JSPI 1936
pp: 1--19 (col.fig.: nil)
PROD. TYPE: COM
ED: Ramesh
PAGN: Suresh – SCAN: Profi

ARTICLE IN PRESS

ELSEVIER

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

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Asymptotic behavior of posterior distribution of the change-point parameter

Andrew L. Rukhin *

Department of Mathematics and Statistics, University of Maryland at Baltimore County, Baltimore, MD 21250, USA

Received 22 June 2000; received in revised form 9 July 2001; accepted 23 July 2001

Abstract

In the asymptotic setting of change-point estimation the behavior of the posterior distribution and of Bayes procedures is studied. The limiting distribution is derived when the prior probabilities converge to geometric probabilities. This distribution is related to the infinite product of random matrices (or affine transforms of the real line). The situation with partially unknown pre- and after-change distributions is investigated. A condition for the limiting law of posterior probabilities to coincide with that for the known pre- and after-change distributions is derived.

MSC: 62F12; 62E20; 62A15

Keywords: Adaptation region; Bayes factor; Change-point analysis; Geometric distribution; Maximum likelihood; Mixture parameter; Prior distribution

1. Introduction and summary

In this paper we investigate the Bayesian aspects of change-point analysis. The Bayesian detection of the change moment was started by Chernoff and Zacks (1964) for a sequence of normal random variables with possibly varying means, and this point of view was pursued further by Broemling (1974). Smith (1975) extended these results to multiparameter models; Cobb (1978) demonstrated that for given ancillary statistics the conditional distribution of the maximum likelihood procedure is approximately the same as the posterior distribution for the change-point parameter against the uniform prior. Ibragimov and Hasminskii (1981) studied the continuous version of the problem (introduced earlier by Shiryaev (1978)) and obtained the exact formulas for the mean squared errors of the Bayes estimator of the change-point for the uniform prior (Pitman estimator) and of the maximum likelihood estimator.

* Tel.: +1-410-455-2408; fax: +1-410-455-1066.
E-mail address: rukhin@math.umbc.edu (A.L. Rukhin).

0378-3758/01/$ - see front matter
PII: S0378-3758(01)00266-X

The articles by Zacks (1983) and by Bhattacharya (1994) (especially Section 3) give a survey of the main results in the estimation of the change-point, and the monographs by Brodsky and Darkhovsky (1993) and by Csorgo and Horvath (1998) summarize the facts obtained in the area of the nonparametric change-point analysis.

The limiting behavior of the posterior distribution of an unknown parameter in the classical situation of a local asymptotically normal experiment is described by the well known Laplace-Bernstein-von Mises Theorem according to which the limiting law is normal. This result is important both to frequentist and Bayesian inference as it leads to approximate confidence (or credible) intervals for the unknown parameter and for asymptotically normal point estimators of this parameter.

In this paper we study the asymptotic behavior of the posterior distribution of the change-point parameter. This problem turns out to be quite different from the classical one. The new limiting law in this situation is shown to be a non-normal distribution expressed through discounted sums financial mathematics and related to the infinite product of random matrices (or affine transforms of the real line). The asymptotic distribution of Bayes estimators in the general framework, including an invariant loss function and some prior distribution approaching a geometric law, is also given when the sample size increases, but the pre- and after-change distributions are fixed.

In Sections 3 and 4 the situation when these distributions are partially unknown, is studied. We explore the asymptotic behavior of the posterior distribution of the change-point and obtain conditions for the weak convergence of the posterior distribution to the same distribution as before. The possibility of such adaptive behavior is also investigated for the procedure based on the estimated values of the pre- (post-) change parameters by only several first (last) observations.

The obtained results are illustrated by the example of normal distributions with both unknown parameters, which is discussed in Section 5.

2. Asymptotic behavior of the posterior distribution and of the Bayes estimator

The change-point parameter, v, is defined when the series of observations has the form

$$\mathbf{x} = (x_1, \ldots, x_v, x_{v+1}, \ldots, x_n),$$

where the random sample $(x_1, \ldots, x_v)$ comes from distribution F, and the second subsample $(x_{v+1}, \ldots, x_n)$ is drawn from distribution G. Here F and G are two different probability distributions with densities f and g. Throughout the paper we assume that the information numbers $K(F,G)$ and $K(G,F)$ are positive and finite.

Denote by $p_v(\mathbf{x}) = \prod_1^v f(x_j) \prod_{v+1}^n g(x_j)$ the corresponding densities, and assume that the prior distribution of the parameter v is given by positive prior probabilities $\lambda_j^{(n)}$ supported by the interval $c_0(n) < j \leqslant c_1(n)$. In other terms we assume firm prior information about the central region, $(c_0(n), c_1(n)]$, where the change-point occurs. Thus, if densities f and g are known, the first $c_0(n) + 1$ and the last $n - c_1(n)$

observations are ancillary statistics since they have given distributions F and G, respectively. Note that the possibilities, $c_0(n)=0$ and $c_1(n)=n$, are not excluded here.

Let $\mathbf{m}$ be the marginal density of $\mathbf{x}$;

$$\mathbf{m}(\mathbf{x})=\sum_{j=c_0(n)+1}^{c_1(n)} \lambda_j^{(n)} p_j(\mathbf{x}).$$

We are interested in the asymptotic posterior distribution of the random change-point parameter N whose distribution is given by probabilities $\lambda_j^{(n)}$, $j=c_0(n)+1,\dots,c_1(n)$. Then the posterior probabilities for the change-point parameter have the form

$$P(N=m|\mathbf{x})=\frac{p_m(\mathbf{x})\lambda_m^{(n)}}{\mathbf{m}(\mathbf{x})}. \tag{1}$$

When v, $c_0(n)+1\leqslant v\leqslant c_1(n)$, is the true value of the change-point parameter, one has

$$\mathbf{m}(\mathbf{x})=p_v(\mathbf{x})\lambda_v^{(n)}\left[1+\sum_{c_0(n)<j<v}\frac{g}{f}(x_{j+1})\cdots\frac{g}{f}(x_v)\frac{\lambda_j^{(n)}}{\lambda_v^{(n)}}\right.$$
$$\left.+\sum_{v<j\leqslant c_1(n)}\frac{f}{g}(x_{v+1})\cdots\frac{f}{g}(x_j)\frac{\lambda_j^{(n)}}{\lambda_v^{(n)}}\right].$$

Therefore

$$P(N=m|\mathbf{x})=\frac{p_m(\mathbf{x})\lambda_m^{(n)}}{p_v(\mathbf{x})\lambda_v^{(n)}}\left[1+\sum_{k=1}^{v-c_0(n)}\tilde{\xi}_1\cdots\tilde{\xi}_k\frac{\lambda_{v-k}^{(n)}}{\lambda_v^{(n)}}+\sum_{k=1}^{c_1(n)-v}\xi_1\cdots\xi_k\frac{\lambda_{v+k}^{(n)}}{\lambda_v^{(n)}}\right]^{-1}. \tag{2}$$

Here and further $\tilde{\xi}_k$ and ξ_k, $k=1,2,\dots$, are mutually independent positive random variables such that the random variables $\tilde{\xi}_k$ are distributed as the likelihood ratio g/f under F, and the distribution of ξ_j is that of f/g under G. Clearly, for all k, $E\log\tilde{\xi}_k=-K(F,G)$, and $E\log\xi_k=-K(G,F)$.

We assume that there exists a positive sequence q_ℓ, $\ell=0,\pm1,\pm2,\dots$ such that as $n\to\infty$, $v-c_0(n)\to\infty$, $c_1(n)-v\to\infty$;

$$\max_{1\leqslant k\leqslant c_1(n)-v}\left|\frac{\lambda_{v+k}^{(n)}}{q_k\lambda_v^{(n)}}-1\right|\to 0 \tag{3}$$

and

$$\max_{1\leqslant k<v-c_0(n)}\left|\frac{\lambda_{v-k}^{(n)}}{q_{-k}\lambda_v^{(n)}}-1\right|\to 0. \tag{4}$$

Conditions (3) and (4) hold if for any fixed k and ℓ

$$\frac{\lambda_k^{(n)}}{\lambda_\ell^{(n)}}\to q_{k-\ell}$$

and this convergence is uniform when $|k-\ell| \leqslant c_1(n) - c_0(n)$. Then for any fixed k, $\lambda_{v+k}^{(n)}/\lambda_v^{(n)} \to q_k$, so that $q_k q_\ell = q_{k+\ell}$, and $q_k = q^k$ for some positive $q(=q_1)$.

