Boosting the First-Hitting-Time Regression Model

Vegard Stikbakke

February 15, 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.

CHAPTER 2

First hitting time regression models

2.1 Survival analysis and time-to-event models

sec:survival

In many fields, it is interesting to consider the lifetime of some entity. A lifetime ends when some event occurs. We are then interested in inferring things about this lifetime, and what it depends upon. In medical fields, this is the field of survival analysis, while in engineering fields, the field is called reliability analysis. In the first case, we consider e.g. the lifetime of patients with some chronic disease, or the length of a hospital stay after some treatment. In the latter, we consider e.g. the time before some component of a system breaks and must be replaced.

The time-to-event T is a continuous, non-negative random variable $T \sim f(t)$, $t > 0$, for some probability density function f. We are particularly interested in two things related to T:

1. The survival function $S(t)$ – the probability of an individual having survived until time t. Note that $S(t) = 1 - F(t)$, where F is the cumulative density function of f.

2. The hazard function $h(t)$ – the probability of the event happening during some (small) interval. Note that this is conditional on surviving until time t, and is defined as $h(t) = \frac{f(t)}{S(t)}$.

Regression

sec:surv-reg

To find out anything interesting, we need to be able to do regression on covariates. Given a sample of n independent observations $\{t_i, \mathbf{x}_i, \delta_i, i = 1, \cdots, n\}$, where individual i has covariates $\mathbf{x}_i$, lifetime t_i and censoring indicator $delta_i$. From Caroni 2017, p. 10, the likelihood is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(t_i|\mathbf{x}_i, \boldsymbol{\theta})^{\delta_i} S(t_i|\mathbf{x}_i, \boldsymbol{\theta})^{1-\delta_i} \tag{2.1}$$

{eq:surv-lik}

Proportional hazards

The most used method for doing regression on survival data is the Cox proportional hazards (PH) regression. It is based on an assumption that is often called the PH property or the PH assumption, namely that

$$h(t|x) = h_0(t)g(\mathbf{x}). \tag{2.2}$$

This property states that at any two time points t_1 and t_2, the ratio between the hazard functions of any two $\mathbf{x}_1$ and $\mathbf{x}_2$ will be the same:

$$\frac{h(t_1|x_1)}{h(t_1|x_2)} = \frac{h(t_2|x_1)}{h(t_2|x_2)} \tag{2.3}$$

This is a large assumption to make, and it will rarely be the case in practice. (Lee and Whitmore 2010) However, many times Cox regression will work well in practice.

2.2 The first hitting time model

sec:fht

Revisiting the examples of the two lifetime settings, it may in both cases be natural to imagine that the event happens as a process reaches a threshold. Then one way to model the time-to-event is to model the process itself, and look at the time it takes for the process to reach this threshold, at which point the event we are interested in is triggered.

Excellent reviews of the FHT are Lee and Whitmore 2006 and Caroni 2017.

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 set, $B \subset \mathcal{Y}$, 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 $P(S < \infty) < 1$.

Typically, one will consider a process with boundary $B = 0$. The event then occurs if and when the process $\{Y\}$ reaches 0 at $y(T)$.

This framework makes for a flexible model, due to many possible choices for the process and the boundary.

The FHT model is attractive because it is a good model for what happens in the underlying process. Additionally, it is more flexible than the PH model. In fact, the PH model may be obtained by constructing the FHT model in a specific way. (Lee and Whitmore 2010.)

As mentioned, the most widely used models in survival analysis are proportional hazards models. These models consider hazard functions over time. Hazard functions say something about the chance of an event happening during some (small) interval.

We now look at a common choice of the process.

Wiener process

sec: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 a process which we would require to be monotonic, typically a physical degradation, i.e. where the damage cannot mend itself, unlike a patient's health. The first-hitting-time that arises from the gamma process is inverse gamma. (Lee and Whitmore 2006, p. 503.)

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

2.3 First hitting time regression based on underlying Wiener process

The first hitting time of the Wiener process (section 2.2) follows an inverse Gaussian distribution (derivation in Chhikara 1988, pp. 23-29):

also derive more clearly in appendix?

$$f(t|y_0, \mu, \sigma^2) = \frac{y_0}{\sqrt{2\pi\sigma^2 t^3}} \exp\left[-\frac{(y_0 + \mu t)^2}{2\sigma^2 t}\right] \tag{2.4}$$

{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.)

