AMSI
AUSTRALIAN MATHEMATICAL SCIENCES INSTITUTE

National Collaboration in the Mathematical Sciences

Operations Research Modelling for the Comparative Analysis of the Emergency Department Length of Stay between Stroke and Cardiac Patients

Hayden Johns

February 28, 2014

1 Abstract

Emergency Departments form a critical part of care for victims of stroke and heart attack, and should be made as efficient as possible. Many statistical methods, including the use of Coxian Phase-Type distributions, can be used to help improve the quality of service provided by hospitals. This project investigates the suitability of Coxian Phase-Type distributions to model Emergency Department length of stay. The results from analysis are mixed, with the fitted distribution being a poor fit to the data, while the characteristics of the model are plausible at capturing Emergency Department processes. Further work is needed to investigate this.

2 Introduction

Stroke and heart attack are both medical emergencies, and it is critical that patients are seen and treated as soon as possible [1] [2]. Therefore, it is important to make Emergency Departments (ED) as efficient and streamlined as possible.

One way of examining ED is through patient *Length of Stay (LoS)*, the time from patient arrival to when they leave. There are a variety of methods of analysing this type of data [3] , one of which is through the use of *Coxian Phase-Type Distributions*

2.1 Coxian Phase-Type Distributions

Coxian Phase-Type Distributions model the time to an event happening (for example, a patient leaving the hospital) as the time it takes a patient to move through n sequential "phases" to reach an *absorption phase* (phase $n+1$). An absorption phase is a phase where they no longer have a chance of moving to any other phase. Depending on the system being modelled, this could correspond to reaching the end of a queue, machine malfunction, or leaving an Emergency Department.

All patients begin in phase 1, and at any point have a chance of moving to the next phase in the sequence, or transitioning directly to the absorption phase. The chance of moving from phase i to phase $i+1$ is described by a parameter λ_i, and the chance of moving directly from phase i to phase $n+1$ is described by a parameter μ_i. In general, the system is described through two vectors, $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)$ and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_{n-1})$. $\lambda_n \equiv 0$ for all n, and is not explicitly referred to.

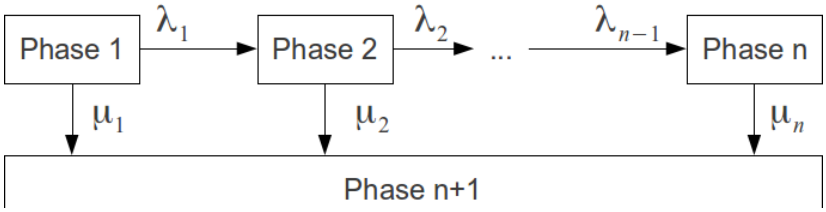

Figure 1: Coxian Phase-Type model

The number of phases n is dependent on the system being modelled. The value of n is conventionally chosen as the minimum number that satisfactorily describes the system to prevent over-fitting and unnecessary complexity. [4]

The Probability Density Function for the time to reach phase $n+1$ is given by the formula $f(t)$:

$$f(t) = \mathbf{p} \exp\{\mathbf{Q}t\} \mathbf{q} \tag{1}$$

where

$$\mathbf{Q} = \begin{bmatrix} -(\lambda_1 + \mu_1) & \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & -(\lambda_2 + \mu_2) & \lambda_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -(\lambda_{n-1} + \mu_{n-1}) & \lambda_{n-1} \\ 0 & 0 & 0 & \cdots & 0 & -\mu_n \end{bmatrix}$$

$$\mathbf{p} = (100\ldots00) \qquad \mathbf{q} = \mathbf{Q1} = \boldsymbol{\mu}^T = (\mu_1\mu_1\ldots\mu_n)^T$$

Coxian Phase Type models are used in statistical analysis in a wide variety of fields, ranging from health care [5] to investigating the length of a student's university enrolment [6] . It is assumed that the number of phases n and the values of $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ are meaningful and reflect some underlying, hidden structure, either reflecting patient characteristics [4] or some standard process [6]. However, there is no guarantee that this assumption is correct.

2.2 Emergency Department structure

Routinely collected Emergency Department records provide a great level of information for statistical analysis. There exists a well defined structure that all patients that enter Emergency Departments go through, first being attended by the triage nurse who determines the urgency of treatment, then being admitted to a bed, and so on.

