Title: A General Class of Additive Parametric Models for Recurrent Event Data

Author: Toby L. Trotta

Thesis Chair: Dr. Russell S. Stocker

Thesis Committee Members: Dr. Yongtao Cao
Dr. Mavis Pararai

We propose a general class of additive parametric models to analyze recurrent event data that uses the effective age process given in Peña and Hollander. We assume a negligible intervention effect is applied to each unit after an observed event which may be classified as perfect, minimal, or partial. Estimators are derived and their asymptotic properties are established under regularity conditions that are formulated using doubly-indexed processes. We investigate the finite sample properties via a simulation study in RStudio. The simulation study shows minimal and decreasing bias in the maximum likelihood estimators, and that the standard errors can be approximated using the inverse observed information matrix. The model and asymptotic properties are used to examine the indolent non-Hodgkin's lymphoma data set from Gonzalez, Peña, and Slate.

CHAPTER 1

INTRODUCTION

Time-to-event analysis is a branch of statistics concerned with the time until a unit experiences an event (failure). Clinicians interested in mortality rates served as the primary application until recent developments have shown useful to disciplines including engineering, social sciences, and economics. For example, engineers may be interested in analyzing the time until a machine malfunctions while hiring managers may be interested in the time until a prospective employee accepts (or declines) an employment offer.

For studies involving the risk of experiencing multiple events of the same nature within the allotted study time, data is classified as "recurrent." Examples in the biomedical sciences include repeat hospitalizations after major surgery, chronic illness flare-ups, or drug and alcohol relapse. The occurrences of protests or social movements after a political event or the number of career transitions within the first five years after graduation may be of interest to sociologists. Engineers may use recurrent event analysis as a maintenance scheduling tool or for environmental monitoring as it pertains to polluting incidents, for example. Given its vast applications across a variety of disciplines, appropriate recurrent event models must be developed.

The primary difficulty researchers experience when analyzing time-to-event data is censoring. A unit is censored when, under any circumstance, the exact time until an event of interest occurs is unobserved. The difficulty this provides is that information regarding the unit's survival time is only partially known. While right-censored data is most common, left- and interval-censored data has been greatly discussed in the literature. Studies prone to right-censoring include analyzing the time until a customer submits a five-year limited warranty claim, a patient's time until discharge within 30 days, or the time until a side effect occurs during a clinical trial, all of which may not occur during the study.

1.1 Literature Review

Various models have been proposed to analyze event times and treatment effects for time-to-event data. The Cox proportional hazards model (2.1) has been widely adopted in epidemiological and clinical research settings due to its handling of censored observations, used to investigate the association between the covariates and event times (see [4] and [23]). Cox considered the partial likelihood approach to eliminate the baseline hazard from the score function for some model regression parameter β. The Andersen-Gill Model [19], an extension of the Cox proportional hazards model, introduces counting and at-risk processes to indicate how many observed events have occurred for the unit and whether the unit is at risk of an observed event at time s, respectively. The Cox proportional hazards model and the Andersen-Gill model are constructed having multiplicative link function of possibly time-dependent covariates. While both the Cox proportional hazards model and Andersen-Gill model assume observations are independent and that censoring is non-informative, the former also assumes the instantaneous relative hazard of experiencing an event between two groups remains proportional over time. When this assumption is violated, an additive hazards model may be appropriate.

Models like Aalen [17] and Lin and Ying [3] incorporate link functions that act additively on the hazard. Aalen [17] established the precedent for non-parametric additive models if, for example, the Cox model assumption of constant hazard over time is violated. Lin and Ying [3] expanded the discussion on Aalen's additive hazard model.

1.2 Recurrent Event Models

The aforementioned models have been further developed to accommodate studies that involve an intervention or repair process after an observed event. Cook and Lawless [1] cover many of the models discussed below and several others recurrent event models in the literature. Repair models by Brown and Proschan [14] described implementing a repair process, where interventions return "perfectly repaired" units with probability p and "minimally repaired" units with probability $1 - p$. In implementing a repair model, we introduce the concept of a unit's effective age which is used to describe the operating age of the unit. The effective age process considers a sequence of event times $0 \equiv S_{i0} < S_{i1} < \dots$ with associated effective age functions $\mathcal{E}_{i1}(s), \mathcal{E}_{i2}(s), \dots$ that may be dependent on $S_{i0}, S_{i1}, \dots$, and are used to capture the influence of previous interventions and account for dependency in gap times [5].

The Peña and Hollander [6] model provides the framework for the effective age process necessary to model recurrent events. Under the Peña and Hollander model, it requires a class of predictable processes $\{\mathcal{E}_i(s) \,|\, 0 \le s \le s \wedge \tau_i\}$, further defined in Chapter 3.

Block, et al. [7] extended the Brown and Proschan model by introducing time-dependent perfect and minimal repair probabilities $p(s)$ and $1 - p(s)$, respectively. In the event a minimal repair occurs after a failure at time t, regardless of the time-dependency of repair probabilities, the minimally repaired item has derived survival probability (1.1):

$$S(s \,|\, t) = \frac{S(s+t)}{S(t)}, \quad s \ge 0. \tag{1.1}$$

Shaked and Shanthikumar [15] extended the Brown and Proschan model to account for multivariate data, first assuming no more than one component can fail at a time, and a second allowing for multiple failures at a time. Whitaker and Samaniego [12] proposed several inference methods under the same framework as Brown and Proschan [14] and discussed the estimation procedures for common lifetime distributions on system lifetimes between successive repairs.

These recurrent event models operate under the assumption that the duration of intervention effects is negligible. Iyer [24] assumed the repair processes take time, useful for reliability analysis in system maintenance and engineering studies. These repair models are another important aspect of reliability theory and have been found useful in many industrial applications (see Brito, Tomazella, and Ferreira [26]).

The literature provides several models discussing perfect, minimal, or partial repairs models with Last and Szekli [13] proposing a model to include destructive intervention effects. In this case, the effective age model at the j^{th} event is defined as

$$A_0 = 0, A_j = (1 - p_j)(A_{j-1} + T_j) \text{ for } j \geq 1$$

where $p_j \leq 1$ is the repair probability, unbounded below for destructive interventions with respect to the gap time, the time between recurring events, $T_j = S_{ij} - S_{ij-1}$.

The researcher must implement an appropriate effective age process for the event of interest, otherwise, any potential intervention processes may overestimate or underestimate the true operating age of the unit. The effective age process implementation and its formal definition will be discussed in Chapter 3.

1.3 Goals of the Thesis

In this thesis, we investigate a class of additive parametric models that use an effective age process provided in the Peña and Hollander [6] models. Our model is an adaptation of the Lin and Ying model (2.4) under a fully parametric baseline hazard function with strictly time-independent covariate values. We begin our discussion by providing the mathematical setting provided by Peña and Hollander [6]. We briefly discuss how the model can be transformed into functions of gap times by the doubly-indexed processes proposed by Stocker and Adekpedjou [22]. We observe that the variance-covariance matrix can be estimated by the inverse of the observed information matrix, in which the properties of the sampling distribution will be examined using RStudio. Finally, we apply the proposed class of models to an indolent non-Hodgkin's lymphoma data set from Gonzalez, Peña, and Slate [8].