Graphical Tools for Censored Survival Data

Debasis Sengupta
Indian Statistical Institute, Calcutta
203, Barrackpore Trunk Road, Calcutta 700 035, India

Abstract

This review article discusses a number of graphical procedures for
Survival Analysis. These include descriptive plots such as Event his-
tory diagrams, scatterplots and plots of estimated curves. Diagnostic
plots for checking model assumptions and comparing two samples are
also discussed, with special reference to the Cox regression model. A
few specific suggestions have also been made for the modification of
some of these plots for better quantitative assessment and suitability
to the human visual system. It is hoped that this work will help build a
comprehensive package for the analysis of survival data, and stimulate
further research to fill the gaps in the current state of the art.

1   Introduction

Graphical methods of examining data have gained considerable popular-
ity over the last two decades. Many analysts feel that simple descriptive
plots allow them to comprehend the overall pattern, if any, which is not un-
derstood easily from a tabular representation. Sometimes these plots help
form conjectures or open up unexpected directions of further investigation.
In the context of exploratory data analysis, graphical representation is an
essential part of model building. Graphical tools are also quite useful in
communicating the findings of an applied statistician to the customer. For
example, an estimated curve or confidence band is best presented through
a plot. Sometimes it is also possible to supplement model-based prediction
by representing the associated uncertainty visually.

Apart from the aspect of presentation, graphical methods can also con-
tribute to a better analysis of the data. For instance, a formal statistical
test may be accompanied by a plot to examine how the data does or does
not conform to the null hypothesis. When the stakes are high, it is often
wiser to also examine the pattern in these plots instead of making a decision
based solely on a single binary outcome of the formal test, or a p-value.
This strategy is particularly relevant when the statistical test is marginally