}$. The transition probability $\mathrm{P}_{ij}^{(m)}$ in this matrix is the probability that a process in state i will be in state j after m additional units of time. That is,

$$\mathrm{P}_{ij}^{(m)} = \Pr\{X_{n+m} = j \mid X_n = i\}, \qquad m \ge 0, i, j \ge 0$$

Of course $\mathrm{P}_{ij}^{(1)} = \mathrm{P}_{ij}$. In order to arrive at state *j* from state *i* after *m+k* time steps, the Markov chain has to occupy certain state after *m* steps, say, *n*, of course *n* may possibly be any state in the state space. Therefore for all states *i, j* and for all $m, k \ge 0$, we have

$$P_{ij}^{(m+k)} = \sum_n P_{in}^{(k)} P_{nj}^{(m)} \tag{2.4}$$

This equation is called the ***Chapman-Kolmogorov*** equation. $P_{in}^{(k)} P_{nj}^{(m)}$ represents the probability that starting in i the process will go to state j in *m+k* transitions through a path which takes it in state *n* at the *k*th transition. Hence summing over all intermediate states *n* we will get probability that process will be in state j after *m+k* transitions.

Now we verify (2.4)

$$\begin{aligned} P_{ij}^{(k+m)} &= \Pr\{X_{k+m} = j \mid X_0 = i\} = \sum_n \Pr\{X_{k+m} = j, X_k = n \mid X_0 = i\} \\ &= \sum_n \Pr\{X_{k+m} = j \mid X_k = n, X_0 = i\} \cdot \Pr\{X_k = n \mid X_0 = i\} \\ &= \sum_n \Pr\{X_{k+m} = j \mid X_k = n\} \cdot \Pr\{X_k = n \mid X_0 = i\} = \sum_n P_{in}^{(k)} P_{nj}^{(m)} \end{aligned}$$

Using this rule we can show that the *m*-step transition probability matrix $\mathrm{P}^{(m)}$ can be calculated by multiplying the one-step transition matrix *m* times with itself: P^{m}.

$$\textit{i.e}\quad P^{(m+k)} = P^{(m)} \cdot P^{(k)} = P^{(k)} \cdot P^{(m)}$$

hence

$$P^{(m)} = P^{m-1}P^{1} = P^{m-2}PP = \ldots = \underbrace{P \cdots P}_{m\,\text{times}} = P^{m}.$$

The Markov chain always performs a transition after each unit of time. The time between transitions is deterministic and is set by the decision-maker based on available data or expert opinion. No transition is the same as a transition to the same state, which is sometimes referred to as a virtual transition.

Note: $P_{ii}^{0} = 1$ and $P_{ij}^{0} = 0,$ if $i \neq j$

2.2 Bridge inspection data

This data contains information about the condition states at which the structure was during inspections (inspection history) and on the year of construction.

State	Structure rating
0	Perfect
1	Very good
2	Good
3	Reasonable
4	Mediocre
5	Bad
6	Very bad

Table 2-1 Condition rating scheme

We see that the information, which comes from deterioration data, is quite subjective. We see condition rating ranging from a perfect (state 0) to a very bad (state 6) through the definition of states. The data comes from the Civil Engineering Division of the Ministry of Transport, Public Works and Water Management in the Netherlands.

The database includes a total of 5986 registered inspection events for 2473 individual superstructures. Ignoring the time between the construction of the bridge and a first inspection, there are 3513 registered transitions between condition states.

The next table shows the count of the transitions from each state to any other state.

From \ To	0	1	2	3	4	5	6
0	520	134	327	111	36	7	0
1	270	128	222	97	36	7	0
2	284	101	368	193	61	9	5
3	94	33	119	131	42	3	1
4	16	14	42	50	17	7	0
5	7	3	4	4	3	0	1
6	1	1	0	3	1	0	0

Table 2-2 Count of transitions

From this table we may observe that states 5 and 6 rarely occur in the database. For determining a transition probability matrix, these states are combined into state 5 to represents a condition 'bad' and 'very bad' (otherwise some transition probabilities might be zero).

Assume that a structure can be in any one of *k>0* discrete states, where state 0 is the initial state. The probability of moving from state *i* to state *j* in a single unit of time is defined as:

$$P_{ij} = \Pr\{X_{t+1} = j \mid X_t = i\} \quad \text{for} \quad i, j = 0,1,2,...,k \qquad (2.5)$$

Of course it must hold that $1 \geq P_{ij} \geq 0$ and $\sum_{j=0}^{k} P_{ij} = 1$. These transition probabilities can be calculated for each possible pair of states and represented by a $(k+1)\times(k+1)$ transition probability matrix:

$$P = \begin{bmatrix} 0.4581 & 0.1181 & 0.2881 & 0.0978 & 0.0317 & 0.0062 \\ 0.3553 & 0.1684 & 0.2921 & 0.1276 & 0.0474 & 0.0092 \\ 0.2782 & 0.0989 & 0.3604 & 0.1890 & 0.0597 & 0.0137 \\ 0.2222 & 0.0780 & 0.2813 & 0.3097 & 0.0993 & 0.0095 \\ 0.1096 & 0.0959 & 0.2877 & 0.3425 & 0.1164 & 0.0479 \\ 0.2857 & 0.1429 & 0.1429 & 0.2500 & 0.1429 & 0.0357 \end{bmatrix} \qquad (2.6)$$

These probabilities were estimated by the formula:

$$\hat{P}_{ij} = N_{ij} \Big/ \sum\nolimits_{j=0}^{k} N_{ij} \text{ ,} \qquad (2.7)$$

where N_{ij} is the number of observed transitions from state *i* to *j*.

Here we want to refer to the paper of Anderson and Goodman (1957). This paper shows how to estimate Markov Chains and derives asymptotic properties of the maximum likelihood estimator in (2.7).

The estimator is very simple: calculate the number of superstructures that moved from state *i* to state *j*. Let this number be N_{ij}, then we enter this number in a two-way table. The estimate of $\hat{P}_{ij}$ is the *ij*-th entry in the table divided by the sum of the entries in the *i*-th row.

3 Chi-Square Test

This chapter gives attention to the Chi-Square Test, which was used to verify the significance of the Markov property. Below we explain this statistical test, its main assumption and corresponding parameters. Below, we will give description of the way, in which we build hypothesis, as well as definition of the degrees of freedom.

3.1 Description of the test

To test the statistical significance of the Markov property assumption, we used a Chi-Square test. All tests of statistical significance involve a comparison between:

- An observed value (for example an observed number of heads in N tosses of a coin; an observed number of certain strings in N strings; and so on.)
- The value that one would expect to find, on average, if nothing other than chance coincidence, mere random variability, were operating in the situation (expected value).

Chi-Squared tests measures the proportionate amount by which each observed frequency deviates from its corresponding expected frequency.

The effect of this will be the formula:

$$\frac{\left(observed\ \ frequency - expected\ \ frequency\right)^2}{expected\ \ frequency} = \frac{(O_i - E_i)^2}{E_i} \tag{3.1}$$

Hence summing over all possible outcomes, according to the formula (3.1), we get

$$Q = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \tag{3.2}$$

The value of the statistic Q is asymptotically Chi-Square distributed; with *n* degrees of freedom

More generally, if Q_i are independently distributed according to a Q distribution with $r_1, r_2, ..., r_k$ degrees of freedom, then

$$Q = \sum_{j=1}^{k} Q_j \tag{3.3}$$

is distributed according to Q with $r = \sum_{j=1}^{k} r_j$ degrees of freedom. (Lowry, internet website)

Figure 3-1 shows the theoretical sampling distribution of a chi-square statistic. In this distribution, the critical value of the chi-square statistic for significance at the P = 0.05 level is Q = 5.99. That is, of all the possible values of a chi-squared variable that might have resulted in this situation, only 5 percent would have been equal to or greater than 5.99. Hence, an observed chi-square value precisely equal to 5.99 could be said to be significant *at* the 0.05 level.

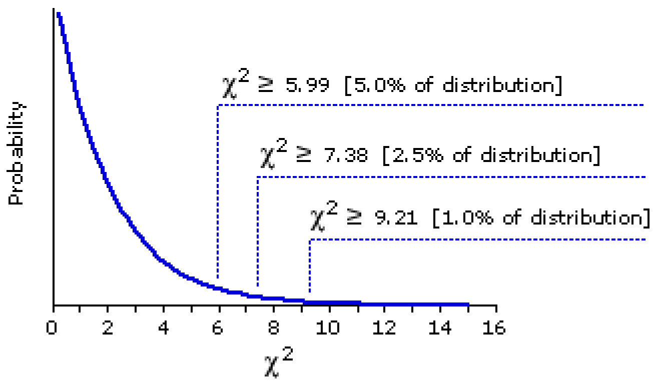

Figure 3-1 Theoretical sampling distribution of chi-square (df=2)

Figure 3-1 presents the theoretical sampling distribution of the chi-squared statistic as well as the critical values of this distribution. It shows also a significance level at the 0.025 (2.5 percent) and 0.01 (1 percent) level.

3.2 Degrees of freedom

The appropriate sampling distribution of a chi-square statistic is determined by a general property called **degrees of freedom**, denoted by **df**. Degrees of freedom, df, is simply an index of the amount of random variability that can be present in a particular situation. Its closest literal translation would be something along the line of "degrees of arbitrariness."

Suppose we have three cells such as the ones shown below, and we are free to plug any integer numbers that we want into them, subject only to the condition that the sum of the three numbers must be equal to certain specified quantity. For purposes of illustration we will set the sum at 10, though it could actually be any positive integer value.

a □ + b □ + c □ =10

We can plug numbers arbitrarily into any two of the cells, but once those cells are plugged in the value of the third is rigidly fixed. Thus, plug 6 into any one cell and 2 into either of the remaining two, and the value of the third is then fixed as 10 - 6 - 2 = 2. So here, with three cells, your degrees of freedom would be equal to two. The same logic then extends to cases where the number of cells is two, four, five, six, and so on. When applying chi-square procedures to situations in which there is only one dimension of categorization, the general principle for determining degrees of freedom is

df = (number of cells) – 1

In situations, in which observed items are sorted according to two different dimensions, the principle for determining the degrees of freedom is
df = (r-1)(c-1)

where r = number of rows and c = number of columns in a contingency table (Lowry, internet website). This table is constructed in a very simple way and an example is given in later analysis.

3.3 Statistical significance

Every time when we perform any test for statistical significance, we need to specify the level of it. In this subparagraph we would like to explain, what this means.

Consider the event $\delta \equiv \{z \geq z_{observed}\}$. A value $0 \leq \alpha \leq 1$ such that $P(\delta) \leq \alpha$ is considered "significant" (i.e., is not simply due to chance) and is known as an alpha value. The probability that a variate would assume a value greater than or equal to the observed value strictly by chance, $P(\delta)$, is known as a p-value.

Depending on the type of data and conventional practices of a given field of study, a variety of different alpha values may be used. Usually the following terminology is used:

- $P(\delta) \geq 5\%$ as "not significant”,
- $1\% \leq P(\delta) \leq 5\%$ as "significant" and
- $P(\delta) < 1\%$ as "highly significant".

In this thesis we write significance level as $P(X \geq Q) = P(\delta) \leq \alpha$, where X denote our variable coming from a Markov chain and Q denote a realization of the χ^2- distribution. If we assume 5% significance level, then for $P(X \geq Q)$ less than 5%, we reject the hypothesis.

Figure 3-2 shows the probability distribution that appears in Figure 3-1 for df = 2 with two other members of the family of chi-square distributions, the ones for df = 3 and df = 4.

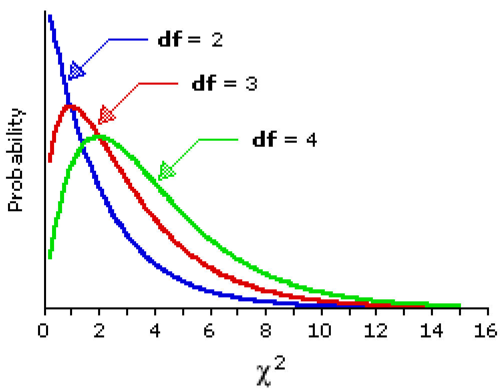

Figure 3-2 Chi-Square sampling distribution for df=2, 3 and 4

In the practical application of chi-square tests, we do not need to examine every detail of the relevant sampling distribution; we have to know the critical value that a calculated value of the chi-square statistic must equal or exceed in order to be judged statistically significant. A full set of such critical values is listed in Appendix A, and a somewhat abbreviated version of the same set is shown here in Table 3-1.

Level of Significance (non-directional test)

df	0.05	0.025	0.01	0.005	0.001
1	3.84	5.02	6.63	7.88	10.83
2	5.99	7.38	9.21	10.6	13.82
3	7.81	9.35	11.34	12.84	16.27
4	9.49	11.14	13.28	14.86	18.47
5	11.07	12.83	15.09	16.75	20.52
--	--	--	--	--	--
10	18.31	20.48	23.21	25.19	29.59
11	19.68	21.92	24.73	26.76	31.26
--	--	--	--	--	--

Table 3-1 Partial Table of Critical values of Chi-Square

For the moment we will confine our attention to Table 3-1 . The first column on the left in this table, labelled df, lists various values of degrees of freedom; the row across the top lists various levels of significance (P = 0.05, 0.025, etc.); and each of the other entries indicates the critical value of chi-square that an observed value must meet or exceed in order to be judged significant at a given level of significance for a given value of df. Thus, for df = 2 the minimum value of chi-square required for significance at the basic P =0.05 level is 5.99; for df = 3 it is 7.81; for df = 4 it is 9.49; and so on. For the more stringent significance level of P = 0.025, the required value of chi-square is 7.38 for df = 2; 9.35 for df = 3; 11.14 for df = 4; and so on.

In applying chi-square and other statistical tests, it is conventional to speak of a given result as being either non-significant, or significant *at* a certain level, or significant *beyond* a certain level. The illustration in below table for two degrees of freedom will give us an idea of what these three phrases mean and how they are used.

If the observed value of chi-square is:	Then it is:
smaller than 5.99	non-significant
equal to 5.99	significant at the .05 level
greater than 5.99	significant beyond the .05 level
equal to 7.38	significant at the .025 level
greater than 7.38	significant beyond the .025 level
equal to 9.21	significant at the .01 level
greater than 9.21	significant beyond the .01 level
etc	etc.

Table 3-2 Illustration of significance for df=2

3.4 The Null Hypothesis and the Research Hypothesis

In this section, we will be use some familiar items of symbolic notation like "=" and "≠". They should be read as follows:

a = b	a and b will not significantly differ; that is, a will be equal to b, within the limits of random variability;
a ≠ b	a and b will significantly differ (there will be a difference between a and b that goes beyond what could be expected on the basis of mere random variability);

When performing a test of statistical significance, it is useful convention to distinguish between

i. The particular hypothesis that the research is seeking to examine; and
ii. The logical antithesis of the research hypothesis (alternative hypothesis);

The second of these items is commonly spoken as the null hypothesis, where the word "null" in this context has the meaning of "zero". A conventional symbolic notation is H_0, which is H for "hypothesis", with a subscripted zero, to denote that it is "null". The research hypothesis sometimes is spoken of as the experimental hypothesis or alternative hypothesis, and is denoted as H_1. In general, the null hypothesis is to the effect that the observed results will not significantly differ from expected values. In other words, that will be equal to expected values, within the limits of random variability:

H_0 : observed value **=** expected value

And the alternative hypothesis is to the effect that there will be a significant difference between the observed and expected results.

H_1 : observed value ≠ expected value

The version of H_1 that we have just given is non-directional, because it affirms that there will be a difference between the observed and expected results, but makes no claim about the particular direction of the difference.

4 Verification of the Markov property

In order to test the Markov property we need to verify if the transition probabilities $\Pr\left\{X_{t+1} = m \mid X_t = j, X_{t-1} = i\right\}$ from the present to future state don't depend on past states. Written in a reduced notation: $P(m \mid j,i)$ = the probability of going to state *m*, given state *j* and *i* which occurred previously to *j* ($i \le j \le m$). If the Markov property holds, then $P_{ijm} = P_{jm}$.

In this chapter we will present three different tests, which we used to test the Markov property:

- A test based on the contingency tables,
- A test to verify if a chain is of given order, and the third,
- A test to verify if the transition probabilities are constant in time.

Each of them has its own interpretation and may give us a different result. Below we a present short description of each of these tests and give a summary of the results. To test the significance of the Markovian assumption, we use the chi-square statistic.
In this chapter we present only the main results. For all analysis we refer the reader to the internship report of the author [Skuriat-Olechnowska, 2005].

4.1 Test based on contingency tables

This case involves an analysis of at least two different three-state transition sequences. Three-state transition sequences consist of three-condition states: past, present and future. These condition states correspond to the bridge condition ratings occurring over three inspections. Possible transition sequences, with the same present and future states, but with different past states, are tracked to determine if there is a difference in occurrence dependent on the past state history. Here we used a simple frequency analysis of sequence occurrence. If there is no significant difference in frequency between the sequences being tracked, this may indicate the Markovian property.

We used the following definitions for the informal analysis to state transition sequences based on frequency probabilities (Scherer and Glagola, 1994):

- State transition sequence (STS): this is a particular three-state sequence occurrence, it involves a past, present and a future state;
- State sequence occurrences (SSO): this refers to the number of times a specified STS appears in the available database; number of (*i,j,m*)
- Two-state occurrence (TSO): TSO refers to the number a specified two-sequence appears in the available database. This sequence involves a past and a present state; number of (*i,j*)
- Frequency probability: This refers to a ratio of SSO over TSO:

$$P(m \mid j,i) = \frac{SSO}{TSO}$$

$$P(m \mid j,i) = \frac{[\sum (m,j,i) occurrences]}{\sum [(j,i) occurrences]}$$

After this short introduction we can perform a data analysis. Below we present only one, from many contingency tables, which was used to test the Markov property. Let's denote "i" as the past state, "j" as the present (now) and "m" as a future state.

State transition sequence (STS) *(*i, j, m*)*	State sequence occurrences (SSO)	Two-state occurrences (TSO)	Probability $\frac{\text{SSO}}{\text{TSO}}$
001	20	218	0.09
101	14	116	0.12
201	16	99	0.16
301	4	29	0.14
401	1	6	0.17
501	2	5	0.4
601	0	1	0

Table 4-1 Frequency of state transition sequences

Table 4-1 is an example frequency table. We can observe here how they are constructed. The first column contains the sequence which we want to consider, the second column shows us the frequency of occurrence of each sequence, while the third one shows the frequency of the remaining sequences for this group. The last column contains probabilities computed as a ratio of state sequence occurrences and two- state occurrences. Because probabilities in the last column look very similar, we can guess that the Markov property holds for this group of sequences (sequences which are going from present state 0 to future state 1 but start from different past states.)

To be sure of our decision, we will perform a test. The general null-hypothesis in this scenario is that the generated frequency distributions for a specified case are from the same distribution. Verifying this null-hypothesis based on the chi-squared statistic gives insight into the concept of accepting the Markovian property for the bridge condition deterioration data.

The contingency table, which we built, looks like the one in Table 4-2.

Category	Sequence occurrences	Nonsequence occurrences	Total
1 2 3 . . . h	N_{11} N_{21} . . . .. N_{h1}	N_{12} N_{22} . . . . N_{h2}	$N_{1.}$ $N_{2}.$ . . . . $N_{h}.$
Total	$N_{.1}$	$N_{.2}$	n

Table 4-2 Two-way contingency table.

In the above table, Category means state transition sequence, i.e. sequence of states for which we want to test the Markov property. The second column, i.e. sequence occurrences, contain the number of these sequences, while third column, denoted as nonsequence occurrences, contain count of all two-state sequences without these denoted in category, in other words, this is the difference between two-state occurrences and state sequence occurrences given in Table 4-1.

The notation is as follows:

- $N_{i.} = \sum_j N_{ij}$ - a row total;
- $N._j = \sum_i N_{ij}$ - a column total;
- $\sum_i N_{i\cdot} = \sum_j N_{\cdot j} = n$ - total number of elements in all rows and columns;

For this kind of tables the chi-squared statistic is given by:

$$Q = \sum_{i,j} \frac{[N_{ij} - n(N_{i\cdot}/n)(N_{\cdot j}/n)]^2}{n(N_{i\cdot}/n)(N_{\cdot j}/n)} \tag{4.1}$$

and has approximately the chi-squared distribution with (h-1) degrees of freedom, because we have h rows and only two columns.

The statistic Q of Eq. (4.1) has intuitive appeal. N_{ij} is the observed number in the *ij*-th cell, and $n(N_{i\cdot}/n)(N_{\cdot j}/n)$ is an estimator of the expected number in the *ij*-th cell when H_0 is true. Thus, Q will tend to be small for H_0 true and large for H_0 false.

Creating the contingency table presented above, the following chi-squared value was calculated:

Value of Q	**Degrees of freedom**	$P(X > Q)$	**Decision for H_0**
7.3194	6	0.2923	not reject

Table 4-3 Value of test for contingency table.

Remember that we assume the significance level to be 0.05. In this case we can not reject the hypothesis. This would indicate that we might assume the Markov property for this group of sequences.

The results for all data, for all tables with significant number of frequencies, the reader can find in my internship report. For most of these tables the hypothesis that the data satisfies the Markov property was not rejected, hence we may assume that the Markov property holds for all data.

4.2 Hypothesis test on the order of the Markov chain

In explaining the derivation of the test, which we would like to use in this paragraph, we follow Anderson and Goodman (1957, pp. 89-109) Assume that the chain is stationary, i.e. that the transition to a certain state doesn't depend on the time at which this transition took place, i.e. $p_{ijm}(t) = p_{ijm}$ for all t.
Denoting the probability of being in state m at time t given that an individual was in state i at time $t-2$ and in j at time $t-1$, by $p_{ijm}(t)$ $(i,j,m = 0,1,2,\ldots,k,\ t = 2,3,\ldots,T)$, we have that

$$p_{ijm} = \Pr\left\{X(t) = m \mid X(t-1) = j, X(t-2) = i\right\}.$$

This formula tells us that the probability of going to the next state depends on the state where it is now and on the past state. When a chain depends on the past state then we say that the chain is of second order.

A first-order chain is a special case of a second-order chain, i.e. a $p_{ijm}(t)$ does not depend on i *(past)* and then

$$p_{ijm} = p_{jm} = \Pr\left\{X(t) = m \mid X(t-1) = j\right\}$$

Now let n_{ijm} be the number of bridges in state i at $t-2$, in j at $t-1$, and in m at t, then the maximum likelihood estimate of p_{ijm} for a stationary chain is

$$\hat{p}_{ijm} = n_{ijm} / \sum_{l=0}^{k} n_{ijl}$$

According to Anderson and Goodman (1957), the test which verifies if the chain is of second order or not, is defined as follows:

Let us define a hypothesis

- H_0 : that the chain is of first-order, this mean that $p_{0jm} = p_{1jm} = p_{2jm} = \ldots = p_{kjm} = p_{jm}$, say, for $j,m = 0,1,2,\ldots,k$.
- H_1 : that it is of second-order.

The chi-squared statistic for the null hypothesis H_0 is

$$Q_j = \sum_{i,m} n^*_{ij} (\hat{p}_{ijm} - \hat{p}_{jm})^2 / \hat{p}_{jm} \qquad (4.2)$$

where

$$\hat{p}_{jm} = \sum_{i=0}^{k} n_{ijm} \Big/ \sum_{i=0}^{k}\sum_{l=0}^{k} n_{ijl} \tag{4.3}$$

and

$$n_{ij}^{*} = \sum_{m} n_{ijm} \tag{4.4}$$

Hence if the H_0 is true, Q_j has a chi-square distribution with $(k-1)^2$ degrees of freedom.
This test relates to the contingency table approach dealing with a given value of the present state j. Hence the hypothesis can be tested separately for each value of *j*. The joint hypothesis that $p_{ijm} = p_{jm}$ for all *i, j, m* = 0,1, 2, …, k , can be obtained by computing the sum

$$Q = \sum_{j=1}^{k} Q_j \quad , \tag{4.5}$$

which has a limiting Chi-square distribution with $k(k-1)^2$ degrees of freedom. Below we present only example of this calculation and the result for it.

past	**present**	**future**						
		0	1	2	3	4	5	6
0	0	111	20	74	9	3	1	0
1	0	46	14	36	18	2	0	0
2	0	31	16	32	14	6	0	0
3	0	11	4	7	6	1	0	0
4	0	3	1	2	0	0	0	0
5	0	3	2	0	0	0	0	0
6	0	1	0	0	0	0	0	0

Table 4-4 Contingency table to test the order of Markov chain.

This table shows the count of transitions for a present state 0 with dependence on any past and future state. We may observe that not every transition is observable from our data, which results in zeros. To calculate the probability matrix p_{ijm} and the corresponding probabilities p_{jm}, we need to know the count of each past state *i*. This was done in the previously mentioned internship.

For the Table 4-4 the variable of the chi-square statistic Q and the result of the test based on a significance level α , here also assumed at 5% level, are presented in Table 4-5.

Q_j	$P(X > Q_j)$	**Degrees of freedom**	**Decision for H_0**
42.3720	0.2153	36	not reject

Table 4-5 Value of test for table 4-4.

Table 4-5 informs us about the result of test. We see that for the given contingency table we can not reject the hypothesis that the chain is of first-order. This is good news for us, because this indicates that the Markov property holds for our data.

4.3 Testing the hypothesis of stationarity

This last test verifies our assumption of stationarity of the Markov chain. In the stationary Markov chain, p_{ij} is the probability that an individual (in our case, superstructure in a bridge) was in state *i* at time *t-1* and then moves to *j* at time *t.* A general alternative to this assumption is that the transition probability depends on *t* denoted by $p_{ij}(t)$ (Anderson and Goodman, 1957).

We test the null hypothesis:

$\mathrm{H}_0: \ p_{ij}(t) \ = \ p_{ij}$ *(t=1,2,3,…, T),* against
$\mathrm{H}_1: \ p_{ij}(t) \ \neq \ p_{ij}$ *(t=1,2,3,…, T).*

Under the alternative hypothesis, the estimates of the transition probabilities at time *t* are:

$$\hat{p}_{ij}(t) = \frac{n_{ij}(t)}{\sum_j n_{ij}(t)} \tag{4.6}$$

where $n_{ij}(t)$ denote the number of individuals in state *i* at *t-1* and *j* at *t*, $i, j = 0,1,2,...,k$ where $\mathrm{k} = 5$ in our case. We combined state 6 into 5 because, we have too small number of transitions to state 6. Now state 5 indicates states 'bad' and 'very bad'.

For a given past state i, the set $\hat{p}_{ij}(t)$ has the same asymptotic distribution as the estimates of multinomial probabilities $p_{ij}(t)$ for T independent samples. Hence a $k \times T$ table, looks like a contingency table and having the same formal appearance, can be used to represent the joint estimates $\hat{p}_{ij}(t)$ for a given *i* and for *j* =1, 2, …, k and *t*=1, 2, …, T.

