Double Barrier Models for Length of Stay in Hospital

by

Julie Horrocks

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics and Actuarial Science

Waterloo, Ontario, Canada, 1999

0-612-44767-7

The University of Waterloo requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

Abstract

Length of stay (LOS) in hospital is a widely used outcome measure in Health Services research, often acting as a surrogate for resource consumption or as a measure of efficiency. Recent activity in the field focuses on modeling the dependence of LOS on covariates, using administrative data collected for the purpose of calculating fees for doctors, or data extracted from medical charts. This problem is a challenging one, due to the high skewness of the distribution of LOS, the presence of multiple destinations (healthy discharge, death in hospital, transfer to another institution) and the unexplained heterogeneity which remains even after all available covariates have been included in the model.

In this thesis, we develop parametric models for LOS that accommodate the skewness of the distribution and allow for multiple destinations. The models are based on the time, T, until a Wiener process with drift (representing a health level process) hits one of two barriers, one representing healthy discharge, the other representing death in hospital. The model is parameterized in terms of the barrier levels and drift, which are allowed to depend on covariates. The parameters of the model are estimated using the method of maximum likelihood. We show how to estimate expected LOS and probability of discharge, and discuss ways of testing hypotheses of interest. An interesting feature of the variable T is that the density and distribution functions are infinite series. We show that the density and its derivatives are absolutely and uniformly convergent, and that regularity conditions are satisfied in the zero drift case for iid observations.

The models can easily be extended to allow the drift parameter to have a mixing distribution, thereby partially addressing the issue of unexplained heterogeneity. While mixture models often require numerical integration in order to estimate parameters, we show that, if the mixing distribution is normal, numerical integration is unnecessary for these models. Also, an extension to handle transfers out of hospital is implemented. Since the decision to transfer is at least partially based on the health level of the patient, transfers cannot be treated as independently censored observations. We develop a model in which patients are transferred with probability p when their health level reaches an intermediate decision barrier. We can then model p as a function of covariates. As before, the parameters of these models are estimated using maximum likelihood, and

we show how to estimate expected LOS and probability of healthy discharge.

This approach to analyzing LOS has many parallels with competing risks analysis, and can be seen as a way of formalizing a competing risk situation. Further work will explore incorporation of time-varying covariates, different distributions for the health level process, and formal measures of goodness of fit.

Acknowledgements

First of all, I would like to thank my supervisor, Mary Thompson, who gave me guidance in all things, support, and encouragement through the rough places. Also thank you to my co-supervisor, Jiahua Chen, and the members of my committee, David Matthews, and Jeanette O'Hara-Hines, for many helpful suggestions. And thank you to my external examiners, Alex Whitmore and John Hirdes, who helped to make my defense a learning experience. I would like to thank all the many faculty and grad students whom I have befriended over the years. Without you, life would have been dull indeed.

Thanks to Darcy for seeing me through the first years of this project, and for invaluable ongoing computing and math advice. Thanks to my parents and siblings for always having faith in me. And finally thanks to Cam, for conversations, emotional support, childcare, and especially for Corinna.

Contents

List of Figures

Chapter 1

Introduction

1.1 Length of Stay in Hospital

Length of stay (LOS) in hospital is a widely used outcome measure in Health Services research, often acting as a surrogate for resource consumption or as a measure of efficiency [4]. Research questions sometimes focus on average length of stay for patients grouped by hospital or physician. For instance Rosenthal et al [34] in a 1997 JAMA article compare severity-adjusted mortality and LOS in teaching versus non-teaching hospitals. In order to ensure fair comparisons between groups of patients, adjustments need to be made for both hospital-level variables such as number of beds and hospital type (chronic versus acute care), and individual-level variables such as age, sex and presence of chronic conditions. Other studies try to determine how patient-level characteristics affect length of stay. Bonuck and Arno [5] found that HIV/AIDS patients with inadequate housing stayed in hospital five days longer on average than those with stable housing. Morris et al [28] model durations of stay in nursing homes as a function of patient characteristics such as age, gender, marital status and general health.

The data used to investigate these questions are often administrative data, whose primary purpose is in calculating fees for doctors. Even after all available covariates have been included in the model, large amounts of heterogeneity remain. This situation remains true even when clinical data are used. For instance, Rosenthal et al [34] studied six diagnoses that represent common causes of hospitalization, and used patients' medical records as the data source. They found that at most 25% of the variation in LOS could be explained by their severity-adjusted models.

There are several features of LOS which make it challenging to model. It has a highly skewed distribution and the presence of outliers is common. Figure 1.1 shows a histogram of LOS for females admitted to hospital with circulatory disorders and myocardial infarction. While ro-

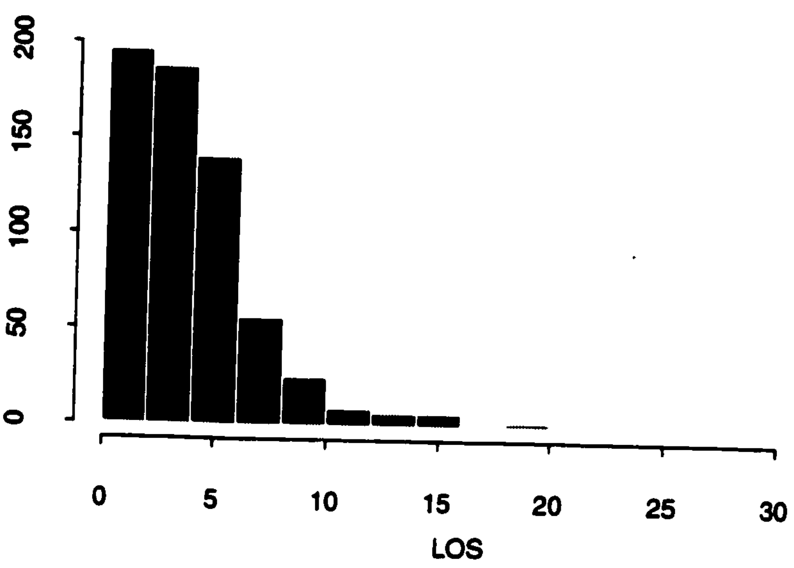


Figure 1.1: Histogram of LOS, Females, APR-DRG=121, Utah Discharge Database

bust methods would decrease the influence of outliers, interest centers on total consumption as estimated by total LOS, and outliers are an important component of the total.

Another interesting feature of LOS data is the existence of multiple destinations, as shown in Figure 1.2. An individual's stay in hospital can end in healthy discharge, death, or transfer to a different institution. So far, this aspect has largely been ignored in health services research. Often only patients who experienced healthy discharge are analyzed, or all hospital stays are analyzed together, ignoring final outcome.

A further challenge is the residual heterogeneity that remains even after all available covariates have been included in the model. For normal linear models, any unexplained variation becomes part of the residual variation and is accounted for in confidence intervals and tests, but does not affect interpretation of parameters. However heterogeneity does affect the interpretation of pa-

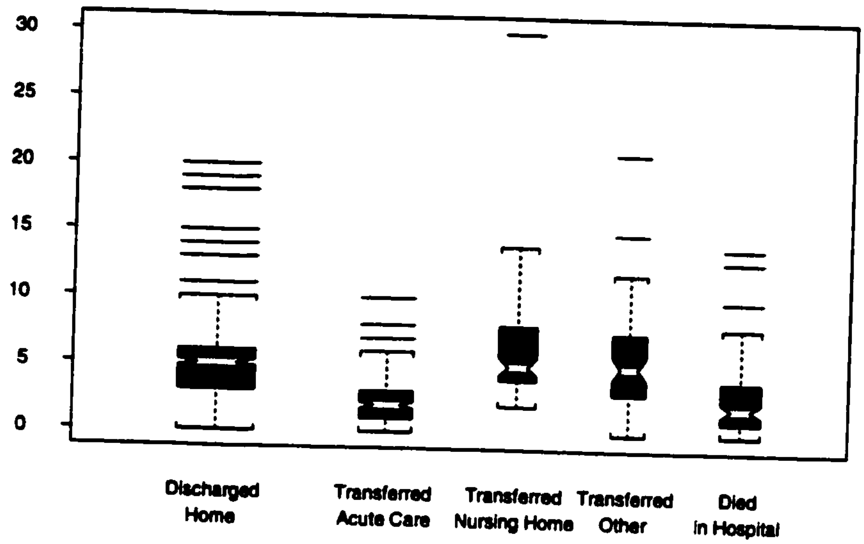


Figure 1.2: LOS by Discharge Destination, Females, APR-DRG=121, Utah Discharge Database

rameters in non-normal models such as the logistic model, as discussed by Neuhaus, Kalbfleisch and Hauck [31]. Moreover, when the response is time to some event, a selection effect occurs because individuals at high risk experience the event of interest sooner [1, 41]. Results must be interpreted with care. The population hazard function, $h(t)$, represents the conditional rate of failure given survival to time t, averaged over the whole population. In the presence of heterogeneity, it cannot be interpreted as an individual's conditional failure rate. This will be discussed in more detail in section 6.1.

In this thesis, we develop models for LOS which deal with multiple destinations, and accommodate the highly skewed distribution of LOS. These models are derived from the waiting time until a Wiener process hits one of two barriers, one barrier representing healthy discharge, the other death in hospital. Transfers to other institutions may be modelled as breach of an intermediate barrier. These models are rich in the many ways that they can accommodate covariates, both individual and hospital-level.

The relationship of these models to competing risks analyses is discussed in Chapter 7. Note that while the competing risks problem is usually presented in terms of failure time and causes of failure, the corresponding ideas for health care data would be time in hospital and discharge status (ie death, discharge home, transfer to another institution). The approach used here is distinct from the classical competing risks problem, as expounded by Daniel Bernoulli in 1760, and as described for instance in the monograph by David and Moeschberger [13], or Chapter 9 of Cox and Oakes [11]. In the classical competing risks analysis, one postulates the existence of several latent failure times, $T_1, T_2, \ldots, T_p$, corresponding to p different causes of failure. Only one of these failure times is actually observed, namely the minimum of $T_1, T_2, \ldots, T_p$. Often, the purpose of a classical competing risks analysis is to estimate the distribution of time to one type of failure, T_i say, once all other causes of failure have been "eliminated". To do this, we must be very precise about what we mean by elimination of a cause of failure. Usually this will involve some very strong assumptions, such as independence of the latent failure times.

In this thesis, we are not concerned with modelling latent times, but consider only the observed time in hospital T. We model the joint distribution of T and D, where D is a discrete variable indicating discharge status. This is similar to the approach described in Kalbfleisch and Prentice ([22], p. 163) and in Cox and Oakes ([11], p. 143).

A brief outline of the thesis is now given. Section 1.2 describes a data set which is used throughout the thesis to demonstrate the various models. In section 1.3 we discuss earlier work relating health status and length of stay in hospital to a Wiener process with a single barrier. In Chapter 2, health status is modelled as a zero-drift Wiener process with two barriers, one representing discharge and the other death. Maximum likelihood is used to estimate the parameters. In Chapter 3, we extend the model to incorporate a drift parameter, μ. Chapters 4 and 5 address technical considerations associated with the models, including uniform convergence of the density functions and their derivatives, and regularity conditions. In Chapter 6 we discuss a mixture model in which the drift parameter μ is itself a random variable from a Normal distribution. This allows the model to account for some excess heterogeneity. We also discuss a model which allows for transfers to another institution. Chapter 7 examines the relationship of this approach to the competing risks analyses expounded by Prentice et al [32]. Finally, Chapter 8 concludes the thesis with a discussion of the models developed, and some areas for further research.

1.2 Data

Throughout this thesis, the Utah Hospital Inpatient Discharge Data File (1996) [39] is used for purposes of illustration. Since January 1992, all licensed hospitals in the State of Utah must, by law, report information on inpatient discharges. This includes personal information about the patient, the services received in hospital and associated charges, and medical information. The Utah Department of Health makes this information available in various formats for public use. Three fixed subsets of the data base are publicly available to researchers at minimal cost, and special requests for additional information are considered. Also an inquiry system is accessible on the internet at http://161.119.100.19/hda/?.

Discharge data is received quarterly from hospitals, and extensive procedures are in place to ensure that the over 50 Utah hospitals report consistent and valid data. Patient confidentiality is assured through various means. Several covariates are available in grouped format only, for instance age and diagnosis codes. Furthermore, where there is a danger that an individual may be identified from the information in the data file, data values are encrypted. Patients treated in small hospitals, or residing in zip-code areas with a small population, or those with sensitive diagnoses (such as HIV) have encrypted information.

Note that no observations are censored, due to the sampling scheme used to construct the database. The database consists of all patients discharged in 1996. Patients who are in hospital on December 31, 1996 and who remain in hospital into 1997 will not be included in this data file, but instead will go into the 1997 data file. On the other hand, patients who were admitted in 1995, but whose hospital stay continued into 1996 and ended in 1996 will appear in the 1996 file. All information will be included on these patients, including procedures performed and diagnoses made in 1995.

In addition to the outcome variables length of stay and discharge status (eg discharged home or to another institution, or died in hospital), a large amount of covariate information is available. Patient level characteristics include agegroup, sex, zip-code and county of residence, marital status and race. The admitting hospital is identified, and we can determine hospital characteristics such as number of beds, type of ownership (government, not-for-profit, investor-owned), affiliation (HMO or free-standing), location, and whether urban or rural, teaching or non-teaching. The type of admission (emergency, urgent, elective, or newborn), the source of admission (eg physician referral, transfer from another institution), and where relevant, the speciality of the admitting physician, are also recorded. Information on the payer (Medicaid, Medicare, HMO, etc.) is also available. All of these variables are known at the time of admission to hospital, and are fixed (ie do not change over time).

The file also contains principal procedure and up to five secondary procedures, and principal diagnosis and up to eight secondary diagnoses. Procedures and diagnoses are recorded in the form of ICD-9-CM codes. ICD-9-CM stands for International Classification of Diseases (9th Revision) Clinical Modification. The International Classification of Diseases is the World Health Organization's system for coding diseases, originally developed in 1975, and subsequently supplemented with codes for surgical and other procedures. There are thousands of these codes, and several systems have been developed for grouping admissions according to ICD-9-CM codes. Three of these systems are included in the database, namely DRG (diagnosis related group), APR-DRG (all-patient-refined DRG) and MDC (major diagnosis category). The MDC system classifies admissions into 25 broad categories based on their ICD-9-CM code for principal diagnosis. The categories relate to the major organ system affected (eg circulatory system), or the type of disorder (eg HIV) or condition (eg pregnancy). The DRG system is a system for classifying patients, (or more precisely admissions), into one of 495 groups, according to information available on the computerized patient chart, namely principal and secondary diagnoses, principal procedure, age, and sex and (for a few DRG's) outcome. This system attempts to group together patients with similar clinical attributes and similar resource consumption, (as measured by LOS and costs). DRG's are used to develop indices of hospital case mix (that is, summary measures of the frequency of types of admissions in a hospital), and as a basis for reimbursement. In some cases, two DRG's are identical except for the presence or absence of complications and comorbidities. A complication or comorbidity is a disease or condition other than the principal diagnosis. While a comorbidity is present on admission to hospital, complications develop while the patient is in hospital. The APR-DRG system is another system, similar to DRG, with 791 categories.

Note that the time of the procedures and diagnoses is not recorded, although this information is presumably available from patient charts. While secondary diagnoses usually represent comorbid conditions present at admission to hospital, sometimes they represent conditions which arose during the hospital stay, for instance myocardial infarction. These events, as well as various procedures, might be important predictors of LOS or costs. If the time of the procedures and conditions which lead to secondary diagnoses were available, they could be used as time-varying covariates. While time-varying covariates are not addressed in this thesis, they are an important area for future research.

The variable "patient severity subclass" (severity), is a measure of the number of comorbidities and complications. It is constructed by the 3M PC-Grouper software which classifies admissions into APR-DRG's, and may, for some DRG's, contain information that is not known at time of admission. Because of the necessity of adjusting for comorbidities in investigative studies, many indices have been developed to quantify severity of disease. Some of these are calculated using chart review data [8, 7], while others are based on claims data, [27]. When data are extracted from charts, researchers can be careful to use only data available on admission to hospital. When indices are constructed from claims data, care must be taken not to include conditions which might have developed during the hospital stay. For instance in a study of prostatectomy [27], myocardial infarction and congestive heart failure were excluded from calculation of a claims-based comorbidity index, since these conditions likely developed after the patient was admitted to hospital. In this thesis, the variable severity is treated as if it were a fixed covariate whose value was known at admission to hospital, and so illustrates how a comorbidity index could be used with these models.

The 1996 data file consists of over 200,000 records. For the figures shown in section 1.1, a small subset of the data base, comprising 632 admissions, was extracted. This is the set of all female patients admitted to a licensed Utah hospital in fiscal year 1996 with an APR-DRG classification of 121, ie Circulatory Disorders with Acute Myocardial Infarction. This data set or a subset thereof is used throughout the thesis to demonstrate the various models.

1.3 Health Status as a Stochastic Process

The idea of modelling the physiological status of an individual as a one-dimensional stochastic process, with an absorbing barrier representing death, can be traced back at least to Sacher and Trucco [35] and probably beyond. If we could quantify health status, it would be a highly multi-dimensional construct, encompassing for instance organ function, mental health, physical conditioning, age, and gender. Suppose now that we can construct a one-dimensional summary of health status, called health level. Let $H(t)$ denote an individual's health level at time t, and consider the stochastic process $\{H(t); t > 0\}$. Further assume that if the individual's overall health improves, the measure of health level increases, whereas if health deteriorates $H(t)$ decreases. If $H(t)$ reaches a very low level, the individual dies.

Eaton and Whitmore [15] extended this idea to model length of stay in hospital, and specifically assumed that $H(t)$ follows a Wiener process with drift $\mu > 0$ and volatility parameter σ^2. A person is postulated to enter hospital when his health process $H(t)$ falls below a certain level. The time an individual enters hospital is taken to be time 0, and health level at this time is arbitrarily set to 0, ie $H(0) = 0$. An individual's length of stay in hospital is then the time when his health process first rises above the level $u > 0$, ie the first time the process reaches a barrier at u. We can use this conceptualization to derive a distribution for the length of stay in hospital, T. It is in fact well known, (see for instance [38] that T has an inverse Gaussian distribution.

Note that Eaton and Whitmore fit this model to some data on psychiatric patients, very few of whom were transferred or died in hospital, so a single barrier was sufficient for their purposes. Beginning in Chapter 2, we will extend this model to allow for multiple destinations.

There are many ways to account for population heterogeneity in this model. The barrier position may vary for different patients, or for patients grouped by hospital or health-care provider. We may be able to model some of this heterogeneity by allowing barrier position to depend on individual level covariates such as marital status. There is some evidence (see [28]) that hospitals tend to discharge patients relatively early if there is a caretaker at home, and being married can be a surrogate for presence of caretaker. Then we might allow the upper barrier to differ for married and unmarried patients. Alternatively, we may allow the upper barrier to depend on type of hospital, whether urban/rural, teaching/non-teaching, etc.

We might also allow the drift and volatility parameters of the underlying Wiener process to vary by individual. For instance we may allow drift to depend on severity of disease, age, sex or other individual-level covariates.

Whitmore [42] extended the model to allow for negative drift, and applied it to data on employment duration. For a Wiener process with negative drift, the first passage time to a barrier above the origin has a defective distribution with a mass of probability at infinity.

In 1983 Whitmore [43] presented a regression model for the inverse gaussian distribution with positive drift parameter μ and upper barrier fixed at $u = 1$. He allowed the drift to depend on the covariates through an inverse link function (ie $1/\mu = x^T\beta$), and allowed for censoring. He used the EM algorithm to find the maximum likelihood estimates of the regression and volatility parameters. Inference was based on asymptotic distributions. He was not able to extend this model to the negative drift case however.

Aalen [1] discussed these approaches and extended the model by allowing the drift parameter to have a mixing distribution. This is discussed further in Chapter 6.

Chapter 2

Double Barrier Model, Zero Drift

In this chapter, we introduce a simple double barrier model for length of stay in hospital. This extends the work of Eaton and Whitmore [15] to allow the latent health level process to end at one of two barriers, representing healthy discharge and death. The problem of allowing for other outcomes will be considered in section 6.2.

A brief outline of the Chapter is now given. In Section 2.1 we examine the distribution of T, the time the health level process first reaches either a barrier representing healthy discharge or a barrier representing death in hospital. We discuss derivation of the cumulative distribution function, density, and subdensities, and examine the form of these functions as well as the hazard and cumulative hazard. In Section 2.2 we discuss maximum likelihood estimation of the parameters, and non-linear functions of the parameters. We extend the model to accommodate covariates, and discuss starting values needed for the iterative estimation procedure. Section 2.3 deals with informal methods of model assessment, while in Section 2.4 we demonstrate the model using some real data.

2.1 Distribution of Time until Breach

In addition to an upper barrier representing healthy discharge, we now postulate the existence of a lower barrier representing death. An individual's length of stay in hospital is modelled as the time, T, until his health level process first reaches one of the two barriers. To simplify the discussion in this section, assume that the process has zero drift. We further assume that the data set represents a fairly homogeneous group of individuals, all of whom have approximately the same health status on admission to hospital, and all of whom share the same upper and lower barriers. These restrictions will be relaxed later, by allowing the barrier parameters to depend on covariates (section 2.2.2), and by introducing a drift parameter (Chapter 3).

To further specify the model, assume that an individual's health level process, $\{H(t), t > 0\}$, is a standard Wiener process, with drift 0 and volatility 1. (Note that taking volatility equal to 1 entails no loss of generality here, since it is a scaling factor and health status is, in any case, a latent (unmeasurable) quantity.) The individual is assumed to enter hospital at time 0, with health level 0, i.e. $H(0) = 0$. The health level process unfolds in the presence of two barriers, as shown in figure 2.1. The upper barrier is identified with a health level of u, and the lower barrier with health level equal to $-\ell$, where $u > 0$ and $\ell > 0$. The random variable T represents the time the process first reaches one of the two barriers, which we equate with length of stay in hospital. We will say that the process ends when it hits one of the two barriers. The distribution of T has no name but is discussed in most intermediate-level textbooks on stochastic processes (eg [17, 23]). Where necessary, we will refer to it as a first passage two barrier (FP2B) distribution.

Let $F(t; u, \ell)$ denote the cumulative distribution function (cdf) of T, so that $F(t; u, \ell)$ is the probability that the process hits one of the two barriers at or before time t. The process may end at time t by reaching the upper barrier u, or the lower barrier $-\ell$. Let $F_\ell(t)$ equal the probability that the process hits the lower barrier at or before time t without reaching the upper barrier. Similarly, let $F_u(t)$ equal the probability that the process hits the upper barrier at or before time t without having hit the lower barrier. Then clearly $F(t; u, \ell) = F_\ell(t) + F_u(t)$.

We may also think of this situation as a multivariate distribution, with two random variables, one continuous and one discrete. This point of view is similar to the model for several types of failure discussed in Cox and Oakes [11, p.143], also discussed by Kalbfleisch and Prentice [22, p. 163] under the name of competing risks model. As before, let the continuous random variable T represent the time until the process hits one of the two barriers. Further, define the discrete random variable D which takes the value u if the process ends at the upper barrier, and which takes the value l if the process ends at the lower barrier. Then $F_\ell(t) = P(T \leq t, D = \mathrm{l})$ while

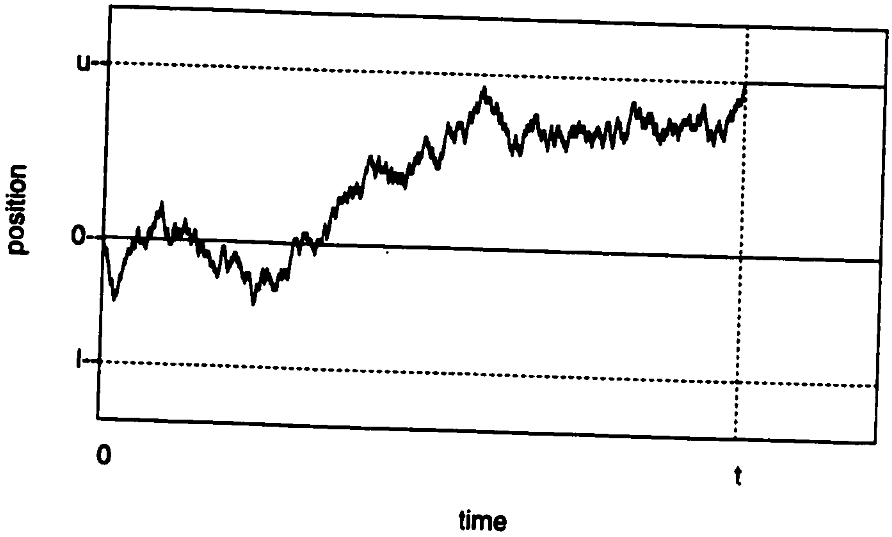


Figure 2.1: Wiener Process starting at 0 with barriers at ℓ and u

$F_u(t) = P(T \leq t, D = \text{u})$. The joint distribution of (T, D) can now be described by

$$F_{T,D}(t,d) = \begin{cases} F_u(t) & \text{if } d = \text{u} \\ F_\ell(t) & \text{if } d = \text{l}. \end{cases}$$

Note that in health services research, we have data on both time in hospital (corresponding to T) and discharge status (corresponding to D), so that we are able to model this joint distribution. $F_\ell(t)$ and $F_u(t)$ are referred to as subdistribution functions, because the limit of each function as t approaches infinity is less than one. The marginal cdf of T is of course $F(t; u, \ell)$ and the probability that the process ultimately ends at the upper barrier is

$$P(D = \text{u}) = \lim_{t \to \infty} F_u(t).$$

Since we are assuming zero drift, these probabilities can be derived via simple reflection arguments. These arguments are heuristic but are quite standard (see for instance [23]). In appendix

B we show that the upper subdistribution function $F_u(t)$ is given by

$$F_u(t) = 2\sum_{k=0}^{\infty}\left\{\Phi\left(\frac{s_k+\ell}{\sqrt{t}}\right) - \Phi\left(\frac{s_k-\ell}{\sqrt{t}}\right)\right\} \tag{2.1}$$

where

$$s_k = -(2k+1)(u+\ell)$$

and $\Phi(x)$ is the standardized normal cumulative distribution function evaluated at x. Similarly, we can show that the lower subdistribution function is

$$F_\ell(t) = 2\sum_{k=0}^{\infty}\left\{\Phi\left(\frac{s_k+u}{\sqrt{t}}\right) - \Phi\left(\frac{s_k-u}{\sqrt{t}}\right)\right\}. \tag{2.2}$$

Note that to get the lower subdistribution function, we only need to reverse the roles of u and ℓ in the upper subdistribution function, ie if we write $F_\ell(t) = g(t;\ell,u)$ then $F_u(t) = g(t;u,\ell)$. This is discussed further in appendix B. Of course, if $u = \ell$, then $F_u(t) = F_\ell(t)$, and $P(D = u) = F_u(\infty) = P(D = l) = F_\ell(\infty) = .5$. We will see throughout the thesis that probabilities associated with the lower subdistribution function have the same functional form as corresponding probabilities associated with the upper subdistribution function. This fact can be exploited when writing the computer programs to estimate parameters. For instance a single module in the computer program can calculate both $F_\ell(t)$ and $F_u(t)$, with only a change of argument. Finally, since $F(t;u,\ell) = F_u(t) + F_\ell(t)$, we find

$$F(t;u,\ell) = 2\sum_{k=0}^{\infty}\left\{\Phi\left(\frac{s_k+\ell}{\sqrt{t}}\right) - \Phi\left(\frac{s_k-\ell}{\sqrt{t}}\right) + \Phi\left(\frac{s_k+u}{\sqrt{t}}\right) - \Phi\left(\frac{s_k-u}{\sqrt{t}}\right)\right\}.$$

Consider for a moment the series

$$F_u(t) = 2\sum_{k=0}^{\infty}\left\{\Phi\left(\frac{s_k+\ell}{\sqrt{t}}\right) - \Phi\left(\frac{s_k-\ell}{\sqrt{t}}\right)\right\}.$$

Let $g_{2k} = s_k + \ell$ and $g_{2k+1} = s_k - \ell$. Then the sequence $\{g_k\}_{k=0}^{\infty}$ is a decreasing sequence of negative numbers. Also, for finite t, the sequence $\{\Phi(g_k/\sqrt{t}\}_{k=0}^{\infty}$ is a decreasing sequence of positive numbers, the limit of which is 0 as $k \to \infty$. Thus the series $F_u(t)$ converges pointwise for finite t, by the alternating series test. Furthermore, as we show in appendix A, this series, grouped as shown, is absolutely convergent for finite t. Similarly, $F_\ell(t)$ is also absolutely convergent. Absolute convergence of $F_\ell(t)$ and $F_u(t)$ imply the absolute convergence of $F(t;u,\ell)$.

There are many other ways of deriving the cdf $F(t;u,\ell)$ or equivalently the survivor function $\mathcal{F}(t;u,\ell) = 1 - F(t;u,\ell)$. For instance, from the joint distribution of $(\max\{H(t) : 0 < t < T\}, \min\{H(t) : 0 < t < T\}, H(T))$, which is given in a 1951 paper by Feller [16], we can derive an alternative form for $F(t;u,\ell)$, namely

$$\begin{aligned} F(t;u,\ell) = 1 - &\sum_{k=-\infty}^{\infty} \left\{\Phi\left(\frac{2k(u+\ell)+u}{\sqrt{t}}\right) - \Phi\left(\frac{2k(u+\ell)-\ell}{\sqrt{t}}\right)\right\} \\ + &\sum_{k=-\infty}^{\infty} \left\{\Phi\left(\frac{2k(u+\ell)+2u+\ell}{\sqrt{t}}\right) - \Phi\left(\frac{2k(u+\ell)+u}{\sqrt{t}}\right)\right\}. \end{aligned}$$

The disturbing fact that these expressions involve infinite series does not turn out to be an insurmountable problem, as in general only a small number of terms are perceptibly different from zero. This is illustrated in Figure 2.2, where for $\ell = 2$, u=1 and $t = 64$, the sum

$$\sum_{k=0}^{\infty}\left\{\Phi\left(\frac{s_k+\ell}{\sqrt{t}}\right) - \Phi\left(\frac{s_k-\ell}{\sqrt{t}}\right)\right\}$$

is the shaded area under the curve. The value of $F_u(t)$ at $t = 64$ is computed as 2 times the shaded area. As $k \to \infty$, the terms of the series approach 0 very quickly, and in fact only the first 6 terms are larger than 10e-5 in absolute value. Since the terms of the series are strictly decreasing in absolute value as $k \to \infty$, any computer programs written to evaluate this function can easily monitor the size of the terms, and cease looping over the summation when some predefined limit is reached. Also, from well-known results for alternating series, we know that the size of the error incurred by truncating the series after n terms is less in absolute value than the $(n+1)^{st}$ term. Again this fact can be used to determine when enough terms of the series, written in alternating series form, have been evaluated. A similar check can be made for the first and second derivatives.

As we show in Chapter 4, $F_u(t)$ can be differentiated term-by-term for finite t, giving the corresponding subdensity,

$$f_u(t) = -t^{-\frac{1}{2}}\sum_{k=0}^{\infty}\left\{\phi\left(\frac{s_k+\ell}{\sqrt{t}}\right)(s_k+\ell) - \phi\left(\frac{s_k-\ell}{\sqrt{t}}\right)(s_k-\ell)\right\},$$

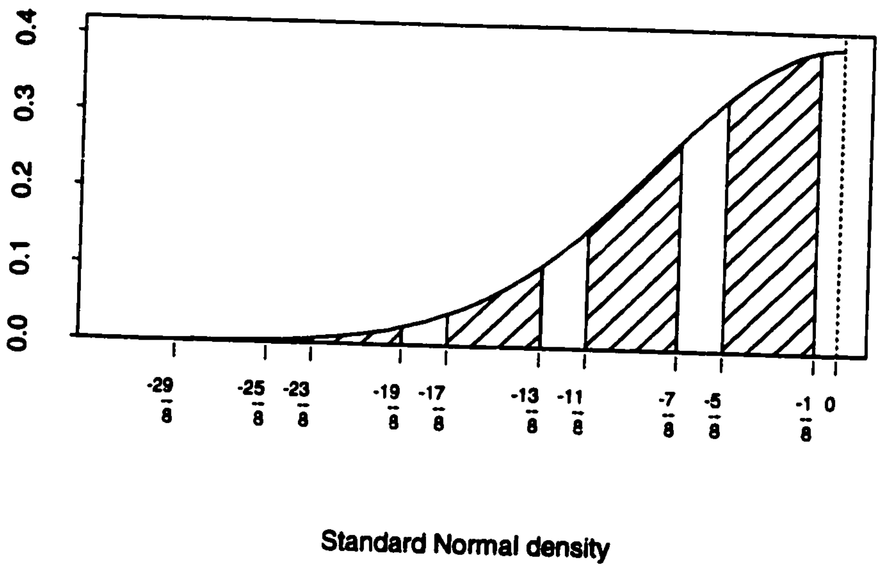


Figure 2.2: $F_u(64)$ is 2 x shaded area under curve

where $\phi(x)$ is the standard normal density evaluated at x. It turns out that we can write this more compactly as

$$f_u(t) = -t^{-\frac{3}{2}} \sum_{k=-\infty}^{\infty} \phi\left(\frac{s_k + \ell}{\sqrt{t}}\right)(s_k + \ell). \tag{2.3}$$

There are many other ways of deriving $f_u(t)$. For instance, in section 3.1, we give the corresponding subdensity, $f_u^{\mu}(t)$, for the case where the health level process is a Wiener process with drift $\mu \neq 0$. This subdensity can be derived from difference equation arguments as outlined in Feller [17] and in appendix C of this thesis. Then letting $\mu \to 0$ in the expression for $f_u^{\mu}(t)$, we get $f_u(t)$.

Similarly we find that the subdensity corresponding to $F_\ell(t)$ is

$$f_\ell(t) = -t^{-\frac{3}{2}} \sum_{k=0}^{\infty} \left\{ \phi\left(\frac{s_k + u}{\sqrt{t}}\right)(s_k + u) - \phi\left(\frac{s_k - u}{\sqrt{t}}\right)(s_k - u) \right\}, \tag{2.4}$$

Figure 2.3: lower subdensity, $f_l(t)$

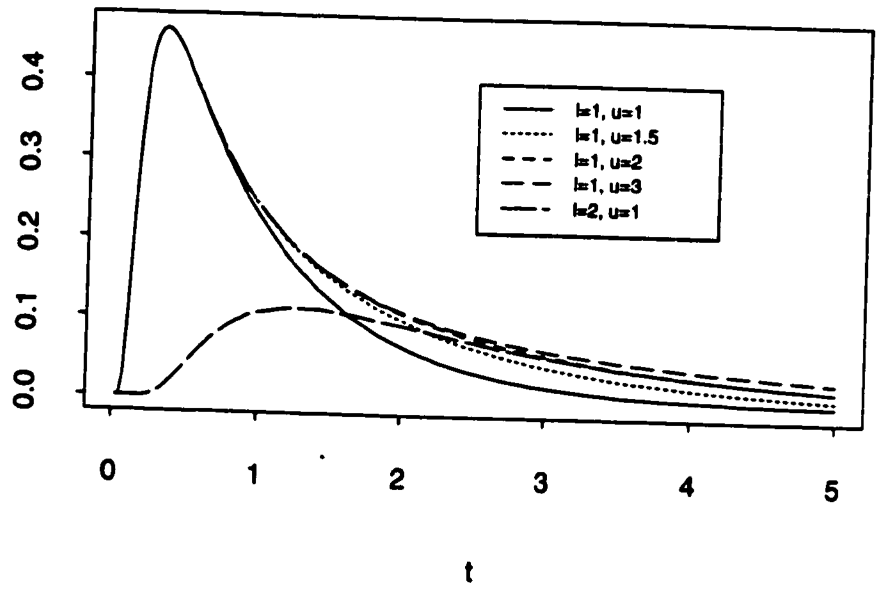
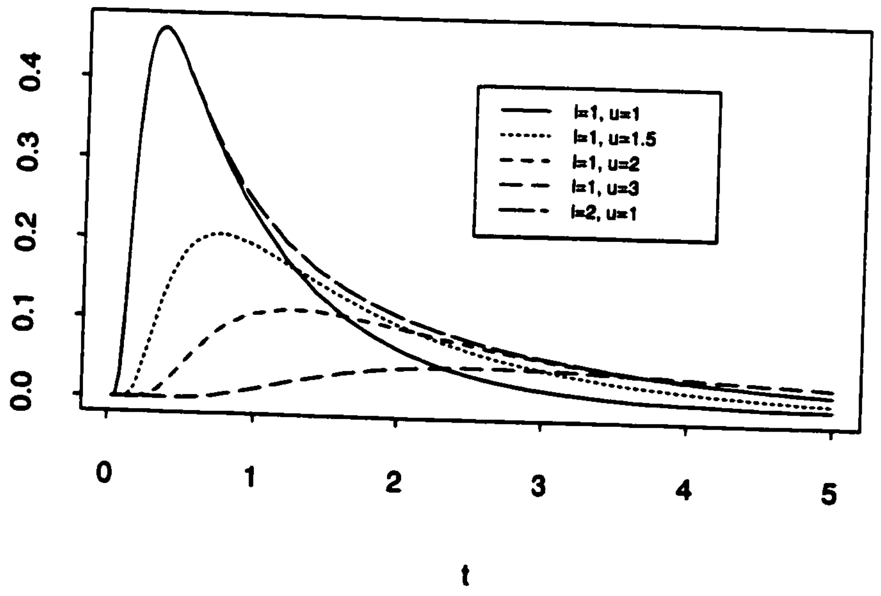


Figure 2.4: upper subdensity, $f_u(t)$



which can also be expressed as

$$f_\ell(t) = -t^{-\frac{3}{2}} \sum_{k=-\infty}^{\infty} \phi\left(\frac{s_k+u}{\sqrt{t}}\right)(s_k+u).$$

Note that once again, we can get $f_\ell(t)$ by reversing the roles of u and ℓ in $f_u(t)$. The two functions $f_u(t)$ and $f_\ell(t)$ are referred to as subdensities because each integrates to something less than 1. Figures 2.3 and 2.4 show plots of the subdensities for various values of the parameters.

Figures 2.5 and 2.6 show the upper and lower subdensities for a given set of parameters plotted on the same plot. Of course when $u = \ell$, the subdensities are identical. If $\ell < u$, then the lower subcdf dominates the upper, since the probability of travelling the shorter distance ℓ by any given time t is larger than the probability of travelling a distance $u > \ell$. This fact is also apparent from examination of the expressions for the subcdf's in 2.1 and 2.2, since, if $u > \ell$, $F_u(t) < F_\ell(t)$ for all t. This imparts a lack of flexibility to the model, in contrast with the drift model discussed in Chapter 3. There, as we shall see, the addition of an extra parameter allows a wider variety of relationships between the two subdensities.

The density of T, the time the process reaches one of the two barriers, is $f(t; u, \ell) = f_u(t)+f_\ell(t)$. A plot of the density for various values of the parameters u and ℓ is shown in Figure 2.7. Note that $f(t; u, \ell) = f(t; \ell, u)$, so in fact we show the density only for $u = 1$ and several values of ℓ.

If we let the lower barrier approach negative infinity, we return to the situation of a single barrier. When the underlying process is a Wiener process with zero drift and volatility σ^2, the distribution of the first passage time to a single barrier at u is the stable law with index 1/2. The density is given by

$$f(t) = \frac{\lambda^{1/2}}{\sqrt{2\pi t^3}} \exp\left(\frac{-\lambda}{2t}\right)$$

where $\lambda = u^2/\sigma^2$. This density, with $u = 1$ and $\sigma^2 = 1$ is also shown in figure 2.7 as the heavy line. This is of course the limiting form of $f(t; 1, \ell)$ as $\ell \to \infty$. Note that the mean and variance of this distribution are infinite. The stable law distributions are discussed briefly in appendix B.

Figure 2.5: subdensities, $u = 1$, $\ell = 2$

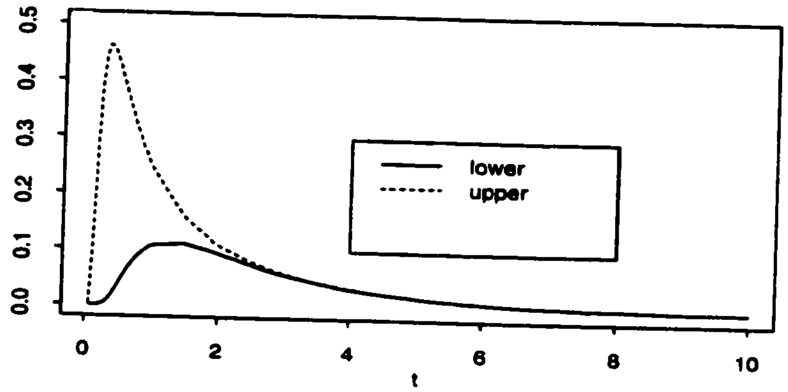


Figure 2.6: subdensities, $u = 2$, $\ell = 1$

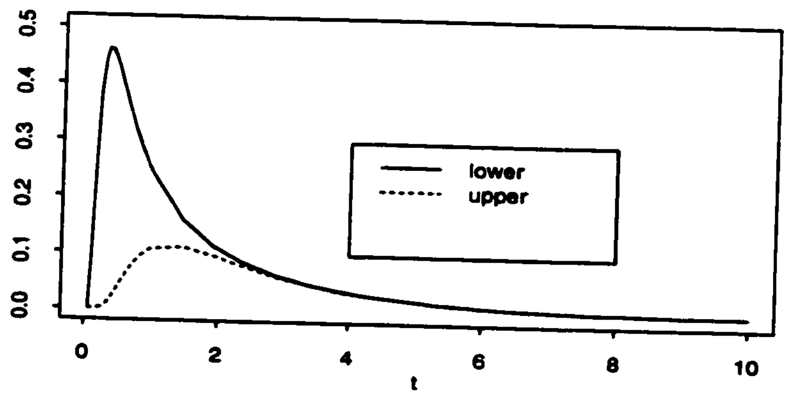
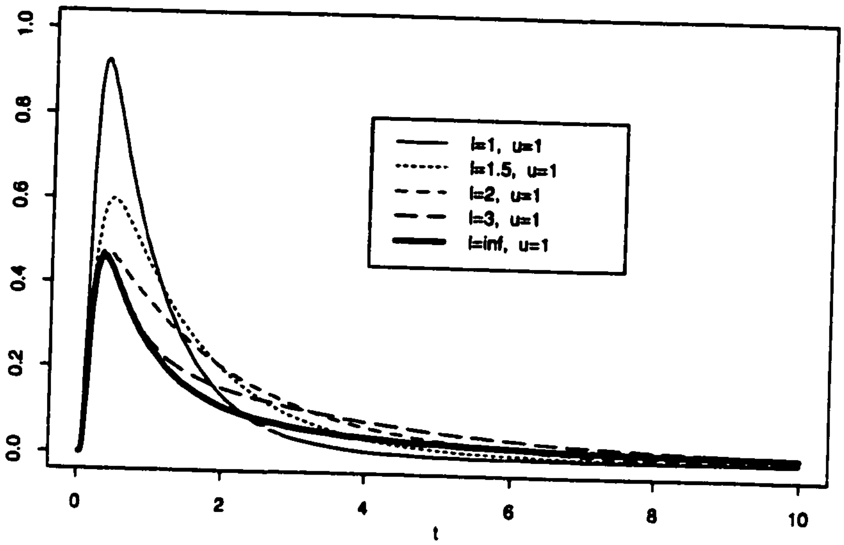


Figure 2.7: density function, $f(t; u, \ell)$



In survival modelling, one often examines the hazard function,

$$h(t) = \lim_{\Delta t \to 0+} \frac{P(t \le T < t + \Delta t | T \ge t)}{\Delta t} = \frac{f(t)}{\mathcal{F}(t)},$$

where $\mathcal{F}(t) = 1 - F(t)$ is the survivor function. Figure 2.8 shows the hazard function for $\ell = 1$ and several values of u. Apparently, the hazard function can be either unimodal or non-decreasing. It evidently approaches a constant as $t \to \infty$, and so for large values of t, it behaves like the exponential distribution.

Finally, we note that this distribution has non-proportional hazards, i.e. there are no two values of $\theta = (u, \ell)$ and no real number α, such that $h(t; \theta_1) = \alpha\ h(t; \theta_2)$. This is illustrated graphically in Figure 2.9, where the log cumulative hazard is plotted against log time. The cumulative hazard $\mathcal{H}(t)$ equals

$$\mathcal{H}(t) = \int_0^t h(s)\, ds = -\log \mathcal{F}(t).$$

If two hazard functions $h_1(t)$ and $h_2(t)$ are proportional, then so are the cumulative hazards,

Figure 2.8: hazard function, $h(t; u, \ell)$

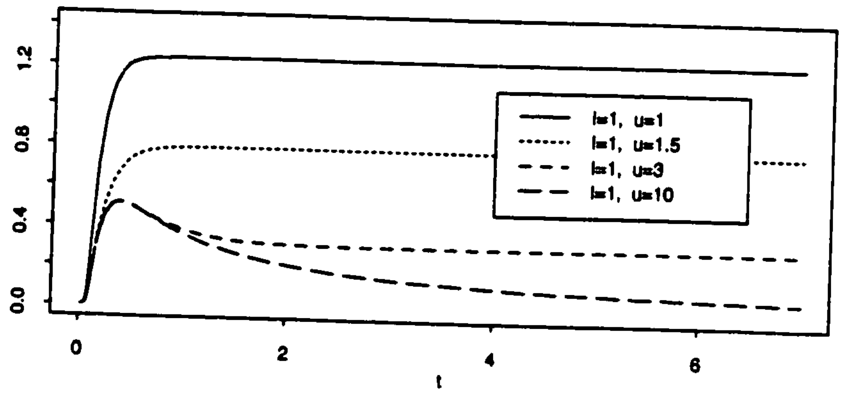
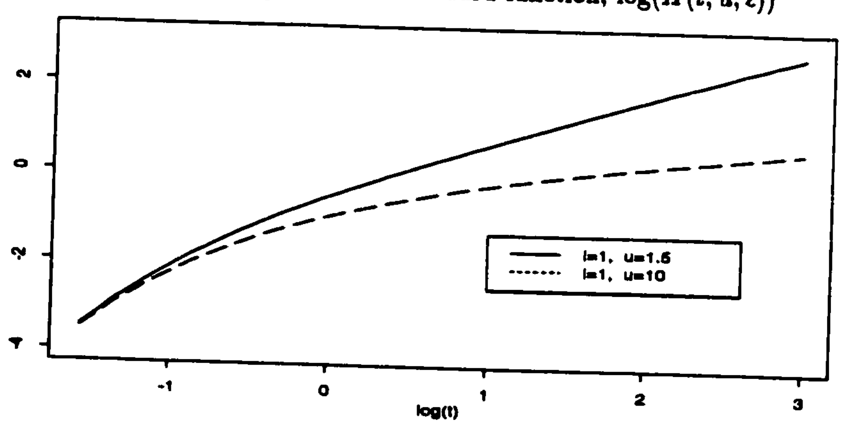


Figure 2.9: log cumulative hazard function, $\log(H(t; u, \ell))$



$\mathcal{H}_1(t)$ and $\mathcal{H}_2(t)$ so that

$$\log(\mathcal{H}_1(t)) = \log(\alpha) + \log(\mathcal{H}_2(t)).$$

Thus a plot of the log of the cumulative hazards against time (or some function thereof) should give parallel curves. Figure 2.9 shows two cumulative hazards for different values of u plotted against log time, and clearly the two curves are not parallel.

For convenience we record here the expected value of T, the marginal distribution of D, and the conditional expectation of T given D, for the simple zero drift model. These quantities will be derived more generally in section 3.2 when we discuss a process with non-zero drift. It will be shown that, for zero drift,

$$\begin{aligned} E(T) &= \quad m_T = \ell u \\ P(D = \mathrm{u}) &= \quad p_D = \ell/(u+\ell) \\ P(D = \mathrm{l}) &= 1 - p_D = u/(u+\ell) \\ E(T|D=\mathrm{u}) &= m_{T/u} = \frac{u(u+2\ell)}{3} \\ E(T|D=\mathrm{l}) &= m_{T/\ell} = \frac{\ell(2u+\ell)}{3}. \end{aligned}$$

Once again, we can get $P(D = \mathrm{l})$ by reversing the roles of u and ℓ in $P(D = \mathrm{u})$. Also, we can get $E(T|D = \mathrm{l})$ by reversing the roles of u and ℓ in $E(T|D = \mathrm{u})$. Finally note that the conditional density of T, given $D = \mathrm{u}$) is

$$\lim_{\Delta t \to 0+} \frac{P(t \le T < t + \Delta t | D = \mathrm{u})}{\Delta t} = \frac{f_u(t)}{P(D = \mathrm{u})}$$

2.2 Maximum Likelihood Estimation

When there are no covariates, individuals make different likelihood contributions according to their ultimate destination, as follows. Individuals who are discharged at time t, ie for whom $[T = t, D = \mathrm{u}]$, make a likelihood contribution, $f_u(t)$, proportional to the probability that the process hits the upper barrier at u for the first time in $[t, t + dt)$ and does so without first going through the lower barrier. For individuals who die at time t, we have $[T = t, D = \mathrm{l}]$. These individuals make a likelihood contribution, $f_\ell(t)$, proportional to the probability that the process hits the lower barrier at $-\ell$ for the first time in $[t, t+dt)$ and does so without first going through u.

It is possible that the data will include individuals whose outcome is not known at the end of the study period. For these censored observations, we know only that the time of the process hits either barrier is greater than the time on study, t. If the censoring mechanism can be considered independent of the health level process, likelihood construction is relatively straightforward. These individuals contribute a term proportional to the probability that T is greater than t, namely $\mathcal{F}(t)$.

Thus for the zero drift model with volatility=1 and no covariates, the log-likelihood is

$$\mathcal{L}(u,\ell;t) = \sum_{\{i:D_i=u\}} \log f_u(t_i) + \sum_{\{i:D_i=\ell\}} \log f_\ell(t_i) + \sum_{\{i:D_i=c\}} \log \mathcal{F}(t_i)$$

where t_i is the time of death or discharge for individual i, and D_i is the discharge status of individual i where now D_i can take values u, ℓ or c depending on whether individual i is discharged, dies or is censored at time t_i.

As is well known, for independent and identically distributed (iid) observations from a distribution belonging to a parametric family of distributions which satisfies appropriate regularity conditions, the maximum likelihood estimator (mle) is consistent, asymptotically efficient, and asymptotically normal. We discuss regularity conditions for the LOS model in Chapter 5.

In the multi-parameter case, let $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p)$ represent the vector of maximum likelihood estimates, let $\mathcal{L}(\theta)$ represent the log likelihood function, and let $I(\hat{\theta})$ represent the observed information, that is the matrix $-\partial^2\mathcal{L}(\theta)/\partial\theta^2$, evaluated at $\hat{\theta}$. If the multi-parameter regularity conditions are satisfied, then $\hat{\theta}$ converges in distribution to a normal distribution with mean vector θ and covariance matrix which can be estimated by $I^{-1}(\hat{\theta})$. This fact can be used to construct confidence intervals for components θ_j as follows: Let v_j be the j^{th} diagonal element of $I^{-1}(\hat{\theta})$ and z_α be the α quantile of the normal distribution. Then a symmetric two-sided $(1-\alpha)\%$ confidence interval for θ_j is $\hat{\theta}_j \pm z_{\alpha/2}\sqrt{v_j}$. This will be referred to in the examples as a normal-theory confidence interval, since it relies on the assumption of asymptotic normality.

Values of u and ℓ that maximize the likelihood, or equivalently the log likelihood, can be found via the Newton Raphson method, or some other optimization routine. Both Splus and SAS have convenient, built-in optimization routines. These routines can also perform constrained optimization, which is relevant here since both u and ℓ must be positive. For the examples tried so far, it has not been necessary to use the constrained optimization routine, as the constraints have been automatically satisfied by the mle's.

Although the routine can run without derivatives, it works much faster if the user supplies a routine for calculating the gradient, i.e. the vector of first derivatives of the log likelihood. In order to get the value of the observed information, which is necessary for inference, the user must supply a routine for calculating the hessian, the matrix of second derivatives of the log likelihood. Derivation of first and second derivatives is straightforward, and the results are given in appendix E.

As mentioned at the beginning of this chapter, we set the volatility parameter $\sigma^2 = 1$, since it is merely a scaling factor. Also as we show in appendix F, it is not possible to simultaneously estimate u, ℓ and σ^2 due to identifiability problems.

2.2.1 Non-linear Functions of the Parameters

Since the health level process is not directly observable, the values of u and ℓ are not readily interpretable. But expected time in hospital and the probability that the hospital stay ends in healthy discharge are quantities of direct interest to the researcher. As we saw in section 2.1, $E(T) = m_T = \ell u$ and $P(D = \mathrm{u}) = p_D = \ell/(u + \ell)$, both of which are non-linear functions of the parameters. To estimate these quantities, we replace u in the above expressions by its maximum likelihood estimate $\hat{u}$, and ℓ by its maximum likelihood estimate $\hat{\ell}$. Because of the invariance of the likelihood under parameter transformation, this gives us respectively the maximum likelihood estimate $\hat{m}_T$ of $E(T)$ and the maximum likelihood estimate $\hat{p}_D$ of $P(D = \mathrm{u})$.

The conditional expected length of stay given outcome will be a useful quantity for informal model assessment, as will be discussed in section 2.3. At first glance, this quantity may not seem to be of much use to planners, since outcome will be unknown when an individual is admitted to hospital. However, depending on the medical condition under study, individuals who ultimately die in hospital may need more expensive daily care than those who are eventually discharged.

Thus a researcher may want to predict the proportion of patients in a cohort who are expected to die, and also predict their length of stay, and multiply this by per diem expenses. As we saw in section 2.1, $E(T|D = \mathrm{u}) = m_{T/u} = \frac{u(u+2\ell)}{3}$ and $E(T|D = \mathrm{l}) = m_{T/\ell} = \frac{\ell(2u+\ell)}{3}$. By the invariance property of maximum likelihood, the maximum likelihood estimates of these quantities are found by replacing u by $\hat{u}$ and ℓ by $\hat{\ell}$.

Of course we would also like confidence intervals for these quantities. For instance, researchers might want to assess whether expected time in hospital is significantly different for two different values of a covariate. Variance estimators and confidence intervals for non-linear functions of the parameters are difficult to derive. For this reason, we will use the bootstrap to construct confidence intervals, which will be illustrated in section 3.5.

2.2.2 Covariates

As discussed in section 1.3, we may allow the barrier levels to vary for different individuals or groups of individuals. Let u_i and ℓ_i denote the upper and lower barriers respectively for individual i. Let

$$\underline{x}_i^u = (x_{i0}^u, x_{i1}^u, \ldots, x_{ip_u}^u)'$$

be the $((p_u + 1) \times 1)$ vector of covariates measured on individual i which are thought to affect the upper barrier, and

$$\underline{x}_i^\ell = (x_{i0}^\ell, x_{i1}^\ell, \ldots, x_{ip_\ell}^\ell)'$$

be the $((p_\ell + 1) \times 1)$ vector of covariates measured on individual i which are thought to affect the lower barrier. Then we may model each barrier level as a function of the relevant covariates, by specifying

$$\begin{aligned} u_i &= g(\underline{\beta}^u, \underline{x}_i^u) \\ \ell_i &= h(\underline{\beta}^\ell, \underline{x}_i^\ell) \end{aligned}$$

where $\underline{\beta}^u$ and $\underline{\beta}^\ell$ are vectors of parameters to be estimated. We may want to choose functions $g(.)$ and $h(.)$ that automatically satisfy the constraints $u_i > 0$ and $\ell_i > 0$; for instance we might take $g(\underline{\lambda}, \underline{x}) = \exp(\underline{\lambda}' * \underline{x})$, where $*$ denotes matrix multiplication.

For the purposes of these preliminary investigations, we take both $g(.)$ and $h(.)$ to be the identity function, so that each barrier level depends on the covariates in a linear fashion:

$$u_i = \underline{\beta}^u * \underline{x}_i^u = \beta_0^u\ x_{i0}^u + \beta_1^u\ x_{i1}^u + \ldots + \beta_{p_u}^u\ x_{ip_u}^u$$
$$\ell_i = \underline{\beta}^\ell * \underline{x}_i^\ell = \beta_0^\ell\ x_{i0}^\ell + \beta_1^\ell\ x_{i1}^\ell + \ldots + \beta_{p_\ell}^\ell\ x_{ip_\ell}^\ell.$$

In the examples shown in this thesis, we will use standard unconstrained optimization routines to find the maximum likelihood estimates of the parameters. Extension of the log-likelihood, gradient and hessian to accommodate covariates is straightforward and is shown in appendix E.

2.2.3 Starting Values

The routines used to find the maximum likelihood estimates are iterative, and it is imperative to have good starting values for the parameters. For a model with no covariates, the following heuristic approach works well in practice. We know that for this model $E(T) = \ell u$ and $P(D = \mathrm{u}) = \ell/(u+\ell)$. We can estimate $E(T)$ by $\bar{t}$, the sample average LOS. We can estimate $P(D = \mathrm{u})$ by the observed proportion of individuals in the sample who were ultimately discharged, which we denote as $\tilde{p}_u$. Then, approximately, $\bar{t} = \ell u$ and $\tilde{p}_u = \ell/(u+\ell)$. Solving these two equations for ℓ and u gives

$$u = \sqrt{\bar{t}\left(\frac{1-\tilde{p}_u}{\tilde{p}_u}\right)}$$
$$\ell = \sqrt{\frac{\bar{t}\,\tilde{p}_u}{1-\tilde{p}_u}}$$

These rough estimates can be used as starting values for u and ℓ.

