ITERATIVE APPROXIMATION OF MARKOV PROCESS PARAMETERS IN A MODEL OF LARGE SCALE BUSINESS PURCHASES

A Thesis presented to

the Faculty of the Graduate School

at the University of Missouri

In Partial Fulfillment

of the Requirements for the Degree

Masters of Applied Mathematics

---

by

Max Highsmith

Dr. Samuel Walsh, Thesis Supervisor

JULY 2020

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled:

ITERATIVE APPROXIMATION OF MARKOV PROCESS PARAMETERS IN MODELS OF LARGE SCALE BUSINESS PURCHASES

presented by Max Highsmith,
a candidate for the degree of Masters of Applied Mathematics and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. Samuel Walsh

Dr. Niladri Syam

Dr. Stephen Montgomery-Smith

ACKNOWLEDGMENTS

I first and foremost acknowledge my thesis adviser, Dr. Samuel Walsh for his guidance and support in allowing me to complete this thesis. I am also grateful to Dr. Niladri Syam and Dr. Stephen Montgomery-Smith for sitting on my thesis committee. I extend my gratitude to the Department of Mathematics at the University of Missouri for their support.

TABLE OF CONTENTS

ACKNOWLEDGMENTS . . . . . **ii**

LIST OF FIGURES . . . . . **iv**

ABSTRACT . . . . . **v**

1 Introduction 1

1.1 Business Purchases 1

2 Background 2

2.1 Markov Models 2

2.2 Maximum Likelihood 5

2.3 Likelihood of Continuous and Discrete Markov Chains 6

2.4 Expectation-Maximization Implementation 8

3 Data and Experiments 11

3.1 Model 11

3.2 Data Set 11

4 Evaluation Of Fit 16

4.1 Conclusion 18

5 Future Work 19

5.1 Future Data 19

5.2 Hidden Markov Models 19

BIBLIOGRAPHY **20**

LIST OF FIGURES

Figure Page

1 small dataset pipeline 12

2 large dataset pipeline 13

ABSTRACT

Large scale business to business purchases often involve a series of vetting stages which must be satisfied prior to their resolution. This paper models the progression of sales projects through a company's sales pipeline via a stochastic process. The pipeline is viewed as a sequential Markov chain wherein each time step permits: failure, no movement, or advancement in a linear ordering of potential states terminating in a successful purchase. The transition probabilities which parameterize this model are determined via time series data of each salesman's progression in selling their products. In theory, the parameters guiding such a Markov process can be estimated analytically via the Maximum Likelihood procedure. However because the state transitions of sales projects can not be provided in real time, the Expectation-Maximization algorithm is applied to estimate these parameters. Our results suggest that more data is needed to generate a reasonable approximation of transition probabilities.

1 Introduction

1.1 Business Purchases

Business-to-business transactions of expensive products are often complex processes. Consequently, many companies use pipelines to represent sales prospects. In such pipelines a sales project is viewed as progressing through a series of states before terminating in either success or failure. In this paper we propose a novel way of modeling such pipelines.

The proposed model is agnostic to the exact business action performed in each of these states, but is contingent on the pipeline involving a progression of sales projects through some number of states, denoted $s_1, ..., s_d$. For the data observed, the progression through these steps is considered to be sequential, meaning that the states can be ordered in a manner such that anytime a sales project changes states, the project can either move forward from state s_n to s_{n+1}, succeed, or fail. A sales project can never move backwards in the pipeline.

For a company, it would be useful to apply existing sales records towards predicting the success or failure of projects in the future. Such knowledge could help in projecting growth or profit as well as guide business strategy. We therefore formulate the following problem: Given the observations of the progression of previous sales projects through a pipeline, can the time-line for future projects be predicted, and, if so, how? To address this question we model the path which a project will take as the actualization of a continuous time stochastic process, wherein each stage of the pipeline is viewed as a possible state of the sales project. The properties of this stochastic process are described by a time dependent transition probability matrix $P(t)$ which is determined by an infinitesimal generator matrix L. Using records of previous sales projects we estimate the component values of L and evaluate the utility of L in retroactively predicting the observed time series data.

In the next section, we state more precisely the definition of these objects, then discuss the details of the model and estimation procedure. In section 3 we apply them to real world data, and in section 4 we perform a statistical analysis of the results. Finally, in section 5, we discuss future directions for this work.

