ARTICLE IN PRESS

Available at www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT

Journal of Statistical Planning and Inference ▮▮▮ (▮▮▮▮) ▮▮▮–▮▮▮

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Consistent estimation in generalized broken-line regression

Ryan Gill[a], Michael Baron[b,c,*,1]

[a]Department of Mathematics, University of Louisville, Louisville, KY 40292, USA
[b]Programs in Mathematical Sciences, University of Texas at Dallas, Richardson, TX 75083, USA
[c]Mathematical Sciences, IBM, T.J. Watson Research Center, Yorktown Heights, NY 10598, USA

Received 28 March 2002; accepted 5 September 2003

Abstract

A change-point model is considered where the canonical parameter of an exponential family drifts from its control value at an unknown time and changes according to a broken-line regression. Necessary and sufficient conditions are obtained for the existence of consistent change-point estimators. When sufficient conditions are met, it is shown that the maximum likelihood estimator of the change point is consistent, unlike the classical abrupt change-point models. Results are extended to the case of nonlinear trends and nonequidistant observations.

Keywords: Canonical parameter; Change-point problem; Consistency; Generalized regression; Maximum likelihood; Natural exponential family

1. Introduction

In the classical parametric change-point model, the data set $\underline{\mathbf{X}}$ consists of two parts, $(\mathbf{X}_1,\ldots,\mathbf{X}_\nu)$ being a sample from distribution $P(\cdot|\theta_0)$ and $(\mathbf{X}_{\nu+1},\ldots,\mathbf{X}_n)$ from distribution $P(\cdot|\theta_1)$. The change point ν is the parameter of interest, and θ_j, $j=0,1$ are known or unknown nuisance parameters. This defines an abrupt change model, which has been well studied during the recent decades. Classical references are e.g. Basseville and Nikiforov (1993), Carlstein (1988), Chernoff and Zacks (1964), Cobb (1978), Hinkley (1970), and Page (1954); also see Bhattacharya (1995), Lai (1995) and Zacks (1983) for surveys of retrospective and sequential change-point estimation methods.

* Corresponding author. Programs in Mathematical Sciences, University of Texas at Dallas, Richardson, TX 75083, USA.

E-mail addresses: rsgill01@erdos.math.louisville.edu (R. Gill), mbaron@utdallas.edu (M. Baron).

[1] Research is supported by the National Science Foundation, grant #9818959.

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2003.09.015

However, abrupt change models are found impractical in a number of applications. In many problems related to quality and process control, global climate, economics, pollution, etc., the distribution parameter goes out of control and changes *gradually* instead of jumping instantly to a new value and stabilizing there.

In this paper, we consider a *gradual change-point model*, where an observation $\mathbf{X}_j \in \mathbb{R}^d$, $j = 1, \ldots, n$, collected at time $t(j)$, has distribution $P(\cdot|\boldsymbol{\theta}_j)$ with

$$\boldsymbol{\theta}_j = \boldsymbol{\theta}_{t(j),v} = \boldsymbol{\theta}_0 + \boldsymbol{\beta}(t(j) - v)^+. \tag{1}$$

Here $\boldsymbol{\theta}_0$ is a *control value* of the parameter, and the system is "in control" until time v. After that, the system drifts away from the control value at the rate of $\boldsymbol{\beta} \in \mathbb{R}^d$ units per one unit of time. For a multiparameter family with $d > 1$, the slope $\boldsymbol{\beta}$ not only determines the speed, but also the direction of change. The sequence of sampling times $\{t(j)\}$ is increasing to $t(\infty) = \lim_{j \to \infty} t(j) \leqslant +\infty$. Placing no additional assumptions on $t(j)$ allows to generalize (1) to nonlinear models (Remark 1 below).

Similar models are studied in Gupta and Ramanayake (2001), for the case of Exponential distributions, and in Hušková (1999), Jarušková (1998), and Siegmund and Zhang (1995) and a few other papers, for the case of a location parameter family. A commonly considered model involves "a trend" and "an error", so that only the mean of observed values changes according to the broken-line regression. However, in reality, almost any sliding of a system from its control state results in a change (typically, increase) of variance, probabilities of large deviations, and other characteristics, in addition to a changing mean. Therefore, we consider the general case of a d-variate natural exponential family

$$\mathrm{d}P_{\boldsymbol{\theta}}(\mathbf{x}) = \mathrm{e}^{\langle \boldsymbol{\theta}, \mathbf{x} \rangle - \psi(\boldsymbol{\theta})} \, \mathrm{d}P_{\mathbf{0}}(\mathbf{x}) \tag{2}$$

with the canonical parameter $\boldsymbol{\theta} \in \mathrm{int}(\Theta)$, where

$$\Theta = \left\{ \boldsymbol{\theta} \in \mathbb{R}^d \,\middle|\, \int_{\mathbb{R}^d} \mathrm{e}^{\langle \boldsymbol{\theta}, \mathbf{x} \rangle} \, \mathrm{d}P_{\mathbf{0}}(x) < \infty \right\}$$

is a convex parameter set, containing the control value $\boldsymbol{\theta}_0$, and $\psi(\boldsymbol{\theta}) = \log \int_{\mathbb{R}^d} \mathrm{e}^{\langle \boldsymbol{\theta}, \mathbf{x} \rangle} \, \mathrm{d}P_{\mathbf{0}}(\mathbf{x})$ is the log Laplace transform of P_0. The function $\psi(\boldsymbol{\theta})$ is analytic and convex on Θ, with $\nabla\psi(\boldsymbol{\theta}) = \mathbf{E}_{\boldsymbol{\theta}}\mathbf{X}$ and $\nabla^2\psi(\boldsymbol{\theta}) = \mathrm{Var}_{\boldsymbol{\theta}}\,\mathbf{X}$ for all $\boldsymbol{\theta} \in \mathrm{int}(\Theta)$ (Brown, 1986).

The model described by (1) and (2) extends the class of *generalized linear models* (McCullagh and Nelder, 1983; Myers et al., 2002) to the case of a broken-line regression. From this point of view, the role of a *link function* is played by $(\nabla\psi)^{-1}(\boldsymbol{\mu})$, where $\boldsymbol{\mu} = \mathbf{E}_{\boldsymbol{\theta}}\mathbf{X}$. Model (1) allows a generalized linear trend to change at an unknown moment. In a more general setting, $t(j)$ represent covariates, not necessarily corresponding to sampling times. In this setting, v is the value of a covariate that separates two different generalized linear trends of the response variable.

Remark 1 (Nonlinear models). Notice that the model defined by (1) is not restricted to linear trends of the parameter. Any *nonlinear trend* can fit into Eq. (1) by an appropriate transformation of the time scale and consequent adjustment of sampling times $t(j)$. For example, consider a system that is suspected to go out of control at time υ and to change at a logarithmic speed, whereas observations are sampled at equidistant intervals, say, every hour. Then the distribution of collected data coincides with the distribution of $\mathbf{X}_1, \mathbf{X}_2, \ldots$, where $\mathbf{X}_j$, observed at the time $t(j) = \log(j)$, has the distribution $P_{\boldsymbol{\theta}_j}$, with $\boldsymbol{\theta}_j$ going out of control at the time $v = \log(\upsilon)$, in accordance with the basic linear model (1).