For starting values of the regression parameters β in models with covariates, we use a sort of mock least squares estimate, which we now describe. A linear least squares model for time in hospital would take the form

$$E(T) = \mu_T = X\beta,$$

and the least squares estimate of β would be

$$\tilde{\beta} = (X'X)^{-1}(X'T).$$

In the models developed here, we specify that the upper barrier is a linear function of covariates, i.e. $u = X\beta$, so that $\beta = (X'X)^{-1}(X'u)$. Of course, u is unobservable. So, to find starting values, we might replace u in the above expression with a function of T, $g(T)$ say, which behaves approximately like u. Since $(X'X)^{-1}(X'\ g(T))$ has the form of a least squares estimate, we refer to it as a mock least squares estimate.

In practice, it seems that to estimate u, we should use only individuals in the data set whose paths ended at the upper barrier, i.e. who were discharged. So to estimate starting values, we divide the data set into two groups, according to discharge status. Let T_u be the vector of times in hospital for individuals who were ultimately discharged, and let X_u be the matrix formed by stacking the row vectors $(\underline{x}_i^u)'$ belonging to those individuals who were ultimately discharged. Then the starting values for the vector of parameters $\underline{\beta}^u = (\beta_0^u, \beta_1^u, \ldots, \beta_{p_u}^u)'$ could be calculated as $(X_u'X_u)^{-1}(X_u'\ g(T_u))$. To find a suitable function $g(T_u)$, recall that $E(T|D = \mathrm{u}) = m_{T/u} = \frac{u(u+2\ell)}{3}$ and $P(D = \mathrm{u}) = p_D = \ell/(u + \ell)$. So then, approximately,

$$u = \sqrt{3m_{T/u}\left(\frac{1 - p_D}{1 + p_D}\right)}.$$

Thus a reasonable candidate for the function $g(T_u)$ is

$$g(T_u) = \sqrt{3T_u\left(\frac{1 - \tilde{p}_u}{1 + \tilde{p}_u}\right)}.$$

Similarly, let T_ℓ be the vector of times in hospital for individuals who died in hospital, and let X_ℓ be the matrix formed by stacking the row vectors $(\underline{x}_i^\ell)'$ for those individuals who died in hospital. Then starting values for $\underline{\beta}^\ell = (\beta_0^\ell, \beta_1^\ell, \ldots, \beta_{p_\ell}^\ell)'$ can be calculated as $(X_\ell'X_\ell)^{-1}(X_\ell'\ h(T_\ell))$, where

$$h(T_\ell) = \sqrt{3T_\ell\left(\frac{\tilde{p}_u}{1+\tilde{p}_u}\right)}.$$

2.3 Assessment of Model Fit

Once a model has been chosen and the maximum likelihood estimates calculated, it is important to assess the fit of the model to the data. We first discuss the scenario where the model involves a small number of discrete covariates, each with a small number of levels, so that the number of distinct combinations of covariate levels or "covariate cells" in the data is small. Then an informal assessment of fit can be made by comparing the average length of stay in each covariate cell to that predicted by the model. These results can be displayed in a table.

The model predictions we use are the maximum likelihood estimates of the expected time in hospital at the appropriate combination of covariate values. Recall that, for individual i, the expected time in hospital m_{T_i} is a simple function of the barrier levels u_i and ℓ_i, and in turn u_i is a function of covariates $\underline{x}_i^u$ and regression parameters $\underline{\beta}^u$, and ℓ_i is a function of covariates $\underline{x}_i^\ell$ and regression parameters $\underline{\beta}^\ell$. Thus

$$m_{T_i} = m_T(\underline{\beta}^u, \underline{\beta}^\ell, \underline{x}_i^u, \underline{x}_i^\ell)$$

and the mle of m_{T_i} which we will use for predictions, is

$$\hat{m}_{T_i} = m_T(\hat{\underline{\beta}}^u, \hat{\underline{\beta}}^\ell, \underline{x}_i^u, \underline{x}_i^\ell).$$

We can also examine observed versus predicted LOS in each covariate by discharge cell. In this case, the necessary predicted values will be

$$\hat{m}_{T/u_i} = m_{T/u}(\hat{\underline{\beta}}^u, \hat{\underline{\beta}}^\ell, \underline{x}_i^u, \underline{x}_i^\ell)$$

and

$$\hat{m}_{T/\ell_i} = m_{T/\ell}(\hat{\underline{\beta}}^u, \hat{\underline{\beta}}^\ell, \underline{x}_i^u, \underline{x}_i^\ell).$$

We can also compare the observed and predicted proportions of individuals who are discharged in each covariate cell. For predicted proportions, we will use

$$\hat{p}_{D_i} = p_D(\underline{\hat{\beta}}^u, \underline{\hat{\beta}}^\ell, \underline{x}_i^u, \underline{x}_i^\ell).$$

These methods are illustrated in detail in section 2.4. If the covariates are continuous, we can still perform this kind of analysis, after dividing the range of each covariate into a small number of non-overlapping intervals.

When a moderate to large number of distinct covariate cells is present in the data, it is best to use graphical means of assessing model fit. We will construct the i^{th} "raw" residual r_i by taking the i^{th} observation in the data set, $(t_i, \underline{x}_i)$, and subtracting the length of stay predicted by the model at these covariate values, from the observed length of stay, giving

$$r_i = t_i - m_T(\underline{\hat{\beta}}^u, \underline{\hat{\beta}}^\ell, \underline{x}_i^u, \underline{x}_i^\ell).$$

The raw residuals can then be plotted against observation number and against each covariate in the data set. These residuals measure the distance between observed and predicted values. Thus large values indicate observations that might warrant further investigation, and systematic patterns in the plots can help to point out some inadequacy in the model specification.

To assess whether the FP2B distribution provides an adequate overall fit to the model, we want to examine residuals with a known distribution, and thus another kind of residual is called for. A general definition of residuals, useful for models which do not belong to the location-scale family of distributions, was given by Cox and Snell [12] in 1968. Suppose that we can express the response for individual i, t_i, as a function of a vector of unknown parameters $\underline{\beta}$, a vector of covariates $\underline{x}_i$, and some quantity ϵ_i where $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are independent and identically distributed (iid) from some well-known distribution. If we can then write ϵ_i as

$$\epsilon_i = g(t_i, \underline{x}_i, \underline{\beta})$$

for some function $g(.)$, then we can define general residuals R_i as

$$R_i = g(t_i, \underline{x}_i, \hat{\underline{\beta}})$$

where $\hat{\underline{\beta}}$ is the mle of $\underline{\beta}$.

The residuals, R_i, will be neither independent, nor identically distributed. However Cox and Snell argue that when the number of parameters is small in relation to the number of observations, the distribution of the residual R_i should not be too far from the distribution of ϵ_i, and at any rate, this approximation will be good enough for graphical assessment. On the other hand, formal tests of goodness of fit based on residuals will probably require more careful methods. For instance, Cox and Snell give some first-order corrections which can be made to the residuals to bring their distribution closer to that of $\underline{\epsilon}$.

If T is a random variable from a distribution with cdf $F(t)$, then both the transformed random variables $Z = F(T)$ and $U = 1 - F(T) = \mathcal{F}(T)$ have a uniform (0,1) distribution. Furthermore, the transformed random variable, $-\log \mathcal{F}(T)$, has a unit exponential distribution.

Suppose that we have a parametric family of distributions with survivor functions $\mathcal{F}(t, \theta)$, and that we fit a regression model with $\theta_i = h(\underline{x}_i, \underline{\beta})$ for some function h. Then we can use $\mathcal{F}(t_i, \underline{x}_i, \underline{\beta})$ (or some transformation thereof) as the required ϵ_i, since they are iid with a uniform (0,1) distribution. This is especially convenient in survival models where the survivor function, $\mathcal{F}(t, \underline{\theta})$, is needed for calculating the contribution of censored individuals to the likelihood, so algorithms for calculating it will already be available. The Cox-Snell residual R_i is defined as $-\log \mathcal{F}(t_i, \underline{x}_i, \hat{\underline{\beta}})$. If the data do come from the distribution with survivor function $\mathcal{F}(t_i, \underline{x}_i, \underline{\beta})$, then the Cox-Snell residuals should have approximately an exponential(1) distribution.

We can use this fact to check graphically that the model developed in this chapter gives a suitable fit to the data under investigation. A quantile plot, with the ordered Cox-Snell residuals plotted against the order statistics of a unit exponential distribution, should give approximately a straight line with unit slope if the fit of the model is adequate. These plots give a graphical measure of overall goodness of fit.

Unlike residuals from normal linear models, the Cox-Snell residuals used in survival analysis do not correspond to a distance. Thus the suggested plots can be hard to interpret. Note that, in the absence of covariates, large survival times will have small values of $\mathcal{F}(t, \theta)$, and so the corresponding residuals will be large. Of course, when covariates are present, the relationship is more complicated. Observations with survival times corresponding to small estimated probabilities of survival will have large residual values. Thus, observations with large values of Cox-Snell residuals will correspond to survival times that are unusually long, given the covariates.

Recently, concern has been expressed in the literature that the Cox-Snell residuals are too lenient for checking distributional assumptions. Baltazar-Aban and Pena [3] show that when the data is from an exponential distribution with no censoring, and an exponential model is fit to the data via maximum likelihood estimation, the ordered Cox-Snell residuals will have a smaller variance than the order statistics of an exponential distribution. Thus the quantile plot will be closer to a straight line, on average, than we would expect if in fact the Cox-Snell residuals did have an exact exponential distribution. They also give some simulation results which indicate that this phenomenon also occurs when the data come from the hypothesized distribution and that distribution is Weibull. The more relevant question of the behaviour of the Cox Snell residuals when the data do not come from the hypothesized distribution is only indirectly addressed in the article. This is an area which needs further research.

If the data includes censored observations, then the corresponding residuals will also be censored. Lawless [25] discusses residuals for censored data. Gentleman and Crowley [19] discuss graphical methods for censored data. They recommend that censored observations be presented on plots using the same plotting symbol as uncensored observations, but with a lighter color value. If this is done, censored observations will be perceived as having somewhat less importance than uncensored observations. They also discuss boxplots and smoothing techniques for censored observations.

2.4 Example

In this section we demonstrate the model on a small subset of the Utah data. This subset consists of females over 80 years old with a recorded APR-DRG of 121 (Circulatory Disorders with Myocardial Infarction), who either died in hospital (n=32) or were discharged home (n=102). On average, members of this subset spent 4.582 days in hospital (median 4 days). As shown in

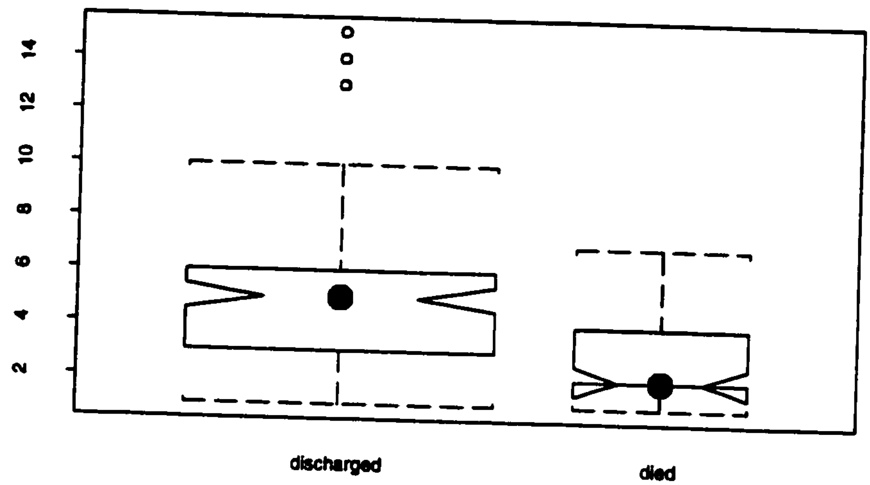


Figure 2.10: LOS by Destination, Females > 80 years, APR-DRG=121

figure 2.10, there is clearly a difference in LOS by discharge status. The mean length of stay in the discharge group is 5.147 days (median 5 days), while in the group who died in hospital, the mean is 2.781 days (median 2 days). However, recall that discharge status is part of the response, and so it would be incorrect to include it as a covariate in the model.

In what follows, we will treat the variable severity as if it represented comorbidities present at admission to hospital (see the discussion in section 1.2). Because of sparsity in the data, we have recoded the covariate severity to three levels (from its original 4 levels). Severity=2 represents no, minor or moderate complication or co-morbidity; Severity=3 represents major complication or co-morbidity; Severity=4 represents extreme complication or co-morbidity. In figure 2.11 we show side-by-side boxplots for each severity by outcome group. Note that the width of each box is an indication of the number of individuals in the data set with the given level of severity and discharge status. The effect of severity is to increase time to discharge, but to decrease time to death, although the effect is less pronounced. On the boxplot the circles indicate the mean LOS

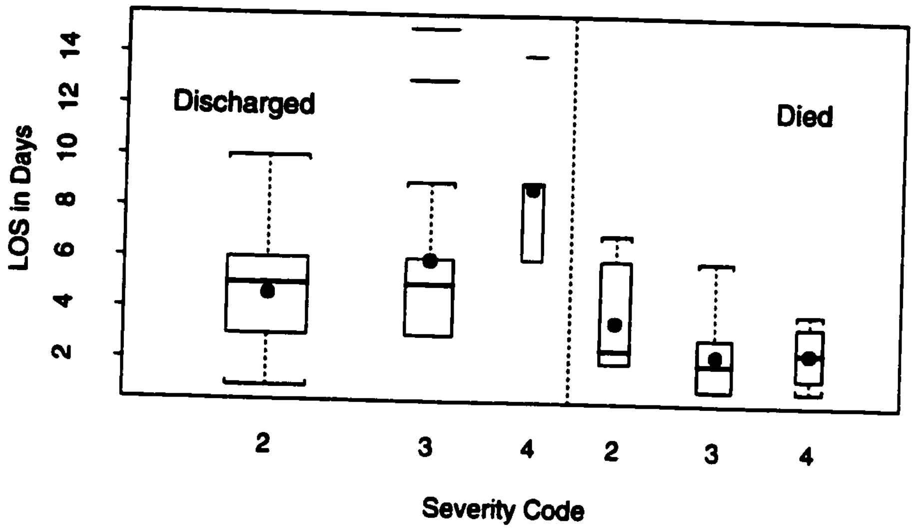


Figure 2.11: LOS by Severity and Outcome

in each group, while the horizontal lines represent the median.

Recall that in the development of the model, we assumed that all individuals in the data set had roughly the same health level upon admission to hospital, which we arbitrarily set to 0. In other words, letting $\{H_i(t), t > 0\}$ represent the health level process for individual i, we have so far assumed that $H_i(0) = 0$ for all individuals. However, individuals do enter hospital with differing severity of disease, so that it is more realistic to let $H_i(0)$ vary across individuals, and to model $H_i(0)$ as a function of covariates. Thus it would appear that we should develop a new and more complicated model. However, as we now show, the existing model can be parameterized to allow for differing health levels at admission to hospital.

Let us make the arbitrary assumption that individuals with low severity enter hospital with a health level of 0. Patient i, who is admitted with higher severity of disease, would then enter hospital with a health level less than 0, say $-h_i$ (where $h_i > 0$). In effect, entering hospital with a high level of severity increases the distance to the upper barrier by h_i, and decreases the distance to the lower barrier by h_i, as compared with someone who enters hospital with a low level of severity. As we show in appendix B, allowing the process to start at $-h_i$ with upper barrier $u_i = b$ and lower barrier $-\ell_i = -a$, (where $a, b, h_i > 0$), is equivalent to starting the process at 0 with barriers at $u_i = b + h_i$ and $-\ell_i = -a + h_i = -(a - h_i)$, in the sense that the time until the process first reaches one of the two barriers has the same distribution function (and subdistribution functions) in each case. Thus we can use our existing model which assumes that all individuals enter hospital with health level 0. But for individual i who enters hospital with increased severity of disease, we allow both the upper and lower barriers to be raised by the same amount, h_i, where h_i will be modeled as a function of covariates, and estimated.

More specifically, to model the effect of severity on the health level of individual i at admission to hospital, we can let

$$\begin{aligned} u_i &= \beta_0^u + h_i \\ \ell_i &= \beta_0^\ell - h_i. \end{aligned}$$

where $h_i = \beta_1 x_i$, and $x_i = 0$ if patient i is admitted to hospital with a low level of severity and $x_i = 1$ if severity of disease is high at admission.

More generally, the model may be parameterized so that the upper and lower barriers are functions of distinct parameters.

$$\begin{aligned} u_i &= \beta_0^u + \beta_1^u x_{i1} \\ \ell_i &= \beta_0^\ell + \beta_1^\ell x_{i1}. \end{aligned}$$

This would be appropriate if severity was thought to affect the level of the barriers, as well as the starting value of the process.

We create two dummy variables to represent severity. Let $x_{i1} = 1$ if individual i exhibits a major complication or comorbidity and 0 otherwise. Let $x_{i2} = 1$ if individual i has an extreme complication or comorbidity and 0 otherwise. Individuals classified as having either no, minor, or moderate complication or comorbidity fall into the baseline category, with both x_{i1} and x_{i2} equal to 0. Then the barrier levels for individual i will be modelled as

$$u_i = \beta_0^u + \beta_1^u x_{i1} + \beta_2^u x_{i2}$$
$$\ell_i = \beta_0^\ell + \beta_1^\ell x_{i1} + \beta_2^\ell x_{i2}$$

The maximum likelihood estimates and their estimated standard errors are shown in the following table.

	β_0^u	β_1^u	β_2^u	β_0^ℓ	β_1^ℓ	β_2^ℓ
Estimate	2.010	0.623	1.639	2.705	-0.803	-1.034
Std Err	0.156	0.324	0.800	0.260	0.342	0.420

The reported standard errors are the square root of the diagonal elements of $I^{-1}(\hat{\theta})$, the inverse of the observed information matrix, evaluated at the mle. The parameter estimates indicate that the more severely ill a patient, the greater the distance to the upper barrier (discharge), and the shorter the distance to the lower barrier (death). This agrees with our intuition.

As figure 2.12 indicates, the overall fit of the model is not good. If the model is correctly specified, the ordered Cox-Snell residuals plotted against the quantiles of the exponential(1) distribution should lie on a straight line with slope 1 and intercept 0. This plot indicates that the fit is not very satisfactory, but it's difficult to look at the plot and say what exactly is going wrong. Some further exploration is required, which we now carry out.

Figure 2.13 shows the raw residuals plotted against observation number. Here we can see no systematic pattern. We see that there are four observations that have fairly large residuals, indicating that the model is underestimating length of stay for these individuals. Closer examination of the data reveals that these individuals were eventually discharged home.

In the following table we show, in each covariate cell, the observed average time in hospital (T), the observed average time for those that were discharged ($T|D = \text{u}$), and the observed average time for those who died in hospital ($T|D = \text{l}$). The last two columns show the observed proportion of patients who were discharged ($P(D = \text{u})$), and the proportion who ultimately died in hospital ($P(D = \text{l})$). The corresponding expected quantities predicted by the model are shown in brackets.

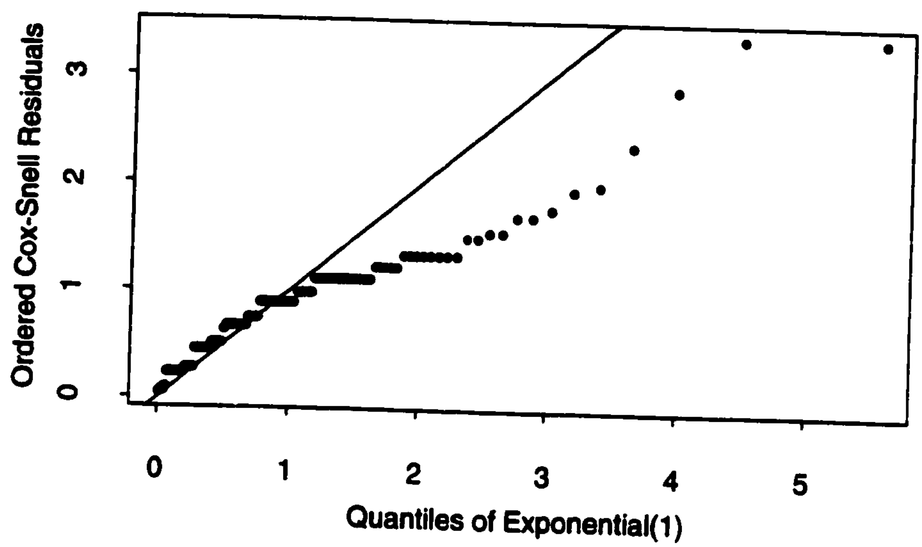


Figure 2.12: Quantile Plot of Cox Snell residuals

Severity	n	T	$T \mid D = u$	$T \mid D = l$	$P(D = u)$	$P(D = l)$
low	82	4.49(5.43)	4.61(4.97)	3.60(6.06)	0.88(0.57)	0.12(0.43)
med	39	4.67(5.01)	5.96(5.65)	2.36(4.54)	0.64(0.42)	0.36(0.58)
high	13	4.92(6.10)	8.80(8.50)	2.50(5.00)	0.38(0.31)	0.62(0.69)

Note that the observations and predictions agree in rank as severity increases. However we see that the model is systematically overestimating both probability of death and time until death in each of the three covariate cells. This lack of fit could be due to some unexplained variability in the data. However the lack of flexibility exhibited by the model with respect to the relationship between the upper and lower subdensities, is probably also partially to blame. As we shall see in section 3.5, incorporating a drift parameter allows more flexibility and results in a better fit overall.

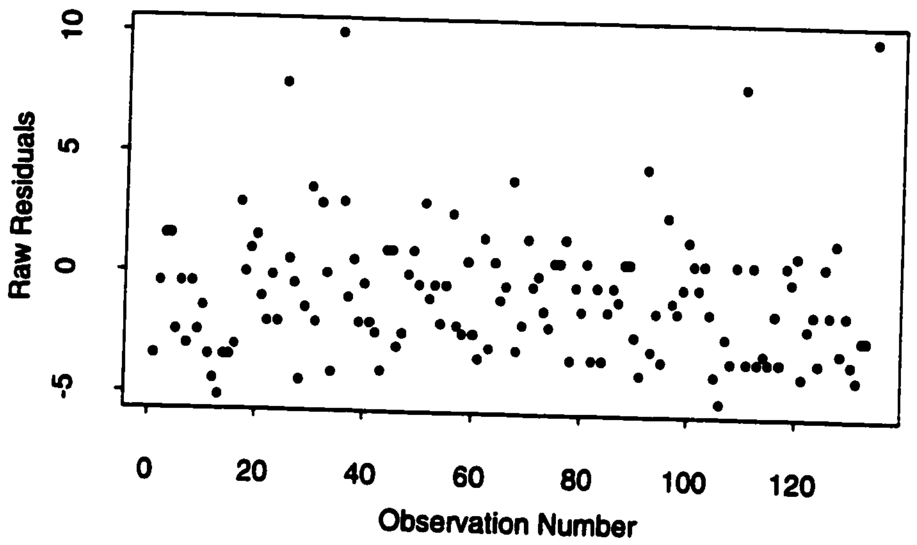


Figure 2.13: Index plot of raw residuals

Chapter 3

Double Barrier Model, Non-Zero Drift

As we saw in Chapter 2, the zero-drift model does not appear to be sufficiently flexible to provide a good fit to the data. In this chapter, we extend the model to allow for the inclusion of a drift parameter, μ.

A brief outline of the Chapter is now given. In section 3.1 we examine the distribution of T, the time of breach of one of the two barriers. We discuss derivation of the subdensities and cdf and examine the form of these functions and the hazard function. In section 3.2 we derive expressions for certain non-linear functions of the parameters that are of special interest to the researcher. In section 3.3 we compare this distribution to the inverse Gaussian distribution. In section 3.4 we discuss maximum likelihood estimation, extend the model to accommodate covariates, and describe starting values for the iterative estimation procedure. Finally in section 3.5 we demonstrate the model using some real data.

3.1 Distribution of Time until Breach

The health level process, $\{H^{\mu}(t), t > 0\}$, is now a Wiener process with drift μ different from zero and volatility $\sigma^2 = 1$. As in Chapter 2, we argue that the volatility parameter is merely a scaling factor, and so no loss of generality is entailed by setting it equal to a constant. Furthermore, as shown in appendix F, we run into identifiability problems if we try to estimate σ^2 in addition to the barrier and drift parameters. As before we take $H(0) = 0$ and postulate the existence of an upper barrier at u and a lower barrier at $-\ell$. We need to find the distribution of T, the time the process first reaches one of the two barriers. We will call this distribution the FP2Bμ distribution, (First Passage time, Two Barrier, drift μ).

Because of a lack of symmetry introduced by the drift parameter, reflection arguments can no longer be used to derive the cdf of T, nor the subdistribution functions. There are various possible alternatives. One general strategy that we will use to find expressions for the subdensities, expectations, and probabilities, exploits the fact that a Wiener process with drift can be constructed as the limit (in a certain sense, see section C.2) of an asymmetric random walk on the integers. We will start with the random walk and set up a difference equation, the solution of which gives the desired quantity in the discrete framework. We next solve the difference equation, and then take the limit of the solution in the manner referred to above, to give the corresponding quantity in the continuous case. This passage to the limit in distribution is justified by the weak invariance principle (see eg [36]).

For example, we can set up and solve a difference equation for the probability that the random walk reaches the lower barrier in exactly n steps, without first reaching the upper barrier. We then take the limit of this solution in the sense described above. This gives an expression for $f_\ell^\mu(t)$, the subdensity corresponding to the event that a Wiener process with drift μ hits the lower barrier $-\ell$ at time t without first hitting the upper barrier u:

$$f_\ell^\mu(t) = -t^{-\frac{3}{2}} e^{\frac{-(\mu t+2\ell)\mu}{2}} \sum_{k=-\infty}^{\infty} \left\{ \phi\left(\frac{s_k+u}{\sqrt{t}}\right)(s_k+u) \right\}. \tag{3.1}$$

This procedure is outlined in detail in appendix C, section C.2, following Feller. Note that the limit as $\mu \to 0$ of this expression is $f_\ell(t)$ (as in equation 2.4), the corresponding subdensity in the zero-drift case.

The subdensity corresponding to the event that the process hits the upper barrier u at time t without first hitting the lower barrier is

$$f_u^\mu(t) = -t^{-\frac{3}{2}} e^{\frac{-(\mu t-2u)\mu}{2}} \sum_{k=-\infty}^{\infty} \left\{ \phi\left(\frac{s_k+\ell}{\sqrt{t}}\right)(s_k+\ell) \right\}. \tag{3.2}$$

Again the limit of this expression as $\mu \to 0$ is $f_u(t)$ (as given in equation 2.3), the corresponding subdensity in the no-drift case. Also note that to get $f_u^\mu(t)$, we can replace u with ℓ, ℓ with u, and μ with $-\mu$ in $f_\ell^\mu(t)$. This is explained in more detail in section C.2. Figures 3.1 and 3.2 show plots of the subdensities for $\ell = 1, u = 2$ and various values of μ. Figures 3.3, 3.4 and 3.5 show the relationship between the two subdensities for various values of the parameters. We can see that there is more flexibility in the relationship between the subdensities, than in the zero-drift case.

The density of T, the time of breach of one of the two barriers is $f(t) = f_u^\mu(t) + f_\ell^\mu(t)$. Figure 3.6 shows the density plotted against time for various values of the parameters. Note that the density may be bi-modal, if the modes of the two subdensities are well separated.

Of course, other methods of deriving the distribution functions and densities exist. Starting with the joint distribution of

$$(\max\{H(t) : 0 < t < T\},\ \min\{H(t) : 0 < t < T\},\ H(T))$$

where $\{H(t), t > 0\}$ is a zero-drift Wiener process, we can apply a Girsanov factor to obtain the corresponding joint distribution in the drift case. (see [33]). This method is particularly helpful in finding the survivor function in the drift case, $\mathcal{F}^\mu(t)$, which we now set down.

$$\mathcal{F}^\mu(t) = \sum_{k=-\infty}^{\infty} \left\{ e^{-\mu c_k} \left[\Phi\left(\frac{c_k + u - \mu t}{\sqrt{t}}\right) - \Phi\left(\frac{c_k - \ell - \mu t}{\sqrt{t}}\right) \right] \right. \tag{3.3}$$

$$\left. - e^{\mu d_k} \left[\Phi\left(\frac{-d_k + u - \mu t}{\sqrt{t}}\right) - \Phi\left(\frac{-d_k - \ell - \mu t}{\sqrt{t}}\right) \right] \right\} \tag{3.4}$$

where $c_k = 2k(u + \ell)$ and $d_k = 2k(u + \ell) + 2u$. Expressions for the subsurvivor functions are now given. The upper subsurvivor function is

$$\mathcal{F}_u^\mu(t) = \sum_{k=-\infty}^{\infty} \operatorname{sign}(\mu) \left\{ e^{-\mu c_k} \Phi\left(\operatorname{sign}(\mu)\left(\frac{c_k + u - \mu t}{\sqrt{t}}\right)\right) - e^{\mu d_k} \Phi\left(\operatorname{sign}(\mu)\left(\frac{-d_k + u - \mu t}{\sqrt{t}}\right)\right) \right\} \tag{3.5}$$

and the lower subsurvivor function is

$$\mathcal{F}_{\ell}^{\mu}(t) = \sum_{k=-\infty}^{\infty} \operatorname{sign}(\mu) \left\{ e^{\mu d_k} \Phi\left(\operatorname{sign}(\mu)\left(\frac{-d_k - \ell - \mu t}{\sqrt{t}}\right)\right) - e^{-\mu c_k} \Phi\left(\operatorname{sign}(\mu)\left(\frac{c_k - \ell - \mu t}{\sqrt{t}}\right)\right)\right\}. \tag{3.6}$$

The presence of $\operatorname{sign}(\mu)$ is necessary to ensure that the series converges.

The hazard function for various values of the parameters is shown in figure 3.7. Some interesting shapes are possible for small values of t, while for larger values the hazard appears to approach a constant value, just as in the zero-drift case. Note that this distribution can have either proportional or non-proportional hazards. Figure 3.8 shows the log cumulative hazard plotted against log time for various values of the parameters. On this scale, proportional hazards will show up as parallel curves. In fact $h(t; a, a, \mu_1) = \alpha h(t; a, a, \mu_2)$ for all $a > 0$.

Figure 3.1: $f_\ell^\mu(t)$

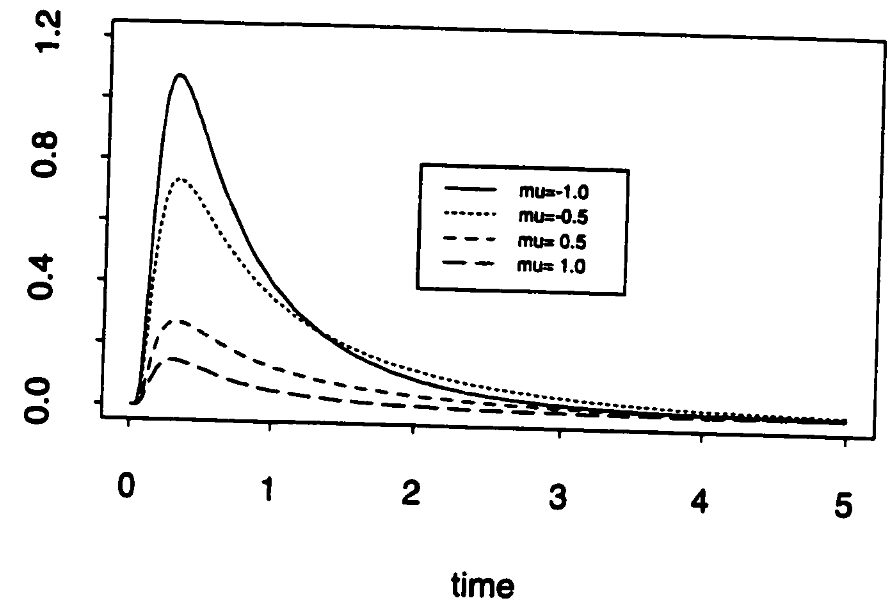


Figure 3.2: $f_u^\mu(t)$

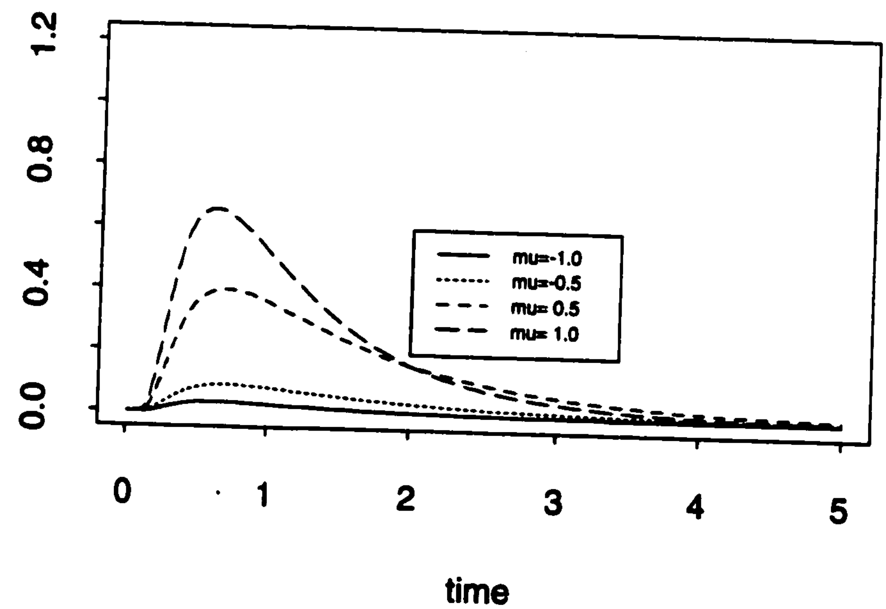


Figure 3.3: subdensities, $u = 2$, $\ell = 1$, $\mu = -.1$

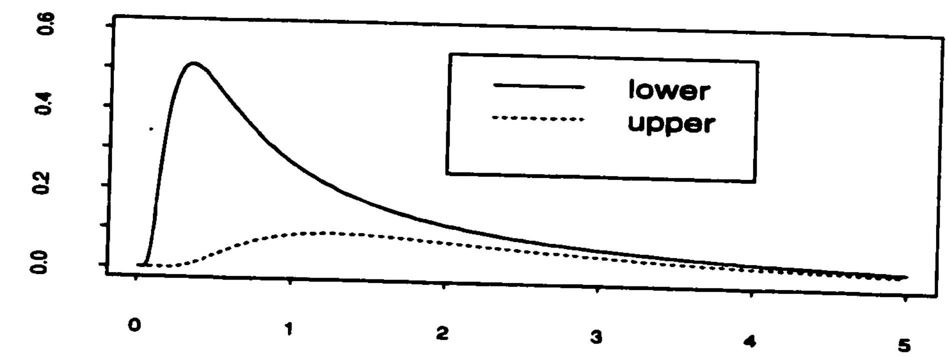


Figure 3.4: subdensities, $u = 2$, $\ell = 1$, $\mu = 0$

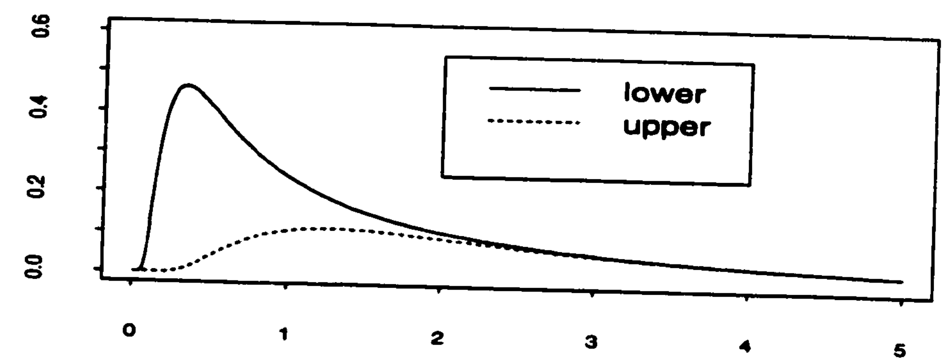


Figure 3.5: subdensities, $u = 2$, $\ell = 1$, $\mu = .6$

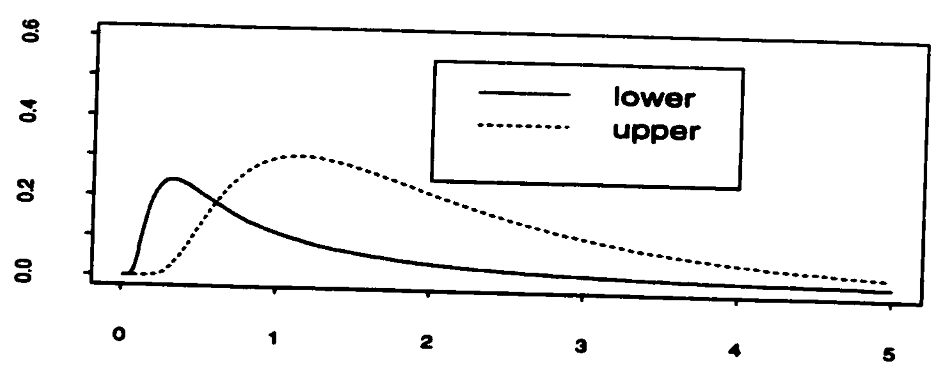


Figure 3.6: $f(u, \ell; t)$

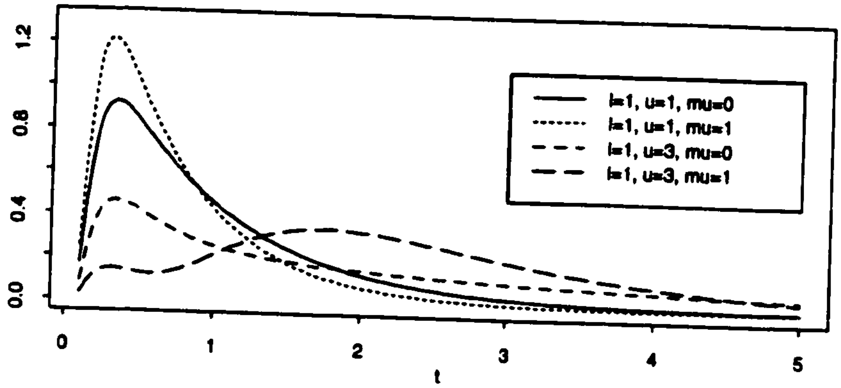


Figure 3.7: $h(u, \ell; t)$

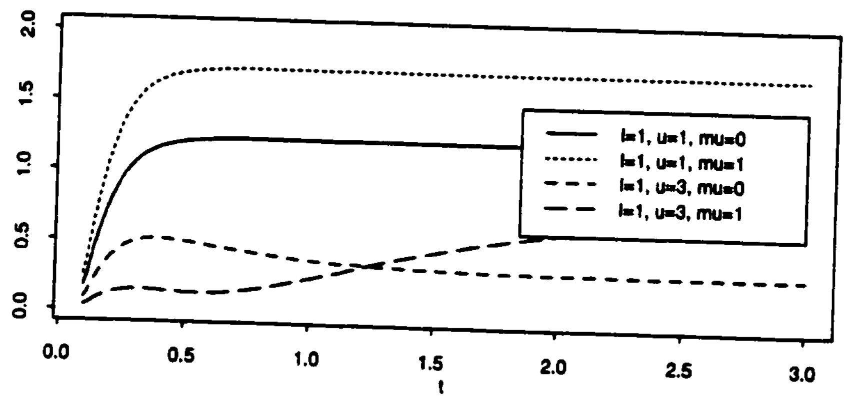
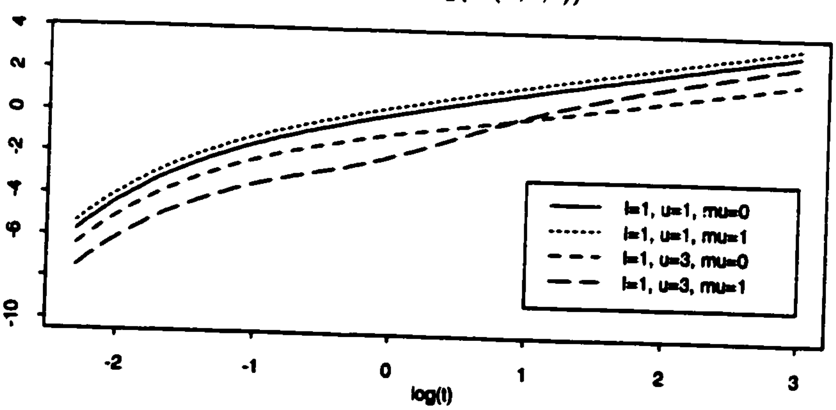


Figure 3.8: $\log(H(u, \ell; t))$



3.2 Non-linear Functions of the Parameters

In this section we discuss several important non-linear functions of the parameters, namely the expected time until breach, $E(T)$; the probability that the upper barrier is ultimately breached, $P(D = \mathrm{u})$; and the conditional expection of time until breach given that the upper barrier is breached, $E(T|D = \mathrm{u})$. For the discrete case, i.e. an asymmetric random walk on the integers with two absorbing barriers, expressions for these quantities are well known and are easily found as solutions to difference equations. To find the corresponding quantities in the continuous case, i.e. a Wiener process with drift and two absorbing barriers, we take the solution to the difference equation and pass to the limit in the manner described in appendix C, section C.2.

As noted below, an expression for $P(D = \mathrm{u})$ is given in Karlin and Taylor [23]. However I have not been able to find expressions for $E(T)$ and $E(T|D)$ in the literature. These quantities have been carefully checked by numerical integration for several values of the parameters, as will be discussed in section 3.2.3.

3.2.1 Probability that $D = \mathrm{l}$

In this section, we show how to find the probability that the process ultimately ends at the lower barrier. Following our general strategy, we start with the discrete case, set up a difference equation for $P(D = \mathrm{l})$, find the solution, and then take the limit as described in appendix C, section C.2.

Consider a random walk on the integers, which takes an upward step with probability p, and a downward step with probability $q = 1 - p$, and with absorbing barriers at 0 and $a > z$. Following Feller [17], let q_z be the probability that the walk, starting from position z, ultimately ends at the lower barrier. Then for $z = 1, 2, \ldots, a - 1$, we have the difference equation

$$q_z = p\, q_{z+1} + q\, q_{z-1}$$

with boundary conditions $q_0 = 1$ and $q_a = 0$. The general solution of this system of difference equations is given by

$$q_z = \begin{cases} \frac{(q/p)^a - (q/p)^z}{(q/p)^a - 1} & p \neq q \\ 1 - z/a & p = q. \end{cases} \tag{3.7}$$

Taking the limit as described in appendix C.2 and setting $\sigma^2 = 1$ gives

$$\begin{cases} \frac{e^{-2\mu\alpha} - e^{-2\mu\zeta}}{e^{-2\mu\alpha} - 1} & \mu \neq 0 \\ 1 - \zeta/\alpha & \mu = 0. \end{cases}$$

To describe the situation where the process starts at 0, and has absorbing barriers at $u > 0$ and $-\ell < 0$, we need only change co-ordinate systems and substitute ℓ for ζ, and $u + \ell$ for α. This gives

$$P(D = \mathrm{l}) = \begin{cases} \frac{e^{-2\mu(u+\ell)} - e^{-2\mu\ell}}{e^{-2\mu(u+\ell)} - 1} & \mu \neq 0 \\ 1 - \ell/(u + \ell) & \mu = 0 \end{cases}$$

or equivalently

$$P(D = \mathrm{l}) = \begin{cases} \frac{e^{-2\mu u} - 1}{e^{-2\mu u} - e^{2\mu\ell}} & \mu \neq 0 \\ u/(u + \ell) & \mu = 0. \end{cases}$$

Note that at $\mu = 0$, the first line in the previous expression gives an indeterminate form, 0/0. A single application of L'Hopital's rule shows that the limit as $\mu \to 0$ of $\frac{e^{-2\mu u} - 1}{e^{-2\mu u} - e^{2\mu\ell}}$ is $u/(u + \ell)$, the same value found by taking the limit of q_z in the prescribed manner when $p = q$. Also note that this agrees with an expression given in Karlin and Taylor [23] which is derived by martingale arguments.

Of course for this model, the only two possible outcomes are breach of the upper barrier and breach of the lower barrier. Thus $P(D = \mathrm{u}) = 1 - P(D = \mathrm{l})$ and is given by

$$P(D = \mathrm{u}) = \begin{cases} \frac{1 - e^{2\mu\ell}}{e^{-2\mu u} - e^{2\mu\ell}} & \mu \neq 0 \\ \ell/(u + \ell) & \mu = 0. \end{cases}$$

Note that we can get $P(D = \mathrm{u})$ from $P(D = \mathrm{l})$ by replacing u with ℓ, ℓ with u and μ with $-\mu$.

3.2.2 Expected Value of T

To find the expected value of T, the time the process first reaches one of the two barriers, consider the same random walk as in the previous subsection, with absorbing barriers at 0 and $a > z$. Let D_z be the expected time until a walk, starting at position z, is absorbed at one of the barriers. Then for $z = 1, 2, \dots, a-1$,

$$\begin{aligned} D_z &= p(D_{z+1}+1) + q(D_{z-1}+1) \\ &= pD_{z+1} + qD_{z-1} + 1 \end{aligned}$$

with boundary conditions $D_0 = 0$ and $D_a = 0$.

As shown in Feller [17], the general solution of this difference equation is given by

$$Dz = \begin{cases} \frac{z}{q-p} - \frac{a}{q-p}\frac{1-(q/p)^z}{1-(q/p)^a} & p \neq q \\ z(a-z) & p = q. \end{cases} \tag{3.8}$$

Now taking the limit of this expression as described in appendix C, section C.2, and setting $\sigma^2 = 1$, we find that the expected number of steps until the particle is absorbed approaches

$$\begin{cases} r\left[\frac{-\zeta}{\mu} + \left(\frac{\alpha}{\mu}\right)\frac{1-\exp(-2\mu\zeta)}{1-\exp(-2\mu\alpha)}\right] & \mu \neq 0 \\ r\left[\zeta(\alpha-\zeta)\right] & \mu = 0. \end{cases}$$

Dividing this by the number of steps per unit time r, gives the expected time until a Wiener process with drift, starting from position ζ, reaches one of the two barriers (at 0 or α).

Making a change of coordinate system to describe the situation where the process starts at 0 and has absorbing barriers at $u > 0$ and $-\ell < 0$ gives

$$E(T) = \begin{cases} -\frac{\ell}{\mu} + \frac{u+\ell}{\mu}\left[\frac{1-e^{-2\mu\ell}}{1-e^{-2\mu(u+\ell)}}\right] & \mu \neq 0 \\ \ell u & \mu = 0 \end{cases}$$

Note that at $\mu = 0$, the first line in the previous expression gives an indeterminate form, 0/0. A single application of L'Hopital's rule gives

$$\lim_{\mu\to 0} E(T) = \lim_{\mu \to 0} \frac{2\ell(u+\ell)[e^{-2\mu(u+\ell)} - e^{-2\mu\ell}]}{1 - e^{-2\mu(u+\ell)}[1 - 2\mu(u+\ell)]}$$

Again this gives 0/0, so another application of L'Hopital's rule is needed. This gives

$$\lim_{\mu\to 0} E(T) = \lim_{\mu\to 0} \frac{-2l(u+\ell)[2(u+\ell)e^{-2\mu(u+\ell)} - 2\ell e^{-2\mu\ell}]}{2(u+\ell)e^{-2\mu(u+\ell)}[2-2\mu(u+\ell)]} = \ell u.$$

This agrees with the value of $E(T)$ found above at $\mu = 0$.

3.2.3 Conditional Expectation of T, given outcome

In this section we discuss the conditional expectation of time in hospital, T, given outcome, that is given knowledge of whether the hospital stay ended in discharge or death. This quantity is useful for purposes of model assessment, and may be of interest to the researchers as discussed in section 2.2.1.

The conditional density of T, given D=u (ie given that the patient is ultimately discharged) is

$$f_{t|u} = \frac{P(t < T \leq t+dt, D = \mathrm{u})}{P(D = \mathrm{u})} = \frac{f_u(t)}{P(D = \mathrm{u})}.$$

Similarly the conditional density of T, given $D = \mathrm{l}$ (ie given that the patient dies in hospital) is

$$f_{t|\ell} = \frac{P(t < T \leq t+dt, D = \mathrm{l})}{P(D = \mathrm{l})} = \frac{f_\ell(t)}{P(D = \mathrm{l})}.$$

Then the conditional expectation of T, given ultimate discharge is

$$E(T|D = \mathrm{u}) = \frac{\int t f_u(t)dt}{P(D = \mathrm{u})}$$

and similarly, the conditional expectation of T, given death in hospital is

$$E(T|D = \mathrm{l}) = \frac{\int t f_\ell(t)dt}{P(D = \mathrm{l})}.$$

Note that

$$\begin{aligned} E(T) = E(E(T|D)) &= E(T|D = \mathrm{u})P(D = \mathrm{u}) + E(T|D = \mathrm{l})P(D = \mathrm{l}) \\ &= \int t f_u(t)dt + \int t f_\ell(t)dt. \end{aligned}$$

Given estimates of the model parameters, numerical integration will give values for the expressions $\int t f_u(t)dt$ and $\int t f_\ell(t)dt$. This can be done in Maple or some other package. Recall that both $f_u(t)$ and $f_\ell(t)$ are infinite series, and so approximations in the form of a finite number of terms of the series must be used. Since the integration is over the interval $(0, \infty)$, a special technique (such as transformation of the variable of integration) must be used, and this is done automatically by Maple.

Numerical integration is very computationally expensive and so closed forms for these conditional expectations will now be derived. As before, we start from the discrete case, and pass to the limit as described in section appendix C, section C.2. The formulae so derived have been found to agree with results obtained by numerical integration for several values of the parameters.

Consider the (by now familiar) random walk on the integers with absorbing barriers at 0 and a. Let $u_{z,n}$ be the probability the walk, starting from position z, ends at the lower barrier at step n. and let q_z be the probability that, starting from z, the lower barrier is ultimately breached. Let m_z be $\sum_{n=0}^{\infty} n\ u_{z,n}$, so that the expected time until breach of the lower barrier, given that this event occurs, is m_z/q_z. As shown in Chung [10], the quantities m_z satisfy the difference equation

$$m_z = p\ m_{z+1} + q\ m_{z-1} + q_z$$

for $z = 1, 2, \ldots, a-1$, with boundary conditions $m_0 = m_a = 0$. The solution to this difference equation when $p \neq q$ is given by

$$\begin{aligned} m_z = & -\frac{2\, w^a a w - w^a w - 2\, w^a a - (w^a)^2}{(w^a - 1)\, p\, (-1 + w)^2} + \\ & \frac{w^a z w + w^z w z - w^z w - w^a z - w^z z - w^z w^a}{(w^a - 1)\, p\, (-1 + w)^2} - \\ & (w^a - w^z)\left(-\frac{2\, w^a a w - w^a w - 2\, w^a a - (w^a)^2}{(w^a - 1)\, p\, (-1 + w)^2} + \frac{-w - w^a}{(w^a - 1)\, p\, (-1 + w)^2}\right)(w^a - 1)^{-1} \end{aligned} \quad (3.9)$$

where $w = q/p$. (Note that Chung gives the difference equation and solution in the more general case where probability of an upward step can vary at each epoch.)

The quantity corresponding to m_z when $p \neq q$ in the continuous case is $\int_0^\infty t\ f_\ell^\mu(t)dt$. Then the expected time until breach of the lower barrier, given that this barrier is hit before the upper barrier, is $\int_0^\infty t\ f_\ell^\mu(t)dt/P(D=1)$. To find $\int_0^\infty t\ f_\ell^\mu(t)dt$, we start with m_z as given in (3.9), and pass to the limit as described in section C.2. We first need to do some algebraic manipulations to eliminate the explicit presence of p in (3.9). Note that

$$p(1-w)=p-q.$$

Using this fact, we can rewrite (3.9) as

$$\left[(p-q)\,(w^a-1)^2\,(-1+w)\right]^{-1}\left[2\,w^a a+w^{a+1}z+2\,w^{a+1+z}a+w^{2a}z-\right.$$
$$\left.w^z z-w^a z+w^{z+1}z+w^{z+a}z-w^{a+1+z}z-w^{2a+1}z-2\,w^{z+a}a-2\,w^{a+1}a\right]. \quad (3.10)$$

We now pass to the limit, which reduces here to the simple process of replacing w^x with $e^{-2\mu x}$, and $(p-q)$ with μ. Then to translate the axes to the situation where the process starts at 0, and has absorbing barriers at u and ℓ, we replace a with $u+\ell$ and z with ℓ. This gives

$$\int_0^\infty t\ f_\ell^\mu(t)dt=$$
$$\left[\mu\left(e^{-2\mu(u+\ell)}-1\right)^2\left(-1+e^{-2\mu}\right)\right]^{-1}\left[2\,e^{-2\mu(u+\ell)}\,(u+\ell)+e^{-2\mu(u+\ell+1)}\ell+\right.$$
$$2\,e^{-2\mu(u+2\ell+1)}\,(u+\ell)+e^{-4\mu(u+\ell)}\ell-e^{-2\mu\ell}\ell-e^{-2\mu(u+\ell)}\ell+e^{-2\mu(\ell+1)}\ell+$$
$$\left.e^{-2\mu(2\ell+u)}\ell-e^{-2\mu(u+2\ell+1)}\ell-e^{-2\mu(2u+2\ell+1)}\ell-2\,e^{-2\mu(2\ell+u)}\,(u+\ell)-2\,e^{-2\mu(u+\ell+1)}\,(u+\ell)\right].$$

Then the conditional expectation of T, given that the process ends at the lower barrier, i.e. $E(T|D=l)$, is $\int_0^\infty t\ f_\ell^\mu(t)dt$ divided by $P(D=l)$ when $\mu\neq 0$. This gives

$$\left[\mu\left(e^{-2\mu(u+l)}-1\right)^2\left(-1+e^{-2\mu}\right)\left(e^{-2\mu u}-1\right)\right]^{-1}\left[-\left(-e^{-2\mu(u+l+1)}+e^{-2\mu l}+e^{-2\mu(u+l)}-\right.\right.$$
$$\left.e^{-2\mu(l+1)}+e^{-2\mu(u+2l+1)}+e^{-2\mu(2u+2l+1)}-e^{-4\mu(u+l)}-e^{-2\mu(u+2l)}\right)\left(e^{-2\mu u-2\mu l}-1\right)e^{2\mu l}l$$
$$\left.-\left(2\,e^{-2\mu(u+2l)}-2\,e^{-2\mu(u+2l+1)}+2\,e^{-2\mu(u+l+1)}-2\,e^{-2\mu(u+l)}\right)\left(e^{-2\mu u-2\mu l}-1\right)e^{2\mu l}\,(u+l)\right].$$

To find $E(T|D=l)$ in the no-drift case, we take the limit of the above expression as $\mu\to 0$. This gives

$$\frac{\ell}{3}(2u+\ell).$$

Similarly one can show that $\int_0^\infty t f_u^\mu(t)dt$ in the drift case is

$$\left[\mu\left(-1+e^{-2\mu(u+\ell)}\right)^2\left(-1+e^{-2\mu}\right)\right]^{-1}\left[e^{-2\mu(u+\ell+1)}(u+\ell)-e^{-2\mu(\ell+1)}\ell-e^{-2\mu}\ell+\right.$$
$$e^{-2\mu\ell}\ell+e^{-2\mu}(u+\ell)-e^{-2\mu(u+\ell)}(u+\ell)+e^{-2\mu(u+\ell+1)}\ell-\ell e^{-2\mu(u+\ell)}+$$
$$(u+\ell)e^{-2\mu\ell}-e^{-2\mu(u+2\ell+1)}(u+\ell)-e^{-2\mu(\ell+1)}(u+\ell)+$$
$$\left.e^{-2\mu(u+2\ell)}(u+\ell)+e^{-2\mu(u+2\ell+1)}\ell-e^{-2\mu(u+2\ell)}\ell-u\right].$$

Then $E(T|D=\mathrm{u})$ is $\int_0^\infty t f_u^\mu(t)dt/P(D=\mathrm{u})$, which gives

$$\left[\mu\left(e^{-2\mu(u+l)}-1\right)^2\left(-1+e^{-2\mu}\right)\left(-1+e^{2\mu l}\right)\right]^{-1}\left[\left(e^{-2\mu}+e^{2\mu(-1+l)}+e^{-2\mu u}+\right.\right.$$
$$\left.e^{-2\mu(u+l)}-e^{-2\mu(u+l+1)}-1-e^{-2\mu(u+1)}\right)\left(e^{-2\mu u}-e^{2\mu l}\right)e^{-2\mu l}l$$
$$+\left(-e^{2\mu(-1+l)}+e^{-2\mu}-e^{-2\mu(u+1)}-e^{-2\mu(u+l)}+e^{-2\mu u}-1+e^{-2\mu(u+l+1)}\right)$$
$$\left.\left(e^{-2\mu u}-e^{2\mu l}\right)e^{-2\mu l}(u+l)+ue^{2\mu l}\left(e^{-2\mu u}-e^{2\mu l}\right)e^{-2\mu l}\right].$$

For zero drift, $E(T|D=\mathrm{u})$ is given by

$$\frac{u}{3}(u+2\ell).$$

As usual, $E(T|D=\mathrm{u})$ is found from $E(T|D=\mathrm{l})$ by reversing the roles of u and ℓ, and replacing μ by $-\mu$.

