EXPLICIT SOLUTIONS IN A SMOOTH CHANGE PROBLEM

ANDREW L. RUKHIN
University of Maryland, Baltimore County

A setting of the change-point estimation problem with a "smooth" change is investigated. A special class of such models, where the change is defined as a gradual shift in the parameter, is considered. We show that these models can be analyzed via explicitly given procedures. It is demonstrated that M-estimators can consistently estimate the change in the parameter structure as well as the additional unknown parameter. Under mild regularity assumptions, asymptotic normality and asymptotic efficiency of these procedures are established.

1. Introduction

The traditional setting of the change-point problem assumes that the change is abrupt and permanent. Clearly there are situations where this assumption is not met, and it is desirable to estimate the time moment at which the stationary character of observations has changed, or to test the hypothesis that such a change has indeed occurred. The possibility of simultaneous consistent estimation of the parameters characterizing the change of means has been demonstrated by Yao and Au (1989), who considered nonlinear regression models with step functions describing the evolution of the mean. Huang and Chang (1993) studied models with smooth change intervention when the observations during the change period have the distributions which are mixtures of the pre- and after- change distributions. Rukhin and Vajda (1997) investigated a general nonlinear regression model for the change in the mean function. See Brodsky and Darkhovsky (1993) and Csorgo and Horvath (1998) for further references.

Here we study a model with a smooth change in the parameter. As in Husková (1996) minimum contrast estimators are used for estimation of the change rate.

2. The model

The following model relates the change estimation to nonlinear parameter estimation. Consider a parametric family of probability distributions P_θ, $\theta \in \Theta$, $\Theta \subset R^p$ with Θ being an open convex subset of Euclidean space. Let $f(\cdot \mid \theta)$ denote the corresponding densities. We assume regularity conditions

Keywords and phrases: asymptotic efficiency; change-point problem; consistency; exponential family; information matrix; M-estimators.

AMS subject classifications: Primary 62F12; Secondary 62E20.

specified, for instance, in Section 7, Chapter 1 of Ibragimov and Hasminski (1981). These conditions guarantee differentiability of the expected values of functions with finite second moment, and they imply the existence of the Fisher information matrix $I(\theta) = E_\theta \ell(X \mid \theta)[\ell(X \mid \theta)]^T$ with $\ell(x \mid t) = \partial \log f(x \mid t)/\partial t = \nabla \log f(x \mid t)$, when $f(x \mid t)$ is a differentiable function of t. Actually in Proposition 3.1, it is assumed that this function is twice differentiable, although a milder condition concerning weak differentiability in t of $s(x \mid t) = \sqrt{f(x \mid t)}$ suffices. In the latter case the information matrix can be defined as $I(t) = 4 \int s(x \mid t)[s(x \mid t)]^T \, dx$, which is supposed to be nonsingular.

In our setting the observations x_j, $j = 1, \ldots, n$, are independent with x_j having a distribution $P_{\gamma+\varphi(j)\theta}$. Here $\varphi(j) = \varphi(j, n)$ is a monotonically increasing sequence in j for any fixed n; for j fixed, $\varphi(j, n) \to 0$ as $n \to \infty$. The form of the sequence $\varphi(j)$ is supposed to be known. In the application to volumetric analysis discussed in Section 4, $\varphi(j, n)$ is proportional to $F(j/n)$, where F is the distribution function on $(0, 1)$.

Our goal is to estimate the parameters γ and θ. In the situation where the initial parameter γ is given, we derive an explicit form of asymptotically normal and efficient estimates of the parameter θ. When γ is unknown, these estimators have a less explicit form, but still can be easily implemented in practice.

3. M-estimators: consistency and asymptotic normality

3.1. Known parameter γ

Let us start with the situation where γ is known. Then θ is a common parameter in a sequence of independent observations; it is assumed that the maximum likelihood estimator,

$$\hat{\theta}(x_1, \ldots, x_n) = \arg\max \sum_1^n \log f(x_j | \gamma + \varphi_j \theta),$$

is uniquely defined.

To prove the asymptotic normality and asymptotic efficiency of this estimator a variety of conditions have been developed. The following assumptions A.I–A.III translate conditions (4.1)–(4.4) from Section 4 of Chapter III in Ibragimov and Hasminski (1981). More general conditions which guarantee the asymptotic normality of the maximum likelihood estimator are given by LeCam (1986), Section 16.3 and by Strasser (1985), Section 80.6.