We also assume that with $\mu = \log q$

$$-K(F,G) < \mu < K(G,F). \tag{5}$$

Then under density p_v, when $n \to \infty$, the distributions of both sums in (2) converge. Indeed, because of (5), the random series $\sum_j q^j \xi_1 \cdots \xi_j$ converges with probability one, if $\mu < K(G,F)$, and the series $\sum_j q^{-j} \tilde{\xi}_1 \cdots \tilde{\xi}_j$ converges with probability one, if $\mu > -K(F,G)$. (The convergence of the series above follows from Lemma 1.7 in Vervaat, 1979.)

Also

$$\left| \sum_{k=1}^{c_1(n)-v} \xi_1 \cdots \xi_k \frac{\lambda_{v+k}^{(n)}}{\lambda_v^{(n)}} - \sum_{k=1}^{\infty} \xi_1 \cdots \xi_k q^k \right|$$

$$\leqslant \sum_{k=1}^{c_1(n)-v} q^k \xi_1 \cdots \xi_k \left| \frac{\lambda_{v+k}^{(n)}}{q^k \lambda_v^{(n)}} - 1 \right| + \sum_{k=c_1(n)-v+1}^{\infty} q^k \xi_1 \cdots \xi_k .$$

By (3) and because of the mentioned series convergence both terms here tend in distribution to 0. The same conclusion holds for the first sum in the right-hand side of (2).

If $v - m \to m_0$,

$$\frac{p_m(\mathbf{x})\lambda_m^{(n)}}{p_v(\mathbf{x})\lambda_v^{(n)}} \to \begin{cases} q^{|m_0|}\xi_1 \cdots \xi_{|m_0|} & m_0 \leqslant 0 \\ q^{-m_0}\tilde{\xi}_1 \cdots \tilde{\xi}_{m_0} & m_0 > 0. \end{cases}$$

Thus, under conditions (3) and (4) the distribution of the posterior probability, $P(N = m|\mathbf{x})$, converges. When v is the true value of the change-point parameter, such that $v - m \to m_0$, then

$$P(N = m|\mathbf{x}) \to \begin{cases} q^{|m_0|} \dfrac{\xi_1 \cdots \xi_{|m_0|}}{1 + \sum_{k=1}^{\infty} \tilde{\xi}_1 \cdots \tilde{\xi}_k q^{-k} + \sum_{k=1}^{\infty} \xi_1 \cdots \xi_k q^k} & m_0 \leqslant 0 \\[2ex] q^{-m_0} \dfrac{\tilde{\xi}_1 \cdots \tilde{\xi}_{m_0}}{1 + \sum_{k=1}^{\infty} \tilde{\xi}_1 \cdots \tilde{\xi}_k q^{-k} + \sum_{k=1}^{\infty} \xi_1 \cdots \xi_k q^k} & m_0 > 0. \end{cases} \tag{6}$$

Since, for example, the sum of the form $\sum_{v<j\leqslant n} \xi_{v+1} \cdots \xi_j \lambda_j^{(n)}/\lambda_v^{(n)}$ does not converge in probability, only the statement about the convergence in distribution can be made.

The dependence of the asymptotic distribution on underlying probability measures F and G and its fairly complicated structure are in contrast with the asymptotic normality of the posterior distribution in the classical parametric setting of local asymptotic normality. Of course, in the latter case the convergence takes place with probability one (Le Cam (1986), Chapter 17, Section 7).

Similar results can be derived for the Bayes estimators of the change-point parameter. Consider a non-negative loss function, $L(\delta, v) = L(\delta - v)$, such that $L(0) = 0$. When f

and g are given, the estimate δ of v, $c_0(n)+1 \leqslant v \leqslant c_1(n)$, in our situation will be based only on $x_{c_0(n)+1}, \ldots, x_{c_1(n)}$. Since most commonly used loss functions are defined for all real values of $d = \delta - v$, it is convenient to allow δ to take all values in the interval $(c_0(n), c_1(n)]$ and to assume that $L(d)$ is a continuous function of d, $-\infty < d < \infty$.

Let $\mathscr{E}$ denote the expectation obtained from the conditional distribution of $\mathbf{x}$ for fixed v and from the prior distribution of v. The Bayes estimator δ_n is defined by the formula

$$\begin{aligned}\delta_n(\mathbf{x}) &= \arg \min_{d:\ c_0(n) < d \leqslant c_1(n)} \mathscr{E}\{L(d-v)|\mathbf{x}\} \\ &= \arg \min_{d:\ c_0(n) < d \leqslant c_1(n)} \sum_v L(d-v)\lambda_v^{(n)} p_v(\mathbf{x}).\end{aligned} \tag{7}$$

For the sake of concreteness one accepts the usual change-point analysis convention that, when the minimizer in (7) is not defined uniquely, the smallest value of d is taken. Then δ_n is uniquely defined with probability 1 (which is always true if L is a strictly convex function).

The argument as above shows that the distribution of $\delta_n - v$ coincides with that of

$$\begin{aligned}\arg \min_{d:\ c_0(n)-v < d \leqslant c_1(n)-v} \Bigg[& L(d) + \sum_{k=1}^{v-c_0(n)-1} L(d+k)\tilde{\xi}_1 \cdots \tilde{\xi}_k \frac{\lambda_{v-k}^{(n)}}{\lambda_v^{(n)}} \\ & + \sum_{k=1}^{c_1(n)-v} L(d-k)\xi_1 \cdots \xi_k \frac{\lambda_{v+k}^{(n)}}{\lambda_v^{(n)}} \Bigg].\end{aligned} \tag{8}$$

Under our convention the expression in brackets in (8) is almost surely a continuous function of d.

Assume that for a positive polynomial $P(d)$ and some positive α_1 and α_2

$$\begin{aligned}&L(d) \leqslant P(d)\mathrm{e}^{\alpha_1 d} \quad d > 0, \\ &L(d) \leqslant P(d)\mathrm{e}^{\alpha_2 |d|} \quad d < 0\end{aligned} \tag{9}$$

and

$$\alpha_1 - K(F,G) < \mu < K(G,F) - \alpha_2. \tag{10}$$

Then the process

$$Y_{vn}(d) = L(d) + \sum_{k=1}^{v-c_0(n)-1} L(k+d)\tilde{\xi}_1 \cdots \tilde{\xi}_k \frac{\lambda_{v-k}^{(n)}}{\lambda_v^{(n)}} + \sum_{k=1}^{c_1(n)-v} L(d-k)\xi_1 \cdots \xi_k \frac{\lambda_{v+k}^{(n)}}{\lambda_v^{(n)}}$$

converges weakly as $n \to \infty$ to

$$Y(d) = L(d) + \sum_{j=1}^{\infty} L(d+j)\tilde{\xi}_1 \cdots \tilde{\xi}_j q^{-j} + \sum_{j=1}^{\infty} L(d-j)\xi_1 \cdots \xi_j q^j. \tag{11}$$

Notice that the condition (10) guarantees the almost sure convergence of the series (11). If (10) does not hold, then $P_v(|\delta_n - v| > z) \to 1$ for any positive z.

Under additional regularity conditions preventing $Y(d)$ from attaining its minimum at infinity, $\delta_n(\mathbf{x}) - v$ converges in distribution to

$$V = \arg\min_d Y(d), \tag{12}$$

provided that V is uniquely defined by (12) with probability one.

We formulate these results.

Theorem 1. *Under conditions* (3), (4), (5) *as* $n \to \infty$, $v - c_0(n) \to \infty$, $c_1(n) - v \to \infty$ *and* $v - m \to m_0$ *the distribution of the posterior probability tends to the distribution in* (6). *Let* δ_n *be the Bayes estimator* (7) *of the change-point parameter for the loss function* $L(d - v)$ *satisfying* (9). *Under the conditions above with* (5) *replaced by* (10), *the limiting distribution of the sequence* $\delta_n - v$ *is given by* (12), *provided that* V *is uniquely defined by* (12) *almost surely.*

According to Theorem 1, $\delta_n - v = \mathrm{O}_P(1)$. This rate for $n \to \infty$ is known to hold in a more general non-parametric setting, see Duembgen (1991), Yao et al. (1994). Theorem 1 shows that this is indeed the best possible rate. The analogue of Theorem 1 for the Bayes estimators when $c_0(n) = 0$, $c_1(n) = n$ is discussed by Rukhin (1997).

