

On Two Estimates Related to the Change-Point Problem

F. Enikeeva^{1*}

¹*The Kharkevich Inst. for Inform. Transmission Probl. of RAS, Moscow, Russia*

Received September 17, 2010; in final form, December 28, 2011

Abstract—We consider the problem of estimating a smooth functional of an unknown signal with discontinuity from Gaussian observations. The signal is a known function depending on an unknown parameter. This problem is closely related to the famous change-point problem. We obtain an asymptotic likelihood ratio process for the noise level tending to 0. Bayesian and maximum likelihood estimates are constructed and their relative efficiency is studied. Some simulation results and conclusions on non-asymptotic behavior of these estimates are presented.

Keywords: change-point problem, white noise model, relative efficiency.

2000 Mathematics Subject Classification: primary 62G05; secondary 62G20, 93E10, 62L12.

DOI: 10.3103/S1066530712010024

1. INTRODUCTION

We consider the classical Gaussian white noise model

$$dX_\varepsilon(t) = f(t)dt + \varepsilon dW(t), \quad t \in [0, 1], \quad (1)$$

where $f(t)$ is an unknown signal, $W(t)$ is the standard Brownian motion, the noise level $\varepsilon > 0$ is known. We assume that the function f is continuous everywhere on $[0, 1]$ except an unknown point τ and depends on an unknown parameter θ , $f(t) \equiv f^\tau(\theta, t)$.

Let $\mathcal{L}: \mathbf{L}_2[0, 1] \rightarrow \mathbb{R}$ be a given smooth functional of f . The goal of this paper is to compare Bayesian and maximum likelihood estimates of $\mathcal{L}[f]$ assuming that the function f is known up to the parameters τ and θ . Let $\hat{\mathcal{L}}(X_\varepsilon)$ be an estimate of $\mathcal{L}[f]$. We will use the quadratic loss function and the mean square risk for measuring the performance of the estimator:

$$R_\varepsilon(\hat{\mathcal{L}}, \mathcal{L}) = \mathbf{E}_{\theta, \tau}(\hat{\mathcal{L}}(X_\varepsilon) - \mathcal{L}[f])^2.$$

The model of observations (1) of the Wiener process with a discontinuous drift was first considered by Ibragimov and Hasminski [10]. Assuming that the function f is known with an unknown discontinuity point τ the authors studied asymptotic efficiencies of Bayesian and maximum likelihood estimates of τ as $\varepsilon \rightarrow 0$. Asymptotic mean-square error of an MLE of the discontinuity point τ was calculated and an approximate value of the quadratic risk of Bayesian estimate was obtained. Later Rubin and Song [13] found an exact representation for the mean-square error of the Bayes estimate of τ in terms of Riemann's zeta function. According to their results, the Bayes procedure of estimating a change point is asymptotically more efficient than the maximum likelihood procedure with the asymptotic relative efficiency $\frac{8}{13}\zeta(3) \approx 0.7397$.

This problem is closely related to the famous change-point problem considered by many authors. The literature on the change-point problem is vast, we refer the reader to the monographs of Csörgő and Horváth [7] on asymptotic theory in the change-point problem, of Brodsky and Darhovskiy [5] on nonparametric methods, of Shiryaev [15] on optimal detection of change in distribution, and many references therein. We also refer to an excellent review article of Bhattacharya [1] that provides historical perspectives of the classical change-point problem.

*E-mail: enikeeva@iitp.ru

In this paper, we construct two estimates of $\mathcal{L}[f]$ for model (1). We compare the asymptotic efficiencies of MLE and Bayesian estimate in the white noise model following the approach of [10]. It turns out that both estimates have the same asymptotic efficiency.

The paper is organized as follows. In Section 2 we give a precise statement of the problem and obtain the asymptotic likelihood ratio process. In Section 3 the results on the relative efficiency of Bayesian and maximum likelihood estimates of the smooth functional are presented. Section 4 contains the results for a sequence version of (1) representing the change in mean of a Gaussian sequence. In Section 5 we present simulation results for different values of signal-to-noise ratio and discuss both asymptotic and non-asymptotic aspects of the problem.

2. LIMITING LIKELIHOOD RATIO PROCESS

It will be easier to work with a stochastic process $Y(t)$ satisfying the stochastic differential equation

$$dY(t) = \frac{1}{\varepsilon} f^\tau(\theta, t) dt + dW(t), \quad t \in [0, 1], \quad (2)$$

where $W(t)$ is the standard Wiener process, $W(0) = 0$, and $\varepsilon > 0$.