In this analysis, time t will be grouped to avoid too wide age discrepancy, which is equivalent to a lack of data. Age groups are defined as follows:

- Group 0: ages between 0 and end at 4 years;
- Group 1: ages starts at 5 and ends at 9 years;
- Group 2: ages starts at 10 and ends at 14 years;
- Group 3: ages starts at 15 and ends at 19 years;
- Group 4: ages starts at 20 and ends at 24 years;
- Group 5: ages starts at 25 and ends at 29 years ;
- etc.
- The last group contain ages above 70 years old, this is Group 14.

In each of the above age groups, we count only transitions which starts at earlier group and end at a given. In other words, Group k contains transitions from Group k-1 to Group k. In this case, we do not consider transitions inside given group, hence we do not take into account Group 0. Group 1 contains transitions from Group 0 to Group 1, and so on.

We see that the discrepancy of groups is quite big and by considering them in this form, we lost information about intermediate states. However this is the only way to get a good approximation of the transition matrices. Below we present a way, how contingency tables should be constructed to perform this test:

t \ j	0	1	.	.	.	k
1 = Group 1	$\hat{p}_{i0}(1)$	$\hat{p}_{i1}(1)$	.	.	.	$\hat{p}_{ik}(1)$
2 = Group 2	$\hat{p}_{i0}(2)$	$\hat{p}_{i1}(2)$	.	.	.	$\hat{p}_{ik}(2)$
.	.	.		.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
T = Group 14	$\hat{p}_{i0}(T)$	$\hat{p}_{i1}(T)$	.	.	.	$\hat{p}_{ik}(T)$

Table 4-6 The contingency table for age dependent probabilities

Group 14, contains only one transition, hence we will build above table only for 13 age groups.

As we see in each row we place an estimated probability of going from a fixed past state i to the other future states for a given age group. This indicates that we have to consider each of six past states separately and for each of them build this kind of table like the one presented in Table 4-6.

To test stationarity, we build a null hypothesis that the random variables represented by the T rows have the same distribution, so that the data are homogeneous in this respect. This is equivalent to the hypothesis that there are k = 6 constants p_{i0}, p_{i1}, . . . , p_{i5}, with $\sum_j p_{ij} = 1$, such that the probability associated with the *j*th column is equal to p_{ij} in all T rows; that is, $p_{ij}(t) = p_{ij}$ for t = Group1, …, Group13.

The chi-square test of homogeneity seems appropriate here, hence in order to test this hypothesis, we calculate

$$Q_i = \sum_{t,j} \frac{n_i(t-1)[\hat{p}_{ij}(t) - \hat{p}_{ij}]^2}{\hat{p}_{ij}} \tag{4.7}$$

where $n_i(t-1) = \sum_j n_{ij}(t)$ and contain count of transitions from state i at time t, i.e. count of transitions from state i inside age Group t . If the null hypothesis is true, Q_i has a limiting Chi-square distribution with $(k-1)(T-1)$ degrees of freedom. Summing over all Q_i we get, that

$$Q = \sum_{i=1}^{k} Q_i = \sum_i \sum_{t,j} \frac{n_i(t-1)[\hat{p}_{ij}(t) - \hat{p}_{ij}]^2}{\hat{p}_{ij}} \qquad (4.8)$$

has the usual limiting chi-square distribution with $k(k-1)(T-1)$ degrees of freedom.

For each of the age groups, we should calculate the probability matrix $\hat{p}_{ij}$ (which we calculated in the same way as the matrix of transition probabilities given in (2.6)) and time dependent probability matrices $\hat{p}_{ij}(t)$, which looks like a contingency table presented in Table 4-6.

We don't want to give here all the results, but only a one of them. Below we present the results of this test.

Past state	Q_j	$P(X > Q_j)$	**Degrees of freedom**	**Decision for H_0**
0	11.6886	1	60	Not reject
1	8.6046	1	60	Not reject
2	90.0564	0.0073	60	reject
3	13.7443	1	60	Not reject
4	33.7709	0.9975	60	Not reject
5	0.6071	1.0000	60	Not reject
Total	**158.4719**	**1**	**360**	Not reject

Table 4-7 Results of test for constant probabilities.

For all values of Q_j, we calculate the probability $P(X > Q_j)$, which should be less than or equal to the significance level α**.** Remember that we reject the hypothesis at the 5% level, we can not reject this hypothesis; this means that the p_{ij} are constant and do not depend on time t, i.e. group of age. In above table we can write 'accept' instead 'not reject', because the p-value $P(X > Q_j)$ of each Q_j is very big, hence we may by more sure in the null hypothesis.

This result indicates that our chain is stationary and our assumption made at the beginning of this analysis was correct. On the other hand, we lost a lot of information coming from transitions inside of the groups. We didn't take them into consideration. For a proper analysis, we should use observations coming from each year, but in our data, we don't have this information. Here, inspections are performed periodically, and we do not observe every transition on it occurs.

4.4 Conclusions

This chapter was committed to present the method of test the Markov property for the DISK database. In the previous pages we have presented 3 different tests, from which two of them verify the Markov property and one of them verifies the assumption of stationarity. Each of them has its own interpretation and gives a different result. The first test, based on the contingency tables, verified the dependence on past states. The second verifies whether a chain is of given order and the third, verifies whether transition probabilities are constant in time.

The first test, which we performed, is based on contingency tables and it partially resulted in Table 4-4. This table indicates that we can not reject the null hypothesis that the generated frequency distributions for a specified case are from the same distribution, hence we may say that we accept it. This would indicate that we might assume the Markov property for transitions from state 0 to 1. Indeed this test depends on how a contingency table is built and it doesn't give us an unambiguous answer to our question about the Markov property. For this reason, the next test was performed.

In the second test, we use a more theoretical assumption to avoid ambiguities. The null hypothesis that a chain is of first order (doesn't depend on past state) was tested against the alternative hypothesis that the chain is of second order (depends on past state). An example in which we performed this test indicates that we can not reject the hypothesis that a chain from the DISK database is of first-order. This means that a transition to the next state doesn't depend on the past states, but only on the present. This indicates that the Markov property holds for the considered example. Of course, to make a decision about the Markov property, we have to consider many more examples. This was done during an internship and is presented in a different report.

The third test verifies the assumption about stationarity of the transition probabilities.
The null hypothesis means that transition probabilities are constant i.e. do not depend on time (in our case, on the age group). The results of this test are fully presented in Table 4-7 and the decision was made that we can't reject the null hypothesis.

Although the three hypothesis tests seem to work well, in practice their applicability might be limited. In the database we observe only transitions from one state to the other one periodically, and we don't know the exact time when this happens. Testing the Markov property we should know the exact times when a transition occurred. Also N_{ij}, the number of transitions from state *i* to state *j* is assumed to be after one year. The second and third tests follow the same restriction. We may claim that these tests are not completely suitable for periodically inspected structures, but they give a useful indication.

5 Model development

From this chapter, we will consider other data. In the previous chapter, we used data for superstructures, now we analyse data containing information about the general condition of the bridge. This data includes a total of 9892 registered transitions between condition states for 6871 individual bridges.

From \ To	0	1	2	3	4	5	6
0	302	935	597	309	46	45	16
1	25	332	1319	606	41	9	4
2	52	117	2127	1140	107	19	2
3	13	58	659	626	83	11	1
4	0	12	92	110	14	7	0
5	4	2	9	17	13	2	0
6	0	5	1	3	0	0	0

Table 5-1 Count of transitions for data for general condition

We decided to make this change, because this data contains more observations, and information about the age of the bridge as well as additional information about covariates, which we would like to use in later analysis.

Here we would like to assume that the Markov property holds for our data and to try to develop a one-year transition probability matrix. To do this we need to introduce some additional assumptions.

Let's assume that the bridge can move only from one state to the next state in one year. This means that, for example, the bridge can not go from state 1 onto state 3 during a one year period; it can only go from state 1 onto state 2. The next establishment, which was made relates to the initial state. We assume that every new bridge starts its deterioration from the perfect state. This means, every new bridge starts from state 0 and during next years deteriorates to higher states, finally ends in the worst condition. Of course, when the bridge reaches the worst state, state 5, it can not go anywhere else, hence it stays in this state with probability one. This is called an absorbing state. Because we want to develop a model which will describe the degradation process in bridges only, we have to assume that there is no maintenance. Hence, we deleted all transitions indicates maintenance. This means that we use the forward transitions and deleted all backward transitions. It's natural, that the bridge can remain in given state during some years, because the degradation process in concrete bridges is very long lasting. The last assumption which was made is stationarity. This allowed us to make simpler calculation.

In this chapter, firstly we will introduce the developed model and later calculate its parameters. We introduce also a definition of state-(in)dependent probabilities, because we use them in the analysis. Naturally, before we introduce a correct model, we will consider two different models, and later decide which of them fits better to our data.

5.1 Model

A Markov chain or Markov Process can be divided in two cases: Discrete-time (transitions are at certain points in time $\Delta t, 2\Delta t, 3\Delta t, \ldots$ where Δt denotes the time interval, say one year) and Continuous-time (where transitions occur at $t \geq 0$).

Continuous time Markov chains may be separated in regular continuous time Markov processes (with exponential waiting times) and Semi-Markov processes (with non-exponentially distributed waiting times having a probability distribution such as gamma, Weibull, …). Combining discrete-time with regular continuous time Markov processes we may observe a relationship which is illustrated as follow:

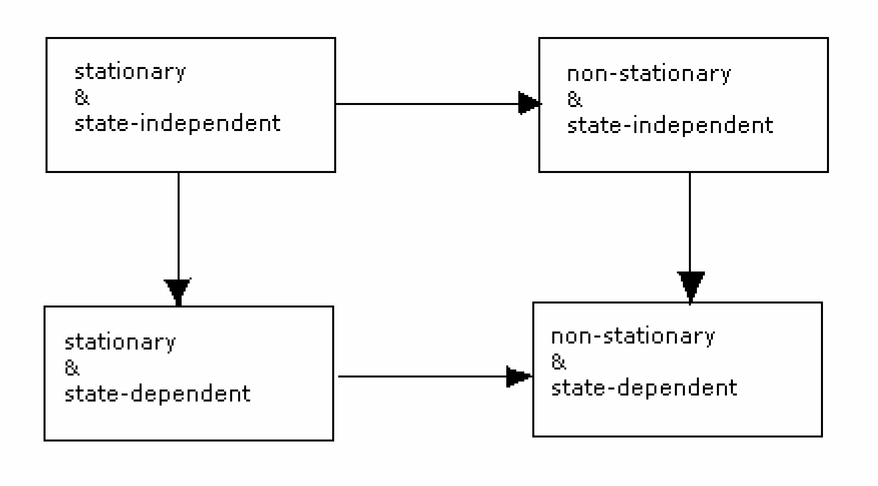

Figure 5-1 Graph of relationship between different Markov models

The first box (upper left) in Figure 5-1 represents the simplest Markov model, which can be considered. The most general and difficult case is presented in lower right corner of the above figure. The remaining two cases represent the mid-level between those two.

In the definition of the relationship between different Markov models, we used the formula 'state-(in)dependent'. This denotes the dependence between transition probabilities and states. We say that a transition probability is state-dependent if its value depends on the state from which a transition starts, i.e. p_{ij} depends on state i.

To estimate a one-year transition probability matrix, we need to make some assumptions. Let's suppose that in one year, the bridge can only make transition to the next state implying there is no repair. Of course, when it enters state number 5, then it will remain there. Hence the Markov chain has five transient states and one absorbing state (state nr 5). With these assumptions and under the principle that a transition to the next state doesn't depend on the state in which the chain is, the transition probability matrix looks like the one presented below:

$$\mathrm{P}=\begin{bmatrix} 1-p & p & 0 & \cdots & 0 \\ 0 & 1-p & p & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & \cdots & 0 & 1-p & p \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} \tag{5.1}$$

As we see, matrix (5.1) reflects all our assumptions. Of course in the rest of this analysis, the assumption about stationary transition probabilities is still valid.

Now we may substitute (5.1) into Figure 5-1 to get general information about the kinds of one-year Markov chains that can be used. This is presented in Figure 5-2.

	Stationary	Non-stationary
State-independent	$\begin{bmatrix} 1-p & p & 0 & \cdots & 0 \\ 0 & 1-p & p & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & \cdots & 0 & 1-p & p \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1-p(t) & p(t) & 0 & \cdots & 0 \\ 0 & 1-p(t) & p(t) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & \cdots & 0 & 1-p(t) & p(t) \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}$
State-dependent	$\begin{bmatrix} 1-p_1 & p_1 & 0 & \cdots & 0 \\ 0 & 1-p_2 & p_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & \cdots & 0 & 1-p_{n-1} & p_{n-1} \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1-p_1(t) & p_1(t) & 0 & \cdots & 0 \\ 0 & 1-p_2(t) & p_2(t) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & \cdots & 0 & 1-p_{n-1}(t) & p_{n-1}(t) \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}$

Figure 5-2 One-step transition probability matrices for four different cases.

The matrix

$$\mathrm{P}=\left\|P_{ij}\right\|=\begin{bmatrix} P_{00} & P_{01} & 0 & \dots & 0 \\ 0 & P_{11} & P_{12} & \dots & 0 \\ 0 & 0 & \dots & \dots & 0 \\ 0 & 0 & \dots & P_{k-1k-1} & P_{k-1k} \\ 0 & 0 & \dots & \dots & 1 \end{bmatrix} \text{ for } i,j=0,1,\dots,k\ . \tag{5.2}$$

will be denoted as a one-step transition matrix.

The m step transition matrix with stationary transition probabilities is given by

$$\mathrm{P}^{\mathrm{m}}=\mathrm{P}_{\mathrm{ij}}(m)=\Pr\left\{X_m=j \mid X_0=i\right\}=(\mathrm{P}^{\mathrm{m}})_{\mathrm{ij}} \tag{5.3}$$

In (2.2) we give a similar formula, but there it denotes a transition probability from state i at time t to state j at time $t+1$, what is simply a transition in one step at given time t. Here formula (5.3) denotes a transition probability in *m* steps.

In our analysis we can interpret *m* as a time between one inspection and the second one. In other words, if we observe a transition from state 2 to state 4 during 5 years, then this is equivalent to the situation when we observe $\Pr\{X_5 = 4 \mid X_0 = 2\}$. To calculate this probability we should use formula (5.3).

5.2 Maximum Likelihood Method to estimate transition probabilities

In this paragraph we would like to introduce the Maximum Likelihood Method. This method we used to estimate transition probabilities in a one-step transition probability matrix for the state-independent case given in (5.1) and for the state-dependent case, given in (5.12).

This method is the most general and popular method of estimation of the unknown parameter θ within a family F of distributions. Because each value of $\theta \in \Theta$ defines a model in F that attaches (potentially) different probabilities (or probability density) to the observed data, the probability of the observed data as a function of θ is called the likelihood function. Every value of θ that has high likelihood corresponds to a model which gives high probability to the observed data. The principle of maximum likelihood estimation is to adopt the model with greatest likelihood, since of all the models under consideration, this is the one that assigns highest probability to the observed data (Coles, 2001).

Now we would like to discuss this subject in greater detail. Assume that $x_1, x_2, ..., x_n$ are independent realizations of a random variable X having probability of getting our bridge deterioration dataset (in terms of continuous variables it will be density function) $f(x;\theta)$. Then, the **likelihood function** is

$$L(\underline{\theta}) = \prod_{i=1}^{n} f(x_i;\underline{\theta}) . \tag{5.4}$$

where $\underline{\theta} = (\theta_1, \theta_2, ..., \theta_d)$. In most cases, working with the likelihood function can be difficult. Therefore, it is often more convenient to use logarithms and to work with the **log-likelihood function**

$$\ell(\underline{\theta}) = \log L(\underline{\theta}) = \sum_{i=1}^{n} \log f(x_i;\underline{\theta}) . \tag{5.5}$$

Both (5.4) and (5.5) can be generalized to the situation where the X_i are independent, but not necessarily with identical distributions. In this case, denoting the probability of getting observation of X_i by $f_i(x_i;\underline{\theta})$, we obtain

$$L(\underline{\theta}) = L(\underline{x} \mid \underline{\theta}) = \prod_{i=1}^{n} f_i(x_i;\underline{\theta}) \tag{5.6}$$

and

$$\ell(\underline{\theta}) = \ell(\underline{x} \mid \underline{\theta}) = \log L(\underline{\theta}) = \sum_{i=1}^{n} \log f_i(x_i; \underline{\theta}) \tag{5.7}$$

The maximum likelihood estimator $\hat{\underline{\theta}}$ of $\underline{\theta}$ is defined as the value of $\underline{\theta}$ that maximizes the likelihood function. Since the logarithm function is monotonic, the log-likelihood takes its maximum at the same point as the likelihood function, so that the maximum likelihood estimator also maximizes the corresponding log-likelihood function.
For some situations it is possible to obtain the maximum likelihood estimator explicitly, generally by differentiating the log-likelihood and equating it to zero.

In the deterioration database observations for different bridges can be considered as independent, but the data for one bridge are realizations of a Markov chain, so they are not independent. Although, the Markov property allows us to calculate the likelihood function for a given bridge as a multiplication of its observations, hence the likelihood function for all data can be estimated in a following way: each observation x_k contains information about the past state, present state and the time between those inspections, which is collected in the triplet $\{i_k, j_k, m_k\}$. The likelihood function is then given by

$$L(\underline{x} \mid \underline{\theta}) = L(x_1, x_2, ..., x_n \mid \underline{\theta}) = \prod_{k=1}^{n} p_{i_k j_k}^{m_k} = \prod_{k=1}^{n} \Pr\left\{X_{m_k} = j_k \mid X_0 = i_k\right\} \tag{5.8}$$

where

- $\underline{\theta} = (p_{01}, p_{12}, ..., p_{45})$,
- $p_{i_k j_k}$ is the probability of going from state i_k to state j_k in one step (year) for observation k ,
- m_k is the number of years between two inspections for observation k ,
- $x_k = \{\text{going from } i_k \text{ to } j_k \text{ in } m_k \text{ steps (years)}\}$ for $i, j = 0,1,2,\ldots,6$, $n = 1,2,\ldots$ are observations for observation k ,
- n is the number of observations.

We have only a 5-dimensional parametric vector $\underline{\theta}$, because we have combined states 6 into 5 to have one state indicating the conditions bad and worse. We can rewrite the above likelihood formula in a more general way by using the whole probability matrix given in (5.2) as

$$L(x_1, x_2, ..., x_n \mid \underline{\theta}) = \prod_{k=1}^{n} (p_{i_k j_k}^{m_k}) = \prod_{k=1}^{n} (\mathrm{P}^{m_k})_{i_k j_k} \ , \tag{5.9}$$

where $(\mathrm{P}^{m_k})_{i_k j_k}$ means that we take matrix P to the power m and consider element $(i_k j_k)$ which coming from observation x_k.
Using the above results, the maximum likelihood estimator for $\underline{\theta}$ is given by

$$\hat{\underline{\theta}} = \max_{\underline{\theta}} L(\underline{x} \mid \underline{\theta}) = \max_{\underline{p}} \prod_{k=1}^{n} (\mathrm{P}^{m_k})_{i_k j_k}$$

For the log-likelihood case this equation is given by:

$$\underline{\hat{\theta}} = \max_{\underline{\theta}} \ \log L(\underline{x}\,|\,\underline{\theta}) \Rightarrow \underline{\hat{\theta}} = \max_{\underline{p}} \sum_{k=1}^{n} \log(\mathrm{P}^{m_k})_{i_k j_k} \qquad (5.10)$$

Equation (5.10) is equivalent with taking the derivative of the log-likelihood $\ell(\underline{\theta}) = \ell(\underline{x}\,|\,\underline{\theta})$ and equating it to zero:

$$\frac{\partial}{\partial \theta_j} \ell(\underline{\theta}) = 0 \quad \text{for} \quad j = 1, 2, ..., d$$

From this equation we derive the maximum likelihood estimator $\underline{\hat{\theta}}$ of $\underline{\theta}$.

For the first case of our analysis, we consider the situation with stationary and state-independent probabilities in the one-step transition matrix P. This means, that the probabilities of going to the next state are the same for each state. This matrix is given by:

$$\mathrm{P} = \begin{bmatrix} 1-p & p & 0 & \cdots & 0 \\ 0 & 1-p & p & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & \cdots & 0 & 1-p & p \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} . \qquad (5.11)$$

To find the maximum likelihood estimator for $\underline{\theta}$ in this simple case, we calculate the log-likelihood for each value of $p \in \{0.01, 0.02, ... ,0.99\}$ which comes from the interval (0, 1) and after this we choose the *p* which results in the highest likelihood. Alternative is numerically maximizing the log-likelihood function. Unfortunately, when we use a numerical procedure for unconstrained maximization, we can get that our estimator for $\underline{\theta}$ is not from the interval (0, 1). Therefore, we apply a simple transformation to assure that the estimator of $\underline{\theta}$ represents a probability.

To obtain values from a certain interval (a, b) for $x \in \mathbb{R}$, the following transformation is very useful:

Knowing that

$$\lim_{x\to\infty} \frac{b\exp(x)+a}{\exp(x)+1} = b \quad \text{and} \quad \lim_{x\to-\infty} \frac{b\exp(x)+a}{\exp(x)+1} = a ,$$

we can take a=0 and b=1 to get a value of *p* from the desired interval (0, 1)

$$p = \frac{b\exp(x)+a}{\exp(x)+1} = \frac{\exp(x)+0}{\exp(x)+1} = \frac{1}{1+\exp(-x)} .$$

Making a simple conversion, we derive a formula for x which is given by:

$$p(1+\exp(-x)) = 1 \Rightarrow \frac{1-p}{p} = \exp(-x) \Rightarrow x = -\log(\frac{1-p}{p}) = \log(\frac{p}{1-p}).$$

This transformation is called the logistic transformation. And now, using the logistic transformation, we optimise the log-likelihood function over x rather than p and compute p given the optimal x.

The above transformation will ensure that our maximum likelihood estimator $\underline{\theta}$ is in the interval (0,1).

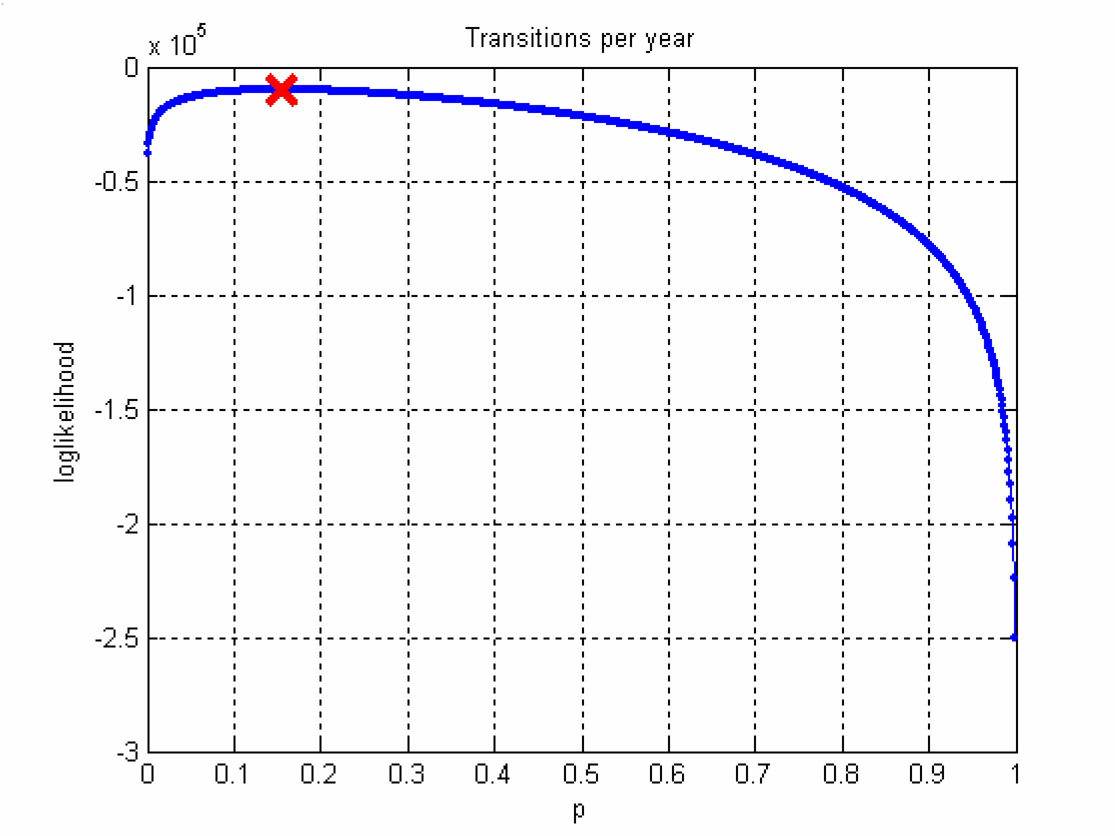

Figure 5-3 Log-likelihood function for state-independent probabilities and its maximum value

Figure 5-3 presents the log-likelihood function as a function of the transition probability *p*. We see that for small values of this probability, the log-likelihood function has higher values than for larger ones. The cross indicates the highest value of the log-likelihood function.

Our analysis gives us a maximum likelihood estimator for $\underline{\theta}$ equal to $\hat{\underline{\theta}}$ = p = 0.155.

For the second case, we consider the situation with stationary and state-dependent probabilities in the one-step transition matrix P. This means that the probabilities of going to the next state can be different for each state; that is,

$$P = \begin{bmatrix} 1-p_1 & p_1 & 0 & \cdots & 0 \\ 0 & 1-p_2 & p_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & \cdots & 0 & 1-p_{n-1} & p_{n-1} \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}. \quad (5.12)$$

For Equation (5.12), the maximum likelihood estimator for $\underline{\theta}$ is given by

$$\hat{\underline{\theta}} = [0.410, 0.310, 0.096, 0.033, 0.114].$$

Table 5-1 summarise the above results for both state-independent and state-dependent transition probabilities.

time	Maximum Likelihood Estimator for $\underline{\theta}$	
	State-independent	State-dependent
All data	p= 0.155	$\underline{\theta}$ =[0.410, 0.310, 0.096, 0.033, 0.114]

Table 5-2 Maximum likelihood estimators

From this table we can observe that, if we would assume state-independence, then all transition probabilities would be constant and equal to 0.155. In the event of state-dependence, these probabilities are different for each state. As we see, the difference in transition probabilities is quite substantial.

5.3 Confidence interval for estimated probabilities

Maximum likelihood parameter estimation is a very popular method. We can find widely applicable approximations for a number of useful probability distributions. These approximations include standard errors and confidence intervals, quality our confidence in the estimators. There are several useful results, which are not difficult to calculate. The framework is given by:

- $x_1, x_2, ..., x_n$ are independent realizations of a random variable *X* having distribution $\mathrm{F} \in F$.
- The family F is indexed by a *d*-dimensional parameter $\underline{\theta}$ and true distribution F has $\underline{\theta} = \underline{\theta}_0$.
- The maximum likelihood estimate of θ_0 is denoted $\hat{\theta}_0$.

