行政院國家科學委員會補助專題研究計畫 ■ 成 果 報 告
□期中進度報告

Nonparametric Bayes estimation in Cox model using Bernstein priors.

(應用伯氏先驗分布在柯斯模型之無母數貝氏估計)

計畫類別：■ 個別型計畫　　□ 整合型計畫
計畫編號：NSC　96－2118－M－032－008－
執行期間：　96 年　8　月　1　日至　97　年　7　月　31　日

計畫主持人： 溫 啟 仲
共同主持人： 張 憶 壽
計畫參與人員： 郭 育 成　黃 詠 詳

成果報告類型(依經費核定清單規定繳交)：■精簡報告　□完整報告

本成果報告包括以下應繳交之附件：
□赴國外出差或研習心得報告一份
□赴大陸地區出差或研習心得報告一份
■出席國際學術會議心得報告及發表之論文各一份
□國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、
　　　　　列管計畫及下列情形者外，得立即公開查詢
　　　　　□涉及專利或其他智慧財產權，□一年□二年後可公開查詢

執行單位： 淡 江 大 學 數 學 系

中　華　民　國　97　年　10　月　15　日

計畫名稱：

Nonparametric Bayes estimation in Cox model using Bernstein priors.
(應用伯氏先驗分布在柯斯模型之無母數貝氏估計)

中文摘要：

對於 Cox 存活分析模型，我們使用在回歸參數上的一有母數先驗分布及在基線累積風險函數上的一無母數伯氏先驗分布來研究其無母數貝氏估計。我們研究提出之先驗分布的支集，建立一有效計算後驗分布的演算法。除此之外，我們進行模擬試驗來評量此貝氏統計方法的數值表現，並與傳統之無母數最大概然估計做一比較。結果顯示此一貝氏統計方法提供 Cox 存活分析模型中的參數較佳之估計，而且此貝氏估計是不易受不同之先驗分布選擇所影響。

英文摘要：

We study the nonparametric Bayesian estimation in the Cox model using a parametric prior for regression parameter and a nonparametric Bernstein polynomial based prior for baseline cumulative hazard function. We study the support of the proposed prior and develop an efficient algorithm for computing posterior distribution. In addition, we conduct simulation studies to evaluate the performance of this method and compare our Bayes estimates with the classical nonparametric maximum likelihood estimates. The results indicate that this Bayesian approach provides better estimates of the parameters in the Cox model and the Bayes estimates seem to be insensitive to the choice of prior.

關鍵詞

Cox model; geometric prior; Markov chain Monte Carlo; random Bernstein polynomials;
柯斯模型; 幾何先驗分布; 馬可夫鏈摩地卡羅; 隨機伯氏多項氏

報告內容

一 前言、研究目的、文獻探討

Nonparametric Bayesian methods in survival analysis require a prior random process that generates a baseline hazard function or cumulative hazard function. Kalbfleisch (1978) used a gamma process to model the cumulative hazard function, Dykstra and Laud (1981) employed a gamma process to model the hazard function, and Hjort (1990) and Damien et al. (1996) used a beta process to model the cumulative hazard function. These prior processes imply an independent increment property for the cumulative hazard functions, which is often unrealistic. Modeling dependence between hazard increments have been studied by several authors, Including Gamerman (1991), Arjas & Gasbarra (1994), and McKeague & Tighiouart (2000, 2002), among others. In particular, McKeague & Tighiouart (2000, 2002) model the hazard rate as a step function where the jump times form a time homogeneous Poisson process and the levels form a Markov random field. An extensive review of the prior random processes used for survival analysis can be found in Sinha & Dey (1997, 1998) and Ibrahim et al. (2001).

