Published in final edited form as:
J Stat Plan Inference. 2009 September 1; 139(9): 3132–3141. doi:10.1016/j.jspi.2009.02.017.

Confidence Estimation via the Parametric Bootstrap in Logistic Joinpoint Regression

Ryan Gill[†,*], **Grzegorz A. Rempala**[‡], and **Michal Czajkowski**[†]
[†]Department of Mathematics, University of Louisville, Louisville, KY 40292 USA
[‡]Department of Biostatistics, Medical College of Georgia, Augusta, GA 30902 USA

Abstract

We consider asymptotic properties of the maximum likelihood and related estimators in a clustered logistic joinpoint model with an unknown joinpoint. Sufficient conditions are given for the consistency of confidence bounds produced by the parametric bootstrap; one of the conditions required is that the true location of the joinpoint is not at one of the observation times. A simulation study is presented to illustrate the lack of consistency of the bootstrap confidence bounds when the joinpoint is an observation time. A removal algorithm is presented which corrects this problem, but at the price of an increased mean square error. Finally, the methods are applied to data on yearly cancer mortality in the United States for individuals age 65 and over.

Keywords

logistic joinpoint regression; confidence estimation; parametric bootstrap; maximum likelihood; mortality trends

1 Introduction

There are a wide variety of statistical methods for analyzing nonlinear models. If one is interested only in summarizing the trends in the data and in obtaining a good flexible non-linear fit, then one may take advantage of many types of spline models existing in the literature (see [10,11,13,14,[20],21,23,36,42] and the references therein). On the other hand, if one's interest lies mainly in estimating and making inferences on the location of structural changes in the underlying model, the model frequently considered is that of a segmented regression in which the knots (also referred to as joinpoints) are unknown (see, for instance, [15,16,25,26, 27,29,31,37]).

Segmented regression models are popular, for instance, as tools in modeling general disease trends and originally have been introduced in the context of epidemiological studies of occupational exposures for modeling threshold limit values in logistic regression models with a single joinpoint (see [18,41]). Subsequently, various multiple joinpoint algorithms have also been applied to disease trend models. For instance, Kim et al. [28] suggested a sequential *backward selection* algorithm for testing the number of joinpoints in a model that uses the

*Corresponding author. E-mail: rsgill01@louisville.edu.

least-squares criteria under squared-error loss and a goodness-of-fit measure based on the *F*-statistic, and applied the algorithm to model U.S. yearly cancer rates. The algorithm has been implemented in the free software *Joinpoint*, version 3.3, (see http://srab.cancer.gov/joinpoint/) which facilitates fitting and testing the model for Gaussian and Poisson regressions. Czajkowski et al. [7] compared the above algorithm with a *forward selection* algorithm in the logistic joinpoint regression setting with multiple joinpoints where model parameters were estimated by maximum likelihood and applied the methods to model longitudinal data on cancer mortality in a cohort of chemical workers. An R package ljr [6] available at http://www.R-project.org [34] has been developed implementing both the backward and forward algorithms for the logistic joinpoint model. Many alternative approaches for selecting the number of change points based on information theory have been also considered in a variety of contexts (see [8,30,32,40,43] and the references therein).

When working with statistics having complicated distributions such as those encountered in segmented regression models, the bootstrap and parametric bootstrap are common tools used for confidence estimation (see, for instance, [4,9,12,22]). In many cases, the bootstrap is effective, but there are also cases where the bootstrap is not consistent when parameters are on the boundary of the parameter space. Several examples are discussed in [1]. In this paper, we are interested in consistency of parametric bootstrap confidence bounds in the context of the logistic joinpoint regression. In particular, we describe herein a situation in which the consistency of the bootstrap confidence bounds fails, in the sense that they are not asymptotically correct. We refer to Section 7.4 of [39] for a general discussion of the bootstrap confidence bounds. The material in this paper discusses the behavior of the bootstrap for a segmented regression model; however, see [24] for a detailed discussion of the asymptotic behavior of the bootstrap for detecting changes in a multiphase linear regression model which differs from ours in that it does not impose continuity constraints.

The paper is organized as follows. In Section 2, we introduce a clustered logistic joinpoint model and discuss maximum likelihood estimation of its parameters. The consistency and asymptotic normality of the maximum likelihood and related estimators are discussed in Section 3. Section 4 gives sufficient conditions for the consistency of the parametric bootstrap, illustrates via simulation a situation when the parametric bootstrap is not consistent, and suggests a removal algorithm to restore consistency for the simulated example. Finally, in Section 5, the methods proposed are applied to data on yearly cancer mortality in the United States for individuals age 65 and over.

2 Clustered Logistic Joinpoint Model

Suppose that $Y_1,\dots, Y_N$ are independent Binomial random variables such that Y_i is the sum of m_i independent Bernoulli random variables each with probability of success p_{0i}. Denote the realizations of these random variables as $y_1,\dots, y_N$, respectively, let $\boldsymbol{y} = [y_1,\dots, y_N]^T$, and let $n=\sum_{i=1}^{N} m_i$. Furthermore, let p_{0i} have the functional form

$$p_{0i}=\frac{e^{\eta_{0i}}}{1+e^{\eta_{0i}}} \tag{1}$$

where

$$\eta_{0i}=\alpha_0+\beta_0 t_i+\delta_0(t_i-\tau_0)_+ +\gamma_0^{T}\boldsymbol{x}_i \tag{2}$$

and $t_+ = \max\{t, 0\}$. Here $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N$ are fixed q-dimensional covariates, and the 'times' $t_1 \le \ldots \le t_N$ are ordered covariates. Also, α_0 is the unknown intercept, τ_0 is the unknown joinpoint, β_0 is the unknown slope coefficient for t_i before τ_0, $\delta_0 \neq 0$ is the change in the slope coefficient after τ_0, and γ_0 is the unknown q-dimensional vector of coefficients for the fixed covariates.

2.1 Maximum Likelihood Estimation

In this section, the maximum likelihood estimate of $\boldsymbol{\theta}_0=[\alpha_0,\beta_0,\tau_0,\delta_0,\gamma_0^{\mathrm{T}}]^{\mathrm{T}}$ is derived. The log-likelihood function for the sample $y_1,\ldots, y_N$ is given by

$$l(\boldsymbol{\theta})=\sum_{i=1}^{N}\left\{\eta_i(\boldsymbol{\theta})y_i - m_i \ln(1+e^{\eta_i(\boldsymbol{\theta})})\right\}$$

where $\boldsymbol{\theta} = [\alpha, \beta, \tau, \delta, \gamma^{\mathrm{T}}]^{\mathrm{T}}$ and

$$\eta_i(\boldsymbol{\theta})=\begin{cases} \alpha+\beta t_i+\gamma^{\mathrm{T}}\boldsymbol{x}_i & \text{if } t_i \le \tau \\ \alpha+\beta t_i+\delta(t_i-\tau)+\gamma^{\mathrm{T}}\boldsymbol{x}_i & \text{if } t_i > \tau \end{cases}.$$