Theorem 5.1 is based on an asymptotic limit law obtained for sample size *n* increases to infinity. They are valid only under regularity conditions, which say something about the existence of the first and second derivative of the log-likelihood function.
We will assume here that these conditions are met so that the accuracy improves as *n* increases.
The following theorem is very useful.

Theorem 5.1 Let $x_1, x_2, ..., x_n$ be independent realizations from a distribution within a parametric family F, and let $\ell(\cdot)$ and $\hat{\underline{\theta}}_0$ denote respectively the log-likelihood function and the maximum likelihood estimator of the *d*-dimensional model parameter $\underline{\theta}_0$. Then under suitable regularity conditions, for large n

$$\hat{\theta}_0 \sim \mathrm{MVN_d}(\underline{\theta}_0, \mathrm{I_E}(\underline{\theta}_0)^{-1}),$$

where

$$\mathrm{I_E}(\underline{\theta}) = \begin{bmatrix} e_{1,1}(\underline{\theta}) & \cdots & & e_{1,d}(\underline{\theta}) \\ \vdots & \ddots & e_{i,j}(\underline{\theta}) & \vdots \\ & e_{j,i}(\underline{\theta}) & \ddots & \\ e_{d,1}(\underline{\theta}) & \cdots & & e_{d,d}(\underline{\theta}) \end{bmatrix}$$

with

$$e_{i,j}(\underline{\theta}) = E\left\{-\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell(X \mid \underline{\theta})\right\} = nE\left\{-\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell(\underline{x} \mid \underline{\theta})\right\}$$

□

$\mathrm{MVN_d}(\underline{\theta}_0, \mathrm{I_E}(\underline{\theta}_0)^{-1})$ denotes the Multivariate Normal Distribution with mean vector $\underline{\theta}_0$ and variance-covariance matrix $\Sigma = \mathrm{I_E}(\underline{\theta}_0)^{-1}$. The joint density function of $\underline{x}$ has the form

$$f(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\underline{x} - \underline{\theta}_0)^T \Sigma^{-1}(\underline{x} - \underline{\theta}_0)\right\}, \underline{x} \in \mathbb{R}^d$$

The matrix $\mathrm{I_E}(\underline{\theta}_0)$, which measures the expected curvature of the log-likelihood surface, is usually referred to as the **expected Fisher information matrix** for a sample of *n* observations.

Theorem 5.1 can be used to obtain approximate confidence intervals for individual components of $\underline{\theta}_0 = (\theta_1, \theta_2, ..., \theta_d)$.
Denoting an arbitrary term in the inverse of $\mathrm{I_E}(\underline{\theta}_0)$ by $\psi_{i,j}$, it follows from the properties of the multivariate normal distribution that, for large *n*,

$$\hat{\theta}_i \sim N(\theta_i, \psi_{i,i}) .$$

Hence, if $\psi_{i,i}$ were known, an approximate $(1-\alpha) = 95\%$ confidence interval for θ_i would be

$$\hat{\theta}_i \pm z_{1-\alpha/2}\sqrt{\psi_{i,i}}$$

where $z_{1-\alpha/2}$ is the $100(1-\alpha)$ quantile of the standard normal distribution.

Because in most cases the true value of $\underline{\theta}_0$ is generally unknown, it is common to approximate the terms of $\mathrm{I_E}(\underline{\theta}_0)$ with those of the **observed Fisher information matrix** for a sample of *n* observations, defined by

$$\mathrm{I_o}(\underline{\theta}) = \begin{bmatrix} -\frac{\partial^2}{\partial\theta_1^2}\ell(\underline{x}\,|\,\underline{\theta}) & \cdots & & -\frac{\partial^2}{\partial\theta_1\partial\theta_d}\ell(\underline{x}\,|\,\underline{\theta}) \\ \vdots & \ddots & -\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell(\underline{x}\,|\,\underline{\theta}) & \vdots \\ & -\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell(\underline{x}\,|\,\underline{\theta}) & \ddots & \\ -\frac{\partial^2}{\partial\theta_d\partial\theta_1}\ell(\underline{x}\,|\,\underline{\theta}) & \cdots & & -\frac{\partial^2}{\partial\theta_d^2}\ell(\underline{x}\,|\,\underline{\theta}) \end{bmatrix}$$

This matrix is evaluated at $\underline{\theta} = \hat{\underline{\theta}}$, which is the maximum likelihood estimator (Coles, 2001). Denoting the terms of the inverse of this matrix by $\tilde{\psi}_{i,j}$ it follows that an approximate $100(1-\alpha)$ confidence interval for θ_i is

$$\hat{\theta}_i \pm z_{1-\alpha/2}\sqrt{\tilde{\psi}_{i,j}}$$

To calculate the first and second derivative of the log-likelihood function $\ell(\cdot)$, we will need the following definition.

Using the product rule, we may derive equation for the derivative of the matrix $\mathrm{P^m}$, which is given by:

$$\frac{\partial}{\partial\theta}\mathrm{P}_\theta{}^m = \sum_{i=0}^{m-1}\mathrm{P}_\theta{}^i \cdot \left(\frac{\partial}{\partial\theta}\mathrm{P}_\theta\right)\cdot \mathrm{P}_\theta{}^{m-1-i}. \tag{5.13}$$

As we see, the factors in the product are matrices and the matrix product is in general non-commutative, and that is why we get formula (5.13) and not simply

$$\frac{\partial}{\partial\theta}\mathrm{P}_\theta{}^m = m\cdot\mathrm{P}_\theta{}^{m-1}\cdot\left(\frac{\partial}{\partial\theta}\mathrm{P}_\theta\right). \tag{5.14}$$

Now we can derive equations for the derivatives of the maximum likelihood estimator of $\underline{\theta}$, which are given by the following:

$$\begin{aligned}\frac{\partial}{\partial\theta_u}\ell(\underline{\theta}) &= \frac{\partial}{\partial\theta_u}\sum_{l=1}^{n}\log(\mathrm{P}^{m_l})_{i_l\,j_l} = \sum_{l=1}^{n}\frac{\partial}{\partial\theta_u}\log(\mathrm{P}^{m_l})_{i_l\,j_l} = \sum_{l=1}^{n}\frac{1}{(\mathrm{P}^{m_l})_{i_l\,j_l}}\frac{\partial}{\partial\theta_u}(\mathrm{P}^{m_l})_{i_l\,j_l} \\ &= \sum_{l=1}^{n}\frac{1}{(\mathrm{P}^{m_l})_{i_l\,j_l}}\left(\sum_{k=0}^{m_l-1}(\mathrm{P}^k)\cdot\frac{\partial}{\partial\theta_u}\mathrm{P}\cdot(\mathrm{P}^{m_l-1-k})\right)_{i_l\,j_l}\end{aligned}, \tag{5.15}$$

where $\frac{\partial}{\partial \theta_u}\mathrm{P}$ has value -1 in point (i,i) and 1 in $(i,i+1)$ for $i=1,2,...,6$ and the rest of the matrix elements are zeros.

We also need the second derivative of the log-likelihood function to calculate the confidence interval:

$$\frac{\partial^2}{\partial\theta_u\partial\theta_v}\ell(\underline{\theta})=\frac{\partial}{\partial\theta_v}\left(\frac{\partial}{\partial\theta_u}\ell(\underline{\theta})\right)=\frac{\partial}{\partial\theta_v}\left(\sum_{l=1}^{n}\frac{1}{(\mathrm{P}^{m_l})_{i_l j_l}}\frac{\partial}{\partial\theta_u}(\mathrm{P}^{m_l})_{i_l j_l}\right)$$

$$=\sum_{l=1}^{n}\frac{(\mathrm{P}^{m_l})_{i_l j_l}\cdot\frac{\partial^2}{\partial\theta_u\partial\theta_v}(\mathrm{P}^{m_l})_{i_l j_l}-\frac{\partial}{\partial\theta_u}(\mathrm{P}^{m_l})_{i_l j_l}\frac{\partial}{\partial\theta_v}(\mathrm{P}^{m_l})_{i_l j_l}}{(\mathrm{P}^{m_l})_{i_l j_l}^{\;2}}$$

Because the expectation of $\frac{\partial^2}{\partial\theta_u\partial\theta_v}(\mathrm{P}^{m_l})_{i_l j_l}$ is equal to zero (Kalbfleisch and Lawless, 1985), we can write that observed information matrix is

$$\mathrm{I_o}(\theta)=\sum_{l=1}^{n}-\frac{\frac{\partial}{\partial\theta_u}(\mathrm{P}^{m_l})_{i_l j_l}\frac{\partial}{\partial\theta_v}(\mathrm{P}^{m_l})_{i_l j_l}}{(\mathrm{P}^{m_l})_{i_l j_l}^{\;2}}. \tag{5.16}$$

Now we can use a Newton-Raphson Method to estimate the maximum likelihood estimator for $\underline{\theta}$, which is the maximum of the function

$$L(x_1,x_2,...,x_n\,|\,\underline{\theta})=\prod_{l=1}^{n}(\mathrm{P}^{m_l})_{i_l j_l}\;.$$

This iterative process follows a set guideline to approximate one root, based on the function, its derivatives and an initial $\underline{\theta}$-value.
It starts at the initial, arbitrarily chosen $\underline{\theta}$-value, and calculates the new one using:

$$\hat{\theta}_{n+1}=\theta_n-\frac{\ell'(\theta_n)}{\ell''(\theta_n)}$$

where $\ell'(\theta_n)=\frac{\partial}{\partial\theta_u}\ell(\underline{\theta})$ and $\ell''(\theta_n)=\frac{\partial^2}{\partial\theta_u\partial\theta_v}\ell(\underline{\theta})$.

Because this is a numerical method, we need to specify some stopping point, at which the process will terminate. Therefore, let’s assume that the process will stop when an error between the new estimator and the previous one will become less than ε, i.e.

$\left\|\hat{\theta}_{n+1}-\hat{\theta}_n\right\|\leq\varepsilon=0.0001$.

After we obtain the maximum likelihood estimate for $\underline{\theta}$, we can compare it with those obtained in the previous paragraph.

The "numerical opt." solution in the state-dependent case denotes the application of MATLAB'S optimization function "fminsearch".

case	State-independent		State-dependent	
	Numerical opt.	Newton-Raphson	Numerical opt.	Newton-Raphson
all data	0.155	0.154	[0.410, 0.310, 0.096, 0.033, 0.114]	[0.410, 0.310, 0.096, 0.033, 0.114]

Table 5-3 Comparison of log-likelihood estimators

Having a good approximation for the maximum likelihood estimator, we can now calculate the standard deviation and the 5th and 95th percentile, (i.e., lower and upper bound of the 95% confidence interval) for $\hat{\underline{\theta}}$. The results are presented in the next table.

case	State-independent				State-dependent				
	$\hat{\underline{\theta}}$	std	5%	95%	state	$\hat{\underline{\theta}}$	std	5%	95%
all data	0.154	0.00171	0.151	0.158	0	0.410	0.00823	0.394	0.427
					1	0.310	0.00532	0.230	0.321
					2	0.096	0.00198	0.092	0.100
					3	0.033	0.00187	0.029	0.036
					4	0.114	0.01435	0.086	0.142

Table 5-4 Maximum-likelihood estimators with confidence intervals for state-independent and state-dependent case

From table 5-3 we see that the smallest standard deviation occurs for the maximum likelihood estimators obtained from all data.

In the next section we will perform a simple statistical test to obtain which model we should use for our bridge deterioration database. We want to decide between a two models: one for state-independent transition probabilities and one for state-dependent transition probabilities. To choose the right model is very important, because this will give us additional information about the appropriate type of the Markov model.

5.4 Testing state independence against state dependence

In this section, we consider testing a simple null hypothesis against a simple alternative hypothesis. Now we will investigate which model fits better to the bridge deterioration data in the DISK database, one with a single state-independent transition probability or one with state-dependent transition probabilities. These two models were presented in the previous section.

The test, which we use here, is simple. Assume that we have a sample that came from one of two different distributions. Our aim is to determine from which one. More precisely, assume that a random sample $X_1, X_2, ..., X_n$ came from the density $f_0(x)$ or $f_1(x)$ and we want to test the null hypothesis H_0 :sample distributed as $f_0(\cdot)$, abbreviated $X_i \sim f_0(\cdot)$, versus the alternative hypothesis $\mathrm{H}_1 : X_i \sim f_1(\cdot)$. Bellows definitions and theorems comes from Mood, Graybill and Boes (1974, pp. 410-442)

Definition 5.1 : Simple likelihood-ratio test

Let $X_1, X_2, ..., X_n$ be a random sample from either $f_0(\cdot)$ or $f_1(\cdot)$. A test Υ of $\mathrm{H}_0 : X_i \sim f_0(\cdot)$ versus $\mathrm{H}_1 : X_i \sim f_1(\cdot)$ is defined to be a *simple likelihood-ratio test* if Υ is defined by

Reject H_0 if $\lambda < k$,

Accept H_0 if $\lambda > k$,

Either accept H_0, reject H_0, or randomize if $\lambda = k$,

where

$$\lambda = \lambda(x_1, x_2, ..., x_n) = \frac{\prod_{i=1}^{n} f_0(x_i)}{\prod_{i=1}^{n} f_1(x_i)} = \frac{L_0(x_1, x_2, ..., x_n)}{L_1(x_1, x_2, ..., x_n)} = \frac{L_0}{L_1} \tag{5.17}$$

and *k* is a nonnegative constant. [$L_j = L_j(x_1, x_2, ..., x_n)$ is the likelihood function for sampling from the density $f_j(\cdot)$.]

□

For each different k we have a different test. For a fixed k the test says to reject H_0 if the ratio of likelihoods is small; that is, reject H_0 if it is more likely (L_1 is large compared to L_0) that the sample comes from $f_1(\cdot)$ rather than from $f_0(\cdot)$. This test has intuitive appeal.

Now we would like to consider a more general problem. Assume that we have a random sample $X_1, X_2, ..., X_n$ from a density $f(x;\theta)$, $\theta \in \underline{\bar{\Theta}}$, and we want to test null hypothesis H_0 : $\theta \in \underline{\bar{\Theta}}_0$ versus alternative hypothesis H_1 : $\theta \in \underline{\bar{\Theta}}_1 = \underline{\bar{\Theta}} - \underline{\bar{\Theta}}_0$, where $\underline{\bar{\Theta}}_0 \subset \underline{\bar{\Theta}}$, $\underline{\bar{\Theta}}_1 \subset \underline{\bar{\Theta}}$ and $\underline{\bar{\Theta}}_0$ and $\underline{\bar{\Theta}}_1$ are disjoint.

Definition 5.2 : Generalized likelihood-ratio

Let $L(\theta; x_1, x_2, ..., x_n)$ be the likelihood function for a sample $X_1, X_2, ..., X_n$ having joint density $f_{x_1,...,x_n}(x_1, ..., x_n; \theta)$ where $\theta \in \underline{\bar{\Theta}}$. The *generalized likelihood-ratio* denoted by λ or λ_n is defined to be

$$\lambda = \lambda(x_1, x_2, ..., x_n) = \lambda_n = \frac{\sup_{\theta \in \underline{\bar{\Theta}}_0} L(\theta; x_1, x_2, ..., x_n)}{\sup_{\theta \in \underline{\bar{\Theta}}} L(\theta; x_1, x_2, ..., x_n)} \tag{5.18}$$

□

We want to mention here, that λ is a function of $x_1, x_2, ..., x_n$, namely $\lambda(x_1, x_2, ..., x_n)$. If we replace observations by their corresponding random variables $X_1, X_2, ..., X_n$, then we write Λ for λ; that is $\Lambda = \lambda(X_1, ..., X_n)$. Λ as a function of the random variables is again a random variable. Because Λ does not depend on unknown parameters we can say that it is a test statistic.

Now we want to make some remarks:

i. We used the same symbol λ to denote the simple likelihood-ratio, but the generalized likelihood-ratio does not reduce to the simple likelihood-ratio for $\underline{\overline{\Theta}} = \{\theta_0, \theta_1\}$
ii. λ given in equation (5.18) necessarily satisfies $0 \le \lambda \le 1$;
iii. The parameter θ can be a vector;
iv. The denominator of Λ is the likelihood function evaluated at the maximum-likelihood estimator;

The values λ of the statistic Λ are used to formulate a test of H_0: $\theta \in \underline{\overline{\Theta}}_0$ versus H_1: $\theta \in \underline{\overline{\Theta}}_1 = \underline{\overline{\Theta}} - \underline{\overline{\Theta}}_0$ by employing the *generalized likelihood-ratio test principle,* which states that H_0 is to be rejected if and only if $\lambda \le \lambda_0$, where λ_0 is some fixed constant satisfying $0 \le \lambda_0 \le 1$. (The constant λ_0 is often specified by fixing the size of the test). The generalized likelihood-ratio test makes sense since λ tends to be small when H_0 is not true, because then the denominator of λ tends to be larger than the numerator.

The distribution of the generalized likelihood-ratio test statistic is intractable, and an approximate test can be obtained using an asymptotic distribution of the generalized likelihood-ratio. The following theorem, which will not be proved because of the advanced character of its proof, gives the asymptotic distribution of the generalized likelihood-ratio.

Theorem 5.2 Let $X_1, X_2, ..., X_n$ be a sample with joint density $f_{X_1,...,X_n}(\cdot,...,\cdot;\theta)$, that satisfies quite general regularity conditions, where $\theta = (\theta_1,...,\theta_k)$. Suppose that the parameter space $\underline{\overline{\Theta}}$ is *k*-dimensional. In testing the hypothesis

$$\mathrm{H}_0 : \theta_1 = \theta_1^{\,0}, ..., \theta_r = \theta_r^{\,0}, \theta_{r+1}, ..., \theta_k \ ,$$

Where $\theta_1^0, ..., \theta_r^0$ are known and $\theta_{r+1}, ..., \theta_k$ are left unspecified, $-2\log\Lambda_n$ is approximately distributed as a chi-square distribution with *r* degrees of freedom when H_0 is true and the sample size *n* is large.

□

We have assumed that $1 \le r \le k$ in the above theorem. If $r = k$, then all parameters are specified and none is left unspecified. The parameter space $\underline{\overline{\Theta}}$ is *k*-dimensional, and because H_0 specifies the value of *r* of the components of $(\theta_1,...,\theta_k)$, the dimension of $\underline{\overline{\Theta}}_0$ is $k - r$. Hence, the degrees of freedom of the asymptotic chi-square distribution in the above theorem can be considered in two ways: first, as the number of parameters specified by H_0 and, second, as the difference in the dimensions of $\underline{\overline{\Theta}}$ and $\underline{\overline{\Theta}}_0$.

Recall that Λ_n is the random variable, which has values

$$\lambda_n = \sup_{\underline{\bar{\Theta}}_0} L(\theta_1,...,\theta_k; x_1,...,x_n) \Big/ \sup_{\underline{\bar{\Theta}}} L(\theta_1,...,\theta_k; x_1,...,x_n) ,$$

which in turn is the generalized likelihood-ratio for a sample of size *n*. $\underline{\bar{\Theta}}_0$ is a subset of $\underline{\bar{\Theta}}$ which is specified by H_0. The generalized likelihood-ratio principle says that H_0 has to be rejected for λ_n small, but on the other hand $-2\log\lambda_n$ increases as λ_n decreases, hence a test that is equivalent to a generalized likelihood-ratio test is one that reject for $-2\log\lambda_n$ large. The above theorem gives us an approximate distribution for the values $-2\log\lambda_n$ when H_0 is true, and hence a test with approximate size α is given by the following:

Reject H_0 if and only if $-2\log\lambda_n > \chi^2_{1-\alpha}(r)$

where $\chi^2_{1-\alpha}(r)$ is the $(1-\alpha)$th quantile of the chi-square distribution with r degrees of freedom. The degrees of freedom r is the dimension of the parameter space representing those components that are specified by the null hypothesis.

The purpose of this section is to determine which model fits better to the DISK database. To do this, we want to use the above theorem. In our model $\theta_0 = (p, p, p, p, p)$, whereas $\theta_1 = (p_0, p_1, p_2, p_3, p_4)$. Based on this assumption, we write the null hypothesis H_0 in the form: H_0: $p_0 = p_1 = p_2 = p_3 = p_4$, where p_i are the parameters of the transition probability matrix. If we use the following reparameterization, H_0 will have the desired form of Theorem 5.2

$$\theta_0 = \frac{p_0}{p_4}, \theta_1 = \frac{p_1}{p_4}, \theta_2 = \frac{p_2}{p_4}, \theta_3 = \frac{p_3}{p_4}, \theta_4 = p_4$$

Now H_0 becomes $H_0: \theta_0 = 1, \theta_1 = 1, \theta_2 = 1, \theta_3 = 1,\ \theta_4$; that is, the first 4 components are specified to be 1 and the remaining one is unspecified. Hence Theorem 5.2 may be applicable, and because of the invariance property of the maximum likelihood estimates, the likelihood-ratio obtained before and after reparameterization are the same; for this reason the asymptotic distribution of $-2\log\lambda_n$ is a chi-squared distribution with 4 degrees of freedom when H_0 is true.

In Chapter 3 we assumed that the significance level α is equal to 0.05, which corresponds to 5%. Using this, we calculate the log-likelihood-ratio, which in this case is equal to $\log\lambda_n$ = -2180.1843 and $-2\log\lambda_n$ = 4360.3686. This number is quite big, which indicates that we should reject hypothesis H_0. Because $\Pr(-2\log\lambda_n > \chi^2_{1-\alpha}(r)) = 1$, we are quite confident about this decision.

In Table 5-4 we show the result of the likelihood ratio test for all data. Number of degrees of freedom is 4.

case	$\log\lambda_n$	$-2\log\lambda_n$	$\Pr(X > -2\log\lambda_n)$	**decision**
All data	-2180.1843	4360.3686	0	reject

Table 5-5 Results from likelihood-ratio test

Figure 5-5 presents a chi-squared distribution with 4 degrees of freedom.

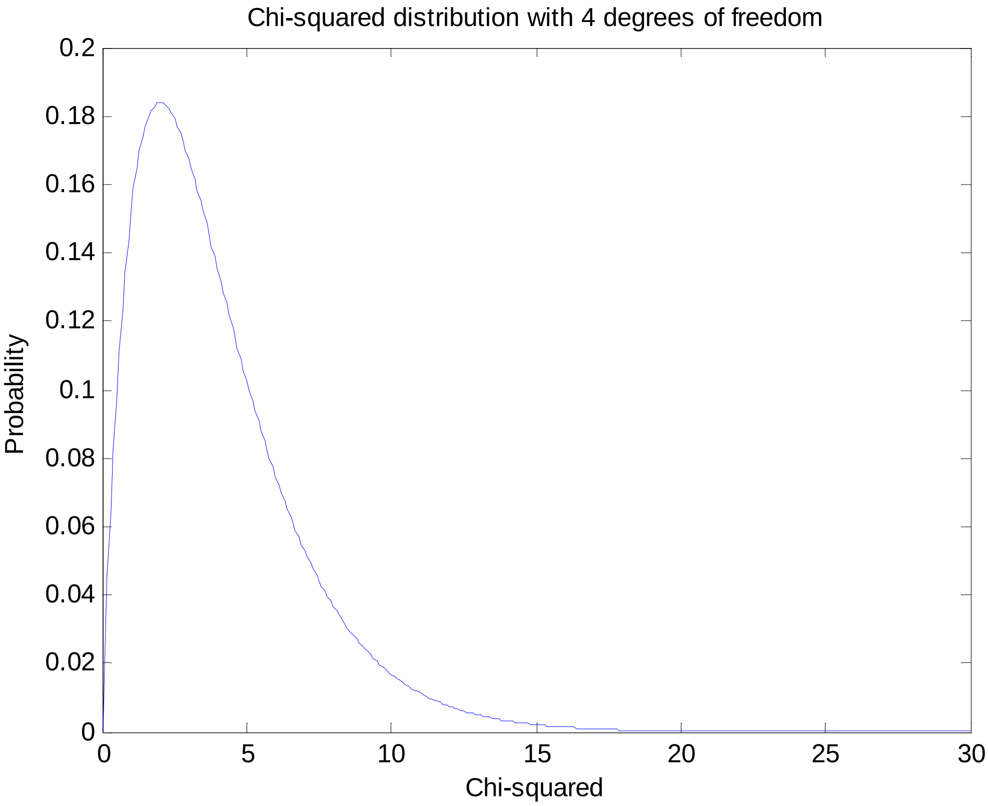

Figure 5-5 Chi-squared distribution with 4 degrees of freedom

We can see that for $\chi^2_{1-\alpha}(r)$ bigger than 18, the value of this distribution is almost equal to zero.

In conclusion, we may clearly say that we should use the model with state-dependent probabilities, i.e. that we should reject the hypothesis $\mathrm{H}_0: \; p_0 = p_1 = p_2 = p_3 = p_4$,which means that all probabilities are the same.

6 Data Analysis

In the former chapter we have developed a model, which describes our data well. Now we would like to consider the influence of some covariates on this model. In this approach, we will separate the bridges into sensible groups and determine the effect of this grouping on the parameters of the Markov model. To recognize the covariate with the largest effect is very important for making correct maintenance decisions.

Usually, data contain covariates of which some of them are important and the others are not. In our data, we consider the covariates, which in our opinion could have influence on the deterioration process and therefore could affect the model parameters. Hence, we take into account the following covariates:

1. Year of construction of the structure (we consider two groups of age): all bridges constructed before 1976 and all bridges constructed after 1976;

2. Location of structure: bridges "in the road" subject to heavy traffic against bridges "over the road" subject to light traffic;

3. Type of bridge: separated into "concrete viaducts" and "concrete bridges" ;

4. Use, which means the type of traffic, which uses the bridge: traffic only with trucks and cars, and a mixture of traffic (also bikes, pedestrians, etc.);

5. Province in which bridge is located:

   5.a Groningen
   5.b Friesland
   5.c Drenthe
   5.d Overijssel
   5.e Gelderland
   5.f Utrecht
   5.g Noord-Holland
   5.h Zuid-Holland
   5.i Zeeland
   5.j Noord-Brabant
   5.k Limburg
   5.l Flevoland
   5.m West-Duitsland

West-Duitsland is not a province of the Netherlands, but a very small number of bridges have been assigned this location due to administrative reasons. The bridges contained in this area are located close to or on the border with Germany. There are only 4 bridges which are assigned to this region. Most likely, the maintenance responsibility for these bridges has changed during the years due to a change in communal borders.