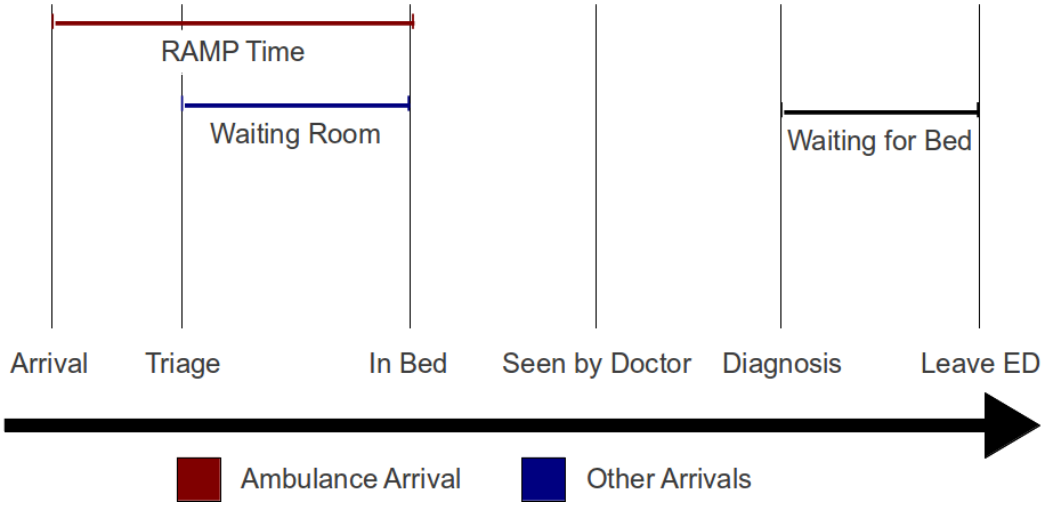

Figure 2: Emergency Department Structure[1].

Emergency Department records include:

- Timestamps for each stage in Emergency Department procedure
- Method of arrival and circumstances of leaving
- Demographic information, including:
  - Age
  - Primary language (and if interpreter was required)
  - Sex
  - Religion
- Triage nurse notes and triage category
- Diagnosis and treatment

This level of structure provides an excellent framework with which to validate the assumption that the parameters related to Coxian Phase-Type distributions are valid.

[1]Points of interest: Waiting Room time (blue line) is a common complaint amongst patients. Two major problems with ED efficiency are RAMP time (red line) where an ambulance waits to offload a patient, and delays in discharge (black line) where a patient must wait for a bed in the main hospital to become available. These are both challenging and important areas of study in and of themselves.

Coxian Phase-Type distributions can be used to model patient Length of Stay data, and the results of this analysis can be compared with patient records and the explicit structure of ED departments to examine the accuracy and interpretability of the Coxian Phase-Type model.

3 Method

This project aims to fit Coxian Phase Type models to real Emergency Department data, and compare results with the expected structure of Emergency Department data. Three years of Emergency Department records were available for examination, from thee metropolitan hospitals based around Melbourne. The records are separated into two data sets with 48590 and 51690 entries, containing all patient entries related to neurological and cardiac problems respectively. In both cases, both patients presenting with common symptoms of neurological or cardiac issues are included, as are people who presented with unrelated symptoms who were diagnosed with a neurological or cardiac issue.

Once a Coxian Phase-Type model has been fitted to the data, the group of patients S_k reaching absorption from phase k can be calculated as described by Marshall and McClean [4]:

$$S_k = \left\{ t_j : m \sum_{i=1}^{k-1} \pi_i < j \leq m \sum_{i=1}^{k} \pi_i \right\} \tag{2}$$

Where m is the number of data points and $t_1, \ldots, t_m$ is the ordered (ascending) Length of Stay time, and π_i is the probability of reaching the absorption phase from phase i, given by:

$$\pi_i = \left(\frac{\mu_i}{\mu_i + \lambda_i} \right) \prod_{j=1}^{i-1} \left(\frac{\lambda_j}{\mu_j + \lambda_j} \right) \tag{3}$$

These patient subsets can be examined for commonalities. If the patients followed a similar process, it provides support for the assumption that the number of phases and the parameter values reflect some underlying structure.

