Published in final edited form as:

J Biopharm Stat. 2015 ; 25(3): 539–547. doi:10.1080/10543406.2014.923726.

OPTIMAL ESTIMATION FOR REGRESSION MODELS ON τ -YEAR SURVIVAL PROBABILITY

Minjung Kwak,

Department of Statistics, Yeungnam University, Gyeongsan, Gyeongbuk 712749, ROK,
mjkwak@yu.ac.kr, Tel: 82-53-810-2321

Jinseog Kim, and

Department of Statistics and Information Science, Dongguk University, Gyeongju 780714, ROK,
jinseog.kim@gmail.com, Tel:82-54-770-2247

Sin-Ho Jung*

Samsung Center Research Institute, Samsung Medical Center, SungKyunKwan University,
Seoul, 135710, ROK. sinho.jung@duke.edu, Tel:82-2-3410-9815

Abstract

A logistic regression method can be applied to regressing the τ -year survival probability to covariates, if there are no censored observations before time τ . But if some observations are incomplete due to censoring before time τ , then the logistic regression cannot be applied. Jung (1996) proposed to modify the score function for logistic regression to accommodate the right censored observations. His modified score function, motivated for a consistent estimation of regression parameters, becomes a regular logistic score function if no observations are censored before time τ . In this paper, we propose a modification of Jung's estimating function for an optimal estimation for the regression parameters in addition to consistency. We prove that the optimal estimator is more efficient than Jung's estimator. This theoretical comparison is illustrated with a real example data analysis and simulations.

Keywords

Censoring distribution; Logistic regression; Nonnegative definite; Survival probability

1 Introduction

Cox (1972) regression based on proportional hazards model has been widely used to relate the survival distribution of a patient with a set of covariates. So, Cox regression method is to investigate the impact of covariates on the patient's survival time over the whole time span. In a certain clinical setting, however, there may be a landmark time τ relevant to determine the patient's disease status. Suppose that, if a cancer patient is free of disease for 5 years after a surgery for tumor resection, then this patient is considered to be completely cured of the cancer. In this case, we may be interested in relating the disease-free probability at time

*Correspondence to: Sin-Ho Jung, sinho.jung@duke.edu.

$\tau = 5$ years to a given set of covariates. The regression analysis of survival probability at a chosen time point is of interest for time-dependent ROC curve methods, refer to e.g. Heagerty and Zheng (2005) and Zheng et al. (2010). If there is no censored observation before time τ , then we can apply a logistic regression method. But when some observations are censored before time τ , we cannot apply the logistic regression method to the survival data.

The censoring mechanism is usually assumed to be independent of the survival time itself. Such an assumption is reasonable, for example, when the survival time is administratively censored. Under the independence assumption, the probability of being observed to survive time τ is a product of two probabilities, probability of not being censored by τ and probability of surviving time τ . Using this property, Jung (1996) adjusted the score function of the logistic regression for the censoring probability, and regressed the survival probability at time τ to covariates.

Jung's approach of correction for the censoring probability is aimed at a consistent estimation of the regression coefficients. In this paper we propose an inference method for an optimal as well as consistent regression estimators. We theoretically proved that the proposed estimators are more efficient than Jung's. This comparison is illustrated by analyzing real example data and simulation studies.

2 An Optimal Regression Method for τ -Year Survival Probability

For patient i ($i = 1, \dots, n$), let T_i be the survival time and $Z_i = (Z_{1i}, \dots, Z_{pi})^T$ the covariate vector. Suppose that, for a time point τ , survival probability, $\pi_i = \text{pr}(T_i \geq \tau)$, satisfies the relation

$$\phi(\pi_i) = \beta_0^T Z_i, \quad (1)$$

where ϕ , called a link function, is a monotonic differentiable function in $[0, 1]$ and β_0 is the true value of the regression parameter $\beta = (\beta_1, \dots, \beta_p)^T$. Among popularly used link functions are logit link $\phi(\pi) = \log\{\pi/(1-\pi)\}$, log-log link $\phi(\pi) = -\log\{-\log(\pi)\}$, and probit, the inverse of standard normal cumulative density function. Doksum and Gasko (1990) extensively review a correspondence between models in binary regression analysis and in survival analysis.

