University of Strathclyde

Department of Management Science

Three Competing Risk Problems in the Study of Mechanical Systems Reliability.

by

Michael Richard Burnham

Supervised by

Prof. Tim Bedford & Prof. Lesley Walls

A thesis presented in fulfilment
of the requirements for the degree of Doctor of Philosophy

Submitted 2010

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed: Date:

Acknowledgements

In completing this thesis I owe a great many debts to a great many people for their support and advice over the last four years. I would like to give particular recognition to those individuals that have helped me most throughout this period of my life.

First and foremost I would like to thank my two supervisors, Tim Bedford and Lesley Walls. Their guidance and support have sustained me on this fascinating journey and it is only through their trust and encouragement in giving me the space to explore this rich landscape of mathematical possibilities that I have been able to forge a thesis that attends to my own peculiar mathematical idiosyncrasies.

I owe a particular debt of gratitude to Isha Dewan for allowing me to study under her guidance for two months at the Indian Statistical Institute in New Delhi during my third year and for introducing me to the third problem presented in this thesis. I also owe Isha my thanks for teaching me about non-parametric statistics, as well as for giving me several important insights, timely advice on several of my drafts and continued support.

I would like to thank Frank Coolen for several useful discussions had during a competing risk meeting in Glasgow in 2009. I owe him special thanks for his suggestion of how to approach solving the third problem of my thesis.

I would also like to extend my thanks to the following people for their help and support, in no particular order:

Olivier Gaudoin for giving me the opportunity to visit Grenoble University for two weeks in my second year in order to study the work on competing risks that he and others had done there; Yann Dijoux for the many helpful conversations about competing risk modelling that we had, as well as for his kindness in showing me around Grenoble; Babakalli Alkali for helping me take my first steps into the world of competing risks; Matthew Revie and Gavin Hardman for their continued

advice and support on all things PhD; Alison, Caroline, Claire and Jackie without whom I would have been unregistered, penniless and starving on the streets; my fellow students with whom I have been able to celebrate in our successes and commiserate in our defeats; the academic staff of the Management Science Department who have given great advice and perspective that has helped me to develop in Science, Philosophy and Mathematics.

Lastly, I would like to thank my father, whose inquiring mind has been both the model and the encouragement that has always led me to look at the world and ponder.

Having reached the end of this particular journey, I have a keener understanding of what Aeschylus meant in saying that we must "suffer into truth".

Contents

List of Figures

List of Tables

Abstract

This thesis considers three problems within the field of competing risks modelling in reliability.

The first problem concerns the question of identifiability within certain subclasses of Doyen and Gaudoin's recently proposed generalised competing risks framework. Bedford and Lindqvist have shown identifiability for one such subclass - a two component series system in which, every time a component fails it is restored to a state "as good as new", while the other component is restored to a state "as bad as old".

In this thesis two different subclasses are shown to be identifiable. The first is a generalisation of the Bedford and Lindqvist example for series systems with n components. The second is an n component series system in which each time a component fails it is restored to a state "as good as new". At the same time the remaining components are restored to a state "as good as new" with probability p (which may depend on both the component being restored and the component that failed), or to a state "as bad as old" with probability $(1-p)$.

The second problem concerns the use of competing risks models to study opportunistic maintenance. Bedford and Alkali proposed the following model - the system exhibits a sequence of warning signals, the inter-arrival times of which are assumed to be independently distributed (but non-identical) exponential random variables. The hazard rate of the time to system failure is modelled as a piece-wise exponential distribution, in which the hazard rate is constant between signals. A sequence of maintenance opportunities occurs according to a homogeneous Poisson process and the first opportunity after the k^{th} signal is used to preventatively maintain the system.

In this thesis closed form expressions for the above model are calculated (subject to some minor technical restrictions) for the marginal distributions of the time to both preventative and corrective maintenance. Also, the sub-distribution of the time to corrective maintenance is calculated.

The third problem concerns the estimation of the marginal distribution for one of two independent competing risks, when knowledge of which risk caused the system shut-down is unknown for some of the observations in the dataset.

In this thesis a new estimator based on the Kaplan-Meier product limit estimator is developed for the above set-up. A re-distribution to the right algorithm is also developed and this is shown to be equivalent to the new estimator. The new estimator is also shown to be consistent.

Notation

This is a list of nomenclature that is used throughout the thesis.

X	-	random variable representing the time to system failure (defined on p.17).
Z	-	random variable representing the time to preventative maintenance (defined on p.17).
T	-	random variable representing the time to system stoppage (minimum of X and Z, defined on p.18).
δ	-	random variable representing the cause of system stoppage (1 if $X < Z$, 0 otherwise. Defined on p.18).
$f_Y(t)$	-	probability density function of a random variable Y.
$F_Y(t)$	-	cumulative distribution function of a random variable Y.
$S_Y(t)$	-	survival function of a random variable Y (defined on p.18).
$S_Y^*(t)$	-	sub-survival function of a random variable Y (defined on p.18).
$S_X^c(t)$	-	normalised sub-survival function of a random variable Y (defined on p.19).
$S(t)$	-	joint survival function of X and Z (defined on p.19).
$\Phi(t)$	-	Phi function (defined on p.19).
$\lambda(t)$	-	initial failure intensity function (defined on p.31).
λ_t	-	conditional failure intensity function (defined on p.31).
A_i	-	i^{th} virtual age (defined on p.32).

Part I

Preliminary Theory

Chapter 1

Introduction

This thesis addresses three problems found in the domain of reliability theory, using probability and statistics. All three problems are described using competing risk models (introduced in Chapters 2 and 4). As such, most of this thesis will have a very theoretical emphasis. The purpose of this chapter is to describe the background against which competing risk models reside. Firstly a definition of reliability will be given along with several motivations for why people are interested in it and what sorts of questions are of importance. The discussion will then turn to describing the data that effectively serves as the "input" to these models. This description will involve discussion of where the raw data comes from and how this raw data is processed into a form that can be input into the models, along with the assumptions that are made to achieve this transformation.

1.1 Defining Reliability

There are many definitions of reliability given in textbooks on the topic, most just differing slightly in their wording. The following definition is given by Rausand and Hoyland and is a reasonable exemplar. Reliability is "The ability of an item to perform a required function, under given environmental and operational conditions and for a stated period of time."[53]
The concept of "item" can be sharpened by considering several sub-categories. There are various different taxonomies, however the one adopted throughout this thesis is taken from Moss[69].

Component	-	an item that is not considered repairable (e.g. bearings, seals, pipes and flanges).
Equipment	-	an item that is an assembly of components that operates to provide a specific function (e.g. Compressors, pumps and valves).
System	-	an item that is a combination of equipment and components (in series or parallel).

It should be noted that the boundary between equipment and system can shift depending on the focus of the particular application of reliability. Usually one would study a system, sometimes one might study a piece of equipment and as such it is possible to think of that piece of equipment as a system in it's own right. This notion of a boundary which delimits an item is a very important one in reliability and will be mentioned again, later in this chapter.

Having discussed what reliability is, it seems natural to consider why people are interested in the study of reliability. Bendell and Cannon[14] give six questions that they consider typical of those that prompt people's interest in the study of reliability.

- How reliable is this system, or what is its probability of surviving a specified number of hours?
- Which is the most reliable of a number of possible designs for a system?
- Is it worth introducing redundancy into a design in case a component fails?
- How is the field performance of an item affected by its conditions of use?
- What will be the warranty, maintenance and logistic support costs for the system in the field?
- Can we improve the reliability and throughput of the system?

Andrews and Moss[4] claim that there are two main areas in which reliability is applied: "Reliability assessment of new designs" and the "analysis of in-service reliability performance". It can be seen that the six questions above can be placed quite neatly into these two categories (although sometimes a question can be asked in either category). It is clear that in the first area there is no direct operational data from which to derive quantitative answers to questions about reliability. Green and Bourne[50] present a pertinent approach to the question of what can be done when such "data are required on an element in a particular application".

1. Is there data available from a "nominally identical element that has been used in a nominally identical application on a previous occasion"?
2. If not, then is there data available from a "generally similar element"?
3. If not then "subdivide the element under consideration" and ask the same two questions above, of these sub-elements.
4. "At any stage of the sub-division process, the possibility of setting up a practical experiment to obtain the required data may be reviewed."
5. "the reliability characteristics of the system and it's various elements can be monitored and recorded during it's initial and subsequent operating life".

Naturally the first three points assume that points four or five were carried out when these "nominally identical" or "generally similar" items were used on these "previous occasions". Thus it can be seen that the primary means of data acquisition is through field experience and sample testing. The secondary means of data acquisition is through access to what are called reliability databases. These are the repositories implicitly referred to when Green and Bourne mention the recording of reliability characteristics in point five (although equally applicable to the results of the experiments in point four). It is possible that such databases are available to an organisation through it's own previous experience and contained within it's own particular brand of reliability database. However in the event that this is not the case there have been a number of different efforts to construct "generic" reliability data sources, which are discussed in Section 1.3.

It is worth noting one of the important features of modern reliability databases highlighted by Cooke[23, 24], that they have different users with different needs. Cooke recognises "at least three types of users":

- "the maintenance engineer interested in measuring and optimising maintenance performance";
- "the component designer interested in optimising component performance";
- "the risk/reliability analyst wishing to predict reliability of complex systems in which the component operates".

Again it can be seen that the three points on this list can be placed quite neatly within the above two categories of "analysis of in-service reliability performance" and the "rdeliability assessment

of new designs", discussed in the previous section. The first two items falling into each category respectively, while the last item falls into both categories.

1.2 Raw Reliability Data

Andrews and Moss[4] state that "most organisations collect data that can be analysed to provide reliability information. These data exist in maintenance and test records, operational logbooks, and other technical information systems". Moss[69] outlines three broad categories of in-service reliability data:

Inventory Data	-	"these define the design and functional characteristics of the item".
Failure Event Data	-	"textual descriptions of each failure event with additional codes to relate to the item's inventory, failure codes and operational and environmental conditions".
Operating State Data	-	"dates and times of failure, operating state prior to failure, active repair and waiting times".

Inventory data also includes a strict definition of the boundary of the item under consideration. The idea of an item boundary was discussed earlier and is important in order to clearly define what is meant by an item failure. The definition of reliability given above included "the ability of an item to perform a required function", it is therefore obvious that the functional characteristics of the item are part of the data that will be necessary in order to determine the reliability of that item. Inventory data is usually something that is developed in the design stage of the in-service reliability database system.

Failure event data is most often derived from a range of sources, this will almost always include records from maintenance logs. Moss[69] highlights one of the difficulties with the assimilation of these data sources into an in-service reliability database, "Maintenance is concerned with testing, repair and replacement of equipment and the records will therefore concentrate on the work done to restore equipment to an acceptable state". It can often be a significant challenge to maintain a high standard of quality when extrapolating from maintenance records to failure event data.

Operating state data can be placed on a time-line, similar to that shown in Figure 1.1, to show the

behaviour of the item throughout the course of it's life. These events would include failures and maintenance activities. Each "point" on the time-line would contain the detailed information about what happened at that particular event. The time axis would typically be in terms of calendar elapsed time at this stage. It is possible to model this time-line using a marked point process, which will be introduced in a later chapter.

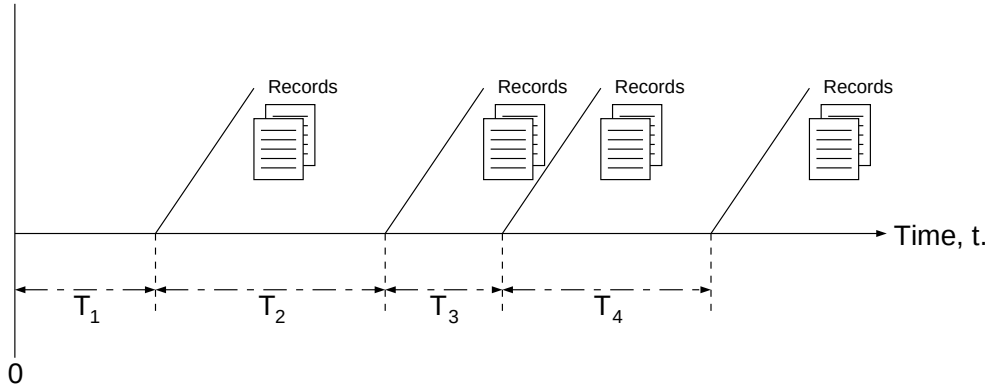


Figure 1.1: An Example Time-line of Operating State Data.

A failure event is usually described in the event history by a failure mode. Failure modes describe "the way in which a component fails, usually from a functional or sub-functional point of view"[23]. They often describe the "symptoms" of what has happened, for example the centre for chemical process safety (CCPS) describes failure modes as "the observed manner of failure. The failure modes describe the loss of required system function(s) that result from failures"[47]. Moss describes a failure mode as "the effect by which the failure is observed"[69]. The term "failure event" is slightly ambiguous in that, as well as being used to refer to failures, it is also frequently invoked to describe the performance of preventative maintenance on that item. This practice will be continued throughout this thesis as well.

Cooke also defines the notions of failure mechanisms which "describe the actual physical processes leading to a failure"[23] and failure causes which describe"the circumstances during design, manufacture or use which have led to a failure"[47]. Moss also refers to a failure cause as defining "the mechanism that led to the failure"[69]. This is another example of where item boundaries are important. As discussed above, it is often possible to sub-divide a system or piece of equipment into several parts and consider these parts separately. It is easy to see that a failure cause at the

component level leads to an observed effect at that level, which can be defined as a failure mode for that component. But at a higher level where the component is a part of a piece of equipment, the previously mentioned failure effect of that component can be considered as the failure cause of the piece of equipment, which can lead to an observed effect in the piece of equipment that can be defined as one of the failure modes of that piece of equipment. This manner of defining failure modes as causes in the next inventory level up in a system heirachy is very useful, but it is also another reason why it is essential to clearly denote the item boundaries in order to prevent confusion and ambiguity.

Mention has already been made of maintenance activities and how these are also often included in failure data. Doyen and Gaudoin[41] define two different types of maintenance activity. They define corrective maintenance (CM) as any maintenance action that "is carried out after a failure and intends to put the system into a state in which it can perform its function again". They define preventive maintenance (PM) as any maintenance action that "is carried out when the system is operating and intends to slow down the wear process and reduce the frequency of occurrence of system failures".

Cooke and Bedford[25] describe four different scheduling types for maintenance:

- **Calendar based** - Planned maintenance activities, not based on observed deterioration of component, but scheduled from calendar time, or some surrogate, such as cycles or cumulative load.
- **Condition based** - The components state is observed to deviate from that which the manufacturer intended, though the components functionality (as required by the system in which it serves) is still maintained.
- **Opportunity based** - Maintenance is undertaken when a suitable opportunity presents itself. If the system is shut down for overhaul of one component, other components can be maintained as well. The decision to maintain a component at an opportunity might or might not be triggered by the condition of the component.
- **Emergency** - Actions to repair a component which is in a state that disables the system. The system cannot function until the repair is done, and the repair activity was not planned beforehand.

When a catastrophic or degraded failure is encountered the system is normally halted (if not

halted already by the failure) for emergency corrective maintenance to take place. Preventative maintenance is more complex and can involve a combination of the remaining three scheduling types.

To consider the event history of a single item in isolation would ignore the existence of event histories for similar if not identical items. Large organisations will often have multiple installations with the same or similar items performing the same tasks. Especially in the case of safety critical systems, failure events can be quite rare for a single item. This leads to the common practice of pooling such event histories together as described by Cooke and Bedford[25]. What happens is that the individual event histories of each item are adjusted or standardised to allow comparisons between event histories. For example, adjustments are made to take account of different environmental conditions and stresses placed on the equipment throughout it's life. This has been done using a variety of different methods, including stress factors or duty factors as discussed in Andrews and Moss[4]. The adjusted data can now be "pooled". The act of pooling assumes that after adjustment, the different event histories for specific items can be considered as different realisations of the same marked point process. The pooled data can be used to estimate all sorts of properties associated with reliability and maintainability.

1.3 Generic Reliability Data

Many of the definitions made in discussing reliability data in the previous section can be carried over to the discussion of generic reliability data, particularly the idea of failure modes, preventive maintenance and corrective maintenance. In-service reliability databases are set up to capture the primary reliability data of functioning systems and equipment. This enables the estimation of all sorts of properties associated with reliability and maintainability. Generic reliability data handbooks on the other hand, are an attempt to provide such estimates to users who either do not have any in-service data of their own or are seeking to benchmark their own data against an external source. These handbooks will usually include inventory data, but will not include failure event data or operating state data.

An early example of such a handbook is the US Defence Department Handbook MIL-HDBK-217E[31] published in 1986, which presents point estimates of the number of failures per 10^6 hours for electronic components. These estimates of the base failure rate were also supplemented by environmental factors and stress factors under a range of different environmental conditions, similar

to the factors discussed at the end of the previous section.

More modern generations of reliability handbooks typically contain ranges of values instead of point estimates and these failure rates would be for individual failure modes rather than just a general failure rate for each item. The OREDA[71] handbook (Offshore Reliability Data) is a typical example of this, presenting failure data for equipment used in the offshore industry. Estimates for lower, upper, mean and standard deviation values of the failure rate for each failure mode are provided, as well as the mean active repair time. The size of the sample population and the length of the period of observation are also included for each equipment class. Other well known examples of such handbooks include CCPS[46] and EIREDA[76] (European safety, reliability and data association).

This generic data is typically derived not just from one organisation's in-service databases, but from the in-service databases of multiple contributing organisations. In order to try and maintain the consistency and quality of the pooled in-service data used to calculate these measures, it is common for generic reliability databases to be constructed around standards, such as the IEC 60300-3-2 standard[54] (International Electrotechnical Commission) for collection of dependability data from the field and the ISO 14224 standard[55] (International Organization for Standardization) for the collection and exchange of reliability and maintenance data.

Akmedjanov[1] and Fragola[48] are two good review papers on the historical development of generic reliability databases.

1.4 Competing Risks Data Structures

This section will introduce the data structure that is the required format for data input into competing risk models and will show how the adjusted pooled data discussed at the end of Section 1.2 is converted into this competing risks data structure.

Users of reliability data frequently want to extract information about distinct types of event in the history. For example, in the case of the maintenance engineer seeking to improve the existing maintenance policies or to compare the impact of maintenance activities between plants, they would want to distinguish between events involving corrective and preventative maintenance. In the case of the component designer seeking to establish a baseline estimate for a new design or perhaps to compare the reliability of different design alternatives, they would want to isolate the reliability characteristics of specific components within the system. Recall that when discussing

generic reliability handbooks, mention was made that failure rate estimates are commonly provided for each failure mode of a given item.

The question is, how should events in the time-line that are caused by other failure modes be handled? Given such an event (Event A say), one cannot simply exclude Event A from the time-line and remove the time that has elapsed since the occurrence of the most recent event preceding Event A (denote this time period T_A). This is because Event A contains implicit information about the failure rate item, namely that the item survived without failure of any kind for the length of time T_A. To ignore this implicit information by removing such events from the time-line would lead to bias in any estimates produced. Instead events are grouped into subsets, typically subsets of failure modes. A mark δ_A is then associated with each failure mode, resulting in an event time-line with a mark denoting the failure mode of each event, as shown in Figure 1.2.

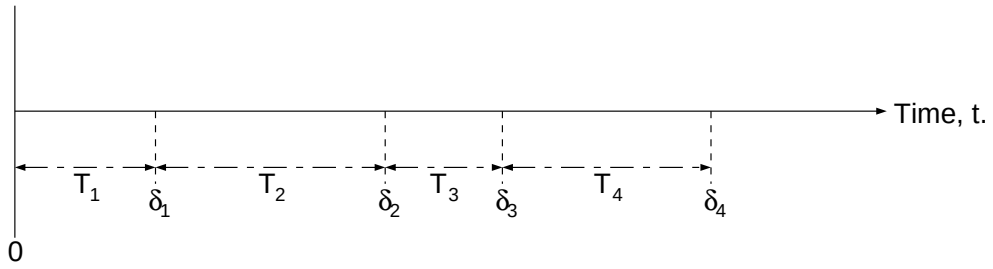


Figure 1.2: An Example Time-line of the Marked Data (a Marked Point Process).

The act of distinguishing between different events in the event history or "points" of the process is described as colouring by Cooke and Bedford[25] and as already mentioned earlier, the resulting process can be described using a marked point process.

Thus the resulting data set is a sequence of times and marks. It is often more useful, while preserving the ordering of the sequence, to use the time between successive events instead of the time since the beginning of the observation period. The resulting sequence of pairs is typically denoted $\{(T_i, \delta_i)\}_{i\geq 1}$. Unless explicitly stated, when referring to "observable data" or "datasets" in this thesis, it is data in the format above that will be intended.

Having introduced the data structure that will serve as the input to the various competing risk models discussed in the remainder of this thesis, a brief outline of the thesis is now given.

1.5 Thesis Outline

Having described the data structure that will be used throughout this thesis, the general outline of the remainder of the thesis will now be sketched. The remainder of this document is divided into two main parts. The first part will introduce the two frameworks of classical competing risks and generalised competing risks.

Chapter 2 will introduce the classical competing risks framework and discuss some of it's important properties. The primary distinction between classical competing risks and generalised competing risks is that classical competing risks is restricted to situations in which the system under study is always restored to a state as good as new after every event.This will be followed by a review of several models developed within this framework for studying the interaction between corrective and preventative maintenance.

In preparation for introducing the generalised competing risks framework, Chapter 3 will introduce the notion of imperfect maintenance models for repairable systems. These models are largely discussed within the context of survival analysis although any extensions made to competing risks will be highlighted as appropriate. Chapter 4 will bring together the two topics of imperfect maintenance models and classical competing risks in a discussion of the recent work by Doyen and Gaudoin [41] in attempting to generalise the classical competing risks framework to include maintenance models with imperfect repair, which is called the generalised competing risks framework.

The second part of this thesis consists of the authors contributions to this field. Three problems within the field of competing risks modelling in reliability are addressed.

The first problem, which is the subject of Chapter 5, concerns the question of identifiability within certain subclasses of the generalised competing risks framework (The notion of identifiability is introduced in Chapter 2 when discussing classical competing risks). The work of this chapter is largely motivated by the work of Bedford and Lindqvist[11] in demonstrating that identifiability is preserved for a small sub-class of generalised competing risk models. The Bedford and Lindqvist model is a two component series system in which, every time a component fails it is restored to a state "as good as new", while the other component is restored to a state "as bad as old". The proof technique developed by Bedford and Lindqvist will be placed in a more general framework and two other sub-classes of the generalised competing risk framework will be shown to be identifiable. The first is a generalisation of the Bedford and Lindqvist example for series systems with n components. The second is an n component series system in which each time a component fails it is restored to

a state "as good as new". At the same time the remaining components are restored to a state "as good as new" with probability p (which may depend on both the component being restored and the component that failed), or to a state "as bad as old" with probability $(1-p)$.

The second problem, which is the subject of Chapter 6, is to calculate closed form expressions for several important functions in the following competing risk model involving opportunistic maintenance that was proposed by Bedford and Alkali[10]. A system exhibits a sequence of warning signals, the inter-arrival times of which are assumed to be independently distributed (but non-identical) exponential random variables. The hazard rate of the time to system failure is modelled as a piece-wise exponential distribution, in which the hazard rate is constant between signals. A sequence of maintenance opportunities occurs according to a homogenous Poisson process and the first opportunity after the k^{th} signal is used to preventively maintain the system. Closed form expressions are calculated (subject to some minor technical restrictions) for the marginal distributions of the time to both preventative and corrective maintenance, as well as for the sub-distribution of the time to corrective maintenance (These are all important functions in the classical competing risks framework that are introduced in Chapter 2). Given the large number of calculations involved, a simulation study will be used to provide extra assurance that the calculations are correct.

The third problem, which is the subject of Chapter 7, concerns the estimation of the marginal distribution for one of two independent competing risks when knowledge of which risk caused the system shut-down is unknown for some of the observations in the dataset. A new statistical estimator is developed, which is a variant of the Kaplan-Meier product limit estimator. The new estimator is shown to be consistent and several examples of it's application on test data are given to illustrate it's manner of use.

Chapter 8 brings together the results of the work done in the earlier chapters of the second part of this thesis. The main results are reiterated and where appropriate, conclusions are drawn and discussions of further work in each problem are discussed. Finally, a discussion of issues surrounding the practical application of competing risk models is given, summarising future directions for research to help reduce the current gap between theory and practice.

There are three annexes in the appendices. Appendix A provides the detail of the extra calculations required for part of the second problem (the subject of Chapter 6), Appendix B provides the MATLAB computer code used during the simulation study for the second problem and Appendix C gives a proof of the extension of the Peterson Bounds for classical competing risk models (discussed

in Section 2.2) to the case of symmetric GCR models using virtual ages (discussed in Section 4.2).

Chapter 2

Classical Competing Risks

This chapter will introduce the main ideas of the classical competing risk framework within the context of system reliability. In the previous chapter a data structure was introduced that will serve as the input for the models introduced in this chapter and Chapter 4. The primary distinction between classical competing risks and generalised competing risks (which will be introduced in Chapter 4) is that classical competing risks is restricted to situations in which the system under study is always restored to a state as good as new after every event (This condition will be formally defined in the next sub-section).

Classical competing risk models have often been used in other domains outside of reliability. There is a strong history of the use of competing risks within medical studies for example, such as Hoel's[51] use of competing risks to model mortality data from animal laboratory experiments. This shared heritage of competing risk modelling often results in alternative terminology being used in a number of the classical papers of competing risk theory, because of the different background of the author. For example, in the medical context one speaks of "the force of mortality" (applied to a person or animal) instead of the hazard rate (applied to a component). The terminology applied throughout this thesis will be that which is commonly used amongst the reliability community and when discussing papers from outside the reliability domain the discussion will frame those ideas using reliability terminology.

Because this chapter focuses on the classical competing risks framework, the concepts and notation will be developed using the language of random variables. When introducing generalised competing risks in Chapter 4, this framework will be recast into the language of point processes.

Consider a system that runs continually. An event is said to occur whenever a maintenance action or failure takes place. Each event is described by a failure mode (introduced in Chapter 1) and the elapsed time between the occurrence of that event and the most recent previous one. Repair times are considered negligible.

2.1 Basic Model Structure

Classical competing risk models assume that the renewal property holds. This means that the times to each event can be modelled using a renewal process, with times between events modelled as independent identically distributed random variables. This is most appropriate for situations in which the costs of maintenance activities is so large that it is imperative to take steps to avoid them as much as possible by ensuring that whenever maintenance is carried out, the system is restored to a state that is as good as new. This assumption will be relaxed in Chapter 4, where a more general competing risk framework involving imperfect maintenance is introduced.

Consider a series system subject to n different failure modes, the vector of n random variables $\left(X^1, \ldots, X^n\right)$ is defined such that the i^{th} random variable X^i denotes the time to the next potential occurrence of system stoppage due to the i^{th} failure mode.

In some cases the behaviour of only one particular failure mode is of interest. The modelling can be simplified by re-coding all the remaining failure modes into one category, thus leaving the failure mode of interest and a "nuisance" failure mode. An important example of only considering two types of event is when studying the interaction between corrective and preventative maintenance. In such cases the time to failure of the system will be denoted by X and the time to preventative maintenance will be denoted by Z. One assumes that these two events cannot occur at the same time, in other words that $P\{X = Z\} = 0$.

The remainder of this chapter will develop the notation of classical competing risks in terms of these two causes, although it should be noted that multivariate analogues of most of the functions introduced in this section can be constructed in an obvious manner for the vector of random variables $\left(X^1, \ldots, X^n\right)$.

In practice, what can actually be observed is the minimum of these two random variables (denoted

by $T = \min(X, Z)$) and which random variable this is, denoted by

$$\delta = I_{\{X<Z\}} = U = \begin{cases} 1 & \text{if } X < Z \\ 0 & \text{otherwise,} \end{cases}$$

where $I\{u\}$ is the indicator function, equal to 1 if u is true and 0 if u is false. An important quantity is the marginal survival function (often just called the marginal) of X, which is defined in terms of the cumulative distribution function of X (F_X) as follows

$$S_X(t) = P\{X > t\} = 1 - P\{X \leq t\} = 1 - F_X(t). \tag{2.1.1}$$

This function describes the probability that the system will have survived (i.e. not experienced a critical failure event) up to (but not including) time t and is therefore a useful function when studying the impact of changes in the maintenance policy upon the reliability of a system. However this function cannot be directly estimated from the observable data for reasons to be discussed in Section 2.2. Given what can be observed it is only possible to directly estimate the sub-survivor function of X, defined as follows

$$S^*_X(t) = P\{X > t, X < Z\}. \tag{2.1.2}$$

One can similarly define marginal and sub-survival functions for Z. The following definition will prove useful in later sections.

Definition 1 (Sub-survival pair)**.** *S^*_X and S^*_Z form a sub-survival pair, written $\{S^*_X, S^*_Z\}$, if*

1. *S^*_X and S^*_Z are non-negative non-increasing real functions with $S^*_X(0) \leq 1$ and $S^*_Z(0) \leq 1$;*
2. *$\lim_{t\to\infty} S^*_X(t) = \lim_{t\to\infty} S^*_Z(t) = 0$;*
3. *$S^*_X(0) = 1 - S^*_Z(0)$.*

Consider a sample observation, $T = t_0$ and $\delta = 1$ say. Because $\delta = 1$ and T is the minimum of X and Z this necessarily implies that $X = T$ in this case, so the event $\{X = t_0, X < Z\}$ is observed (Note that the event $\{X = t_0\}$ is not observed from the data). By symmetry the case of $T = t_0$ and $\delta = 0$ is the event $\{Z = t_0, Z < X\}$. This implies that given a sample of M pairs of observations

$\{(T_i, \delta_i)\}_{1 \leq i \leq M}$, these sub-survivor functions can be estimated by using a variant of the empirical distribution function as follows

$$\hat{S}_X^*(t) = \frac{1}{M} \sum_{i=1}^{M} I\left\{T_i > t, \delta_i = 1\right\}$$
$$\hat{S}_Z^*(t) = \frac{1}{M} \sum_{i=1}^{M} I\left\{T_i > t, \delta_i = 0\right\}.$$

Unlike the survival function, the sub-survival function does not equal 1 at the origin, so $1 - S_X^*(t)$ is not a probability distribution function. Instead $1 - S_X^*(t)$ is often referred to as the sub-distribution function, denoted by $F_X^*(t)$. The sub-survival function can be normalised so that it equals 1 at the origin, by dividing by the constant $S_X^*(0)$. It turns out that this normalised sub-survival function is equal to the conditional sub-survivor function, which is the probability that $X > t$ conditioned on the event that $X < Z$. This function is defined as follows

$$S_X^c(t) = P\left\{X > t | X < Z\right\} = \frac{P\left\{X > t, X < Z\right\}}{P\left\{X < Z\right\}} = S_X^*(t)/S_X^*(0). \tag{2.1.3}$$

It is also possible to consider the joint survivor function of X and Z

$$S(t) = P\{X > t, Z > t\}. \tag{2.1.4}$$

Another function, which is of use during model selection, is the probability of censoring beyond time t. That is, the probability that at time t since the component went into service, the next event will be a preventative maintenance action

$$\Phi(t) = P\{Z < X | Z > t, X > t\}. \tag{2.1.5}$$

This function is closely related to the sub-survivor functions as can be seen

$$\Phi(t) = \frac{P\{Z < X, Z > t\}}{P\{Z > X > t\} + P\{X > Z > t\}} = \frac{S_Z^*(t)}{S_X^*(t) + S_Z^*(t)}. \tag{2.1.6}$$

This necessarily implies that $\Phi(t)$ can also be directly estimated from observable data, just like the functions S_X^*, S_Z^*. These three functions are often useful when choosing between particular classes of models. The first two functions can often be used too as will be discussed in the next

two sections.

Section 2.2 will discuss why the functions S_X and S_Z are not identifiable from observable data and how S_X^* and S_Z^* can sometimes be used to reject certain models. Section 2.3 will present several reliability models developed to describe the interaction between corrective and preventative maintenance and will show how the behaviour of $\Phi(t)$ can sometimes be used in model selection.

2.2 Identifiability Problem

As discussed above, each potential event is modelled as a random variable. This means that the occurrence of one event will prevent the observation of the remaining random variable, which models the other potential event time. This remaining random variable is said to be censored by the first random variable and the resulting pattern of censoring creates problems when trying to estimate the survival functions of X and Z. The random variables X and Z are often referred to as latent random variables in this context.

Each latent random variable represents the time to system stoppage due to a given cause. The presence of censorship ensures that only one such random variable will be observed and the other will be unobserved. As Crowder puts it, "in a sense they cease to exist, if they ever did"[30]. This question as to the existence of non-observable constructs is quite important. Statistics is usually conducted from the positivist tradition of scientific philosophy which, as Popper puts it, views "truth as manifest"[75]. This viewpoint would cast doubt on the existence of a construct that was not manifest, as in this case with the censored latent random variables. An alternative philosophical perspective that will be adopted throughout this thesis, is that of critical realism. In summarising part of the position of critical realism, Bhaskar says that the basic principle of the realist philosophy of science is "that perception gives us access to things and experimental activity access to structures that exist independently of us"[15]. In other words critical realism views reality as objective, but unlike positivism, critical realism asserts that there are layers to reality, not all of which are easy to access or even directly observable. Thus although some may object to the inclusion of elements that model non-observable constructs, this practice is viewed by the author as reasonable, especially given the added explanatory power that such modelling yields in this case. The impact of this censoring behaviour on the estimation of the marginal distributions of the latent random variables was addressed by Tsiatis[82]. Noting that users of competing risks would frequently assume that X and Z were independent, Tsiatis states that these users would construct

a model using the observable data and that they would record "expressions of satisfaction" that there was a reasonable agreement between the computed value of the sub-survival functions and their empirical estimates. Tsiatis[82] showed that it was not possible to test the validity of this assumption using existing data, by showing that given any estimate of a sub-survival pair, an independent model could always be found such that the model sub-survival pair would agree with the estimated sub-survival pair. In other words, the observable data is not sufficient to identify a unique model. Crowder[29] showed that given any empirical sub-survivor pair estimated from a dataset, there are an infinite set of possible models for which the model sub-survival pair would agree with the estimated sub-survival pair.

This feature of the competing risk framework is called the non-identifiability problem. Some results have been developed which quantify the level of uncertainty due to non-identifiability. Peterson[73] showed that the survival functions of X and Z are bounded point-wise in the following way

$$S_X^*(t) + S_Z^*(0) \geq S_X(t) \geq S_X^*(t) + S_Z^*(t). \tag{2.2.1}$$

Crowder[29] was able to show that the difference between the survivor function of X and the Peterson upper bound is a non-decreasing function in t. Bedford and Meilijson[13] improved these results by including a subtle case not considered by Peterson. They were also able to generate a confidence interval based on these bounds which could be used as the basis for a statistical test on the goodness of fit of any particular model given the data. However, the confidence intervals are quite conservative[9].

The modern approach to competing risks in reliability has largely been to select a class of models relying on considerations other than the observable data (this will be discussed in Subsection 2.3.11). It may be possible to reject some model classes from consideration based on Bedford and Meilijson's goodness of fit test, but because of the conservative nature of this test it is often necessary to consider other sources of evidence. For example, model selection might be based on expert judgement gained by eliciting information from the maintenance crew, or perhaps based on an understanding of the physical processes involved. An example of such a study can be found in [17] where the assumption of independence for two different wear mechanisms on a cylinder liner is made on the grounds that one of the physical wear processes occurs on the outside of a cylinder liner and the other occurs on the inside. In such cases it may not prove difficult to mount a challenge against the validity of assumptions like this one, but the point is to recognise that any assumptions

made to limit the class of applicable models will almost always be based on evidence of this sort. Although some may choose to argue that only statistical evidence based on observable data should be used to justify choices of model selection (in order to satisfy objective notions of scientific rigour found within much of the natural sciences for example), in this case such evidence does not and indeed cannot exist. In order to proceed with model building, these inherent weaknesses in the chain of argument for a given model are unavoidable. This reasoning is apparent in existing studies, as exemplified by the following quotation from Langseth and Lindqvist[58]

> *"due to non-identifiability of competing risks problems we cannot really infer that [the LBL model] combined with the [Brown and Proschan] model is the correct model for the present dataset. We can only conclude that the data does not reject this combined model...To be advocating this model we would need to discuss the assumptions of [the repair alert model] with domain experts".*

The direction of much research over recent years has been to identify assumptions which are very reasonable and are general enough to include a wide variety of models while also preserving identifiability. A number of models have been developed in this spirit, usually by focusing on extra knowledge gained from the maintenance policies and practices implemented for the particular system under study. Several of these models will be discussed in the next section.
One of the significant open questions that this thesis will explore in Chapter 5.3 is to what extent the non-identifiability problem is present for systems in which the repair regime is not assumed to be perfect.

2.3 Some Competing Risks Reliability Models

This section contains a series of competing risks models developed in reliability which model the interaction of preventative and corrective maintenance. For each model, the detail of the model is discussed and where appropriate, intuitive interpretations of the various model components are given. Two important properties that will also be discussed are model identifiability and the behaviour of $\Phi(t)$. The concluding sub-section will discuss the question of model selection.

2.3.1 Independent Model

This model assumes that X and Z are independent. The application of this model to studying the interaction between preventative and corrective maintenance is questionable, because it implies that preventative maintenance is not dependant on the failure events that have occurred. This implies that if all preventative maintenance activities were to cease, there would be no impact on the rate of occurrence of critical failures. However, applying this model in the study of critical failures caused by one of a number of components arranged in a series system is frequently useful and such models will be discussed when considering generalised competing risks in Chapter 4.

It has been shown by several authors independently[70, 67] that if X and Z are independent, non-atomic and share essential suprema then their marginal distributions can be uniquely identified. Van der Weide and Bedford[83] showed that given a pair of sub-distribution functions and assuming independence, although the survival functions can be calculated it is possible that one of the random variables will have an atom at infinity. These results imply that the marginal distributions can always be identified assuming independence, although as previously mentioned, this assumption will need to be justified outside of the given sub-survival pair.

An important result presented by Bedford and Meilijson[12] showed that assuming exponential distributions for both X and Z, there is an interval of potential values for the failure rate of X and that the lower bound of this interval corresponds to the case when X and Z are independent. Put another way, any other model involving dependence between X and Z would give a higher estimate of the failure rate for X than the independent model in this case. These results demonstrate the pitfalls of poor model selection by showing the bias that can be incurred when incorrectly assuming independence.

2.3.2 Mixed Exponential Model

This model due to Bunea, Bedford and Cooke[20] is a special case of the independent model, where the distributions of X and Z are specified. The distribution for X is assumed to be a mixture of two exponential distributions, while Z is also exponentially distributed. The model can be viewed as a combination of two distinct sub-populations with the same maintenance regime for both. Because this model is a sub-set of the independent model it is unique assuming independence, but this model can not be fitted to every dataset. One of the features of this model shown by Bunea, Bedford and Cooke is that $\Phi(t)$ is minimal at the origin and continually increasing in t.

2.3.3 Dependent Copula Model

The dependent copula model uses a copula to model the dependence between two variables by the use of a single parameter.

By selecting a copula family, it is possible to perform a sensitivity analysis based on the degree of dependence between the two variables. Both Zheng and Klein[84], and Bunea and Bedford[19] presented empirical evidence that the degree of dependence was more important than the choice of copula. Using this approach, Bunea and Bedford[19] studied the sensitivity of optimum maintenance costs with regard to model selection (or more specifically, with regard to the degree of dependence between the two variables). They showed that the optimum maintenance costs varied considerably under different degrees of dependence.

Zheng and Klein[84] showed that the Kaplan-Meier estimator [56] used in the case of the independent model can be generalised to give a consistent estimator based on an assumption about the underlying copula of X and Z. They also showed that assuming the copula is known, the marginal distributions of the resulting model are identifiable.

2.3.4 Random Signs Model

This model, discussed by Cooke[26], assumes that the preventative maintenance is related to the failure time by

$$Z = X - \xi \tag{2.3.1}$$

where the sign of ξ is independent of X. The random signs model says that whether X is censored or not is independent of X (note though that if X is censored then the value of Z may depend on the age of the system).

The idea behind this model is that engineers should seek to perform preventative maintenance as close as possible to the time just before the occurrence of failure. If the engineers are unsuccessful then corrective maintenance is performed instead. The probability of successfully performing this preventative maintenance action is equal to the probability that ξ is positive.

Cooke also proves the following theorem[26] regarding the existence of a random signs model for a given sub-survival pair.

Theorem 1. *Let $\{S_1^*, S_2^*\}$ be a pair of continuous strictly monotonic sub-survival functions; then the following are equivalent:*

There exist random variables ξ and X with the sign of ξ independent of X such that

1.
$$S_1^*(t) = P\{X > t \cap \xi < 0\}, \tag{2.3.2}$$
$$S_2^*(t) = P\{X - \xi > t \cap \xi > 0\}. \tag{2.3.3}$$

2. *For all $t > 0$,*
$$\frac{S_1^*(t)}{S_1^*(t)} > \frac{S_2^*(t)}{S_2^*(t)}. \tag{2.3.4}$$

This theorem says that if Equation 2.3.4 holds then the data set can be described by a random signs model. However, this model will not be uniquely defined because the distribution of ξ given $\xi < 0$ is not knowable and hence the distribution of Z is not identifiable. However, as can be seen from the definition of random signs

$$\frac{S_X^*(t)}{S_X^*(0)} = \frac{P\{X > t, X < Z\}}{P\{X < Z\}} = P\{X > t\} = S_X(t), \tag{2.3.5}$$

so the distribution of X is identifiable. One of the properties of the random signs model is that $\Phi(t)$ is maximal at the origin.

2.3.5 LBL Model

This model presented by Langseth and Lindqvist[59] is a special case of the random signs model. As in random signs, X and the sign of ξ are independent. However, given that $\xi > 0$ and $X = x$, then the LBL model assumes that

$$P\{Z \leq z | Z < X, X = x\} = \frac{H(z)}{H(x)} \qquad 0 \leq z \leq x, \tag{2.3.6}$$

where

$$H(t) = \int_0^t \lambda_X(u)du \tag{2.3.7}$$

and λ_X is the conditional intensity function of X (conditional intensity functions will be introduced in Chapter 3). From this equation it can be seen that the conditional density function of Z given both $Z < X$ and $X = x$ is proportional to the conditional intensity function of X, the underlying failure process. Langseth and Lindqvist further demonstrate that both marginal distributions for

X and Z are identifiable, which is noteworthy given the absence of uniqueness for the marginal distribution of Z in the random signs model. Note that because this model is a sub-set of the random signs model, the Φ function is also maximal at the origin.

2.3.6 Repair Alert Model

This model is proposed by Lindqvist et. al.[63] and is another special case of the random signs model. This model generalises the LBL model by specifying a general function $G(t)$ in place of $H(t)$. $G(t)$ is an increasing function defined on $[0, +\infty)$ and satisfying $G(0) = 0$.

The function $G(t)$ is taken as being reflective of the alertness of the maintenance team. Lindqvist et. al are also able to establish identifiability for the marginal distributions of this model class. Note again that because this model is a sub-set of the random signs model, the Φ function is maximal at the origin.

2.3.7 Random Clipping Model

The random clipping model is discussed by Cooke[23, 24]. This model assumes that X is exponentially distributed and that there is a variable $U = X - W$ where $W > 0$ is a random variable independent from X. U represents the time that a warning signal is given off by some part of the system, to signal that critical failure is imminent.

Cooke showed that conditional on $U > 0$, U has the same distribution as X. This allows these warning signals to be used to estimate important properties which would normally be estimated using X. Clearly this is quite useful given that U could be estimated more directly than X.

The questions of identifiability and whether the ϕ function has a distinctive shape, are still open.

2.3.8 Delay Time Model

The delay time model[22] has been applied to study maintenance within the context of ordinary survival analysis. This model assumes that a warning signal is given by the system at time U, after which there is another period V until the component fails at time $X = U + V$.

Bedford[9] suggested a competing risks version of this model by considering preventative maintenance as $Z = U + W$. W is the time taken by the maintenance crew to observe the equipment in it's degraded state and repair it. It is assumed that U, V and W are mutually independent.

Not much has been done with this model to date, so the question of identifiability is still an open one, however Hokstad and Jensen[52] showed that $\Phi(t)$ is constant under the additional assumption that the latent variables are exponentially distributed.

2.3.9 Alert Delay Model

This model, presented by Dijoux and Gaudion[36], assumes that

$$Z = pX + \epsilon \tag{2.3.8}$$

$$X = pX + (1-p)X \tag{2.3.9}$$

where X and ϵ are exponentially distributed and independent from each other, $p \in [0,1]$. Motivated mainly from the fact that existing models did not fit all the possible data sets generated in practical applications, this model is similar to the delay time model except pX and $(1-p)X$ are not independent. The interpretation of the model is therefore very similar, with pX representing the time that the system issues a warning signal and ϵ representing the delay needed by the maintenance crew to observe the warning and carry out the preventative maintenance. Dijoux and Gaudion showed that the marginal distributions of the alert delay model are identifiable for $p \neq 1$. They also showed that if X and ϵ are both exponentially distributed then $\Phi(t)$ is strictly increasing.

2.3.10 Weiner Process Model

Lindqvist et. al. [62] model the degradation of the system over time as Brownian motion with positive drift μ and variance σt. This is a stochastic process $\{W(t) : t \geq 0\}$ such that

- $W(0) = 0$;
- $\{W(t) : t \geq 0\}$ has stationary and independent increments;
- For all $t > 0$, $W(t) \sim N(\mu, \sigma^2 t)$.

They consider two events, The first event is a signal given off by the system at the first time that the degradation process reaches a certain fixed threshold value (say s). The second event is a system failure which occurs at the first time that the degradation process reaches a second, higher fixed threshold value (say c). The first model uses a constant probability q that the system maintainers

will observe the signal given off by the system and perform preventative maintenance. If the signal is not observed then the degradation process continues on until the second threshold is first reached, at which point a critical failure occurs and corrective maintenance must be performed. This model is extended to consider a random level s for the signal given off by the system. In both cases, sub-distribution functions and likelihood functions are derived.

This is an alternative approach to most of the other models because in this case the underlying degradation of the system is explicitly modelled and this is what is used to describe the dependence between preventative and corrective maintenance. The idea of signals that can act as triggers for the observant maintenance engineer that spots them is the same idea behind the random signs model. Indeed Lindqvist et. al. show that the basic model with fixed signal threshold is a random signs model.

2.3.11 Model Selection

A useful property of the model classes discussed in this section is that several of them carry a different "signature" in the shape of their respective Φ functions. A dataset can be used to directly estimate the Φ function and the resulting shape can be studied to reject certain model classes from consideration. Indeed, Dijoux and Gaudion[36] cite this as one of the primary motivations for developing the Alert Delay model (subsection 2.3.9). This is a useful first step during model selection, although several models classes are too general to allow for distinct "signature" shapes in their Φ functions, such as the Copula model and the Weiner process model. There are also several models in the list that remain under-developed, such as the delay time model and the random clipping model.

2.4 Summary

In this chapter the classical competing risks framework has been introduced. The necessary notation was developed and the identifiability problem was also introduced. Several existing models developed within this framework were also discussed. The next chapter will introduce the idea of imperfect repair models in the case of univariate survival analysis. The classical competing risks models developed in this chapter and the univariate maintenance models developed in the next chapter will then be used in Chapter 4 when introducing the generalised competing risks

framework.

Chapter 3

Univariate Maintenance Models

As discussed in Section 2.1, classical competing risks models assume the renewal property holds. In practice this property is often too restrictive for many situations faced in reliability. This section will begin with a brief introduction to point processes, defining the various pieces of notation that will be needed in order to discuss the maintenance models that follow. Several maintenance models developed within survival analysis will be presented and where applicable their previous use within the field of competing risks will be discussed. Particular emphasis will be placed on the work of Doyen and Gaudoin in both generalising maintenance models to competing risks[41] and in their construction of two classes of maintenance models, called *arithmetic reduction of virtual age* and *arithmetic reduction of intensity*[40]. For a more general review of maintenance models see either [74] or [61]. A thorough text on the subject is [5].

One of the problems inherent in classical competing risks and discussed in Section 2.1, is the identifiability problem. It will be seen that this problem is not always present when one considers alternative maintenance models to that of complete renewal. For example, Bedford and Lindqvist[11] consider a series system of several components under complete renewal, minimal repair and partial repair (minimal repair will be defined shortly and partial repair will be defined in section 5.3). They demonstrate that the marginal survival distributions for each of the components in this model are identifiable for partial repair but not for minimal repair or complete renewal. The possibility that identifiability problems are not present when considering certain alternative maintenance models to that of complete renewal will be discussed further in Chapter 5.

3.1 Point Processes

A *point process* is a set $\{N(t)\}$ of random variables referenced by $t \in [0, \infty)$, where $N(b) - N(a) = N(a, b]$ is a finite non-negative integer $\forall b > a \geq 0$ with probability 1, and $N(0) = 0$.

For this chapter a distinction will not be made between corrective and preventative maintenance. The number of events that have occurred by time t will be denoted by a point process $N(t)$. The corresponding arrival times will be denoted by $Y_1, Y_2, \ldots$. The time between the i^{th} and $(i+1)^{th}$ arrivals will be denoted by X_i. The history of the process up to time t will be denoted by H_t. This history is simply the value of $N(y)$ for $0 < y < t$. The distribution of the point process $N(t)$ can be completely defined by specifying the *conditional failure intensity function* (often just referred to as the intensity function)

$$\lambda_t = \lim_{\delta \to 0} \frac{P\{N(t, t+\delta] = 1 | H_t\}}{\delta}. \tag{3.1.1}$$

It is assumed before the first failure that the conditional failure intensity function is equal to the *initial intensity function* $\lambda(t)$ which is a known continuous function of t. For ease of exposition the conditioning on the history of the point process will be suppressed in the notation from now on. In all the models considered here, repairs are assumed to take negligible time. With this terminology in place it is now possible to describe complete renewal in terms of point processes. The conditional failure intensity for complete renewal is

$$\lambda_t = \lambda(t - Y_{N_t}). \tag{3.1.2}$$

As can be seen, this means that the conditional failure intensity is reset to $\lambda(0)$ every time an event occurs.

Another very common maintenance model is the minimal repair model mentioned earlier, which is sometimes called "as bad as old". This is where the system is restored to a functioning state with the same reliability characteristics as a system of the same age as this system was at the time of failure. The conditional failure intensity for minimal repair is just

$$\lambda_t = \lambda(t). \tag{3.1.3}$$

For a comprehensive treatment of point processes within the context of reliability see [45] or [81].

3.2 Brown and Proschan Model

This model is presented in [18] and assumes that when a system is repaired at failure it is restored to one of two states. With probability p the system is restored to "as good as new" (the familiar complete renewal), with probability $1-p$ the system is restored to "as bad as old" (minimal repair). In this way the age of the system is modelled as a sequence of repeated "as bad as old" repairs until the system age is reset by an "as good as new" repair. Clearly if $p = 1$ then this is just the classical renewal process. A more general imperfect repair model was considered by Block et. al.[16] where they allowed the probability p to be time dependent.