3.3 Relation to Inverse Gaussian Distribution

A closely related distribution is the inverse Gaussian. Consider a Wiener process with drift $\mu > 0$ and volatility σ^2, and let T be the time the process, starting from 0, first reaches a barrier at $u > 0$. Then T has an inverse Gaussian distribution, with density

$$g(t;u,\mu,\sigma^2)=\frac{u}{\sigma\sqrt{2\pi t^3}}\exp\left(\frac{-(u-\mu t)^2}{2\sigma^2 t}\right)$$

(see for instance [9]). All positive and negative moments of this distribution exist. Note that the drift parameter for the FP2Bμ distribution discussed in this chapter can be any real number.

However, if the drift parameter for the inverse Gaussian distribution is less than 0, while $u > 0$, we get a defective distribution with a mass of probability at infinity.

Parameterized as above, the inverse Gaussian distribution has identifiability problems, since two distinct values of the parameter vector can give the same value of the density. There are many ways of reparameterizing to reduce the parameter space to two dimensions. Most simply, we can set $\sigma^2 = 1$. In this case the inverse Gaussian distribution has mean u/μ and variance u/μ^3.

We can argue probabilistically that $F^\mu(t; u, \ell, \mu)$, the survivor function of the FP2Bμ distribution, is dominated by the cdf of the inverse Gaussian distribution, which we will denote $G(t; u, \mu)$. Of course $F^\mu(t; u, \ell, \mu)$ gives the probability, p say, that the process hits one of the two barriers before time t. For the inverse Gaussian distribution with mean $\mu > 0$, the lower barrier has been removed, and $G(t; u, \mu)$, the probability of reaching the upper barrier before time t, must be smaller than p (because paths that used to be stopped by the bottom barrier can now continue on). Thus, when $\mu > 0$, $F^\mu(t; u, \ell, \mu) > G(t; u, \mu)$, and so $1 - F^\mu(t; u, \ell, \mu) < 1 - G(t; u, \mu)$. Thus, if $T_1 \sim FP2B(u, \ell, \mu)$ and $T_2 \sim IG(u, \mu)$,

$$a^{-1}E(T_1^a) = \int t^{a-1}(1 - F^\mu(t; u, \ell, \mu))\, dt < \int t^{a-1}(1 - G(t; u, \mu))\, dt = a^{-1}E(T_2^a) \qquad (3.11)$$

where a is any real number. For a Wiener process with drift $-\mu < 0$, the time until hitting a single barrier at $-u < 0$ also has an $IG(u, \mu)$ distribution. Since all moments of the inverse Gaussian distribution exist, ie $E(T_2^a) < \infty$ for all a, equation 3.11 ensures that all moments of the FP2Bμ distribution exist, for $\mu \neq 0$, $u < \infty$, $\ell < \infty$.

3.4 Maximum Likelihood Estimation

For the model with non-zero drift and volatility=1, the log-likelihood is given by

$$\mathcal{L}(u, \ell, \mu; t) = \sum_{\{i:D_i=u\}} \log f_u^\mu(t_i) + \sum_{\{i:D_i=l\}} \log f_\ell^\mu(t_i) + \sum_{\{i:D_i=c\}} \log \mathcal{F}^\mu(t_i)$$

where, as before, D_i is the discharge status of individual i, which takes values u, l or c depending on whether individual i is discharged, dies or is independently censored at time t_i.

As before, we use optimization routines to find the values of u, ℓ and μ which maximize the log-likelihood. To find the maximum likelihood estimates of mean time in hospital, probability of ultimate discharge, and conditional expectation of time in hospital given destination, we substitute the maximum likelihood estimates of the parameters u, ℓ and μ into the expressions given in section 3.2. Since $P(D = \text{u})$, $E(T)$ and $E(T|D = \text{u})$ are all non-linear functions of the parameters u, ℓ and μ, the bootstrap will be used to construct confidence intervals.

When both destination and time of breach are known, there are no problems with identifiability, that is distinct values of the parameter vector $\theta = (u, \ell, \mu)$ will give rise to distinct distributions of the observables. However, certain configurations of the data potentially give estimability problems. If the same number of individuals breach the upper barrier as breach the lower barrier, and $u = \ell$, then $\theta_1 = (a, a, \mu)$ and $\theta_2 = (a, a, -\mu)$ will give identical values of the distribution. It is unlikely in any practical situation that u will exactly equal ℓ. However, in the rare case where this occurs, it is easily shown analytically that if we restrict the parameter space so that $u = \ell$, the maximum likelihood estimate of μ is zero. Thus there is still a unique maximum likelihood estimate, namely $\hat{\theta} = (a, a, 0)$.

3.4.1 Covariates

We may allow the barrier levels to vary for different individuals or groups of individuals. We can allow these parameters to depend on covariates exactly as described before.

In the context of health care data, the drift parameter μ might be interpreted as the propensity to get well. Some of the heterogeneity in the data may be explained by allowing μ to depend on individual-level variables, such as age, gender and severity of disease at admission. This approach is particularly useful with detailed individual data, for instance on physical conditioning, or comorbidities. In the example shown in section 3.5, we allow the drift parameter to be a linear function of some individual covariates,

$$\mu_i = \beta_0^\mu \; x_{i0}^\mu + \beta_1^\mu \; x_{i1}^\mu + \ldots + \beta_{p_\mu}^\mu \; x_{ip_\mu}^\mu$$

where $(x_{i0}^\mu, x_{i1}^\mu, \ldots, x_{ip_\mu}^\mu)'$ is a vector of covariates measured on individual i, thought to affect the drift parameter, and $\beta_0^\mu, \beta_1^\mu, \ldots \beta_{p_\mu}^\mu$ are parameters to be estimated.

We again use maximum likelihood to estimate the parameters. Extension of the log-likelihood, gradient and hessian to accommodate covariates in the drift case is straightforward and is shown in appendix E.

3.4.2 Starting Values

As always, good starting values for the estimation procedure are important. For a model with no covariates which includes only the parameters u, ℓ and μ, experience has shown that it is usually adequate to run the same data through the zero-drift software, and use these estimates as starting values for the barrier parameters in the drift case. For starting values for the drift parameters, a value of zero seems to suffice.

For a drift model involving covariates, it seems to be adequate to start with mock least squares estimates of the barrier parameters as described in section 2.2.3, and use a starting value of zero for all the drift parameters.

3.5 Example 1

In this section we fit the drift model to the same data set discussed in section 2.4. Here we will allow the drift governing an individual's health status process to be a function of the covariate severity. As before, let $x_{i1} = 1$ if individual i exhibits a major complication or comorbidity, and let $x_{i2} = 1$ if individual i has an extreme complication or comorbidity. Individuals with no, minor, or moderate complication or comorbidity have both x_{i1} and x_{i2} equal to 0.

Thus the drift for individual i is

$$\mu_i = \beta_0^\mu + \beta_1^\mu \; x_{i1} + \beta_2^\mu \; x_{i2}$$

The following table shows the maximum likelihood estimates and standard errors for this model. The last two lines show two 95% confidence intervals. The first is a normal-theory symmetric confidence interval constructed using the displayed standard error. The second is a bootstrap confidence interval, which will be discussed below.

	u	ℓ	β_0^μ	β_1^μ	β_2^μ
Estimate	3.573	1.661	0.654	-0.291	-0.582
Std Err	0.234	0.146	0.071	0.091	0.138
Normal	(3.15, 3.99)	(1.40, 1.92)	(0.52, 0.79)	(-0.47, -0.11)	(-0.86, -0.31)
Bootstrap	(3.15, 4.14)	(1.50, 1.85)	(0.53, 0.82)	(-0.49, -0.12)	(-0.97, -0.34)

The fit of the model is illustrated graphically in figure 3.9. The ordered Cox-Snell residuals

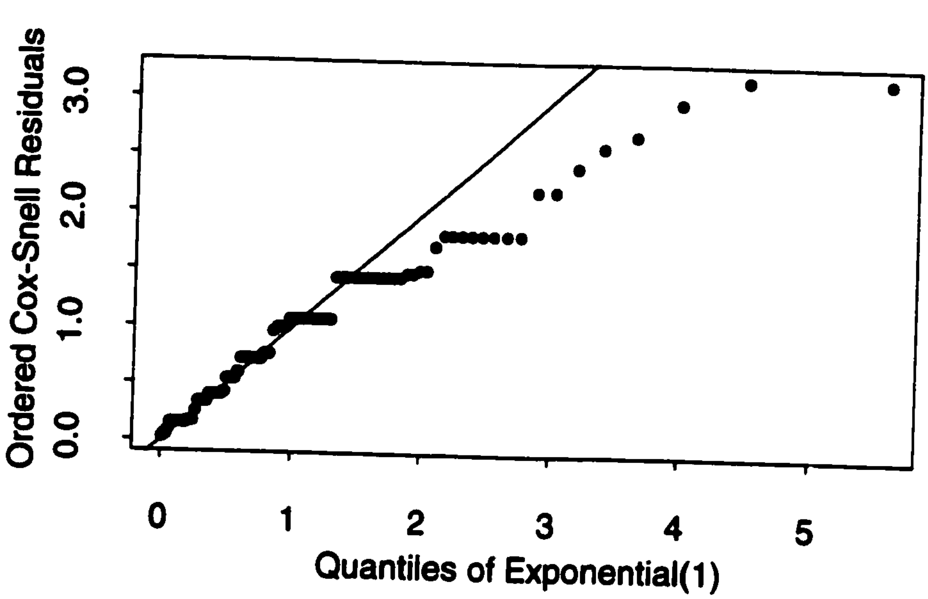


Figure 3.9: Quantile Plot of Cox Snell residuals, Example 1

plotted against the quantiles of the exponential(1) distribution should lie on a straight line with slope 1 and intercept 0. and this plot indicates that the fit is better than the zero-drift model, although there is still room for improvement. Figure 3.10 shows that four observations have large raw residuals. Later we will refit the model, without these observations.

The next table shows the estimated drift in each severity group. The observed average time in hospital for those who were discharged is given in the column labelled $T|D = \mathrm{u}$. The observed average LOS for those who died in hospital is given in column $T|D = \mathrm{l}$. Column T gives the overall observed average time in hospital. The observed proportion of patients who were discharged is

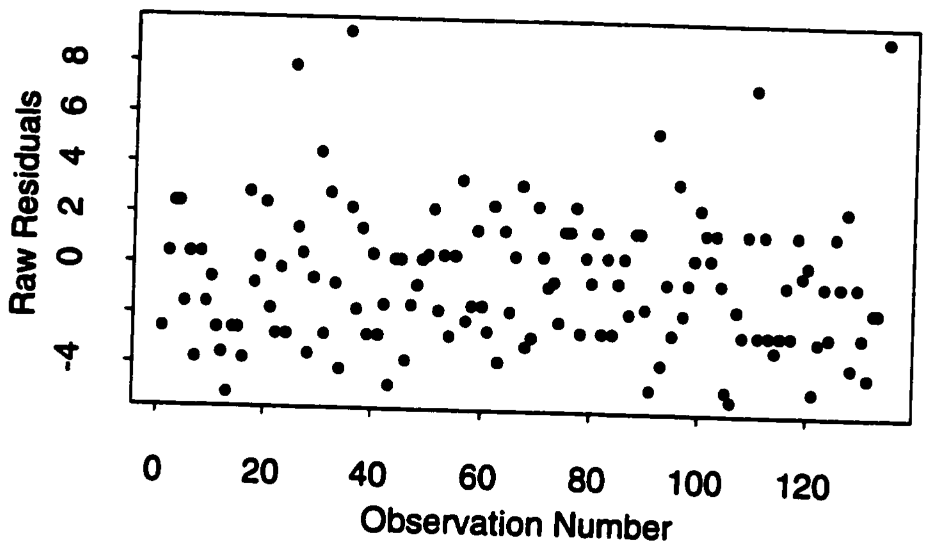


Figure 3.10: Index plot of raw residuals

given in column $P(D = \text{u})$, and the observed proportion who died in hospital is given in column $P(D = \text{l})$. The corresponding values predicted by the model are shown in brackets.

Severity	n	$\hat{\mu}$	T	$T\|D = \text{u}$	$T\|D = \text{l}$	$P(D = \text{u})$	$P(D = \text{l})$
low	82	(0.654)	4.49(4.56)	4.61(4.82)	3.60(2.45)	0.88(0.89)	0.12(0.11)
med	39	(0.363)	4.67(5.76)	5.96(6.59)	2.36(3.64)	0.64(0.72)	0.36(0.28)
high	13	(0.072)	4.92(6.14)	8.80(8.12)	2.50(4.81)	0.38(0.40)	0.62(0.60)

Here, the upper barrier is estimated to be 3.573, and the lower barrier to be -1.661. The model indicates that severity has a significant effect on the drift of the health status process, with the sicker patients having a less pronounced propensity to get well. The estimated drift is 0.654 for the *low* severity group, 0.363 for the *medium* severity group, and 0.072 for the *high* severity group. In all severity groups, the drift is estimated to be positive; thus even in the more severely ill group, the patients tend to drift toward the upper barrier. The model predicts that, as severity increases, the drift decreases towards zero.

Note that the model predicts that as drift decreases, the expected time to hitting the upper barrier increases, which seems reasonable. However, expected time to hitting the lower barrier also increases with decreasing drift, which may seem somewhat counterintuitive. Examination of the subdensities may help to explain the situation. The subdensities for the low and high severity groups are shown in figures 3.11 and 3.12. The upper subdensities are shown as a solid line, and the lower as a dotted line. The mean LOS, given the upper barrier is reached, is shown as a solid vertical line, while mean LOS given the lower barrier is reached is shown as a dotted vertical line. Due to the heavy tails, the means are to the right of the modes.

We see in figure 3.11, that when the drift is high, the probability that the process ends at the lower barrier is quite small, as indicated by the small area under the dotted curve. However, given that that the process does end at the lower barrier, it will do so relatively early (2.45 days on average). This seems strange, but recall that the lower barrier is closer to the origin than the upper barrier, so (despite the upward drift) some of the paths will breach the lower barrier early in the observation period. Paths that move upward in the early time intervals, however, are unlikely to come down as far as the lower barrier later on.

Figure 3.12 shows the situation when the drift is reduced to almost 0. First, it now takes longer to hit either barrier than it did when the drift was higher, since now the paths tend to wobble about the horizontal axis. Since the upper barrier is farther away (3.573) than the lower barrier (1.661), and the drift is almost 0, it is less likely that the process will end at the upper barrier. Also it takes, on average, longer to hit this barrier (8.12 days) than to hit the lower barrier (4.82 days)

We can see that the drift model is doing a good job of predicting some features of the data. The probability of each outcome (discharge or death) is well predicted by this model, in contrast to the zero-drift model. The predicted time in hospital in the group of patients who were ultimately discharged (shown in the column labeled $T|D = \text{u}$) is reasonably close to the observed average. However, for those who ultimately die in hospital predicted time in hospital is quite different from the observed average (column $T|D = \text{l}$). For this group, the model predicts that time until death increases with severity of illness, whereas in fact the observed average LOS is highest for the *low* severity group, and lowest for the *medium* severity group.

Figure 3.11: Sub-Densities, Low Severity Group, $\mu = 0.654$

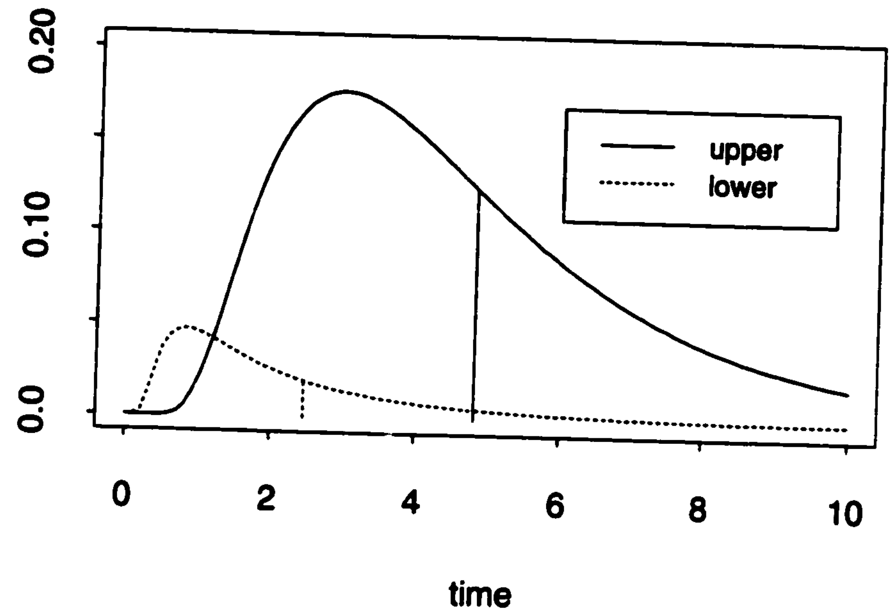
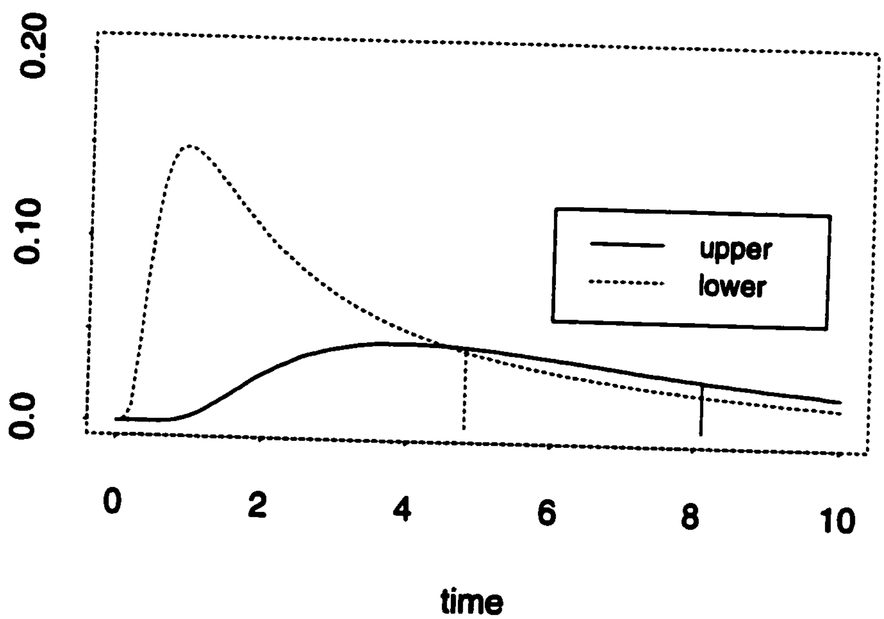


Figure 3.12: Sub-Densities, High Severity Group, $\mu = 0.072$



The overall time in hospital is also not well predicted, due to the lack of fit in the *low* severity group.

There seems to be a trade-off here, between providing a good fit to probability of outcome, time in hospital given discharge, and time in hospital given death. In this case the model does a good job of predicting probability of outcome and time until discharge. Time until death is less well predicted. Note that the group of patients who died is much smaller (n=32) than the group who were discharged (n=102). So the model is doing a good job of making predictions for the larger group.

An assumption implicit in this model is that, within each severity group, all patients share the same value of the drift parameter and the barrier parameters. It would seem that this assumption is not supported by the data. Perhaps an important covariate is missing.

The previous table gives normal theory confidence intervals, and bootstrap confidence intervals constructed from the 2.5% and 9.75% quantiles of a bootstrap distribution of 1000 iterations. The confidence intervals are compared graphically in figure 3.13. Here we can see that the bootstrap intervals are nearly symmetric, and close in length to the normal-theory confidence intervals. This gives some reassurance that the estimates are asymptotically normal, an issue which will be discussed in Chapters 4 and 5.

We rely on the bootstrap to construct confidence intervals for the predicted length of stay and proportion of patients discharged in each covariate group, since these are non-linear functions of the parameters. These are shown in the next table.

Severity	*low*	*med*	*high*
E(T)	4.560	5.757	6.142
	(4.15, 5.03)	(5.09, 6.57)	(4.11, 7.16)
P(D=u)	0.887	0.717	0.401
	(0.83, 0.93)	(0.54, 0.84)	(0.08, 0.67)

In particular, we note that the confidence intervals for the *high* severity group are rather large, perhaps due to the relatively small number of patients (n=13) in this group.

Recall that figure 3.10 showed four large raw residuals. In fact these correspond to the four

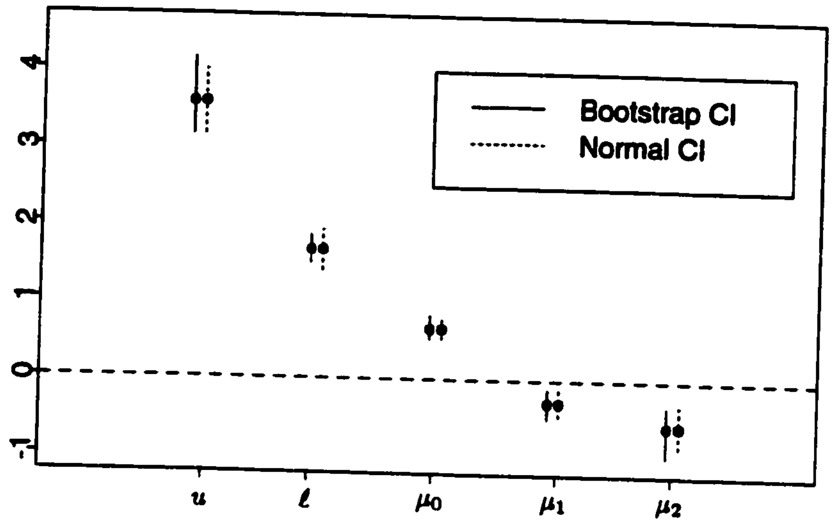


Figure 3.13: Confidence Intervals

largest values of LOS in the data set. We now refit the model, removing these four observations. As shown in figure 3.14, excluding these four values does not seem to improve the fit of the model. Note that the raw residuals are not identically distributed, and a few large values are not necessarily indicative of a bad fit.

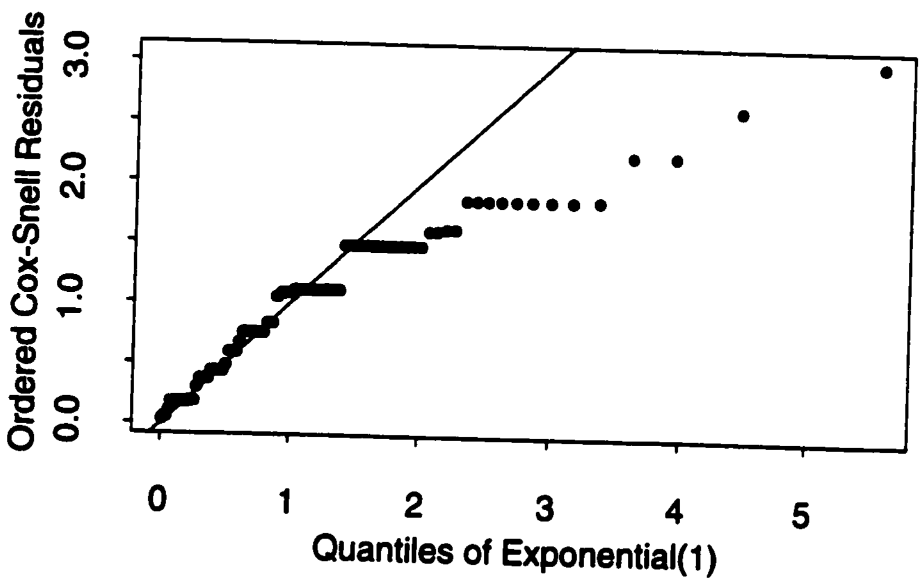


Figure 3.14: Quantile Plot of Cox Snell residuals, four largest observations removed

3.6 Example 2

We now investigate another subset of the data consisting of individuals with renal failure (DRG = 316), 75 years or older, who were either discharged home ($n = 78$) or died in hospital ($n = 22$). Here because of sparsity in the data we have coded severity at two levels. Let $x_{i1} = 1$ if individual i has a high severity score (major or extreme comorbidities and complications), and $x_{i1} = 0$ otherwise. Side by side boxplots of severity by outcome group are shown in figure 3.15.

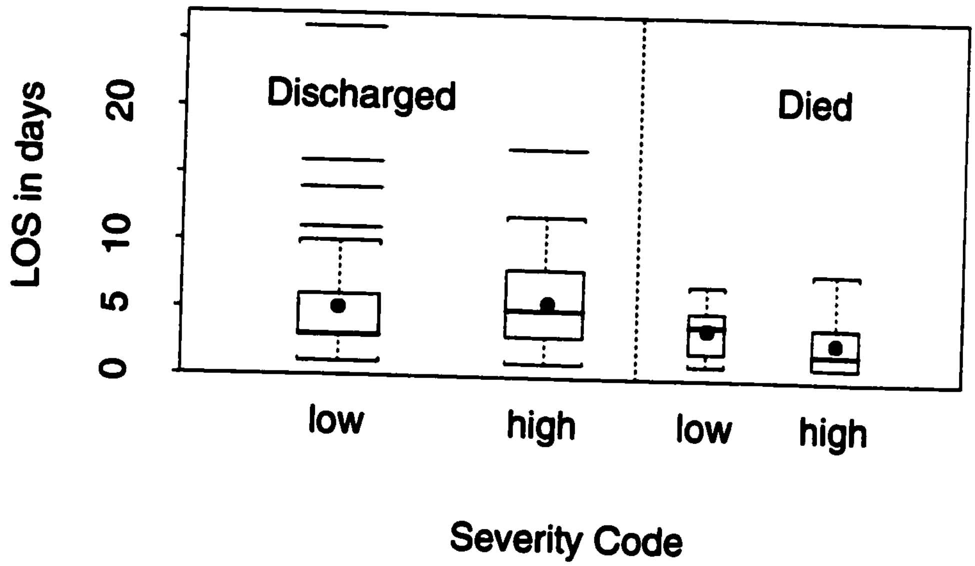


Figure 3.15: LOS by Severity and Outcome

An earlier investigation showed that severity does not significantly affect the drift parameter. The model considered here is

$$u_i = \beta_0^u + \beta_1^u x_{i1}$$
$$\ell_i = \beta_0^\ell + \beta_1^\ell x_{i1}.$$

The following table shows the maximum likelihood estimates and standard errors for this model.

	β_0^u	β_1^u	β_0^ℓ	β_1^ℓ	μ
Estimate	2.592	0.623	2.247	-0.554	0.381
Std Err	0.240	0.327	0.297	0.348	0.059

All individuals in the data set are estimated to have the same positive drift. Severity has a marginal effect on the upper and lower barriers. High severity increases the apparent distance to the upper barrier, and decreases the distance to the lower barrier. As discussed earlier, this situation could reflect differing health levels at admission to hospital.

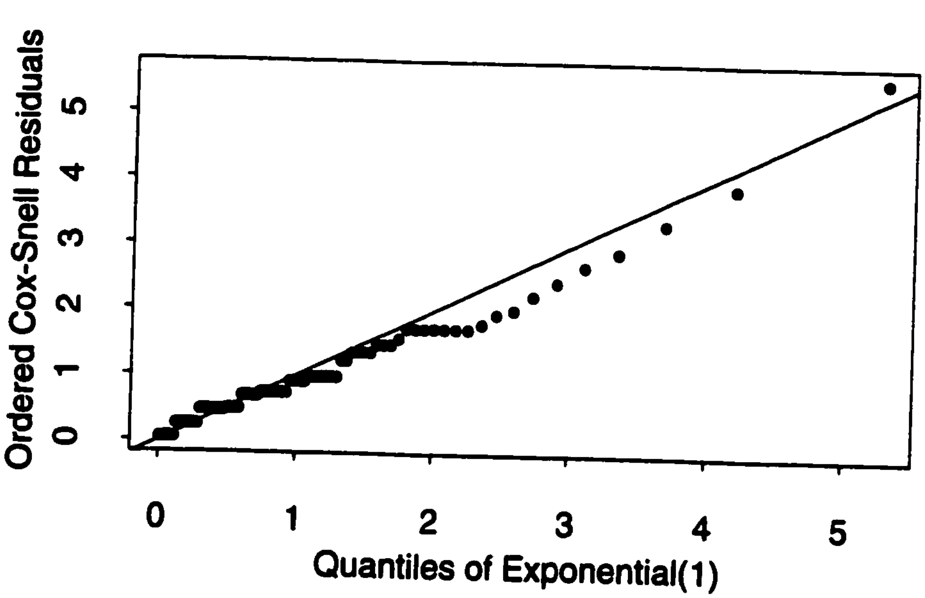


Figure 3.16: Quantile Plot of Cox Snell residuals

The fit of the model is illustrated graphically in figure 3.16. The plot indicates a good fit to the data. The observed and predicted values are shown in the following table. Predicted values are given in parentheses.

Severity	n	T	$T\|D = u$	$T\|D = l$	$P(D = u)$	$P(D = l)$
low	48	4.85(4.78)	5.08(4.85)	3.75(4.35)	0.84(0.84)	0.17(0.16)
high	52	4.81(5.11)	5.53(5.69)	2.85(3.47)	0.73(0.74)	0.27(0.26)

The predictions are quite good as well, although the model is overestimating $E(T|D = l)$ somewhat.

3.7 Example 3

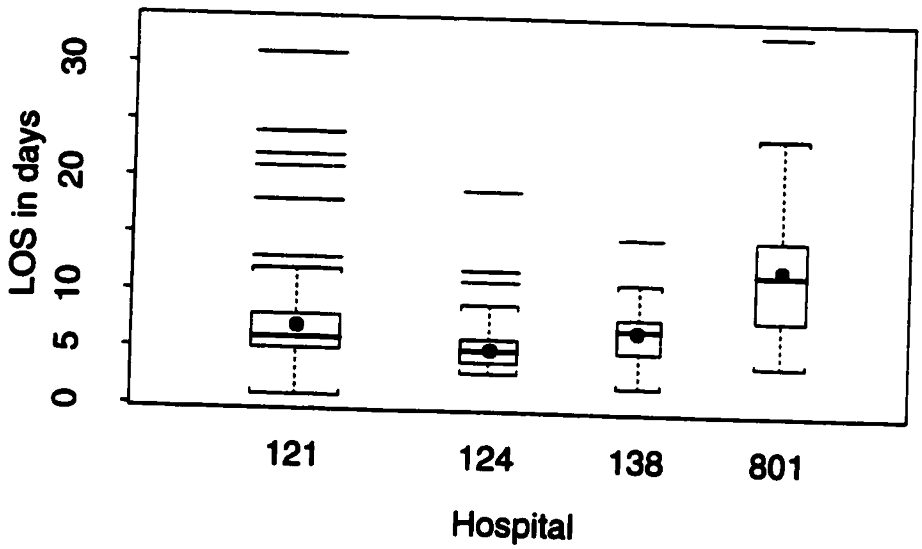


Figure 3.17: LOS by hospital, Males, $\geq$ 70, DRG=107

Our final example for this section investigates a surgical DRG, coronary bypass with cardiac catheter (DRG 107). This subset of patients are males, age 70 or over. Here we investigate whether LOS distribution differs between four large hospitals, adjusting for severity of disease, and also age of patient.

Severity is coded at two levels, and the variable sev_i equals 1 if individual i had major or extreme complications or comorbidities and 0 otherwise. Age of patient is coded in five year age groups. There are three indicator variables for the four hospitals. Thus the variable $hosp_{ji}$ equals

1 if individual i stayed at hospital j, and 0 otherwise. The baseline hospital is a rural hospital, whereas the other three hospitals are urban. Here we let the drift parameter depend on age and severity. The upper barrier is allowed to differ for the different hospitals. All patients in the data set are assumed to share the same lower barrier. The upper barrier and drift are

$$u_i = \beta_0^u + \beta_1^u hosp_{1i} + \beta_2^u hosp_{2i} + \beta_3^u hosp_{3i}$$
$$\mu_i = \beta_0^\mu + \beta_1^\mu sev_i + \beta_2^\mu age_i.$$

The estimated coefficients for this model are now given.

	β_0^u	β_1^u	β_2^u	β_3^u	ℓ	β_0^μ	β_1^μ	β_2^μ
Estimate	10.985	-4.394	-5.855	-4.075	2.175	2.260	-0.145	-0.081
Std Err	0.716	0.620	0.652	0.739	0.244	0.276	0.047	0.016

The fit of the model is illustrated graphically in figure 3.18.

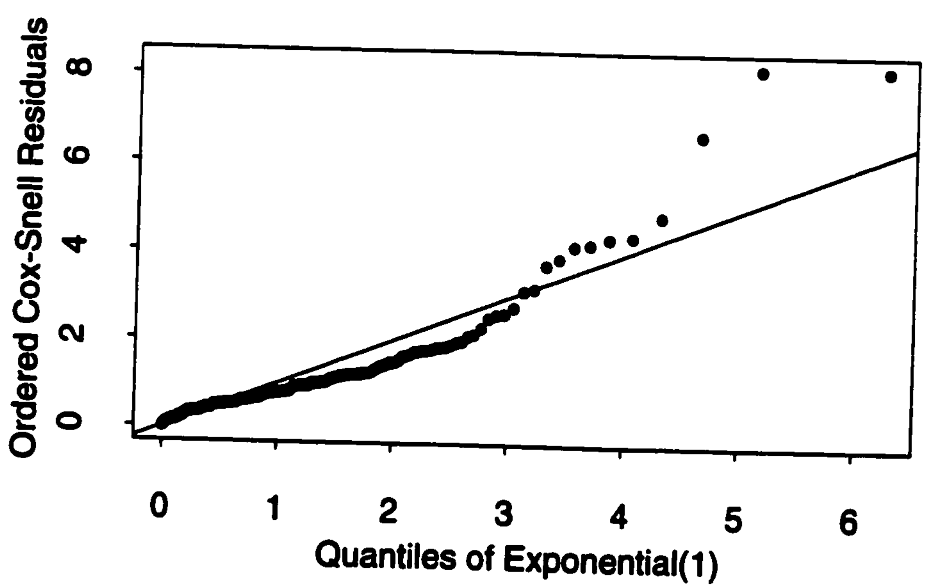


Figure 3.18: Quantile Plot of Cox Snell residuals

The baseline hospital is estimated to have a much higher upper barrier than the three urban hospitals. Increasing severity lowers the drift parameter, and increasing age likewise decreases drift toward zero. The Cox-Snell plot shows that the fit of the model is not good for the observations with large residuals. Further investigation is needed.

Chapter 4

Properties of the Zero Drift FP2B Distribution

In this chapter we study the properties of the FP2B distribution with drift equal to 0, using the series representation developed in the previous Chapter. (An alternate representation will be explored in Chapter 5.) We establish that the density of the FP2B distribution can be obtained by differentiating the cumulative distribution function (cdf) discussed in the previous Chapter term by term. Also the various derivatives can be obtained by differentiating an appropriate series term by term.

We will focus on the upper density and its derivatives, and the corresponding cdf. The results and proofs for the lower barrier are exactly parallel. Also we discuss here the derivatives with respect to u of the upper density. The results and proofs for the derivatives with respect to ℓ and the mixed derivatives are similar and in most cases slightly simpler. Unless otherwise indicated, we assume both u and ℓ to be finite. Where not stated, the limits of summation over k run from 0 to ∞. We will use $\phi(x)$ to denote the standard normal density evaluated at x, and $\Phi(x)$ to denote the standard normal cdf.

We will make use of three commonly known theorems, which are stated here without proof. Conditions for interchanging differentiation and summation, i.e. conditions that allow us to differentiate a series term by term, are given in the following theorem, from Bressoud [6]:
Theorem *Let $F(x) = f_1(x) + f_2(x) + f_3(x) + \ldots$ be an infinite series for which each summand, $f_k(x)$, is differentiable at every point in an open interval I containing a. If $\sum_{k=1}^{\infty} f_k'(x)$ converges*

uniformly over I, then $F(x)$ is differentiable at a and

$$F'(a) = \sum_{k=1}^{\infty} f_k'(a)$$

Likewise, sufficient conditions for interchanging integration over a finite interval and summation are given in the following theorem, also from Bressoud [6].

Theorem *Let $f_1(x)+f_2(x)+f_3(x)+\dots$ be uniformly convergent over the interval $[a,b]$, converging to $F(x)$. If each $f_k(x)$ is integrable over $[a,b]$, then so is $F(x)$ and*

$$\int_a^b F(x)\, dx = \sum_{k=1}^{\infty} \int_a^b f_k(x)\, dx$$

We will also use the well-known Weierstrass M-test, which is given here as stated in Apostol [2].

Theorem (Weierstrass M-test). *Let $\{M_n\}$ be a sequence of nonnegative numbers such that*

$$|f_n(x)| \le M_n, \text{ for } n = 1, 2, \dots, \text{ and for every } x \text{ in } S.$$

Then $\sum f_n(x)$ converges uniformly on S if $\sum M_n$ converges.

Corollary *Under the same assumptions given for the Weierstrass M-test, $f(x) = \sum f_n(x)$ is bounded if $\sum M_n$ converges .*

Proof:

$$f(x) = \sum f_n(x) \le \sum |f_n(x)| \le \sum M_n$$

which converges by assumption. **Q.E.D.**

4.1 The FP2B Density

In this section we will show that

- $f_u(t; u, \ell)$ is truly the upper subdensity by showing that it can be obtained by differentiating the upper cdf term by term,
- for small t, $f_u(t; u, \ell)$ behaves like $t^{-3/2}\phi\left(\frac{u}{\sqrt{t}}\right)$.

The probability that the process hits the upper barrier before time t, and does so without first hitting the lower barrier is

$$F_u(t;u,\ell) = 2\sum_k \left\{\Phi\left(\frac{s_k+\ell}{\sqrt{t}}\right) - \Phi\left(\frac{s_k-\ell}{\sqrt{t}}\right)\right\} \tag{4.1}$$

where

$$s_k = -(2k+1)(u+\ell).$$

Note that $u,\ell > 0$ and $s_k < 0$. In this section, we will write the series in (4.1) as $F_u(t;u,\ell) = 2\sum_k F_{uk}(t;u,\ell)$ where $F_{uk}(t;u,\ell) = \Phi\left(\frac{s_k+\ell}{\sqrt{t}}\right) - \Phi\left(\frac{s_k-\ell}{\sqrt{t}}\right)$. We will define $F_{uk}(0;u,\ell) = 0$, which makes $F_{uk}(t;u,\ell)$ continuous at 0.

Theorem 4.1 *For $u,\ell > 0$, $F_u(t;u,\ell) = \sum_k F_{uk}(t;u,\ell)$ can be differentiated term by term with respect to t, for all $t \geq 0$, giving*

$$f_{uk}(t;u,\ell) = -t^{-3/2}\left\{(s_k+\ell)\,\phi\left(\frac{s_k+\ell}{\sqrt{t}}\right) - (s_k-\ell)\,\phi\left(\frac{s_k-\ell}{\sqrt{t}}\right)\right\}.$$

Proof: For $u,\ell > 0$, each function $F_{uk}(t;u,\ell)$ is differentiable for each t in $(0,\infty)$, and the derivative with respect to t of $2F_{uk}(t;u,\ell)$ is

$$f_{uk}(t;u,\ell) = 2\frac{\partial F_{uk}(t;u,\ell)}{\partial t} = -t^{-3/2}\left\{(s_k+\ell)\,\phi\left(\frac{s_k+\ell}{\sqrt{t}}\right) - (s_k-\ell)\,\phi\left(\frac{s_k-\ell}{\sqrt{t}}\right)\right\}.$$

The function $F_{uk}(t;u,\ell)$ has a right hand derivative at $t=0$, and this right hand derivative is easily seen to be 0, the limit as $t \to 0$ of $f_{uk}(t;u,\ell)$. To show that we can differentiate the series $\sum_k F_{uk}(t;u,\ell)$ with respect to t term by term, it suffices to show that $\sum_k f_{uk}(t;u,\ell)$ is uniformly convergent in t. We will do this using the Weierstrass M-test.

Let g_k equal the first summand of $f_{uk}(t;u,\ell)$, i.e. $g_k = -t^{-3/2}(s_k+\ell)\,\phi(\frac{s_k+\ell}{\sqrt{t}})$. Note that g_k is positive, and achieves its maximum over $t \in (0,\infty)$ at $t = \frac{1}{3}(s_k+\ell)^2$. Let M_k denote the maximum value of this function over $t \in (0,\infty)$. Clearly, M_k is $O(k^{-2})$ and so $\sum_k M_k$ converges. By the Weierstrass M-test, the series $-t^{-3/2}\sum_k(s_k+\ell)\,\phi(\frac{s_k+\ell}{\sqrt{t}})$ converges uniformly in t, for $t \in (0,\infty)$. A similar argument applies to $-t^{-3/2}\sum_k(s_k-\ell)\,\phi(\frac{s_k-\ell}{\sqrt{t}})$, and so $\sum_k f_{uk}(t;u,\ell)$ converges uniformly in t, for $t \in (0,\infty)$. Note that we have also shown that the series $\sum_k f_{uk}(t;u,\ell)$ converges absolutely. Also, by the corollary to the Weierstrass M-test, the function $|f_u(t;u,\ell)|$ is bounded.

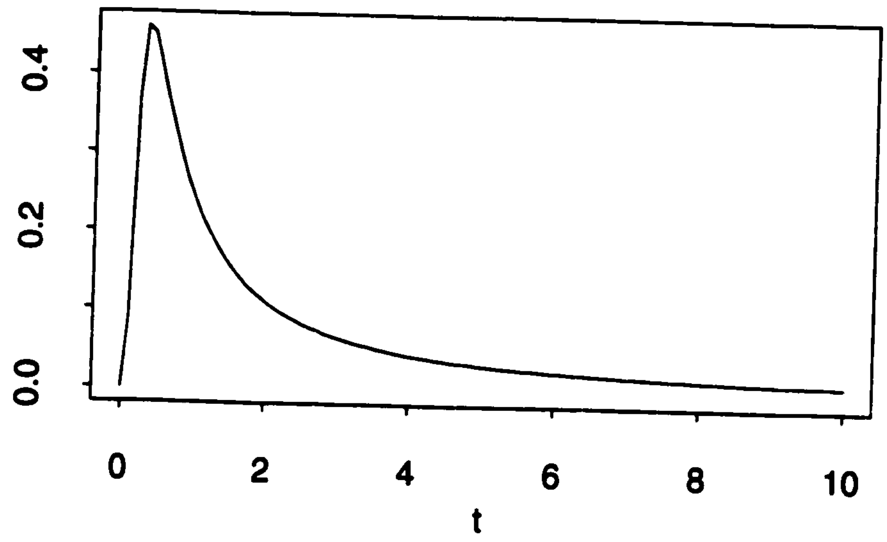


Figure 4.1: The first summand of $f_{uk}(t;u,\ell)$ plotted against t for $u=1, \ell=1, k=0$

Thus we have shown that the series $2\sum_k F_{uk}(t;u,\ell)$ can be differentiated term by term, with respect to t for $t \in [0,\infty)$. **Q.E.D.**

Since the function

$$f_u(t;u,\ell) = -t^{-3/2} \sum_k \left\{ (s_k+\ell)\,\phi\left(\frac{s_k+\ell}{\sqrt{t}}\right) - (s_k-\ell)\,\phi\left(\frac{s_k-\ell}{\sqrt{t}}\right) \right\}. \quad (4.2)$$

is obtained by differentiating the upper subcdf, it is a subdensity function. Similarly we can show that $f_\ell(t;u,\ell)$ can be obtained by differentiating the lower subcdf. Thus $f(t;u,\ell) = f_u(t;u,\ell) + f_\ell(t;u,\ell)$ is a density.

Although it is the case that for finite T,

$$\int_0^T \sum_k f_{uk}(t;u,\ell)\,dt = \sum_k F_{uk}(T;u,\ell),$$

we cannot conclude that the same is true for T replaced by ∞. In other words, while uniform convergence is enough to ensure that a series can be integrated term by term over a finite interval, it is not enough in the case of an unbounded interval. In fact, for the series investigated in this chapter we can show that term by term integration with respect to t over $[0, \infty)$ is not valid. In particular, for the upper subdensity,

$$\int_0^\infty f_u(t;u,\ell)\,dt = \int_0^\infty \sum_k f_{uk}(t;u,\ell)\,dt \neq \sum_k \int_0^\infty f_{uk}(t;u,\ell)\,dt. \tag{4.3}$$

Note that the left hand side of (4.3) equals

$$\int_0^\infty f_u(t;u,\ell)\,dt = \lim_{t\to\infty} F_u(t;u,\ell) = P(D = \mathrm{u})$$

where $P(D = \mathrm{u})$ is the probability that the process ends at the upper barrier. As discussed earlier, we know from difference equation arguments that for the no-drift model, $P(D = \mathrm{u}) = \frac{\ell}{\ell+u}$. However the right hand side of (4.3) equals

$$\sum_k \int_0^\infty f_{uk}(t;u,\ell)\,dt = \sum_k \lim_{t\to\infty} F_{uk}(t;u,\ell) = \sum_k \lim_{t\to\infty} \left\{ \Phi\left(\frac{s_k+\ell}{\sqrt{t}}\right) - \Phi\left(\frac{s_k-\ell}{\sqrt{t}}\right) \right\} = 0.$$

Thus we have shown that term-by-term integration of the series is not valid.

Theorem 4.2 *For $u, \ell > 0$, the function $f_u(t;u,\ell)$ is positive for $t > 0$ and integrates to $F_u(\infty;u,\ell)$.*

Proof: Since $f_u(t;u,\ell)$ is an upper subdensity, it is clear that it is non-negative and integrates to $F_u(\infty;u,\ell)$. It is also easily shown that $f_u(t;u,\ell)$ is continuous in t, since it is the sum of a uniformly convergent series of continuous functions. It will be shown in theorems 4.3 and 5.3 that $f_u(t;u,\ell)$ is strictly positive for small t and for large t. To prove that $f_u(t;u,\ell)$ is strictly positive for intermediate t is more difficult, but an argument using essentially the self-similarity of the Wiener process can be constructed.

Theorem 4.3 *For fixed $u, \ell > 0$ and sufficiently small t,*

$$ct^{-3/2}\phi\left(\frac{u}{\sqrt{t}}\right) \le f_u(t;u,\ell) \le Ct^{-3/2}\phi\left(\frac{u}{\sqrt{t}}\right)$$

for some $c, C > 0$. In particular, $f_u(t;u,\ell)$ approaches 0 as $t \to 0$.

Proof: The first summand of the first term of $f_u(t;u,\ell)$ is $t^{-3/2}u\phi\left(\frac{u}{\sqrt{t}}\right)$. This summand dominates $f_u(t;u,\ell)$ as $t \to 0$, i.e.

$$\frac{-t^{-3/2}(u+2\ell)\phi\left(\frac{u+2\ell}{\sqrt{t}}\right)+\sum_{k=2}^{\infty} f_{uk}}{t^{-3/2}u\phi\left(\frac{u}{\sqrt{t}}\right)}$$

approaches 0 as $t \to 0$. This is true since it can be shown that $\phi\left(\frac{u}{\sqrt{t}}\right)$ can be factored out of the numerator, leaving terms which converge to a sum which goes to 0 as $t \to 0$. Thus as $t \to 0$, $f_u(t;u,\ell)$ behaves like $t^{-3/2}u\phi\left(\frac{u}{\sqrt{t}}\right)$. In particular $f_u(t;u,\ell)$ approaches 0 as $t \to 0$, and for small enough t, we can find constants c and C such that $ct^{-3/2}\phi\left(\frac{u}{\sqrt{t}}\right) \le f_u(t;u,\ell) \le Ct^{-3/2}\phi\left(\frac{u}{\sqrt{t}}\right)$.
Q.E.D.

4.2 First Derivatives

In this section we will show that

- for each fixed t and $\ell > 0$, the series $f_u(t;u,\ell) = \sum f_{uk}(t;u,\ell)$ can be differentiated term by term with respect to u, for $u > 0$, giving $\sum \frac{\partial}{\partial u} f_{uk}(t;u,\ell) = \frac{\partial}{\partial u} f_u(t;u,\ell)$;
- for each fixed $u, \ell > 0$, the series $\frac{\partial}{\partial u} f_u(t;u,\ell) = \sum \frac{\partial}{\partial u} f_{uk}(t;u,\ell)$ is uniformly convergent in t, for $t \ge 0$;
- for small t, $\frac{\partial}{\partial u} f_u(t;u,\ell)$ behaves like $t^{-5/2}\phi\left(\frac{u}{\sqrt{t}}\right)$; and
- for $u, \ell > 0$, $\int_0^\infty \frac{\partial}{\partial u} f_u(t;u,\ell)\,dt = \frac{\partial}{\partial u}\int_0^\infty f_u(t;u,\ell)\,dt$. This will partly satisfy regularity conditions required for consistency and asymptotic normality of the maximum likelihood estimate, as will be discussed in Chapter 5.

We will define $f_{uk}(0;u,\ell) = 0$ and $\frac{\partial}{\partial u} f_{uk}(0;u,\ell) = 0$ so that these functions are continuous at 0.
Theorem 4.4 *For each fixed $t \ge 0$ and $\ell > 0$, the series $f_u(t;u,\ell) = \sum_k f_{uk}(t;u,\ell)$ can be differentiated term by term with respect to u, for $u > 0$;*

Proof: Assume t and ℓ are fixed positive real numbers. Then $f_{uk}(t;u,\ell)$ is differentiable for each u in $[0,\infty)$ and the derivative of $f_{uk}(t;u,\ell)$ is

$$\frac{\partial f_{uk}(t;u,\ell)}{\partial u} = t^{-\frac{5}{2}}(2k+1)\left\{\phi\left(\frac{s_k+\ell}{\sqrt{t}}\right)\left[t-(s_k+\ell)^2\right]-\phi\left(\frac{s_k-\ell}{\sqrt{t}}\right)\left[t-(s_k-\ell)^2\right]\right\}.$$

The function $f_{uk}(t;u,\ell)$ has a right hand derivative at $t = 0$, and this right hand derivative is easily seen to be 0, the limit as $t \to 0$ of $\frac{\partial f_{uk}(t;u,\ell)}{\partial u}$. To show that we can differentiate the series $\sum_k f_{uk}(t;u,\ell)$ with respect to u term by term, we only need to show that $\sum_k \frac{\partial}{\partial u} f_{uk}(t;u,\ell)$ is uniformly convergent in u. We will do this using the Weierstrass M-test.

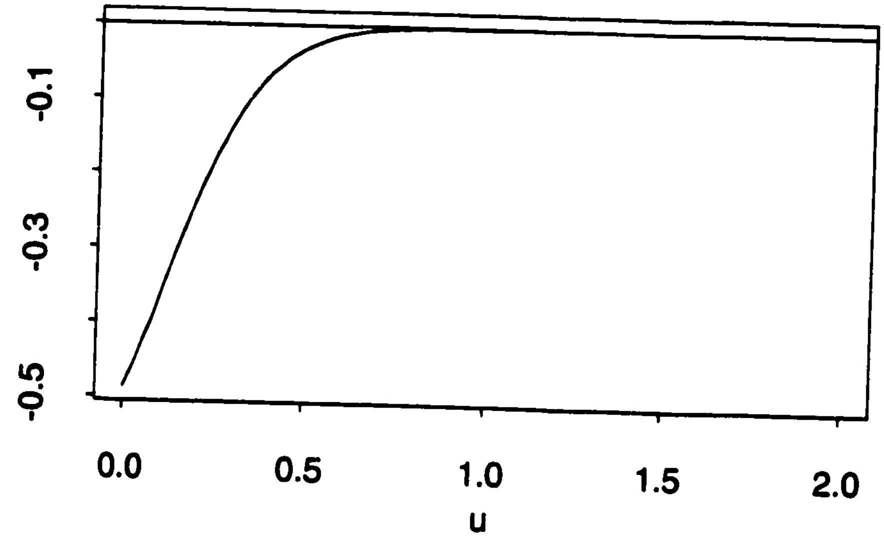


Figure 4.2: The first summand of $\frac{\partial}{\partial u} f_{uk}(t;u,\ell)$ plotted against u for $t = 1, \ell = 1, k = 1$

Let $g_k(t;u,\ell)$ represent the first summand of $\frac{\partial}{\partial u} f_{uk}(t;u,\ell)$, i.e.

$$g_k(t;u,\ell) = t^{-\frac{5}{2}}(2k+1)\phi\left(\frac{s_k+\ell}{\sqrt{t}}\right)\left[t-(s_k+\ell)^2\right]$$

The limit of $g_k(t;u,\ell)$ as $u \to \infty$ is 0. For large enough k, (specifically for $(s_k+\ell)^2 > 3t$), the functions $g_k(t;u,\ell)$ are negative-valued and monotonically increasing for u in $(0,\infty)$.

Let M_k equal the maximum over $u \in (0,\infty)$ of $|g_k(t;u,\ell)|$. For k large enough, this maximum occurs at $u = 0$ and

$$M_k = t^{-5/2}(2k+1)\phi\left(\frac{2k\ell}{\sqrt{t}}\right)\left[(2kl)^2 - t\right].$$

for $k = 0, 1, 2, \ldots$. Using the integral test, we will now show that the series $\sum M_k$ converges. M_k is a sequence of positive terms, which is decreasing for large enough k, say $k > k^*$. Let

$$M(x) = t^{-5/2}(2x+1)\phi\left(\frac{2x\ell}{\sqrt{t}}\right)\left[(2xl)^2 - t\right].$$

Then the integral

$$\int_{k^*}^{\infty} M(x)\, dx$$

converges to

$$\frac{1}{2}\frac{\phi\left(\frac{2k^*\ell}{\sqrt{t}}\right)(2k^*\ell^2(2k^*+1)+t)}{\ell^2 t^{3/2}},$$

which is finite for finite t. Thus the series $\sum M_k$ converges by the integral test, and the series $\sum |g_k(t; u, \ell)|$ converges uniformly in u by the Weierstrass M-test. A similar argument applies to the second summand in $\frac{\partial}{\partial u} f_{uk}(t; u, \ell)$, and so the series $\sum \frac{\partial}{\partial u} f_{uk}(t; u, \ell)$ converges absolutely, and uniformly in u for $u \in (0, \infty)$.

Thus, we have shown that conditions for term-wise differentiation with respect to u of the series $\sum_k f_{uk}(t; u, \ell)$ are satisfied. **Q.E.D.**

Theorem 4.5 *For $u, \ell > 0$, the series $\frac{\partial}{\partial u} f_u(t; u, \ell) = \sum_k \frac{\partial}{\partial u} f_{uk}(t; u, \ell)$ is uniformly convergent in t, for $t \geq 0$.*

Proof: Again let $g_k(t; u, \ell)$ represent the first summand in $\frac{\partial}{\partial u} f_u(t; u, \ell)$. For each $k = 0, 1, 2, \ldots$, the function $g_k(t; u, \ell)$ approaches 0 both as $t \to 0$ and as $t \to \infty$. It has a minimum over $t \in [0, \infty)$ at $t = \frac{3-\sqrt{6}}{3}(s_k + \ell)^2$, where it has a negative value, and a maximum at $t = \frac{3+\sqrt{6}}{3}(s_k + \ell)^2$, where it is positive-valued.

The function $|g_k(t; u, \ell)|$ achieves a maximum over $t \in [0, \infty)$ at $t = \frac{3-\sqrt{6}}{3}(s_k + \ell)^2$. Let M_k denote the maximum value of this function. Clearly, M_k is $O(k^{-2})$, and so the series $\sum_{k=0}^{\infty} M_k$ converges. By the Weierstrass M-test, the series $\sum |g_k(t; u, \ell)|$ converges uniformly in t. A similar argument applies to the second summand in $\frac{\partial}{\partial u} f_u(t; u, \ell)$. Thus $\sum_k \frac{\partial}{\partial u} f_{uk}(t; u, \ell)$ is absolutely, and uniformly convergent in t, for $t \geq 0$. Also, $|\frac{\partial}{\partial u} f_u(t; u, \ell)|$ is bounded. **Q.E.D.**

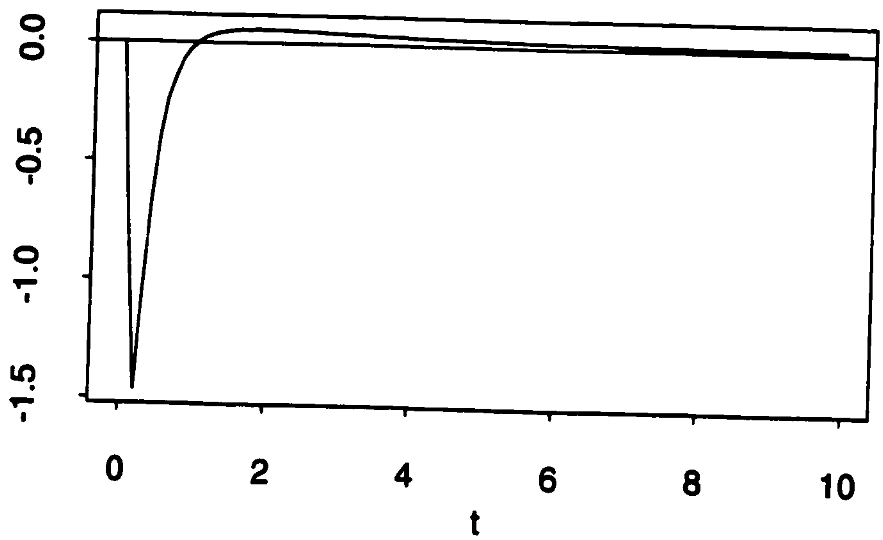


Figure 4.3: The first summand of $\frac{\partial}{\partial u} f_{uk}(t; u, \ell)$ plotted against t for $u = 1, \ell = 1, k = 0$

Theorem 4.6 *For fixed $u, \ell > 0$ and sufficiently small t,*

$$ct^{-5/2}\phi\left(\frac{u}{\sqrt{t}}\right) \leq \frac{\partial}{\partial u} f_u(t; u, \ell) \leq Ct^{-5/2}\phi\left(\frac{u}{\sqrt{t}}\right)$$

for some $c, C > 0$.