2 Background

2.1 Markov Models

We begin by defining a stochastic process $\{X(t)\}$ as a collection of random variables defined on a common probability space $(\Omega, \mathcal{F}, P)$, where Ω is a sample space, $\mathcal{F}$ is a σ-algebra on Ω, and P is a probability measure. The random variables are indexed by a set $\mathbb{T}$ and all take values in a measurable state space (S, Σ) (see, for example, [1]). A filtration $\{\mathcal{F}_t\}$ is an increasing sequence of σ-algebras on $(\Omega, \mathcal{F})$ where each $t \in \mathbb{R}^+$. $\{X(t)\}$ is said to be *adapted to* $\{\mathcal{F}_t\}$ if $X(t) : \Omega \to \Sigma$ is a $(\mathcal{F}_t, \Sigma)$-measurable function for each $t \in \mathbb{T}$. A stochastic process is said to have the *Markov property* if the conditional probability distribution of future states of a process depends only upon the present state and not on the sequence of events that preceded it, meaning

$$P(X(t) \in A|\mathcal{F}_s) = P(X(t) \in A|X(s)), \quad \text{for all } A \in \Sigma \text{ and for all } s, t \in \mathbb{T} \text{ with } s < t. \tag{1}$$

In the case where Σ is discrete and $\mathbb{T} = \mathbb{N}$ this reduces to

$$P(X(t) = s_n|X(t-1), X(t-2), ..., X(0)) = P(X(t) = s_n|X(t-1)). \tag{2}$$

A Markov process is considered *time-homogeneous* if

$$P(X(t+\tau) = j|X(t) = i) = P(X(\tau)|X(0) = i), \text{ for all } t, \tau \in \mathbb{T} \text{ with } t + \tau \in \mathbb{T} \tag{3}$$

A stochastic process satisfying the Markov property is called a *Markov process*. A Markov process having a countable state space S is referred to as a *Markov chain*. If a Markov chain's index set $\mathbb{T}$ is discrete the Markov chain is a *discrete-time Markov chain* and if a Markov chain's index set $\mathbb{T}$ is continuous then the Markov chain is a *Markov jump-process*. The behavior of a Markov jump-process $X(t)$, $t \in \mathbb{T}$ with state space $S = \{s_1 ... s_d\}$ is identified via a unique time-dependent transition matrix $P = (p_{ij}(t))_{ij}$ such that,

$$p_{ij}(t) = P(X(t) = s_j | X(0) = s_i). \tag{4}$$

In the cases where the limit

$$L = \lim_{t \to 0} \frac{P(t) - I_d}{t} \tag{5}$$

exists, the transition matrix can be expressed as a matrix exponential

$$P(t) = \exp(tL), \tag{6}$$

where $L = (l_{ij})_{ij}$ is referred to as the transition rate matrix or infinitesimal generator matrix. A $d \times d$ matrix L will generate a $d \times d$ transition matrix via (6) if and only if L satisfies the following conditions:

1. all off-diagonal entries are nonnegative,

2. the sum over each row is equal to zero.

We will denote

$$\mathcal{G} = \left\{ L = (l_{ij})_{ij} \in \mathbb{R}^{d \times d} : l_{ij} \geq 0, \quad \forall i \neq j, \quad l_{ii} = -\sum_{j \neq i} l_{ij}, \quad i, j \in \{1, ..., d\} \right\} \tag{7}$$

as the set of all infinitesimal generator matrices for Markov processes with d states.

In practice, the stage of a sales project within a business pipeline can only be observed in discrete intervals. We are therefore concerned with determining the generator matrix L given a finite sampling of observations $Y = \{y_0 = X(t_0), ..., y_n = X(t_n)\}$. This poses a difficulty because, when only given a time series Y, it is not immediately obvious if Y arises from a finite sampling of a jump-process or if Y instead arises from a discrete-time Markov chain that cannot be embedded into a corresponding jump-process. Let the set

$$\mathcal{B} = \left\{ P \in \mathbb{R}^{d \times d} : \exists L \in G, P = \exp(L) \right\} \tag{8}$$

denote the set of transition matrices P that have infinitesimal generators. While it is clear that $\mathcal{B}$ is a subset of all stochastic matrices, the question of what characterizes the elements of $\mathcal{B}$ is a well-explored, unsolved problem in mathematics referred to as the embedding problem; see [2]. An additional difficulty arises in prediction of L, the matrix exponential is not an injective function when L has complex valued eigenvalues, meaning that for some $P \in \mathcal{B}$,

$$P = \exp(L) = \exp(K), \quad L \neq K, \quad L, K \in \mathcal{G}. \tag{9}$$

It is apparent from both of these issues that approximating a generator is a difficult, and sometimes impossible task. Instead, our goal is to determine a matrix L which fits a finite sampling L as well as is possible, in some sense to be made precise shortly. To explore the strategies for predicting L when such an approximation is feasible requires the clarification of some terms used in statistics.