The Brown and Proschan model was extended to competing risks by Langseth and Lindqvist[59]. They combined this model with the repair alert model, with a view to treating the parameter p of the Brown and Proschan model (along with a parameter from the repair alert model) as a measure of the efficiency of the maintenance crew.

Renewal processes are appropriate when considering a system which is completely repaired each time maintenance is carried out or a failure occurs. Minimal repair is a good approximation when dealing with a system consisting of lots of different components and only one or a small number of components are repaired. The Brown and Proschan model contains both these cases, allowing for the possibility that repairs are either to one, a small number or all of the system components. It should be noted that the possibility of a significant number of system components being repaired is not allowed in this model. For example if half the components in the system were repaired while the remaining half were not. This model does not go much further then, than the classical competing risks set-up.

3.3 Kijima Models

The Kijima model was discussed in [57] and uses the idea of a virtual age process. If the system has the virtual age $V_{n-1} = y$ immediately after the $(n-1)^{th}$ repair, then for time $t > 0$, the n^{th} failure-time X_n is assumed to have the survival function

$$P\{X_n > t\} = P\{X > y + t | X > y\} = \frac{S(y+t)}{S(y)}, \tag{3.3.1}$$

where $S(t)$ is the survival function of the failure-time of a new system and it is assumed that $V_0 = 0$. The quality of general repair is represented as a sequence of random variables A_n taking a value

between 0 and 1, where A_n denotes the quality of the nth repair. Two models are constructed, referred to as model I and model II respectively

I. $V_n = V_{n-1} + A_n X_n$;

II. $V_n = A_n(V_{n-1} + X_n)$.

The first model does not allow a repair to improve on the work of previous repairs. The n^{th} repair can only restore the system to a state as good as the state immediately after the $(n-1)^{th}$ repair. The second model does allow this and in this case the value of A_n reflects the extent and quality of the repair on the whole system. A repair in the second model is a minimal repair when $A_n = 1$ and a perfect repair when $A_n = 0$.

This model class can be applied to a rich set of situations, greatly expanding the domain of problems that can be modelled with the Brown and Proschan model above.

The A_i are usually treated as random variables and a specific parametric form is chosen. This allows for inclusion of parameters that can be estimated from the data and can yield useful information about the quality of the maintenance work performed, much like the p parameter in the Brown Proschan model.

3.4 Dorado-Hollander-Sethuraman Model

This model was developed by Dorado, Hollander and Sethuraman in [39]. The model is based around the following two sequences $\{A_j\}_{j\geq 1}$ and $\{\Theta_j\}_{j\geq 1}$, these being the effective ages and life supplements respectively.

The effective ages are such that $A_1 = 0$ and $A_j \geq 0$. The life supplements are such that $\Theta_1 = 1$ and $\Theta_j \in (0, 1]$. Also the following inequality holds to ensure that the effective ages are always less than the actual ages

$$A_j \leq A_{j-1} + \Theta_{j-1} X_{j-1}, \tag{3.4.1}$$

where the X_j's are the inter-arrival times as before. The model class is fixed by considering joint survivor functions of the inter-arrival times as follows

$$P\{X_j > t | A_j, \Theta_j, X_1, \ldots, X_{j-1}\} = S_{\Theta_j}^{A_j}(t) = \frac{S(\Theta_j t + A_j)}{S(A_j)}. \tag{3.4.2}$$

Both the Brown-Proschan and Kijima models can be considered as special cases of this model. The Kijima model is recovered by specifying all $\Theta_j = 1$, the Brown-Proschan model is recovered by specifying all $\Theta_j = 1$ and by restricting each A_j to taking a value of either 0 or 1. Effectively, this model is similar to the virtual age reduction models like the Kijima model, except rather than just reducing the actual age by an additive sequence, the actual age is reduced by a linear equation.

The strength of this model is that it generalises other models such as the Kijima models. However, this is at the cost of increased complexity in the way that the reduction of the actual age is calculated. It is necessary to specify the distributions for both sequences, $\{A_j\}_{j\geq 1}$ and $\{\Theta_j\}_{j\geq 1}$. The issue of how to make sensible assumptions about parametric forms for two different sequences, neither of which is observed in isolation, is certainly not trivial. It is certainly easier to imagine how one might make useful assumptions about the parametric form of a single sequence, as in the Kijima model. The issue of how to separate out the scaling and shifting effects of the impact of maintenance on the virtual age of the system is not treated by Dorado, Hollander and Sethuraman.

3.5 Malik's Proportional Age Reduction Model

This model devised by Malik[66] is a subset of the Kijima model. Malik proposed the idea of an "improvement factor" described by a parameter ρ. This factor described the extent to which a system had been restored to it's original state. Using the notation established for the Kijima model, the virtual age V_{i+1} is defined as

$$V_{i+1} = V_i + \rho\,(Y_{i+1} - Y_i)\,, \tag{3.5.1}$$

with $V_0 = 0$. The conditional failure intensity is

$$\lambda_t = \lambda(t - \rho Y_{N_t}). \tag{3.5.2}$$

Shin, Lim and Lie[79] developed a procedure to estimate ρ for certain parametric families of non-homogeneous poisson processes using Malik's proportional age reduction model.

3.6 Trend Renewal Processes

In [43] Elvebakk, Lindqvist and Heggland discuss a new model which generalises the following well known property of non-homogeneous Poisson processes. If $T_1, T_2 \ldots$ is a non-homogeneous Poisson process with intensity $\lambda(t)$, then the stochastic process $\Lambda(T_1), \Lambda(T_2), \ldots$ is a homogeneous Poisson process (where $\Lambda(t)$ is just the cumulative intensity defined as $\Lambda(t) = \int_0^t \lambda(s)ds$).

Trend-renewal processes are defined as any stochastic process for which the corresponding process transformed by $\Lambda(t)$ is a renewal process, defined by some distribution function $F(t)$ (to obtain uniqueness, $F(t)$ is normally restricted to having an expected value equal to 1). In this way the trend renewal process is defined by specifying the intensity function $\lambda(t)$ and the distribution of the time transformed renewal process.

The idea of this model is that given a system where a major component is repaired but none of the surrounding components are, it may seem reasonable to model this as a renewal process and then simulate the effect of the aging environment by speeding up the rate at which the system ages with a time transformation as above.

Lindqvist says that a noticeable difference between trend-renewal processes and virtual age models "is that the virtual age type model usually requires that the virtual age process is observable. Such observations may, however, often be lacking in real data"[61]. This is not a problem for trend renewal processes which do not use virtual ages.

3.7 Generalized Proportional Intensity Model

This model is discussed in [72] and is defined by the following equation for the intensity function

$$\lambda_t = \lambda(t) \left\{ \prod_{i=1}^{U(t)} r_i \right\} \left\{ \prod_{j=1}^{V(t)} s_j \right\} exp(x(t)\gamma). \tag{3.7.1}$$

Considering each part in turn, $\lambda(t)$ is a baseline intensity function as usual. The exponential term is a standard way of including regression in the model, where $x(t)$ is a vector of explanatory variables and γ is an unknown parameter vector of regression coefficients. This regression term allows for the estimation of a system's general properties using data from several identical systems by including information about the variation in the operating conditions and environment of each system. Note that similar regression terms can be included for this purpose in all of the models of this chapter.

Ignoring the regression term, what remains is a multiplicative scaling of the initial intensity function. In this way these models can be seen as a special case of the Dorado, Hollander and Sethuraman model presented above where the A_i are all zero. The Θ_i are divided into two classes, r_i and s_j, which denote the intensity scaling factors for preventative and corrective maintenance respectively. $U(t)$ and $V(t)$ are two point processes that count the occurrences of preventative and corrective maintenance respectively. The minimal repair model can be recovered by ignoring the regression term and setting all the r_i and s_j equal to one.

3.8 Chan and Shaw Model

An idea put forward by Chan and Shaw[21] was that the effect of maintenance can be viewed as reducing the intensity of the failure process. Given a function describing the initial intensity, if a maintenance event was to occur at a time t, then the change to the intensity function could be modelled as a proportional reduction to the intensity of the failure process for the time after t. This equates to the following

$$\lambda_t = (1 - \rho)\lambda(t). \tag{3.8.1}$$

This model was not followed up by Chan and Shaw, they merely proposed it.

3.9 Arithmetic Reduction of Intensity Models

The next two subsections will discuss the work by Doyen and Gaudoin[40] in constructing two classes of maintenance models which contain several of the models already discussed, as well as presenting the possibility of several new models. The first of these two model classes is the arithmetic reduction of intensity model class (hereafter referred to as ARI models). The following is a description of the general structure of the ARI models. Let B be a function of the previous failure history before time t (note that B is restricted as to the form that it can take).

$$\lambda_t = \lambda(t) - \sum_{i=1}^{N_t} B(i, T_1, \cdots, T_i). \tag{3.9.1}$$

The conditional failure intensity of the ARI(1) model has the following form

$$\lambda_t = \lambda(t) - \rho\lambda(Y_{N_t}). \tag{3.9.2}$$

The conditional failure intensity of the ARI(m) model has the form

$$\lambda_t = \lambda(t) - \rho \sum_{j=0}^{Min(m-1,N_t-1)} (1-\rho)^j \lambda(Y_{N_t-j}). \tag{3.9.3}$$

And lastly, the conditional failure intensity of the ARI(∞) model has the form

$$\lambda_t = \lambda(t) - \rho \sum_{j=0}^{N_t-1} (1-\rho)^j \lambda(Y_{N_t-j}). \tag{3.9.4}$$

The idea proposed by Chan and Shaw in Subsection 3.8 corresponds to the ARI(∞) model above. One of the interesting properties of both model classes is the parameter of the model class (m in the case of ARI models). For both model classes Doyen and Gaudoin describe these parameters as the amount of memory possessed by each maintenance model. The main motivation for the construction of these model classes is that like the Brown and Proschan model discussed earlier, there is a parameter which expresses the efficiency of the maintenance carried out.

3.10 Arithmetic Reduction of Virtual Age Models

The second model class in Doyen and Gaudoin's[40] classification scheme is the arithmetic reduction of virtual age model class (hereafter referred to as ARA models). The following is a description of the general structure of the ARA models. Let C be a function of the previous failure history before time t (note that C is restricted as to the form that it can take).

$$\lambda_t = \lambda(t - \sum_{i=1}^{N_t} C(i, T_1, \cdots, T_i)). \tag{3.10.1}$$

The conditional failure intensity of the ARA(1) model has the following form

$$\lambda_t = \lambda(t - \rho Y_{N_t}). \tag{3.10.2}$$

The conditional failure intensity of the ARA(n) model has the form

$$\lambda_t = \lambda(t - \rho \sum_{j=0}^{Min(n-1, N_t-1)} (1-\rho)^j Y_{N_t-j}). \tag{3.10.3}$$

And lastly, the conditional failure intensity of the $ARA(\infty)$ model has the form

$$\lambda_t = \lambda(t - \rho \sum_{j=0}^{N_t-1} (1-\rho)^j Y_{N_t-j}). \tag{3.10.4}$$

Malik's proportional age reduction model corresponds to an ARA(1) model, as do the complete renewal and minimal maintenance models (with ρ equal to 1 and 0 respectively). A subset of the Kijima models (those for which all the A_i are equal to $1 - \rho$) corresponds to an ARA(n) model. The weakness of both the ARA and ARI model classes is that ρ is a quantity that can change from event to event. ρ will effectively only be an average of the efficiency of the maintenance carried out. It is also conceivably possible for there to be a systematic non-zero trend in the behaviour of ρ. For example, maintenance staff getting better at their job through time. It is not clear how useful a point estimate is in such cases due to the lack of ability to model the trend in ρ over time.

3.11 Summary

These models demonstrate means by which imperfect repair can be incorporated into reliability modelling. In the next chapter a framework developed by Doyen and Gaudoin[41] for incorporating univariate maintenance models into a competing risks set-up will be presented. Several models discussed in this section will be seen to fit within this framework. However, it should be noted that this is not the only possible means of extending competing risk models to include imperfect repair. For example, the Dorado-Hollander-Sethuraman model could be used to develop a similar framework.

Chapter 4

Generalised Competing Risks (GCR)

In this section the class of generalised competing risk models is introduced. This class can be used to model failure data for a system with multiple failure modes as in the classical case, but the renewal assumption can be relaxed. This allows for more generality and a wider range of applications. The classification presented here is based on that by Doyen and Gaudoin[41]. This work certainly seems to be the most advanced generalisation of the competing risks framework by using maintenance models. However there are several other maintenance models which are more general and may also be capable of being used to extend the competing risks framework.

After defining the generalised competing risks framework, attention will be restricted to those models which employ virtual ages to describe the maintenance effects. In order to build a generalised competing risks model using virtual age maintenance models it is necessary to specify a classical virtual age model for both corrective maintenance and preventative maintenance effects. It is also necessary to describe the dependency between corrective maintenance and preventative maintenance by using a classical competing risks model.

To date, the only model really explored in this framework is that of the conditionally independent generalised competing risks model, which will be discussed in Section 4.4.

4.1 Basic Model Structure

Thus far X and Z have been treated as random variables under classical competing risks. In order to allow for imperfect repair, it is necessary to keep track of the order in which events have occurred. To this end, for all $k \geq 1$, the time to the next corrective maintenance after the k^{th} maintenance action is denoted by X_{k+1} and the time to the next preventative maintenance after the k^{th} maintenance action is denoted by Z_{k+1}. As before, only one of these events will be observed and this is denoted by $T_{k+1} = \min(X_{k+1}, Z_{k+1})$. Which one of these events is observed is denoted by

$$\delta_{k+1} = \begin{cases} 0 & \text{if the } (k+1)^{th} \text{ maintenance is a corrective maintenance event } (X_{k+1} < Z_{k+1}) \\ 1 & \text{if the } (k+1)^{th} \text{ maintenance is a preventative maintenance event } (Z_{k+1} \leq X_{k+1}). \end{cases} \tag{4.1.1}$$

X_1, Z_1, T_1 and δ_1 are all defined as in classical competing risks. Figure 4.1 is a graphical representation of a GCR model. The upper part of the diagram shows the observable quantities T and δ. The lower part shows the quantities X and Z, with the grey ending of the arrows indicating the fact that the actual values are censored and all that is known is that the quantities are at least as large as the minimum of the two latent variables in each case.

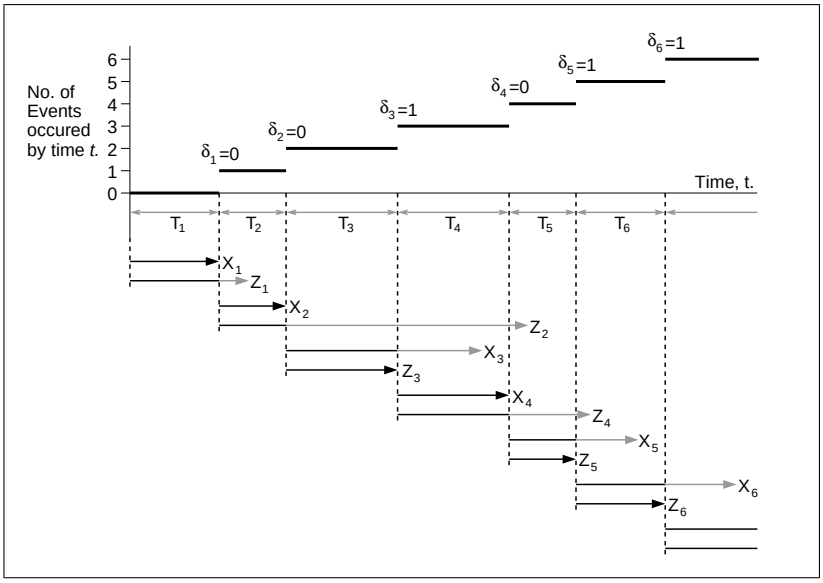


Figure 4.1: An Example GCR Model.

The generalised sub-survival functions are defined as

$$S^*_{X_{k+1}}(x|\mathbf{T}_k, \boldsymbol{\delta}_k) = P\{X_{k+1} > x, X_{k+1} < Z_{k+1}|\mathbf{T}_k, \boldsymbol{\delta}_k\} \tag{4.1.2}$$

$$S^*_{Z_{k+1}}(z|\mathbf{T}_k, \boldsymbol{\delta}_k) = P\{Z_{k+1} > z, Z_{k+1} \leq X_{k+1}|\mathbf{T}_k, \boldsymbol{\delta}_k\}, \tag{4.1.3}$$

where the bold symbols $\mathbf{T}_k, \boldsymbol{\delta}_k$ represent vectors of length (k) containing the previous random variables $T_1, \ldots, T_k$ and $\delta_1, \ldots, \delta_k$. The generalised conditional marginal survivor functions are defined in a similar manner

$$S_{X_{k+1}}(x|\mathbf{T}_k, \boldsymbol{\delta}_k) = P\{X_{k+1} > x|\mathbf{T}_k, \boldsymbol{\delta}_k\} \tag{4.1.4}$$

$$S_{Z_{k+1}}(z|\mathbf{T}_k, \boldsymbol{\delta}_k) = P\{Z_{k+1} > z|\mathbf{T}_k, \boldsymbol{\delta}_k\}. \tag{4.1.5}$$

Lastly, the generalised conditional joint survivor functions are defined as

$$S_{k+1}(x, z|\mathbf{T}_k, \boldsymbol{\delta}_k) = P\{X_{k+1} > x, Z_{k+1} > z|\mathbf{T}_k, \boldsymbol{\delta}_k\}. \tag{4.1.6}$$

Again, $S^*_{X_1}$, $S^*_{Z_1}$, S_{X_1}, S_{Z_1} and S_1 are all defined as in the classical case.

4.2 Symmetric GCR Models Using Virtual Ages (Symmetric GVA)

Symmetric GCR models using virtual ages are a particular sub-class of GCR models which make certain assumptions about the relationships between the first set of random variables (X_1, Z_1, T_1 and δ_1) and the remaining sets of variables. The model interpretation in this context is that after the k^{th} maintenance the system behaves like a new system that has survived without failure to the effective age A_k. It can be seen that classical competing risk models are a special case of symmetric GVA models where the virtual ages are all zero.

For symmetric GVA models, the marginal distributions for the time to first failure are also not identifiable, as in the case of classical competing risks. A useful extension of the Peterson bounds (introduced in Section 2.2 on Page 20) from classical competing risk models to symmetric GVA models is given in Appendix C.

Assumption 1 (The Virtual Age property). *There exists a sequence of effective ages* $\{A_k\}_{k\geq 1}$,

$A_0 = 0$ *such that immediately after the* k^{th} *maintenance the system behaves like a new system that has survived without failure to the effective age* A_k*, where* A_k *is a deterministic function of* $(\boldsymbol{T}_k, \boldsymbol{\delta}_k)$.

For this thesis, the condition that the effective ages are deterministic functions of $(\mathbf{T}_k, \boldsymbol{\delta}_k)$ is too restrictive. More general functions will be allowed in chapter 5, however the resulting models will still be described as belonging to the class of GCR models using virtual ages.
Given the generalised joint survival function for the first failure, the virtual age property means that the remaining generalised joint survival functions can be described in terms of the first one and the effective ages after each failure. The model as described above is stated in probability language as follows

$$P\{X_{k+1} > x, Z_{k+1} > z|\mathbf{T}_k, \boldsymbol{\delta}_k\} = P\{X > A_k + x, Z > A_k + z|X > A_k, Z > A_k, A_k\}. \quad (4.2.1)$$

Where (X, Z) is a random couple with the same distribution as (X_1, Z_1). The joint survival function is then

$$S_{k+1}(x, z|\mathbf{T}_k, \boldsymbol{\delta}_k) = \frac{S_1(A_k + x, A_k + z)}{S_1(A_k, A_k)}. \quad (4.2.2)$$

4.3 Asymmetric GCR Models Using Virtual Ages (Asymmetric GVA)

Asymmetric GCR models using virtual ages extend the idea of symmetric GVA models by considering distinct sequences of effective ages, one for each failure mode. This concept is difficult to ascribe a meaning in the context of preventative and corrective maintenance, so use shall be restricted to modelling systems of n components where the i^{th} failure mode corresponds to the failure of the i^{th} component. The model interpretation in this context is that after the k^{th} maintenance, the components will experience different repair actions and will therefore each behave like a new component that has survived without failure to a different effective age depending on the component in question.

Assumption 2 (The asymmetric Virtual Age property)**.** *There exist* m *sequences of effective ages* $\left\{A_k^i\right\}_{k\geq 1}$, $A_0^i = 0$ *(for all* i *from* 1 *to* m*) such that after the* k^{th} *maintenance the system behaves like a similar system whose components have survived without failure to the corresponding effective*

ages $(A_k^1, \ldots, A_k^m)$*, where the* $A_k^i s$ *are deterministic functions of* $(\boldsymbol{T}_k, \boldsymbol{\delta}_k)$*.*

Again, given the generalised joint survival function for the first failure, the asymmetric virtual age property means that the remaining generalised joint survival functions can be described in terms of this first one and the effective ages after each failure. The model as described above is stated in probability language as follows

$$\begin{aligned} &P\{Z_{k+1}^1 > z^1, \ldots, Z_{k+1}^m > z^m | \mathbf{T}_k, \boldsymbol{\delta}_k\} \\ &= P\{Z^1 > A_k^1 + z^1, \ldots, Z^m > A_k^m + z^m | Z^1 > A_k^1, \ldots, Z^m > A_k^m, A_k^1, \ldots, A_k^m\}. \end{aligned}$$

Where $(Z^1, \ldots, Z^m)$ is a random m-tuple with the same distribution as $(Z_1^1, \ldots, Z_1^m)$. The joint survival function is then

$$S_{k+1}(z^1, z^2, \ldots, z^m | \mathbf{T}_k, \boldsymbol{\delta}_k) = \frac{S_1(A_k^1 + z^1, \ldots, A_k^m + z^m)}{S_1(A_k^1, \ldots, A_k^m)}.$$

4.4 The Conditionally Independent Generalised Competing Risks Model (CIGCR)

Dijoux, Doyen and Gaudoin[35] have applied their generalised competing risks framework to what they call a conditionally independent generalised competing risks model. Given a sequence of pairs of variables $\{(X_k, Z_k)\}_{k \geq 1}$, these pairs are said to be conditionally independent if the following property holds for all $k \geq 0$, $\forall x \geq 0$ and $\forall z \geq 0$:

$$S_{k+1}(x, z | \mathbf{T}_k, \boldsymbol{\delta}_k) = S_{X_{k+1}}(x | \mathbf{T}_k, \boldsymbol{\delta}_k) S_{Z_{k+1}}(z | \mathbf{T}_k, \boldsymbol{\delta}_k). \tag{4.4.1}$$

Using this model, they study the case when the conditional marginal survivor functions are exponentially distributed and develop closed form expressions (using maximum likelihood) for the estimates of the parameters. This exponential model has four parameters, the two initial failure rates of the system when it is as good as new, plus two multiplicative constants that represent the delay of corrective maintenance through the occurrence of preventative maintenance and the acceleration of preventative maintenance through the occurrence of corrective maintenance.

4.5 Summary

In this chapter the work done by Doyen and Gaudoin to generalise the competing risk framework using virtual age models has been introduced. As mentioned in the conclusion of Chapter 3, it is also possible to use other types of maintenance models to generalise the classical case of complete renewal, but this possibility is not discussed any further in this thesis.

In the first chapter of the second part of this thesis, several asymmetric GCR models using virtual ages will be examined. Expanding on the work done by Bedford and Lindqvist[11], the joint survival distributions of these specific model classes will be shown to be identifiable given a single realisation of the process, under certain conditions.

Part II

Results and Examples

Chapter 5

Identifiability of Several Asymmetric GVA Models

When introducing asymmetric GVA models in Chapter 4, it was noted that in order for an asymmetric GVA model to be constructed it was necessary to choose particular virtual age models for each of the competing risk variables, as well as the dependencies between those variables. In this chapter several asymmetric GVA models will be presented with assumptions made about the maintenance regime that allow the virtual age models to be fully described. The aim of this chapter will be to demonstrate that in these particular cases (subject to some technical constraints) it is not necessary to make assumptions about the nature of the dependencies between the risk variables in order to have a uniquely identifiable joint distribution for the model from a single realisation of the process. The work presented in this chapter is a series of extensions to the work done by Bedford and Lindqvist[11] which was mentioned in Chapter 3.

This chapter will begin by sketching some of the preliminary theory that will prove useful in demonstrating identifiability. Following this, the two different models will each be introduced and formally defined. In each case the interpretation of the various model components will be discussed and a proof of identifiability of the joint distribution will be given.

5.1 Identifiability

In Section 2.2 the identifiability problem for the marginal distributions in classical competing risks was discussed. Doyen and Gaudoin[41] pointed out that the symmetric GVA framework (introduced in Section 4.2) is also subject to the problem of non-identifiability. However, the work of Bedford and Lindqvist[11] demonstrates that certain models classes which are sub-classes of the asymmetric GCR framework using virtual ages (introduced in section 4.3) are identifiable, in that the joint distribution is identifiable from a single realisation of the process. Note that in the case of classical competing risks, the identifiability of all the marginal distributions is equivalent to the identifiability of the joint distribution.

This section will present a justification for these different types of behaviour. Attention will be restricted to GVA models of a two component system using virtual ages (both symmetric and asymmetric models will be considered). In these cases the state of the system at a given moment in time can be expressed by the virtual ages of the two components, V_0 and V_1. The state space of the system can then be represented as a point on the following diagram, with the two axes denoting the virtual ages of the two components.

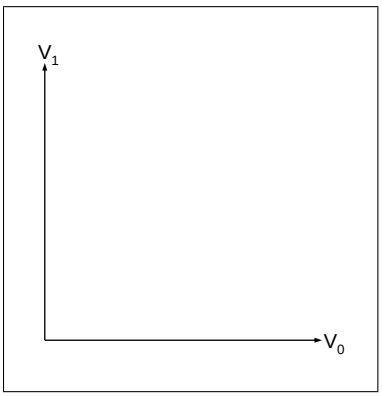


Figure 5.1: Template for the State Space Diagram.

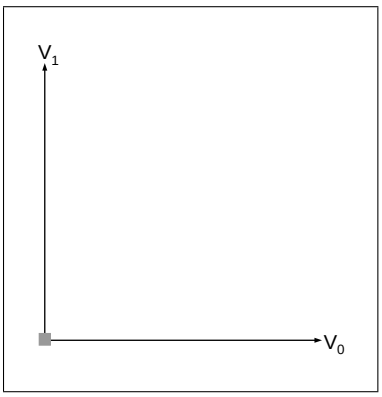
These diagrams will be used extensively throughout this chapter to discuss the behaviour of different classes of competing risk models. First consider the simplest case, that of classical competing risks. In this case the system starts in a state of "as good as new" which means that the virtual ages of both of the components are zero (this is depicted in Figure 5.2(a)). As the system is in use, the two virtual ages increase with time (after an interval of time t_0, the two virtual ages of the components will have increased by t_0). This implies that at any given elapsed time t since the

previous maintenance event, both V_0 and V_1 equal t (this is depicted in Figure 5.2(b)). When an event occurs the system is again completely restored, setting both the virtual ages back to zero (this is depicted in Figure 5.2(c)). This process will continue in the same way. It can be seen that the state of the system is then restricted to following the path of the diagonal line defined by $V_0 = V_1$ (this line is depicted in Figure 5.2(c)). The joint survival function of the two latent variables representing the current age of each of the components, takes values on the whole positive quadrant represented in the diagram. However it is clear from the above discussion that the state of the system will never leave the diagonal line. Thus the set of potential joint survival functions consists of all those that agree with the observable data on that line. The behaviour of such joint survival functions is unrestricted by the observable data outside that set. Thus, given any particular infinite sequence or path through the state space (often referred to as a single realisation of the process), such a sequence would be found to be in agreement with a large number of possible joint survival functions (which is exactly the problem of non-identifiability).

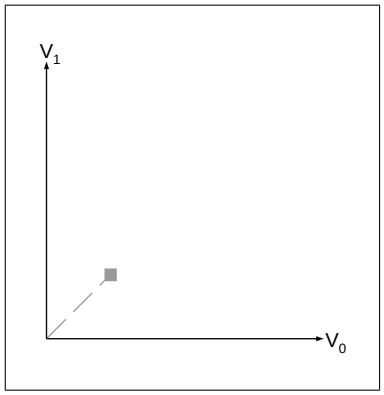
In the case of the symmetric GCR framework using virtual ages, both virtual ages start at zero (this is depicted in Figure 5.3(a)). As the system is in use, the two virtual ages increase in time such that at a given elapsed time t since the previous maintenance event, both V_0 and V_1 equal t (this is depicted in Figure 5.3(b)). When an event occurs the system is repaired, but not necessarily completely restored. All that is assumed is that the virtual ages of the two components are still equal after the maintenance activity. This means that the state of the system will still be represented by some point (a, a) on the diagonal line $V_0 = V_1$ (this is depicted in Figure 5.3(c)). The process will continue in a similar fashion, with the two virtual ages increasing with time at the same rate as before, such that at some time $b - a$ later the system state is represented by the point (b, b) (this is depicted in Figure 5.3(d)).

Again this pattern of behaviour will restrict the possible states of the system to those on the diagonal line and the set of potential joint survival functions will consist of all those that agree with the observable data on that line. The behaviour of such joint survival functions will be unrestricted by the observable data outside that set. Thus, given any particular infinite sequence or path through the state space (often referred to as a single realisation of the process), such a sequence would be found to be in agreement with a large number of possible joint survival functions (which is exactly the problem of non-identifiability).

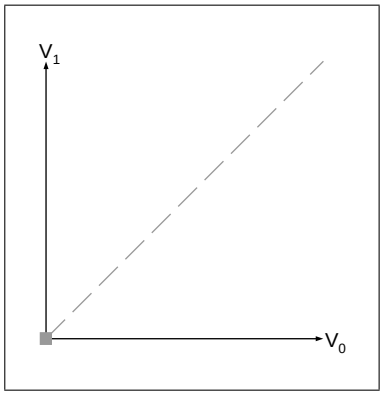
In the case of the asymmetric GCR framework using virtual ages, the possible range of behaviour



(a) The Starting State of the Model.



(b) The state of the Model After Time t.

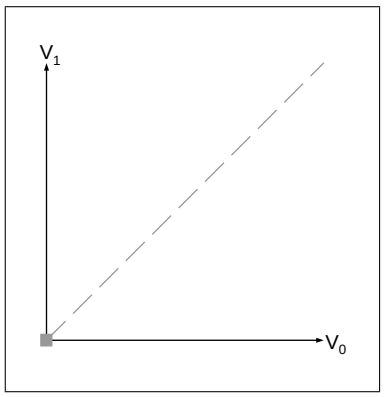


(c) The state of the Model Immediately After the Occurrence of a Maintenance Activity.

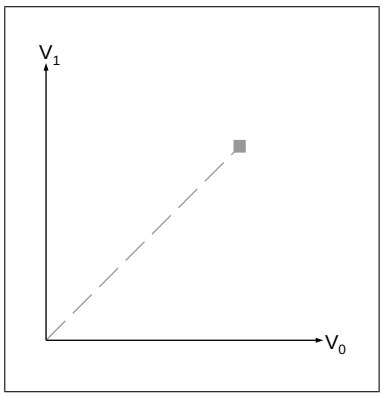
Figure 5.2: State Space Diagrams Depicting the Behaviour of the Classical Competing Risks Framework.

is quite different. It will be seen in Sections 5.3 and 5.4 that in the two asymmetric GCR models considered, both cases allow for the possibility that the system state can potentially reach any point in the entire state space. Consider the simple case of the Bedford and Lindqvist model[11] for two components, discussed in more detail in Section 5.3. In this case the maintenance strategy applied is that the component that fails is perfectly repaired, while the other component is minimally repaired. Figure 5.4 is a sequence of state space diagrams depicting an example realisation of this model.

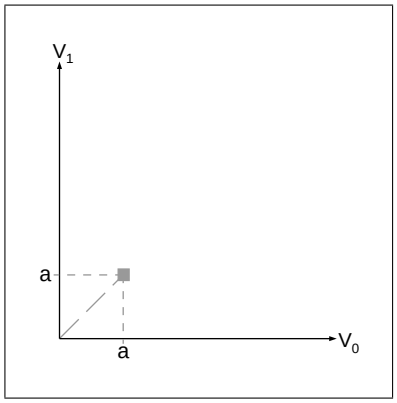
Both virtual ages start at zero (this is depicted in Figure 5.4(a)). As the system is in use, the two virtual ages increase in time such that at a given elapsed time t since the previous maintenance event, both V_0 and V_1 equal t (this is depicted in Figure 5.4(b)). When one of the components fails, the component corresponding to V_0 say, this component is completely restored and $V_0 = 0$.



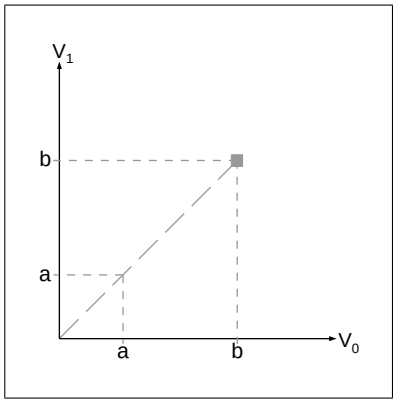
(a) The Starting State of the Model.



(b) The State of the Model After Time t.



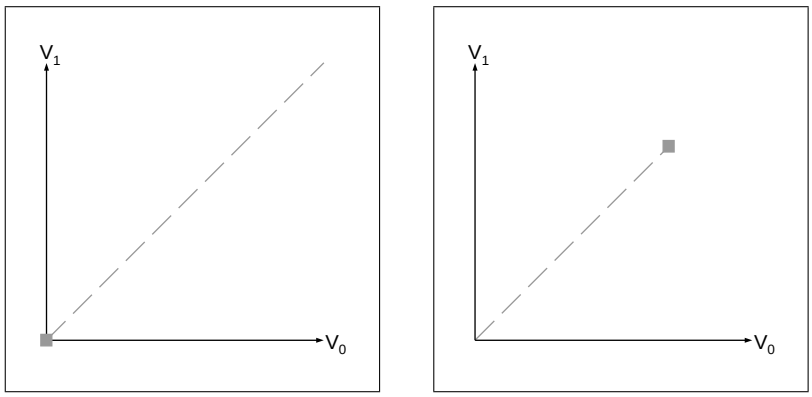
(c) The State of the Model Immediately After the Occurrence of a Maintenance Activity.



(d) The State of the Model After a Further Time $b - a$.

Figure 5.3: State Space Diagrams Depicting the Behaviour of the Symmetric GCR Framework Using Virtual Ages.

However, V_1 remains unaffected by the maintenance activity and the resulting system state is no longer confined to the set of states with $V_0 = V_1$ (this is depicted in Figure 5.4(c)). The process will continue with each component aging with time and the path traced by the state of the system will be parallel to the line described by $V_0 = V_1$, but not on that line (this is depicted in Figure 5.4(d)). This behaviour means that the the set of potential joint survival functions will consist of all those that agree with the observable data but the observable data can now potentially occur at any point in the state space diagram, not just on the diagonal line $V_0 = V_1$. To explain why this leads to the identifiability of the joint survival function, it will be necessary to utilise the notion of ergodicity introduced in the next section.



(a) The Starting State of the Model.

(b) The State of the Model After Time t.

(c) The State of the Model Immediately After the Occurrence of a Maintenance Activity.

(d) The State of the Model After a Brief Further Time.

Figure 5.4: State Space Diagrams Depicting the Behaviour of the Asymmetric GCR Framework Using Virtual Ages.

5.2 Ergodicity of Discrete Time Markov Chains With a General State Space

The purpose of this section is to introduce some preliminary theory about ergodicity of discrete-time Markov chains with a general state space. Some basic results from Doob[38] and a result from a paper by Bedford and Lindqvist[11] will be presented. This theory will be used in identifiability proofs for several sub-classes of the asymmetric GVA framework in later sections.

5.2.1 Some Notation and Conventions

Consider a set X of possible states ξ, with a Borel field $\mathscr{F}_X$ over X. The set X will be referred to as a *state space*) and the states ξ will be referred to as *points* of that state space. The function $p : X \times \mathscr{F}_X \longmapsto [0,1]$ will be called a *stochastic transition function* if $\forall \xi \in X$ and $\forall A \in \mathscr{F}_X$:

I. $p(\xi, A)$ for fixed ξ determines a probability measure on the set A.

II. $p(\xi, A)$ for fixed A determines a function of ξ that is measurable with respect to the field $\mathscr{F}_X$.

The n-step transition probabilities (which are also stochastic transition functions) are given by

$$\begin{aligned} p^{(1)}(\xi, A) &= p(\xi, A) \\ p^{(n+1)}(\xi, A) &= \int_X p^{(n)}(\eta, A) p(\xi, d\eta). \end{aligned}$$

The set Ω will denote the set of all sequences $\omega = (\xi_1, \xi_2, \ldots)$ of elements of the state space, where $\xi_j \in X$. Given such an $\omega \in \Omega$, the function $x_n : \Omega \longmapsto X$ is defined to be the coordinate function of the n^{th} element of ω, that is $x_n(\omega) = \xi_n$.
Given an initial probability distribution $p : \mathscr{F}_X \longmapsto [0,1]$, the probability of being in $A \in \mathscr{F}_X$ at time n is given recursively by

$$\begin{aligned} P\{x_n(\omega) \in A\} &= p(A) && (n = 1) \\ &= \textstyle\int_X p^{(n-1)}(\eta, A) p(d\eta) && (n > 1). \end{aligned}$$

If this probability is independent of n then the x_n process is strictly stationary and $p(A)$ is called a *stationary absolute probability distribution* $\forall A \in \mathscr{F}_X$. The following condition on the behaviour of $p(\xi, A)$ as A becomes small will be important in the rest of this chapter.

Definition 1 (Doeblin's Condition)**.** *There exists a finite-valued measure ϕ on $\mathscr{F}_X$ with $\phi(X) > 0$, an integer $m \geq 1$ and a constant $\epsilon > 0$ such that $\forall \xi \in X$ and $\forall A \in \mathscr{F}_X$*

$$\textit{if } \phi(A) \leq \epsilon \textit{ then } p^{(m)}(\xi, A) \leq 1 - \epsilon.$$

5.2.2 Some Results on Ergodicity

What follows are a series of useful definitions and results proven by Doob[38]. From now on the Doeblin condition will be assumed to hold. This section is concluded with a result from a paper by Bedford and Lindqvist[11], presented in the form of a theorem.

A set $E \in \mathscr{F}_X$ is called an *invariant set* if $\forall \xi \in E$ and $\forall n \in \mathbb{N}^{(>0)}$

$$p^{(n)}(\xi, E) = 1.$$

Informally, once the process has entered E, the process will not be able to enter states not in E. E is called *minimally invariant* if for all proper sub-sets E_0 of E, if E_0 is invariant then $\phi(E_0) = 0$. Doob shows that under the Doeblin condition there are a countable number $E_1, E_2, \ldots$ of disjoint minimally invariant sets in X. For the purposes of proving identifiability of the transition densities, only those processes with just one minimally invariant set covering the whole state space will be of interest. Such a process is said to be *ergodic*. The main results of Doob[38] that will be used are his theorems 5.7, 6.1 and 6.2, which are stated here without proof.

Theorem 2 (Doob's Theorem 5.7 - Special case)**.** *Under the Doeblin condition, if the process is ergodic then the following limit defines a stationary absolute probability distribution $\forall \xi \in X$, independent of ξ*

$$\pi(E) = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} p^{(m)}(\xi, E)$$

for all $E \in \mathscr{F}_X$.

Theorem 3 (Doob's Theorem 6.2 - Special case)**.** *Let $f : X \longmapsto \mathscr{R}^{(\geq 0)}$ be a measurable function with respect to $\mathscr{F}_X$, with*

$$\int_X |f(\xi)|\, \pi(d\xi) < \infty.$$

Then under the Doeblin condition, for any initial distribution of probabilities,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} f(x_m)$$

exists with probability 1 and also,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} f(x_m) = \int_X f(\xi)\pi(d\xi) = \mathbb{E}\{f(x_1)\}$$

with probability 1.

Theorem 4 (Bedford and Lindqvist)**.** *Given an ergodic Markov chain with continuous transition densities that satisfy the Doeblin condition, the transition densities are identifiable from a single realisation of the process.*

Proof. Given that there is at most one invariant set which covers the whole space X, by Theorem 2 there is a single stationary distribution π. Hence from Theorem 3, for any Borel set A and for any starting point X_0

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \mathbf{1}_{\{A\}}(X_i) = \pi(A),$$

where f has been chosen as $\mathbf{1}_{\{A\}}$, the characteristic function for the set A. Suppose now that X_0 has the stationary distribution and consider the Markov chain $\{Y_n\}_{(N\geq 0)} = \{(X_n, X_{n-1})\}_{(N\geq 0)}$ on the state space X^2.

Since this Markov chain is also ergodic, for all measurable sets A and B the probability $q(A, B) = P(X_0 \in A, X_1 \in B)$ can be determined in the same way as π. For any starting point Y_0

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \mathbf{1}_{\{(A,B)\}}(X_i, X_{i-1}) = q(A, B).$$

The probability $q(A, B)$ can be written in terms of the transition probabilities and the stationary distribution as $q(A, B) = \int_A p(x, B)d\pi(x)$. Since the transition densities are continuous this necessarily implies

$$p(x, B) = \lim_{\epsilon\to 0} \frac{q((x-\epsilon, x+\epsilon), B)}{\pi((x-\epsilon, x+\epsilon))}.$$

This implies that the transition probabilities can be written in terms of the transition densities and the stationary distribution as $p(x, B) = \int_B p(x, y)d\pi(y)$. Again, since the transition densities are continuous

$$p(x, y) = \lim_{\epsilon\to 0} \frac{q(x, (y-\epsilon, y+\epsilon))}{\pi((y-\epsilon, y+\epsilon))}.$$

Thus the transition densities of the Markov chain are identifiable from a single sequence, as required.

□

5.3 The Bedford and Lindqvist Model With n Components

The models considered in this section are a generalisation of those discussed by Bedford and Lindqvist[11]. They considered a model for two component systems with the following repair strategy, called partial repair. When one component fails that component is replaced or restored to a state of as good as new while the other component is given minimal repair (these two terms were defined in Chapter 3). This model can be described using the asymmetric GVA framework introduced in Section 4.3. The generalisation made in this section is to consider systems with n components. The repair policy is such that if the system fails due to component i say, that component is fully renewed as before while the remaining components are all minimally repaired. Again this model can be described using the asymmetric GVA framework. The state of the system is captured by the virtual ages of each of the components.

5.3.1 Markov Chain Representation

In this subsection a Markov chain representation of the Bedford and Lindqvist model with n components will be introduced. This representation will facilitate the proof that the joint distribution of this model is identifiable from a single realisation. Let $\underline{B}_i$ denote the i^{th} point in the Markov chain. The basic property of a Markov chain is that

$$P\{\underline{B}_{i+1} = \underline{j} | \underline{B}_i = \underline{k}, \underline{B}_{i-1} = \underline{k}_{i-1}, , \underline{X}_1 = \underline{k}_1, \underline{B}_0 = \underline{k}_0\} = P\{\underline{B}_{i+1} = \underline{j} | \underline{B}_i = \underline{k}\}.$$

Thus the state at the i^{th} point should only depend on the the state of the previous point. In general, the sequences of virtual age models are deterministic functions of the entire past of of the process. However, it will be seen that the Bedford and Lindqvist model satisfies the Markov property.
The state of any given point in the Markov chain can be defined as the following $n+1$ dimensional vector using the state of the n component asymmetric GVA framework

$$\underline{B}_i = \left(B_i^1, B_i^2, B_i^3, \ldots, B_i^{n+1}\right) = \left(T_i, \delta_i, A_i^1, \ldots, A_i^{\delta_i - 1}, A_i^{\delta_i + 1}, \ldots, A_i^n\right).$$

Recall that T_i is the time between the i^th and $(i+1)^th$ failure event, $\delta_i \in \{1, \cdots n\}$ denotes the component that was the cause of system failure and the set $A_i^1, \ldots, A_i^{\delta_i - 1}, A_i^{\delta_i + 1}, \ldots, A_i^n$ denotes the virtual ages of each of the components (except the one that caused the failure) immediately

after the i^th failure event. Care must be taken with describing the state space at this point, because of the repair strategy employed. Assume that at the $(i+1)^{th}$ event, the j^{th} component was the cause of system failure. It should be noted that the virtual ages of the components that didn't fail cannot be less than the time T_i, the value of the first component of the state.

The new state space takes the following form

$$\mathscr{B} = \begin{cases} \left(B^1, B^2, \ldots, B^{n+1}\right): & B^1 \in [0, +\infty) \\ & B^2 \in \{1, \cdots, n\} \\ & B^j \in [B^1, +\infty), \forall j \in \{3, \cdots, n+1\}. \end{cases}$$

Having shown how to construct the Markov chain from the asymmetric GVA model, the two representations will be shown to be isomorphic by demonstrating how to re-construct the asymmetric GVA model from the Markov chain representation.

Consider the sequence of states $\{\underline{B}_i\}_{(i\geq 0)}$ such that for each i, $\underline{B}_i \in \mathscr{B}$. The corresponding asymmetric GVA model can be constructed in the following way. The T_i and δ_i directly equate to the first and second components respectively of the vector for the i^{th} element in the sequence. For each j from 1 to n, the virtual age A_i^j is reconstructed using the following formula

$$A_i^j = \begin{cases} B_i^{j+2} & j < \delta_i \\ 0 & j = \delta_i \\ B_i^{j+1} & j > \delta_i. \end{cases}$$

It will now be enough to demonstrate that the Markov chain representation is identifiable from a single sequence.

5.3.2 Identifiability of the Markov Chain Representation

Following the same manner of argument given by Bedford and Lindqvist[11], it is necessary to impose the following technical condition to ensure identifiability. "The probability of failure of all n components, each at least once, within any time interval is bounded away from zero, independently of the age of the component". This condition will be referred to as the BL (Bedford and Lindqvist) condition. In order to define this condition mathematically, the following conventions and definitions will be useful.

For all $\delta > 0$ and for all $i \in \{1, \ldots, n\}$, the following sets are defined

$$D_i^\delta = \begin{cases} \left(B^1, B^2, \ldots, B^{n+1}\right): & B^1 \in [0, \delta) \\ & B^2 \in \{i\} \\ & B^j \in [B^1, \delta), \forall j \in \{3, \cdots, n+1\}. \end{cases}$$

The union of these sets for all i from 1 to n will be denoted by D^δ. D^δ represents the set of states for which all the components have failed at least once within a time interval δ.

An interval set about a point $z = \left(z^1, \cdots, z^{n+1}\right) \in \mathscr{B}$ will be defined in the following way for all $\delta > 0$

$$C_z^\delta = \begin{cases} \left(B^1, B^2, \ldots, B^{n+1}\right): & B^1 \in (\max(z^1 - \delta, 0), z^1 + \delta) \\ & B^2 \in \left\{z^2\right\} \\ & B^j \in (\max(z^j - \delta, 0), z^j + \delta), \forall j \in \{3, \cdots, n+1\}. \end{cases}$$

There are n zeros, one for each type. For all $i \in \{1, \ldots, n\}$, the i^{th} zero is written as

$$0_i = (0, i, 0, \cdots, 0).$$

These states represent the potential starting states of the process, at which point all the virtual ages are zero and no failures have happened yet. The following condition is imposed to preserve the logical structure of the problem. For all $i, j \in \{1, \ldots, n\}$ and for all $x \in \mathscr{B}$ the following equality holds

$$p\left(0_i, x\right) = p\left(0_j, x\right).$$

This condition means that the transition probability from one zero to any other state x is the same as the transition probability from any of the other zeroes to x, thus enforcing equivalent behaviour for transitions from any zero states. This preserves the notational convenience of having these multiple zero states while ensuring that they can be treated as equivalent. Having made these definitions, the BL condition can be described mathematically as follows.

Definition 2 (BL Condition)**.** *For all $\delta > 0$ there exists $\eta > 0$ and $k > 0$, such that $\forall s \in \mathscr{B}$*

$$p^{(k)}\left(s, D^\delta\right) > \eta.$$

This condition is the mathematical representation of the condition that "The probability of failure of all n components, each at least once, within any time interval is bounded away from zero, independently of the age of the component". The set D^{δ} represents the set of states that satisfy this event. Using this condition it can be shown that the Doeblin condition (definition 1 on pg.52) holds for this Markov chain.

Theorem 5 (Establishing Doeblin's condition)**.** *Given the BL condition holds for the n component Markov chain defined above and that the original intensity functions of the asymmetric GVA model are continuous, the Doeblin condition is satisfied.*

Proof. The assumption of continuity in the original intensity functions yields continuity in all components of each vector argument except the second component of each vector argument (which is discrete).

Applying the BL condition with s chosen to be zero implies that for all $\alpha > 0$, there exists $\eta^0 > 0$ and there exists $k > 0$ such that

$$p^{(k)}\left(0, D^{\alpha}\right) > \eta^0.$$

This implies that there exists $\gamma > 0$ and there exists $z_0 \in D^{\alpha}$ such that $p^{(k)}\left(0, z_0\right) > \gamma$ (since the transition densities cannot all be zero within a set whose transition probability is non-zero). Because the transition probabilities are continuous in all but the second components of each of their arguments, this implies that there exists $\eta > 0$ and there exists some interval set $C_{z_0}^{\eta} \subset Z^{\alpha}$ such that for all $b \in C_{z_0}^{\eta}$

$$p^{(k)}\left(0, b\right) > \eta.$$

Also by continuity, for all $\epsilon > 0$ there exists $\delta > 0$ such that for all $b \in C_{z_0}^{\eta}$ and for all $x \in D^{\delta}$

$$\left|p^{(k)}\left(x, b\right) - p^{(k)}\left(0, b\right)\right| < \epsilon.$$

Which necessarily implies

$$\epsilon + p^{(k)}\left(x, b\right) > p^{(k)}\left(0, b\right) > \eta.$$

Choosing $\epsilon = \frac{\eta}{2}$, we have

$$p^{(k)}\left(x, b\right) > \frac{\eta}{2}. \tag{5.3.1}$$

A second application of the BL condition using the same δ from above implies that there exists $\mu > 0$ and there exists $h > 0$ such that for all $s \in \mathscr{B}$

$$p^{(h)}\left(s, D^{\delta}\right) > \mu. \tag{5.3.2}$$

Combining the two results (Equations 5.3.1, and 5.3.2) allows the $(k+h)$-step transition density to be estimated. For all $s \in \mathscr{B}$ and for all $b \in C_{z_0}^{\eta}$ there exists $\eta'' > 0$ such that

$$\begin{aligned} p^{(k+h)}(s,b) &= \textstyle\int_{\mathscr{B}} p^{(k)}(x,b)p^{(h)}(s,dx) \\ &\geq \textstyle\int_{D^{\delta}} p^{(k)}(x,b)p^{(h)}(s,dx) \\ &> \tfrac{\eta}{2}\mu = \eta''. \end{aligned}$$

Given this bound, for any measurable set $A \subset \mathscr{B}$ the $(k+h)$-step transition probability is bounded in the following way for all $s \in \mathscr{B}$:

$$\begin{aligned} p^{(k+h)}(s,A) &= 1 - p^{(k+h)}(s,A^c) \leq 1 - p^{(k+h)}(s, C^{\eta} \setminus A) \\ &< 1 - \eta''\phi(C^{\eta} \setminus A) \leq 1 - \eta''\left[\phi(C^{\eta}) - \phi(A)\right]. \end{aligned}$$

Here ϕ is the obvious product measure of Lebesque measure and uniform measure over the state space. Taking

$$\epsilon = \frac{\phi(C^{\eta})\,\eta''}{1+\eta''}$$

and remembering the first part of the Doeblin condition, gives

$$\begin{aligned} p^{(k+h)}(s,A) &< 1 - \eta''\left[\phi(C^{\eta}) - \phi(A)\right] = 1 - (1+\eta'')\epsilon + \eta''\phi(A) \\ &< 1 - (1+\eta'')\epsilon + \eta''\epsilon = 1 - \epsilon \end{aligned}$$

and Doeblin's condition is satisfied. □

As stated previously, one of the consequences of Doeblins condition is that $\mathscr{B}$ can be decomposed into a countable number of minimally invariant non-null sets and one transient set. Any two minimally invariant sets will either be disjoint or will differ by at most a set of measure zero.
In this case, for any $\alpha > 0$ consider the set D^{α}. This set is not minimally invariant because it is

always possible for the system to survive to some time longer than α before the next system failure. This means there is at most one invariant set which covers the whole space $\mathscr{B}$. From Theorem 4 in Section 5.2.2 it is concluded that the transition densities of the Markov chain are identifiable from a single sequence. Because the Markov chain representation and the n component assymmetric GVA framework are isomorphic, this necessarily implies that the joint distribution for the n component assymmetric GVA framework is also identifiable from a single realisation of the process.