The limiting distribution of $\delta_n - v$ for the zero-one loss and the uniform prior distribution has been derived by Hinkley (1970). Notice that for this loss function the processes $Y_{vn}(d)$ and $Y(d)$ do not have continuous trajectories. Other settings of the asymptotic estimation of the change-point with converging pre-change and after-change distributions have been extensively studied (see Ritov, 1990; Csorgo and Horvath, 1998).

In Section 3 we restrict our attention to the (truncated) geometric prior distributions with probabilities proportional to q^j, $j = c_0(n) + 1, \ldots, c_1(n)$ for some positive q. These prior distributions are important because of Theorem 1. Values of q larger than 1 are of interest in problems where the change-point is more likely to happen at the end of the observation period, as in many quality control type situations. Although conditions (3) and (4) are rather restrictive and the prior geometric distributions are not always adequate, without these conditions, for example, under Poisson or binomial prior distributions, the limiting law for the posterior distribution may not exist at all, or it may depend on the limiting value of v/n.

A helpful way of looking at the marginal density $m(\mathbf{x})$ is by introducing the sequence of 2×2 independent random matrices of the form

$$Y_k = \begin{pmatrix} \xi_k & \lambda_{k-1} \\ 0 & 1 \end{pmatrix}$$

which can be interpreted as affine transformations of the real line $z : z \to \xi_k z + \lambda_{k-1}$. One has

$$Y_1 Y_2 \cdots Y_n = \begin{pmatrix} \xi_1 \cdots \xi_n & \sum_{k=0}^{n} \xi_1 \cdots \xi_k \lambda_k \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \xi_1 \cdots \xi_n & m(\mathbf{x}) \big/ \prod g(x_j) \\ 0 & 1 \end{pmatrix}.$$

Thus under assumptions of Theorem 1 one can derive the limiting behavior of the ratio $m(\mathbf{x})/\prod g(x_j)$ as a corollary of a limit theorem for products of random matrices. An analysis of the convergence of the series as above can be found in Grenander (1964) and Vervaat (1979). As a matter of fact, the mentioned result concerning the divergence of $\delta_n - v$, when (5) is violated, can be deduced from one of these limit theorems (see for example Grincevicius, 1975).

Let

$$\Pi = \Pi(q) = \sum_{k=1}^{\infty} \xi_1 \cdots \xi_k q^k, \tag{13}$$

where $\xi_j,\ j = 1, 2, \ldots$ are i.i.d. positive random variables. Under condition $E \log \xi_i < -\log q$, the series defining Π converges with probability one.

The stochastically discounted sums (13) (the so-called *perpetuities*), which determine the limiting distribution of posterior distribution and of Bayes rules in the change-point problem, also appear in problems of life insurance and finance (Dufresne, 1990). They are important in the study of sequential Shiryaev-type procedures to detect the change-point, see Shiryaev (1978), Pollak (1985).

3. Asymptotic behavior of the posterior distribution with the unknown parameters: finite nuisance parameter case

In this section a situation, when distributions F and G are known only up to a nuisance parameter, is considered. To start, we look at the asymptotic behavior of the posterior distribution of the change-point parameter when the pre-change distribution belongs to a finite family $\{F_\beta,\ \beta = 1, \ldots, B\}$, and the after-change distribution is a member of the family

$\{G_\gamma,\ \gamma = 1, \ldots, \Gamma\}$. The estimation problem of the parameter γ was described by Nikiforov (1995) as the *isolation* problem for the change-point parameter.

We assume that the nuisance parameters β and γ are independent and have positive prior probabilities π_β and τ_γ. Let

$$\hat{p}_v(\mathbf{x}) = \left[\sum_\beta \pi_\beta \prod_1^v f_\beta(x_j)\right] \left[\sum_\gamma \tau_\gamma \prod_{v+1}^n g_\gamma(x_j)\right]$$

be the conditional likelihood function of $\mathbf{x}$ for given v. Then for geometric prior probabilities

$$\hat{\mathbf{m}}(\mathbf{x}) \propto \sum_k q^k \hat{p}_k(\mathbf{x})$$

is the marginal density of $\mathbf{x}$, and the posterior distribution is defined by the formula

$$P(N = m|\mathbf{x}) = \frac{q^m \hat{p}_m(\mathbf{x})}{\hat{\mathbf{m}}(\mathbf{x})}. \tag{14}$$

Our goal is to find the limiting distribution of (14) for a fixed pair (β_0, γ_0) of true nuisance parameter values. In particular it is of interest to determine conditions under which it coincides with that of (6) with distributions of ξ and $\tilde{\xi}$ evaluated under these parametric values. We will assume that for all γ and all β;

$$E^F_{\beta_0}\left[\log\frac{g_\gamma}{g_{\gamma_0}}(X)\right]^2<\infty,\quad E^F_{\beta_0}\left[\log\frac{f_\beta}{f_{\beta_0}}(X)\right]^2<\infty \tag{15}$$

$$E^G_{\gamma_0}\left[\log\frac{g_\gamma}{g_{\gamma_0}}(X)\right]^2<\infty,\quad E^G_{\gamma_0}\left[\log\frac{f_\beta}{f_{\beta_0}}(X)\right]^2<\infty. \tag{16}$$

Also assume that

$$\frac{\sqrt{c_1(n)-c_0(n)}}{\min[c_1(n),n-c_0(n)]}\to 0 \tag{17}$$

and

$$\min[c_0(n),n-c_1(n)]\to\infty. \tag{18}$$

Let v be the true value of the change-point parameter, so that

$$p_v(\mathbf{x})=\prod_1^v f_{\beta_0}(x_j)\prod_{v+1}^n g_{\gamma_0}(x_j).$$

Then $\hat{\mathbf{m}}(\mathbf{x})/[q^v\hat{p}_v(\mathbf{x})]$ has the distribution of

$$1+\sum_1^{v-c_0(n)-1}\frac{\hat{p}_{v-k}(\mathbf{x})}{\hat{p}_v(\mathbf{x})}q^{-k}+\sum_1^{c_1(n)-v}\frac{\hat{p}_{v+k}(\mathbf{x})}{\hat{p}_v(\mathbf{x})}q^k. \tag{19}$$

To derive the asymptotic distribution of the first sum in (19) we study the behavior of

$$\frac{\hat{p}_{v-k}(\mathbf{x})}{\hat{p}_v(\mathbf{x})}=\frac{\hat{p}_{v-k}(\mathbf{x})}{p_{v-k}(\mathbf{x})}\frac{p_v(\mathbf{x})}{\hat{p}_v(\mathbf{x})}\frac{p_{v-k}(\mathbf{x})}{p_v(\mathbf{x})}=\frac{\hat{p}_{v-k}(\mathbf{x})}{p_{v-k}(\mathbf{x})}\frac{p_v(\mathbf{x})}{\hat{p}_v(\mathbf{x})}\prod_{v-k+1}^v\frac{g_{\gamma_0}(x_j)}{f_{\beta_0}(x_j)} \tag{20}$$

for $1\leqslant k<v-c_0(n)$. The distribution of the last factor in the right-hand side of (20) is that of the product $\tilde{\xi}_1\cdots\tilde{\xi}_k$, with the distribution of $\tilde{\xi}_j$ now being that of the likelihood ratio g_{γ_0}/f_{β_0} under F_{β_0}. Thus we need to determine the asymptotic behavior of two Bayes factors $p_{v-k}(\mathbf{x})/\hat{p}_{v-k}(\mathbf{x})$ and $p_v(\mathbf{x})/\hat{p}_v(\mathbf{x})$ for H_0: $\beta=\beta_0,\ \gamma=\gamma_0$.