2.2 Maximum Likelihood

We define a *parameter* θ as a mathematical object that influences the output or behavior of another mathematical object q_θ, where θ is viewed as being unchanging. We call the probability space $(\Omega, \mathcal{F}, P)$ an experiment and we call an $\omega \in \Omega$ an *outcome*. Note that in our model we can view the observation of a Markov process over a collection of times $\mathbb{T}$ as an experiment and we can view

$$\omega = \{y_i = X(t_i), t_i \in \mathbb{T}\} \in \Omega \tag{10}$$

as an outcome of that experiment. Here $y_i = X(t_i)$ denotes the sales project was observed to be in state y_i at time t_i. Because our model is observing sales data for processes that have already occurred, ω is already fixed.

Let there be a finite non-negative measure μ on our probability space $(\Omega, \mathcal{F}, P)$. We say that P is *absolutely continuous* with respect to μ, denoted $P \ll \mu$, if $\mu(A) = 0$ implies $P(A) = 0$ for $A \in \mathcal{F}$. By the Radon-Nikodym Theorem if $P \ll \mu$ then there is an $f \in L^1(\Omega, d\mu)$ such that

$$P(A) = \int_A f d\mu, \text{ for all } A \in \mathcal{F}. \tag{11}$$

The function f is referred to as the *Radon-Nikodym derivative* of P with respect to μ; see [3]. In the case where we have a discrete-time Markov process we will use the counting measure for μ and in the case where we have a Markov jump-process we will use the Lebesgue measure for μ. Recall from the discussion of Markov models that $P \in \mathcal{G}$ is generated by a matrix L, which can be viewed as a parameter of P. Furthermore, from (11) we see that the selection of P determines f, thus we can view P, and consequently L, as a parameter of the function f. We now define the *likelihood*

distribution function given ω as

$$\mathcal{L}(L) = f_L(\omega). \tag{12}$$

The goal of the maximum likelihood method is to select L which yields the highest value for $\mathcal{L}(L)$, which we will denote $\hat{L}$. In the case where μ is the Lebesgue measure we have a continuous likelihood distribution function denoted $\mathcal{L}_c(L)$ and when μ is the counting measure, we have a discrete likelihood distribution function $\mathcal{L}_d(L)$.

2.3 Likelihood of Continuous and Discrete Markov Chains

For a Markov jump-process $X(t)$ with state space $S = \{s_1, ..., s_d\}$, and time interval $\mathbb{T} = [0, T]$ we let the random variable $R(T) = (R_i(T))_i$ be the total time a process has spent in state i from time $t = 0$ to time $t = T$, denoted

$$R_i(T) = \int_0^T \chi(X(s) = s_i) ds. \tag{13}$$

We also define the random variable $N = (N_{ij}(T))_{ij}$ to be the number of transitions from state s_i to s_j which have occurred from time $t = 0$ to time $t = T$. Derivations in Dempster, Laird and Rubin [4] show that for a jump-process the continuous time likelihood function $\mathcal{L}_c$ is given by

$$\mathcal{L}_c(L) = \prod_{\substack{i,j=1 \\ i \neq j}}^{d} l_{ij}^{N_{ij}(T)} \exp(-l_{ij} R_i(T)). \tag{14}$$

Because the log function is monotone and we are interested in the value of $\hat{L}$ that maximizes $\mathcal{L}(L)$, not the actual value of $\mathcal{L}(L)$, we can instead solve the far simpler

equation

$$\log(\mathcal{L}_c(L)) = \sum_{\substack{i,j=1 \\ i \neq j}}^{d} [N_{ij}(T)\log(l_{ij}) - l_{ij}R_i(T)]. \tag{15}$$