5.4 The Repair Replace Model With n Components

In some cases, repair at the system level can be viewed as replacement at the component level. This next model is an attempt to describe just such a situation. Consider a series system of n components subject to the following repair strategy; every time that a component fails the system is repaired by replacing certain components. It shall be assumed that the probability of failure of the system only depends on the ages of each of the components. During a specified period of operation of this system the following data can be collected each time a failure of any sort occurs.

- The time since the previous failure.
- The component that failed.
- The components that were replaced to repair the system immediately after the failure.

This augmented dataset is what shall be meant by "failure data" in this section.

It shall be assumed that the probability of a component being replaced will depend only on which component it is and on which component failed. This can be represented by a matrix of probabilities with entries $p_{ij} \in (0,1)$ for $i \neq j$ denoting the probability that component i is replaced given component j failed. In the case of $i = j$ it will be assumed that the component that fails is always replaced. It should be noted that systems with any $p_{ij} = 0$ or 1 are not considered. This means that the maintenance model for a component is based on the same mathematical representation as the Brown-Proschan model for imperfect repair introduced in Section 3.2. However the repair replace imperfect repair model has a different interpretation for it's parameters.

The system can be described using the asymmetric GVA framework, where the time to failure of component j is modelled as the j^{th} risk. The requirement that the sequence of virtual ages be a deterministic function of the past can be relaxed without loss, as described by Doyen and Gaudoin[41].

5.4.1 Markov Chain Representation

In this subsection a Markov chain representation of the repair replace model with n components will be introduced. This representation will facilitate the proof that the joint distribution of this model is identifiable from a single realisation. As in the case of the Bedford and Linqvist model, it will be seen that the repair replace model satisfies the Markov property and can be modelled using a Markov chain representation. The state of any given point in the Markov chain can be defined in the same way as in the previous section

$$\underline{B}_i = \left(B_i^1, B_i^2, B_i^3, \ldots, B_i^{n+1}\right) = \left(T_i, \delta_i, A_i^1, \ldots, A_i^{\delta_i - 1}, A_i^{\delta_i + 1}, \ldots, A_i^n\right).$$

As before $\delta_i \in \{1, \cdots n\}$ and denotes the component which was the cause of system failure. Again, care must be taken with notation at this point because of the repair strategy employed. Assume that at the $(i+1)^{th}$ event, the j^{th} component was the cause of system failure. The virtual ages of the components that didn't fail and weren't replaced cannot be less than the time T_i, the value of the first component of the state. The virtual ages of the components that didn't fail and were replaced are of course zero.

This leads to a subtly different state space than in the previous section which takes the following form

$$\mathscr{B} = \left\{ \begin{array}{ll} \left(B^1, B^2, \ldots, B^{n+1}\right): & B^1 \in [0, +\infty) \\ & B^2 \in \{1, \cdots, n\} \\ & B^j \in \left\{\{0\} \cup [B^1, +\infty)\right\}, \forall j \in \{3, \cdots, n+1\}. \end{array} \right.$$

Having shown how to construct the Markov chain from the asymmetric GVA model, it can be seen that the method of re-constructing the asymmetric GVA model from the Markov chain representation in the previous section is exactly the same for this model as well. This implies that the two representations are isomorphic and that it will therefore be enough to demonstrate that the Markov chain representation is identifiable from a single sequence.

5.4.2 Identifiability of the Markov Chain Representation

Following the same manner of argument given by Bedford and Lindqvist[11], it is necessary to impose the following technical condition to ensure identifiability. "The probability of failure of a component just after any other system failure, is bounded away from zero independently of the age

of the component". This condition will be referred to as the RR (repair replace) condition. In order to define this condition mathematically, the following conventions and definitions will be useful.
For all $\delta > 0$ and for all $i \in \{1, \ldots, n\}$, the following sets are defined

$$D_i^\delta = \begin{cases} \left(B^1, B^2, \ldots, B^{n+1}\right): & B^1 \in [0, \delta) \\ & B^2 \in \{i\} \\ & B^j \in \left\{\{0\} \cup [B^1, +\delta)\right\}, \forall j \in \{3, \cdots, n+1\}. \end{cases}$$

The union of these sets for all i from 1 to n will be denoted by D^δ. The set of states D^δ represents those states for which *all* the components have failed at least once within a time interval δ.
For all $\delta > 0$ and for all $i \in \{1, \ldots, n\}$, the following sets are defined

$$N_i^\delta = \begin{cases} \left(B^1, B^2, \ldots, B^{n+1}\right): & B^1 \in [0, \delta) \\ & B^2 \in \{i\} \\ & B^j \in \left\{\{0\} \cup [B^1, +\infty)\right\}, \forall j \in \{3, \cdots, n+1\}. \end{cases}$$

The union of these sets for all i from 1 to n will be denoted by N^δ. The set of states N^δ represents those states for which the failure of a component has occurred within time δ since the previous failure event.
For all $\delta > 0$ and for all $i \in \{1, \ldots, n\}$, the following sets are defined

$$Z_i^\delta = \begin{cases} \left(B^1, B^2, \ldots, B^{n+1}\right): & B^1 \in [0, \delta) \\ & B^2 \in \{i\} \\ & B^j \in \{0\}, \forall j \in \{3, \cdots, n+1\}. \end{cases}$$

The union of these sets for all i from 1 to n will be denoted by Z^δ. The set of states Z^δ represents those states for which the failure of a component has occurred within time δ since the previous failure and all the components were replaced.
An interval set about a point $z = \left(z^1, \cdots, z^{n+1}\right) \in \mathscr{B}$ will be defined in the following way for all $\delta > 0$

$$C_z^\delta = \begin{cases} \left(B^1, B^2, \ldots, B^{n+1}\right): & B^1 \in (\max(z^1 - \delta, 0), z^1 - \delta) \\ & B^2 \in \left\{z^2\right\} \\ & B^j \in (\max(z^j - \delta, 0), z^j - \delta), \forall j \in \{3, \cdots, n+1\}. \end{cases}$$

There are n zeros, one for each type. For all $i \in \{1, \ldots, n\}$, the i^{th} zero is written as

$$0_i = (0, i, 0, \cdots, 0).$$

The following condition is imposed to preserve the logical structure of the problem. For all $i, j \in \{1, \ldots, n\}$ and for all $x \in \mathscr{B}$ the following equality holds

$$p\left(0_i, x\right) = p\left(0_j, x\right).$$

Having made these definitions, the RR condition can be described mathematically as follows.

Definition 3 (RR Condition)**.** *For all $\delta > 0$ there exists $\eta > 0$, such that for all $s \in \mathscr{B}$*

$$p^{(1)}\left(s, N^{\delta}\right) > \eta.$$

Using this condition it can be shown that the Doeblin condition (Definition 1 on pg.52) holds for this Markov chain.

Theorem 6 (Establishing Doeblin's condition)**.** *Given the RR condition holds for the n component Markov chain defined above and that the original intensity functions of the asymmetric GVA model are continuous, the Doeblin condition is satisfied for this problem.*

Proof. The assumption of continuity in the original intensity functions yields continuity in all components of each vector argument except the second component of each vector argument (which is discrete).

First, an application of the RR condition with s chosen to be zero implies that for all $\alpha > 0$ there exists $\eta^0 > 0$ such that

$$p^{(1)}\left(0, N^{\alpha}\right) > \eta^0.$$

Note that for all $j \in \{1, \ldots, n\}$, if component j has failed the probability that every component in the system is replaced (regardless of the age of the components) is

$$\prod_{i=1}^{n} p_{ij} > 0.$$

Remembering that:

- N^{α} represents the set of states for which the failure of a component has occurred within time α since the previous failure;
- Z^{α} represents the set of states for which the failure of a component has occurred within time α since the previous failure AND all the components were replaced;

it is easy to see that

$$p^{(1)}\left(0, Z_j^{\alpha}\right) = p^{(1)}\left(0, N_j^{\alpha}\right) \prod_{i=1}^{n} p_{ij} > 0.$$

since $Z^{\alpha} \subset D^{\alpha}$, this implies

$$p^{(1)}\left(0, D^{\alpha}\right) \geq p^{(1)}\left(0, Z^{\alpha}\right) > 0.$$

Using the same argument from the previous section, this implies that there exists $\gamma > 0$ and there exists $z_0 \in D^{\alpha}$ such that $p^{(1)}\left(0, z_0\right) > \gamma$ (since the transition densities cannot all be zero within a set whose transition probability is non-zero). Because the transition probabilities are continuous in all but the second components of each of their arguments, this implies that there exists $\eta > 0$ and there exists some interval set $C_{z_0}^{\eta}$ such that for all $b \in C_{z_0}^{\eta}$

$$p^{(1)}\left(0, b\right) > \eta.$$

Also by continuity, for all $\epsilon > 0$ there exists $\delta > 0$ such that for all $b \in C_{z_0}^{\eta}$ and for all $x \in D^{\delta}$

$$\left| p^{(1)}\left(x, b\right) - p^{(1)}\left(0, b\right) \right| < \epsilon.$$

Which necessarily implies

$$\epsilon + p^{(1)}\left(x, b\right) > p^{(1)}\left(0, b\right) > \eta.$$

Choosing $\epsilon = \frac{\eta}{2}$, we have

$$p^{(1)}\left(x, b\right) > \frac{\eta}{2}. \tag{5.4.1}$$

A second application of the RR condition using the same δ from above implies that there exists $\mu > 0$ such that for all $s \in \mathscr{B}$

$$p^{(1)}\left(s, N^{\delta}\right) > \mu.$$

Using the same argument as above, this implies that there exists $\mu' > 0$ such that

$$p^{(1)}\left(s, Z_j^\delta\right) = p^{(1)}\left(s, N_j^\delta\right) \prod_{i=1}^{n} p_{ij} > \mu',$$

which leads to

$$p^{(1)}\left(0, D^\delta\right) \geq p^{(1)}\left(0, Z^\delta\right) > \mu'. \tag{5.4.2}$$

Combining the two results (Equations 5.4.1, and 5.4.2) allows the 2-step transition density to be estimated. For all $s \in \mathscr{B}$ and for every $b \in C_{z_0}^\eta$ there exists $\eta'' > 0$ such that

$$\begin{aligned} p^{(2)}(s,b) &= \textstyle\int_{\mathscr{B}} p^{(1)}(x,b) p^{(1)}(s,dx) \\ &\geq \textstyle\int_{D^\delta} p^{(1)}(x,b) p^{(1)}(s,dx) \\ &> \tfrac{\eta}{2}\mu' = \eta''. \end{aligned}$$

Given this bound, for any measureable set $A \subset \mathscr{B}$ the 2-step transition probability is bound in the following way for all $s \in \mathscr{B}$:

$$\begin{aligned} p^{(2)}(s,A) &= 1 - p^{(2)}\left(s, A^c\right) \leq 1 - p^{(2)}\left(s, C^\eta \setminus A\right) \\ &< 1 - \eta'' \phi\left(C^\eta \setminus A\right) \leq 1 - \eta''\left[\phi\left(C^\eta\right) - \phi(A)\right]. \end{aligned}$$

Here ϕ is the obvious product measure of Lebesque measure and uniform measure over the state space. Taking

$$\epsilon = \frac{\phi\left(C^\eta\right)\eta''}{1+\eta''}$$

and remembering the first part of the Doeblin condition, gives

$$\begin{aligned} p^{(2)}(s,A) &< 1 - \eta''\left[\phi\left(C^\eta\right) - \phi(A)\right] = 1 - (1+\eta'')\epsilon + \eta''\phi(A) \\ &< 1 - (1+\eta'')\epsilon + \eta''\epsilon = 1 - \epsilon \end{aligned}$$

and Doeblin's condition is satisfied. □

As stated previously, one of the consequences of Doeblins condition is that $\mathscr{B}$ can be decomposed into a countable number of minimally invariant non-null sets and one transient set. Any two minimally invariant sets will either be disjoint or will differ by at most a set of measure zero.

In this case, for any $\alpha > 0$ consider the set D^{α}. This set is not minimally invariant because it is always possible for the system to survive to some time longer than α before the next system failure. This means there is at most one invariant set which covers the whole space $\mathscr{B}$. From Theorem 4 in Section 5.2.2 It is concluded that the transition densities of the Markov chain are identifiable from a single sequence. Because the Markov chain representation and the n component assymmetric GVA framework are isomorphic, this necessarily implies that the joint distribution for the n component assymmetric GVA framework is also identifiable from a single realisation of the process.

5.5 An Example of the Repair Replace Model

The previous two sections have demonstrated the identifiability of two different models and in each case this was dependant on certain conditions. The natural question that arises is whether these conditions are so restrictive that none of the practical situations that one would try to model, actually satisfy these conditions. Bedford and Lindqvist[11] addressed this question for the case of the Bedford and Lindqvist model with two components. This section will seek to apply their strategy to the case of the Repair Replace model with n components.

The adopted strategy is to assume a commonly used form for the marginal distributions of failure for each component and then to show that, coupled with the repair replace strategy, the resulting model satisfies the conditions required for identifiability of the joint distribution. This will successfully demonstrate that a commonly applied reliability model satisfies the conditions required for identifiability.

The proposed example will assume independence between the times to failure of all of the components. The resulting marginal distributions will each be assumed to follow exponential distributions for the time to first failure. In this case the sequence of times to failure of the first component will be denoted by $Z_1, \cdots$ and the sequence of survival functions for the times between consecutive failures of the first component will be denoted by $S_1, \cdots$ (this is all the notation that will be needed for the purposes of this example). For all $t > 0$, the survival function for the first time to failure of the first component is equal to

$$S_1(t) = e^{-\gamma t},$$

where $\gamma > 0$. The exponential distribution is widely used in reliability theory for modelling the time to first failure of a component.

This example will now be shown to satisfy The RR Condition necessary for identifiability of the joint distribution, which was introduced in Subsection 5.4.2 on Page 63. Recall that for all $\epsilon > 0$, N^{ϵ} represents the set of states for which the failure of a component has occurred within time ϵ since the previous failure. For any $s \in \mathscr{B}$, assuming that s is the $(i-1)^{th}$ failure

$$\begin{aligned} p^{(1)}(s, N^{\epsilon}) &= P\{T_i < \epsilon\} \geq P\{Z_i < \epsilon\} = 1 - S_i(\epsilon | A_{i-1}) \\ &= 1 - \frac{S_1(A_{i-1} + \epsilon)}{S_1(A_{i-1})} = 1 - \frac{e^{-\gamma(A_{i-1}+\epsilon)}}{e^{-\gamma A_{i-1}}} = 1 - e^{-\gamma\epsilon} > 0, \end{aligned}$$

as required.

5.6 Summary

In this chapter, the joint distributions of two different generalised competing risk models with asymmetric virtual ages have been shown to be identifiable. The first model was a generalisation of the work done by Bedford and Lindqvist[11] and involved a system of n components in series with the following repair strategy, the component that caused the system to fail is replaced while the remaining components are given minimal repair to restore them to a working state. The second model involved a system of n components in series with the following repair strategy, the component that caused the system to fail is replaced while the remaining components are either given minimal repair to restore them to a working state or they are replaced. The decision to minimally repair or completely replace these remaining components is modelled by a constant probability that depends only on which component caused the system failure and which component is being considered for replacement. An example was given for the second model to demonstrate that the conditions of this model allow for it to be applied in situations of practical interest.

Chapter 6

Opportunistic Maintenance and the Random Signals Model

The work presented in this chapter is an extension of the work done by Bedford and Alkali[10] on opportunistic maintenance (introduced in Chapter 1) in competing risk models.

In Section 2.3 a discussion of existing competing risk models for studying the interaction of preventative and corrective maintenance was given. The modelling in these cases only explicitly takes account of condition-based and calendar-based preventative maintenance. In many situations this would not be unreasonable as Bedford and Cooke[25] point out that "There is considerable overlap between the notions of condition-based and opportunity-based maintenance". But in some cases opportunistic maintenance plays a more significant role than that of condition-based or calendar-based maintenance. For example Allen et. al.[2] describe the maintenance policy at the Stanford Linear Accelerator centre as being "a modification of the run-to-fail" policy, which relies primarily on opportunistic maintenance.

First a brief overview will be given of the opportunistic maintenance models introduced by Bedford and Alkali[10]. A detailed description of the Random Signals model will then be given, followed by the calculations used to derive expressions for the marginal distributions of the time to both preventative and corrective maintenance, as well as the sub-distribution of the time to corrective maintenance. Finally, in view of the copious amounts of algebra involved in the aforementioned calculations, a small simulation study is given and the empirical results generated are shown to be in agreement with those results predicted by the calculations.

6.1 Some Opportunistic Maintenance Models

This section will present three competing risk models which explicitly include opportunistic maintenance. The three models discussed are all found in Bedford and Alkali[10].

6.1.1 Opportunity Age Replacement Model

In this model opportunities for carrying out preventative maintenance occur randomly following a Poisson process. The maintenance policy is that the first opportunity after an elapsed time of system operation t_0, is used to maintain the system. This model has the virtue of being quite simple and therefore easy to apply. However, this model does not allow for system maintainer's knowledge of the condition of the system to influence the maintenance decision.

6.1.2 Opportunity Alert Model

Again, in this model opportunities occur randomly following a Poisson process. The system gives off a warning signal at some random time, and the first opportunity after this signal is used to maintain the system. This model allows for the maintainer's knowledge about the condition of the system to be included in the form of a warning signal observed by the engineer.

6.1.3 Signal Opportunity Model

In this model the system gives off a series of warning signals and again, opportunities occur randomly following a Poisson process. The hazard rate of the failure time is modelled as a piece-wise constant exponential distribution, in which the hazard rate is constant between signals and increases at the occurrence of each successive signal. The first opportunity after the k^{th} signal is used to maintain the system. The advantage of this model over the previous one is that the failure characteristics of the system are explicitly modelled.

This chapter will focus on a variant of the signal opportunity model, given that this model is more general than the opportunity alert model and includes knowledge of the condition of the system, which the opportunity age replacement model does not. Closed form expressions will be given (subject to some minor technical restrictions) for the marginal distributions of the time to both preventative and corrective maintenance, as well as for the sub-distribution of the time to corrective maintenance.

6.2 The Random Signals Model

This section will introduce the detailed workings of the Random Signals model and the notation used to describe it.

6.2.1 Notation and Model Set-up

There are n signals that occur, with the signal times denoted by $S_1, S_2, \ldots, S_n$ and with $S_0 = 0$. For each i from 0 to $n-1$, the inter-signal time $S_{i+1} - S_i$ is assumed to be exponentially distributed with parameter γ_i (EXP(γ_i)). It will be notationally convenient to define $\gamma_n = 0$. The different inter-signal times are independent, though not identically distributed. For each i from 0 to $n-1$, the system has a constant hazard rate λ_i in the period between S_i and S_{i+1}. After S_n, the system has a constant hazard rate λ_n. It will often be convenient to use the following notational shorthand $\delta_i = \gamma_i + \lambda_i$.

The censoring arises through opportunistic maintenance. Assume opportunities for maintenance occur according to a Poisson process with parameter ρ. The opportunistic maintenance strategy used is that the first opportunity after the k^{th} signal (k fixed and strictly less than n) is used to preventively maintain the system. Hence the system lifetime is censored if the first opportunity after the k^{th} signal is earlier than the time to failure.

Let $X_{S_0,S_1\ldots,S_n}$ denote the time to failure conditional on the occurrence of n signals with the set of parameter pairs $(\gamma_0, \lambda_0), \cdots, (\gamma_n, \lambda_n)$ (this will often be abbreviated to X as in previous chapters). Similarly, let $X_{S_k,S_{k+1}\ldots,S_n}$ denote the time to failure conditional on the occurrence of $n-k$ signals with the set of parameter pairs $(\gamma_k, \lambda_k), \cdots, (\gamma_n, \lambda_n)$ and assuming that $S_k = 0$.

As before, Z will denote the time to preventative maintenance. Note that the conditional distribution function of Z given $S_k = r$ is

$$F_{Z|S(k)}(t,r) = \begin{cases} 1 - e^{-\rho[t-r]} & t \geq r, \\ 0 & \text{otherwise.} \end{cases} \tag{6.2.1}$$

Thus, the conditional density function of Z given S_k is

$$f_{Z|S(k)}(t,r) = \begin{cases} \rho e^{-\rho[t-r]} & t \geq r, \\ 0 & \text{otherwise.} \end{cases} \tag{6.2.2}$$

Figure 6.1 illustrates the structure of the model and it's main components as defined above.

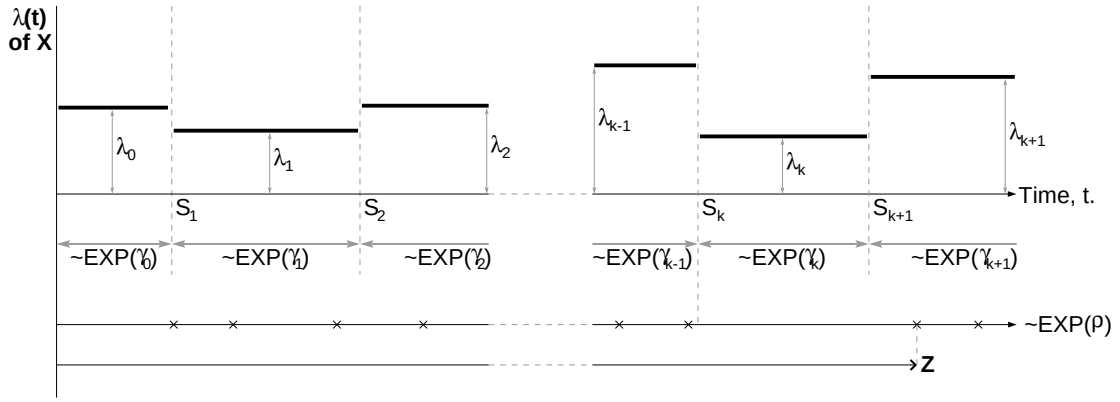


Figure 6.1: The Random Signals Model.

The two following observations will prove useful in calculating the various closed form results of this chapter. Below is the expression for the density function of the sum of k exponentially distributed random variables given that the γ_j are distinct for all j from 0 to k.

$$f_{S(k)}(x) = \left[\prod_{i=0}^{k-1} \gamma_i\right] \sum_{j=0}^{k-1} \frac{e^{-\gamma_j x}}{\prod_{u\neq j, u=0}^{k-1}(\gamma_u - \gamma_j)}. \tag{6.2.3}$$

This expression is a well-known result and a very elegant proof by Balazs can be found in [6].
The second observation, which can be seen by observing Figure 6.1, is that the conditional distribution for X given the values of $S(1)\ldots S(n)$ is

$$F_{X|S(1),\ldots,S(n)}(t|r_1,\ldots,r_n) = 1 - e^{-\{\sum_{i=0}^{j-1}[r_{i+1}-r_i]\lambda_i + [t-r_j]\lambda_j\}} \qquad \text{for } r_j \leq t \leq r_{j+1}.$$

Gathering r_j terms together yields

$$F_{X|S(1),\ldots,S(n)}(t|r_1,\ldots,r_n) = 1 - e^{-\{\sum_{i=1}^{j}[\lambda_{i-1}-\lambda_i]r_i + \lambda_j t\}} \qquad \text{for } r_j \leq t \leq r_{j+1}. \tag{6.2.4}$$

6.3 Main Results of the Random Signals Model

This section will present the calculations used to determine expressions for the marginal distributions of both X and Z as well as the sub-distribution of X.

6.3.1 The Marginal Distribution of X

The calculations given in this subsection will yield a closed form expression for the marginal distribution of X under the following assumption:

- For all i from 0 to n and for all j from 0 to n, $i \neq j$ implies that $\delta_i \neq \delta_j$.

These calculations will use several lemmas found in Appendix A.2. The marginal distribution of X is calculated as follows

$$F_X(t) = P\{X \leq t\} = \sum_{j=0}^{n} P\{X \leq t, j \text{ signals have occurred by time } t\}.$$

By conditioning on the values of the signals and remembering that for each term in the summation, $S(1), \cdots, S(j)$ are all less than t while $S(j+1)$ is greater than t (except when $j = n$, in which case $S(1), \cdots, S(n)$ are all less than t), this necessarily implies

$$\begin{aligned} F_X(t) = \sum_{j=0}^{n-1} \int_0^t \cdots \int_0^t \int_t^{+\infty} & P\{X \leq t | S(1) = r_1, \cdots, S(j+1) = r_{j+1}\} \\ & \times f_{S(j+1)|S(1),\ldots,S(j)}(r_{j+1}) dr_{j+1} \cdots f_{S(2)|S(1)}(r_2) dr_2 f_{S(1)}(r_1) dr_1 \\ + \int_0^t \cdots \int_0^t & P\{X \leq t | S(1) = r_1, \cdots, S(n) = r_n\} \\ & \times f_{S(n)|S(1),\ldots,S(n-1)}(r_n) dr_n \cdots f_{S(2)|S(1)}(r_2) dr_2 f_{S(1)}(r_1) dr_1, \end{aligned}$$

where $f_{S(j+1)|S(1),\ldots,S(j)}(t)$ is the conditional density function of $S(j+1)$ given the preceding signal times $S(1), \ldots, S(j)$. It is clear that given the dependence on the most recent previous signal time, $f_{S(j+1)|S(j)}(t)$ is independent of all the other previous signal times and is equal to $\gamma_j e^{-\gamma_j [r_{j+1} - r_j]}$. This and the application of Equation 6.2.4 (found on Page 71) necessarily implies

$$\begin{aligned} F_X(t) = \sum_{j=0}^{n-1} \int_0^t \cdots \int_{r_{j-1}}^t \int_t^{+\infty} & \left[1 - e^{-\{\sum_{i=1}^{j} [\lambda_{i-1} - \lambda_i] r_i + \lambda_j t\}}\right] \\ & \times \gamma_j e^{-\gamma_j [r_{j+1} - r_j]} dr_{j+1} \ldots \gamma_1 e^{-\gamma_1 [r_2 - r_1]} dr_2 \gamma_0 e^{-\gamma_0 r_1} dr_1 \\ + \int_0^t \cdots \int_{r_{n-1}}^t & \left[1 - e^{-\{\sum_{i=1}^{n} [\lambda_{i-1} - \lambda_i] r_i + \lambda_n t\}}\right] \\ & \times \gamma_{n-1} e^{-\gamma_{n-1} [r_n - r_{n-1}]} dr_n \ldots \gamma_1 e^{-\gamma_1 [r_2 - r_1]} dr_2 \gamma_0 e^{-\gamma_0 r_1} dr_1. \end{aligned}$$

Grouping the r_j's outside of the bracketed terms together and then multiplying out the brackets necessarily implies

$$\begin{aligned}
F_X(t) = &\sum_{j=0}^{n-1} \int_0^t \cdots \int_{r_{j-1}}^t \int_t^{+\infty} \gamma_j e^{-\gamma_j r_{j+1}} dr_{j+1} \gamma_{j-1} e^{-[\gamma_{j-1}-\gamma_j] r_j} dr_j \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
&- \sum_{j=0}^{n-1} \int_0^t \cdots \int_{r_{j-1}}^t \int_t^{+\infty} e^{-\{\sum_{i=1}^{j} [\lambda_{i-1}-\lambda_i] r_i + \lambda_j t\}} \\
&\qquad \times \gamma_j e^{-\gamma_j r_{j+1}} dr_{j+1} \gamma_{j-1} e^{-[\gamma_{j-1}-\gamma_j] r_j} dr_j \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
&+ \int_0^t \cdots \int_{r_{n-1}}^t \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
&- \int_0^t \cdots \int_{r_{n-1}}^t e^{-\{\sum_{i=1}^{n} [\lambda_{i-1}-\lambda_i] r_i + \lambda_n t\}} \\
&\qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1.
\end{aligned}$$

Again grouping the r_j's together and remembering that $\delta_j = \gamma_j + \lambda_j$ necessarily implies

$$\begin{aligned}
F_X(t) = &\sum_{j=0}^{n-1} \int_0^t \cdots \int_{r_{j-1}}^t \int_t^{+\infty} \gamma_j e^{-\gamma_j r_{j+1}} dr_{j+1} \gamma_{j-1} e^{-[\gamma_{j-1}-\gamma_j] r_j} dr_j \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
&- \sum_{j=0}^{n-1} e^{-\lambda_j t} \int_0^t \cdots \int_{r_{j-1}}^t \int_t^{+\infty} \gamma_j e^{-\gamma_j r_{j+1}} dr_{j+1} \\
&\qquad \times \gamma_{j-1} e^{-[\delta_{j-1}-\delta_j] r_j} dr_j \ldots \gamma_0 e^{-[\delta_0-\delta_1] r_1} dr_1 \\
&+ \int_0^t \cdots \int_{r_{n-1}}^t \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
&- e^{-\lambda_n t} \int_0^t \cdots \int_{r_{n-1}}^t \gamma_{n-1} e^{-[\delta_{n-1}-\lambda n] r_n} dr_n \\
&\qquad \times \gamma_{n-2} e^{-[\delta_{n-2}-\delta_{n-1}] r_{n-1}} dr_{n-1} \ldots \gamma_0 e^{-[\delta_0-\delta_1] r_1} dr_1.
\end{aligned}$$

Integrating the two inner-most integrals in the two summations (those with respect to dr_{j+1}, from t to $+\infty$) necessarily implies

$$\begin{aligned}
F_X(t) = &\sum_{j=0}^{n-1} e^{-\gamma_j t} \int_0^t \cdots \int_{r_{j-1}}^t \gamma_{j-1} e^{-[\gamma_{j-1}-\gamma j]r_j} dr_j \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&- \sum_{j=0}^{n-1} e^{-\lambda_j t} e^{-\gamma_j t} \int_0^t \cdots \int_{r_{j-1}}^t \gamma_{j-1} e^{-[\delta_{j-1}-\delta_j]r_j} dr_j \ldots \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1 \\
&+ \int_0^t \cdots \int_{r_{n-1}}^t \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}]r_{n-1}} dr_{n-1} \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&- e^{-\lambda_n t} \int_0^t \cdots \int_{r_{n-1}}^t \gamma_{n-1} e^{-[\delta_{n-1}-\lambda n]r_n} dr_n \\
&\qquad \times \gamma_{n-2} e^{-[\delta_{n-2}-\delta_{n-1}]r_{n-1}} dr_{n-1} \ldots \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1.
\end{aligned}$$

Lemma 1 found on Page 164 of Appendix A.2 necessarily implies

$$\begin{aligned}
F_X(t) = 1 &- \sum_{j=0}^{n-1} e^{-\delta_j t} \int_0^t \cdots \int_{r_{j-1}}^t \gamma_{j-1} e^{-[\delta_{j-1}-\delta_j]r_j} dr_j \ldots \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1 \\
&- e^{-\lambda_n t} \int_0^t \cdots \int_{r_{n-1}}^t \gamma_{n-1} e^{-[\delta_{n-1}-\lambda n]r_n} dr_n \\
&\qquad \times \gamma_{n-2} e^{-[\delta_{n-2}-\delta_{n-1}]r_{n-1}} dr_{n-1} \ldots \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1.
\end{aligned}$$

Remembering that $\gamma_n = 0$ (and thus $\lambda_n = \delta_n$) necessarily implies

$$F_X(t) = 1 - \sum_{j=0}^{n} e^{-\delta_j t} \int_0^t \cdots \int_{r_{j-1}}^t \gamma_{j-1} e^{-[\delta_{j-1}-\delta_j]r_j} dr_j \ldots \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1.$$

Finally, Lemma 2 found on Page 165 of Appendix A.2 necessarily implies

$$F_X(t) = 1 - \sum_{j=0}^{n} \left[\prod_{r=0}^{j-1} \gamma_r \right] \sum_{i=0}^{j} \frac{e^{-\delta_i t}}{\prod_{\substack{u=0 \\ u \neq i}}^{j} (\delta_u - \delta_i)}. \tag{6.3.1}$$

6.3.2 The Marginal Distribution of Z

The calculations given in this subsection will yield a closed form expression for the marginal distribution of Z under the following assumptions:

- For all i from 0 to n and for all j from 0 to n, $i \neq j$ implies that $\gamma_i \neq \gamma_j$;

- For all i from 0 to n, $\gamma_i \neq \rho$.

The marginal distribution of Z is calculated as follows by conditioning on the value of $S(k)$

$$F_Z(t) = P\{Z \leq t\} = \int_0^\infty P\{Z \leq t | S(k) = r\} f_{S(k)}(r) dr.$$

Equations 6.2.1 (found on Page 70) and 6.2.3 (found on Page 71)necessarily imply

$$F_Z(t) = \int_0^t \left[1 - e^{-\rho[t-r]}\right] \left[\prod_{i=0}^{k-1} \gamma_i\right] \sum_{j=0}^{k-1} \frac{e^{-\gamma_j r}}{\prod_{u \neq j, u=0}^{k-1} (\gamma_u - \gamma_j)} dr.$$

Bringing the bracketed term inside the summation necessarily implies

$$F_Z(t) = \int_0^t \left[\prod_{i=0}^{k-1} \gamma_i\right] \sum_{j=0}^{k-1} \left(\frac{e^{-\gamma_j r}}{\prod_{u \neq j, u=0}^{k-1} (\gamma_u - \gamma_j)} - e^{-\rho t} \frac{e^{-[\gamma_j - \rho] r}}{\prod_{u \neq j, u=0}^{k-1} (\gamma_u - \gamma_j)} \right) dr.$$

Bringing the integral inside the summation necessarily implies

$$F_Z(t) = \left[\prod_{i=0}^{k-1} \gamma_i\right] \sum_{j=0}^{k-1} \left(\frac{\int_0^t e^{-\gamma_j r} dr}{\prod_{u \neq j, u=0}^{k-1} (\gamma_u - \gamma_j)} - e^{-\rho t} \frac{\int_0^t e^{-[\gamma_j - \rho] r} dr}{\prod_{u \neq j, u=0}^{k-1} (\gamma_u - \gamma_j)} \right).$$

Solving the integrals necessarily implies

$$\begin{aligned} F_Z(t) &= \left[\prod_{i=0}^{k-1} \gamma_i\right] \sum_{j=0}^{k-1} \left(\frac{1 - e^{-\gamma_j t}}{\gamma_j \prod_{u \neq j, u=0}^{k-1} (\gamma_u - \gamma_j)} - e^{-\rho t} \frac{e^{-[\gamma_j - \rho] t} - 1}{[\rho - \gamma_j] \prod_{u \neq j, u=0}^{k-1} (\gamma_u - \gamma_j)} \right) \\ &= \left[\prod_{i=0}^{k-1} \gamma_i\right] \sum_{j=0}^{k-1} \left(\frac{1 - e^{-\gamma_j t}}{\gamma_j \prod_{u \neq j, u=0}^{k-1} (\gamma_u - \gamma_j)} - \frac{e^{-\gamma_j t} - e^{-\rho t}}{[\rho - \gamma_j] \prod_{u \neq j, u=0}^{k-1} (\gamma_u - \gamma_j)} \right). \end{aligned}$$

6.3.3 The Sub-distribution of X

The calculations given in this subsection will yield a closed form expression for the sub-distribution of X under the following assumptions:

- For all i from 0 to $k-1$ and for all j from 0 to $k-1$, $i \neq j$ implies that $\delta_i \neq \delta_j$;
- For all i from k to n and for all j from k to n, $i \neq j$ implies that $\delta_i \neq \delta_j$;
- For all i from 0 to $k-1$ and for all j from k to n, $\delta_i \neq \delta_j + \rho$.

The sub-distribution of X is calculated as follows

$$F_X^*(t) = P\{X \leq t, X < Z\} = \sum_{w=0}^{n} P\{X \leq t, X < Z, w \text{ signals have occurred by time } t\}.$$

By conditioning on the values of the signals and remembering that for each term in the summation, $S(1), \cdots, S(w)$ are all less than t while $S(w+1)$ is greater than t (except when $w = n$, in which case $S(1), \cdots, S(n)$ are all less than t), this necessarily implies

$$\begin{aligned} F_X^*(t) = \sum_{w=0}^{n-1} \int_0^t \cdots \int_0^t \int_t^{+\infty} & P\{X \leq t, X < Z | S(1) = r_1, \cdots, S(w+1) = r_{w+1}\} \\ & \times f_{S(w+1)|S(1),\ldots,S(w)}(r_{w+1}) dr_{w+1} \ldots f_{S(2)|S(1)}(r_2) dr_2 f_{S(1)}(r_1) dr_1 \\ + \int_0^t \cdots \int_0^t & P\{X \leq t, X < Z | S(1) = r_1, \cdots, S(n) = r_n\} \\ & \times f_{S(n)|S(1),\ldots,S(n-1)}(r_n) dr_n \ldots f_{S(2)|S(1)}(r_2) dr_2 f_{S(1)}(r_1) dr_1. \end{aligned}$$

where $f_{S(w+1)|S(1),\ldots,S(w)}(t)$ is the conditional density function of $S(w+1)$ given the preceding signal times $S(1), \ldots, S(w)$. It is clear that given the dependence on the most recent previous signal time, $f_{S(w+1)|S(w)}(t)$ is independent of all the other previous signal times and is equal to $\gamma_w e^{-\gamma_w [r_{w+1} - r_w]}$.

Also note that if $w < k$ then conditioning on $S(w) < t < S(w+1)$ necessarily implies that $Z > S(k) \geq S(w+1) > t$, so if $X \leq t$, then $X < Z$. This necessarily implies

$$\begin{aligned} F_X^*(t) = \sum_{w=0}^{k-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} & P\{X \leq t | S(1) = r_1, \cdots, S(w+1) = r_{w+1}\} \\ & \times \gamma_w e^{-\gamma_w [r_{w+1} - r_w]} dr_{w+1} \ldots \gamma_1 e^{-\gamma_1 [r_2 - r_1]} dr_2 \gamma_0 e^{-\gamma_0 r_1} dr_1 \\ + \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} & P\{X \leq t, X < Z | S(1) = r_1, \cdots, S(w+1) = r_{w+1}\} \\ & \times \gamma_w e^{-\gamma_w [r_{w+1} - r_w]} dr_{w+1} \ldots \gamma_1 e^{-\gamma_1 [r_2 - r_1]} dr_2 \gamma_0 e^{-\gamma_0 r_1} dr_1 \\ + \int_0^t \cdots \int_{r_{n-1}}^t & P\{X \leq t, X < Z | S(1) = r_1, \cdots, S(n) = r_n\} \\ & \times \gamma_{n-1} e^{-\gamma_{n-1} [r_n - r_{n-1}]} dr_n \ldots \gamma_1 e^{-\gamma_1 [r_2 - r_1]} dr_2 \gamma_0 e^{-\gamma_0 r_1} dr_1. \end{aligned}$$

By grouping the r_w's together and conditioning on the value of Z between r_k and $+\infty$ in the second summation and the last term, this necessarily implies

$$
\begin{aligned}
F_X^*(t) = \sum_{w=0}^{k-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} & P\{X \le t | S(1) = r_1, \cdots, S(w+1) = r_{w+1}\} \\
& \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
+ \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} \int_{r_k}^{+\infty} & \\
& \times P\{X \le t, X < z | S(1) = r_1, \cdots, S(w+1) = r_{w+1}, Z = z\} f_{Z|S(k)}(z, r_k) dz \\
& \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
+ \int_0^t \cdots \int_{r_{n-1}}^t \int_{r_k}^{+\infty} & P\{X \le t, X < z | S(1) = r_1, \cdots, S(n) = r_n, Z = z\} f_{Z|S(k)}(z, r_k) dz \\
& \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1,
\end{aligned}
$$

where $f_{Z|S(k)}$ is the conditional density function of Z given $S(k)$. Note that conditional on the value of $S(k)$, the probability that $X < z$ is independent of Z. By splitting the integrals over z into two parts (z between r_k and t, z between t and $+\infty$) and applying Equation 6.2.2 (found on

Page 70) for the conditional density of Z given $S(k)$, this necessarily implies

$$
\begin{aligned}
F_X^*(t) = \sum_{w=0}^{k-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} & P\{X \le t | S(1) = r_1, \cdots, S(w+1) = r_{w+1}\} \\
& \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1} - \gamma_w] r_w} dr_w \cdots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1 \\
+ \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} \int_{r_k}^t & P\{X < z | S(1) = r_1, \cdots, S(w+1) = r_{w+1}\} \rho e^{-\rho[z - r_k]} dz \\
& \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1} - \gamma_w] r_w} dr_w \cdots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1 \\
+ \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} & P\{X \le t | S(1) = r_1, \cdots, S(w+1) = r_{w+1}\} \int_t^{+\infty} \rho e^{-\rho[z - r_k]} dz \\
& \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1} - \gamma_w] r_w} dr_w \cdots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1 \\
+ \int_0^t \cdots \int_{r_{n-1}}^t \int_{r_k}^t & P\{X < z | S(1) = r_1, \cdots, S(n) = r_n\} \rho e^{-\rho[z - r_k]} dz \\
& \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2} - \gamma_{n-1}] r_{n-1}} dr_{n-1} \cdots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1 \\
+ \int_0^t \cdots \int_{r_{n-1}}^t & P\{X \le t | S(1) = r_1, \cdots, S(n) = r_n\} \int_t^{+\infty} \rho e^{-\rho[z - r_k]} dz \\
& \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2} - \gamma_{n-1}] r_{n-1}} dr_{n-1} \cdots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1.
\end{aligned}
$$

The integrals over z in the second summation and the penultimate term can be split up into $n - k$ terms (where $r_u < z < r_{u+1}$ for u between k and $n - 1$) and a final term integrating z from r_n to

$+\infty$. This necessarily implies

$$
\begin{aligned}
F_X^*(t) = \sum_{w=0}^{k-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} & P\{X \le t | S(1) = r_1, \cdots, S(w+1) = r_{w+1}\} \\
& \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1} - \gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1 \\
+ \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} & \\
& \times \sum_{u=k}^{w-1} \left[\int_{r_u}^{r_{u+1}} P\{X < z | S(1) = r_1, \cdots, S(u+1) = r_{u+1}\} \rho e^{-\rho[z - r_k]} dz \right] \\
& \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1} - \gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1 \\
+ \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} \int_{r_w}^t & P\{X < z | S(1) = r_1, \cdots, S(w+1) = r_{w+1}\} \rho e^{-\rho[z - r_k]} dz \\
& \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1} - \gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1 \\
+ \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} & P\{X \le t | S(1) = r_1, \cdots, S(w+1) = r_{w+1}\} \int_t^{+\infty} \rho e^{-\rho[z - r_k]} dz \\
& \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1} - \gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1 \\
+ \int_0^t \cdots \int_{r_{n-1}}^t \sum_{u=k}^{n-1} & \left[\int_{r_u}^{r_{u+1}} P\{X < z | S(1) = r_1, \cdots, S(u+1) = r_{u+1}\} \rho e^{-\rho[z - r_k]} dz \right] \\
& \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2} - \gamma_{n-1}] r_{n-1}} dr_{n-1} \ldots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1 \\
+ \int_0^t \cdots \int_{r_{n-1}}^t \int_{r_n}^t & P\{X < z | S(1) = r_1, \cdots, S(n) = r_n\} \rho e^{-\rho[z - r_k]} dz \\
& \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2} - \gamma_{n-1}] r_{n-1}} dr_{n-1} \ldots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1 \\
+ \int_0^t \cdots \int_{r_{n-1}}^t & P\{X \le t | S(1) = r_1, \cdots, S(n) = r_n\} \int_t^{+\infty} \rho e^{-\rho[z - r_k]} dz \\
& \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2} - \gamma_{n-1}] r_{n-1}} dr_{n-1} \ldots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1.
\end{aligned}
$$

Applying Equation 6.2.4 (found on Page 71) to each term necessarily implies

$$
\begin{aligned}
F_X^*(t) = \sum_{w=0}^{k-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} & \left[1 - e^{-\{\sum_{i=1}^{w}[\lambda_{i-1}-\lambda_i]\times r_i + \lambda_w t\}}\right] \\
& \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w]r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
+ \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} \sum_{u=k}^{w-1} & \left[\int_{r_u}^{r_{u+1}} \left[1 - e^{-\{\sum_{i=1}^{u}[\lambda_{i-1}-\lambda_i]r_i + \lambda_u z\}}\right] \rho e^{-\rho[z-r_k]} dz\right] \\
& \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w]r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
+ \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} & \int_{r_w}^t \left[1 - e^{-\{\sum_{i=1}^{w}[\lambda_{i-1}-\lambda_i]r_i + \lambda_w z\}}\right] \rho e^{-\rho[z-r_k]} dz \\
& \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w]r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
+ \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} & \left[1 - e^{-\{\sum_{i=1}^{w}[\lambda_{i-1}-\lambda_i]r_i + \lambda_w t\}}\right] \int_t^{+\infty} \rho e^{-\rho[z-r_k]} dz \\
& \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w]r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
+ \int_0^t \cdots \int_{r_{n-1}}^t \sum_{u=k}^{n-1} & \left[\int_{r_u}^{r_{u+1}} \left[1 - e^{-\{\sum_{i=1}^{u}[\lambda_{i-1}-\lambda_i]r_i + \lambda_u z\}}\right] \rho e^{-\rho[z-r_k]} dz\right] \\
& \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}]r_{n-1}} dr_{n-1} \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
+ \int_0^t \cdots \int_{r_{n-1}}^t \int_{r_n}^t & \left[1 - e^{-\{\sum_{i=1}^{n}[\lambda_{i-1}-\lambda_i]r_i + \lambda_n z\}}\right] \rho e^{-\rho[z-r_k]} dz \\
& \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}]r_{n-1}} dr_{n-1} \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
+ \int_0^t \cdots \int_{r_{n-1}}^t & \left[1 - e^{-\{\sum_{i=1}^{n}[\lambda_{i-1}-\lambda_i]r_i + \lambda_n t\}}\right] \int_t^{+\infty} \rho e^{-\rho[z-r_k]} dz \\
& \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}]r_{n-1}} dr_{n-1} \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1.
\end{aligned}
$$

All that remains is some rather lengthy algebra. These remaining calculations have been included for completeness, but relegated to the appendix in Section A.1 on Page 139. The result of these

calculations is stated below.

$$
\begin{aligned}
F_X^*(t) = 1 &- \sum_{u=0}^{k-1} \left[\prod_{r=0}^{u-1} \gamma_r \right] \sum_{j=0}^{u} \frac{e^{-\delta_j t}}{\prod_{\substack{u=0 \\ u \neq j}}^{u} (\delta_u - \delta_j)} \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r \right] \frac{1}{\lambda_u + \rho} \sum_{q=0}^{k-1} \frac{\rho[\delta_q - \lambda_u - \rho]\left(1 - e^{-\delta_q t}\right) - \delta_q \lambda_u e^{-\delta_q t}}{\delta_q \left[\prod_{\substack{w=k \\ w \neq q}}^{u} (\delta_w - \delta_q + \rho) \right] \left[\prod_{\substack{v=0 \\ v \neq q}}^{k-1} (\delta_v - \delta_q) \right]} \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r \right] \frac{1}{\lambda_u + \rho} \sum_{q=k}^{u} \frac{\rho[\delta_q - \lambda_u]\left(1 - e^{-[\delta_q + \rho]t}\right) - \lambda_u[\delta_q + \rho] e^{-[\delta_q + \rho]t}}{[\delta_q + \rho] \left[\prod_{\substack{w=k \\ w \neq q}}^{u} (\delta_w - \delta_q) \right] \left[\prod_{\substack{v=0 \\ v \neq q}}^{k-1} (\delta_v - \delta_q - \rho) \right]}.
\end{aligned}
\tag{6.3.2}
$$

6.4 Simulation

Because of the complexity of the calculations involved in deriving the expressions given in the previous section, a simulation study was conducted to present a test of validity by showing close agreement between the empirical results of the simulation and the analytical results of the calculations. In order to simulate multiple instances of the random variables X and Z introduced in Section 6.2, a simulation program was written using the MATLAB software package. The program code is included in Appendix B found on Page 174. This section will describe how the simulation study was carried out and will also present the results of that study.

6.4.1 Empirical Sampling of X and Z

The following procedure was used to generate empirical samples of X and Z. First, sample values of the inter-arrival times for the signals $S_1, S_2, \ldots, S_n$ were simulated using the inverse transformation method described on Page 644 of "Introduction to Probability Models" by Ross[78]. This means that using MATLAB to simulate a random number u, between 0 and 1, this number was transformed into an instance of an exponential random variable t with parameter ν by the following operation

$$t = -\frac{\ln u}{\nu}.$$

A similar application of the inverse transformation method was used to calculate a sample instance of the inter-arrival time between S_k and Z. The sum of the inter-arrival times between $S_0, S_1, S_2, \ldots, S_k$ and the inter-arrival time between S_k and Z yields the value of a sample instance of Z.