One has

$$\begin{aligned}\frac{\hat{p}_v(\mathbf{x})}{p_v(\mathbf{x})}&=\sum_\beta\pi_\beta\prod_1^v\frac{f_\beta}{f_{\beta_0}}(x_j)\sum_\gamma\tau_\gamma\prod_{v+1}^n\frac{g_\gamma}{g_{\gamma_0}}(x_j)\\&=\sum_\beta\pi_\beta\exp\left\{\sum_1^v\log\frac{f_\beta}{f_{\beta_0}}(x_j)\right\}\sum_\gamma\tau_\gamma\exp\left\{\sum_{v+1}^n\log\frac{g_\gamma}{g_{\gamma_0}}(x_j)\right\}.\end{aligned}$$

By the law of large numbers for $\beta\neq\beta_0$;

$$\sum_1^v\log\frac{f_\beta}{f_{\beta_0}}(x_j)\to-\infty$$

as $v \to \infty$, and for $\gamma \neq \gamma_0$ as $n - v \to \infty$;

$$\sum_{v+1}^{n} \log \frac{g_\gamma}{g_{\gamma_0}}(x_j) \to -\infty.$$

Therefore, under the density $p_v(\mathbf{x})$ as $n \to \infty$;

$$\hat{p}_v(\mathbf{x})/p_v(\mathbf{x}) \to \pi_{\beta_0}\tau_{\gamma_0}. \tag{21}$$

For $v - k \to \infty$,

$$\begin{aligned}\frac{\hat{p}_{v-k}(\mathbf{x})}{p_{v-k}(\mathbf{x})} &= \sum_\beta \pi_\beta \exp\left\{\sum_{1}^{v-k} \log \frac{f_\beta}{f_{\beta_0}}(x_j)\right\} \sum_\gamma \tau_\gamma \exp\left\{\sum_{v-k+1}^{n} \log \frac{g_\gamma}{g_{\gamma_0}}(x_j)\right\} \\ &= [\pi_{\beta_0} + \mathrm{o}_P(1)] \sum_\gamma \tau_\gamma \exp\left\{\sum_{v-k+1}^{v} \log \frac{g_\gamma}{g_{\gamma_0}}(x_j) + \sum_{v+1}^{n} \log \frac{g_\gamma}{g_{\gamma_0}}(x_j)\right\}.\end{aligned}$$

Since, for $v < j$, observations x_j have density g_{γ_0}, the Central Limit Theorem shows that for any $\gamma \neq \gamma_0$ because of (16)

$$\sum_{v+1}^{n} \log \frac{g_\gamma}{g_{\gamma_0}}(x_j) = -(n-v)K(G_{\gamma_0}, G_\gamma) + \mathrm{O}_P(\sqrt{n-v}).$$

When $v - k + 1 \leqslant j \leqslant v$, observations x_j have density f_{β_0}, so that for $k \to \infty$;

$$\sum_{v-k+1}^{v} \log \frac{g_\gamma}{g_{\gamma_0}}(x_j) = kE_{\beta_0}^F \log \frac{g_\gamma}{g_{\gamma_0}}(X) + \Delta_k^\gamma,$$

where $\max_{1 \leqslant k \leqslant v - c_0(n)} k^{-1/2}\Delta_k^\gamma = \mathrm{O}_P(1)$, as the maximum of normalized partial sums converges in distribution by (15). It follows that

$$\max\nolimits_{1 \leqslant k < v - c_0(n)} \Delta_k^\gamma = \mathrm{O}_P(\sqrt{v - c_0(n)}).$$

Thus for $1 \leqslant k < v - c_0(n)$;

$$\begin{aligned}\sum_{v-k+1}^{v} \log \frac{g_\gamma}{g_{\gamma_0}}(x_j) + \sum_{v+1}^{n} \log \frac{g_\gamma}{g_{\gamma_0}}(x_j) &= kE_{\beta_0}^F \log \frac{g_\gamma}{g_{\gamma_0}}(X) - (n-v)K(G_{\gamma_0}, G_\gamma) \\ &\quad + \mathrm{O}_P(\sqrt{v - c_0(n)}) + \mathrm{O}_P(\sqrt{n-v}).\end{aligned}$$

Because of (21), one has the following representation for the distribution of $\hat{p}_{v-k}(\mathbf{x}) p_v(\mathbf{x})/[p_{v-k}(\mathbf{x})\hat{p}_v(\mathbf{x})]$;

$$\left[1 + \sum_{\gamma \neq \gamma_0} \frac{\tau_\gamma}{\tau_{\gamma_0}} \exp\left\{kE_{\beta_0}^F \log \frac{g_\gamma}{g_{\gamma_0}}(X)\right\} \exp\{-(n-v)K(G_{\gamma_0}, G_\gamma) + \mathrm{O}_P(\sqrt{v - c_0(n)}) + \mathrm{O}_P(\sqrt{n-v})\}\right][1 + \mathrm{o}_P(1)].$$

It follows that the distribution of the first sum in (19) has the limiting law $\sum_1^\infty q^{-k} \tilde{\xi}_1 \cdots \tilde{\xi}_k$, provided that for any $\gamma \neq \gamma_0$

$$\sum_{1 \leqslant k < v - c_0(n)} q^{-k} \tilde{\xi}_1 \cdots \tilde{\xi}_k \exp\left\{kE_{\beta_0}^F \log \frac{g_\gamma}{g_{\gamma_0}}(X)\right\} < \infty$$

and

$$\exp\{-(n-v)K(G_{\gamma_0},G_\gamma)+\mathrm{O}_P(\sqrt{v-c_0(n)})+\mathrm{O}_P(\sqrt{n-v})\}\to 0.$$

The former condition holds if

$$-\mu-K(F_{\beta_0},G_{\gamma_0})+E^F_{\beta_0}\log\frac{g_\gamma}{g_{\gamma_0}}(X)=-\mu-K(F_{\beta_0},G_\gamma)<0, \tag{22}$$

and the second condition follows from (17) as

$$\frac{\sqrt{v-c_0(n)}}{n-v}\leqslant\frac{\sqrt{c_1(n)-c_0(n)}}{n-c_1(n)}\to 0.$$

Thus under condition (22), the desired convergence in distribution of the first sum in (19) takes place.

We consider now the second sum in (19) which corresponds to k, $1\leqslant k\leqslant c_1(n)-v$. To determine the limiting behavior of

$$\frac{\hat{p}_{v+k}(\mathbf{x})}{\hat{p}_v(\mathbf{x})}=\frac{\hat{p}_{v+k}(\mathbf{x})}{p_{v+k}(\mathbf{x})}\frac{p_v(\mathbf{x})}{\hat{p}_v(\mathbf{x})}\prod_{v+1}^{v+k}\frac{f_{\beta_0}}{g_{\gamma_0}}(x_j)$$

it suffices to look only at

$$\begin{aligned}\frac{\hat{p}_{v+k}(\mathbf{x})}{p_{v+k}(\mathbf{x})}&=\sum_\beta\pi_\beta\prod_1^{v+k}\frac{f_\beta}{f_{\beta_0}}(x_j)\sum_\gamma\tau_\gamma\prod_{v+k+1}^{n}\frac{g_\gamma}{g_{\gamma_0}}(x_j)\\&=\sum_{\beta,\gamma}\pi_\beta\tau_\gamma\exp\left\{\sum_1^{v+k}\log\frac{f_\beta}{f_{\beta_0}}(x_j)+\sum_{v+k+1}^{n}\log\frac{g_\gamma}{g_{\gamma_0}}(x_j)\right\}.\end{aligned}$$

As before, one obtains as according to (18) $n-v-k\geqslant n-c_1(n)\to\infty$,

$$\max_{\gamma\neq\gamma_0}\sum_{v+k+1}^{n}\log\frac{g_\gamma}{g_{\gamma_0}}(x_j)\to-\infty$$

and for a fixed $\beta\neq\beta_0$ as $n\to\infty$;

$$\begin{aligned}\sum_1^{v+k}\log\frac{f_\beta}{f_{\beta_0}}(x_j)&=-vK(F_{\beta_0},F_\beta)+kE^G_{\gamma_0}\log\frac{f_\beta}{f_{\beta_0}}(X)\\&\quad+\mathrm{O}_P(\sqrt{v})+\mathrm{O}_P(\sqrt{c_1(n)-v}).\end{aligned}$$

Because of (17)

$$\frac{\sqrt{c_1(n)-v}}{v}\leqslant\frac{\sqrt{c_1(n)-c_0(n)}}{c_0(n)}\to 0$$

and the limiting distribution of $\sum_{1\leqslant k\leqslant c_1(n)-v}q^k(\hat{p}_{v+k})/\hat{p}_v(\mathbf{x})$ is that of $\sum_1^\infty q^k\xi_1\cdots\xi_k$ provided that for any $\beta\neq\beta_0$

$$\mu-K(G_{\gamma_0},F_{\beta_0})+E^G_{\gamma_0}\log\frac{f_\beta}{f_{\beta_0}}(X)=\mu-K(G_{\gamma_0},F_\beta)<0. \tag{23}$$

As in Section 2, after dividing the numerator and the denominator of (14) by $q^v \hat{p}_v(\mathbf{x})$ one obtains when, say, $v - m \to m_0 > 0$

$$P(N = m|\mathbf{x}) = \frac{q^{m-v}\hat{p}_m(\mathbf{x})/\hat{p}_v(\mathbf{x})}{1 + \sum_{c_0(n)<j<v} \hat{p}_j(\mathbf{x})/\hat{p}_v(\mathbf{x})q^{j-v} + \sum_{v<j\leqslant c_1(n)} \hat{p}_j(\mathbf{x})/\hat{p}_v(\mathbf{x})q^{j-v}}$$

$$\to q^{-m_0}\tilde{\xi}_1 \cdots \tilde{\xi}_{m_0}\left[\sum_1^\infty q^{-k}\tilde{\xi}_1 \cdots \tilde{\xi}_k + \sum_1^\infty q^k \xi_1 \cdots \xi_k\right]^{-1} \tag{24}$$

provided that the inequalities (22) and (23) hold for all values β, γ of the pre- and post-change parameters, i.e. if

$$-\min_{\beta,\gamma} K(F_\beta, G_\gamma) < \mu < \min_{\beta,\gamma} K(G_\gamma, F_\beta). \tag{25}$$

Thus, we have established the following result.