It is useful to consider the *jth super log-likelihood function*

$$\begin{aligned} l_j(\boldsymbol{\theta}) &= \sum_{i=1}^{j}\left\{(\alpha+\beta t_i+\gamma^{\mathrm{T}}\boldsymbol{x}_i)y_i - m_i \ln(1+e^{\alpha+\beta t_i+\gamma^{\mathrm{T}}\boldsymbol{x}_i})\right\} \\ &\quad+\sum_{i=j+1}^{N}\left\{(\alpha+\beta t_i+\delta(t_i-\tau)+\gamma^{\mathrm{T}}\boldsymbol{x}_i)y_i - m_i \ln(1+e^{\alpha+\beta t_i+\delta(t_i-\tau)+\gamma^{\mathrm{T}}\boldsymbol{x}_i})\right\} \end{aligned} \tag{3}$$

for $j = 1,\ldots,N - 1$. Note that

$$l_j(\boldsymbol{\theta})=l(\boldsymbol{\theta}) \text{ if } \tau \in [t_j, t_{j+1}]$$

and l_j is infinitely differentiable with respect to $\boldsymbol{\theta}$, but l is not differentiable at $\tau = t_i$ for $i = 1,\ldots,N$ if $\delta \neq 0$. Letting $\boldsymbol{\phi}(\boldsymbol{\theta}) = [\alpha, \beta, \alpha - \delta\tau, \beta + \delta, \gamma^{\mathrm{T}}]^{\mathrm{T}}$, we can express (3) as

$$l_j(\boldsymbol{\theta})=\sum_{i=1}^{N}\left\{\phi^{\mathrm{T}}(\boldsymbol{\theta})z_i^{(j)}y_i - m_i \ln\left(1+e^{\phi^{\mathrm{T}}(\boldsymbol{\theta})\boldsymbol{z}_i^{(j)}}\right)\right\}$$

where $\mathbf{z}_i^{(j)}=[1-\upsilon_i^{(j)},(1-\upsilon_i^{(j)})t_i,\upsilon_i^{(j)},\upsilon_i^{(j)}t_i,\boldsymbol{x}_i^{\mathrm{T}}]^{\mathrm{T}}$ and

$$\upsilon_i^{(j)}=\begin{cases} 0 & \text{if } i \le j \\ 1 & \text{if } i > j \end{cases}.$$

This is a logistic regression model with the design matrix $\boldsymbol{Z}^{(j)}=[\boldsymbol{z}_1^{(j)},\ldots,\boldsymbol{z}_N^{(j)}]^{\mathrm{T}}$ and the canonical parameter vector $\boldsymbol{\phi}(\boldsymbol{\theta})$.

The function l_j is an infinitely differentiable function with respect to $\boldsymbol{\phi}$. Differentiating l_j with respect to $\boldsymbol{\phi}$, we have

$$\frac{\partial l_j}{\partial \phi}=\sum_{i=1}^{N} \boldsymbol{z}_i^{(j)} \left\{ y_i - m_i \rho_i^{(j)}(\boldsymbol{\theta}) \right\} \tag{4}$$

where $\rho_i^{(j)}(\boldsymbol{\theta})=e^{\eta_i^{(j)}(\boldsymbol{\theta})} / \left(1+e^{\eta_i^{(j)}(\boldsymbol{\theta})}\right)$ with $\eta_j^{(j)}(\boldsymbol{\theta})=\phi^{\mathrm{T}}(\boldsymbol{\theta})\boldsymbol{z}_i^{(j)}$. Setting (4) to zero, dividing by n, and simplifying, we obtain

$$\sum_{i=1}^{N} u_i \boldsymbol{z}_i^{(j)} \left(\frac{y_i}{m_i} - \rho_i^{(j)}(\boldsymbol{\theta}) \right) = \mathbf{0} \tag{5}$$

where $u_i = m_i/n$. If the matrix $\boldsymbol{Z}^{(j)}=[\boldsymbol{z}_1^{(j)}, \ldots, \boldsymbol{z}_N^{(j)}]^{\mathrm{T}}$ is of full rank, then the solution to (5) is the unique maximizer of l_j since

$$-\frac{\partial^2 l_j}{\partial \phi \partial \phi^{\mathrm{T}}}=n(\boldsymbol{Z}^{(j)})^{\mathrm{T}} \boldsymbol{U} \boldsymbol{W}^{(j)}(\boldsymbol{\theta}) \boldsymbol{Z}^{(j)}$$

is positive definite, where $\boldsymbol{W}^{(j)}(\boldsymbol{\theta})$ is a diagonal matrix with diagonal elements $\rho_i^{(j)}(\boldsymbol{\theta})(1 - \rho_i^{(j)}(\boldsymbol{\theta}))i = 1,\ldots,N$ and

$$U \text{ is a diagonal matrix with diagonal elements } u_1, \ldots, u_N. \tag{6}$$

For each $j = 1,\ldots,N - 1$, denote the maximizer of l_j as

$$\widehat{\boldsymbol{\theta}}_n^{(j)}=\left[\widehat{\alpha}_n^{(j)}, \widehat{\beta}_n^{(j)}, \widehat{\tau}_n^{(j)}, \widehat{\delta}_n^{(j)}, (\widehat{\boldsymbol{\gamma}}_n^{(j)})^{\mathrm{T}}\right]^{\mathrm{T}}. \tag{7}$$

Using the invariance property of the method of maximum likelihood (for example, see [33]) $\widehat{\boldsymbol{\theta}}_n^{(j)}=\phi^{-1}(\widehat{\phi(\boldsymbol{\theta}_n^{(j)})})$ where $\widehat{\phi(\boldsymbol{\theta}_n^{(j)})}$ is the solution to (5).

The following algorithm can be used to compute the maximum likelihood estimate (MLE) $\hat{\boldsymbol{\theta}}_n$ of $\hat{\boldsymbol{\theta}}_0$.

1. For $j = 1,\ldots,N - 1$, compute $\widehat{\boldsymbol{\theta}}_n^{(j)}$. If $\widehat{\tau}_n^{(j)} \in (t_j, t_j+1)$, then compute $l(\widehat{\boldsymbol{\theta}}_n^{(j)})$; otherwise, the MLE of τ_0 is not in (t_j, t_{j+1}) and there is no need to evaluate l at any $\boldsymbol{\theta}$ such that $\tau \in (t_j, t_{j+1})$.
2. For $j = 2,\ldots,N - 1$, fix $\tau = t_j$ so that the model is equivalent to logistic regression with covariates 1, t_i, $(t_i - t_j)+$, and $\boldsymbol{x}_i$, $i = 1,\ldots,N$. Then fit a logistic regression model to obtain respective possible estimates $\tilde{\alpha}_n^{(j)}, \tilde{\beta}_n^{(j)}, \tilde{\delta}_0^{(j)}$, and $\tilde{\gamma}_0^{(j)}$ of α_0, β_0, δ_0, and γ_0. Denote the possible estimate of $\boldsymbol{\theta}_0$ as

$$\tilde{\boldsymbol{\theta}}_n^{(j)}=[\tilde{\alpha}_n^{(j)}, \tilde{\beta}_n^{(j)}, t_j, \tilde{\delta}_n^{(j)}, (\tilde{\gamma}_n^{(j)})^{\mathrm{T}}]^{\mathrm{T}},$$

and compute $l(\widehat{\boldsymbol{\theta}}_n^{(j)})$.

3. The MLE of $\boldsymbol{\theta}_0$ is the value of $\boldsymbol{\theta}$ which maximizes $l(\boldsymbol{\theta})$ among the values at which we evaluate l in steps 1 and 2.

See [7] for a detailed description of an algorithm for computing the MLE in the logistic joinpoint regression model with multiple joinpoints.

3 Asymptotic properties

In this section, the consistency and the asymptotic distribution of the maximum likelihood estimator are established. Consider the function $\boldsymbol{F}^{(j)} : \mathbb{R}^N \times \mathbb{R}^{4+q} \to \mathbb{R}^{4+q}$ such that

$$\boldsymbol{F}^{(j)}(\boldsymbol{p},\boldsymbol{\theta})=\sum_{i=1}^{N} u_i \boldsymbol{z}_i^{(j)}\left(p_i-\rho_i^{(j)}(\boldsymbol{\theta})\right),\ j=1,\ldots,N-1$$

where $\boldsymbol{p} = [p1,\ldots, p_N]^{\mathrm{T}}$ and $p_i \in (0, 1)$ for $i = 1,\ldots,N$. In particular, we will be interested in values of k such that

$$\tau_0 \in [t_k, t_{k+1}]\,. \tag{8}$$

Throughout this paper, k shall be reserved for values such that (8) holds. If $\tau_0 \in (t_k, t_{k+1})$, then k is unique. If $\tau_0 = t_\ell$ for some ℓ, then (8) holds for both $k = \ell - 1$ and $k = \ell$.

Letting $\boldsymbol{p}_0 = [p_{01},\ldots, p_{0N}]^{\mathrm{T}}$, the following lemma uses the Implicit Function Theorem (see, for example, [35]) to express $\boldsymbol{\phi}(\boldsymbol{\theta})$ as a function of $\boldsymbol{p}$. Proofs of lemmas and theorems are deferred to Appendix A.