The last covariate, province, has a lot of sub-covariates and this can result in not enough data being available to estimate the model parameters for each of them individually. Taking this into account, we decided to modify this covariate by grouping it according to population density or proximity to the sea. Hence the last covariate can be considered as two separate covariates:

6. Population density
   - 6.a Higher population density:
     - 6.a.1. Utrecht
     - 6.a.2. Noord-Holland
     - 6.a.3. Zuid-Holland
     - 6.a.4. Noord-Brabant
     - 6.a.5. Limburg
   - 6.b Lower population density:
     - 6.b.1. Groningen
     - 6.b.2. Friesland
     - 6.b.3. Drenthe
     - 6.b.4. Overijssel
     - 6.b.5. Gelderland
     - 6.b.6. Zeeland
     - 6.b.7. Flevoland
     - 6.b.8. West-Duitsland

7. Proximity to the sea
   - 7.a Close to the sea
     - 7.a.1. Groningen
     - 7.a.2. Friesland
     - 7.a.3. Noord-Holland
     - 7.a.4. Zuid-Holland
     - 7.a.5. Zeeland
     - 7.a.6. Flevoland
   - 7.b Inland
     - 7.b.1. Drenthe
     - 7.b.2. Overijssel
     - 7.b.3. Gelderland
     - 7.b.4. Utrecht
     - 7.b.5. Noord-Brabant
     - 7.b.6. Limburg
     - 7.b.7. West-Duitsland

This grouping of provinces has been done ad hoc.

As we see, covariates 6 and 7, share some sub-covariates. Hence, we should decide which of them we want to use in the analysis. Of course, to start with we will do the calculations for both of them. In a later analysis, we will decide which of them has bigger influence on the model parameters and therefore also on the deterioration process.

6.1 Influence of the covariates on the model parameters

Previously, we specified the covariates which we will take into account in the analysis. Hence, we can now present the estimated parameters of the model for each of them. In the previous chapter, we decided to use the model with state-dependent probabilities, so now we will present results only for this model. First, we mention the parameters for all data, and we add the new estimated parameters. All of them are summarized Table 6-1.

Type of data		Estimated parameters
All data		**[0.4104, 0.3102, 0.0963, 0.0328, 0.1143]**
Construction year	Built before 1976	[0.8026, 0.4135, 0.1027, 0.0335, 0.0816]
	Built after 1976	[0.3988, 0.2542, 0.0842, 0.0305, 0.2073]
Location	"in the road"- heavy traffic	[0.4490, 0.3115, 0.0966, 0.0328, 0.1140]
	" over the road"- light traffic	[0.3180, 0.3017, 0.0949, 0.0329, 0.1150]
Type of bridge	"concrete viaducts"	[0.3933, 0.2976, 0.0946, 0.0327, 0.1332]
	"concrete bridges"	[0.5883, 0.3867, 0.1025, 0.0332, 0.0493]
Type of traffic	Only with trucks and cars	[0.4193, 0.3029, 0.0951, 0.0326, 0.1152]
	Mixture of traffic	[0.3491, 0.3675, 0.1020, 0.0340, 0.1096]

Table 6-1 Influence of different covariates on the estimated parameters

Taking into account each of the covariates separately, we observe quite a difference in estimated parameters. For example: for covariate 'Construction year', we see that the probability of going from state 0 to state 1 is 0.8026 for bridges constructed before 1976 while for younger bridges it is only 0.3988. The same situation occurs with the probability of going from state 1 into state 2. The situation changes for the other probabilities. Obviously, for younger bridges, the transition probability to the next state is less than the one for older bridges. To summarize, we can say that bridges younger than 28 years deteriorate slower at the beginning, but with age this deterioration accelerates. On the other hand, bridges older than 28 years, which still are in the states 0 or 1 deteriorate very fast at the beginning, but later this process slows down.

Comparing the estimated parameters for the covariate 'Construction year' with those estimated for all data, we see that between parameters for older bridges and for all data the difference is quite substantial, particularly for the probability in initial states. The difference between parameters for all data and for bridges younger than 28 years is not very large, but there still is a difference. A similar analysis can be made for the other covariates, but we leave this for the reader.

Here we would like to mention that from the above table it follows that the covariate "Concrete viaduct" has a small influence on the transition probabilities. The same can be said about the covariate "Over the road- light traffic" and "Only truck and cars". The difference between estimated probabilities is very small. We observe the same effect for the pattern of the expected condition, which is calculated in the next section.

Table 6-2 shows the difference between estimated transition probabilities for the last covariate "Province", for each of the sub-covariates separately. We observe that for some provinces we get that the same transition probabilities are equal to zero or one. This indicates a lack of data for the specific sub-covariate.

Type of data		Estimated parameters
All data		**[0.4104, 0.3102, 0.0963, 0.0328, 0.1143]**
Province in which bridge is located	Groningen	[0.2912, 0.9476, 0.2608, 0.0050, 0.0000]
	Friesland	[0.2285, 0.3324, 0.0952, 0.0546, 0.0000]
	Drenthe	[0.4425, 0.3712, 0.1739, 0.0805, 0.0000]
	Overijssel	[0.2843, 0.2350, 0.2201, 0.0282, 0.0637]
	Gelderland	[0.2321, 0.3743, 0.1964, 0.0599, 0.0957]
	Utrecht	[0.8176, 0.4492, 0.0624, 0.0000, 0.0001]
	Noord-Holland	[0.5625, 0.2183, 0.0932, 0.0075, 0.1509]
	Zuid-Holland	[0.5324, 0.2211, 0.0629, 0.0351, 0.1231]
	Zeeland	[0.5575, 0.6394, 0.0512, 0.0000, 0.8807]
	Noord-Brabant	[0.5622, 0.3578, 0.0926, 0.0356, 0.3216]
	Limburg	[0.5059, 0.4049, 0.1290, 0.0263, 0.0000]
	Flevoland	[1.0000, 0.3461, 0.2899, 0.0688, 0.0000]
	West-Duitsland	[0.9785, 0.8198, 0.0000, 0.0183, 1.0000]

Table 6-2 Estimated parameters for covariate "province"

Type of data		Estimated parameters
All data		**[0.4104, 0.3102, 0.0963, 0.0328, 0.1143]**
Population density	Higher	[0.5398, 0.2790, 0.0789, 0.0255, 0.1422]
	Lower	[0.2677, 0.3856, 0.1817, 0.0477, 0.0847]
Proximity to the sea	Close to the sea	[0.4739, 0.2501, 0.0787, 0.0233, 0.1176]
	Inland	[0.3465, 0.3725, 0.1223, 0.0416, 0.1111]

Table 6-3 Estimated parameters for Population density and Proximity to the sea

Table 6-3 shows the differences in estimated transition probabilities for the grouped covariate "Province". We show the results for both groups "Population density" and "Proximity to the sea" to show the differences between them, compared to the probabilities estimated for all data. We see that the difference between higher and lower population density is quite large, but compared to the results for all data, it is not so visual (there is substantial difference in the first and the last probability for higher population density, and first, third and the last probability for lower population density). Comparing the estimated parameters for all data to the parameters for the covariate "Proximity to the sea", we also see strong differences. But we can not compare the covariates "Population density" and "Proximity to the sea" with each other, because they are partly selected on the basis of the same provinces.

6.2 Expected condition at given time

Having estimated the parameters for each of these covariates, we would like to know how quickly they converge to state 5, i.e. how rapidly they fail. To calculate this, let's introduce a short definition of failure. A component is said to fail, when its deterioration process exceeds a certain level, in our case state 5. The next purpose of this analysis is to calculate the expected condition for a given age, and to predict the time when bridges with given covariates achieve their failure state 5.

The expected condition at a given age is simply given by:

$$E\{X(t)\} = \sum_{j=0}^{n} j \Pr\{X(t) = j\}, \quad (6.1)$$

where $X(t)$ denotes the state of a Markov chain at given time t, and $j \in \{0,1,2,3,4,5\}$ is the number of each state. The probability inside the summation is the probability that a chain is in state j at time t. It is calculated under the assumption that each of the bridges starts from state i at time zero:

$$\Pr\{X(t) = j\} = \sum_{i=0}^{5} \Pr\{X(t) = j \mid X(0) = i\} \Pr\{X(0) = i\} \quad (6.2)$$

where $\Pr\{X(t) = j \mid X(0) = i\}$ denotes probabilities coming from the one-step probability matrix and t is an age of this bridge.

Let the time at which the failure level is crossed, be denoted by the lifetime T (in years). Then, due to the Markov model, the cumulative lifetime distribution can be written as:

$$F(t) = \Pr\{T \le t\} = \Pr\{X(t) \ge y\}, \quad (6.3)$$

where y denotes the failure level, i.e. state 5. Because we have only 6 states and the last one is 5, we will calculate $\Pr\{X(t) = y\}$.

The probability of failure per year, denoted by $q_t, t = 1,2,3,\ldots$, follows immediately from (6.3) :

$$q_t = F(t) - F(t-1) \quad . \quad (6.4)$$

We present the results for each covariate separately below and, to compare, for all data together, i.e. not divided into covariates.

We may observe that these figures, especially Figures 6-1 and 6-2, which show the expected condition and lifetime distribution for both models from chapter 5 respectively, confirm our results from that chapter. There, we decided to use the model with state-dependent transition probabilities, because this model fits better to our data. The same result we see in the figure below. The solid line indicates the expected condition for the state-dependent model and the dashed line is for the model with state-independent probabilities. We see that for state-dependent model, deterioration is slower than for the state-independent one, especially for the transition to the higher states. The state-dependent model seems to be more reasonable.

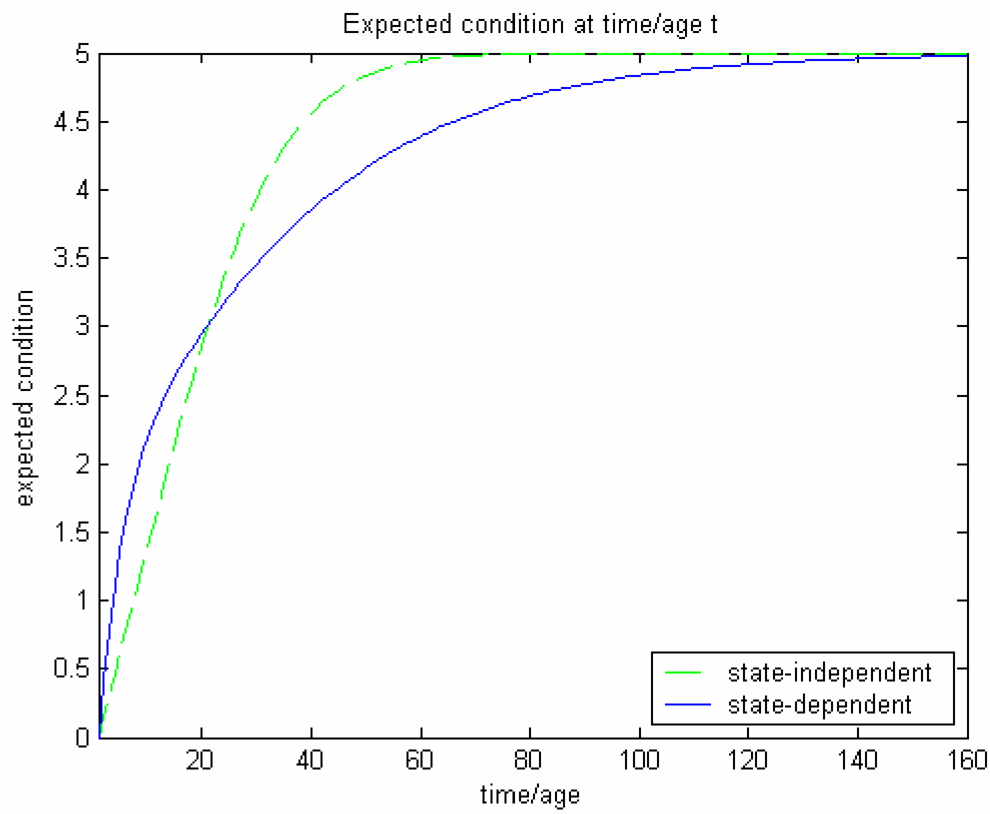

Figure 6-1 Expected condition at time t for all data

The annual probability of failure shows us how fast bridges reach the failure state 5, given that the bridge started in state 0, which is the initial state. From Figure 6-2 we observe that the curve for the state-independent model is less wide than for the state-dependent model, is more concentrated around 27 years and reaches a maximum probability of 0.033, whereas the curve for the state-dependent model reaches its maximum with probability 0.017 what is almost half as large. This lifetime distribution for the state-dependent model reaches the higher probability of failure about 35 years, which is more than years for the state-independent model. The model with state-dependent transition probabilities reaches the failure state at an age 8 years older than state-independent model with lower probability, and a wider spread. We see that, while the failure probability of the state-independent model is almost zero at 100 years, the state-dependent model has still a probability equal to 0.004 to reach a failure state 5.

In below Table we present expected lifetime and 90% confidence intervals for both of the above lifetime distributions.

	5% lower	Expected Lifetime	5% upper
State-independent	15	33,26	57
State-dependent	19	53,83	119

Table 6-4 Expected Lifetime and 90% confidence intervals for both models

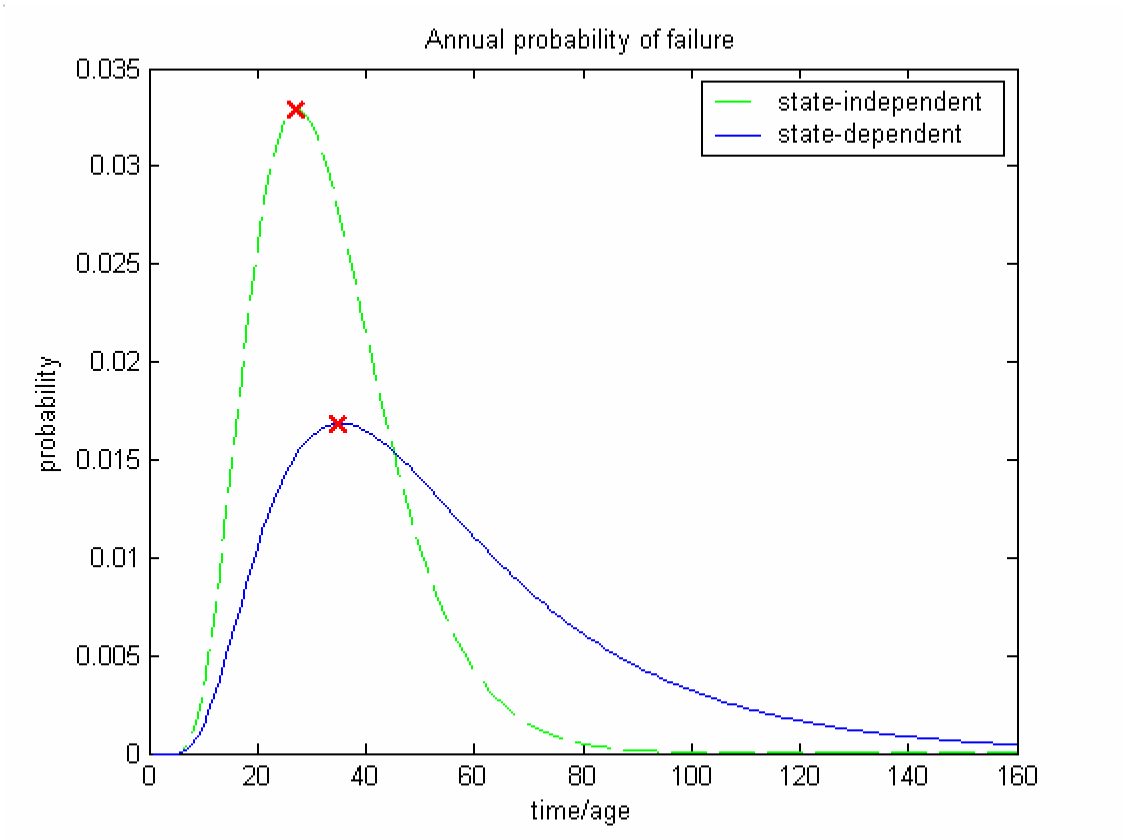

Figure 6-2 Annual probability of failure for all data

For the rest of this analysis we decided to use the model with state-dependent transition probabilities, as this model seems to be more reasonable and, as we decided in chapter 5, fits better to our data.

Figures 6-3 and 6-4 show the differences between the expected condition at a given time as well as the lifetime distribution for covariate "Construction year". This covariate is divided into two groups, built before 1976 and built after 1976 and we will compare these two groups.

As we see below, the age has influence on probabilities of transition from one state to the other. We may conclude that the covariate "Built before 1976" influences the model parameters, especially for initials transitions, slightly more than the "Built after 1976". Figure 6-3 shows the difference in expected conditions, which visualizes the influence of this covariate on the model parameters. We may say that younger bridges deteriorate slower (this is a reasonable result), whereas older ones deteriorate slightly faster. The annual probability of failure, presented in Figure 6-4, seems to be very similar. Hence, we may say that in both groups, it takes the same time for the bridges to reach state 5.

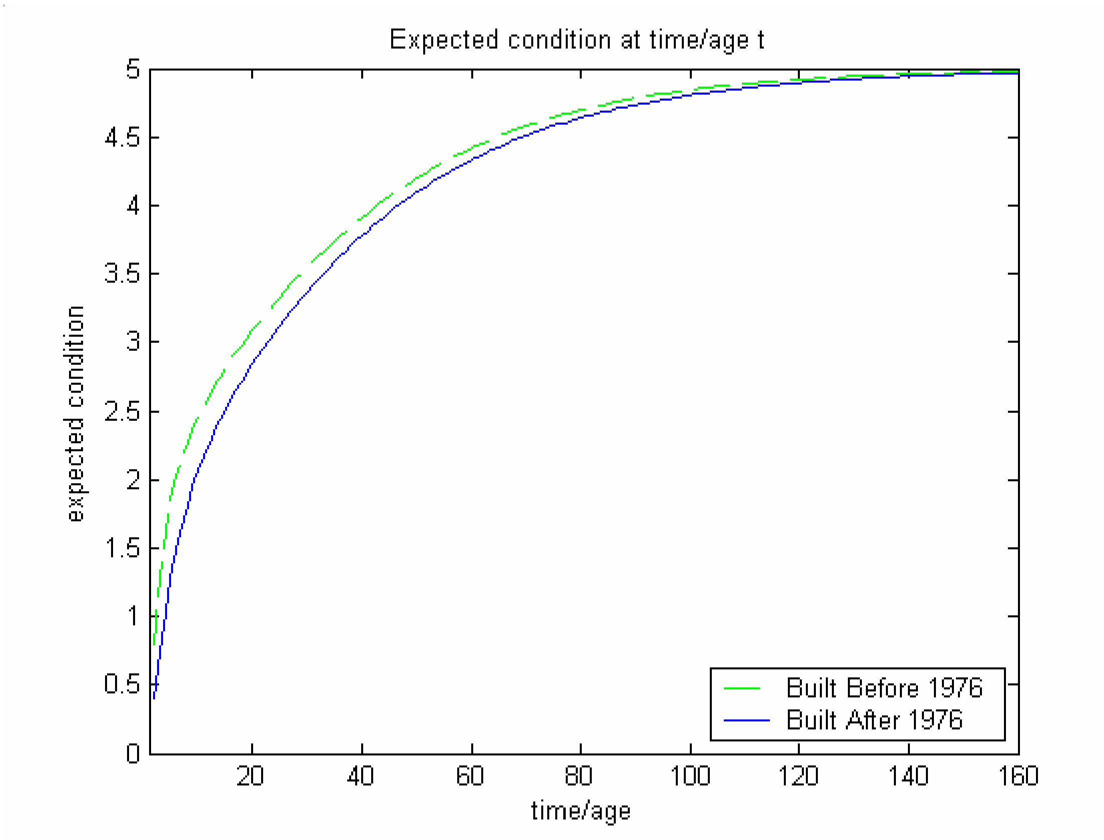

Figure 6-3 Expected condition at time t for covariate "Construction year"

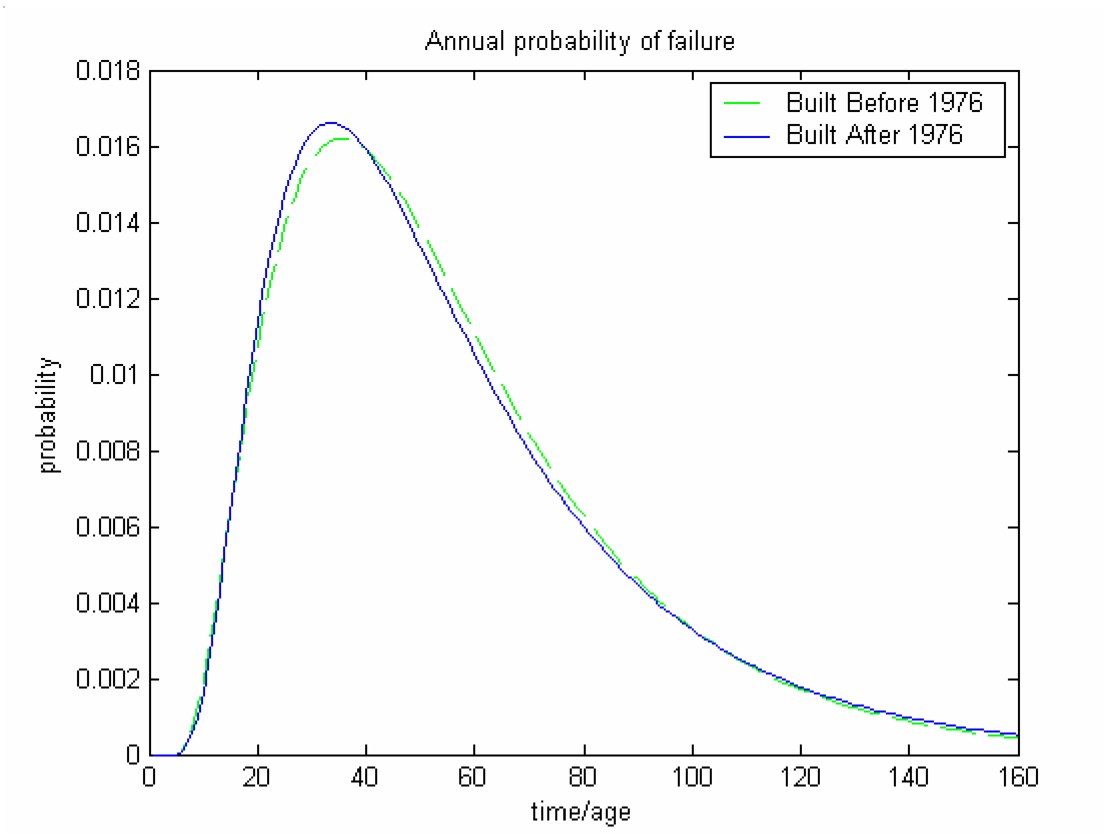

Figure 6-4 Annual probability of failure for covariate" Construction year"

Figures 6-5 and 6-6 show the difference between the expected condition at a given time for the covariate "Location of structure" and the corresponding annual probabilities of failure. The same results are presented for the covariate "Type of traffic" in Figures 6-7 and 6-8. These covariates were divided into two groups (each of them separately): "In the road-heavy traffic" and "Over the road-light traffic", and "Only trucks and cars" and "Mixture of traffic", respectively.

In these figures, we may observe that these covariates have a similar behaviour for each group. There is a slight difference for covariate "Location", but this difference is so small that it doesn't influence the model parameters. The same can be said about the second covariate "Type of traffic", presented in Figures 6-7 and 6-8. In these figures, especially in Figure 6-7, we see that the curves of the expected condition at given time for both groups of the covariate "Type of traffic", coincide. This indicates that these groups do not differ and we may claim that they have no influence on the model parameters. These results are similar to those presented at the beginning in Table 6-1.