More recently, Chang et al. (2005) discussed the use of Bernstein polynomials in specifying prior for cumulative hazard functions and provided a smooth Bayes estimate of a convex cumulative hazard function. They introduced the Bernstein priors, which have large supports, select only smooth cumulative hazard function, and can easily incorporate certain geometric information into the priors. For integers $0 \le i \le k$, let $\varphi_{ik}(t) = C_i^k t^i (1-t)^{k-i}$, where $C_i^k = k!/(i!(k-i)!)$. Then $\{\varphi_{ik} \mid i = 0,...,k\}$ is the so-called Bernstein basis for polynomials of degree up to *k*. We note that these Bernstein bases provide simple polynomial approximations to continuous functions on compact subsets of R^1 and efficient representation for polynomial curves in terms of geometrically important properties.

Our goal is to study the nonparametric Bayesian estimation in the Cox model using a parametric prior for the regression coefficient and a nonparametric Bernstein prior for the baseline cumulative hazard function. We note that Chang et al. (2005) discussed the survival models without covariates. We propose available MCMC algorithms to generate samples from joint posterior distributions of regression coefficients and baseline cumulative hazard functions and conduct simulation studies to evaluate the performance of this method. In addition, we compare our nonparametric Bayes estimates with the classical nonparametric maximum likelihood estimates (NPMLE).

二 研究方法

1) Bernstein polynomials

Let Λ be a continuous function on $[0,\tau]$. Let $b_{ik}=\Lambda(i\tau/k)$ for $i=0,...,k,$ and $b_k=(b_{0k},...,b_{kk})$. Then $\sum_{i=0}^{k} b_{ik}\varphi_{ik}(t/\tau)$, denoted by $\Lambda_{b_k}(t),$ is the *k*th order Bernstein polynomial of Λ. We know from the Bernstein–Weierstrass approximation theorem that Λ_{b_k} converges uniformly to Λ on $[0,\tau]$ as *k* goes to infinity. Much of the geometry of Λ is preserved by its Bernstein polynomials and very much of the geometry of a Bernstein polynomial can be read off from its coefficients. These observations suggest the possibility of introducing a prior on the class of the baseline cumulative hazard functions by assigning a probability measure on the coefficients (k,b_k) through the Bernstein polynomials.

Let $\Lambda_a(t)=\sum_{i=0}^{k} a_i\varphi_{ik}(t/\tau)$. The following proposition concerns the relation between the shape of Λ_a and its coefficients $a_0,a_1,...,a_k$. Proposition 1 provides sufficient conditions under which Λ_a enjoys certain geometric properties, and Proposition 2 characterizes certain geometric properties of a continuous function in terms of its Bernstein polynomials. Notice that the linear combination of functions Λ_a in each statement of Proposition 1 still satisfies the same conditions.

Proposition 1

1. If $a_0\le a_1\le ...\le a_k$, then $\Lambda_a^{'}(\cdot)\ge 0$ on $[0,\tau]$.

2. If $a_{i+1}-2a_i+a_{i-1}\ge 0$ for every *i=1,…,k-1*, then $\Lambda_a^{''}(\cdot)\ge 0$ on $[0,\tau]$.

Let $I_k=\{\Lambda_a: 0=a_0\le a_1\le\cdots\le a_k\}$ and $I_k^{(c)}=\{\Lambda_a:\Lambda_a\in I_k, a_{i+1}+a_{i-1}\ge 2a_i,\ \text{for } i=1,\cdots,k-1\}$.

Then we have

Proposition 2

1. *Let D consist of linear combinations of elements in* $\bigcup_{k=1}^{\infty} I_k$ *, with non-negative*

coefficients. Then the closure of D in uniform norm is precisely the set of all increasing and continuous functions Λ *on* $[0,\tau]$ *with* $\Lambda(0)=0.$

2. *Let* $D^{(c)}$ *consist of linear combinations of elements in* $\bigcup_{k=2}^{\infty} I_k^{(c)}$ *, with non-negative coefficients. Then the closure of* $D^{(c)}$ *in uniform norm is precisely the set of all increasing and convex functions* Λ *on* $[0,\tau]$ *with* $\Lambda(0)=0.$

2) Cox's proportional hazards model based on Bernstein polynomials

The Cox's proportional hazards model assumes that the cumulative hazard rate of a subject with *p*-dimensional covariate *z* at time *t* is given by

$$e^{\beta' z}\Lambda(t), \tag{1}$$

where $\beta \in R^p$ is the regression coefficient of covariate and Λ is the baseline hazard function. The goal is to study the nonparametric Bayes estimator of (β,Λ).