3.1 Data Selection

Due to the general nature of the data, for the purposes of this project it was necessary to select a subset of patients. It was possible to draw on the experience of Dr. Ian

Mosely to help select a subset of the neurological data. Ultimately, only patients that presented with neurological issues and were diagnosed with a "Brain Attack" (complementing the idea of a Heart Attack) were selected, where a Brain Attack was defined as one of the following:

- Stroke and stroke-type diagnosis
  - Cerebral Infarction (ischaemic stroke)
  - Intracerebral Haemorrhage (haemorrhagic stroke)
  - Stroke (cause unknown)
  - Transient Ischaemic Attack (mini-stroke)
- Bleeding within skull
  - Intracranial Haemmorhage
  - Subdural Haematoma
  - Subarachnoid Haemmorhage
- Intracranial Space-occupying lesion (tumor inside skull, e.g. brain cancer)

Unfortunately, there was no access to a clinician with any significant experience with cardiac diseases, so the cardiac data set was not examined in close detail. The subset of the neurological data set became the focus of investigation.

3.2 Fitting Coxian Phase-Type Distribution to data

To fit a Coxian Phase-Type distribution to the data, a *Maximum Likelihood Estimate* was undertaken. A Maximum Likelihood Estimate is a method of calculating parameters for a model. It chooses the parameters that maximise the probability of getting the observed result. Marshall and McClean performed this calculation for a given number of phases n my selecting $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ to minimise, data $\mathbf{t}$, the function $\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\lambda}; \mathbf{t})$ [4]:

$$\mathcal{L}\left(\mu_1, \mu_2, \ldots, \mu_n, \lambda_1, \lambda_2, \ldots, \lambda_{n-1}; \mathbf{t}\right) = -\sum_{i=1}^{m} \log\left[\mathbf{p} \exp\left\{\mathbf{Q} t_i\right\} \mathbf{q}\right] \tag{4}$$

This is known as the *(negative) log-likelihood function.*

Where m is the number of data entries, $\mathbf{t} = (t_1, t_2, \ldots, t_m)$ and $\mathbf{p}$,$\mathbf{q}$ and $\mathbf{Q}$ are defined as in equation 1. To select the number of phases, n begins at 1 and increments,

fitting an n phase model at each point, and fitting halts when there is no substantial decrease in the value of $\mathcal{L}$ (in this project defined as being a decrease of less than $\epsilon = 0.05\%$). The second-to-last model is selected, as the final model showed little to no improvement to justify the additional complexity.

A Maximum Likelihood Estimate allows the fitting of an n-phase model to be treated as a $2n-1$ dimensional optimisation problem, which can be solved in a number of programs, including MATLAB. This process was automated in a MATLAB script, using the fmincon() function and MATLAB's in-built interior-point algorithm to perform the Maximum Likelihood Estimate. See appendix B for the MATLAB code used.

4 Results

In addition to the final results, the implementation process produced noteworthy challenges worthy of further investigation.

4.1 Local Minimum and Maximum values in Maximum Likelihood Estimate

The function $\mathcal{L}$ suffers from local minimum and maximum values, meaning that the found solution was heavily dependent on the chosen starting point for the search. This was overcome by running from multiple random starting locations and comparing the found results from each iteration.

However, it is possible for the minimum found value to occur in multiple places. In this case, two different sets of values for $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ can produce identical distributions:

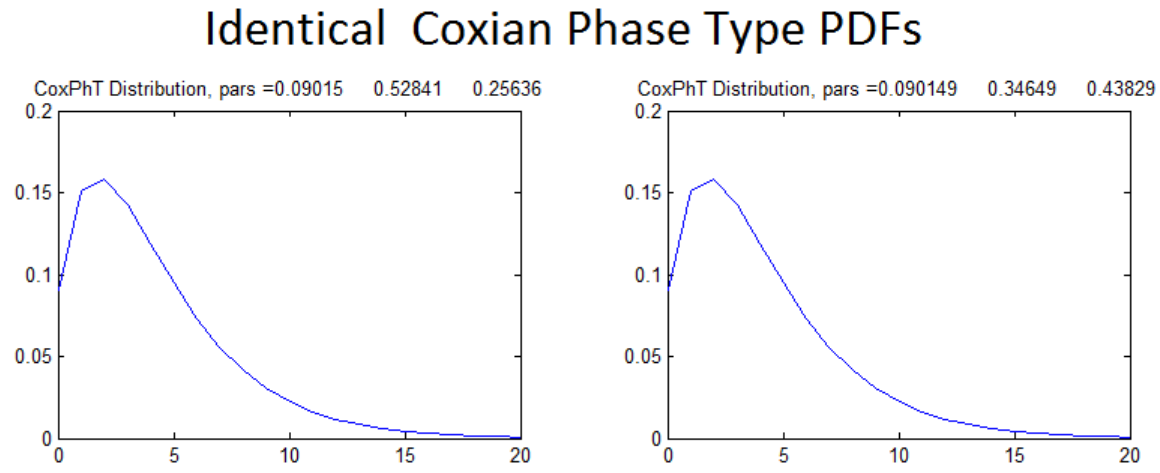

Figure 3: Identical distributions, with $\mu_1 = 0.09$, $\mu_2 = 0.53$, $\lambda_1 = 0.26$ and $\mu_1 = 0.09$, $\mu_2 = 0.34$, $\lambda_1 = 0.43$

While the non-uniqueness of optimal parameter choices does not impact the final distribution, the existence of multiple, equally viable models raises concerns regarding the appropriateness of interpreting these models. Further investigation is warranted to determine the impact that this has on model interpretation.

4.2 Fitting to data

The method for fitting Coxian Phase-Type models to the chosen data found the system was adequately described as a four-phase system (All fits can be found in appendix A), with parameter values $\mu_1 = 1.81 \times 10^{-11}$, $\mu_2 = 0.000189$, $\mu_3 = 2.69 \times 10^{-11}$, $\mu_4 = 0.00374$, $\lambda_1 = 0.032758$, $\lambda_2 = 0.032567$, $\lambda_3 = 0.00374$:

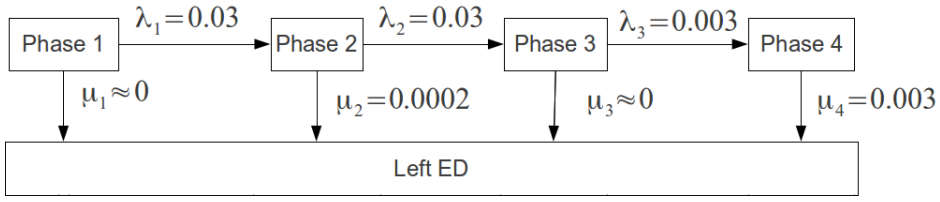

Figure 4: Fitted Emergency Department model

Notably, in phases 1 and 3 the rate of transition to the absorption phase is effectively zero. As such the system appears to have two "active" phases where patients can leave,

and two "holding" phases that serve to describe processes that take time but do not directly lead to a patient leaving Emergency Department (for example, admittance to a bed but waiting to be seen by a doctor). However, the fitted distribution is a poor visual fit against the data:

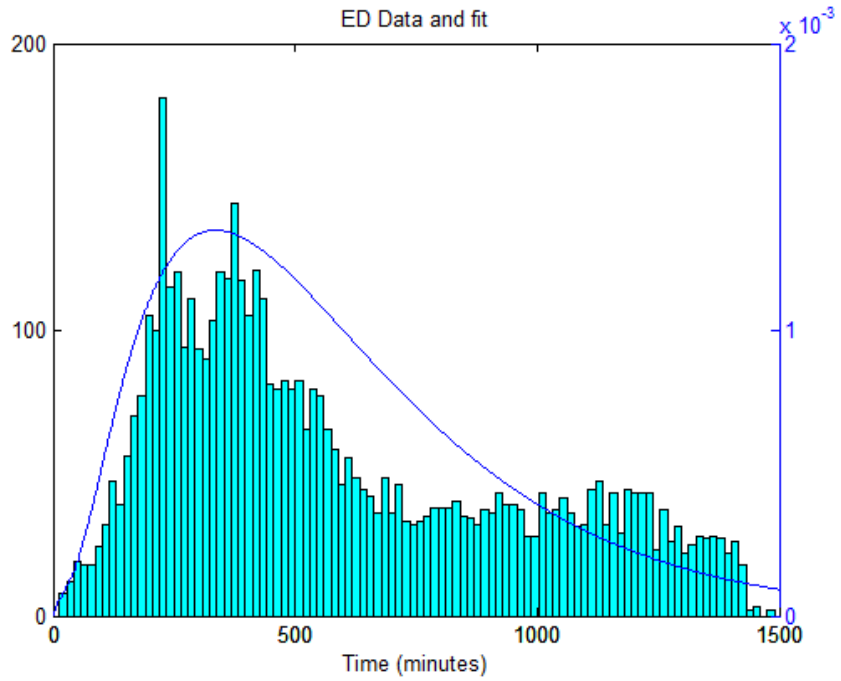

Figure 5: Coxian Phase-Type distribution against Length of Stay data

This fit could potentially be improved with a higher phase model, which would allow greater flexibility in capturing the shape of the data. While the Coxian Phase-Type distribution found is a poor fit, parameter values seem plausible in describing the explicit structure of Emergency Department: The first event in Emergency Department process is when a patient is seen the triage nurse, and are sent elsewhere (other hospital departments, a general practitioner or home), or they enter the main ED process, resulting in additional time spent before departure. This was supported by Emergency Department records, which indicated that patients who left ED in phase 2 were, in the case of early symptom onset, rushed into intensive care, operating rooms or sent to another hospital. Patients with later onset of symptoms corresponded to referrals and transfers from other hospitals, and as such were sped through Emergency. Patients exiting in phase 4 corresponded to all other patients, and made up the body of ED admissions.

5 Future works

Further investigation is needed to explore and confirm these results, and this project will be continued as an Honours research project. Of concern is improving the distribution fit, comparing results with other methods of analysis, and comparing results for cardiac and neurological patients. Ultimately, this project aims to provide useful clinical information to medical practitioners to assist them in delivering a better quality of service. To assist with this, the National Stroke Foundation has awarded an honours research grant.

6 Conclusion

This project fitted Coxian Phase-Type distributions to real Emergency Department data to examine the assumptions made in model interpretation. While the distribution was a poor visual fit, the number of phases and the patients associated with exiting at each phase reasonably reflects the explicit structure of Emergency Departments. Further work is needed to explore the validity of use for Coxian Phase Type distributions, and this project will become the basis for an Honours research project

7 Acknowledgments

This project was co-supervised by Dr. Babak Abassi and Professor Leonid Churilov. This project also heavily drew from the experience of Dr. Ian Mosley and Professor Adele Marshall, both of whom provided practical advice that was critical to this project.

References

[1] A. Meretoja, K. Mahsa, J. Saver, T. Tatlisumak, M. Parsons, K. Markku, S. Davis, G. Donnan, and L. Churilov, "Stroke thrombolysis. save a minute - save a day.." To be published in *Stroke*, 2013.

[2] G. De Luca, H. Suryapranata, J. P. Ottervanger, and E. M. Antman, "Time delay to treatment and mortality in primary angioplasty for acute myocardial infarction: Every minute of delay counts," *Circulation*, vol. 109, no. 1225, p. 1223, 2004.

[3] A. Marshall, "Length of stay-based patient flow models: Recent developments and future directions," *Health Care Management Science*, vol. 8, pp. 213–220, 2005.

[4] A. H. Marshall and S. I. McClean, "Using coxian phase-type distributions to identify patient characteristics for duration of stay in hospital," *Health Care Management Science*, vol. 7, no. 4, pp. 285–289, 2004.

[5] A. H. Marshall, B. Shaw, and S. I. McClean, "Estimating the costs for a group of geriatric patients using the coxian phase-type distribution," *Statistics in medicine*, vol. 26, no. 13, pp. 2716–2729, 2007.

[6] A. H. Marshall, M. Zenga, and S. Giordano, "Modelling students' length of stay at university using coxian phase-type distributions.," *International Journal of Statistics & Probability*, vol. 2, no. 1, 2013.