The function $f^\tau(\theta, t)$ is defined as

$$f^\tau(\theta, t) = \begin{cases} f_1(\theta_1, t), & 0 \leq t < \tau, \\ f_2(\theta_2, t), & \tau < t \leq 1, \end{cases} \quad (3)$$

where $\theta = (\theta_1, \theta_2)$ and τ are unknown parameters that belong to some compact sets, $\theta \in \Theta = \Theta_1 \times \Theta_2 \subset \mathbb{R}^2$, $\tau \in T = [t_0, 1 - t_0] \subset (0, 1)$, $t_0 > 0$. Setting $t_0 > 0$, we assume that the change-point τ is separated from 0 and 1 so that the change in the data really occurred. The functions $f_1(\theta_1, \cdot)$ and $f_2(\theta_2, \cdot)$ are defined on some open subsets of \mathbb{R} that contain $[0, \tau]$ and $[\tau, 1]$, respectively.

At the point $t = \tau$ we define $f^\tau(\theta, \tau \pm) = \lim_{t \rightarrow \tau \pm} f^\tau(\theta, t)$. Denote by $\Delta = f^\tau(\theta, \tau -) - f^\tau(\theta, \tau +) \equiv f_1(\theta_1, \tau) - f_2(\theta_2, \tau)$ the jump size at the point τ assuming that $\Delta \neq 0$.

In fact, $f^\tau(\theta, \cdot)$ depends on θ_1 on $[0, \tau]$ and on θ_2 on $[\tau, 1]$. Thus, by abuse of notation we will write $\frac{\partial f_i}{\partial \theta}(\theta_i, t)$ meaning the value of the partial derivative of $f_i(x, t)$ with respect to x at $x = \theta_i$.

Let $\mathcal{F} \subset \mathbf{L}_2[0, 1]$ be a linear space such that for any fixed parameters θ and τ the function $f^\tau(\theta, \cdot) \in \mathcal{F}$ satisfies the following condition.

Condition F. Assume that

- (a) The functions $f_i(\theta_i, t)$, $i = 1, 2$, are continuous in t in their domains and in θ_i on Θ_i , respectively.
- (b) For any x in a neighborhood of θ , $f^\tau(x, t)$ has a bounded derivative $\frac{\partial f^\tau}{\partial t}(\theta, t)$ for all $t \in [0, 1]$ except $t = \tau$.
- (c) $f_i(\theta, t)$, $i = 1, 2$, are differentiable with respect to θ at θ_i 's such that

$$\lim_{\delta \rightarrow 0} \frac{1}{\delta} \left\| f_i(\theta_i + \delta, \cdot) - f_i(\theta_i, \cdot) - \frac{\partial f_i}{\partial \theta}(\theta_i, \cdot) \delta \right\|_{\mathbf{L}_2[0, 1]} = 0.$$

Our goal is to estimate a smooth functional $\mathcal{L}[f^\tau(\theta, \cdot)]$ of the signal $f^\tau(\theta, \cdot)$. Below the conditions on the functional \mathcal{L} are specified.

Condition L. Let $\theta \in \Theta$ and $\tau \in T$ be fixed. The functional $\mathcal{L}: \mathcal{F} \subset \mathbf{L}_2[0, 1] \rightarrow \mathbb{R}$ is Fréchet differentiable at $f^\tau(\theta, \cdot) \in \mathcal{F}$.

Condition L'. The value $\mathcal{L}[f^\tau(\theta, \cdot)]$ of the functional $\mathcal{L}: \mathcal{F} \subset \mathbf{L}_2[0, 1] \rightarrow \mathbb{R}$ at $f^\tau(\theta, \cdot)$ is differentiable with respect to τ in a neighborhood of θ .

Let $\widehat{\tau}_{\text{mle}}^\varepsilon$ be an MLE of τ based on observations (2). Let $\widehat{\tau}_b^\varepsilon$ be a Bayes estimate of τ based on observations (2), where τ has some positive prior distribution on $[0, 1]$. The analysis of quadratic errors of these two estimates is based on the properties of the stochastic process

$$V(t) = \exp(B(t) - |t|/2), \tag{4}$$

where $B(t)$ is the two-sided Brownian motion defined by

$$B(t) = \begin{cases} W_1(t), & t \geq 0, \\ W_2(-t), & t < 0. \end{cases} \tag{5}$$

Here $W_i(t)$, $t \geq 0$, $i = 1, 2$, are independent standard Wiener processes with $W_i(0) = 0$. In fact, if $\Delta = f_1(\theta_1, \tau) - f_2(\theta_2, \tau)$, then the process $V(\Delta^2 t)$ is a limiting process for the likelihood ratio of τ corresponding to the observations $Y(t)$ [10].