Remark 2 (Possible location of a change point). Classical abrupt change-point models assume that the change occurs at one of sampling times. That is, $v = t(j)$ for some j. This is caused by the fact that any value of v between $t(j)$ and $t(j+1)$ generates the same joint distribution of $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$. Hence, allowing v to be a continuous parameter yields an *unidentifiable* (ill-posed) model.

Unlike the classical models, any real value of v in $(-\infty, t(n)]$ results in a different distribution of data. Thus, v is not necessarily the time epoch when an observation has been collected. It is possible that two consecutive observations are collected at times $v - s$ and $v + t$, but no data point is sampled at the time v.

The case $v = t(n)$ is special; it means that all the observed data follow the same "in-control" distribution, i.e., a change has not occurred before the end of our sampling period. In this case, it is impossible to *predict* when it will occur (except for a Bayesian model with a prior distribution of v). Any value of $v \geqslant t(n)$ yields the same joint distribution of data, therefore, we let $v \in (-\infty, t(n)]$ for identifiability purposes.

Remark 3 (Sampling times). Finally, we no longer assume that sampling occurs at equidistant times. The sequence of sampling times $\{t(j)\}$ is almost arbitrary, but it is assumed to be known which is crucial for estimating the change point.

In this paper, we investigate the possibility of estimating the change-point parameter v consistently. In the abrupt change-point model, there does not exist a consistent change-point estimator (Hinkley, 1970). Indeed, the parameter θ stabilizes at a new value θ_1, and increasing a sample size by obtaining remote observations from the distribution $P(\cdot|\theta_1)$ contributes little information about v. Conversely, under model (1), the parameter drifts further away from its control value, making occurrence of a change more and more apparent. As a result, consistent estimators of the change point *exist* for certain exponential families.

The maximum likelihood estimation procedure under model (1) is derived in Section 2. Section 3 establishes a necessary condition for the existence of a consistent estimator. Sufficient conditions for the existence of consistent estimators are obtained in Section 4. The proof is constructive; it is shown that under these conditions, the maximum likelihood estimator is consistent.

2. Maximum likelihood estimation

For the regression-type gradual single change-point model where $\boldsymbol{\theta}_0$ and $\boldsymbol{\beta}$ are known, consider a data set $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ where $\mathbf{x}_j$ is the observed value of $\mathbf{X}_j$ for $j = 1, \ldots, n$.

Let $\{f(\cdot|\boldsymbol{\theta})\,|\,\boldsymbol{\theta}\in\Theta\}$ be a family of densities, under a reference measure μ, associated with $\{P_{\boldsymbol{\theta}}\,|\,\boldsymbol{\theta}\in\Theta\}$. Let $z\leqslant t(n)$ and let n_z be the number of observations taken at or before time z. Then the likelihood function of $\mathbf{x}$ given z is

$$L(\mathbf{x}|z,\boldsymbol{\theta}_0,\boldsymbol{\beta}) = f(\mathbf{x}_1|\boldsymbol{\theta}_0)\cdots f(\mathbf{x}_{n_z}|\boldsymbol{\theta}_0)f(\mathbf{x}_{n_z+1}|\boldsymbol{\theta}_{t(n_z+1),z})\cdots f(\mathbf{x}_n|\boldsymbol{\theta}_{t(n),z})$$
$$= \prod_{j=1}^{n} f(\mathbf{x}_j|\boldsymbol{\theta}_0) \prod_{j=n_z+1}^{n} \frac{f(\mathbf{x}_j|\boldsymbol{\theta}_{t(j),z})}{f(\mathbf{x}_j|\boldsymbol{\theta}_0)}, \tag{3}$$

where

$$\boldsymbol{\theta}_{t,z} = \boldsymbol{\theta}_0 + \boldsymbol{\beta}(t-z)^+, \tag{4}$$

random vectors $\mathbf{X}_j$, $j=1,\ldots,n$, are independent, and $\prod_k^n \equiv 1$ if $k>n$.

This model may not be well defined for some values of z in $(-\infty,t(n)]$. Let $\mathscr{V}_n$ be the set of z for which (2) along with (4) are well defined for all $j=1,\ldots,n$. From the convexity of Θ, $\mathscr{V}_n$ is connected and $\mathscr{V}_n=(t(n)-\tau,\infty)$, where

$$\tau = \sup\{t\,|\,\boldsymbol{\theta}_0+\boldsymbol{\beta}t\in\mathrm{int}(\Theta)\}.$$

Indeed, since Θ is convex and $\boldsymbol{\theta}_0\in\mathrm{int}(\Theta)$, it is necessary and sufficient for

$$\{\boldsymbol{\theta}_0+\boldsymbol{\beta}(t(j)-z)^+\,|\,j=1,\ldots,n\}\subset\mathrm{int}(\Theta)$$

that $\boldsymbol{\theta}_0+\boldsymbol{\beta}(t(n)-z)^+\in\mathrm{int}(\Theta)$. The latter implies that $(t(n)-z)^+<\tau$, i.e., $z>t(n)-\tau$. All such z constitute the set $\mathscr{V}_n$.

Then, from (3), the maximum likelihood estimate of v is

$$\hat{v}_{\mathrm{MLE}} = \hat{v}_{\mathrm{MLE}}(n) = \operatorname*{arg\,max}_{z\in\mathscr{V}_n,\, z\leqslant t(n)} L(\underline{\mathbf{X}}|z,\boldsymbol{\theta}_0,\boldsymbol{\beta}) = \operatorname*{arg\,max}_{z\in\mathscr{V}_n,\, z\leqslant t(n)} g(z)$$

(see Section 1, Remark 2), where

$$g(z) = \log L(\underline{\mathbf{X}}|z,\boldsymbol{\theta}_0,\boldsymbol{\beta}) - \log L(\underline{\mathbf{X}}|z,\boldsymbol{\theta}_0,\mathbf{0})$$
$$= \sum_{j=n_z+1}^{n} \{(t(j)-z)\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - \varphi(t(j)-z)+\varphi(0)\} \tag{5}$$

and $\varphi(t)=\psi(\boldsymbol{\theta}_0+\boldsymbol{\beta}t)$. Thus, the maximum likelihood estimate can be computed from a cumulative sums process based on independent log-likelihood ratios. In general, it may not be unique, therefore we consider a set of all maximum likelihood estimates of v

$$\mathscr{M}_n = \left\{ z\leqslant t(n)\,\middle|\, g(z)=\max_{w\in\mathscr{V}_n} g(w)\right\}.$$

Lemma 1. *The function* $g(z)$ *is*

(a) *analytic on* $\mathscr{V}_n\setminus\{t(1),\ldots,t(n)\}$;
(b) *continuous on* $\mathscr{V}_n$;
(c) *strictly concave down on* $(\inf\mathscr{V}_n, t(1)]$ *and* $[t(j),t(j+1)]$ *for any* $j=1,\ldots,n-1$.

Proof. Since φ is analytic on $(t(i), t(i+1))$ and obviously any linear function is analytic, the sum of $(n-i)$ functions is analytic and (a) follows.

By (a), it suffices to show that $g(z)$ is continuous at all the sampling times $t(i)$. For $i = 1, \ldots, n$,

$$\begin{aligned}\lim_{z \uparrow t(i)} g(z) &= \sum_{j=(i-1)+1}^{n} \lim_{z \uparrow t(i)} \{(t(j)-z)\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - [\varphi(t(j)-z) - \varphi(0)]\} \\ &= \sum_{j=i}^{n} \{(t(j)-t(i))\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - [\varphi(t(j)-t(i)) - \varphi(0)]\} \\ &= 0 + \sum_{j=i+1}^{n} \{(t(j)-t(i))\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j - [\varphi(t(j)-t(i)) - \varphi(0)]\} \\ &= g(t(i)).\end{aligned}$$

Similarly, $g(z) \to g(t(i))$ as $z \downarrow t(i)$, for $i = 0, 1, \ldots, n-1$. This proves continuity of $g(z)$.