Lemma 3.1

Suppose that $\boldsymbol{F}^{(k)}(\boldsymbol{p}, \boldsymbol{\theta}) = \boldsymbol{0}$ and (8) holds. If $\boldsymbol{Z}^{(k)}$ is full rank, then in an open neighborhood of $(\boldsymbol{p}_0, \boldsymbol{\theta}_0)$, say $B_{\boldsymbol{p}0}$, there exists a function $\boldsymbol{g}^{(k)}$ such that $\boldsymbol{g}^{(k)}(\boldsymbol{p}) = \boldsymbol{\phi}(\boldsymbol{\theta})$ for $(\boldsymbol{p}, \boldsymbol{\theta}) \in B_{\boldsymbol{p}0}$ and the second partial derivatives of $\boldsymbol{g}^{(k)}$ exist and are continuous.

Implicit differentiation can be used to obtain the explicit form for the partial derivatives of $\boldsymbol{g}^{(k)}$. Differentiating both sides of $\boldsymbol{F}^{(k)}(\boldsymbol{p}, \boldsymbol{\theta}) = \boldsymbol{0}$ with respect to $\boldsymbol{p}^{\mathrm{T}}$, we obtain

$$(\boldsymbol{Z}^{(k)})^{\mathrm{T}}\boldsymbol{U}=(\boldsymbol{Z}^{(k)})^{\mathrm{T}}\boldsymbol{U}\boldsymbol{W}^{(k)}(\boldsymbol{\theta})\boldsymbol{Z}^{(k)}\boldsymbol{G}^{(k)}(\boldsymbol{p})$$

Where $\boldsymbol{G}^{(k)}(\boldsymbol{p})=\dfrac{\partial \boldsymbol{g}^{(k)}(\boldsymbol{p})}{\partial \boldsymbol{p}^{\mathrm{T}}}$ is the matrix of partial derivatives of each component of $\boldsymbol{g}$ with respect to $\boldsymbol{p}^{\mathrm{T}}$. If $\boldsymbol{Z}^{(k)}$ is full rank, then we have

$$\boldsymbol{G}^{(k)}(\boldsymbol{p})=\left((\boldsymbol{Z}^{(k)})^{\mathrm{T}}\boldsymbol{U}\boldsymbol{W}^{(k)}(\boldsymbol{\theta})\boldsymbol{Z}^{(k)}\right)^{-1}(\boldsymbol{Z}^{(k)})^{\mathrm{T}}\boldsymbol{U}.$$

Let $\hat{\boldsymbol{p}}_n = [y_1/m_1,\ldots, y_N/m_N]^{\mathrm{T}}$. Lemma 3.1 can be used to show that the maximizer of l_k is a strongly consistent estimator of $\boldsymbol{\theta}_0$ for any k such that the model is correctly specified. The following lemma states this result for the case when each of the m_i's increases linearly, although the linearity can be relaxed to deal with differing rates of divergence. Let $\widehat{\boldsymbol{\theta}}_n^{(k)}$ be as defined by (7).

Lemma 3.2

For sufficiently large n_0*, suppose there exist constants* $0 < c_1 < c_2$ *such that* $nc_1 \leq m_{\mathrm{i}} \leq nc_2$ *for* $i = 1,\ldots,N$ *when* $n \geq n_0$*. If* $\boldsymbol{Z}^{(k)}$ *is full rank and (8) holds, then* $\widehat{\boldsymbol{\theta}}_n^{(k)} \to \boldsymbol{\theta}_0$ *a.s. as* $n \to \infty$*.*

In this lemma and the following theorems, N is taken to be fixed. This setting is reasonable for analyzing data retrospectively when there is a large amount of data at each observation time or in each time-specific cluster of observations (see, for example, [7]). Since N is fixed, few assumptions are required to achieve consistent estimators. In practice, the assumption that $\boldsymbol{Z}^{(k)}$ is full rank for values of k such that (8) holds is needed so that the parameters are identifiable. In other settings in which N is allowed to increase to infinity, additional assumptions are required on the spacing of the covariates and/or the location of the change-point parameter to ensure that there is enough data in each part of the input space to estimate all of the parameters consistently. However, the specific assumptions required depends on the particular way that N increases. For discussion about typical assumptions for some other change-point problems, see [2] and [32].

Under the conditions of Lemma 3.2, it is seen that the estimator of $\boldsymbol{\theta}_0$ based on k such that the model is correctly specified is consistent. Next, it must be shown that, for each j such that $\tau_0 \notin [t_j, t_{j+1}]$, the maximum value of l_j is less than the maximum value of l_k. To prove this, we introduce the *saturated likelihood function* $s : \mathbb{R}^N \to \mathbb{R}$ defined by

$$s(\boldsymbol{p})=\sum_{i=1}^{N}\left\{\ln\left(\frac{p_i}{1-p_i}\right)y_i+m_i\ln(1-p_i)\right\} \tag{9}$$

which provides a parameter for each distinct set of observed covariates. Using this concept, it can be shown that the maximum likelihood estimator (maximizer of l) is strongly consistent. As before, let $\hat{\boldsymbol{\theta}}_n$ denote the maximum likelihood estimator of $\boldsymbol{\theta}_0$.

Theorem 3.1

Under the conditions of Lemma 3.2 for each k such that (8) holds, it follows that $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ *a.s. as* $n\to\infty$*.*

Next, we examine the asymptotic distribution of $\hat{\boldsymbol{\theta}}_n$. Since consistency has been established, it suffices to consider the distribution(s) of $\widehat{\boldsymbol{\theta}}_n^{(k)}$ with k given by (8).

There are two cases which must be examined. First consider the more complicated one when $\tau_0 = t_\ell$ for some ℓ. In this case, we must consider the joint behavior of $\widehat{\boldsymbol{\theta}}_n^{(\ell-1)}$ and $\widehat{\boldsymbol{\theta}}_n^{(\ell)}$. Denote by $\boldsymbol{\Phi}(\omega)=\frac{\partial\phi^{-1}(\omega)}{\partial\omega^{\mathrm{T}}}$ the $(q+4)\times(q+4)$ matrix of partial derivatives of the components of $\boldsymbol{\phi}^{-1}$. Let $\mathscr{A}(\mathbf{0},\boldsymbol{S})$ denote the (multivariate) normal distribution with zero mean vector and covariance matrix $\boldsymbol{S}$. Also, let $\Rightarrow$ denote convergence in distribution. The following theorem proves the asymptotic normality of the joint distribution of these quantities. The proof is given in Appendix A.

Theorem 3.2

For sufficient large n, suppose there exist constants $0 < c_1 < c_2$ *such that* $nc_1 \le m_i \le nc2$. *If* $\tau_0 = t_\ell$ for some ℓ and $\mathbf{Z}^{(k)}$ *is full rank for* $k = \ell - 1$ and $k = \ell$, then

$$\sqrt{n}\left(\begin{bmatrix} \widehat{\boldsymbol{\theta}}_n^{(\ell-1)} \\ \widehat{\boldsymbol{\theta}}_n^{(\ell)} \end{bmatrix} - \begin{bmatrix} \boldsymbol{\theta}_0 \\ \boldsymbol{\theta}_0 \end{bmatrix}\right) \Rightarrow \mathcal{N}(\mathbf{0}, \psi \mathbf{G}\mathbf{U}^{-1}\mathbf{W}^{(\ell)}\mathbf{G}^{\mathrm{T}}\psi^{\mathrm{T}}) \tag{10}$$

as $n \to \infty$ *where* $\boldsymbol{G}=\boldsymbol{G}(\boldsymbol{p}_0)=\begin{bmatrix} \boldsymbol{G}^{(\ell-1)}(\boldsymbol{p}_0) \\ \boldsymbol{G}^{(\ell)}(\boldsymbol{p}_0) \end{bmatrix}, \psi=\begin{bmatrix} \boldsymbol{\Phi}^{(\ell-1)} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{\Phi}^{(\ell)} \end{bmatrix}, \boldsymbol{\Phi}^{(k)}=\boldsymbol{\Phi}(\boldsymbol{g}^{(k)}(\boldsymbol{p}_0))$ and $\boldsymbol{W}^{(k)} = \boldsymbol{W}^{(k)}(\boldsymbol{\theta}_0)$ *for* $k = \ell - 1$ and $k = \ell$, and $\boldsymbol{U}$ *is defined by (6).*

The result for the second case where $\tau_0 \in (t_k, t_{k+1})$ for some k is given in Theorem 3.3. The proof is omitted since it can be proven in a manner similar to Theorem 3.2. Note that $\widehat{\boldsymbol{\theta}}_n=\widehat{\boldsymbol{\theta}}_n^{(k)}$ *a.s.* when n is sufficiently large, as shown in the proof of Theorem 3.1.