Theorem 2. *The posterior probabilities* (14) *against geometric prior distribution under conditions* (15), (16), (17) *and* (25) *converge in distribution to* (24).

An analogue of Theorem 2 is also true for the profile likelihood function

$$\bar{p}_v(\mathbf{x}) = \max_\beta \prod_1^v f_\beta(x_j) \max_\gamma \prod_{v+1}^n g_\gamma(x_j).$$

Under the natural analogue of the condition (10) which takes into account the constants α_1, α_2, the process Y_{vn} in (19) weakly converges to the process $Y(d)$ defined in (11). Also then $\delta_n - v$ converges in distribution to (12). The same is true for the Bayes estimator based on the maximum likelihood elimination of the nuisance parameters, which is a traditional method.

4. Asymptotic behavior of posterior distribution with the unknown parameters: continuous nuisance parameters

It is of some interest to investigate richer distributions families. Assume a regular parametric family $\{f(x,\beta),\ \beta \in \mathbf{B}\}$ with $\mathbf{B}$ being an open subset of $\mathbf{R}^p$ as the model for possible pre-change distributions, and let the family $\{g(x,\gamma),\ \gamma \in \boldsymbol{\Gamma},\ \boldsymbol{\Gamma} \subset \mathbf{R}^q\}$ describe all possible after-change distributions. Here regularity is understood in terms of Section 1.7 of Ibragimov and Hasminskii (1981) or Section 2.2 of Roussas (1972). More general conditions can be found in Section 16.3 of Le Cam (1986).

Assuming that β_0, γ_0 are true values of the nuisance parameters, let $I_1 = I_1(\beta_0) = E^F_{\beta_0} \times \phi_0(X)\phi_0^T(X) = -\{E^F_{\beta_0}\partial^2/(\partial\beta_i\partial\beta_k)\log f(X,\beta_0)\}$ denote the Fisher information matrix for the family $\{f(.,\beta)\}$. Here $\phi_0(x) = (\partial/\partial\beta)\log f(x,\beta_0)$ is the gradient of $\log f$. Similarly with $\psi_0(x) = (\partial/\partial\gamma)\log g(X,\gamma_0)$ $I_2 = I_2(\gamma_0) = E^G_{\gamma_0}\psi_0(X)\psi_0^T(X) = -\{E^G_{\gamma_0}\partial^2/(\partial\gamma_i\partial\gamma_k) \times \log g(X,\gamma_0)\}$ is the Fisher information matrix for the second family. Also we denote $\Psi^F_k = I_2^{-1/2}\sum_1^k \psi_0(X_j)$, $\Phi^F_k = I_1^{-1/2}\sum_1^k \phi_0(X_j)$ with independent random variables X_j having distribution F_{β_0}, and $\Psi^G_k = I_2^{-1/2}\sum_1^k \psi_0(X_j)$, $\Phi^G_k = I_1^{-1/2}\sum_1^k \phi_0(X_j)$, have a similar meaning. As $k \to \infty$ both $\Phi_k^F/\sqrt{k}$ and $\Psi_k^G/\sqrt{k}$ have asymptotic standard normal distributions.

We establish now a result similar to the one used in Theorem 2 for the posterior distribution (14) when with some positive prior densities $\pi(\beta)$ and $\tau(\gamma)$;

$$\hat{p}_v(\mathbf{x}) = \int_{\mathbf{B}} \prod_1^v f_\beta(x_j)\pi(\beta)\,\mathrm{d}\beta \int_{\boldsymbol{\Gamma}} \prod_{v+1}^n g_\gamma(x_j)\tau(\gamma)\,\mathrm{d}\gamma.$$

The first proposition deals with the asymptotic behavior of the distribution of Bayes factors $p_{v-k}(\mathbf{x})/\hat{p}_{v-k}(\mathbf{x})$ and $p_v(\mathbf{x})/\hat{p}_v(\mathbf{x})$ for the null hypothesis $\beta=\beta_0,\ \gamma=\gamma_0$.

Proposition 1. *Assume that*

$$\frac{c_1(n)-c_0(n)}{\min[c_1(n), n-c_0(n)]} \to 0. \tag{26}$$

Then for a fixed k, $k < v - c_0(n)$, one has the following representation of distributions $\hat{p}_{v-k}(\mathbf{x})p_v(\mathbf{x})/[p_{v-k}(\mathbf{x})\hat{p}_v(\mathbf{x})]$ and $\hat{p}_{v+k}(\mathbf{x})p_v(\mathbf{x})/[p_{v-k}(\mathbf{x})\hat{p}_v(\mathbf{x})]$. With independent p-dimensional standard normal vector Z_1 and q-dimensional standard normal vector Z_2,

$$\begin{aligned}
&\log \frac{\hat{p}_{v-k}(\mathbf{x})}{p_{v-k}(\mathbf{x})}\frac{p_v(\mathbf{x})}{\hat{p}_v(\mathbf{x})} \\
&\stackrel{d}{=} \frac{1}{2(n-v+k)}\langle \Psi_k^F, \Psi_k^F\rangle + \frac{\sqrt{v-k}}{n-v}\langle Z_2, \Psi_k^F\rangle - \frac{k}{2(n-v+k)}\langle Z_2, Z_2\rangle \\
&\quad + \frac{k}{2v}\langle Z_1, Z_1\rangle - \frac{\sqrt{v-k}}{v}\langle Z_1, \Phi_k^F\rangle - \frac{1}{2v}\langle \Phi_k^F, \Phi_k^F\rangle \\
&\quad + \mathrm{O}_P\left(\frac{\sqrt{k}}{n-v+k}\right) + \mathrm{O}_P\left(\frac{\sqrt{k}}{v}\right) + \mathrm{O}_P\left(\frac{1}{\sqrt{n-v}}\right) + \mathrm{O}_P\left(\frac{1}{\sqrt{v-k}}\right).
\end{aligned}$$

Similarly for a fixed k, $k < c_1(n) - v$

$$\begin{aligned}
&\log \frac{\hat{p}_{v+k}(\mathbf{x})}{p_{v+k}(\mathbf{x})}\frac{p_v(\mathbf{x})}{\hat{p}_v(\mathbf{x})} \\
&\stackrel{d}{=} \frac{1}{2(n-v+k)}\langle \Phi_k^G, \Phi_k^G\rangle + \frac{\sqrt{v-k}}{n-v+k}\langle Z_1', \Phi_k^G\rangle + \frac{k}{2(n-v+k)}\langle Z_1', Z_1'\rangle \\
&\quad - \frac{k}{2(v+k)}\langle Z_2', Z_2'\rangle - \frac{\sqrt{v}}{v+k}\langle Z_2', \Psi_k^G\rangle - \frac{1}{2(n-v+k)}\langle \Psi_k^G, \Psi_k^G\rangle \\
&\quad + \mathrm{O}_P\left(\frac{\sqrt{k}}{n-v+k}\right) + \mathrm{O}_P\left(\frac{\sqrt{k}}{v+k}\right) + \mathrm{O}_P\left(\frac{1}{\sqrt{n-v+k}}\right) + \mathrm{O}_P\left(\frac{1}{\sqrt{v}}\right),
\end{aligned}$$

where Z_1' and Z_2' have the same meaning as above.

