Working Title

First Author Second Author

January 30, 2018

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

sec:intro

In this thesis, we work with boosting for regression in the first-hitting-time model. First-hitting-time is a model in survival analysis which serves as an alternative to the proportional hazards model, typically known as Cox regression. Developments in FHT regression are relatively recent, and there has to our knowledge been no attempt at tackling it in the high-dimensional case, in which boosting is an appropriate choice of method.

PART I

The First Part

CHAPTER 2

First hitting time regression models

sec:fht

2.1 First hitting time models

An FHT model has two main components.

1. A stochastic process $\{Y(t), t \in \mathcal{T}, y \in \mathcal{Y}\}$, with $Y(0) = y_0$.
2. A boundary, or threshold, $B \subset \mathcal{Y}$, where $y_0 \notin \mathcal{B}$

The first hitting time is the first time the process reaches the boundary set. Formally, the FHT is a stochastic variable S, which is defined as

$$S = \inf\{t \colon Y(t) \in \mathcal{B}\}$$

Note that it is possible that $P(S < \infty) < 1$.

This framework makes for a flexible model, due to many possible choices for the process and the boundary. We now look at a common choice of the process.

Wiener process

wiener

The Wiener process, also known as the standard Brownian motion process, is a process which is continuous in time and space, and has the properties (Caroni 2017, p. 61) that

- $Y(t)$ has independent increments, such that $Y(t_2)-Y(t_1)$ and $Y(t_4)-Y(t_3)$ are independent for any disjoint intervals, and
- for any interval (t_1, t_2),
$$Y(t_2) - Y(t-1) \sim N(\mu(t_2 - t_1), \sigma^2(t_2 - t_1)).$$

This process may both increase and decrease. However, if we want a monotonic restriction on the movement of the process, we may use a gamma process.

Gamma process

The gamma process is suitable for modelling something which we would require to be monotonic, typically a physical degradation, i.e. where the damage cannot mend itself. (Caroni 2017, p. 59)

Other choices of processes include Markov chain state models, the Bernoulli process, and the Ornstein-Uhlenbeck process.

2.2 First hitting time regression based on underlying Wiener process

The first hitting time of the Wiener process (section 2.1) process follows an inverse Gaussian distribution (derivation in Chhikara 1988, pp. 23-29):

$$f(t|y_0, \mu, \sigma^2) = \frac{y_0}{\sqrt{2\pi\sigma^2 t^3}} \exp\left[-\frac{(y_0 + \mu t)^2}{2\sigma^2 t}\right] \tag{2.1}$$

{eq:fht-ig}

If μ is positive, $Y(t) \leq 0$ is not certain to occur. Note also that this parameterization is over-parameterized, because Y has an arbitrary scale, so we can without loss of generality let $\sigma^2 = 1$.

While μ and y_0 have simple interpretations in terms of the underlying process, they do not in terms of the lifetime distribution. The mean lifetime is $\frac{y_0}{|\mu|}$, and the variance is $\frac{y_0}{|\mu|^3}$. (Caroni 2017, p. 62.)

We may introduce effects from covariates by allowing μ and y_0 to depend on covariates $\mathbf{x}$. Suitable models are

$$\mu = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{u} \tag{2.2}$$

$$\ln y_0 = \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{v} \tag{2.3}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are vectors of regression coefficients. $\mathbf{u}$ and $\mathbf{v}$ may be either completely different, partially different, or equal. In the latter case, we use $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$ and $\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{x}$ instead, for clarity.

Appendices

Bibliography

caroni2017

caroni2017 Caroni, C. "First Hitting Time Regression Models". In: *First Hitting Time Regression Models*. John Wiley & Sons, Inc., 2017, pp. 55–79.

chhikara1988

chhikara1988 Chhikara, R. *The Inverse Gaussian Distribution: Theory: Methodology, and Applications*. Statistics: A Series of Textbooks and Monographs. Taylor & Francis, 1988.