Theorem 3.3

For sufficient large n, suppose there exist constants $0 < c_1 < c_2$ *such that* $nc_1 \le m_i \le nc_2$. *If* $\tau_0 \in (t_k, t_{k+1})$ *for some k and* $\mathbf{Z}^{(k)}$ *is full rank, then*

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \Rightarrow \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Phi}^{(k)}\boldsymbol{G}^{(k)}\boldsymbol{U}^{-1}\boldsymbol{W}^{(k)}(\boldsymbol{G}^{(k)})^{\mathrm{T}}(\boldsymbol{\Phi}^{(k)})^{\mathrm{T}}\right)$$

as $n \to \infty$ *where* $\boldsymbol{G}^{(k)}=\boldsymbol{G}^{(k)}(\boldsymbol{p}0),\ \boldsymbol{\Phi}^{(k)}=\boldsymbol{\Phi}(\boldsymbol{g}^{(k)}(\boldsymbol{p}_0)),\ \boldsymbol{W}^{(k)}=\boldsymbol{W}^{(k)}(\boldsymbol{\theta}_0)$, *and* $\boldsymbol{U}$ *is defined by (6).*

4 Consistency of Bootstrap Confidence Bounds

Suppose we use the parametric bootstrap to generate Binomial random variables Y_{ib}^* with sizes m_i and probabilities of success $\hat{p}_{0i}$ for $i = 1,\ldots,N$ and $b = 1,\ldots, B$ where B is the number of bootstrap samples. Set $\boldsymbol{Y}_b^*=[Y_{1b}^*, \ldots, Y_{Nb}^*]^{\mathrm{T}}$ and let $\widehat{\boldsymbol{\theta}}_{nb}^*$ denote the bth bootstrap replication of $\hat{\boldsymbol{\theta}}_n$, with P and P^* being the probability measures under $\boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\theta}}_n^*$, respectively.

4.1 Consistency when $\tau_0 \ne t_\ell$ for any ℓ

We shall consider the consistency of the *bootstrap percentile method*. For simplicity, we describe the one-sided interval for a parameter ζ_0, but the results extend to two-sided intervals with proper modifications. Set $K_B(x)=P^*(\widehat{\zeta}_n^* \le x)$. Then the bootstrap estimator of the upper bound of the α-level one-sided percentile confidence interval for ζ_0 is $\zeta_{BP}=K_B^{-1}(\alpha)$. The following lemma gives sufficient conditions for the consistency of bootstrap confidence intervals of this form. For a proof, see Theorem 7.9 of [39].

Lemma 4.1—*Let* $H_n(x)=P(\sqrt{n}(\widehat{\zeta}_n - \zeta_0) \le x)$ and $\widehat{H}_B(x)=P^*(\sqrt{n}(\widehat{\zeta}_n^* - \widehat{\zeta}_n) \le x)$. *Suppose that*

1. $\sup_x|H_n(x) - H\hat{}B(x)| = op(1)$
2. $\sup_x|H_n(x) - \Psi(x)| = o(1)$ *for some continuous, strictly increasing and symmetric about zero distribution* $\Psi(x)$

Then $P(\zeta_{BP} \le \zeta_0) \to 1 - \alpha$ *as* $n \to \infty$.

Thus, in order to prove consistency of bootstrap confidence intervals for $\mathbf{a}^T\theta_0$ where $\mathbf{a}$ is a fixed non-zero $(q + 4)$-dimensional vector, we need to check conditions (1) and (2). If $\tau_0 \in (t_k, t_{k+1})$, Theorem 3.3 verifies condition (2) of Lemma 4.1. So, for the case when $\tau_0 \in (t_k, t_{k+1})$, we next verify condition (1) in the following result.

Theorem 4.1—*Under the conditions of Theorem 3.3, it follows that*

$$\sup_x \left| P(\sqrt{n}(\boldsymbol{a}^T\hat{\boldsymbol{\theta}}_n - \boldsymbol{a}^T\boldsymbol{\theta}_0) \le x) - P^*(\sqrt{n}(\boldsymbol{a}^T\hat{\boldsymbol{\theta}}_n^* - \boldsymbol{a}^T\hat{\boldsymbol{\theta}}_n) \le x) \right| \to 0 \; a.s. \tag{11}$$

as $n \to \infty$.

Hence, from the last theorem, we see that the conditions for bootstrap confidence interval consistency given in Lemma 4.1 hold when $\tau_0 \in (t_k, t_{k+1})$ since (11) implies convergence in probability. Continuity of Ψ will not hold if $\tau_0 = t_\ell$ for some ℓ as shown by Theorem 3.2.

4.2 Simulation when $\tau_0 = t_\ell$ for some ℓ

To illustrate the behavior of the bootstrap estimates and confidence bounds when $\tau_0 = t_\ell$ for some ℓ, we perform the following simulation study. Suppose that we have equally-spaced observation times $t_i = i$ for $i = 1,\ldots,N = 7$, no additional covariates ($q = 0$), true coefficient values $\alpha_0 = \beta_0 = 0$ and $\delta_0 = 0.2$, and a joinpoint at $\tau_0 = 4$ for the model specified by (1) and (2). For various choices of $m_1 = \ldots = m_7$, we simulate R = 1, 000, 000 data sets and compute the estimate of τ_0 and the bootstrap estimate of τ_0 for each data set.

One simple way to see that Theorem 4.1 does not hold for this case is to use our simulation to estimate the proportion of times that $\hat{\tau}_n$ equals $\tau_0 = 4$ and compare this with the proportion of times that the bootstrap estimate $\widehat{\tau}_n^*$ equals τ_0. These quantities, denoted by $\hat{P}(\hat{\tau}_n = 4)$ and $\widehat{P}^*(\widehat{\tau}_n^*=4)$, as well as empirical estimates of the mean square error $\widehat{MSE}(\widehat{\tau}_n)=E[(\widehat{\tau}_n-\tau_0)^2]$ and the bootstrap estimate $\widehat{MSE}^*(\widehat{\tau}_n^*)=E^*[(\widehat{\tau}_n^*-\widehat{\tau}_n)^2]$ of this quantity are reported in Table 1 for m_i = 100, 10^5, 10^7, and 10^9 using the contributed R package ljr [6]. Clearly, the bootstrap underestimates the true probability that $\hat{\tau}_n = 4$ as seen in columns 2 and 4 of Table 1 and overestimates the true MSE of $\hat{\tau}_n$ as seen in columns 3 and 5.

The estimate $\hat{P}(\hat{\tau}_n = 4)$ = .2020 when $m_i = 10^9$ agrees with the theoretical value suggested by Theorem 3.2. As $n \to \infty$, note that

$$\sqrt{n}\left(\begin{bmatrix} \widehat{\tau}_n^{(3)} \\ \widehat{\tau}_n^{(4)} \end{bmatrix} - \begin{bmatrix} 4 \\ 4 \end{bmatrix}\right) \Rightarrow \mathcal{N}(\mathbf{0}, \boldsymbol{S})$$

where

$$\boldsymbol{A}=\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

and

$$\boldsymbol{S}=\boldsymbol{A}\psi\boldsymbol{G}\boldsymbol{U}^{-1}\boldsymbol{W}^{(4)}\boldsymbol{G}^T\psi^T\boldsymbol{A}^T=\begin{bmatrix} 2127.2073 & 638.1622 \\ 638.1622 & 2167.1859 \end{bmatrix}.$$

Thus, for large n, we have

$$P(\widehat{\tau}_n=4)=P(\widehat{\tau}_n^{(3)}>4 \text{ and } \widehat{\tau}_n^{(4)}<4) \to .2019703;$$

this bivariate normal probability can be computed using the contributed R package mvtnorm [17].