Finally, we have

$$g''(z) = \sum_{j=i+1}^{n} \{-\varphi''(t(j)-z)\} < 0$$

which proves (c). □

The last part of the proof was based on Lemma 2 below which follows from Theorem 1.13(iv) of Brown (1986).

Lemma 2. *For any minimal d-dimensional standard exponential family* (2), $\nabla^2\psi(\boldsymbol{\theta})$ *is strictly positive definite for all* $\boldsymbol{\theta} \in \Theta$.

3. Consistent estimation: necessary condition

It will be seen that the possibility of estimating the change-point parameter v consistently depends on the exponential family (2) and the sequence of sampling times $\{t(j)\}$. Unboundedness of $\nabla\psi(\theta) = \mathbf{E}_{\boldsymbol{\theta}}\mathbf{X}$ is the key necessary condition for the existence of consistent estimators.

The study of consistency refers to a *sequence* $\{\mathbf{X}_j, j \geqslant 1\}$, where $\mathbf{X}_j$ are distributed according to (1) and (2). Assumptions of Theorems 1–3 guarantee existence of such a sequence. For an unbounded sequence of sampling times $\{t(j)\}$, Theorems 1 and 2 imply that $\tau = +\infty$, thus the parameter set for the change point v is $(-\infty, +\infty)$.

The proof is based on the analysis of likelihood ratios

$$\rho_{y,z} = L(\underline{\mathbf{X}} | y, \boldsymbol{\theta}_0, \boldsymbol{\beta}) / L(\underline{\mathbf{X}} | z, \boldsymbol{\theta}_0, \boldsymbol{\beta}) = \exp\{g(y) - g(z)\}. \tag{6}$$

Theorem 1. *Under model* (1), (2) *with* $\boldsymbol{\theta}_0 + \boldsymbol{\beta}\lambda \in \operatorname{int}(\Theta)$ *for all* $\lambda \geqslant 0$, *there does not exist a consistent estimator of a change point* v, *if*

(a) *the function* $\varphi'(\lambda) = \boldsymbol{\beta}^{\mathrm{T}} \nabla \psi(\boldsymbol{\theta}_0 + \boldsymbol{\beta}\lambda)$ *is bounded from above*;
(b) $\liminf_{j\to\infty} \{t(j+1) - t(j)\} > 0$ *for a sequence of sampling times* $t(j)$.

Proof. The proof is based on the following contradiction. We will show that existence of a consistent estimator yields $\mathbf{E}_{v=u}(\rho_{u,v}) \to \infty$ as $n \to \infty$ for $u < v$. On the other hand, boundedness of $\mathbf{E}_{v=u}(\rho_{u,v})$ follows from conditions (a) and (b).

Assume existence of a consistent estimator $T_n(\mathbf{X}_1, \dots, \mathbf{X}_n)$. Then, for any u,

$$P_{v=u}(|T_n - u| < \varepsilon) \to 1$$

as $n \to \infty$. So pick u and v with $u < v$ and consider the likelihood ratio

$$\rho = \rho_{u,v}(\mathbf{X}) = \frac{\mathrm{d}P_{v=u}}{\mathrm{d}P_{v=v}}(\mathbf{X})$$

which satisfies

$$\mathbf{E}_v h(\mathbf{X})\rho(\mathbf{X}) = \mathbf{E}_u h(\mathbf{X}) \tag{7}$$

for any μ-measurable function $h = h(\mathbf{x})$. Using (7) with the indicator function $h_1(\mathbf{x}) = I(\{\mathbf{x} : |T_n - u| < \varepsilon\})$, we have

$$\mathbf{E}_v(h_1\rho) = \mathbf{E}_u(h_1) = P_u(|T_n - u| < \varepsilon) \to 1 \tag{8}$$

as $n \to \infty$. Also,

$$\mathbf{E}_v(h_1) = P_v(|T_n - u| < \varepsilon) \to 0 \tag{9}$$

and

$$\operatorname{Var}_v h_1 = (\mathbf{E}_v h_1)(1 - \mathbf{E}_v h_1) \to 0. \tag{10}$$

Since $\mathbf{E}_v\rho = 1$, (8) and (9) imply that

$$\operatorname{Cov}_v(h_1, \rho) = \mathbf{E}_v h_1 \rho - \mathbf{E}_v h_1 \mathbf{E}_v \rho \to 1. \tag{11}$$

Then, from (10) and (11),

$$\operatorname{Var}_v \rho = \frac{\operatorname{Cov}_v^2(h_1, \rho)}{\operatorname{Corr}_v^2(h_1, \rho) \operatorname{Var}_v h_1} \to \infty, \tag{12}$$

because the correlation coefficient is bounded between -1 and 1. Now using (7) with $h = \rho$ along with (12), one has

$$\mathbf{E}_u\rho = \mathbf{E}_v\rho^2 = \operatorname{Var}_v \rho + (\mathbf{E}_v\rho)^2 \to \infty. \tag{13}$$

Now, we will find a contradiction with (13) by obtaining a uniform upper bound for $\mathbf{E}_u\rho$. Condition (b) guarantees existence of positive N and δ such that $t(j+1) - t(j) > \delta$ for all $j \geqslant N$. Then, let $r = \lceil 2((v-u)/\delta) \rceil$. Also, let

$$M_j = \sup\{\varphi'(w) \mid t(j) - u \leqslant w \leqslant t(j) + v - 2u\}$$

and

$$m_j = \inf\{\varphi'(w) \mid t(j) - v \leqslant w \leqslant t(j) - u\}.$$

Since φ is convex, φ' is nondecreasing, and $M_j \leqslant m_k$ as long as $t(j) - 2u \leqslant t(k) - 2v$. Therefore, for any k, $M_{N+kr+j} \leqslant m_{N+(k+1)r+j}$, because

$$t(N + (k + 1)r + j) - 2v \geqslant t(N + kr + j) + r\delta - 2v \geqslant t(N + kr + j) - 2u.$$

From (5) and (6), we have

$$\begin{aligned}
\log \mathbf{E}_u \rho &= A + \sum_{j=s}^{n} \{\log \mathbf{E}_u \mathrm{e}^{(v-u)\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j} - [\varphi(t(j) - u) - \varphi(t(j) - v)]\} \\
&= A + \sum_{j=s}^{n} \{[\varphi(t(j) + v - 2u) - \varphi(t(j) - u)] \\
&\quad + [\varphi(t(j) - u) - \varphi(t(j) - v)]\} \\
&\leqslant A + (v - u) \sum_{j=s}^{n} (M_j - m_j) \\
&\leqslant A + (v - u)N \left(\sup_w \varphi'(w) - m_s \right) \\
&\quad + (v - u) \sum_{j=0}^{r-1} \left\{ \sum_{k=0}^{\infty} (M_{N+kr+j} - m_{N+(k+1)r+j}) - m_{N+j} \right\} \\
&\leqslant A + (v - u) \left(N \sup_w \varphi'(w) - N m_s - r m_N \right) < \infty,
\end{aligned}$$

where $s = n_v + 1$ and

$$\begin{aligned}
A &= \sum_{j=n_u+1}^{n_v} \{\log \mathbf{E}_u \mathrm{e}^{(t(j)-u)\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j} - [\varphi(t(j) - u) - \varphi(0)]\} \\
&= \sum_{j=n_u+1}^{n_v} \{\varphi(2(t(j) - u)) - 2\varphi(t(j) - u) + \varphi(0)\}.
\end{aligned}$$

This contradicts (13) and shows that there is no consistent estimator of v. □

Condition (a) of Theorem 1 implies boundedness of $\varphi'(\lambda)$ along at least one direction in $\mathbb{R}^d$ that is not orthogonal to $\boldsymbol{\beta}$. Two most typical situations are considered in the following corollaries.