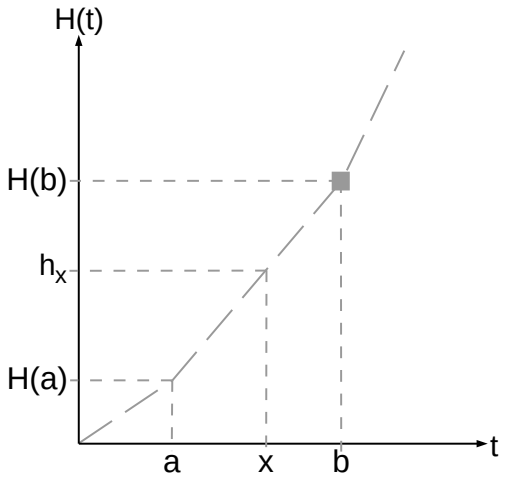


Figure 6.2: A Graph Demonstrating the Inverse Transformation of a Sample Instance of the Cumulative Hazard Function of X.

To construct a sample instance of X, another application of the inverse transformation method was used (with $\nu = 1$) to calculate a sample instance h_x of the cumulative hazard rate of X. The cumulative hazard function $H(t)$ of X was constructed and this was used to inverse transform h_x into a sample instance of X. To construct $H(t)$, it was only necessary to explicitly calculate it's value at each signal time. Because the hazard rate is constant between signals it was enough to apply the following equation for linear interpolation to calculate the sample instance of X,

$$x = a + \lambda(h_X - H(a)),$$

where a and b are two consecutive signal times such that $H(a) \leq h_X < H(b)$ and λ is the constant hazard rate between a and b. This can be seen clearly in Figure 6.2, which depicts the graph of the cumulative hazard function of X against time t.

6.4.2 Computation of Empirical and Analytical Expressions

The discussion will now turn to how the sampling procedure for X and Z discussed in the previous sub-section was used to develop empirical estimates of the analytical expressions calculated in Section 6.3 and also how these estimates were compared with the analytical expressions. The purpose of this study was merely to provide additional assurance that no arithmetic or logical

errors had been made during the calculations of the previous section. As such, when discussing the selection of simulation parameters it was not considered necessary to provide highly rigorous justification.

Firstly, a large number of sample pairs of X and Z were generated. The number of samples to generate was chosen to be $M = 20000$, because this value provided a large enough sample to allow for good statistical comparisons without requiring excessive computational time.

Several different sequences of gamma and lambda parameters were also chosen, along with values of n, k and ρ. The parameter sets that were chosen are listed in Table 6.1. These parameter sets were chosen based on commonly encountered values whilst also attempting to provide a range of different behaviours in order to reduce the possibility of coincidental good agreement for a specific configuration of parameters (in other words, to try to avoid the possibility that although the analytical expressions are incorrect, they coincidentally happen to yield the correct results for a specific set of parameters).

Table 6.1: A Table to show the Different Parameter Sets Used in the Simulation Study.

Parameter Set No.	$\boldsymbol{n}$	$\boldsymbol{k}$	$\boldsymbol{\rho}$	$\boldsymbol{\alpha}$ (in order)	$\boldsymbol{\lambda}$ (in order)
1	5	3	0.20	0.50,0.33,0.23,0.14	0.01,0.02,0.03,0.09,0.04
2	5	3	0.20	0.50,0.33,0.23,0.14	0.05,0.03,0.01,0.04,0.06
3	5	3	0.20	0.17,0.24,0.63,0.80	0.05,0.03,0.01,0.04,0.06
4	5	3	0.20	0.17,0.24,0.31,0.12	0.05,0.03,0.01,0.04,0.06
5	5	3	0.61	0.17,0.24,0.31,0.12	0.06,0.03,0.05,0.04,0.01
6	7	2	0.40	0.17,0.24,0.31,0.12,0.41,0.44	0.06,0.03,0.05,0.04,0.02,0.10,0.08
7	3	2	0.70	0.41,0.44	0.02,0.10,0.08

Having constructed a sequence $\{x_i, z_i\}_{1 \leq i \leq M}$ of M sample pairs based on a given parameter set, an array of time values was selected. These values were chosen uniformly between 0 and q, where both q and the granularity of the array were chosen in order to capture the significant behaviour of the four functions under consideration without requiring excessive computational time. The values of the analytical expressions were calculated for each of the time points in the array. The empirical estimates for these analytical expressions were also calculated for each time point t in the array,

using the following expressions,

$$\hat{F}_X(t) = \frac{1}{M}\sum_{i=1}^{M} I\{x_i < t\}$$
$$\hat{F}_Z(t) = \frac{1}{M}\sum_{i=1}^{M} I\{z_i < t\}$$
$$\hat{F}^*_X(t) = \frac{1}{M}\sum_{i=1}^{M} I\{x_i < t, x_i < z_i\},$$

where the indicator function I equals 1 if the arguments within the brackets are all true and 0 otherwise. Figure 6.3 is a graph depicting all three analytical expressions for the first set of parameter values in Table 6.1.

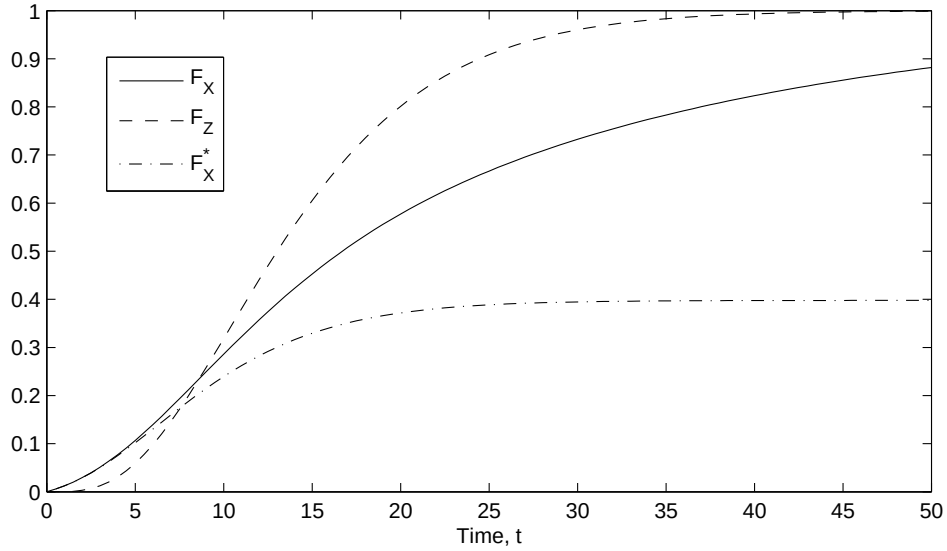
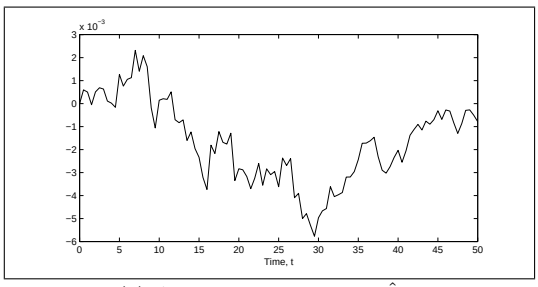


Figure 6.3: A Graph of the Marginal Distribution Functions of Both X and Z, and Also the Sub-distribution Function of X, for the First Parameter Set Used in the Simulation Study.

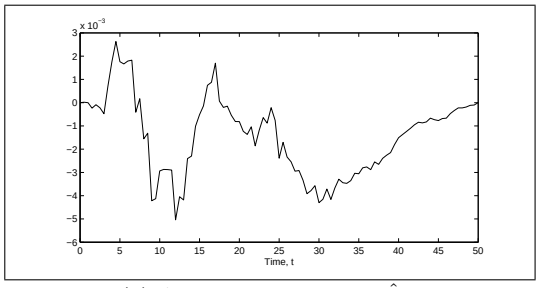
At this point several elementary visual confirmations of correctness can be taken from these types of graphs. Firstly, the two marginal distributions are both increasing functions with asymptotes at infinity equal to 1. Secondly, the sub-distribution function of X is an increasing function with an asymptote between 0 and 1 as t tends to infinity. In this case the asymptotic value to three decimal places for the sub-distribution function of X is $\lim_{t\to+\infty} S^*_X(t) = 0.401$, which indeed is between 0 and 1.

6.4.3 Comparison of Empirical and Analytical Expressions

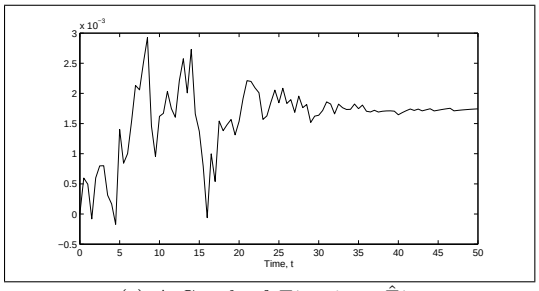
As a first step towards comparing the empirical and analytical expressions computed in the previous sub-section, graphs were constructed of the difference between each corresponding pair, similar to those depicted in Figure 6.4. It can be seen that these graphs appear to depict a "white-noise" process, centred around zero and with the axes scaled by 10^{-3} - which is exactly what one would expect if the analytical expressions were correct.



(a) A Graph of F_X minus $\hat{F}_X$.



(b) A Graph of F_Z minus $\hat{F}_Z$.



(c) A Graph of F_X^* minus $\hat{F}_X^*$.

Figure 6.4: A Series of Graphs Showing the Differences Between the Empirical and Analytical Expressions Computed in the Simulation Study.

In order to more rigorously assess the agreement between the corresponding empirical and analytical expressions, the Kolmogorov-Smirnov test was used. This test is based on the following statistic for a given distribution function $F(t)$and it's empirical counterpart $\hat{F}_n(t)$, where n is the number of samples used to construct the empirical distribution function,

$$D_n = \sup_{t \in [0,q]} \left| \hat{F}_n(t) - F(t) \right|.$$

The test works by rejecting the null hypothesis (that $\hat{F}_n$ converges to F over the domain $[0, q]$) at level α if

$$\sqrt{n} D_n > K_\alpha$$

where K_α is the critical value of the Kolmogorov distribution K (in other words, given α the value of K_α is such that $P\{K \leq K_\alpha\} = 1 - \alpha$). The details of this test can be found in Chapter 30 of Stuart and Ord[80].

Table 6.2: A Table to Show the Values of the Kolmgorov Statistic for each Parameter Set Used in the Simulation Study.

Parameter Set No.	$\boldsymbol{n^{1/2}D_n}$**for** $\boldsymbol{F_X}$	$\boldsymbol{n^{1/2}D_n}$**for** $\boldsymbol{F_Z}$	$\boldsymbol{n^{1/2}D_n}$**for** $\boldsymbol{F_X^*}$
1	0.64	0.67	0.63
2	0.89	0.91	0.85
3	0.77	0.71	0.53
4	0.96	0.72	0.43
5	0.54	0.51	0.61
6	0.66	0.50	0.62
7	0.49	0.56	0.66

For each parameter set used in the study, the Kolmogorov statistics for the three analytical expressions were computed and these are contained in Table 6.2. Given that n was very large, the asymptotic values of the Kolmogorov distribution as n tends to infinity were used for the critical values of K_α. Table 6.2 contains several critical values for different levels of α, which were taken from Lindley and Scott[60].

Table 6.3: A Table to Show Several Critical Values of the Asymptotic Kolmogorov Distribution as n Tends to Infinity (Values shown are to three decimal places).

Significance Level ($\boldsymbol{\alpha}$)	**Critical Value** ($\boldsymbol{K_\alpha}$)
5%	1.358
1%	1.628
0.1%	1.949

It can be seen that none of the analytical expressions are rejected at these significance levels. This implies that the null hypothesis that the empirical functions converge to the analytic functions as the sample size tends to infinity is not rejected at these significance levels, which in turn implies that there is statistical evidence to suggest that the analytical expressions are correct.

6.5 Summary

This chapter has discussed the Random Signals model, which attempts to model opportunistic maintenance. The detail of this model was introduced and then the following analytical expressions were developed for the marginal functions of both X and Z, as well as the sub-distribution function of X.

$$
\begin{aligned}
F_X(t) &= 1 - \sum_{j=0}^{n} \left[\prod_{r=0}^{j-1} \gamma_r \right] \sum_{i=0}^{j} \frac{e^{-\delta_i t}}{\prod_{\substack{u=0 \\ u \neq i}}^{j} (\delta_u - \delta_i)}, \\
F_Z(t) &= \left[\prod_{i=0}^{k-1} \gamma_i \right] \sum_{j=0}^{k-1} \left(\frac{1 - e^{-\gamma_j t}}{\gamma_j \prod_{u \neq j, u=0}^{k-1} (\gamma_u - \gamma_j)} - \frac{e^{-\gamma_j t} - e^{-\rho t}}{[\rho - \gamma_j] \prod_{u \neq j, u=0}^{k-1} (\gamma_u - \gamma_j)} \right), \\
F_X^*(t) &= 1 - \sum_{u=0}^{k-1} \left[\prod_{r=0}^{u-1} \gamma_r \right] \sum_{j=0}^{u} \frac{e^{-\delta_j t}}{\prod_{\substack{u=0 \\ u \neq j}}^{u} (\delta_u - \delta_j)} \\
&\quad + \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r \right] \frac{1}{\lambda_u + \rho} \sum_{q=0}^{k-1} \frac{\rho[\delta_q - \lambda_u - \rho] \left(1 - e^{-\delta_q t}\right) - \delta_q \lambda_u e^{-\delta_q t}}{\delta_q \left[\prod_{\substack{w=k \\ w \neq q}}^{u} (\delta_w - \delta_q + \rho) \right] \left[\prod_{\substack{v=0 \\ v \neq q}}^{k-1} (\delta_v - \delta_q) \right]} \\
&\quad + \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r \right] \frac{1}{\lambda_u + \rho} \sum_{q=k}^{u} \frac{\rho[\delta_q - \lambda_u] \left(1 - e^{-[\delta_q + \rho]t}\right) - \lambda_u [\delta_q + \rho] e^{-[\delta_q + \rho]t}}{[\delta_q + \rho] \left[\prod_{\substack{w=k \\ w \neq q}}^{u} (\delta_w - \delta_q) \right] \left[\prod_{\substack{v=0 \\ v \neq q}}^{k-1} (\delta_v - \delta_q - \rho) \right]}.
\end{aligned}
$$

Given the number and complexity of the algebraic manipulations required to reach these results, a simulation study was conducted and this provided good graphical and statistical evidence that these calculations were correct.

Chapter 7

Variant of the Kaplan Meier Estimator (for Missing Data)

This chapter addresses the problem of trying to estimate latent variables for two independent competing risks when knowledge of which risk caused the system shut-down is unknown for some of the observations in the dataset.

Firstly, a review of attempts to address the missing data problem is presented. This is followed by a derivation of the Kaplan-Meier product limit estimator which is given to serve as a useful means of comparison with the newly proposed estimator. The remainder of this chapter will present the proposed variant of the Kaplan-Meier product limit estimator for the case of missing data. This estimator is first developed formally, then a redistribution to the right algorithm is also developed. These two approaches are demonstrated to be equivalent and some examples of their application are given. Finally, the estimator is shown to be consistent under certain conditions and an example satisfying these conditions is given.

7.1 Review of Competing Risk Models With Missing Data

In many situations there may be great difficulty in unambiguous identification of a failure mode for a given event. This problem has received quite a bit of attention in the past and the following discussion will seek to compare and contrast several of the approaches used. Little and Rubin[64] define several classes of mechanisms that lead to missing data. They consider a complete dataset

$Y = y_{ij}$ (a set of i observations of a vector of j components), with $M = (M_{ij})$ defined as the missing data indicator. Y is divided into two subsets Y_{OBS} and Y_{MIS} which denote the observed and missing components of Y respectively. The mechanism is characterised by the conditional distribution of M given Y, say $f(M|Y, \phi)$, where ϕ denotes unknown parameters. If "missingness" does not depend on Y, i.e.

$$f(M|Y, \phi) = f(M|\phi) \text{ for all } Y, \phi$$

the missing data mechanism is called missing completely at random (MCAR). If "missingness" only depends on Y_{OBS}, i.e.

$$f(M|Y, \phi) = f(M|Y_{OBS}, \phi) \text{ for all } Y_{MIS}, \phi$$

then the missing data mechanism is called missing at random (MAR). If "missingness" depends on any of the values of Y_{MIS} then the missing data mechanism is called not missing at random (NMAR). This type of mechanism usually requires knowledge of the parametric form to make any reasonable progress.

The first person to really explore the question of competing risks with missing failure modes is Dinse[37]. Given a system that can experience $J = \{1, \cdots, k\}$ different failure modes, for the i^{th} event define T_i as the time until a failure occurs and $\delta_i \in J$ as the failure mode of the event. Dinse identified four different types of missing data in this competing risk setup:

I. $\{T_i = t, \delta_i = j\}$

II. $\{T_i > t\}$

III. $\{T_i > t, \delta_i = j\}$

IV. $\{T_i = t\}$

The first type corresponds to the complete data setup and the second type corresponds to typical right censoring. The third type is a little unusual in that the failure occurrence is right-censored but the failure mode is known in advance. The fourth type is of most relevance in this discussion and is when failure times are known but the failure mode of the corresponding event is not known. In this case Dinse treated the missing data of type four as MAR. Dinse developed an EM algorithm to add pseudo-data, turning observations of type 3 into those of type I, in order to calculate estimators from the likelihood function.

Miyakawa[68] Looks at a competing risk setup with two independent latent variables X and Z, and with missing δ_i's. Again, it is assumed that the "missingness" mechanism is MAR. Maximum likelihood equations are developed in the case of exponential marginals giving closed form estimators for the parameters λ_X and λ_Z. In the more general non-parametric case an EM algorithm is applied using several different forms of estimator for the missing data. The results are compared in a simulation study, which shows that the normal Product Limit estimator has a much higher bias and mean squared error than the other newly proposed estimators in this paper.

Lo[65] also considers a competing risk model with two independent failure modes. In this case a missing data indicator ϵ_i is introduced for the i^{th}observation. The probability $P(\epsilon_i = 1)$ is modelled as a constant, so the mechanism of "missingness" is NMAR. Lo shows that the set of maximum likelihood estimators and the set of self-consistent estimators are equal and infinite, although not all estimators in this set are consistent. Two new estimators are proposed which are proven to be consistent and to convergence to a Gaussian process.

Goetghebeur and Ryan[49] consider a proportional hazards competing risk model with two failure modes and develop a set of estimators. In this case the "missingness" mechanism is taken to be MAR and the probability of ϵ_i is said to depend only on T_i.

Dewanji and Sengupta[34] consider a general missing cause pattern where the observed data consists of the failure time and a set of i possible causes (from the main set of J causes), denoted by the random variable G. They also include the possibility of general right censoring ($\delta_i = 0$ and no knowledge of G or j) for each observation. Firstly under the MAR assumption the EM algorithm is used, the pseudo-dataset is taken to be the complete dataset with the failure causes fully known. Under quite general assumptions about the data the EM algorithm is shown to converge to a global maximum. Secondly, it is assumed that the observed masking set is independent of the actual cause of failure and the failure time. By applying Bayes theorem

$$P[G = g|J = j] = \frac{P[J = j|G = g]P[G = g]}{\sum_{g' \ni j} P[J = j|G = g']P[G = g']}.$$

It is argued that in certain circumstances experimentalists are able to use laboratory testing to ascertain estimates of $P[J = j|G = g]$. Assuming such estimates are given, a Nelson-Aalen estimator is developed using Martingale Theory. Both cases assume that the "missingness mechanism is MAR, in the second case the extra assumption of time independence is also made. Flehinger et. al.[44] also consider a competing risk setup with general missing cause pattern. In this case a

second stage of analysis is explicitly introduced into the notational framework, the idea being that a subset of the items with non-singleton masking sets are taken to a second stage of analysis (typically some form of laboratory testing) where the true cause of failure is determined for these items. It is assumed that the competing risks are all independent and that the "missingness" mechanism does not depend on time, although it may depend on the underlying cause. Again, Bayes formula is applied to study the diagnostic probabilities $P_{G|J}$. Maximum likelihood estimators are obtained using an EM algorithm to estimate the parameters $P_{G|J}$ and P_J.

Craiu and Duchesne[28] also look at a general pattern of missing data with n competing risks and the presence of possible right censoring. Their approach is based on parametric modeling and again, Bayes formula is applied to study the diagnostic probabilities $P_{G|J}$. An EM algorithm is developed to generate maximum likelihood estimates of the cause-specific hazard functions. The assumption of MAR is not made in this framework and it is pointed out that the framework provides for the possibility of carrying out several important statistical tests on assumptions made in other models. Examples of such include

- $P_{g|j}(t) = P_{g|j}$ (masking is independent of the failure time);
- $P_{g|j} = P_g$ (masking is independent of the actual failure mode);
- $P_{\{1\}|1} = \cdots = P_{\{J\}|J}$ (probability of identification at first stage is independent of actual failure mode).

Craiu and Duchesne apply this framework to both a piece-wise constant hazards model and a Weibull model in a simulation study to show the quality of the resulting estimators.

In both the work by Flehinger et. al. and the work by Craiu and Duchesne, a second stage analysis is considered. The criteria for selecting those cases taken to a second stage analysis is assumed to depend only on the observed data in both cases. This seems strange, especially given the extra notational burden imposed by the second stage analysis, because it would seem to imply that elements taken to the second stage analysis are equivalent to elements fully identified at the first stage.

Racine and Hoel [77] consider a slightly different problem. They consider the case when doctors are uncomfortable with attributing cause of death to a given failure mode. This situation is not quite the same as above in that knowledge of the failure mode is incomplete, not completely missing. They consider the case of two independent failure modes and rather than force classification of

death by risk 1 or death by risk 2, they use the following weighted indicator:

$$P_i = \begin{cases} 0 & \text{if } R_1 \text{ is definitely not the COD (Cause of Death);} \\ q_1 & \text{if } R_1 \text{ is probably not the COD;} \\ q_2 & \text{if } R_1 \text{ is probably the COD;} \\ 1 & \text{if } R_1 \text{ is definitely the COD.} \end{cases}$$

This indicator is used to construct a Kaplan-Meier type estimator and demonstrate using a simulation study that forcing classification can lead to considerable bias (which is significantly greater than for the new estimator).

7.2 Derivation of the Kaplan-Meier Product Limit Estimator

This section will present a derivation of the Kaplan-Meier product limit estimator. The derivation is heavily influenced by that given in Deshpande and Purohit[32], as well as a series of lecture notes prepared by Dewan[33].

7.2.1 The Formal Derivation

Consider a set of independent identically distributed failure times $X_1, \cdots, X_n$ and a set of independent identically distributed censoring times $Z_1, \cdots, Z_n$ which represent the times at which the failure times would be right censored if the item in question has not already failed beforehand. Assume that for each i, X_i and Z_i are independent. One can observe the pairs (T_i, δ_i), where $T_i = \min(X_i, Z_i)$ and $\delta_i = I(X_i < Z_i)$.

The Product Limit estimator can be derived in the following manner. Place the T_i in order such that:

$$T_{(1)} < T_{(2)} < \cdots < T_{(n)}.$$

One can consider the real line as being partitioned into a series of intervals by the $T_{(i)}$ such that for all i from 1 to n, the half-open interval $I_i = [T_{(i-1)}, T_{(i)})$, where $T_{(0)} = 0$. The following definitions will prove useful during the derivation.

$T_{(i)}$ = the right endpoint of I_i.

n = the number of items under observation at the start.

$R(t)$ = the number of items that have survived
(their failure times have not occurred or been censored) to t^-.

n_i = the number of items that have survived to the beginning of I_i,
equal to $R(T_{(i)}) = n - (i-1)$.

d_i = the number of items that fail (failure times that occur) during I_i
(by definition these occur at $T_{(i)}$).

p_i = $P\{\text{item survives through } I_i | \text{that item had survived to beginning of } I_i\}$.

q_i = $1 - p_i$.

Assuming that there are no ties among the observations, the following serve as natural estimators for p_i and q_i.

$$\hat{p_i} = \left\{ \begin{array}{ll} 1 - \frac{1}{n_i} & \text{if } \delta_{(i)} = 1 \\ 1 & \text{if } \delta_{(i)} = 0 \end{array} \right\}, \text{ and } \hat{q_i} = \frac{1}{n_i}.$$

So for $t > T_{(i)}$, the survivor function can be estimated by

$$\hat{S}(t) = \prod_{(i|T_{(i)} \leq t)} \hat{p_i} = \prod_{(i|T_{(i)} \leq t)} \left(1 - \frac{1}{n_i}\right)^{\delta_{(i)}} = \prod_{(i|T_{(i)} \leq t)} \left(\frac{n-i}{n-i+1}\right)^{\delta_{(i)}}. \tag{7.2.1}$$

In the event that there are ties amongst uncensored failure times, one can assume that they have occurred in a distinct sequence with infinitesimally small distances between time points. This leads to the following contribution to the product limit estimator

$$\left(1 - \frac{d_i}{n_i}\right).$$

In the event that there are ties between censored and uncensored observations, one assumes that the uncensored observations occur before the censored observations. Lastly, if the final observation $T_{(n)}$ is censored then $\hat{S}(t) > 0$ for $t > T_{(n)}$, so the following definition is made.

$$\lim_{t \to +\infty} \hat{S}(t) = 0.$$

So by considering the ordering of distinct $T_{(i)}$'s ($T'_{(1)} < T'_{(2)} < \cdots < T'_{(m)}$) (with $m \leq n$), the Product Limit estimator with possible ties between observations can be defined as follows, for $t > T'_{(i)}$

$$\hat{S}(t) = \prod_{(i|T'_{(i)} \leq t)} \left(1 - \frac{d_i}{n_i}\right). \tag{7.2.2}$$

7.2.2 The Re-distribution to the Right Algorithm

Another way to derive the Product limit estimator was proposed by Efron[42] and is called the re-distribution to the right algorithm. This algorithm is a means of constructing the probability mass function. As before the case of no ties will be considered first and the case of ties will be developed second. Given the set of n pairs (T_i, δ_i), again consider the ordering

$$T_{(1)} < T_{(2)} < \cdots < T_{(n)}$$

defined in the previous sub-section.

Step 1: for each $T_{(i)}$ place probability mass of $\frac{1}{n}$ at $T_{(i)}$.

Step 2: proceeding in an iterative fashion consider the first observation ($i = 1$).

Step 3: if $\delta_{(i)} = 0$, re-distribute the probability mass at $T_{(i)}$ equally, at each of the $T_{(j)}$ such that $j > i$.

Step 4: repeat step 3 for each i in ascending order.

Let k equal the largest integer less than or equal to n such that $\delta_{(k)} = 1$. If $k < n$ then the density function is undefined for all $t > T_{(k)}$. In the event that there are ties between censored and uncensored observations, assume that the uncensored observations occur before the censored observations. Lastly, it should be noted that the assumption regarding ties between censored and uncensored observations is intuitively obvious in this derivation. If one assumes the contrary, that the uncensored observations occur after the censored observations, then some of the mass at the censored observations would be redistributed to those uncensored observations, i.e. not be re-distributed at all. Intuitively this would be incorrect because it is known that the censored lifetime variables are greater than or equal to the tied uncensored observations. From this algorithm,

censored observations re-distribute their mass to the right, only uncensored observations keep their mass at the observation time.

7.3 Derivation of the Proposed Estimator for Missing Data

This section will present a formal derivation of a variant of the Kaplan-Meier product limit estimator with missing data regarding the cause of failure for some of the observations. A redistribute to the right algorithm will also be developed.

7.3.1 The Problem Definition

Consider a set of independent identically distributed lifetimes $X_1, \cdots, X_n$ and a set of independent identically distributed censoring times $Z_1, \cdots, Z_n$ which represent the times at which the lifetimes would be right censored if they have not already occurred beforehand. Assume that for each i, X_i and Z_i are independent. Under ordinary right censoring one can observe the n pairs (T_i, δ_i), where $T_i = \min(X_i, Z_i)$ and $\delta_i = I(X_i < Z_i)$. However in the case of missing data one only observes the n pairs (T_i, ϵ_i) and the set $\{\delta_i : \epsilon_i = 1\}$, where

$$\epsilon_i = \begin{cases} 1 & \text{if } \delta_i \text{ is observed} \\ 0 & \text{otherwise.} \end{cases}$$

The task is to find a suitable estimator, similar to the product limit estimator, for the marginal distribution of the failure times in this case. Motivated by a consideration of the re-distribution to the right algorithm it seems reasonable to consider the following strategy. Assume that δ_i is known for all observations except the j^{th} such observation. There are two extreme possibilities, which are to regard δ_j as censored or uncensored, the two corresponding product limit estimators could then be calculated in the usual manner. It is clear that a good estimator should lie between these two cases. Thinking in terms of the re-distribution to the right algorithm, the two extreme cases correspond to the choice of redistributing all or none of the mass to the right of T_j. So it seems natural to consider the possibility of re-distributing a portion of the mass to the right and leaving the remaining portion at T_j. The nature of this proportion will be discussed more thoroughly in Subsection 7.7.1. For example, this remaining portion may be a fixed constant β, or it may depend on T_j (denoted $\beta(T_j)$). These ideas will be formalised and generalised to k such missing indicator

variables in the next subsection.

7.3.2 The Formal Derivation

The missing data version of the Product Limit estimator can be derived in the following manner, considering first the case of no ties between failure times. Place the T_i in order such that

$$T_{(1)} < T_{(2)} < \cdots < T_{(n)}.$$

As before, one can consider the real line as being partitioned into a series of intervals I_i. The following extra definitions will prove useful during the derivation.

$$\begin{aligned}
l_i &= \text{the number of items that are lost with no indication of failure or censoring during } I_i \text{ (by definition these occur at } T_{(i)}). \\
\beta(T_{(i)}) &= P\{\delta^*_{(i)} = 1 | \epsilon_{(i)} = 0, T_{(i)}\}, \text{ where } \delta^*_{(i)} \text{ is defined to be the value of the unobserved indicator variable.} \\
u_i &= P\{\text{item survives through } I_i | \text{that item had survived to beginning of } I_i\}. \\
v_i &= 1 - u_i.
\end{aligned}$$

Note that $\beta(T_{(i)})$ is the remaining portion that will not be redistributed in the case of a missing data observation, as discussed above. In the more general case where $\delta_{(i)}$ might take on several values from a set $J = \{0, \cdots, r\}$ with $r > 1$, $\beta(T_{(i)})$ would need to be defined for each possible value of J.

In the case of no ties, the following serve as natural estimators for u_i and v_i.

$$\hat{v_i} = \left\{ \begin{array}{ll} \frac{1}{n_i} & \text{if } \delta_{(i)} = 1, \epsilon_{(i)} = 1 \\ 0 & \text{if } \delta_{(i)} = 0, \epsilon_{(i)} = 1 \\ \frac{\beta(T_{(i)})}{n_i} & \epsilon_{(i)} = 0 \end{array} \right\}, \text{ and } \hat{u_i} = \left\{ \begin{array}{ll} 1 - \frac{1}{n_i} & \text{if } \delta_{(i)} = 1, \epsilon_{(i)} = 1 \\ 1 & \text{if } \delta_{(i)} = 0, \epsilon_{(i)} = 1 \\ 1 - \frac{\beta(T_{(i)})}{n_i} & \epsilon_{(i)} = 0 \end{array} \right\}.$$

So for $t > T_{(i)}$ the survival function can be estimated by

$$\begin{aligned}\hat{S}(t) &= \prod_{(i|T_{(i)}\leq t)} \hat{u_i} \\ &= \prod_{(i|T_{(i)}\leq t)} \left(1-\frac{\beta(T_{(i)})}{n_i}\right)^{1-\epsilon_{(i)}} \left(1-\frac{d_i}{n_i}\right)^{\epsilon_{(i)}\delta_{(i)}} \\ &= \prod_{(i|T_{(i)}\leq t)} \left(1-\frac{\beta(T_{(i)})}{n_i}\right)^{1-\epsilon_{(i)}} \left(1-\frac{1}{n_i}\right)^{\epsilon_{(i)}\delta_{(i)}} . \qquad (7.3.1)\end{aligned}$$

In the event that there are ties amongst uncensored failure times, one can assume that they have occurred in a distinct sequence with infinitesimally small distances between time points. This leads to the following contribution to the product limit estimator

$$\left(1-\frac{d_i}{n_i}\right).$$

In the event that there are ties between censored and uncensored observations, one assumes that the uncensored observations occur before the censored observations as in the case of the normal Product Limit estimator. The remaining cases involve ties with missing data and censored or uncensored observations. In this case one assumes that the uncensored observations occur before the missing data observations, which occur before the censored observations. This will be seen to be quite sensible when considered from the perspective of the re-distribution to the right algorithm in Subsection 7.3.3. Assume that m individuals are alive just before time t and at time t, d individuals die and l individuals are lost without any indication as to the cause (death or censoring). Using the above ordering, the resulting contribution to the product limit estimator due to events occurring at t will be the following

$$\begin{aligned}&\left(1-\frac{1}{m}\right)\cdots\left(1-\frac{1}{(m-d+1)}\right)\left(1-\frac{\beta(T_{(i)})}{(m-d)}\right)\cdots\left(1-\frac{\beta(T_{(i)})}{(m-d-l+1)}\right) \\ = \;& \frac{m-1}{m}\cdots\frac{m-d}{(m-d+1)}\times\prod_{k=1}^{l}\left(1-\frac{\beta(T_{(i)})}{(m-d+1-k)}\right) \\ = \;& \left(1-\frac{d}{m}\right)\prod_{k=1}^{l}\left(1-\frac{\beta(T_{(i)})}{(m-d+1-k)}\right).\end{aligned}$$

Lastly, if the final observation $T_{(n)}$ is censored or involves missing data then $\hat{S}(t)>0$ for $t>T_{(n)}$,

so the following definition is made.

$$\lim_{t\to+\infty} \hat{S}(t) = 0. \tag{7.3.2}$$

So by considering the ordering of distinct $T_{(i)}$'s ($T'_{(1)} < T'_{(2)} < \cdots < T'_{(m)}$) (with $m \leq n$), the general case of the product limit estimator for missing data with possible ties between observations can be defined as follows, for $t > T'_{(i)}$

$$\hat{S}(t) = \prod_{(i|T'_{(i)}\leq t)} \left(1 - \frac{d_i}{n_i}\right) \left(\prod_{k=1}^{l_i} \left(1 - \frac{\beta(T'_{(i)})}{(n_i - d_i + 1 - k)}\right)\right). \tag{7.3.3}$$

7.3.3 The Re-distribution to the Right Algorithm

This subsection will present an algorithm to construct the probability mass function in this case. As before the case of no ties will be developed first, then a justification of the presumed ordering of tied events will be given.

Given the set of n observations of the form (T_i, ϵ_i) when $\epsilon = 0$ and $(T_i, \epsilon_i, \delta_i)$ when $\epsilon = 1$, again consider the ordering

$$T_{(1)} < T_{(2)} < \cdots < T_{(n)}$$

defined in the previous sub-section.

Step 1: for each $T_{(i)}$ place probability mass of $\frac{1}{n}$ at $T_{(i)}$.

Step 2: proceeding in an iterative fashion consider the first observation ($i = 1$).

Step 3: if $\epsilon_{(i)} = 1$ and $\delta_{(i)} = 0$, re-distribute the probability mass at $T_{(i)}$ equally, at each of the $T_{(j)}$ such that $j > i$.

Step 4: if $\epsilon_{(i)} = 0$, re-distribute the proportion $1 - \beta(T_{(i)})$ of the probability mass at $T_{(i)}$ equally, at each of the $T_{(j)}$ such that $j > i$.

Step 5: repeat steps 3 and 4 for each i in ascending order.

Let k equal the largest integer less than or equal to n such that $\epsilon_{(k)} = 1$ and $\delta_{(k)} = 1$. If $k < n$ then then the density function is undefined for all $t > T_k$.

As mentioned in subsection 7.3 in the case of ties, assume that the uncensored observations occur before the missing data observations, which occur before the censored observations.This assumption is intuitively obvious in this derivation. Assume that a tie has occurred between the three different types of observation. If one assumes an ordering without censored observations being considered as last, then some or all of the mass initially placed at a censored observation will not be re-distributed at all, which is not correct.

If missing data observations are assumed to occur before uncensored observations, then a portion of the mass initially placed at a missing data observation would be re-distributed to the right only as far as the tied uncensored observations and similarly this would mean that that proportion, which should be re-distributed, is in fact not re-distributed at all.

So both alternatives are intuitively incorrect because in the first case the resulting density function would indicate that censored data are treated in the same way as either missing or uncensored data. In the second case missing data does not re-distribute as much mass as it should.

7.4 Equivalence of the Two Derivations for the Proposed Estimator

This subsection will establish that the formal derivation of the Kaplan-Meier product limit estimator with missing data is equivalent to the re-distribution to the right algorithm derived in the previous subsection.

First a general expression for the probability mass will be derived.

Theorem 7. *For k from 1 to n, after the k^{th} iteration through steps three and four of the proposed redistribution to the right algorithm,*

I. The probability mass at $T_{(k)}$ will equal:

- *If $\epsilon_{(k)} = 1$ and $\delta_{(k)} = 0$, then the probability mass present at $T_{(k)}$ will equal 0.*
- *If $\epsilon_{(k)} = 1$ and $\delta_{(k)} = 1$, then the probability mass present at $T_{(k)}$ will equal*

$$\frac{1}{n}\prod_{j=1}^{k-1}\left(\frac{(n-j+1-\beta(T_{(j)}))}{n-j}\right)^{1-\epsilon_{(j)}}\left(\frac{(n-j+1)}{n-j}\right)^{\epsilon_{(j)}(1-\delta_{(j)})}.$$

- *If $\epsilon_{(k)} = 0$, then the probability mass present at $T_{(k)}$ will equal*

$$\frac{\beta(T_{(k)})}{n} \prod_{j=1}^{k-1} \left(\frac{(n-j+1-\beta(T_{(j)}))}{n-j} \right)^{1-\epsilon_{(j)}} \left(\frac{(n-j+1)}{n-j} \right)^{\epsilon_{(j)}(1-\delta_{(j)})}.$$

II. For any r such that $k < r \leq n$, the probability mass at $T_{(r)}$ will equal:

$$\frac{1}{n} \prod_{j=1}^{k} \left(\frac{(n-j+1-\beta(T_{(j)}))}{n-j} \right)^{1-\epsilon_{(j)}} \left(\frac{(n-j+1)}{n-j} \right)^{\epsilon_{(j)}(1-\delta_{(j)})}.$$

In the event of ties the contribution from each observation is calculated using the order of uncensored observations, missing data observations and then censored observations. The resulting probability mass at that point will be the sum of these contributions.

Proof. First the case of no ties will be considered. It is clear from the algorithm that probability mass is only placed at points of observation, so it is only necessary to consider the probability mass at those points. When iterating through steps three and four of the algorithm, probability mass at each observation point $T_{(i)}$ is re-distributed. In every case no probability mass is re-distributed to an earlier point in time. Also, if probability mass is re-distributed to the right, this is done equally to every observation. Thus at each iteration of the algorithm it is only necessary to consider the probability mass at that observation point $T_{(i)}$ and the probability mass at any one of the observation points to the right of $T_{(i)}$. The proof will proceed in an inductive fashion, the inductive step being each iteration through steps three and four of the proposed algorithm.

For the base case there are three possible values for the probability mass at $T_{(1)}$:

- If $\epsilon_{(1)} = 1$ and $\delta_{(1)} = 0$, then the algorithm will redistribute all of the probability mass at $T_{(1)}$ to the right, leaving zero probability mass at $T_{(1)}$ as required.
- If $\epsilon_{(1)} = 1$ and $\delta_{(1)} = 1$, then the algorithm will not redistribute any of the initial probability mass placed at $T_{(1)}$, leaving $\frac{1}{n}$ probability mass at $T_{(1)}$ as required
- If $\epsilon_{(1)} = 0$, then the algorithm will re-distribute the proportion $1 - \beta(T_{(1)})$ of the probability mass at $T_{(1)}$ to the right, leaving $\frac{\beta(T_{(i)})}{n}$ probability mass at $T_{(1)}$ as required.

For any r such that $k < r \leq n$, there are three possible values for the probability mass at $T_{(r)}$:

- If $\epsilon_{(1)} = 1$ and $\delta_{(1)} = 0$, then the algorithm will redistribute all of the $\frac{1}{n}$ probability mass equally among the remaining $n-1$ observation points. So the probability mass at $T_{(r)}$ will

equal

$$\frac{1}{n}\left(1+\frac{1}{n-1}\right)=\frac{1}{(n-1)}.$$

From the formula given in the statement of the theorem, the probability mass at $T_{(r)}$ is equal to

$$\frac{1}{n}\left(\frac{(n-1+1)}{n-1}\right)=\frac{1}{n-1},$$

as required.

- If $\epsilon_{(1)}=1$ and $\delta_{(1)}=1$, then the algorithm will not redistribute any of the initial probability mass placed at $T_{(1)}$, leaving $\frac{1}{n}$ probability mass at $T_{(r)}$. From the formula given in the statement of the theorem, the probability mass at $T_{(r)}$ is equal to $\frac{1}{n}$, as required.

- If $\epsilon_{(1)}=0$, then the algorithm will re-distribute the proportion $1-\beta(T_{(1)})$ of the probability mass $\frac{1}{n}$ at $T_{(1)}$ among the remaining $n-1$ observation points. So the probability mass at $T_{(r)}$ will equal

$$\frac{1}{n}\left(1+\frac{1-\beta(T_{(1)})}{n-1}\right)=\frac{1}{n}\frac{n-1+1-\beta(T_{(1)})}{n-1}=\frac{1}{n}\frac{n-\beta(T_{(1)})}{n-1}.$$

From the formula given in the statement of the theorem, the probability mass at $T_{(r)}$ is equal to

$$\frac{1}{n}\left(\frac{(n-1+1-\beta(T_{(1)}))}{n-1}\right)=\frac{1}{n}\frac{(n-\beta(T_{(1)}))}{n-1},$$

as required.

For the inductive step of the proof, for any r such that $i<r\leq n$, consider the probability mass at $T_{(r)}$ before the $(i+1)^{th}$ iteration through steps three and four of the algorithm. By the inductive hypothesis, this will be equal to

$$\frac{1}{n}\prod_{j=1}^{i}\left(\frac{(n-j+1-\beta(T_{(j)}))}{n-j}\right)^{1-\epsilon_{(j)}}\left(\frac{(n-j+1)}{n-j}\right)^{\epsilon_{(j)}(1-\delta_{(j)})}.$$

The final step will be to consider the probability mass at each $T_{(r)}$ after the $(i+1)^{th}$ iteration through steps three and four of the algorithm. Note again that for all s such that $i+1<s\leq n$, the probability mass at each $T_{(s)}$ will be the same. Thus it will only be necessary to consider the probability mass at $T_{(i+1)}$ and any one of the observations points $T_{(s)}$.

Considering first the probability mass at $T_{(i+1)}$ after the $(i+1)^{th}$ iteration through steps three and four of the algorithm, there are three cases:

- If $\epsilon_{(i+1)} = 1$ and $\delta_{(i+1)} = 0$, then the algorithm will redistribute all of the probability mass at $T_{(i+1)}$ to the right, leaving zero probability mass at $T_{(i+1)}$ as required.

- If $\epsilon_{(i+1)} = 1$ and $\delta_{(i+1)} = 1$, then the algorithm will not redistribute any of the probability mass placed at $T_{(i+1)}$, leaving probability mass at $T_{(i+1)}$ equal to

$$\frac{1}{n}\prod_{j=1}^{i}\left(\frac{(n-j+1-\beta(T_{(j)}))}{n-j}\right)^{1-\epsilon_{(j)}}\left(\frac{(n-j+1)}{n-j}\right)^{\epsilon_{(j)}(1-\delta_{(j)})},$$

  as required.

- If $\epsilon_{(i+1)} = 0$, then the algorithm will re-distribute the proportion $1-\beta(T_{(i+1)})$ of the probability mass at $T_{(i+1)}$ to the right, leaving the proportion $\beta(T_{(i+1)})$ of probability mass at $T_{(i+1)}$. So the probability mass at $T_{(i+1)}$ will equal

$$\frac{\beta(T_{(i+1)})}{n}\prod_{j=1}^{i}\left(\frac{(n-j+1-\beta(T_{(j)}))}{n-j}\right)^{1-\epsilon_{(j)}}\left(\frac{(n-j+1)}{n-j}\right)^{\epsilon_{(j)}(1-\delta_{(j)})},$$

  as required.

Finally, considering the probability mass at $T_{(s)}$ after the $(i+1)^{th}$ iteration through steps three and four of the algorithm, there are three cases:

- If $\epsilon_{(i+1)} = 1$ and $\delta_{(i+1)} = 0$, then the algorithm will redistribute all of the probability mass at $T_{(i+1)}$ equally among the remaining $n-i-1$ observation points. So the probability mass at $T_{(s)}$ will equal

$$\begin{aligned}
&\left(1+\frac{1}{n-i-1}\right)\frac{1}{n}\prod_{j=1}^{i}\left(\frac{(n-j+1-\beta(T_{(j)}))}{n-j}\right)^{1-\epsilon_{(j)}}\left(\frac{(n-j+1)}{n-j}\right)^{\epsilon_{(j)}(1-\delta_{(j)})}\\
=\ &\left(\frac{n-i}{n-i-1}\right)\frac{1}{n}\prod_{j=1}^{i}\left(\frac{(n-j+1-\beta(T_{(j)}))}{n-j}\right)^{1-\epsilon_{(j)}}\left(\frac{(n-j+1)}{n-j}\right)^{\epsilon_{(j)}(1-\delta_{(j)})}.
\end{aligned}$$

  From the assumption about $\epsilon_{(i+1)}$ and $\delta_{(i+1)}$, this is just equal to

$$\frac{1}{n}\prod_{j=1}^{i+1}\left(\frac{(n-j+1-\beta(T_{(j)}))}{n-j}\right)^{1-\epsilon_{(j)}}\left(\frac{(n-j+1)}{n-j}\right)^{\epsilon_{(j)}(1-\delta_{(j)})},$$

as required.

- If $\epsilon_{(i+1)} = 1$ and $\delta_{(i+1)} = 1$, then the algorithm will not redistribute any of the probability mass placed at $T_{(i+1)}$, leaving probability mass at $T_{(s)}$ equal to

$$\frac{1}{n}\prod_{j=1}^{i}\left(\frac{(n-j+1-\beta(T_{(j)}))}{n-j}\right)^{1-\epsilon_{(j)}}\left(\frac{(n-j+1)}{n-j}\right)^{\epsilon_{(j)}(1-\delta_{(j)})}.$$

From the assumption about $\epsilon_{(i+1)}$ and $\delta_{(i+1)}$, this is just equal to

$$\frac{1}{n}\prod_{j=1}^{i+1}\left(\frac{(n-j+1-\beta(T_{(j)}))}{n-j}\right)^{1-\epsilon_{(j)}}\left(\frac{(n-j+1)}{n-j}\right)^{\epsilon_{(j)}(1-\delta_{(j)})},$$

as required.

- If $\epsilon_{(i+1)} = 0$, then the algorithm will re-distribute the proportion $1-\beta(T_{(i+1)})$ of the probability mass at $T_{(i+1)}$ among the remaining $n-i-1$ observation points. So the probability mass at $T_{(s)}$ will equal

$$\begin{aligned}
&\left(1+\frac{1-\beta(T_{(i+1)})}{n-i-1}\right)\frac{1}{n}\prod_{j=1}^{i}\left(\frac{(n-j+1-\beta(T_{(j)}))}{n-j}\right)^{1-\epsilon_{(j)}}\left(\frac{(n-j+1)}{n-j}\right)^{\epsilon_{(j)}(1-\delta_{(j)})}\\
=\;&\left(\frac{n-i-\beta(T_{(i+1)})}{n-i-1}\right)\frac{1}{n}\prod_{j=1}^{i}\left(\frac{(n-j+1-\beta(T_{(j)}))}{n-j}\right)^{1-\epsilon_{(j)}}\left(\frac{(n-j+1)}{n-j}\right)^{\epsilon_{(j)}(1-\delta_{(j)})}.
\end{aligned}$$

From the assumption about $\epsilon_{(i+1)}$ and $\delta_{(i+1)}$, this is just equal to

$$\frac{1}{n}\prod_{j=1}^{i+1}\left(\frac{(n-j+1-\beta(T_{(j)}))}{n-j}\right)^{1-\epsilon_{(j)}}\left(\frac{(n-j+1)}{n-j}\right)^{\epsilon_{(j)}(1-\delta_{(j)})},$$

as required.

So by induction the theorem is correct in the case of no ties. In the event of ties the algorithm assumes that the tied observations are infinitesimally separated, using the ordering of uncensored observations, missing data observations and then censored observations to calculate the individual contributions to the probability mass at that point. The probability mass at such a point is then just the sum of the individual contributions from each observation point as stated in the theorem. □

The following theorem will establish the equivalence of the re-distribution to the right algorithm and the formally derived expression.

Theorem 8. *The re-distribution to the right algorithm is equivalent to the formally derived expression in Equation 7.3.3.*

Proof. It is clear that in both cases probability mass is only present at discrete points (a subset of the time points for the actual observations). It will be enough to consider the mass at each such point, verifying equivalence in an inductive fashion. First the case of no ties will be considered. The base case can be considered at the same time as the inductive step by defining $T_{(0)} = 0$. It can be seen that for the base case ($i = 1$) no probability mass is present at a time prior to the first observation $T_{(1)}$ and the formal expression defines the survival function to be 1 for the interval between 0 and the first observation.