A framework for Bayesian survival analysis would specify a prior on the space of unknown regression coefficients and unknown baseline cumulative hazard functions whose restrictions to the study period $[0,\tau]$ are of the form

$$\Lambda(t;k,b_k)=\sum_{i=0}^{k} b_{ik}\varphi_{ik}(t/\tau), \tag{2}$$

where *k* is a positive integer and $b_k=(b_{0k},b_{1k},...,b_{kk})$. When the order of the polynomial needs not to be stressed, we denote $\Lambda(t;k,b_k)$ by $\Lambda_{b_k}(t)$.

The prior can then be specified by assigning a probability measure π on the $\theta=(\beta,k,b_k)$. A convenient way to elicit the prior is to consider

$$\pi(\beta,k,b_k)=\pi_1(\beta)\pi_2(k)\pi_3(b_k\,|\,k). \tag{3}$$

Specifically, we take π_1 to be a *p* dimensional multivariate normal density, with mean μ and covariance matrix c_0W, where μ and c_0 are specified scalars and *W* is a known *pxp* diagonal matrix. In practice, we can take *W* to be a diagonal matrix consisting of the sample variance of the covariates. Since larger values of c_0 make π_1 more noninformative, we conducted a sensitivity analysis with two values of c_0 to examine the impact on posterior quantities of interest.

Let *q* be a density with support containing the value of the true baseline cumulative hazard rate at time τ . Let $b_{0,k}=0,\ \ b_{k,k}$ be generated from *q*, and $b_{1,k}\le b_{2,k}\le\cdots\le b_{k-1,k}$ be the order statistics of a random sample of size $k-1$ from $unif\left(0,b_{k,k}\right)$; this joint distribution of $\left(b_{0,k},\cdots b_{k,k}\right)$ is the conditional density of $\pi_3\left(b_{0,k},b_{1,k},\cdots b_{k,k}\mid k\right)$ to be used in the prior. In view of Proposition 1 and Proposition 2, we know that the associated $\Lambda_{b_k}\left(\cdot\right)$ is smooth and increasing on $\left[0,\tau\right]$.

In practice, we set the hyper-parameter μ in π_1 being the maximum likelihood estimate of the regression coefficient and *q* being $unif\left(\Lambda^*-\varepsilon,\Lambda^*+\varepsilon\right)$, where Λ^* is the maximum likelihood estimate of the baseline cumulative hazard function at τ and ε is a suitable positive number.

3) Bayesian Inference

Let $T_i, C_i,$ and Z_i denote, respectively, the event time, censoring time, and covariate of the *i*-th subject. Assume that (T_i, C_i, Z_i) are independent and identically distributed for $i=1,\dots,n,$ and that the data may be subject to right censoring. Then the actually observed data is $\left\{\left(X_1,\delta_1,Z_1\right),\dots,\left(X_n,\delta_n,Z_n\right)\right\}$, where $X_i=\min(T_i,C_i)$ is the study time and $\delta_i=I\left(T_i\le C_i\right)$ is the event indicator. Assume that T_1 and C_1 are conditionally independent given Z_1 and that the distribution of (C_1,Z_1) has nothing to do with $\theta=\left(\beta,k,b_k\right)$. In accordance with (1) and (2), given $\theta=\left(\beta,k,b_k\right)$, the likelihood function for $\left\{\left(X_1,\delta_1,Z_1\right),\dots,\left(X_n,\delta_n,Z_n\right)\right\}$ is