Corollary 1 (Equally spaced observations). *If condition* (a) *of Theorem* 1 *holds and* $t(j) = a + bj$ *for some* $a \geqslant 0$ *and* $b > 0$, *then there is no consistent estimator of* v.

Corollary 2 (Univariate case). *Let $d=1$ and let $t(j)$ satisfy condition* (b) *of Theorem* 1. *If β is positive* (*negative*), *and $\psi'(\theta)$ is bounded from above* (*below*), *then there is no consistent estimator of v.*

To illustrate the result of Theorem 1, we consider the family of exponential distributions that satisfies condition (a). This example is also applicable to Beta$(\alpha,1)$, single-parameter Pareto(σ), Laplace(λ), and Weibull(λ,τ) (known τ) families of distributions because they have the same canonical representation (2).

Example 1 (Inconsistency of $\hat{v}$ for exponential distributions). The family of Exponential (λ) distributions can be written in the form (2) with $\psi(\theta)=-\log(1-\theta)$, $\theta=1-\lambda\in(-\infty,1)$. Let $\theta_0=0$, $\beta=-1$, and $t(j)=j$, so that observations $X_1,\ldots,X_n$ are collected at equally spaced times. Consider an estimator $\tilde{v}=\arg\max_{z=1,\ldots,n} L(\underline{\mathbf{X}}|z,0,1)$.

In order to show that $\tilde{v}$ is not consistent, it is sufficient to show that $P(\log\rho_{v-1,v}<0)$ does not converge to 1 as $n\to\infty$. From (5),

$$\log\rho_{v-1,v}=\sum_{j=v}^{n}\left(-X_j+\log\frac{j-v+2}{j-v+1}\right)=\log\prod_{j=0}^{n-v}Y_j+\log(n-v+2), \tag{14}$$

where $Y_j=\exp\{-X_{v+j}\}$ has Beta$(1+j,1)$ distribution.

For any distinct positive $a_1,\ldots,a_N$, the product of independent Beta $(a_j,1)$ random variables has the cumulative distribution function

$$F(t)=\prod_{j=1}^{N}a_j\sum_{k=1}^{N}\frac{(-1)^{N-1}}{a_k}\prod_{\substack{i=1\\ i\neq k}}^{N}\left(\frac{1}{a_k-a_i}\right)t^{a_k}$$

for $0<t<1$ (see Gill, 2002 for the proof). With $a_j=1+j$, $N=n-v+1$, and $t=(n-v+2)^{-1}$, from (14) we have

$$\begin{aligned}P(\log\rho_{v-1,v}<0)&=F\left(\frac{1}{n-v+2}\right)=F\left(\frac{1}{N+1}\right)\\&=N!\sum_{k=1}^{N}\frac{(-1)^{N-1}}{k}\prod_{\substack{i=1\\ i\neq k}}^{N}\left(\frac{1}{k-i}\right)\left(\frac{1}{1+N}\right)^{k}\\&=1-\sum_{k=0}^{N}\frac{N!}{k!(N-k)!}\left(-\frac{1}{1+N}\right)^{k}\\&=1-\left(1-\frac{1}{1+N}\right)^{n}.\end{aligned}$$

The last probability converges to $1-\mathrm{e}^{-1} \approx 0.6321$ as $N \to \infty$. This result is supported by a Monte Carlo study in Gill (2002). Thus, $\hat{v}_{\mathrm{MLE}}$ is not consistent since $P(\log \rho_{v-1,v} < 0)$ does not converge to 1.

4. Consistent estimation: sufficient conditions

This section gives sufficient conditions on the family $P_{\boldsymbol{\theta}}$ and sampling times $t(j)$, under which, unlike the classical abrupt change model, the maximum likelihood estimator $\hat{v}_{\mathrm{MLE}}(n)$ of the change point v is consistent. A direct proof of consistency requires it to be shown that

$$P(|\hat{v}_{\mathrm{MLE}}(n) - v| > \varepsilon) \to 0 \tag{15}$$

as $n \to \infty$. If $\hat{v}_{\mathrm{MLE}}(n)$ is not defined uniquely, then (15) must hold for *any* sequence of maximum likelihood estimates $z_n \in \mathscr{M}_n$.

It will be convenient to express (15) in terms of the following sets:

$$U_{z,k,\gamma} = \{\mathbf{x} \,|\, g(k) - g(z) \leqslant \gamma\}, \tag{16}$$

$$U_A = \bigcup_{z \in A} U_{z,\hat{v}_{\mathrm{MLE}},0} = \{\mathbf{x} \,|\, A \cap \mathscr{M}_n \neq \emptyset\},$$

$$D^-_{z,\gamma} = \left\{\mathbf{x} \,\middle|\, \lim_{y \uparrow z} g'(y) \leqslant \gamma\right\}, \tag{17}$$

$$D^+_{z,\gamma} = \left\{\mathbf{x} \,\middle|\, \lim_{y \downarrow z} g'(y) \geqslant \gamma\right\}, \tag{18}$$

$$V_N = \{i \in \mathbb{Z}^+ \,|\, n - N \leqslant i < n \text{ and } t(i+1) - t(i) < 1\},$$

$$W_N = \{i \in \mathbb{Z}^+ \,|\, n - N \leqslant i < n \text{ and } t(i+1) - t(i) \geqslant 1\}.$$

First, we restate (15) in the new notations as

$$P(U_{(-\infty, v-\varepsilon)} \cup U_{(v+\varepsilon, t(n)]}) \to 0 \tag{19}$$

as $n \to \infty$. Then, the sets involved in (19) are connected with simpler sets D^- and D^+ by the following two lemmas.

Lemma 3. *For any* $a \in [t(i), t(i+1))$ *and* $b \in (t(i), t(i+1)]$, *one has* $U_{[t(i),b)} \subset D^-_{b,0}$ *and* $U_{(a,t(i+1)]} \subset D^+_{a,0}$.

Lemma 4. *For any positive integer* $N < n$,

$$U_{(t(n-N),t(n)]} \backslash U_{t(n-N),v,\gamma} \text{ is a subset of } \bigcup_{i=n-N}^{n-1} D^+_{t(i),\gamma_i^*},$$

where

$$\gamma_i^* = \begin{cases} \gamma/N & \textit{if } i \in V_N, \\ 0 & \textit{if } i \in W_N. \end{cases}$$

The proofs of these and further lemmata are given in the appendix.

According to Lemma 3, the maximum likelihood estimator $\hat{v}_{\text{MLE}}$ cannot belong to the interval $[t(i), b)$ if the left derivative of g at b is positive, and it cannot belong to $(a, t(i+1)]$ if the right derivative of g at a is negative. Also, by Lemma 4, if $\hat{v}_{\text{MLE}}$ is in the interval $(t(n-N), t(n)]$ and $g(v) > g(t(n-N)) + \gamma$, then the right derivative of g must be at least γ_i^* at least one sampling time $t(i)$, where $i \geqslant n-N$.

