MACHINE LEARNING FOR HEALTH

Survival Analysis: The Role of Time in Clinical Prediction

Censoring, risk, longitudinal trajectories, and models that ask when

TECHNICAL HANDOUT · SESSION 11

Professor Flávio Luiz Seixas
Graduate Program in Computing

Academic teaching material based exclusively on the PDFs available in the project

Table of Contents

Introduction

When an outcome is the time to an event, reducing the problem to a binary class loses information and handles incomplete observations incorrectly. Survival analysis represents the event, time, and censoring jointly. [1]

This session presents Kaplan–Meier estimation, Cox regression, longitudinal data, and machine-learning extensions. The definitions are accompanied by the assumptions available in the local PDFs. [1]

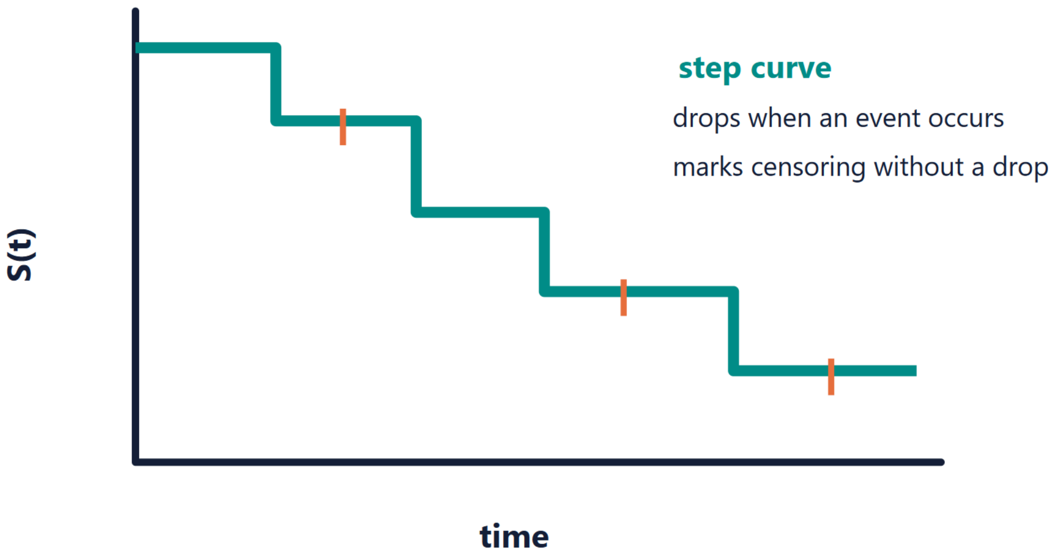

Figure 1 — Stepwise survival curve. Source: original illustration based on [1].

1. Event, time origin, and censoring

Survival analysis models the time from a defined starting point to an event such as recurrence, disease onset, or death. Time zero must be comparable across individuals, and the event needs an operational definition [1].

Censoring occurs when the exact event time is not observed during follow-up. A patient may leave the study, be lost to follow-up, or reach the end of the period without an event. A censored observation indicates that the individual remained event-free until a certain time; it is not equivalent to a definitive absence of the event [1].

KEY CONCEPT
Survival analysis models the time from a defined starting point to an event such as recurrence, disease onset, or death.

2. Kaplan–Meier and hazard

The Kaplan–Meier estimator calculates survival as a product of conditional probabilities across event times. Its step curve changes when events occur and keeps censored individuals in the risk set only while they remain observable [1].

The hazard function describes the occurrence of the event over time among those who remain at risk. Survival and hazard are related but are not interchangeable [1].

3. Cox regression

Cox regression relates covariates to hazard through hazard ratios while leaving the shape of the baseline hazard unspecified. The usual interpretation assumes proportionality: the relative effect of covariates remains constant over time. When this assumption fails, time-dependent effects or other models should be considered [1].

4. Longitudinal data, competing risks, and ML

Longitudinal data record repeated measurements and can be represented by timelines that locate events and measurements. The choice between continuous and discrete representations depends on the type and availability of the measurements [1].

Tree, forest, boosting, and neural-network algorithms have been adapted for censoring and time-to-event outcomes. Complex models can represent nonlinearities but still depend on temporal definitions, validation, and appropriate metrics [1].

The QMD calls for a discussion of competing risks. Because the available PDFs do not adequately support their modeling, this extension is not developed in the handout.

SOURCE LIMITATION
Competing risks and survival-specific metrics have limited coverage in the available PDFs.

Session summary

Event, time origin, and censoring: survival analysis models the time from a defined starting point to an event such as recurrence, disease onset, or death.
Kaplan–Meier and hazard: the kaplan–meier estimator calculates survival as a product of conditional probabilities across event times.
Cox regression: cox regression relates covariates to hazard through hazard ratios while leaving the shape of the baseline hazard unspecified.
Longitudinal data, competing risks, and ML: longitudinal data record repeated measurements and can be represented by timelines that locate events and measurements.

References

[1] Bellazzi, R.; Dagliati, A.; Nicora, G. Predicting Medical Outcomes. In: Intelligent Systems in Medicine and Health. Springer, 2022. External link

Local file: pdfs/3-Intelligent Systems/XI-Predicting Medical Outcomes.pdf · PDF pages used: 14, 15, 16, 17, 19, 25.