$$\prod_{i=1}^{n}\left(\left[\lambda_{b_k}\left(X_i\right)e^{\beta' z_i}\right]^{\delta_i}\exp\left[-\Lambda_{b_k}\left(X_i\right)e^{\beta' z_i}\right]\right),$$

where $\lambda_{b_k}\left(\cdot\right)$ is the derivative of $\Lambda_{b_k}^{'}\left(\cdot\right)$. The posterior density v of $\left(\beta,k,b_k\right)$ given data is proportional to

$$\left[\prod_{i=1}^{n}\left(\left[\lambda_{b_k}\left(X_i\right)e^{\beta' z_i}\right]^{\delta_i}\exp\left[-\Lambda_{b_k}\left(X_i\right)e^{\beta' z_i}\right]\right)\right]\pi_1\left(\beta\right)\pi_2\left(k\right)\pi_3\left(b_k \mid k\right).$$

We propose an algorithm to generate samples from the posterior distribution for the proposed prior. Since the parameter space consists of subspaces of different dimension, we use the reversible jump Metropolis-Hastings (RJMH) algorithm, as discussed in Green (1995), to generate posterior distribution for inference. The details are not presented here.

三 結果與討論

（1）Support of the prior

The following proposition regards the support of the prior.

Proposition 3 Assume π_1 has support *B*, $\pi_2\left(k\right)>0$ for every $k=1,2,\cdots$, and the conditional density $\pi_3\left(b_k \mid k\right)$ of $\pi\left(\cdot\,|\mathrm{B}\times\left\{k\right\}\times B_k\right)$ has support B_k on an infinite subsequence of k. Let Λ be a given increasing and continuous function on $[0,\tau]$ with $\Lambda\left(0\right)=0$ and $\beta\in B$. Then $\pi\{(\beta',k,b_k)\in \mathrm{B}\times B:\ \|\Lambda_{b_k}-\Lambda\|_\infty+\|\beta'-\beta\| < \varepsilon\}>0$ for every $\varepsilon>0$. Here $\|\cdot\|_\infty$ denotes the sup-norm over $[0,\tau]$ and $\|\cdot\|$ denotes the Euclidean norm.

Proof. Let $\Lambda_{(k)}(t)=\sum_{i=0}^{k}\Lambda(i\tau/k)\varphi_{i,k}(t/\tau)$. Let $\varepsilon>0$ be given. Using the Bernstein-Weierstrass Theorem, see, for example, Altomare and Campiti, 1994, we can choose a large k_1 so that $|\Lambda_{(k_1)}(t)-\Lambda(t)|\le\varepsilon/3$ for every $t\in[0,\tau]$. Combining this with the fact that $\|\Lambda_{b_k}-\Lambda\|_\infty\le \max_{i=0,\cdots k}|b_{i,k}-\Lambda(i\tau/k)|$ for $b_k=\left(b_{0,k},b_{1,k},\cdots b_{k,k}\right)\in B_k$, we get

$$\pi\{(\beta',k,b_k)\in \mathrm{B}\times B:\ \|\Lambda_{b_k}-\Lambda\|_\infty+\|\beta'-\beta\|\ \ <\varepsilon\}$$

$$\geq \pi\{(\beta',k,b_k)\in \mathrm{B}\times B:\ \|\Lambda_{b_k}-\Lambda_{(k_1)}\|_\infty<\frac{\varepsilon}{3},\|\beta'-\beta\|\ \ <\frac{\varepsilon}{3}\}$$

$$\geq \pi\{(\beta',k_1,b_k)\in \mathrm{B}\times\{k_1\}\times \mathrm{B}_{k_1}:\ \max_{i=0,\cdots,k_1}|b_{i,k_1}-\Lambda(i\tau/k_1)\|<\frac{\varepsilon}{3},\|\beta'-\beta\|\ \ <\frac{\varepsilon}{3}\},$$

which is positive, because of the assumptions on $\pi_2(k_1)$ and $\pi_1(\beta')$. This completes the proof.