Lemmas 3 and 4 allow estimation of the probability in (15) by means of sets D^- and D^+. The latter have a rather simple structure determined by the first derivative of the log-likelihood ratio function $g(z)$.

Lemma 5 (Monotone properties of $D^-_{z,\gamma}$ and $D^+_{z,\gamma}$). (a) *For any* z, $D^-_{z,\gamma}$ *is increasing in* γ.
(b) *For any* z, $D^+_{z,\gamma}$ *is decreasing in* γ.
(c) *For any* γ *and* $i < n$, $D^-_{z,\gamma}$ *is increasing in* z *on* $(t(i), t(i+1)]$.
(d) *For any* γ *and* $i < n$, $D^+_{z,\gamma}$ *is decreasing in* z *on* $[t(i), t(i+1))$.

In order to estimate the probability in (15), we will derive upper bounds for $P(U_{z,v,\gamma})$, $P(D^-_{z,\gamma})$, and $P(D^+_{z,\gamma})$. Here we use the *Chernoff inequality*

$$P(Y \geqslant y) \leqslant \exp(-uy) M_Y(u) \tag{20}$$

that holds for all $u > 0$ where $M_Y(u)$, the moment generating function of the random variable Y, exists (e.g. Ross, 1994).

Chernoff bounds for the mentioned probabilities will be expressed in terms of the following integrals,

$$I(a, b; u) = \int_{(1-u)a+ub}^{a} \varphi''(w)(a-w)\,\mathrm{d}w,$$

$$J(a, b; u) = (1-u) I(a, b; u) + u I(b, a; 1-u),$$

$$K(a, b) = \int_a^b \varphi''(w)\,\mathrm{d}w = \varphi'(b) - \varphi'(a).$$

Lemma 6 (Chernoff bound for $U_{z,v,\gamma}$). *For sufficiently small* $u > 0$ *and any* $v, z \in \mathscr{V}_n$,

$$P(U_{z,v,\gamma}) \leqslant \exp\left\{ u\gamma - \sum_{j=n_m+1}^{n} J((t(j)-v)^+, (t(j)-z)^+; u) \right\},$$

where $m = \min\{v, z\}$.

Lemma 7 (Chernoff bound for $D^-_{z,\gamma}$). *For sufficiently small $u>0$ and any $v,z\in\mathscr{V}_n$,*

$$P(D^-_{z,\gamma}) \leqslant \exp\left\{ u\gamma - \sum_{j=n_z^-+1}^{n} [uK(t(j)-v+u, t(j)-z) \right.$$
$$\left. + I(t(j)-v, t(j)-v+1; u)] \right\},$$

where $n_z^- = \lim_{t\uparrow z} n_t$ is the number of observations collected before time z.

Lemma 8 (Chernoff bound for $D^+_{z,\gamma}$). *For sufficiently small $u>0$ and any $v,z\in\mathscr{V}_n$,*

$$P(D^+_{z,\gamma}) \leqslant \exp\left\{ -u\gamma - \sum_{j=n_z+1}^{n} [uK(t(j)-z, t(j)-v-u) \right.$$
$$\left. + I(t(j)-v, t(j)-v-1; u)] \right\}.$$

It remains to derive a lower bound for the integrals I, J, and K that appear in Lemmas 6–8.

Lemma 9. *If $\varphi''(w)\geqslant C$ for any $w\in[a,b]$ and some $C>0$, then*

$$I(a,b;u) \geqslant Cu^2(a-b)^2/2,$$

$$J(a,b;u) \geqslant Cu(1-u)(a-b)^2/2$$

and

$$K(a,b) \geqslant C(b-a).$$

We are now ready to prove the main result of this section.

Theorem 2. *Consider model* (1), (2) *with $\boldsymbol{\theta}_0+\boldsymbol{\beta}\lambda\in\operatorname{int}(\Theta)$ for all $\lambda\geqslant 0$ and $t(\infty)=\lim_{j\to\infty} t(j)=+\infty$. Suppose that $\varphi''(w)\geqslant C$ for any $w\geqslant 0$ and some $C>0$. Then the maximum likelihood estimator of a change point is consistent.*

Proof. Based on Lemmas 3–9, we will prove that

$$\begin{aligned} P(|\hat{v}_{\mathrm{MLE}}(n)-v|>\varepsilon) &= P(U_{(-\infty,v-\varepsilon)} \cup U_{(v+\varepsilon,t(n)]}) \\ &\leqslant P(U_{(-\infty,v-\varepsilon)}) + P(U_{(v+\varepsilon,t(n-\lfloor n^\alpha\rfloor)]}) + P(U_{(t(n-\lfloor n^\alpha\rfloor),t(n)]}) \end{aligned} \tag{21}$$

converges to 0 for any $\varepsilon > 0$ by showing that each probability in (21) converges to 0 for some $\alpha \in (0,1)$.

First, we notice that for any $i \leqslant n_{v-\varepsilon}$ and $b \in (t(i), t(i+1)]$,

$$\log P(U_{[t(i),b)}) \leqslant \log P(D^{-}_{b,0}) \tag{22}$$

$$\leqslant -\frac{\varepsilon}{2}\sum_{j=i+1}^{n} K(t(j)-v+\varepsilon/2, t(j)-b) \tag{23}$$

$$\leqslant -\frac{C\varepsilon}{2}(n-i)\left(v-b-\frac{\varepsilon}{2}\right) \tag{24}$$

$$\leqslant -\frac{C\varepsilon^2}{4}(n-n_{v-\varepsilon}),$$

if $b \leqslant v-\varepsilon$. Here, (22) follows from Lemma 3, (23) follows from Lemma 7 with $z=b$ and $u=\varepsilon/2$, and (24) follows from Lemma 9.

A similar inequality holds for $P(U_{(-\infty,t(1))})$. Then,

$$P(U_{(-\infty,v-\varepsilon)}) \leqslant P(U_{(-\infty,t(1))}) + \sum_{i=1}^{n_{v-\varepsilon}-1} P(U_{[t(i),t(i+1))}) + P(U_{[t(n_{v-\varepsilon}),v-\varepsilon)})$$

$$\leqslant (n_{v-\varepsilon}+1)\exp\left\{-\frac{C\varepsilon^2}{4}(n-n_{v-\varepsilon})\right\} \to 0, \tag{25}$$

as $n \to \infty$.

Consider the second term in (21). Similarly to (22)–(24),

$$\log P(U_{(a,t(i+1)]}) \leqslant \log P(D^{+}_{a,0})$$

$$\leqslant -\frac{\varepsilon}{2}\sum_{j=i+1}^{n} K(t(j)-a, t(j)-v-\varepsilon/2)$$

$$\leqslant -\frac{C\varepsilon}{2}(n-i)(a-v-\varepsilon/2) \leqslant -\frac{C\varepsilon^2}{4}n^{\alpha},$$

by Lemmas 3 and 8, for any $i \in [n_{v+\varepsilon}, n-\lfloor n^{\alpha}\rfloor - 1]$ and $a \in (t(i), t(i+1)]$ such that $a \geqslant v+\varepsilon$. Hence,

$$P(U_{(v+\varepsilon,t(n-\lfloor n^{\alpha}\rfloor)]}) \leqslant P(U_{(v+\varepsilon,t(n_{v+\varepsilon}+1)]}) + \sum_{j=n_{v+\varepsilon}+1}^{n-\lfloor n^{\alpha}\rfloor-1} P(U_{(t(i),t(i+1)]})$$

$$\leqslant (n-\lfloor n^{\alpha}\rfloor - n_{v+\varepsilon})\exp\left\{-\frac{C\varepsilon^2}{4}n^{\alpha}\right\} \to 0, \tag{26}$$

as $n \to \infty$.