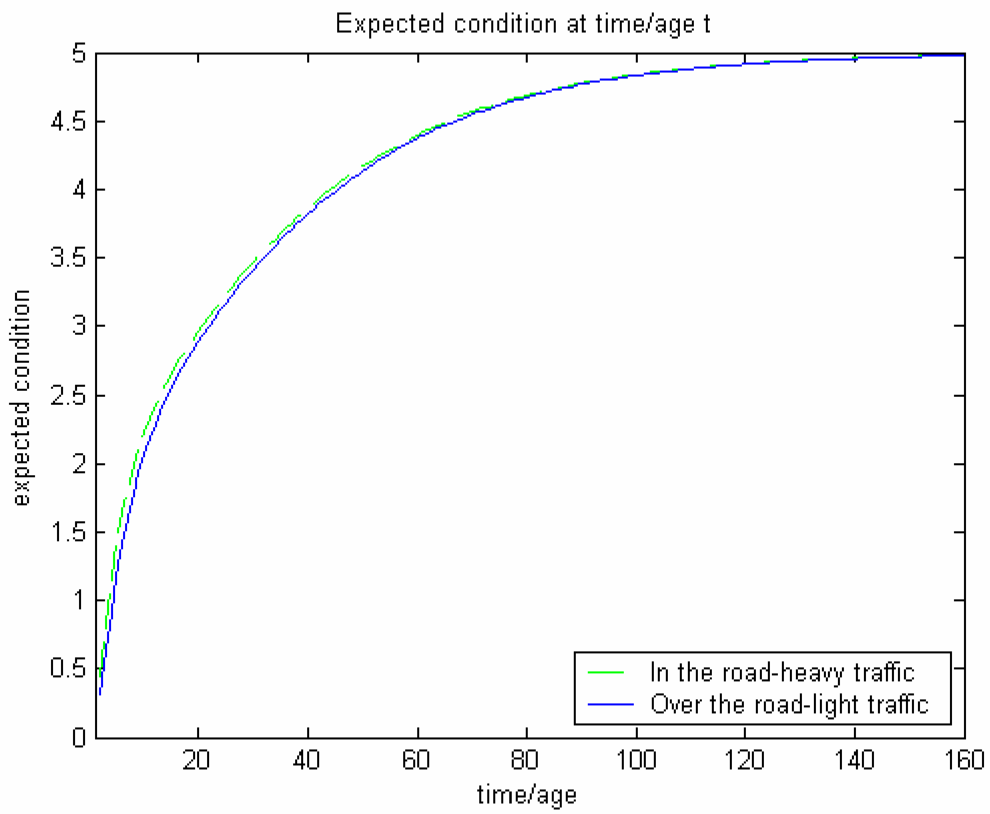

Figure 6-5 Expected condition at time t for covariate "Location of structure"

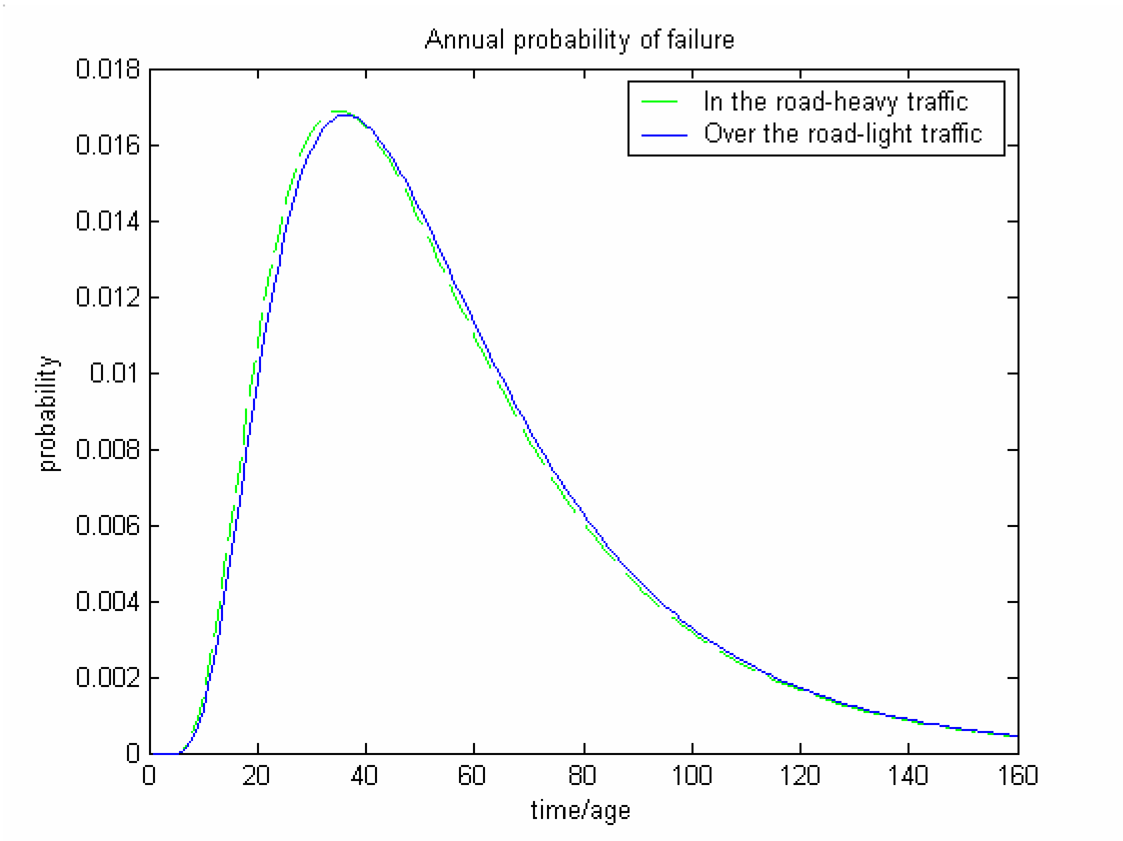

Figure 6-6 Annual probability of failure for covariate „Location of structure"

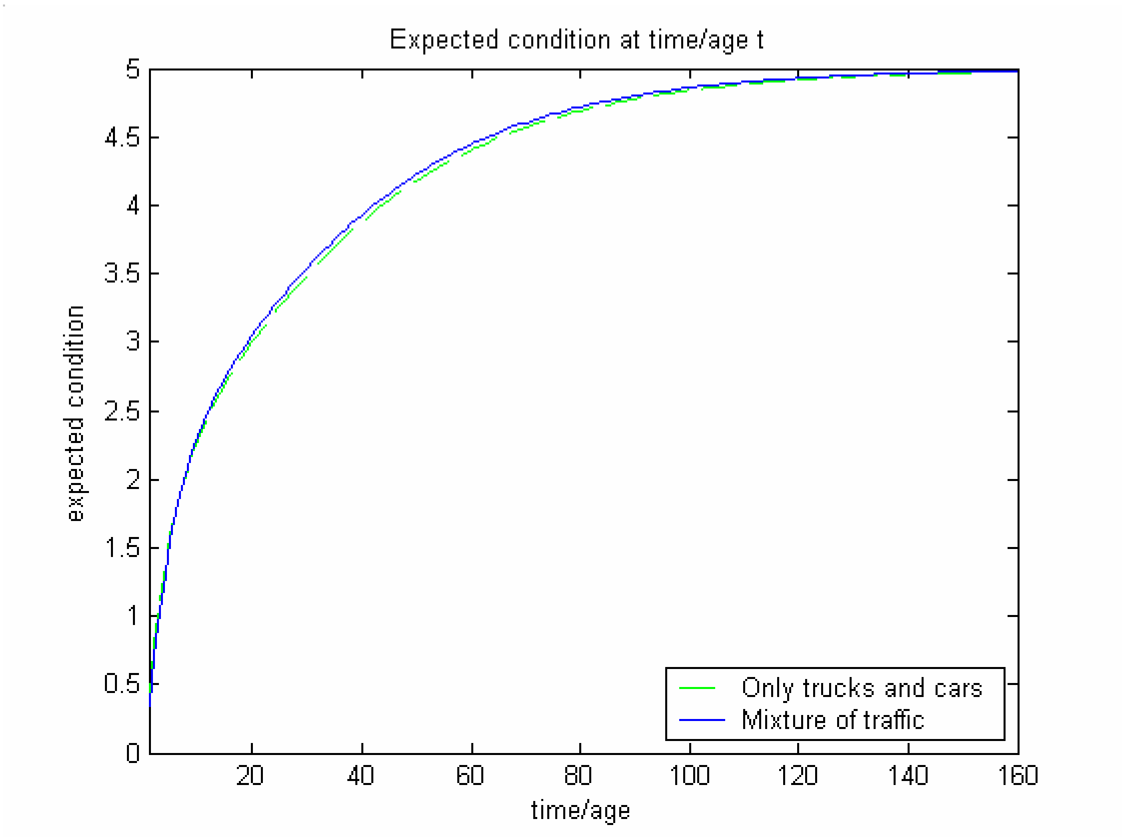

Figure 6-7 Expected condition at time t for covariate "Type of traffic"

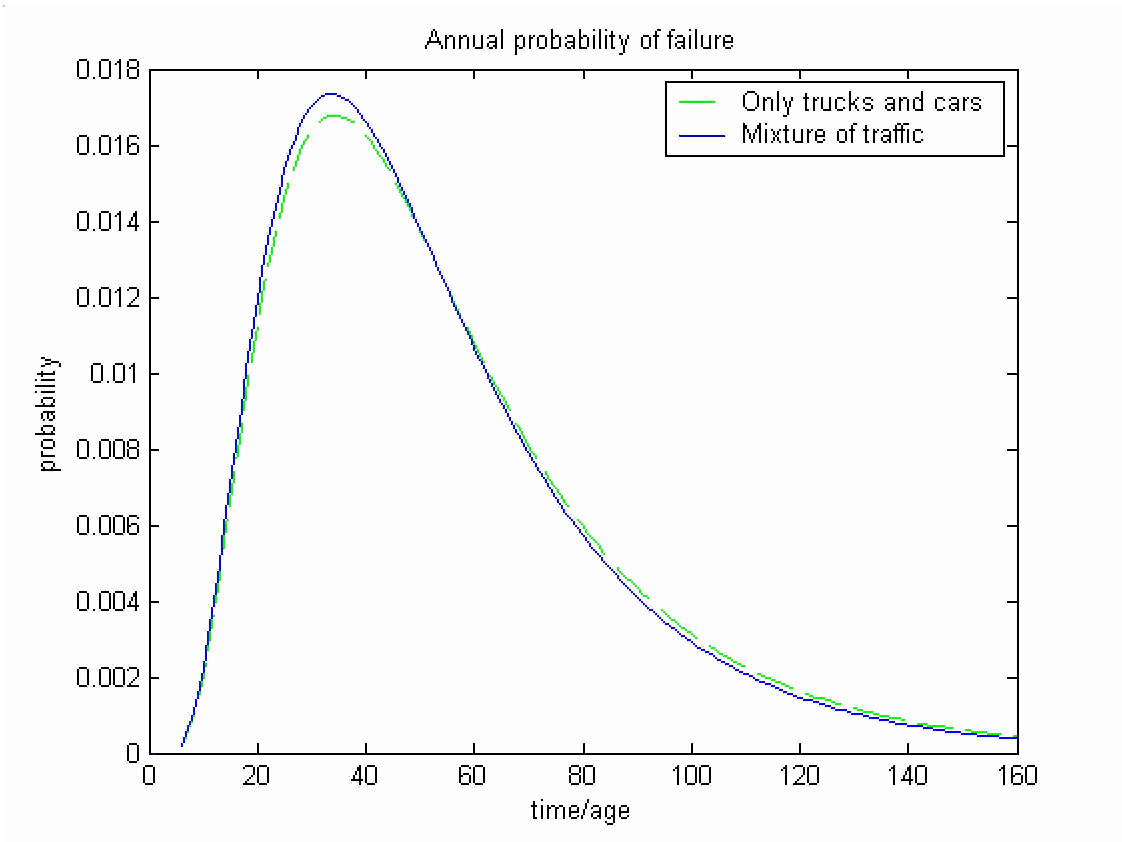

Figure 6-8 Annual probability of failure for covariate "Type of traffic"

Now, we would like to present the results of computations of the expected condition at a given time and the annual probability of failure for the covariates "Type of bridge" (Figures 6-9 and Figure 6-10) and "Population density" (Figures 6-11 and Figure 6-12).
The covariate "Population density" represents the grouping of covariate "Province" in the way shown at the beginning of this chapter. We decide to use this covariate instead of using another type of grouping, because for this covariate we have more data points and the estimated parameters will be more accurate.

As we see, these covariates differ between their sub-covariates, which may indicate influence on the model parameters. We observe that in Figure 6-11 there is quite an essential difference between the covariates "higher population density" and "lower population density". The first of them deteriorates slower than the second one, which can be clearly observed in the transition probabilities in the higher states.
The covariate "Type of bridge – concrete bridge" has a lower probability of failure for younger bridges, especially up to 55 years, that the covariate "Type of bridge – concrete viaducts"; this is presented in Figure 6-10. The same covariate has quite high transition probabilities for bridges which still are in the states 0 or 1, this has influence on the expected condition at given time, especially during the firsts years.
The figures for the remaining covariates as well as figures with additional curve for all data to better interpretation and comparison are presented in Appendix B and Appendix C.

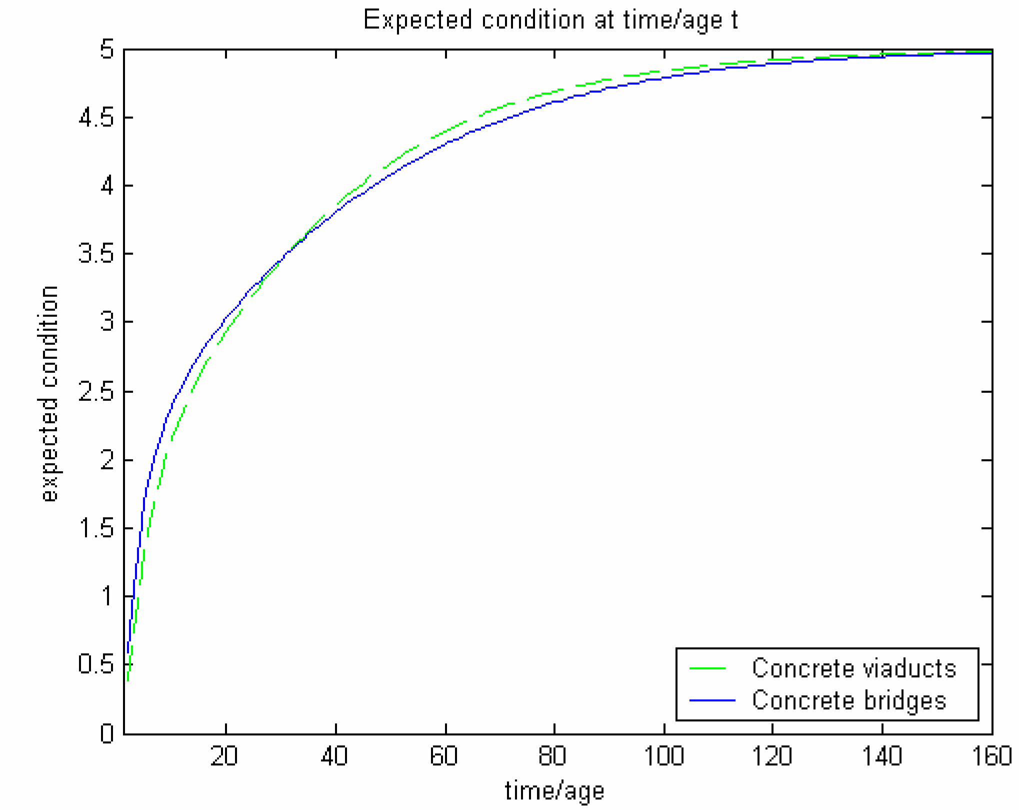

Figure 6-9 Expected condition at time t *for covariate "Type of bridge"*

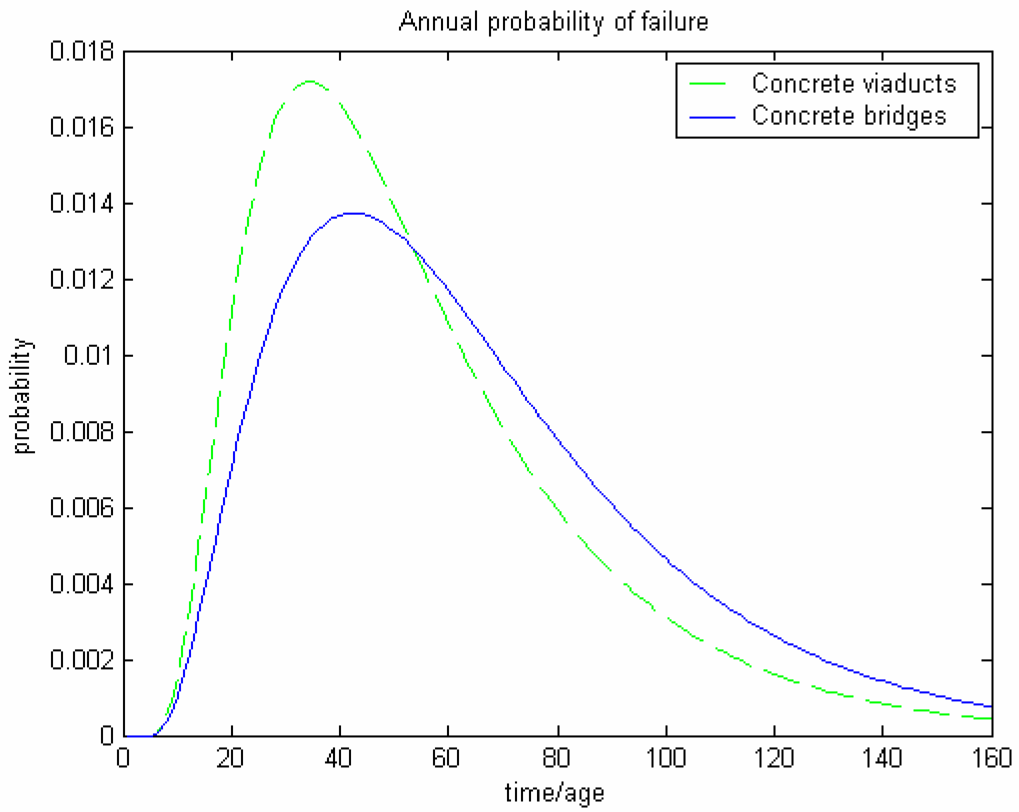

Figure 6-10 Annual probability of failure for covariate "Type of bridge"

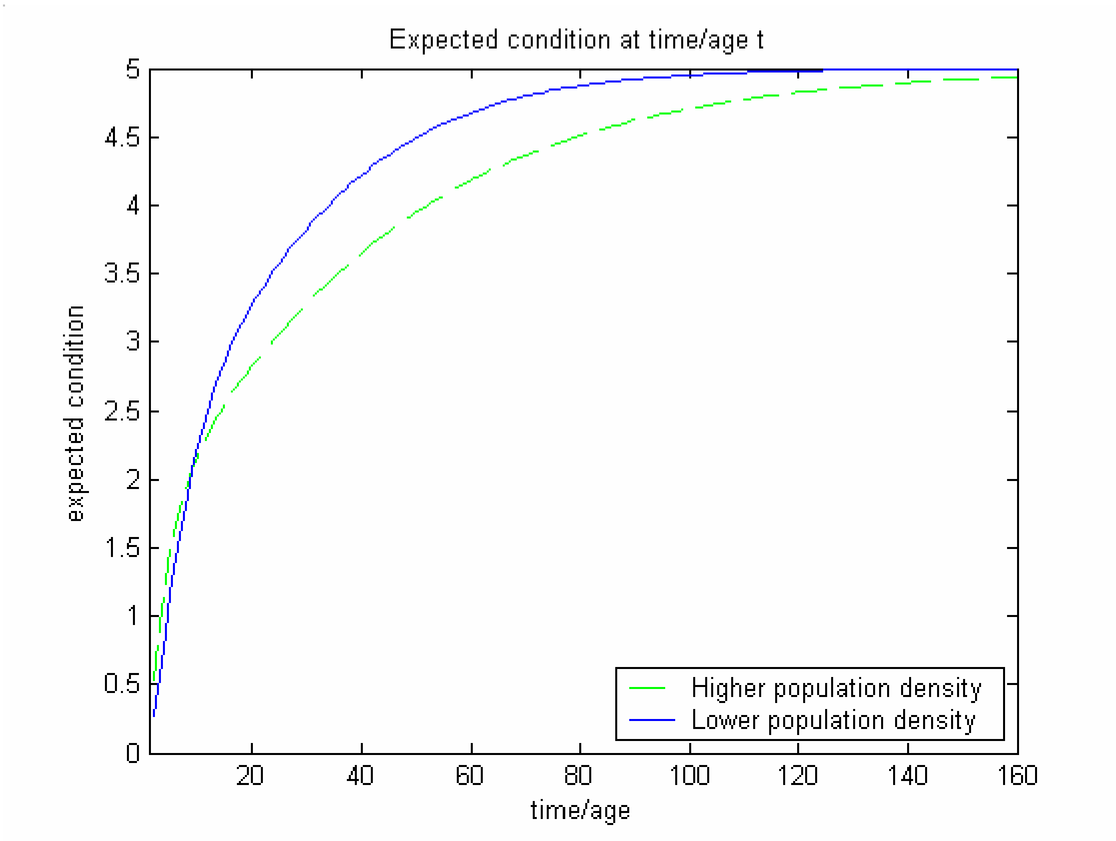

Figure 6-11 Expected condition at time t for covariate "Population density"

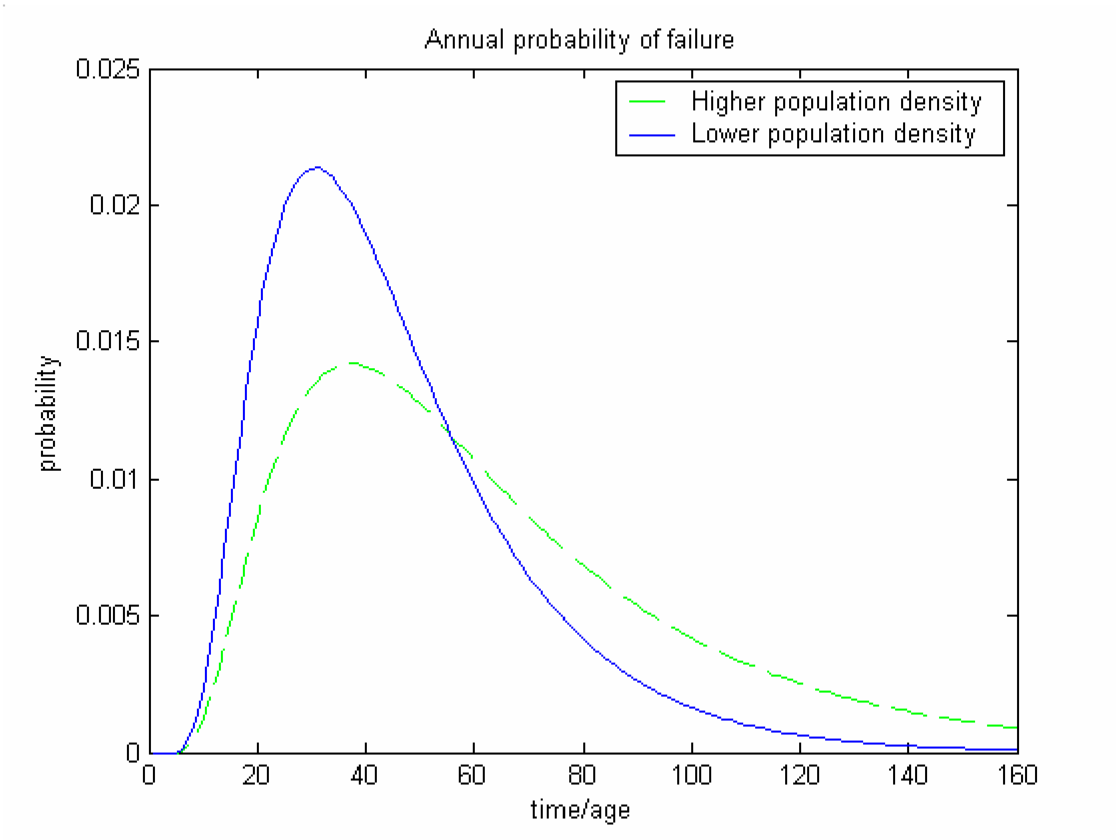

Figure 6-12 Annual probability of failure for covariate "Population density"

6.3 Logistic regression

In this section we would like to investigate the influence of the covariates on the model parameters, i.e. the transition probabilities, using logistic regression.

Logistic regression allows us to predict a discrete outcome from a set of variables that may be continuous, discrete, dichotomous, or a mix of any of these. Generally, the dependent or response variable is dichotomous, such as success/failure or presence/absence and, in our case, transition to the next state or not. Hence it can take value 1 with probability p, or the value 0 with probability $1-p$. Our Markov chain can make each transition with probability p or it may remain in given state with probability $1-p$. This makes its parameters independent binary variables. In logistic regression the relationship between the covariates and response variables is given by the formula:

$$p = \frac{\exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n)}{1 + \exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n)}, \tag{6.5}$$

where α is the constant of the equation showing us what the dependent variable p will be, when all the independent variables are zero. The other parameters $\underline{\beta} = (\beta_1, ..., \beta_n)$, are the coefficients of the covariates, representing the amount the dependent variable p changes when the independent variables changes 1 unit. The logistic regression equation can be rewritten as:

$$\log\left[\frac{p}{1-p}\right] = \alpha + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n \tag{6.6}$$

The goal of logistic regression is to correctly predict the influence of the independent (predictor) variables (covariates) on the dependent variable (transition probability) for individual cases. A model is created that includes all predictor variables that may be useful in predicting the response variable.

The process by which coefficients are tested for significance involves several different techniques. One of them is a Wald Test. This test is a way of testing whether the parameters associated with a group of covariates are zero. Hence it tests the statistical significance of each coefficient β_i in the model. For each coefficient β_i, there will be an associated parameter, z, which is called a Z- statistic, which is:

$$z = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}, \tag{6.7}$$

where $\hat{\beta}_i$ is the estimated value for the coefficient β_i and $SE(\hat{\beta}_i)$ denotes standard error of this estimated coefficient and is defined as the positive square root of the variance of these estimated coefficient.

If for a particular covariate, or group of covariates, the Wald test is significant, then we would conclude that the parameters associated with these covariates are not zero, so that the covariates should be included in the model. If the Wald test is not significant than these covariates can be omitted from the model.

The ratio given in (6.7) is the ratio of the estimated coefficient to its estimated standard error. Under the null hypothesis, the Wald statistic follows a standard normal distribution. Having standard error, we can compute a 95% confidence interval for the coefficient, being

$$\hat{\beta} \pm z_{1-\alpha/2} SE(\hat{\beta}),$$

where the $z_{1-\alpha/2}$ denotes the $100(1-\alpha)$ percentile of the standard normal distribution. Under the null hypothesis that the coefficient is equal to zero, we can reject the hypothesis if the two-tailed p-value of the Wald test $\Pr\left\{ |z| > \frac{\hat{\beta}}{SE(\hat{\beta})} \right\}$ is less than 5% (Hosmer et al., 1999).

A model developed in this paper requires 5 different transition probabilities, which are presented in (5.12). We define each of these probabilities in terms of a function of the covariates as follows:

$$p_i = p_i(\underline{\beta_i}\underline{x}') = \frac{\exp(\underline{\beta_i}\underline{x}')}{1+\exp(\underline{\beta_i}\underline{x}')} \qquad i = 1,2,3,4,5 \tag{6.8}$$

The vector $\underline{x}$ contains the covariates which is for the j^{th} object equal to $\underline{x}_j = (1, x_{j_1}, x_{j_2}, ..., x_{j_5})$. Obviously, each element of the vector $\underline{x}$ has dichotomous distribution, i.e. it is equal to 1 if given covariates appear for given object or 0 if not. There are also 5 logistic regressions, for each of the model parameters, having parametric vector $\underline{\beta_i} = (\beta_{i_0}, \beta_{i_1}, ..., \beta_{i_5})$, where β_{i_0} is equivalent to the constant of the equation and the rest are coefficients of covariates. Large positive (negative) values of $\underline{\beta_i}\underline{x}'$ yield large (small) transition probabilities.

6.3.1 Estimating the regression coefficients

The goal is to estimate the parametric vectors $\underline{\beta_i} = (\beta_{i_0}, \beta_{i_1}, ..., \beta_{i_5})$. The model involving the transition probabilities is given by:

$$\begin{bmatrix} 1-p_1(\underline{\beta_1}\underline{x}') & p_1(\underline{\beta_1}\underline{x}') & 0 & \cdots & 0 \\ 0 & 1-p_2(\underline{\beta_2}\underline{x}') & p_2(\underline{\beta_2}\underline{x}') & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & \cdots & 0 & 1-p_5(\underline{\beta_5}\underline{x}') & p_5(\underline{\beta_5}\underline{x}') \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} \tag{6.9}$$

By substituting (6.8) into (6.9), we get the one-step transition probability matrix from which we can estimate the vectors $\underline{\beta_i} = (\beta_{i_0}, \beta_{i_1}, ..., \beta_{i_5})$. The estimates are obtained by maximizing the likelihood function using the transition probability matrix in (6.9).

To simplify the notation, we restrict the evaluation of the likelihood function to the model with state-independent probabilities given in (5.11).

The overall likelihood is thus given by:

$$L(x_1, x_2, ..., x_n \mid p) = \prod_{k=1}^{n} (\mathrm{P}^{m_k})_{i_k\, j_k} ,$$

where P represents the one-step probability matrix and the observations $x_1, x_2, ..., x_n$ contain information about transitions and the times between them. Maximizing the likelihood is equivalent to finding the maximum of the log-likelihood function, which is represented by the following formula:

$$\ell(\underline{\mathrm{x}}|p) \; = \; \log L(\underline{\mathrm{x}}|p) = \sum_{k=1}^{n} \log (\mathrm{P}^{m_k})_{i_k\, j_k} . \tag{6.10}$$

To calculate the vector of coefficients $\underline{\beta} = (\beta_0, \beta_1, ..., \beta_5)$, we should take the derivative of the log-likelihood with respect to this vector and equate it to zero.