Denote by $\mathbf{1}_A(u)$ the indicator function of an event A.

A.I. With

$$J_n^2 = J^2(n, \theta) = \sum_1^n \varphi_j^2 I(\gamma + \varphi_j \theta)$$

denoting the Fisher information matrix about θ, for any $\epsilon > 0$ and $u \in R^p$

$$\limsup_{n\to\infty} \sup_{\Theta} \sum_{j=1}^{n} E_\theta[\varphi_j u^T J_n^{-1} \ell_j(X \mid \gamma + \varphi_j \theta)]^2 \\ \times \mathbf{1}_{(\epsilon,\infty)} |\varphi_j u^T J_n^{-1} \ell_j(X \mid \gamma + \varphi_j \theta)| = 0.$$

A.II. With $\|A\|$ denoting the spectral norm of matrix A,

$$\sup_{n,\theta_1,\theta_2} \|J^{-1}(n,\theta_1) J^2(n,\theta_2) J^{-1}(n,\theta_1)\| < \infty$$

and

$$\limsup_{n\to\infty} \operatorname{tr}\left(J_n^{-2}\right) = 0.$$

A.III. For a sequence $\xi_n \to \infty$ with $\tilde{s}(x \mid t) = \nabla s(x \mid t) = \nabla\sqrt{f(x \mid t)}$,

$$\sup_{\theta} \sup_{|u|\le \xi_n} \frac{1}{|u|^2} \sum_{j=1}^{n} \int \left[\left[\tilde{s}(x \mid \gamma + \varphi_j(\theta + J_n^{-1}u)) - \tilde{s}(x \mid \gamma + \varphi_j \theta) \right]^T J_n^{-1} u \right]^2 dx \\ \to 0$$

and for some $\beta > 0$

$$\inf_{\theta} \inf_{|u|\ge \xi_n} \frac{1}{\|J_n\|^\beta} \sum_{j=1}^{n} \int \left[s(x \mid \gamma + \varphi_j(\theta + J_n^{-1}u)) - s(x \mid \gamma + \varphi_j \theta) \right]^2 dx > 0.$$

Under conditions A.I–A.III the random vector $J_n[\hat{\theta} - \theta]$ is asymptotically normal with zero mean and the identity covariance matrix $\mathbf{I}$. Also $\hat{\theta}$ is asymptotically efficient in the sense that for any unimodal symmetric loss function W

$$\lim_{n\to\infty} E_\theta W\big(J_n(\hat{\theta} - \theta)\big) = (2\pi)^{-p/2} \int W(u) \exp\{-\|u\|^2/2\}\, du \\ \le \liminf_{n\to\infty} E_\theta W\big(J_n(\delta_n - \theta)\big)$$

for any regular sequence of estimators δ_n (Ibragimov and Hasminski, 1981, Chapter III, Section 4, p. 191).

In particular, the following matrix inequality holds,

$$\liminf_{n\to\infty} J_n E_\theta (\delta_n - \theta)(\delta_n - \theta)^T J_n \ge \mathbf{I}.$$

Thus, the maximum likelihood estimator (as well as the Bayes estimator against a smooth positive prior density) is asymptotically efficient as it has the "smallest" limiting covariance matrix.

The following (approximate maximum likelihood) estimator possesses the same properties under the following conditions. Denote by $D(x \mid \gamma)$ the Hessian (the matrix of second partial derivatives) of the function $f(x \mid \gamma)$. Note that in one-dimensional case the following condition (3.1) means that

$$\int \frac{f''(x|\gamma_1) f'(x|\gamma)}{f(x|\gamma)}\, dx < \infty.$$

Proposition 3.1. *Assume that* A.I–A.III *are valid and for any* γ_1

$$\int D(x \mid \gamma_1)\ell(x \mid \gamma)\, dx < \infty. \tag{3.1}$$

Then, under the conditions

$$\frac{\sum_j \varphi_j^3}{\sum_j \varphi_j^2} \to 0, \tag{3.2}$$

and

$$\sum_{j=1}^{n} \varphi_j^2 \to \infty, \tag{3.3}$$

the estimator of θ

$$\delta(x_1, \ldots, x_n) = \left[\sum_j \varphi_j^2\right]^{-1} I^{-1}(\gamma) \sum_j \varphi_j \ell(x_j \mid \gamma),$$

is asymptotically normal and asymptotically efficient.