Proof. We look only at the case $1 \leqslant k < v - c_0(n)$. To determine the asymptotic behavior of the first Bayes factor we write it as

$$\frac{\hat{p}_{v-k}(\mathbf{x})}{p_{v-k}(\mathbf{x})} = \int_{\mathbf{B}} \prod_{1}^{v-k} \frac{f(x_j,\beta)}{f(x_j,\beta_0)} \pi(\beta)\,\mathrm{d}\beta \int_{\boldsymbol{\Gamma}} \prod_{v-k+1}^{n} \frac{g(x_j,\gamma)}{g(x_j,\gamma_0)} \tau(\gamma)\,\mathrm{d}\gamma$$

$$= \int_{\mathbf{B}} \int_{\boldsymbol{\Gamma}} \exp\left\{ \sum_{1}^{v-k} \log \frac{f(x_j,\beta)}{f(x_j,\beta_0)} + \sum_{v-k+1}^{n} \log \frac{g(x_j,\gamma)}{g(x_j,\gamma_0)} \right\} \pi(\beta)\tau(\gamma)\,\mathrm{d}\beta\,\mathrm{d}\gamma.$$

Let $\hat{\beta}_{v-k}$ denote the maximum likelihood estimator of β on the basis of $x_1,\ldots,x_{v-k}$. Then

$$\int_{\mathbf{B}} \exp\left\{ \sum_{1}^{v-k} \log \frac{f(x_j,\beta)}{f(x_j,\beta_0)} \right\} \pi(\beta)\,\mathrm{d}\beta = \exp\left\{ \sum_{1}^{v-k} \log \frac{f(x_j,\hat{\beta}_{v-k})}{f(x_j,\beta_0)} \right\} \pi(\beta_0)$$

$$\int_{\mathbf{B}} \exp\left\{ -\frac{v-k}{2} \langle I_1(\beta - \hat{\beta}_{v-k}), \beta - \hat{\beta}_{v-k} \rangle \right\} \mathrm{d}\beta [1 + \mathrm{o}_P(1)].$$

The proof of Theorem 8.1, Chapter 1 in Ibragimov and Hasminskii (1981) gives

$$\sum_{1}^{v-k} \log \frac{f(x_j,\hat{\beta}_{v-k})}{f(x_j,\beta_0)} = \left\langle (\hat{\beta}_{v-k} - \beta_0), \sum_{1}^{v-k} \phi_0(x_j) \right\rangle + \mathrm{O}_P\left(\frac{1}{\sqrt{v-k}} \right)$$

$$= \frac{1}{2(v-k)} \left\langle I_1^{-1} \sum_{1}^{v-k} \phi_0(x_j), \sum_{1}^{v-k} \phi_0(x_j) \right\rangle + \mathrm{O}_P\left(\frac{1}{\sqrt{v-k}} \right).$$

The same argument with $k=0$ shows that

$$\log \int_{\mathbf{B}} \prod_{1}^{v} \frac{f(x_j,\beta)}{f(x_j,\beta_0)} \pi(\beta)\,\mathrm{d}\beta = \frac{1}{2v} \left\langle I_1^{-1} \sum_{1}^{v} \phi_0(x_j), \sum_{1}^{v} \phi_0(x_j) \right\rangle + \mathrm{O}_P\left(\frac{1}{\sqrt{v}} \right).$$

By combining these results one obtains the following expression for the distribution of the logarithm of the first factor of $\hat{p}_{v-k}(\mathbf{x}) p_v(\mathbf{x}) / \hat{p}_v(\mathbf{x}) p_{v-k}(\mathbf{x})$,

$$\log \int_{\mathbf{B}} \prod_{1}^{v-k} \frac{f(x_j,\beta)}{f(x_j,\beta_0)} \pi(\beta)\,\mathrm{d}\beta - \log \int_{\mathbf{B}} \prod_{1}^{v} \frac{f(x_j,\beta)}{f(x_j,\beta_0)} \pi(\beta)\,\mathrm{d}\beta$$

$$= \frac{k}{2v(v-k)} \left\langle I^{-1} \sum_{1}^{v-k} \phi_0(x_j), \sum_{1}^{v-k} \phi_0(x_j) \right\rangle$$

$$- \frac{1}{v} \left\langle I^{-1} \sum_{1}^{v-k} \phi_0(x_j), \sum_{v-k+1}^{v} \phi_0(x_j) \right\rangle$$

$$- \frac{1}{2v} \left\langle I^{-1} \sum_{v-k+1}^{v} \phi_0(x_j), \sum_{v-k+1}^{v} \phi_0(x_j) \right\rangle + \mathrm{O}_P\left(\frac{1}{\sqrt{v-k}} \right).$$

By the Central Limit Theorem, the distribution above is of the form

$$\frac{k}{2v}\langle Z_1, Z_1\rangle - \frac{\sqrt{v-k}}{v}\langle Z_1, \Phi_k^F\rangle - \frac{1}{2v}\langle \Phi_k^F, \Phi_k^F\rangle + \mathrm{O}_P\left(\frac{\sqrt{k}}{v}\right) + \mathrm{O}_P\left(\frac{1}{\sqrt{v-k}}\right).$$

Now we turn to the second factor corresponding to $\max_\gamma \sum_{v-k+1}^n \log g_\gamma/g_{\gamma_0}(x_j) - \max_\gamma \sum_{v+1}^n \log g_\gamma/g_{\gamma_0}(x_j)$. The maximum likelihood estimator $\hat{\gamma}_{n-v+k} = \arg\max_\gamma \prod_{v-k+1}^n g(x_j, \gamma)$, based on the last $n-v+k$ observations, can be interpreted as an M-estimator of the parameter γ, when the true distribution is the mixture of densities $f(x, \beta_0)$ and $g(x, \gamma_0)$. The first density corresponding to $x_{v-k+1}, \ldots, x_v$ has the weight

$$w = \frac{k}{n-v+k} \leqslant \frac{v-c_0(n)}{n-c_0(n)}$$

and the second density corresponding to $x_{v+1}, \ldots, x_n$ has the weight $1-w$. Since under condition (26) $w \leqslant [c_1(n)-c_0(n)]/[n-c_0(n)] \to 0$, one obtains as before for a fixed k

$$\max_\gamma \sum_{v-k+1}^n \log g(x_j,\gamma) - \sum_{v-k+1}^n \log g(x_j, \gamma_0)$$
$$= \frac{n-v+k}{2}\langle \Lambda(\hat{\gamma}_{n-v+k}-\gamma_0), \hat{\gamma}_{n-v+k}-\gamma_0\rangle + \mathrm{O}_P\left(\frac{1}{\sqrt{n-v+k}}\right).$$

In this formula

$$\Lambda = (1-w)I_2 - wI_{12},$$

where

$$I_{12} = \left\{E_{\beta_0}^F \frac{\partial^2}{\partial\gamma_i\partial\gamma_k}\log g(X, \gamma_0)\right\},$$

so that

$$\Lambda^{-1} \sim I_2^{-1} + wI_2^{-1}[I_{12}+I].$$

According to stochastic expansion of M-estimators,

$$\sqrt{n-v+k}\,\Lambda[\hat{\gamma}_{n-v+k}-\gamma_0] = \frac{1}{\sqrt{n-v+k}}\sum_{v-k+1}^n \psi_0(x_j) + \mathrm{O}_P\left(\frac{1}{\sqrt{n-v+k}}\right).$$

(see for example Bednarski et al., 1991). Thus for any fixed k;

$$\max_\gamma \sum_{v-k+1}^n \log\frac{(x_j,\gamma)}{g(x_j,\gamma_0)} - \max_\gamma \sum_{v+1}^n \log\frac{g(x_j,\gamma)}{g(x_j,\gamma_0)}$$
$$= \frac{1}{2(n-v+k)}\left\langle \Lambda^{-1}\left[\sum_{v-k+1}^v \psi_0(x_j) + \sum_{v+1}^n \psi_0(x_j)\right], \sum_{v-k+1}^v \psi_0(x_j) + \sum_{v+1}^n \psi_0(x_j)\right\rangle$$
$$- \frac{1}{2(n-v)}\left\langle \Lambda^{-1}\sum_{v+1}^n \psi_0(x_j), \sum_{v+1}^n \psi_0(x_j)\right\rangle + \mathrm{O}_P\left(\frac{1}{\sqrt{n-v}}\right).$$

The Central Limit Theorem shows that this distribution has the form

$$\frac{1}{2(n-v+k)}\langle \Psi_k^F, \Psi_k^F\rangle + \frac{\sqrt{v-k}}{n-v+k}\langle Z_2, \Psi_k^F\rangle - \frac{k}{2(n-v+k)}\langle Z_2, Z_2\rangle + \mathrm{O}_P\left(\frac{k}{n-v+k}\right) + \mathrm{O}_P\left(\frac{1}{\sqrt{v-k}}\right).$$

This leads to the first formula of the Proposition 1.

