Honours Topics

Supervisor: Małgorzata O'Reilly

Any pure research or applied topic of your interest in Probability theory, Stochastic/Statistical Modelling, Markov Chains, Simulation, or other areas of Operations Research. If you are interested in analysis involving uncertainty, or constructing models for unpredictable real-life systems, or inventing algorithms and coding them, I can find an interesting project for you. Please see the examples below.

Phase-Type Distributions and their Applications

(currently available)

Summary: First, recall the exponential distribution definition. We say that a random variable X follows *exponential* distribution with (rate) parameter $\lambda > 0$, and write $X \sim \text{Exp}(\lambda)$, when its density function $f(t)$ is

$$f(t) = \begin{cases} \lambda e^{\lambda t} & \text{when } t \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Recall that the cumulative distribution function $F(t)$, $f(t) = F^{'}(t)$, is then given by

$$F(t) = P(X \leq t) = \begin{cases} 1 - e^{\lambda t} & \text{when } t \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Phase-Type (PH) distribution is a generalization of the exponential distribution, which is defined as follows. Let $\{J(t)\}$ be an *absorbing* continuous-time Markov Chain with state space $\mathcal{S} = \{0, 1, \ldots, m\}$, where 0 denotes an absorbing state, and rates matrix $\mathbf{Q} = [q_{ij}]$ partitioned as

$$\mathbf{Q} = \begin{bmatrix} 0 & \mathbf{0} \\ \underline{\mathbf{t}} & \mathbf{T} \end{bmatrix},$$

where the (square) matrix $\mathbf{T} = [q_{ij}]_{i,j \in \mathcal{S} \setminus \{0\}}$ records the transition rates between the nonabsorbing states, and the (column) vector $\underline{\mathbf{t}} = [q_{i,0}]_{i \in \mathcal{S} \setminus \{0\}}$ records the transition rates from nonabsorbing states to the absorbing state 0. Also, let $\underline{\boldsymbol{\alpha}}$ be the initial distribution (row) vector $\underline{\boldsymbol{\alpha}} = [\alpha_i]_{i \in \mathcal{S}}$.

We say that a random variable X follows **Phase-Type (PH)** distribution with parameters $\underline{\boldsymbol{\alpha}}$, $\mathbf{T}$, and $\underline{\mathbf{t}}$, and write $X \sim \text{PH}(\underline{\boldsymbol{\alpha}}, \mathbf{T}, \mathbf{t})$, when its density function $f(t)$ is given by

$$f(t) = \begin{cases} \underline{\boldsymbol{\alpha}} e^{\mathbf{T}t} \underline{\mathbf{t}} & \text{when } t \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

The cumulative distribution function is then given by

$$F(t) = P(X \leq t) = \begin{cases} 1 - \underline{\boldsymbol{\alpha}} e^{\mathbf{T}t} \underline{\mathbf{1}} & \text{when } t \geq 0 \\ 0 & \text{otherwise,} \end{cases}$$

where $\underline{\mathbf{1}}$ denotes a column vector of ones of appropriate size.

As an example of application, if X records the duration of *lifetime* (eg of a gene), then we interpret $f(t)$ as the density of the lifetime ending at the precise time point t, $F(t)$ as the probability of it ending before time t, and $(1 - F(t))$ as the probability of surviving at least t units of time.

PH distribution has various interesting applications and connections to a wide range of models and problems. As example, it can be used to approximate any positive-valued distribution. The beauty of the PH distribution is that various complex problems (eg general sum of exponentially distributed random variables) can be solved in a

simple and elegant way. This project will focus on the review of the literature in the theory of PH distribution and its applications, and will include some numerical work.

Reference:

[1] G. Latouche, V. Ramaswami. Introduction to Matrix Analytic Methods in Stochastic Modelling, 1st edition. Chapter 2: PH Distributions; ASA SIAM, 1999.

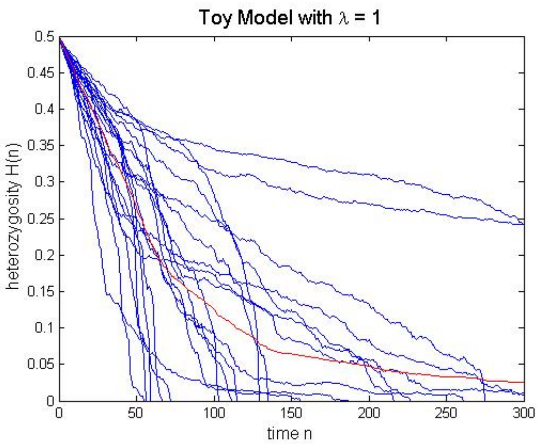
Stochastic Models for the Conservation of Endangered Species.