If all survival times are observed, maximum likelihood estimator of β is obtained by solving the equation

$$\sum_{i=1}^n \{I(T_i \geq \tau) - \pi(\beta^T Z_i)\} \frac{\pi'(\beta^T Z_i) Z_i}{\pi(\beta^T Z_i) \{1 - \pi(\beta^T Z_i)\}} = 0,$$

where π is the inverse of the link function ϕ and $\pi'(\phi) = \pi(\phi)/\phi$. However, in most studies observing lifetime, some survival times are censored. If a survival time is censored before time τ , then it is not clear whether this subject will survive beyond τ or not. In this case, the regular maximum likelihood method can not be used to estimate the regression parameters β

of model (1). Let $X_i = \min(T_i, C_i)$ with C_i , censoring time, which is independent of T_i and δ_i be the event indicator taking 1 if T_i is observed and 0 otherwise. To simplify our discussion, we assume that C_i are independent and identically distributed with $G(t) = \text{pr}(C_i \geq t)$. Then noting that $E\{I(X_i \geq \tau)\} = \text{pr}(T_i \geq \tau)G(\tau)$, Jung (1996) proposed a modified estimating equation, $\tilde{U}(\beta) = 0$, where

$$\tilde{U}(\beta) = \sum_{i=1}^n \{I(X_i \geq \tau) - \hat{\theta}\pi(\beta^T Z_i)\} \frac{\pi'(\beta^T Z_i)Z_i}{\pi(\beta^T Z_i)\{1 - \pi(\beta^T Z_i)\}} \quad (2)$$

and $\hat{\theta}$ is the Kaplan-Meier estimate of $\theta = G(\tau)$ that is obtained by exchanging the role of censoring and event. He proves that the estimator is consistent and asymptotically normal under moderate assumption. In fact, estimating function (2) is oriented to consistency of regression estimator, rather than to efficiency. In this article, we propose an optimal estimating function that can improve the efficiency while maintaining consistency.

Let $\varepsilon(\beta) = \{\varepsilon_1(\beta), \dots, \varepsilon_n(\beta)\}^T$, $\varepsilon_i(\beta) = I(X_i \geq \tau) - \hat{\theta}\pi(\beta^T Z_i)$,

$$D(\beta) = -\frac{\partial \varepsilon(\beta)}{\partial \beta} = \hat{\theta} \{\pi'(\beta^T Z_1)Z_1, \dots, \pi'(\beta^T Z_n)Z_n\}^T$$

and $V = \text{cov}\{\varepsilon(\beta_0)\}$. Noting, from estimating function (2), that $\varepsilon(\beta)$ plays the role of residual vector, we consider the estimating function

$$U(\beta) = D^T V^{-1} \varepsilon(\beta),$$

where $D = D(\beta_0)$. Similarly to Jung (1996), we can show that $\hat{\beta}$, the solution of $U(\beta) = 0$, is consistent and asymptotically normal with mean β_0 and with covariance matrix $(D^T V^{-1} D)^{-1}$.

Now let's prove that $\hat{\beta}$ is an optimal estimator. For an $n \times p$ full rank matrix H , we may consider an estimating function $U_H(\beta) = H^T \varepsilon(\beta)$. Note that

$$H^T = \left\{ \frac{\pi'(\beta^T Z_1)Z_1}{\pi(\beta^T Z_1)\{1 - \pi(\beta^T Z_1)\}}, \dots, \frac{\pi'(\beta^T Z_n)Z_n}{\pi(\beta^T Z_n)\{1 - \pi(\beta^T Z_n)\}} \right\}$$

for $\tilde{U}$ and $H^T = D^T V^{-1}$ for U . Also, using similar arguments to Jung (1996), we can show that the solution $\hat{\beta}_H$ to the equation $U_H(\beta) = 0$ is consistent and approximately normal with mean β_0 and with covariance matrix

$$(H^T D)^{-1} H^T V H (D^T H)^{-1}.$$

Further, Appendix A shows that $\text{cov}(\hat{\beta}_H) - \text{cov}(\hat{\beta})$ is nonnegative definite, meaning that, for any linear combination of β , $\hat{\beta}$ provides more efficient estimator than $\hat{\beta}_H$.