（2）Simulation studies

These simulation studies serve to show that the numerical performance of the estimates is excellent and the Bayes estimate are insensitive to the choice of the prior. There are 100 replicates in each simulation study scenario and each replicate is a random sample with sample size 30 or 50. We set $\beta=0$ or 1; $\Lambda(t)=t^{1.5}$; $\tau=2$; the censoring variable C_1 is exponentially distributed with mean 2; the distribution of the corvariate Z_1 is uniform{0; 1}.To specify the prior distribution, we set $k_0=10$; $\alpha=6$; $\varepsilon=1$, and π_1 has variance c_0 with c_0 being 1 or 100. We note that the mean μ of π_1 and Λ^* in π_3 are specified by the maximum likelihood estimates.

Viewing the posterior as a distribution on regression coefficients and baseline cumulative hazard function through (2), we calculate the Bayes estimate, the sample mean $(\hat{\beta},\hat{\Lambda})$ from the posterior distribution; namely

$$(\hat{\beta},\hat{\Lambda}(\cdot))=\left(\frac{1}{J}\sum_{j=1}^{J}\beta^{(j)},\frac{1}{J}\sum_{j=1}^{J}\Lambda_{b(j)}(\cdot)\right),$$

where $(\beta^{(1)},b^{(1)}),(\beta^{(2)},b^{(2)}),\cdots$ are chosen randomly from $\mathrm{B}\times B$ according to the posterior distribution.

Table 1 and Table 2 summarize the results of the simulation studies. Table 1 gives the results with sample size 30 and Table 2 gives the results with sample size 50. The first column of Table 1 lists the true values of the regression parameter and the second column indicates the value c_0 used in our prior. The third, fourth and fifth columns report respectively the averaged bias (Bias), sample standard deviation (SD) and sample mean-squared error (MSE) of the 100 estimates of β. The performance of baseline cumulative hazard is carried out by comparing $\hat{\Lambda}$ with the true baseline cumulative hazard function Λ in sup-norm $\|\hat{\Lambda}-\Lambda_0\|_\infty$ and the averaged square error

$$\mathrm{ASE}(\hat{\Lambda})=\frac{1}{\mathrm{n}}\sum_{i=1}^{n}[\hat{\Lambda}(X_i)-\Lambda(X_i)]^2,$$

with n being 30 or 50. The sixth column of Table 1 reports the average of $\|\hat{\Lambda}-\Lambda_0\|_\infty$ of the 100 estimates; the final column gives the mean averaged squared error (MASE) of the 100 estimates. Numbers in brackets are the corresponding results obtained using maximum likelihood approach. The entries in Table 2 bear the same meaning as those in Table1.

(3) Results

It is clear from Table 1 and Table 2 that our Bayes estimates seem to be insensitive to the choice of c_0. In the estimation of regression parameter, both maximum likelihood method and Bayesian method work nicely, although our method seems to perform a little better in terms of mean squared error; in the estimation of baseline cumulative hazard, our method outperforms.

Table 1. Simulation study with sample size $n = 30$. Numbers in brackets are estimates using maximum likelihood approach, others are using our method.

True value	c_0	Bias	SD	MSE	$\|\| \hat{\Lambda} - \Lambda_0 \|\|_\infty$	MASE
$\beta = 0$	1	-0.0264	0.5042	0.2549	0.7813	0.1314
	100	-0.0550	0.5001	0.2531	0.7867	0.1340
		[0.0484]	[0.5380]	[0.2918]	[1.2351]	[0.1829]
$\beta = 1$	1	0.0309	0.4719	0.2236	0.6403	0.0558
	100	-0.0052	0.4405	0.1940	0.6274	0.0537
		[0.1240]	[0.5287]	[0.2949]	[0.9381]	[0.0818]

Table 2. Simulation study with sample size $n = 50$. Numbers in brackets are estimates using maximum likelihood approach, others are using our method.