The value $\hat{L}$ which maximizes $\mathcal{L}_c(L)$ will be where the partial derivatives of $\log(\mathcal{L}_c(L))$ with respect to l_{ij} vanish, thus taking the partials we see

$$\frac{d\log(\mathcal{L}_c(L))}{dL} = 0 \iff \hat{l}_{ij}(t) = \frac{N_{ij}(T)}{R_i(T)}. \tag{16}$$

This means that evaluating $\hat{L}$ reduces to determining our experiment's values for $R(T)$ and $N(T)$. While $\hat{L}$ can be be determined analytically if provided with the values of $R(T)$ and $N(T)$, in practice this is problematic because continuous time information about the progression of a sales project through a pipeline is impossible to obtain. Therefore, we should not expect to have the explicit values of $R(T)$ or $N(T)$. More realistically the progression of such projects will be observed through discrete, incomplete observations made on a daily or weekly basis. We know from Metzner [5], that the evaluation of the likelihood function in the discrete case is

$$\mathcal{L}_d(L) = \prod_{k=0}^{n-1} p_{y_k, y_{k+1}}(\tau), \tag{17}$$

where $\tau_k = t_{k+1} - t_k$ is the time between two consecutive observations. Supposing that the process $X(t)$ has been observed over equidistant points $t_n = n\tau$ and $n = 0, ..., N$. We define the frequency matrix

$$C = (c_{ij})_{ij} \quad c_{ij} = \sum_{n=0}^{N-1} \chi(x(t_n) = i)\chi(X(t_{n+1}) = j). \tag{18}$$

For all data in this paper the intervals between discrete observations remains constant, so we can assume $\tau_k = \tau$ for all k. Thus the discrete likelihood reduces to

$$\mathcal{L}_d(L) = \prod_{i,j=1}^{d} [p_{ij}(\tau)]^{c_{ij}}. \tag{19}$$

However, even with uniformly sized observation intervals taking the derivative of our discrete log likelihood function yields

$$\begin{aligned} &\tfrac{d}{dL} \log(\mathcal{L}_d(L)) = \sum_{n=1}^{\infty} \sum_{k=1}^{n} \frac{\tau^n}{n} (L^T)^{k-1} Z (L^T)^{n-k}, \\ &\text{where } Z = (z_{ij})_{ij} z_{ij} = \tfrac{c_{ij}}{\exp(\tau L)_{ij}}. \end{aligned} \tag{20}$$

Solving for when these partials are all zero is an analytically intractable problem, and consequently this formula cannot be used to determine $\hat{L}$ in the discrete case. We circumvent this issue by assuming that the Markov chain can be modeled via a continuous time Markov process and then use the Expectation-Maximization algorithm to iteratively approximate the random variables $N(T)$ and $R(T)$. By obtaining approximations of $N(T)$ and $R(T)$, we can use (16) to determine $\hat{L}$.

2.4 Expectation-Maximization Implementation

Following Metzner et al. [5], we use an iterative approximation scheme to maximize $\mathcal{L}(L)$. This iterative approximation is broken into an expectation step and a maximization step.

Initialization

To begin the expectation maximization algorithm requires a starting estimate for $\hat{L}$, denoted L_0. To obtain this estimate we create a virtual transition matrix

$$\hat{P}_{\text{virt}} = (\hat{p}_{ij})_{ij}, \quad \hat{p}_{ij} = \frac{c_{ij}}{\sum_{j=1}^{d} c_{ij}}. \tag{21}$$

From this virtual transition matrix we determine the off-diagonal entries,

$$L_0 = (l_{ij})_{ij}, \quad l_{ij} = \frac{log(\hat{P}_{\text{virt}})}{\tau} \quad i \neq j. \tag{22}$$

We set the diagonal entries L_0 to the corresponding negative row sums defined in (7). Once an estimate is made for L_0 we alternate between the expectation and maximization step until our estimate for $\hat{L}$ ceases to change.

Maximization Step

Maximization involves selecting the maximally likely Markov process generator matrix $\hat{L}$ given random variables $R(T)$ and $N(T)$. Because we are evaluating a continuous time likelihood this reduces to evaluating the generator $\hat{L}$ using (16).