Proof: As in theorem 4.3, we argue that the first summand of the first term of $\frac{\partial}{\partial u} f_{uk}(t; u, \ell)$ dominates as $t \to 0$. **Q.E.D.**

To show that regularity conditions (for consistency and asymptotic normality of estimates) are satisfied, we need to prove further properties of the subdensities and their derivatives. For example, if $f(t; \theta)$ is a density with parameter θ, we need to find an integrable function $g(t)$ such that

$$\left|\frac{\partial f(t;\theta)}{\partial \theta}\right| \leq g(t)$$

However this turns out to be difficult to do in the case of parameters $\theta = (u, \ell)$. While we can find analytically the location and value of the maximum of each function $\left|\frac{\partial f_k(t;\theta)}{\partial \theta}\right|$ it is not possible to determine the location and value of the maximum of the summed function $\frac{\partial f(t;\theta)}{\partial\theta} = \sum \frac{\partial f_k(t;\theta)}{\partial\theta}$. Thus we discuss, in this chapter, weaker conditions that can be shown using only the representations for $f_u(t;u,\ell)$ and $f_\ell(t;u,\ell)$ analogous to (4.2), leaving the full examination of regularity conditions to Chapter 5.

An important sufficient condition for consistency of the maximum likelihood estimate of a parameter θ is that the expectation of the score is 0, ie

$$E_\theta\left(\frac{\partial}{\partial\theta}\log(\mathcal{L}(\theta;t))\right) = 0.$$

where E_θ denotes expectation with respect to θ, and $\mathcal{L}(\theta;t)$ represents the likelihood. For the FP2B models considered here, we can write the log likelihood (for a single individual) as

$$\mathcal{L}(u,\ell;t) = \delta\log f_u(t;u,\ell) + (1-\delta)\log f_\ell(t;u,\ell)$$

where $\delta = 1$ if the individual was discharged, and $\delta = 0$ if the individual died in hospital. Then the score function (for a single individual) is

$$S(u,\ell;t) = \delta\left(\frac{\partial}{\partial\theta}\log f_u(t;u,\ell)\right) + (1-\delta)\left(\frac{\partial}{\partial\theta}\log f_\ell(t;u,\ell)\right)$$

where $\theta = (u,\ell)$. The expected value of the score is then

$$E_\theta(S(u,\ell;t)) = \int_0^\infty \left\{\left(\frac{\partial}{\partial\theta}\log f_u(t;u,\ell)\right) f_u(t;u,\ell) + \left(\frac{\partial}{\partial\theta}\log f_\ell(t;u,\ell)\right) f_\ell(t;u,\ell)\right\}\, dt,$$

because the probability that $T \in [t,t+dt)$ and $\delta = 1$ is $f_u(t;u,\ell)dt$, while the probability that $T \in [t,t+dt)$ and $\delta = 0$ is $f_\ell(t;u,\ell)dt$. Then the above equals

$$\int_0^\infty \frac{\partial}{\partial\theta} f_u(t;u,\ell)\, dt + \int_0^\infty \frac{\partial}{\partial\theta} f_\ell(t;u,\ell)\, dt,$$

which equals 0, as long as differentiation and integration can be interchanged. More precisely, the expected value of the score equals 0 if differentiation with respect to θ and integration over the interval $(0,\infty)$ can be interchanged for both $f_u(t;u,\ell)$ and $f_\ell(t;u,\ell)$.

We will now show that this interchange is valid for the upper subdensity of the FP2B distribution and the component u of θ. Because $f_u(t;\theta)$ in the FP2B case is an infinite series which cannot be integrated term by term over $(0, \infty)$, the situation is rather delicate. We will need the following theorem.

Theorem 4.7 *Let $\psi(x)$ be an integrable function defined on $(-\infty, u]$, $u > 0$, whose derivative $\psi'(x)$ exists on $(-\infty, u)$ and satisfies $\int_{-\infty}^{0} |\psi'(v)|dv < \infty$; and let $\Psi(x) = \int_{-\infty}^{x} \psi(v)dv$. Then*

$$\lim_{T\to\infty} \sum_{k=0}^{\infty} \left\{ \Psi\left(\frac{s_k+\ell}{\sqrt{T}}\right) - \Psi\left(\frac{s_k-\ell}{\sqrt{T}}\right) \right\} = \left(\frac{\ell}{u+\ell}\right) \Psi(0)$$

where $s_k = -(2k+1)(u+\ell)$ and $u, \ell, T > 0$.

Proof: Let $t_k = -2k(u+\ell)$. Note that $t_k - u = s_k + \ell$, and $t_k + u = s_m - \ell$, where $m = k-1$. Then

$$\Psi(0) = \int_{-\infty}^{0} \psi(v)dv = \sum_k \left\{ \Psi\left(\frac{s_k+\ell}{\sqrt{T}}\right) - \Psi\left(\frac{s_k-\ell}{\sqrt{T}}\right) + \Psi\left(\frac{t_k+u}{\sqrt{T}}\right) - \Psi\left(\frac{t_k-u}{\sqrt{T}}\right) \right\} - \left[\Psi\left(\frac{u}{\sqrt{T}}\right) - \Psi(0) \right]. \tag{4.4}$$

For each k,

$$\Psi\left(\frac{s_k+\ell}{\sqrt{T}}\right) - \Psi\left(\frac{s_k-\ell}{\sqrt{T}}\right) \approx \psi\left(\frac{s_k+\ell}{\sqrt{T}}\right) \frac{2\ell}{\sqrt{T}}$$

and

$$\Psi\left(\frac{t_k+u}{\sqrt{T}}\right) - \Psi\left(\frac{t_k-u}{\sqrt{T}}\right) \approx \psi\left(\frac{s_k+\ell}{\sqrt{T}}\right) \frac{2u}{\sqrt{T}} \approx \left[\Psi\left(\frac{s_k+\ell}{\sqrt{T}}\right) - \Psi\left(\frac{s_k-\ell}{\sqrt{T}}\right) \right] \frac{u}{\ell}.$$

Then

$$\Psi\left(\frac{s_k+\ell}{\sqrt{T}}\right) - \Psi\left(\frac{s_k-\ell}{\sqrt{T}}\right) + \Psi\left(\frac{t_k+u}{\sqrt{T}}\right) - \Psi\left(\frac{t_k-u}{\sqrt{T}}\right) \approx \left[\Psi\left(\frac{s_k+\ell}{\sqrt{T}}\right) - \Psi\left(\frac{s_k-\ell}{\sqrt{T}}\right) \right] \left[1 + \frac{u}{\ell}\right]. \tag{4.5}$$

We show below that, as $T \to \infty$, we can take the approximation to be exact. Substituting expression 4.5 into expression 4.4, taking the limit as $T \to \infty$ and rearranging gives

$$\left(\frac{\ell}{u+\ell}\right) \Psi(0) = \sum_k \left\{ \Psi\left(\frac{s_k+\ell}{\sqrt{T}}\right) - \Psi\left(\frac{s_k-\ell}{\sqrt{T}}\right) \right\}.$$

We now show that the error in the approximation goes to 0 as $T \to \infty$. For each k,

$$\left|\Psi\left(\frac{s_k+\ell}{\sqrt{T}}\right)-\Psi\left(\frac{s_k-\ell}{\sqrt{T}}\right)-\psi\left(\frac{s_k+\ell}{\sqrt{T}}\right)\frac{2\ell}{\sqrt{T}}\right| \le \frac{1}{2}\psi'\left(\frac{\theta_k}{\sqrt{T}}\right)\left(\frac{2\ell}{\sqrt{T}}\right)^2$$

where $\theta_k \in (s_k-\ell, s_k+\ell)$ and

$$\left|\Psi\left(\frac{t_k+u}{\sqrt{T}}\right)-\Psi\left(\frac{t_k-u}{\sqrt{T}}\right)-\frac{u}{\ell}\left\{\Psi\left(\frac{s_k+\ell}{\sqrt{T}}\right)-\Psi\left(\frac{s_k-\ell}{\sqrt{T}}\right)\right\}\right| \le \frac{1}{2}\psi'\left(\frac{\zeta_k}{\sqrt{T}}\right)\left(\frac{2u}{\sqrt{T}}\right)^2 + \left(\frac{u}{2\ell}\right)\psi'\left(\frac{\theta_k}{\sqrt{T}}\right)\left(\frac{2\ell}{\sqrt{T}}\right)^2$$

where $\zeta_k \in (t_k-u, t_k+u)$. Then the error in approximating the sum in 4.4 by $\sum_k\left\{\Psi\left(\frac{s_k+\ell}{\sqrt{T}}\right)-\Psi\left(\frac{s_k-\ell}{\sqrt{T}}\right)\right\}[1+\frac{u}{\ell}]$ is less than or equal to

$$\frac{K_1}{T}\sum_k\psi'\left(\frac{\theta_k}{\sqrt{T}}\right)+\frac{K_2}{T}\sum_k\psi'\left(\frac{\zeta_k}{\sqrt{T}}\right) \le \frac{K_3}{\sqrt{T}}\int_{-\infty}^{0}\psi'(w)\,dw,$$

which approaches 0 as $T \to \infty$, as long as the integral in the above expression converges. Thus we can ignore the error in the approximation as long as $\int_{-\infty}^{0}\psi'(w)\,dw < \infty$. **Q.E.D.**

Note that we have also proved the following corollary:

Corollary:

$$\lim_{T\to\infty}\frac{2\ell}{\sqrt{T}}\sum_k\psi\left(\frac{s_k+\ell}{\sqrt{T}}\right)=\lim_{T\to\infty}\frac{2\ell}{\sqrt{T}}\sum_k\psi\left(\frac{s_k-\ell}{\sqrt{T}}\right)=\left(\frac{\ell}{u+\ell}\right)\Psi(0)$$

As an aside, we can use theorem 4.7 to show that

$$\int_0^\infty f_u(t;u,\ell)\,dt=\lim_{t\to\infty}F_u(t;u,\ell)=\frac{\ell}{u+\ell},$$

without recourse to difference equation arguments. Recall that, for the upper subdistribution function $F_u(t;u,\ell)$, we have

$$\frac{1}{2}F_u(t;u,\ell)=\sum_k\left\{\Phi\left(\frac{s_k+\ell}{\sqrt{t}}\right)-\Phi\left(\frac{s_k-\ell}{\sqrt{t}}\right)\right\}.$$

We apply theorem 4.7, taking the function $\Psi(x)$ to be the standard normal cdf $\Phi(x)$. As $t \to \infty$, the sum in the above expression approaches $\frac{\ell}{u+\ell}\Phi(0)$ where $\Phi(0)=1/2$. Therefore, $\lim_{t\to\infty}F_u(t;u,\ell)=\frac{\ell}{u+\ell}$.

Theorem 4.8 *For $u, \ell > 0$,*

$$\frac{\partial}{\partial u}\int_0^\infty f_u(t;u,\ell)\,dt = \int_0^\infty \frac{\partial}{\partial u} f_u(t;u,\ell)\,dt \tag{4.6}$$

taking each $\int_0^\infty$ to mean $\lim_{T\to\infty}\int_0^T$.

Proof: From difference equation arguments, we know that $\int_0^\infty f_u(t;u,\ell)\,dt = \frac{\ell}{u+\ell}$, so $\frac{\partial}{\partial u}\int_0^\infty f_u(t;u,\ell)\,dt = \frac{\partial}{\partial u}\left(\frac{\ell}{u+\ell}\right) = -\frac{\ell}{(u+\ell)^2}$. The right hand side of equation 4.6 is

$$\begin{aligned}\int_0^\infty \frac{\partial}{\partial u} f_u(t;u,\ell)\,dt &= \lim_{T\to\infty}\int_0^T \sum_k \frac{\partial}{\partial u} f_{uk}(t;u,\ell)\,dt \\ &= \lim_{T\to\infty}\sum_k \int_0^T \frac{\partial}{\partial u} f_{uk}(t;u,\ell)\,dt \\ &= -2\lim_{T\to\infty}\sum_k \left(\frac{2k+1}{\sqrt{T}}\right)\left\{\phi\left(\frac{s_k+\ell}{\sqrt{T}}\right) - \phi\left(\frac{s_k-\ell}{\sqrt{T}}\right)\right\}\end{aligned} \tag{4.7}$$

Note that interchange of integration over the bounded interval $(0,T)$ and summation in equation (4.7) is valid since $\sum \frac{\partial}{\partial u} f_{uk}(t;u,\ell)$ is uniformly convergent in t, as we proved in theorem 4.5. Now since $(2k+1) = -s_k/(u+\ell)$ the last expression equals

$$\begin{aligned}\left(\frac{2}{u+\ell}\right)\lim_{T\to\infty}\sum_k \Bigg\{&\left[\left(\frac{s_k+\ell}{\sqrt{T}}\right)\phi\left(\frac{s_k+\ell}{\sqrt{T}}\right) - \left(\frac{s_k-\ell}{\sqrt{T}}\right)\phi\left(\frac{s_k-\ell}{\sqrt{T}}\right)\right] \\ &- \frac{\ell}{\sqrt{T}}\left[\phi\left(\frac{s_k+\ell}{\sqrt{T}}\right) - \phi\left(\frac{s_k-\ell}{\sqrt{T}}\right)\right] \\ &- \frac{2\ell}{\sqrt{T}}\,\phi\left(\frac{s_k-\ell}{\sqrt{T}}\right)\Bigg\}\end{aligned}$$

We now apply theorem 4.7 to the first line of the above, taking $\Psi(x)$ to be $x\phi(x)$. Since the value of this function at 0 is 0, the first line in the above approaches 0 as $T\to\infty$. For the second line, we take $\Psi(x)$ to be $\phi(x)$ whose value at 0 is $1/\sqrt{2\pi}$. Thus the second line approaches 0 as well. By the corollary, the last line approaches $-\frac{\ell}{2(u+\ell)}$. Thus the RHS of expression 4.6 equals $-\frac{\ell}{(u+\ell)^2}$ as desired. **Q.E.D.**

4.3 Second Derivatives

In this section we will show that

- for each fixed $t \geq 0$ and $\ell > 0$, the series $\frac{\partial}{\partial u} f_u(t; u, \ell) = \sum \frac{\partial}{\partial u} f_{uk}(t; u, \ell)$ can be differentiated term by term with respect to u, for $u > 0$, giving $\sum \frac{\partial^2}{\partial u^2} f_{uk}(t; u, \ell) = \frac{\partial^2}{\partial u^2} f_u(t; u, \ell)$;
- for each fixed $u, \ell > 0$, the series $\frac{\partial^2}{\partial u^2} f_u(t; u, \ell) = \sum \frac{\partial^2}{\partial u^2} f_{uk}(t; u, \ell)$ is uniformly convergent in t, for $t \geq 0$;
- for small t, $\frac{\partial^2}{\partial u^2} f_u(t; u, \ell)$ behaves like $t^{-7/2}\phi\left(\frac{u}{\sqrt{t}}\right)$; and
- for $u, \ell > 0$, $\frac{\partial^2}{\partial u^2} \int_0^\infty f_u(t; u, \ell)\, dt = \int_0^\infty \frac{\partial^2}{\partial u^2} f_u(t; u, \ell)\, dt$, thus partially satisfying the regularity conditions required for consistency and asymptotic normality of the maximum likelihood estimate, as will be discussed in Chapter 5.

We will define $\frac{\partial^2}{\partial u^2} f(0; u, \ell) = 0$, making this function continuous at 0. All the proofs are exactly analogous to those in the previous section.

Theorem 4.9 *For each fixed $t \geq 0$ and $\ell > 0$, the series $\frac{\partial}{\partial u} f_u(t; u, \ell) = \sum_k \frac{\partial}{\partial u} f_{uk}(t; u, \ell)$ can be differentiated term by term with respect to u, for $u > 0$;*

Proof: Assume t and ℓ are fixed. Then $\frac{\partial}{\partial u} f_{uk}(t; u, \ell)$ is differentiable for each u in $[0, \infty)$ and its derivative is

$$\frac{\partial^2}{\partial u^2} f_{uk}(t; u, \ell) = t^{-\frac{7}{2}}(2k+1)^2 \left\{ \phi\left(\frac{s_k + \ell}{\sqrt{t}}\right)(s_k + \ell)\left[3t - (s_k + \ell)^2\right] - \phi\left(\frac{s_k - \ell}{\sqrt{t}}\right)(s_k - \ell)\left[3t - (s_k - \ell)^2\right] \right\}.$$

The function $\frac{\partial f_{uk}(t; u, \ell)}{\partial u}$ has a right hand derivative at $t = 0$, and this right hand derivative is easily seen to be 0, the limit as $t \to 0$ of $\frac{\partial^2}{\partial u^2} f_{uk}(t; u, \ell)$. To show that we can differentiate the series $\sum_k \frac{\partial}{\partial u} f_{uk}(t; u, \ell)$ with respect to u term by term, we only need to show that $\sum_k \frac{\partial^2}{\partial u^2} f_{uk}(t; u, \ell)$ is uniformly convergent in u. We will do this using the Weierstrass M-test.

Let $g_k(t; u, \ell)$ represent the first summand of $\frac{\partial^2}{\partial u^2} f_{uk}(t; u, \ell)$, i.e.

$$g_k(t; u, \ell) = t^{-\frac{7}{2}}(2k+1)^2 \phi\left(\frac{s_k + \ell}{\sqrt{t}}\right)(s_k + \ell)\left[3t - (s_k + \ell)^2\right]$$

The limit of $g_k(t; u, \ell)$ as $u \to \infty$ is 0. For large enough k, (specifically for $k > \frac{3+\sqrt{6}}{4\ell^2} t$), the functions $g_k(t; u, \ell)$ are positive-valued and monotonically decreasing for u in $(0, \infty)$.

Let M_k equal the maximum over $u \in (0, \infty)$ of $|g_k(t; u, \ell)|$. For large enough k, this maximum occurs at $u = 0$ and

$$M_k = t^{-7/2}(2k+1)^2 \phi\left(\frac{2k\ell}{\sqrt{t}}\right) 2k\ell(4k^2\ell^2 - 3t)$$

Using the integral test, we will now show that the series $\sum M_k$ converges. The sequence $\{M_k\}$ is positive for $k > \sqrt{3t}/2l$. Furthermore, it is decreasing for large enough k (say $k > k^*$), because the term in ϕ will dominate all other terms in the expression for M_k. The integral

$$\int_{k^*}^{\infty} M(x)\, dx$$

converges to

$$\frac{1}{2} \frac{\phi\left(\frac{2k^*\ell}{\sqrt{t}}\right) \{(16k^{*4} + 16k^{*3} + 4k^{*2})\ell^4 + (4tk^{*2} - t)\ell^2 + 2t^2\}}{\ell^3 t^{5/2}}$$

which is finite for finite t and ℓ. Thus the series $\sum M_k$ converges by the integral test, and the series $\sum |g_k(t; u, \ell)|$ converges uniformly in u by the Weierstrass M-test. A similar argument applies to the second summand in $\frac{\partial^2}{\partial u^2} f_{uk}(t; u, \ell)$, and so the series $\sum \frac{\partial^2}{\partial u^2} f_{uk}(t; u, \ell)$ converges absolutely, and uniformly in u for $u \in (0, \infty)$.

Thus, we have shown that conditions for term-wise differentiation with respect to u of the series $\sum_k \frac{\partial}{\partial u} f_{uk}(t; u, \ell)$ are satisfied. **Q.E.D.**

Theorem 4.10 *For $u, \ell > 0$, the series $\frac{\partial^2}{\partial u^2} f_u(t; u, \ell) = \sum_k \frac{\partial^2}{\partial u^2} f_{uk}(t; u, \ell)$ is uniformly convergent in t, for $t \geq 0$.*

Proof: Again let $g_k(t; u, \ell)$ represent the first summand in $\frac{\partial^2}{\partial u^2} f_u(t; u, \ell)$. For each $k = 0, 1, 2, \ldots$, the function $g_k(t; u, \ell)$ approaches 0 both as $t \to 0$ and as $t \to \infty$. It has a minimum over $t \in [0, \infty)$ at $t = \frac{5-\sqrt{10}}{15}(s_k + \ell)^2$, where it has a negative value, and a maximum at $t = \frac{5+\sqrt{10}}{15}(s_k + \ell)^2$, where it is positive-valued.

The function $|g_k(t; u, \ell)|$ achieves a maximum over $t \in [0, \infty)$ at $t = \frac{5-\sqrt{10}}{15}(s_k + \ell)^2$. Let M_k denote the maximum value of this function. Clearly, M_k is $O(k^{-2})$, and so the series $\sum_{k=0}^{\infty} M_k$ converges. By the Weierstrass M-test, the series $\sum |g_k(t; u, \ell)|$ converges uniformly in t. A similar argument applies to the second summand in $\frac{\partial^2}{\partial u^2} f_u(t; u, \ell)$. Thus $\sum_k \frac{\partial^2}{\partial u^2} f_{uk}(t; u, \ell)$ is absolutely, and uniformly convergent in t, for $t \geq 0$. Also, $|\frac{\partial^2}{\partial u^2} f_u(t; u, \ell)|$ is bounded. **Q.E.D.**

Theorem 4.11 *For fixed $u, \ell > 0$ and sufficiently small t,*

$$ct^{-7/2}\phi\left(\frac{u}{\sqrt{t}}\right) \leq \frac{\partial^2}{\partial u^2} f_u(t; u, \ell) \leq Ct^{-7/2}\phi\left(\frac{u}{\sqrt{t}}\right)$$

for some $c, C > 0$.

Proof: As in theorem 4.3, we argue that the first summand of the first term of $\frac{\partial^2}{\partial u^2} f_{uk}(t; u, \ell)$ dominates as $t \to 0$. **Q.E.D.**

A further condition for consistency of the maximum likelihood estimate of a parameter θ is that the variance of the score equal minus the Fisher information, ie

$$E_\theta\left(S(\theta; t)^2\right) = -E_\theta\left(\frac{\partial^2}{\partial \theta^2}\mathcal{L}(\theta; t)\right).$$

For the FP2B models, the score for a single individual, when squared, equals

$$S(u, \ell; t)^2 = \left(\frac{\partial}{\partial \theta}\mathcal{L}(u, \ell; t)\right)^2 =$$

$$\left(\delta\frac{\partial}{\partial \theta}\log f_u(t; u, \ell)\right)^2 + 2\delta(\delta - 1)\left(\frac{\partial}{\partial \theta}\log f_u(t; u, \ell)\right)\left(\frac{\partial}{\partial \theta}\log f_\ell(t; u, \ell)\right) + \left((1 - \delta)\frac{\partial}{\partial \theta}\log f_\ell(t; u, \ell)\right)^2$$

where $\theta = (u, \ell)$ and where $\delta = 1$ if the individual is discharged and $\delta = 0$ if the individual dies in hospital. Note that the middle term vanishes, since $\delta(1 - \delta)$ is always 0. Then

$$\begin{aligned} E_\theta\left(S(u, \ell; t)^2\right) &= \int_0^\infty \left(\frac{\partial}{\partial \theta}\log f_u(t; u, \ell)\right)^2 f_u(t; u, \ell)\, dt + \int_0^\infty \left(\frac{\partial}{\partial \theta}\log f_\ell(t; u, \ell)\right)^2 f_\ell(t; u, \ell)\, dt \\ &= \int_0^\infty \frac{\frac{\partial^2}{\partial \theta^2} f_u(t; u, \ell)}{f_u(t; u, \ell)}\, dt + \int_0^\infty \frac{\frac{\partial^2}{\partial \theta^2} f_\ell(t; u, \ell)}{f_\ell(t; u, \ell)}\, dt. \end{aligned} \tag{4.8}$$

Also, for the FP2B model,

$$\frac{\partial^2}{\partial \theta^2}\mathcal{L}(u, \ell; t) = \delta\frac{\partial^2}{\partial \theta^2}\log f_u(t; u, \ell) + (1 - \delta)\frac{\partial^2}{\partial \theta^2}\log f_\ell(t; u, \ell).$$

It can be shown that

$$
\begin{aligned}
&-E_\theta\left(\frac{\partial^2}{\partial\theta^2}\mathcal{L}(u,\ell;t)\right) = \\
&-\int_0^\infty \frac{\partial^2}{\partial\theta^2}f_u(t;u,\ell)\,dt + \int_0^\infty \frac{\frac{\partial^2}{\partial\theta^2}f_u(t;u,\ell)}{f_u(t;u,\ell)}\,dt - \int_0^\infty \frac{\partial^2}{\partial\theta^2}f_\ell(t;u,\ell)\,dt + \int_0^\infty \frac{\frac{\partial^2}{\partial\theta^2}f_\ell(t;u,\ell)}{f_\ell(t;u,\ell)}\,dt.
\end{aligned}
\tag{4.9}
$$

If we can differentiate twice under the integral, then

$$
\int \frac{\partial^2}{\partial\theta^2}f_u(t;u,\ell)\,dt + \int \frac{\partial^2}{\partial\theta^2}f_\ell(t;u,\ell)\,dt = \frac{\partial^2}{\partial\theta^2}\int (f_u(t;u,\ell) + f_\ell(t;u,\ell))\,dt
$$

equals 0 since $f(t;u,\ell) = f_u(t;u,\ell) + f_\ell(t;u,\ell)$ is a density and so integrates to 1. In that case, equation 4.8 will equal equation 4.9. Thus if we can differentiate each subdensity twice with respect to θ under the integral with respect to t on $(0,\infty)$, the variance of the score will equal minus the Fisher information.

We now show that we can differentiate $f_u(t;u,\ell)$ twice with respect to component u of θ under the integral.

Theorem 4.12 *For $u,\ell > 0$,*

$$
\frac{\partial^2}{\partial u^2}\int_0^\infty f_u(t;u,\ell)\,dt = \int_0^\infty \frac{\partial^2}{\partial u^2}f_u(t;u,\ell)\,dt \tag{4.10}
$$

Proof: From theorem 4.8, we have $\frac{\partial}{\partial u}\int_0^\infty f_u(t;u,\ell)\,dt = -\frac{\ell}{(u+\ell)^2}$ so $\frac{\partial^2}{\partial u^2}\int_0^\infty f_u(t;u,\ell)\,dt = \frac{2\ell}{(u+\ell)^3}$. Now

$$
\begin{aligned}
\int_0^\infty \frac{\partial^2}{\partial u^2}f_u(t;u,\ell)\,dt &= \lim_{T\to\infty}\int_0^T \sum_k \frac{\partial^2}{\partial u^2}f_{uk}(t;u,\ell)\,dt \\
&= \lim_{T\to\infty}\sum_k \int_0^T \frac{\partial^2}{\partial u^2}f_{uk}(t;u,\ell)\,dt \\
&= \lim_{T\to\infty}\sum_k -2T^{-3/2}(2k+1)^2\left\{\phi\left(\frac{s_k+\ell}{\sqrt{T}}\right)(s_k+\ell) - \phi\left(\frac{s_k-\ell}{\sqrt{T}}\right)(s_k-\ell)\right\}
\end{aligned}
\tag{4.11}
$$

Note that interchange of integration and summation in equation 4.11 is valid since $\sum \frac{\partial^2}{\partial u^2}f_{uk}(t;u,\ell)$ is uniformly convergent in t, as we proved in theorem 4.10. Noting that $(2k+1) = -s_k/(u+\ell)$

we can write the above as

$$\lim_{T\to\infty} -\frac{2}{(u+\ell)^2}\sum_k \left\{\left[\Psi_1\left(\frac{s_k+\ell}{\sqrt{T}}\right) - \Psi_1\left(\frac{s_k-\ell}{\sqrt{T}}\right)\right]\right.$$
$$-\frac{2\ell}{\sqrt{T}}\left[\Psi_2\left(\frac{s_k+\ell}{\sqrt{T}}\right) - \Psi_2\left(\frac{s_k-\ell}{\sqrt{T}}\right)\right]$$
$$+\frac{\ell^2}{T}\left[\Psi_3\left(\frac{s_k+\ell}{\sqrt{T}}\right) - \Psi_3\left(\frac{s_k-\ell}{\sqrt{T}}\right)\right]$$
$$\left.-2\frac{2\ell}{\sqrt{T}}\Psi_2\left(\frac{s_k-\ell}{\sqrt{T}}\right)\right\}$$

where $\Psi_1(x) = x^3\phi(x)$, $\Psi_2(x) = x^2\phi(x)$, and $\Psi_3(x) = x\phi(x)$. Note that $\Psi_1(0) = \Psi_2(0) = \Psi_3(0) = 0$, so that, applying theorem 4.7, the first three lines of the above expression approach 0 as $T \to \infty$. Let $\Psi_4(x) = \int_{-\infty}^{x} \Psi_2(s)ds = \Phi(x) - x\phi(x)$. Then by the corollary, the last line of the above equals $-2\frac{\ell}{u+\ell}\Psi_4(0) = -\frac{\ell}{u+\ell}$. So finally the RHS of equation 4.10 equals $\frac{2\ell}{(u+\ell)^3}$ as desired.
Q.E.D.

4.4 Third Derivatives

In this section we will show that

- for each fixed $t \geq 0$ and $\ell > 0$, the series $\frac{\partial^2}{\partial u^2}f_u(t;u,\ell) = \sum \frac{\partial^2}{\partial u^2}f_{uk}(t;u,\ell)$ can be differentiated term by term with respect to u, for $u > 0$, giving $\sum \frac{\partial^3}{\partial u^3}f_{uk}(t;u,\ell) = \frac{\partial^3}{\partial u^3}f_u(t;u,\ell)$;
- for each fixed $u, \ell > 0$, the series $\frac{\partial^3}{\partial u^3}f_u(t;u,\ell) = \sum \frac{\partial^3}{\partial u^3}f_{uk}(t;u,\ell)$ is uniformly convergent in t, for $t \geq 0$; and
- for small t, $\frac{\partial^3}{\partial u^3}f_u(t;u,\ell)$ behaves like $t^{-9/2}\phi\left(\frac{u}{\sqrt{t}}\right)$.

We will define $\frac{\partial^3}{\partial u^3}f(0;u,\ell) = 0$, making this function continuous at 0. All the proofs are exactly analogous to those in the previous section.

Theorem 4.13 *For each fixed $t \geq 0$ and $\ell > 0$, the series $\frac{\partial^2}{\partial u^2}f_u(t;u,\ell) = \sum_k \frac{\partial^2}{\partial u^2}f_{uk}(t;u,\ell)$ can be differentiated term by term with respect to u, for $u > 0$;*

Proof: Assume t and ℓ are fixed. Then $\frac{\partial^2}{\partial u^2} f_{uk}(t;u,\ell)$ is differentiable for each u in $[0,\infty)$ and its derivative is

$$\frac{\partial^3}{\partial u^3} f_{uk}(t;u,\ell) = t^{-\frac{9}{2}}(2k+1)^3 \left\{ \phi\left(\frac{s_k+\ell}{\sqrt{t}}\right) \left[6(s_k+\ell)^2 t - (s_k+\ell)^4 - 3t^2\right] - \phi\left(\frac{s_k-\ell}{\sqrt{t}}\right) \left[6(s_k+\ell)^2 t - (s_k+\ell)^4 - 3t^2\right] \right\}.$$

The function $\frac{\partial^2 f_{uk}(t;u,\ell)}{\partial u^2}$ has a right hand derivative at $t=0$, and this right hand derivative is easily seen to be 0, the limit as $t \to 0$ of $\frac{\partial^3}{\partial u^3} f_{uk}(t;u,\ell)$. To show that we can differentiate the series $\sum_k \frac{\partial^2}{\partial u^2} f_{uk}(t;u,\ell)$ with respect to u term by term, we only need to show that $\sum_k \frac{\partial^3}{\partial u^3} f_{uk}(t;u,\ell)$ is uniformly convergent in u. We will do this using the Weierstrass M-test.

Let $g_k(t;u,\ell)$ represent the first summand of $\frac{\partial^3}{\partial u^3} f_{uk}(t;u,\ell)$. We now show that for large enough k, the functions $g_k(t;u,\ell)$ are negative-valued and monotonically increasing for u in $(0,\infty)$. Differentiating $g_k(t;u,\ell)$ with respect to u and setting the result equal to 0 shows that $g_k(t;u,\ell)$ has five critical points. Three of these are always negative. The remaining two are

$$\frac{-2k\ell + c_1\sqrt{t}}{2k+1}, \quad \frac{-2k\ell + c_2\sqrt{t}}{2k+1}$$

where c_1 and c_2 are positive constants. These two critical points will be negative for large enough k. By differentiating again, we can show that c_1 is a minimum, and that $g_k(t;u,\ell)$ is negative valued at this minimum, and approaches 0 as $u \to 0$. Thus $g_k(t;u,\ell)$ is negative and monotonically increasing on $u \in (0,\infty)$ for large enough k.

Let M_k equal the maximum over $u \in (0,\infty)$ of $|g_k(t;u,\ell)|$. For large enough k, this maximum occurs at $u=0$. The sequence M_k is positive and decreasing for large enough k (say $k > k^*$), because the term in ϕ will dominate all other terms in the expression for M_k. The integral

$$\int_{k^*}^{\infty} M(x)\, dx$$

converges to

$$\frac{\phi\left(\frac{2k^*\ell}{\sqrt{t}}\right) h(t,k^*)}{t^{7/2}\ell^4}$$

where $h(t, k^*)$ is a sixth degree polynomial in t and k. This expression is finite for finite t and ℓ. Thus the series $\sum M_k = \sum |g_k(t; 0, \ell)|$ converges by the integral test, and so the series $\sum |g_k(t; u, \ell)|$ converges uniformly in u by the Weierstrass M-test. A similar argument applies to the second summand in $\frac{\partial^3}{\partial u^2} f_{uk}(t; u, \ell)$, and so the series $\sum \frac{\partial^3}{\partial u^2} f_{uk}(t; u, \ell)$ converges absolutely, and uniformly in u for $u \in (0, \infty)$.

Thus, we have shown that conditions for term-wise differentiation with respect to u of the series $\sum_k \frac{\partial^2}{\partial u^2} f_{uk}(t; u, \ell)$ are satisfied. **Q.E.D.**

Theorem 4.14 *For $u, \ell > 0$, the series $\frac{\partial^3}{\partial u^2} f_u(t; u, \ell) = \sum_k \frac{\partial^2}{\partial u^2} f_{uk}(t; u, \ell)$ is uniformly convergent in t, for $t \geq 0$.*

Proof: Again let $g_k(t; u, \ell)$ represent the first summand in $\frac{\partial^3}{\partial u^2} f_u(t; u, \ell)$. For each $k = 0, 1, 2, \ldots$, the function $g_k(t; u, \ell)$ approaches 0 both as $t \to 0$ and as $t \to \infty$.

The function $|g_k(t; u, \ell)|$ achieves a maximum over $t \in [0, \infty)$ at $t = c(s_k + \ell)^2$ where c is a positive constant. Let M_k denote the maximum value of this function. Clearly, M_k is $O(k^{-2})$, and so the series $\sum_{k=0}^{\infty} M_k$ converges. By the Weierstrass M-test, the series $\sum |g_k(t; u, \ell)|$ converges uniformly in t. A similar argument applies to the second summand in $\frac{\partial^3}{\partial u^2} f_u(t; u, \ell)$. Thus $\sum_k \frac{\partial^3}{\partial u^2} f_{uk}(t; u, \ell)$ is absolutely, and uniformly convergent in t, for $t \geq 0$. Also, $|\frac{\partial^3}{\partial u^2} f_u(t; u, \ell)|$ is bounded. **Q.E.D.**

Theorem 4.15 *For fixed $u, \ell > 0$ and sufficiently small t,*

$$ct^{-9/2}\phi\left(\frac{u}{\sqrt{t}}\right) \leq \frac{\partial^3}{\partial u^2} f_u(t; u, \ell) \leq Ct^{-9/2}\phi\left(\frac{u}{\sqrt{t}}\right)$$

for some $c, C > 0$.

Proof: As in theorem 4.3, we argue that the first summand of the first term of $\frac{\partial}{\partial u} f_{uk}(t; u, \ell)$ dominates as $t \to 0$. **Q.E.D.**

Chapter 5

Properties of the Zero Drift FP2B Distribution, Sin Representation

An alternate representation for the densities is available [17, p. 359]. This series converges slowly for small t, and so is not as useful for computational purposes as the representation discussed in Chapter 4, at least not for the distribution of times encountered in hospital length of stay. However, it converges quickly for large t and so is more helpful in verifying regularity conditions. The equivalence of the two series is a consequence of the theory of theta functions [17, p. 370].

In this chapter, using the new representation, we show that all moments of the zero drift FP2B distribution exist, and examine the behaviour of $f_u(t; u, \ell)$ and its first three derivatives for t large. Finally, we establish that regularity conditions for the FP2B distribution (in the case of iid observations) are satisfied in the zero drift case. Here we focus attention on the upper subdensity, as the results and proofs for the complete density can be reduced to proofs for the upper and lower subdensities. As we point out below, regularity in the drift case will follow as an easy consequence. Unless otherwise indicated, we assume both u and ℓ to be finite.

In the notation used in the thesis, the alternate representation for the upper subdensity is

$$f_u^{\mu}(t; u, \ell, \mu) = \frac{\pi}{(u+\ell)^2} \exp\left(\frac{-(\mu t + 2u)\mu}{2}\right) \sum_{k=1}^{\infty} k \exp\left(\frac{-k^2\pi^2 t}{2(u+\ell)^2}\right) \sin\left(\frac{\pi k u}{u+\ell}\right).$$

This series, known in diffusion theory as Furth's formula for first passages, is uniformly convergent in t [17, p. 359]. For the zero drift case, the series reduces to

$$f_u^0(t;u,\ell) = \frac{\pi}{(u+\ell)^2}\sum_{k=1}^{\infty} k\exp\left(-k^2 t\lambda\right)\ \sin(k\gamma) \tag{5.1}$$

where $\lambda = \pi^2/(2(u+\ell)^2)$, and $\gamma = \pi u/(u+\ell)$. Note that, for $u,\ \ell > 0$, both λ and γ are strictly positive, and furthermore $0 < \gamma < \pi$. Also, note that because the function $g(t) = \exp\left(\frac{-(\mu t + 2u)\mu}{2}\right)$ approaches 0 very quickly as $t \to \infty$, and its integral $\int_0^\infty g(t)\ dt$ likewise converges quickly, regularity conditions for the drift case will follow easily from the zero drift case. For the remainder of this chapter, the zero drift subdensity will be denoted as $f_u(t;u,\ell)$, without the superscript 0.

5.1 Moments of the Zero Drift FP2B distribution

We prove in two steps that moments of all values exist for the upper subdensity.

Theorem 5.1 *The integral*

$$\int_{t_0}^{\infty} t^p f_u(t;u,\ell)dt$$

converges for all p, where $t_0 = 1/(2\lambda)$.

Proof:

$$\begin{aligned} f_u(t;u,\ell) &= \frac{\pi}{(u+\ell)^2}\sum_{k=1}^{\infty} k\exp\left(-k^2 t\lambda\right)\ \sin(k\gamma) \\ &\le \frac{\pi}{(u+\ell)^2}\sum_{k=1}^{\infty} k\exp\left(-k^2 t\lambda\right) = \frac{\pi}{(u+\ell)^2}\sum_{k=1}^{\infty} g(k) \end{aligned}$$

where $g(k) = k\exp\left(-k^2 t\lambda\right)$. The function $g(k)$ achieves a maximum on $(0,\infty)$ at $k = 1/\sqrt{2t\lambda}$. Then for $t > 1/(2\lambda)$, $g(k)$ is monotonically decreasing on $(1,\infty)$. We can then apply lemma D.0.1 (see appendix D), which gives

$$\sum_{k=1}^{\infty} g(k) \le \int_1^\infty g(k)\ dk + g(1)$$

$$= \exp(-t\lambda)\left(1 + \frac{1}{2\lambda t}\right)$$

Thus $f_u(t;u,\ell) \leq \exp(-t\lambda)\left(1+\frac{1}{2\lambda t}\right)$ and $t^p f_u(t;u,\ell) \leq t^p \exp(-t\lambda)\left(1+\frac{1}{2\lambda t}\right)$. The integral of this last function with respect to t over (t_0,∞) converges for all p. Thus we have shown that

$$\int_{t_0}^{\infty} t^p f_u(t;u,\ell)dt$$

converges for all p, where $t_0 = 1/(2\lambda) > 0$. **Q.E.D.**

Theorem 5.2 *All positive moments of the zero drift FP2B distribution exist.*

Proof: We showed in theorem 5.1 that $\int_{t_0}^{\infty} t^p f_u(t;u,\ell)dt$ converges for all p, where $t_0 = 1/(2\lambda)$. As shown in theorem 4.1, $f_u(t;u,\ell)$ is bounded on $(0,\infty)$. Thus the integral of $t^p f_u(t;u,\ell)$ over $[0,t_0]$ is finite for all $p>0$. Similar statements can be made about $f_\ell(t;u,\ell)$. Thus all positive moments of the FP2B distribution exist. **Q.E.D.**

5.2 Regularity Conditions

The following specification of regularity conditions for consistency and asymptotic normality of maximum likelihood estimates from iid observations comes from Serfling [37] and is for single-parameter models. The extension to multiple parameter models is straightforward.

Regularity Conditions. Consider Θ to be an open interval (not necessarily finite) in R. We assume:

(R1) For each $\theta \in \Theta$, the derivatives

$$\frac{\partial \log f(x;\theta)}{\partial\theta},\ \frac{\partial^2 \log f(x;\theta)}{\partial\theta^2},\ \frac{\partial^3 \log f(x;\theta)}{\partial\theta^3}$$

exist, all x;

(R2) For each $\theta_0 \in \Theta$, there exist functions $g(x)$, $h(x)$ and $H(x)$ (possibly depending on θ_0) such that for θ in a neighborhood $N(\theta_0)$ the relations

$$\left|\frac{\partial f(x;\theta)}{\partial\theta}\right| \leq g(x),\ \left|\frac{\partial^2 f(x;\theta)}{\partial\theta^2}\right| \leq h(x),\ \left|\frac{\partial^3 \log f(x;\theta)}{\partial\theta^3}\right| \leq H(x)$$

hold, all x, and

$$\int g(x)dx < \infty,\ \int h(x)dx < \infty,\ E_\theta\{H(X)\} < \infty, \quad \text{for } \theta \in N(\theta_0);$$

(R3) For each $\theta \in \Theta$,

$$0 < E_\theta \left\{ \left(\frac{\partial \log f(X;\theta)}{\partial \theta} \right)^2 \right\} < \infty.$$

Here E_θ denotes expectation taken with respect to θ.

Thus we will prove in the zero drift case that

- the first, second, and third derivatives of $\log f_u(t; u, \ell)$ exist;
- There exists a function $g(t)$ such that

$$\left| \frac{\partial}{\partial u} f_u(t; u, \ell) \right| \le g(t)$$

  for all $t \in (0, \infty)$ and $\int_0^\infty g(t)\ dt < \infty$;
- There exists a function $h(t)$ such that

$$\left| \frac{\partial^2}{\partial u^2} f_u(t; u, \ell) \right| \le h(t)$$

  for all $t \in (0, \infty)$ and $\int_0^\infty h(t)\ dt < \infty$;
- There exists a function $H(t)$ such that

$$\left| \frac{\partial^3}{\partial u^3} \log\ f_u(t; u, \ell) \right| \le H(t)$$

  for all $t \in (0, \infty)$ and $E_u\{H(T)\} < \infty$; and
- For each $u > 0$,

$$0 < E_u \left\{ \left(\frac{\partial \log f_u(T; u, \ell)}{\partial u} \right)^2 \right\} < \infty.$$

Here E_u can be taken to be expectation with respect to the upper subdensity. It can be shown that the regularity conditions of Serfling can be reduced to conditions like these.

We first prove several theorems about the behaviour of $f_u(t; u, \ell)$ and its derivatives when t is large.

Theorem 5.3 *For t large,*

$$f_u(t;u,\ell) \geq \epsilon \exp(-t\lambda)$$

for some $\epsilon > 0$ and where $\lambda = \pi^2/(2(u+\ell)^2)$.

Proof:

$$f_u(t;u,\ell) = \sum_{k=1}^{\infty} f_{uk} = \frac{\pi}{(u+\ell)^2} \sum_{k=1}^{\infty} k \exp\left(-k^2 t\lambda\right) \ \sin(k\gamma).$$

We will argue that the first term of this series dominates for large t, ie

$$\frac{\sum_{k=2}^{\infty} f_{uk}}{f_{u1}} \to 0 \quad \text{as} \quad t \to \infty.$$

Write $\frac{\sum_{k=2}^{\infty} f_{uk}}{f_{u1}}$ as $c \sum a_k b_k$ where

$$a_k = \frac{k e^{-k^2\lambda} \sin(k\gamma)}{e^{-\lambda} \sin(\gamma)}$$

and

$$b_k = \frac{e^{-k^2\tau\lambda}}{e^{-\tau\lambda}}$$

where $\tau = t-1$. Clearly $\sum a_k$ converges. We now show that $\lim_{\tau\to\infty} \sum b_k = 0$. First note that $b_k = \exp(-\tau\lambda(k^2-1))$ is monotonically decreasing for $k \in (0,\infty)$. Then

$$\sum_{k=2}^{\infty} e^{-\tau\lambda(k^2-1)} \leq \int_1^{\infty} e^{-\tau\lambda(k^2-1)}\, dk = \frac{1-\Phi\left(\sqrt{2\tau\lambda}\right)}{e^{-\tau\lambda}\sqrt{2\tau\lambda}}$$

The limit of this expression as $\tau \to \infty$ is $0/0$. We now apply L'Hopital's rule. Differentiating the numerator and denominator of the last expression with respect to τ gives

$$\frac{\phi\left(\sqrt{2\tau\lambda}\right)\ \tau^{-1/2}}{e^{-\tau\lambda}(\tau^{-1/2} - 2\lambda\tau^{1/2})} = \frac{1}{1-2\lambda\tau}$$

which approaches 0 as $\tau \to \infty$. Thus $\lim_{\tau\to\infty} \sum_{k=2}^{\infty} a_k b_k = \lim_{\tau\to\infty} \sum_{k=2}^{\infty} f_{uk}/f_{u1} = 0$. Thus the first term of the series, $f_{u1} = k_1 \exp(-t\lambda)\sin(\gamma)$, dominates for large t. Note that, for finite t, f_{u1} is strictly positive, because $0 < \gamma < \pi$. So for t large, $f_u(t;u,\ell) > \epsilon \exp(-t\lambda)$, for some $\epsilon > 0$.
Q.E.D.

Theorem 5.4 *For t sufficiently large,*

$$\left|\frac{\partial}{\partial u} f_u(t; u, \ell)\right| \leq k_1 t \exp(-t\lambda)$$

for some $k_1 > 0$ and where $\lambda = \pi^2/(2(u+\ell)^2)$.

Proof:

$$\left|\frac{\partial}{\partial u} f_u(t; u, \ell)\right| = \left|\frac{\pi}{(u+\ell)^5} \sum_{k=1}^{\infty} k \exp\left(-k^2 t\lambda\right)\right.$$

$$\left\{t\pi^2 k^2 \sin\left(k\gamma\right) - 2(u+\ell)^2 \sin\left(k\gamma\right)\right.$$

$$\left. +\pi k \ell(u+\ell) \cos\left(k\gamma\right)\right\}\Big| \quad (5.2)$$

$$\leq \frac{\pi}{(u+\ell)^5} \sum_{k=1}^{\infty} k \exp\left(-k^2 t\lambda\right) \left\{t\pi^2 k^2 + 2(u+\ell)^2 + \pi k \ell(u+\ell)\right\}$$

where $\gamma = \pi u/(u+\ell)$. For large t, it suffices to examine the behaviour of the factor involving t. So consider the sum

$$\sum_{k=1}^{\infty} k^3 t \exp(-k^2 t\lambda).$$

Note that the function $g_0(k) = k^3 t \exp(-k^2 t\lambda)$ achieves a maximum on $k \in (0, \infty)$ at $k^* = \sqrt{3/(2t\lambda)}$, and monotonically decreases for $k > k^*$. Then for $t > 3/(2\lambda)$, $g_0(k)$ is monotonically decreasing on $(1, \infty)$. For functions of this shape we have

$$\sum_{1}^{\infty} g_0(k) \leq \int_1^{\infty} g_0(k)\, dk + g_0(1)$$

(see lemma D.0.1, in appendix D). Thus, for $t > 3/(2\lambda)$

$$\sum_{k=1}^{\infty} k^3 t \exp(-k^2 t\lambda) \leq \int_1^{\infty} k^3 t \exp(-k^2 t\lambda)\, dk + t \exp(-t\lambda)$$

$$= \frac{1}{2\lambda^2 t^2} \exp(-t\lambda)(t\lambda + 1) + t \exp(-t\lambda)$$

$$= \exp(-t\lambda) \left(\frac{1}{2\lambda t} + \frac{1}{2\lambda^2 t^2} + t\right).$$

For large t, this last expression is $O(t\exp(-t\lambda))$. Therefore, we can show that for large t, $f_u(t;u,\ell) < k_1 t\exp(-t\lambda)$, for some $k_1 > 0$. **Q.E.D.**

Theorem 5.5 *For large t,*

$$\left|\frac{\partial^2}{\partial u^2} f_u(t;u,\ell)\right| \leq k_2 t^2 \exp(-t\lambda)$$

for some constant $k_2 > 0$.

Proof:

$$\left|\frac{\partial^2}{\partial u^2} f_u(t;u,\ell)\right| = \Bigg|\frac{1}{(u+\ell)^8}\sum_{k=1}^{\infty}\exp\left(-k^2 t\lambda\right)$$
$$\{t^2k^5\pi^5\sin(k\gamma) + t[-7k^3\pi^3\sin(k\gamma)(u+\ell)^2 + 2k^4\pi^4\cos(k\gamma)\ell(u+\ell)]$$
$$+6k\pi\sin(k\gamma)(u+\ell)^4 - 2k^3\pi^3\sin(k\gamma)[u\ell^3 + u^2\ell^2 + \ell^4] - 6k^2\pi^2\cos(k\gamma)(u+\ell)^4\}\Bigg| \tag{5.3}$$

$$\leq \frac{1}{(u+\ell)^8}\sum_{k=1}^{\infty}\exp\left(-k^2 t\lambda\right)\{t^2k^5\pi^5 + t[-7k^3\pi^3(u+\ell)^2 + 2k^4\pi^4\ell(u+\ell)]$$
$$+6k\pi(u+\ell)^4 - 2k^3\pi^3[u\ell^3 + u^2\ell^2 + \ell^4] - 6k^2\pi^2(u+\ell)^4\} \tag{5.4}$$

For large t, it suffices to examine the behaviour of the factor involving t^2. So consider the sum

$$\sum_{k=1}^{\infty} g_0(k) = \sum_{k=1}^{\infty} k^5 t^2 \exp(-k^2 t\lambda).$$

Each summand $g_0(k)$ achieves a maximum over $k \in (0,\infty)$ at $k^* = \sqrt{\frac{5}{2t\lambda}}$, and monotonically decreases thereafter. Then for $t > 5/(2\lambda)$, $g_0(k)$ is monotonically decreasing on $(1,\infty)$. By lemma D.0.1, for $t > 5/(2\lambda)$, the sum above is less than or equal to

$$\exp(-t\lambda)\left(\frac{1}{t\lambda^3} + \frac{1}{\lambda^2} + \frac{t}{2\lambda} + t^2\right).$$

For large t, the above function is $O(t^2\exp(-t\lambda))$. Thus we can find a constant k_2 such that $\left|\frac{\partial^2}{\partial u^2} f_u(t;u,\ell)\right| \leq k_2 t^2 \exp(-t\lambda)$. **Q.E.D.**

Theorem 5.6 *For large t,*

$$\left|\frac{\partial^3}{\partial u^3} f_u(t;u,\ell)\right| \leq k_3 t^3 \exp(-t\lambda)$$

for some constant $k_3 > 0$.

Proof: We can show that

$$\left|\frac{\partial^3}{\partial u^3} f_u(t; u, \ell)\right| \le \left|\frac{1}{(u+\ell)^{11}} \sum_{k=1}^{\infty} \exp\left(-k^2 t\lambda\right) t^3 k^7 \pi^7\right|$$

plus some terms in smaller powers of t, which will be inconsequential for large t. Consider the sum

$$\sum_{k=1}^{\infty} g_0(k) = \sum_{k=1}^{\infty} k^7 t^3 \exp(-k^2 t\lambda).$$

Each summand $g_0(k)$ achieves a maximum over $k \in (0, \infty)$ at $k^* = \sqrt{\frac{7}{2t\lambda}}$, and monotonically decreases thereafter. Then for $t > 7/(2\lambda)$, $g_0(k)$ is monotonically decreasing on $(1, \infty)$. By lemma D.0.1, for $t > 7/(2\lambda)$, the sum above is less than or equal to

$$\exp(-t\lambda)\left(\frac{3}{t\lambda^4} + \frac{3}{\lambda^3}\frac{3t}{2\lambda^2} + \frac{t^2}{2\lambda} + t^3\right).$$

For large t, the above function is $O(t^3 \exp(-t\lambda))$. Thus we can find a constant k_3 such that $\left|\frac{\partial^3}{\partial u^3} f_u(t; u, \ell)\right| \le k_3 t^3 \exp(-t\lambda)$. **Q.E.D.**

We now show that regularity conditions are satisfied.

Theorem 5.7 *The first derivative of* $\log f_u(t; u, \ell)$ *with respect to u exists for* $t \in (0, \infty)$.

Proof: First, note that

$$\frac{\partial}{\partial u} \log f_u(t; u, \ell) = \frac{\frac{\partial}{\partial u} f_u(t; u, \ell)}{f_u(t; u, \ell)}.$$

Also, as we showed in theorem 4.5, $\frac{\partial}{\partial u} f_u(t; u, \ell)$ is bounded for $t \in (0, \infty)$. We showed in theorem 4.2 $f(t; u, \ell)$ is greater than 0 for $t > 0$. Thus, the derivative of $\log f_u(t; u, \ell)$ with respect to u exists for $t \in (0, \infty)$. **Q.E.D.**

Theorem 5.8 *The second derivative of* $\log f_u(t; u, \ell)$ *with respect to u exists for* $t \in (0, \infty)$.

Proof:

$$\frac{\partial^2}{\partial u^2} \log f_u(t; u, \ell) = \frac{\frac{\partial^2}{\partial u^2} f_u(t; u, \ell)}{f_u(t; u, \ell)} - \left(\frac{\frac{\partial}{\partial u} f_u(t; u, \ell)}{f_u(t; u, \ell)}\right)^2$$

We showed above that $\frac{\partial}{\partial u} f_u(t;u,\ell)/f_u(t;u,\ell)$ exists, so we only need to prove that the first summand is less than ∞. Also, as shown in theorem 4.10, $\frac{\partial^2}{\partial u^2} f_u(t;u,\ell)$ is bounded on $(0,\infty)$. Since $f_u(t;u,\ell)$ is greater than 0 for $t>0$, the second derivative of $\log f_u(t;u,\ell)$ with respect to u exists for $t \in (0,\infty)$. **Q.E.D.**

Theorem 5.9 *The third derivative of* $\log f_u(t;u,\ell)$ *with respect to* u *exists for* $t \in (0,\infty)$.

Proof:

$$\frac{\partial^3}{\partial u^3} \log f_u(t;u,\ell) = \frac{\frac{\partial^3}{\partial u^3} f_u(t;u,\ell)}{f_u(t;u,\ell)} - 3\frac{\frac{\partial^2}{\partial u^2} f_u(t;u,\ell)}{f_u(t;u,\ell)}\frac{\frac{\partial}{\partial u} f_u(t;u,\ell)}{f_u(t;u,\ell)} + 2\left(\frac{\frac{\partial}{\partial u} f_u(t;u,\ell)}{f_u(t;u,\ell)}\right)^3 \tag{5.5}$$

We only need to prove that the first summand is less than ∞, since the other summands have been shown to be less than ∞ above. First, $\frac{\partial^3}{\partial u^3} f_u(t;u,\ell)$ is bounded on $(0,\infty)$, as shown in theorem 4.14. Since $f_u(t;u,\ell)$ is greater than 0 for $t>0$, the third derivative of $\log f_u(t;u,\ell)$ with respect to u exists for $t \in (0,\infty)$. **Q.E.D.**

Theorem 5.10 *There exists a function* $g(t)$ *such that*

$$\left|\frac{\partial}{\partial u} f_u(t;u,\ell)\right| \le g(t)$$

for all $t \in (0,\infty)$ *and* $\int_0^\infty g(t)\, dt < \infty$.