4.3 Removal Algorithm

Herein we discuss a modified estimator of τ_0 for which bootstrap consistency holds as a remedy for the lack of consistency of the parametric bootstrap when $\tau_0 = t_i$ for some i. First, compute $\hat{\boldsymbol{\theta}}_n$. Then remove the observation(s) at the observation time t_ℓ which is closest to $\hat{\tau}_n$ and re-fit the MLE without the observation(s), denoting the result as

$$\widetilde{\boldsymbol{\theta}}_n=[\widetilde{\alpha}_n, \widetilde{\beta}_n, \widetilde{\tau}_n, \widetilde{\delta}_n, (\widetilde{\gamma}_n)^{\mathrm{T}}]^{\mathrm{T}}.$$

If the matrix $\boldsymbol{Z}^{(k)}$, with k defined by (8), is still full rank without the observation(s) and m_i satisifies the condition in Lemma 3.2, then bootstrap consistency holds for this algorithm even if $\tau_0 = t_\ell$ for some ℓ (since these observations will be removed with probability 1 as $n \to \infty$).

As an illustration of the effect of using this algorithm, we use the same simulated data sets considered in Section 4.2 and compare the results for the bootstrap based on the method of maximum likelihood with the bootstrap based on the removal algorithm. Letting $\hat{\tau}_{(.95)}$ and $\widetilde{\tau}_{(.95)}$ represent the 95th percentiles of the empirical distributions of $\hat{\tau}_n$ and $\widetilde{\tau}_n$, Table 2 computes the estimated tail probabilities $\widehat{P}^*(\widehat{\tau}_n^* - \widehat{\tau}_n \geq \widehat{\tau}_{(.95)})$ and $\widehat{P}^*\left((\widetilde{\tau}_n)^* - \widetilde{\tau}_n \geq \widetilde{\tau}_{(.95)}\right)$ for the bootstrap method. Clearly, the bootstrap overestimates the 95th percentile of the distribution of $\hat{\tau}_n - \tau_0$, but provides a good estimate of $\widetilde{\tau}_n - \tau_0$ even for relatively small sample sizes. However, the tradeoff for obtaining accurate confidence bounds is the loss of information caused by discarding part of the data. This also can be seen in Table 2 by observing that the estimates of the mean square errors for $\widetilde{\tau}_n$ are higher than those for $\hat{\tau}_n$.

5 Example

We now apply the method of clustered logistic joinpoint regression to model yearly cancer mortality in the United States for individuals age 65 and over during the period 1979–1998. The data set was obtained from the CDC wonder database [5], and it includes y_i – the number of deaths in the ith observed year due to neoplasms (ICD-9 codes 140–239), m_i –the population during the ith observed year, and $t_i = 1978.5 + i$ – the midpoint of the ith observed year. So, we use the model given by (1) and (2) with $q = 0$. The observed cancer mortality rates y_i/m_i are plotted versus time in Figure 1.

Table 3 gives the parameter estimates for each of the unknown parameters and the fitted probabilities based on maximum likelihood are illustrated by the solid curve in Figure 1. The estimated joinpoint $\hat{\tau}_n = 1993.686$ supports previous findings which attribute the decrease in cancer mortality to improvements in prevention, detection, and treatments [19].

Next, we obtain estimated 95% confidence bounds for τ based on the parametric bootstrap. Using the estimates listed in Table 3 as our model parameters, we generate $R = 100000$ bootstrap samples and compute the estimate of τ for each sample. Then our estimated confidence bounds are the 2.5 and 97.5 percentiles of the bootstrap distribution of τ. In this manner, we obtain the confidence bounds (1993.371, 1993.930).

Note that the 15th observed year t_{15} = 1993.5 falls within this interval. Thus, there might be a problem with consistency of the bootstrap as discussed in Section 4. We can attempt to remedy this possible bias by repeating our analysis with i = 15 removed from the data set. When the model is fit without the 15th observation, the estimate of the joinpoint changes slightly to $\widehat{\tau_n}=1993.595$. The estimated coefficients change very little (see Table 3) and the fitted probabilities based on the removal algorithm are illustrated by the dashed curve in Figure 1.

The 95% estimated confidence bounds with observation 15 omitted are (1993.318,1993.851). While it is true that the interval still contains 1993.5, this value is no longer in our data set since it has been removed, and it seems reasonable to claim that it is unlikely that we are dealing with a situation where $\tau_0 = t_\ell$ for some ℓ and, thus, the consistency of the bootstrap confidence bounds is more plausible.

6 Summary and Conclusion

After presenting details necessary for the computation of the maximizer of the super log-likelihood functions and the MLE of the parameters in the clustered logistic joinpoint model, we considered the asymptotic properties of these estimators. Sufficient conditions for the consistency of the MLE were given. Asymptotic normality of the MLE was also shown under the same conditions as long as the true location of the joinpoint had not been at one of the observation times. Under this latter proviso we also showed the consistency of the bootstrap confidence bounds.

However, if the true location of the joinpoint was one of the observation times, then it was shown that the joint distribution of the maximizers of indices corresponding to the neighboring intervals was asymptotically normal. A simulation study was performed to illustrate the lack of consistency of the bootstrap method in generating confidence bounds in this case due to asymptotic bias. It was also shown that we could remove this bias and obtain consistent estimates via a *removal algorithm* at the cost of a higher MSE.

Finally, the model and the methods were used to analyze yearly cancer mortality in the United States for individuals age 65 and over. The bootstrap confidence interval included one observation time, so the model was refit without that observation. There is only a slight change in the resulting fitted model, but the consistency of the bootstrap is more plausible with the second fit. An R package ljr [6] capable of fitting these models is available in the contributed packages at http://www.R-project.org [34].

A Proofs

Proof of Lemma 3.1

Since $\boldsymbol{F}^{(k)}$ is infinitely differentiable and $\boldsymbol{F}^{(k)}(\boldsymbol{p}_0, \boldsymbol{\theta}_0) = \boldsymbol{0}$, it remains to verify that the determinant of $\frac{\partial \boldsymbol{F}^{(k)}}{\partial \phi^{\mathrm{T}}}\Big|_{(\boldsymbol{p}0,\boldsymbol{\theta}_0)} = -(\boldsymbol{Z}^{(k)})^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Z}^{(k)}$ is not zero in order to apply the Implicit Function Theorem. Here $\boldsymbol{D} = \boldsymbol{U}\boldsymbol{W}^{(k)}(\boldsymbol{\theta}_0)$ is a $N \times N$ diagonal matrix with diagonal elements $d_i = m_i p_{0i}(1 - p_{0i})/n$ for $i = 1,\ldots,N$. Since $d_i \in (0, 1)$, $\frac{\partial \boldsymbol{F}^{(k)}}{\partial \phi^{\mathrm{T}}}\Big|_{(\boldsymbol{p}0,\boldsymbol{\theta}_0)}$ has a nonzero determinant if and only if $\boldsymbol{Z}^{(k)}$ is full rank.

Proof of Lemma 3.2

The strong law of large numbers implies that $\hat{\boldsymbol{p}}_n \to \boldsymbol{p}_0$ *a.s.* as $n \to \infty$. For all $\boldsymbol{\theta}$, the left side of (5) converges to $\boldsymbol{F}^{(k)}(\boldsymbol{p}_0, \boldsymbol{\theta})$ *a.s.* The continuity of $\boldsymbol{F}^{(k)}$ implies that.

$$\mathbf{0}=\lim_{n\to\infty}\boldsymbol{F}^{(k)}(\widehat{\boldsymbol{p}}_0,\widehat{\boldsymbol{\theta}}_n^{(k)})=\boldsymbol{F}^{(k)}(\widehat{\boldsymbol{p}}_0,\lim_{n\to\infty}\widehat{\boldsymbol{\theta}}_n^{(k)}).$$

It is clear that θ_0 is the unique value of $\boldsymbol{\theta}$ such that $\boldsymbol{F}^{(k)}(\boldsymbol{p}_0, \boldsymbol{\theta}) = \mathbf{0}$ by the same argument as the one used to show that the solution to (5) is a unique maximizer of l_j when y_i/n_i is replaced by p_i. Thus, it follows that $\widehat{\boldsymbol{\theta}}_n^{(k)} \to \boldsymbol{\theta}_0$ *a.s.* as $n\to\infty$.