The cumulative distribution function of the FHT is (from Xiao et al. 2015, p. 7)

$$F(t|\mu, \sigma^2, y_0) = \Phi\left[(-\frac{y_0 + \mu t)}{\sqrt{\sigma^2 t}}\right] + \exp\left(-\frac{2y_0\mu}{\sigma^2}\right)\Phi\left[\frac{\mu t - y_0}{\sqrt{\sigma^2 t}}\right], \tag{2.5}$$

{eq:cumulative}

where $\Phi(x)$ is the cumulative of the standard normal, i.e.,

$$\Phi(x) = \int_{-\infty}^{x} \exp(-y^2/2)/\sqrt{2\pi}\, \mathrm{d}y. \tag{2.6}$$

Regression

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

$$\begin{aligned} \mu &= \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x} \\ \ln y_0 &= \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{x} \end{aligned} \tag{2.7}$$

{eq:coeffs}

where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are vectors of regression coefficients. If we deem suitable, we let elements in $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ be zero, e.g. if we think some elements in $\mathbf{x}$ are relevant for y_0 but not μ or vice versa.

2.4 Likelihood

sec:lik

In section 2.1, we stated the likelihood of lifetime regression models in (2.1). For an inverse gaussian FHT this then becomes (inserting (2.4) and (2.5) into (2.1), and since $S(t) = 1 - F(t)$)

$$\begin{aligned} L(\boldsymbol{\theta}) = &\left(\frac{y_0}{\sqrt{2\pi\sigma^2 t^3}} \exp\left[-\frac{(y_0 + \mu t)^2}{2\sigma^2 t} \right] \right)^{\delta_i} \\ &\times \left[1 - \Phi\left(-\frac{y_0 + \mu t}{\sqrt{\sigma^2 t}} \right) - \exp\left(-\frac{2y_0\mu}{\sigma^2} \right) \Phi\left(\frac{\mu t - y_0}{\sqrt{\sigma^2 t}} \right) \right]^{1-\delta_i} \end{aligned} \tag{2.8}$$

{eq:fht-lik}

Since we let $\sigma^2 = 1$, this simplifies to

$$\begin{aligned} L(\boldsymbol{\theta}) = &\left(\frac{y_0}{\sqrt{2\pi t^3}} \exp\left[-\frac{(y_0 + \mu t)^2}{2t} \right] \right)^{\delta_i} \\ &\times \left[1 - \Phi\left(-\frac{y_0 + \mu t}{\sqrt{t}} \right) - \exp(-2y_0\mu) \Phi\left(\frac{\mu t - y_0}{\sqrt{t}} \right) \right]^{1-\delta_i} \end{aligned} \tag{2.9}$$

We can now substitute (2.7) into this. This is often fitted with numerical maximum likelihood methods.

2.5 The `threg` package

There exists an R package `threg` for fitting regression with inverse gaussian FHT, described in Xiao et al. 2015. We provide a small example here, which is the one described in the help pages of the package.

```
library(threg)
data("lkr")
lkr$f.treatment2=factor(lkr$treatment2)
# head(lkr)
fit <- threg(Surv(weeks, relapse) ~ f.treatment2|f.treatment2, data=lkr)
fit
```

Which provides the following output

```
Call:
threg(formula = Surv(weeks, relapse) ~ f.treatment2 | f.treatment2,
    data = lkr)


                          coef  se(coef)         z       p
lny0: (Intercept)    2.0097844 0.1705141 11.786620 0.0e+00
lny0: f.treatment21 -1.2739233 0.2441633 -5.217504 1.8e-07
  mu: (Intercept)   -0.5886165 0.1340126 -4.392246 1.1e-05
  mu: f.treatment21  0.5888365 0.1535081  3.835866 1.3e-04

Log likelihood =-104.64, AIC =217.28
```

Here we fit an inverse gaussian FHT model where both y_0 and μ have an intercept term and an indicator term for treatment 2.

Appendices

Bibliography

caroni2017 [1] Caroni, C. *First Hitting Time Regression Models*. John Wiley & Sons, Inc., 2017.

chhikara1988 [2] Chhikara, R. *The Inverse Gaussian Distribution: Theory: Methodology, and Applications*. Statistics: A Series of Textbooks and Monographs. Taylor & Francis, 1988.

lee2010 [3] Lee, M.-L. T. and Whitmore, G. A. "Proportional hazards and threshold regression: their theoretical and practical connections". In: *Lifetime Data Analysis* 16.2 (Apr. 2010), pp. 196–214.

lee2006 [4] Lee, M.-L. T. and Whitmore, G. A. "Threshold Regression for Survival Analysis: Modeling Event Times by a Stochastic Process Reaching a Boundary". In: *Statist. Sci.* 21.4 (Nov. 2006), pp. 501–513.

threg [5] Xiao, T. et al. "The R Package threg to Implement Threshold Regression Models". In: *Journal of Statistical Software, Articles* 66.8 (2015), pp. 1–16.