Proof. Taylor's formula with remainder shows that for some γ_1

$$\begin{aligned} E_\theta \ell(x_j \mid \gamma) &= \int f(x \mid \gamma + \varphi_j \theta)\ell(x \mid \gamma)\, dx \\ &= \int f(x \mid \gamma)\ell(x \mid \gamma)\, dx + \varphi_j \left[\int \nabla f(x \mid \gamma)\ell(x \mid \gamma)\, dx\right]\theta^T \\ &\qquad + \frac{\varphi_j^2}{2}\theta^T \left[\int D(x \mid \gamma_1)\ell(x \mid \gamma)\, dx\right]\theta \\ &= \varphi_j I(\gamma)\theta + O(\varphi_j^2). \end{aligned}$$

Here $\nabla f(x \mid \gamma) = f(x \mid \gamma)\ell(x \mid \gamma)$ is the gradient of f, and the term $O(\varphi_j^2)$ is uniform in j, i.e., it is bounded above by $C\varphi_j^2$ with a constant C which does not depend on j (but depends on θ and γ.) Similarly, the covariance matrix of $\ell(x_j|\gamma)$ has the form

$$\mathrm{Cov}\big(\ell(x_j|\gamma)\big) = I(\gamma) + O(\varphi_j),$$

where the term $O(\varphi_j)$ is uniform in j. Thus, according to (3.2)

$$E_\theta \delta(x_1, \ldots, x_n) = \theta + O\left(\frac{\sum_j \varphi_j^3}{\sum_j \varphi_j^2}\right) = \theta + o(1),$$

and

$$\operatorname{Cov}(\delta) = \frac{1}{(\sum_j \varphi_j^2)^2} I^{-1}(\gamma)\Big[\sum_{j=1}^{n} \varphi_j^2 \operatorname{Cov}(\ell(x_j|\gamma))\Big] I^{-1}(\gamma)$$
$$= \frac{1}{\sum_j \varphi_j^2} I^{-1}(\gamma) + O\Big(\frac{\sum_j \varphi_j^3}{[\sum_j \varphi_j^2]^2}\Big).$$

Therefore under condition (3.3), $\operatorname{Cov}(\delta) \to 0$, and because of A.I–A.III the Feller–Lindeberg Theorem applies. Therefore, $[\sum_1^n \varphi_j^2]^{1/2} I^{1/2}(\gamma)[\delta - \theta]$, as well as $J_n[\delta - \theta]$, are asymptotically standard normal, and δ is asymptotically efficient. □

3.2. Unknown parameter γ

If γ is unknown, one has to estimate it. For this purpose we introduce the following condition.

(C) For a sequence of i.i.d. random variables y_j, $j = 1, \dots, n$, with y_j having distribution P_γ, there exists a consistent estimator $\gamma^\star(y_1, \dots, y_n)$ of γ, which is robust in the following sense: when the distribution of x_j is $P_{\gamma_0 + \Delta_j}$, where $\max_j \Delta_j \to 0$, then $\gamma^\star(x_1, \dots, x_n) \to \gamma_0$ in probability.

Under general regularity conditions similar to those in Section 65 of Borovkov (1998), M-estimators satisfy the condition (C). To see that, let $\rho(x \mid \gamma)$ be a contrast function, i.e., for all $\gamma \neq \gamma_0$

$$E_{\gamma_0} \rho(X|\gamma) < E_{\gamma_0} \rho(X|\gamma_0).$$

We will assume that $\rho(x \mid \gamma)$ is uniformly integrable and continuous in γ for almost all x. Let $\Delta(\gamma)$ denote a ball of radius Δ centered at γ.

To verify condition (C) we need the following assumptions.