Proof: As we showed in theorem 4.5, $\frac{\partial}{\partial u} f_u(t;u,\ell)$ is bounded for $t \in (0,\infty)$. Thus we only need to show that $\frac{\partial}{\partial u} f_u(t;u,\ell)$ is dominated in absolute value in the right tail by an integrable function $g(t)$. We showed in theorem 5.4 that $\left|\frac{\partial}{\partial u} f_u(t;u,\ell)\right| \le k_1 t \exp(-t\lambda)$, which is clearly integrable with respect to t on (t_0,∞), where $t_0 > 0$. **Q.E.D.**

Theorem 5.11 *There exists a function* $h(t)$ *such that*

$$\left|\frac{\partial^2}{\partial u^2} f_u(t;u,\ell)\right| \le h(t)$$

for all $t \in (0,\infty)$ *and* $\int_0^\infty h(t)\, dt < \infty$

Proof: As we showed in theorem 4.10, $\frac{\partial^2}{\partial u^2} f_u(t;u,\ell)$ is bounded for $t \in (0,\infty)$. In theorem 5.5 we showed that, for large t, $\left|\frac{\partial^2}{\partial u^2} f_u(t;u,\ell)\right| \le k_2 t^2 \exp(-\lambda t)$ which is clearly integrable with respect to t on (t_0,∞), for $t_0 > 0$. Thus $\left|\frac{\partial^2}{\partial u^2} f_u(t;u,\ell)\right|$ is dominated in the right tail by an integrable function. **Q.E.D.**

Theorem 5.12 *There exists a function $H(t)$ such that*

$$\left|\frac{\partial^3}{\partial u^3}\log\ f_u(t;u,\ell)\right| \le H(t)$$

for all $t \in (0,\infty)$ and $E_u\{H(t)\} < \infty$

Proof: From equation 5.5 we have

$$\left|\frac{\partial^3}{\partial u^3}\log f_u(t;u,\ell)\right| \le \frac{\left|\frac{\partial^3}{\partial u^3}f_u(t;u,\ell)\right|}{f_u(t;u,\ell)} + 3\frac{\left|\frac{\partial^2}{\partial u^2}f_u(t;u,\ell)\right|}{f_u(t;u,\ell)}\frac{\left|\frac{\partial}{\partial u}f_u(t;u,\ell)\right|}{f_u(t;u,\ell)} + 2\frac{\left|\frac{\partial}{\partial u}f_u(t;u,\ell)\right|^3}{f_u(t;u,\ell)^2} \quad (5.6)$$

As we showed in theorem 4.14, $\frac{\partial^3}{\partial u^3}f_u(t;u,\ell)$ is bounded for $t \in (0,\infty)$. We showed in theorem 4.3 that, for small t, $f_u(t;u,\ell) \ge ct^{-3/2}\phi\left(\frac{u}{\sqrt{t}}\right)$. Also for small t, $\frac{\partial}{\partial u}f_u(t;u,\ell) \le C_1t^{-5/2}\phi\left(\frac{u}{\sqrt{t}}\right)$ and $\frac{\partial^2}{\partial u^2}f_u(t;u,\ell) \le C_1t^{-7/2}\phi\left(\frac{u}{\sqrt{t}}\right)$, as was shown in theorems 4.6 and 4.11. Then, for small t, the second summand, when multiplied by $f_u(t;u,\ell)$, is less than or equal to $Ct^{-9/2}\phi\left(\frac{u}{\sqrt{t}}\right)$ and so is bounded. This is also true for the third summand. Thus we only need to ensure that each summand in (5.6), when multiplied by $f_u(t;u,\ell)$, is dominated in the right tail by an integrable function.

Using theorem 5.6, we have that the first summand multiplied by $f_u(t;u,\ell)$ is

$$\left|\frac{\partial^3}{\partial u^3}f_u(t;u,\ell)\right| \le k_1t^3\exp(-t\lambda)$$

for some $k_1 > 0$. From theorems 5.3, 5.4 and 5.5, the second summand, when multiplied by $f_u(t;u,\ell)$ is

$$3\frac{\left|\frac{\partial^2}{\partial u^2}f_u(t;u,\ell)\right|\left|\frac{\partial}{\partial u}f_u(t;u,\ell)\right|}{f_u(t;u,\ell)} \le \frac{(k_2t^2\exp(-t\lambda))\ (k_3t\exp(-t\lambda))}{\epsilon\exp(-t\lambda)} = k_4t^3\exp(-t\lambda)$$

for some $k_2, k_3, k_4 > 0$. Finally, from theorems 5.3 and 5.4, the third summand, when multiplied by $f_u(t;u,\ell)$ is

$$2\frac{\left|\frac{\partial}{\partial u}f_u(t;u,\ell)^3\right|}{f_u(t;u,\ell)^2} \le \frac{k_5\left(t^3\exp(-t\lambda)\right)^3}{\left(\epsilon\exp(-t\lambda)\right)^2} = k_6t^3\exp(-t\lambda)$$

for some $k_5, k_6 > 0$. Since the integral

$$\int_{t_0}^{\infty} t^3\exp(-t\lambda)\ dt$$

converges for $t_0 > 0$, we have proved the desired result. **Q.E.D.**

Theorem 5.13

$$0 < E_u\left\{\left(\frac{\partial \log f_u(t;u,\ell)}{\partial u}\right)^2\right\} < \infty.$$

Proof: Note that

$$E_u\left\{\left(\frac{\partial \log f_u(t;u,\ell)}{\partial u}\right)^2\right\} = \int_0^\infty \left(\frac{\partial f_u(t;u,\ell)}{\partial u}\right)^2 \frac{1}{f_u(t;u,\ell)}\, dt \tag{5.7}$$

First, as noted previously, both $f_u(t;u,\ell)$ and $\frac{\partial f_u(t;u,\ell)}{\partial u}$ are bounded. To show that the integral in 5.7 converges on $(0,\infty)$, we need to show that (1) the integral converges for large t and (2) the integrand is bounded as $t \to 0$.

We first show that the integral converges for large t. We saw above that, for large t, $\frac{\partial f_u(t;u,\ell)}{\partial u} \le ct\exp(-\lambda t)$. Also, in theorem 5.3, we showed that $f_u(t;u,\ell) > \epsilon\exp(-t\lambda)$, for some $\epsilon > 0$. Thus $\left(\frac{\partial f_u(t;u,\ell)}{\partial u}\right)^2 \frac{1}{f_u(t;u,\ell)} \le kt^2\exp(-\lambda t)$ for some constant $k > 0$, and so the integral in 5.7 converges for large t.

Now for t small, we need to show that the integrand in 5.7 is bounded. We showed in theorem 4.3 that, for small t, $f_u(t;u,\ell) \ge ct^{-3/2}\phi\left(\frac{u}{\sqrt{t}}\right)$. Also for small t, $\frac{\partial}{\partial u}f_u(t;u,\ell) \le C_1 t^{-5/2}\phi\left(\frac{u}{\sqrt{t}}\right)$ and $\frac{\partial^2}{\partial u^2}f_u(t;u,\ell) \le C_1 t^{-7/2}\phi\left(\frac{u}{\sqrt{t}}\right)$, as was shown in theorems 4.6 and 4.11. Then, for small t, the integrand in 5.7 is less than or equal to $Ct^{-7/2}\phi\left(\frac{u}{\sqrt{t}}\right)$ and so is bounded.
Q.E.D.

Chapter 6

Extensions

6.1 Mixture Models

In this section, we discuss the effect of heterogeneity in the population on the interpretation of the hazard function in survival analysis. We then extend the model developed in Chapter 3 to a mixture model that recognizes this heterogeneity.

An unavoidable characteristic of administrative data in health services research is heterogeneity. Often there are very few covariates available that are truly helpful in predicting LOS. Excess heterogeneity can affect the interpretation of the model, particularly the hazard rate. Recall that the hazard function at time t gives the instantaneous risk of failure at time t given survival up to time t. When the response is time to death, individuals in a heterogeneous population that are at high risk will die sooner. Due to this selection effect, we cannot interpret the population hazard as the risk to any single individual over time.

The simplest example of a heterogeneous population consists of two homogeneous subgroups of individuals. Suppose both groups have constant hazards, with the hazard of the "frail" group being higher than the hazard for the "robust" group. Suppose that both of these subgroups are initially of the same size, so that the frail group makes up half the population. One can show that the hazard for the whole population decreases monotonically over time, even though the hazard in each group remains constant. This is because those at high risk (the frail group) are failing early, resulting in a higher proportion of the population belonging to the robust group. Thus over time, the population hazard will approach the hazard of the more robust group, i.e. over time the population hazard will be pulled downwards. However, the population hazard does not represent the survival experience of any individual member of the population.

As another illustration, consider a population where one group of individuals has an increasing hazard, but the other group is "immune", so has a constant hazard rate of 0. In this case, one can show that the population hazard initially increases to some maximum value, and thereafter decreases. This shape of population hazard is typical of data on time to breakdown of marriage. This phenomenon has been interpreted as showing that an individual marriage is at high risk of breakdown at some time (the "seven year itch"), but if the marriage survives beyond this time, the chances of breakdown will decrease. In fact, the shape of the hazard could be due to heterogeneity in the population. Some individuals could be immune to marriage breakdown due to religious beliefs, while others could be at ever-increasing risk of divorce. For other examples of "heterogeneity's ruses", see the paper by Vaupel and Yashin [41].

Any real population will be far more complicated than a mixture of just two homogeneous populations, and it is even possible that no two individuals will share exactly the same hazard function. To address the problem of heterogeneity, mixed models, or frailty models as they are sometimes called, have been developed for survival data [24]. The canonical frailty model is the proportional frailty model. Here the hazard rate for individual j in group i is of the form

$$h_{ij}(t) = Z_i h_0(t) \exp(\beta x_{ij})$$

where $h_0(t)$ is a baseline hazard common to all individuals, x_{ij} is a vector of covariates for the individual j in group i, and Z_i, the "frailty" for group i, is a random variable from some distribution. The regression parameters β, and the parameters of the frailty distribution are estimated by some method. Note that the groups may be of size 1, i.e. that each individual has their own distinct hazard rate. If the variance of the frailty distribution is found to be significantly different from 0, this gives evidence that heterogeneity is indeed present in the population and that a frailty model is in fact necessary. Frailty distributions that have been investigated in the literature include the gamma, positive stable, inverse Gaussian, and log-normal, and extensions have been developed to allow for the presence of immune individuals in the population.

The models discussed in this thesis can be extended to give a mixture model for LOS, which recognizes the presence of heterogeneity in the population. Again assume that health status follows a Wiener process with drift μ and volatility $\sigma^2 = 1$, but allow the drift parameter μ to be a random variable from some distribution. This approach was investigated by Aalen [1] in the case of a single barrier, where LOS has an inverse Gaussian distribution.

For the case of two barriers, let μ be distributed as a Normal random variable with mean θ and variance ν^2. Then the density of LOS is given by

$$\int_{-\infty}^{\infty} f_u^{\mu}(t)\frac{1}{\sqrt{2\pi\nu^2}}e^{-\frac{(\mu-\theta)^2}{2\nu^2}}d\mu + \int_{-\infty}^{\infty} f_\ell^{\mu}(t)\frac{1}{\sqrt{2\pi\nu^2}}e^{-\frac{(\mu-\theta)^2}{2\nu^2}}d\mu \tag{6.1}$$

The first of these integrals is

$$t^{-3/2}(2\pi\nu^2)^{-1/2}\left(\sum_{k=-\infty}^{k=\infty} c_k\,\phi\left(\frac{c_k}{\sqrt{t}}\right)\right)\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}\left[(\mu t-2u)\mu+\frac{(\mu-\theta)^2}{\nu^2}\right]\right)d\mu \tag{6.2}$$

where $c_k = 2k(u+\ell)+u$. The integral in 6.2 can be evaluated by completing the square to give

$$(t\nu^2+1)^{-1/2}\exp\left(-\frac{1}{2\nu^2}\left[\theta^2-\frac{(u\nu^2+\theta)^2}{t\nu^2+1}\right]\right)\int_{-\infty}^{\infty}\exp\left(-\frac{(w-z)^2}{2\nu^2}\right)dw$$

where $z = (u\nu^2+\theta)/\sqrt{t\nu^2+1}$, and we have made the change of variable $w = \mu\sqrt{t\nu^2+1}$. The integral in the last expression is the kernel of the normal cdf, which integrates to $(2\pi\nu^2)^{1/2}$. Finally we get

$$[t^3(t\nu^2+1)]^{-1/2}\exp\left(\frac{-(\theta^2 t-u^2\nu^2-2u\theta)}{2(t\nu^2+1)}\right)\sum_{k=-\infty}^{k=\infty} c_k\,\phi\left(\frac{c_k}{\sqrt{t}}\right)$$

which we denote as $f_u^M(t)$. The second integral in 6.1 is handled in exactly the same way to give

$$f_\ell^M(t) = [t^3(t\nu^2+1)]^{-1/2}\exp\left(\frac{-(\theta^2 t-\ell^2\nu^2-2\ell\theta)}{2(t\nu^2+1)}\right)\sum_{k=-\infty}^{k=\infty} e_k\,\phi\left(\frac{e_k}{\sqrt{t}}\right)$$

where $e_k = 2k(u+\ell)+\ell$. Then the density for the distribution of LOS is the sum of the two subdensities

$$f^M(t) = f_\ell^M(t) + f_u^M(t)$$

Note that, while many mixture models require numerical integration to evaluate the densities, numerical integration is not necessary here. The subdensities for $\ell = 1, u = 2, \nu = 1$ and various values of θ are shown in figures 6.1 and 6.2.

The above model allows each individual to have their own distinct drift.

This model can be easily extended to accommodate covariates. As an example, let the drift parameter μ come from a Normal distribution with mean θ where now θ is allowed to depend linearly on a covariate

$$\theta_i = \theta_0 + \theta_1 x_i$$

Estimation in this case is straightforward.

6.2 Transfers

So far, we have only considered two outcomes, death and discharge. However, any hospital database will include information on individuals who are transferred to another institution. From the point of view of the hospital administrator, transfers are an important component of costs and resource consumption, and so it is imperative to include them in the model.

In general, patients are transferred from one acute care institution to another because the new institution has specialized facilities and staff that are better suited to the needs of the patient. Sometimes, patients are transferred to another acute care institution because it is closer to their family or area of residence. Patients may be transferred to another type of institution (such as a chronic care or nursing home) because the patient needs an alternate (less intense) level of care. In the United States, the patient's insurance policy may also dictate that the patient be transferred to a cheaper facility at the earliest possible opportunity.

In any case it seems reasonable to assume that decision to transfer depends at least partially on the health status of the patient, so it would not be correct to treat transfers as independently censored observations. Thus another approach is developed here.

It seems reasonable to postulate that individuals become eligible for transfer to another acute care institution once their health status first reaches a moderate level, w, where $0 < w < u$. Once this level is achieved, an individual may be transferred to another acute care institution

Figure 6.1: $f_l^M(t)$

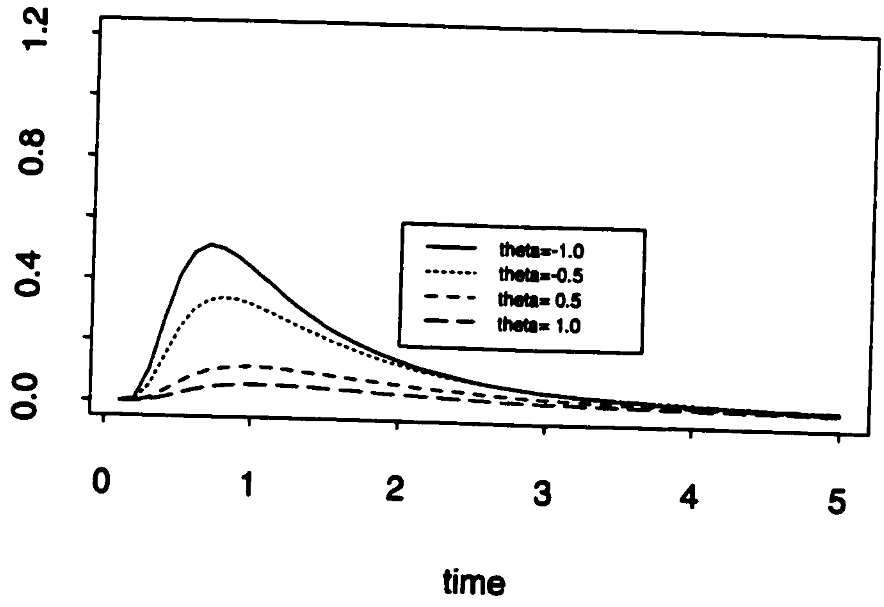
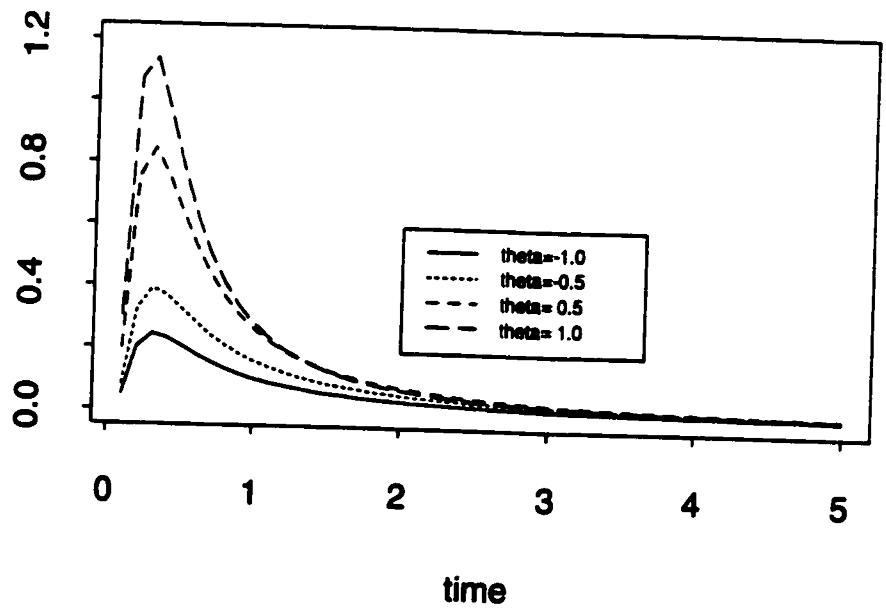


Figure 6.2: $f_u^M(t)$



depending on his needs, availability of a bed, availability of transport, etc. Given enough up-to-date information, it might be possible to determine whether or not an individual would be transferred. However given the scanty information available to health services researchers, it seems more useful to model this decision as a random event.

We now develop this approach in terms of the health level process $H(t)$. The situation is shown in figure 6.3, with discharge barrier u, transfer decision barrier w and lower barrier $-\ell$ corresponding to death in hospital. Let t^* be the time the health level process $H(t)$ first reaches a

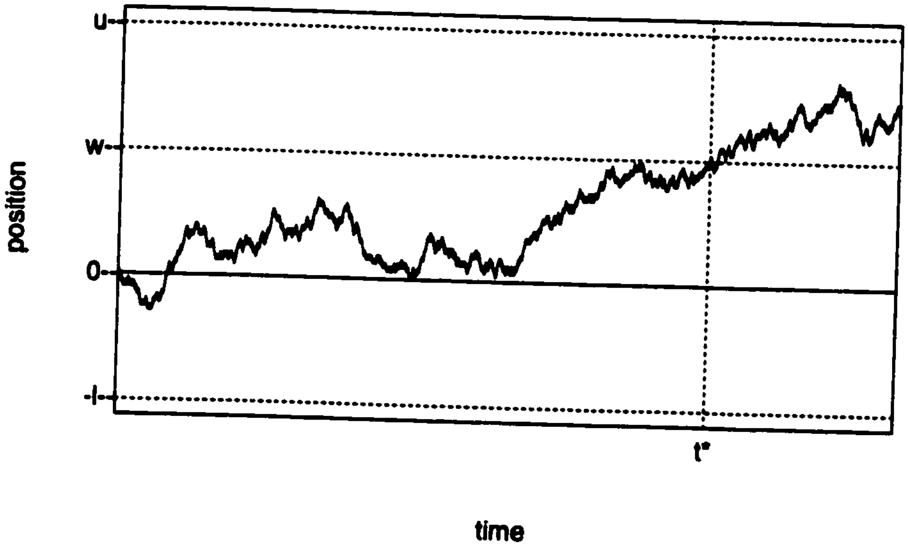


Figure 6.3: Decision barrier

barrier at w. With probability p, the process is terminated at t^*, corresponding to the event that the individual is transferred to some other institution. With probability $q = 1 - p$, the process continues until it is either absorbed at the upper barrier u, (the patient is discharged), or it is absorbed at the lower barrier $-\ell$, (the patient dies in hospital). Note that in this simple model, only the first visit of the process to w potentially triggers a transfer. It is also possible that the process could be censored by some independent mechanism, such as end of study period.

We now derive the distribution of T, the time in hospital, under this new model. The event that the hospital stay ends in an interval of time $(t, t+dt)$ can be considered as the union of the four disjoint events:

1. The health level process first hits $-\ell$ in $(t, t+dt)$, without going through w.
2. The process first hits w in $(t, t+dt)$, without going through $-\ell$, and (with probability p) is terminated.
3. The process hits w in (t_1, t_1+dt) where $t_1 < t$, continues on (with probability q), and hits the barrier u in $(t, t+dt)$, without hitting $-\ell$ along the way.
4. The process hits w in (t_1, t_1+dt) where $t_1 < t$, continues on (with probability q), and hits the barrier $-\ell$ in $(t, t+dt)$, without hitting u along the way

Since there are now three barriers, we need new notation that makes missing a barrier explicit. Throughout this section, we will assume the more general drift model is used, so dependence on μ will be suppressed. Write $f_a^{(b)}(t)$ for the subdensity corresponding to the event that a Wiener process with drift and two barriers a and b, hits a for the first time in an interval around time t, without first going through b. Note that one of a, b will be positive, (representing the upper barrier), and the other negative (representing the lower barrier). Given values for a and b, an expression for $f_a^{(b)}(t)$ can be calculated as shown in section 3.1. If $a > 0$, and $b < 0$, then calculate the upper subdensity as in equation 3.2 with a replacing u and $|b|$ replacing ℓ . Similarly, if $a < 0$, and $b > 0$, then calculate the lower subdensity as in equation 3.1 with $|a|$ replacing ℓ and b replacing u .

Now consider item 1, the event that the process first hits the lower barrier $-\ell$ in an interval around time t, without hitting the barrier at w. In the new notation, this event has probability proportional to $f_{-\ell}^{(w)}(t)$.

The event described in item 2 is only slightly more complicated. This is the event that the process hits the barrier at w for the first time in an interval of time around t, without hitting the lower barrier at $-\ell$. (Note that we don't need to worry about the process hitting u, since $u > w$ and if the process is hitting w for the first time, it will not have encountered u.) The event in question has probability proportional to $f_w^{(-\ell)}(t)$. This must be multiplied by p to take account of the fact that only a proportion p of these paths actually terminate when they reach w.

We now consider the probability of the event described in item 3, i.e. the probability that the process is terminated at u in an interval of time around t. Note that all paths which terminate at u must have first hit w, but not been terminated at w. Also these paths must have avoided the lower barrier at $-\ell$. Thus we need to consider all paths that reached u at time t without going through $-\ell$, and then take only the proportion that made it through the intermediate barrier at w. The probability corresponding to this event is proportional to is $q\ f_u^{(-\ell)}(t)$.

Item 4 is the event that the process terminates at $-\ell$ in $(t, t = dt)$, having previously passed through the intermediate barrier w. We first confine attention to all paths that do not hit u. Of these, we need paths that go to w and eventually down to $-\ell$. (Note that the paths may wobble around w several times, but may not hit u before $-\ell$). But this is the set of *all* paths that go to $-\ell$ (without hitting u), minus those that go to $-\ell$ without hitting w. Again we need to take only the proportion q of these paths which were not terminated at w. Thus the event described in Item 4 has probability proportional to $q(f_{-\ell}^{(u)}(t) - f_{-\ell}^{(w)}(t))$.

Finally the probability that the hospital stay ends in the interval $(t, t + dt)$ is proportional to

$$f_{-\ell}^{(w)}(t) + p\ f_w^{(-\ell)}(t) + q\ f_u^{(-\ell)}(t) + q[f_{-\ell}^{(u)}(t) - f_{-\ell}^{(w)}(t)]$$

which can be written as

$$p\ f_{-\ell}^{(w)}(t) + p\ f_w^{(-\ell)}(t) + q\ f_u^{(-\ell)}(t) + q f_{-\ell}^{(u)}(t).$$

This is of course the density of the random variable T, time in hospital when transfers are possible.

The likelihood contributions for the various individuals in the data set are now as follows: Individuals who are transferred at time t, corresponding to item 2, contribute $pf_w^{(-\ell)}(t)$. Individuals who are discharged at time t, corresponding to item 3, contribute $qf_u^{(-\ell)}(t)$. Individuals who die at time t may or may not have passed through the intermediate barrier. Thus the event "death at time t" is the union of the two events described in items 1 and 4. These individuals thus contribute $f_{-\ell}^{(w)}(t) + q[f_{-\ell}^{(u)}(t) - f_{-\ell}^{(w)}(t)]$. This can be more succinctly written as $pf_{-\ell}^{(w)}(t) + qf_{-\ell}^{(u)}(t)$.

Ultimately, assuming u and ℓ are finite, one of the three mutually exclusive outcomes (death in hospital, discharge home, or transfer) must occur. We can define the discrete random variable D which takes values l, u, or w according to whether the individual dies in hospital, is discharged or transferred. Then the probability of ultimate discharge is

$$P(D = \mathrm{u}) = \lim_{T\to\infty} \int_0^T q f_u^{(-\ell)}(t)\, dt = \lim_{T\to\infty} q F_u^{(-\ell)}(T)$$

The probability of death in hospital is

$$\begin{aligned} P(D = \mathrm{l}) &= \lim_{T\to\infty} \int_0^T q f_{-\ell}^{(u)}(t) + p f_{-\ell}^{(w)}(t)\, dt \\ &= \lim_{T\to\infty} q F_{-\ell}^{(u)}(T) + p F_{-\ell}^{(w)}(T) \end{aligned}$$

Finally, the probability of transfer is

$$P(D = \mathrm{w}) = \lim_{T\to\infty} \int_0^T p f_w^{(-\ell)}(t)\, dt = \lim_{T\to\infty} p F_w^{(-\ell)}(T)$$

It is also possible to accommodate independently censored individuals, for instance those lost to follow-up or still in hospital at the end of the study. We can represent the event [censored at time t] as the union of the two events, [censored at time t without passing through w] and [censored at time t having passed through w at time $t_1 < t$]. Note that we cannot observe when, or even if, an individual who is not transferred passes through w.

Here again we will need some new notation. Write $\mathcal{F}^{(b,c)}(t)$ for the probability that neither the barrier at b, nor the barrier at c is breached by time t, i.e. that the time of breach is greater than t. This can be calculated as in equation 3.3. Then the probability that an individual is censored at time t without passing through w, is the probability that neither the barrier at w or $-\ell$ has been breached by time t. In the new notation, this probability is $\mathcal{F}^{(-\ell,w)}(t)$.

Individuals who are censored at time t, having previously passed through w are more complicated. The probability that an individual hits w for the first time at time t_1 and is not transferred is $q f_w^{(-\ell)}(t_1)$. Given the occurrence of this event, the probability that he has not hit the upper or lower barrier by time t is then $\mathcal{F}^{(-(\ell+w),u-w)}(t - t_1)$. Thus the probability that an individual is censored at time t, having previously passed through w is

$$q \int_0^t f_w^{(-\ell)}(t_1)\, \mathcal{F}^{(-(\ell+w),u-w)}(t-t_1)\, dt_1.$$

We will denote this integral as $\mathcal{I}^{u,w,-\ell}(t)$. Thus individuals who are censored at time t contribute $\mathcal{F}^{-\ell,w}(t) + \mathcal{I}^{u,w,-\ell}(t)$ to the likelihood.

Unfortunately, the integral $\mathcal{I}^{u,w,-\ell}(t)$ has no closed form in general. Thus calculation of the likelihood would involve a numerical integration, and this would have to be repeated at each step of the optimization procedure which finds the maximum likelihood estimates. Since this is a very time consuming operation, we would like to find another expression for the probability of being censored at time t.

We do this by noting that someone who is not censored at time t was either discharged, transferred or died sometime before time t, ie their hospital stay was less than t. This event has probability

$$\begin{aligned} P(T < t) &= \int_0^t \left[q f_u^{(-\ell)}(s) + p f_{-\ell}^{(w)}(s) + q f_{-\ell}^{(u)}(s) + p f_w^{(-\ell)}(s) \right] ds \\ &= q F_u^{(-\ell)}(t) + p F_{-\ell}^{(w)}(t) + q F_{-\ell}^{(u)}(t) + p F_w^{(-\ell)}(t) \qquad (6.3) \end{aligned}$$

where $F_a^{(b)}(t)$ is the probability that the barrier at a is hit before time t, without first hitting the barrier at b. Write $F_{a,b}(t)$ for the probability that one of the barriers at a or b was breached before time t, and note that $F_{a,b}(t) = 1 - \mathcal{F}^{(a,b)}(t)$, so an expression is available. Of course $F_a^{(b)}(t) + F_b^{(a)}(t) = F_{a,b}(t)$. Now the complement of the above event, the probability of being censored at time t, has probability $1 - qF_{u,-\ell}(t) - pF_{w,-\ell}(t)$. Thus the log-likelihood for the transfer model is

$$\begin{aligned} \mathcal{L}(u, -\ell, w, p, \mu) = &\sum_{\{i:D_i=u\}} \log[q f_u^{(-\ell)}(t_i)] + \sum_{\{i:D_i=l\}} \log[p f_{-\ell}^{(w)}(t_i) + q f_{-\ell}^{(u)}(t_i)] \\ &+ \sum_{\{i:D_i=w\}} \log[p f_w^{(-\ell)}(t_i)] + \sum_{\{i:D_i=c\}} \log[1 - qF_{u,-\ell}(t_i) - pF_{w,-\ell}(t_i)] \end{aligned}$$

where D_i is u, l, w, or c depending on whether individual i is discharged, dies, is transferred or is censored at time t_i. The parameters to be estimated in this model are $u, -\ell, w, p$ and μ.

As before, we can contemplate modelling these parameters as functions of covariates. We can allow the barrier levels to depend on covariates exactly as before. We can also allow p, the probability of transfer given that the decision barrier at w is breached, to depend on covariates. Since p must always be between 0 and 1, a logistic model for p would be appropriate, ie $\log(p/(1-p)) = x\beta$. However for the preliminary investigations described in this thesis, we have used a linear model for p, i.e. $p = x\beta$, which, at least so far, has always given an estimate of p in the admissible range. The extension to covariates is straightforward and is described in appendix E.3.

The Cox-Snell residuals for this model are easily constructed. Recall that T, the time in hospital in the presence of transfers, has survivor function $1 - qF_{u,-\ell}(t) - pF_{w,-\ell}(t)$. The Cox-Snell residual for individual i is then

$$R_i = -\log[1 - qF_{u_i,-\ell_i}(t_i) - pF_{w_i,-\ell_i}(t_i)]$$

where $u_i = (\hat{\underline{\beta}}^u)'\underline{x}_i^u$, $\ell_i = (\hat{\underline{\beta}}^\ell)'\underline{x}_i^\ell$, and $w_i = (\hat{\underline{\beta}}^w)'\underline{x}_i^w$. Note that the function $F_{a,b}(t)$ depends also on the drift parameter which is estimated as $\mu_i = (\hat{\underline{\beta}}^\mu)'\underline{x}_i^\mu$.

The expectation of time in hospital, and the probability that the patient is discharged, transferred or died in hospital, are important quantities. Also the expected time in hospital given outcome is useful for model assessment. These quantities can be found as simple functions of the corresponding quantities in the model without transfers.

First, consider the model with only two barriers, and no mechanism for transfer. Let $T_{u,\ell}$ be the random variable representing time until breach of one of the two barriers $-\ell < 0$ or $u > 0$ in a Wiener process with drift. Also, let $D_{u,\ell}$ be a discrete variable which takes the value l if the lower barrier at $-\ell$ is breached, and which takes the value u if the upper barrier at u is breached. If we define functions corresponding to $E(T_{u,\ell})$, $P(D_{u,\ell} = \text{u})$ and $E(T_{u,\ell}|D_{u,\ell} = \text{u})$ then we can specify all the necessary quantities discussed above for the transfer model in terms of these functions. Let

$$h_1(u, \ell, \mu) = -\frac{\ell}{\mu} + \frac{u+\ell}{\mu}\left[\frac{1 - e^{-2\mu\ell}}{1 - e^{-2\mu(u+\ell)}}\right]$$

$$h_2(u,\ell,\mu) = \frac{1-e^{2\mu\ell}}{e^{-2\mu u}-e^{2\mu\ell}},$$

and

$$h_3(u,\ell,\mu) = \left[\mu\left(-1+e^{-2\mu(u+\ell)}\right)^2\left(-1+e^{-2\mu}\right)\right]^{-1}\left[-\left(-e^{-2\mu(u+\ell+1)}+e^{-2\mu(\ell+1)}+\right.\right.$$
$$e^{-2\mu(u+\ell)}+e^{-2\mu}-e^{-2\mu\ell}-e^{-2\mu(u+2\ell+1)}+e^{-2\mu(u+2\ell)}\Big)\ell-\Big(e^{-2\mu(\ell+1)}-e^{-2\mu\ell}$$
$$\left.\left.+e^{-2\mu(u+2\ell+1)}-e^{-2\mu(u+2\ell)}-e^{-2\mu}+e^{-2\mu(u+\ell)}-e^{-2\mu(u+\ell+1)}\right)(u+\ell)-u\right].$$

Then

$$\begin{aligned}
E(T_{u,\ell}) &= \int_0^\infty t\left[f_u^{(\ell)}(t)+f_\ell^{(u)}(t)\right]\,dt &&= h_1(u,\ell,\mu)\\
P(D_{u,\ell}=\mathrm{u}) &= \int_0^\infty f_u^{(\ell)}(t)\,dt &&= h_2(u,\ell,\mu)\\
P(D_{u,\ell}=\mathrm{l}) &= \int_0^\infty f_\ell^{(u)}(t)\,dt &&= h_2(\ell,u,-\mu)\\
E(T_{u,\ell}|D_{u,\ell}=\mathrm{u}) &= \frac{\int_0^\infty t f_u^{(\ell)}(t)\,dt}{P(D_{u,\ell}=\mathrm{u})} &&= \frac{h_3(u,\ell,\mu)}{h_2(u,l,\mu)}\\
E(T_{u,\ell}|D_{u,\ell}=\mathrm{l}) &= \frac{\int_0^\infty t f_\ell^{(u)}(t)\,dt}{P(D_{u,\ell}=l)} &&= \frac{h_3(\ell,u,-\mu)}{h_2(\ell,u,-\mu)}
\end{aligned}$$

Now consider the model which allows for transfer to another institution, and let $T_{u,\ell,w}$ represent the time until either discharge, death, or transfer, and let $D_{u,\ell,w}$ take the value u, l, or w depending on whether discharge, death, or transfer has occurred. Then

$$
\begin{aligned}
E(T_{u,\ell,w}) &= \int_0^\infty t\left[qf_u^{(\ell)}(t) + qf_\ell^{(u)}(t) + pf_\ell^{(w)}(t) + pf_w^{(\ell)}(t)\right]\, dt \\
&= qE(T_{u,\ell}) + pE(T_{w,\ell}) \qquad\qquad = q\ h_1(u,\ell,\mu) + p\ h_1(w,\ell,\mu) \\
P(D_{u,\ell,w} = \mathrm{u}) &= \int_0^\infty qf_u^{(\ell)}(t)\, dt \\
&= qP(D_{u,\ell} = \mathrm{u}) \qquad\qquad = q\ h_2(u,\ell,\mu) \\
P(D_{u,\ell,w} = \mathrm{l}) &= \int_0^\infty qf_\ell^{(u)}(t) + pf_\ell^{(w)}(t)\, dt \\
&= qP(D_{u,\ell} = \mathrm{l}) + pP(D_{w,\ell} = \mathrm{l}) \qquad\qquad = q\ h_2(\ell,u,-\mu) + p\ h_2(\ell,w,-\mu) \\
P(D_{u,\ell,w} = \mathrm{w}) &= \int_0^\infty pf_w^{(\ell)}(t)\, dt \\
&= pP(D_{w,\ell} = \mathrm{w}) \qquad\qquad = p\ h_2(w,\ell,\mu)
\end{aligned}
$$

Finally

$$
\begin{aligned}
E(T_{u,\ell,w}|D = \mathrm{u}) &= \frac{P(T_{u,\ell,w} \in (t,t+dt), D_{u,\ell,w} = \mathrm{u})}{P(D_{u,\ell,w} = \mathrm{u})} = \frac{\int_0^\infty tqf_u^{(\ell)}(t)\, dt}{qP(D_{u,\ell} = \mathrm{u})} \\
&= \frac{q\ h_3(u,\ell,\mu)}{q\ h_2(u,\ell,\mu)}
\end{aligned}
$$

$$
\begin{aligned}
E(T_{u,\ell,w}|D = l) &= \frac{P(T_{u,\ell,w} \in (t,t+dt), D_{u,\ell,w} = l)}{P(D_{u,\ell,w} = l)} = \frac{\int_0^\infty t\left[qf_\ell^{(u)}(t) + pf_\ell^{(w)}\right]\, dt}{qP(D_{u,\ell} = l) + pP(D_{w,\ell} = l)} \\
&= \frac{q\ h_3(\ell,u,-\mu) + p\ h_3(\ell,w,-\mu)}{q\ h_2(\ell,u,-\mu) + p\ h_2(\ell,w,-\mu)}
\end{aligned}
$$

$$
\begin{aligned}
E(T_{u,\ell,w}|D = \mathrm{w}) &= \frac{P(T_{u,\ell,w} \in (t,t+dt), D_{u,\ell,w} = \mathrm{w})}{P(D_{u,\ell,w} = \mathrm{w})} = \frac{\int_0^\infty tpf_w^{(\ell)}(t)\, dt}{pP(D_{w,\ell} = \mathrm{w})} \\
&= \frac{p\ h_3(w,\ell,\mu)}{p\ h_2(w,\ell,\mu)}
\end{aligned}
$$

6.3 Starting Values

For the transfer model, no simple relationship exists between $P(D = u)$ and $E(T|D = u)$ which can be exploited to get starting values for the regression parameters. For starting values we again use a mock least squares estimate, as described in section 2.2.3, but with $g(T_u)$ now equal to the simple function $\sqrt{T_u}$. Recall that $g(T_u)$ is chosen to behave approximately like u. This choice of $g(.)$ can be rationalized by the self-similarity of the Wiener process, which remains the same in distribution if we change the time scale by a factor of c and the space scale by a factor of $\sqrt{c}$.

Specifically, for individuals who were discharged, we form the vector T_u of observed LOS, and the matrix X_u of covariates thought to affect the upper barrier. Then we use $(X_u' X_u)^{-1}(X_u' \sqrt{T_u})$ as the starting values for the vector of regression parameters $\underline{\beta}^u = (\beta_0^u, \beta_1^u, \dots, \beta_{p_u}^u)'$. A similar strategy is used to find starting values for $\underline{\beta}^t$ and $\underline{\beta}^w$.

Again we can use zero as the starting value for the drift parameters. When there are no covariates on the parameter p, an obvious starting value is to use the proportion of patients in the data set who were transferred. This will be a very rough estimate of p, which represents the proportion of individuals who passed through the transfer barrier and were transferred. Nevertheless it seems to work well in practice. If covariates are present on the parameter p, then it suffices to use the starting value discussed, and repeat it for each covariate.

6.4 Example

In this section we illustrate the transfer model using a larger subset of the data, comprising females of all ages admitted with an APR-DRG of 121 (Circulatory Disorders with MI), a total of 445 cases. Here we test whether type of payer affects probability of transfer to another acute care hospital, adjusting for severity of disease. Type of payer is coded at three levels (government insurer, private insurer and managed care). Let $z_{i1} = 1$ if individual i has a government insurance plan (Medicare, Medicaid, or other government) and let $z_{i2} = 1$ if individual i has managed care. Baseline individuals, with z_{i1} and z_{i2} both equal to 0, have private insurance plans (Blue Cross, Blue Shield or other private) . As before, let $x_{i1} = 1$ if individual i exhibits a major complication or comorbidity, and let $x_{i2} = 1$ if individual i has an extreme complication or comorbidity.

Individuals with no, minor, or moderate complication or comorbidity have both x_{i1} and x_{i2} equal to 0.

Thus the drift for individual i is

$$\mu_i = \beta_0^\mu + \beta_1^\mu\, x_{i1} + \beta_2^\mu\, x_{i2}$$

while the probability of transfer, given the health process reaches the intermediate barrier at w is given by

$$p_i = \beta_0^p + \beta_1^p\, z_{i1} + \beta_2^p\, z_{i2}$$

The following table shows the maximum likelihood estimates and standard errors for this model.

	u	l	β_0^μ	β_1^μ	β_2^μ	w	β_0^p	β_1^p	β_2^p
Estimate	3.743	1.832	0.829	-0.364	-0.687	2.212	0.488	-0.239	-0.169
Std Err	0.140	0.099	0.043	0.056	0.068	0.115	0.076	0.080	0.102

Increasing severity reduces the upward drift, although drift remains positive for all covariate groups. The conditional probability of transfer given that the intermediate barrier is reached is estimated to be highest for individuals with private insurance, somewhat lower for those with managed care (though not significantly so), and significantly lower for patients with government insurance.

The observed and predicted values are shown in the following table for the least severely ill patients. The higher severity groups have less than 10 individuals in some covariate cells and are not shown here.

Severity	Payer	n	$P(D=\mathrm{u})$	$P(D=l)$	$P(D=\mathrm{w})$
low	private	37	0.49(0.48)	0.03(0.05)	0.49(0.47)
low	managed	39	0.67(0.65)	0.00(0.05)	0.33(0.30)
low	gov't	218	0.69(0.72)	0.05(0.05)	0.26(0.24)

Severity	Payer	n	T	$T\|D=\mathrm{u}$	$T\|D=l$	$T\|D=\mathrm{w}$
low	private	37	2.70(3.34)	3.33(4.29)	4.00(2.14)	2.00(2.46)
low	managed	39	2.77(3.63)	3.35(4.29)	. (2.15)	1.62(2.46)
low	gov't	218	4.02(3.76)	4.66(4.29)	3.54(2.16)	2.42(2.46)

The model is doing a very good job of predicting probability of each outcome (discharge, death, transfer). Note that none of the 39 individuals in the managed care group died in the original admitting hospital. The (unconditional) probability of transfer, $P(D = \mathrm{w})$, is highest for individuals with private insurance, and lowest for patients with government insurance, with managed care falling in between. Note that the expected time in hospital given $D = \mathrm{u}$ or $D = \mathrm{w}$ is the same for each payer, since payer is not assumed to affect the drift or barrier levels, only the probablity of transfer. So given transfer, payer will not affect expected time in hospital. However the probability of transfer is used to calculate expected LOS given $D = \mathrm{l}$ so it changes slightly.

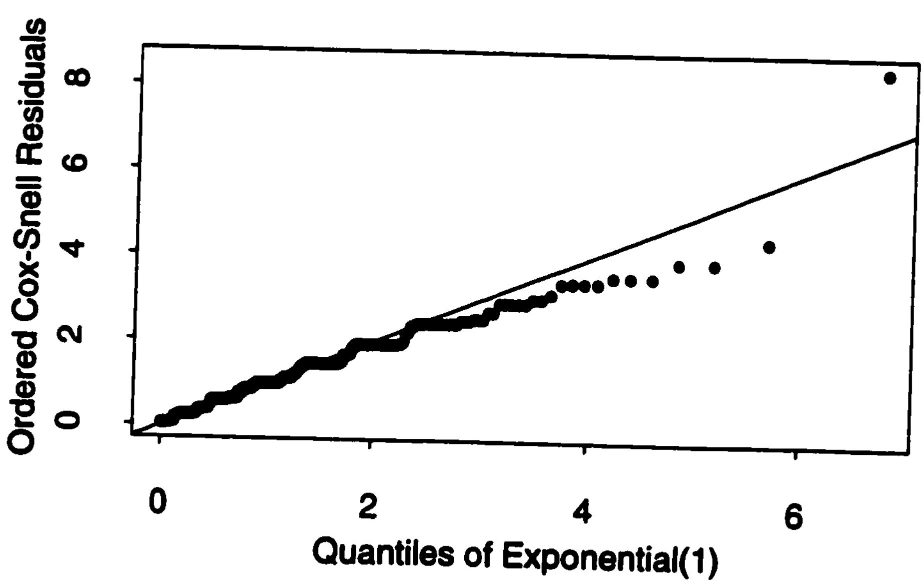


Figure 6.4:

The model appears to fit quite well, except for the approximately 10 points that appear in the top right quadrant of the plot. Some investigation shows that these points belong, for the most part, to individuals who were discharged from hospital, and had long stays in hospital (7 days or more). The model predictions are much lower than the observed LOS for these individuals. Again, this could be due to a missing covariate. However, the model does a good job of prediction for the bulk of the data.

Chapter 7

Comparison with Competing Risks in Survival Analysis

In this chapter, we briefly review some of the common models used in survival analysis where the response variable is time to a single type of failure. We then discuss competing risks, in which there are several types of failure, and outline two approaches to the problem. We discuss the relationship between certain competing risks models and FP2B models. In the final section, we fit some of the competing risks models to the data on females with circulatory disorders and myocardial infarction and compare the estimates obtained to those from the FP2B model discussed in section 3.5.

7.1 Common models used in Survival Analysis

Survival analysis is concerned with modelling T, the time until some event, often a failure. Usually some of the observations are censored, which means that the actual failure time is unobserved, but is known to be greater than an observed censoring time. The presence of censoring greatly complicates the analysis in many cases, and so special methods of analysis have been developed for failure time data.

There are many possible censoring schemes. In a random censoring mechanism, for individual i we observe $X_i = \min(T_i, C_i)$ where both T_i and C_i are random variables, with T_i being time until failure and C_i a censoring time. This design includes many clinical trials where the end of the study is fixed, but patients are accrued randomly over time. It also includes studies in which the individual censoring times are fixed in advance. In Type II censoring, the study is ended after a pre-determined number of failures have been observed. All these schemes are examples of independent censoring mechanisms. Censoring is independent if an individual who is censored at time t is representative of all other individuals who share the same covariate values, and who survive to time t.

As usual, let $f(t)$ denote the density of T and $F(t)$ its cumulative distribution function. The survivor function $\mathcal{F}(t) = 1 - F(t) = P(T \geq t)$ gives the probability of surviving past time t. An important concept in survival analysis is the hazard function, $h(t)$, defined as

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{\mathcal{F}(t)},$$

where $P(A)$ denotes the probability of event A. The value of the hazard at time t gives the instantaneous probability of failure at time t, given that an individual has survived to time t. Another useful function is the integrated or cumulative hazard

$$\mathcal{H}(t) = \int_0^t h(s)\, ds.$$

These definitions lead to the relationships

$$h(t) = -\frac{d}{dt} \log(\mathcal{F}(t))$$

$$\mathcal{F}(t) = \exp\left(-\int_0^t h(u)\, du\right) = \exp(-\mathcal{H}(t))$$

and

$$\mathcal{H}(t) = -\log \mathcal{F}(t)$$

In the presence of censoring, the survivor function of T in a homogeneous population can be estimated by the Kaplan Meier or product-limit estimate, $\hat{\mathcal{F}}(t)$. This is a non-parametric maximum likelihood estimate. Suppose that the distinct observed failure times in the data set are $t_1 < t_2 < \ldots < t_p$. Then

$$\hat{\mathcal{F}}(t) = \prod_{k|t_k<t} \left(1 - \frac{d_k}{n_k}\right) \tag{7.1}$$

where d_k is the number of failures at time t_k and n_k is the number of individuals still at risk (ie those who have not failed and are uncensored) just before time t_k. Clearly, $\hat{\mathcal{F}}(t)$ is a step function, with $\hat{\mathcal{F}}(0) = 1$ and jumps at the observed failure times.

Suppose we have data on time to failure for n independent individuals, and an independent censoring mechanism. Then the parametric likelihood is

$$\prod_{i=1}^{n} f(t_i)^{\delta_i} \mathcal{F}(t_i)^{1-\delta_i} = \prod_{i=1}^{n} h(t_i)^{\delta_i} \mathcal{F}(t_i),$$

where t_i is the time of either failure or censoring for individual i, and where δ_i equals 1 if individual i is actually observed to fail and 0 if individual i is censored at time t_i. Note that, in the last expression, while only uncensored individuals contribute a factor of $h(t_i)$ to the likelihood, all individuals, both censored and uncensored, contribute a factor of $\mathcal{F}(t_i)$. The likelihood can be written

$$\prod_{i=1}^{n} h(t_i)^{\delta_i} \exp\left[-\int_0^{t_i} h(u)\, du\right]. \tag{7.2}$$

Parametric families of distributions commonly used to model survival data are the exponential, gamma, Weibull, log-normal, or log-logistic. We note in passing that for the exponential distribution, the hazard is constant over time, i.e. $h(t) = \lambda$, and the survivor function is $\mathcal{F}(t) = \exp(-\lambda t)$. The Weibull distribution has hazard $h(t) = \lambda^p p t^{p-1}$ and survivor function $\mathcal{F}(t) = \exp(-(\lambda t)^p)$.

There are two broad classes of models in common use to study the effect of covariates on failure time, proportional hazard models and accelerated failure time models. In accelerated failure time (AFT) models, the covariates act multiplicatively on T, and so the effect of a covariate is to accelerate (or decelerate) the time until failure. Letting $Y = \log(T)$, an AFT model for Y is linear in the parameters and has the form

$$Y = x\beta + W$$

where W is a random variable from some specified distribution. Usually W is chosen so that $\exp(W)$ has one of the following distributions: exponential, gamma, Weibull, log-normal, or log-logistic. The hazard for the AFT model has the form

$$h(t, x, \beta) = h_0(t \exp(-x\beta)) \exp(-x\beta)$$

and the survivor function is

$$\mathcal{F}(t, x, \beta) = \exp\left[-\int_0^{t \exp(-x\beta)} h_0(u)\, du\right]$$

where $h_0(t)$ is the baseline hazard function of $\exp(W)$, ie $h_0(t)$ is the hazard function of T when $x = 0$. The likelihood for an AFT model is formed as in equation (7.2) with $h(t)$ parameterized as shown above. We note that the exponential AFT model has hazard $h(t, x, \beta) = \lambda \exp(-x\beta)$, while the hazard for the Weibull AFT model is $h(t, x, beta) = \lambda^p p t^{p-1} \exp(-x\beta p)$. Note that if an intercept is included in the model (ie the first covariate has the value 1 for all individuals in the data set), then we can, without loss of generality, take $\lambda = 1$. The parameterization used by Splus has $\lambda = 1$ and a scale parameter $\sigma = 1/p$.

In the proportional hazards model, the covariates act multiplicatively on the hazard, usually through the relationship

$$h(t, x, beta) = h_0(t) \exp(x\beta),$$

where x is a vector of covariates, β is a vector of regression parameters, and $h_0(t)$ is a baseline hazard function, the hazard for an individual with $x = 0$. The survivor function is then given by

$$\mathcal{F}(t, x, \beta) = [\mathcal{F}_0(t)]^{\exp(x\beta)}$$

where

$$\mathcal{F}_0(t) = \exp\left[-\int_0^t h_0(u)\, du\right].$$

The most popular proportional hazards model is the Cox model, a semi-parametric model in which the baseline hazard $h_0(t)$ is left unspecified. The β parameters of this model are estimated using the method of partial likelihood. There is some overlap of AFT and proportional hazard models, in that Weibull AFT models (which include exponential models as a special case) can have proportional hazards.

7.2 Competing Risks Models

Often we know not only the time of failure, but we can also designate the failure as having one of several possible modes, or attribute the failure to one of several possible causes. For instance a machine, consisting of several components, may fail when any one of its components fails. A person's cause of death might be heart disease or cancer. The multiple modes of failure are termed competing risks.

There are two distinct approaches to analyzing this kind of data. The classical approach, going back to Daniel Bernoulli, postulates the existence of m random variables $T_1, T_2, \ldots, T_m$, where T_j is the time at which failure mode j occurs. We observe only the first of these failure times, $T = \min(T_1, T_2, \ldots, T_m)$, and also which failure type has occurred. The remaining failure times are latent, or unobserved. With this kind of data, the joint distribution of the T_i's can be estimated only in very special circumstances, for instance if the T_i's are multivariate normal [30] or the $T_i's$ are independent from a distribution with non-vanishing right tail [29].

Typically the problem posed in the classical approach is to estimate the marginal distribution of T_j if one or more of the other causes of failure could be "eliminated". Often one assumes that the T_i are independent. In the case of a machine with independent components, it may make sense to consider elimination of a failure type. However with human subjects and causes of death, it is unlikely that failure types are independent. As pointed out by Cox and later studied in detail by Tsiatis, with the type of data available in the competing risks situation, it is impossible to distinguish between a model with dependent T_j and one with independent T_j. Even if independence is assumed, a mechanism for cause-removal would still need to be specified, likely involving other assumptions (for instance, that individuals who would have died from the eliminated cause are subject to the same risk of death from the remaining causes as the rest of the population).

The other approach to competing risks was expounded by Prentice et al [32]. In this approach, we model the time to failure T, and a discrete random variable D which represents cause of failure. Suppose that there are m causes of failure, so that D takes values in the set $\{1, 2, \dots, m\}$. The cause-specific hazard for cause j is defined as

$$h_j(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t, D = j \mid T \geq t)}{\Delta t} = \frac{f_j(t)}{\mathcal{F}(t)}$$

where

$$f_j(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t, D = j\ )}{\Delta t}.$$

We can also define cause-specific subsurvivor functions

$$\mathcal{F}_j(t) = P(T \geq t, D = j) = \int_t^\infty f_j(s)\, ds$$

Note that these three functions, because they are probabilities, are additive:

$$h(t) = \sum_{j=1}^{m} h_j(t).$$

$$f(t) = \sum_{j=1}^{m} f_j(t).$$

$$\mathcal{F}(t) = \sum_{j=1}^{m} \mathcal{F}_j(t).$$

The marginal distribution of D is given by

$$\pi_j = P(D = j) = \mathcal{F}_j(0) \qquad j = 1, 2, \dots, m$$

Typically the question of interest in this approach to competing risks is the effect of covariates on the time until failure from cause j. Suppose that on each of n individuals, $i = 1, 2, \dots, n$, we observe time of failure or censoring t_i, the censoring indicator δ_i, the cause of failure j_i and a vector of covariates x_i. The parametric likelihood is then

$$\prod_{i=1}^{n} [h_{j_i}(t_i, x_i, \beta)]^{\delta_i} \mathcal{F}(t_i, x_i, \beta)$$

where β is a vector of parameters to be estimated. We now perform some manipulations which allow us to show that, as long as each of the cause-specific hazards is parameterized individually, the likelihood factors into m terms, one for each cause of failure.

Define the pseudo-survivor functions

$$\mathcal{G}_j(t, x, \beta) = \exp\left[-\int_0^t h_j(s, x, \beta)\, ds\right]$$

and note that the survivor function

$$\mathcal{F}(t, x, \beta) = \exp\left[-\int_0^t \sum_{j=1}^{m} h_j(s, x, \beta)\, ds\right] = \prod_{j=1}^{m} \mathcal{G}_j(t, x, \beta).$$

(The pseudo-survivor functions do not have an interpretation as probabilities, in general). Then the likelihood can be written as

$$\prod_{i=1}^{n} [h_{j_i}(t_i, x_i, \beta)]^{\delta_i} \mathcal{F}(t_i, x_i, \beta) = \prod_{i=1}^{n} [h_{j_i}(t_i, x_i, \beta)]^{\delta_i} \prod_{j=1}^{m} \mathcal{G}_j(t_i, x_i, \beta)$$

$$= \prod_{i=1}^{n} [h_{j_i}(t_i, x_i, \beta)]^{\delta_i} \prod_{j=1}^{m} \exp\left[-\int_0^{t_i} h_j(s, x_i, \beta)\, ds\right].$$

This factors into m terms, one for each cause, with the j^{th} term given by

$$\prod_{i=1}^{n} [h_j(t_i, x_i, \beta)]^{\delta_{ij}} \exp\left[-\int_0^{t_i} h_j(s, x_i, \beta)\, ds\right] \tag{7.3}$$

where δ_{ij} equals 1 if $j_i = j$ and if individual i is uncensored, and 0 otherwise. Note that all individuals, whether censored, failed from cause j, or failed from other causes, contribute to the last factor of the likelihood. Thus the j^{th} factor of the likelihood has the same form as the likelihood for a single cause of failure given in equation (7.2). In fact, equation 7.3 gives the same value as we would get by treating failures due to causes other than cause j as censored, and constructing the likelihood as in 7.2. Suppose that we parameterize the model so that for each

j, $h_j(t, x, \beta)$ is a function of a vector of parameters β_j and that β_j and β_k have no elements in common. The likelihood factor for the j^{th} cause then becomes

$$\prod_{i=1}^{n} [h_j(t_i, x_i, \beta_j)]^{\delta_{ij}} \exp\left[-\int_0^{t_i} h_j(s, x_i, \beta_j)\, ds\right] \tag{7.4}$$

and we may estimate the cause-specific parameters (β_j) separately for each cause of failure, by maximizing equation (7.4).

A useful function for summarising the survival experience of individuals is the cause-specific subdistribution function $F_j(t)$, the probability that an individual fails from cause j *and* does so before time t:

$$F_j(t) = P(T \leq t, D = j) = \int_0^t f_j(s)\, ds = \int_0^t h_j(s)\mathcal{F}(s)\, ds.$$

Note that

$$F(t) = \sum_{j=1}^{m} F_j(t)$$

$$\pi_j = P(D = j) = F_j(\infty)$$

and

$$F_j(t) = \pi_j - \mathcal{F}_j(t)$$

The function $F_j(t)$ goes by various names in the literature, namely cumulative incidence function (Kalbfleisch and Prentice), absolute cause-specific risk, crude incidence curve, and cause-specific failure probability.

A point of much confusion to non-statisticians is how to estimate survival probabilities in the presence of competing risks. A common practice is, for the j^{th} cause of failure, to form the Kaplan Meier estimate obtained by treating all failure times from causes other than j as censored. It can be shown that the Kaplan-Meier estimate obtained in this way in fact estimates the pseudo-survivor function $\mathcal{G}_j$. Plots of $\mathcal{G}_j(t)$ versus time are useful in that they illustrate the cause-specific hazard functions on a certain scale. However they should not be interpreted as probabilities. As we saw above, $F_j(t)$ does represent a probability, namely the probability that an individual will fail prior to time t from cause j. Unfortunately, the common statistical packages do not have routines for estimating $F_j(t)$.

