STOCKHOLM UNIVERSITY
DEPARTMENT OF MATHEMATICS
Div. of Mathematical Statistics

Comparison of Methods to Estimate Hazard Rates

Leo G. Levenius

January 5, 2026

A central part of life insurance and survival analysis is to estimate the *hazard rate*

$$\alpha(t) \coloneqq \lim_{\Delta t \to 0} \frac{1}{\Delta t} \mathbb{P}(t \leq T < t + \Delta t \mid T \geq t), \tag{1}$$

where T is the time of the event, e.g., death. Given $\alpha(\cdot)$, we can derive many results like the *survival function*

$$S(t) \coloneqq \mathbb{P}(T > t) = \exp\left\{ \int_0^t \alpha(s)\, ds \right\} \tag{2}$$

and the expected time of death

$$\mathbb{E}[T] = \int_0^\infty S(t)\, dt. \tag{3}$$

An interesting idea for a thesis would be to compare some different methods for estimating the hazard rate. A popular non-parametric model is the *Nelson–Aalen estimator* for the cumulative hazard rate $A(t) = \int_0^t \alpha(s)\, ds$,

$$\hat{A}(t) = \sum_{T_j \leq t} \frac{1}{Y(T_j)}, \tag{4}$$

where $Y(t)$ is the number of people alive, or *at risk*, just before t. Another is Sir David Cox's *proportional hazard model*,

$$\alpha(t) = \alpha_0(t) \exp\{\beta^\top \mathbf{x}\}, \tag{5}$$

where $\alpha_0(\cdot)$ is a baseline hazard rate and $\mathbf{x}$ is a covariate vector corresponding to characteristics influencing the hazard rate, e.g., male/female, smoker/non-smoker, diabetic/non-diabetic.

The methods would be applied to simulated data with known true hazard rates that the estimates can be compared to, where the true hazard rates can vary in complexity. It is also possible to use real data, but this will not be provided.

The project can accommodate large variations in difficulty and complexity, depending on the student's ambitions. An example would be to replace $\beta^\top \mathbf{x}$ in (5) with a *gradient boosting machine* (GBM) to capture non-linearity and inter-covariate dependencies. It could be useful if the student has taken the courses MT5011 and MT5012, but they are far from compulsory. The methods used are described in more detail in the graduate course MT7052 and book *Survival and Event History Analysis* (2008) by Aalen *et al.*