Hyperbolastic Models for Survival Data

Zoran Bursac[1], Mohammad Tabatabai[2], D. Keith Williams[1], and Karan P. Singh[3]
University of Arkansas for Medical Sciences[1]
Cameron University[2]
University of North Texas Health Sciences Center[3]

Abstract

Modeling the outcomes that factor in the element of time has been extensively studied and applied in a wide range of medical and biological studies. In continuation of our work on a family of hyperbolastic growth models, new hyperbolastic survival models are introduced. The models are utilized for analysis of a published survival data set, and the results are compared with those obtained from commonly used survival models (e.g. Cox, Weibull, log-logistic and lognormal). With the same number of parameters as in the classical models, the new models perform better in terms of values of log-likelihood and survival probability prediction. The new hyperbolastic survival models produce flexible hazards, accommodate a mix of covariates and may be a useful predictive tool in various research fields that employ time-to-event data.

Keywords: Hyperbolastic Survival Model, Cox Model, Weibull Model, Log-likelihood, Log-logistic Model, Lognormal Model.

1. Introduction

Survival data is generally considered as data that involves time to some event, for instance death, manifestation of disease or recurrent disease event. Survival data concepts have been widely applied in biomedicine, engineering, social sciences, business and other fields [2,3]. As a continuation of our work on the family of hyperbolastic models [1,6] in this paper we continue the innovative application of hyperbolic functions to survival models. Here we introduce two new parametric survival models, called hyperbolastic survival models (H1 and H2). These two-parameter models exhibit variety of flexible hazard shapes. The application of these two models is demonstrated on Hosmer and Lemeshow HMO HIV data set with two covariates [2]. Also performance is compared with several commonly used models, like Cox, Weibull, lognormal and log-logistic.

2. Hyperbolastic Distribution 1

The proposed cumulative distribution function and probability density functions are, respectively:

$$F(t) = 1 - e^{-\lambda[\mathrm{ArcSinh}(\gamma t)]^{\gamma}}$$

and

$$f(t) = \frac{\lambda\gamma^2 e^{-\lambda \mathrm{ArcSinh}(\gamma t)^{\gamma}} \mathrm{ArcSinh}(\gamma t)^{-1+\gamma}}{\sqrt{1+\gamma^2 t^2}}$$

where time variable *t*, and scale λ and shape γ like parameters are

$$t \geq 0, \lambda > 0, \gamma > 0.$$

The survivorship function is therefore

$$S(t) = e^{-\lambda[\mathrm{ArcSinh}(\gamma t)]^{\gamma}}$$

and the hazard function

$$h(t) = \frac{\lambda\gamma^2 \mathrm{ArcSinh}(\gamma t)^{-1+\gamma}}{\sqrt{1+\gamma^2 t^2}}.$$

Several shapes of the hazard function are presented in the Figure 1 for different values of parameter γ while λ is held constant at 1.

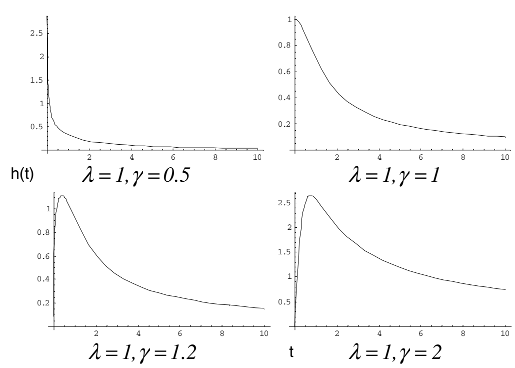

Figure 1. Hazard functions of hyperbolastic distribution 1.

3. Hyperbolastic Distribution 2

The proposed cumulative distribution function and probability density functions are, respectively:

$$F(t) = 1 - e^{\left[\lambda\left(1 - t^{\gamma}/\mathrm{ArcSinh}\left(t^{\gamma}\right)\right)\right]}$$

and

$$f(t) = -\lambda \times e^{\left(1 - \frac{t^{\gamma}}{\mathrm{ArcSinh}\left(t^{\gamma}\right)}\right)} \times \left(\frac{\gamma t^{-1+2\gamma}}{\sqrt{1+t^{2\gamma}}\,\mathrm{ArcSinh}\left(t^{\gamma}\right)^{2}} - \frac{\gamma t^{-1+\gamma}}{\mathrm{ArcSinh}\left(t^{\gamma}\right)} \right)$$

where time variable *t*, scale λ and shape γ like parameters are

$t > 0, \lambda > 0, \gamma > 0.$

The survivorship function is therefore

$$S(t) = e^{\left[\lambda\left(1 - t^{\gamma}/\mathrm{ArcSinh}\left(t^{\gamma}\right)\right)\right]}$$

and the hazard function

$$h(t) = -\lambda \times \left(\frac{\gamma t^{-1+2\gamma}}{\sqrt{1+t^{2\gamma}}\,\mathrm{ArcSinh}\left(t^{\gamma}\right)^{2}} - \frac{\gamma t^{-1+\gamma}}{\mathrm{ArcSinh}\left(t^{\gamma}\right)} \right).$$

Several shapes of the hazard function are presented in the Figure 2 for different values of parameter γ while λ is held constant at 1.

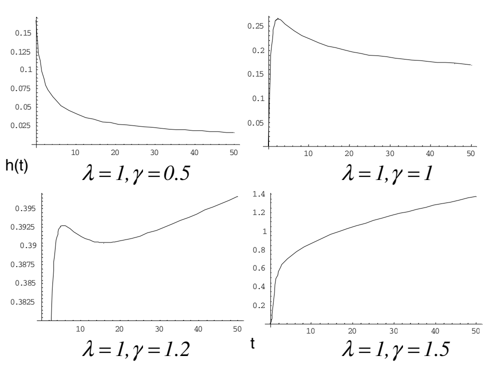

Figure 2. Hazard functions of hyperbolastic distribution 2.