Recall the classical competing risks framework where $T = \min(T_1, \ldots, T_m)$. If the T_j are independent, then $\mathcal{G}_j$ has an interpretation as the survivor function of the latent variable T_j. Then $f(t)$ is the density of the minimum observed time, T. However, in many cases it is not reasonable to assume independence, and as mentioned above, this assumption cannot be checked.

Suppose that we want to quantify the survival experience for a homogeneous group of individuals. Note that this might be a subset of the data consisting of all individuals who share the same covariate values. To estimate the pseudo-survivor function, $\mathcal{G}_j$, we treat failure times due to causes other than j as censored and form the Kaplan Meier estimate, which we will denote $\hat{\mathcal{G}}_j$. Let $t_{j1} < t_{j2} < \ldots < t_{jK}$ be the distinct observed failure times for cause j. Then

$$\hat{\mathcal{G}}_j(t) = \prod_{k|t_{jk}<t} \left(1 - \frac{d_{jk}}{n_{jk}}\right)$$

where d_{jk} is the number of failures due to cause j at time t_{jk} and n_{jk} is the total number of individuals in the data set still at risk just before time t_{jk}. It's easy to see that, as long as there are no ties among observed failure times from different causes,

$$\hat{\mathcal{F}}(t) = \prod_{j=1}^{m} \hat{\mathcal{G}}_j(t)$$

where $\hat{\mathcal{F}}(t)$, given in equation (7.1), is the regular Kaplan Meier estimate formed by ignoring cause of failure. Then a non-parametric estimate of $F_j(t)$ is

$$\hat{F}_j(t) = \sum_{k|t_{jk}<t} \frac{d_{jk}}{n_{jk}} \hat{\mathcal{F}}(t_{jk}).$$

Then we can plot both $\hat{\mathcal{G}}_j(t)$ and $\hat{F}_j(t)$ against time t. While $\hat{F}_j(t)$ can be used to predict the probability that individuals will fail from cause j before time t, $\hat{\mathcal{G}}_j(t)$ can only be used for model assessment.

Suppose now that we are modeling the effects of covariates using an accelerated failure time model. Maximization of the j^{th} component of the likelihood given in equation (7.3) will give estimates $\hat{\beta}_j$ of the parameters. An estimate of the j^{th} cause-specific hazard function is

$$\hat{h}_j(t) = h_j(t, x, \hat{\beta}_j) = h_{0j}(t \exp(-x\hat{\beta}_j)) \exp(-x\hat{\beta}_j)$$

and of the j^{th} pseudo-survivor function is

$$\hat{\mathcal{G}}_j(t) = \mathcal{G}(t, x, \hat{\beta}_j) = \exp\left[-\int_0^{t\exp(-x\hat{\beta}_j)} h_{0j}(s)\, ds\right]$$

where $h_{0j}(t)$ is the baseline hazard for an individual who fails from cause j. (A more restrictive model would have $h_{0j} = h_0$ for all j). Then an estimate of the overall survivor function is

$$\hat{\mathcal{F}}(t) = \mathcal{F}(t, x, \hat{\beta}_j) = \prod_{j=1}^{m} \hat{\mathcal{G}}_j(t)$$

and an estimate of the cumulative incidence function is

$$\hat{F}_j(t) = F_j(t, x, \hat{\beta}) = \int_0^t \hat{h}_j(s)\hat{\mathcal{F}}(s)\, ds.$$

These functions are easily calculated when each cause-specific hazard is constant over time, i.e. has the form of an exponential hazard. To allow dependence on covariates, we will use an exponential AFT model. We will take $h_{0j}(t) = 1$, which involves no loss of generality if an intercept is included in the model. Then, for an individual with covariate values given by the vector x,

$$h_j(t) = \exp(-x\beta_j)$$

$$\mathcal{G}_j(t) = \exp\{-t\exp(-x\beta_j)\}$$

$$\mathcal{F}(t) = \exp(-ct)$$

$$F_j(t) = c^{-1}\exp(-x\beta_j)\{1 - \exp(-ct)\}$$

$$P(D = j) = F_j(\infty) = c^{-1}\exp(-x\beta_j)$$

where $c = \sum_{j=1}^{m} \exp(-x\beta_j)$. Note that, T the time to failure (ignoring cause) has survivor function $\exp(-tc)$ and so has an exponential distribution with mean c^{-1}. Also notice that the cumulative incidence functions $F_j(t)$ are proportional, differing by a factor of $\exp(-x\beta_j)$. Then the conditional distribution function of T given $D = j$ is

$$F_{T|D}(t|j) = \frac{F_j(t)}{P(D=j)} = 1 - \exp(-ct)$$

which is exponential, and furthermore has no dependence on j. Thus, the conditional distributions of time to failure given outcome are identical (for a given value of the covariate vector). This implies that T and D are independent. This is a feature of proportional risks models, which are models in which the hazard functions for separate competing risks are proportional. Finally,

$$\mathrm{E}(T|D=j) = \mathrm{E}(T) = \int_0^\infty \mathcal{F}(t)dt = c^{-1}.$$

To estimate these functions, we just substitute the estimated value $\hat{\beta}$ for β. The Cox-Snell residual for individual i is calculated as

$$r_i = -\log(\mathcal{F}(t_i, x_i, \hat{\beta})) = \hat{c_i}\ t_i$$

where $\hat{c_i} = \sum_{j=1}^{m} \exp(-x_i\hat{\beta}_j)$.

Another competing risk model specifies that each cause-specific hazard has the form of a Weibull AFT hazard. We will allow a different scale parameter p_j for each of the j outcomes. This ensures that the model is not a proportional risks model. Then

$$h_j(t) = p_j\ t^{p_j - 1} \exp(-x\beta_j p_j)$$

$$\mathcal{G}_j(t) = \exp(-t^{p_j} \exp(-x\beta_j p_j))$$

$$\mathcal{F}(t) = \exp\left(\sum_{j=1}^{m} -t^{p_j} \exp(-x\beta_j p_j)\right)$$

$$F_j(t) = p_j \exp(-x\beta_j p_j) \int_0^t s^{p_j-1} \exp\left(\sum_{j=1}^{m} -s^{p_j} \exp(-x\beta_j p_j)\right) ds$$

Again, to estimate these quantities, we just substitute the estimated value $\hat{\beta}$ for β. Note that $\hat{F}_j(t)$ involves an integral which must be evaluated numerically. $\hat{F}_j(\infty)$, the predicted probability of failure from cause j, can also be found numerically. The survivor function for the conditional distribution of T given that $D = j$ is

$$\mathcal{F}_{T|D}(t|j) = 1 - \frac{F_j(t)}{P(D=j)},$$

and the expected time until failure, given that failure occurs due to cause j is

$$\mathrm{E}(T|D=j) = \int_0^\infty \mathcal{F}_{T|D}(s|j)\, ds.$$

To estimate this last function, note that we must integrate under a curve whose equation has no closed form. Fortunately most numerical integrators can handle this situation.

The Cox-Snell residual for individual i is

$$r_i = \sum_{j=1}^{m} t^{\hat{p}_j} \exp(-x\hat{\beta}_j \hat{p}_j).$$

7.3 Competing risks and FP2B models

The foregoing approach to competing risks, which models the joint distribution of T and D, is clearly similar in many respects to the FP2B models for length of stay outlined in this thesis. In the context of length of stay, time until exit from hospital corresponds to time until failure. The two types of exits, healthy discharge and death, correspond to the causes of failure. Recall that the value of the subdistribution function $F_u(t)$ gives the probability that an individual is discharged by time t, while $F_\ell(t)$ gives the probability that an individual dies in hospital, and does so before time t. $F_u(t)$ and $F_\ell(t)$ are, in the terminology of competing risks, cumulative incidence functions, corresponding to $F_j(t), j = 1, 2$. As we saw in Chapters 2 and 3, these functions arise naturally in the development of the model. They can easily be estimated and plotted by plugging the values of the estimated parameters into equations 3.5 and 3.6. Plots of the cause-specific hazard functions can also be made.

Note that the subdensities for the different outcomes of the FP2B model cannot be estimated separately, since all the parameters are involved in both the upper and lower subdensities. While the covariate x might only affect the upper barrier u, so that the associated parameter, β^{*} say, is only involved in the upper barrier, u appears in both the upper and lower subdensity. The FP2B model fully specifies the joint density of T and D, while competing risks models, generally speaking, only specify the separate subdensities, and not the relationship between them.

The classical competing risks approach does not seem very palatable for modelling LOS in hospital, as it seems somewhat absurd to consider the time someone would be released from hospital, if they didn't die first. Certainly one could not assume independence of the latent random variables time to death and time to discharge.

One advantage of the FP2B model, though not particularly applicable to LOS in hospital, is that we now have a mechanism for removing a cause, in that we can think of letting one of the barriers approach infinity. If the drift is positive, and the lower barrier is removed, the resulting distribution is inverse Gaussian. If the drift is negative and the lower barrier is removed, the resulting distribution is defective (has a mass of probability at infinity), but nevertheless has been successfully used in modelling (see [43]). Predictions of expected time to breach, and probability of breach of the remaining barrier, would be possible if one were willing to assume that, after removal of one barrier, the remaining barrier and the drift parameter remained unchanged.

7.4 Example 1: Exponential competing risks model

We now fit some of the competing risk models to the same data set fitted in sections 2.4 and 3.5.

Here we examine the exponential competing risk model. In the following table we show, in each covariate cell, the observed time in hospital (T), the observed time for those that were discharged ($T|D = \text{u}$), and the observed time for those who died in hospital ($T|D = \text{l}$). The last two columns show the observed proportion of patients who were discharged ($P(D = \text{u})$), and the proportion who ultimately died in hospital ($P(D = \text{l})$). The corresponding expected quantities predicted by the model are shown in brackets.

Severity	n	T	$T \mid D = \mathrm{u}$	$T \mid D = \mathrm{l}$	$P(D = \mathrm{u})$	$P(D = \mathrm{l})$
low	82	4.49	4.61(4.49)	3.60(4.49)	0.88(0.88)	0.12(0.12)
med	39	4.67	5.96(4.67)	2.36(4.67)	0.64(0.64)	0.36(0.36)
high	13	4.92	8.80(4.92)	2.50(4.92)	0.38(0.38)	0.62(0.62)

First we note that the model is fitting the probabilities of each outcome exactly. Note that the expected time in hospital, for a given level of the covariates, is the same for both outcome groups, since this is a proportional risks model. The expected times in hospital, given outcome, are not very close to those observed, but are in fact exactly equal to the average unconditional LOS. A plot of the Cox Snell residuals for this model further indicates that the fit is not very satisfactory. In general, low LOS are overpredicted by the model, while high LOS are underpredicted.

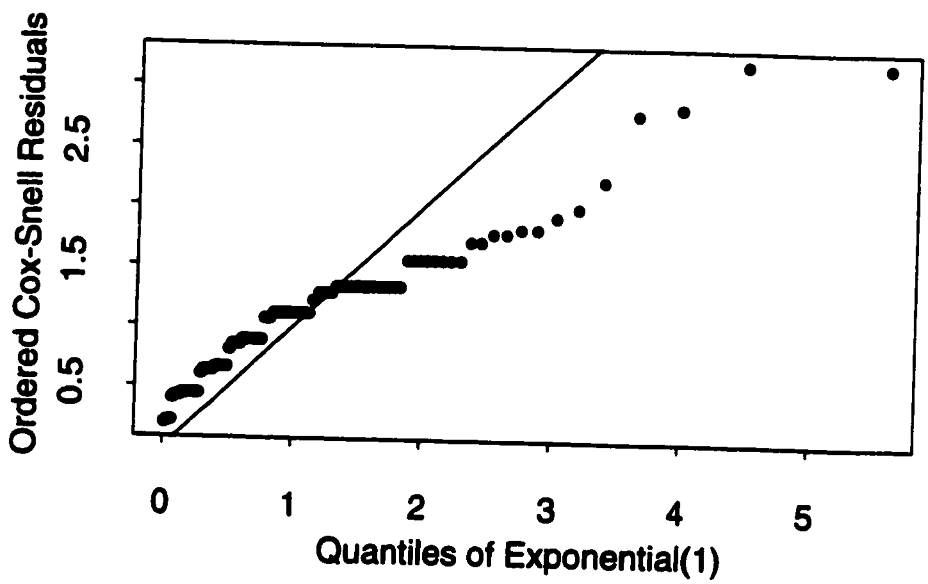


Figure 7.1: Exponential Competing Risks model

7.5 Example 2: Weibull competing risks model

Here we fit the Weibull competing risks model described above to the same data. The model predictions are shown in the following table.

Severity	n	T	$T \mid D = \mathrm{u}$	$T \mid D = \mathrm{l}$	$P(D = \mathrm{u})$	$P(D = \mathrm{l})$
low	82	4.49	4.61(4.59)	3.60(2.89)	0.88(0.88)	0.12(0.12)
med	39	4.67	5.96(5.96)	2.36(3.63)	0.64(0.61)	0.36(0.39)
high	13	4.92	8.80(6.85)	2.50(3.98)	0.38(0.37)	0.62(0.63)

Again, the predicted probabilities of each outcome are almost exactly equal to the observed proportion of patients in each outcome group. Length of stay given each outcome is well predicted by the model for low and medium severity patients who were discharged. The model is underestimating LOS both for high severity patients who are ultimately discharged and low severity patients who die in hospital by about 20%. On the other hand, it overestimates LOS for those who die in hospital in both the medium and high severity groups by more than 50%.

The Cox-Snell residuals are shown in figure 7.2. While observations with large LOS are underpredicted by this model, the overall fit is much improved over the exponential model. We now compare the Weibull competing risks model to the FP2B model fit in section 3.5. A comparison of the quantile plot of the Cox-Snell residuals for the two models indicates that the Weibull competing risks model gives a somewhat better fit. There is still substantial lack of fit for observations with high residuals, corresponding to those patients who remain in hospital an unusually long length of time, given the covariates.

We now compare the predicted times and probabilites of each outcome given by the two models. The Weibull competing risks model does a better job of predicting probability of outcome, especially in the medium severity group. The Weibull model also does a better job of predicting LOS, except for patients with high severity who were ultimately discharged. This group of patients are observed to have long lengths of stay on average, which are underpredicted by the Weibull model.

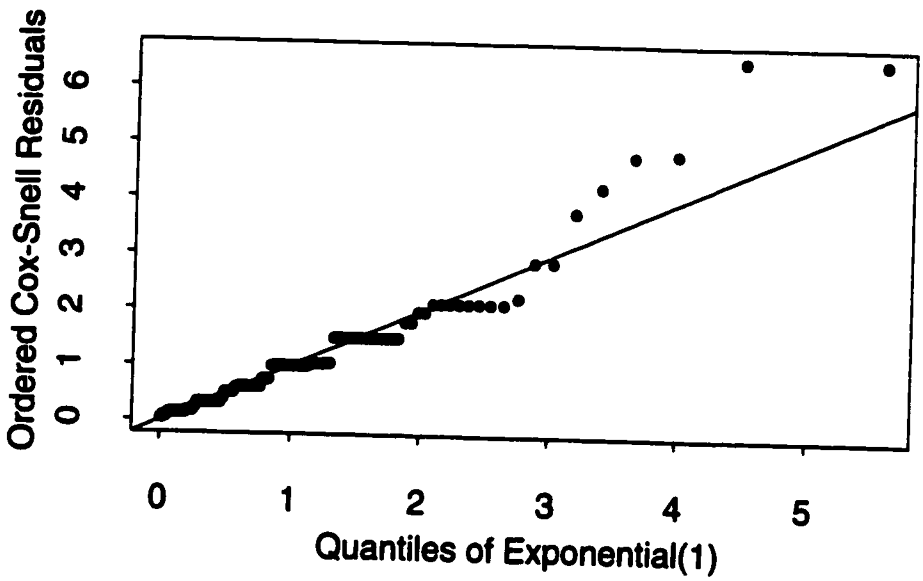


Figure 7.2: Weibull Competing Risks model

Chapter 8

Discussion and Further Work

8.1 Discussion

In this thesis, we have presented statistical models for time to breach of an upper or lower barrier in a latent Wiener process, which we call FP2B models. The motivating application for this work is length of stay in hospital, for patients with diseases and procedures and for agegroups in which a substantial number of deaths occur in hospital. There are many other possible applications, for instance

- time until graduate students graduate or withdraw,
- time until CD4 counts in AIDS patients reach an upper or lower threshold,
- time until cancer patients go into remission or die,
- in breast cancer patients: time from beginning of treatment to "progression" (cancer starts to grow again) or "shrinkage" (cancer shrinks by a specified amount, usually 50%).

Because cancer and AIDS patients are not monitored continuously, the latter three applications would involve interval censoring, which would be an interesting area for further research.

The parameters of the FP2B models are the barrier levels and drift. Since these are unobservable, we have given expressions for observable non-linear quantities that are of direct interest to the researcher, namely expected time in hospital, and the probability that the hospital stay ends in healthy discharge. In health services applications, where LOS is sometimes used as a surrogate for resource consumption, it is important to estimate the mean expected time, rather than the median, since total LOS can be reconstructed from the mean. The parameters of the model have been estimated using the method of maximum likelihood. We have shown that regularity conditions are satisfied, for the zero drift FP2B distribution with iid observations.

The subdensities of this model are infinite series, which we have shown are absolutely convergent and uniformly convergent with respect to the parameters and to the variable t over $(0, \infty)$. The computed likelihood is necessarily an approximation, in that only a finite number of terms of the series can be included. It would seem to be a good approximation, since the terms of the series drop off like $\exp(-k^2)$ as $k \to \infty$. However, small inaccuracies in the likelihood could lead to large inaccuracies in the estimated parameters, if the likelihood is flat. While more work is needed to investigate this issue, we have demonstrated that the approach is workable and that the maximum likelihood estimates obtained seem reasonable.

In administrative data files, LOS in hospital is recorded in days, and for the DRG's investigated in this study, LOS typically takes values in the set $1, 2, \ldots, 35$. Thus a discrete time model might provide a somewhat better fit to these data. But for other applications, in which time to the event in question is recorded in continuous time, a continuous model such as developed here, would be necessary.

A simple discrete time model would be easy to construct. For instance, we could model the health level process as a random walk on the integers, which takes jumps of size one in either the positive or negative direction at discrete time intervals of one day. Then the distribution of time to hitting a lower barrier is well known, and in fact is given in equation C.8 of the thesis. Expressions for the mean, conditional means, and probability of each outcome are given in equations 3.8, 3.10 and 3.7. More realistic but less tractable discrete time models would allow other distributions for jump size. Further investigation would be needed to determine whether the resulting estimates for LOS parameters would be very different from their corresponding estimates in the continuous framework.

We can think of the continuous FP2B model as an approximation to the discrete time model. Note that the subdensities for the simple discrete time model involve the binomial coefficients, which, for large values of LOS, would have to be approximated using for instance Stirling's approximation. Thus even this discrete time model would involve an approximation. Also, any analytical investigations would presumably have to invoke this approximation as well.

As discussed in section 1.2, the data investigated in this thesis do not include any right censored observations. However right censoring may be present in other applications. In this thesis, we use maximum likelihood to estimate the parameters of the FP2B models. Independently censored observations can easily be accommodated; individuals who are censored at time t contribute a factor of $\mathcal{F}(t)$ to the likelihood, where $\mathcal{F}$ is the survivor function of the FP2B distribution. All the software written to support this thesis can accommodate independently censored observations. Under mild conditions on the (independent) censoring (see [22]), maximum likelihood estimates will be consistent and asymptotically normal.

However, good estimates of the mean of a distribution in finite samples are notoriously difficult to obtain when a high proportion of observations are censored (see [21]). The same sorts of difficulties would apply to the estimate of P(D=u) since this functional also depends heavily on the tails of the cdf.

We have shown that the FP2B models have many similarities to competing risks models in which the joint distribution of time to failure and cause of failure is modelled. Indeed, the FP2B models can be thought of as special parametric models for two competing risks. While the most common competing risks analysis leaves the relationship between the two subdensities unspecified, the FP2B models specify the relationship between the subdensities, due to the fact that the same parameters appear in both subdensities. This may be a disadvantage in that the usual competing risks model is more flexible. However the FP2B model will presumably be more efficient, if the data do indeed arise from the kind of mechanism which motivates the FP2B distribution.

Note though that the usual competing risks analysis parameterizes each of the outcomes separately. If the data did come from the mechanism described in this thesis, that is with an underlying process hitting one of two barriers, then it would seem to be inappropriate to allow separate parameters for each outcome.

A non-statistical issue is that the idea of an underlying one-dimensional stochastic process with two barriers may provide a conceptual framework for certain research problems that allows researchers to think clearly about their hypotheses and how to test them. In the following discussion, we will use the idea of internal and external covariates as described in [22]. An internal covariate is defined to be a covariate which can only be measured at time t if the individual survives to time t. On the other hand, an external covariate is not directly involved with the underlying health level process.

It seems reasonable that internal variables might affect the health level process through the drift parameter, (which we might interpret as an individual's propensity to get well) while external variables might only affect the barrier levels. However, it may be difficult to classify a variable as truly external. For instance it seems that payer (Medicare/Medicaid/HMO) would be unlikely to affect an individual's health level process. However, the group of patients served by a particular payer are distinct. Since individuals insured by Medicaid tend to be disadvantaged, we may find that this disadvantage shows up as an effect on the drift parameter. External hospital-level covariates such as type (rural/urban, teaching/non-teaching), or size of hospital, might conceivably affect barrier levels, but again the appropriate model depends on the hypothesis under investigation. To test whether rural hospitals keep patients longer than necessary, it would be natural to place an indicator variable for location (rural/urban) on the upper barrier. If this variable were thought actually to affect quality of care, it might be placed on the drift parameter.

It seems natural that internal attributes of the individual patient, such as sex, age, and presence of co-morbidities, will directly affect the health status process and an individual's propensity to get well, and so should be put on the drift parameter. However, some internal attributes are important indicators of health level status at admission to hospital, for instance co-morbidities present at admission. As discussed in Chapter 3, differing health status levels at admission to hospital can be modelled by allowing both barrier levels to be raised (or lowered) by some estimated value. It is not possible to distinguish whether a change in apparent barrier level is due to a change in starting value, or an actual change in barrier level.

It should be noted that patients with like covariates all share the same drift, regardless of outcome (death or discharge). If we wish to model a differential effect of an internal covariate on outcome, then it would be necessary to put the covariate on the barriers as well. However, this raises the question of why such a differential effect exists. For instance, homelessness is an internal attribute which would presumably affect propensity to get well, as well as starting level. In addition, a differential effect on outcome could be present because hospitals were holding homeless patients longer out of compassion, or because drift varies across individuals in a way not explained by the model.

We have developed two extensions in Chapter 6. The first extension, a parametric mixture model, recognizes the presence of heterogeneity in the data that remains even after all relevant covariates have been included in the model. We have shown that if we allow the drift parameter to have a normal mixing distribution, the likelihood is quite tractable in that it does not involve numerical integration.

The second extension, a transfer model, gives a way of dealing with a third outcome, transfer to another institution. We postulate the existence of an intermediate barrier, and model probability of transfer given breach of this barrier as a function of covariates. Note that while independently censored observations are easily accommodated by the earlier models, transfers cannot be treated as independent censoring, since the decision to transfer at least partially depends on the health level of the individual. In the example given in the thesis, we modelled transfer to another acute care institution. We assumed that the transfer barrier level was greater than the starting position of the health level process, but less than the discharge barrier. This was felt to be appropriate since myocardial infarction is a catastrophic event, and presumably patients are not transferred until their condition stabilizes somewhat. In other applications, it may be more appropriate to have the transfer barrier below the starting position, for instance for individuals transferred to a nursing home.

8.2 Further Work

A methodology for dealing with time-varying covariates should be developed, so that information on procedures and diagnoses that develop while the patient is in hospital can be included in the modelling process. In other applications, we might want to incorporate information from a covariate or marker process. One approach would be to extend the work of Whitmore, Crowder, and Lawless [44]. They use a bivariate Wiener process to model the joint distribution of a marker process and an unobservable degradation process. In this approach, it is assumed that a unit fails when the degradation process crosses a single barrier. Then time until failure will have an inverse Gaussian distribution. They find that information from the marker process can improve predictions and increase the efficiency of estimation of the degradation process parameters when censoring is present. This increase can be substantial if the correlation between the two processes is high.

In this thesis, the underlying health level process is assumed to be a Wiener process, with independent increments, but it is possible that another process with dependent increments might be more realistic, for instance the Ornstein-Uhlenbeck process.

The mixture model developed in Chapter 6 recognizes unstructured heterogeneity in the population, but it is important to develop random effects models which accommodate the known hierarchical structure of the population, since patients are clustered within hospital. A methodology for incorporating sampling weights would be useful for some studies. Formal techniques of model assessment need to be addressed. Finally, the models could be extended to deal with recurrent events, which arise in health services research as multiple admissions.

Appendix A

Absolute convergence of F(t)

In this appendix we show that the series $F_u(t)$ is absolutely convergent for finite $t \geq 0$, where

$$F_u(t) = 2\sum_{k=0}^{\infty}\left\{\Phi\left(\frac{s_k+\ell}{\sqrt{t}}\right) - \Phi\left(\frac{s_k-\ell}{\sqrt{t}}\right)\right\}$$

with $s_k = -(2k+1)(u+l)$, and $\ell, u > 0$.

Let t be non-negative and finite. Consider the series $\sum_{k=0}^{\infty} a_k$ with $a_k = \Phi\left(\frac{s_k+\ell}{\sqrt{t}}\right)$. For $-(s_k+\ell)/\sqrt{t} \geq 2$, ie for $k \geq (\sqrt{t}+\ell/2)/(u+l) - 1/2$, we have

$$a_k = \Phi\left(\frac{s_k+\ell}{\sqrt{t}}\right) = \frac{1}{\sqrt{2\pi}}\int_{-\frac{s_k+\ell}{\sqrt{t}}}^{\infty} e^{-u^2/2}du$$

$$< \frac{1}{\sqrt{2\pi}}\int_{-\frac{s_k+\ell}{\sqrt{t}}}^{\infty} e^{-u}du = \frac{1}{\sqrt{2\pi}}\exp\left(\frac{s_k+\ell}{\sqrt{t}}\right) = b_k.$$

If $\sum b_k$ converges, then $\sum a_k$ converges by the comparison test.

We now use the integral test to show that $\sum b_k$ converges. We can write b_k as $\alpha\exp(-k\theta)$ for positive constants $\alpha = (2\pi)^{-1/2}\exp(-u/\sqrt{t})$ and $\theta = 2(u+l)/\sqrt{t}$. Now $\alpha\exp(-x\theta)$ is a positive, continuous function of x, that is non-increasing for $x \geq 0$ and

$$\int_0^{\infty} \alpha e^{-x\theta}dx = \frac{\alpha}{\theta}$$

which is finite for finite t. Therefore $\sum b_k$ converges by the integral test. It follows that $\sum a_k = \sum \Phi(\frac{s_k+\ell}{\sqrt{t}})$ converges by the comparison test.

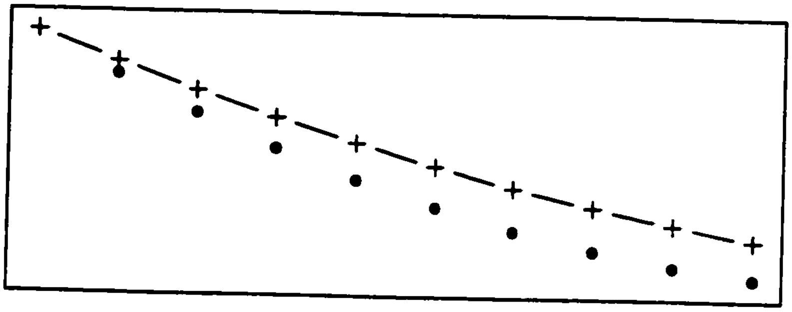


Figure A.1: The sequence a_k is represented by dots, the sequence b_k by plus signs, and the function $\alpha e^{-x\theta}$ by a smooth line.

Similarly, one can show that $\sum_{k=0}^{\infty}\Phi(\frac{s_k-\ell}{\sqrt{t}})$, $\sum_{k=0}^{\infty}\Phi(\frac{s_k+u}{\sqrt{t}})$, and $\sum_{k=0}^{\infty}\Phi(\frac{s_k-u}{\sqrt{t}})$ all converge. Then

$$F_u(t) = 2\sum_{k=0}^{\infty}\left\{\Phi\left(\frac{s_k+\ell}{\sqrt{t}}\right) - \Phi\left(\frac{s_k-\ell}{\sqrt{t}}\right)\right\}$$

converges absolutely, as does

$$F_\ell(t) = 2\sum_{k=0}^{\infty}\left\{\Phi\left(\frac{s_k+u}{\sqrt{t}}\right) - \Phi\left(\frac{s_k-u}{\sqrt{t}}\right)\right\}.$$

Appendix B

Derivation of distribution, Zero Drift

A Wiener process with zero drift is a continuous time, continuous state-space stochastic process $\{H(t); t \geq 0\}$ with continuous sample paths. Here the index t represents time. The increments $H(t) - H(s)$ exhibit the properties of stationarity, independence and normality. More specifically,

1. the distribution of $H(t) - H(s)$ depends on $t - s$, (and not t or s),
2. $H(t_j) - H(s_j)$ is independent of $H(t_k) - H(s_k)$ whenever the intervals (s_j, t_j) and (s_k, t_k) are disjoint, and
3. $H(t) - H(s)$ has a normal distribution with mean 0 and variance $\sigma^2(t - s)$.

The parameter σ^2 is referred to as the volatility or variance parameter. Property 1 is sometimes called the property of "temporal homogeneity". Additionally, the zero-drift Wiener process is spatially homogeneous, since the distribution of $H(t) - H(s)$ does not depend on $H(s)$.

In this appendix, we consider a Wiener process $\{H(t); t \geq 0\}$ which starts at 0, (ie $H(0) = 0$), so that $H(t) \sim N(0, \sigma^2 t)$. We will derive an expression for $F_u(T)$, the probability that the process first hits an upper barrier at $u > 0$ at some time before time T, without having hit a lower barrier at $-\ell < 0$. We will do this by constructing the set of paths of the process that reach u before time T, without first hitting $-\ell$. We will then find the probability associated with all paths in the set.

Start with the set of all paths that reach level u before time T. Now this will include some paths that reached $-\ell$ before T and then went on to u. These paths must be removed from the set. But in doing so, we will have removed paths that go to u, then $-\ell$ then u, which should be part of the set, so we must add these paths back in. Then again, paths that go to $-\ell$, u, $-\ell$, u must be removed, etc. Writing P[A] for the probability of event A, we have

$$\begin{aligned} F_u(T) &= \text{P[all paths that reach } u \text{ before time } T \text{ without hitting } -\ell] \\ &= \text{P[all paths that reach } u \text{ before time } T] \\ &- \text{P[paths that go to } -\ell\text{, then } u \text{ before } T] \\ &+ \text{P[paths that go to } u, -\ell, u \text{ before } T] \\ &- \text{P[paths that go to } -\ell, u, -\ell, u \text{ before } T] + \ldots. \end{aligned} \tag{B.1}$$

To find each of these probabilities, we use a reflection argument. For instance, the first probability on the right hand side of equation (B.1) is the probability of all paths that hit u at some time T_1 where $0 < T_1 < T$. But the event $[H(T_1) = u]$ is the union of the two disjoint events $[H(T_1) = u, H(T) > u]$ and $[H(T_1) = u, H(T) < u]$. Furthermore there is a 1:1 correspondence between paths with $H(T_1) = u, H(T) > u$ and those with $H(T_1) = u, H(T) < u$. This is shown

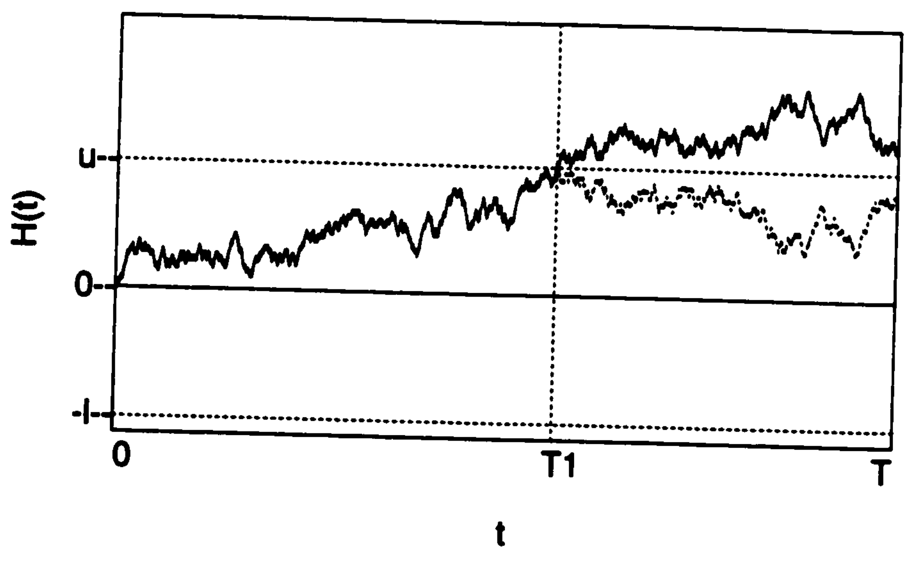


Figure B.1: Sample path reflected about $H(t) = u$

in figure B.1 where the path which ends above u at time T (depicted by the solid line) is reflected about the line $H(t) = u$, to give a corresponding path which ends below u at time T (represented by the dotted line). Because of the properties of the Wiener process and the symmetry of the normal distribution, these events have equal probability. Therefore,

$$\begin{aligned} P[H(T_1) = u] &= 2P[H(T_1) = u,\ H(T) > u] \\ &= 2P[H(T) > u] \\ &= 2\Phi\left(\frac{-u}{\sigma\sqrt{T}}\right) \end{aligned}$$

where $\Phi(x)$ is the cdf of the standard normal distribution evaluated at x. The second equality is true because all paths that reach a position greater than u must have passed through u at a previous time.

The second term of equation (B.1) is only slightly more complicated. We need the probability

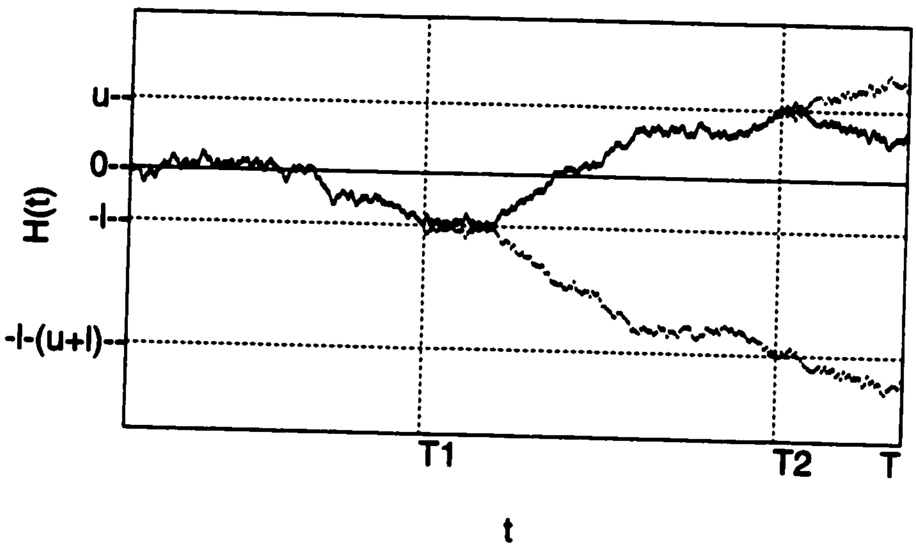


Figure B.2: Sample path reflected about $H(t) = u$ and $H(t) = -\ell$

of all paths that hit $-\ell$ at some time T_1, then u at some time T_2, where $0 < T_1 < T_2 < t$. As shown in figure B.2, among paths with $H(T_1) = -\ell$ and $H(T_2) = u$ there is a 1:1 correspondence between those that end below u at time T, and those that end above u at time T. Thus

$$P[H(T_1) = -\ell, H(T_2) = u] = 2P[H(T_1) = -\ell, H(T_2) = u, H(T) > u]$$

This type of path is difficult to deal with, so we reflect it about the line $H = -\ell$ to give a path with $H(T_1) = -\ell$, $H(T_2) = -\ell - (u+\ell)$, $H(T) < -\ell - (u+\ell)$. The probability associated with paths of this type is just

$$P[H(T) < -\ell - (u+\ell)] = 2\Phi\left(\frac{-\ell - (u+\ell)}{\sigma\sqrt{T}}\right)$$

since all paths that reach $-\ell - (u+\ell)$ must have previously passed through the levels $H = -\ell$ and $H = -\ell - (u+\ell)$. Thus P[paths that go to ℓ then u before time T] $= 2\Phi\left(\frac{-\ell-(u+\ell)}{\sigma\sqrt{T}}\right)$.

The same sorts of arguments show that the probability of all paths that go to u then $-\ell$ a total of k times, and then to u, all before time T, is $2\Phi\left(\frac{-u-2k(u+\ell)}{\sigma\sqrt{T}}\right)$ for $k = 0, 1, 2, 3, \ldots$. Recall that these types of paths are added to the set in equation (B.1). Similarly, the probability of all paths that go to $-\ell$ then u a total of $k+1$ times before T is $2\Phi\left(\frac{-\ell-(2k+1)(u+\ell)}{\sigma\sqrt{T}}\right)$ for $k = 0, 1, 2, 3, \ldots$. Recall that paths like this must get removed from the set, so we subtract these probabilities.

In total then, the probability that the process reaches the upper barrier u for the first time before time T and does so without going through $-\ell$ is

$$F_u(T) = 2\sum_{k=0}^{\infty}\left\{\Phi\left(\frac{-u-2k(u+\ell)}{\sigma\sqrt{T}}\right) - \Phi\left(\frac{-\ell-(2k+1)(u+\ell)}{\sigma\sqrt{T}}\right)\right\}.$$

The lower sub-distribution function, $F_\ell(t)$, can be derived in an analogous manner. However, we can also argue that, due to the symmetry of the process, the time to reach a barrier at -ℓ is the same as the time to reach a barrier at ℓ. Thus we can get $F_\ell(t)$ by reversing the roles of u and ℓ in $F_u(t)$. More precisely, if we denote $F_u(t)$ as the function $g(t,u,\ell)$, then$F_\ell(t) = g(t,\ell,u)$. Thus

$$F_\ell(T) = 2\sum_{k=0}^{\infty}\left\{\Phi\left(\frac{-\ell-2k(u+\ell)}{\sigma\sqrt{T}}\right) - \Phi\left(\frac{-u-(2k+1)(u+\ell)}{\sigma\sqrt{T}}\right)\right\}.$$

Let $F(T)$ denote the probability that the process breaches one of the two barriers before time T. Then since $F(T) = F_u(T) + F_\ell(T)$, we get

$$F(T) = 2\sum_{k=0}^{\infty}\left\{\Phi\left(\frac{s_k+\ell}{\sigma\sqrt{T}}\right) - \Phi\left(\frac{s_k-\ell}{\sigma\sqrt{T}}\right) + \Phi\left(\frac{s_k+u}{\sigma\sqrt{T}}\right) - \Phi\left(\frac{s_k-u}{\sigma\sqrt{T}}\right)\right\}$$

where $s_k = -(2k+1)(u+\ell)$. Note that in the body of the thesis, we take $\sigma^2 = 1$.

Theorem B.1 *Let $T_{a,b,x}$ denote the time that a Wiener process, $\{H(t), t \geq 0\}$, starting at $H(0) = x$, first reaches either a barrier at a or b, where $a < x < b$. Then $T_{a,b,0}$ has the same distribution as $T_{a+c,b+c,c}$*

Proof: If we retrace our steps through the proceeding derivation of the upper sub-distribution function of $T_{u,-\ell,0}$, it is clear that the theorem follows from the spatial homogeneity of the Wiener process. Let $H_c(t)$ be a Wiener process that starts at c. Then

$$H_c(t) - H_c(0) = H_c(t) - c \sim N(0,t).$$

To find the upper sub-distribution function of $T_{u+c,-\ell+c,c}$ we need the probability of all paths that reach $u+c$ before time t without first hitting $-\ell+c$. As before we can decompose this into an alternating series of probabilities.

$$\begin{aligned} F_u(T) &= \text{P[all paths that reach } u+c \text{ before time } T \text{ without hitting } -\ell+c] \\ &= \text{P[all paths that reach } u+c \text{ before time } T] \\ &- \text{P[paths that go to } -\ell+c, \text{ then } u+c \text{ before } T] \\ &+ \text{P[paths that go to } u+c, -\ell+c, u+c \text{ before } T] + \ldots. \end{aligned}$$

The first addend in the series equals

$$2P(H_c(t) > u+c) = 2P(H_c(t) - c > u) = 2\Phi\left(\frac{-u}{\sqrt{t}}\right)$$

exactly the value of the first addend in expression B.1. All the other terms in the new series have the same probability as their corresponding terms in B.1. Thus the upper sub-distribution function of $T_{u,-\ell,0}$ is identical to the upper sub-distribution function of $T_{u+c,-\ell+c,c}$. Similar arguments apply to the lower sub-distribution function and the distribution function itself. **Q.E.D.**

B.1 Stable Laws

As before, consider a zero-drift Wiener process with volatility σ^2 which starts at 0. Now let the lower barrier at $-\ell \to -\infty$ so that only a single barrier at u remains, and let T_u be the first passage time to this barrier. Then $P(T_u < t)$ is given by the first term in expression B.1 so

$$P(T_u < t) = 2\Phi\left(-\frac{u}{\sigma\sqrt{t}}\right).$$

Differentiating gives the density

$$f(t; u, \sigma) = \frac{u}{\sqrt{2\pi t^3}} \exp\left(-\frac{u^2}{2\sigma^2 t}\right) = t^{-3/2} u \, \phi\left(\frac{u}{\sigma\sqrt{t}}\right).$$

This distribution is a stable law with index 1/2 (see below). Note that if we let $u \to \infty$, leaving a barrier at $-\ell < 0$, the first passage time $-\ell$ has the same density given above, with u replaced by ℓ. Thus for the first passage time to a barrier at $a \neq 0$, the density is given by

$$f(t; a, \sigma) = \frac{|a|}{\sqrt{2\pi t^3}} \exp\left(-\frac{a^2}{2\sigma^2 t}\right) = t^{-3/2} |a| \, \phi\left(\frac{a}{\sigma\sqrt{t}}\right).$$

The stable laws are characterized as follows. $X, X_1, X_2, \ldots X_n$ have a stable distribution if $\sum_{i=1}^{n} X_i$ has the same distribution as $a_n X + b_n$ where $a_n > 0$ and b_n are arbitrary constants. Clearly the normal distribution is stable. The limiting distribution of sums of iid random variables must be a stable law.

The characteristic function of a random variable X with stable distribution can be written as ([14])

$$E[e^{iXt}] = \begin{cases} \exp\left(-|ct|^{\alpha} \exp(-i \, \frac{\pi}{2} \, \beta \, \mathrm{sgn}(t)) + i\delta t\right) & \alpha \neq 1 \\ \exp\left(-|ct| - i(2\beta/\pi)ct \, \log|ct| + i\delta t\right) & \alpha = 1 \end{cases}$$

where $-\infty < t < \infty, 0 < \alpha \leq 2$, $|\beta| \leq \min(\alpha, 2-\alpha)$, $c > 0$, $-\infty < \delta < \infty$. The parameter α is referred to as the index or characteristic exponent. The parameter β is a skewness parameter, with $\beta = 0$ corresponding to a symmetric distribution. c and δ are respectively scale and location parameters.

The stable laws include the normal family of distributions ($\alpha = 2$) and the Cauchy distributions

($\alpha = 1$), as well as the first passage time distributions discussed above ($\alpha = 1/2$). These are the only three cases in which the density has a closed form. However, an expression for the density in terms of an infinite series is known for all stable laws. Assuming the random variable X has been standardized to have $c = 1,\ \delta = 0$, the density of X is given by ([18], p. 549)

$$f(x;\alpha,\beta) = \begin{cases} \frac{1}{\pi x}\sum_{j=1}^{\infty}\frac{\Gamma(1+j\alpha)}{j!}(-x^{-\alpha})^j \sin\left[\frac{j\pi}{2}(\beta-\alpha)\right] & x > 0, 0 < \alpha < 1 \\ \frac{1}{\pi x}\sum_{j=1}^{\infty}\frac{\Gamma(1+j\alpha^{-1})}{j!}(-x)^j \sin\left[\frac{j\pi}{2\alpha}(\beta-\alpha)\right] & x > 0, 1 < \alpha < 2. \end{cases}$$

To find the density for $x < 0$, note that $f(-x;\alpha,\beta) = f(x;\alpha,-\beta)$.

If a stable law has characteristic exponent α, then it has finite absolute moments of order γ where $0 < \gamma < \alpha$. All absolute moments of order $\geq \alpha$ are infinite ([20], p. 182). Of the stable laws, only the normal has finite variance. Both the Cauchy and the stable law with $\alpha = 1/2$ have both infinite variance and infinite expectation.

Appendix C

Derivation of Distribution, Non-Zero Drift

C.1 Solution to Difference Equation

In this appendix we give the derivation of the subdensities for the drift case. Following Feller ([17]), we start with a random walk, and construct and solve a difference equation for the probability of breach of one of the barriers at epoch n. As the number of steps per unit time gets large, while the size of the steps gets small, the random walk becomes, in the limit, a Wiener process. Taking the limit of the probability of breach at time n in the same way, gives the subdensity we seek.

Consider independent random variables $Z_1, Z_2, \ldots$, where $P(Z_i = 1) = p$ and $P(Z_i = -1) = 1 - p$. Let $X_0 = 0$, and $X_n = X_{n-1} + Z_n$. We can visualize the stochastic process $\{X_n\}_{n=0}^{\infty}$ as the path of a particle which starts at position $z > 0$ at time 0 and performs a random walk on the integers. At each unit of time the particle takes either an upward step of unit length with probability p, or a downward step of unit length with probability q, where $p + q = 1$. Suppose that there are absorbing barriers at positions $a > z$ and 0.

A more colorful representation of this situation is that of a gambler, who starts with an initial sum of money z dollars, playing against an adversary whose initial capital is $a - z$ dollars. The gambler wins a hand with probability p, in which case he takes a dollar from the adversary. He loses with probability q, in which case he must give a dollar to his adversary. The game continues until one of the players is ruined, that is until one player's capital is reduced to 0. In this case

X_n represents the gambler's fortune at the n^{th} hand of the game.

Let $u_{z,n}$ be the probability that the walk, starting from position z, breaches the lower barrier at the n^{th} step, (without first reaching the upper barrier). We can set up a system of difference equations to describe the situation:

$$u_{z,n+1} = p\ u_{z+1,n} + q\ u_{z-1,n} \qquad z = 1, 2, \ldots, a-1 \qquad n = 0, 1, 2, \ldots \tag{C.1}$$

with boundary conditions

$$\begin{aligned} u_{0,0} &= 1 & & \\ u_{a,n} &= 0 & n &= 0, 1, 2, \ldots \\ u_{0,n} &= 0 & n &= 1, 2, 3, \ldots \\ u_{z,0} &= 0 & z &= 1, 2, \ldots, a. \end{aligned} \tag{C.2}$$

This system can be solved using the method of generating functions. The generating functions for $u_{z,n}$ is

$$U_z(s) = \sum_{n=0}^{\infty} u_{z,n} s^n$$

Transforming the sequences in equation C.1 to their respective generating functions gives

$$\frac{U_z(s) - u_{z,0}}{s} = pU_{z+1}(s) + qU_{z-1}(s)$$

and since $u_{z,0} = 0$ for $z = 1, 2, \ldots, a$

$$U_z(s) = psU_{z+1}(s) + qsU_{z-1}(s) \tag{C.3}$$

This is itself a second-order homogeneous difference equation in the variable $U_z(s)$. The initial conditions are given by

$$U_0(s) = 1 \qquad U_a(s) = 0 \tag{C.4}$$

since

$$U_0(s) = u_{0,0} + u_{0,1}s + u_{0,2}s^2 + \ldots + u_{0,n}s^n + \ldots$$

and

$$U_a(s) = u_{a,0} + u_{a,1}s + u_{a,2}s^2 + \ldots + u_{a,n}s^n + \ldots$$

and, as specified in C.2, $u_{0,0} = 1$, but all other coefficients in the two power series equal zero. General difference equation theory suggests that we try solutions of the form $U_z(s) = \lambda^z(s)$, which upon substitution into C.3 gives

$$\lambda^z = ps\lambda^{z+1} + qs\lambda^{z-1}$$

Dividing through by λ^{z-1} and rearranging, we get

$$\lambda^2 - \frac{1}{ps}\lambda + \frac{q}{p} = 0$$

This equation has roots

$$\lambda_1(s) = \frac{1 + \sqrt{1 - 4pqs^2}}{2ps}$$

and

$$\lambda_2(s) = \frac{1 - \sqrt{1 - 4pqs^2}}{2ps}$$

The general solution will then be of the form

$$U_z(s) = A(s)\lambda_1^z(s) + B(s)\lambda_2^z(s) \tag{C.5}$$

where $A(s)$ and $B(s)$ are determined by the boundary conditions C.4. We find that

$$A(s) = \frac{-\lambda_2^a(s)}{\lambda_1^a(s) - \lambda_2^a(s)}$$

$$B(s) = \frac{\lambda_1^a(s)}{\lambda_1^a(s) - \lambda_2^a(s)}$$

so that

$$U_z(s) = \frac{\lambda_1^a(s)\lambda_2^z(s) - \lambda_1^z(s)\lambda_2^a(s)}{\lambda_1^a(s) - \lambda_2^a(s)}$$

or equivalently, because $\lambda_1(s)\lambda_2(s) = q/p$,

$$U_z(s) = \left(\frac{q}{p}\right)^z \frac{\lambda_1^{a-z}(s) - \lambda_2^{a-z}(s)}{\lambda_1^a(s) - \lambda_2^a(s)} \tag{C.6}$$

A little algebra allows us to rewrite C.6 as a function independent of λ_1

$$U_z(s) = \frac{\lambda_2^z}{1 - \lambda_2^{2a}\left(\frac{p}{q}\right)^a} - \frac{\left(\frac{q}{p}\right)^{z-a} \lambda_2^{2a-z}}{1 - \lambda_2^{2a}\left(\frac{p}{q}\right)^a}$$

Then expanding the denominators as geometric series, ie

$$\frac{1}{1 - \lambda_2^{2a}\left(\frac{p}{q}\right)^a} = \sum_{k=0}^{\infty}\left[\lambda_2^{2a}\left(\frac{p}{q}\right)^a\right]^k$$

allows us to write

$$U_z(s) = \sum_{k=0}^{\infty}\left(\frac{p}{q}\right)^{ka} \lambda_2^{2ka+z} - \sum_{k=1}^{\infty}\left(\frac{p}{q}\right)^{ka-z} \lambda_2^{2ka-z} \tag{C.7}$$

We still need to derive an explicit expression for λ_2. Consider another random walk, starting at $z > 0$ as before, but now with only a single absorbing barrier at 0. It turns out that $\lambda_2^z(s)$ is the generating function of $w_{z,n}$, the probability that the walk ends at the barrier at epoch n, as we will now show.

The difference equation for the single barrier situation has the same form as before, namely

$$w_{z,n+1} = p\ w_{z+1,n} + q\ w_{z-1,n} \qquad z = 1, 2, 3, \ldots \qquad n = 0, 1, 2, \ldots$$

but with different boundary conditions

$$\begin{aligned} w_{0,0} &= 1 \\ w_{0,n} &= 0 \qquad n = 1, 2, 3, \ldots \\ w_{z,0} &= 0 \qquad z = 1, 2, 3, \ldots \end{aligned}$$

Transforming the difference equation in terms of generating functions $W_z(s) = \sum_{n=0}^{\infty} w_{z,n}s^n$ gives

$$W_z(s) = psW_{z+1}(s) + qsW_{z-1}(s)$$

with boundary condition $W_0(s) = 1$. The same particular solutions $\lambda_1(s)$ and $\lambda_2(s)$ still apply but the boundary condition now dictates that $A(s) = 0$ and $B(s) = 1$ so that the general solution of equation C.5 is given by

$$W_z(s) = \sum_{n=0}^{\infty} w_{z,n} s^n = \lambda_2^z(s)$$

Now in fact, $w_{z,n}$ can be found by simple combinatorial arguments (eg [17]) as

$$w_{z,n} = \frac{z}{n} \binom{n}{\frac{(n+z)}{2}}^* p^{(n-z)/2} q^{(n+z)/2}$$

where the symbol $()^*$ denotes the usual combinatorial symbol if n and z are of the same parity, and zero otherwise. Substituting the expression for λ_2^z into the generating function C.7 gives

$$U_z(s) = \sum_{k=0}^{\infty} \left[\left(\frac{p}{q}\right)^{ka} \sum_{n=0}^{\infty} w_{2ka+z,n} s^n \right] - \sum_{k=1}^{\infty} \left[\left(\frac{p}{q}\right)^{ka-z} \sum_{n=0}^{\infty} w_{2ka-z,n} s^n \right]$$

$$= \sum_{n=0}^{\infty} s^n \left[\sum_{k=0}^{\infty} \left(\frac{p}{q}\right)^{ka} w_{2ka+z,n} - \sum_{k=1}^{\infty} \left(\frac{p}{q}\right)^{ka-z} w_{2ka-z,n} \right]$$

The n^{th} coefficient of the series gives the quantity we seek, namely $u_{z,n}$, the probability that a walk starting at z reaches the bottom barrier at epoch n:

$$u_{z,n} = \sum_{k=0}^{\infty} \left(\frac{p}{q}\right)^{ka} w_{2ka+z,n} - \sum_{k=1}^{\infty} \left(\frac{p}{q}\right)^{ka-z} w_{2ka-z,n} \tag{C.8}$$

C.2 Limiting Distribution

In fact, $w_{z,n} = \frac{z}{n} P(X_n = \frac{n+z}{2})$ where $X_n \sim Bin(n, q)$. Then since $X_n \xrightarrow{D} Y$ where $Y \sim N(nq, \sqrt{nqp})$, we have that, as $n \to \infty$,

$$\begin{aligned} w_{z,n} &\sim \frac{z}{n} \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{(\frac{n+z}{2} - nq)^2}{2npq}\right) \\ &= \frac{z}{n} \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{(z + n(p-q))^2}{8npq}\right). \end{aligned}$$

We began this section with a random walk that takes a single step of length one at each time epoch. Consider what happens as the steps become smaller and smaller in size, and at the same time the number of steps per unit time interval gets larger and larger. If in the limit we let δ, the size of the steps go to zero, r, the number of steps per time epoch go to infinity, and p, the probability of an upward step go to 1/2 in such a way that

$$\begin{aligned}(p-q)\delta r &\rightarrow \mu \\ 4pq\delta^2 r &\rightarrow \sigma^2\end{aligned}$$

where μ and σ^2 are constants, we get a Wiener process with drift μ and volatility σ^2.

We want to find the probability that the limiting process, starting from position $\zeta > 0$, reaches the lower barrier at 0, in a small interval around time t, without breaching the upper barrier at $\alpha > \zeta$. We must adjust the starting position, z, and upper barrier a, and the total number of steps n in the discrete walk so that

$$z \sim \zeta/\delta \qquad a \sim \alpha/\delta \qquad n \sim tr$$

Also, we take the limits so that

$$2p \sim 1 + \frac{\mu\delta}{\sigma^2} \qquad 2q \sim 1 - \frac{\mu\delta}{\sigma^2}$$

and

$$\left(\frac{p}{q}\right)^{\frac{\nu}{\delta}} \sim \left(1 - \frac{2\mu/\sigma^2}{1/\delta}\right)^{-\frac{\nu}{\delta}} \rightarrow e^{\frac{2\mu\nu}{\sigma^2}}$$

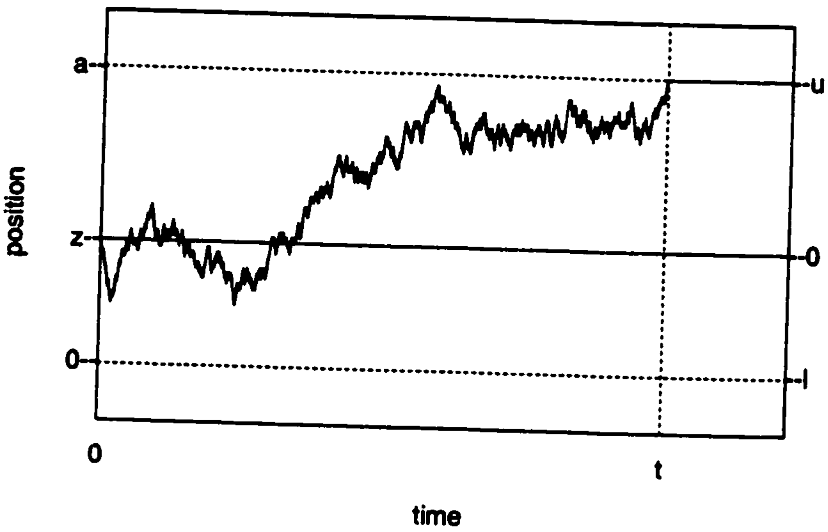
Taking the limit of expression C.8 in the way described gives

$$\frac{2}{r}\frac{1}{\sqrt{2\pi\sigma^2 t^3}}e^{-\frac{(\mu t+2\zeta)\mu}{2\sigma^2}}\sum_{k=0}^{\infty}(\zeta+2k\alpha)e^{-\frac{(\zeta+2k\alpha)^2}{2\sigma^2 t}} - \sum_{k=1}^{\infty}(2k\alpha-\zeta)e^{-\frac{(\zeta-2k\alpha)^2}{2\sigma^2 t}} \tag{C.9}$$

This is the probability that the process ends at the lower barrier in an interval of length $2/r$ around t. The corresponding subdensity is

$$\frac{1}{\sqrt{2\pi\sigma^2 t^3}}e^{-\frac{(\mu t+2\zeta)\mu}{2\sigma^2}}\sum_{k=-\infty}^{\infty}(\zeta+2k\alpha)e^{-\frac{(\zeta+2k\alpha)^2}{2\sigma^2 t}} \tag{C.10}$$

Now we need to change co-ordinate systems in order to describe the situation where the process starts at 0, and the absorbing barriers are at $u > 0$ and $-\ell < 0$. As shown in figure C.2, a distance



of ζ in the old system corresponds to a distance of ℓ in the new system, while a distance of a in the old system corresponds to a distance of $u + \ell$ in the new system. Making these substitutions in C.10 gives

$$-\frac{1}{\sqrt{2\pi\sigma^2 t^3}} e^{-\frac{(\mu t + 2\ell)\mu}{2\sigma^2}} \sum_{k=-\infty}^{\infty} (-\ell - 2k(u+\ell)) e^{-\frac{(-\ell-2k(u+\ell))^2}{2\sigma^2 t}}. \tag{C.11}$$

Setting $\sigma^2 = 1$ and using notation established in the main body of the thesis gives $f_\ell^\mu(t)$

$$-t^{-\frac{3}{2}} e^{\frac{-(\mu t + 2\ell)\mu}{2}} \sum_{k=-\infty}^{\infty} (s_k + u)\, \phi\left(\frac{s_k + u}{\sqrt{t}}\right)$$

The other subdensity, $f_u^\mu(t)$, is derived in a similar fashion. Again we start with the discrete case, and think of a gambler with initial capital z. In this case we want the probability of the event that the gambler wins at epoch n. But this is just the event that his adversary, with initial capital $a - z$, loses at epoch n. Thus we need to replace z in the above argument with $a - z$. Also the adversary wins each hand with probability q, so the roles of p and q in the above argument must be reversed. Eventually we get that $f_u^\mu(t)$ is

$$-t^{-\frac{3}{2}}e^{\frac{-(\mu t-2u)\mu}{2}}\sum_{k=-\infty}^{\infty}\left\{(s_k+\ell)\phi\left(\frac{s_k+\ell}{\sqrt{t}}\right)\right\}$$

Appendix D

Integral Bounds

We use the following lemma to establish integral bounds on infinite series.