Expectation Step

The expectation step is a bit more involved. As outlined by Metzner [5], due to the Markov property and the assumption of a homogeneous jump-process, the expectation

for the parameters $R(T)$ and $N(T)$ can be expressed as the summations

$$E[R_i(t)|Y] = \sum_{k=1}^{d}\sum_{h=1}^{d} c_{kh}E[R_i(\tau)|X(\tau) = s_h, X(0) = s_k],$$
$$E[N_{ij}(t)|Y] = \sum_{k=1}^{d}\sum_{h=1}^{d} c_{kh}E[N_{ij}(\tau)|X(\tau) = s_h, X(0) = s_k], \tag{23}$$

where s_l, $s_k \in S$. The individual conditional expectations on the right can be decomposed by identities from probability theory to yield

$$E[R_i(t)|X(t) = s_l, X(0) = s_k] = \frac{E[R_i(t)\chi(X(t)=s_l)|X(0)=s_k]}{p_{kl}(t)},$$
$$E[N_{ij}(t)|X(t) = s_l, X(0) = s_k] = \frac{E[N_{ij}(t)\chi(X(t)=s_l)|X(0)=s_k]}{p_{kl}(t)}. \tag{24}$$

For computational purposes we consider the numerators on the right as auxiliary functions defined as

$$M^i_{kl}(t) = E[R_i(t)\chi(X(t) = s_l)|X(0) = s_k],$$
$$F^{ij}_{kl}(t) = E(N_{ij}(t)\chi(X(t) = s_l)|X(0) = s_k]. \tag{25}$$

The functions M and F satisfy a system of differential equations.

$$\tfrac{d}{dt}M^i_k(t) = M^i_k(t)L + A^i_k(t), M^i_k(0) = 0, A^i_k(t) = p_{ki}(t)e_i,$$
$$\tfrac{d}{dt}F^{ij}_k(t) = M^i_k(t)L + B^{ij}_k(t), F^{ij}_k(0) = 0, B^{ij}_k(t) = l_{ij}p_{ki}(t)e_j. \tag{26}$$

We solve these systems numerically and apply the approximate solutions to equations (24), (23) to calculate the expected values for $N(T)$ and $R(T)$.

3 Data and Experiments

3.1 Model

We assume that a sales pipeline can be viewed as a time-homogeneous Markov jump process where s_i represents the ith possible stage in the pipeline. The progression of a sales project X will be indexed according to the time $t \in \mathbb{T}$ that has passed since the project began. The statement $X(t) = s_i$ indicates that the sales project was in stage s_i at time t. As discussed in the Section 2, the Markov jump process is guided by probability matrix $P(t) = (p_{ij}(t))_{ij}$, which can be determined by an infinitesimal generator matrix L. The process is assumed to be sequential, meaning that for all $t \in \mathbb{T}$, $p_{ij}(t) = 0$ if $j \leq i$. We view L and $P(t)$ as unknowns to be determined. We only have a finite sampling of the progression of the sales project, thus

$$Y = \{y_0 = X(t_0), ..., y_n = X(t_n)\} \tag{27}$$

represents the recorded stages of our project at the finite set of times $\{t_0, ..., t_n\} \in \mathbb{T}$. Using the finite sampling of observed progression of X through the pipeline we retrospectively predict the maximally likely value for L as described in 2, and consequently $P(t)$. Once these matrices are determined, we can make predictions about the progression of future sales data at given times according to the probabilities of $P(t)$.

3.2 Data Set

The data used in this paper comes from the sales records of two separate products from the same company. They were obtained by Prof. Niladri Syam and provided to the author in an anonymous format due to non-disclosure agreements. The first product follows the simplest pipeline possible with three states: Start, Closure and Failure. Data collection for the first pipeline began on 7/20/2015 and concluded

12/7/2015. A total of 162 projects were observed across four different salesmen. The second product uses a more involved pipeline with six stages: Qualifying Project Bidders, Quoting Bid List, Buying Opportunity, Renegotiating, Closure, and Failure. For the second pipeline the average value of individual transactions is $2 million. Data collection for the second pipeline began on 2/1/2017 and concluded 8/21/2017. A total of 30 projects were observed across nine different salesman. For both pipelines observations are provided daily.