The inductive step is as follows. For all i from 1 to n, assume the agreement of the two estimates up to the i^{th} time point and assume that $T_{(i)}$ is an uncensored observation. For all r such that $T_{(i-1)} \leq r < T_{(i)}$ and for all t such that $T_{(i)} \leq t < T_{(i+1)}$, this implies that

$$\begin{aligned}
\hat{S}(r) - \hat{S}(t) &= \hat{S}(r)\left[1 - \left(1 - \frac{1}{n_i}\right)\right] \\
&= \frac{1}{n_i}\hat{S}(r) \\
&= \frac{1}{n_i}\prod_{(j|T_{(j)}\leq t)}\left(1 - \frac{\beta(T_{(j)})}{n_j}\right)^{1-\epsilon_{(j)}}\left(1 - \frac{1}{n_j}\right)^{\epsilon_{(j)}\delta_{(j)}} \\
&= \frac{1}{n-i+1}\prod_{j=1}^{i-1}\left(\frac{n-j+1-\beta(T_{(j)})}{n-j+1}\right)^{1-\epsilon_{(j)}}\left(\frac{n-j}{n-j+1}\right)^{\epsilon_{(j)}\delta_{(j)}}.
\end{aligned}$$

The above expression is the probability mass present at $T_{(i)}$ according to the formally derived expression. This will now be shown to agree with the probability mass assigned to $T_{(i)}$ by the redistribution to the right algorithm. First the exponents of the various factors in the product are

manipulated as follows.

$$\begin{aligned}
\frac{\hat{S}(r)}{n_i} &= \frac{1}{n-i+1}\prod_{j=1}^{i-1}\left(\frac{n-j+1-\beta(T_{(j)})}{n-j+1}\right)^{1-\epsilon_{(j)}}\left(\frac{n-j}{n-j+1}\right)^{\epsilon_{(j)}\delta_{(j)}} \\
&= \frac{1}{n-i+1}\prod_{j=1}^{i-1}\left(n-j+1-\beta(T_{(j)})\right)^{1-\epsilon_{(j)}}(n-j+1)^{\epsilon_{(j)}-1}\left(\frac{n-j+1}{n-j}\right)^{-\epsilon_{(j)}\delta_{(j)}} \\
&= \frac{1}{n-i+1}\prod_{j=1}^{i-1}\left(n-j+1-\beta(T_{(j)})\right)^{1-\epsilon_{(j)}}\left(\frac{n-j+1}{n-j}\right)^{\epsilon_{(j)}[1-\delta_{(j)}]}\left(\frac{1}{n-j+1}\right)(n-j)^{\epsilon_{(j)}} \\
&= \frac{1}{n-i+1}\prod_{j=1}^{i-1}\left(\frac{n-j+1-\beta(T_{(j)})}{n-j}\right)^{1-\epsilon_{(j)}}\left(\frac{n-j+1}{n-j}\right)^{\epsilon_{(j)}[1-\delta_{(j)}]}\frac{n-j}{n-j+1}.
\end{aligned}$$

The last factor in the product can be seen to cancel with successive terms so that the above expression is equal to

$$\begin{aligned}
&\frac{1}{n-i+1}\frac{n-i+1}{n}\prod_{j=1}^{i-1}\left(\frac{n-j+1-\beta(T_{(j)})}{n-j}\right)^{1-\epsilon_{(j)}}\left(\frac{n-j+1}{n-j}\right)^{\epsilon_{(j)}[1-\delta_{(j)}]} \\
= \;& \frac{1}{n}\prod_{j=1}^{i-1}\left(\frac{n-j+1-\beta(T_{(j)})}{n-j}\right)^{1-\epsilon_{(j)}}\left(\frac{n-j+1}{n-j}\right)^{\epsilon_{(j)}[1-\delta_{(j)}]},
\end{aligned}$$

as required. Assuming instead that $T_{(i)}$ is a missing data observation implies that

$$\begin{aligned}
\hat{S}(r)-\hat{S}(t) &= \hat{S}(r)\left[1-\left(1-\frac{\beta(T_{(i)})}{n_i}\right)\right] \\
&= \frac{\beta(T_{(i)})}{n_i}\hat{S}(r) \\
&= \frac{\beta(T_{(i)})}{n_i}\prod_{(j|T_{(j)}\le t)}\left(1-\frac{\beta(T_{(j)})}{n_j}\right)^{1-\epsilon_{(j)}}\left(1-\frac{1}{n_j}\right)^{\epsilon_{(j)}\delta_{(j)}} \\
&= \frac{\beta(T_{(i)})}{n-i+1}\prod_{j=1}^{i-1}\left(\frac{n-j+1-\beta(T_{(j)})}{n-j+1}\right)^{1-\epsilon_{(j)}}\left(\frac{n-j}{n-j+1}\right)^{\epsilon_{(j)}\delta_{(j)}}.
\end{aligned}$$

The probability mass at a missing data observation differs from the probability mass at an uncensored observation by a multiplicative factor of $\beta(T_{(i)})$, so by employing the previous argument, the two estimates can be seem to agree for this case as well.

Clearly if $T_{(i)}$ is a censored observation then there is no probability mass distributed at $T_{(i)}$ by the re-distribution to the right algorithm. Equally there is no contribution made to the survival function in the formally derived expression. This establishes agreement between the two estimates

for all t such that $T_{(i)} \leq t < T_{(i+1)}$. By induction this leads to the agreement of the two estimates on the domain for which the formal expression of the survival function is defined. For the case of tied observations, it is enough to recognise that both the formal expression and re-distribute to the right algorithm are arrived at by considering the tied observations to be infinitesimally separated. These tied observations are also ordered in the same way in each estimate, so the same argument employed for the case of no ties can be applied to yield agreement of the two estimates on the domain of definition for the formal expression of the survival function. □

7.5 Some Illustrative Examples

The following two examples are given to illustrate the use of the estimator and to demonstrate the agreement between the redistribution to the right algorithm and the formal derivation. The first example is one proposed by F. Coolen [27] which is a situation involving only one missing data observation and assuming $\beta = 0.5$, a constant.

Example 1. *Consider the following dataset (where a $^+$ indicates a right censored observation and a $^-$ indicates a missing data observation)*

$$1, 2^+, 3^-, 4, 5^+, 6.$$

This can be rewritten using the notation established previously

$T_{(i)}$	1	2^+	3^-	4	5^+	6
$\epsilon_{(i)}$	1	1	0	1	1	1
$\delta_{(i)}$	1	0		1	0	1.

Applying Equation 7.3.1 from the formal derivation yields

$$\begin{aligned}
\hat{S}(0) &= 1 \\
\hat{S}(1) &= \prod_{(i|T_{(i)} \leq 1)} \left(1 - \frac{\beta(T_{(i)})}{n_{(i)}}\right)^{1-\epsilon_{(i)}} \left(1 - \frac{1}{n_{(i)}}\right)^{\delta_{(i)}\epsilon_{(i)}} = \left(1 - \frac{1}{6}\right) = \frac{5}{6} \\
\hat{S}(2) &= \prod_{(i|T_{(i)} \leq 2)} \left(1 - \frac{\beta(T_{(i)})}{n_{(i)}}\right)^{1-\epsilon_{(i)}} \left(1 - \frac{1}{n_{(i)}}\right)^{\delta_{(i)}\epsilon_{(i)}} = \left(1 - \frac{1}{6}\right) = \frac{5}{6}
\end{aligned}$$

$$\hat{S}(3) = \prod_{(i|T_{(i)}\leq 3)} \left(1 - \frac{\beta(T_{(i)})}{n_{(i)}}\right)^{1-\epsilon_{(i)}} \left(1 - \frac{1}{n_{(i)}}\right)^{\delta_{(i)}\epsilon_{(i)}} = \left(1 - \frac{1}{6}\right)\left(1 - \frac{1}{2\times 4}\right) = \frac{5}{6} \times \frac{7}{8} = \frac{35}{48}$$

$$\hat{S}(4) = \prod_{(i|T_{(i)}\leq 4)} \left(1 - \frac{\beta(T_{(i)})}{n_{(i)}}\right)^{1-\epsilon_{(i)}} \left(1 - \frac{1}{n_{(i)}}\right)^{\delta_{(i)}\epsilon_{(i)}} = \left(1 - \frac{1}{6}\right)\left(1 - \frac{1}{2\times 4}\right)\left(1 - \frac{1}{3}\right)$$

$$= \frac{5}{6} \times \frac{7}{8} \times \frac{2}{3} = \frac{70}{144}$$

$$\hat{S}(5) = \prod_{(i|T_{(i)}\leq 5)} \left(1 - \frac{\beta(T_{(i)})}{n_{(i)}}\right)^{1-\epsilon_{(i)}} \left(1 - \frac{1}{n_{(i)}}\right)^{\delta_{(i)}\epsilon_{(i)}} = \left(1 - \frac{1}{6}\right)\left(1 - \frac{1}{2\times 4}\right)\left(1 - \frac{1}{3}\right)$$

$$= \frac{5}{6} \times \frac{7}{8} \times \frac{2}{3} = \frac{70}{144}$$

$$\hat{S}(6) = \prod_{(i|T_{(i)}\leq 6)} \left(1 - \frac{\beta(T_{(i)})}{n_{(i)}}\right)^{1-\epsilon_{(i)}} \left(1 - \frac{1}{n_{(i)}}\right)^{\delta_{(i)}\epsilon_{(i)}}$$

$$= \left(1 - \frac{1}{6}\right)\left(1 - \frac{1}{2\times 4}\right)\left(1 - \frac{1}{3}\right)\left(1 - \frac{1}{1}\right) = 0.$$

This agrees with Coolen's calculations[27].

The next example tests the new estimator with numerous missing data observations and ties between the different types of observations. It will be assumed that $\beta = 0.75$, a constant. The formal equation will be shown to agree with the re-distribute to the right algorithm.

Example 2. *Consider the following dataset (where a $^+$ indicates a right censored observation and a $^-$ indicates a missing data observation)*

$$9, 13, 13^+, 18, 20^-, 23, 23, 23, 28^+, 45^+, 48, 53^+, 53^+, 74, 74, 74^-, 74^-, 74^-, 74^+, 90, 102^-, 161^+.$$

The ties can be removed and using the notation established previously, this dataset can be rewritten in the following way

$T'_{(i)}$	9	13	18	20	23	28	45	48	53	74	90	102	161
n_i	22	21	19	18	17	14	13	12	11	9	3	2	1
d_i	1	1	1	0	3	0	0	1	0	2	1	0	0
l_i	0	0	0	1	0	0	0	0	0	3	0	1	0.

Using the formula for the new estimator found in Equation 7.3.1, the survival function equals

$$
\begin{aligned}
\hat{S}(0) &= 1 \\
\hat{S}(9) &= \left(1-\frac{1}{22}\right) = \frac{21}{22} \\
\hat{S}(13) &= \frac{21}{22} \times \left(1-\frac{1}{21}\right) = \frac{21}{22} \times \frac{20}{21} = \frac{10}{11} \\
\hat{S}(18) &= \frac{10}{11} \times \left(1-\frac{1}{19}\right) = \frac{10}{11} \times \frac{18}{19} \\
\hat{S}(20) &= \frac{10}{11} \times \frac{18}{19} \times \left(1-\frac{0.75}{18-0+1-1}\right) = \frac{10}{11} \times \frac{18}{19} \times \frac{69}{72} = \frac{10}{11} \times \frac{69}{76} = \frac{5}{11} \times \frac{69}{38} \\
\hat{S}(23) &= \frac{10}{11} \times \frac{69}{76} \times \left(1-\frac{3}{17}\right) = \frac{10}{11} \times \frac{69}{76} \times \frac{14}{17} = \frac{5}{11} \times \frac{69}{19} \times \frac{7}{17} \\
\hat{S}(48) &= \frac{10}{11} \times \frac{69}{76} \times \frac{14}{17} \times \left(1-\frac{1}{12}\right) = \frac{10}{11} \times \frac{69}{76} \times \frac{14}{17} \times \frac{11}{12} = \frac{10}{12} \times \frac{69}{76} \times \frac{14}{17} = \frac{5}{4} \times \frac{23}{19} \times \frac{7}{17} \\
\hat{S}(74) &= \frac{10}{12} \times \frac{69}{76} \times \frac{14}{17} \times \left(1-\frac{2}{9}\right)\left(1-\frac{0.75}{(9-2+1-1)}\right)\left(1-\frac{0.75}{(9-2+1-2)}\right) \\
&\quad \times \left(1-\frac{0.75}{(9-2+1-3)}\right) \\
&= \frac{10}{12} \times \frac{69}{76} \times \frac{14}{17} \times \frac{7}{9} \times \frac{25}{28} \times \frac{21}{24} \times \frac{17}{20} = \frac{7}{12} \times \frac{69}{76} \times \frac{25}{36} \times \frac{21}{24} = \frac{7}{4} \times \frac{23}{76} \times \frac{25}{36} \times \frac{7}{8} \\
\hat{S}(90) &= \frac{7}{12} \times \frac{69}{76} \times \frac{25}{36} \times \frac{21}{24} \times \left(1-\frac{1}{3}\right) = \frac{7}{12} \times \frac{69}{76} \times \frac{25}{36} \times \frac{21}{24} \times \frac{2}{3} \\
&= \frac{7}{12} \times \frac{23}{38} \times \frac{25}{36} \times \frac{21}{24} = \frac{7}{12} \times \frac{23}{38} \times \frac{25}{36} \times \frac{7}{8} \\
\hat{S}(102) &= \frac{7}{12} \times \frac{23}{38} \times \frac{25}{36} \times \frac{21}{24}\left(1-\frac{0.75}{2}\right) = \frac{7}{12} \times \frac{23}{38} \times \frac{25}{36} \times \frac{21}{24} \times \frac{5}{8} = \frac{7}{24} \times \frac{23}{38} \times \frac{25}{36} \times \frac{35}{32}.
\end{aligned}
$$

Applying the formula from the re-distribute to the right algorithm, the masses of the probability density function equal

$$
\begin{aligned}
\hat{f}(9) &= \frac{1}{22} \\
\hat{f}(13) &= \frac{1}{22} \\
\hat{f}(18) &= \frac{1}{22} \times \left(\frac{22-3+1}{22-3}\right) = \frac{1}{22} \times \frac{20}{19} \\
\hat{f}(20) &= \frac{3}{4} \times \frac{1}{22} \times \frac{20}{19} = \frac{3}{88} \times \frac{20}{19} \\
\hat{f}(23) &= 3 \times \frac{1}{22} \times \frac{20}{19} \times \left(\frac{22-5+1-0.75}{22-5}\right) = \frac{3}{22} \times \frac{20}{19} \times \frac{69}{68} = \frac{3}{11} \times \frac{5}{19} \times \frac{69}{34} \\
\hat{f}(48) &= \frac{1}{11} \times \frac{5}{19} \times \frac{69}{34} \times \left(\frac{22-9+1}{22-9}\right)\left(\frac{22-10+1}{22-10}\right) = \frac{1}{11} \times \frac{5}{19} \times \frac{69}{34} \times \frac{14}{13} \times \frac{13}{12} \\
&= \frac{1}{11} \times \frac{5}{19} \times \frac{69}{34} \times \frac{14}{12}
\end{aligned}
$$

$$
\begin{aligned}
\hat{f}(74) &= 2 \times \frac{1}{22} \times \frac{20}{19} \times \frac{69}{68} \times \frac{14}{12} \times \left(\frac{22-12+1}{22-12}\right)\left(\frac{22-13+1}{22-13}\right) \\
&+ \frac{3}{4} \times \frac{1}{22} \times \frac{20}{19} \times \frac{69}{68} \times \frac{14}{12} \times \left(\frac{22-12+1}{22-12}\right)\left(\frac{22-13+1}{22-13}\right) \\
&+ \frac{3}{4} \times \frac{1}{22} \times \frac{20}{19} \times \frac{69}{68} \times \frac{14}{12} \times \left(\frac{22-12+1}{22-12}\right)\left(\frac{22-13+1}{22-13}\right)\left(\frac{22-16+1-0.75}{22-16}\right) \\
&+ \frac{3}{4} \times \frac{1}{22} \times \frac{20}{19} \times \frac{69}{68} \times \frac{14}{12} \times \left(\frac{22-12+1}{22-12}\right)\left(\frac{22-13+1}{22-13}\right)\left(\frac{22-16+1-0.75}{22-16}\right) \\
&\quad \times \left(\frac{22-17+1-0.75}{22-17}\right) \\
&= \frac{11}{4} \times \frac{1}{22} \times \frac{20}{19} \times \frac{69}{68} \times \frac{14}{12} \times \frac{11}{10} \times \frac{10}{9} + \frac{3}{4} \times \frac{1}{22} \times \frac{20}{19} \times \frac{69}{68} \times \frac{14}{12} \times \frac{11}{10} \times \frac{10}{9} \times \frac{6.25}{6} \\
&+ \frac{3}{4} \times \frac{1}{22} \times \frac{20}{19} \times \frac{69}{68} \times \frac{14}{12} \times \frac{11}{10} \times \frac{10}{9} \times \frac{7-0.75}{6} \times \frac{6-0.75}{5} \\
&= \frac{3}{4} \times \frac{1}{22} \times \frac{20}{19} \times \frac{69}{68} \times \frac{14}{12} \times \frac{11}{10} \times \frac{10}{9}\left(\frac{11}{3} + \frac{25}{24} + \frac{25}{24} \times \frac{21}{20}\right) \\
&= \frac{5}{19} \times \frac{69}{68} \times \frac{14}{4} \times \frac{1}{18}\left(\frac{11}{3}\frac{20}{20}\frac{8}{8} + \frac{25}{24}\frac{20}{20} + \frac{25}{24} \times \frac{21}{20}\right) \\
&= \frac{5}{19} \times \frac{69}{68} \times \frac{14}{4} \times \frac{1}{18}\left(\frac{2260}{24}\frac{1}{20} + \frac{525}{24} \times \frac{1}{20}\right) \\
&= \frac{5}{19} \times \frac{69}{68} \times \frac{7}{4} \times \frac{1}{9} \times \frac{1}{20} \times \frac{2785}{24} = \frac{5}{19} \times \frac{23}{68} \times \frac{7}{32} \times \frac{557}{36} \\
\hat{f}(90) &= \frac{1}{22} \times \frac{20}{19} \times \frac{69}{68} \times \frac{14}{13} \times \frac{13}{12} \times \frac{11}{10} \times \frac{10}{9} \times \frac{25}{24} \times \frac{21}{20} \times \left(\frac{22-18+1-0.75}{22-18}\right) \\
&\quad \times \left(\frac{22-19+1}{22-19}\right) \\
&= \frac{1}{22} \times \frac{20}{19} \times \frac{69}{68} \times \frac{14}{13} \times \frac{13}{12} \times \frac{11}{10} \times \frac{10}{9} \times \frac{25}{24} \times \frac{21}{20} \times \frac{17}{16} \times \frac{4}{3} \\
&= \frac{69}{68} \times \frac{14}{12} \times \frac{1}{18} \times \frac{25}{24} \times \frac{21}{19} \times \frac{17}{16} \times \frac{4}{3} \\
&= \frac{23}{68} \times \frac{7}{6} \times \frac{17}{18} \times \frac{25}{32} \times \frac{7}{19} \\
\hat{f}(102) &= \frac{3}{4} \times \frac{1}{22} \times \frac{20}{19} \times \frac{69}{68} \times \frac{14}{13} \times \frac{13}{12} \times \frac{11}{10} \times \frac{10}{9} \times \frac{25}{24} \times \frac{21}{20} \times \frac{17}{16} \times \frac{4}{3} \\
&= \frac{3}{4} \times \frac{23}{68} \times \frac{7}{6} \times \frac{17}{18} \times \frac{25}{32} \times \frac{7}{19} \\
&= \frac{23}{68} \times \frac{7}{8} \times \frac{17}{18} \times \frac{25}{32} \times \frac{7}{19}.
\end{aligned}
$$

The resulting survival function equals

$$
\begin{aligned}
\hat{S}(9) &= 1-\frac{1}{22}=\frac{21}{22}\\
\hat{S}(13) &= \frac{21}{22}-\frac{1}{22}=\frac{20}{22}=\frac{10}{11}\\
\hat{S}(18) &= \frac{10}{11}-\frac{1}{22}\times\frac{20}{19}=\frac{10}{11}\left(1-\frac{1}{19}\right)=\frac{10}{11}\times\frac{18}{19}\\
\hat{S}(20) &= \frac{10}{11}\times\frac{18}{19}-\frac{3}{88}\times\frac{20}{19}=\frac{10}{11}\times\frac{36}{38}-\frac{3}{11}\times\frac{5}{38}=\frac{5}{11}\times\frac{69}{38}\\
\hat{S}(23) &= \frac{5}{11}\times\frac{69}{38}-\frac{3}{11}\times\frac{5}{19}\times\frac{69}{34}=\frac{5}{11}\times\frac{69}{38}-\frac{5}{11}\times\frac{69}{38}\times\frac{3}{17}=\frac{5}{11}\times\frac{69}{38}\frac{14}{17}\\
\hat{S}(48) &= \frac{5}{11}\times\frac{69}{19}\times\frac{7}{17}-\frac{1}{11}\times\frac{5}{19}\times\frac{69}{34}\times\frac{14}{12}=\frac{15}{11}\times\frac{23}{19}\times\frac{7}{17}-\frac{15}{11}\times\frac{23}{19}\times\frac{7}{17}\times\frac{1}{12}\\
&= \frac{15}{11}\times\frac{23}{19}\times\frac{7}{17}\times\frac{11}{12}=\frac{15}{12}\times\frac{23}{19}\times\frac{7}{17}\\
\hat{S}(74) &= \frac{5}{4}\times\frac{23}{19}\times\frac{7}{17}-\frac{5}{19}\times\frac{23}{68}\times\frac{7}{32}\times\frac{557}{36}=\frac{5}{19}\times\frac{23}{68}\times\frac{7}{1}-\frac{5}{19}\times\frac{23}{68}\times\frac{7}{32}\times\frac{557}{36}\\
&= \frac{5}{19}\times\frac{23}{68}\times 7\left(\frac{1}{32}\times\frac{32\times 36}{36}-\frac{1}{32}\times\frac{557}{36}\right)=\frac{5}{19}\times\frac{23}{68}\times\frac{7}{32}\times\frac{1152-557}{36}\\
&= \frac{23}{76}\times\frac{5}{68}\times\frac{595}{36}\times\frac{7}{8}=\frac{119}{68}\times\frac{23}{76}\times\frac{25}{36}\times\frac{7}{8}=\frac{7}{4}\times\frac{23}{76}\times\frac{25}{36}\times\frac{7}{8}\\
\hat{S}(90) &= \frac{7}{4}\times\frac{23}{76}\times\frac{25}{36}\times\frac{7}{8}-\frac{23}{68}\times\frac{7}{6}\times\frac{17}{18}\times\frac{25}{32}\times\frac{7}{19}\\
&= \frac{7}{8}\times\frac{23}{38}\times\frac{25}{36}\times\frac{7}{8}-\frac{23}{68}\times\frac{7}{6}\times\frac{68}{36}\times\frac{25}{32}\times\frac{7}{38}\\
&= \frac{7}{8}\times\frac{23}{38}\times\frac{25}{36}\times\frac{7}{8}-\frac{7}{8}\times\frac{23}{38}\times\frac{25}{36}\times\frac{7}{8}\times\frac{1}{3}\\
&= \frac{7}{8}\times\frac{23}{38}\times\frac{25}{36}\times\frac{7}{8}\times\frac{2}{3}=\frac{7}{12}\times\frac{23}{38}\times\frac{25}{36}\times\frac{7}{8}\\
\hat{S}(102) &= \frac{7}{12}\times\frac{23}{38}\times\frac{25}{36}\times\frac{7}{8}-\frac{23}{68}\times\frac{7}{8}\times\frac{17}{18}\times\frac{25}{32}\times\frac{7}{19}\\
&= \frac{7}{24}\times\frac{23}{38}\times\frac{25}{36}\times\frac{7}{4}-\frac{25}{32}\times\frac{23}{68}\times\frac{7}{38}\times\frac{68}{36}\times\frac{7}{8}\\
&= \frac{7}{24}\times\frac{23}{38}\times\frac{25}{36}\times\frac{7}{4}-\frac{7}{32}\times\frac{23}{38}\times\frac{25}{36}\times\frac{7}{8}\\
&= \frac{7}{24}\times\frac{23}{38}\times\frac{25}{36}\times\frac{7}{4}-\frac{7}{24}\times\frac{23}{38}\times\frac{25}{36}\times\frac{7}{4}\times\frac{3}{8}\\
&= \frac{7}{24}\times\frac{23}{38}\times\frac{25}{36}\times\frac{7}{4}\times\frac{5}{8}=\frac{7}{24}\times\frac{23}{38}\times\frac{25}{36}\times\frac{35}{32}.
\end{aligned}
$$

This agrees with the values found previously (using Equation 7.3.1).

7.6 Some Preliminary Theory on Counting Processes

This section will present some background material for the proof found in the next section. The material in this section is heavily reliant on that by Andersen et. al.[3] on counting processes and

the Nelson-Aalen estimator. In the rest of this section, whenever page numbers are given these will refer to Andersen et. al.[3]. The following definitions will prove useful.

A *probability space* $\mathscr{P} = (\Omega, \mathscr{F}, P)$ is a set Ω with a system of subsets $\mathscr{F}$ of Ω which form a σ-algebra. This space is also equipped with a measure

$$P : \mathscr{F} \longmapsto [0, 1]$$

with $P(\Omega) = 1$. The space $(\Omega, \mathscr{F})$ without a probability measure is called a *measurable space*.

Given two measurable spaces $(U, \mathscr{U})$ and $(V, \mathscr{V})$, a function from U to V is called a *measurable function* when the pre-image of that function maps any element in $\mathscr{V}$ into an element in $\mathscr{U}$.

Let $\mathscr{T}$ be a fixed continuous time interval of the form $[0, \tau)$ or $[0, \tau]$ for a given end time τ, $0 < \tau \leq \infty$.

Given a probability space $\mathscr{P} = (\Omega, \mathscr{F}, P)$ a *filtration*

$$\{\mathscr{F}_t : t \in \mathscr{T}\}$$

is a collection of sub-σ-algebras with the following properties:

$$\mathscr{F}_s \subseteq \mathscr{F}_t \subseteq \mathscr{F} \quad \text{for all } s < t \qquad \text{(increasing)},$$

$$\mathscr{F}_s = \bigcap_{t>s} \mathscr{F}_t \quad \text{for all } s \qquad \text{(right continuous)}.$$

A filtration can be thought of as a history of discrete events over continuous time. At any given point in time some outcomes are known and some are not. As time passes, knowledge about what has happened grows, equally the history about these discrete events increases. A filtration is a collection of σ-algebras indexed by time. Each sub-σ-algebra $\mathscr{F}_t$ represents the growing knowledge of events, up to and including time $t \in \mathscr{T}$, and is a subset of $\mathscr{F}$.

7.6.1 Stochastic Processes

Given a probability space $\mathscr{P} = (\Omega, \mathscr{F}, P)$ a *stochastic process* X is a collection of random variables indexed by a set $\mathscr{T}$, $\{X(t) : t \in \mathscr{T}\}$. The rest of this subsection will introduce several different kinds of stochastic process.

A stochastic process X is said to be *adapted* to a filtration $\mathscr{F}$ if $X(t)$ is $\mathscr{F}_t$-measurable for all $t \in \mathscr{T}$.

A stochastic process X is said to be *predictable* if $X(t)$ is $\mathscr{F}_{t-}$-measurable for all $t \in \mathscr{T}$.
A stochastic process X is said to be *cadlag* if it's sample paths $(X(t,\omega) : t \in \mathscr{T})$ are right-continuous with left-hand limits for almost all ω. Expressing this formally

$$X(t+,\omega) = \lim_{s \downarrow t} X(s,\omega) \text{ exists and equals } X(t,\omega),$$
$$X(t-,\omega) = \lim_{s \uparrow t} X(s,\omega) \text{ exists.}$$

A stochastic process X is said to be a *finite variation process* if X is cadlag with sample paths $(X(t,\omega) : t \in \mathscr{T})$ of locally bounded variation for almost all ω. Expressing this formally

$$\int_{[0,t]} |dX(s)| < \infty \quad \text{for all } t \in \mathscr{T} \text{ and for almost all } \omega \in \Omega,$$

where such an integral is always taken to be a path-wise Lebesque-Stieltjes integral for a given $\omega \in \Omega$. This property of stochastic processes ensures that stochastic integrals are well defined.
The definitions given above for adapted and cadlag stochastic processes facilitate the following two definitions of counting processes. Given a probability space $\mathscr{P}$ and a suitable filtration $\{\mathscr{F}_t : t \in \mathscr{T}\}$, a *counting process* N is an adapted, cadlag stochastic process with the following properties:

- $N(0) = 0$;
- sample paths are piecewise constant and non-decreasing, with jumps of size $+1$ only;
- $N(t) < \infty$ almost surely for all $t \in \mathscr{T}$.

Given a probability space $\mathscr{P}$ and a suitable filtration $\{\mathscr{F}_t : t \in \mathscr{T}\}$, a *multivariate counting process*

$$\mathbf{N} = (N_1, \cdots, N_k)$$

is a vector of counting processes, such that no two components of the vector jump simultaneously.

7.6.2 Multiplicative Intensity Model

Given an intensity process $\lambda(t)$ (introduced in Chapter 3.1, the *cumulative intensity process* is defined as

$$\Lambda(t) = \int_0^t \lambda(s) ds.$$

The definitions made so far allow for the introduction of the multiplicative intensity model.

Definition 2 (Aalen's Multiplicative Intensity Model (p.176-179))**.** *Let* $(\Omega, \mathscr{F})$ *be a measurable space equipped with a probability measure* P *and with a filtration* $(\mathscr{F}_t, t \in \mathscr{T})$
Let $\boldsymbol{N} = ((N_1(t), \cdots, N_k(t)); t \in \mathscr{T})$ *be a multivariate counting process defined on* $(\Omega, \mathscr{F})$ *and adapted to the filtration* $(\mathscr{F}_t, t \in \mathscr{T})$*, such that* $\boldsymbol{N}$ *satisfies the following property: For all* h *from* 1 *to* k*, the intensity process* $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_k)$ *is given by*

$$\lambda_h(t) = \alpha_h(t) Y_h(t).$$

Here α_h *is a non-negative deterministic function depending on* P*, and* Y_h *is a predictable process which is observable (it does not depend on* P*).*

In the context of this chapter α_h will refer to the individual hazard rate of an item due to risk h and $Y_h(t)$ will refer to the number of items at risk at time t. This will be discussed more fully in the next section.
The following definition will introduce the Nelson-Aalen estimator, which will be used to define the Kaplan-Meier estimator.

Definition 3 (Nelson-Aalen Estimator (p.177-179))**.** *Given a multivariate counting process* $\boldsymbol{N} = ((N_1(t), \cdots, N_k(t)); t \in \mathscr{T})$ *with intensity process* $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_k)$ *satisfying the multiplicative intensity model* $\lambda_h(t) = \alpha_h(t) Y_h(t)$*, the Nelson-Aalen estimator is an estimator of*

$$A_h(t) = \int_0^t \alpha_h(s) ds.$$

The Nelson-Aalen estimator is equal to

$$\hat{A}_h(t) = \int_0^t \frac{1}{Y_h(s)} dN_h(s),$$

where such an integral is just the ordinary path-wise Lebesgue-Stieltjes integral over $[0, t]$ *for all* $t \in \mathscr{T}$*.*

Having introduced the Nelson-Aalen estimator, the following theorems will show that it is consistent and that it can be used to derive the Kaplan-Meier estimator, which is then also shown to be consistent. In order to present these theorems the following defintion will be required.

Definition 4 (Convergence in Probability)**.** *A sequence of random variables* $\{X_n\}$ *converges in probability to* X *if for all* $\epsilon > 0$

$$\lim_{n \to +\infty} P\left\{|X_n - X| \geq \epsilon\right\} = 0.$$

This will be denoted as $X_n \xrightarrow{P} X$.

Theorem 9 (Consistency of the Nelson-Aalen Estimator (p.190, Theorem IV.1.1.))**.** *Consider a sequence of counting processes* $\boldsymbol{N}^{(n)} = \left(N_1^{(n)}(t), \cdots, N_k^{(n)}(t)\right)$, $n = 1, 2, \cdots$, *each satisfying the multiplicative intensity model* $\lambda_h^{(n)}(t) = \alpha_h(t) Y_h^{(n)}(t)$, *with the* α_h *being the same for all* n. *Define* $J_h^{(n)}(t) = I\left\{Y_h^{(n)}(t) > 0\right\}$ *and assume that* $J_h^{(n)}/Y_h^{(n)}$ *is locally bounded for each* h *and* n *(where* $J_h^{(n)}/Y_h^{(n)}$ *is interpreted as zero whenever* $Y_h^{(n)} = 0$*). Let* $t \in \mathscr{T}$ *and assume that, as* $n \to +\infty$,

$$\int_0^t \frac{J_h^{(n)}(s)}{Y_h^{(n)}(s)} \alpha_h(s) ds \xrightarrow{P} 0$$

and

$$\int_0^t (1 - J_h^{(n)}(s)) \alpha_h(s) ds \xrightarrow{P} 0.$$

Then as $n \to +\infty$,

$$\sup_{s \in [0,t]} \left| \hat{A}_h^{(n)}(s) - A_h^{(n)}(s) \right| \xrightarrow{P} 0.$$

The proof of this theorem is omitted, but can be found on page 190 of Andersen et. al.[3]. On page 256 of Andersen et. al.[3], the Kaplan-Meier estimator is derived directly from the Nelson-Aalen estimator using results on product-integrals. Because the Nelson-Aalen estimator is a step-function, the Kaplan-Meier estimator can be written as

$$\hat{S}_h(t) = \prod_{s \leq t} \left(1 - \Delta \hat{A}_h(s)\right).$$

The next theorem establishes the consistency of the Kaplan-Meier estimator.

Theorem 10 (Consistency of the Kaplan-Meier Estimator (p.261, Theorem IV.3.1.))**.** *Consider a sequence of counting processes* $\boldsymbol{N}^{(n)} = \left(N_1^{(n)}(t), \cdots, N_k^{(n)}(t)\right)$, $n = 1, 2, \cdots$, *each satisfying the multiplicative intensity model* $\lambda_h^{(n)}(t) = \alpha_h(t) Y_h^{(n)}(t)$, *with the* α_h *being the same for all* n. *Define* $J_h^{(n)}(t) = I\left\{Y_h^{(n)}(t) > 0\right\}$ *and assume that* $J_h^{(n)}/Y_h^{(n)}$ *is locally bounded for each* h *and* n *(where* $J_h^{(n)}/Y_h^{(n)}$ *is interpreted as zero whenever* $Y_h^{(n)} = 0$*). Let* $t \in \mathscr{T}$ *be such that* $S(t) > 0$ *and assume*

that, as $n \to +\infty$,

$$\int_0^t \frac{J_h^{(n)}(s)}{Y_h^{(n)}(s)} \alpha_h(s) ds \xrightarrow{P} 0$$

and

$$\int_0^t (1 - J_h^{(n)}(s)) \alpha_h(s) ds \xrightarrow{P} 0.$$

Then as $n \to +\infty$,

$$\sup_{s \in [0,t]} \left| \hat{S}_h^{(n)}(s) - S_h^{(n)}(s) \right| \xrightarrow{P} 0.$$

Proof. This result is an immediate consequence of Theorem 9 and the continuity of product-integrals, as remarked by Andersen et. al.[3]. □

The next theorem, stated without proof, is a common extension of Slutsky's theorem which can be found as Proposition 6.1, part a, on page 104 of Basu[8]. This theorem implies that the sum of two consistent estimators is itself consistent.

Theorem 11 (Extension of Slutsky's Theorem)**.** *Consider a sequence of random variables* $\{X_n\}$ *and a sequence of random variables* $\{Y_n\}$*, such that* $X_n \xrightarrow{P} X$ *and* $Y_n \xrightarrow{P} Y$*. Then*

$$X_n + Y_n \xrightarrow{P} X + Y.$$

This sequence of results means that consistency for the proposed estimator for missing data can be established if that estimator can be shown to satisfy the multiplicative intensity model presented here for the Nelson-Aalen estimator. This will be the done in the next section.

7.6.3 Martingales and Localisation

A cadlag adapted stochastic process M is called a *martingale* if for all $t \in \mathscr{T}$,

$$E(|M(t)|) < \infty$$

and M satisfies the martingale property,

$$E(M(t)|\mathscr{F}_s) = M(s)$$

for all $s \leq t$. Martingales can be thought of informally as random "white noise" processes. If

instead of satisfying the martingale property, such a process M satisfies

$$E(M(t)|\mathscr{F}_s) \geq M(s)$$

for all $s \leq t$, then it is called a *sub-martingale*. A martingale M is called *square integrable* if

$$sup_{\tau \in \mathscr{T}} E(M(t)^2) < \infty.$$

Attention is now turned to the idea of localisation, which is based on stopping times. A stopping time is a random variable T, which takes values in $\mathscr{T}$, such that

$$\{T \leq t\} \in \mathscr{F}_t \quad \text{for all } t \in \mathscr{T}.$$

A *localizing sequence* of stopping times is a non-decreasing sequence of stopping times $\{T_n\}_{n\geq 0}$ such that

$$P\{T_n \geq t\} \to 1 \quad \text{as } n \to \infty \quad \text{for all } t \in \mathscr{T}.$$

A stochastic process X is then said to have a property *locally* if a localizing sequence exists such that, for each n, the process $I\{T_n > 0\} M^{T_n}$ is said to have that property (where $M^{T_n}(t) = M(\min(t, T_n))$ and I is the usual indicator function).
A stochastic process X is *locally bounded* if there exists a localizing sequence $\{T_n\}_{n\geq 0}$ and a sequence of constants $\{c_n\}_{n\geq 0}$ such that, for each n,

$$\sup_{t \leq T_n} |X(t)| \leq c_n \quad \text{almost surely on the event } \{T_n > 0\}.$$

A martingale M is called a *local square integrable martingale* if there exists an increasing sequence of stopping times $\{T_n\}_{n\geq 0}$ such that

$$P\{T_n \geq t\} \to 1 \quad \text{as } n \to \infty \quad \text{for all } t \in \mathscr{T}$$

and the stopped processes

$$I\{T_n > 0\} M^{T_n}$$

are square integrable martingales for each n.

7.6.4 Decomposition and Stochastic Integration Theorems

The following theorem is a version of the famed Doob-Meyer decomposition theorem that says that a local sub-martingale can be uniquely decomposed into two component parts, one a predictable process (called a *compensator*), the other a random "white noise" process (a martingale).

Theorem 12 (p.66-67, Doob-Meyer Decomposition)**.** *Suppose X is a local sub-martingale. The process $\tilde{X}$ is called a compensator of X and is a predictable, cadlag finite variation process such that $M = X - \tilde{X}$ is a local martingale, zero at time zero. The compensator $\tilde{X}$ is unique for X.*

Because counting processes are local sub-martingales, the Doob-Meyer decomposition ensures the existence of a unique compensator. It turns out that this compensator is actually the cumulative intensity function for that counting process.

Theorem 13 (p.73-74, Doob-Meyer Decomposition for Counting Processes)**.** *Suppose N is a counting process with cumulative intensity function Λ. Then $M = N - \Lambda$ is a local square integrable martingale, zero at time zero.*

The following theorem is a useful remark about the properties of the integral of a predictable process with respect to a martingale.

Theorem 14 (p.71, Theorem II.3.1. - special case)**.** *Suppose M is a finite variation local square integrable martingale and H is a locally bounded predictable process. Then $\int H dM$ is a local square integrable martingale.*

7.6.5 Some Useful Results

This sub-section contains a pair of results that will prove useful in the next section. The first result is a theorem about convergence in probability of a sequence of stochastic processes to a deterministic function. The theorem says that if the stochastic process X converges in probability to deterministic reasonably well-behaved function f, then if X can be bounded by a reasonably well-behaved process k_δ, the stochastic integral of X will converge in probability to the integral of f.

Theorem 15 (Gill's Theorem (p.85, Theorem II.5.3.))**.** *Suppose that for some sequence $X^{(n)}$ of processes*

$$X^{(n)}(s) \xrightarrow{P} f(s) \qquad as\ n \to \infty$$

for almost all $s \in \mathscr{T}$, where the deterministic function f satisfies

$$\int_0^\tau |f(s)|ds < \infty.$$

Suppose also that for all $\delta > 0$, there exists a process k_δ with

$$\int_0^\tau k_\delta(s)ds < \infty,$$

such that

$$\liminf_{n\to\infty} P\left\{|X^{(n)}(s)| \leq k_\delta(s) \text{ for all } s\right\} \geq 1 - \delta.$$

Then

$$\sup_\tau \left| \int_0^\tau X^{(n)}(s)ds - \int_0^\tau f(s)ds \right| \xrightarrow{P} 0.$$

Theorem 16 (Lenglart's Inequality - continuous version (p.86, Equation 2.5.18))**.** *Suppose N is a counting process with continuous compensator Λ and local square integrable martingale M such that $M = N - \Lambda$. For any $\eta > 0$ and for any $\delta > 0$*

$$P\left\{\sup_{\tau\in\mathscr{T}} |M| > \eta\right\} \leq \frac{\delta}{\eta^2} + P\left\{\Lambda(\tau) > \delta\right\}.$$

This last result is just the application of Equation 2.4.3 on page 74 of Andersen et. al.[3] to Lenglart's Inequality stated in Equation 2.5.18 on Page 86 of Andersen et. al.[3].

7.7 Consistency of the Proposed Estimator

This section will present a proof of consistency for the proposed estimator. Before doing so, a discussion will be given regarding the impact on $\beta(t_i)$ of the different assumptions about the mechanisms that cause data to be missing. The consistency proof will rely heavily on the theory presented in Section 7.6 on counting processes and the Nelson-Aalen estimator.

7.7.1 Different Missing Data Mechanisms

When applying this estimator it will be necessary to specify the form of

$$\beta(t_i) = P\{\delta_i = 1|\epsilon_i = 0, T_i = t_i\}.$$

Applying Little and Rubin's[64] classification of "missingness" mechanisms introduced in Section 7.1, there are four sets of assumptions to consider:

I. ϵ_i is independent of both T_i and δ_i for each i.

II. ϵ_i is dependent only on T_i for each i.

III. ϵ_i is dependent only on δ_i for each i.

IV. ϵ_i is dependent on both T_i and δ_i for each i.

In the first case the "missingness" mechanism is not dependent on observed or unobserved data and is therefore missing completely at random (MCAR). In the second case the "missingness" mechanism is missing at random (MAR) because ϵ depends on the observable data T_i. In both the last two cases the "missingness" mechanism depends on data that is unobserved and therefore data is not missing at random (NMAR). In these last two cases it will be necessary to assume a parametric form for $P\{\delta_i = 1|\epsilon_i = 0, T_i = t_i\}$ in order to proceed.

In both the first two cases ϵ_i and δ_i are assumed to be independent, which implies that $P\{\delta_i = 1|\epsilon_i = 0, T_i = t_i\} = P\{\delta_i = 1|T_i = t_i\}$ so the distribution of the missing δ_i is the same as that of the observed δ_i, but the latter probability distribution is still rather difficult to estimate. Assuming further that δ_i and T_i are independent implies that $P\{\delta_i = 1|\epsilon_i = 0, T_i = t_i\} = P\{\delta_i = 1\}$ which is a constant and can be estimated by

$$\hat{P}\{\delta_i = 1\} = \frac{\sum_{i=1}^{n} I\{\delta_i = 1, \epsilon_i = 1\}}{\sum_{i=1}^{n} I\{\epsilon_i = 1\}}.$$

7.7.2 Proof of Consistency

This subsection will present a proof for the consistency of the proposed estimator with no ties (derived in Section 7.3) under the assumption that $\hat{\beta}$ is a predictable, bounded, consistent estimator of the probability β and that for all $t \in \mathscr{T}$, $\beta(t) > 0$. As remarked in Subsection 7.7.1, if the "missingness" mechanism is presumed to be MAR or MCAR then this probability can be estimated. This proof will rely heavily on work by Andersen et. al.[3] on counting processes and the Nelson-Aalen estimator discussed in Section 7.6. In the rest of this section, whenever page numbers are given these will refer to Andersen et. al.[3].

First it will be necessary to describe the proposed estimator using the counting process language from the previous section. The competing risk set-up discussed in Section 7.3 involved two risks,

with the possibility that the mark describing which risk was responsible for failure could be missing. This can be recast into a competing risks model with four risks

I. Failure observations,

II. missing data observations which were actually failures,

III. Censored observations,

IV. missing data observations which were actually censors.

Given an increasing sequence indexed by n, of sets of observation times $T_1, \cdots, T_n$ with corresponding indicators ϵ_i and δ_i (those δ_i being unobserved for which $\epsilon_i = 0$), this leads to the following sequence of four-component multivariate counting processes

$$\mathbf{N}^{(n)} = \left(\left(N_1^{(n)}(t), N_2^{(n)}(t), N_3^{(n)}(t), N_4^{(n)}(t) \right); t \in \mathscr{T} \right),$$

such that

$$\begin{aligned}
N_1^{(n)}(t) &= \sum_{i=0}^{n} I\left\{T_i \leq t, \delta_i = 1, \epsilon_i = 1\right\}, \\
N_2^{(n)}(t) &= \sum_{i=0}^{n} I\left\{T_i \leq t, \delta_i = 1, \epsilon_i = 0\right\}, \\
N_3^{(n)}(t) &= \sum_{i=0}^{n} I\left\{T_i \leq t, \delta_i = 0, \epsilon_i = 1\right\}, \\
N_4^{(n)}(t) &= \sum_{i=0}^{n} I\left\{T_i \leq t, \delta_i = 0, \epsilon_i = 0\right\}.
\end{aligned}$$

It should be noted that only the counting processes $N_1^{(n)}(t)$,$N_3^{(n)}(t)$ and $N_{2+4}^{(n)}(t) = N_2^{(n)}(t)+N_4^{(n)}(t)$ are actually observable. Considering the cause-specific hazard functions $\alpha_1(t)$,$\alpha_2(t)$,$\alpha_3(t)$ and $\alpha_4(t)$, the sequence of intensity functions can be described simply as

$$\boldsymbol{\lambda}^{(n)}(t) = \left(\alpha_1(t) Y^{(n)}(t), \alpha_2(t) Y^{(n)}(t), \alpha_3(t) Y^{(n)}(t), \alpha_4(t) Y^{(n)}(t) \right),$$

where $Y^{(n)}(t) = \sum_{i=1}^{n} I\left\{T_i > t,\right)$ are the remaining items still at risk. Note that the intensity process of the counting process $N_{2+4}^{(n)}(t)$, generated by superimposing $N_2^{(n)}(t)$ and $N_4^{(n)}(t)$, is just the sum of $\lambda_2^{(n)}(t)$ and$\lambda_4^{(n)}(t)$.

Assuming for the moment that all the processes are observable, then the sequence of counting processes of interest are $N_{1+2}^{(n)}(t)$. By Definition 3 in the previous section, the sequence of Nelson-Aalen estimates would estimate the following quantity

$$A_{1+2}(t) = \int_0^t \alpha_{1+2}(s)ds = A_1(t) + A_2(t).$$

Clearly $A_1(t)$ can be estimated by the standard Nelson-Aalen estimator and from Theorem 9 this estimate is consistent. What remains to be shown is how to estimate $A_2(t)$ and that such an estimate is itself consistent. Once this is established, the sum of two consistent estimators is itself consistent by Theorem 11, and the estimate of $A_{1+2}(t)$ can then be used to derive the Kaplan-Meier estimator. From the same argument of Theorem 10 it would follow immediately that this estimator is consistent. All that will remain is to note that the resulting Kaplan-Meier estimator is the same as that derived in Section 7.3.

Given that $\beta(t) = P\{\delta = 1|\epsilon = 0, T = t\}$, this necessarily implies

$$\lambda_2^{(n)}(t) = \beta(t)\lambda_{2+4}^{(n)}(t).$$

Theorem 13 necessarily implies that

$$M_{2+4}^{(n)}(t) = N_{2+4}^{(n)}(t) - \int_0^t \alpha_{2+4}(s)Y^{(n)}(s)ds = N_{2+4}^{(n)}(t) - \int_0^t \frac{\alpha_2(s)}{\beta(s)}Y^{(n)}(s)ds$$

are local square integrable martingales. Following the same argument as Andersen et. al., this implies that the following can be written symbolically

$$dN_{2+4}^{(n)}(t) = \frac{\alpha_2(t)}{\beta(t)}Y^{(n)}(t)dt + dM_{2+4}^{(n)}(t).$$

This leads naturally to the following sequence of estimates of $A_2(t)$

$$\hat{A}_2^{(n)}(t) = \int_0^t \frac{\hat{\beta}^{(n)}(s)}{Y^{(n)}(s)}dN_{2+4}^{(n)}(s), \tag{7.7.1}$$

where $\hat{\beta}^{(n)}$ are a sequence of estimators for β based on the first n elements of observable data. The following sequence of indicator processes is introduced $J^{(n)}(t) = I\left\{Y^{(n)}(t) > 0\right\}$, facilitating the

definition below

$$A_2^{*(n)}(t) = \int_0^t \alpha_{2+4}(s)\hat{\beta}^{(n)}(s)J^{(n)}(s)ds.$$

Because $N_{2+4}^{(n)}$ may only jump when $Y^{(n)}$ is positive, Equation 7.7.1 can be rewritten as

$$\hat{A}_2^{(n)}(t) = \int_0^t \frac{\hat{\beta}^{(n)}(s)J^{(n)}(s)}{Y^{(n)}(s)} dN_{2+4}^{(n)}(s),$$

where $J^{(n)}(s)/Y^{(n)}(s)$ is interpreted as zero whenever $Y^{(n)}(s) = 0$. This implies that

$$\begin{aligned}\hat{A}_2^{(n)}(t) - A_2^{*(n)}(t) &= \int_0^t \frac{\hat{\beta}^{(n)}(s)J^{(n)}(s)}{Y(^{(n)}s)} dN_{2+4}^{(n)}(s) - \int_0^t \frac{\hat{\beta}^{(n)}(s)J^{(n)}(s)}{Y^{(n)}(s)}\alpha_{2+4}(s)Y^{(n)}(s)ds \\ &= \int_0^t \frac{\hat{\beta}^{(n)}(s)J^{(n)}(s)}{Y^{(n)}(s)} dM_{2+4}^{(n)}(s). \end{aligned} \tag{7.7.2}$$

On page 178 of Andersen et. al.[3], the processes $J^{(n)}/Y^{(n)}$ are shown to be predictable. Because $\hat{\beta}^{(n)}$ are assumed to be predictable, this necessarily implies that $\hat{\beta}^{(n)}J^{(n)}/Y^{(n)}$ are predictable processes.

On page 179 of Andersen et. al.[3], it is stated that if the stochastic processes $Y^{(n)}$ take on integer values (which in this case they clearly do) then $J^{(n)}/Y^{(n)}$ are locally bounded. By assumption $\hat{\beta}^{(n)}$ are locally bounded, so $\hat{\beta}^{(n)}J^{(n)}/Y^{(n)}$ are also locally bounded.