Lemma D.0.1 *Let $g(x)$ be a non-negative function that decreases monotonically on $[a, \infty)$. Let $\{g_k\}$ be the sequence formed from $g(x)$ at integer values of x. Then*

$$\sum_{k=a}^{\infty} g_k \leq \int_a^{\infty} g(x)\, dx + g(a)$$

Proof:

$$\sum_{k=a+1}^{\infty} g_k \leq \int_a^{\infty} g(x)\, dx$$

so that

$$\sum_{k=a}^{\infty} g_k \leq \int_a^{\infty} g(x)\, dx + g(a).$$

Q.E.D.

Appendix E

Derivatives

This appendix records the derivatives for all the models discussed in the thesis and outlines the incorporation of covariates. To simplify notation, we write f_{ui} for $f_u(t_i)$, $f_{\ell i}$ for $f_\ell(t_i)$ and $\mathcal{F}_i$ for $\mathcal{F}(t_i)$. Where limits are not indicated, the sums over k extend from $-\infty$ to ∞.

E.1 Zero Drift Model

The log likelihood for the zero-drift model without covariates is

$$\mathcal{L}(u,\ell;t_1,t_2,\ldots,t_n) = \sum_{i=1}^{n} \mathcal{L}_i = \sum_{\{i:D_i=\mathrm{u}\}} \mathcal{L}_{ui} + \sum_{\{i:D_i=\mathrm{l}\}} \mathcal{L}_{\ell i} + \sum_{\{i:D_i=\mathrm{c}\}} \mathcal{L}_{ci} \tag{E.1}$$

where D_i is u, l or c according to whether individual i was discharged, died or was censored. Here $\mathcal{L}_i$ is the contribution to the likelihood from individual i, and

$$\mathcal{L}_i = \begin{cases} \mathcal{L}_{ui} = \log(f_{ui}) & \text{if } D_i = \mathrm{u}, \\ \mathcal{L}_{\ell i} = \log(f_{\ell i}) & \text{if } D_i = \mathrm{l}, \\ \mathcal{L}_{ci} = \log(\mathcal{F}_i) & \text{if } D_i = \mathrm{c} \end{cases} \tag{E.2}$$

where

$$f_{ui} = t_i^{-\frac{3}{2}} \sum_k \phi\left(\frac{c_k}{\sqrt{t_i}}\right) c_k$$

$$f_{\ell i} = t_i^{-\frac{3}{2}} \sum_k \phi\left(\frac{e_k}{\sqrt{t_i}}\right) e_k$$

$$\mathcal{F}_i = \sum_{k=0}^{\infty}\left\{\Phi\left(\frac{s_k+\ell}{\sqrt{t}}\right) - \Phi\left(\frac{s_k-\ell}{\sqrt{t}}\right) + \Phi\left(\frac{s_k+u}{\sqrt{t}}\right) - \Phi\left(\frac{s_k-u}{\sqrt{t}}\right)\right\}. \quad \text{(E.3)}$$

and $c_k = -(s_k+\ell)$ and $e_k = -(s_k+u)$. As before, $s_k = -(2k+1)(u+\ell)$, so that

$$c_k = 2k(u+\ell)+u \qquad e_k = 2k(u+\ell)+\ell. \quad \text{(E.4)}$$

For this model, u and ℓ are the parameters to be estimated.

In what follows, we will need the derivatives of the subdensities, so we record them now. Write $\theta = (\theta_1,\theta_2)' = (u,\ell)'$ for the vector of parameters to be estimated. Then, for $j = 1, 2$, we have

$$\frac{\partial f_{ui}}{\partial\theta_j} = t_i^{-\frac{3}{2}}\sum_k \frac{\partial c_k}{\partial\theta_j}\,\phi\left(\frac{c_k}{\sqrt{t_i}}\right)\left[1-\frac{c_k^2}{t_i}\right]$$

$$\frac{\partial f_{\ell i}}{\partial\theta_j} = t_i^{-\frac{3}{2}}\sum_k \frac{\partial e_k}{\partial\theta_j}\,\phi\left(\frac{e_k}{\sqrt{t_i}}\right)\left[1-\frac{e_k^2}{t_i}\right]$$

where

$$\frac{\partial c_k}{\partial\theta_1} = \frac{\partial c_k}{\partial u} = 2k+1 \qquad \frac{\partial e_k}{\partial\theta_1} = \frac{\partial e_k}{\partial u} = 2k$$

$$\frac{\partial c_k}{\partial\theta_2} = \frac{\partial c_k}{\partial \ell} = 2k \qquad \frac{\partial e_k}{\partial\theta_2} = \frac{\partial e_k}{\partial \ell} = 2k+1.$$

The gradient, $G(u,\ell)$, ie the vector of first derivatives of $\mathcal{L}(u,\ell;t_1,t_2,\ldots,t_n)$, is a (2 x 1) vector, $\left(\frac{\partial\mathcal{L}}{\partial u}\;\;\frac{\partial\mathcal{L}}{\partial\ell}\right)'$. We can write

$$G(u,\ell) = \sum_{i=1}^{n} G_i = \sum_{\{i:D_i=u\}} G_{ui} + \sum_{\{i:D_i=l\}} G_{\ell i}$$

where G_i is the contribution from individual i and

$$G_i = \begin{cases} G_{ui} & \text{if } D_i = \text{u}, \\ G_{\ell i} & \text{if } D_i = \text{l} \end{cases}$$

where

$$G_{ui} = \begin{pmatrix} \frac{\partial\mathcal{L}_{ui}}{\partial u} \\ \frac{\partial\mathcal{L}_{ui}}{\partial\ell} \end{pmatrix} \qquad G_{\ell i} = \begin{pmatrix} \frac{\partial\mathcal{L}_{\ell i}}{\partial u} \\ \frac{\partial\mathcal{L}_{\ell i}}{\partial\ell}. \end{pmatrix}$$

For $j = 1, 2$,

$$\frac{\partial \mathcal{L}_{ui}}{\partial \theta_j} = \frac{1}{f_{ui}} \frac{\partial f_{ui}}{\partial \theta_j} = \frac{\sum_k \frac{\partial c_k}{\partial \theta_j} \phi\left(\frac{c_k}{\sqrt{t_i}}\right)\left[1 - \frac{c_k^2}{t_i}\right]}{\sum_k \phi\left(\frac{c_k}{\sqrt{t_i}}\right) c_k}$$

$$\frac{\partial \mathcal{L}_{\ell i}}{\partial \theta_j} = \frac{1}{f_{\ell i}} \frac{\partial f_{\ell i}}{\partial \theta_j} = \frac{\sum_k \frac{\partial e_k}{\partial \theta_j} \phi\left(\frac{e_k}{\sqrt{t_i}}\right)\left[1 - \frac{e_k^2}{t_i}\right]}{\sum_k \phi\left(\frac{e_k}{\sqrt{t_i}}\right) e_k}. \tag{E.5}$$

The Hessian, $H(u, \ell)$, ie the matrix of second derivatives of $\mathcal{L}(u, \ell; t_1, t_2, \ldots, t_n)$, is a (2 x 2) matrix. The element in the j^{th} row and j'^{th} column of H is $\frac{\partial^2 \mathcal{L}}{\partial \theta_j \partial \theta_{j'}}$. We can write

$$H(u, \ell) = \sum_{i=1}^{n} H_i = \sum_{\{i:D_i=\mathrm{u}\}} H_{ui} + \sum_{\{i:D_i=\mathrm{l}\}} H_{\ell i}$$

where H_i is the contribution from individual i and

$$H_i = \begin{cases} H_{ui} & \text{if } D_i = \mathrm{u}, \\ H_{\ell i} & \text{if } D_i = \mathrm{l} \end{cases}$$

where

$$H_{ui} = \begin{pmatrix} \frac{\partial^2 \mathcal{L}_{ui}}{\partial u^2} & \frac{\partial^2 \mathcal{L}_{ui}}{\partial u \partial \ell} \\ \frac{\partial^2 \mathcal{L}_{ui}}{\partial u \partial \ell} & \frac{\partial^2 \mathcal{L}_{ui}}{\partial \ell^2} \end{pmatrix} \qquad H_{\ell i} = \begin{pmatrix} \frac{\partial^2 \mathcal{L}_{ui}}{\partial u^2} & \frac{\partial^2 \mathcal{L}_{ui}}{\partial u \partial \ell} \\ \frac{\partial^2 \mathcal{L}_{ui}}{\partial u \partial \ell} & \frac{\partial^2 \mathcal{L}_{ui}}{\partial \ell^2} \end{pmatrix}$$

Then for $j = 1, 2$,

$$\frac{\partial^2 \mathcal{L}_{ui}}{\partial \theta_j \partial \theta_{j'}} = \left[\frac{1}{f_{ui}} \frac{\partial^2 f_{ui}}{\partial \theta_j \partial \theta_{j'}} - \frac{1}{f_{ui}^2} \frac{\partial f_{ui}}{\partial \theta_j} \frac{\partial f_{ui}}{\partial \theta_{j'}}\right]$$

$$\frac{\partial^2 \mathcal{L}_{\ell i}}{\partial \theta_j \partial \theta_{j'}} = \left[\frac{1}{f_{\ell i}} \frac{\partial^2 f_{\ell i}}{\partial \theta_j \partial \theta_{j'}} - \frac{1}{f_{\ell i}^2} \frac{\partial f_{\ell i}}{\partial \theta_j} \frac{\partial f_{\ell i}}{\partial \theta_{j'}}\right], \tag{E.6}$$

where the second derivatives of the subdensities are

$$\frac{\partial^2 f_{ui}}{\partial \theta_j \partial \theta_{j'}} = t_i^{-\frac{3}{2}} \sum_k \frac{\partial c_k}{\partial \theta_j} \frac{\partial c_k}{\partial \theta_{j'}} \phi\left(\frac{c_k}{\sqrt{t_i}}\right) \frac{c_k}{t_i} \left[\frac{c_k^2}{t_i} - 3\right]$$

$$\frac{\partial^2 f_{\ell i}}{\partial \theta_j \partial \theta_{j'}} = t_i^{-\frac{3}{2}} \sum_k \frac{\partial e_k}{\partial \theta_j} \frac{\partial e_k}{\partial \theta_{j'}} \phi\left(\frac{e_k}{\sqrt{t_i}}\right) \frac{e_k}{t_i} \left[\frac{e_k^2}{t_i} - 3\right].$$

The third derivatives of the subdensities are needed to prove that regularity conditions are satisfied, so they are recorded here.

$$\frac{\partial^3 f_{ui}}{\partial\theta_j\partial\theta_{j'}\partial\theta_{j''}} = t_i^{-\frac{5}{2}} \sum_k \frac{\partial c_k}{\partial\theta_j}\frac{\partial c_k}{\partial\theta_{j'}}\frac{\partial c_k}{\partial\theta_{j''}} \phi\left(\frac{c_k}{\sqrt{t_i}}\right)\left[\frac{6c_k^2}{t_i} - \frac{c_k^4}{t_i^2} - 3\right]$$

$$\frac{\partial^3 f_{\ell i}}{\partial\theta_j\partial\theta_{j'}\partial\theta_{j''}} = t_i^{-\frac{5}{2}} \sum_k \frac{\partial e_k}{\partial\theta_j}\frac{\partial e_k}{\partial\theta_{j'}}\frac{\partial e_k}{\partial\theta_{j''}} \phi\left(\frac{e_k}{\sqrt{t_i}}\right)\left[\frac{6e_k^2}{t_i} - \frac{e_k^4}{t_i^2} - 3\right]$$

Extension of this model to allow the barrier levels u and ℓ to depend on covariates is straightforward. We may allow each barrier to depend on a different set of covariates. Let

$$\underline{xu}_i = (xu_{i0}, xu_{i1}, \ldots, xu_{ip_u})'$$

be the $((p_u + 1) \times 1)$ vector of covariates measured on individual i which are thought to affect the upper barrier, and

$$\underline{x\ell}_i = (x\ell_{i0}, x\ell_{i1}, \ldots, x\ell_{ip_\ell})'$$

be the $((p_\ell + 1) \times 1)$ vector of covariates measured on individual i which are thought to affect the lower barrier. Since now the barriers may vary by individual, write u^i, l^i for the barrier levels for individual i. The simplest conceivable model allows each barrier level to be a linear function of the relevant covariates, ie

$$u^i = \underline{u}' * \underline{xu}_i = u_0\, xu_{i0} + u_1\, xu_{i1} + \ldots + u_p\, xu_{ip_u}$$

$$\ell^i = \underline{\ell}' * \underline{x\ell}_i = \ell_0\, x\ell_{i0} + \ell_1\, x\ell_{i1} + \ldots + \ell_p x\ell_{ip_\ell} \qquad \text{(E.7)}$$

where $\underline{u} = (u_0, u_1, \ldots, u_{p_u})'$ and $\underline{\ell} = (\ell_0, \ell_1, \ldots, \ell_{p_\ell})'$, are the parameters of the model, and $*$ denotes matrix multiplication.

Denote the log likelihood for the model with covariates as $\mathcal{L}^*(u, \ell)$. This will have the same form as $\mathcal{L}(u, \ell)$ in equations E.1, E.2, E.3, and E.4, but with u and ℓ expanded as in equation E.7.

The gradient for this model

$$G^*(u, \ell) = \sum_{i=1}^{n} G_i^* = \sum_{\{i:D_i=u\}} G_{ui}^* + \sum_{\{i:D_i=l\}} G_{\ell i}^*$$

is now a vector of length $q = (p_u + p_\ell + 2)$ where G^*_{ui} and $G^*_{\ell i}$ have j^{th} elements

$$\frac{\partial \mathcal{L}^*_{ui}}{\partial \lambda_j} \quad \text{and} \quad \frac{\partial \mathcal{L}^*_{\ell i}}{\partial \lambda_j}$$

respectively, where $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_q) = (\underline{u} \,|\, \underline{\ell}) = (u_0, u_1, \ldots, u_{p_u}, \ell_0, \ell_1, \ldots, \ell_{p_\ell})$.

Note that we can write G^*_{ui} and $G^*_{\ell i}$ succinctly as

$$G^*_{ui} = \begin{pmatrix} \frac{\partial \mathcal{L}_{ui}}{\partial u} \otimes \underline{xu}_i \\ \hline \frac{\partial \mathcal{L}_{ui}}{\partial \ell} \otimes \underline{x\ell}_i \end{pmatrix} \qquad G^*_{\ell i} = \begin{pmatrix} \frac{\partial \mathcal{L}_{\ell i}}{\partial u} \otimes \underline{xu}_i \\ \hline \frac{\partial \mathcal{L}_{\ell i}}{\partial \ell} \otimes \underline{x\ell}_i \end{pmatrix}$$

where the partials are as specified in equation E.5, and $\otimes$ denotes Kronecker product.

The Hessian for the model with covariates,

$$H^*(u, \ell) = \sum_{i=1}^{n} H^*_i = \sum_{\{i:D_i=u\}} H^*_{ui} + \sum_{\{i:D_i=l\}} H^*_{\ell i}$$

is now a q x q matrix with $(j, j')^{th}$ element $\frac{\partial^2 \mathcal{L}}{\partial \lambda_j \partial \lambda_{j'}}$. This can be written succinctly as

$$H^*_{ui} = \left(\begin{array}{c|c} \frac{\partial^2 \mathcal{L}_{ui}}{\partial u^2} \otimes (\underline{xu}_i * \underline{xu}'_i) & \frac{\partial^2 \mathcal{L}_{ui}}{\partial u \partial \ell} \otimes (\underline{xu}_i * \underline{x\ell}'_i) \\ \hline \frac{\partial^2 \mathcal{L}_{ui}}{\partial u \partial \ell} \otimes (\underline{xu}_i * \underline{x\ell}'_i) & \frac{\partial^2 \mathcal{L}_{ui}}{\partial \ell^2} \otimes (\underline{x\ell}_i * \underline{x\ell}'_i) \end{array}\right)$$

and

$$H^*_{\ell i} = \left(\begin{array}{c|c} \frac{\partial^2 \mathcal{L}_{\ell i}}{\partial u^2} \otimes (\underline{xu}_i * \underline{xu}'_i) & \frac{\partial^2 \mathcal{L}_{\ell i}}{\partial u \partial \ell} \otimes (\underline{xu}_i * \underline{x\ell}'_i) \\ \hline \frac{\partial^2 \mathcal{L}_{\ell i}}{\partial u \partial \ell} \otimes (\underline{xu}_i * \underline{x\ell}'_i) & \frac{\partial^2 \mathcal{L}_{\ell i}}{\partial \ell^2} \otimes (\underline{x\ell}_i * \underline{x\ell}'_i) \end{array}\right)$$

where the partials are as written in equation E.6.

E.2 Drift Model

The log likelihood for the drift model without covariates is

$$\mathcal{L}^\mu(u, \ell, \mu; t_1, t_2, \ldots, t_n) = \sum_{i=1}^{n} \mathcal{L}^\mu_i = \sum_{\{i:D_i=u\}} \mathcal{L}^\mu_{ui} + \sum_{\{i:D_i=l\}} \mathcal{L}^\mu_{\ell i} + \sum_{\{i:D_i=c\}} \mathcal{L}^\mu_{ci} \tag{E.8}$$

where D_i = u if individual i was discharged from hospital, D_i = l if individual i was observed to die in hospital. D_i = c if individual c was censored. Here

$$\mathcal{L}_i^\mu = \begin{cases} \mathcal{L}_{ui}^\mu = \log(f_{ui}^\mu) & \text{if } D_i = \text{u}, \\ \mathcal{L}_{\ell i}^\mu = \log(f_{\ell i}^\mu) & \text{if } D_i = \text{l}, \\ \mathcal{L}_{ci}^\mu = \log(\mathcal{F}_i^\mu) & \text{if } D_i = \text{c} \end{cases} \tag{E.9}$$

and

$$f_{ui}^\mu = t_i^{-\frac{3}{2}} e^{\frac{-(\mu t_i - 2u)\mu}{2}} \sum_k \phi\left(\frac{c_k}{\sqrt{t_i}}\right) c_k$$

$$f_{\ell i}^\mu = t_i^{-\frac{3}{2}} e^{\frac{-(\mu t_i + 2\ell)\mu}{2}} \sum_k \phi\left(\frac{e_k}{\sqrt{t_i}}\right) e_k \tag{E.10}$$

$$\mathcal{F}_i^\mu = e^{-c_k \mu} \left\{ \Phi\left(\frac{c_k + u - \mu t_i}{\sqrt{t_i}}\right) - \Phi\left(\frac{c_k - l - \mu t_i}{\sqrt{t_i}}\right) \right\} - \tag{E.11}$$

$$e^{d_k \mu} \left\{ \Phi\left(\frac{-d_k + u - \mu t_i}{\sqrt{t_i}}\right) - \Phi\left(\frac{-d_k - l - \mu t_i}{\sqrt{t_i}}\right) \right\} \tag{E.12}$$

For this model, we have to estimate the parameters u, ℓ, and μ.

Let

$$y(u, t_i) = t_i^{-\frac{3}{2}} e^{\frac{-(\mu t_i - 2u)\mu}{2}}$$

$$y(\ell, t_i) = t_i^{-\frac{3}{2}} e^{\frac{-(\mu t_i + 2\ell)\mu}{2}}.$$

Then the partial derivatives of the sub-densities can be expressed as

$$\frac{\partial f_{ui}^\mu}{\partial u} = y(u, t_i) \sum_k \phi\left(\frac{c_k}{\sqrt{t_i}}\right) \left\{ \mu c_k + \frac{\partial c_k}{\partial u} \left[1 - \frac{c_k^2}{t_i}\right] \right\}$$

$$\frac{\partial f_{ui}^\mu}{\partial \ell} = y(u, t_i) \sum_k \frac{\partial c_k}{\partial \ell} \phi\left(\frac{c_k}{\sqrt{t_i}}\right) \left[1 - \frac{c_k^2}{t_i}\right]$$

$$\frac{\partial f_{ui}^\mu}{\partial \mu} = y(u, t_i)(u - \mu t_i) \sum_k \phi\left(\frac{c_k}{\sqrt{t_i}}\right) c_k$$

$$\frac{\partial f_{\ell i}^\mu}{\partial u} = y(\ell, t_i) \sum_k \frac{\partial e_k}{\partial u} \phi\left(\frac{e_k}{\sqrt{t_i}}\right) \left[1 - \frac{e_k^2}{t_i}\right]$$

$$\frac{\partial f_{\ell i}^\mu}{\partial \ell} = y(\ell, t_i) \sum_k \phi\left(\frac{e_k}{\sqrt{t_i}}\right) \left\{ -\mu e_k + \frac{\partial e_k}{\partial \ell} \left[1 - \frac{e_k^2}{t_i}\right] \right\}$$

$$\frac{\partial f^{\mu}_{\ell i}}{\partial \mu} = y(\ell, t_i)(-l - \mu t_i) \sum_k \phi\left(\frac{e_k}{\sqrt{t_i}}\right) e_k.$$

The gradient, $G^{\mu}(u, \ell)$, ie the vector of first derivatives of $\mathcal{L}^{\mu}(u, \ell, \mu; t_1, t_2, \ldots, t_n)$, is a (3 x 1) vector, $\left(\frac{\partial \mathcal{L}^{\mu}}{\partial u} \; \frac{\partial \mathcal{L}^{\mu}}{\partial \ell} \; \frac{\partial \mathcal{L}^{\mu}}{\partial \mu}\right)'$. We can write

$$G^{\mu}(u, \ell, \mu) = \sum_{i=1}^{n} G^{\mu}_i = \sum_{\{i:D_i=\mathrm{u}\}} G^{\mu}_{ui} + \sum_{\{i:D_i=\mathrm{l}\}} G^{\mu}_{\ell i}$$

where

$$G^{\mu}_i = \begin{cases} G^{\mu}_{ui} & \text{if } D_i = \mathrm{u}, \\ G^{\mu}_{\ell i} & \text{if } D_i = \mathrm{l} \end{cases}$$

and

$$G^{\mu}_{ui} = \begin{pmatrix} \frac{\partial \mathcal{L}^{\mu}_{ui}}{\partial u} \\ \frac{\partial \mathcal{L}^{\mu}_{ui}}{\partial \ell} \\ \frac{\partial \mathcal{L}^{\mu}_{ui}}{\partial \mu} \end{pmatrix} = \begin{pmatrix} \frac{1}{f^{\mu}_{ui}} \frac{\partial f^{\mu}_{ui}}{\partial u} \\ \frac{1}{f^{\mu}_{ui}} \frac{\partial f^{\mu}_{ui}}{\partial \ell} \\ \frac{1}{f^{\mu}_{ui}} \frac{\partial f^{\mu}_{ui}}{\partial \mu} \end{pmatrix} = \begin{pmatrix} \frac{\sum_k \phi\left(\frac{e_k}{\sqrt{t_i}}\right)\left\{\mu e_k + \frac{\partial e_k}{\partial u}\left[1 - \frac{e_k^2}{t_i}\right]\right\}}{\sum_k \phi\left(\frac{e_k}{\sqrt{t_i}}\right) e_k} \\ \frac{\sum_k \frac{\partial e_k}{\partial \ell} \phi\left(\frac{e_k}{\sqrt{t_i}}\right)\left[1 - \frac{e_k^2}{t_i}\right]}{\sum_k \phi\left(\frac{e_k}{\sqrt{t_i}}\right) e_k} \\ (u - \mu t_i) \end{pmatrix}$$

$$G^{\mu}_{\ell i} = \begin{pmatrix} \frac{\partial \mathcal{L}^{\mu}_{\ell i}}{\partial u} \\ \frac{\partial \mathcal{L}^{\mu}_{\ell i}}{\partial \ell} \\ \frac{\partial \mathcal{L}^{\mu}_{\ell i}}{\partial \mu} \end{pmatrix} = \begin{pmatrix} \frac{1}{f^{\mu}_{\ell i}} \frac{\partial f^{\mu}_{\ell i}}{\partial u} \\ \frac{1}{f^{\mu}_{\ell i}} \frac{\partial f^{\mu}_{\ell i}}{\partial \ell} \\ \frac{1}{f^{\mu}_{\ell i}} \frac{\partial f^{\mu}_{\ell i}}{\partial \mu} \end{pmatrix} = \begin{pmatrix} \frac{\sum_k \frac{\partial e_k}{\partial u} \phi\left(\frac{e_k}{\sqrt{t_i}}\right)\left[1 - \frac{e_k^2}{t_i}\right]}{\sum_k \phi\left(\frac{e_k}{\sqrt{t_i}}\right) e_k} \\ \frac{\sum_k \phi\left(\frac{e_k}{\sqrt{t_i}}\right)\left\{-\mu e_k + \frac{\partial e_k}{\partial \ell}\left[1 - \frac{e_k^2}{t_i}\right]\right\}}{\sum_k \phi\left(\frac{e_k}{\sqrt{t_i}}\right) e_k} \\ (-\ell - \mu t_i) \end{pmatrix} \tag{E.13}$$

The Hessian, $H^{\mu}(u, \ell, \mu)$, is now a (3 x 3) matrix. Let $\theta = (\theta_1, \theta_2, \theta_3) = (u, \ell, \mu)$. Then the element in the j^{th} row and j'^{th} column of H^{μ} is $\frac{\partial^2 \mathcal{L}^{\mu}}{\partial \theta_j \partial \theta_{j'}}$. We can write

$$H^{\mu}(u, \ell, \mu) = \sum_{i=1}^{n} H^{\mu}_i = \sum_{\{i:D_i=\mathrm{u}\}} H^{\mu}_{ui} + \sum_{\{i:D_i=\mathrm{l}\}} H^{\mu}_{\ell i}$$

where

$$H_i^\mu = \begin{cases} H_{ui}^\mu & \text{if } D_i = \text{u}, \\ H_{\ell i}^\mu & \text{if } D_i = \text{l} \end{cases}$$

and

$$H_{ui}^\mu = \begin{pmatrix} \frac{\partial^2 \mathcal{L}_{ui}^\mu}{\partial u^2} & \frac{\partial^2 \mathcal{L}_{ui}^\mu}{\partial u \partial \ell} & \frac{\partial^2 \mathcal{L}_{ui}^\mu}{\partial u \partial \mu} \\ \frac{\partial^2 \mathcal{L}_{ui}^\mu}{\partial \ell \partial u} & \frac{\partial^2 \mathcal{L}_{ui}^\mu}{\partial \ell^2} & \frac{\partial^2 \mathcal{L}_{ui}^\mu}{\partial \ell \partial \mu} \\ \frac{\partial^2 \mathcal{L}_{ui}^\mu}{\partial \mu \partial u} & \frac{\partial^2 \mathcal{L}_{ui}^\mu}{\partial \mu \partial \ell} & \frac{\partial^2 \mathcal{L}_{ui}^\mu}{\partial \mu^2} \end{pmatrix} \qquad H_{\ell i} = \begin{pmatrix} \frac{\partial^2 \mathcal{L}_{\ell i}^\mu}{\partial u^2} & \frac{\partial^2 \mathcal{L}_{\ell i}^\mu}{\partial u \partial \ell} & \frac{\partial^2 \mathcal{L}_{\ell i}^\mu}{\partial u \partial \mu} \\ \frac{\partial^2 \mathcal{L}_{\ell i}^\mu}{\partial \ell \partial u} & \frac{\partial^2 \mathcal{L}_{\ell i}^\mu}{\partial \ell^2} & \frac{\partial^2 \mathcal{L}_{\ell i}^\mu}{\partial \ell \partial \mu} \\ \frac{\partial^2 \mathcal{L}_{\ell i}^\mu}{\partial \mu \partial u} & \frac{\partial^2 \mathcal{L}_{\ell i}^\mu}{\partial \mu \partial \ell} & \frac{\partial^2 \mathcal{L}_{\ell i}^\mu}{\partial \mu^2} \end{pmatrix}$$

where for $j = 1, 2, 3$,

$$\begin{aligned} \frac{\partial^2 \mathcal{L}_{ui}^\mu}{\partial \theta_j \partial \theta_{j'}} &= \left[\frac{1}{f_{ui}^\mu} \frac{\partial^2 f_{ui}^\mu}{\partial \theta_j \partial \theta_{j'}} - \left(\frac{1}{f_{ui}^\mu} \right)^2 \frac{\partial f_{ui}^\mu}{\partial \theta_j} \frac{\partial f_{ui}^\mu}{\partial \theta_{j'}} \right] \\ \frac{\partial^2 \mathcal{L}^\mu{}_{\ell i}}{\partial \theta_j \partial \theta_{j'}} &= \left[\frac{1}{f_{\ell i}^\mu} \frac{\partial^2 f_{\ell i}^\mu}{\partial \theta_j \partial \theta_{j'}} - \left(\frac{1}{f_{\ell i}^\mu} \right)^2 \frac{\partial f_{\ell i}^\mu}{\partial \theta_j} \frac{\partial f_{\ell i}^\mu}{\partial \theta_{j'}} \right]. \end{aligned} \tag{E.14}$$

The second derivatives of f_u^μ are

$$\begin{aligned} \frac{\partial^2 f_{ui}^\mu}{\partial u^2} &= y(u, t_i) \sum_k \phi\left(\frac{c_k}{\sqrt{t_i}}\right) \left\{ \mu^2 c_k + 2\mu \frac{\partial c_k}{\partial u}\left[1 - \frac{c_k^2}{t_i}\right] + \left(\frac{\partial c_k}{\partial u}\right)^2 \frac{c_k}{t_i}\left[\frac{c_k^2}{t_i} - 3\right] \right\} \\ \frac{\partial^2 f_{ui}^\mu}{\partial u \partial \ell} &= y(u, t_i) \sum_k \phi\left(\frac{c_k}{\sqrt{t_i}}\right) \frac{\partial c_k}{\partial \ell} \left\{ \frac{\partial c_k}{\partial u} \frac{c_k}{t_i}\left[\frac{c_k^2}{t_i} - 3\right] + \mu\left[1 - \frac{c_k^2}{t_i}\right] \right\} \\ \frac{\partial^2 f_{ui}^\mu}{\partial u \partial \mu} &= y(u, t_i) \sum_k \phi\left(\frac{c_k}{\sqrt{t_i}}\right) \left\{ (u - \mu t_i)\left(\mu c_k + \frac{\partial c_k}{\partial u}\left[1 - \frac{c_k^2}{t_i}\right]\right) + c_k \right\} \\ \frac{\partial^2 f_{ui}^\mu}{\partial \ell^2} &= y(u, t_i) \sum_k \phi\left(\frac{c_k}{\sqrt{t_i}}\right) \left(\frac{\partial c_k}{\partial \ell}\right)^2 \frac{c_k}{t_i}\left[\frac{c_k^2}{t_i} - 3\right] \\ \frac{\partial^2 f_{ui}^\mu}{\partial \ell \partial \mu} &= y(u, t_i)(u - \mu t_i) \sum_k \phi\left(\frac{c_k}{\sqrt{t_i}}\right) \frac{\partial c_k}{\partial \ell}\left[1 - \frac{c_k^2}{t_i}\right] \\ \frac{\partial^2 f_{ui}^\mu}{\partial \mu^2} &= y(u, t_i)\left[(u - \mu t_i)^2 - t_i\right] \sum_k \phi\left(\frac{c_k}{\sqrt{t_i}}\right) c_k. \end{aligned}$$

The second derivatives of f_ℓ^μ are

$$\begin{aligned} \frac{\partial^2 f_{\ell i}^\mu}{\partial u^2} &= y(\ell, t_i) \sum_k \phi\left(\frac{e_k}{\sqrt{t_i}}\right) \left(\frac{\partial e_k}{\partial u}\right)^2 \frac{e_k}{t_i}\left[\frac{e_k^2}{t_i} - 3\right] \\ \frac{\partial^2 f_{\ell i}^\mu}{\partial u \partial \ell} &= y(\ell, t_i) \sum_k \phi\left(\frac{e_k}{\sqrt{t_i}}\right) \frac{\partial e_k}{\partial u} \left\{ \frac{\partial e_k}{\partial \ell} \frac{e_k}{t_i}\left[\frac{e_k^2}{t_i} - 3\right] - \mu\left[1 - \frac{e_k^2}{t_i}\right] \right\} \end{aligned}$$

$$\frac{\partial^2 f_{\ell i}^{\mu}}{\partial u \partial \mu} = y(\ell, t_i)(-\ell - \mu t_i) \sum_k \phi\left(\frac{e_k}{\sqrt{t_i}}\right) \frac{\partial e_k}{\partial u}\left[1 - \frac{e_k^2}{t_i}\right]$$

$$\frac{\partial^2 f_{\ell i}^{\mu}}{\partial \ell^2} = y(\ell, t_i) \sum_k \phi\left(\frac{e_k}{\sqrt{t_i}}\right)\left\{\mu^2 e_k - 2\mu \frac{\partial e_k}{\partial \ell}\left[1 - \frac{e_k^2}{t_i}\right] + \left(\frac{\partial e_k}{\partial \ell}\right)^2 \frac{e_k}{t_i}\left[\frac{e_k^2}{t_i} - 3\right]\right\}$$

$$\frac{\partial^2 f_{\ell i}^{\mu}}{\partial \ell \partial \mu} = y(\ell, t_i) \sum_k \phi\left(\frac{e_k}{\sqrt{t_i}}\right)\left\{(\ell + \mu t_i)\left(\mu e_k - \frac{\partial e_k}{\partial \ell}\left[1 - \frac{e_k^2}{t_i}\right]\right) - e_k\right\}$$

$$\frac{\partial^2 f_{\ell i}^{\mu}}{\partial \mu^2} = y(\ell, t_i)\left[(\ell + \mu t_i)^2 - t_i\right] \sum_k \phi\left(\frac{e_k}{\sqrt{t_i}}\right) e_k.$$

Extending the model to include covariates is again straightforward. As before, let

$$\underline{xu}_i = (xu_{i0}, xu_{i1}, \ldots, xu_{ip_u})'$$

$$\underline{x\ell}_i = (x\ell_{i0}, x\ell_{i1}, \ldots, x\ell_{ip_\ell})'$$

be the vector of covariates measured on individual i which are thought to affect the upper and lower barriers respectively, and now let

$$\underline{x\mu}_i = (x\mu_{i0}, x\mu_{i1}, \ldots, x\mu_{ip_\mu})'$$

be the $((p_\mu + 1) \times 1)$ vector of covariates measured on individual i which are thought to affect the drift, μ. The simplest conceivable model allows each of u, ℓ and μ to be a linear function of the relevant covariates, ie

$$u^i = \underline{u}' * \underline{xu}_i = u_0\, xu_{i0} + u_1\, xu_{i1} + \ldots + u_{p_u}\, xu_{ip_u}$$

$$\ell^i = \underline{\ell}' * \underline{x\ell}_i = \ell_0\, x\ell_{i0} + \ell_1\, x\ell_{i1} + \ldots + \ell_{p_\ell} x\ell_{ip_\ell}$$

$$\mu^i = \underline{\mu}' * \underline{x\mu}_i = \mu_0\, x\mu_{i0} + \mu_1\, x\mu_{i1} + \ldots + \mu_{ip_\mu} x\mu_{ip_\mu} \tag{E.15}$$

where $\underline{u} = (u_0, u_1, \ldots, u_{p_u})'$, $\underline{\ell} = (\ell_0, \ell_1, \ldots, \ell_{p_\ell})'$ and $\underline{\mu} = (\mu_0, \mu_1, \ldots, \mu_{p_\mu})'$ are the parameters of the model.

Denote the log likelihood for the drift model with covariates as $\mathcal{L}^{\mu *}(u, \ell)$. This will have the same form as $\mathcal{L}^{\mu}(u, \ell)$ in equations E.8, E.9, and E.11, but with u, ℓ, and μ expanded as in equation E.15.

The gradient for the model with covariates

$$G^{\mu*}(u,\ell,\mu)=\sum_{i=1}^{n}G_i^{\mu*}=\sum_{\{i:D_i=u\}}G_{ui}^{\mu*}+\sum_{\{i:D_i=l\}}G_{\ell i}^{\mu*}$$

is now a vector of length $q=(p_u+p_\ell+p_\mu+3)$ where $G_{ui}^{\mu*}$ and $G_{\ell i}^{\mu*}$ have j^{th} elements

$$\frac{\partial\mathcal{L}_{ui}^{\mu*}}{\partial\lambda_j}\quad\text{and}\quad\frac{\partial\mathcal{L}_{\ell i}^{\mu*}}{\partial\lambda_j}$$

respectively, where

$$\lambda=(\lambda_1,\lambda_2,\ldots,\lambda_q)'=(\underline{u}'\,|\,\underline{\ell}')'=(u_0,u_1,\ldots,u_{p_u},\ell_0,\ell_1,\ldots,\ell_{p_\ell},\mu_0,\mu_1,\ldots,\mu_{p_\mu})'.$$

Note that we can write $G_{ui}^{\mu*}$ and $G_{\ell i}^{\mu*}$ succinctly as

$$G_{ui}^{\mu*}=\begin{pmatrix}\frac{\partial\mathcal{L}_{ui}^{\mu}}{\partial u}\otimes\underline{xu}_i\\ \hline \frac{\partial\mathcal{L}_{ui}^{\mu}}{\partial\ell}\otimes\underline{x\ell}_i\\ \hline \frac{\partial\mathcal{L}_{ui}^{\mu}}{\partial\mu}\otimes\underline{x\mu}_i\end{pmatrix}\qquad G_{\ell i}^{\mu*}=\begin{pmatrix}\frac{\partial\mathcal{L}_{\ell i}^{\mu}}{\partial u}\otimes\underline{xu}_i\\ \hline \frac{\partial\mathcal{L}_{\ell i}^{\mu}}{\partial\ell}\otimes\underline{x\ell}_i\\ \hline \frac{\partial\mathcal{L}_{\ell i}^{\mu}}{\partial\mu}\otimes\underline{x\mu}_i\end{pmatrix}$$

where the partials are as specified in equation E.13.

The Hessian for this model

$$H^{\mu*}(u,\ell,\mu)=\sum_{i=1}^{n}H_i^{\mu*}=\sum_{\{i:D_i=u\}}H_{ui}^{\mu*}+\sum_{\{i:D_i=l\}}H_{\ell i}^{\mu*}$$

is now a q x q matrix with $(j,j')^{th}$ element $\frac{\partial^2\mathcal{L}^{\mu*}}{\partial\lambda_j\partial\lambda_{j'}}$. This can be written succinctly as

$$H_{ui}^{\mu*}=\left(\begin{array}{c|c|c}\frac{\partial^2\mathcal{L}_{ui}^{\mu}}{\partial u^2}\otimes(\underline{xu}_i*\underline{xu}_i') & \frac{\partial^2\mathcal{L}_{ui}^{\mu}}{\partial u\partial\ell}\otimes(\underline{xu}_i*\underline{x\ell}_i') & \frac{\partial^2\mathcal{L}_{ui}^{\mu}}{\partial u\partial\mu}\otimes(\underline{xu}_i*\underline{x\mu}_i')\\ \hline \frac{\partial^2\mathcal{L}_{ui}^{\mu}}{\partial\ell\partial u}\otimes(\underline{x\ell}_i*\underline{xu}_i') & \frac{\partial^2\mathcal{L}_{ui}^{\mu}}{\partial\ell^2}\otimes(\underline{x\ell}_i*\underline{x\ell}_i') & \frac{\partial^2\mathcal{L}_{ui}^{\mu}}{\partial\ell\partial\mu}\otimes(\underline{x\mu}_i*\underline{x\ell}_i')\\ \hline \frac{\partial^2\mathcal{L}_{ui}^{\mu}}{\partial\mu\partial u}\otimes(\underline{x\mu}_i*\underline{xu}_i') & \frac{\partial^2\mathcal{L}_{ui}^{\mu}}{\partial\mu\partial\ell}\otimes(\underline{x\mu}_i*\underline{x\ell}_i') & \frac{\partial^2\mathcal{L}_{ui}^{\mu}}{\partial\mu^2}\otimes(\underline{x\mu}_i*\underline{x\mu}_i')\end{array}\right)$$

and

$$H_{\ell i}^{\mu*}=\left(\begin{array}{c|c|c}\frac{\partial^2\mathcal{L}_{\ell i}^{\mu}}{\partial u^2}\otimes(\underline{xu}_i*\underline{xu}_i') & \frac{\partial^2\mathcal{L}_{\ell i}^{\mu}}{\partial u\partial\ell}\otimes(\underline{xu}_i*\underline{x\ell}_i') & \frac{\partial^2\mathcal{L}_{\ell i}^{\mu}}{\partial u\partial\mu}\otimes(\underline{xu}_i*\underline{x\mu}_i')\\ \hline \frac{\partial^2\mathcal{L}_{\ell i}^{\mu}}{\partial\ell\partial u}\otimes(\underline{x\ell}_i*\underline{xu}_i') & \frac{\partial^2\mathcal{L}_{\ell i}^{\mu}}{\partial\ell^2}\otimes(\underline{x\ell}_i*\underline{x\ell}_i') & \frac{\partial^2\mathcal{L}_{\ell i}^{\mu}}{\partial\ell\partial\mu}\otimes(\underline{x\mu}_i*\underline{x\ell}_i')\\ \hline \frac{\partial^2\mathcal{L}_{\ell i}^{\mu}}{\partial\mu\partial u}\otimes(\underline{x\mu}_i*\underline{xu}_i') & \frac{\partial^2\mathcal{L}_{\ell i}^{\mu}}{\partial\mu\partial\ell}\otimes(\underline{x\mu}_i*\underline{x\ell}_i') & \frac{\partial^2\mathcal{L}_{\ell i}^{\mu}}{\partial\mu^2}\otimes(\underline{x\mu}_i*\underline{x\mu}_i')\end{array}\right)$$

where the partials are as written in equation E.14.

E.3 Transfer model

Here we record the derivatives for the transfer model. Write $f_{ui}^{(\ell)}$ for $f_{ui}^{(\ell)}$. As stated in section(?), the log-likelihood for the transfer model without covariates is

$$\mathcal{L}(u,\ell,w,p,\mu) = \sum_{\{i:D_i=\mathrm{u}\}} \log\left[q\, f_{ui}^{(\ell)}\right] + \sum_{\{i:D_i=\mathrm{l}\}} \log\left[p\, f_{\ell i}^{(w)} + q\, f_{\ell i}^{(u)}\right] +$$

$$\sum_{\{i:D_i=\mathrm{w}\}} \log\left[p\, f_{wi}^{(\ell)}\right] + \sum_{\{i:D_i=\mathrm{c}\}} \log\left[1 - qF_{u,\ell}(t_i) - pF_{w,\ell}(t_i)\right]$$

where D_i is u, l, w, or c depending on whether individual i is discharged, dies, is transferred or is censored at time t_i. For this model, we want to estimate the parameter vector $\boldsymbol{\theta} = (\theta_1,\theta_2,\theta_3,\theta_4,\theta_5)' = (u,\ell,w,p,\mu)'$, where the parameter u represents the value of the upper barrier, ℓ represents the lower barrier, w the transition barrier, and μ is the drift parameter for the underlying health level process. The parameter $p = 1 - q$ is the probability that an individual who passes through the transfer barrier at w is actually transferred. Let $y(u,t) = t^{-\frac{3}{2}} e^{\frac{-(\mu t - 2u)\mu}{2}}$ and define the function

$$f(u,l,\mu,t) = y(u,t) \sum_k \phi\left(\frac{c_k}{\sqrt{t}}\right) c_k.$$

where $c_k = 2k(u+\ell) + u$. Here and in the rest of the section, sums over k run from $-\infty$ to ∞. Then

$$\begin{aligned} f_{ui}^{(\ell)} &= f(u,l,\mu,t_i) \\ f_{wi}^{(\ell)} &= f(w,l,\mu,t_i) \\ f_{\ell i}^{(u)} &= f(l,u,-\mu,t_i) \\ f_{\ell i}^{(w)} &= f(l,w,-\mu,t_i) \end{aligned}$$

Also define

$$F(u,l,\mu,t) = \sum_k e^{-c_k\mu}\left[\Phi\left(\frac{cu_k}{\sqrt{t}}\right) - \Phi\left(\frac{c\ell_k}{\sqrt{t}}\right)\right] - e^{d_k\mu}\left[\Phi\left(\frac{du_k}{\sqrt{t}}\right) - \Phi\left(\frac{d\ell_k}{\sqrt{t}}\right)\right]$$

where $cu_k = 2k(u+\ell) + u - \mu t$, $c\ell_k = 2k(u+\ell) - \ell - \mu t$, $du_k = -2k(u+\ell) - u - \mu t$, and $d\ell_k = -2k(u+\ell) - 2u - \ell - \mu t$.

The gradient, $G(u,\ell,\mu,w,p)$ has components

$$G(u,\ell,\mu,w,p) = \sum_{\{i:D_i=u\}} G_{ui} + \sum_{\{i:D_i=l\}} G_{\ell i} + \sum_{\{i:D_i=w\}} G_{wi} + \sum_{\{i:D_i=c\}} G_{ci}.$$

Here G_{ui} is a vector of length 5 with j^{th} element, a_j say, where

$$a_j = \frac{\partial}{\partial\theta_j}\log\left[qf_{ui}^{(\ell)}\right] = \left(\frac{1}{qf_{ui}^{(\ell)}}\right)\left(\frac{\partial\, qf_{ui}^{(\ell)}}{\partial\theta_j}\right).$$

Recall that $q = (1-p)$ is a parameter, so does not cancel in general in the above expression. Define the functions

$$\begin{aligned}
G_1(u,\ell,\mu,t) &= \frac{\partial f_u^{(\ell)}}{\partial u} = y(u,t)\sum_k \phi\left(\frac{c_k}{\sqrt{t}}\right)\left\{\mu c_k + (2k+1)\left[1 - \frac{c_k^2}{t}\right]\right\} \\
G_2(u,\ell,\mu,t) &= \frac{\partial f_u^{(\ell)}}{\partial \ell} = y(u,t)\sum_k 2k\,\phi\left(\frac{c_k}{\sqrt{t}}\right)\left[1 - \frac{c_k^2}{t}\right] \\
G_3(u,\ell,\mu,t) &= \frac{\partial f_u^{(\ell)}}{\partial \mu} = (u-\mu t)f(u,l,\mu,t)
\end{aligned} \tag{E.16}$$

Then the elements of G_{ui} are given by

$$\begin{aligned}
a_1 &= \frac{\frac{\partial}{\partial u} q\, f_{ui}^{(\ell)}}{q\, f_{ui}^{(\ell)}} = \frac{G_1(u,\ell,\mu,t_i)}{f(u,\ell,\mu,t_i)} \\
a_2 &= \frac{\frac{\partial}{\partial \ell} q\, f_{ui}^{(\ell)}}{q\, f_{ui}^{(\ell)}} = \frac{G_2(u,\ell,\mu,t_i)}{f(u,\ell,\mu,t_i)} \\
a_3 &= \frac{\frac{\partial}{\partial \mu} q\, f_{ui}^{(\ell)}}{q\, f_{ui}^{(\ell)}} = \frac{G_3(u,\ell,\mu,t_i)}{f(u,\ell,\mu,t_i)} = (u-\mu t_i) \\
a_4 &= \frac{\frac{\partial}{\partial w} q\, f_{ui}^{(\ell)}}{q\, f_{ui}^{(\ell)}} = 0
\end{aligned}$$

$$a_5 = \frac{\frac{\partial}{\partial p} q\ f_{ui}^{(\ell)}}{q\ f_{ui}^{(\ell)}} = -1/q$$

Note that $y(u,t)$ appears in both the numerator and denominator of a_1 and a_2, so some further simplification is possible.

G_{wi} is a vector of length 5 with j^{th} element, b_j say, where

$$b_j = \frac{\partial}{\partial \theta_j} \log \left[p f_{wi}^{(\ell)} \right] = \left(\frac{1}{p f_{wi}^{(\ell)}} \right) \left(\frac{\partial\, p f_{wi}^{(\ell)}}{\partial \theta_j} \right).$$

Then the elements of G_{wi} can be written as

$$b_1 = \frac{\frac{\partial}{\partial u} q\ f_{wi}^{(\ell)}}{q\ f_{wi}^{(\ell)}} = 0$$

$$b_2 = \frac{\frac{\partial}{\partial \ell} q\ f_{wi}^{(\ell)}}{q\ f_{wi}^{(\ell)}} = \frac{G_2(w,\ell,\mu,t_i)}{f(w,\ell,\mu,t_i)}$$

$$b_3 = \frac{\frac{\partial}{\partial \mu} q\ f_{wi}^{(\ell)}}{q\ f_{wi}^{(\ell)}} = \frac{G_3(w,\ell,\mu,t_i)}{f(w,\ell,\mu,t_i)} = (w - \mu t_i)$$

$$b_4 = \frac{\frac{\partial}{\partial w} q\ f_{wi}^{(\ell)}}{q\ f_{wi}^{(\ell)}} = \frac{G_1(w,\ell,\mu,t_i)}{f(w,\ell,\mu,t_i)}$$

$$b_5 = \frac{\frac{\partial}{\partial p} q\ f_{wi}^{(\ell)}}{q\ f_{wi}^{(\ell)}} = 1/p$$

$G_{\ell i}$ has j^{th} element

$$\frac{\partial}{\partial \theta_j} \log \left[p f_{\ell i}^{(w)} + q f_{\ell i}^{(u)} \right] = \frac{\frac{\partial}{\partial \theta_j} [p f_{\ell i}^{(w)} + q f_{\ell i}^{(u)}]}{q f_{\ell i}^{(u)} + p f_{\ell i}^{(w)}}$$

$G_{\ell i}$ can be expressed as $D1u + D1w$ where $D1u$ is a vector of length 5 with j^{th} element c_j say where

$$c_j = \left(\frac{1}{q f_{\ell i}^{(u)} + p f_{\ell i}^{(w)}} \right) \left(\frac{\partial}{\partial \theta_j}\ q f_{\ell i}^{(u)} \right),$$

and $D1w$ is a vector of length 5 with j^{th} element d_j say where

$$d_j = \left(\frac{1}{q f_{\ell i}^{(u)} + p f_{\ell i}^{(w)}} \right) \left(\frac{\partial}{\partial \theta_j}\ p f_{\ell i}^{(w)} \right).$$

The elements of $D1u$ can be written as

$$c_1 = \frac{\frac{\partial}{\partial u} q\, f_{\ell i}^{(u)}}{q\, f_{\ell i}^{(u)} + p\, f_{\ell i}^{(w)}} = \frac{q\, G_2(\ell, u, -\mu, t_i)}{q\, f(\ell, u, -\mu, t_i) + p\, f(\ell, w, -\mu, t_i)}$$

$$c_2 = \frac{\frac{\partial}{\partial \ell} q\, f_{\ell i}^{(u)}}{q\, f_{\ell i}^{(u)} + p\, f_{\ell i}^{(w)}} = \frac{q\, G_1(\ell, u, -\mu, t_i)}{q\, f(\ell, u, -\mu, t_i) + p\, f(\ell, w, -\mu, t_i)}$$

$$c_3 = \frac{\frac{\partial}{\partial \mu} q\, f_{\ell i}^{(u)}}{q\, f_{\ell i}^{(u)} + p\, f_{\ell i}^{(w)}} = \frac{-q\, G_3(\ell, u, -\mu, t_i)}{q\, f(\ell, u, -\mu, t_i) + p\, f(\ell, w, -\mu, t_i)}$$

$$c_4 = \frac{\frac{\partial}{\partial w} q\, f_{\ell i}^{(u)}}{q\, f_{\ell i}^{(u)} + p\, f_{\ell i}^{(w)}} = 0$$

$$c_5 = \frac{\frac{\partial}{\partial p} q\, f_{\ell i}^{(u)}}{q\, f_{\ell i}^{(u)} + p\, f_{\ell i}^{(w)}} = \frac{-f(\ell, u, -\mu, t_i)}{q\, f(\ell, u, -\mu, t_i) + p\, f(\ell, w, -\mu, t_i)}$$

and the elements of $D1w$ are

$$d_1 = \frac{\frac{\partial}{\partial u} p\, f_{\ell i}^{(w)}}{q\, f_{\ell i}^{(u)} + p\, f_{\ell i}^{(w)}} = 0$$

$$d_2 = \frac{\frac{\partial}{\partial \ell} p\, f_{\ell i}^{(w)}}{q\, f_{\ell i}^{(u)} + p\, f_{\ell i}^{(w)}} = \frac{p\, G_1(\ell, w, -\mu, t_i)}{q\, f(\ell, u, -\mu, t_i) + p\, f(\ell, w, -\mu, t_i)}$$

$$d_3 = \frac{\frac{\partial}{\partial \mu} p\, f_{\ell i}^{(w)}}{q\, f_{\ell i}^{(u)} + p\, f_{\ell i}^{(w)}} = \frac{-p\, G_3(\ell, w, -\mu, t_i)}{q\, f(\ell, u, -\mu, t_i) + p\, f(\ell, w, -\mu, t_i)}$$

$$d_4 = \frac{\frac{\partial}{\partial w} p\, f_{\ell i}^{(w)}}{q\, f_{\ell i}^{(u)} + p\, f_{\ell i}^{(w)}} = \frac{p\, G_2(\ell, w, -\mu, t_i)}{q\, f(\ell, u, -\mu, t_i) + p\, f(\ell, w, -\mu, t_i)}$$

$$d_5 = \frac{\frac{\partial}{\partial p} p\, f_{\ell i}^{(w)}}{q\, f_{\ell i}^{(u)} + p\, f_{\ell i}^{(w)}} = \frac{f(\ell, w, -\mu, t_i)}{q\, f(\ell, u, -\mu, t_i) + p\, f(\ell, w, -\mu, t_i)}$$

Note that the minus sign in c_3 and d_3 appear because we have replaced μ with $-\mu$ in the function G_3, but still desire the derivative with respect to positive μ.