Finally, consider the last probability in (21). Let $\gamma = (C\varepsilon^2/8)(n - n_{v+\varepsilon})$ and $U_\alpha = U_{t(n-\lfloor n^\alpha \rfloor),v,\gamma}$. From Lemma 4,

$$
\begin{aligned}
P(U_{(t(n-\lfloor n^\alpha \rfloor),t(n)]}) &\leqslant P(U_\alpha) + P(U_{(t(n-\lfloor n^\alpha \rfloor),t(n)]} \setminus U_\alpha) \\
&\leqslant P(U_\alpha) + P\left(\bigcup_{i=n-\lfloor n^\alpha \rfloor}^{n-1} D^+_{t(i),\gamma_i^*} \right) \\
&\leqslant P(U_\alpha) + \sum_{i \in V_{\lfloor n^\alpha \rfloor}} P(D^+_{t(i),\gamma/\lfloor n^\alpha \rfloor}) + \sum_{i \in W_{\lfloor n^\alpha \rfloor}} P(D^+_{t(i),0}). \qquad (27)
\end{aligned}
$$

We will show that each term in (27) converges to 0. From Lemma 6 with $z = t(n - \lfloor n^\alpha \rfloor)$ and $u = 1/2$ and Lemma 9,

$$
\begin{aligned}
\log P(U_\alpha) &\leqslant \frac{1}{2}\gamma - \sum_{j=n_v+1}^{n} J\left(t(j) - v, (t(j) - t(n - \lfloor n^\alpha \rfloor))^+ ; \frac{1}{2} \right) \\
&\leqslant \frac{1}{2}\gamma - \frac{C}{8} \sum_{j=n_v+1}^{n} (t(n - \lfloor n^\alpha \rfloor))^+ - v)^2 \\
&\leqslant \frac{1}{2}\gamma - \frac{C\varepsilon^2}{8}(n - n_{v+\varepsilon}) = -\frac{C\varepsilon^2}{16}(n - n_{v+\varepsilon}) \qquad (28)
\end{aligned}
$$

for sufficiently large n. Therefore, as $n \to \infty$,

$$
P(U_\alpha) = \mathrm{O}\left(\exp\left\{ -\frac{C\varepsilon^2 n}{16} \right\} \right) \to 0. \qquad (29)
$$

Next, for any $i \in V_{\lfloor n^\alpha \rfloor}$, from Lemma 8 with $z = t(i)$ and $u = 1/2$,

$$
\log P(D^+_{t(i),\gamma/\lfloor n^\alpha \rfloor}) \leqslant -\frac{\gamma}{2\lfloor n^\alpha \rfloor} \leqslant -\frac{C\varepsilon^2(n - n_{v+\varepsilon})}{16 n^\alpha},
$$

so that

$$
\sum_{i \in V_{\lfloor n^\alpha \rfloor}} P(D^+_{t(i),\gamma/\lfloor n^\alpha \rfloor}) \leqslant n^\alpha \exp\left\{ -\frac{C\varepsilon^2(n - \lfloor n^\alpha \rfloor)}{16 n^\alpha} \right\} \to 0, \qquad (30)
$$

as $n \to \infty$.

The last term in (27) vanishes if the set $W_{\lfloor n^\alpha \rfloor}$ is empty. Otherwise, let η be the cardinality of $W_{\lfloor n^\alpha \rfloor}$, and let w_i be the ith smallest element of $W_{\lfloor n^\alpha \rfloor}$. Then, applying Lemma 8 with $z = t(w_i)$, $\gamma = 0$, $u = \varepsilon/2$ and Lemma 9,

$$
\sum_{i \in W_{\lfloor n^\alpha \rfloor}} P(D^+_{t(i),0}) = \sum_{i=1}^{\eta} P(D^+_{t(w_i),0})
$$

$$\leqslant \sum_{i=1}^{\eta} \exp\left\{ -\frac{\varepsilon}{2} \sum_{j=w_i+1}^{n} K\left(t(j) - t(w_i), t(j) - v - \frac{\varepsilon}{2} \right) \right\}$$

$$\leqslant \eta \exp\left\{ \frac{-C\varepsilon}{2} \left(t(n - \lfloor n^\alpha \rfloor) - v - \frac{\varepsilon}{2} \right) \right\} \to 0, \tag{31}$$

as $n \to \infty$ because $\eta \leqslant n^\alpha$ and $t(w_i) \geqslant t(n - \lfloor n^\alpha \rfloor) \to \infty$.

From (29)–(31), we see that all three terms in (27) converge to 0. Combining this result with (25) and (26) shows that the probability in (21) also converges to 0, which proves consistency of $\hat{v}_{\mathrm{MLE}}$. □

As seen from the next theorem, the maximum likelihood estimator of the change point is always consistent in the case when $t(\infty) < \infty$, with no restrictions on the exponential family. The assumption of bounded sampling times is rather impractical; however, generalized broken-line regression models with $\lim t(i) < \infty$ can be used for covariates other than times of observations. Also, according to Remark 1 of Section 1, this model is equivalent to a nonlinear model with $t(\infty) = \infty$, where the parameter $\boldsymbol{\theta}_j$ leaves its control state and converges to a new value.

Theorem 3. *If* $t(\infty) < \infty$ *and* $\boldsymbol{\theta}_0 + \boldsymbol{\beta}\lambda \in \mathrm{int}(\Theta)$ *for all* $0 \leqslant \lambda \leqslant t(\infty)$, *then the maximum likelihood estimator of a change point is consistent.*

Proof. Similarly to (21) and (27),

$$P(|\hat{v}_{\mathrm{MLE}} - v| > \varepsilon) \leqslant P(U_{(-\infty,z)}) + P(U_{[z,v-\varepsilon)}) + P(U_{(v+\varepsilon,t(n-\lfloor n^\alpha \rfloor))})$$
$$+ P(U_\alpha) + \sum_{i \in V_{\lfloor n^\alpha \rfloor}} P(D^+_{t(i),\gamma_i^*}) + \sum_{i \in W_{\lfloor n^\alpha \rfloor}} P(D^+_{t(i),0}).$$

Boundedness of the sequence $t(j)$ implies, by Lemma 2, that $\varphi''(t)$ is separated from 0 on any closed interval that is contained in $[0, \tau)$. Let $\delta = \min\{\tau/2, 1\}$ and $z = \min\{v - \varepsilon, t(\infty) - \tau + \delta\}$. Then, for any $v \in [z, t(\infty)]$ and $t \leqslant t(\infty)$,

$$0 \leqslant (t - v)^+ \leqslant (t(\infty) - z)^+ = (\tau - \delta)^+ < \tau,$$

and $\varphi''((t - v)^+) = \boldsymbol{\beta}^{\mathrm{T}} \nabla^2 \psi(\boldsymbol{\theta}_0 + \boldsymbol{\beta}(t - v)^+)\boldsymbol{\beta} \geqslant C$ is separated from 0 by some $C > 0$. Repeating the proof of Theorem 2, we obtain that

$$P(U_{[z,v-\varepsilon)}), P(U_{(v+\varepsilon,t(n-\lfloor n^\alpha \rfloor)]}), P(U_\alpha), \sum_{i \in V_{\lfloor n^\alpha \rfloor}} P(D^+_{t(i),\gamma_i^*}) \to 0,$$

as $n \to \infty$, because (25), (26), (29), and (30) do not require that $t(\infty) < \infty$. Also,

$$\sum_{i \in W_{\lfloor n^\alpha \rfloor}} P(D^+_{t(i),0}) \leqslant P(W_{\lfloor n^\alpha \rfloor} \neq \emptyset) \to 0$$

in the case of convergent $\{t(n)\}$, because $t(i+1)-t(i)<1$ for all $i \geqslant n-\lfloor n^{\alpha} \rfloor$, if n is sufficiently large.