Since V is an $n \times n$ matrix, obtaining its inverse matrix in the optimal estimating equation looks problematic, especially when sample size n is large. However, because of its unique structure, the inverse matrix can be easily obtained. From Appendix B, we have

$$V = A - \gamma \pi^{\otimes 2}, \quad (3)$$

where $A = \theta \times \text{diag}\{\pi_1(1-\theta\pi_1), \dots, \pi_n(1-\theta\pi_n)\}$, $\pi = (\pi_1, \dots, \pi_n)^T$, $\gamma = \theta^{-2} \int_0^\tau Y^{-1}(t) d\Lambda(t)$, $\Lambda(t) = -\log G(t)$, $Y(t) = \sum_{i=1}^n I(X_i \geq t)$, and $a^{\otimes 2} = aa^T$ for a vector a . Since A is a diagonal matrix, $\gamma \pi^{\otimes 2}$ in (3) reflects the dependency between residuals. Noting that $\gamma = O_p(n^{-1})$, we see that the dependency gets weaker as n increases. Now, by applying the Bartlett (1951) equality to (3), we have

$$V^{-1} = A^{-1} + \frac{\gamma}{1 - \gamma \pi^T A^{-1} \pi} (A^{-1} \pi)^{\otimes 2}. \quad (4)$$

Hence, inversion of V involves only that of diagonal matrix A .

Using Newton-Raphson's method, solution to the equation $U(\beta) = 0$ at iteration k is obtained by

$$\hat{\beta}^{(k)} = \hat{\beta}^{(k-1)} + (D^T V^{-1} D)^{-1} U(\hat{\beta}^{(k-1)}).$$

Since D and V include unknown parameters β_0 , θ and $\Lambda(t)$, we replace them with $\hat{\beta}^{(k-1)}$, $\hat{\theta}$ and Nelson's (1969) estimate $\hat{\Lambda}(t) = \int_0^t Y^{-1}(s) dN(s)$. Here, $N(t) = \sum_{i=1}^n N_i(t)$ and $N_i(t) = (1 - \delta_i)I(X_i \leq t)$. Replacing the parameters included in the weight $D_T V^{-1}$ with their consistent estimates does not change asymptotic property of the resulting estimator $\hat{\beta}$.

3 Example

The proposed optimal estimating method is applied to real data. Phenobarbital was studied in the treatment of children with febrile seizure in a double blinded clinical trial (Farwell et al., 1990). Two hundred seventeen eligible febrile seizure children were randomized to either a two-year prophylactic phenobarbital treatment or a placebo arm. A variable of interest was the time to the recurrent seizure from the index febrile seizure. The number of prior febrile seizures (NPF) and the age at the index seizure (AGE) were considered important covariates for the time to the recurrent seizure. We relate three covariates, treatment (TRT), NPF and AGE to the probability of staying seizure-free for two years when the treatment was tapered. In this data, 29 observations were censored before 2 years. From Table 1, we observe that the standard errors of the regression estimates by the optimal method are smaller than those by Jung's (1996) method. The AGE variable is significant by the optimal method at 5% significance level, but not by Jung's method. From the analysis results, we may interpret that the fewer the number of prior febrile seizures and the older when the index seizure occurs, the more likely to remain seizure-free for two years. The treatment is not significant. These results are based on logistic regression model, but regression models using other link functions give similar results.

4 Simulation Studies

Extensive simulation studies were conducted to investigate finite-sample properties of the proposed optimal estimating method which is based on asymptotic theory.

In the first simulation, covariate Z_i was generated from $U(0, 1)$ distribution. Given Z_i , T_i was generated from an exponential distribution with hazard rate $\exp(-\beta Z_i)$. With log-log link $\varphi(\pi) = -\log\{-\log(\pi)\}$ and $\pi_i = \text{pr}(T_i \leq \tau)$, this survival distribution entails the model