Finally, G_{ci} has j^{th} element

$$\frac{\partial}{\partial \theta_j} \log\left[1 - qF_{u,\ell}(t_i) - pF_{w,\ell}(t_i)\right] = \frac{\frac{\partial}{\partial \theta_j}\left[-qF_{u,\ell}(t_i) - pF_{w,\ell}(t_i)\right]}{1 - qF_{u,\ell}(t_i) - pF_{w,\ell}(t_i)}$$

G_{ci} can be expressed as $-D1Fu - D1Fw$ where $D1Fu$ is a vector of length 5 with j^{th} element

e_j say where

$$e_j = \left(\frac{1}{1 - qF_{u,\ell}(t_i) - pF_{w,\ell}(t_i)}\right)\left(\frac{\partial}{\partial\theta_j}\, qF_{u,\ell}(t_i)\right),$$

and $D1Fw$ is a vector of length 5 with j^{th} element f_j say where

$$f_j = \left(\frac{1}{1 - qF_{u,\ell}(t_i) - pF_{w,\ell}(t_i)}\right)\left(\frac{\partial}{\partial\theta_j}\, pF_{w,\ell}(t_i)\right).$$

Define the functions

$$\begin{aligned}
GF_1(u,\ell,\mu,t) =& \frac{\partial}{\partial u}F_{u,\ell}(t) = \Big\{-2k\mu e^{-c_k\mu}\left[\Phi\left(\frac{cu_k}{\sqrt{t_i}}\right) - \Phi\left(\frac{c\ell_k}{\sqrt{t_i}}\right)\right] + \\
& e^{-c_k\mu}\left[\frac{2k+1}{\sqrt{t_i}}\phi\left(\frac{cu_k}{\sqrt{t_i}}\right) - \frac{2k}{\sqrt{t_i}}\phi\left(\frac{c\ell_k}{\sqrt{t_i}}\right)\right] \\
& (2k+2)\mu e^{d_k\mu}\left[\Phi\left(\frac{du_k}{\sqrt{t_i}}\right) - \Phi\left(\frac{d\ell_k}{\sqrt{t_i}}\right)\right] - \\
& e^{d_k\mu}\left[\frac{-2k-1}{\sqrt{t_i}}\phi\left(\frac{du_k}{\sqrt{t_i}}\right) + \frac{2k+2}{\sqrt{t_i}}\phi\left(\frac{d\ell_k}{\sqrt{t_i}}\right)\right]\Big\} \\
GF_2(u,\ell,\mu,t) =& \frac{\partial}{\partial \ell}F_{u,\ell}(t) = \Big\{-2k\mu e^{-c_k\mu}\left[\Phi\left(\frac{cu_k}{\sqrt{t_i}}\right) - \Phi\left(\frac{c\ell_k}{\sqrt{t_i}}\right)\right] + \\
& e^{-c_k\mu}\left[\frac{2k}{\sqrt{t_i}}\phi\left(\frac{cu_k}{\sqrt{t_i}}\right) - \frac{2k-1}{\sqrt{t_i}}\phi\left(\frac{c\ell_k}{\sqrt{t_i}}\right)\right] - \\
& 2k\mu e^{d_k\mu}\left[\Phi\left(\frac{du_k}{\sqrt{t_i}}\right) - \Phi\left(\frac{d\ell_k}{\sqrt{t_i}}\right)\right] - \\
& e^{d_k\mu}\left[-\frac{2k}{\sqrt{t}} - \phi\left(\frac{du_k}{\sqrt{t}}\right) - \frac{-2k-1}{\sqrt{t}}\phi\left(\frac{d\ell_k}{\sqrt{t}}\right)\right]\Big\} \\
GF_3(u,\ell,\mu,t) =& \frac{\partial}{\partial \mu}F_{u,\ell}(t) = \Big\{-c_k e^{-c_k\mu}\left[\Phi\left(\frac{cu_k}{\sqrt{t}}\right) - \Phi\left(\frac{c\ell_k}{\sqrt{t}}\right)\right] + \\
& e^{-c_k\mu}\left[-\phi\left(\frac{cu_k}{\sqrt{t}}\right)\sqrt{t} + \phi\left(\frac{c\ell_k}{\sqrt{t}}\right)\sqrt{t}\right] - \\
& d_k e^{d_k\mu}\left[\Phi\left(\frac{du_k}{\sqrt{t}}\right) - \Phi\left(\frac{d\ell_k}{\sqrt{t}}\right)\right] - \\
& e^{d_k\mu}\left[-\phi\left(\frac{du_k}{\sqrt{t}}\right)\sqrt{t} + \phi\left(\frac{d\ell_k}{\sqrt{t}}\right)\sqrt{t}\right]\Big\}
\end{aligned}$$

Then the elements of $D1Fu$ can be written as

$$\begin{aligned}
e_1 &= \frac{q}{a}\frac{\partial}{\partial u}F_{u,\ell}(t_i) = \frac{q}{a}GF_1(u,\ell,\mu,t_i) \\
e_2 &= \frac{q}{a}\frac{\partial}{\partial \ell}F_{u,\ell}(t_i) = \frac{q}{a}GF_2(u,\ell,\mu,t_i)
\end{aligned}$$

$$e_3 = \frac{q}{a}\frac{\partial}{\partial\mu}F_{u,\ell}(t_i) = \frac{q}{a}GF_3(u,\ell,\mu,t_i)$$
$$e_4 = \frac{q}{a}\frac{\partial}{\partial w}F_{u,\ell}(t_i) = 0$$
$$e_5 = \frac{1}{a}\frac{\partial}{\partial p}qF_{u,\ell}(t_i) = -F(u,\ell,\mu,t_i)$$

where $a = 1/(1 - qF_{u,\ell}(t_i) - pF_{w,\ell}(t_i))$. The elements of $D1Fw$ can be written as

$$f_1 = \frac{p}{a}\frac{\partial}{\partial u}F_{w,\ell}(t_i) = 0$$
$$f_2 = \frac{p}{a}\frac{\partial}{\partial \ell}F_{w,\ell}(t_i) = \frac{p}{a}GF_2(w,\ell,\mu,t_i)$$
$$f_3 = \frac{p}{a}\frac{\partial}{\partial\mu}F_{w,\ell}(t_i) = \frac{p}{a}GF_3(w,\ell,\mu,t_i)$$
$$f_4 = \frac{p}{a}\frac{\partial}{\partial w}F_{w,\ell}(t_i) = \frac{p}{a}GF_1(2,\ell,\mu,t_i)$$
$$f_5 = \frac{1}{a}\frac{\partial}{\partial p}pF_{w,\ell}(t_i) = F(w,\ell,\mu,t_i)$$

The Hessian $H(u,\ell,\mu,w,p)$ has components

$$H(u,\ell,\mu,w,p) = \sum_{\{i:D_i=u\}} H_{ui} + \sum_{\{i:D_i=l\}} H_{\ell i} + \sum_{\{i:D_i=w\}} H_{wi} + \sum_{\{i:D_i=c\}} H_{ci}.$$

Here H_{ui} is a 5 x 5 symmetric matrix with the element in row j, column k given by

$$\frac{\partial^2}{\partial\theta_j\partial\theta_k}\log\left[qf_{ui}^{(\ell)}\right] = \frac{\partial}{\partial\theta_j}\left(\frac{\partial}{\partial\theta_k}\log\left[qf_{ui}^{(\ell)}\right]\right) = \frac{\frac{\partial^2}{\partial\theta_j\partial\theta_k}qf_{ui}^{(\ell)}}{qf_{ui}^{(\ell)}} - \left(\frac{\frac{\partial}{\partial\theta_j}qf_{ui}^{(\ell)}}{qf_{ui}^{(\ell)}}\right)\left(\frac{\frac{\partial}{\partial\theta_k}qf_{ui}^{(\ell)}}{qf_{ui}^{(\ell)}}\right)$$

which can be expressed as

$$H_{ui} = D2u - G_{ui} * G'_{ui}$$

where $*$ denotes matrix multiplication, and G_{ui} is given in equation (E.16). Here $D2u$ is a 5 x 5 symmetric matrix with elements $a_{j,k}$ say where

$$a_{j,k} = \left(\frac{1}{qf_{ui}^{(\ell)}}\right)\left(\frac{\partial^2}{\partial\theta_j\theta_k}\,qf_{ui}^{(\ell)}\right).$$

Define the functions

$$H_1(u,\ell,\mu,t) = \frac{\partial^2 f_u^{(\ell)}}{\partial u \partial u} = y(u,t)\sum_k \phi\left(\frac{c_k}{\sqrt{t}}\right)\left\{\mu^2 c_k + 2\mu(2k+1)\left[1-\frac{c_k^2}{t}\right] + (2k+1)^2\frac{c_k}{t}\left[\frac{c_k^2}{t}-3\right]\right\}$$

$$H_2(u,\ell,\mu,t) = \frac{\partial^2 f_u^{(\ell)}}{\partial \ell \partial u} = y(u,t)\sum_k \phi\left(\frac{c_k}{\sqrt{t}}\right) 2k\left\{(2k+1)\frac{c_k}{t}\left[\frac{c_k^2}{t}-3\right] + \mu\left[1-\frac{c_k^2}{t}\right]\right\}$$

$$H_3(u,\ell,\mu,t) = \frac{\partial^2 f_u^{(\ell)}}{\partial \ell \partial \ell} = y(u,t)\sum_k \phi\left(\frac{c_k}{\sqrt{t}}\right)(2k)^2\frac{c_k}{t}\left[\frac{c_k^2}{t}-3\right]$$

$$H_4(u,\ell,\mu,t) = \frac{\partial^2 f_u^{(\ell)}}{\partial \mu \partial u} = y(u,t)\sum_k \phi\left(\frac{c_k}{\sqrt{t}}\right)\left\{(u-\mu t)\left(\mu c_k + (2k+1)\left[1-\frac{c_k^2}{t}\right]\right) + c_k\right\}$$

$$H_5(u,\ell,\mu,t) = \frac{\partial^2 f_u^{(\ell)}}{\partial \mu \partial \ell} = y(u,t)(u-\mu t)\sum_k \phi\left(\frac{c_k}{\sqrt{t}}\right) 2k\left[1-\frac{c_k^2}{t}\right]$$

$$H_6(u,\ell,\mu,t) = \frac{\partial^2 f_u^{(\ell)}}{\partial \mu \partial \mu} = ((u-\mu t)^2 - t)\ f(u,\ell,\mu,t)$$

Then the unique elements of $D2u$ are

$$a_{1,1} = \frac{\frac{\partial^2}{\partial u \partial u} q\ f_{ui}^{(\ell)}}{q\ f_{ui}^{(\ell)}} = \frac{H_1(u,\ell,\mu,t_i)}{f(u,\ell,\mu,t_i)}$$

$$a_{2,1} = \frac{\frac{\partial^2}{\partial \ell \partial u} q f_{ui}^{(\ell)}}{q\ f_{ui}^{(\ell)}} = \frac{H_2(u,\ell,\mu,t_i)}{f(u,\ell,\mu,t_i)}$$

$$a_{2,2} = \frac{\frac{\partial^2}{\partial \ell \partial \ell} q\ f_{ui}^{(\ell)}}{q\ f_{ui}^{(\ell)}} = \frac{H_3(u,\ell,\mu,t_i)}{f(u,\ell,\mu,t_i)}$$

$$a_{3,1} = \frac{\frac{\partial^2}{\partial \mu \partial u} q\ f_{ui}^{(\ell)}}{q\ f_{ui}^{(\ell)}} = \frac{H_4(u,\ell,\mu,t_i)}{f(u,\ell,\mu,t_i)}$$

$$a_{3,2} = \frac{\frac{\partial^2}{\partial \mu \partial \ell} q\ f_{ui}^{(\ell)}}{q\ f_{ui}^{(\ell)}} = \frac{H_5(u,\ell,\mu,t_i)}{f(u,\ell,\mu,t_i)}$$

$$a_{3,3} = \frac{\frac{\partial^2}{\partial \mu \partial \mu} q\ f_{ui}^{(\ell)}}{q\ f_{ui}^{(\ell)}} = \frac{H_6(u,\ell,\mu,t_i)}{f(u,\ell,\mu,t_i)} = (u-\mu t_i)^2 - t_i$$

$$a_{4,1} = \frac{\frac{\partial^2}{\partial w \partial u} q\ f_{ui}^{(\ell)}}{q\ f_{ui}^{(\ell)}} = 0 \qquad a_{4,2} = \frac{\frac{\partial^2}{\partial w \partial \ell} q\ f_{ui}^{(\ell)}}{q\ f_{ui}^{(\ell)}} = 0$$

$$a_{4,3} = \frac{\frac{\partial^2}{\partial w \partial \mu} q\ f_{ui}^{(\ell)}}{q\ f_{ui}^{(\ell)}} = 0 \qquad a_{4,4} = \frac{\frac{\partial^2}{\partial w \partial w} q\ f_{ui}^{(\ell)}}{q\ f_{ui}^{(\ell)}} = 0$$

$$a_{5,1} = \frac{\frac{\partial}{\partial p}\frac{\partial}{\partial u} q\ f_{ui}^{(\ell)}}{q\ f_{ui}^{(\ell)}} = \frac{-\frac{\partial}{\partial u} f_{ui}^{(\ell)}}{q\ f_{ui}^{(\ell)}} = \frac{-G_1(u,\ell,\mu,t_i)}{q f(u,\ell,\mu,t_i)}$$

$$a_{5,2} = \frac{\frac{\partial}{\partial p}\frac{\partial}{\partial \ell} q f_{ui}^{(\ell)}}{q\ f_{ui}^{(\ell)}} = \frac{-\frac{\partial}{\partial \ell} f_{ui}^{(\ell)}}{q\ f_{ui}^{(\ell)}} = \frac{-G_2(u,\ell,\mu,t_i)}{qf(u,\ell,\mu,t_i)}$$

$$a_{5,3} = \frac{\frac{\partial}{\partial p}\frac{\partial}{\partial \mu} q\ f_{ui}^{(\ell)}}{q\ f_{ui}^{(\ell)}} = \frac{-\frac{\partial}{\partial \mu} f_{ui}^{(\ell)}}{q\ f_{ui}^{(\ell)}} = \frac{-G_3(u,\ell,\mu,t_i)}{qf(u,\ell,\mu,t_i)}$$

$$a_{5,4} = \frac{\frac{\partial}{\partial p}\frac{\partial}{\partial w} q\ f_{ui}^{(\ell)}}{q\ f_{ui}^{(\ell)}} = 0$$

$$a_{5,5} = \frac{\frac{\partial}{\partial p}\frac{\partial}{\partial p} q\ f_{ui}^{(\ell)}}{q\ f_{ui}^{(\ell)}} = \frac{-\frac{\partial}{\partial p} f_{ui}^{(\ell)}}{q\ f_{ui}^{(\ell)}} = 0$$

The last row of the Hessian can be simplified somewhat. Note that

$$\frac{\partial^2}{\partial p \partial \theta_k} \log\left[q f_{ui}^{(\ell)}\right] = \frac{\partial}{\partial p}\left(\frac{q \frac{\partial}{\partial \theta_k} f_{ui}^{(\ell)}}{q f_{ui}^{(\ell)}}\right)$$

equals 0 for $\theta_k \in \{u, \ell, \mu, w\}$. Thus letting $h_{j,k}$ denote the element in row j, column k of H_{ui} we have

$$h_{5,1} = h_{5,2} = h_{5,3} = h_{5,4} = 0$$

which saves some computation. Also note that

$$h_{5,5} = \frac{\partial^2}{\partial p \partial p} \log\left[q f_{ui}^{(\ell)}\right] = \frac{\partial}{\partial p}\left(\frac{-f_{ui}^{(\ell)}}{q f_{ui}^{(\ell)}}\right) = -1/(q^2)$$

Persons who are transferred contribute the term H_{wi}, a 5 x 5 symmetric matrix with elements

$$\frac{\partial^2}{\partial \theta_j \partial \theta_k} \log(p f_{wi}^{(\ell)})$$

which can be expressed as

$$H_{wi} = D2w - G_{wi} * G_{wi},$$

where $D2w$ is a 5 x 5 symmetric matrix with elements $b_{j,k}$ say where

$$b_{j,k} = \left(\frac{1}{p f_{wi}^{(\ell)}}\right)\left(\frac{\partial^2}{\partial \theta_j \theta_k}\ p f_{wi}^{(\ell)}\right).$$

Then the unique elements of $D2w$ are

$$b_{1,1} = \frac{\frac{\partial^2}{\partial u \partial u} p\, f_{wi}^{(\ell)}}{p\, f_{wi}^{(\ell)}} = 0 \qquad b_{2,1} = \frac{\frac{\partial^2}{\partial \ell \partial u} p\, f_{wi}^{(\ell)}}{p\, f_{wi}^{(\ell)}} = 0$$

$$b_{2,2} = \frac{\frac{\partial^2}{\partial \ell \partial \ell} p\, f_{wi}^{(\ell)}}{p\, f_{wi}^{(\ell)}} = \frac{H_3(w,\ell,\mu,t_i)}{f(w,\ell,\mu,t_i)}$$

$$b_{3,1} = \frac{\frac{\partial^2}{\partial \mu \partial u} p\, f_{wi}^{(\ell)}}{p\, f_{wi}^{(\ell)}} = 0$$

$$b_{3,2} = \frac{\frac{\partial^2}{\partial \mu \partial \ell} p\, f_{wi}^{(\ell)}}{p\, f_{wi}^{(\ell)}} = \frac{H_5(w,\ell,\mu,t_i)}{f(w,\ell,\mu,t_i)}$$

$$b_{3,3} = \frac{\frac{\partial^2}{\partial \mu \partial \mu} p\, f_{wi}^{(\ell)}}{p\, f_{wi}^{(\ell)}} = \frac{H_6(w,\ell,\mu,t_i)}{f(w,\ell,\mu,t_i)} = (w - \mu t_i)^2 - t_i$$

$$b_{4,1} = \frac{\frac{\partial^2}{\partial w \partial u} p\, f_{wi}^{(\ell)}}{p\, f_{wi}^{(\ell)}} = 0$$

$$b_{4,2} = \frac{\frac{\partial^2}{\partial w \partial \ell} p\, f_{wi}^{(\ell)}}{p\, f_{wi}^{(\ell)}} = \frac{H_2(w,\ell,\mu,t_i)}{f(w,\ell,\mu,t_i)}$$

$$b_{4,3} = \frac{\frac{\partial^2}{\partial w \partial \mu} p\, f_{wi}^{(\ell)}}{p\, f_{wi}^{(\ell)}} = \frac{H_4(w,\ell,\mu,t_i)}{f(w,\ell,\mu,t_i)}$$

$$b_{4,4} = \frac{\frac{\partial^2}{\partial w \partial w} p\, f_{wi}^{(\ell)}}{p\, f_{wi}^{(\ell)}} = \frac{H_1(w,\ell,\mu,t_i)}{f(w,\ell,\mu,t_i)}$$

$$b_{5,1} = \frac{\frac{\partial}{\partial p}\frac{\partial}{\partial u} p\, f_{wi}^{(\ell)}}{p\, f_{wi}^{(\ell)}} = 0$$

$$b_{5,2} = \frac{\frac{\partial}{\partial p}\frac{\partial}{\partial \ell} p f_{wi}^{(\ell)}}{p\, f_{wi}^{(\ell)}} = \frac{\frac{\partial}{\partial \ell} f_{wi}^{(\ell)}}{p\, f_{wi}^{(\ell)}} = \frac{G_2(w,\ell,\mu,t_i)}{pf(w,\ell,\mu,t_i)}$$

$$b_{5,3} = \frac{\frac{\partial}{\partial p}\frac{\partial}{\partial \mu} p\, f_{wi}^{(\ell)}}{p\, f_{wi}^{(\ell)}} = \frac{\frac{\partial}{\partial \mu} f_{wi}^{(\ell)}}{p\, f_{wi}^{(\ell)}} = \frac{G_3(w,\ell,\mu,t_i)}{pf(w,\ell,\mu,t_i)}$$

$$b_{5,4} = \frac{\frac{\partial}{\partial p}\frac{\partial}{\partial w} p\, f_{wi}^{(\ell)}}{p\, f_{wi}^{(\ell)}} = \frac{\frac{\partial}{\partial w} f_{wi}^{(\ell)}}{p\, f_{wi}^{(\ell)}} = \frac{G_1(w,\ell,\mu,t_i)}{pf(w,\ell,\mu,t_i)}$$

$$b_{5,5} = \frac{\frac{\partial}{\partial p}\frac{\partial}{\partial p} p\, f_{wi}^{(\ell)}}{p\, f_{wi}^{(\ell)}} = \frac{\frac{\partial}{\partial p} f_{wi}^{(\ell)}}{p\, f_{wi}^{(\ell)}} = 0$$

Again, the last row of the Hessian can be simplified somewhat. Note that

$$\frac{\partial^2}{\partial p \partial \theta_k} \log\left[p f_{wi}^{(\ell)}\right] = \frac{\partial}{\partial p}\left(\frac{p \frac{\partial}{\partial \theta_k} f_{wi}^{(\ell)}}{p f_{wi}^{(\ell)}}\right)$$

equals 0 for $\theta_k \in \{u, \ell, \mu, w\}$. Thus letting $h_{j,k}$ denote the element in row j, column k of H_{wi} we have

$$h_{5,1} = h_{5,2} = h_{5,3} = h_{5,4} = 0$$

which saves some computation. Also note that

$$h_{5,5} = \frac{\partial^2}{\partial p \partial p} \log\left[p f_{wi}^{(\ell)}\right] = \frac{\partial}{\partial p}\left(\frac{f_{wi}^{(\ell)}}{p f_{wi}^{(\ell)}}\right) = -1/(p^2)$$

Persons who die at time t_i contribute the term $H_{\ell i}$, a 5 x 5 symmetric matrix with elements

$$\frac{\partial^2}{\partial\theta_j \partial\theta_k} \log\left[q f_{\ell i}^{(u)} + p f_{\ell i}^{(w)}\right]$$

which can be expressed as

$$H_{\ell i} = (D2\ell u + D2\ell w) - G_{\ell i} * G'_{\ell i}$$

where $D2\ell u$ is a 5 x 5 symmetric matrix with elements $c_{j,k}$ say where

$$c_{j,k} = \left(\frac{1}{q f_{\ell i}^{(u)} + p f_{\ell i}^{(w)}}\right)\left(\frac{\partial^2}{\partial\theta_j \theta_k}\, q f_{\ell i}^{(u)}\right).$$

and where $D2\ell w$ is a 5 x 5 symmetric matrix with elements $d_{j,k}$ say where

$$d_{j,k} = \left(\frac{1}{q f_{\ell i}^{(u)} + p f_{\ell i}^{(w)}}\right)\left(\frac{\partial^2}{\partial\theta_j \theta_k}\, p f_{\ell i}^{(w)}\right).$$

Then the unique elements of $D2\ell u$ are

$$c_{1,1} = \frac{\frac{\partial^2}{\partial u \partial u} q\, f_{\ell i}^{(u)}}{q\, f_{\ell i}^{(u)} + p\, {}_w f_{\ell i}} = \frac{q H_3(\ell, u, -\mu, t_i)}{q f(\ell, u, -\mu, t_i) + p f(\ell, w, -\mu, t_i)}$$

$$c_{2,1} = \frac{\frac{\partial^2}{\partial \ell \partial u} q\, f_{\ell i}^{(u)}}{q\, f_{\ell i}^{(u)} + p\, {}_w f_{\ell i}} = \frac{q H_2(\ell, u, -\mu, t_i)}{q f(\ell, u, -\mu, t_i) + p f(\ell, w, -\mu, t_i)}$$

$$c_{2,2} = \frac{\frac{\partial^2}{\partial \ell \partial \ell} q\, f_{\ell i}^{(u)}}{q\, f_{\ell i}^{(u)} + p\, {}_w f_{\ell i}} = \frac{q H_1(\ell, u, -\mu, t_i)}{q f(\ell, u, -\mu, t_i) + p f(\ell, w, -\mu, t_i)}$$

$$c_{3,1} = \frac{\frac{\partial^2}{\partial \mu \partial u} q\, f_{\ell i}^{(u)}}{q\, f_{\ell i}^{(u)} + p\, {}_w f_{\ell i}} = \frac{-q H_5(\ell, u, -\mu, t_i)}{q f(\ell, u, -\mu, t_i) + p f(\ell, w, -\mu, t_i)}$$

$$c_{3,2} = \frac{\frac{\partial^2}{\partial\mu\partial\ell} q\ f_{\ell i}^{(u)}}{q\ f_{\ell i}^{(u)} + p\ {}_w f_{\ell i}} = \frac{-qH_4(\ell, u, -\mu, t_i)}{qf(\ell, u, -\mu, t_i) + pf(\ell, w, -\mu, t_i)}$$

$$c_{3,3} = \frac{\frac{\partial^2}{\partial\mu\partial\mu} q\ f_{\ell i}^{(u)}}{q\ f_{\ell i}^{(u)} + p\ {}_w f_{\ell i}} = \frac{qH_6(\ell, u, -\mu, t_i)}{qf(\ell, u, -\mu, t_i) + pf(\ell, w, -\mu, t_i)}$$

$$c_{4,1} = \frac{\frac{\partial^2}{\partial w\partial u} q\ f_{\ell i}^{(u)}}{q\ f_{\ell i}^{(u)} + p\ {}_w f_{\ell i}} = 0 \qquad c_{4,2} = \frac{\frac{\partial^2}{\partial w\partial\ell} q\ f_{\ell i}^{(u)}}{q\ f_{\ell i}^{(u)} + p\ {}_w f_{\ell i}} = 0$$

$$c_{4,3} = \frac{\frac{\partial^2}{\partial w\partial\mu} q\ f_{\ell i}^{(u)}}{q\ f_{\ell i}^{(u)} + p\ {}_w f_{\ell i}} = 0 \qquad c_{4,4} = \frac{\frac{\partial^2}{\partial w\partial w} q\ f_{\ell i}^{(u)}}{q\ f_{\ell i}^{(u)} + p\ {}_w f_{\ell i}} = 0$$

$$c_{5,1} = \frac{\frac{\partial}{\partial p}\frac{\partial}{\partial u} q\ f_{\ell i}^{(u)}}{q\ f_{\ell i}^{(u)} + p\ {}_w f_{\ell i}} = \frac{-\frac{\partial}{\partial u} f_{\ell i}^{(u)}}{q\ f_{\ell i}^{(u)} + p\ {}_w f_{\ell i}} = \frac{-G_2(\ell, u, -\mu, t_i)}{q\ f_{\ell i}^{(u)} + p\ {}_w f_{\ell i}}$$

$$c_{5,2} = \frac{\frac{\partial}{\partial p}\frac{\partial}{\partial \ell} q\ f_{\ell i}^{(u)}}{q\ f_{\ell i}^{(u)} + p\ {}_w f_{\ell i}} = \frac{-\frac{\partial}{\partial \ell} f_{\ell i}^{(u)}}{q\ f_{\ell i}^{(u)} + p\ {}_w f_{\ell i}} = \frac{-G_1(\ell, u, -\mu, t_i)}{q\ f_{\ell i}^{(u)} + p\ {}_w f_{\ell i}}$$

$$c_{5,3} = \frac{\frac{\partial}{\partial p}\frac{\partial}{\partial \mu} q\ f_{\ell i}^{(u)}}{q\ f_{\ell i}^{(u)} + p\ {}_w f_{\ell i}} = \frac{-\frac{\partial}{\partial \mu} f_{\ell i}^{(u)}}{q\ f_{\ell i}^{(u)} + p\ {}_w f_{\ell i}} = \frac{G_3(\ell, u, -\mu, t_i)}{q\ f_{\ell i}^{(u)} + p\ {}_w f_{\ell i}}$$

$$c_{5,4} = \frac{\frac{\partial}{\partial p}\frac{\partial}{\partial w} q\ f_{\ell i}^{(u)}}{q\ f_{\ell i}^{(u)} + p\ {}_w f_{\ell i}} = 0$$

$$c_{5,5} = \frac{\frac{\partial}{\partial p}\frac{\partial}{\partial p} q\ f_{\ell i}^{(u)}}{q\ f_{\ell i}^{(u)} + p\ {}_w f_{\ell i}} = \frac{-\frac{\partial}{\partial p} f_{\ell i}^{(u)}}{q\ f_{\ell i}^{(u)} + p\ {}_w f_{\ell i}} = 0$$

Then the unique elements of $D2\ell w$ are

$$d_{1,1} = \frac{\frac{\partial^2}{\partial u\partial u} p\ f_{\ell i}^{(w)}}{q\ f_{\ell i}^{(u)} + p\ {}_w f_{\ell i}} = 0 \qquad d_{2,1} = \frac{\frac{\partial^2}{\partial\ell\partial u} p\ f_{\ell i}^{(w)}}{q\ f_{\ell i}^{(u)} + p\ {}_w f_{\ell i}} = 0$$

$$d_{2,2} = \frac{\frac{\partial^2}{\partial\ell\partial\ell} p\ f_{\ell i}^{(w)}}{q\ f_{\ell i}^{(u)} + p\ {}_w f_{\ell i}} = \frac{pH_1(\ell, w, -\mu, t_i)}{qf(\ell, u, -\mu, t_i) + pf(\ell, w, -\mu, t_i)}$$

$$d_{3,1} = \frac{\frac{\partial^2}{\partial\mu\partial u} p\ f_{\ell i}^{(w)}}{q\ f_{\ell i}^{(u)} + p\ {}_w f_{\ell i}} = 0$$

$$d_{3,2} = \frac{\frac{\partial^2}{\partial\mu\partial\ell} p\ f_{\ell i}^{(w)}}{q\ f_{\ell i}^{(u)} + p\ {}_w f_{\ell i}} = \frac{-pH_4(\ell, w, -\mu, t_i)}{qf(\ell, u, -\mu, t_i) + pf(\ell, w, -\mu, t_i)}$$

$$d_{3,3} = \frac{\frac{\partial^2}{\partial\mu\partial\mu} p\ f_{\ell i}^{(w)}}{q\ f_{\ell i}^{(u)} + p\ {}_w f_{\ell i}} = \frac{pH_6(\ell, w, -\mu, t_i)}{qf(\ell, u, -\mu, t_i) + pf(\ell, w, -\mu, t_i)}$$

$$d_{4,1} = \frac{\frac{\partial^2}{\partial w\partial u} p\ f_{\ell i}^{(w)}}{q\ f_{\ell i}^{(u)} + p\ {}_w f_{\ell i}} = 0$$

$$d_{4,2} = \frac{\frac{\partial^2}{\partial w \partial l} p \, f_{li}^{(w)}}{q \, f_{li}^{(u)} + p \, {}_w f_{li}} = \frac{pH_2(\ell, w, -\mu, t_i)}{qf(\ell, u, -\mu, t_i) + pf(\ell, w, -\mu, t_i)}$$

$$d_{4,3} = \frac{\frac{\partial^2}{\partial w \partial \mu} p \, f_{li}^{(w)}}{q \, f_{li}^{(u)} + p \, {}_w f_{li}} = \frac{-pH_5(\ell, w, -\mu, t_i)}{qf(\ell, u, -\mu, t_i) + pf(\ell, w, -\mu, t_i)}$$

$$d_{4,4} = \frac{\frac{\partial^2}{\partial w \partial w} p \, f_{li}^{(w)}}{q \, f_{li}^{(u)} + p \, {}_w f_{li}} = \frac{pH_3(\ell, w, -\mu, t_i)}{qf(\ell, u, -\mu, t_i) + pf(\ell, w, -\mu, t_i)}$$

$$d_{5,1} = \frac{\frac{\partial}{\partial p} \frac{\partial}{\partial u} p \, f_{li}^{(w)}}{q \, f_{li}^{(u)} + p \, {}_w f_{li}} = \frac{-\frac{\partial}{\partial u} f_{li}^{(w)}}{q \, f_{li}^{(u)} + p \, {}_w f_{li}} = 0$$

$$d_{5,2} = \frac{\frac{\partial}{\partial p} \frac{\partial}{\partial l} p \, f_{li}^{(w)}}{q \, f_{li}^{(u)} + p \, {}_w f_{li}} = \frac{\frac{\partial}{\partial l} f_{li}^{(w)}}{q \, f_{li}^{(u)} + p \, {}_w f_{li}} = \frac{G_1(\ell, w, -\mu, t_i)}{q \, f_{li}^{(u)} + p \, {}_w f_{li}}$$

$$d_{5,3} = \frac{\frac{\partial}{\partial p} \frac{\partial}{\partial \mu} p \, f_{li}^{(w)}}{q \, f_{li}^{(u)} + p \, {}_w f_{li}} = \frac{\frac{\partial}{\partial \mu} f_{li}^{(w)}}{q \, f_{li}^{(u)} + p \, {}_w f_{li}} = \frac{-G_3(\ell, w, -\mu, t_i)}{q \, f_{li}^{(u)} + p \, {}_w f_{li}}$$

$$d_{5,4} = \frac{\frac{\partial}{\partial p} \frac{\partial}{\partial w} p \, f_{li}^{(w)}}{q \, f_{li}^{(u)} + p \, {}_w f_{li}} = \frac{\frac{\partial}{\partial w} f_{li}^{(w)}}{q \, f_{li}^{(u)} + p \, {}_w f_{li}} = \frac{G_2(\ell, w, -\mu, t_i)}{q \, f_{li}^{(u)} + p \, {}_w f_{li}}$$

$$d_{5,5} = \frac{\frac{\partial}{\partial p} \frac{\partial}{\partial p} p \, f_{li}^{(w)}}{q \, f_{li}^{(u)} + p \, {}_w f_{li}} = \frac{-\frac{\partial}{\partial p} f_{li}^{(w)}}{q \, f_{li}^{(u)} + p \, {}_w f_{li}} = 0$$

Persons who are censored at time t_i contribute the term H_{ci}, a 5 x 5 symmetric matrix with elements

$$\frac{\partial^2}{\partial \theta_j \partial \theta_k} \log[-qF(u, l, t) + pF(w, l, t)]$$

which can be expressed as

$$H_{ci} = (-D2Fu - D2Fw) - G_{ci} * G'_{ci}$$

where $D2Fu$ is a 5 x 5 symmetric matrix with elements $e_{j,k}$ say where

$$e_{j,k} = \left(\frac{1}{1 - qF(u, l, t) + pF(w, l, t)} \right) \left(\frac{\partial^2}{\partial \theta_j \theta_k} \, qF(u, l, t) \right).$$

and where $D2Fw$ is a 5 x 5 symmetric matrix with elements $f_{j,k}$ say where

$$f_{j,k} = \left(\frac{1}{1 - qF(u, l, t) + pF(w, l, t)} \right) \left(\frac{\partial^2}{\partial \theta_j \theta_k} \, pF(w, l, t) \right).$$

Define the functions

$$
\begin{aligned}
HF_1(u,\ell,\mu,t) = \Bigg\{ & 4k^2\mu^2 e^{-c_k\mu}\left[\Phi\left(\frac{cu_k}{\sqrt{t_i}}\right) - \Phi\left(\frac{c\ell_k}{\sqrt{t_i}}\right)\right] - \\
& 4k\mu e^{-c_k\mu}\left[\phi\left(\frac{cu_k}{\sqrt{t_i}}\right)\frac{2k+1}{\sqrt{t_i}} - \frac{2k}{\sqrt{t_i}}\phi\left(\frac{c\ell_k}{\sqrt{t_i}}\right)\right] + \\
& e^{-c_k\mu}\left[\frac{(2k+1)^2}{t_i^{3/2}} - \phi\left(\frac{cu_k}{\sqrt{t_i}}\right)(cu_k) + \frac{4k^2}{t_i^{3/2}}\phi\left(\frac{c\ell_k}{\sqrt{t_i}}\right)(c\ell_k)\right] - \\
& (2k+2)^2\mu^2 e^{d_k\mu}\left[\Phi\left(\frac{du_k}{\sqrt{t_i}}\right) - \Phi\left(\frac{d\ell_k}{\sqrt{t_i}}\right)\right] - \\
& 2(2k+2)\mu e^{d_k\mu}\left[\phi\left(\frac{du_k}{\sqrt{t_i}}\right)\frac{-2k-1}{\sqrt{t_i}} + \phi\left(\frac{d\ell_k}{\sqrt{t_i}}\right)\frac{2k+2}{\sqrt{t_i}}\right] - \\
& e^{d_k\mu}\left[-\phi\left(\frac{du_k}{\sqrt{t_i}}\right)(du_k)\frac{(-2k-1)^2}{t_i^{3/2}} + \phi\left(\frac{d\ell_k}{\sqrt{t_i}}\right)(d\ell_k)\frac{(2k+2)^2}{t_i^{3/2}}\right]\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
HF_2(u,\ell,\mu,t) = \Bigg\{ & 4k^2\mu^2 e^{-c_k\mu}\left[\Phi\left(\frac{cu_k}{\sqrt{t_i}}\right) - \Phi\left(\frac{c\ell_k}{\sqrt{t_i}}\right)\right] - \\
& 2k\mu e^{-c_k\mu}\left[\frac{2k}{\sqrt{t_i}}\phi\left(\frac{cu_k}{\sqrt{t_i}}\right) - \frac{2k-1}{\sqrt{t_i}}\phi\left(\frac{c\ell_k}{\sqrt{t_i}}\right)\right] - \\
& 2k\mu e^{-c_k\mu}\left[\frac{2k+1}{\sqrt{t_i}} - \phi\left(\frac{cu_k}{\sqrt{t_i}}\right)\frac{2k}{\sqrt{t_i}} + \phi\left(\frac{c\ell_k}{\sqrt{t_i}}\right)\right] \\
& e^{-c_k\mu}\frac{k(2k+1)}{t_i^{3/2}}\left[-2\phi\left(\frac{cu_k}{\sqrt{t_i}}\right)(cu_k) + 2\phi\left(\frac{c\ell_k}{\sqrt{t_i}}\right)(c\ell_k)\frac{(2k-1)k}{t_i^{3/2}}\right] - \\
& 2(2k+2)\mu^2 k e^{d_k\mu}\left[\Phi\left(\frac{du_k}{\sqrt{t_i}}\right) - \Phi\left(\frac{d\ell_k}{\sqrt{t_i}}\right)\right] - \\
& (2k+2)\mu e^{d_k\mu}\left[\frac{-2k}{\sqrt{t_i}}\phi\left(\frac{du_k}{\sqrt{t_i}}\right) - \frac{-2k-1}{\sqrt{t_i}}\phi\left(\frac{d\ell_k}{\sqrt{t_i}}\right)\right] - \\
& 2k\mu e^{d_k\mu}\left[\frac{-2k-1}{\sqrt{t_i}}\phi\left(\frac{du_k}{\sqrt{t_i}}\right) + \frac{2k+2}{\sqrt{t_i}}\phi\left(\frac{d\ell_k}{\sqrt{t_i}}\right)\right] - \\
& e^{d_k\mu}\left[\frac{2k(-2k-1)}{t_i^{3/2}}\phi\left(\frac{du_k}{\sqrt{t_i}}\right)(du_k) - \phi\left(\frac{d\ell_k}{\sqrt{t_i}}\right)(d\ell_k)\frac{(-2k-1)(2k+2)}{t_i^{3/2}}\right]\Bigg\}
\end{aligned}
$$

$$
HF_3(u,\ell,\mu,t) = \Bigg\{ 4k^2\mu^2 e^{-c_k\mu}\left[\Phi\left(\frac{cu_k}{\sqrt{t_i}}\right) - \Phi\left(\frac{c\ell_k}{\sqrt{t_i}}\right)\right] -
$$

$$4k\mu e^{-c_k\mu}\left[\frac{2k}{\sqrt{t_i}}\phi\left(\frac{cu_k}{\sqrt{t_i}}\right)-\phi\left(\frac{c\ell_k}{\sqrt{t_i}}\right)\frac{2k-1}{\sqrt{t_i}}\right]+$$
$$e^{-c_k\mu}\left[\frac{-4k^2}{t_i^{3/2}}\phi\left(\frac{cu_k}{\sqrt{t_i}}\right)(cu_k)+\frac{(2k-1)^2}{t_i^{3/2}}\phi\left(\frac{c\ell_k}{\sqrt{t_i}}\right)(c\ell_k)\right]-$$
$$4k^2\mu^2e^{d_k\mu}\left[\Phi\left(\frac{du_k}{\sqrt{t_i}}\right)-\Phi\left(\frac{d\ell_k}{\sqrt{t_i}}\right)\right]-$$
$$4k\mu e^{d_k\mu}\left[\frac{-2k}{\sqrt{t_i}}\phi\left(\frac{du_k}{\sqrt{t_i}}\right)-\frac{-2k-1}{\sqrt{t_i}}\phi\left(\frac{d\ell_k}{\sqrt{t_i}}\right)\right]-$$
$$e^{d_k\mu}\left[\frac{-4k^2}{t_i^{3/2}}\phi\left(\frac{du_k}{\sqrt{t_i}}\right)(du_k)+\phi\left(\frac{d\ell_k}{\sqrt{t_i}}\right)(d\ell_k)\frac{(-2k-1)^2}{t_i^{3/2}}\right]\Bigg\}$$

$$HF_4(u,\ell,\mu,t)=\Bigg\{-2ke^{-c_k\mu}\left[\Phi\left(\frac{cu_k}{\sqrt{t_i}}\right)-\Phi\left(\frac{c\ell_k}{\sqrt{t_i}}\right)\right]+$$
$$2k\mu c_ke^{-c_k\mu}\left[\Phi\left(\frac{cu_k}{\sqrt{t_i}}\right)-\Phi\left(\frac{c_k-\ell-\mu t_i}{\sqrt{t_i}}\right)\right]-$$
$$2k\mu e^{-c_k\mu}\left[-\phi\left(\frac{cu_k}{\sqrt{t_i}}\right)\sqrt{t_i}+\phi\left(\frac{c\ell_k}{\sqrt{t_i}}\right)\sqrt{t_i}\right]-$$
$$c_ke^{-c_k\mu}\left[\frac{2k+1}{\sqrt{t_i}}\phi\left(\frac{cu_k}{\sqrt{t_i}}\right)-\frac{2k}{\sqrt{t_i}}\phi\left(\frac{c\ell_k}{\sqrt{t_i}}\right)\right]+$$
$$e^{-c_k\mu}\left[\frac{2k+1}{\sqrt{t_i}}\phi\left(\frac{cu_k}{\sqrt{t_i}}\right)(cu_k)-\frac{2k}{\sqrt{t_i}}\phi\left(\frac{c\ell_k}{\sqrt{t_i}}\right)(c\ell_k)\right]-$$
$$(2k+2)e^{d_k\mu}\left[\Phi\left(\frac{du_k}{\sqrt{t_i}}\right)-\Phi\left(\frac{d\ell_k}{\sqrt{t_i}}\right)\right]-$$
$$(2k+2)\mu d_ke^{d_k\mu}\left[\Phi\left(\frac{du_k}{\sqrt{t_i}}\right)-\Phi\left(\frac{d\ell_k}{\sqrt{t_i}}\right)\right]-$$
$$(2k+2)\mu e^{d_k\mu}\left[-\phi\left(\frac{du_k}{\sqrt{t_i}}\right)\sqrt{t_i}+\phi\left(\frac{d\ell_k}{\sqrt{t_i}}\right)\sqrt{t_i}\right]-$$
$$d_ke^{d_k\mu}\left[\frac{-2k-1}{\sqrt{t_i}}\phi\left(\frac{du_k}{\sqrt{t_i}}\right)+\frac{2k+2}{\sqrt{t_i}}\phi\left(\frac{d\ell_k}{\sqrt{t_i}}\right)\right]-$$
$$e^{d_k\mu}\left[\frac{-2k-1}{\sqrt{t_i}}\phi\left(\frac{du_k}{\sqrt{t_i}}\right)(du_k)+\frac{2k+2}{\sqrt{t_i}}\phi\left(\frac{d\ell_k}{\sqrt{t_i}}\right)(d\ell_k)\right]\Bigg\}$$

$$
\begin{aligned}
HF_5(u,\ell,\mu,t) = \Bigg\{ & -2ke^{-c_k\mu}\left[\Phi\left(\frac{cu_k}{\sqrt{t_i}}\right)-\Phi\left(\frac{c\ell_k}{\sqrt{t_i}}\right)\right]+ \\
& 2k\mu c_k e^{-c_k\mu}\left[\Phi\left(\frac{cu_k}{\sqrt{t_i}}\right)-\Phi\left(\frac{c\ell_k}{\sqrt{t_i}}\right)\right]- \\
& 2k\mu e^{-c_k\mu}\left[-\phi\left(\frac{cu_k}{\sqrt{t_i}}\right)\sqrt{t_i}+\phi\left(\frac{c\ell_k}{\sqrt{t_i}}\right)\sqrt{t_i}\right]- \\
& c_k e^{-c_k\mu}\left[\frac{2k}{\sqrt{t_i}}\phi\left(\frac{cu_k}{\sqrt{t_i}}\right)-\frac{2k-1}{\sqrt{t_i}}\phi\left(\frac{c\ell_k}{\sqrt{t_i}}\right)\right]+ \\
& e^{-c_k\mu}\left[\frac{2k}{\sqrt{t_i}}\phi\left(\frac{cu_k}{\sqrt{t_i}}\right)(cu_k)-\frac{2k-1}{\sqrt{t_i}}\phi\left(\frac{c\ell_k}{\sqrt{t_i}}\right)(c\ell_k)\right]- \\
& 2ke^{d_k\mu}\left[\Phi\left(\frac{du_k}{\sqrt{t_i}}\right)-\Phi\left(\frac{d\ell_k}{\sqrt{t_i}}\right)\right]- \\
& 2k\mu d_k e^{d_k\mu}\left[\Phi\left(\frac{du_k}{\sqrt{t_i}}\right)-\Phi\left(\frac{d\ell_k}{\sqrt{t_i}}\right)\right] \\
& -2k\mu e^{d_k\mu}\left[-\phi\left(\frac{du_k}{\sqrt{t_i}}\right)\sqrt{t_i}+\phi\left(\frac{d\ell_k}{\sqrt{t_i}}\right)\sqrt{t_i}\right]- \\
& d_k e^{d_k\mu}\left[\sqrt{-2k}\sqrt{t_i}\phi\left(\frac{du_k}{\sqrt{t_i}}\right)-\frac{-2k-1}{\sqrt{t_i}}\phi\left(\frac{d\ell_k}{\sqrt{t_i}}\right)\right]- \\
& e^{d_k\mu}\left[\frac{-2k}{\sqrt{t_i}}\phi\left(\frac{du_k}{\sqrt{t_i}}\right)(du_k)-\frac{-2k-1}{\sqrt{t_i}}\phi\left(\frac{d\ell_k}{\sqrt{t_i}}\right)(d\ell_k)\right]\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
HF_6(u,\ell,\mu,t) = \Bigg\{ & c_k^2 e^{-c_k\mu}\left[\Phi\left(\frac{cu_k}{\sqrt{t_i}}\right)-\Phi\left(\frac{c\ell_k}{\sqrt{t_i}}\right)\right]- \\
& 2c_k e^{-c_k\mu}\left[-\phi\left(\frac{cu_k}{\sqrt{t_i}}\right)\sqrt{t_i}+\phi\left(\frac{c\ell_k}{\sqrt{t_i}}\right)\sqrt{t_i}\right]+ \\
& e^{-c_k\mu}\left[-\phi\left(\frac{cu_k}{\sqrt{t_i}}\right)(cu_k)\sqrt{t_i}+\phi\left(\frac{c\ell_k}{\sqrt{t_i}}\right)(c\ell_k)\sqrt{t_i}\right]- \\
& d_k^2 e^{d_k\mu}\left[\Phi\left(\frac{du_k}{\sqrt{t_i}}\right)-\Phi\left(\frac{d\ell_k}{\sqrt{t_i}}\right)\right]- \\
& 2d_k e^{d_k\mu}\left[-\phi\left(\frac{du_k}{\sqrt{t_i}}\right)\sqrt{t_i}+\phi\left(\frac{d\ell_k}{\sqrt{t_i}}\right)\sqrt{t_i}\right]- \\
& e^{d_k\mu}\left[-\phi\left(\frac{du_k}{\sqrt{t_i}}\right)(du_k)\sqrt{t_i}+\phi\left(\frac{d\ell_k}{\sqrt{t_i}}\right)(d\ell_k)\sqrt{t_i}\right]\Bigg\}
\end{aligned}
$$

Then the elements of $D2Fu$ are

$$
\begin{aligned}
e_{1,1} &= \frac{q}{a}HF_1(u,\ell,\mu,t_i) \\
e_{2,1} &= \frac{q}{a}HF_2(u,\ell,\mu,t_i) \\
e_{2,2} &= \frac{q}{a}HF_3(u,\ell,\mu,t_i)
\end{aligned}
$$

$$
\begin{aligned}
e_{3,1} &= \frac{q}{a} HF_4(u, \ell, \mu, t_i) \\
e_{3,2} &= \frac{q}{a} HF_5(u, \ell, \mu, t_i) \\
e_{3,3} &= \frac{q}{a} HF_6(u, \ell, \mu, t_i) \\
e_{4,1} &= e_{4,2} = e_{4,3} = e_{4,4} = 0 \\
e_{5,1} &= \frac{-GF_1(u, \ell, \mu, t_i)}{a} \\
e_{5,2} &= \frac{-GF_2(u, \ell, \mu, t_i)}{a} \\
e_{5,3} &= \frac{-GF_3(u, \ell, \mu, t_i)}{a} \\
e_{5,4} &= = e_{5,5} = 0
\end{aligned}
$$

where $b = 1 - qF_{u,\ell}(t_i) - pFw, \ell(t_i)$. Then the unique elements of $D2Fw$ are

$$
\begin{aligned}
f_{1,1} &= f_{2,1} = f_{3,1} = f_{4,1} = f_{5,1} = 0 \\
f_{2,2} &= \frac{p}{a} HF_3(w, \ell, \mu, t_i) \\
f_{3,2} &= \frac{p}{a} HF_5(w, \ell, \mu, t_i) \\
f_{3,3} &= \frac{p}{a} HF_6(w, \ell, \mu, t_i) \\
f_{4,2} &= \frac{p}{a} HF_2(w, \ell, \mu, t_i) \\
f_{4,3} &= \frac{p}{a} HF_4(w, \ell, \mu, t_i) \\
f_{4,4} &= \frac{p}{a} HF_1(w, \ell, \mu, t_i) \\
f_{5,2} &= \frac{GF_2(w, \ell, \mu, t_i)}{a} \\
f_{5,3} &= \frac{GF_3(w, \ell, \mu, t_i)}{a} \\
f_{5,4} &= \frac{GF_1(w, \ell, \mu, t_i)}{a} \\
f_{5,5} &= 0
\end{aligned}
$$

Appendix F

Identifiability

In this appendix we show that it is not possible to estimate all four parameters u, ℓ, μ and σ^2 because of identifiability problems. Let $\theta = (u, \ell, \mu, \sigma^2)$. From equation C.11, the sub-density corresponding to the event that a Wiener process with drift μ and volatility σ^2 hits the lower barrier at $-\ell < 0$ at time t without breaching the upper barrier at $u > 0$ is

$$f_\ell(t;\theta) = -\frac{1}{\sqrt{2\pi\sigma^2 t^3}}\, e^{-\frac{(\mu t + 2\ell)\mu}{2\sigma^2}} \sum_{k=-\infty}^{\infty} (s_k + u)\, e^{-\frac{(s_k+u)^2}{2\sigma^2 t}}$$

where $s_k = -(2k+1)(u+\ell)$. Similarly, the sub-density corresponding to the event that the process hits the upper barrier at time t without breaching the lower barrier is

$$f_u(t;\theta) = -\frac{1}{\sqrt{2\pi\sigma^2 t^3}}\, e^{-\frac{(\mu t - 2u)\mu}{2\sigma^2}} \sum_{k=-\infty}^{\infty} (s_k + \ell)\, e^{-\frac{(s_k+\ell)^2}{2\sigma^2 t}}.$$

If no observations are censored, the joint frequency function of the observations is

$$\mathcal{L}(\theta; t) = \prod_{\{i:D_i=u\}} f_u(t_i;\theta) \prod_{\{i:D_i=\ell\}} f_\ell(t_i;\theta).$$

Let

$$\begin{aligned} \theta_1 &= (\ell, u, \mu, \sigma^2) \\ \theta_2 &= (m\ell, mu, m\mu, (m\sigma)^2) \end{aligned}$$

where m is any real number. Then it's easy to see that

$$f_\ell(t;\theta_1) = f_\ell(t;\theta_2)$$

and

$$f_u(t;\theta_1) = f_u(t;\theta_2)$$

and so

$$\mathcal{L}(\theta_1;t) = \mathcal{L}(\theta_2;t).$$

Thus we have shown that distinct values of the parameter vector, $\theta_1 \neq \theta_2$, can give rise to exactly the same frequency function for the data.

Note that this result holds for the zero drift model where μ is fixed at 0, as well as the more general drift model.

Bibliography

[1] Aalen, O. (1994). Effects of frailty in survival analysis. *Statistics in Medicine*, **3**, 227-243.

[2] Apostol, T.M. (1974). *Mathematical Analysis.* London: Addison-Wesley.

[3] Baltazar-Aban I. and Pena, E.A. (1995). Properties of Hazard-Based Residuals and Implications in Model Diagnostics. *Journal of the American Statistical Association*, **90**, 185-197.

[4] Bardsley, M. Coles, J. and Jenkins, L. (ed.) (1989) *DRG's and health care: The management of case mix* London: King Edward's Hospital Fund.

[5] Bonuck, K.A. and Arno, P.S. (1997). Social and Medical Factors Affecting Hospital Discharge of Persons with HIV/AIDS. *Journal of Community Health*, **4**, 225-232.

[6] Bressoud, D. (1994). *A Radical Approach to Real Analysis.* Mathematical Association of America.

[7] Brewster A.C., Karlin, B. G., Hyde, L,A., Jacobs, C.M. (1984). MEDISGRPS: A clinically based approach to classifying hospital patients at admission. *Inquiry*, **23**, 377-387.

[8] Charlson, M.E., Pompei, P., Ales, K.L., MacKenzie, C.R. (1987). A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation. *J. Chron Dis*, **40**, 373-383.

[9] Chhikara, R. S. and Folks, J. L. (1989). *The Inverse Gaussian Distribution* New York: Marcel Dekker, Inc.

[10] Chung, Kai Lai. (1960) *Markov Chains with Stationary Transition Probabilities* Berlin: Springer-Verlag.

[11] Cox, D. R. and Oakes, D. (1984). *The Analysis of Survival Data.* London: Chapman and Hall.

[12] Cox, D. R. and Snell, E. J. (1968). A General Definition of Residuals. *J. R. Statist. Soc. B*, **45**, 248-75.

[13] David, H. A. and Moeschberger, M. L. (1978). *The Theory of Competing Risks.* England: Charles Griffin and Company.

[14] DuMouchel, W. H. (1973). On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distribution. *Annals of Statistics*, **1**, 948-957.

[15] Eaton, W. W. and Whitmore, G. A. (1997). Length of stay as a stochastic process: A general approach and application to hospitalization for schizophrenia. *Journal of Mathematical Sociology*, **5**, 273-292.

[16] Feller, W. (1951). The asymptotic distribution of the range of sums of independent random variables. *Annals of Mathematical Statistics*, **22**, 427-432.

[17] Feller, W. (1986). *An Introduction to Probability Theory and its Applications, Volume 1* New York: John Wiley and Sons.

[18] Feller, W. (1986). *An Introduction to Probability Theory and its Applications, Volume 2* New York: John Wiley and Sons.

[19] Gentleman, R. and Crowley, J. (1991). Graphical Methods for Censored Data. *Journal of the American Statistical Association*, **86**, 678-683.

[20] Gnedenko, B. V. and Kolmogorov, A. N. (1954). Limit Distributions for Sums of Independent Random Variables. Reading: Addison-Wesley Publishing Company, Inc.

[21] Kalbfleisch, J.G. *Probability and Statistical Inference, Volume 2* New York: Springer-Verlag.

[22] Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data.* New York: John Wiley and Sons.

[23] Karlin, S. and Taylor, H. M. (1975). *A First Course in Stochastic Processes.* New York: Academic Press.

[24] Klein, J. P. and Moeschberger, M. L. (1997). *Survival Analysis.* New York: Springer Verlag.

[25] Lawless, J.F. (1982). *Statistical Models and Methods for Lifetime Data.* New York: John Wiley and Sons.

[26] Liu, K., Coughlin, T. and McBride, T. (1991). Predicting nursing-home admission and length of stay; A duration analysis. *Medical Care*, **29**, 125-141.

[27] Malenka, D. J., McLerran, D., Roos, N., Fisher, E.S., Wennberg, J.E. (1994). Using Administrative Data to Describe Casemix: A comparison with the Medical Record. *J Clin Epidemiol* **47**, 1027-1032.

[28] Morris C. N., Norton E. C. and Zhou X. H. (1994). Parametric analysis of nursing home usage. In *Case Studies in Biometry.* ed. Lange, N., Ryan, L., Billard, L., Brillinger, D., Conquest, L. and Greenhouse, J. New York: John Wiley and Sons.

[29] Nadas, A. (1970). On estimating the distribution of a random vector when only the smallest co-ordinate is observable. *Technometrics* **12** 923-924.

[30] Nadas, A. (1971). The distribution of the identified minimum of a normal pair determines the distribution of the pair. *Technometrics* **13** 201-202.

[31] Neuhaus J. M., Kalbfleisch J.D., Hauck W.W. (1991). A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *Int Stat Rev* **59** 25-36.

[32] Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Flournoy, N., Farewell, V. T. and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, **34**, 541-554.

[33] Redekop, J. (1995). *Extreme-value distributions for generalizations of Brownian motion.* Unpublished PhD Thesis. University of Waterloo.

[34] Rosenthal, G. E., Harper, D. O., Quinn, L. M. and Cooper, G. S. (Aug. 13, 1997). Severity-adjusted mortality and length of stay in teaching and nonteaching hospitals; results of a regional study. *Journal of the American Medical Association*, **278**, 485-490.

[35] Sacher, G. A. and Trucco, E. (1962). The stochastic theory of mortality. *Annals New York Academy of Sciences*, **96**, 985-1007.

[36] Sen, P. K., and Singer, J. M. (1993). *Large Sample Methods in Statistics*. New York: Chapman and Hall.

[37] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. New York: John Wiley and Sons.

[38] Seshadri, V. (1993). *The inverse Gaussian distribution : a case study in exponential families*. Oxford: Clarendon Press.

[39] Utah Health Data Committee/Office of Health Data Analysis, Utah Department of Health. (1997). *Utah Hospital Inpatient Discharge Data File (1996)*. Salt Lake City: Utah Department of Health.

[40] Utah Health Data Committee/Office of Health Data Analysis, Utah Department of Health. (1997). *Utah Hospital Discharge Database 1996, Public-Use Data File, User's Manual, Version 1*. Salt Lake City: Utah Department of Health.

[41] Vaupel, J.W. and Yashin, A.I. (1985). Heterogeneity's ruses: Some surprising effects of selection on population dynamics. *American Statistician*, **39**, 176-185.

[42] Whitmore, G. A. (1979). An inverse Gaussian model for labour turnover. *Journal of the Royal Statistical Society, Series A*, **142**, 468-478.

[43] Whitmore, G. A. (1983). A regression method for censored inverse Gaussian data. *Canadian Journal of Statistics*, 11, 305-315.

[44] Whitmore, G. A., Crowder, M. J., and Lawless, J. F. (1998). Failure Inference From a Marker Process Based on a Bivariate Wiener Model. *Lifetime Data Analysis* 4 229-251.