C.I. For any $\Delta > 0$ one can find $\epsilon > 0$ such that

$$\min_{t \notin \Delta(\gamma)} E_\gamma[\rho(X \mid t) - \rho(X \mid \gamma)] \leq -\epsilon.$$

C.II. With $\rho^\Delta(x \mid \gamma) = \max_{t \in \Delta(\gamma)} \rho(x \mid t)$, for any $\alpha > 0$ there exists $\Delta > 0$ such that for all $|t - \gamma| > \alpha$

$$E_{\gamma_0}[\rho^\Delta(X \mid t) - \rho(X \mid \gamma_0)] < -\epsilon < 0.$$

C.III. For any positive Δ and any $t \in \Theta$

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} E_{\gamma_0}[\rho^\Delta(x_j \mid t) - \rho(x_j \mid \gamma_0)] \mathbf{1}_{(\tau,\infty)}[\rho^\Delta(x_j \mid t) - \rho(x_j \mid \gamma_0)] = 0.$$

Define the $\widehat{M}$-estimator $\gamma^\star$ as the maximizer of $\sum_{j=1}^n \rho(x_j \mid \gamma)$,

$$\gamma^\star(x_1, \ldots, x_n) = \arg\max \sum_{j=1}^n \rho(x_j|\gamma).$$

This estimator is known to be (strongly) consistent when $x_1, \ldots, x_n$ form a random sample. It is also asymptotically normal under mild regularity conditions on $\rho(x \mid \gamma)$.

Lemma 3.1. *If independent random variables $x_j, j = 1, \ldots, n$ have distributions $P_{\gamma_0+\Delta_j}$ with $\max_j \Delta_j \to 0$, then under conditions* C.I, C.II *and* C.III

$$\gamma^\star \to \gamma_0 \quad \textit{almost surely},$$

so that condition (C) *is satisfied.*

Proof. We will show that for any neighborhood $\alpha(\gamma_0)$ of γ_0 with probability one

$$\limsup_{n\to\infty} \sup_{\gamma\notin\alpha(\gamma_0)} \frac{1}{n}\sum_{j=1}^n [\rho(x_j \mid \gamma) - \rho(x_j \mid \gamma_0)] < -\epsilon$$

for some positive ϵ. This will prove Lemma 3.1.

For the Δ from condition C.II, one can find a finite covering of $\Theta \setminus \alpha(\gamma_0)$ by balls $\Delta(t_k)$, $k = 1, \ldots, K$. Then

$$\begin{aligned}\sup_{\gamma\notin\alpha(\gamma_0)} \frac{1}{n}\sum_{j=1}^n [\rho(x_j \mid \gamma) - \rho(x_j \mid \gamma_0)] &\le \max_k \frac{1}{n}\sum_{j=1}^n \sup_{t\in\Delta(t_k)} [\rho(x_j \mid t) - \rho(x_j \mid \gamma_0)] \\ &\le \max_k \frac{1}{n}\sum_{j=1}^n \sup_{t\in\Delta(t_k)} [\rho(x_j \mid t) - \rho(x_j \mid \gamma_0)].\end{aligned}$$

Since

$$\begin{aligned}&P\big(\gamma \notin \alpha(\gamma_0)\big) \\ &= P\bigg(\sup_{\gamma\notin\alpha(\gamma_0)} \frac{1}{n}\sum_{j=1}^n [\rho(x_j|\gamma) - \rho(x_j|\gamma_0)] > \sup_{\gamma\in\alpha(\gamma_0)} \frac{1}{n}\sum_{j=1}^n [\rho(x_j|\gamma) - \rho(x_j|\gamma_0)]\bigg) \\ &\le P\bigg(\sum_{j=1}^n \sup_{\gamma\notin\alpha(\gamma_0)} \frac{1}{n}[\rho(x_j \mid \gamma) - \rho(x_j \mid \gamma_0)] > 0\bigg),\end{aligned}$$

it suffices to show that for any $t \notin \alpha(\gamma_0)$

$$P\bigg(\frac{1}{n}\sum_{j=1}^n [\rho^\Delta(x_j \mid t) - \rho(x_j \mid \gamma_0)] > 0\bigg) \to 0. \tag{3.4}$$

Notice that our conditions imply that for any $t \notin \alpha(\gamma_0)$ one can find $\Delta > 0$ and $N > 0$ such that for some positive ϵ