$$\phi(\pi_i) = \beta_0 + \beta_1 Z_i,$$

where $\beta_0 = -\log(\tau)$ and $\beta_1 = \beta$. Censoring variables were generated from $U(c_0, c_0 + c_1)$. For a given β value, the constant c_1 was chosen for 30% censoring before $\tau = 1$ by $U(0, c_1)$. With the chosen c_1 value fixed now, c_0 was chosen for 10% or 20% censoring before $\tau = 1$ by $U(c_0, c_0 + c_1)$. For each combination of β and censoring proportion, 1000 random samples of $\{(X_i, \delta_i, Z_i), i = 1, \dots, n\}$ were generated with $n = 100, 200$ or 300 . For each sample, $(\hat{\beta}_0, \hat{\beta}_1)$ and its variance were calculated using Jung's (1996) method and the optimal method. Table 2 reports bias and standard error (SE) of $\hat{\beta}_1$, and empirical power (PWR) for $H_0 : \beta_1 = 0$, which was calculated as the proportion of the 1000 testings with a nominal level 0.05 that rejected H_0 . The regression estimates by the both methods have very small bias, especially under $\beta_1 = 0$. Under $H_1 : \beta_1 = 2$, the optimal method seems to have a slightly smaller bias than Jung's method. When $\beta_1 = 0$, we observe that the empirical powers by both methods are fairly close to the nominal level 0.05. When $\beta_1 = 2$, optimal estimation method seems to be more powerful than Jung's method. For example, with $n = 100$ and 30% censoring, the optimal method has an over twice higher power than that of Jung's (0.278 vs. 0.136). Difference in power between two methods increases as censoring proportion increases. Note that, if there is no censoring before τ , the two estimating methods are identical. Standard error of optimal method is always smaller than that of Jung's method.

In the second simulation, we want to check the robustness of estimation methods against misspecification of link function. Two covariates Z_{1i} and Z_{2i} were generated from Bernoulli(1/2) and $U(0,1)$ respectively. Given Z_{1i} and Z_{2i} , T_i was generated from an exponential distribution with hazard rate $\exp(-\beta_1 Z_{1i} - \beta_2 Z_{2i})$. With log-log link $\varphi(\pi) = -\log\{-\log(\pi)\}$ and $\pi_i = \text{pr}(T_i \leq \tau)$, this survival distribution entails the model

$$\phi(\pi_i) = \beta_0 + \beta_1 Z_{1i} + \beta_2 Z_{2i},$$

where $\beta_0 = -\log(\tau)$. We introduce another covariate in the regression model, since regression models with a single covariate will do not depend on the choice of a link function under H_0 . Censoring variable was similarly generated as in the first simulation set. For each sample, regression estimates and their standard errors were calculated using the popular logit link, a wrong choice. Table 3 reports standard error of $\hat{\beta}_1$ and empirical power for $H_0 : \beta_1 = 0$ with $\tau = 1$ and $n = 100$. Empirical powers by both methods are fairly close to the nominal level 0.05 under the null hypothesis, although Jung's method is slightly conservative under heavy censoring (30%). When $\beta_1 = 2$, optimal estimation method is always more powerful

than Jung's method. As in the first simulation, standard error of optimal method is always smaller than that of Jung's method and the difference in power between two methods increases as censoring proportion increases.

5 Concluding Remarks

The optimal regression method of the τ -year survival probability is easy to understand and does not require any assumption on the form of the underlying survival distribution. Its coefficients can be estimated by assigning optimal weights to the regression residual terms. This approach is proved to be more efficient than the approach proposed by Jung (1996). This theoretical comparison is further illustrated by a real data analysis and simulations.

We have assumed that the censoring distribution is free of covariates. This assumption holds for most well conducted clinical trials where the major cause of censored is administrative, but it may be violated for some retrospective studies. If the censoring distribution depends on a discrete covariate, we may partition the whole data into a number of strata identified by the values of the covariate, so that the censoring times within each stratum have an identical distribution. Suppose that there are K strata, and within stratum k ($k = 1, \dots, K$) censoring times are independent and identically distributed with survivor function $G_k(t)$. Let $\{(X_{ki}, \delta_{ki}, Z_{ki}), i = 1, \dots, n_k\}$ denote the data from stratum k , where $\sum_{k=1}^K n_k = n$. Then the optimal estimating function $U(\beta)$ is extended to

$$\bar{U}(\beta) = \sum_{k=1}^K D_k^T V_k^{-1} \varepsilon_k(\beta),$$

where $\varepsilon_k(\beta) = \{\varepsilon_{k1}(\beta), \dots, \varepsilon_{k,n_k}(\beta)\}^T$, $\varepsilon_{ki}(\beta) = I(X_{ki} \leq \tau) - \hat{\theta}_k \pi(\beta^T Z_{ki})$,

$$D_k(\beta) = \frac{\partial \varepsilon_k(\beta)}{\partial \beta} = \hat{\theta}_k \{\pi'(\beta^T Z_{k1}) Z_{k1}, \dots, \pi'(\beta^T Z_{k,n_k}) Z_{k,n_k}\}^T,$$