To calculate the MLE with respect to the coefficient vector $\underline{\beta} = (\beta_0, \beta_1, ..., \beta_5)$, we use the chain rule. This rule tells us that the derivative of the function $f(u)$ is the derivative of f, evaluated at that u, times the derivative of u. Hence, if $y = f(g(x))$ and $u = g(x)$, then

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx} . \tag{6.11}$$

Applying the above rule to (6.10) we have that

$$\begin{aligned}
\frac{\partial}{\partial \beta_l} \ell(\underline{\mathrm{x}}|p) &= \frac{\partial \ell(\underline{\mathrm{x}}|p)}{\partial p} \cdot \frac{\partial p}{\partial \beta_l} = \frac{\partial}{\partial p} \sum_{k=1}^{n} \log (\mathrm{P}^{m_k})_{i_k\, j_k} \cdot \frac{\partial p}{\partial \beta_l} \\
&= \sum_{k=1}^{n} \frac{\partial}{\partial p} \log (\mathrm{P}^{m_k})_{i_k\, j_k} \cdot \frac{\partial p}{\partial \beta_l} \\
&= \sum_{k=1}^{n} \frac{1}{(\mathrm{P}^{m_k})_{i_k\, j_k}} \frac{\partial}{\partial p} (\mathrm{P}^{m_k})_{i_k\, j_k} \cdot \frac{\partial p}{\partial \beta_l}
\end{aligned}$$

Using formula (5.13) we get

$$\frac{\partial}{\partial \beta_l} \ell(\underline{\mathrm{x}}|p) = \sum_{k=1}^{n} \frac{1}{(\mathrm{P}^{m_k})_{i_k\, j_k}} \left(\sum_{b=0}^{m_k - 1} (\mathrm{P}^b) \cdot \frac{\partial}{\partial p} \mathrm{P} \cdot (\mathrm{P}^{m_k - 1 - b}) \right)_{i_k\, j_k} \cdot \frac{\partial p}{\partial \beta_l} \tag{6.12}$$

When the expression of (6.8) is substituted in (6.12), we find that the derivative of log-likelihood with respect to $\underline{\beta} = (\beta_0, \beta_1, ..., \beta_5)$ is given by:

$$\frac{\partial}{\partial\beta_l}\ell(\underline{\mathrm{x}}|p)=\sum_{k=1}^{n}\frac{1}{(\mathrm{P}^{m_k})_{i_k j_k}}\left(\sum_{b=0}^{m_k-1}(\mathrm{P}^b)\cdot\frac{\partial}{\partial p}\mathrm{P}\cdot(\mathrm{P}^{m_k-1-b})\right)_{i_k j_k}\cdot\frac{\partial p}{\partial\beta_l}$$

$$=\sum_{k=1}^{n}\frac{1}{(\mathrm{P}^{m_k})_{i_k j_k}}\left(\sum_{b=0}^{m_k-1}(\mathrm{P}^b)\cdot\frac{\partial}{\partial p}\mathrm{P}\cdot(\mathrm{P}^{m_k-1-b})\right)_{i_k j_k}\cdot\frac{\partial}{\partial\beta_l}\left(\frac{\exp(\beta\underline{z}')}{1+\exp(\beta\underline{z}')}\right)$$

$$=\sum_{k=1}^{n}\frac{1}{(\mathrm{P}^{m_k})_{i_k j_k}}\left(\sum_{b=0}^{m_k-1}(\mathrm{P}^b)\cdot\frac{\partial}{\partial p}\mathrm{P}\cdot(\mathrm{P}^{m_k-1-b})\right)_{i_k j_k}\cdot\frac{z_l\exp(\beta\underline{z}')}{(1+\exp(\beta\underline{z}'))^2}$$

To calculate the variance and standard deviation of these parameters we need to know a Fisher information matrix which is defined in Theorem 5.1. This matrix is described as expected value of the second derivative of the log-likelihood function. This second derivative is given by the formula:

$$\frac{\partial^2}{\partial\beta_l\partial\beta_j}\ell(\underline{\mathrm{x}}|p)=\frac{\partial}{\partial\beta_j}\left(\frac{\partial}{\partial\beta_l}\ell(\underline{\mathrm{x}}|p)\right)=\frac{\partial}{\partial\beta_j}\left(\sum_{k=1}^{n}\frac{1}{(\mathrm{P}^{m_k})_{i_k j_k}}\frac{\partial}{\partial p}(\mathrm{P}^{m_k})_{i_k j_k}\cdot\frac{\partial p}{\partial\beta_l}\right)$$

$$=\sum_{k=1}^{n}\frac{\partial}{\partial\beta_j}\left(\frac{1}{(\mathrm{P}^{m_k})_{i_k j_k}}\frac{\partial}{\partial p}(\mathrm{P}^{m_k})_{i_k j_k}\cdot\frac{\partial p}{\partial\beta_l}\right)$$

$$=\sum_{k=1}^{n}\frac{\partial}{\partial\beta_j}\left(\frac{\frac{\partial}{\partial p}(\mathrm{P}^{m_k})_{i_k j_k}}{(\mathrm{P}^{m_k})_{i_k j_k}}\right)\cdot\frac{\partial p}{\partial\beta_l}+\frac{\frac{\partial}{\partial p}(\mathrm{P}^{m_k})_{i_k j_k}}{(\mathrm{P}^{m_k})_{i_k j_k}}\cdot\frac{\partial^2 p}{\partial\beta_l\partial\beta_j}$$

$$=\sum_{k=1}^{n}\frac{\partial}{\partial p}\left(\frac{\frac{\partial}{\partial p}(\mathrm{P}^{m_k})_{i_k j_k}}{(\mathrm{P}^{m_k})_{i_k j_k}}\right)\cdot\frac{\partial p}{\partial\beta_j}\cdot\frac{\partial p}{\partial\beta_l}+\frac{\frac{\partial}{\partial p}(\mathrm{P}^{m_k})_{i_k j_k}}{(\mathrm{P}^{m_k})_{i_k j_k}}\cdot\frac{\partial^2 p}{\partial\beta_l\partial\beta_j}$$

$$=\sum_{k=1}^{n}\left(\frac{(\mathrm{P}^{m_k})_{i_k j_k}\cdot\frac{\partial^2}{\partial p^2}(\mathrm{P}^{m_k})_{i_k j_k}-\frac{\partial}{\partial p}(\mathrm{P}^{m_k})_{i_k j_k}\frac{\partial}{\partial p}(\mathrm{P}^{m_k})_{i_k j_k}}{{(\mathrm{P}^{m_k})_{i_k j_k}}^2}\cdot\frac{\partial p}{\partial\beta_j}\cdot\frac{\partial p}{\partial\beta_l}+\frac{\frac{\partial}{\partial p}(\mathrm{P}^{m_k})_{i_k j_k}}{(\mathrm{P}^{m_k})_{i_k j_k}}\cdot\frac{\partial^2 p}{\partial\beta_l\partial\beta_j}\right)$$

We can rewrite the result more clearly as follows:

$$\frac{\partial^2}{\partial\beta_l\partial\beta_j}\ell(\underline{\mathrm{x}}|p)=$$

$$=\sum_{k=1}^{n}\left(\frac{-\left(\frac{\partial}{\partial p}(\mathrm{P}^{m_k})_{i_k j_k}\right)^2}{{(\mathrm{P}^{m_k})_{i_k j_k}}^2}\cdot\frac{z_l z_j\exp(2\beta\underline{z}')}{(1+\exp(\beta\underline{z}'))^4}+\frac{\frac{\partial}{\partial p}(\mathrm{P}^{m_k})_{i_k j_k}}{(\mathrm{P}^{m_k})_{i_k j_k}}\cdot\frac{z_l z_j\exp(\beta\underline{z}')-z_l z_j\exp(3\beta\underline{z}')}{(1+\exp(\beta\underline{z}'))^4}\right)\quad(6.13)$$

The negative value of the second derivative of the log-likelihood given in (6.13) is known as the observed Fisher information matrix, and denoted as

$$\mathbf{I}(\beta) = -\frac{\partial^2 \ell(\underline{x} \mid p)}{\partial \beta_l \partial \beta_j}. \tag{6.14}$$

The estimated covariance matrix of the coefficients is the inverse of (6.14) evaluated at $\hat{\beta}$ and is

$$\text{Var}\ (\hat{\beta}) = \mathbf{I}(\hat{\beta})^{-1} \tag{6.15}$$

The estimator of the standard error, denoted by $SE(\hat{\beta})$, is the positive square root of the variance estimator in (6.15).

Below we present the results of this analysis. Table 6-4 and its continuation in Table 6-5 show the estimated beta coefficients (for 5 covariates) for each of the transition probabilities from (6.9) together with the 95% confidence interval. These tables also contain information about the Wald test, z-statistic and p-value for this test. Assuming a 5% significance level, we can reject the null-hypothesis that the estimated beta coefficient is equal to zero for a *p*-value less than the 5%. Confidence intervals and p-values are presented in brackets below the corresponding estimated values.

The covariates, which were taken into account, are:

- Location (over the road)
- Type of bridge (Concrete bridge)
- Type of traffic (Mixture of traffic)
- Population density (high)
- Construction year (built before 1976)

We only use these selected covariates, because they were involved in the logistic regression approach (independent variables are denoted as 1 if a given object belongs to this covariate or 0 if not; hence, for example, covariate "Location-over the road" is a reference group and is denoted as 1, while covariate "Location-in the road" is denoted as 0).

As we see, only covariates "Type of bridge-concrete bridge", "Population density-high", and "Construction year-Built before 1976" have influence on the model parameters. Looking more carefully, we observe that covariate "Type of bridge - concrete bridge" influences only the first two transition probabilities, i.e. transitions from state 0 to state 1, and from state 1 into state 2. Their confidence intervals do not contain zero and the Wald test has values 4.92 and 2.54 with p-values 0.00 and 0.01 respectively. The notation 0.00 denotes that this is a very small number, close to zero, but not equal to zero. The next covariate "Population density-high" influences all transition probabilities, while the covariate "Construction year-Built before 1976" influences the first three parameters and the last one. This is especially visible for the first transition, from state 0 into state 1. For this parameter, the beta coefficient corresponding with covariate "Construction year" has value 1.76, which results in a high transition probability.

	Model for					
	p_1		p_2		p_3	
	Coeff.	z	Coeff.	z	Coeff.	z
Intercept	-1.13 (-1.29 -0.97)	-13.8 (0.0)	-0.92 (-1.05 -0.8)	-14.25 (0.0)	-1.76 (-1.88 -1.64)	-28.94 (0.0)
Location (Over the road)	-0.14 (-0.36 0.07)	-1.29 (0.19)	-0.10 (-0.26 0.06)	-1.25 (0.21)	-0.11 (-0.24 0.02)	-1.68 (0.09)
Type of bridge (Concrete bridges)	**0.9026*** (0.54 1.26)	4.92 (0.00)	**0.21*** (0.05 0.38)	2.54 (0.01)	-0.07 (-0.19 0.05)	-1.15 (0.25)
Type of traffic (Mixture of traffic)	-0.06 (-0.33 0.21)	-0.43 (0.66)	**0.20*** (0.00 0.38)	2.00 (0.04)	0.049 (-0.08 0.18)	0.72 (0.47)
Population density (High)	**0.15*** (0.13 0.18)	13.08 (0)	**-0.03*** (-0.05 -0.02)	-4.23 (0.0)	**-0.10*** (-0.11 -0.09)	-15.49 (0.0)
Construction year (built before 1976)	**1.76*** (1.13 2.39)	5.47 (0.0)	**0.69*** (0.57 0.81)	11.42 (0)	**0.27*** (0.17 0.37)	5.24 (0.0)

Table 6-5 Estimated Coefficient and z-Score for covariates Location, Type of use, Use, Population density and Construction year.

	Model for			
	p_4		p_5	
	Coeff.	z	Coeff.	z
Intercept	-3.22 (-3.51 -2.94)	-22.19 (0.0)	-1.93 (-2.58 -1.28)	-5.82 (0.0)
Location (Over the road)	0.01 (-0.31 0.34)	0.08 (0.93)	0.01 (-0.88 0.90)	0.023 (0.98)
Type of bridge (Concrete bridges)	-0.03 (-0.33 0.27)	-0.19 (0.85)	-0.38 (-1.31 0.54)	-0.81 (0.42)
Type of traffic (Mixture of traffic)	0.01 (-0.32 0.34)	0.063 (0.94)	0.08 (-0.89 1.05)	0.165 (0.87)
Population density (High)	**-0.06*** (-0.089 -0.03)	-4.287 (0.0)	**0.08*** (0.012 0.15)	2.29 (0.02)
Construction year (built before 1976)	0.21 (-0.07 0.48)	1.44 (0.15)	**-0.82*** (-1.47 -0.17	-2.47 (0.013)

Table 6-6 Estimated Coefficient and z-Score for covariates Location, Type of use, Use, Population density and Construction year, continuation.

* indicates statistical significance

6.4 Conclusions

In this chapter, we assessed the influence of the covariates on the model parameters, which are the transition probabilities of the Markov chain modelling bridge deterioration. From the above analysis we may conclude that some covariates have significant impact on transition probabilities and some of them not. At the beginning of this analysis, we estimated the model parameters for each of the covariates separately and then compared them to the model parameters for all data without taking into account the covariates. The results are given in tables 6-1, 6-2 and 6-3. In Table 6-2, we present transition probabilities for the last covariance "Province" in which we consider each of these provinces separately. For some of the 13 provinces we did not have enough data to estimate the transition probabilities and for this reason we decided to group these covariates. Hence, in Table 6-3 we present the results for grouped provinces. We choose to use a grouping: "Proximity to the sea" and "Population density". For the subsequent analysis we have decided to use the covariate "Population density".

Given the values of the transition probabilities, we calculated the expected condition at given time and the corresponding probability distribution of the time to reach a failure state (state5). These results are presented in figures to better interpret the differences between each group in given covariates. The visual interpretation confirms our decision from Chapter 5, to use the model with state-dependent transition probabilities. Therefore, we used only this model to visualize the differences in covariates.

In the last part of this chapter we analysed the influences of the covariates on the transition probabilities using logistic regression. Results are presented in Tables 6-5 and Table 6-6. From these tables, we conclude that only three covariates have statistically significant influence on the transition probabilities. These are: "Type of bridge - concrete bridge", "Population density-high", and "Construction year (built before 1976)".

7 Uncertainty in conditions

Every time when an inspection was made, the inspector gave his condition rate according to his/her point of view and interpretation. However, the evaluation of the inspectors is subjective and sometimes inconsistent. We believe that taking this subjectivity into account would be important and results in a better Markov model.

Let's assume that we inspect one object. When the inspector inspects the deterioration process, he/she ranks the condition on scale from 0 to 5 (6), denoting the states from new to very bad (see Table 2-1). Let's assume that he/she classified our object into state 3. A few years later, another inspector performed inspection, and his rate is also 3 or even 2, but no maintenance was performed. This may indicate that during those few years, the deterioration did not progress or it progressed but the first or second inspector classified it wrong. It is possible that during the first inspection, the bridge actually was in state 2 and then passed into state 3, or it was in state 2 and remained there, and so on…. We can guess what exactly happens and we still are not sure of the condition after each inspection. In other words, we would like to know what is the probability of observing state i, given the true condition is j, and what the difference is between the observed process and the actual one.

The following analysis is done based on Hidden Markov Models (Lystig *et al*., 2002), where the main assumption is that the hidden process is Markov, while the observed process is not. Hence, assume that we have one or more independent sequences of T observations $Y_1, Y_2, ..., Y_T$. We claim that there exists an unobserved sequence $X_1, X_2, ..., X_T$ of the true Markov process. To reduce the number of possibilities, we assume that an observed process of states can differ from the actual one only by a small amount. We relate the original process $X_1, X_2, ..., X_T$ to $Y_1, Y_2, ..., Y_T$ through simple assumptions, which are:

$$\Pr\{X_t \mid X_{t-1}, X_{t-2}, ..., X_1, X_0\} = \Pr\{X_t \mid X_{t-1}\} \tag{7.1}$$

$$Y_t = X_t + \varepsilon_t \tag{7.2}$$

This means that the actual process forms a first order Markov chain and the observed process differs from the actual one by the inspector error ε_t. This error indicates a difference in condition rating and can take values from the set $\{-1,0,1\}$. The chain takes on one of s discrete values at each of the T time points of observation. Hence, the observed process is discrete as well as the actual one. Although the actual process is Markov, the observed process is not. Each realization of the observed process at given time points, is dependent upon all of the preceding values of the observed process at previous time points.

The actual Markov model is defined by two sets of probabilities. First, the $s \times s$ matrix $\mathrm{P_t}$ of transition probabilities between the actual states from time $t-1$ to time t, with p_{ij}^t denoting $\Pr\{X_t = j \mid X_{t-1} = i\}$. Second the probability distribution of the 3 possible error outcomes denoted by $\Pr\{\varepsilon_t = k\}$, where $k \in \{-1,0,1\} = \mathrm{S}$. We assume here that X and ε are independent.

Complete specification of the actual process involves parameterizing each of the sets of probabilities mentioned above. For instance, in the previous definitions, both the transition probabilities and the collection of inspection error have been indexed by t. This provides for the possible inclusion of time-varying covariates, affecting either the transitions between the actual states or the behaviour of the observed process, in the model. It is also straightforward to generalize the deterioration model to the case of multiple independent observation sequences with a slight complication in notation. For clarity of presentation, we will show the results for a single observation sequence and later introduce a more general model.

Let's denote the probability of going from state i to state j in t years as:

$$P_{ij}(t) = \Pr\{X(t) = j \mid X(0) = i\}, \text{ where } i, j = 1, 2, \ldots, s \qquad (7.3)$$

We are interested in finding the probability of observing state k, when the true state is j, which simply is $\Pr\{Y(t) = k, X(t) = j\}$. At the beginning, assume that we inspect only one object and we inspect once. Hence,

$$\begin{aligned}
\Pr\{Y(t) = k, X(t) = j\} &= \Pr\{Y(t) = k \mid X(t) = j\} \Pr\{X(t) = j\} \\
&= \sum_{i=1}^{s} \Pr\{Y(t) = k \mid X(t) = j\} \Pr\{X(t) = j \mid X(0) = i\} \Pr\{X(0) = i\} \\
&= \sum_{i=1}^{s} \Pr\{Y(t) = k \mid X(t) = j\} \Pr\{X(0) = i\} P_{ij}(t)
\end{aligned}$$

where $\Pr\{Y(t) = k \mid X(t) = j\}$ denotes probability of observing state *k* given the actual process is in state *j*.

Summing over all j, we will get the likelihood of observing state k, at time $t \geq 0$, given that we have only one observation for a given object; that is:

$$L(Y) = \Pr\{Y(t) = k\} = \sum_{j=1}^{s} \sum_{i=1}^{s} \Pr\{Y(t) = k \mid X(t) = j\} \Pr\{X(0) = i\} P_{ij}(t)$$

Now let's assume that we have two observations and the one-step transition probability matrix is given by (5.12), indicating state-dependent probabilities. For this case, the probability of observing two different states is:

$$\Pr\{Y(t_i) = k, Y(t_{i-1}) = j\} = \Pr\{Y(t_i) = k \mid Y(t_{i-1}) = j\} \Pr\{Y(t_{t-1}) = j\}$$

And making use of (7.2) we will get that:

$$\begin{aligned}
&\Pr\{Y(t_i) = k, Y(t_{i-1}) = j\} = \Pr\{X(t_i) = k - \varepsilon_{t_i}, X(t_{i-1}) = j - \varepsilon_{t_{i-1}}\} \\
&= \sum_{m \in S} \sum_{l \in S} \Pr\{X(t_i) = k - m, X(t_{i-1}) = j - l\} \Pr\{\varepsilon_{t_i} = m\} \Pr\{\varepsilon_{t_{i-1}} = l\} \\
&= \sum_{m \in S} \sum_{l \in S} \Pr\{X(t_i) = k - m \mid X(t_{i-1}) = j - l\} \Pr\{X(t_{i-1}) = j - l\} \Pr\{\varepsilon_{t_i} = m\} \Pr\{\varepsilon_{t_{i-1}} = l\} \\
&= \sum_{m \in S} \sum_{l \in S} \Pr\{X(t_{i-1}) = j - l\} \Pr\{\varepsilon_{t_i} = m\} \Pr\{\varepsilon_{t_{i-1}} = l\} P_{j-l,k-m}(t_i - t_{i-1})
\end{aligned}$$

where

$$\Pr\{X(t_{i-1}) = j-l\} = \sum_{i=1}^{s} \Pr\{X(0)=i\}\Pr\{X(t_{i-1}) = j-l \mid X(0)=i\}$$
$$= \sum_{i=1}^{s} \Pr\{X(0)=i\} P_{i,j-l}(t_{i-1})$$

Of course, if $k-\varepsilon_{t_i}, j-\varepsilon_{t_{i-1}} < 0$ or $k-\varepsilon_{t_i}, j-\varepsilon_{t_{i-1}} > 5$ then $P_{j-l,k-m}(t_i - t_{i-1}) = 0$.

At this moment we are ready to try to derive formula for the probability of observing state k, when the true state is j having more observations than two. To simplify the notation we will use Y_i instead $Y(t_i)$. Of course we still consider observations for only one object:

$$\Pr\{Y_n = k_n, Y_{n-1} = k_{n-1}, ..., Y_2 = k_2, Y_1 = k_1, Y_0 = k_0\}$$
$$= \Pr\{Y_n = k_n \mid Y_{n-1} = k_{n-1}, ..., Y_2 = k_2, Y_1 = k_1, Y_0 = k_0\} \cdot \Pr\{Y_{n-1} = k_{n-1}, ..., Y_2 = k_2, Y_1 = k_1, Y_0 = k_0\}$$
$$= \Pr\{Y_n = k_n \mid Y_{n-1} = k_{n-1}, ..., Y_2 = k_2, Y_1 = k_1, Y_0 = k_0\} \cdot$$
$$\Pr\{Y_{n-1} = k_{n-1} \mid Y_{n-2} = k_{n-2}, ..., Y_2 = k_2, Y_1 = k_1, Y_0 = k_0\} \cdot \Pr\{Y_{n-2} = k_{n-2}, ..., Y_2 = k_2, Y_1 = k_1, Y_0 = k_0\}$$
$$= .... =$$
$$= \Pr\{Y_n = k_n \mid Y_{n-1} = k_{n-1}, ..., Y_2 = k_2, Y_1 = k_1, Y_0 = k_0\}$$
$$\cdot \Pr\{Y_{n-1} = k_{n-1} \mid Y_{n-1} = k_{n-2}, ..., Y_2 = k_2, Y_1 = k_1, Y_0 = k_0\} \cdot ... \cdot \Pr\{Y_1 = k_1 \mid Y_0 = k_0\} \Pr\{Y_0 = k_0\}$$

In terms of the actual process, this equation reads as

$$\Pr\{Y_n = k_n, Y_{n-1} = k_{n-1}, ..., Y_2 = k_2, Y_1 = k_1, Y_0 = k_0\}$$
$$= \sum_{\varepsilon_n} ... \sum_{\varepsilon_0} \Pr\{X_n = k_n - \varepsilon_n \mid X_{n-1} = k_{n-1} - \varepsilon_{n-1}, ..., X_2 = k_2 - \varepsilon_2, X_1 = k_1 - \varepsilon_1, X_0 = k_0 - \varepsilon_0\}$$
$$\cdot \Pr\{X_{n-1} = k_{n-1} - \varepsilon_{n-1} \mid Y_{n-2} = k_{n-2} - \varepsilon_{n-2}, ..., X_2 = k_2 - \varepsilon_2, X_1 = k_1 - \varepsilon_1, X_0 = k_0 - \varepsilon_0\}$$
$$\cdot ... \cdot \Pr\{X_1 = k_1 - \varepsilon_1 \mid X_0 = k_0 - \varepsilon_0\} \cdot \Pr\{X_0 = k_0 - \varepsilon_0\} \cdot \Pr\{\varepsilon_n\} \cdot \Pr\{\varepsilon_{n-1}\} \cdot ... \cdot \Pr\{\varepsilon_1\} \cdot \Pr\{\varepsilon_0\}$$

We know that the actual process is Markov (7.1), hence

$$\Pr\{Y_n = k_n, Y_{n-1} = k_{n-1}, ..., Y_2 = k_2, Y_1 = k_1, Y_0 = k_0\}$$
$$= \sum_{\varepsilon_n} ... \sum_{\varepsilon_0} \Pr\{X_n = k_n - \varepsilon_n \mid X_{n-1} = k_{n-1} - \varepsilon_{n-1}\} \cdot \Pr\{X_{n-1} = k_{n-1} - \varepsilon_{n-1} \mid X_{n-2} = k_{n-2} - \varepsilon_{n-2}\}$$
$$\cdot ... \cdot \Pr\{X_1 = k_1 - \varepsilon_1 \mid X_0 = k_0 - \varepsilon_0\} \Pr\{X_0 = k_0 - \varepsilon_0\} \cdot \Pr\{\varepsilon_n\} \cdot \Pr\{\varepsilon_{n-1}\} \cdot ... \cdot \Pr\{\varepsilon_1\} \cdot \Pr\{\varepsilon_0\}$$
$$= \sum_{\varepsilon_n} ... \sum_{\varepsilon_0} \Pr\{\varepsilon_0\} \Pr\{X_0 = k_0 - \varepsilon_0\} \cdot \prod_{i=1}^{n} \Pr\{X_i = k_i - \varepsilon_i \mid X_{i-1} = k_{i-1} - \varepsilon_{i-1}\} \cdot \Pr\{\varepsilon_i\} \qquad (7.4)$$

If we denote the likelihood function for a given object as:

$$L_i(Y_0^i, ..., Y_n^i \mid \underline{p}) = \Pr\{Y^i{}_n = k_n, Y^i{}_{n-1} = k_{n-1}, ..., Y^i{}_2 = k_2, Y^i{}_1 = k_1, Y^i{}_0 = k_0\}$$

Then the likelihood function for all objects is given by:

$$L(\underline{Y}^1,\ldots,\underline{Y}^m \mid \underline{p}) = \prod_{i=1}^{m} L_i(Y_0^i,\ldots,Y_n^i \mid \underline{p}) \tag{7.5}$$

Hence, according to (7.5) and (7.4) we get:

$$L(\underline{Y}^1,\ldots,\underline{Y}^m \mid \underline{p}) =$$

$$\prod_{i=1}^{m}\left(\sum_{\varepsilon_n}\cdots\sum_{\varepsilon_0}\Pr\{\varepsilon_0\}\Pr\left\{X_0^i = k_0-\varepsilon_0\right\}\cdot\prod_{j=1}^{n}\Pr\left\{X_j^i = k_j-\varepsilon_j \mid X_{j-1}^i = k_{j-1}-\varepsilon_{j-1}\right\}\cdot\Pr\left\{\varepsilon_j\right\}\right) \tag{7.6}$$

In our case, the chain always starts from state 0, this means $\Pr\{\varepsilon_0 = 0\} = 1$ implying $\Pr\{X_0 = k_0\} = 1$ if and only if $k_0 = 0$.