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} E_{\gamma_0}[\rho^{\Delta}(x_j \mid t) - \rho(x_j \mid \gamma_0)]\mathbf{1}_{(-N,\infty)}(\rho^{\Delta}(x_j \mid t) - \rho(x_j \mid \gamma_0)) < -\epsilon.$$

The strong law of large numbers implies that

$$P\Big(\frac{1}{n}\sum_{j=1}^{n}[\rho^{\Delta}(x_j \mid t) - \rho(x_j \mid \gamma_0)]\mathbf{1}_{(-N,\infty)}(\rho^{\Delta}(x_j \mid t) - \rho(x_j \mid \gamma_0)) > 0\Big) \to 0,$$

which proves (3.4). □

Now we turn to the estimation problem of (γ, θ). Our main result for the case where γ is unknown is the following.

Theorem 3.1. *Under condition* (C), *let the independent random variables* x_j, $j = 1, \ldots, n$ *have distributions* $P_{\gamma+\varphi_j\theta}$. *There exists a consistent estimator* $(\hat{\gamma}^\star, \hat{\theta}^\star)$ *of the pair* (γ, θ) (*explicitly defined in* (3.5) *below*). *Under assumptions* A.I–A.III *with* θ *replaced by* (γ, θ), *this pair can be chosen to be asymptotically normal and asymptotically efficient.*

Proof. According to condition (C) with $\Delta_j = \varphi_j\theta$, $j = 1, \ldots, m$, one can find a consistent estimator $\gamma^\star = \gamma^\star(x_1, \ldots, x_m)$ on the basis of the first $m = o(n)$, $m \to \infty$, observations $x_1, \ldots, x_m$. This estimator can be used as a plug-in estimate of γ in the rule δ from Proposition 3.1, i.e.,

$$\theta^\star(x_1, \ldots, x_n) = \frac{1}{\sum_j \varphi_j^2} I^{-1}(\gamma^\star) \sum_{j=m+1}^{n} \varphi_j \ell(x_j \mid \gamma^\star).$$

Then the resulting estimator $\theta^\star$ is also consistent, and asymptotically normal.

We use this pair $(\gamma^\star, \theta^\star)$ as a one-step approximation to the likelihood equation in the Newton–Raphson method. Namely, put

$$(3.5) \qquad \begin{pmatrix} \hat{\gamma}^\star \\ \hat{\theta}^\star \end{pmatrix} = \begin{pmatrix} \gamma^\star \\ \theta^\star \end{pmatrix} + \mathbf{L}^{-1}(n|\gamma^\star, \theta^\star) \sum_j \begin{pmatrix} \ell(x_j \mid \gamma^\star + \varphi_j\theta^\star) \\ \varphi_j \ell(x_j \mid \gamma^\star + \varphi_j\theta^\star) \end{pmatrix}.$$

Then

$$\begin{aligned}
&\mathbf{L}^{1/2}(n \mid \gamma,\theta)\begin{pmatrix}\hat{\gamma}^\star-\gamma\\ \hat{\theta}^\star-\theta\end{pmatrix}\\
&\quad=\mathbf{L}^{1/2}(n \mid \gamma,\theta)\begin{pmatrix}\gamma^\star-\gamma\\ \theta^\star-\theta\end{pmatrix}\\
&\qquad+\mathbf{L}^{1/2}(n \mid \gamma,\theta)\mathbf{L}^{-1}(n \mid \gamma^\star,\theta^\star)\sum_j\begin{pmatrix}\ell(x_j \mid \gamma^\star+\varphi_j\theta^\star)\\ \varphi_j\ell(x_j \mid \gamma^\star+\varphi_j\theta^\star)\end{pmatrix}\\
&\quad=\mathbf{L}^{1/2}(n \mid \gamma,\theta)\begin{pmatrix}\gamma^\star-\gamma\\ \theta^\star-\theta\end{pmatrix}+\mathbf{L}^{-1/2}(n \mid \gamma,\theta)\sum_j\begin{pmatrix}\ell(x_j \mid \gamma+\varphi_j\theta)\\ \varphi_j\ell(x_j|\gamma+\varphi_j\theta)\end{pmatrix}\\
&\qquad-\mathbf{L}^{1/2}(n \mid \gamma,\theta)\mathbf{L}^{-1}(n \mid \gamma,\theta)\mathbf{L}(n|\gamma_1,\theta_1)\begin{pmatrix}\gamma^\star-\gamma\\ \theta^\star-\theta\end{pmatrix}\\
&\quad=\mathbf{L}^{-1/2}(n \mid \gamma,\theta)\sum_j\begin{pmatrix}\ell(x_j \mid \gamma+\varphi_j\theta)\\ \varphi_j\ell(x_j \mid \gamma+\varphi_j\theta)\end{pmatrix}+o_P(1).
\end{aligned}$$