(currently available)

Co-supervisors:
Dr Sophie Hautphenne, Melbourne University
A/Prof. Barbara Holland

Summary: In populations of endangered species, management strategies referred to as *genetic rescue* have been advocated in order to help avoid extinction. An example of considerable concern in the Australian context is the conservation management of Tasmanian Devils suffering from the *Devil Facial Tumour Disease* (DFTD), which puts them in danger of extinction. An important factor in this context is the ability to assess the impact of conservation efforts. Conservation strategies have been used with the hope of increasing the genetic diversity of the wild population, but this remains a challenging problem. This project will focus on models for the numerical assessment of conservation strategies, which will assist in these efforts. This project will involve reviewing the literature in the area and numerical experimentation using simulation models.

You will have an opportunity to be part of a rich collaborative environment and interact with mathematicians and biologists studying the wild populations of Tasmanian Devils.



Markovian Branching Processes

(currently available)

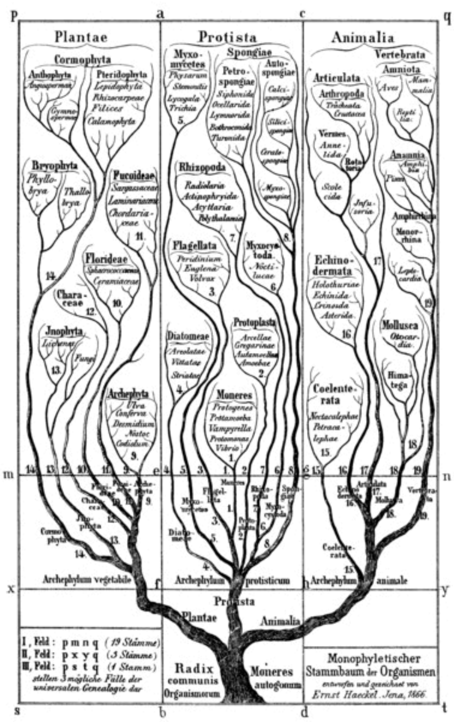
Co-supervisor: A/Prof. Barbara Holland

This project will focus on the literature review in the theory and applications of the Markovian Branching Processes, and a discussion of their potential applications in phylogenetics.

"Branching processes can be seen as stochastic processes describing the evolution of a population of individuals which reproduce and die independently, according to some specific probability distributions. These processes may be classified into several categories."[1]

"The continuous-time Markovian Multitype Branching Process (ctMMTBP) (Athreya and Ney 1971; Harris 1963) can be used for modeling a variety of phenomena. A special case of a ctMMTBP occurs when branch points are restricted to generate at most two new offspring in which case it is known as the binary branch point ctMMTBP, a process that was used in Kontoleon (2005) and Pinelis (2003) to model biological phenomena such as phylogenetic processes and macroevolution."[2]

"One of the fundamental problems in biology is concerned with deciphering and understanding the nature of evolution. The results of evolution can be seen through the diversity of life found on earth today. The relationships between species can be ascertained using a variety of biological and statistical techniques. These relationships can be pictorially represented on a tree diagram called a phylogenetic tree."[3]



References:

1. Sophie Hautphenne, An Algorithmic Look at Phase-Controlled Branching Processes, PhD Thesis, Université Libre de Bruxelles, 2009.

2. Bean N.G., Kontoleon N., Taylor P.G. Markovian trees: Properties and algorithms (2008). *Annals of Operations Research*, 160(1):31–50.

3. Nectarios Kontoleon, The Markovian Binary Tree: A Model of the Macroevolutionary Process, PhD Thesis, The University of Adelaide, 2006.