Hence we can write (7.6) in the simpler form

$$L(\underline{Y}^1,\ldots,\underline{Y}^m \mid \underline{p}) = \prod_{i=1}^{m}\left(\sum_{\varepsilon_n}\cdots\sum_{\varepsilon_0}\prod_{j=1}^{n}\Pr\left\{X_j^i = k_j-\varepsilon_j \mid X_{j-1}^i = k_{j-1}-\varepsilon_{j-1}\right\}\cdot\Pr\left\{\varepsilon_j\right\}\right) \tag{7.7}$$

Now, to estimate a new one-step probability matrix $\mathbf{P}$ with new transition probabilities p_{ij}, we need to maximize the log-likelihood function of (7.7); that is,

$$\begin{aligned}
\ell(\underline{Y}^1,\ldots,\underline{Y}^m \mid \underline{p}) &= \log L(\underline{Y}^1,\ldots,\underline{Y}^m \mid \underline{p}) \\
&= \log\left(\prod_{i=1}^{m}\left(\sum_{\varepsilon_n}\cdots\sum_{\varepsilon_0}\prod_{j=1}^{n}\Pr\left\{X_j^i = k_j-\varepsilon_j \mid X_{j-1}^i = k_{j-1}-\varepsilon_{j-1}\right\}\cdot\Pr\left\{\varepsilon_j\right\}\right)\right) \\
&= \sum_{i=1}^{m}\log\left(\sum_{\varepsilon_n}\cdots\sum_{\varepsilon_0}\prod_{j=1}^{n}\Pr\left\{X_j^i = k_j-\varepsilon_j \mid X_{j-1}^i = k_{j-1}-\varepsilon_{j-1}\right\}\cdot\Pr\left\{\varepsilon_j\right\}\right) \\
&= \sum_{i=1}^{m}\log\left(\sum_{\varepsilon_n}\cdots\sum_{\varepsilon_0}\prod_{j=1}^{n}P_{k_{j-1}-\varepsilon_{j-1},k_j-\varepsilon_j}(t_j-t_{j-1})\cdot\Pr\left\{\varepsilon_j\right\}\right)
\end{aligned}$$

To maximize this function with respect to vector $\underline{p}$, we need to calculate the first derivative of the above equation and then equate it to zero:

$$\begin{aligned}
\frac{\partial}{\partial p_u}\ell(\underline{Y}^1,\ldots,\underline{Y}^m \mid \underline{p}) &= \frac{\partial}{\partial p_u}\sum_{i=1}^{m}\log\left(\sum_{\varepsilon_n}\cdots\sum_{\varepsilon_0}\prod_{j=1}^{n}P_{k_{j-1}-\varepsilon_{j-1},k_j-\varepsilon_j}(t_j-t_{j-1})\cdot\Pr\left\{\varepsilon_j\right\}\right) \\
&= \sum_{i=1}^{m}\frac{\partial}{\partial p_u}\log\left(\sum_{\varepsilon_n}\cdots\sum_{\varepsilon_0}\prod_{j=1}^{n}P_{k_{j-1}-\varepsilon_{j-1},k_j-\varepsilon_j}(t_j-t_{j-1})\cdot\Pr\left\{\varepsilon_j\right\}\right) \\
&= \sum_{i=1}^{m}\frac{1}{A}\frac{\partial}{\partial p_u}A,
\end{aligned} \tag{7.8}$$

where

$$A = \sum_{\varepsilon_n}\cdots\sum_{\varepsilon_0}\prod_{j=1}^{n}P_{k_{j-1}-\varepsilon_{j-1},k_j-\varepsilon_j}(t_j-t_{j-1})\cdot\Pr\left\{\varepsilon_j\right\}$$

and

$$\frac{\partial}{\partial p_u} A = \frac{\partial}{\partial p_u} \sum_{\varepsilon_n} \cdots \sum_{\varepsilon_0} \prod_{j=1}^{n} P_{k_{j-1}-\varepsilon_{j-1}, k_j-\varepsilon_j}(t_j - t_{j-1}) \cdot \Pr\left\{\varepsilon_j\right\}$$

$$= \sum_{\varepsilon_n} \cdots \sum_{\varepsilon_0} \frac{\partial}{\partial p_u} \left(\prod_{j=1}^{n} P_{k_{j-1}-\varepsilon_{j-1}, k_j-\varepsilon_j}(t_j - t_{j-1}) \cdot \Pr\left\{\varepsilon_j\right\} \right) \quad (7.9)$$

Unfortunately, the above result can only be calculated numerically. Hence, we need to derive an algorithm to calculate the derivative with respect to vector $\underline{p}$, and determine a new one–step probability matrix $\mathbf{P}$.

At this moment we don't know how to calculate the above derivative. Hence, this problem is still open to be solved. Due to time constraint, the incorporation of inspection errors into the Markov deterioration model is left for future research.

8 Conclusions

In this thesis we conducted an analysis on deterioration modelling for bridges in the Netherlands. The data was collected in the database called DISK and provided by the Civil Engineering Division of Rijkswaterstaat. Because the data contained a limited number of condition states of the bridges, we decided to use the Markov chain to modelling the deterioration. The Markov deterioration model is condition-based; hence it is flexible in adapting it to inspection data (visual). Unfortunately, we could not observe the exact time of transitions. Hence, we adapt a Markov chain with interval censoring. Interval censoring means that we do not know the exact time of an event. In our context this means that we do not know the time when the bridge moves from one state to another state. A transition probability was defined as the probability that a bridge will move from one state to another (same or worse one). We assumed that no maintenance is possible.

The Markov property entails that the probability of deteriorating to another state doesn't depend on the history of the process, but only depends on the last condition. Hence, we would like to use the Markov process. Firstly, we verified whether the data supports the Markov property. Next, using Markov processes we built a transition probability matrix. This step resulted in two different models represented by one-year transition probability matrices. These matrices described deterioration process during one year. This involved a decision about which kind of model should be used for later analysis. When we chose the model, we took into account external parameters, which could influence the transition probabilities. We evaluated the expected condition at a given time and identified variables that are predictive for the transition probabilities. They are called covariates. The last step of analysis was to take into account subjectivity in inspections.

From all of these steps we can draw the following conclusions:

- Testing the Markov property was done by three different tests. The first test was based on contingency tables. An example of this test was given for transitions from state 0 into state 1. In this test we tested hypothesis that two series of condition data are from the same distribution. One of these series is a condition sequence of a present and future state and one is a condition sequence contained information about a past, present and future state. This test resulted in the decision that we can accept the hypothesis that the generated frequency is from the same distribution, because we can not reject it. This would indicate that we might assume the Markov property for transitions from state 0 to 1. This test depends on the way of building a contingency table and it doesn't give us an unambiguous answer to our question about the Markov property. For this reason, the next test was performed. The second test tests the null hypothesis, that a chain is of first order (doesn't depend on past state) against the alternative hypothesis that the chain is of second order (depend on past state). An example for which we performed this test indicated that we can not reject the null hypothesis. Hence, for the considered example the Markov property holds. Of course to make decisions about the Markov property we have to consider many more examples. This was done in other work of this author and the results indicated that we can assume the Markov property in the bridge deterioration data.
  The last test takes into account the assumption about the stationarity of transition probabilities. We considered 13 age groups to verify if transition probabilities depend on

time (age of the bridges). This test resulted in the decision that we can assume stationarity, which means that transition probabilities do not depend on time.

- Use of a Markov chain in modelling deterioration process required the defining of transition probabilities between the discrete condition states which are collected in matrix. The initial form of the matrix can be simplified since it is accepted that deterioration is a one-way process. Also improvement of the condition ratings can not be accomplished without maintenance; thus, all elements which indicated a backward process were assumed to be zero. The next assumption based on the fact that during a given period of time, i.e., a one-year, only a single transition of state is allowed. All these assumptions resulted in developing two deterioration models, one with state-independent and one with state-dependent transitions probabilities. The likelihood ratio test enabled us to decide which model fits best to our data. We chose the model with state-dependent probabilities as for this model the likelihood function has the maximum value.

- External parameters, which may influence the transition probabilities, are also considered. These parameters were chosen based on the use and location of the bridges. This was: "Construction year (built before and after 1976)", "Type of traffic (Mixture of traffic and Only truck and cars)", "Location (heavy traffic-in the road and light traffic-over the road)", "Type of bridge (concrete bridge and concrete viaduct)" and "Population density (Higher and Lower)". This set of covariates was proposed by the Civil Engineering Division of Rijkswaterstaat. The covariate "Population density" was selected from the more general covariate "Province", because we did not have enough data to estimate the effect of each province separately. To perform analysis, firstly we estimated the model parameters for each of these covariates separately and then compared them to the model parameters for all data without taking into account the covariates. Later we estimated the expected condition at a given time and the probability distribution of the time to reach the failure state. Because the results are presented in figures we could observe differences and similarities between each group of given sub-covariates and two developed models. The visual interpretation confirms our recommendation to use the model with state-dependent transition probabilities. We concluded that some of these covariates have significant impact on transition probabilities and some of them not, for example big influence indicate the covariate "Population density" while covariate "Location" has no effect. Hence, after the visual analysis, we performed logistic regression analysis to determine which covariates have more impact on the transition probabilities. The result is that only three covariates have a statistically significant influence on the transition probabilities. They are: "Type of bridge - concrete bridge", "High population density", and "Construction year (Bridges built before 1976)".

- In the last chapter, we noted that the evaluation of the inspectors is subjective and sometimes inconsistent. Hence, we tried incorporating this subjectivity into the Markov model and derived a general formula for the probability of observing a certain state given the true state. We think that calculating the derivative of the log-likelihood function, which is needed to estimate a new one-step probability matrix, could be done only numerically. Hence, we left this topic for future research due to time constraints.

We make the following recommendations for future research:

- Develop an algorithm which solves the formula for the derivative of the log-likelihood function taking into account the subjectivity of the inspectors given in Chapter 7. This introduces the possibility of inspection errors into the Markov deterioration model.

- Implement a method to solve the optimization problem for the covariate coefficients estimated in Chapter 6. Due to lack of time we used standard MATLAB optimization function, but it would be good to implement this by self and compare the results.

- Choose another combination of covariates in modelling dependence on transition probabilities and look on the changes. Maybe introducing different reference covariates give better results. In this report we use only covariates: "Location (Over the road)", "Type of bridge(Concrete bridge)", "Type of traffic(Mixture of traffic)", "Population density(high)" and "Construction year(built before 1976)". To improve logistic regression we should incorporate all covariates (the remaining ones) into the model. This will involve a change of data format.

- To test stationarity, we used groups of 5 years. There are 1748 useful transitions. To improve this it would be interesting to introduce another grouping and to compare the results. The best solution would be to use a one year grouping, but we do not have enough data for this analysis.

- Repeat tests for Markov property with a new deterioration data used in Chapter 5 and later. We didn't do this due to time constraints.

- Model used in this analysis does not include maintenance. It would be interesting to incorporate this into Markov model.

9 References

T. W. Anderson, Leo A. Goodman, *Statistical inference about Markov Chains,* Ann. Math. Stat., Vol. 28 (1957), pp.89-109 (M R. 18.944)

A. M. Mood, F. A. Graybill, D. C. Boes, *Introduction to the theory of statistics; Third edition*, Tokyo 1974, pp. 410-442.

William T. Scherer, Douglas M. Glagola, *Markovian Models for bridge maintenance management,* Journal of Transportation Engineering, Vol.120, No. 1, January/February, 1994, pages 37-51.

Stuart Coles, *An Introduction to Statistical modelling of Extreme Values*, London: Springer 2001.

Richard Lowry, *Concepts and Applications of Inferential Statistics*, professor of psychology, Vassar College, Poughkeepsie, NY USA. Internet website: http://vassun.vassar.edu/~lowry/webtext.html

Theodore C. Lystig, James P. Hughes, *Exact Computation of the Observed Information Matrix for Hidden Markov Models*, Journal of Computational and Graphical Statistics, Volume 11, Number 3 (2002), pages 678-689

David W. Hosmer, Stanley Lemeshow, *Applied survival analysis: regression modelling time to event data*, New York; Wiley, 1999

A. van Beek, G. C. M Gaal, J.M. van Noortwijk, J. D. Bakker, *Validation model for service life prediction of concrete structures*, 2nd International RILEM workshop on Life Prediction and Aging Management of Concrete Structures, Paris, France, 5-6 May 2003, pages 257-267

M. K. Skuriat-Olechnowska, *Test of hypothesis whether the Markov property holds for bridge deterioration data*, Internship Report, TU Delft, April 2005

J. D. Kalbfleisch, J. F. Lawless, *The analysis of panel data under a Markov assumption*, Journal of the American Statistical Association 80, year 1985, pages 863-871.

Golabi, K., R.B. Kulkarni, and G.B. Way (1982). *A state-wise pavement management system*. Interfaces, 12:5-21.

Golabi, K. , and R. Shepard (1997). *Pontis: A system for maintenance optimization and improvement of US bridge networks*. Interfaces, 27:71-88.

Thompson, P. D., E. P. Small, M. Johnson, and A. R. Marshall (1998). *The Pontis bridge management system*, Structural Engineering International*, IABSE*, 8(4):303-308.

Wang, K. C. P., and J. P. Zaniewski (1996). 20/30 *Hindsight: The new pavement optimization in the Arizona state highway network*. Interfaces, 26:77-89.

Appendices

Appendix A: Critical values of Chi-Square

	Level of Significance				
df	**.05**	**.025**	**.01**	**.005**	**.001**
1	3.84	5.02	6.63	7.88	10.83
2	5.99	7.38	9.21	10.60	13.82
3	7.81	9.35	11.34	12.84	16.27
4	9.49	11.14	13.28	14.86	18.47
5	11.07	12.83	15.09	16.75	20.51
6	12.59	14.45	16.81	18.55	22.46
7	14.07	16.01	18.48	20.28	24.32
8	15.51	17.53	20.09	21.95	26.12
9	16.92	19.02	21.67	23.59	27.88
10	18.31	20.48	23.21	25.19	29.59
11	19.68	21.92	24.73	26.76	31.26
12	21.03	23.34	26.22	28.30	32.91
13	22.36	24.74	27.69	29.82	34.53
14	23.68	26.12	29.14	31.32	36.12
15	25.00	27.49	30.58	32.80	37.70
16	26.30	28.85	32.00	34.27	39.25
17	27.59	30.19	33.41	35.72	40.79
18	28.87	31.53	34.81	37.16	42.31
19	30.14	32.85	36.19	38.58	43.82
20	31.41	34.17	37.57	40.00	45.31
21	32.67	35.48	38.93	41.40	46.80
22	33.92	36.78	40.29	42.80	48.27
23	35.17	38.08	41.64	44.18	49.73
24	36.42	39.36	42.98	45.56	51.18
25	37.65	40.65	44.31	46.93	52.62
26	38.89	41.92	45.64	48.29	54.05
27	40.11	43.19	46.96	49.65	55.48
28	41.34	44.46	48.28	50.99	56.89
29	42.56	45.72	49.59	52.34	58.30
30	43.77	46.98	50.89	53.67	59.70

Appendix B: Expected condition for covariates

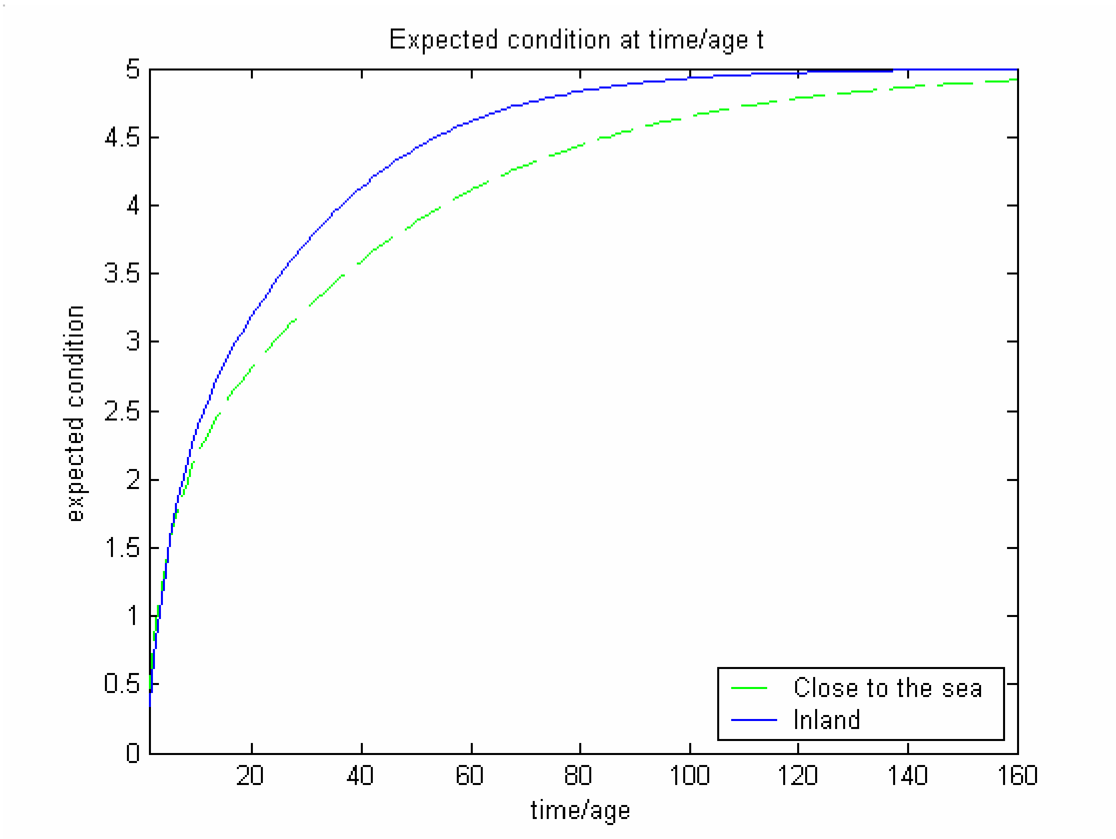

Figure B-1 Expected condition at time t for covariate "Proximity to the sea"

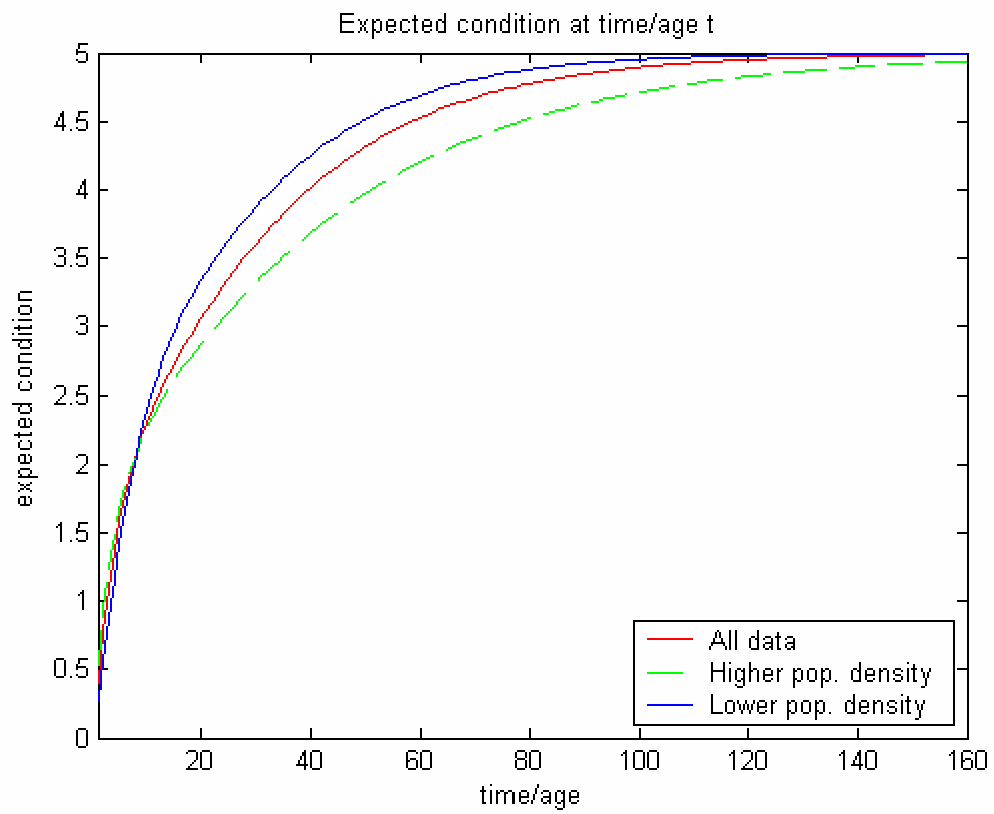

Figure B-2 Expected condition at time t for covariate "population density"

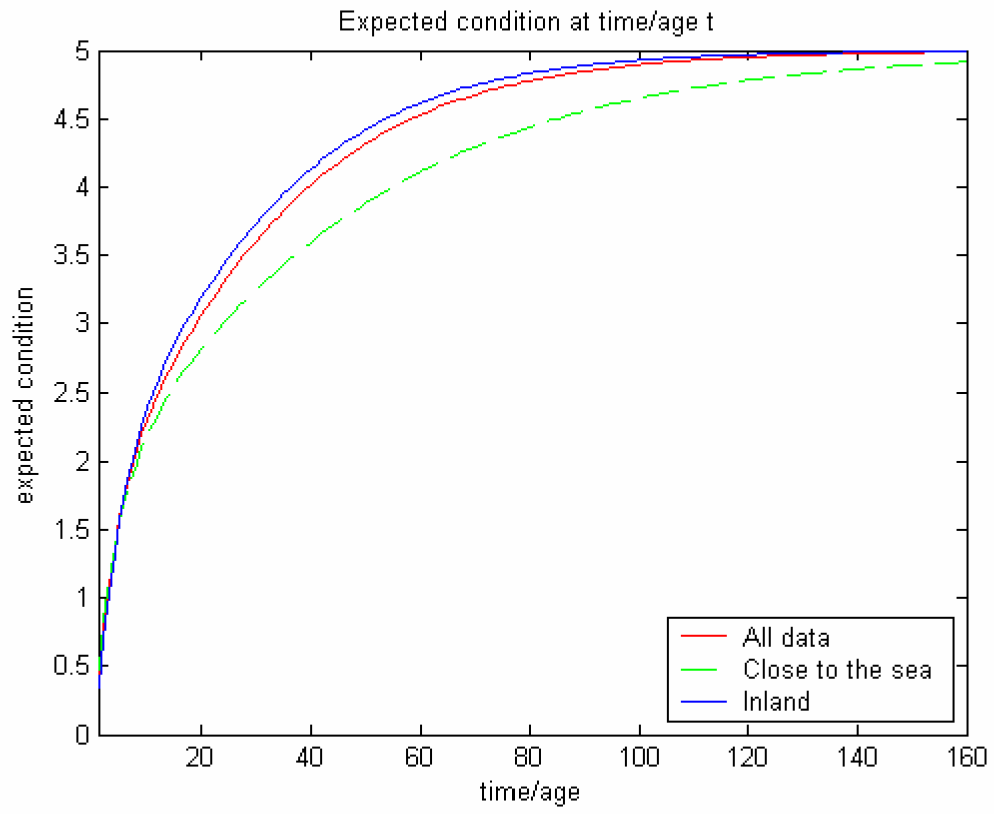

Figure B-3 Expected condition at time t for covariate "Proximity to the sea"