Having established the necessary conditions in the last two paragraphs, Theorem 14 can now be applied to the result in Equation 7.7.2 (assuming that the stochastic processes $M_{2+4}^{(n)}$ are of finite variation), which necessarily implies that

$$\hat{A}_2^{(n)}(t) - A_2^{*(n)}(t), \tag{7.7.3}$$

are local square integrable martingales. The compensator of $\hat{A}_2^{(n)}(t) - A_2^{*(n)}(t)$ is equal to

$$\int_0^t \frac{\hat{\beta}^{(n)}(s)J^{(n)}(s)}{Y^{(n)}(s)}\alpha_{2+4}(s)ds.$$

The following three expressions will now be established:

$$\int_0^t \frac{\hat{\beta}^{(n)}(s)J^{(n)}(s)}{Y^{(n)}(s)}\alpha_{2+4}(s)ds \xrightarrow{P} 0 \tag{7.7.4}$$

$$\int_0^t \left(1 - J^{(n)}(s)\right)\hat{\beta}^{(n)}(s)\alpha_{2+4}(s)ds \xrightarrow{P} 0 \tag{7.7.5}$$

$$\int_0^t \left(\beta(s) - \hat{\beta}^{(n)}(s)\right)\alpha_{2+4}(s)ds \xrightarrow{P} 0. \tag{7.7.6}$$

Having assumed that $A_2(t) < \infty$ for all $t \in \mathscr{T}$ and because $Y^{(n)}$ is integer valued, there exists a constant c not depending on n such that $Y^{(n)}(t) < c$ implies $Y^{(n)}(t) = 0$. This necessarily implies that

$$\frac{J^{(n)}(t)}{Y^{(n)}(t)}\alpha_{2+4}(t) \quad \text{is bounded by } \frac{1}{c}\alpha_{2+4}(t), \tag{7.7.7}$$

$$\left(1 - J^{(n)}(t)\right)\alpha_{2+4}(t) \quad \text{is bounded by } \alpha_{2+4}(t). \tag{7.7.8}$$

Assuming that $Y^{(n)}(t) \xrightarrow{P} \infty$ and given that $\hat{\beta}^{(n)}$ are locally bounded, Theorem 15 in the previous section can be applied (in both cases with $f \equiv 0$) to the expressions on the left-hand side of Equations 7.7.7 and 7.7.8, establishing the truth of Equations 7.7.4 and 7.7.5. Equation 7.7.6 is established by the assumed consistency of $\hat{\beta}$ and the assumption that $A_2(t) < \infty$ for all $t \in \mathscr{T}$. Applying Theorem 16 in the previous section to Equation 7.7.3 necessarily implies that for all $t \in \mathscr{T}$, for all $\eta > 0$ and for all $\delta > 0$

$$P\left\{\sup_{s\in[0,t]}\left|\hat{A}_2^{(n)}(t) - A_2^{*(n)}(t)\right| > \eta\right\} \leq \frac{\delta}{\eta^2} + P\left\{\int_0^t \frac{\hat{\beta}^{(n)}(s)J^{(n)}(s)}{Y^{(n)}(s)}\alpha_{2+4}(s)ds > \delta\right\}.$$

Equation 7.7.4 necessarily implies that as $n \to +\infty$,

$$\sup_{s\in[0,t]}\left|\hat{A}_2^{(n)}(t) - A_2^{*(n)}(t)\right| \xrightarrow{P} 0.$$

Note that the following expression

$$\begin{aligned}\left|A_2^{*(n)}(t) - A_2(t)\right| &= \int_0^t \left(\beta(s) - J^{(n)}\hat{\beta}^{(n)}(s)\right)\alpha_{2+4}(s)ds \\ &= \int_0^t \left(\beta(s) - \hat{\beta}^{(n)}(s) + \hat{\beta}^{(n)}(s) - J^{(n)}\hat{\beta}^{(n)}(s)\right)\alpha_{2+4}(s)ds \\ &= \int_0^t \left(\beta(s) - \hat{\beta}^{(n)}(s)\right)\alpha_{2+4}(s)ds + \int_0^t \left(1 - J^{(n)}(s)\right)\hat{\beta}^{(n)}(s)\alpha_{2+4}(s)ds,\end{aligned}$$

converges in probability to zero by Equations 7.7.5 and 7.7.6. By the triangle inequality this necessarily implies

$$\sup_{s\in[0,t]}\left|\hat{A}_2^{(n)}(t) - A_2(t)\right| \xrightarrow{P} 0,$$

and the estimator is consistent. The sum of two consistent estimators is itself consistent by Theorem 11 and the estimate of $A_{1+2}(t)$ is equal to

$$\hat{A}_{1+2}(t) = \int_0^t \frac{\hat{\beta}^{(n)}(s)}{Y^{(n)}(s)}dN_{2+4}^{(n)}(s) + \int_0^t \frac{1}{Y^{(n)}(s)}dN_1^{(n)}(s),$$

Which, because $N_1^{(n)}$ and $N_{2+4}^{(n)}$ give mass one at each of their successive jumps and zero mass everywhere else, is just

$$\hat{A}_{1+2}(t) = \sum_{j:T_j\leq t}\left(\frac{\hat{\beta}(T_j)}{Y^{(n)}(T_j)}\right)^{(1-\epsilon)} + \sum_{j:T_j\leq t}\left(\frac{1}{Y^{(n)}(T_j)}\right)^{\epsilon_j\delta_j}.$$

Applying the same arguments described in the previous subsection, the resulting Kaplan-Meier estimator is equal to

$$\hat{S}(t) = \prod_{s\leq t}\left(1 - \Delta\hat{A}_{1+2}(s)\right) = \prod_{j:T_j\leq t}\left(1 - \frac{\hat{\beta}(T_j)}{Y^{(n)}(T_j)}\right)^{(1-\epsilon)}\left(1 - \frac{1}{Y^{(n)}(T_j)}\right)^{\epsilon_j\delta_j}.$$

This is identical to Equation 7.3.1 which was the Kaplan-Meier estimator derived in Section 7.3. From the same argument of Theorem 10 it follows immediately that this estimator is consistent as well.

7.7.3 Application of the Consistency Proof

This subsection will present a simple application of the proof in the previous sub-section, for the case where the "missingness" mechanism is not dependent on observed or unobserved data and is therefore missing completely at random (MCAR), and also that δ_i and T_i are independent.
As discussed in sub-section 7.7.1, these assumptions necessarily imply that $P\{\delta_i = 1|\epsilon_i = 0, T_i = t_i\} = P\{\delta_i = 1\}$ which is a Bernoulli random variable and as such, it is well known (A thorough demonstration can be found in Baltagi[7] on pages 15 and 19) that this can be estimated by

$$\hat{\beta}^{(n)} = \frac{\sum_{i=1}^{n} I\{\delta_i = 1, \epsilon_i = 1\}}{\sum_{i=1}^{n} I\{\epsilon_i = 1\}},$$

and that this estimator is consistent. It can be seen that $\hat{\beta}^{(n)}$ are bounded by 1, and because $\hat{\beta}^{(n)}$ are independent of time t this is a predictable process. The resulting estimator takes the form

$$\hat{S}(t) = \prod_{j:T_j \leq t} \left(1 - \frac{d_j}{(1 - l_j)n_j}\right)^{(1-\epsilon)} \left(1 - \frac{1}{n_j}\right)^{\epsilon_j \delta_j},$$

where d_j, l_j and n_j are as defined in Sections 7.2.1 and 7.3.2. Having satisfied the conditions of the theorem proven in the previous sub-section, this estimator is consistent.

7.8 Summary

In this chapter a new statistical estimator was proposed for the estimation of marginal distributions from classical competing risks data under the assumption of independence and in the presence of some occasional missing data on the cause of system failure. The proposed estimator is a variant of the Kaplan-Meier product limit estimator. A redistribution to the right algorithm, similar to the one for the traditional Kaplan-Meier estimator, was developed and this was shown to be equivalent to the formally derived estimator. Some illustrative examples were given and the estimator was shown to be consistent.

Chapter 8

Conclusions and Further Work

This chapter will draw together the various strands of this thesis and summarise what has been achieved. The Contribution of this thesis is to theoretical knowledge of competing risks in reliability, through the examination of three problems identified from the literature. These developments are summarised and further work arising from them is discussed in Sections 8.1, 8.2 and 8.3. Finally, a general discussion of the difficulties and some of the potential barriers that exist for analysts when attempting to apply competing risk modelling is given in Section 8.4, with a brief mention of those areas whose development is considered by the author to be of most significance in helping to close the gap between practice and theory.

8.1 Identifiability of Several GNVA Models

In Chapter 5 two subclasses of the asymmetric GCR framework using virtual ages were studied. Following a line of argument developed by Bedford and Lindqvist [11], the joint distributions of these model classes were both shown to be identifiable subject to some technical conditions.

By generalising the proof framework provided by Bedford and Lindqvist it is now quite straightforward to consider other subclasses of the asymmetric GCR framework using virtual ages which satisfy certain properties. One of the main limitations of the work in this chapter is that only those virtual age models that allow the possibility for the component to be restored to a state as good as new can be considered. This goes back to the requirement of ergodicity that every element in the state space can potentially be visited from any other starting element in the state

space. Unfortunately, as is clear from the review given of univariate maintenance models, most of the existing virtual age models do not allow for that possibility. An open question of considerable interest to the author is whether or not an alternative proof technique that does not utilise the notion of ergodicity can be brought to bear on this identifiability problem. Such a technique would greatly enhance the range of useful results that could be generated in this area.

8.2 The Random Signals Model

In chapter 6 a model developed by Bedford and Alkali[10] for opportunistic maintenance in a competing risks framework was explored and closed form expressions were derived for the marginal functions for preventative and corrective maintenance, as well as for the sub-distribution function for corrective maintenance. In view of the large number of algebraic manipulations involved, a simulation study was conducted to validate the correctness of these expressions.

It is believed by the author that the natural next step would be to calculate an expression for the sub-distribution function for preventative maintenance. Also, as discussed in Subsection 2.3.11 it is often useful to study the shape of the Φ function when considering model selection. Calculating an expression for the Φ function in this case would also be useful. Another interesting problem would be to consider the modelling of opportunistic preventative maintenance in the presence of deterministic calender-based preventative maintenance. The deterministic preventative maintenance would act as a right-censoring on the existing model.

8.3 Variant of the Kaplan Meier Estimator (for Missing Data)

In Chapter 7 a new estimator based on the Kaplan-Meier product limit estimator was derived. This estimator explicitly handles datasets with missing causes for system halts. A variant of the well-known distribution to the right algorithm was also developed and shown to be equivalent to the formal closed-form expression. The use of this estimator is demonstrated using some simple example datasets and the estimator is then shown to be consistent using some techniques from Martingale theory.

In the current mathematical literature there are many useful properties of statistical estimators.

Having demonstrated that this estimator is consistent, it is believed by the author that among the most promising candidate properties to consider for further work would be to try and demonstrate that the estimator possesses asymptotic normality.

Asymptotic normality can be defined in the following manner. Given a sequence of observations $\{O_i\}_{i>0}$, one can consider the sequence of estimators $\left\{\hat{\Theta}_i\right\}_{i>0}$, where $\hat{\Theta}_n$ uses the first n observations to estimate the true parameter Θ_0. The sequence of estimators is said to possess asymptotic normality if the following property holds

$$\sqrt{n}(\hat{\Theta}_n - \Theta_0) \sim N(0, \sigma) \tag{8.3.1}$$

for some variance σ.

8.4 Model Selection and The Theory-Practice Gap

This thesis has focused on theoretical developments in competing risks modelling for mechanical systems reliability. This section will attempt to bring together several of the important points regarding the application of these theoretical models in practice.

In Section 2.3 a number of classical competing risk models for reliability were introduced and two important points were made about model selection. Firstly, that the Φ function has a signature shape for several reliability models and this can be used as a criterion for model selection. Secondly, model selection is hampered by the lack of theoretical development for several classical competing risk models regarding basic properties of the Φ function and whether the model class is identifiable. These gaps should be explored in order to better facilitate the development of guidance for model selection.

In Section 2.2 a statistical goodness of fit test developed by Bedford and Meilijson was introduced. This too can be used in the process of model selection, although the test can be quite conservative. Certainly the question of identifiability for the marginal distributions complicates the process of model validation significantly in the case of classical competing risks. One of the really promising aspects of the work discussed in Chapter 5 is that there are classes of maintenance models which can be used in the generalised competing risks framework for which identifiability of the marginal distributions is preserved. Of course this relies on the appropriateness of those particular maintenance models for the situation under study. Much of the above mentioned promise will depend on

future work to uncover more classes of maintenance models for which identifiability of the marginal distributions is preserved and the frequency with which such maintenance models fit the behaviour encountered in practice.

Another important aspect of these models is that often they have been constructed to answer specific questions, such as the Dependent Copula model, or they have been constructed around a particular engineering scenario. The discussion of opportunistic maintenance models in competing risks in Chapter 6 is an example of this latter construction, as are the Mixed Exponential and Random Clipping models discussed in Section 2.3. This can also be a useful guide in model selection, by rejecting those models for which the heuristic interpretation of the engineering scenario underpinning them is not compatible with the circumstances that one is trying to model.

It should be noted that many of these models have been developed by academics and are not widely known in the reliability community. A highly significant factor in this lack of awareness is the difficulty in model selection. This is yet another reason why it is important to improve the theory for existing models that are under-developed, in order to enhance the potential for useful guidance in model selection.

It should also be noted that even in situations where suitable data collection processes are in place, much of the theoretical discussion does not take into account the inherent lack of consistency and completeness often encountered in raw data when first approaching a situation in practice. Chapter 1 introduced several of the important assumptions that have to be made in order to arrive at a data structure that can be modelled using competing risks, while glossing over some of the problems that need to be overcome at those early stages in data cleaning.

Consider some of the following examples of potential problems often encountered in practice. Sometimes system failures will only be detected at regular intervals or the accuracy of the time-stamps on the various records may be suspect. It is also difficult in some situations to decide exactly when a system has failed. For example, the CCPS[46] group failure modes according to three different categories:

- **catastrophic** - A failure that is both sudden and causes termination of one or more fundamental functions,
- **degraded** - A failure that is gradual or partial,
- **incipient** - An imperfection in the state or condition of equipment such that a degraded or

catastrophic failure can be expected to result if corrective action is not taken.

It may prove very problematic to try to decide the exact moment that the system fails in the case of a degraded failure mode type.

Also, some failure mechanisms act in calendar time, for example weathering or chemical erosion. But some failure mechanisms act only when the system is functioning, for example wear of moving parts caused by rubbing or mechanical torque. In most cases a choice of time measurement (whether it is actual time, functional time or some other alternative) is made based on which failure mechanism is the most prevalent factor.

The important point to notice is that in each of these scenarios a choice is being made. Often, the most that can be done in these cases is to try and make that choice sensibly and consistently (which is another reason why the standards for reliability data collection discussed in Section 1.3 are very important). Sometimes help can be found from theoretical improvements such as the work done in Chapter 7, which allows for estimation using datasets in which the data describing the cause of system failure is sometimes missing.

There is certainly much work to be done in classical competing risks to improve the usability of the existing models and to aid analysts in areas where the raw data encountered is perhaps not quite in the ideal condition one would like. There is also significant work still to be done in generalised competing risk modelling by considering different classes of maintenance models that occur in practice and developing the properties of these models. Much of this landscape is still unexplored.

Bibliography

[1] F.M. Akhmedjanov, *Reliability databases: State-of-the-art and perspectives*, Ris National Laboratory, 2001.

[2] C.W. Allen, S. Anderson, R. Erickson, W. Linebarger, J.C. Sheppard, and M. Stanek, *Opportunistic or event-driven maintenance at the stanford linear accelerator center*, 1997.

[3] P.K. Andersen, . Borgan, R.D. Gill, and N. Keiding, *Statistical models based on counting processes*, Springer Series in Statistics, Springer-Verlag, New York, 1993.

[4] J.D. Andrews and T.R. Moss, *Reliability and risk assessment*, 2nd ed., Professional Engineering Pub., London, 2002.

[5] H.E. Ascher and H. Feingold, *Repairable systems reliability: Modelling, inference, misconceptions and their causes*, Marcel Dekker, New York, 1984.

[6] M. Balazs, *Sum of independent exponentials*, `http://www.math.bme.hu/~balazs/sumexp.pdf`, [Online; accessed 22nd September 2010].

[7] B.H. Baltagi, *Econometrics*, Springer Texts in Business and Economics, Springer-Verlag, Berlin, 2011.

[8] A.K. Basu, *Meaure theory and probability*, Prentice Hall of India, New Delhi, 1999.

[9] T. Bedford, *Competing risk modelling in reliability*, Modern Statistical and Mathematical Methods in Reliability, World Scientific, Singapore, 2006, pp. 1–16.

[10] T. Bedford and B.M. Alkali, *Competing risk and opportunistic informative maintenance modelling*, Proceedings of the IMech E Part O: Journal of Risk and Reliability **223** (2009), no. 4, 363–372.

[11] T. Bedford and B.H. Lindqvist, *The identifiability problem for repairable systems subject to competing risks*, Advances in Applied Probability **36** (2004), no. 3, 774–790.

[12] T. Bedford and I. Meilijson, *The marginal distributions of lifetime variables which right censor each other*, Ims Lecture Notes Monograph Series (H. Koul and J. Deshpande, eds.), vol. 27, Institute of Mathematical Statistics, 1995.

[13] ———, *A characterization of marginal distributions of (possibly dependent) lifetime variables which right censor each other*, The Annals of Statistics **25** (1997), no. 4, 1622–1645.

[14] A. Bendell and A.G. Cannon, *Reliability data banks*, Elsevier Applied Science, London, New York, 1991.

[15] R. Bhaskar, *A realist theory of science*, Leeds books Ltd., 1975.

[16] H.W. Block, W.S. Borges, and T.H. Savits, *Age-dependent minimal repair*, Journal of Applied Probability **22** (1985), 370–385.

[17] D. Bocchetti, M. Giorgio, M. Guida, and G. Pulcini, *A competing risk model for the reliability of cylinder liners in ship diesel engines*, [Presented at ENBIS 2007, Glasgow].

[18] M. Brown and F. Proschan, *Imperfect repair*, Journal of Applied Probability **20** (1983), 851–859.

[19] C. Bunea and T. Bedford, *The effect of model uncertainty on maintenance optimization*, IEEE TRANSACTIONS ON RELIABILITY **51** (2002), no. 4, 486–493.

[20] C. Bunea, R. Cooke, and B.H. Lindqvist, *Competing risk perspective over reliability databases*, MMR 2002 (Trondheim), 2002, pp. 131–134.

[21] J.K. Chan and L. Shaw, *Modeling repairable systems with failure rates that depend on age and maintenance*, IEEE TRANSACTIONS ON RELIABILITY **42** (1993), no. 4, 566–571.

[22] A. Christer, *A review of delay time analysis for modelling plant maintenance*, Stochastic Models in Reliability and Maintenance, Springer, 2002.

[23] R. Cooke, *The design of reliability data bases, part i: Review of standard design concepts*, Reliability Engineering and System Safety **51** (1996), 137–146.

[24] ———, *The design of reliability data bases, part ii: Competing risk and data compression*, Reliability Engineering and System Safety **51** (1996), 209–223.

[25] R. Cooke and T. Bedford, *Reliability databases in perspective*, IEEE TRANSACTIONS ON RELIABILITY **51** (2002), no. 3, 294–310.

[26] R.M. Cooke, *The total time on test statistic and age-dependent censoring*, Statistics and Probability Letters **18** (1993), 307–312.

[27] F. Coolen, *An example of the redistribution to the right algorithm (private communication)*, 2009.

[28] R.V. Craiu and T. Duchesne, *Inference based on the em algorithm for the competing risk model with masked causes of failure*, Biometrika **91** (2004), no. 3, 545–558.

[29] M. Crowder, *On the identifiability crisis in competing risks analysis*, Scandinavian Journal of Statistics **18** (1991), 223–233.

[30] ———, *Classical competing risks*, chapman and Hall/CRC, 2001.

[31] US Defense Department, *Mil-hdbk-217e: Military handbook, reliability prediction of electronic equipment*, US Department of Defense, 1986.

[32] J.V. Deshpande and S.G. Purohit, *Life time data: Statistical models and methods*, Series on Quality, Reliability and Engineering Statistics, vol. 11, World Scientific Publishing, Singapore, 2005.

[33] I. Dewan, *Some lectures on statistics (unpublished)*, 2010.

[34] A. Dewanji and D. Sengupta, *Estimation of competing risks with general missing pattern in failure types*, Biometrics **59** (2003), no. 4, 1063–1070.

[35] Y. Dijoux, L. Doyen, and O. Gaudoin, *Conditionally independent generalized competing risks for maintenance analysis*, MMR 2007 (Glasgow), 2007.

[36] Y. Dijoux and O. Gaudoin, *The alert-delay competing risks model for maintenance analysis*, Journal of Statistical Planning and Inference **139** (2009), 1587–1603.

[37] G.E. Dinse, *Nonparametric estimation for partially complete time and type of failure data*, Biometrics **38** (1982), no. 2, 417–431.

[38] J.L. Doob, *Stochastic processes*, John Wiley and Sons, 1953.

[39] C. Dorado, M. Hollander, and J. Sethuraman, *Nonparametric estimation for a general repair model*, The Annals of Statistics **25** (1997), no. 3, 1140–1160.

[40] L. Doyen and O. Gaudoin, *Classes of imperfect repair models based on reduction of failure intensity or virtual age*, Reliability Engineering and System Safety **84** (2004), 45–56.

[41] ———, *Imperfect maintenance in a generalized competing risk framework*, Journal of Applied Probability **43** (2006), no. 3, 825–839.

[42] B. Efron, *The two sample problem with censored data*, Proceedings of the fifth Berkeley Symposium (Berkeley, CA), vol. 4, University of California Press, 1967, pp. 831–853.

[43] G. Elvebakk, B.H. Lindqvist, and K. Heggland, *The trend-renewal process for statistical analysis of repairable systems*, Technometrics **45** (2003), no. 1, 31–44.

[44] B.J. Flehinger, B. Reiser, and E. Yashchin, *Parametric modeling for survival with competing risks and masked failure causes*, Lifetime Data Analysis **8** (2002), 177–203.

[45] T. Fleming and D. Harrington, *Counting processes and survival analysis*, John Wiley and Sons, 1991.

[46] Center for Chemical Process Safety, *Guidelines for process equipment reliability data with data tables*, American Institute of Chemical Engineers, New York, 1989.

[47] ———, *Guidelines for improving plant reliability through data collection and analysis*, American Institute of Chemical Engineers, New York, 1998.

[48] J. Fragola, *Reliability and risk analysis data base development: An historical perspective*, Reliability Engineering and System Safety **51** (1996), 125–136.

[49] E. Goetghebeur and L. Ryan, *Analysis of competing risks survival data when some failure types are missing*, Biometrika **82** (1995), no. 4, 821–833.

[50] A.E. Green and A.J. Bourne, *Reliability technology*, Wiley-interscience, 1972.

[51] D.G. Hoel, *A representation of mortality data by competing risks*, Biometrics **28** (1972), no. 2, 475–488.

[52] P. Hokstad and Jensen, *Predicting the failure rate for components that go through a degradation state*, Safety and Reliability (S. Lydersen, G.K. Hansen, and H.A. Sandtorv, eds.), Balkema, Rotterdam, 1998, pp. 389–396.

[53] A. Hoyland and M. Rausand, *System reliability theory: Models and statistical methods*, John Wiley and Sons, 1994.

[54] IEC, *Iec 60300-3-2: Dependability management, part 3 - application guide, section 2 - collection of dependability data from the field*, International Electrotechnical Commission, Geneva, 1993.

[55] ISO, *Iso 14224: Petroleum, petrochemical and natural gas industries - collection and exchange of reliability and maintenance data for equipment*, International Organization for Standardization, Geneva, 2006.

[56] E.L. Kaplan and P. Meier, *Nonparametric estimation from incomplete observations*, Journal of the American Statistical Association **53** (1958), no. 282, 457–481.

[57] M. Kijima, *Some results for repairable systems with general repair*, Journal of Applied Probability **26** (1989), 89–102.

[58] H. Langseth and Lindqvist B.H., *Competing risks for repairable systems: A data study*, Journal of Statistical Planning and Inference **136** (2006), 1687–1700.

[59] H. Langseth and B.H. Lindqvist, *A maintenance model for components exposed to several failure mechanisms and imperfect repair*, Mathematical and Statistical Methods in Reliability. Quality, Reliability and Engineering Statistics. (K. Doksum and B.H. Lindqvist, eds.), World Scientific, Singapore, 2003, pp. 415–430.

[60] D.V. Lindley and W.F. Scott, *New cambridge statistical tables*, 2nd ed., Cambridge University Press, Cambridge, 1995.

[61] B.H. Lindqvist, *On the statistical modelling and analysis of repairable systems*, Statistical Science **21** (2006), no. 4, 532–551.

[62] B.H. Lindqvist and G. Skogsrud, *Modelling of dependent competing risks by first passage times of wiener processes*, IIE Transactions **41** (2009), 72–80.

[63] B.H. Lindqvist, B. Støve, and H. Langseth, *Modelling of dependence between critical failure and preventive maintenance: The repair alert model*, Journal of Statistical Planning and Inference **136** (2006), 1701–1717.

[64] R.J. Little and D.B. Rubin, *Statistical analysis with missing data*, 2nd ed., Wiley Series in Probability and Statistics, John Wiley and Sons, New Jersey, 2002.

[65] S. Lo, *Estimating a survival function with incomplete cause-of-death data*, Journal of Multivariate Analysis **39** (1991), 217–235.

[66] M.A.K. Malik, *Reliable preventative maintenance scheduling*, AIIE Transactions **11** (1979), 221–228.

[67] D.R. Miller, *A note on independence of multivariate lifetimes in competing risk models*, Annals of Statistics **5** (1977), 576–579.

[68] M. Miyakawa, *Analysis of incomplete data in competing risks model*, IEEE Transactions on Reliability **R-33** (1984), no. 4, 293–296.

[69] T.R. Moss, *The reliability data handbook*, Professional Engineering Pub., London, 2005.

[70] A. Ndas, *On estimating the distribution of a random vector when only the smallest coordinate is observable*, Technometrics **12** (1970), no. 4, 923–924.

[71] OREDA Participants, *Oreda (offshore reliability data) handbook, volume i and ii*, 5th ed., Det Norske Veritas, Trondhein, 2009.

[72] D.F. Percy and B.M. Alkali, *Scheduling preventive maintenance for oil pumps using generalized proportional intensities models*, International Transactions in Operational Research **14** (2007), 547–563.

[73] A.V. Peterson, *Bounds for a joint distribution function with fixed sub-distribution functions: Application to competing risks*, Proc. Nat. Acad. Sci. **73** (1976), no. 1, 11–13.

[74] H. Pham and H. Wang, *Imperfect maintenance*, European Journal of Operational Research **94** (1996), 425–438.

[75] K.R. Popper, *Conjectures and refutations: The growth of scientific knowledge*, 2nd ed., Routledge classics, 2002.

[76] H. Procaccia, P. Aufort, and S. Arsenis, *The european indistry reliability data bank*, 3rd ed., Crete University Press, 1998.

[77] A.H. Racine-Poon and D.G. Hoel, *Nonparametric estimation of the survival function when cause of death is uncertain*, Biometrics **40** (1984), no. 4, 1151–1158.

[78] S.M. Ross, *Introduction to probability models*, 8th ed., Academic Press, San Diego, 2003.

[79] I. Shin, T.J. Lim, and C.H. Lie, *Estimating parameters of intensity function and maintenance effect for repairable unit*, Reliability Engineering and System Safety **54** (1996), 1–10.

[80] A. Stuart and J.K. Ord, *Kendall's advanced theory of statistics vol. 2: Classical inference and relationship*, 5th ed., Edward Arnold, London, 1991.

[81] W.A. Thompson, *Point process models with applications to safety and reliability*, Chapman and Hall, 1988.

[82] A. Tsiatis, *A nonidentifiability aspect in the problem of competing risks*, Proceedings of the National Academy of Science, USA **72** (1975), 20–22.

[83] J.A.M. Van der Weide and T. Bedford, *Competing risks and eternal life*, Safety and Reliability (Proceedings of Esrel '98) (S. Lydersen, G.K. Hansen, and H.A. Sandtorv, eds.), vol. 2, Balkema, Rotterdam, 1998, pp. 1359–1364.

[84] M. Zheng and J.P. Klein, *Estimates of marginal survival for dependent competing risks based on an assumed copula*, Biometrika **82** (1995), no. 1, 127–138.

Part III

Appendix

Appendix A

Extra Calculations for the Random Signals Model

A.1 Remaining Calculations for the Sub-distribution of X

This section details the algebra that leads to the result given by Equation 6.3.2 in Section 6.3.3. For ease of exposition, the expression for the Sub-distribution of X calculated so far on Page 80

will be separated into three parts and each part will be considered separately as follows:

$$
\begin{aligned}
F_X^*(t) &= L_1(t) + L_2(t) + L_3(t),\\
L_1(t) &= \sum_{w=0}^{k-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} \left[1 - e^{-\{\sum_{i=1}^{w}[\lambda_{i-1}-\lambda_i]\times r_i + \lambda_w t\}}\right]\\
&\qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1,\\
L_2(t) &= \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} \sum_{u=k}^{w-1} \left[\int_{r_u}^{r_{u+1}} \left[1 - e^{-\{\sum_{i=1}^{u}[\lambda_{i-1}-\lambda_i] r_i + \lambda_u z\}}\right] \rho e^{-\rho[z-r_k]} dz\right]\\
&\qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1\\
&\quad + \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} \int_{r_w}^t \left[1 - e^{-\{\sum_{i=1}^{w}[\lambda_{i-1}-\lambda_i] r_i + \lambda_w z\}}\right] \rho e^{-\rho[z-r_k]} dz\\
&\qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1\\
&\quad + \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} \left[1 - e^{-\{\sum_{i=1}^{w}[\lambda_{i-1}-\lambda_i] r_i + \lambda_w t\}}\right] \int_t^{+\infty} \rho e^{-\rho[z-r_k]} dz\\
&\qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1,\\
L_3(t) &= \int_0^t \cdots \int_{r_{n-1}}^t \sum_{u=k}^{n-1} \left[\int_{r_u}^{r_{u+1}} \left[1 - e^{-\{\sum_{i=1}^{u}[\lambda_{i-1}-\lambda_i] r_i + \lambda_u z\}}\right] \rho e^{-\rho[z-r_k]} dz\right]\\
&\qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1\\
&\quad + \int_0^t \cdots \int_{r_{n-1}}^t \int_{r_n}^t \left[1 - e^{-\{\sum_{i=1}^{n}[\lambda_{i-1}-\lambda_i] r_i + \lambda_n z\}}\right] \rho e^{-\rho[z-r_k]} dz\\
&\qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1\\
&\quad + \int_0^t \cdots \int_{r_{n-1}}^t \left[1 - e^{-\{\sum_{i=1}^{n}[\lambda_{i-1}-\lambda_i] r_i + \lambda_n t\}}\right] \int_t^{+\infty} \rho e^{-\rho[z-r_k]} dz\\
&\qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1.
\end{aligned}
$$

A.1.1 The First Group of Terms

This subsection will present a series of steps in the calculation of $L_1(t)$.By multiplying out the inner-most brackets, this necessarily implies

$$\begin{aligned}L_1(t) = &\sum_{w=0}^{k-1}\int_0^t \dots \int_{r_{w-1}}^t \int_t^{+\infty} \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1}\gamma_{w-1}e^{-[\gamma_{w-1}-\gamma_w]r_w}dr_w \dots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1}dr_1 \\ &- \sum_{w=0}^{k-1}\int_0^t \dots \int_{r_{w-1}}^t \int_t^{+\infty} e^{-\{\sum_{i=1}^{w}[\lambda_{i-1}-\lambda_i]\times r_i+\lambda_w t\}} \\ &\qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1}\gamma_{w-1}e^{-[\gamma_{w-1}-\gamma_w]r_w}dr_w \dots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1}dr_1.\end{aligned}$$

It should be noted that in the second summation, the inner-most integral (the integral with respect to dr_{w+1}) can be brought to the right of the initial exponential terms as follows

$$\begin{aligned}L_1(t) = &\sum_{w=0}^{k-1}\int_0^t \dots \int_{r_{w-1}}^t \int_t^{+\infty} \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1}\gamma_{w-1}e^{-[\gamma_{w-1}-\gamma_w]r_w}dr_w \dots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1}dr_1 \\ &- \sum_{w=0}^{k-1}\int_0^t \dots \int_{r_{w-1}}^t e^{-\{\sum_{i=1}^{w}[\lambda_{i-1}-\lambda_i]\times r_i+\lambda_w t\}} \\ &\qquad \times \int_t^{+\infty} \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1}\gamma_{w-1}e^{-[\gamma_{w-1}-\gamma_w]r_w}dr_w \dots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1}dr_1.\end{aligned}$$

Solving the inner-most integrals in each summation and moving the resulting exponential terms outside the remaining integrals necessarily implies

$$\begin{aligned}L_1(t) = &\sum_{w=0}^{k-1} e^{-\gamma_w t}\int_0^t \dots \int_{r_{w-1}}^t \gamma_{w-1}e^{-[\gamma_{w-1}-\gamma_w]r_w}dr_w \dots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1}dr_1 \\ &- \sum_{w=0}^{k-1} e^{-\delta_w t}\int_0^t \dots \int_{r_{w-1}}^t e^{-\{\sum_{i=1}^{w}[\lambda_{i-1}-\lambda_i]\times r_i\}} \\ &\qquad \times \gamma_{w-1}e^{-[\gamma_{w-1}-\gamma_w]r_w}dr_w \dots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1}dr_1.\end{aligned}$$

A.1.2 The Second Group of Terms

This subsection will present a series of steps in the calculation of $L_2(t)$. By multiplying out the inner-most bracketed differences, this necessarily implies

$$
\begin{aligned}
L_2(t) = &\sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} \sum_{u=k}^{w-1} \int_{r_u}^{r_{u+1}} \rho e^{-\rho[z-r_k]} dz \\
&\qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
&- \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} \sum_{u=k}^{w-1} \left[\int_{r_u}^{r_{u+1}} e^{-\{\sum_{i=1}^{u} [\lambda_{i-1}-\lambda_i] r_i + \lambda_u z\}} \rho e^{-\rho[z-r_k]} dz \right] \\
&\qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
&+ \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} \int_{r_w}^t \rho e^{-\rho[z-r_k]} dz \\
&\qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
&- \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} \int_{r_w}^t e^{-\{\sum_{i=1}^{w} [\lambda_{i-1}-\lambda_i] r_i + \lambda_w z\}} \rho e^{-\rho[z-r_k]} dz \\
&\qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
&+ \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} \int_t^{+\infty} \rho e^{-\rho[z-r_k]} dz \\
&\qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1, \\
&- \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} e^{-\{\sum_{i=1}^{w} [\lambda_{i-1}-\lambda_i] r_i + \lambda_w t\}} \int_t^{+\infty} \rho e^{-\rho[z-r_k]} dz \\
&\qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1.
\end{aligned}
$$

The positive summations can all be grouped together to yield an integral of z from r_k to ∞. The resulting integral is that of the conditional density function of Z given $S(k) = r$ over the entire

domain of definition (which is just equal to one).

$$
\begin{aligned}
L_2(t) = &\sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
&- \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} \sum_{u=k}^{w-1} \left[\int_{r_u}^{r_{u+1}} e^{-\{\sum_{i=1}^{u} [\lambda_{i-1}-\lambda_i] r_i + \lambda_u z\}} \rho e^{-\rho[z-r_k]} dz \right] \\
&\qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
&- \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} \int_{r_w}^t e^{-\{\sum_{i=1}^{w} [\lambda_{i-1}-\lambda_i] r_i + \lambda_w z\}} \rho e^{-\rho[z-r_k]} dz \\
&\qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
&- \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} e^{-\{\sum_{i=1}^{w} [\lambda_{i-1}-\lambda_i] r_i + \lambda_w t\}} \int_t^{+\infty} \rho e^{-\rho[z-r_k]} dz \\
&\qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1.
\end{aligned}
$$

Grouping together the z's within the last three summations necessarily implies

$$
\begin{aligned}
L_2(t) = &\sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
&- \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} \sum_{u=k}^{w-1} \left[e^{-\{\sum_{i=1}^{u} [\lambda_{i-1}-\lambda_i] r_i - \rho r_k\}} \rho \int_{r_u}^{r_{u+1}} e^{-[\rho+\lambda_u] z} dz \right] \\
&\qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
&- \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} e^{-\{\sum_{i=1}^{w} [\lambda_{i-1}-\lambda_i] r_i - \rho r_k\}} \rho \int_{r_w}^t e^{-[\rho+\lambda_w] z} dz \\
&\qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
&- \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} e^{-\{\sum_{i=1}^{w} [\lambda_{i-1}-\lambda_i] r_i - \rho r_k\}} e^{-\lambda_w t} \int_t^{+\infty} \rho e^{-\rho z} dz \\
&\qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1.
\end{aligned}
$$

Solving those integrals over z in the last three summations necessarily implies

$$
\begin{aligned}
L_2(t) = &\sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
&- \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} \sum_{u=k}^{w-1} \left[\frac{\rho e^{-\{\sum_{i=1}^{u} [\lambda_{i-1}-\lambda_i] r_i - \rho r_k\}}}{\rho + \lambda_u} \left[e^{-[\rho+\lambda_u] r_u} - e^{-[\rho+\lambda_u] r_{u+1}} \right] \right] \\
&\qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
&- \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} \frac{\rho e^{-\{\sum_{i=1}^{w} [\lambda_{i-1}-\lambda_i] r_i - \rho r_k\}}}{\rho + \lambda_w} \left[e^{-[\rho+\lambda_w] r_w} - e^{-[\rho+\lambda_w] t} \right] \\
&\qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
&- \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} e^{-\{\sum_{i=1}^{w} [\lambda_{i-1}-\lambda_i] r_i - \rho r_k\}} e^{-[\rho+\lambda_w] t} \\
&\qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1.
\end{aligned}
$$

Multiplying out the inner-most brackets in the second and third summations and rearranging the

exponential terms necessarily implies

$$
\begin{aligned}
L_2(t) = &\sum_{w=k}^{n-1} \int_0^t \dots \int_{r_{w-1}}^t \int_t^{+\infty} \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \dots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
&- \sum_{w=k}^{n-1} \int_0^t \dots \int_{r_{w-1}}^t \int_t^{+\infty} \sum_{u=k}^{w-1} \left[\frac{\rho e^{\rho r_k}}{\rho + \lambda_u} e^{-\{\sum_{i=1}^{u-1} [\lambda_{i-1}-\lambda_i] r_i + [\lambda_{u-1}-\lambda_u] r_u + [\rho+\lambda_u] r_u\}} \right] \\
&\qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \dots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
&+ \sum_{w=k}^{n-1} \int_0^t \dots \int_{r_{w-1}}^t \int_t^{+\infty} \sum_{u=k}^{w-1} \left[\frac{\rho e^{\rho r_k}}{\rho + \lambda_u} e^{-\{\sum_{i=1}^{u} [\lambda_{i-1}-\lambda_i] r_i + [\rho+\lambda_u] r_{u+1}\}} \right] \\
&\qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \dots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
&- \sum_{w=k}^{n-1} \int_0^t \dots \int_{r_{w-1}}^t \int_t^{+\infty} \frac{\rho e^{\rho r_k}}{\rho + \lambda_w} e^{-\{\sum_{i=1}^{w-1} [\lambda_{i-1}-\lambda_i] r_i + [\lambda_{w-1}-\lambda_w] r_w + [\rho+\lambda_w] r_w\}} \\
&\qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \dots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
&+ \sum_{w=k}^{n-1} e^{-[\rho+\lambda_w] t} \int_0^t \dots \int_{r_{w-1}}^t \int_t^{+\infty} \frac{\rho e^{\rho r_k}}{\rho + \lambda_w} e^{-\{\sum_{i=1}^{w} [\lambda_{i-1}-\lambda_i] r_i\}} \\
&\qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \dots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
&- \sum_{w=k}^{n-1} e^{-[\rho+\lambda_w] t} \int_0^t \dots \int_{r_{w-1}}^t \int_t^{+\infty} e^{\rho r_k} e^{-\{\sum_{i=1}^{w} [\lambda_{i-1}-\lambda_i] r_i\}} \\
&\qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \dots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1.
\end{aligned}
$$

Merging the second and fourth summations, multiplying the sixth summation top and bottom by

$\rho + \lambda_w$ and merging the sixth summation with the fifth summation necessarily implies

$$
\begin{aligned}
L_2(t) = & \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
& - \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} \sum_{u=k}^{w} \left[\frac{\rho e^{\rho r_k}}{\rho + \lambda_u} e^{-\{\sum_{i=1}^{u-1} [\lambda_{i-1}-\lambda_i] r_i + [\rho+\lambda_{u-1}] r_u\}} \right] \\
& \qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
& + \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} \sum_{u=k}^{w-1} \left[\frac{\rho e^{\rho r_k}}{\rho + \lambda_u} e^{-\{\sum_{i=1}^{u} [\lambda_{i-1}-\lambda_i] r_i + [\rho+\lambda_u] r_{u+1}\}} \right] \\
& \qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
& + \sum_{w=k}^{n-1} e^{-[\rho+\lambda_w] t} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} \frac{[\rho - (\rho + \lambda_w)] e^{\rho r_k}}{\rho + \lambda_w} e^{-\{\sum_{i=1}^{w} [\lambda_{i-1}-\lambda_i] r_i\}} \\
& \qquad \times \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1.
\end{aligned}
$$

It should be noted that in the last three summations, the inner-most integrals (the integral with respect to dr_{w+1}) can be brought to the right of the initial exponential terms as follows

$$
\begin{aligned}
L_2(t) = & \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \int_t^{+\infty} \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
& - \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \sum_{u=k}^{w} \left[\frac{\rho e^{\rho r_k}}{\rho + \lambda_u} e^{-\{\sum_{i=1}^{u-1} [\lambda_{i-1}-\lambda_i] r_i + [\rho+\lambda_{u-1}] r_u\}} \right] \\
& \qquad \times \int_t^{+\infty} \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
& + \sum_{w=k}^{n-1} \int_0^t \cdots \int_{r_{w-1}}^t \sum_{u=k}^{w-1} \left[\frac{\rho e^{\rho r_k}}{\rho + \lambda_u} e^{-\{\sum_{i=1}^{u} [\lambda_{i-1}-\lambda_i] r_i + [\rho+\lambda_u] r_{u+1}\}} \right] \\
& \qquad \times \int_t^{+\infty} \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
& - \sum_{w=k}^{n-1} e^{-[\rho+\lambda_w] t} \int_0^t \cdots \int_{r_{w-1}}^t \frac{\lambda_w e^{\rho r_k}}{\rho + \lambda_w} e^{-\{\sum_{i=1}^{w} [\lambda_{i-1}-\lambda_i] r_i\}} \\
& \qquad \times \int_t^{+\infty} \gamma_w e^{-\gamma_w r_{w+1}} dr_{w+1} \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w] r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1.
\end{aligned}
$$

These inner-most integrals can then be solved and the resulting exponential terms can be placed

outside of the other integral terms in each summation as follows

$$\begin{aligned}
L_2(t) = &\sum_{w=k}^{n-1} e^{-\gamma_w t} \int_0^t \cdots \int_{r_{w-1}}^t \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w]r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&- \sum_{w=k}^{n-1} e^{-\gamma_w t} \int_0^t \cdots \int_{r_{w-1}}^t \sum_{u=k}^{w} \left[\frac{\rho e^{\rho r_k}}{\rho + \lambda_u} e^{-\{\sum_{i=1}^{u-1}[\lambda_{i-1}-\lambda_i]r_i + [\rho+\lambda_{u-1}]r_u\}} \right] \\
&\qquad \times \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w]r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&+ \sum_{w=k}^{n-1} e^{-\gamma_w t} \int_0^t \cdots \int_{r_{w-1}}^t \sum_{u=k}^{w-1} \left[\frac{\rho e^{\rho r_k}}{\rho + \lambda_u} e^{-\{\sum_{i=1}^{u}[\lambda_{i-1}-\lambda_i]r_i + [\rho+\lambda_u]r_{u+1}\}} \right] \\
&\qquad \times \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w]r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&- \sum_{w=k}^{n-1} e^{-[\rho+\delta_w]t} \int_0^t \cdots \int_{r_{w-1}}^t \frac{\lambda_w e^{\rho r_k}}{\rho + \lambda_w} e^{-\{\sum_{i=1}^{w}[\lambda_{i-1}-\lambda_i]r_i\}} \\
&\qquad \times \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w]r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1.
\end{aligned}$$

For further ease of exposition, the following definition is introduced.

Definition 5. *For any non-negative real number t, let the equation $G(w,u,k,t)$ be defined as follows, where $w \geq u \geq k \geq 1$ are all non-negative integers.*

$$\begin{aligned}
G(w,u,k,t) = &\int_0^t \int_{r_1}^t \cdots \int_{r_{w-1}}^t \left[e^{\rho r_k} e^{-\{\sum_{i=1}^{u-1}[\lambda_{i-1}-\lambda_i]r_i + [\rho+\lambda_{u-1}]r_u\}} \right] \\
&\qquad \times \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w]r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1.
\end{aligned}$$

Moving the summations over u outside of the remaining integrals and applying the above definition necessarily implies

$$\begin{aligned}
L_2(t) = &\sum_{w=k}^{n-1} e^{-\gamma_w t} \int_0^t \cdots \int_{r_{w-1}}^t \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w]r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&- \sum_{w=k}^{n-1} \sum_{u=k}^{w} \frac{\rho e^{-\gamma_w t}}{\rho + \lambda_u} G(w,u,k,t) \\
&+ \sum_{w=k}^{n-1} \sum_{u=k}^{w-1} \frac{\rho e^{-\gamma_w t}}{\rho + \lambda_u} G(w,u+1,k,t) \\
&- \sum_{w=k}^{n-1} \frac{\lambda_w e^{-[\rho+\delta_w]t}}{\rho + \lambda_w} \int_0^t \cdots \int_{r_{w-1}}^t e^{\rho r_k} e^{-\{\sum_{i=1}^{w}[\lambda_{i-1}-\lambda_i]r_i\}} \\
&\qquad \times \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w]r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1.
\end{aligned}$$

A.1.3 The Third Group of Terms

This subsection will present a series of steps in the calculation of $L_3(t)$. By multiplying out the inner-most bracketed differences, this necessarily implies

$$
\begin{aligned}
L_3(t) = &\int_0^t \cdots \int_{r_{n-1}}^t \sum_{u=k}^{n-1} \int_{r_u}^{r_{u+1}} \rho e^{-\rho[z-r_k]} dz \\
&\qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}]r_{n-1}} dr_{n-1} \cdots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&- \int_0^t \cdots \int_{r_{n-1}}^t \sum_{u=k}^{n-1} \left[\int_{r_u}^{r_{u+1}} e^{-\{\sum_{i=1}^{u}[\lambda_{i-1}-\lambda_i]r_i + \lambda_u z\}} \rho e^{-\rho[z-r_k]} dz \right] \\
&\qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}]r_{n-1}} dr_{n-1} \cdots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&+ \int_0^t \cdots \int_{r_{n-1}}^t \int_{r_n}^t \rho e^{-\rho[z-r_k]} dz \\
&\qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}]r_{n-1}} dr_{n-1} \cdots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&- \int_0^t \cdots \int_{r_{n-1}}^t \int_{r_n}^t e^{-\{\sum_{i=1}^{n}[\lambda_{i-1}-\lambda_i]r_i + \lambda_n z\}} \rho e^{-\rho[z-r_k]} dz \\
&\qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}]r_{n-1}} dr_{n-1} \cdots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&+ \int_0^t \cdots \int_{r_{n-1}}^t \int_{t}^{+\infty} \rho e^{-\rho[z-r_k]} dz \\
&\qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}]r_{n-1}} dr_{n-1} \cdots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&- \int_0^t \cdots \int_{r_{n-1}}^t e^{-\{\sum_{i=1}^{n}[\lambda_{i-1}-\lambda_i]r_i + \lambda_n t\}} \int_t^{+\infty} \rho e^{-\rho[z-r_k]} dz \\
&\qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}]r_{n-1}} dr_{n-1} \cdots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1.
\end{aligned}
$$

The positive terms can all be grouped together to yield an integral of z from r_k to ∞. The resulting integral is that of the conditional density function of Z given $S(k) = r$ over the entire domain of

definition (which is just equal to one).

$$
\begin{aligned}
L_3(t) = & \int_0^t \dots \int_{r_{n-1}}^t \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \dots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
& - \int_0^t \dots \int_{r_{n-1}}^t \sum_{u=k}^{n-1} \left[\int_{r_u}^{r_{u+1}} e^{-\{\sum_{i=1}^{u} [\lambda_{i-1}-\lambda_i] r_i + \lambda_u z\}} \rho e^{-\rho[z-r_k]} dz \right] \\
& \qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \dots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
& - \int_0^t \dots \int_{r_{n-1}}^t \int_{r_n}^t e^{-\{\sum_{i=1}^{n} [\lambda_{i-1}-\lambda_i] r_i + \lambda_n z\}} \rho e^{-\rho[z-r_k]} dz \\
& \qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \dots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
& - \int_0^t \dots \int_{r_{n-1}}^t e^{-\{\sum_{i=1}^{n} [\lambda_{i-1}-\lambda_i] r_i + \lambda_n t\}} \int_t^{+\infty} \rho e^{-\rho[z-r_k]} dz \\
& \qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \dots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1.
\end{aligned}
$$

Grouping together the z's within the last three terms necessarily implies

$$
\begin{aligned}
L_3(t) = & \int_0^t \dots \int_{r_{n-1}}^t \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \dots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
& - \int_0^t \dots \int_{r_{n-1}}^t \sum_{u=k}^{n-1} \left[e^{-\{\sum_{i=1}^{u} [\lambda_{i-1}-\lambda_i] r_i - \rho r_k\}} \rho \int_{r_u}^{r_{u+1}} e^{-[\rho+\lambda_u] z} dz \right] \\
& \qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \dots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
& - \int_0^t \dots \int_{r_{n-1}}^t e^{-\{\sum_{i=1}^{n} [\lambda_{i-1}-\lambda_i] r_i - \rho r_k\}} \rho \int_{r_n}^t e^{-[\rho+\lambda_n] z} dz \\
& \qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \dots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1 \\
& - \int_0^t \dots \int_{r_{n-1}}^t e^{-\{\sum_{i=1}^{n} [\lambda_{i-1}-\lambda_i] r_i - \rho r_k\}} e^{-\lambda_n t} \int_t^{+\infty} \rho e^{-\rho z} dz \\
& \qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \dots \gamma_0 e^{-[\gamma_0-\gamma_1] r_1} dr_1.
\end{aligned}
$$

Solving those integrals over z in the last three terms necessarily implies

$$
\begin{aligned}
L_3(t) = & \int_0^t \dots \int_{r_{n-1}}^t \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \dots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1 \\
& - \int_0^t \dots \int_{r_{n-1}}^t \sum_{u=k}^{n-1} \left[e^{-\{\sum_{i=1}^{u} [\lambda_{i-1}-\lambda_i] r_i - \rho r_k\}} \frac{\rho}{\rho + \lambda_u} [e^{-[\rho+\lambda_u] r_u} - e^{-[\rho+\lambda_u] r_{u+1}}] \right] \\
& \qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \dots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1 \\
& - \int_0^t \dots \int_{r_{n-1}}^t e^{-\{\sum_{i=1}^{n} [\lambda_{i-1}-\lambda_i] r_i - \rho r_k\}} \frac{\rho}{\rho + \lambda_n} [e^{-[\rho+\lambda_n] r_n} - e^{-[\rho+\lambda_n] t}] \\
& \qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \dots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1 \\
& - \int_0^t \dots \int_{r_{n-1}}^t e^{-\{\sum_{i=1}^{n} [\lambda_{i-1}-\lambda_i] r_i - \rho r_k\}} e^{-[\rho+\lambda_n] t} \\
& \qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \dots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1 .
\end{aligned}
$$

Multiplying out the inner-most brackets in the second and third terms and rearranging the exponential terms necessarily implies

$$
\begin{aligned}
L_3(t) = & \int_0^t \dots \int_{r_{n-1}}^t \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \dots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1 \\
& - \int_0^t \dots \int_{r_{n-1}}^t \sum_{u=k}^{n-1} \left[\frac{\rho e^{\rho r_k}}{\rho + \lambda_u} e^{-\{\sum_{i=1}^{u-1} [\lambda_{i-1}-\lambda_i] r_i + [\lambda_{u-1}-\lambda_u] r_u + [\rho+\lambda_u] r_u\}} \right] \\
& \qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \dots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1 \\
& + \int_0^t \dots \int_{r_{n-1}}^t \sum_{u=k}^{n-1} \left[\frac{\rho e^{\rho r_k}}{\rho + \lambda_u} e^{-\{\sum_{i=1}^{u} [\lambda_{i-1}-\lambda_i] r_i + [\rho+\lambda_u] r_{u+1}\}} \right] \\
& \qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \dots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1 \\
& - \int_0^t \dots \int_{r_{n-1}}^t \frac{\rho e^{\rho r_k}}{\rho + \lambda_n} e^{-\{\sum_{i=1}^{n-1} [\lambda_{i-1}-\lambda_i] r_i + [\lambda_{n-1}-\lambda_n] r_n + [\rho+\lambda_n] r_n\}} \\
& \qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \dots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1 \\
& + e^{-[\rho+\lambda_n] t} \int_0^t \dots \int_{r_{n-1}}^t \frac{\rho e^{\rho r_k}}{\rho + \lambda_n} e^{-\{\sum_{i=1}^{n} [\lambda_{i-1}-\lambda_i] r_i\}} \\
& \qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \dots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1 \\
& - e^{-[\rho+\lambda_n] t} \int_0^t \dots \int_{r_{n-1}}^t e^{\rho r_k} e^{-\{\sum_{i=1}^{n} [\lambda_{i-1}-\lambda_i] r_i\}} \\
& \qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \dots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1 .
\end{aligned}
$$

Merging the second and fourth terms, multiplying the sixth term top and bottom by $\rho + \lambda_n$ and then merging the sixth term with the fifth term necessarily implies

$$\begin{aligned}
L_3(t) = &\int_0^t \cdots \int_{r_{n-1}}^t \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \cdots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1 \\
&- \int_0^t \cdots \int_{r_{n-1}}^t \sum_{u=k}^{n} \left[\frac{\rho e^{\rho r_k}}{\rho + \lambda_u} e^{-\{\sum_{i=1}^{u-1} [\lambda_{i-1} - \lambda_i] r_i + [\rho + \lambda_{u-1}] r_u\}} \right] \\
&\qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \cdots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1 \\
&+ \int_0^t \cdots \int_{r_{n-1}}^t \sum_{u=k}^{n-1} \left[\frac{\rho e^{\rho r_k}}{\rho + \lambda_u} e^{-\{\sum_{i=1}^{u} [\lambda_{i-1} - \lambda_i] r_i + [\rho + \lambda_u] r_{u+1}\}} \right] \\
&\qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \cdots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1 \\
&+ e^{-[\rho + \lambda_n] t} \int_0^t \cdots \int_{r_{n-1}}^t \frac{[\rho - (\rho + \lambda_n)] e^{\rho r_k}}{\rho + \lambda_n} e^{-\{\sum_{i=1}^{n} [\lambda_{i-1} - \lambda_i] r_i\}} \\
&\qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \cdots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1.
\end{aligned}$$

For further ease of exposition, the following definition is introduced.

