SEMIPARAMETRIC SURVIVAL ANALYSIS VIA BERNSTEIN POLYNOMIALS

Renato Valladares Panaro[1], Adriana dos Santos Lima[2], Silvio Cabral Patricio[3]

Introduction

According to Lorentz (2012), Bernstein's polynomials (BP) were introduced by Bernstein around 1912 as an alternative to extreme value theorem proof. The Weierstrass' extreme value theorem, in turn, has been an important tool and has been used for many applications in calculus and analysis. The theorem is widely used to state that any continuous function over an interval [a, b] in R is limited and, in addition to, there is a maximum value and a minimum value in that interval.

First, Bernstein showed there exists two reals k and K such that $k \leq B_m(x) \leq K$. So, the mathematician proved that if $f(x)$ is uniformly continuous on $[0,1]$ then $\lim_{m \to \infty} B_m(x) = f(x)$. Similarly, kernel functions, approximation splines and Bernstein's Polynomials can used to approximate functions.

In 2012, Osman and Gosh proposed baseline risk non-parametric modelling for survival proportional hazards regression model. The authors approximate baseline risk function using BP and provide, among other results, proofs on asymptotics. The likelihood log-concavity property shown in this article leads to less costly computational procedures to find bayesian estimators and guarantees the uniqueness of maximum likelihood estimator.

Bayesian inference isn't straightforward as numerical optimization methods already implemented in R that ease frequentist approach. However, this was elegantly done based on gibbs sampling and Adaptive Rejection Metropolis Sampling (ARMS) algorithm.

Goals

Present proportional hazards model's base risk estimates and related functions based on Bernstein Polynomials (BP). Consequently, log likelihood is based on non-parametric BP estimates and parametric estimates alternatively to Cox's model partial likelihood. For this,

[1] Universidade Federal de Minas Gerais (UFMG), renatovp@ufmg.br
[2] Universidade Federal de Minas Gerais (UFMG), adrianalima@ufmg.br
[3] Universidade Federal de Minas Gerais (UFMG), silviocp@ufmg.br

an application is presented making use of larynx dataset available at KMSurv package (Klein et al, 2012). Most relevant results were briefly explained and displayed throughout this text.

Methods

The model's proposition takes into account some assumptions in order to obtain good properties and compliance with the existing literature.

Notations and Censorship Mechanism

This text considers random right censorship without cure fraction, assuming there is a failure time τ such that $\tau = \inf\{t : P(T > t) = 0\}$, $\tau < \infty$. Random variable T_i denotes time to a given event of interest, subject to right random censorship C_i. For each subject, we observe (t_i, Δ_i), $t_i = \min(T_i, C_i)$, $\Delta_i = I(T_i \leq C_i)$. Moreover, the cumulative hazard function is denoted by $H(t) = -\log S(t)$, the hazard by $h(t) = \dot{H}(t) = -\frac{d}{dt}\log S(t)$.

Risk Function Using the Polynomial

Intuitively, an approximation for cumulative risk $H(\frac{k}{m}\tau)$ when $n \to \infty$ using BP is,

$$B(t; m; H) = \sum_{k=0}^{m} H(\tfrac{k}{m}\tau)\binom{m}{k}(t/\tau)^k(1-t/\tau)^{m-k}.$$

Deriving the function $B(\cdot)$, we have

$$\dot{B}(t; m; H) = \sum_{k=0}^{m} H(\tau\tfrac{k}{m})\frac{\Gamma(m+1)}{\Gamma(k)\Gamma(m-k-1)}(t/\tau)^k(1-t/\tau)^{m-k}\left\{\frac{1}{\tau} - (\frac{t}{\tau^2})(1-\frac{t}{\tau})^{-1}\frac{(m-k)}{k}\right\}$$

$$= \sum_{k=1}^{m}\left\{H(\tau\tfrac{k}{m}) - H(\tau\tfrac{k-1}{m})\right\}\frac{f_{\beta)}(t/\tau; m-k+1)}{\tau} = \sum_{k=1}^{m}\gamma_k g_{m,k}(t).$$

Now, risk is approximately $h(t_i|x_i) = \sum_{k=1}^{m}\gamma_k g_{m,k}(t)$ and cumulative risk is $H(t_i|x_i) = \sum_{k=1}^{m}\gamma_k G_{m,k}(t)$. According to Osman and Gosh (2012), $m = \sqrt{n}$ and $\tau = \max_i(t_i)$ are assumed in practical situations.

Inference

Unlike the Cox's model, the proposed approach does not use partial likelihood. For Cox proportional hazards theoretical background see Klein and Moeschberger (2006).

$$l(\gamma, \beta) = \sum_{i=1}^{n}\{\Delta_i \log[h(t_i \mid x_i)] - H(t_i \mid x_i)\} = \sum_{i=1}^{n}\left\{\Delta_i \log[h_m(t_i, \gamma)e^{x_i'\beta}] - H_m(t_i, \gamma)e^{x_i'\beta}\right\}$$

where $h_m(t, \gamma) = \sum_{k=1}^{m}\gamma_k g_{m,k}(t)$, given by the previous expression.