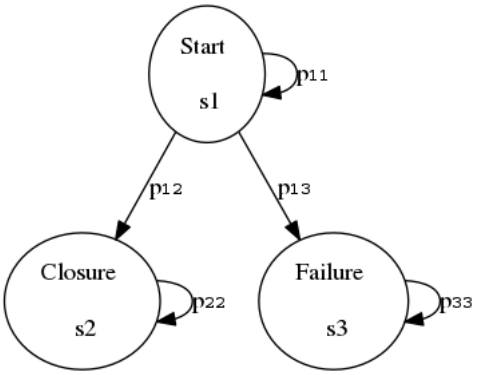

Figure 1: small dataset pipeline

Time Series Concatenation

The approximation method given by Metzner et al. [5] and explained in Chapter 2 has been structured to evaluate parameters of a Markov jump-process given a time-series observation of state changes $Y = \{y_1, ..., y_n\}$. However a sales company is rarely only making one sale at a time and our data regards multiple sales projects observed over the course of different time periods that frequently overlap. Furthermore, due to the sequential nature of both of our pipelines, a sales project ceases to progress once it has reached either the success or failure state. This means that if the parameters of the Markov jump-process are evaluated on any single true time series there will necessarily be insufficient information to categorize the model, since one project cannot transition to both the failure and success state, yet transitions to either is clearly possible. The

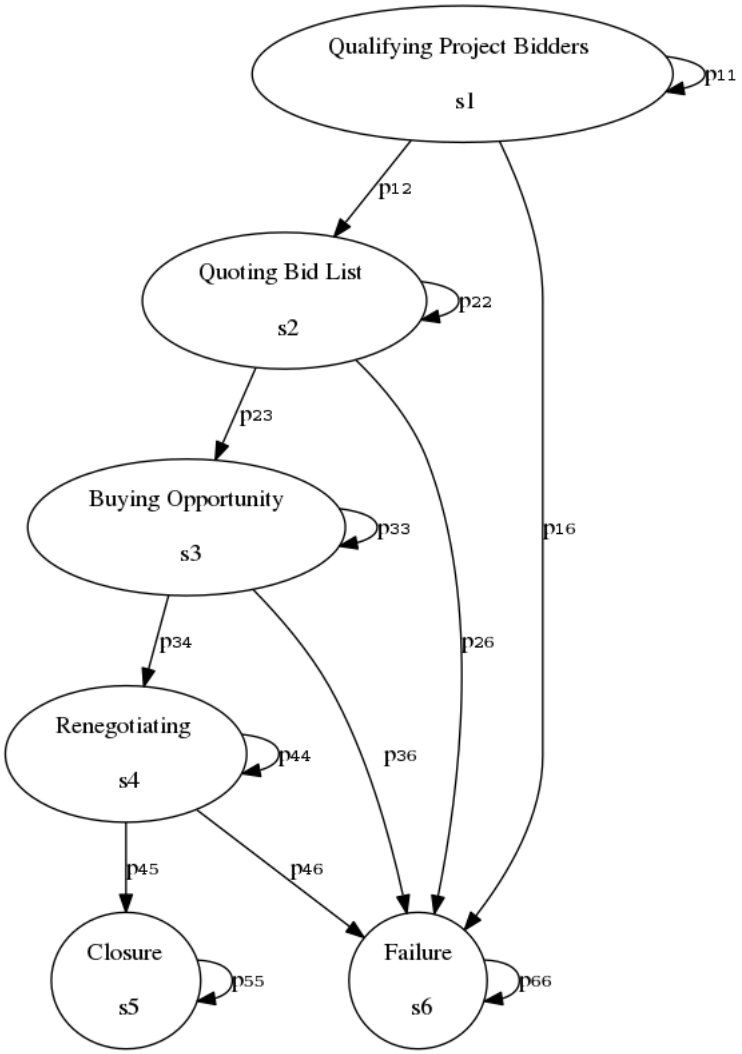

Figure 2: large dataset pipeline

purpose of this model is to make meaningful predictions about the progression of future sales projects using the information provided about the progression of previous sales projects, so ideally the model would utilize all previous sales project data rather than the records of just one project.

To circumvent these issues, we concatenate the time series data for each sales project to form one complete time series, disregarding all transitions from the absorbing states to the start state which would arise from concatenation. These transitions are ignored to prevent noise contaminating the frequency matrix C defined in (18). when performing the Expectation-Maximization algorithm. We also capture the ab-

sorbing nature of the failure and success states by interpreting the sales projects which have reached either absorbing state as remaining in the same state for the duration of the projects observation.