Definition 6. *For any non-negative real number t and for any non-negative integer $n \geq 2$, let the equation $F(n, u, k, t)$ be defined as follows, where $n \geq u \geq k \geq 1$ are all non-negative integers.*

$$\begin{aligned}
F(n, u, k, t) = &\int_0^t \int_{r_1}^t \cdots \int_{r_{n-1}}^t \left[e^{\rho r_k} e^{-\{\sum_{i=1}^{u-1} [\lambda_{i-1} - \lambda_i] r_i + [\rho + \lambda_{u-1}] r_u\}} \right] \\
&\qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \cdots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1.
\end{aligned}$$

Moving the summations over u outside of the remaining integrals and applying the above definition necessarily implies

$$\begin{aligned}
L_3(t) = &\int_0^t \cdots \int_{r_{n-1}}^t \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \cdots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1 \\
&- \sum_{u=k}^{n} \frac{\rho}{\rho + \lambda_u} F(n, u, k, t) \\
&+ \sum_{u=k}^{n-1} \frac{\rho}{\rho + \lambda_u} F(n, u+1, k, t) \\
&- \frac{\lambda_n e^{-[\rho + \lambda_n] t}}{\rho + \lambda_n} \int_0^t \cdots \int_{r_{n-1}}^t e^{\rho r_k} e^{-\{\sum_{i=1}^{n} [\lambda_{i-1} - \lambda_i] r_i\}} \\
&\qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}] r_{n-1}} dr_{n-1} \cdots \gamma_0 e^{-[\gamma_0 - \gamma_1] r_1} dr_1.
\end{aligned}$$

A.1.4 Merging the three Group of Terms

This subsection will present a series of steps in the calculation of $F_X^*(t)$. Rearranging the order of terms necessarily implies

$$
\begin{aligned}
F_X^*(t) = &\sum_{w=0}^{k-1} e^{-\gamma_w t} \int_0^t \cdots \int_{r_{w-1}}^t \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w]r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&+ \sum_{w=k}^{n-1} e^{-\gamma_w t} \int_0^t \cdots \int_{r_{w-1}}^t \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w]r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&+ \int_0^t \cdots \int_{r_{n-1}}^t \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}]r_{n-1}} dr_{n-1} \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&- \sum_{w=0}^{k-1} e^{-\delta_w t} \int_0^t \cdots \int_{r_{w-1}}^t e^{-\{\sum_{i=1}^w [\lambda_{i-1}-\lambda_i]\times r_i\}} \\
&\qquad \times \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w]r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&- \sum_{w=k}^{n-1} \frac{\lambda_w e^{-[\rho+\delta_w]t}}{\rho+\lambda_w} \int_0^t \cdots \int_{r_{w-1}}^t e^{\rho r_k} e^{-\{\sum_{i=1}^w [\lambda_{i-1}-\lambda_i]r_i\}} \\
&\qquad \times \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w]r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&- \frac{\lambda_n e^{-[\rho+\lambda_n]t}}{\rho+\lambda_n} \int_0^t \cdots \int_{r_{n-1}}^t e^{\rho r_k} e^{-\{\sum_{i=1}^n [\lambda_{i-1}-\lambda_i]r_i\}} \\
&\qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}]r_{n-1}} dr_{n-1} \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&- \sum_{w=k}^{n-1} \sum_{u=k}^{w} \frac{\rho e^{-\gamma_w t}}{\rho+\lambda_u} G(w,u,k,t) - \sum_{u=k}^{n} \frac{\rho}{\rho+\lambda_u} F(n,u,k,t) \\
&+ \sum_{w=k}^{n-1} \sum_{u=k}^{w-1} \frac{\rho e^{-\gamma_w t}}{\rho+\lambda_u} G(w,u+1,k,t) + \sum_{u=k}^{n-1} \frac{\rho}{\rho+\lambda_u} F(n,u+1,k,t).
\end{aligned}
$$

Applying Lemma 1 to the first three lines and separating out the last term of the last summation

in each of the last two lines necessarily implies

$$
\begin{aligned}
F_X^*(t) =& 1 - \sum_{w=0}^{k-1} e^{-\delta_w t} \int_0^t \cdots \int_{r_{w-1}}^t e^{-\{\sum_{i=1}^{w}[\lambda_{i-1}-\lambda_i]\times r_i\}} \\
&\qquad \times \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w]r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&- \sum_{w=k}^{n-1} \frac{\lambda_w e^{-[\rho+\delta_w]t}}{\rho+\lambda_w} \int_0^t \cdots \int_{r_{w-1}}^t e^{\rho r_k} e^{-\{\sum_{i=1}^{w}[\lambda_{i-1}-\lambda_i]r_i\}} \\
&\qquad \times \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w]r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&- \frac{\lambda_n e^{-[\rho+\lambda_n]t}}{\rho+\lambda_n} \int_0^t \cdots \int_{r_{n-1}}^t e^{\rho r_k} e^{-\{\sum_{i=1}^{n}[\lambda_{i-1}-\lambda_i]r_i\}} \\
&\qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}]r_{n-1}} dr_{n-1} \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&- \sum_{w=k}^{n-1} \sum_{u=k}^{w} \frac{\rho e^{-\gamma_w t}}{\rho+\lambda_u} G(w,u,k,t) - \sum_{u=k}^{n-1} \frac{\rho}{\rho+\lambda_u} F(n,u,k,t) - \frac{\rho}{\rho+\lambda_n} F(n,n,k,t) \\
&+ \sum_{w=k}^{n-1} \sum_{u=k}^{w-1} \frac{\rho e^{-\gamma_w t}}{\rho+\lambda_u} G(w,u+1,k,t) + \sum_{u=k}^{n-2} \frac{\rho}{\rho+\lambda_u} F(n,u+1,k,t) \\
&+ \frac{\rho}{\rho+\lambda_{n-1}} F(n,n,k,t).
\end{aligned}
$$

Exchanging the order of the double summations in the last two lines necessarily implies

$$
\begin{aligned}
F_X^*(t) =& 1 - \sum_{w=0}^{k-1} e^{-\delta_w t} \int_0^t \cdots \int_{r_{w-1}}^t e^{-\{\sum_{i=1}^{w}[\lambda_{i-1}-\lambda_i]\times r_i\}} \\
&\qquad \times \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w]r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&- \sum_{w=k}^{n-1} \frac{\lambda_w e^{-[\rho+\delta_w]t}}{\rho+\lambda_w} \int_0^t \cdots \int_{r_{w-1}}^t e^{\rho r_k} e^{-\{\sum_{i=1}^{w}[\lambda_{i-1}-\lambda_i]r_i\}} \\
&\qquad \times \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w]r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&- \frac{\lambda_n e^{-[\rho+\lambda_n]t}}{\rho+\lambda_n} \int_0^t \cdots \int_{r_{n-1}}^t e^{\rho r_k} e^{-\{\sum_{i=1}^{n}[\lambda_{i-1}-\lambda_i]r_i\}} \\
&\qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}]r_{n-1}} dr_{n-1} \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&- \sum_{u=k}^{n-1} \left[\sum_{w=u}^{n-1} \frac{\rho e^{-\gamma_w t}}{\rho+\lambda_u} G(w,u,k,t) + \frac{\rho}{\rho+\lambda_u} F(n,u,k,t) \right] - \frac{\rho}{\rho+\lambda_n} F(n,n,k,t) \\
&+ \sum_{u=k}^{n-2} \left[\sum_{w=u+1}^{n-1} \frac{\rho e^{-\gamma_w t}}{\rho+\lambda_u} G(w,u+1,k,t) + \frac{\rho}{\rho+\lambda_u} F(n,u+1,k,t) \right] \\
&+ \frac{\rho}{\rho+\lambda_{n-1}} F(n,n,k,t).
\end{aligned}
$$

Applying Lemma 3 to the last two lines necessarily implies

$$
\begin{aligned}
F_X^*(t) =& 1 - \sum_{w=0}^{k-1} e^{-\delta_w t} \int_0^t \cdots \int_{r_{w-1}}^t e^{-\{\sum_{i=1}^{w}[\lambda_{i-1}-\lambda_i]\times r_i\}} \\
&\qquad \times \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w]r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&- \sum_{w=k}^{n-1} \frac{\lambda_w e^{-[\rho+\delta_w]t}}{\rho+\lambda_w} \int_0^t \cdots \int_{r_{w-1}}^t e^{\rho r_k} e^{-\{\sum_{i=1}^{w}[\lambda_{i-1}-\lambda_i]r_i\}} \\
&\qquad \times \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w]r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&- \frac{\lambda_n e^{-[\rho+\lambda_n]t}}{\rho+\lambda_n} \int_0^t \cdots \int_{r_{n-1}}^t e^{\rho r_k} e^{-\{\sum_{i=1}^{n}[\lambda_{i-1}-\lambda_i]r_i\}} \\
&\qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}]r_{n-1}} dr_{n-1} \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&- \sum_{u=k}^{n-1} \frac{\rho}{\rho+\lambda_u} F(u,u,k,t) - \frac{\rho}{\rho+\lambda_n} F(n,n,k,t) \\
&+ \sum_{u=k}^{n-2} \frac{\rho}{\rho+\lambda_u} F(u+1,u+1,k,t) + \frac{\rho}{\rho+\lambda_{n-1}} F(n,n,k,t).
\end{aligned}
$$

Absorbing the last terms in the last two lines into their respective summations and replacing the functions $F(u,u,k,t)$ and $F(u+1,u+1,k,t)$ by their respective definitions necessarily implies

$$
\begin{aligned}
F_X^*(t) =& 1 - \sum_{w=0}^{k-1} e^{-\delta_w t} \int_0^t \cdots \int_{r_{w-1}}^t e^{-\{\sum_{i=1}^{w}[\lambda_{i-1}-\lambda_i]\times r_i\}} \\
&\qquad \times \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w]r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&- \sum_{w=k}^{n-1} \frac{\lambda_w e^{-[\rho+\delta_w]t}}{\rho+\lambda_w} \int_0^t \cdots \int_{r_{w-1}}^t e^{\rho r_k} e^{-\{\sum_{i=1}^{w}[\lambda_{i-1}-\lambda_i]r_i\}} \\
&\qquad \times \gamma_{w-1} e^{-[\gamma_{w-1}-\gamma_w]r_w} dr_w \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&- \frac{\lambda_n e^{-[\rho+\lambda_n]t}}{\rho+\lambda_n} \int_0^t \cdots \int_{r_{n-1}}^t e^{\rho r_k} e^{-\{\sum_{i=1}^{n}[\lambda_{i-1}-\lambda_i]r_i\}} \\
&\qquad \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}]r_{n-1}} dr_{n-1} \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&- \sum_{u=k}^{n} \frac{\rho}{\rho+\lambda_u} \int_0^t \int_{r_1}^t \cdots \int_{r_{u-1}}^t \left[e^{\rho r_k} e^{-\{\sum_{i=1}^{u-1}[\lambda_{i-1}-\lambda_i]r_i + [\rho+\lambda_{u-1}]r_u\}} \right] \\
&\qquad \times \gamma_{u-1} e^{-\gamma_{u-1} r_u} dr_u \gamma_{u-2} e^{-[\gamma_{u-2}-\gamma_{u-1}]r_{u-1}} dr_{u-1} \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&+ \sum_{u=k}^{n-1} \frac{\rho}{\rho+\lambda_u} \int_0^t \int_{r_1}^t \cdots \int_{r_u}^t \left[e^{\rho r_k} e^{-\{\sum_{i=1}^{u}[\lambda_{i-1}-\lambda_i]r_i + [\rho+\lambda_u]r_{u+1}\}} \right] \\
&\qquad \times \gamma_u e^{-\gamma_u r_{u+1}} dr_{u+1} \gamma_{u-1} e^{-[\gamma_{u-1}-\gamma_u]r_u} dr_u \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1.
\end{aligned}
$$

By grouping all the r_i's close to their respective integrals, this necessarily implies

$$\begin{aligned}
F_X^*(t) =& 1 - \sum_{w=0}^{k-1} e^{-\delta_w t} \int_0^t \dots \int_{r_{w-1}}^t \gamma_{w-1} e^{-[\delta_{w-1}-\delta_w]r_w} dr_w \dots \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1 \\
& - \sum_{w=k}^{n-1} \frac{\lambda_w e^{-[\rho+\delta_w]t}}{\rho+\lambda_w} \int_0^t \dots \int_{r_{w-1}}^t e^{\rho r_k} \gamma_{w-1} e^{-[\delta_{w-1}-\delta_w]r_w} dr_w \dots \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1 \\
& - \frac{\lambda_n e^{-[\rho+\lambda_n]t}}{\rho+\lambda_n} \int_0^t \dots \int_{r_{n-1}}^t e^{\rho r_k} \\
& \qquad \times \gamma_{n-1} e^{-[\delta_{n-1}-\lambda_n]r_n} dr_n \gamma_{n-2} e^{-[\delta_{n-2}-\delta_{n-1}]r_{n-1}} dr_{n-1} \dots \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1 \\
& - \sum_{u=k}^{n} \frac{\rho}{\rho+\lambda_u} \int_0^t \int_{r_1}^t \dots \int_{r_{u-1}}^t e^{\rho r_k} \\
& \qquad \times \gamma_{u-1} e^{-[\rho+\delta_{u-1}]r_u} dr_u \gamma_{u-2} e^{-[\delta_{u-2}-\delta_{u-1}]r_{u-1}} dr_{u-1} \dots \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1 \\
& + \sum_{u=k}^{n-1} \frac{\rho}{\rho+\lambda_u} \int_0^t \int_{r_1}^t \dots \int_{r_u}^t e^{\rho r_k} \\
& \qquad \times \gamma_u e^{-[\rho+\delta_u]r_{u+1}} dr_{u+1} \gamma_{u-1} e^{-[\delta_{u-1}-\delta_u]r_u} dr_u \dots \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1.
\end{aligned}$$

For all the integrals except those on the first line, the order of integration can be reversed, which necessarily implies

$$\begin{aligned}
F_X^*(t) =& 1 - \sum_{w=0}^{k-1} e^{-\delta_w t} \int_0^t \dots \int_{r_{w-1}}^t \gamma_{w-1} e^{-[\delta_{w-1}-\delta_w]r_w} dr_w \dots \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1 \\
& - \sum_{w=k}^{n-1} \frac{\lambda_w e^{-[\rho+\delta_w]t}}{\rho+\lambda_w} \int_0^t \int_0^{r_w} \dots \int_0^{r_2} e^{\rho r_k} \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1 \dots \gamma_{w-1} e^{-[\delta_{w-1}-\delta_w]r_w} dr_w \\
& - \frac{\lambda_n e^{-[\rho+\lambda_n]t}}{\rho+\lambda_n} \int_0^t \int_0^{r_n} \dots \int_0^{r_2} e^{\rho r_k} \\
& \qquad \times \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1 \dots \gamma_{n-2} e^{-[\delta_{n-2}-\delta_{n-1}]r_{n-1}} dr_{n-1} \gamma_{n-1} e^{-[\delta_{n-1}-\lambda_n]r_n} dr_n \\
& - \sum_{u=k}^{n} \frac{\rho}{\rho+\lambda_u} \int_0^t \int_0^{r_u} \dots \int_0^{r_2} e^{\rho r_k} \\
& \qquad \times \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1 \dots \gamma_{u-2} e^{-[\delta_{u-2}-\delta_{u-1}]r_{u-1}} dr_{u-1} \gamma_{u-1} e^{-[\rho+\delta_{u-1}]r_u} dr_u \\
& + \sum_{u=k}^{n-1} \frac{\rho}{\rho+\lambda_u} \int_0^t \int_0^{r_{u+1}} \dots \int_0^{r_2} e^{\rho r_k} \\
& \qquad \times \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1 \dots \gamma_{u-1} e^{-[\delta_{u-1}-\delta_u]r_u} dr_u \gamma_u e^{-[\rho+\delta_u]r_{u+1}} dr_{u+1}.
\end{aligned}$$

The following definition is introduced.

Definition 7. *For any non-negative integers u and k, let the equation $H(u,k)$ be defined as follows, where $u \geq k \geq 1$.*

$$H(u,k) = \int_0^{r_u} \dots \int_0^{r_2} e^{\rho r_k} \gamma_0 e^{-[\delta_0 - \delta_1] r_1} dr_1 \dots \gamma_{u-2} e^{-[\delta_{u-2} - \delta_{u-1}] r_{u-1}} dr_{u-1}.$$

Applying this definition necessarily implies

$$\begin{aligned} F_X^*(t) =& 1 - \sum_{w=0}^{k-1} e^{-\delta_w t} \int_0^t \dots \int_{r_{w-1}}^t \gamma_{w-1} e^{-[\delta_{w-1} - \delta_w] r_w} dr_w \dots \gamma_0 e^{-[\delta_0 - \delta_1] r_1} dr_1 \\ & - \sum_{w=k}^{n-1} \frac{\lambda_w e^{-[\rho + \delta_w] t}}{\rho + \lambda_w} \int_0^t H(w,k) \gamma_{w-1} e^{-[\delta_{w-1} - \delta_w] r_w} dr_w \\ & - \frac{\lambda_n e^{-[\rho + \lambda_n] t}}{\rho + \lambda_n} \int_0^t H(n,k) \gamma_{n-1} e^{-[\delta_{n-1} - \lambda_n] r_n} dr_n \\ & - \sum_{u=k}^{n} \frac{\rho}{\rho + \lambda_u} \int_0^t H(u,k) \gamma_{u-1} e^{-[\rho + \delta_{u-1}] r_u} dr_u \\ & + \sum_{u=k}^{n-1} \frac{\rho}{\rho + \lambda_u} \int_0^t H(u+1,k) \gamma_u e^{-[\rho + \delta_u] r_{u+1}} dr_{u+1}. \end{aligned}$$

Changing the dummy index in the second summation from w to u, remembering that $\gamma_n = 0$ and

applying Lemma 4 necessarily implies

$$
\begin{aligned}
F_X^*(t) = 1 &- \sum_{w=0}^{k-1} e^{-\delta_w t} \int_0^t \cdots \int_{r_{w-1}}^t \gamma_{w-1} e^{-[\delta_{w-1}-\delta_w]r_w} dr_w \ldots \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1 \\
&- \sum_{u=k}^{n} \left[\prod_{r=0}^{u-2} \gamma_r \right] \frac{\lambda_u e^{-[\rho+\delta_u]t}}{\rho+\lambda_u} \sum_{q=0}^{k-1} \frac{\int_0^t e^{-[\delta_q-\delta_{u-1}-\rho]r_u} \gamma_{u-1} e^{-[\delta_{u-1}-\delta_u]r_u} dr_u}{\left[\prod_{\substack{w=k \\ w\neq q}}^{u-1} (\delta_w - \delta_q + \rho) \right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q) \right]} \\
&- \sum_{u=k}^{n} \left[\prod_{r=0}^{u-2} \gamma_r \right] \frac{\lambda_u e^{-[\rho+\delta_u]t}}{\rho+\lambda_u} \sum_{q=k}^{u-1} \frac{\int_0^t e^{-[\delta_q-\delta_{u-1}]r_u} \gamma_{u-1} e^{-[\delta_{u-1}-\delta_u]r_u} dr_u}{\left[\prod_{\substack{w=k \\ w\neq q}}^{u-1} (\delta_w - \delta_q) \right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q - \rho) \right]} \\
&- \sum_{u=k}^{n} \left[\prod_{r=0}^{u-2} \gamma_r \right] \frac{\rho}{\rho+\lambda_u} \sum_{q=0}^{k-1} \frac{\int_0^t e^{-[\delta_q-\delta_{u-1}-\rho]r_u} \gamma_{u-1} e^{-[\delta_{u-1}+\rho]r_u} dr_u}{\left[\prod_{\substack{w=k \\ w\neq q}}^{u-1} (\delta_w - \delta_q + \rho) \right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q) \right]} \\
&- \sum_{u=k}^{n} \left[\prod_{r=0}^{u-2} \gamma_r \right] \frac{\rho}{\rho+\lambda_u} \sum_{q=k}^{u-1} \frac{\int_0^t e^{-[\delta_q-\delta_{u-1}]r_u} \gamma_{u-1} e^{-[\delta_{u-1}+\rho]r_u} dr_u}{\left[\prod_{\substack{w=k \\ w\neq q}}^{u-1} (\delta_w - \delta_q) \right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q - \rho) \right]} \\
&+ \sum_{u=k}^{n-1} \left[\prod_{r=0}^{u-1} \gamma_r \right] \frac{\rho}{\rho+\lambda_u} \sum_{q=0}^{k-1} \frac{\int_0^t e^{-[\delta_q-\delta_u-\rho]r_{u+1}} \gamma_u e^{-[\delta_u+\rho]r_{u+1}} dr_{u+1}}{\left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q + \rho) \right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q) \right]} \\
&+ \sum_{u=k}^{n-1} \left[\prod_{r=0}^{u-1} \gamma_r \right] \frac{\rho}{\rho+\lambda_u} \sum_{q=k}^{u} \frac{\int_0^t e^{-[\delta_q-\delta_u]r_{u+1}} \gamma_u e^{-[\delta_u+\rho]r_{u+1}} dr_{u+1}}{\left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q) \right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q - \rho) \right]}.
\end{aligned}
$$

Rearranging the order of the summations and solving the integrals in all but the first line necessarily

implies

$$
\begin{aligned}
F_X^*(t) = 1 &- \sum_{w=0}^{k-1} e^{-\delta_w t} \int_0^t \cdots \int_{r_{w-1}}^t \gamma_{w-1} e^{-[\delta_{w-1}-\delta_w]r_w} dr_w \ldots \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1 \\
&- \sum_{u=k}^{n} \left[\prod_{r=0}^{u-2} \gamma_r\right] \frac{\lambda_u e^{-[\rho+\delta_u]t}}{\rho+\lambda_u} \sum_{q=0}^{k-1} \frac{\gamma_{u-1}\left[1-e^{-[\delta_q-\delta_u-\rho]t}\right]}{[\delta_q-\delta_u-\rho]\left[\prod_{\substack{w=k\\w\neq q}}^{u-1}(\delta_w-\delta_q+\rho)\right]\left[\prod_{\substack{v=0\\v\neq q}}^{k-1}(\delta_v-\delta_q)\right]} \\
&- \sum_{u=k}^{n} \left[\prod_{r=0}^{u-2} \gamma_r\right] \frac{\rho}{\rho+\lambda_u} \sum_{q=0}^{k-1} \frac{\gamma_{u-1}\left[1-e^{-\delta_q t}\right]}{\delta_q\left[\prod_{\substack{w=k\\w\neq q}}^{u-1}(\delta_w-\delta_q+\rho)\right]\left[\prod_{\substack{v=0\\v\neq q}}^{k-1}(\delta_v-\delta_q)\right]} \\
&+ \sum_{u=k}^{n-1} \left[\prod_{r=0}^{u-1} \gamma_r\right] \frac{\rho}{\rho+\lambda_u} \sum_{q=0}^{k-1} \frac{\gamma_{u}\left[1-e^{-\delta_q t}\right]}{\delta_q\left[\prod_{\substack{w=k\\w\neq q}}^{u}(\delta_w-\delta_q+\rho)\right]\left[\prod_{\substack{v=0\\v\neq q}}^{k-1}(\delta_v-\delta_q)\right]} \\
&- \sum_{u=k}^{n} \left[\prod_{r=0}^{u-2} \gamma_r\right] \frac{\lambda_u e^{-[\rho+\delta_u]t}}{\rho+\lambda_u} \sum_{q=k}^{u-1} \frac{\gamma_{u-1}\left[1-e^{-[\delta_q-\delta_u]t}\right]}{[\delta_q-\delta_u]\left[\prod_{\substack{w=k\\w\neq q}}^{u-1}(\delta_w-\delta_q)\right]\left[\prod_{\substack{v=0\\v\neq q}}^{k-1}(\delta_v-\delta_q-\rho)\right]} \\
&- \sum_{u=k}^{n} \left[\prod_{r=0}^{u-2} \gamma_r\right] \frac{\rho}{\rho+\lambda_u} \sum_{q=k}^{u-1} \frac{\gamma_{u-1}\left[1-e^{-[\delta_q+\rho]t}\right]}{[\delta_q+\rho]\left[\prod_{\substack{w=k\\w\neq q}}^{u-1}(\delta_w-\delta_q)\right]\left[\prod_{\substack{v=0\\v\neq q}}^{k-1}(\delta_v-\delta_q-\rho)\right]} \\
&+ \sum_{u=k}^{n-1} \left[\prod_{r=0}^{u-1} \gamma_r\right] \frac{\rho}{\rho+\lambda_u} \sum_{q=k}^{u} \frac{\gamma_{u}\left[1-e^{-[\delta_q+\rho]t}\right]}{[\delta_q+\rho]\left[\prod_{\substack{w=k\\w\neq q}}^{u}(\delta_w-\delta_q)\right]\left[\prod_{\substack{v=0\\v\neq q}}^{k-1}(\delta_v-\delta_q-\rho)\right]}.
\end{aligned}
$$

Note that because $\gamma_n = 0$ it is possible to include the n^{th} terms in the fourth and seventh summations because these terms would be zero. Bringing the remaining exponential factors (and constant factors) inside the inner summations, gathering theγ_{u-1}'s into the product of γ_r's to the left and absorbing the product terms in the denominators of the second and fifth summations necessarily

implies

$$
\begin{aligned}
F_X^*(t) = 1 &- \sum_{w=0}^{k-1} e^{-\delta_w t} \int_0^t \cdots \int_{r_{w-1}}^t \gamma_{w-1} e^{-[\delta_{w-1}-\delta_w]r_w} dr_w \ldots \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1 \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r\right] \frac{1}{\rho+\lambda_u} \sum_{q=0}^{k-1} \frac{\lambda_u \left[e^{-[\rho+\delta_u]t} - e^{-\delta_q t}\right]}{\left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q + \rho)\right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q)\right]} \\
&- \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r\right] \frac{1}{\rho+\lambda_u} \sum_{q=0}^{k-1} \frac{\rho \left[1 - e^{-\delta_q t}\right]}{\delta_q \left[\prod_{\substack{w=k \\ w\neq q}}^{u-1} (\delta_w - \delta_q + \rho)\right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q)\right]} \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r\right] \frac{1}{\rho+\lambda_u} \sum_{q=0}^{k-1} \frac{\rho\gamma_u \left[1 - e^{-\delta_q t}\right]}{\delta_q \left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q + \rho)\right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q)\right]} \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r\right] \frac{1}{\rho+\lambda_u} \sum_{q=k}^{u-1} \frac{\lambda_u \left[e^{-[\rho+\delta_u]t} - e^{-[\delta_q+\rho]t}\right]}{\left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q)\right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q - \rho)\right]} \\
&- \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r\right] \frac{1}{\rho+\lambda_u} \sum_{q=k}^{u-1} \frac{\rho \left[1 - e^{-[\delta_q+\rho]t}\right]}{[\delta_q+\rho] \left[\prod_{\substack{w=k \\ w\neq q}}^{u-1} (\delta_w - \delta_q)\right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q - \rho)\right]} \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r\right] \frac{1}{\rho+\lambda_u} \sum_{q=k}^{u} \frac{\rho\gamma_u \left[1 - e^{-[\delta_q+\rho]t}\right]}{[\delta_q+\rho] \left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q)\right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q - \rho)\right]}.
\end{aligned}
$$

Multiplying the second, third, fifth and sixth summations top and bottom by δ_q, $(\delta_u - \delta_q + \rho)$,

$(\delta_q + \rho)$ and $(\delta_q - \delta_u)$ respectively, this necessarily implies

$$
\begin{aligned}
F_X^*(t) = 1 &- \sum_{w=0}^{k-1} e^{-\delta_w t} \int_0^t \cdots \int_{r_{w-1}}^t \gamma_{w-1} e^{-[\delta_{w-1}-\delta_w]r_w} dr_w \ldots \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1 \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r\right] \frac{1}{\rho+\lambda_u} \sum_{q=0}^{k-1} \frac{\lambda_u \delta_q \left[e^{-[\rho+\delta_u]t} - e^{-\delta_q t}\right]}{\delta_q \left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q + \rho)\right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q)\right]} \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r\right] \frac{1}{\rho+\lambda_u} \sum_{q=0}^{k-1} \frac{\rho[\delta_q - \delta_u - \rho]\left[1 - e^{-\delta_q t}\right]}{\delta_q \left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q + \rho)\right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q)\right]} \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r\right] \frac{1}{\rho+\lambda_u} \sum_{q=0}^{k-1} \frac{\rho\gamma_u \left[1 - e^{-\delta_q t}\right]}{\delta_q \left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q + \rho)\right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q)\right]} \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r\right] \frac{1}{\rho+\lambda_u} \sum_{q=k}^{u-1} \frac{\lambda_u [\delta_q + \rho]\left[e^{-[\rho+\delta_u]t} - e^{-[\delta_q+\rho]t}\right]}{[\delta_q + \rho]\left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q)\right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q - \rho)\right]} \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r\right] \frac{1}{\rho+\lambda_u} \sum_{q=k}^{u-1} \frac{\rho[\delta_q - \delta_u]\left[1 - e^{-[\delta_q+\rho]t}\right]}{[\delta_q + \rho]\left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q)\right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q - \rho)\right]} \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r\right] \frac{1}{\rho+\lambda_u} \sum_{q=k}^{u} \frac{\rho\gamma_u \left[1 - e^{-[\delta_q+\rho]t}\right]}{[\delta_q + \rho]\left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q)\right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q - \rho)\right]}.
\end{aligned}
$$

Both the third and fourth summations and the sixth and seventh summations can be merged. The different exponential factors in the second and fifth summations can also be separated out and this

necessarily implies

$$
\begin{aligned}
F_X^*(t) = 1 &- \sum_{w=0}^{k-1} e^{-\delta_w t} \int_0^t \cdots \int_{r_{w-1}}^t \gamma_{w-1} e^{-[\delta_{w-1}-\delta_w]r_w} dr_w \ldots \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1 \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r \right] \frac{1}{\rho+\lambda_u} \sum_{q=0}^{k-1} \frac{\lambda_u \delta_q e^{-[\rho+\delta_u]t}}{\delta_q \left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q + \rho) \right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q) \right]} \\
&- \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r \right] \frac{1}{\rho+\lambda_u} \sum_{q=0}^{k-1} \frac{\lambda_u \delta_q e^{-\delta_q t}}{\delta_q \left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q + \rho) \right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q) \right]} \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r \right] \frac{1}{\rho+\lambda_u} \sum_{q=0}^{k-1} \frac{\rho[\delta_q - \lambda_u - \rho]\left[1 - e^{-\delta_q t}\right]}{\delta_q \left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q + \rho) \right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q) \right]} \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r \right] \frac{1}{\rho+\lambda_u} \sum_{q=k}^{u-1} \frac{\lambda_u [\delta_q + \rho] e^{-[\rho+\delta_u]t}}{[\delta_q + \rho] \left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q) \right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q - \rho) \right]} \\
&- \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r \right] \frac{1}{\rho+\lambda_u} \sum_{q=k}^{u-1} \frac{\lambda_u [\delta_q + \rho] e^{-[\delta_q+\rho]t}}{[\delta_q + \rho] \left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q) \right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q - \rho) \right]} \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r \right] \frac{1}{\rho+\lambda_u} \sum_{q=k}^{u-1} \frac{\rho[\delta_q - \lambda_u]\left[1 - e^{-[\delta_q+\rho]t}\right]}{[\delta_q + \rho] \left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q) \right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q - \rho) \right]}.
\end{aligned}
$$

The $\lambda_u e^{-[\rho+\delta_u]t}$ factors can be taken outside the inner summations in the second and fifth summations. Also, both the third and fourth summations and the sixth and seventh summations can be

merged. This necessarily implies

$$
\begin{aligned}
F_X^*(t) = 1 &- \sum_{w=0}^{k-1} e^{-\delta_w t} \int_0^t \dots \int_{r_{w-1}}^t \gamma_{w-1} e^{-[\delta_{w-1}-\delta_w]r_w} dr_w \dots \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1 \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r\right] \frac{\lambda_u e^{-[\rho+\delta_u]t}}{\rho+\lambda_u} \sum_{q=0}^{k-1} \frac{1}{\left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q + \rho)\right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q)\right]} \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r\right] \frac{1}{\rho+\lambda_u} \sum_{q=0}^{k-1} \frac{\rho[\delta_q - \lambda_u - \rho]\left[1 - e^{-\delta_q t}\right] - \lambda_u \delta_q e^{-\delta_q t}}{\delta_q \left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q + \rho)\right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q)\right]} \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r\right] \frac{\lambda_u e^{-[\rho+\delta_u]t}}{\rho+\lambda_u} \sum_{q=k}^{u-1} \frac{1}{\left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q)\right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q - \rho)\right]} \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r\right] \frac{1}{\rho+\lambda_u} \sum_{q=k}^{u-1} \frac{\rho[\delta_q - \lambda_u]\left[1 - e^{-[\delta_q+\rho]t}\right] - \lambda_u[\delta_q+\rho]e^{-[\delta_q+\rho]t}}{[\delta_q+\rho]\left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q)\right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q - \rho)\right]}.
\end{aligned}
$$

Applying Equation A.2.3 to the second and fourth summations (this equation can be found on Page 172 and was established during the proof of Lemma 3), this necessarily implies

$$
\begin{aligned}
F_X^*(t) = 1 &- \sum_{w=0}^{k-1} e^{-\delta_w t} \int_0^t \dots \int_{r_{w-1}}^t \gamma_{w-1} e^{-[\delta_{w-1}-\delta_w]r_w} dr_w \dots \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1 \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r\right] \frac{1}{\rho+\lambda_u} \sum_{q=0}^{k-1} \frac{\rho[\delta_q - \lambda_u - \rho]\left[1 - e^{-\delta_q t}\right] - \lambda_u \delta_q e^{-\delta_q t}}{\delta_q \left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q + \rho)\right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q)\right]} \\
&- \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r\right] \frac{\lambda_u e^{-[\rho+\delta_u]t}}{\rho+\lambda_u} \frac{1}{\left[\prod_{\substack{w=k \\ w\neq u}}^{u} (\delta_w - \delta_u)\right] \left[\prod_{\substack{v=0 \\ v\neq u}}^{k-1} (\delta_v - \delta_u - \rho)\right]} \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r\right] \frac{1}{\rho+\lambda_u} \sum_{q=k}^{u-1} \frac{\rho[\delta_q - \lambda_u]\left[1 - e^{-[\delta_q+\rho]t}\right] - \lambda_u[\delta_q+\rho]e^{-[\delta_q+\rho]t}}{[\delta_q+\rho]\left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q)\right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q - \rho)\right]}.
\end{aligned}
$$

Note that third and fourth summation can be merged by treating the third summation as the term

$q = u$ in the fourth inner summation. This necessarily implies

$$\begin{aligned}
F_X^*(t) = 1 &- \sum_{w=0}^{k-1} e^{-\delta_w t} \int_0^t \cdots \int_{r_{w-1}}^t \gamma_{w-1} e^{-[\delta_{w-1}-\delta_w]r_w} dr_w \ldots \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1 \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r\right] \frac{1}{\rho+\lambda_u} \sum_{q=0}^{k-1} \frac{\rho[\delta_q - \lambda_u - \rho]\left[1-e^{-\delta_q t}\right] - \lambda_u \delta_q e^{-\delta_q t}}{\delta_q \left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q + \rho)\right]\left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q)\right]} \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r\right] \frac{1}{\rho+\lambda_u} \sum_{q=k}^{u} \frac{\rho[\delta_q - \lambda_u]\left[1-e^{-[\delta_q+\rho]t}\right] - \lambda_u[\delta_q+\rho]e^{-[\delta_q+\rho]t}}{[\delta_q+\rho]\left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q)\right]\left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q - \rho)\right]}.
\end{aligned}$$

Lastly, An application of Lemma 2 to the first summation necessarily implies

$$\begin{aligned}
F_X^*(t) = 1 &- \sum_{w=0}^{k-1} e^{-\delta_w t} \left[\prod_{r=0}^{w-1} \gamma_r\right] \sum_{j=0}^{w} \frac{e^{-[\delta_j-\delta_w]t}}{\prod_{\substack{u=0 \\ u\neq j}}^{w} (\delta_u - \delta_j)} \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r\right] \frac{1}{\rho+\lambda_u} \sum_{q=0}^{k-1} \frac{\rho[\delta_q - \lambda_u - \rho]\left[1-e^{-\delta_q t}\right] - \lambda_u \delta_q e^{-\delta_q t}}{\delta_q \left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q + \rho)\right]\left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q)\right]} \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r\right] \frac{1}{\rho+\lambda_u} \sum_{q=k}^{u} \frac{\rho[\delta_q - \lambda_u]\left[1-e^{-[\delta_q+\rho]t}\right] - \lambda_u[\delta_q+\rho]e^{-[\delta_q+\rho]t}}{[\delta_q+\rho]\left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q)\right]\left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q - \rho)\right]}.
\end{aligned}$$

Tidying up the exponential terms in the first summation then leads to

$$\begin{aligned}
F_X^*(t) = 1 &- \sum_{w=0}^{k-1} \left[\prod_{r=0}^{w-1} \gamma_r\right] \sum_{j=0}^{w} \frac{e^{-\delta_j t}}{\prod_{\substack{u=0 \\ u\neq j}}^{w} (\delta_u - \delta_j)} \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r\right] \frac{1}{\rho+\lambda_u} \sum_{q=0}^{k-1} \frac{\rho[\delta_q - \lambda_u - \rho]\left[1-e^{-\delta_q t}\right] - \lambda_u \delta_q e^{-\delta_q t}}{\delta_q \left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q + \rho)\right]\left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q)\right]} \\
&+ \sum_{u=k}^{n} \left[\prod_{r=0}^{u-1} \gamma_r\right] \frac{1}{\rho+\lambda_u} \sum_{q=k}^{u} \frac{\rho[\delta_q - \lambda_u]\left[1-e^{-[\delta_q+\rho]t}\right] - \lambda_u[\delta_q+\rho]e^{-[\delta_q+\rho]t}}{[\delta_q+\rho]\left[\prod_{\substack{w=k \\ w\neq q}}^{u} (\delta_w - \delta_q)\right]\left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q - \rho)\right]},
\end{aligned}$$

as required.

A.2 Some Lemmas Used in the Calculations for the Random Signals Model

Lemma 1. *The following identity holds for all i, a non-negative integer.*

$$\int_0^t \int_{r_1}^t \cdots \int_{r_{i-1}}^t \gamma_{i-1} e^{-\gamma_{i-1} r_i} dr_i \gamma_{i-2} e^{-[\gamma_{i-2}-\gamma_{i-1}]r_{i-1}} dr_{i-1} \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1$$
$$= 1 - \sum_{j=0}^{i-1} e^{-\gamma_j t} \int_0^t \int_{r_1}^t \cdots \int_{r_{j-1}}^t \gamma_{j-1} e^{-[\gamma_{j-1}-\gamma_j]r_j} dr_j \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1.$$

Proof. In the case where $i = 0$ this is obviously true. In the case where $i = 1$

$$\int_0^t \gamma_0 e^{-\gamma_0 r_1} dr_1 = \left(e^{-\gamma_0 0} - e^{-\gamma_0 t}\right) = 1 - e^{-\gamma_0 t},$$

as required. Proceeding now by induction (the base case being $i = 2$), consider the left-hand side

$$\int_0^t \int_{r_1}^t \gamma_1 e^{-\gamma_1 r_2} dr_2 \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1$$
$$= \int_0^t \left(e^{-\gamma_1 r_1} - e^{-\gamma_1 t}\right) \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1$$
$$= \int_0^t \gamma_0 e^{-\gamma_0 r_1} dr_1 - e^{-\gamma_1 t} \int_0^t \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1$$
$$= 1 - e^{-\gamma_0 t} - e^{-\gamma_1 t} \int_0^t \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1,$$

as required. Moving to the inductive step, for all $i \geq 2$ (assuming the truth of the case for i), consider the left-hand side of the identity for $i + 1$. Solving the innermost integral (the integral with respect to r_{i+1} from r_i to t) necessarily implies

$$\int_0^t \int_{r_1}^t \cdots \int_{r_i}^t \gamma_i e^{-\gamma_i r_{i+1}} dr_{i+1} \gamma_{i-1} e^{-[\gamma_{i-1}-\gamma_i]r_i} dr_i \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1$$
$$= \int_0^t \int_{r_1}^t \cdots \int_{r_{i-1}}^t \left(e^{-\gamma_i r_i} - e^{-\gamma_i t}\right) \gamma_{i-1} e^{-[\gamma_{i-1}-\gamma_i]r_i} dr_i \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1$$
$$= \int_0^t \int_{r_1}^t \cdots \int_{r_{i-1}}^t \gamma_{i-1} e^{-\gamma_{i-1} r_i} dr_i \gamma_{i-2} e^{-[\gamma_{i-2}-\gamma_{i-1}]r_{i-1}} dr_{i-1} \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1$$
$$- e^{-\gamma_i t} \int_0^t \int_{r_1}^t \cdots \int_{r_{i-1}}^t \gamma_{i-1} e^{-[\gamma_{i-1}-\gamma_i]r_i} dr_i \ldots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1.$$

Applying the inductive hypothesis (the truth of the identity for i) necessarily implies

$$\begin{aligned}
&\int_0^t \int_{r_1}^t \dots \int_{r_{i-1}}^t \gamma_{i-1} e^{-\gamma_{i-1} r_i} dr_i \gamma_{i-2} e^{-[\gamma_{i-2}-\gamma_{i-1}]r_{i-1}} dr_{i-1} \dots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&\quad - e^{-\gamma_i t} \int_0^t \int_{r_1}^t \dots \int_{r_{i-1}}^t \gamma_{i-1} e^{-[\gamma_{i-1}-\gamma_i]r_i} dr_i \dots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&= 1 - \sum_{j=0}^{i-1} e^{-\gamma_j t} \int_0^t \int_{r_1}^t \dots \int_{r_{j-1}}^t \gamma_{j-1} e^{-[\gamma_{j-1}-\gamma_j]r_j} dr_j \dots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&\quad - e^{-\gamma_i t} \int_0^t \int_{r_1}^t \dots \int_{r_{i-1}}^t \gamma_{i-1} e^{-[\gamma_{i-1}-\gamma_i]r_i} dr_i \dots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\
&= 1 - \sum_{j=0}^{i} e^{-\gamma_j t} \int_0^t \int_{r_1}^t \dots \int_{r_{j-1}}^t \gamma_{j-1} e^{-[\gamma_{j-1}-\gamma_j]r_j} dr_j \dots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1,
\end{aligned}$$

which is equal to the right-hand side of the identity, as required. □

Lemma 2. *The following identity holds for all i, a non-negative integer.*

$$\begin{aligned}
&\int_0^t \int_{r_1}^t \dots \int_{r_{i-1}}^t \gamma_{i-1} e^{-[\delta_{i-1}-\delta_i]r_i} dr_i \dots \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1 \\
&= \left[\prod_{r=0}^{i-1} \gamma_r\right] \sum_{j=0}^{i} \frac{e^{-[\delta_j-\delta_i]t}}{\prod_{\substack{u=0 \\ u\neq j}}^{i} (\delta_u - \delta_j)}.
\end{aligned}$$

Proof. In the case where $i = 0$ this is obviously true. Proceeding by induction with the base case being $i = 1$, integration yields

$$\int_0^t \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1 = \frac{\gamma_0 \left[1 - e^{-[\delta_0-\delta_1]t}\right]}{\delta_0 - \delta_1},$$

and the formula of the lemma yields

$$\gamma_0 \left[\frac{e^{-[\delta_0-\delta_1]t}}{\delta_1 - \delta_0} + \frac{1}{\delta_0 - \delta_1}\right],$$

as required. Moving to the inductive step, for all $i \geq 1$, assuming the truth of the case for i and by using a change of integration

$$\begin{aligned}
&\int_0^t \int_{r_1}^t \dots \int_{r_i}^t \gamma_i e^{-[\delta_i-\delta_{i+1}]r_{i+1}} dr_{i+1} \dots \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1 \\
&= \int_0^t \int_0^{r_{i+1}} \dots \int_0^{r_2} \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1 \dots \gamma_i e^{-[\delta_i-\delta_{i+1}]r_{i+1}} dr_{i+1}.
\end{aligned}$$

Applying the inductive hypothesis (the truth of the identity for i) and then proceeding to solve the remaining integral necessarily implies

$$\begin{aligned}
&\int_0^t \int_0^{r_{i+1}} \dots \int_0^{r_2} \gamma_0 e^{-[\delta_0-\delta_1]r_1} dr_1 \dots \gamma_i e^{-[\delta_i-\delta_{i+1}]r_{i+1}} dr_{i+1} \\
&= \left[\prod_{r=0}^{i} \gamma_r\right] \int_0^t \sum_{j=0}^{i} \frac{e^{-[\delta_j-\delta_i]r_{i+1}}}{\prod_{\substack{u=0 \\ u\neq j}}^{i}(\delta_u-\delta_j)} e^{-[\delta_i-\delta_{i+1}]r_{i+1}} dr_{i+1} \\
&= \left[\prod_{r=0}^{i} \gamma_r\right] \sum_{j=0}^{i} \frac{\int_0^t e^{-[\delta_j-\delta_{i+1}]r_{i+1}} dr_{i+1}}{\prod_{\substack{u=0 \\ u\neq j}}^{i}(\delta_u-\delta_j)} \\
&= \left[\prod_{r=0}^{i} \gamma_r\right] \sum_{j=0}^{i} \frac{e^{-[\delta_j-\delta_{i+1}]t}-1}{(\delta_{i+1}-\delta_j)\prod_{\substack{u=0 \\ u\neq j}}^{i}(\delta_u-\delta_j)} \\
&= \left[\prod_{r=0}^{i} \gamma_r\right] \sum_{j=0}^{i} \frac{e^{-[\delta_j-\delta_{i+1}]t}}{\prod_{\substack{u=0 \\ u\neq j}}^{i+1}(\delta_u-\delta_j)} - \left[\prod_{r=0}^{i} \gamma_r\right] \sum_{j=0}^{i} \frac{1}{\prod_{\substack{u=0 \\ u\neq j}}^{i+1}(\delta_u-\delta_j)}.
\end{aligned}$$

All that remains is to equate the second sum in the final expression above with the $i+1^{th}$ term in the sum of the right-hand side of the identity to complete the proof, in other words the following must be shown to be true

$$-\sum_{j=0}^{i} \frac{1}{\prod_{\substack{u=0 \\ u\neq j}}^{i+1}(\delta_u-\delta_j)} = \frac{1}{\prod_{u=0}^{i}(\delta_u-\delta_{i+1})}.$$

The proof now follows exactly the same line of argument as Balazs' proof[6] of the formula for the sum of independent exponential random variables (included here for completeness). It remains to show that

$$\sum_{j=0}^{i+1} \frac{1}{\prod_{\substack{u=0 \\ u\neq j}}^{i+1}(\delta_u-\delta_j)} = 0.$$

By multiplying both top and bottom, the left hand side can be rewritten as

$$\sum_{j=0}^{i+1} \frac{1}{\prod_{\substack{u=0 \\ u\neq j}}^{i+1}(\delta_u-\delta_j)} = \sum_{j=0}^{i+1} \frac{\prod_{\substack{l=0 \\ l\neq j}}^{i+1} \prod_{\substack{u=0 \\ u\neq l}}^{i+1}(\delta_u-\delta_l)}{\prod_{l=0}^{i+1} \prod_{\substack{u=0 \\ u\neq l}}^{i+1}(\delta_u-\delta_l)}.$$

The denominator can be taken outside of the sum and is clearly non-zero, so the above expression is zero if and only if

$$\sum_{j=0}^{i+1} \prod_{\substack{l=0 \\ l\neq j}}^{i+1} \prod_{\substack{u=0 \\ u\neq l}}^{i+1}(\delta_u-\delta_l) = 0. \tag{A.2.1}$$