Proof of Theorem 3.1

In view of Lemma 3.2, it suffices to show that, with probability 1, $l(\widehat{\boldsymbol{\theta}}_n^{(k)})>l(\widehat{\boldsymbol{\theta}}_n^{(j)})$ if $\tau_0 \in [t_k, t_{k+1}]$ but $\tau_0 \notin [t_j, t_{j+1}]$ From Lemma 3.2, we have. $\widehat{\boldsymbol{\theta}}_n^{(k)} \to \boldsymbol{\theta}_0$ *a.s.* Using the continuity of $\rho_i^{(k)}$, it follows that $p_i^{(k)}(\widehat{\boldsymbol{\theta}}_n^{(k)}) \to p_i^{(k)}(\boldsymbol{\theta}_0)=p0_i$ *a.s.*. for $i = 1,\ldots,N$. Hence, we have $l(\widehat{\boldsymbol{\theta}}_n^{(k)}) \to s(\boldsymbol{p}_0)$ *a.s.* as $n\to\infty$.

Let $\mathcal{S}=\{(p_1,\ldots,p_N)\big|p_i\in[0,1], i=1,\ldots,N\}$ be the set of points considered by the saturated model and let

$$\mathcal{A}_\ell=\left\{(p_1,\ldots,p_N)\in\mathcal{S}\middle|\tau\in[t_\ell,t_{\ell+1}],\ p_i=p_i^{(\ell)}(\boldsymbol{\theta}),\ i=1,\ldots,N\right\}$$

be the set of fitted probabilities corresponding to restricting $\tau \in [t_k, tk+1]$. Clearly $\mathscr{A}_k \subset \mathscr{S}$ for all k. Furthermore, $\boldsymbol{p}_0 \in \mathscr{A}_k$ if $\tau_0 \in [t_k, t_{k+1}]$ and $\boldsymbol{p}_0 \notin \mathscr{A}_j$ if $\tau_0 \notin [t_j, t_{j+1}]$. Thus, $\max_{\boldsymbol{\theta}\in\mathscr{A}_j} l(\boldsymbol{\theta}) < s(\boldsymbol{p}_0)$ and $l(\widehat{\boldsymbol{\theta}}_n^{(k)}) \to s(\boldsymbol{p}_0)$ *a.s.* so that $\max_{\boldsymbol{\theta}\in\mathscr{A}_j} l(\boldsymbol{\theta}) < \max_{\boldsymbol{\theta}\in\mathscr{A}_k} l(\boldsymbol{\theta})$ with probability 1.

Consequently, if k is unique, then $\hat{\tau}_n \in (t_k, t_{k+1})$ *a.s.* as $n \to \infty$. If k is not unique, then $\hat{\tau}_n \in (t_{\ell-1}, t_{\ell+1})$ *a.s.* as $n \to\infty$, and we have $\widehat{\boldsymbol{\theta}}_n^{(\ell-1)} \to \boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\theta}}_n^{\ell} \to \boldsymbol{\theta}_0$ *a.s.* as $n \to \infty$ by Lemma 3.2. In either case, it follows that $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ *a.s.* as $n \to \infty$.

Proof of Theorem 3.2

Take any $\mathbf{a} \in \mathbb{R}^{2(q+4)}$ and let $\boldsymbol{g}(\boldsymbol{p})=\begin{bmatrix}\boldsymbol{g}^{(\ell-1)}(\boldsymbol{p})\\ \boldsymbol{g}^{(\ell)}(\boldsymbol{p})\end{bmatrix}$. Using a multivariate Taylor series expansion, we obtain

$$\begin{aligned}\boldsymbol{a}^{\mathrm{T}}\left(\begin{bmatrix}\phi(\widehat{\boldsymbol{\theta}}_n^{(\ell-1)})\\ \phi(\widehat{\boldsymbol{\theta}}_n^{(\ell)})\end{bmatrix}-\begin{bmatrix}\phi(\boldsymbol{\theta}_0)\\ \phi(\boldsymbol{\theta}_0)\end{bmatrix}\right)&=\mathbf{a}^{\mathrm{T}}(\boldsymbol{g}(\widehat{\boldsymbol{p}}_n)-\boldsymbol{g}(\boldsymbol{p}_0))\\ &=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{G}(\boldsymbol{p}_0)(\widehat{\boldsymbol{p}}_n-\boldsymbol{p}_0)+\tfrac{1}{2}(\widehat{\boldsymbol{p}}_n-\boldsymbol{p}_0)^{\mathrm{T}}\boldsymbol{H}(\tilde{\boldsymbol{p}}_n)(\widehat{\boldsymbol{p}}_n-\boldsymbol{p}_0)\end{aligned} \tag{12}$$

where $\boldsymbol{H}(\tilde{\boldsymbol{p}}_n)=\dfrac{\partial\boldsymbol{a}^{\mathrm{T}}\boldsymbol{G}(\tilde{\boldsymbol{p}}_n)}{\partial\boldsymbol{p}}$ is the Hessian matrix of $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{g}$ evaluated at a point $\hat{\boldsymbol{p}}_n$ on the segment connecting $\hat{\boldsymbol{p}}_n$ and $\boldsymbol{p}_0$. Note that

$$\sqrt{n}(\widehat{\boldsymbol{p}}_n-\boldsymbol{p}_0)\Rightarrow\mathcal{N}(\mathbf{0},\boldsymbol{U}^{-1}\boldsymbol{W}^{(\ell)}(\boldsymbol{p}_0)) \tag{13}$$

as $n \to \infty$. Since $\boldsymbol{H}(\hat{\boldsymbol{p}}_n)$ is symmetric, we can apply the singular value decomposition to obtain $\boldsymbol{H}(\hat{\boldsymbol{p}}_n) = \boldsymbol{C}(\hat{\boldsymbol{p}}_n)\boldsymbol{\Lambda}(\hat{\boldsymbol{p}}_n)\boldsymbol{C}(\hat{\boldsymbol{p}}_n)^{\mathrm{T}}$ where $\boldsymbol{C}(\cdot)$ is orthogonal and $\boldsymbol{\Lambda}(\cdot)$ is diagonal with entries $\lambda_1 \geq \ldots \geq \lambda_{2(q+4)}$. Note that λ_1 is bounded as $n \to \infty$. Thus, (13) implies that

$$\frac{\sqrt{n}}{2}(\widehat{\boldsymbol{p}}_n - \boldsymbol{p}_0)^{\mathrm{T}}\boldsymbol{H}(\tilde{\boldsymbol{p}}_n)(\widehat{\boldsymbol{p}}_n - \boldsymbol{p}_0) \leq \frac{\lambda_1}{\sqrt{n}}\| \sqrt{n}(\widehat{\boldsymbol{p}}_n - \boldsymbol{p}_0)\|^2 = o_P(1). \tag{14}$$

Hence, (12), (14), and Slutsky's Theorem (see [3]) imply that

$$\sqrt{n}\boldsymbol{a}^{\mathrm{T}}\left(\begin{bmatrix} \phi(\widehat{\boldsymbol{\theta}}_n^{(\ell-1)}) \\ \phi(\widehat{\boldsymbol{\theta}}_n^{(\ell)}) \end{bmatrix} - \begin{bmatrix} \phi(\boldsymbol{\theta}_0) \\ \phi(\boldsymbol{\theta}_0) \end{bmatrix}\right) \Rightarrow \mathcal{N}(\boldsymbol{0}, \boldsymbol{a}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{U}^{-1}\boldsymbol{W}^{(\ell)}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{a})$$

as $n \to \infty$. So, the Cramér-Wold Criterion (see, for example, [38]) implies that

$$\sqrt{n}\left(\begin{bmatrix} \phi(\widehat{\boldsymbol{\theta}}_n^{(\ell-1)}) \\ \phi(\widehat{\boldsymbol{\theta}}_n^{(\ell)}) \end{bmatrix} - \begin{bmatrix} \phi(\boldsymbol{\theta}_0) \\ \phi(\boldsymbol{\theta}_0) \end{bmatrix}\right) \Rightarrow \mathcal{N}(0, \boldsymbol{G}\boldsymbol{U}^{-1}\boldsymbol{W}^{(\ell)}\boldsymbol{G}^{\mathrm{T}}) \tag{15}$$

as $n \to \infty$. Since $\widehat{\boldsymbol{\theta}}_n^{(k)} = \phi^{-1}(\phi(\widehat{\boldsymbol{\theta}}_n^{(k)}))$ for $k = \ell-1, \ell$ and $\boldsymbol{\Phi}(\omega) = \frac{\partial \phi^{-1}(\omega)}{\partial \omega^{\mathrm{T}}}$, a similar argument can be used to show that (15) implies (10).