$V_k = \text{cov}\{\varepsilon_k(\beta_0)\}$, and $\hat{\theta}_k$ is the Kaplan-Meier estimator of $G_k(\tau)$ from $\{(X_{ki}, 1 - \delta_{ki}), i = 1, \dots, n_k\}$. It is easy to show that this estimating function has similar asymptotic properties as those of $U(\beta)$. If the censoring distribution depends on a continuous covariate, we may consider partitioning the range of the covariate values into certain number of intervals using some cutoff values, so that the censoring distribution is relatively free of covariates within each interval, and apply the above stratified analysis method.

In this paper our focus has been on optimality of the regression parameters on the survival probability at fixed time point. Unlike time dependent covariates which might be changed at time of patient's death, we assumed that covariates are fixed for each patients and this assumption is reasonable since our consideration is limited on a pre-specified time point instead of following through until a patient die or censored.

The regression method does not require heavy computations, so that all the computations were done on generic personal computer using Fortran 77 codings. Random numbers were

generated using CM-library subroutines. The Fortran codings will be made available from the authors upon request.

Acknowledgements

We would like to thank the editor, the associate editor and the reviewers for valuable comments that greatly improved the presentation of the article. This work was supported by the SungKyunKwan University Research Grant.

Appendix A

Nonnegative definiteness of $\text{cov}(\hat{\beta}_H) - \text{cov}(\hat{\beta})$

In order to prove that $\text{cov}(\hat{\beta}_H) - \text{cov}(\hat{\beta}) = (H^T D)^{-1} H^T V H (D^T H)^{-1} - (D^T V^{-1} D)^{-1}$ is nonnegative definite, it suffices to show that $M = D^T V^{-1} D - D^T H (H^T V H)^{-1} H^T D$ is nonnegative definite. It is easy to show that M is the covariance matrix of $\{D^T V^{-1} - D^T H (H^T V H)^{-1} H^T\} \varepsilon(\beta_0)$, so that M should be nonnegative definite.

Appendix B

$V = A - \gamma \pi^{\otimes 2}$

Let $V = (\sigma_{ij})_{i,j=1, \dots, n}$. Then,

$$\sigma_{ii} = \text{var}\{I(X_i \geq \tau) - \hat{\theta} \pi_i\} = \text{var}\{I(X_i \geq \tau)\} + \pi_i^2 \text{var}(\hat{\theta}) - 2\pi_i \text{cov}\{I(X_i \geq \tau), \hat{\theta}\}.$$

Here,

$$\text{var}\{I(X_i \geq \tau)\} = \theta \pi_i (1 - \theta \pi_i) \quad (\text{A1})$$

and

$$\text{var}(\hat{\theta}) = \gamma \quad (\text{A2})$$

by Greenwood (1926). Also, by the martingale representation of a Kaplan-Meier estimator,

$$\hat{\theta} - \theta = -\theta \sum_{j=1}^n \xi_j, \quad (\text{A3})$$

where $\xi_i = -\sum_{j=1}^n \int_0^\tau Y^{-1}(t) \{dN_j(t) - Y_j(t) d\Lambda(t)\}$. Since ξ_i are asymptotically independent,

$$\text{cov}\{I(X_i \geq \tau), \hat{\theta}\} = \text{cov}\{I(X_i \geq \tau), -\theta \xi_i\}. \quad (\text{A4})$$

It can be easily shown that the righthand side of (A4) equals $\gamma \pi_i$. Hence, from (A1), (A2) and (A4), we have

$$\sigma_{ii} = \theta\pi_i(1 - \theta\pi_i) - \gamma\pi_i^2. \quad (\text{A5})$$

Similarly, it can be shown that, for $i \neq j$,

$$\sigma_{ij} = -\gamma\pi_i\pi_j. \quad (\text{A6})$$

From (A5) and (A6), we have $V = A - \gamma\pi\pi^{\otimes 2}$.