What follows is a series of transformations which involves changing the ordering of the differences between the δ's (leading to changes of sign in the factors of the products). Firstly, for each term in the summation on the left-hand side below, those product terms for which $u = j$ are separated out into two products, those for which $l < j$ and those for which $l > j$.

$$\sum_{j=0}^{i+1}\prod_{\substack{l=0\\l\neq j}}^{i+1}\prod_{\substack{u=0\\u\neq l}}^{i+1}(\delta_u-\delta_l)=\sum_{j=0}^{i+1}\left[\prod_{\substack{l=0\\l\neq j}}^{i+1}\prod_{\substack{u=0\\u\neq l\\u\neq j}}^{i+1}(\delta_u-\delta_l)\right]\left[\prod_{\substack{l=0\\l<j}}^{i+1}(\delta_j-\delta_l)\right]\left[\prod_{\substack{l=0\\l>j}}^{i+1}(\delta_j-\delta_l)\right].$$

The double product can be split into two double products in the following way

$$\sum_{j=0}^{i+1}\prod_{\substack{l=0\\l\neq j}}^{i+1}\prod_{\substack{u=0\\u\neq l}}^{i+1}(\delta_u-\delta_l)=\sum_{j=0}^{i+1}\left[\prod_{\substack{l=0\\l\neq j}}^{i+1}\prod_{\substack{u=0\\u>l\\u\neq j}}^{i+1}(\delta_u-\delta_l)\right]\left[\prod_{\substack{l=0\\l\neq j}}^{i+1}\prod_{\substack{u=0\\u<l\\u\neq j}}^{i+1}(\delta_u-\delta_l)\right]\left[\prod_{\substack{l=0\\l<j}}^{i+1}(\delta_j-\delta_l)\right]\left[\prod_{\substack{l=0\\l>j}}^{i+1}(\delta_j-\delta_l)\right].$$

By multiplying each term in the second double product and the right-most product by minus one this necessarily implies

$$\sum_{j=0}^{i+1}\prod_{\substack{l=0\\l\neq j}}^{i+1}\prod_{\substack{u=0\\u\neq l}}^{i+1}(\delta_u-\delta_l)=\sum_{j=0}^{i+1}\left[\prod_{\substack{l=0\\l\neq j}}^{i+1}\prod_{\substack{u=0\\u>l\\u\neq j}}^{i+1}(\delta_u-\delta_l)\right]^2(-1)^{i(i+1)/2}\left[\prod_{\substack{l=0\\l<j}}^{i+1}(\delta_j-\delta_l)\right]\left[\prod_{\substack{l=0\\l>j}}^{i+1}(\delta_l-\delta_j)\right](-1)^{n-j}.$$

By factorising those products of minus signs that do not depend on j outside the summation, it is immaterial whether they result in a positive or negative sign. By changing the dummy index in the right-most product from l to u this necessarily implies

$$\sum_{j=0}^{i+1}\prod_{\substack{l=0\\l\neq j}}^{i+1}\prod_{\substack{u=0\\u\neq l}}^{i+1}(\delta_u-\delta_l)=\pm\sum_{j=0}^{i+1}\left[\prod_{\substack{l=0\\l\neq j}}^{i+1}\prod_{\substack{u=0\\u>l\\u\neq j}}^{i+1}(\delta_u-\delta_l)\right]^2\left[\prod_{\substack{l=0\\l<j}}^{i+1}(\delta_j-\delta_l)\right]\left[\prod_{\substack{u=0\\u>j}}^{i+1}(\delta_u-\delta_j)\right](-1)^{-j}.$$

By absorbing the right-most product into one of the double products (the right-most product being

equal to the case of $l = j$ in the double product), this necessarily implies

$$\sum_{j=0}^{i+1}\prod_{\substack{l=0\\l\neq j}}^{i+1}\prod_{\substack{u=0\\u\neq l}}^{i+1}(\delta_u-\delta_l)=\pm\sum_{j=0}^{i+1}\left[\prod_{\substack{l=0\\l\neq j}}^{i+1}\prod_{\substack{u=0\\u>l\\u\neq j}}^{i+1}(\delta_u-\delta_l)\right]\left[\prod_{l=0}^{i+1}\prod_{\substack{u=0\\u>l\\u\neq j}}^{i+1}(\delta_u-\delta_l)\right]\left[\prod_{\substack{l=0\\l<j}}^{i+1}(\delta_j-\delta_l)\right](-1)^{-j}.$$

Lastly, by absorbing the right-most product into the second double product (the right-most product being equal to the case of $u = j$ in the second double product), the resulting product can be factorised outside the summation which necessarily implies

$$\sum_{j=0}^{i+1}\prod_{\substack{l=0\\l\neq j}}^{i+1}\prod_{\substack{u=0\\u\neq l}}^{i+1}(\delta_u-\delta_l)=\pm\left[\prod_{l=0}^{i+1}\prod_{\substack{u=0\\u>l}}^{i+1}(\delta_u-\delta_l)\right]\sum_{j=0}^{i+1}\left[\prod_{\substack{l=0\\l\neq j}}^{i+1}\prod_{\substack{u=0\\u>l\\u\neq j}}^{i+1}(\delta_u-\delta_l)\right](-1)^{-j}.$$

This is zero if and only if

$$\sum_{j=0}^{i+1}\prod_{\substack{l=0\\l\neq j}}^{i+1}\prod_{\substack{u=0\\u>l\\u\neq j}}^{i+1}(\delta_u-\delta_l)\,(-1)^{-j}=0. \tag{A.2.2}$$

As pointed out by Balaz[6], for each j the double product in equation A.2.2 is the expression for an $(i+1)\times(i+1)$ Vandermonde determinant of the form

$$\begin{vmatrix}
1 & \delta_0 & \delta_0^2 & \dots & \delta_0^i\\
1 & \delta_1 & \delta_1^2 & \dots & \delta_1^i\\
\vdots & \vdots & \vdots & \ddots & \vdots\\
1 & \delta_{j-1} & \delta_{j-1}^2 & \dots & \delta_{j-1}^i\\
1 & \delta_{j+1} & \delta_{j+1}^2 & \dots & \delta_{j+1}^i\\
\vdots & \vdots & \vdots & \ddots & \vdots\\
1 & \delta_{i+1} & \delta_{i+1}^2 & \dots & \delta_{i+1}^i
\end{vmatrix}$$

and this means that equation A.2.2 is just the expansion of the following $(i+2)\times(i+2)$ determinant

$$\begin{vmatrix}
1 & 1 & \delta_0 & \delta_0^2 & \dots & \delta_0^{i+1}\\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots\\
1 & 1 & \delta_{i+1} & \delta_{i+1}^2 & \dots & \delta_{i+1}^{i+1}
\end{vmatrix}$$

with respect to it's second column using Liebnitz's well-known equation. It is an elementary fact of linear algebra that the determinant of a square matrix with two identical columns is zero which necessarily implies that equation A.2.2 is zero as required. □

Lemma 3. *The following identity holds for any non-negative real number t and for any non-negative integers n, u, k, such that $n > u \geq k \geq 1$.*

$$F(n,u,k,t) = F(u,u,k,t) - \sum_{i=u}^{n-1} e^{-\gamma_i t} G(i,u,k,t).$$

Proof. Recall Definition 6 on Page 151

$$F(n,u,k,t) = \int_0^t \int_{r_1}^t \dots \int_{r_{n-1}}^t \left[e^{\rho r_k} e^{-\{\sum_{i=1}^{u-1}[\lambda_{i-1}-\lambda_i]r_i + [\rho+\lambda_{u-1}]r_u\}} \right] \times \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}]r_{n-1}} dr_{n-1} \dots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1.$$

Given that $n > u$ this necessarily implies that the inner-most integral (the integral with respect to r_n) can be brought to the right of the bracketed group of exponential terms in the following manner

$$\begin{aligned} F(n,u,k,t) = \int_0^t \int_{r_1}^t \dots \int_{r_{n-2}}^t & \left[e^{\rho r_k} e^{-\{\sum_{i=1}^{u-1}[\lambda_{i-1}-\lambda_i]r_i + [\rho+\lambda_{u-1}]r_u\}} \right] \\ & \times \int_{r_{n-1}}^t \gamma_{n-1} e^{-\gamma_{n-1} r_n} dr_n \\ & \times \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}]r_{n-1}} dr_{n-1} \dots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1. \end{aligned}$$

Solving this inner-most integral necessarily implies

$$\begin{aligned} F(n,u,k,t) = \int_0^t \int_{r_1}^t \dots \int_{r_{n-2}}^t & \left[e^{\rho r_k} e^{-\{\sum_{i=1}^{u-1}[\lambda_{i-1}-\lambda_i]r_i + [\rho+\lambda_{u-1}]r_u\}} \right] \\ & \times \left[e^{-\gamma_{n-1} r_{n-1}} - e^{-\gamma_{n-1} t} \right] \\ & \times \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}]r_{n-1}} dr_{n-1} \dots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1. \end{aligned}$$

Multiplying out the brackets, cancelling exponentials in the first group of integrals and moving

$e^{-\gamma_{n-1}t}$ outside the integrals in the second group necessarily implies

$$\begin{aligned}F(n,u,k,t) = &\int_0^t \int_{r_1}^t \dots \int_{r_{n-2}}^t \left[e^{\rho r_k} e^{-\{\sum_{i=1}^{u-1}[\lambda_{i-1}-\lambda_i]r_i+[\rho+\lambda_{u-1}]r_u\}}\right] \\ &\times \gamma_{n-2} e^{-\gamma_{n-2}r_{n-1}} dr_{n-1} \gamma_{n-3} e^{-[\gamma_{n-3}-\gamma_{n-2}]r_{n-2}} dr_{n-2} \dots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1 \\ &- e^{-\gamma_{n-1}t} \int_0^t \int_{r_1}^t \dots \int_{r_{n-2}}^t \left[e^{\rho r_k} e^{-\{\sum_{i=1}^{u-1}[\lambda_{i-1}-\lambda_i]r_i+[\rho+\lambda_{u-1}]r_u\}}\right] \\ &\times \gamma_{n-2} e^{-[\gamma_{n-2}-\gamma_{n-1}]r_{n-1}} dr_{n-1} \dots \gamma_0 e^{-[\gamma_0-\gamma_1]r_1} dr_1.\end{aligned}$$

Recalling the definition of $G(w,u,k,t)$ (Definition 5 on Page 147), it can be seen that

$$F(n,u,k,t) = F(n-1,u,k) - e^{-\gamma_{n-1}t} G(n-1,u,k,t).$$

Applying the same argument recursively to $F(n-1,u,k)$ necessarily implies

$$F(n,u,k,t) = F(u,u,k) - \sum_{i=u}^{n-1} e^{-\gamma_i t} G(i,u,k,t).$$

□

Lemma 4. *The following identity holds for any non-negative integers u and k, such that $u \geq k \geq 1$.*

$$\begin{aligned}H(u,k) = &\left[\prod_{r=0}^{u-2} \gamma_r\right] \sum_{q=0}^{k-1} \frac{e^{-[\delta_q - \delta_{u-1} - \rho]r_u}}{\left[\prod_{\substack{w=k \\ w\neq q}}^{u-1} (\delta_w - \delta_q + \rho)\right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q)\right]} \\ &+ \left[\prod_{r=0}^{u-2} \gamma_r\right] \sum_{q=k}^{u-1} \frac{e^{-[\delta_q - \delta_{u-1}]r_u}}{\left[\prod_{\substack{w=k \\ w\neq q}}^{u-1} (\delta_w - \delta_q)\right] \left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1} (\delta_v - \delta_q - \rho)\right]}.\end{aligned}$$

Proof. For the base case $u = k$, by Definition 7 on Page 155 this necessarily implies

$$H(k,k) = \int_0^{r_k} \dots \int_0^{r_2} e^{\rho r_k} \gamma_0 e^{-[\delta_0 - \delta_1]r_1} dr_1 \dots \gamma_{k-2} e^{-[\delta_{k-2} - \delta_{k-1}]r_{k-1}} dr_{k-1}.$$

Reversing the order of integration necessarily implies

$$H(k,k) = e^{\rho r_k} \int_0^{r_k} \int_{r_1}^{r_k} \dots \int_{r_{k-2}}^{r_k} \gamma_{k-2} e^{-[\delta_{k-2} - \delta_{k-1}]r_{k-1}} dr_{k-1} \dots \gamma_0 e^{-[\delta_0 - \delta_1]r_1} dr_1.$$

Applying Lemma 2 necessarily implies

$$H(k,k) = e^{\rho r_k}\left[\prod_{r=0}^{k-2}\gamma_r\right]\sum_{j=0}^{k-1}\frac{e^{-[\delta_j-\delta_{k-1}]r_k}}{\prod_{\substack{u=0\\u\neq j}}^{k-1}(\delta_u-\delta_j)}$$
$$= \left[\prod_{r=0}^{k-2}\gamma_r\right]\sum_{j=0}^{k-1}\frac{e^{-[\delta_j-\delta_{k-1}-\rho]r_k}}{\prod_{\substack{u=0\\u\neq j}}^{k-1}(\delta_u-\delta_j)},$$

as required. Assuming the truth of the case $u = m$ (where $m \geq k$) necessarily implies

$$H(m+1,k) = \int_0^{r_{m+1}}\dots\int_0^{r_2} e^{\rho r_k}\gamma_0 e^{-[\delta_0-\delta_1]r_1}dr_1\dots\gamma_{m-1}e^{-[\delta_{m-1}-\delta_m]r_m}dr_m.$$

By applying the inductive hypothesis, this necessarily implies

$$H(m+1,k) = \int_0^{r_{m+1}} H(m,k)\gamma_{m-1}e^{-[\delta_{m-1}-\delta_m]r_m}dr_m$$
$$= \left[\prod_{r=0}^{m-2}\gamma_r\right]\sum_{q=0}^{k-1}\frac{\int_0^{r_{m+1}} e^{-[\delta_q-\delta_{m-1}-\rho]r_m}\gamma_{m-1}e^{-[\delta_{m-1}-\delta_m]r_m}dr_m}{\left[\prod_{\substack{w=k\\w\neq q}}^{m-1}(\delta_w-\delta_q+\rho)\right]\left[\prod_{\substack{v=0\\v\neq q}}^{k-1}(\delta_v-\delta_q)\right]}$$
$$+\left[\prod_{r=0}^{m-2}\gamma_r\right]\sum_{q=k}^{m-1}\frac{\int_0^{r_{m+1}} e^{-[\delta_q-\delta_{m-1}]r_m}\gamma_{m-1}e^{-[\delta_{m-1}-\delta_m]r_m}dr_m}{\left[\prod_{\substack{w=k\\w\neq q}}^{m-1}(\delta_w-\delta_q)\right]\left[\prod_{\substack{v=0\\v\neq q}}^{k-1}(\delta_v-\delta_q-\rho)\right]}.$$

Placing the γ_{m-1}'s outside the summations and cancelling exponential terms in the numerator necessarily implies

$$H(m+1,k) = \left[\prod_{r=0}^{m-1}\gamma_r\right]\sum_{q=0}^{k-1}\frac{\int_0^{r_{m+1}} e^{-[\delta_q-\delta_m-\rho]r_m}dr_m}{\left[\prod_{\substack{w=k\\w\neq q}}^{m-1}(\delta_w-\delta_q+\rho)\right]\left[\prod_{\substack{v=0\\v\neq q}}^{k-1}(\delta_v-\delta_q)\right]}$$
$$+\left[\prod_{r=0}^{m-1}\gamma_r\right]\sum_{q=k}^{m-1}\frac{\int_0^{r_{m+1}} e^{-[\delta_q-\delta_m]r_m}dr_m}{\left[\prod_{\substack{w=k\\w\neq q}}^{m-1}(\delta_w-\delta_q)\right]\left[\prod_{\substack{v=0\\v\neq q}}^{k-1}(\delta_v-\delta_q-\rho)\right]}.$$

solving the integrals necessarily implies

$$H(m+1,k) = \left[\prod_{r=0}^{m-1}\gamma_r\right]\sum_{q=0}^{k-1}\frac{1-e^{-[\delta_q-\delta_m-\rho]r_{m+1}}}{[\delta_q-\delta_m-\rho]\left[\prod_{\substack{w=k\\w\neq q}}^{m-1}(\delta_w-\delta_q+\rho)\right]\left[\prod_{\substack{v=0\\v\neq q}}^{k-1}(\delta_v-\delta_q)\right]}$$
$$+\left[\prod_{r=0}^{m-1}\gamma_r\right]\sum_{q=k}^{m-1}\frac{1-e^{-[\delta_q-\delta_m]r_{m+1}}}{[\delta_q-\delta_m]\left[\prod_{\substack{w=k\\w\neq q}}^{m-1}(\delta_w-\delta_q)\right]\left[\prod_{\substack{v=0\\v\neq q}}^{k-1}(\delta_v-\delta_q-\rho)\right]}.$$

Expanding the numerators and absorbing product factors in the denominators necessarily implies

$$
\begin{aligned}
&= \left[\prod_{r=0}^{m-1}\gamma_r\right]\sum_{q=0}^{k-1}\frac{e^{-[\delta_q-\delta_m-\rho]r_{m+1}}}{\left[\prod_{\substack{w=k\\w\neq q}}^{m}(\delta_w-\delta_q+\rho)\right]\left[\prod_{\substack{v=0\\v\neq q}}^{k-1}(\delta_v-\delta_q)\right]} \\
&\quad+\left[\prod_{r=0}^{m-1}\gamma_r\right]\sum_{q=k}^{m-1}\frac{e^{-[\delta_q-\delta_m]r_{m+1}}}{\left[\prod_{\substack{w=k\\w\neq q}}^{m}(\delta_w-\delta_q)\right]\left[\prod_{\substack{v=0\\v\neq q}}^{k-1}(\delta_v-\delta_q-\rho)\right]} \\
&\quad-\left[\prod_{r=0}^{m-1}\gamma_r\right]\sum_{q=0}^{k-1}\frac{1}{\left[\prod_{\substack{w=k\\w\neq q}}^{m}(\delta_w-\delta_q+\rho)\right]\left[\prod_{\substack{v=0\\v\neq q}}^{k-1}(\delta_v-\delta_q)\right]} \\
&\quad-\left[\prod_{r=0}^{m-1}\gamma_r\right]\sum_{q=k}^{m-1}\frac{1}{\left[\prod_{\substack{w=k\\w\neq q}}^{m}(\delta_w-\delta_q)\right]\left[\prod_{\substack{v=0\\v\neq q}}^{k-1}(\delta_v-\delta_q-\rho)\right]}.
\end{aligned}
$$

The proof is complete once the following identity is established

$$
\begin{aligned}
&\sum_{q=0}^{k-1}\frac{1}{\left[\prod_{\substack{w=k\\w\neq q}}^{m}(\delta_w-\delta_q+\rho)\right]\left[\prod_{\substack{v=0\\v\neq q}}^{k-1}(\delta_v-\delta_q)\right]}+\sum_{q=k}^{m-1}\frac{1}{\left[\prod_{\substack{w=k\\w\neq q}}^{m}(\delta_w-\delta_q)\right]\left[\prod_{\substack{v=0\\v\neq q}}^{k-1}(\delta_v-\delta_q-\rho)\right]} \\
&\quad=\frac{-1}{\left[\prod_{\substack{w=k\\w\neq m}}^{m}(\delta_w-\delta_m)\right]\left[\prod_{\substack{v=0\\v\neq m}}^{k-1}(\delta_v-\delta_m-\rho)\right]}. \qquad \text{(A.2.3)}
\end{aligned}
$$

This is equivalent to proving

$$
0=\sum_{q=0}^{k-1}\frac{1}{\left[\prod_{\substack{w=k\\w\neq q}}^{m}(\delta_w-\delta_q+\rho)\right]\left[\prod_{\substack{v=0\\v\neq q}}^{k-1}(\delta_v-\delta_q)\right]}+\sum_{q=k}^{m}\frac{1}{\left[\prod_{\substack{v=0\\v\neq q}}^{k-1}(\delta_v-\delta_q-\rho)\right]\left[\prod_{\substack{w=k\\w\neq q}}^{m}(\delta_w-\delta_q)\right]}.
$$

Consider the following sequence for i from 0 to m

$$
\begin{aligned}
\delta'_i &= \delta_i \qquad \text{for } i \text{ from } 0 \text{ to } k-1, \\
\delta'_i &= \delta_i+\rho \qquad \text{for } i \text{ from } k \text{ to } m.
\end{aligned}
$$

By replacing each incidence of δ_i by the corresponding expression involving δ'_i, this necessarily

implies

$$
\begin{aligned}
0 &= \sum_{q=0}^{k-1} \frac{1}{\left[\prod_{\substack{w=k \\ w\neq q}}^{m}(\delta'_w - \delta'_q)\right]\left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1}(\delta'_v - \delta'_q)\right]} + \sum_{q=k}^{m} \frac{1}{\left[\prod_{\substack{v=0 \\ v\neq q}}^{k-1}(\delta'_v - \delta'_q)\right]\left[\prod_{\substack{w=k \\ w\neq q}}^{m}(\delta'_w - \delta'_q)\right]} \\
&= \sum_{q=0}^{k-1} \frac{1}{\left[\prod_{\substack{v=0 \\ v\neq q}}^{m}(\delta'_v - \delta'_q)\right]} + \sum_{q=k}^{m} \frac{1}{\left[\prod_{\substack{w=0 \\ w\neq q}}^{m}(\delta'_w - \delta'_q)\right]} \\
&= \sum_{q=0}^{m} \frac{1}{\left[\prod_{\substack{v=0 \\ v\neq q}}^{m}(\delta'_v - \delta'_q)\right]}.
\end{aligned}
$$

From Equation A.2.1 on Page 166 (which was established during the proof of Lemma 2), it is evident that this sum is zero, as required. □

Appendix B

Simulation Details for the Kaplan-Meier Estimator With Missing Data

This chapter contains the MATLAB code used in the simulation study discussed in Section 6.4 of Chapter 6. The Functions are grouped under sub-headings for easy reference.
The code can be used by calling *"mySimulationRun"*, which first calls *"validateParams"* to ensure that the parameters conform to the conditions required for the analytical calculations to be true. Then the function *"createObservation"* is called repeatedly to generate one sample pair of X and Z each call. After having built a list of sample pairs, the analytical and empirical functions are called to calculate each version of the marginals of both X and Z, as well as each version of the sub-distribution for X. The value of the asymptote of the sub-distribution is also calculated using the function *"calcXLessZ"* and finally, the function *"calcKolmogorovSmirnovStats"*is called to calculate the value of the Kolmogorov-Smirnov statistic for each analytic and empirical function pair.

B.1 mySimulationRun

```
function [kolmStats, myXLessZ, myMatFuncs] = ...
    mySimulationRun(totalTime, timeSlice, lambdas, gammas, k, rho, observations)
```

```
%Simulation Input:
%   Integer of total time to consider
%   Integer of the total number of time Slices Per Unit of Time
%   Integer of number of signals (implicit)
%   Double array of Lamdas
%   Double array of gammas
%   Integer of kth signal at which next maintenance oportunity is taken
%   Double of rho
%   Integer of observations of the process to generate
%Simulation Output:
%   output as a 7 by (timeStep*totalTime) matrix.
%   column 1 being the values of time, the remaining columns being
%   the values of the eight functions in the order given below.
%       array of empirical estimate of the marginal distribution of X
%       array of empirical estimate of the marginal distribution of Z
%       array of empirical estimate of the sub-distribution of X
%       array of calculated value of the marginal distribution of X
%       array of calculated value of the marginal distribution of Z
%       array of calculated value of the sub-distribution of X


    %validate the input parameters
    validateParams(totalTime, timeSlice, lambdas, gammas, k, rho, observations)

    obs = double(observations);

    oportunityNum = double(k);

    totalT = double(totalTime);
    tSlice = double(timeSlice);
    totalTimePoints = (totalT * tSlice) + 1;
```

```
    myTime = 0 : 1 / tSlice : totalT;

    myX = zeros(obs, 1);
    myZ = zeros(obs, 1);
    for index = 1 : obs
        [myX(index), myZ(index)] = ...
            createObservation(lambdas, gammas, oportunityNum, rho);
    end

    %Calculate output results from the simulation
    myMatFuncs = zeros(totalTimePoints, 9);
    myMatFuncs(:, 1) = myTime;
    myMatFuncs(:, 2) = calcMargX(lambdas, gammas, myTime);
    myMatFuncs(:, 3) = calcMargZ(gammas, oportunityNum, rho, myTime);
    myMatFuncs(:, 4) = calcSubDistX(lambdas, gammas, oportunityNum, rho, myTime);
    myMatFuncs(:, 5) = calcEmpMargX(myX, myTime);
    myMatFuncs(:, 6) = calcEmpMargZ(myZ, myTime);
    myMatFuncs(:, 7) = calcEmpSubDistX(myX, myZ, myTime);

    myXLessZ = calcXLessZ(lambdas, gammas, oportunityNum, rho);
    kolmStats = calcKolmogorovSmirnovStats(myMatFuncs, obs);
end
```

B.2 validateParams

```
function validateParams(totalTime, timeSlice, ...
    lambdas, gammas, k, rho, observations)
%Validate Input params:

    %verify that totalTime is an Integer greater than or equal to 1
    if(isinteger(totalTime) ~= 1 || totalTime <= 0)
        error('faulty totalTime value')
```

```
end
%verify that timeSlice is an Integer greater than or equal to 1
if(isinteger(timeSlice) ~= 1 || timeSlice <= 0)
    error('faulty timeSlice value')
end
%verify that lambdas and gammas are both arrays of the same size (and
%that this is greater than or equal to 2
gammaLength = length(gammas);
if(gammaLength ~= length(lambdas))
    error('number of gammas is not equal to number of lambdas')
end
if(gammaLength < 2)
    error('number of gammas and lambdas is too small')
end
%verify that the last element of gammas is 0
if(gammas(end) ~= 0)
    error('last gamma value is non-zero')
end
%verify that k is an integer greater than or equal to 1 and less than
%the size of gammas (and lambdas)
if(isinteger(k) ~= 1 || k <= 0 || k >= gammaLength)
    error('faulty k value')
end
%verify that observations is an Integer greater than or equal to 1
if(isinteger(observations) ~= 1 || observations <= 0)
    error('faulty observations value')
end

%verify that the gammas are distinct
%calculate the matrix of the differences between any pair of gammas
valGammaDiffs = ones(gammaLength);
```

```
for rowIndex = 1 : gammaLength
    for colIndex = 1 : gammaLength
        if(rowIndex ~= colIndex)
            valGammaDiffs(rowIndex, colIndex) = ...
                gammas(rowIndex) - gammas(colIndex);
        end
    end
end

valTemp = ones(1);
for rowIndex = 1 : gammaLength
    for colIndex = 1 : gammaLength
        valTemp = valTemp .* valGammaDiffs(rowIndex, colIndex);
    end
end
if(valTemp == 0)
    error('gammas are not distinct')
end

%verify that the deltas are distinct

%calculate deltas
valDeltas = lambdas + gammas;
valDeltaDiffs = ones(gammaLength);
for rowIndex = 1 : gammaLength
    for colIndex = 1 : gammaLength
        if(rowIndex ~= colIndex)
            valDeltaDiffs(rowIndex, colIndex) = ...
                valDeltas(rowIndex) - valDeltas(colIndex);
        end
    end
```

```
    end

    valTemp = 1;
    for rowIndex = 1 : gammaLength
        for colIndex = 1 : gammaLength
            valTemp = valTemp .* valDeltaDiffs(rowIndex, colIndex);
        end
    end
    if(valTemp == 0)
        error('deltas are not distinct')
    end

    %verify that rho is distinct from the gammas
    for index = 1 : gammaLength
        if(gammas(index) - rho == 0)
            error('rho is not distinct from the gammas')
        end
    end

    %verify that the first k deltas are distinct from rho plus the last
    %(n - k) deltas
    for firstIndex = 1 : double(k)
        for lastIndex = (double(k) + 1) : gammaLength
            if(valDeltas(firstIndex) - valDeltas(lastIndex) - rho == 0)
                error('rho is not distinct from the (deltai - deltaj)s')
            end
        end
    end
end
```

B.3 createObservation

```
function [myX, myZ] = createObservation(lambdas, gammas, k, rho)
%Construct Simulated Data

    %the number of signals (icluding the zero signal at the origin)
    numSignals = length(lambdas);

    %Construct a vector of exponential random variables
    %This is a sequence of observed Signal times
    mySignalVars = zeros(numSignals, 1);
    for index = 2 : numSignals
        mySignalVars(index) = - log(rand(1)) ./ gammas(index - 1);
    end
    %Construct a vector of cumulative hazard function values
    %This is the sequence of values of the cumulative hazard
    %at each Signal time
    myCumHaz = zeros(numSignals, 1);
    for index = 2 : numSignals
        myCumHaz(index) = ...
            myCumHaz(index - 1) + (mySignalVars(index) .* lambdas(index - 1));
    end

    %Construct a sample value to be taken from the cumulative hazard
    myHazSample = - log(rand(1));

    %Construct X Value by using the sample from the cumulative hazard
    myX = zeros(1);

    %Construct absolute signal times
    myCumSigs = zeros(numSignals, 1);
    for index = 2 : numSignals
```

```
    myCumSigs(index) = myCumSigs(index - 1) + mySignalVars(index);
end

%if the sample value is between the cumulative hazard at S(1) and the
%cumulative hazard at S(0) = 0
if(myHazSample < myCumHaz(2) && myHazSample >= myCumHaz(1))
    myX = ((myHazSample - myCumHaz(1)) ./ lambdas(1)) + myCumSigs(1);
end

%if the sample value is between the cumulative hazard at S(index) and
%S(index - 1)
for index = 2 : (numSignals - 1)
    if(myHazSample < myCumHaz(index + 1) && myHazSample >= myCumHaz(index))
        myX = ((myHazSample - myCumHaz(index)) ./ ...
            lambdas(index)) + myCumSigs(index);
    end
end
%if the sample value is greater than the cumulative hazard at S(n)
if(myHazSample >= myCumHaz(numSignals))
    myX = ((myHazSample - myCumHaz(numSignals)) ./ ...
        lambdas(numSignals)) + myCumSigs(numSignals);
end

%Construct time to Kth Signal
myKSignal = myCumSigs(k + 1);

%Construct Z Value
myZ = zeros(1);
while (myZ < myKSignal)
    myZ = myZ - (log(rand(1)) ./ rho);
end
```

```
end
```

B.4 Analytical Functions

B.4.1 calcMargX

```
function margX = calcMargX(lambdas, gammas, t)
%calcMargX - Calculate the Marginal Distribution of X

    %the number of signals (icluding the zero signal at the origin)
    numSignals = length(lambdas);

    %calculate deltas
    deltas = lambdas + gammas;

    %calculate the matrix of the differences between any pair of deltas
    deltaDiffs = ones(numSignals);
    for rowIndex = 1 : numSignals
        for colIndex = 1 : numSignals
            if(rowIndex ~= colIndex)
                deltaDiffs(rowIndex, colIndex) = ...
                    deltas(rowIndex) - deltas(colIndex);
            end
        end
    end

    %calculate the matrix of products of the differences
    productDifferences = ones(numSignals);
    for colIndex = 1 : numSignals
        productDifferences(1, colIndex) = deltaDiffs(1, colIndex);
        for rowIndex = 2 : numSignals
            productDifferences(rowIndex, colIndex) = ...
```

```
            productDifferences(rowIndex - 1, colIndex) .* ...
            deltaDiffs(rowIndex, colIndex);
    end
end

%calculate the vector of gamma products
gammaProducts = ones(numSignals, 1);
for index = 2 : numSignals
        gammaProducts(index) = gammaProducts(index - 1) .* gammas(index - 1);
end

%calculate the vector of multiplicative factors for the sum of
%exponential terms
expFactors = zeros(numSignals, 1);
for colIndex = 1 : numSignals
    for rowIndex = colIndex : numSignals
        expFactors(colIndex) = ...
            expFactors(colIndex) + (gammaProducts(rowIndex) ./ ...
            productDifferences(rowIndex, colIndex));
    end
end

%calculate the marginal at each time point
totalTimePoints = length(t);
margX = ones(totalTimePoints, 1);
for time = 1 : totalTimePoints
    for index = 1 : numSignals
        margX(time) = margX(time) - ...
            (expFactors(index) .* exp(-deltas(index) .* t(time)));
    end
end
```

```
end
```

B.4.2 calcMargZ

```
function margZ = calcMargZ(gammas, k, rho, t)
%calcMargZ - Calculate the Marginal Distribution of Z

    %calculate the matrix of the differences between any pair of gammas
    gammaDiffs = ones(k);
    for rowIndex = 1 : k
        for colIndex = 1 : k
            if(rowIndex ~= colIndex)
                gammaDiffs(rowIndex, colIndex) = ...
                    gammas(rowIndex) - gammas(colIndex);
            end
        end
    end

    %calculate the vector of products of the differences
    productGammaDifferences = ones(k, 1);
    for colIndex = 1 : k
        productGammaDifferences(colIndex) = gammaDiffs(1, colIndex);
        for index = 2 : k
            productGammaDifferences(colIndex) = ...
                productGammaDifferences(colIndex) * gammaDiffs(index, colIndex);
        end
    end

    %calculate the vector of gamma products
    gammaProducts = ones(1);
    for index = 1 : k
            gammaProducts = gammaProducts .* gammas(index);
```

```
    end

    %calculate the marginal at each time point
    totalTimePoints = length(t);
    margZ = zeros(totalTimePoints, 1);
    for time = 1 : totalTimePoints
        for index = 1 : k
            temp = ((1 - exp(-gammas(index) * t(time))) / ...
                (gammas(index) * productGammaDifferences(index)));
            margZ(time) = margZ(time) + temp;
            temp = ((exp(-gammas(index) * t(time)) - exp(-rho * t(time))) / ...
                ((rho - gammas(index)) * productGammaDifferences(index)));
            margZ(time) = margZ(time) - temp;
        end
        margZ(time) = margZ(time) * gammaProducts;
    end
end
```

B.4.3 calcSubDistX

```
function subDistX = calcSubDistX(lambdas, gammas, k, rho, t)
%calcSubDistX - Calculate the Sub-Distribution of X
    %the number of signals (icluding the zero signal at the origin)
    numSignals = length(lambdas);

    %calculate deltas
    deltas = lambdas + gammas;

    %calculate the matrix of the differences between any pair of deltas
    deltaDiffs = ones(numSignals);
    for rowIndex = 1 : numSignals
        for colIndex = 1 : numSignals
```

```
        if(rowIndex ~= colIndex)
            deltaDiffs(rowIndex, colIndex) = ...
                deltas(rowIndex) - deltas(colIndex);
        end
    end
end

%calculate the matrix of the first products of the delta differences
deltaFirstProdDiffs = ones(numSignals);
for colIndex = 1 : k
    deltaFirstProdDiffs(1, colIndex) = deltaDiffs(1, colIndex);
    for rowIndex = 2 : k
        deltaFirstProdDiffs(rowIndex, colIndex) = ...
            deltaFirstProdDiffs(rowIndex - 1, colIndex) .* ...
            deltaDiffs(rowIndex, colIndex);
    end
end

%calculate the vector of products of the delta differences
deltaLastProdDiffs = ones(numSignals);
for colIndex = (k + 1) : numSignals
    deltaLastProdDiffs((k + 1), colIndex) = deltaDiffs((k + 1), colIndex);
    for rowIndex = (k + 2) : numSignals
        deltaLastProdDiffs(rowIndex, colIndex) = ...
            deltaLastProdDiffs(rowIndex - 1, colIndex) .* ...
            deltaDiffs(rowIndex, colIndex);
    end
end

%calculate the matrix of products of the delta differences with plus
%rho
```

```
plusRhoProdDiffs = ones(numSignals);
for colIndex = 1 : k
    plusRhoProdDiffs(k + 1, colIndex) = deltaDiffs(k + 1, colIndex) + rho;
    for rowIndex = (k + 2) : numSignals
        plusRhoProdDiffs(rowIndex, colIndex) = ...
            plusRhoProdDiffs(rowIndex - 1, colIndex) .* ...
            (deltaDiffs(rowIndex, colIndex) + rho);
    end
end

%calculate the matrix of products of the delta differences with minus
%rho
minusRhoProdDiffs = ones(numSignals);
for colIndex = (k + 1) : numSignals
    minusRhoProdDiffs(1, colIndex) = deltaDiffs(1, colIndex) - rho;
    for rowIndex = 2 : k
        minusRhoProdDiffs(rowIndex, colIndex) = ...
            minusRhoProdDiffs(rowIndex - 1, colIndex) .* ...
            (deltaDiffs(rowIndex, colIndex) - rho);
    end
end

%calculate the vector of gamma products
gammaProducts = ones(numSignals, 1);
for index = 2 : numSignals
        gammaProducts(index) = gammaProducts(index - 1) .* gammas(index - 1);
end

%calculate the vector of multiplicative factors for the sum of
%exponential terms
expFactors = zeros(k, 1);
```

```
for colIndex = 1 : k
    for rowIndex = colIndex : k
        expFactors(colIndex) = ...
            expFactors(colIndex) + (gammaProducts(rowIndex) ./ ...
            deltaFirstProdDiffs(rowIndex, colIndex));
    end
end

%calculate the marginal at each time point
totalTimePoints = length(t);
subDistX = ones(totalTimePoints, 1);
temp = zeros(1);

for time = 1 : totalTimePoints
    for indexj = 1 : k
        subDistX(time) = subDistX(time) - (expFactors(indexj) .* ...
            exp(-deltas(indexj) .* t(time)));
    end
    for indexu = (k + 1) : numSignals
        for indexq = 1 : k
            temp = temp + ...
                (((rho .* (deltas(indexq) - lambdas(indexu) - rho) .* ...
                (1 - exp(-deltas(indexq) .* t(time)))) - ...
                (deltas(indexq) .* lambdas(indexu) .* ...
                exp(-deltas(indexq) .* t(time)))) ./ ...
                (deltas(indexq) .* plusRhoProdDiffs(indexu,indexq) .* ...
                deltaFirstProdDiffs(k,indexq)));
        end
        for indexq = (k + 1) : indexu
            temp = temp + (((rho .* (deltas(indexq) - lambdas(indexu)) .* ...
                (1 - exp(-(deltas(indexq) + rho) .* t(time)))) - ...
```

```
                    ((deltas(indexq) + rho) .* lambdas(indexu) .* ...
                    exp(-(deltas(indexq) + rho) .* t(time)))) ./ ...
                    ((deltas(indexq) + rho) .* minusRhoProdDiffs(k,indexq) .* ...
                    deltaLastProdDiffs(indexu,indexq)));
            end
            subDistX(time) = subDistX(time) + (gammaProducts(indexu) .* ...
                temp ./ (lambdas(indexu) + rho));
            temp = 0;
        end
    end
end
```

B.5 Empirical Calculations

B.5.1 calcEmpMargX

```
function empMargX = calcEmpMargX(myX, t)
%calcEmpMargX - Calculate the Empirical Marginal Distribution of X

    totalTimePoints = length(t);
    sampleNumb = length(myX);
    empMargX = zeros(totalTimePoints, 1);
    for timeIndex = 1 : totalTimePoints
        empMargX(timeIndex) = sum(myX < t(timeIndex))/ sampleNumb;
    end
end
```

B.5.2 calcEmpMargZ

```
function empMargZ = calcEmpMargZ(myZ, t)
%calcEmpMargZ - Calculate the Empirical Marginal Distribution of Z

    totalTimePoints = length(t);
```

```
    sampleNumb = length(myZ);
    empMargZ = zeros(totalTimePoints, 1);
    for timeIndex = 1 : totalTimePoints
        empMargZ(timeIndex) = sum(myZ < t(timeIndex)) / sampleNumb;
    end
end
```

B.5.3 calcEmpSubDistX

```
function empSubDistX = calcEmpSubDistX(myX, myZ, t)
%calcEmpSubDistX - Calculate the Empirical Sub-Distribution of X

    totalTimePoints = length(t);
    sampleNumb = length(myX);
    empSubDistX = zeros(totalTimePoints, 1);
    for timeIndex = 1 : totalTimePoints
        empSubDistX(timeIndex) = ...
            sum((myX < t(timeIndex)) & (myX < myZ)) / sampleNumb;
    end
end
```

B.6 calcXLessZ

```
function XLessZ = calcXLessZ(lambdas, gammas, k, rho)
%calcXLessZ - Calculate the probability that X is less than Z using the
%Sub-Distribution of X
    %the number of signals (icluding the zero signal at the origin)
    numSignals = length(lambdas);

    %calculate deltas
    deltas = lambdas + gammas;
```

```
%calculate the matrix of the differences between any pair of deltas
deltaDiffs = ones(numSignals);
for rowIndex = 1 : numSignals
    for colIndex = 1 : numSignals
        if(rowIndex ~= colIndex)
            deltaDiffs(rowIndex, colIndex) = ...
                deltas(rowIndex) - deltas(colIndex);
        end
    end
end

%calculate the matrix of the first products of the delta differences
deltaFirstProdDiffs = ones(numSignals);
for colIndex = 1 : k
    deltaFirstProdDiffs(1, colIndex) = deltaDiffs(1, colIndex);
    for rowIndex = 2 : k
        deltaFirstProdDiffs(rowIndex, colIndex) = ...
            deltaFirstProdDiffs(rowIndex - 1, colIndex) .* ...
            deltaDiffs(rowIndex, colIndex);
    end
end

%calculate the vector of products of the delta differences
deltaLastProdDiffs = ones(numSignals);
for colIndex = (k + 1) : numSignals
    deltaLastProdDiffs((k + 1), colIndex) = deltaDiffs((k + 1), colIndex);
    for rowIndex = (k + 2) : numSignals
        deltaLastProdDiffs(rowIndex, colIndex) = ...
            deltaLastProdDiffs(rowIndex - 1, colIndex) .* ...
            deltaDiffs(rowIndex, colIndex);
    end
```

```
end

%calculate the matrix of products of the delta differences with plus
%rho
plusRhoProdDiffs = ones(numSignals);
for colIndex = 1 : k
    plusRhoProdDiffs(k + 1, colIndex) = deltaDiffs(k + 1, colIndex) + rho;
    for rowIndex = (k + 2) : numSignals
        plusRhoProdDiffs(rowIndex, colIndex) = ...
            plusRhoProdDiffs(rowIndex - 1, colIndex) .* ...
            (deltaDiffs(rowIndex, colIndex) + rho);
    end
end

%calculate the matrix of products of the delta differences with minus
%rho
minusRhoProdDiffs = ones(numSignals);
for colIndex = (k + 1) : numSignals
    minusRhoProdDiffs(1, colIndex) = deltaDiffs(1, colIndex) - rho;
    for rowIndex = 2 : k
        minusRhoProdDiffs(rowIndex, colIndex) = ...
            minusRhoProdDiffs(rowIndex - 1, colIndex) .* ...
            (deltaDiffs(rowIndex, colIndex) - rho);
    end
end

%calculate the vector of gamma products
gammaProducts = ones(numSignals, 1);
for index = 2 : numSignals
        gammaProducts(index) = gammaProducts(index - 1) .* gammas(index - 1);
end
```

```
%calculate the vector of multiplicative factors for the sum of
%exponential terms
expFactors = zeros(k, 1);
for colIndex = 1 : k
    for rowIndex = colIndex : k
        expFactors(colIndex) = expFactors(colIndex) + ...
            (gammaProducts(rowIndex) ./ ...
            deltaFirstProdDiffs(rowIndex, colIndex));
    end
end

%calculate the probability that X is less than Z
XLessZ = ones(1);
temp = zeros(1);
for indexu = (k + 1) : numSignals
    for indexq = 1 : k
        temp = temp + ...
            ((rho .* (deltas(indexq) - lambdas(indexu) - rho)) ./ ...
            (deltas(indexq) .* plusRhoProdDiffs(indexu,indexq) .* ...
            deltaFirstProdDiffs(k,indexq)));
    end
    for indexq = (k + 1) : indexu
        temp = temp + ...
            ((rho .* (deltas(indexq) - lambdas(indexu))) ./ ...
            ((deltas(indexq) + rho) .* minusRhoProdDiffs(k,indexq) .* ...
            deltaLastProdDiffs(indexu,indexq)));
    end
    XLessZ = ...
        XLessZ + (gammaProducts(indexu) .* temp ./ (lambdas(indexu) + rho));
    temp = 0;
```

```
    end
end
```

B.7 calcKolmogorovSmirnovStats

```
%Calculation of the Kolmogorov Smirnov statistic for each pair of empirical
%and analytical functions
function kolmStats = calcKolmogorovSmirnovStats(myMatFuncs, observations)

%calculate the supremum (maximum of the differences between each function
%pair

    kolmStats = zeros(3,1);
    kolmStats(1) = max(abs(myMatFuncs(:,5) - myMatFuncs(:,2)));
    kolmStats(2) = max(abs(myMatFuncs(:,6) - myMatFuncs(:,3)));
    kolmStats(3) = max(abs(myMatFuncs(:,7) - myMatFuncs(:,4)));

    kolmStats = kolmStats .* sqrt(observations);
end
```

Appendix C

Peterson Bounds for GVA Models

This chapter presents an interesting extension of the Peterson Bounds for classical competing risk models (discussed in Section 2.2) to the case of symmetric GCR models using virtual ages (discussed in Section 4.2). In Classical CR the marginal distributions of the components cannot be identified, however it is well-known that the joint survival function is bounded by the Peterson bounds[73] like so

$$\left[S_X^* + S_Z^*\right]\left[\max\left(x, z\right)\right] \leq S(x, z) \leq S_X^*(x) + S_Z^*(z).$$

A similar result will now be demonstrated for the class of symmetric GVA models introduced in Section 4.2. The proof will follow a very similar line to Peterson's original proof.

Theorem 17. *Given a joint survival function S_1 for a symmetric GVA model with fixed sub-survival functions $S_{X_1}^*$ and $S_{Z_1}^*$, the joint survival function is bounded point-wise in the following manner.*

$$\left[S_{X_1}^* + S_{Z_1}^*\right]\left(\max[x, z] + A_k\right) \leq S_1(A_k + x, A_k + z) \leq S_{X_1}^*(A_k + x) + S_{Z_1}^*(A_k + z).$$

Proof. For the $(k+1)^{th}$ survival function

$$\begin{aligned} S_{k+1}(x, z|\mathbf{T}_k, \boldsymbol{\delta}_k) &= P(X_{k+1} > x, Z_{k+1} > z|\mathbf{T}_k, \boldsymbol{\delta}_k) \\ &= P(X_{k+1} > x, Z_{k+1} > z, X_{k+1} > Z_{k+1}|\mathbf{T}_k, \boldsymbol{\delta}_k) \\ &\quad + P(X_{k+1} > x, Z_{k+1} > z, X_{k+1} \leq Z_{k+1}|\mathbf{T}_k, \boldsymbol{\delta}_k). \end{aligned}$$

Upper bounds for these two terms are

$$\begin{aligned}P(X_{k+1} > x, Z_{k+1} >& z, X_{k+1} > Z_{k+1}|\mathbf{T}_k, \boldsymbol{\delta}_k) \\ &\leq P(Z_{k+1} > z, X_{k+1} > Z_{k+1}|\mathbf{T}_k, \boldsymbol{\delta}_k) = S^*_{Z_{k+1}}(z|\mathbf{T}_k, \boldsymbol{\delta}_k)\end{aligned}$$

and

$$\begin{aligned}P(X_{k+1} > x, Z_{k+1} >& z, X_{k+1} \leq Z_{k+1}|\mathbf{T}_k, \boldsymbol{\delta}_k) \\ &\leq P(X_{k+1} > x, X_{k+1} \leq Z_{k+1}|\mathbf{T}_k, \boldsymbol{\delta}_k) = S^*_{X_{k+1}}(x|\mathbf{T}_k, \boldsymbol{\delta}_k).\end{aligned}$$

Lower bounds for these terms are

$$\begin{aligned}P(X_{k+1} > x, Z_{k+1} >& z, X_{k+1} > Z_{k+1}|\mathbf{T}_k, \boldsymbol{\delta}_k) \\ &\geq P(X_{k+1} > \max[x,z], Z_{k+1} > \max[x,z], X_{k+1} > Z_{k+1}|\mathbf{T}_k, \boldsymbol{\delta}_k) \\ &= P(Z_{k+1} > \max[x,z], X_{k+1} > Z_{k+1}|\mathbf{T}_k, \boldsymbol{\delta}_k) \\ &= S^*_{Z_{k+1}}(\max[x,z]|\mathbf{T}_k, \boldsymbol{\delta}_k)\end{aligned}$$

and

$$\begin{aligned}P(X_{k+1} > x, Z_{k+1} >& z, X_{k+1} \leq Z_{k+1}|\mathbf{T}_k, \boldsymbol{\delta}_k) \\ &\geq P(X_{k+1} > \max[x,z], Z_{k+1} > \max[x,z], X_{k+1} \leq Z_{k+1}|\mathbf{T}_k, \boldsymbol{\delta}_k) \\ &= P(X_{k+1} > \max[x,z], X_{k+1} \leq Z_{k+1}|\mathbf{T}_k, \boldsymbol{\delta}_k) \\ &= S^*_{X_{k+1}}(\max[x,z]|\mathbf{T}_k, \boldsymbol{\delta}_k).\end{aligned}$$

These bounds all lead to the following inequalities, for all $k \geq 1$

$$\left[S^*_{X_{k+1}} + S^*_{Z_{k+1}}\right](\max[x,z]) \leq S_{k+1}(x,z) \leq S^*_{X_{k+1}}(x) + S^*_{Z_{k+1}}(z).$$

But $S^*_{X_{k+1}}$ can be written in terms of $S^*_{X_1}$

$$\begin{aligned} S^*_{X_{k+1}} &= P(X_{k+1} > x, X_{k+1} \leq Z_{k+1} | \mathbf{T}_k, \boldsymbol{\delta}_k) \\ &= P(X_1 > A_k + x, X_1 \leq Z_1 | X_1 > A_k, Z_1 > A_k, A_k) \\ &= \frac{S^*_{X_1}(A_k + x)}{S_1(A_k, A_k)}. \end{aligned}$$

Similarly, $S^*_{Z_{k+1}}$ can be written in terms of $S^*_{Z_1}$. Also note that

$$\begin{aligned} S_{k+1} &= P(X_{k+1} > x, Z_{k+1} > z | \mathbf{T}_k, \boldsymbol{\delta}_k) \\ &= P(X_1 > A_k + x, Z_1 > A_k + z | X_1 > A_k, Z_1 > A_k, A_k) \\ &= \frac{S_1(A_k + x, A_k + z)}{S_1(A_k, A_k)}. \end{aligned}$$

Combining these equalities with the above bounds and cancelling out the denominators gives the desired result. □

This work could be extended in a similar fashion as the work of Bedford and Meilijson[13], discussed in Section 2.2. They were able to give sharp bounds by including a subtle case not considered by Peterson. They were also able to generate a confidence interval based on these bounds which could be used as the basis for a statistical test on the goodness of fit of any particular model given the data. In this case, once a virtual ages model has been chosen, it may be possible to develop a similar goodness of fit test for that particular symmetric GVA model.