Unusable Equation in ML estimation

Due to the sequential nature of the assumed Markov model, evaluation of equations (24) is not always possible. For example, in the larger dataset, by nature of our model there will never be observed transitions from s_4 to s_2, because Renegotiating occurs later in the pipeline than Quoting Bid List. This means that the probability of transition $p_{4,2} = 0$. Because the purpose of equation (24) is to determine $N_{ij}(T)$ and $R_i(T)$, and we know both of these values to be 0 by nature of the model, we forgo evaluation and just assume the left hand expectation to be zero.

Results

Small Pipeline

1000 iterations of the Expectation Maximization Algorithm over our smaller time series data yields the infinitesimal generator matrix

$$\hat{L} = \begin{pmatrix} -0.18485 & 0.15845 & 0.026409 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

We then use (6) with $t = 1$ to determine the transition probability matrix

$$P(1) = \begin{pmatrix} 0.831229 & 0.144667 & 0.0241118 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Large Pipeline

1000 iterations of the Expectation Maximization Algorithm over our larger time series

data yields the infinitesimal generator matrix

$$\hat{L} = \begin{pmatrix} -0.0416 & 0.0333449 & 0.0069759 & 0.00132047 & 0 & 0 \\ 0 & -0.0293365 & 0.0176403 & 0.00152776 & 0 & 0.0101684 \\ 0 & 0 & -0.029825 & 0.0250268 & 0 & 0.00479815 \\ 0 & 0 & 0 & -0.273357 & 0.0154754 & 0.0118603 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

We then use (6) with $t = 1$ to determine the transition probability matrix

$$P(1) = \begin{pmatrix} 0.959214 & 0.0321832 & 0.0070144 & 0.00138753 & 0.0000105 & 0.00019 \\ 0 & 0.971090 & 0.0171262 & 0.0016995 & 0.0000127 & 0.01007 \\ 0 & 0 & 0.970615 & 0.02432 & 0.00019 & 0.00487 \\ 0 & 0 & 0 & 0.973034 & 0.0152658 & 0.0116997 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Observe that the expectation maximization procedure described in Section 2 provides an iterative method for determining off diagonal entries of the infinitesimal generator matrix, while the diagonals are computed using (7). The final two rows of both the generator matrices are all zeros because success and failure are absorbing states with no probability of transition taking place after they are reached. For the larger dataset entries to the right of the transition matrix diagonal become higher as t increases, corresponding with the intuition that probability of transitioning to later stages in the pipeline goes up as more time passes. Entries to the left of both transition matrices diagonals are all 0 because the system being modeled is a sequential

system and consequently cannot transition backwards in the pipeline.

4 Evaluation Of Fit

In determining the effectiveness of the proposed model, the first question that arises is whether the data is effectively described via a homogeneous Markov process. To address this issue we employ a statistical evaluation of the transition probability matrix by building off of the family of chi squared tests identified in Bickenback; see [6]. Bickenback proposes an evaluation of homogeneity over time by dividing a sample into Z non-overlapping time periods. The authors then evaluate transition matrices for each of the sub-samples, as well as the entire sample. A chi-square test is then used to determine if the transition matrices for each sub-sample differ significantly from a transition matrix arising from consideration of the entire space. If the Markov process is time-homogeneous then the difference should be small. The chi-square statistic Q_Z is determined as follows

$$Q_Z = \sum_{z=1}^{Z} \sum_{i=1}^{N} \sum_{B_j} n_i(z) \frac{p_{ij}(z)}{p_{ij}} \tag{28}$$

Where Z is defined to be the number of time partitions, and N is the number of distinct states in the Markov chain so, $n_i(z) = \sum_j N_{ij}$ and $B_j = \{j : p_{ij} > 0\}$.

This statistic would have degrees of freedom equal to the number of summands in Q_Z except those where $n_i(z) = 0$. The number of degrees of freedom is defined to be

$$g = \sum_i a_i(b_i - 1) - (b_i - 1), \tag{29}$$

where b_i is the number of positive entries in the ith row of the matrix for the entire sample and a_i is the number of sub-samples in which observations for the i-th row are available. While this metric could be applied to our data out of the box by par-

titioning according to time intervals, there is strong reason to suspect it would be ineffective. First, due to the sequential nature of our model an individual observation of a sales project can never move backwards in the pipeline. This means that time intervals which occurred earlier would be more inclined to contain transitions from lower indexed stages and later time intervals would contain transitions from higher indexed stages in our concatenated model. Secondly, our model contains two absorbing states: Failure and Success, which are assumed to indefinitely transition to themselves. Thus later time intervals in the partition would contain a high quantity of observations which will never transition.