Let $\hat{\gamma}$, $\hat{\theta}$ denote the maximum likelihood estimators. Then because of the known results concerning the asymptotic behavior of maximum likelihood estimators under local asymptotic normality conditions (Theorem 8.1 of Chapter I in Ibragimov and Hasminski (1981)), the estimators $\hat{\gamma}$, $\hat{\theta}$ are jointly asymptotically normal, and

$$\mathbf{L}^{1/2}(n \mid \gamma,\theta)\begin{pmatrix}\hat{\gamma}-\gamma\\ \hat{\theta}-\theta\end{pmatrix}=\mathbf{L}^{-1/2}(n \mid \gamma,\theta)\sum_j\begin{pmatrix}\ell(x_j \mid \gamma+\varphi_j\theta)\\ \varphi_j\ell(x_j \mid \gamma+\varphi_j\theta)\end{pmatrix}+o_P(1).$$

It follows that

$$\mathbf{L}^{1/2}(n \mid \gamma,\theta)\begin{pmatrix}\hat{\gamma}^\star-\hat{\gamma}\\ \hat{\theta}^\star-\hat{\theta}\end{pmatrix}\to 0,$$

and $\mathbf{L}^{1/2}(n \mid \gamma,\theta)\begin{pmatrix}\hat{\gamma}^\star-\gamma\\ \hat{\theta}^\star-\theta\end{pmatrix}$ is asymptotically normal with zero mean and identity covariance matrix. □

Consistency part of Theorem 3.1 holds for M-estimators of (γ,θ), which are defined as the solution of the equation, $\sum_j \psi_j(x_j \mid \gamma+\varphi_j\theta)=0$, as well as for the $\hat{M}$-estimators, which maximize $\sum_j \rho_j(x_j \mid \gamma+\varphi_j\theta)$. When independent observations $x_1,\dots,x_n$ have different distributions for different contrast functions ρ_j or ψ_j, it is assumed that for $(\gamma,\theta)\neq(\gamma_0,\theta_0)$,

$$E_{\gamma_0+\varphi_j\theta_0}[\rho_j(X_j \mid \gamma+\varphi_j\theta)-\rho_j(X_j \mid \gamma_0+\varphi_j\theta_0)]<0,$$

or the (unique) solution of $E_{\gamma_0+\varphi_j\theta_0}\psi_j(X_j \mid \gamma+\varphi_j\theta)=0$ is $(\gamma,\theta)=(\gamma_0,\theta_0)$. More precisely, for $\hat{M}$-estimators with $s=(\gamma,\theta)$, let

$$\kappa_i^{\Delta}(x,s)=\sup_{u\in\Delta(s)}\rho_i(x \mid u),$$

$$\eta_i^{\Delta}(x,s) = \kappa_i^{\Delta}(x,s) - \rho_i(x \mid s).$$

The following conditions B.I–B.II and C.I–C.II with $\rho = \rho_j$, guarantee that the $\widehat{M}$-estimator of (γ, θ) is consistent. Also the condition (C) holds then with $\gamma^\star$ being the $\hat{M}$-estimator (or M-estimator).

B.I. For any $\delta > 0$ and t, $|t - (\gamma, \theta)| > \delta$, one can find $\Delta > 0$ and N such that with

$$b_n(t) = \sum_{j=1}^{n} E_{\gamma+\varphi_j\theta} |\eta_i^{\Delta}(X_j, t)| \mathbf{1}_{(-N,\infty)} \big(\eta_i^{\Delta}(X_j, t)\big) \to \infty,$$

one has

$$\limsup_{n\to\infty} \frac{1}{b_n(t)} \sum_{j=1}^{n} E_{\gamma+\varphi_j\theta} \eta_i^{\Delta}(X_j, t) \mathbf{1}_{(-N,\infty)} \big(\eta_i^{\Delta}(X_j, t)\big) < -\epsilon,$$