The remaining proof for the case, $1 \leqslant k < c_1(n) - v$, is quite similar.

Now we can establish the convergence to (6) if

$$E_{\beta_0}^F \log \psi_0(X) = 0; \quad E_{\gamma_0}^G \log \phi_0(X) = 0. \tag{27}$$

The range of the change-point values, described by conditions (28) below, is satisfied by sequences $c_0(n), c_1(n)$, such that for some ρ, $0 < \rho < 1$

$$c_0(n) = \rho n + \varepsilon_0(n), \quad c_1(n) = \rho n + \varepsilon_1(n),$$

with $\varepsilon_i(n) = o(n^{1/2})$, $i = 0, 1$. When (27) does not hold, the range of possible change-point values becomes narrower.

Theorem 3. *In the situation of Proposition* 1 *if* (26) *and* (27) *are satisfied, then the asymptotic distribution of* (14) *is that of* (6). *The same conclusion holds without assuming* (27) *if*

$$\frac{(c_1(n) - c_0(n))^2}{\min[c_1(n), n - c_0(n)]} \to 0. \tag{28}$$

Proof. Proposition 1 shows that when the condition (27) is satisfied, both $\hat{p}_{v-k}(\mathbf{x})p_v(\mathbf{x})/[p_{v-k}(\mathbf{x})\hat{p}_v(\mathbf{x})]$ and $\hat{p}_{v+k}(\mathbf{x})p_v(\mathbf{x})/[p_{v+k}(\mathbf{x})\hat{p}_v(\mathbf{x})]$ tend to 1. The argument as in Theorem 2 shows that the convergence of the posterior distribution to (6) takes place. Indeed under (27) both normalized sums, $\Phi_k^G/\sqrt{k}$ and $\Psi_k^F/\sqrt{k}$, have asymptotic normal distributions with zero means, and, for example, $\max_{1 \leqslant k < v - c_0(n)} \Psi_k^F/\sqrt{k} = \mathrm{O}_P(1)$, so that $\max_{1 \leqslant k < v - c_0(n)} \Psi_k^F = \mathrm{O}_P(\sqrt{v - c_0(n)})$. Then because of (26)

$$\max_{1 \leqslant k < v - c_0(n)} \frac{\langle \Psi_k^F, \Psi_k^F\rangle}{n - v + k} = \mathrm{O}_P\left(\frac{v - c_0(n)}{n - c_0(n)}\right) \to 0.$$

A similar conclusion holds for other terms in the formula for the distribution of $\hat{p}_{v-k}(\mathbf{x})p_v(\mathbf{x})/[p_v(\mathbf{x})\hat{p}_v(\mathbf{x})]$ in Proposition 1.

Under condition (28)

$$\max_{1 \leqslant k < v - c_0(n)} \frac{\langle \Psi_k^F, \Psi_k^F\rangle}{n - v + k} = \mathrm{O}_P\left(\frac{k^2 ||E_{\beta_0}^F \log \psi_0(X)||^2}{n - v + k}\right) \to 0$$

along with other terms in the last formula of Proposition 1.

As in Section 2, the same result holds for the profile likelihood function

$$\bar{p}_v(\mathbf{x}) = \max_{\beta\in\mathbf{B}} \prod_1^v f_\beta(x_j) \max_{\gamma\in\Gamma} \prod_{v+1}^n g_\gamma(x_j).$$

The set $(c_0(n), c_1(n)]$ of the values of the change point under the conditions (26) and (27) (or only (28)) can be described as the *adaptation* region. Indeed for any such parametric value v the asymptotic posterior distribution is the same as (8) (which uses the knowledge of β_0 and γ_0). In this situation the Bayes estimator $\hat{v}$ exhibits adaptive behavior, i.e. if β and γ are viewed as nuisance parameters, it is performing in an asymptotically optimal way for any unknown (but fixed) values of these nuisance parameters. A similar result for the zero-one loss and densities $f(x,\beta)$ and $g(x,\gamma)$ belonging to the same exponential family was obtained by Hu and Rukhin (1995).

An alternative procedure estimates the parameter β by the maximum likelihood (or by the Bayes) estimator $\tilde{\beta}(x_1,\ldots,x_{c_0(n)})$ using observations $x_1,\ldots,x_{c_0(n)}$ only; similarly the parameter γ is estimated from observations $x_{c_1(n)+1},\ldots,x_n$ by $\tilde{\gamma}(x_{c_1(n)+1},\ldots,x_n)$. These estimated parametric values are used in the formula (14) to obtain

$$\tilde{p}_v(\mathbf{x}) = \prod_{c_0(n)+1}^{v} f(x_j,\tilde{\beta}) \prod_{v+1}^{c_1(n)} g(x_j,\tilde{\gamma})$$

and $\tilde{\mathbf{m}}(\mathbf{x}) = \sum q^k \tilde{p}_k(\mathbf{x})$.

Theorem 4. *Under conditions of Theorem* 2 *the estimated posterior probabilities*

$$P(N=m|\mathbf{x}) = \frac{q^m \tilde{p}_m(\mathbf{x})}{\tilde{\mathbf{m}}(\mathbf{x})}$$

converge in distribution to (6).

Proof. As before, one obtains the formula for the distribution of $\tilde{\mathbf{m}}(\mathbf{x})/[q^v \tilde{p}_v(\mathbf{x})]$

$$1 + \sum_1^{v-c_0(n)-1} \frac{\tilde{p}_{v-k}(\mathbf{x})}{\tilde{p}_v(\mathbf{x})} q^{-k} + \sum_1^{c_1(n)-v} \frac{\tilde{p}_{v+k}(\mathbf{x})}{\tilde{p}_v(\mathbf{x})} q^k. \tag{29}$$

When $1 \leqslant k < v - c_0(n)$,

$$\frac{\tilde{p}_{v-k}(\mathbf{x})}{\tilde{p}_v(\mathbf{x})} = \prod_{v-k+1}^{v} \frac{g(x_j,\tilde{\gamma})}{f(x_j,\tilde{\beta})} = \prod_{v-k+1}^{v} \frac{g(x_j,\tilde{\gamma})}{g(x_j,\beta_0)} \frac{f(x_j,\beta_0)}{f(x_j,\tilde{\beta})}.$$

To demonstrate the weak convergence of the first sum in (29) notice that for example

$$\tilde{\beta} - \beta_0 = \frac{1}{c_0(n)} I_1^{-1} \sum_1^{c_0(n)} \phi_0(x_j) + \mathrm{O}_P\left(\frac{1}{c_0(n)}\right),$$

so that

$$\sum_{v-k+1}^{v} \log \frac{f(x_j,\tilde{\beta})}{f(x_j,\beta_0)} = \left\langle \sum_{v-k+1}^{v} \phi_0(x_j), \tilde{\beta} - \beta_0 \right\rangle + \mathrm{O}_P\left(\frac{1}{\sqrt{c_0(n)}}\right)$$

has the distribution of

$$\frac{1}{\sqrt{c_0(n)}} \langle \Phi_k^F, \tilde{Z}_1 \rangle + \mathrm{O}_P\left(\frac{||\Phi_k^F||}{c_0(n)}\right)$$

with p-dimensional independent standard normal vector $\tilde{Z}_1$ independent of Φ_k^F. Now convergence follows as in the proof of Theorem 3.

The estimator

$$\tilde{\tilde{\delta}}_n(\mathbf{x}) = \arg \min_{d:\ c_0(n) < d \leqslant c_1(n)} \sum_v L(d-v) q^v \tilde{p}_v(\mathbf{x})$$

is adaptive in a wider region of possible change-point parameter values than (14).

5. Example

In this section we illustrate the results by the change-point detection problem when both the pre-change and after-change data are normal, i.e. $\mathbf{x}_1 = (x_1, \ldots, x_v)$ is a normal random vector whose components are independent and have a normal distribution $N(\theta_\beta, \sigma_\beta^2)$, and the components of the vector $\mathbf{x}_2 = (x_{v+1}, \ldots, x_n)$ are also normal $N(\mu_\gamma, \kappa_\gamma^2)$.