Proof of Theorem 4.1

Simple modifications can be made to the arguments given in Theorem 3.2 to show that

$$\sqrt{n}\boldsymbol{S}_n^{-1/2}(\widehat{\boldsymbol{\theta}}_n^{*} - \widehat{\boldsymbol{\theta}}_n) \Rightarrow \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})\ a.s.$$

as $n \to \infty$, where

$$\boldsymbol{S}_n = \boldsymbol{\Phi}^{(k)}(\boldsymbol{g}^{(k)}(\widehat{p}_n))\boldsymbol{G}^{(k)}(\widehat{\boldsymbol{p}}_n)\boldsymbol{U}^{-1}\boldsymbol{W}^{(k)}(\widehat{\boldsymbol{\theta}}_n)(\boldsymbol{G}^{(k)}(\widehat{\boldsymbol{p}}_n))^{\mathrm{T}}(\boldsymbol{\Phi}^{(k)}(\boldsymbol{g}^{(k)}(\widehat{\boldsymbol{p}}_n)))^{\mathrm{T}}.$$

As $n \to \infty$, we have $\hat{\boldsymbol{p}}_n \to \boldsymbol{p}_0$ *a.s.* and $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ *a.s.* so that $\boldsymbol{S}_n \to \boldsymbol{S}_0$ *a.s.* where

$$\boldsymbol{S}_0 = \boldsymbol{\Phi}^{(k)}\boldsymbol{G}^{(k)}\boldsymbol{U}^{-1}\boldsymbol{W}^{(k)}(\boldsymbol{G}^{(k)})^{\mathrm{T}}(\boldsymbol{\Phi}^{(k)})^{\mathrm{T}}.$$

Thus, by Slutsky's Theorem, we have

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n^{*} - \widehat{\boldsymbol{\theta}}_n) = \boldsymbol{S}_n^{1/2}\left(\sqrt{n}\boldsymbol{S}_n^{-1/2}(\widehat{\boldsymbol{\theta}}_n^{*} - \widehat{\boldsymbol{\theta}}_n)\right) \Rightarrow \boldsymbol{S}_n^{1/2}\mathcal{N}(\boldsymbol{0}, \boldsymbol{I}) \quad a.s.$$

as $n \to \infty$. Thus for any non zero vector $\boldsymbol{a} \in R^{q+4}$ with probability one $\sqrt{n}\boldsymbol{a}^{\mathrm{T}}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ and $\sqrt{n}\boldsymbol{a}^{\mathrm{T}}(\widehat{\boldsymbol{\theta}}_n^{*} - \widehat{\boldsymbol{\theta}}_n)$ have the same weak limit. The triangle inequality and the fact that in this case the weak convergence is uniform, yield (11).

Acknowledgments

The research was partially sponsored by the National Cancer Institute under grant R15 CA106248-02. The authors thank an anonymous reviewer and the associate editor for their helpful comments which helped to improve the paper.

References

1. Andrews DWK. Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. Econometrica 2000;68(2):399–405.
2. Bai J. Estimation of a change point in multiple regression models. The Review of Economics and Statistics 1997;79(4):551–563.
3. Bilodeau, M.; Brenner, D. Theory of Multivariate Statistics. New York: Springer-Verlag; 1999.
4. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. Statistics in Medicine 2000;19:1141–1164. [PubMed: 10797513]
5. Centers for Disease Control and Prevention, National Center for Health Statistics. Compressed Mortality File 1979–1998. CDC WONDER On-line Database, compiled from Compressed Mortality File CMF 1968–1988, Series 20, No. 2A, 2000 and CMF 1989–1998, Series 20, No. 2E. 2003. Accessed at http://wonder.cdc.gov/cmf-icd9.html on March 23, 2008
6. Czajkowski M, Gill R, Rempala G. ljr: Logistic Joinpoint Regression. 2007 R package version 1.0-1.
7. Czajkowski M, Gill R, Rempala G. Model selection in logistic joinpoint regression with applications to analyzing cohort mortality patterns. Statistics in Medicine 2008;27:1508–1526. [PubMed: 17676590]
8. Chen, J.; Gupta, AK. Parametric Statistical Change Point Analysis. Boston: Birkhauser; 2000.
9. Davison, AC.; Hinkley, DV. Bootstrap Methods and their Applications. Cambridge: Cambridge University Press; 1997.
10. de Boor, C. A Practical Guide to Splines. Berlin: Springer; 1978.
11. Dierckx, P. Curve and Surface Fitting with Splines. Oxford: Clarendon; 1993.
12. Efron, B.; Tibshirani, R. An Introduction to the Bootstrap. New York: Chapman and Hall; 1993.
13. Eilers PHC, Marx BD. Flexible smoothing with *B*-splines and penalties. Statistical Science 1996;11 (2):89–102.
14. Eubank, RI. Spline Smoothing and Nonparametric Regression. New York: Marcel Dekker; 1988.
15. Feder PI. On asymptotic distribution theory in segmented regression problems – identified case. Annals of Statistics 1975;3:49–83.
16. Gallant AR, Fuller WA. Fitting segmented polynomial regression models whose joinpoints have to be estimated. Journal of the American Statistical Association 1973;68:144–147.
17. Genz A, Bretz F, Hothorn T. mvtnorm: Multivariate Normal and T Distribution. 2008 R package version 0.8–3.
18. Gössl C, Küchenhoff H. Bayesian analysis of logistic regression with an unknown change point and covariate measurement error. Statistics in Medicine 2001;20:3109–3121. [PubMed: 11590636]
19. Grady, DUS. Cancer Death Rates Are Found to Be Falling. New York Times. 2007 Oct 15. Accessed at http://www.nytimes.com/2007/10/15/us/15cancer.html on March 23, 2008
20. Green, PJ.; Silverman, BW. Nonparametric Regression and Generalized Linear Models. London: Chapman & Hall; 1994.
21. Gu, C. Smoothing Spline ANOVA Models. New York: Springer; 2002.
22. Hall, P. The Bootstrap and Edgeworth Expansion. New York: Spring-Verlag; 1992.
23. Hastie, T.; Tibshirani, R.; Friedman, J. Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York: Springer; 2001.
24. Hušková M, Picek J. Bootstrap in detection of changes in linear regression. Sankhyā 2005;67:200–226.
25. Jandhyala VK, MacNeill IB. Tests for parameter changes at unknown times in linear regression models. Journal of Statistical Planning and Inference 1991;27:291–316.