Thus, it remains to prove that $P(U_{(-\infty,z)})=P(U_{[t(\infty)-\tau,z)}) \to 0$. Notice that $z \leqslant t(\infty) - \tau/2$, hence $t(n)>z$ and $n_z^- < n$ for sufficiently large n. From Lemmas 3 and 7 with $u=\varepsilon/2$,

$$\log P(U_{[t(\infty)-\tau,z)}) \leqslant \log P(D_{z,0}^-) \leqslant -\frac{\varepsilon}{2} \sum_{j=n_z^-+1}^{n} K(t(j)-v+\varepsilon/2, t(j)-z). \quad (32)$$

Here $n-n_z^- \to 0$ and

$$K(t(j)-v+\varepsilon/2, t(j)-z) \geqslant M(j,\varepsilon)(v-\varepsilon/2-z) \geqslant M(j,\varepsilon)\varepsilon/2,$$

where $M(j,\varepsilon)=\min\{\varphi''(t) \mid t(j)-v+\varepsilon/2 \leqslant t \leqslant t(j)-z)$. We have

$$t(j)-z \leqslant \max\{\tau-\delta, t(\infty)-v-\varepsilon\} \leqslant \max\{\tau-\delta, \tau-\varepsilon\}$$

and for sufficiently large j,

$$t(j)-v+\varepsilon/2 \geqslant t(j)-t(\infty)+\varepsilon/2 > \varepsilon/4.$$

Hence,

$$M(j,\varepsilon) \geqslant \min_{[\varepsilon/4, \tau-\min\{\varepsilon,\delta\}]} \varphi''(t) > 0.$$

Thus, sum (32) diverges to ∞, so that $P(U_{[t(\infty)-\tau,z)}) \to 0$ which concludes the proof of the theorem. □

Remark 4 (Confidence estimation of the change point). Proofs of Theorems 2 and 3 allow confidence estimation of the change-point parameter v. Chernoff bounds (25), (26), (30), and (31) obtained in the proof of Theorem 2, along with (21) and (27), estimate the probability that $\hat{v}_{\mathrm{MLE}}$ does not belong to the ε-neighborhood of v. Subtracting their sum from 1 gives a lower bound for the coverage probability of the interval $\hat{v}_{\mathrm{MLE}} \pm \varepsilon$. As often happens, this coverage probability depends on the unknown parameter v. However, under assumption that the change point occurred, say, in a large interval $[a,b]$, one can estimate the confidence level and use $\hat{v}_{\mathrm{MLE}} \pm \varepsilon$ as the confidence interval for v.

5. Conclusion

We showed that unlike the classical *abrupt* change-point problems, it may be possible to estimate the change-point parameter consistently in the case of *gradual* changes. For the existence of consistent estimators, it is sufficient to assume an exponential family with $\varphi''(\theta)$ separated from 0. In this case, the maximum likelihood estimator is consistent. It was also shown that no consistent estimators exist when $\varphi'(\theta)$ is bounded from above. In stochastic modeling, an absolute majority of commonly used exponential families fall into either one or the other category.

For example, according to our results, consistent estimation of a change-point parameter is possible for Normal($\mu, 1$) and Poisson(λ) families with any β and Gamma($\alpha, 1$) family with $\beta > 0$, but not in the case of Binomial(n, p), Beta($\alpha, 1$), Exponential(λ) or Normal($0, \sigma^2$) families.

In the case of a bounded sequence of sampling times, no further assumptions are needed, and the maximum likelihood estimator is consistent under any exponential family. It is also consistent in the situation when the canonical parameter gradually drifts from its control state and converges to another value. This includes the case when the parameter approaches its the finite boundary of its domain, e.g. in the case of Gamma family with an expectation tending to infinity.

Appendix A

A.1. Proof of Lemma 3

Let $\mathbf{x} \in U_{[t(i),b)}$. Then the maximum value of g on $[t(i), b]$ is attained at $z \in [t(i), b) \cap \mathscr{M}_n$. Since g is concave down on $(t(i), t(i+1))$, it follows that

$$\lim_{y \uparrow b} g'(y) < \lim_{y \downarrow z} g'(y) \leqslant 0.$$

Thus, $\mathbf{x} \in D^-_{b,0}$ so that $U_{[t(i),b)} \subset D^-_{b,0}$.

Similarly, let $\mathbf{x} \in U_{(a,t(i+1)]}$. Then the maximum value of g on $[a, t(i+1)]$ is attained at $z \in (a, t(i+1)] \cap \mathscr{M}_n$, and

$$\lim_{y \downarrow a} g'(y) > \lim_{y \uparrow z} g'(y) \geqslant 0.$$

Thus, $\mathbf{x} \in D^+_{a,0}$ so that $U_{(a,t(i+1)]} \subset D^+_{a,0}$.

A.2. Proof of Lemma 4

Suppose $\mathbf{x} \notin \bigcup_{i=n-N}^{n-1} D^+_{t(i),\gamma_i^*}$. Then $\lim_{y \downarrow t(i)} g'(y) < \gamma_i^*$ for all $i \in [n-N, n-1]$ which implies that $g'(z) < \gamma_i^*$ for $z \in (t(i), t(i+1))$. Hence,

$$g(z) - g(t(i)) < \gamma_i^*(z - t(i))$$

for $z \in [t(i), t(i+1)]$.

To show that $\mathbf{x} \notin U_{(t(n-N),t(n)]} \setminus U_{t(n-N),v,\gamma}$, it suffices to show that $\mathbf{x} \notin U_{(t(n-N),t(n)]}$ if $\mathbf{x} \notin U_{t(n-N),v,\gamma}$. The latter implies that

$$g(v) - g(t(n-N)) > \gamma.$$

Choose arbitrary $j \in [n-N, n-1]$ and $z \in [t(j), t(j+1)]$. Then

$$g(v) - g(z) = g(v) - g(t(n-N)) + \sum_{i=n-N}^{j-1} \{g(t(i)) - g(t(i+1))\}$$

$$
\begin{aligned}
&+ g(t(j)) - g(z) \\
&> \gamma - \sum_{i=n-N}^{j-1} \gamma_i^*(t(i+1) - t(i)) - \gamma_j^*(z - t(j)) \\
&= \gamma - \sum_{i \in V_N \cap [n-N, j-1]} \frac{\gamma}{N}(t(i+1) - t(i)) - \gamma_j^*(z - t(j)) \\
&> \gamma - \sum_{i \in V_N \cap [n-N, j-1]} \frac{\gamma}{N} - \gamma_j^*(z - t(j)) \\
&\geqslant \gamma - \gamma(j - n + N + 1)/N \geqslant 0,
\end{aligned}
$$

so that $\mathbf{x} \notin U_{(t(n-N),t(n)]}$. The lemma is proved.