Appendix

A Table of parameter estimates for Emergency Department Length of Stay

There was no significant improvement in $\mathcal{L}$ for 5 phases, so the 4 phase model is selected.

n	$\mathcal{L}$	Parameter Values
1	38119	$\mu_1 = 0.0017$
2	37224	$\mu_1 = 7.45 \times 10^{-14}$, $\mu_2 = 0.0034$
		$\lambda_1 = 0.0034$
3	37152	$\mu_1 = 1.75 \times 10^{-5}$, $\mu_2 = 6.22 \times 10^{-11}$
		$\mu_3 = 0.0037$
		$\lambda_1 = 0.004$, $\lambda_2 = 0.0169$
4	37145	$\mu_1 = 1.81 \times 10^{-11}$, $\mu_2 = 0.0002$
		$\mu_3 = 2.69 \times 10^{-11}$, $\mu_4 = 0.004$
		$\lambda_1 = 0.0328$, $\lambda_2 = 0.0326$
		$\lambda_3 = 0.0037$
5	37142	$\mu_1 = 6.37 \times 10^{-11}$, $\mu_2 = 0.0001$
		$\mu_3 = 0.00022$, $\mu_4 = 1.8 \times 10^{-10}$
		$\mu_5 = 0.0037$
		$\lambda_1 = 0.0477$, $\lambda_2 = 0.0475$
		$\lambda_3 = 0.0474$, $\lambda_4 = 0.0037$

B MATLAB code

B.1 Parameter fitting script

```
clc;
clear;

% Coxian Phase Type Paramater Fitting Script

%% Load data
```

```
t=csvread('input.csv',1,0);
t=round(t(1:size(t,1),1));


%% Actual routime


options.MaxFunEvals=1000;
options.GradObj='off';
options.UseParallel='always';
options.Display='none';
%options.FunValCheck='on';
options.Algorithm = 'interior-point';


results=[];

%continuing from before
results=csvread('5phaserun.csv');
results=padarray(results,[0,2],0,'post');

subplot(1,2,1);
hist(t,100);
xlim([0,3000]);
title('Generated data')

epsilon=5e-4; %improvement to stop at.

i=0;
Lnew=1e10; %Initial (-)log likelihood value, can be arbitrarily large
improvement=1;

'fiting paramaters'

%% Fit paramaters to data

while improvement>epsilon
```

```
%while i<100 alternatively, specify phases to run to
    i=i+1;
    j=0;
    parameters=[];
    while j<30
        fprintf('Phase %i Run %i\n',i,j+1);
        startpoint=0.5*rand(1,2*i-1);
        disp('Starting at');
        disp(startpoint);
        try
            [par,L,exitflag]=fmincon(@(Par) coxphlikelihood(Par,t),...
                                startpoint,[],[],[],[],zeros(1,2*i-1), ...
                                ones(1,2*i-1),[],options);
        catch
            exitflag=-99;
        end
        if exitflag>0
            j=j+1;
            parameters=[parameters;[L,par]];
            disp('Found solution at:');
            disp(par);
        else
            disp('failed to converge');
        end

    end
    if i>1
        results=padarray(results,[0,2],0,'post');
    end
    results=[results;parameters];

    Lold=Lnew;
    Lnew=min(parameters(:,1));
    improvement=(Lold-Lnew)/Lold ;

    fprintf('Min (-)Log Likelihood found: %i, improvement of %f %',Lnew,improvement*
```

```
end
```

B.2 Likelihood objective function

```
function L=coxphlikelihood(Par,t)
        L=-sum(log(coxph(Par,t)));
end
```

B.3 Coxian Phase-Type distribution calculation

```
function f=coxph(Par,t)

    %Par = [M(1), M(2), ... M(n), L(1), L(2), ..., L(n-1)]

    %t is either a scalar value or a vector. Will return either scalar or
    %vector of results

    mu=Par(1:ceil(length(Par)/2));
    lambda=[Par(ceil(length(Par)/2)+1:length(Par)),0];

    j=length(mu);

    p=[1,zeros(1,j-1)];

    T=diag(-(lambda+mu))+...
      [zeros(j-1,1),diag(lambda(1:j-1));zeros(1,j)];

    for i=1:length(t)
        f(i)=max([p*expm(T*t(i))*transpose(mu),0]);
    end
end
```