Past topics examples:

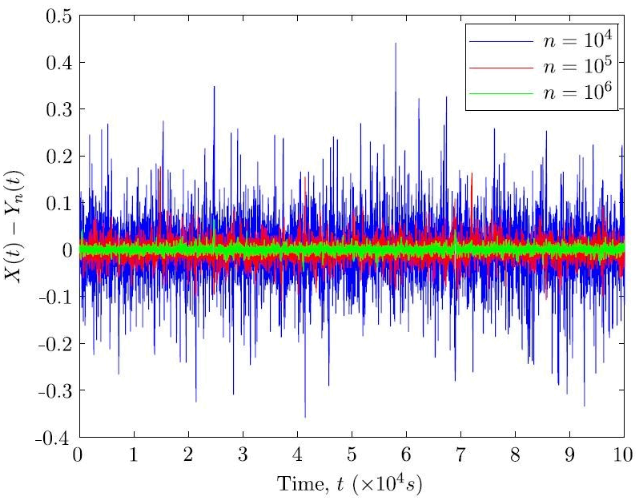
Stochastic Fluid Models and their Applications

(Mr Adrian Tanner, Honours, work in progress)

Summary: Stochastic Fluid Models (SFMs) are a class of models with a two-dimensional state-space $\{\varphi(t), X(t)\}$ consisting of a phase $\varphi(t) \in \mathcal{S}$ and a level $X(t) \in \mathbb{R}$. The phase variable $\varphi(t)$ is often used to describe the state of some physical environment that we wish to model. Simple two-phase examples are on/off mode of a switch in a telecommunications buffer, peak/off-peak period in a telephone network, or wet/dry season in reservoir modeling. The level variable $X(t)$ is used to model some continuous performance measure of the system, eg the amount of data in a buffer, the level of water in a reservoir, or the revenue earned by time t. The model assumes that the transitions between phases occur according to some underlying continuous-time Markov Chain (CTMC) with some state space $\mathcal{S}$ and rates matrix $\mathbf{T} = [T_{ij}]$, with $T_{ij} = dP(X(t) = j|X(0) = i)/dt$. Furthermore, the rate $c_i = dX(t)/dt$ of increase of the fluid level at time t depends on the phase $\varphi(t) = i$ at time t, and so the Markov Chain is the process that *drives* the fluid level at time t.

SFMs offer greater flexibility than traditional CTMCs. Originally, SFMs were inspired by problems in telecommunications. However, it has quickly become evident that SFMs have tremendous application potential in many other areas, including industrial as well as biological systems. Theoretical advances and novel solution techniques for real-world problems of considerable engineering and/or environmental significance have been developed in the recent years. The application areas include telecommunication networks and computer systems, risk processes in insurance, manufacturing systems, hydro-power generation, as well as environmental problems, such as modelling of coral reef resilience, or the spread of forest fires.

The aim of the project is to review the literature in the area of applications of SFMs and construct numerical examples using efficient algorithms and simulations. This project will suit a student who is interested in stochastic modeling, analysis and algorithmic approaches (and coding).



Queueing Models for Bed Queue in Emergency Department

(Mr Jarrad Clark, Honours, 2015/2016)

Summary: Modern hospital is a highly complex and unpredictable system, which cannot be managed efficiently using intuitive methods. Instead, we require sophisticated tools in the form of efficient algorithms developed using appropriate mathematical modeling. Compelling clinical evidence indicates that when mathematical modeling is used in hospitals, significant savings can be made that have a positive outcome to the patients.

The aim of this project was to review the literature in the area of modelling the patients flow in Bed Queue in Emergency Department. The project also involved construction of models and the derivation of various performance measures based on the analysis of a large set of data.

Time Series Analysis: Growth Season Onset Dealy Problem
(Mr Axiom Dowling, Honours, 2015)

Summary: MODIS is a scientific instrument launched into Earth orbit by NASA in 1999 on board the Terra Satellite, and in 2002 on board the Aqua Satellite. A full set of images at the various wavelengths of the entire Earth is downloaded to NASA base stations every 8 days, and is then available at no charge to anybody with Internet access. The Earth is divided into a large number of pixels. The data recorded for a given pixel can be represented as a time series $Y(t)$ of signals at times t.

This has opened the door to monitoring the environment on a global scale in a far more comprehensive and effective way than has ever been possible in the past. However, this data contains significant levels of noise and data impairments of a stochastic nature, which is impossible to mitigate without the use of analytical tools. Consequently, the task of developing appropriate methods for the analysis of MODIS satellite data has been a focus of international research in the recent years.

Stochastic models for the satellite data analysis is a crucial component of the methodology being developed for detecting changes in the environment, or predicting the risk of future scenarios. Models are used to simulate the real-life data, and are an important part of change detection algorithms in which the parameters are fitted to the models from the analysis of the data, and then various metrics are used to detect any deviations of the new inputs from these parameters.