A.3. Proof of Lemma 5

Statements (a) and (b) are obvious, whereas (c) and (d) follow from the concavity of $g(z)$ that implies that $g'(z)$ is a decreasing function.

A.4. Proof of Lemma 6

It follows from (5) and (6) that

$$P(U_{z,v,\gamma}) = P(Y \geqslant y), \tag{A.1}$$

where

$$Y = \sum_{j=n_m+1}^{n} [(t(j) - z)^+ - (t(j) - v)^+]\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_j$$

and

$$y = -\gamma + \sum_{j=n_m+1}^{n} [\varphi((t(j) - z)^+) - \varphi((t(j) - v)^+)].$$

Since $\mathbf{E}_{\boldsymbol{\theta}} \exp\{\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{X}\} = \exp\{\psi(\boldsymbol{\theta} + \boldsymbol{\lambda}) - \psi(\boldsymbol{\theta})\}$ (see Brown, 1986), the moment generating function of Y can be computed as

$$
\begin{aligned}
M_Y(u) &= \mathbf{E} \exp(uY) \\
&= \prod_{j=n_m+1}^{n} \exp\{\varphi((1-u)(t(j) - v)^+ + u(t(j) - v)^+) - \varphi((t(j) - v)^+)\}.
\end{aligned}
$$

According to Chernoff inequality (20),

$$P(Y > y) \leqslant \exp\left\{ u\gamma - \sum_{j=n_m+1}^{n} \alpha_j(z) \right\} \tag{A.2}$$

with

$$\begin{aligned}\alpha_j(z) &= (1-u)\varphi((t(j)-v)^+) + u\varphi((t(j)-z)^+) \\ &\quad - \varphi((1-u)(t(j)-v)^+ + u(t(j)-z)^+) \\ &= (1-u)\int_{(1-u)(t(j)-v)^+ + u(t(j)-z)^+}^{(t(j)-v)^+} \varphi''(w)((t(j)-v)^+ - w)\,\mathrm{d}w \\ &\quad + u\int_{(1-u)(t(j)-v)^+ + u(t(j)-z)^+}^{(t(j)-z)^+} \varphi''(w)((t(j)-z)^+ - w)\,\mathrm{d}w, \\ &= J((t(j)-v)^+, (t(j)-z)^+; u),\end{aligned}$$

where we used the integral remainder form of Taylor's theorem as given in Thompson (1989).

A.5. Proof of Lemma 7

From (5) and (17),

$$P(D^-_{z,\gamma}) = P(Y \geqslant y), \tag{A.3}$$

where

$$Y = \sum_{j=n_z^-+1}^{n} \boldsymbol{\beta}^{\mathrm{T}} X_{t(j)}$$

and

$$y = -\gamma + \sum_{j=n_z^-+1}^{n} \varphi'(t(j)-z).$$

Then, from Chernoff inequality (20),

$$P(Y \geqslant y) \leqslant \exp\left\{ u\gamma - \sum_{j=n_z^-+1}^{n} \alpha_j(z) \right\}, \tag{A.4}$$

where

$$\begin{aligned}\alpha_j(z) &= \varphi(t(j)-v) - \varphi(t(j)-v+u) + u\varphi'(t(j)-z) \\ &= u\int_{t(j)-v+u}^{t(j)-z} \varphi''(w)\,\mathrm{d}w + \int_{t(j)-v}^{t(j)-v+u} \varphi''(w)(w-t(j)+v)\,\mathrm{d}w. \\ &= uK(t(j)-v+u, t(j)-z) + I(t(j)-v, t(j)-v+1; u).\end{aligned}$$

A.6. Proof of Lemma 8

This proof essentially repeats the steps in the proof of Lemma 7 with

$$Y = -\sum_{j=n_z+1}^{n} \boldsymbol{\beta}^{\mathrm{T}} X_{t(j)}$$

and

$$y = \gamma - \sum_{j=n_z+1}^{n} \varphi'(t(j) - z).$$

A.7. Proof of Lemma 9

If $\varphi''(w) \geqslant C$ on $[a, b]$, then

$$I(a, b; u) \geqslant C \int_{(1-u)a+ub}^{a} (a - w)\,\mathrm{d}w = \frac{1}{2} C u^2 (a - b)^2. \tag{A.5}$$

The inequality about $J(a, b; u)$ follows directly from (A.5), and the last statement of the lemma is obvious.

References

Basseville, M., Nikiforov, I.V., 1993. Detection of Abrupt Changes: Theory and Application. PTR Prentice-Hall, Inc., Englewood Cliffs, NJ.

Bhattacharya, P.K., 1995. Some aspects of change-point analysis. Change-Point Problems, IMS Lecture Notes—Monograph Series, Vol. 23. Institute of Mathematical Statistics, Hayward, CA, pp. 28–56.

Brown, L.D., 1986. Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory, Lecture Notes—Monograph Series, Vol. 9. Institute of Mathematical Statistics, Hayward, CA.

Carlstein, E., 1988. Nonparametric estimation of a change-point. Ann. Statist. 16, 188–197.

Chernoff, H., Zacks, S., 1964. Estimating the current mean of a normal distribution which is subject to changes in time. Ann. Math. Statist. 35, 999–1028.

Cobb, G.W., 1978. The problem of the Nile: conditional solution to a change-point problem. Biometrika 62, 243–251.

Gill, R., 2002. Introduction to generalized broken-line regression. Ph.D. Dissertation, The University of Texas at Dallas.

Gupta, A.K., Ramanayake, A., 2001. Change points with linear trends for the exponential distribution. J. Statist. Plann. Inference 93, 181–195.

Hinkley, D.V., 1970. Inference about the change-point in a sequence of random variables. Biometrika 57, 1–17.

Hušková, M., 1999. Gradual changes versus abrupt change. J. Statist. Plann. Inference 76, 109–125.

Jarušková, D., 1998. Testing appearance of linear trend. J. Statist. Plann. Inference 70, 263–276.

Lai, T.L., 1995. Sequential changepoint detection in quality control and dynamical systems. J. R. Statist. Soc. B 57, 613–658.

McCullagh, P., Nelder, J.A., 1983. Generalized Linear Models. Chapman & Hall, London.

Myers, R.H., Montgomery, D.C., Vining, G.G., 2002. Generalized Linear Models. With Applications in Engineering and the Sciences. Wiley-Interscience (John Wiley & Sons), New York.

Page, E.S., 1954. Continuous inspection schemes. Biomterika 41, 100–115.
Ross, R., 1994. A First Course in Probability. Macmillan, New York.
Siegmund, D.O., Zhang, H., 1995. Confidence regions in broken line regression. Change-Point Problems, IMS Lecture Notes—Monograph Series, Vol. 23. Institute of Mathematical Statistics, Hayward, CA, pp. 292–316.
Thompson, H.B., 1989. Taylor's theorem using the generalized Riemann integral. Amer. Math. Monthly 96, 346–350.
Zacks, S., 1983. Survey of classical and Bayesian approaches to the change-point problem: fixed sample and sequential procedures in testing and estimation. Recent Adv. in Statist. Academic Press, New York, pp. 245–269.