On one hand, maximum likelihood estimation consists in maximizing the expression above with respect to $(\gamma, \beta)'$. On the other hand, in consonance with Ibrahim et al (2001), bayesian estimators are calculated for each iteration risk curve recurring to Adaptive Rejection Metropolis Sampling (ARMS) within Gibbs sampling to find posterior conditional distributions. One can't know, but ARMS requires a grid delimiting the aimed density support. HI package (Petris & Tardella, 2013) provides an ARMS routine, in addition to that, a Jacobian variable transformation to push all variables in [0,1].

Results and Discussion

For comparison, Cox model was fitted. Centered but not scaled data were used due to avoid overflow in the argument to the exponential function. These actions do lead to numerical stability as mentioned at survival package details (Therneau & Grambsh, 2000). Table 1 shows coefficient estimates for proportional hazards model using partial likelihood (coxph) and Bernstein Polynomials, both bayesian and frequentist approach.

Table1 – Estimate comparison between BP estimates and Cox's model.

Parameter	Estimate		
	Coxph	Posterior mean (BP)	MLE (BP)
β_1	0.02	0.02	0.02
β_2	0.14	0.16	0.17
β_3	0.64	0.65	0.66
β_4	1.71	1.80	1.80
γ_1	-	0.11	0.10
γ_2	-	0.16	-0.24
γ_3	-	0.09	0.00
γ_4	-	0.10	0.08
γ_5	-	0.14	0.27
γ_6	-	0.23	0.00
γ_7	-	0.29	0.75
γ_8	-	0.29	0.00
γ_9	-	0.25	0.00
γ_{10}	-	0.28	0.00

In table 1, parametric estimates are close if compared to Cox`s Model, but polynomial coefficients differ. Besides the differences between posterior and maximum likelihood

estimations for γ_k, $k = 1, \cdots, 10$ seen on table 1, figure 1 shows that approximate curves using BP are close to Breslow non-parametric approximation curve.

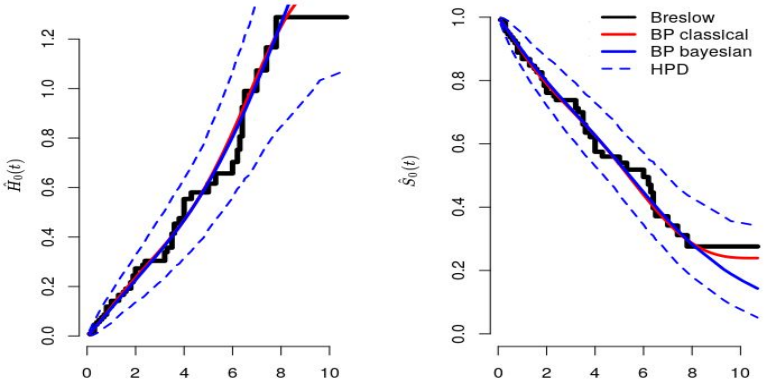

Figure 1 – Base cumulative risk and base survival curves.

For a given set of covariates is also possible to compute the survival function for a patient, in figure 2 the patient is supposed to have 77 years old, similar to coxph.

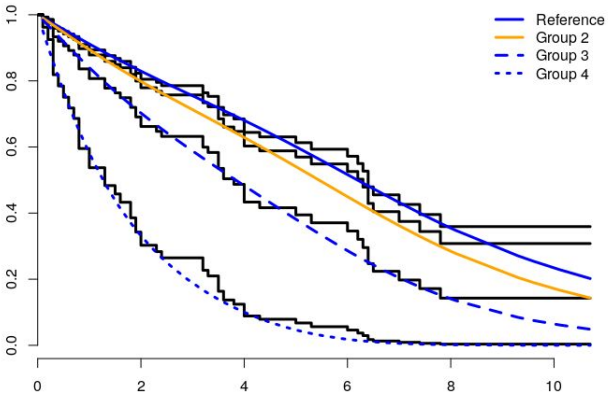

Figure 2 – Survival curves by groups for a given 77 years old patient.

Conclusion

The proposed methodology has three characteristics: (a) the risk rate estimator has the solution of a strictly convex optimization problem for a wide range of applications which is computationally attractive; (b) The model is shown to encompass the proportional risk structure; (c) Asymptotic properties, including consistency, are established under a set of light regularity; (d) in conclusion, the figure 1 and 2 shows continuous approximates to hazard and survival functions.

References

Giovanni Petris and Luca Tardella; original C code for ARMS by Wally Gilks. (2013). HI: Simulation from distributions supported by nested hyperplanes. R package version 0.4. https://CRAN.R-project.org/package=HI

Ibrahim, J. G., Chen, M. H., & Sinha, D. (2001). Bayesian survival analysis. Springer Science & Business Media.

Klein, J. P., & Moeschberger, M. L. (2006). Survival analysis: techniques for censored and truncated data. Springer Science & Business Media.

Lorentz, G. G. (2012). Bernstein polynomials. American Mathematical Soc..

Original by Klein, Moeschberger and modifications by Jun Yan (2012). KMsurv: Data sets from Klein and Moeschberger (1997), Survival Analysis. R package version 0.1-5. https://CRAN.R-project.org/package=KMsurv

Osman, M., & Ghosh, S. K. (2012). Nonparametric regression models for right-censored data using Bernstein polynomials. Computational Statistics & Data Analysis, 56(3), 559-573.

Terry M. Therneau, Patricia M. Grambsch (2000). _Modeling Survival Data: Extending the Cox Model_. Springer, New York. ISBN 0-387-98784-3.