True value	c_0	Bias	SD	MSE	$\|\| \hat{\Lambda} - \Lambda_0 \|\|_\infty$	MASE
$\beta = 0$	1	-0.0369	0.3615	0.1320	0.6980	0.0643
	100	-0.0371	0.3699	0.1382	0.7089	0.0663
		[-0.0018]	[0.3769]	[0.1420]	[0.9973]	[0.0863]
$\beta = 1$	1	-0.0207	0.3486	0.1220	0.5872	0.0347
	100	-0.0314	0.3301	0.1100	0.6003	0.0353
		[0.0522]	[0.4085]	[0.1696]	[0.9300]	[0.0542]

（4）Concluding remarks

We have proposed a Bayesian approach to Cox's regression using a parametric prior for regression parameter and a nonparametric Bernstein polynomial based prior for baseline cumulative hazard function. Our simulation studies indicate that the numerical performance of this method is more excellent than the classical maximum likelihood method and that our Bayes estimates seem to be insensitive to the choice of prior.

参考文獻

1. Altomare, F. & Campiti, M. (1994). *Korovkin-type approximation theory and its applica-tion*. W. de Gruyter, Berlin.
2. Arjas, E. & Gasbarra, D. (1994). Nonparametric Bayesian inference from right censored survival data using the Gibbs sampler. Statist. Sinica 4, 505-524.
3. Chang, I. S., Hsiung, C. A., Wu, Y. J., and Yang, C. C. (2005). Bayesian survival analysis using Bernstein polynomials. Scand. J. Statist. 32, 447-466.
4. Damien, P., Laud, P. W. and Smith, A. F. M. (1996). Nonparametric Inference Based on Beta Processes. Scand. J. Statist. 23, 27-36,
5. Dykstra, R. L. & Laud, P. (1981). A Bayesian nonparametric approach to reliability, Ann. Stat. 9, 356-367.
6. Gammerman, D. (1991). Dynamic Bayesian models for survival data. Applied Statist. 40,63-79.
7. Green, P. G. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82, 711-732.
8. Hjort, N. L. (1990). Nonparametric Bayes estimators based on beta process in models for life history data. Ann. Statist. 18, 1259-1294.
9. Ibrahim, J. G., Chen, M. H. & Sinha D. (2001). Bayesian survival analysis. Springer-Verlag, New York.
10. Kalbfleisch, J. D. (1978) Nonparametric Bayesian analysis of survival time data, J. Royal Stat. Soc. Ser. B 40, 214-221.
11. McKeague, I. W. & Tighiouart, M. (2000). Bayesian estimators for conditional hazard functions. Biometrics 56, 1007-1015.
12. Ibrahim, J. G., Chen, M. H. \& Sinha D. (2001). Bayesian survival analysis. Springer-Verlag, New York.

計畫成果自評

完成原計畫預期之工作項目詳列如下

- ☐ Study the support of the proposed joint prior distributions of the regression coefficient and the baseline cumulative hazard.
- ☐ Develop efficient algorithms for computing the joint posterior distributions of the regression parameter and the baseline cumulative hazard.
- ☐ Conduct simulation studies to evaluate the numerical performance of our methods.
- ☐ Compare our nonparametric Bayes estimates with the nonparametric maximum likelihood estimates.

研究成果之學術或應用價值

We provided a Bayesian approach to Cox's regression model using a parametric prior for regression parameter and a nonparametric Bernstein polynomial based prior for baseline cumulative hazard function. Simulation studies show that this Bayesian approach can provide better estimates of the parameters than the classical nonparametric maximum likelihood estimates.

是否適合在學術期刊發表

We will submit our study to an appropriate journal.