The data being modeled via a Markov chain in Bickenback [6] contains a spatial dimension and consequently homogeneity of the spatial dimension is evaluated by partitioning according to space rather than time yielding the test statistic

$$Q_R = \sum_{r=1}^{R} \sum_{i=1}^{N} \sum_{B_j} n_i(r) \frac{p_{ij}(r)}{p_{ij}} \tag{30}$$

Where R is the spatial partition and Q_R has degrees of freedom equal to the number of summands in Q_R except those where $n_i(t) = 0$. While our data does not contain a spatial component to partition across, we can partition across the collection of sales projects. Consequently we divide our set of sales projects into groups and perform the chi-squared test to evaluate if homogeneity is preserved across the different groups of sales projects. It is worth noting that all of these chi-squared tests are fundamentally comparing multinomial distributions in the form of transition row matrices not Markov processes. This means that the tests, acting under the assumption that sub-samples of the data follow Markov process, evaluate the probability that each sub-sample contains the same underlying Markov process. While this does not explicitly identify if the transitions are, in fact, Markovian, it is an effective metric because if homogeneity is not preserved across groups of sales projects then the tran-

sition matrix has low predictive power, and thus no utility to our original problem formulation.

We test the null hypothesis $H_0 : p_{ij}(t) = p_{ij}$ against the alternative $H_a : p_{ij}(t) \neq p_{ij}$ and test statistic

$$Q_S = \sum_{s=1}^{S} \sum_{i=1}^{N} \sum_{B_j} n_i(s) \frac{p_{ij}(s)}{p_{ij}}. \tag{31}$$

Here s refers to a group of sales projects.

We partition our small pipeline data set into $S = 4$ groups, according to the 4 salesmen who presented the data. Using the transition matrix at our most refined granularity of one day we get a chi squared score of 137.6 with degrees of freedom 6. This grants us a p-value $p < 0.00001$ which is sufficient to reject the null hypothesis, leaving us with the hypothesis that the subsamples do not arise from the same Markov process.

We partition our large data set into $S = 3$ groups. Using the transition matrix at our most refined granularity of one day we get a chi squared score of 2254.5 with degrees of freedom = 28. This grants us a p-value $p < 0.00001$ which is sufficient to reject the null hypothesis, leaving us with the hypothesis that the subsamples do not arise from the same Markov process.

4.1 Conclusion

Based on these results our model demonstrates little utility in predicting the progression of sales projects. It may be the case that the behavior of this companies business pipeline is not guided by a Markov jump process. It is also possible that the small amount of data available, relative to the large space of components to be estimated in $\hat{L}$ prediction, is insufficient to make accurate approximations. In future work, consideration of alternative methods for evaluating the infinitesimal generator matrix such as quadratic optimization or the resolvent method may provide improved

approximations. Experiments on larger datasets may also provide more conclusive results.

5 Future Work

5.1 Future Data

Data used in this paper is provided through a third party entity that is still actively selling their product, consequently sales data is still being collected. If our models inability to make generalizable predictions is due to the limited amount of data currently available, future data may provide more meaningful insight. It may also reveal that it is important, for example, to separate out individual salespeople, or to incorporate seasonal effects.

5.2 Hidden Markov Models

The data provided also provided rankings by sales staff identifying different criteria potential buyers satisfied relevant to each stage in the pipeline. It may be possible to construct a hidden Markov model wherein the scores at these stages are treated as observation distributions.

Bibliography

[1] John Lamperti. Stochastic processes: as survey of the mathematical theory. *Springer-Verlag*, 427:1–2, 1977.

[2] D.T. Crommelin and E. Vanden-Eijnden. Fitting timeseries by continuous-time markov chains: A quadratic programming approach. *Journal of Computational Physics*, 217:782–805, 2006.

[3] Leonid B. Koralov and Yakov G. Sinai. *Theory of Probability and Random Processes*, volume 1. Springer, Heidelberg, Germany, second edition, 2012.

[4] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *J.R. Statist*, 39:1–38, 1977.

[5] P. Metzner, E. Dittmer, T. Jahnke, and Ch. Schutte. Generator estimation of markov jump processes. *Journal of Computational Physics*, 227:353–375, 2007.

[6] F. Bickenbach and B. Eckhardt. Markov or not markov- this should be a question. Kiel Working Paper No. 1086, 2001.