The aim of this project was to review the literature in this area. The project also involved mathematical analysis, modeling, algorithms (and coding), as well as work with the real data and experimentation with different modeling and change detection ideas.

Consecutive K-out-of-N Systems and Applications
(Ms Eman Ali, Masters, 2013)

Summary: A linear consecutive k-out-of-n:F system is a system of n components, arranged in a line, which fails when at least k consecutive components fail. A linear consecutive k-out-of-n:G system is a system of n components, arranged in a line, which works when at least k consecutive components work. There also exist circular k-out-of-n systems and other topologies and generalizations. The aim of the project was to explore and review the current literature in the area. Examples of linear consecutive-k-out-of-n:F systems:

- A telecommunication system with n relay stations (satellites or ground stations) which fails when at least 2 consecutive stations fail,
- An oil pipeline system with n pump stations which fails when at least 2 consecutive pump stations are down.

Markov Models for Microsatellite Mutation
(Mr Tristan Stark, Honours, 2013)

Summary: Markov Chains is the most important class of models in Probability Theory, due to their modeling potential and numerical tractability. The aim of this project was to review the literature concerning the application of Markov Chains in Phylogenetics, the study of evolutionary relation among groups of organisms (e.g. species, populations), for example see:

Calabrese, P., Sainudiin, R., Models of Microsatellite Evolution, in Statistical Methods in Molecular Evolution, R. Nielsen, Editor. 2005, Springer, p. 289-306.

Stochastic Fluid Model for Deteriorating Systems
(Mr Andrew Haigh, Honours, 2012)

Summary: Markovian-modulated models are a class of models with a two-dimensional state-space consisting of a phase and a level. The phase variable is often used to describe the state of some physical environment that we want to model. Simple two-phase examples are on/off mode of a switch in a telecommunications buffer, peak/off-peak period in a telephone network, or wet/dry season in reservoir modeling. The model assumes that the transitions between phases occur according to some underlying continuous-time Markov Chain. Furthermore, the rate of increase of the fluid level at time t depends on the phase at time t, and so the Markov Chain is the process that drives the fluid level at time t. The aim of the project was to explore and review the current literature in the area.

Use of Γ for IMRT Quality Assurance
(Mr Andrew McGrath, Honours, 2012)

Summary: Intensity modulated radiation therapy (IMRT), requires some sort of measurement of the output fields. The commissioning of treatment planning systems routinely requires the comparison of measured and calculated dose distributions. A commonly used method for IMRT Quality Assurance has been the one in which parameter Gamma (Γ) is used to determine whether a particular therapy session has been successful or not. Gamma is a function of the difference between the measured dose distribution and the distribution calculated by the planning software. That is, it is the measure of the difference between the delivery and plan. A standard criterion for passing or failing the therapy has been to pass the therapy whenever some proportion of measurement points, say 90%, pass. However, there exist considerable problems with such criterion. Notably, an alternative, interesting criterion for evaluations with parameter Gamma, proposed by Chappell, Wen and Nicolau, has been used for IMRT Quality Assurance at the Holman Clinic of Royal Hobart Hospital for a number of years. The aim of the project was to analyze the existing set of data and compare the new criterion with the previously used criterion.

Markov Chains for Modelling Maintenance of Deteriorating Systems
(Ms Nawal S. Alqahtani, Masters, 2012)

Summary: High reliability associated with low costs is an important requirement for operating mechanical systems, such as aircrafts, or bridges. Failures can be very costly. These systems are subject to deterioration that may result in a failure. Thus, inspections and maintenance are carried out to prevent deterioration and failures. Improving the maintenance strategies has a great affect on cost savings. In order to quantify the influence of maintenance on reliability and costs, we need mathematical models. The aim of this project was to provide a review of the literature on state diagrams for probabilistic maintenance models.

Probabilistic Modeling of Manufacturing Systems
(Mr Robert Lillico, Honours, 2008)

Summary: The project involved a manufacturing problem at the Cadbury Chocolate Factory. The aim was to develop a suitable model for the chocolate production line. The specific goal was to determine the optimum size of the excess buffer that collects chocolates at the moments of downtime. We used a continuous-time Markov Chains and simulation techniques amongst other approaches, to solve this problem.