To obtain the maximum likelihood estimate with respect to the parameters we maximized the log-likelihood function of the following general form

$$L(\beta) = \sum_{i=1}^{n} c_i \, ln\left[f\left(t_i, \beta, x_i\right)\right] + \left(1 - c_i\right) ln\left[S\left(t_i, \beta, x_i\right)\right]$$

by substituting in the appropriate probability density and survivorship functions [2].

4. Application

To examine the performance of the hyperbolastic survival models, we applied them to Homer and Lemeshow HMO HIV data set [2]. We used two covariates, age and drug. Hyperbolastic models were compared to Cox (C), Weibull (W), lognormal (LN) and log-logistic (LL) models.

Table 1. Parameter estimates for six models and their standard errors.

Model	*Age*	*Drug*
H1	0.0983(0.0183)	1.0704(0.2502)
H2	0.1026(0.0183)	1.1403(0.2532)
C	0.0915(0.0185)	0.9411(0.2555)
W	0.1081(0.0183)	1.2499(0.2594)
LN	0.1004(0.0185)	1.1439(0.2565)
LL	0.1067(0.0189)	1.2404(0.2604)

Log-likelihood functions of hyperbolastic as well as comparison models were coded to be comparable with respect to number of parameters (shape and scale plus two covariates) and time to death was used instead of log(time).

Table 2. Ranking of models based on log-likelihood estimates.

Rank	*Model*	*Log-likelihood*
1	H2	259.8
2	H1	259.9
3	LN	261.1
4	LL	263.4
5	W	263.5
6	C	281.7

Optimization of log likelihood functions was performed using SAS PROC NLP and Mathematica [4,5]. We used Newton-Raphson optimization algorithm with COV=2 option which invokes inverse Hessian variance-covariance matrix. Table 1 shows the parameter estimates for six models for age and drug covariates along with their standard errors. While the parameter estimates are fairly comparable the log-likelihood estimates are slightly lower for H2 and H1 models as compared to the other models which is shown in Table 2. Finally we plot hazard functions for each of the hyperbolastic models by time and age, separately for IV drug users and non-users. We can see from Figures 3 and 4 that hazard of dying is higher for IV drug users and that it increases with age regardless of the time. With respect to time dimension, hazard at first increases and then goes down, following shape similar to some hazards of the lognormal distribution.

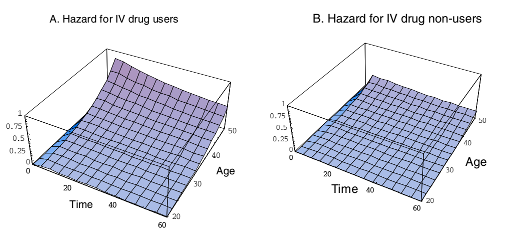

Figure 3. H1 hazards for HMO HIV data

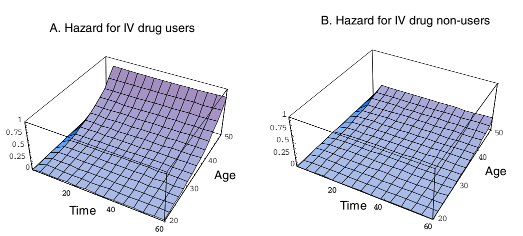

Figure 4. H2 hazards for HMO HIV data

5. Discussion and Conclusions

Proposed hyperbolastic models are simple. With two parameters (shape and scale like) they are comparable to some standard models (e.g. Weibull) [3]. They can be easily coded and optimized in several readily available software packages. Here we demonstrated two, SAS and Mathematica [4,5]. They have flexible hazard shapes in particular H2 model. They include hazard shapes of Weibull, lognormal and log-logistic models plus some additional shapes (Figure 2) [3]. These models can be applied with proportional hazard, accelerated failure time and proportional odds models. They also allow mix of covariates which is a necessary feature in modern research.

Main feature of these models is the flexible hazard functions. They offer an alternative approach to the classic models when the prior knowledge of the hazard function for the problem at hand is not known, and researcher has to choose from several models. It is our recommendation that researcher compare the fit and prediction of proposed models with standard ones before deciding which one to use.

References

1. Bursac Z, Tabatabai M, Williams DK. (2006). Non-linear Hyperbolastic Growth Models and Applications in Cranofacial and Stem Cell Growth. 2005 Proceedings of the American Statistical Association, Biometrics Section [CD-ROM], Alexandria, VA: American Statistical Association, 190-197.
2. Hosmer DW and Lemeshow S. (1999). Applied Survival Analysis: Regression Modeling of Time to Event Data. Wiley.
3. Lee ET. (1992). Statistical Methods for Survival Data Analysis. Wiley.
4. Mathematica v4. (2002). Wolfram Research Inc. Champaign, IL.
5. SAS v9. (2004). SAS Institute Inc. Cary, NC.
6. Tabatabai M, Williams DK, Bursac Z. (2005). Hyperbolastic Models: Theory and Application. Theoretical Biology and Medical Modeling, 2(14).

Contact Information

The SAS code to fit H1 and H2 models to HMO HIV data is available online at http://www.uams.edu/biostat/bursac/Hyperbolastic.htm or will be provided as requested.

Zoran Bursac
Biostatistics
Fay W. Boozman College of Public Health
University of Arkansas for Medical Sciences
4301 W. Markham, Slot 781
Little Rock, AR 72205
Tel. (501) 526-6723
E-mail. zbursac@uams.edu