References

- Bartlett MS. An inverse matrix adjustment arising in discriminant analysis. *Annals of Mathematical Statistics*. 1951; 22:107–111.
- Cox DR. Regression models and life-tables (with discussion). *Journal of Royal Statistical Society B*. 1972; 34:187–202.
- Doksum KA, Gasko M. On a correspondence between models in binary regression analysis and in survival analysis. *International Statistical Review*. 1990; 58:243–252.
- Farwell JR, Lee YJ, Hirtz DG, Sulzbacher SI, Ellenberg JH, Nelson KB. Phenobarbital for Febrile Seizures-Effects on Intelligence and on Seizure Recurrence. *New England Journal of Medicine*. 1990; 322:364–369. [PubMed: 2242106]
- Greenwood, M. Reports on Public Health and Medical Subjects. Vol. 33. London: Her Majesty's Stationary Office; 1926. The natural duration of cancer; p. 1-26.
- Heagerty PJ, Zheng Y. Survival model predictive accuracy and dROC curves. *Biometrics*. 2005; 61:92–105. [PubMed: 15737082]
- Jung SH. Regression analysis for long-term survival rate. *Biometrika*. 1996; 83:227–232.
- Nelson W. Hazard plotting for incomplete failure data. *Journal of Quality Technology*. 1969; 1:2752.
- Zheng Y, Cai T, Stanford JL, Feng Z. Semiparametric models of time-dependent predictive values of prognostic biomarkers. *Biometrics*. 2010; 66:50–60. [PubMed: 19397579]

Table 1

Logistic regression of 2-year seizure probability using Phenobarbital Trial data

	Jung's method			Optimal method		
	$\hat{\beta}$	$se(\hat{\beta})$	P-value	$\hat{\beta}$	$se(\hat{\beta})$	P-value
Intercept	0.322	0.593	0.588	0.276	0.585	0.637
TRT	0.238	0.355	0.501	0.243	0.352	0.490
NPF	-0.689	0.239	0.004	-0.706	0.234	0.003
AGE	0.057	0.030	0.053	0.061	0.029	0.037

Table 2

Bias and standard error (SE) of $\hat{\beta}_1$ and empirical power (PWR) for $H_0 : \beta_1 = 0$

n	Censoring within τ	Jung's method			Optimal method		
		Bias	SE	PWR	Bias	SE	PWR
(a) Under $H_0: \beta_1 = 0$							
100	10%	0.042	0.616	0.054	0.042	0.613	0.058
	20%	0.003	0.715	0.033	0.004	0.708	0.045
	30%	0.014	0.804	0.034	0.015	0.795	0.041
200	10%	-0.015	0.430	0.047	-0.015	0.428	0.048
	20%	0.035	0.495	0.045	0.035	0.493	0.051
	30%	0.003	0.554	0.050	0.005	0.550	0.056
300	10%	0.012	0.350	0.040	0.012	0.350	0.042
	20%	-0.001	0.400	0.046	-0.001	0.399	0.049
	30%	0.006	0.449	0.037	0.006	0.447	0.040
(b) Under $H_1: \beta_1 = 2$							
100	10%	0.126	1.078	0.534	0.101	1.017	0.558
	20%	0.137	1.395	0.292	0.080	1.227	0.380
	30%	0.146	1.696	0.136	0.108	1.426	0.278
200	10%	0.037	0.719	0.846	0.027	0.698	0.853
	20%	0.059	0.880	0.707	0.060	0.829	0.724
	30%	0.145	1.099	0.542	0.097	0.982	0.590
300	10%	0.044	0.573	0.964	0.040	0.561	0.965
	20%	0.040	0.709	0.865	0.038	0.673	0.878
	30%	0.065	0.848	0.732	0.050	0.785	0.769

Table 3

Standard Error (SE) of $\hat{\beta}_1$ and empirical power (PWR) for $H_0 : \beta_1 = 0$ when a wrong link is used

Censoring within τ	$\beta_1 = 0$						$\beta_1 = .5$					
	Jung			Optimal			Jung			Optimal		
	SE	PWR	SE	SE	PWR	PWR	SE	PWR	SE	SE	PWR	PWR
10%	0.524	0.044	0.519	0.052	0.485	0.329	0.483	0.334				
20%	0.639	0.037	0.624	0.051	0.558	0.243	0.551	0.251				
30%	0.772	0.026	0.729	0.037	0.642	0.195	0.626	0.206				