Under the default non-informative (uniform) prior for all nuisance parameters the conditional density of $\mathbf{x}$ for given v, $\min[v, n-v] \geqslant 2$, takes the form

$$\begin{aligned}
\hat{p}_v(\mathbf{x}) &= \frac{1}{(2\pi)^{n/2}} \int\!\!\int \frac{\exp\{-\sum_1^v (x_j-\theta)^2/2\sigma^2\}}{\sigma^v} \, \mathrm{d}\theta \frac{\mathrm{d}\sigma}{\sigma} \\
&\quad \times \int\!\!\int \frac{\exp\{-\sum_{v+1}^n (x_j-\mu)^2/2\kappa^2\}}{\kappa^{n-v}} \, \mathrm{d}\mu \frac{\mathrm{d}\kappa}{\kappa} \\
&= \frac{2^{(n-6)/2} \Gamma((v-1)/2) \Gamma((n-v-1)/2)}{(2\pi)^{n/2-1} \sqrt{v(n-v)} s_v^{v-1} s_{n-v}^{n-v-1}}.
\end{aligned}$$

Here $\bar{x}_k = (x_1 + \cdots + x_k)/k$, $\bar{x}_{n-k} = (x_{k+1} + \cdots + x_n)/(n-k)$, $s_k^2 = \sum_1^k (x_j - \bar{x}_k)^2$ and $s_{n-k}^2 = \sum_{k+1}^n (x_j - \bar{x}_{n-k})^2$.

The profile likelihood function has the form

$$\begin{aligned}
\bar{p}_v(\mathbf{x}) &= \frac{\exp\{-\sum_1^v (x_j - \bar{x}_v)^2/2s_v^2/v - \sum_{v+1}^n (x_j - \bar{x}_{n-v})^2/2s_{n-v}^2/(n-v)\}}{(2\pi)^{n/2}(s_v^2/v)^{v/2}(s_{n-v}^2/(n-v))^{(n-v)/2}} \\
&= \frac{\mathrm{e}^{-n/2}}{(2\pi)^{n/2}(s_v^2/v)^{v/2}(s_{n-v}^2/(n-v))^{(n-v)/2}}.
\end{aligned}$$

These posterior distributions have been used in the change-point detection in the problem of weight differences measurements made at the National Institute of Standards and Technology (NIST) between 1975 and 1988 (see Pollak et al., 1993).

In this data set obtained from the mass calibration at NIST with $n = 217$, a change in the measuring process was surmised to occur during this time period and its detection was of importance. This data suggests not only the change in the mean, but also the change in the variance. As a matter of fact, the latter is more pronounced than the means change. Indeed for $1 \leqslant c_0 \leqslant 20$, $c_1 = 217 - c_0$ the ratio of the estimates of standard deviations obtained from "training" samples of the first c_0 and the last c_0

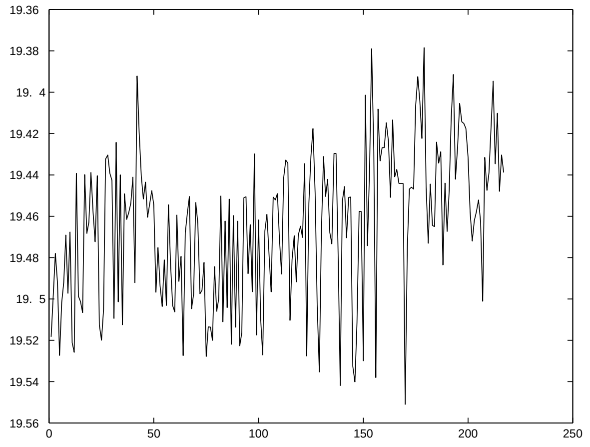

Fig. 1. The NIST weight differences measurements data.

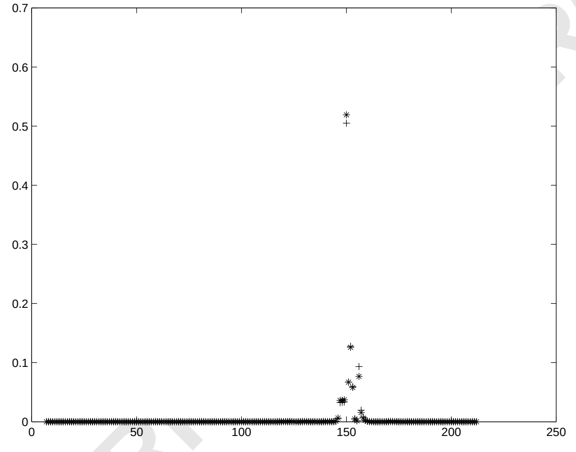

Fig. 2. The posterior probabilities obtained from non-informative prior distributions (line marked by $+$), and from the profile likelihood function (line marked by $*$).

observations is about 1.5 whereas the standardized difference between the means is only about 0.8.

The posterior mean and the posterior mode for both distributions lead to the same conclusion ($\nu = 151$) as reached by Pollak et al. (1993), who used sequential surveillance scheme of Shiryaev-Roberts. Actually the posterior probability of this value calculated from non-informative prior distributions is about 0.5049.., while for the profile likelihood based posterior this probability is 0.5191... For $q = 1$ both posterior distributions are depicted in Fig. 2. This picture says more eloquently than numbers what are the chances for the change-point to occur anywhere in the observation period especially when compared with raw data in Fig. 1.

Acknowledgements

This research was supported by von Humboldt Foundation. The author is grateful to Michael Baron for many valuable comments.

References

Bednarski, T., Clarke, B.R., Kolkiewicz, W., 1991. Statistical expansions and locally uniform Frechet differentiablity. J. Aust. Math. Soc. Ser. A 50, 88–97.

Bhattacharya, P.K., 1994. Some aspects of change-point analysis. In: Mueller, H.-G., Carlstein, E., Siegmund, D. (Eds.), Change-Point Problems. Institute of Mathematical Statistics, Hayward, CA.

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

Broemling, L.D., 1974. Bayesian inferences about a changing sequence of random variables. Commun. Stat. 3, 243–255.

Chernoff, H., Zacks, S., 1964. Estimating the current mean of a normal distribution which is subject to changes in time. Ann. Math. Stat. 35, 999–1018.

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

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

Duembgen, L., 1991. The asymptotic behavior of some nonparametric change-point estimators. Ann. Stat. 19, 1471–1495.

Dufresne, D., 1990. The distribution of a perpetuity, with applications to risk theory and pension funding. Scand. Actuarial J. :39–79.

Grenander, U., 1964. Stochastic groups. Ark. Math. 4, 189–207.

Grincevicius, A.K., 1975. Limit theorems for products of random linear transformations on the line. Lithuanian Math. Trans. 15, 558–579.

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

Hu, I., Rukhin, A.L., 1995. A lower bound for error probability in change-point estimation. Statistica Sinica 5, 319–331.

Ibragimov, I.A., Hasminskii, R.Z., 1981. Statistical Estimation: Asymptotic Theory. Springer, New York.

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

Nikiforov, I.V., 1995. A generalized change-detection problem. IEEE Transactions on Information Theory 41, 171–187.

Pollak, M., 1985. Optimal detection of a change in distribution. Ann. Stat. 13, 206–227.

Pollak, M., Croarkin, C., Hagwood, C., 1993. Surveillance schemes with applications to mass calibration. Technical Report NISTIR 5158, National Institute of Standards and Technology, Gaitherburg, MD.

Ritov, Y., 1990. Asymptotic efficient estimation of the change point with unknown distributions. Ann. Stat. 18, 1829–1839.

Roussas, G.G., 1972. Contiguity of Probability Measures: Some Applications to Statistics. Cambridge University Press, Cambridge, UK.

Rukhin, A.L., 1997. Change-point estimation under asymmetric loss. Stat. Decisions 15, 141–163.

Shiryaev, A.N., 1978. Optimal Stopping Rules. Springer, New York.

Smith, A.F.M., 1975. A Bayesian approach to inference about a change-point in a sequence of random variables. Biometrika 62, 407–416.

Vervaat, W., 1979. On the stochastic difference equation and a representation of non-negative infinitely divisible random variables. Adv. Appl. Probab. 11, 750–783.

Yao, Y., Huang, D., Davis, R.A., 1994. On almost sure behavior of change-point estimators. In: Siegmund, D., Carlstein, E., Mueller, H. (Eds.), Proceedings on Change-Point Problems. Institute of Mathematical Statistics, Hayward, CA.

Zacks, S., 1983. Survey of classical and Bayesian approaches to the change-point problem: fixed sample and sequential procedures of testing and estimation. In: Rizvi, M.H., Siegmund, D. (Eds.), Recent Advances in Statistics, Festschrift for H. Chernoff. Academic Press, New York.