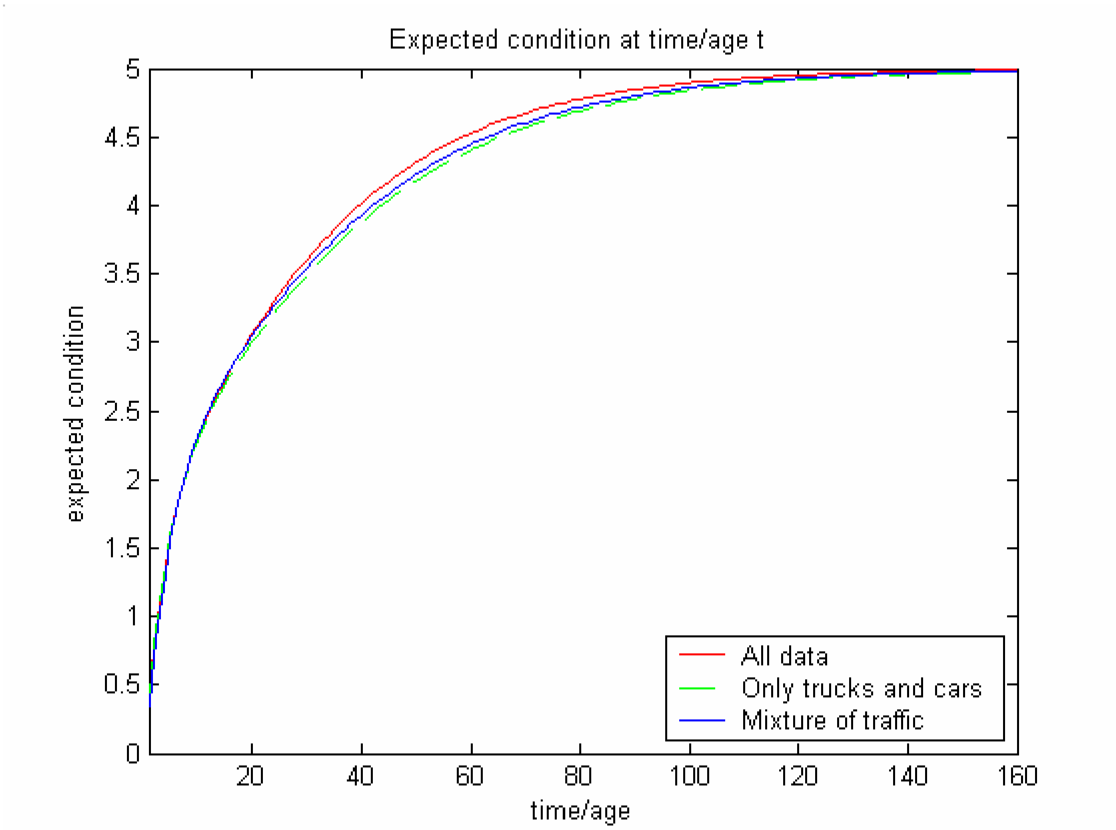

Figure B-4 Expected condition at time t *for covariate "Type of traffic"*

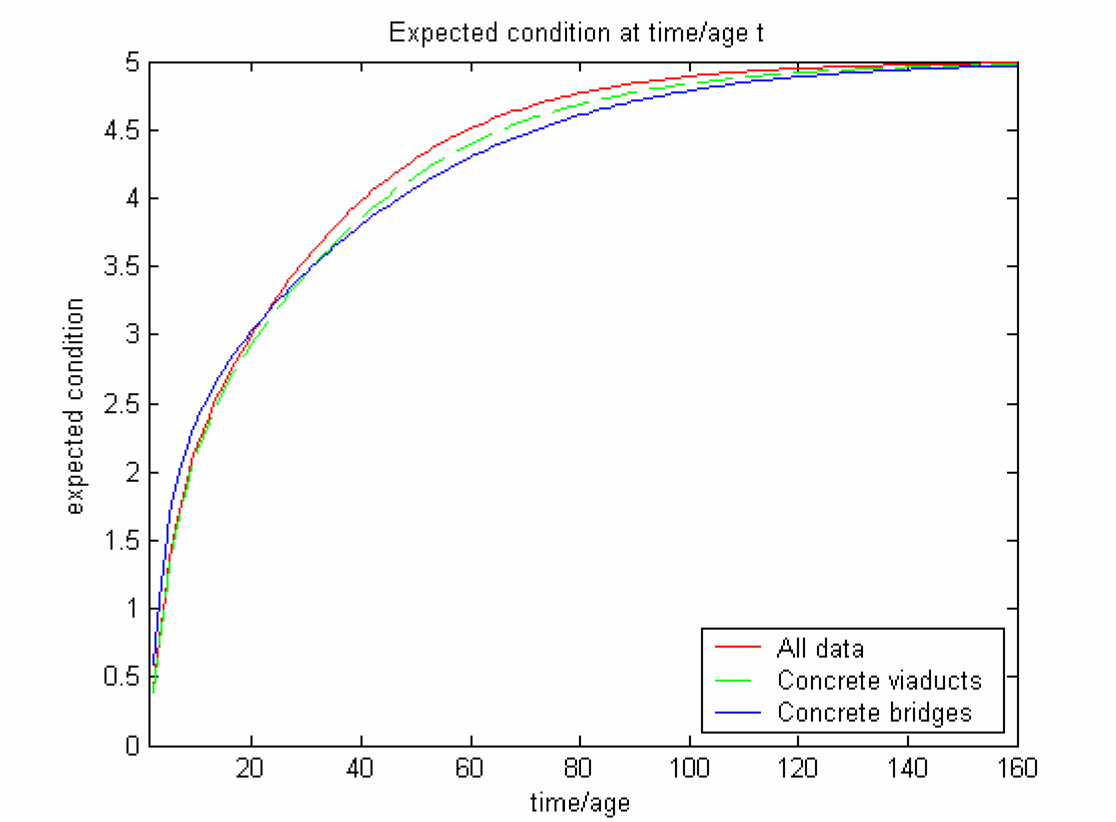

Figure B-5 Expected condition at time t *for covariate "Type of use"*

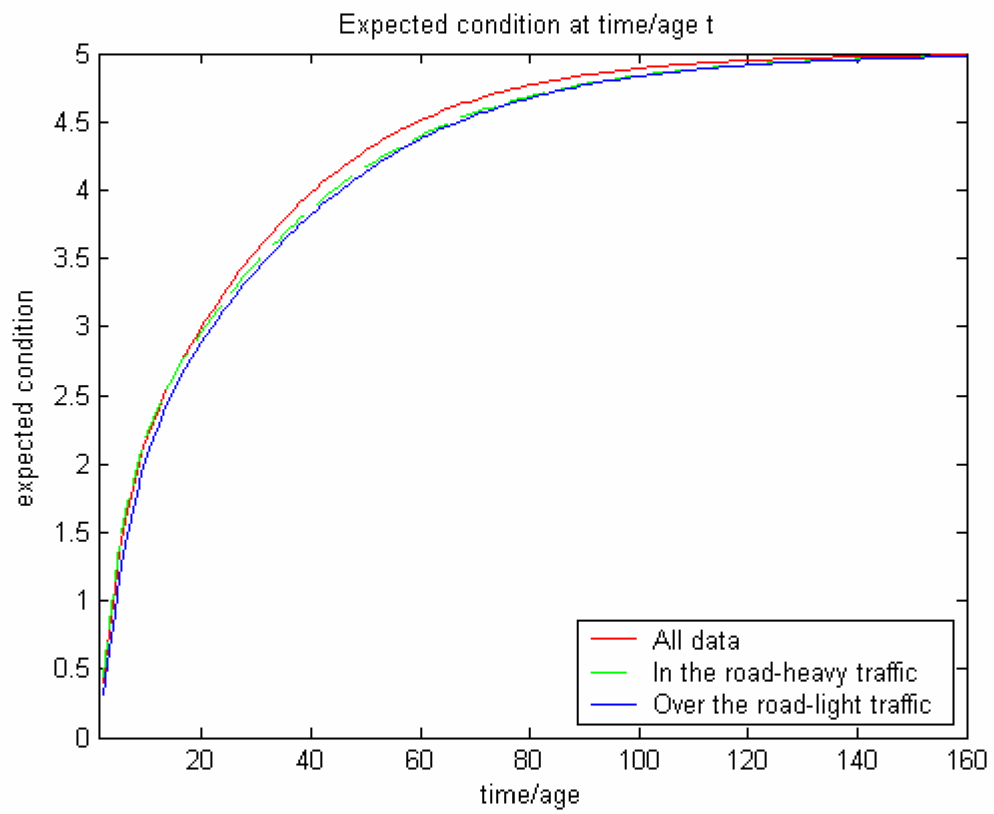

Figure B-6 Expected condition at time t for covariate "Location"

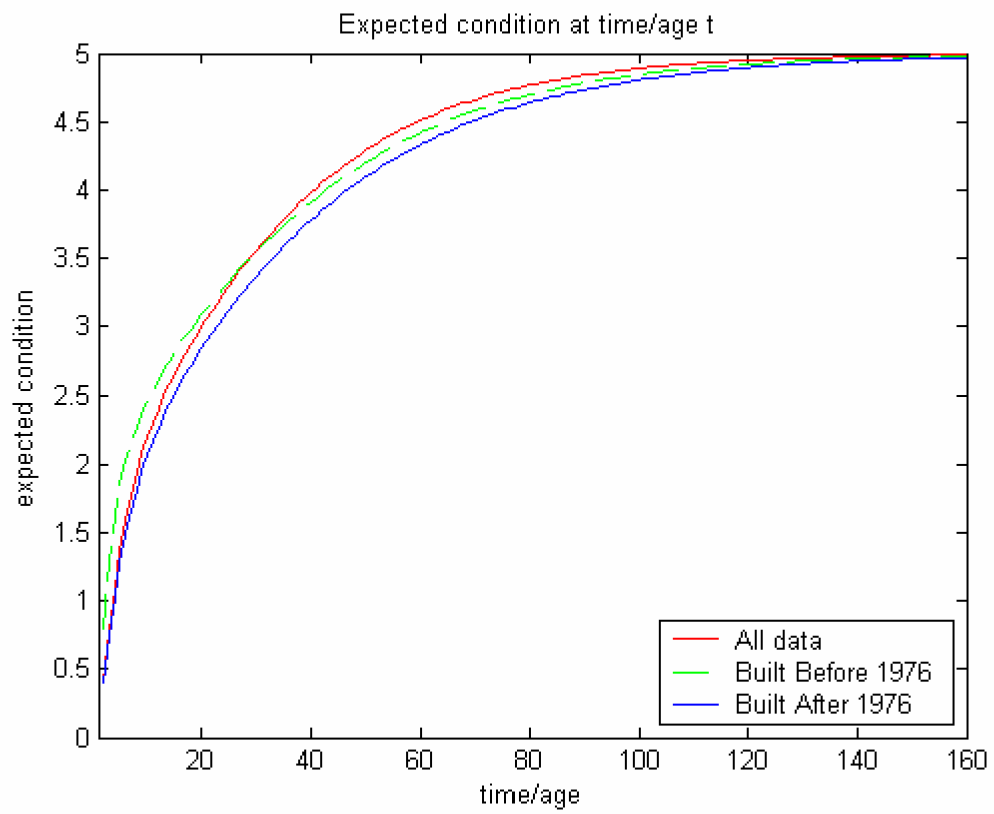

Figure B-7 Expected condition at time t for covariate "Construction year"

Appendix C: Annual probability of failure

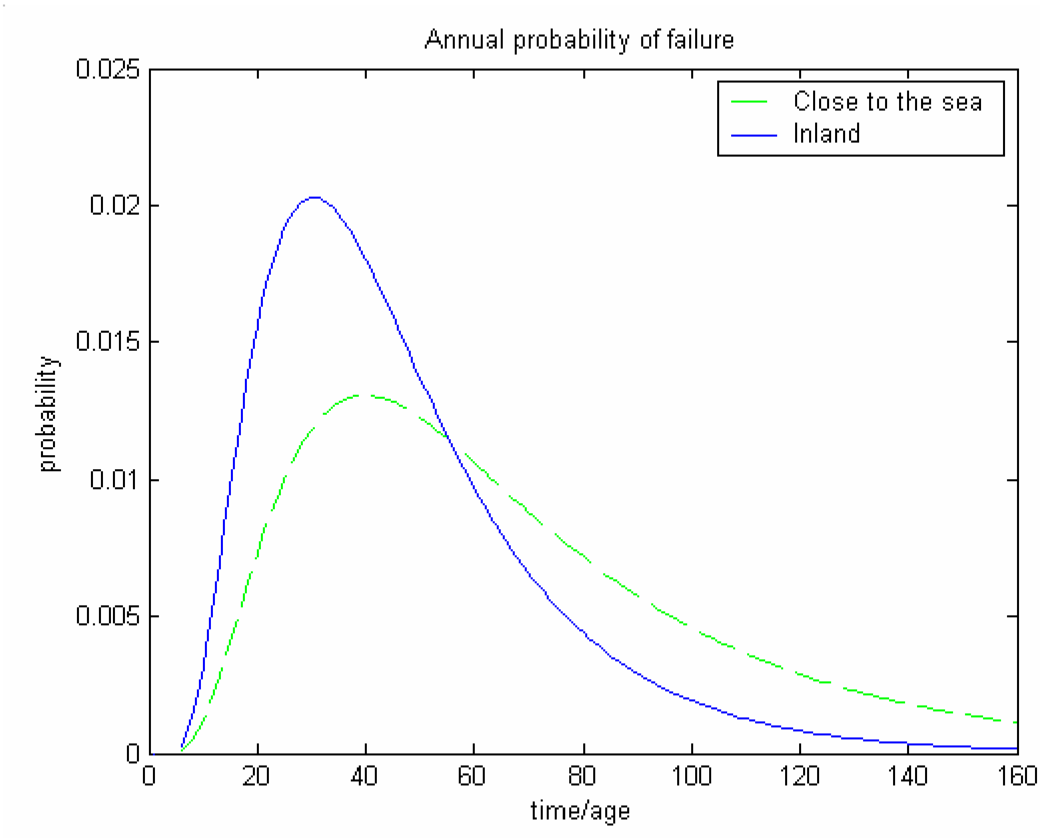

Figure C-1 Annual probability of failure for covariate "Proximity to the sea"

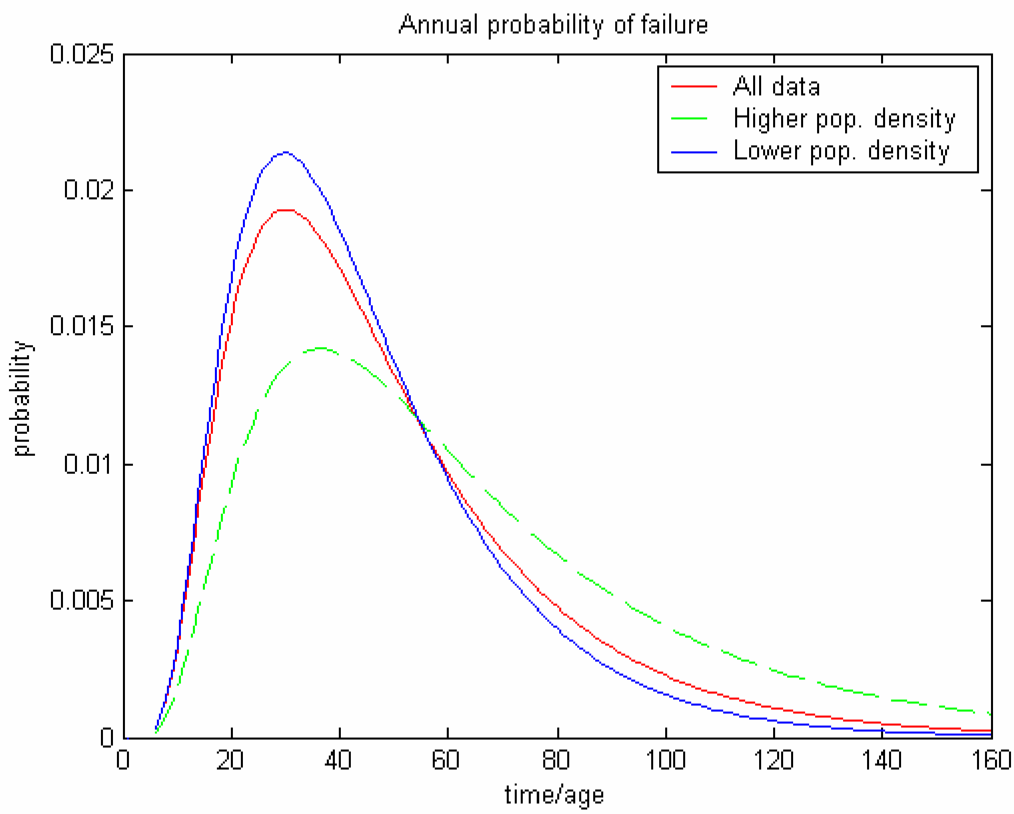

Figure C-2 Annual probability of failure for covariate "Population density" with all data

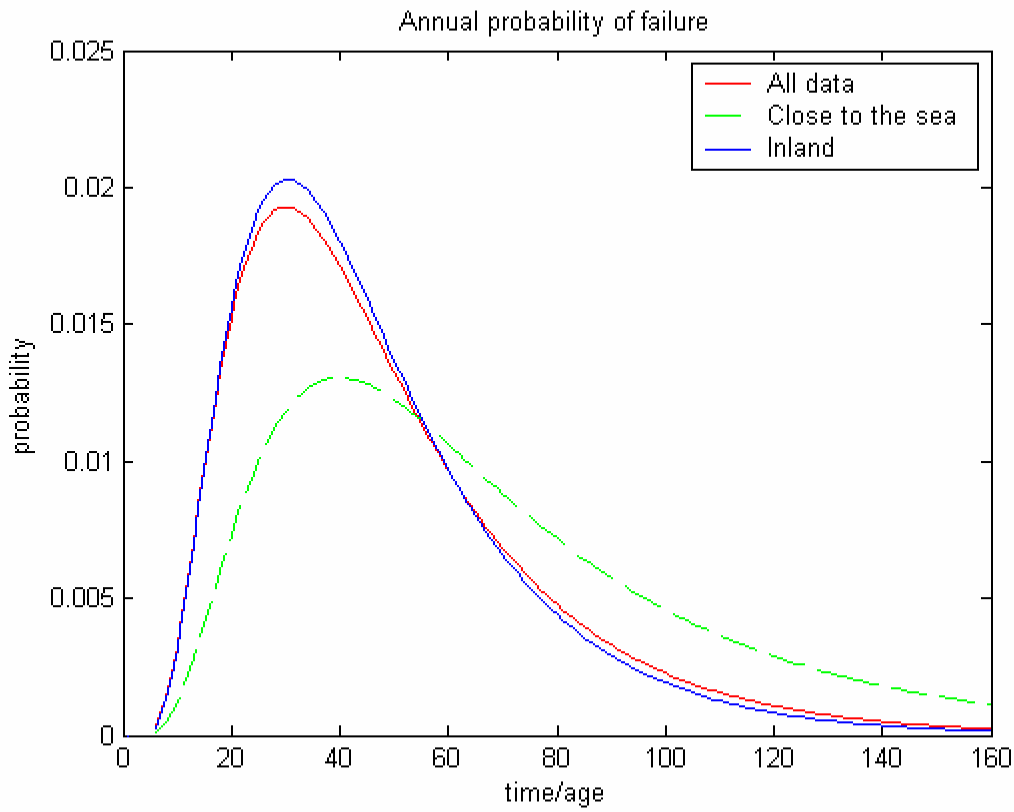

Figure C-3 Annual probability of failure for covariate "Proximity to the sea"

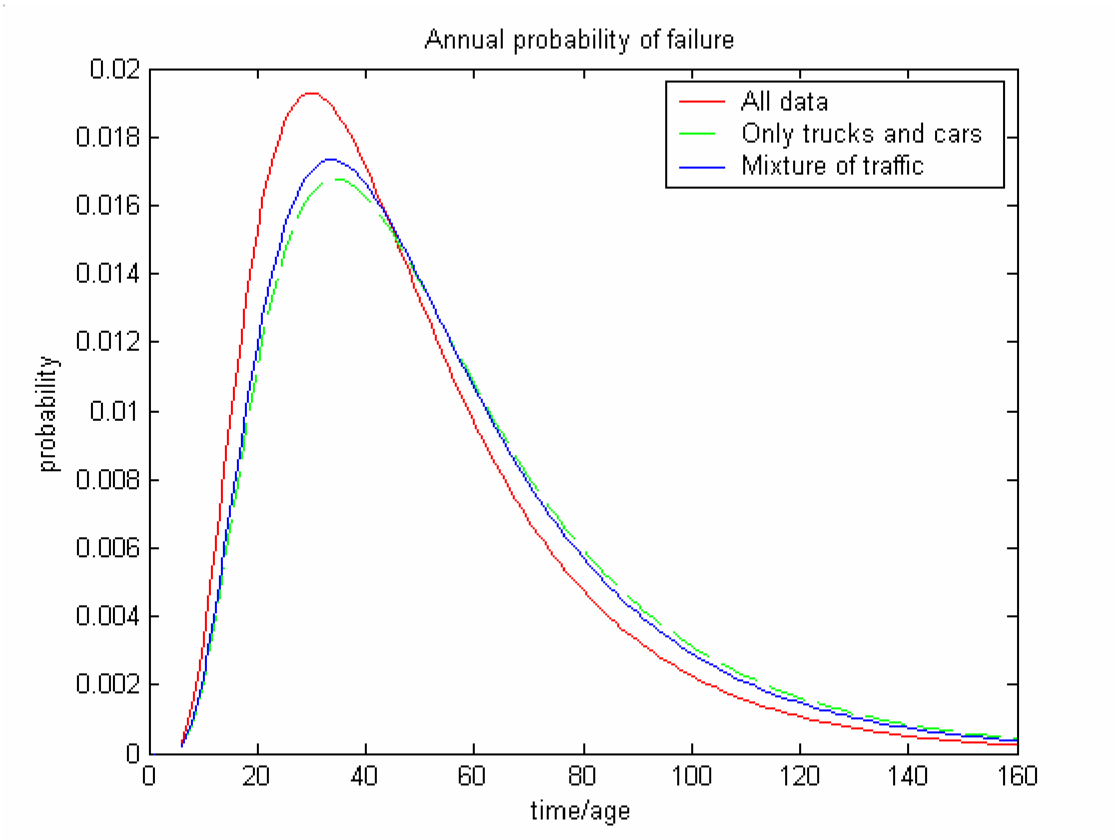

Figure C-4 Annual probability of failure for covariate "Type of traffic"

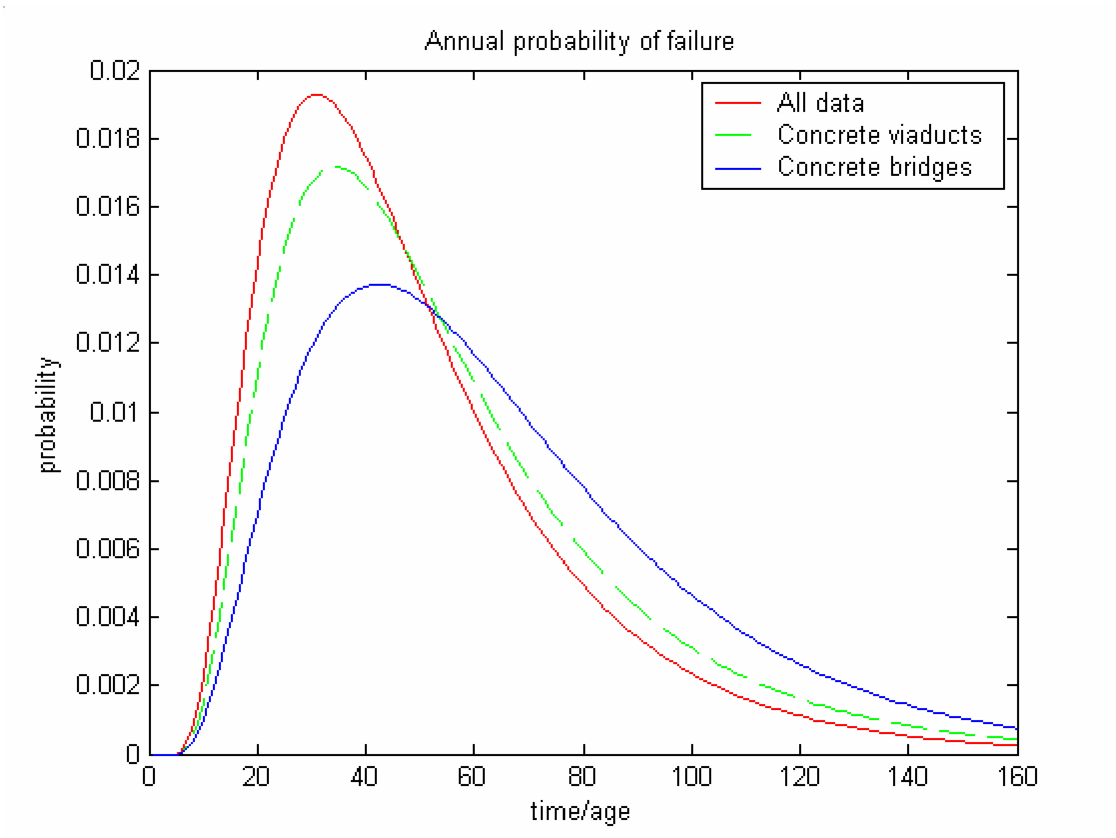

Figure C-5 Annual probability of failure for covariate "Type of use"

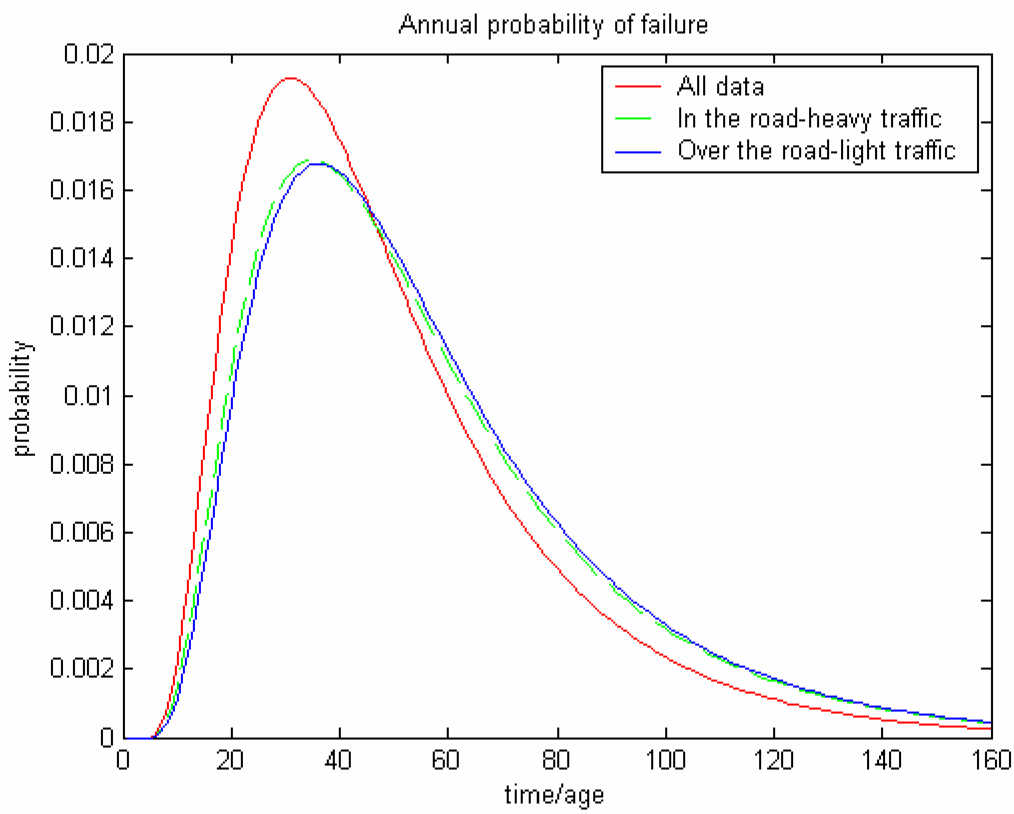

Figure C-6 Annual probability of failure for covariate "Location"

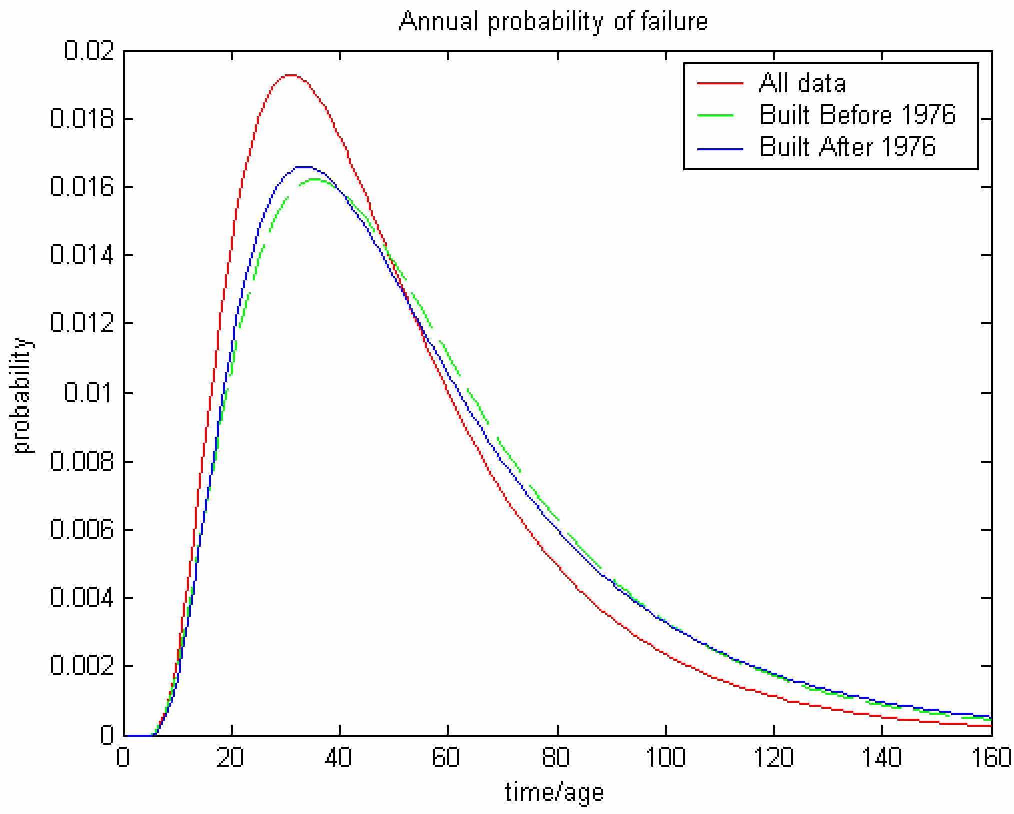

Figure C-7 Annual probability of failure for covariate "Construction year"