B.II. For any $\tau > 0$ and t ,

$$\limsup_{n\to\infty} \frac{1}{b_n(t)} \sum_{j=1}^{n} E_{\gamma+\varphi_j\theta} \eta_i^{\Delta}(X_j, t) \mathbf{1}_{(\tau b_n(t),\infty)} \big(\eta_i^{\Delta}(X_j, t)\big) = 0.$$

Thus conditions B.I and B.II can be used as a replacement for the condition C.III. While B.II is similar in spirit to C.III, to elucidate the role of condition B.I assume that for some $c > 0$,

$$c b_n(\gamma, \theta) \le \sum_{j=1}^{n} E_{\gamma_0+\varphi_j\theta_0} [\rho(x_j \mid \gamma + \varphi_j\theta) - \rho(x_j \mid \gamma_0 + \varphi_j\theta_0)].$$

Then B.I implies that the sum in the right-hand side tends to infinity for any parametric values γ, θ different from the true values γ_0, θ_0.

For the maximum likelihood estimator, when $\rho(x \mid t) = \log f(x \mid t)$, this fact means that

$$\sum_{j=1}^{n} E_{\gamma_0+\varphi_j\theta_0} \log \frac{f(x_j \mid \gamma_0 + \varphi_j\theta_0)}{f(x_j \mid \gamma + \varphi_j\theta)} \to \infty,$$

i.e., the series formed by information numbers diverges.

Denote by

$$\mathbf{L}(n \mid \gamma, \theta) = \begin{pmatrix} \sum_{j=1}^{n} I(\gamma + \varphi_j\theta) & \sum_{j=1}^{n} \varphi_j I(\gamma + \varphi_j\theta) \\ \sum_{j=1}^{n} \varphi_j I(\gamma + \varphi_j\theta) & \sum_{j=1}^{n} \varphi_j^2 I(\gamma + \varphi_j\theta) \end{pmatrix}$$

the sum of information matrices for x_j about two parameters γ and θ. The divergence above holds if the minimal eigenvalue of $\mathbf{L}(n|\gamma, \theta)$ tends to infinity,

$$\lambda_{\min}\big(\mathbf{L}(n|\gamma,\theta)\big) \to \infty. \tag{3.6}$$

If (3.6) holds and the minimal eigenvalue, $\lambda_{\min}\big(I(\gamma+\varphi_j\theta)\big)$, is bounded from below by a positive constant,

$$\inf_\theta \lambda_{\min}\big(I(\gamma+\varphi_j\theta)\big) > \lambda > 0,$$

then, by taking $\theta = 0$, one obtains

$$\lambda_{\min}\begin{pmatrix} n & \sum_{j=1}^n \varphi_j \\ \sum_{j=1}^n \varphi_j & \sum_{j=1}^n \varphi_j^2 \end{pmatrix} \to \infty.$$

Thus, in this situation, B.I with $\bar\varphi = n^{-1}\sum_{j=1}^n \varphi_j$ implies that

$$\frac{\sum_{j=1}^n (\varphi_j - \bar\varphi)^2}{1+\sum_{j=1}^n \varphi_j^2/n} \to \infty. \tag{3.7}$$

To implement (3.5), observe that the inverse of the matrix $\mathbf{L}(n \mid \gamma, 0)$ is needed. In the simplest case when $\theta = 0$, it has the form

$$\mathbf{L}^{-1}(n \mid \gamma,0) = \frac{1}{[n\sum_j \varphi_j^2 - (\sum_j \varphi_j)^2]}\begin{pmatrix} \sum_{j=1}^n \varphi_j^2 I^{-1}(\gamma) & -\sum_{j=1}^n \varphi_j I^{-1}(\gamma) \\ -\sum_{j=1}^n \varphi_j I^{-1}(\gamma) & nI^{-1}(\gamma) \end{pmatrix}.$$

Notice that an asymptotically optimal test of the null hypothesis $\theta = 0$ can be taken to have the critical region $\{\theta^\star > C\}$.