26. Jandhyala VK, MacNeill IB. Iterated partial sum sequences of regression residuals and tests for changepoints with continuity constraints. Journal of the Royal Statistical Society B 1997;59:147–156.
27. Jarušková D. Testing appearance of linear trend. Journal of Statistical Planning and Inference 1998;70:263–276.
28. Kim H-J, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. Statistics in Medicine 2000;19:335–351. [PubMed: 10649300]
29. Kim H-J, Fay MP, Yu B, Barrett MJ, Feuer EJ. Comparability of segmented regression models. Biometrics 2004;60:1005–1014. [PubMed: 15606421]
30. Kim H-J, Yu B, Feuer EJ. Selecting the number of change-points in segmented linear regression. Statistica Sinica. preprint.
31. Kim J, Kim H-J. Asymptotic results in segmented multiple regression. Journal of Multivariate Analysis. preprint.
32. Liu J, Wu S, Zidek JV. On segmented multivariate regression. Statistica Sinica 1997;7:497–525.
33. Pawitan, Y. In All Likelihood: Statistical Modelling and Inference Using Likelihood. Oxford: Oxford University Press; 2001.
34. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2008. ISBN 3-900051-07-0, URL http://www.R-project.org
35. Rempala GA, Szatzschneider K. Bootstrapping Parametric Models of Mortality. Scandinavian Actuarial Journal 2004;1:53–78.
36. Ruppert D. Selecting the number of knots for penalized splines. Journal of Computational Graphics and Statistics 11(4):735–757.
37. Seber, GAF.; Wild, CJ. Nonlinear Regression. New York: Wiley; 1989.
38. Serfling, RJ. Approximation Theorems of Mathematical Statistics. New York: Wiley; 1980.
39. Shao, J. Mathematical Statistics. New York: Springer; 1999.
40. Tiwari RC, Cronin KA, Davis W, Feuer EJ. Bayesian model selection for join point regression with application to age-adjusted cancer rates. Applied Statistics 2005;54:919–939.
41. Ulm K. A statistical method for assessing a threshold in epidemiological studies. Statistics in Medicine 1991;20:341–349. [PubMed: 2028118]
42. Wahba, G. Spline Models for Observational Data. Philadelphia: Society for Industrial and Applied Mathematics; 1990.
43. Yao Y-C. Estimating the number of change-points via Schwarz' criterion. Statistics and Probability Letters 1988;6:181–189.

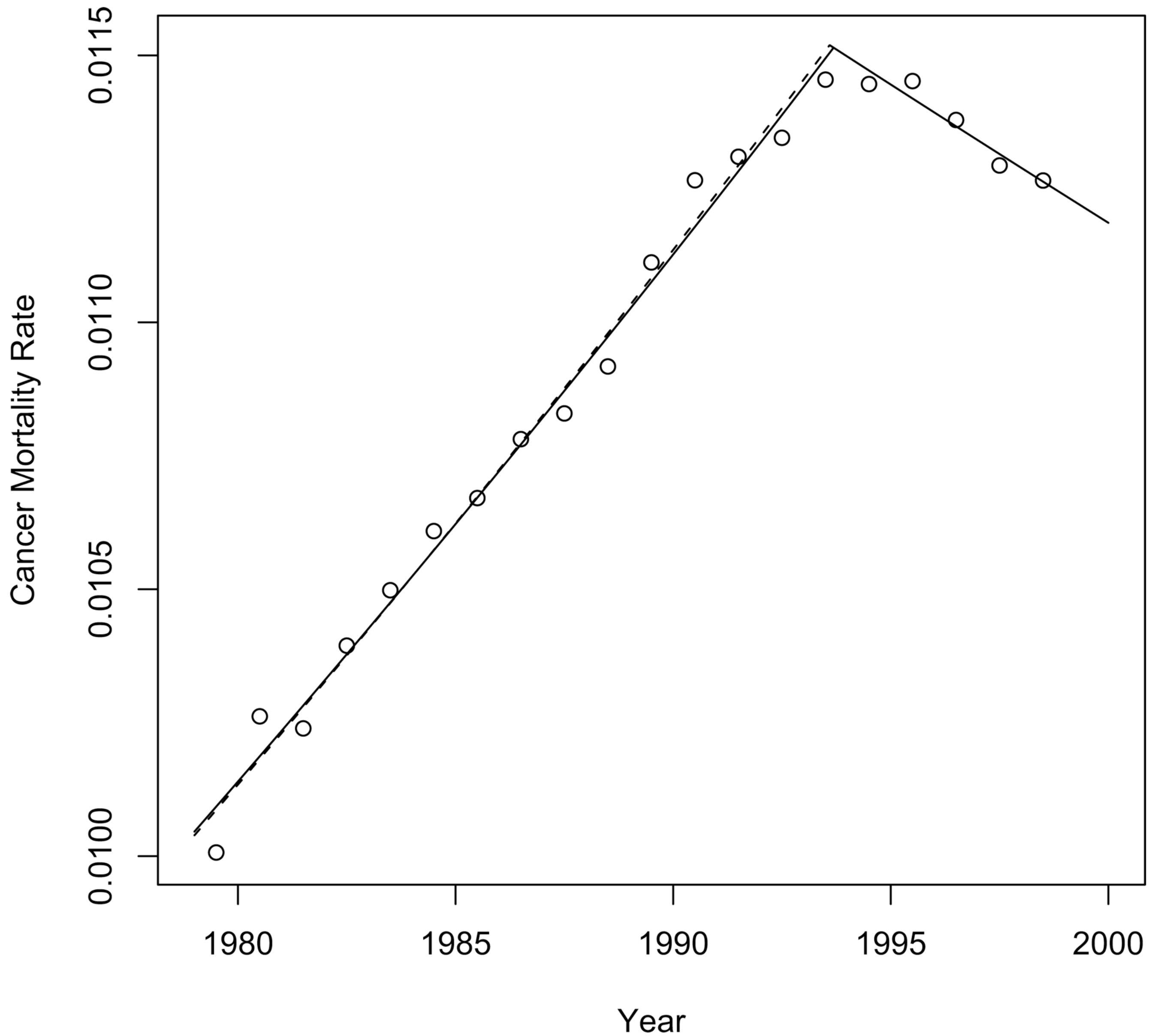

Figure 1.
Observed US yearly cancer mortality rates for individuals age 65 and over. The solid line gives the fitted model based on all of the data. The dashed line gives the fitted model with the 15th observation (where t_{15} = 1993.5) removed.

Table 1

Simulation results for the bootstrap method applied to the MLE.

m_i, $i = 1,\ldots, 7$	$\hat{P}(\hat{\tau}_n = \tau_0)$	$\widehat{MSE}(\hat{\tau}_n)$	$\hat{P}^*(\hat{\tau}_n^* = \tau_0)$	$\widehat{MSE}^*(\hat{\tau}_n^*)$
10^3	.2007	3.967×10^{-1}	.1097	4.779×10^{-1}
10^5	.2023	2.792×10^{-3}	.1157	3.545×10^{-3}
10^7	.2020	2.899×10^{-5}	.1155	3.698×10^{-5}
10^9	.2020	3.122×10^{-7}	.1160	3.972×10^{-7}

Table 2

Comparison of results for the MLE and the estimates based on the removal algorithm.

m_i, $i = 1, \ldots, 7$	$\hat{P}*(\hat{\tau}_n^* - \hat{\tau}_n \geq \hat{\tau}_{(.95)})$	$\widehat{MSE}(\hat{\tau}_n)$	$\hat{P}*((\hat{\tau}_n^-)^* - \hat{\tau}_n^- \geq \hat{\tau}_{(.95)}^-)$	$\widehat{MSE}^-(\hat{\tau}_n^-)$
10^3	.0625	3.967×10^{-1}	.0496	7.910×10^{-1}
10^5	.0670	2.792×10^{-3}	.0500	4.738×10^{-3}
10^7	.0676	2.899×10^{-5}	.0499	4.730×10^{-5}
10^9	.0676	3.122×10^{-7}	.0499	5.061×10^{-7}

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

Parameter estimates in the logistic joinpoint regression model for the US yearly 65+ cancer mortality data.

Parameter	Estimate	Value	Estimate	Value
α_0	$\hat{\alpha}_n$	−23.17397	$\hat{\alpha}_n^-$	−23.44145
β_0	$\hat{\beta}_n$	.0093903	$\hat{\beta}_n^-$	.0095251
δ_0	$\hat{\delta}_n$	−.0140205	$\hat{\delta}_n^-$	−.0141553
τ_0	$\hat{\tau}_n$	1993.686	$\hat{\tau}_n^-$	1993.595