4. Example and application

Assume that the density, $f(\cdot \mid \theta)$, of P_θ, belongs to an exponential family, i.e.,

$$f(x \mid \theta) = \exp\{\theta^T x - \chi(\theta)\}.$$

In this situation the maximum likelihood estimators $\hat\gamma$ and $\hat\theta$ satisfy the following simultaneous equations,

$$\sum_{j=1}^n x_j = \sum_{j=1}^n \nabla\chi(\hat\gamma + \varphi_j\hat\theta),$$

$$\sum_{j=1}^n \varphi_j x_j = \sum_{j=1}^n \varphi_j \nabla\chi(\hat\gamma + \varphi_j\hat\theta).$$

Notice that in this case statistics $\sum x_j$ and $\sum \varphi_j x_j$ are sufficient for parameters γ and θ.

In the particular case when f is in the family of multivariate normal densities with identity covariance matrix, $\chi(t) = ||t||^2/2$, so that

$$\hat\gamma(x_1,\ldots,x_n) = \frac{1}{n\sum_j \varphi_j^2 - (\sum_j \varphi_j)^2}\left[\sum_{j=1}^n \varphi_j^2 \sum_{j=1}^n x_j - \sum_{j=1}^n \varphi_j \sum_{j=1}^n \varphi_j x_j\right],$$

$$\hat\theta(x_1,\ldots,x_n) = \frac{1}{n\sum_j \varphi_j^2 - (\sum_j \varphi_j)^2}\left[n\sum_{j=1}^n \varphi_j x_j - \sum_{j=1}^n \varphi_j \sum_{j=1}^n x_j\right].$$

Both of these normally distributed estimators are unbiased. The condition for their consistency is given by (3.7), which can be verified directly by evaluation of the corresponding covariance matrices.

An application of smooth-change models appears in volumetric analysis where the chemical analysis is performed by measuring the volume of a solution needed to react completely with the substance of interest. In the so-called titration method, an analyte is added progressively to a reagent (titrant) solution until all analyte is consumed (i.e., the equivalence point is reached.) The data, consisting of subsequent readings of the analyte in the solution, can be interpreted as as observations from the family of distributions in our model with $\varphi(j,n) = \phi_n F(j/n)$ with a distribution function F defined on the unit interval $[0,1]$. (The shape of this function is determined by the strength of the acid and of the base, and can be assumed known.) Commonly, the data is believed to be log-normally distributed and the uniform distribution function F, $F(x) = x$ is accepted.

The estimators $\hat{\gamma}$ and $\hat{\theta}$ above, with $x_1, \ldots, x_n$ being replaced by the logarithms of observed analyte's content, turn out to be unbiased and efficient provided that $\phi_n \to 0$, which implies conditions (3.2), (3.3) and (3.7).

Acknowledgements. This research was supported by von Humboldt Foundation. The author is grateful to A. A. Borovkov for interesting discussion. Excellent editorial work of M. Moore is also acknowledged.

REFERENCES

Borovkov, A.A. (1998). *Mathematical Statistics.* Nauka, Novosibirsk. (in Russian).

Brodsky, B.E. and Darkhovsky, B.S. (1993). *Nonparametric Methods in Change-Point Problems.* Kluwer, Dordrecht.

Csorgo, M. and Horvath, L. (1998). *Limit Theorems in Change-Point Analysis.* Wiley, New York.

Husková, M. (1996). Tests and estimators for change point based on M-statistics. *Statist. Decisions* 14, 115–136.

Huang, W.-T., Chang, Y.-P. (1993). Nonparametric estimation in change-point models. *J. Statist. Plan. Inference* 35, 335–347.

Ibragimov, I.A. and Hasminski, R.Z. (1981). *Statistical Estimation: Asymptotic Theory.* Springer-Verlag, New York.

LeCam, L., (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer, New York.

Rukhin, A.L., Vajda, I. (1997). Change-point estimation as a nonlinear regression problem, *Statistics* 30, 181–200.

Strasser, W. (1985). *Mathematical Theory of Statistics.* de Gruyter, Berlin.

Yao, Y.C and Au, S.T. (1989). Least-squares estimation of a step function. *Sankhyā Ser.* A 51, 370–381.

Andrew L. Rukhin
Department of Mathematics and Statistics
University of Maryland, Baltimore County
Baltimore, MD 21250
USA
rukhin@math.umbc.edu