Recall that θ and τ are defined on the compact sets $\Theta = \Theta_1 \times \Theta_2 \subset \mathbb{R}^2$ and $T \subset (0, 1)$, respectively. Following the approach of [10] we will fix the unknown parameters $\theta = (\theta_1, \theta_2)$ and τ and work with the local parameters $h = (h_1, h_2)$ and u . Let $h = (x - \theta)/\varepsilon$ and $u = (t - \tau)/\varepsilon^2$, where $x \in \Theta$, $t \in T$. Introduce the normalized sets $\Theta_\varepsilon = \Theta_\varepsilon^1 \times \Theta_\varepsilon^2$, where $\Theta_\varepsilon^i = \varepsilon^{-1}(\Theta_i - \theta_i)$, $T_\varepsilon = \varepsilon^{-2}(T - \tau)$ such that $h \in \Theta_\varepsilon$, $u \in T_\varepsilon$. Let $\mathbf{P}_{\theta, \tau}$ be the measure generated by the process (2) and $Z_{\theta, \tau}^\varepsilon(h, u)$ be the likelihood ratio of θ and τ based on this process,

$$Z_{\theta, \tau}^\varepsilon(h, u) = \frac{d\mathbf{P}_{\theta + \varepsilon h, \tau + \varepsilon^2 u}}{d\mathbf{P}_{\theta, \tau}}(Y(t)).$$

Lemma 1. *Let $H = H_1 \times H_2 \subset \Theta_\varepsilon$ and $U \subset T_\varepsilon$ be compact sets and Condition F be satisfied. The distribution of the log-likelihood ratio process $\log \tilde{Z}_{\theta, \tau}^\varepsilon(h, u)$ as $\varepsilon \rightarrow 0$ converges uniformly over $(h, u) \in H \times U$ to the distribution of the process*

$$\log Z_{\theta, \tau}^0(h, u) = \frac{1}{2}(Z_1^2 + Z_2^2) - \frac{1}{2}I_1^2 \left(h_1 - \frac{Z_1}{I_1} \right)^2 - \frac{1}{2}I_2^2 \left(h_2 - \frac{Z_2}{I_2} \right)^2 + \log V(\Delta^2 u), \tag{6}$$

where Z_1 and Z_2 are independent $\mathcal{N}(0, 1)$, the process $V(u)$ defined in (4) is independent of Z_1 and Z_2 , and

$$I_1 = \left(\int_0^\tau \left| \frac{\partial f_1}{\partial \theta}(\theta_1, t) \right|^2 dt \right)^{1/2}, \quad I_2 = \left(\int_\tau^1 \left| \frac{\partial f_2}{\partial \theta}(\theta_2, t) \right|^2 dt \right)^{1/2}.$$

Proof. From Girsanov's theorem (see [10], Appendix II, Theorem 1) it follows that the likelihood ratio for the measures generated by $Y(t)$ with the parameters θ, τ and $\theta + \varepsilon h, \tau + \varepsilon^2 u$ satisfies

$$\log Z_{\theta, \tau}^\varepsilon(h, u) = \log \frac{d\mathbf{P}_{\theta + \varepsilon h, \tau + \varepsilon^2 u}}{d\mathbf{P}_{\theta, \tau}}(Y) = s_\varepsilon(h, u) - \frac{1}{2}r_\varepsilon^2(h, u), \tag{7}$$

where

$$s_\varepsilon(h, u) = \frac{1}{\varepsilon} \int_0^1 (f^{\tau + \varepsilon^2 u}(\theta + \varepsilon h, t) - f^\tau(\theta, t)) dW(t),$$

$$r_\varepsilon^2(h, u) = \frac{1}{\varepsilon^2} \int_0^1 (f^{\tau + \varepsilon^2 u}(\theta + \varepsilon h, t) - f^\tau(\theta, t))^2 dt.$$

Using the same approach as in Lemma 7.2.1 of [10] and Condition F[c] it is not difficult to show that uniformly over the compact set $H \times U$

$$r_\varepsilon^2(h, u) = \Delta^2 |u| + h_1^2 \int_0^\tau \left| \frac{\partial f_1}{\partial \theta}(\theta_1, t) \right|^2 dt + h_2^2 \int_\tau^1 \left| \frac{\partial f_2}{\partial \theta}(\theta_2, t) \right|^2 dt + o(1)$$

$$= \Delta^2 |u| + h_1^2 I_1^2 + h_2^2 I_2^2 + o(1), \quad \varepsilon \rightarrow 0.$$

Consider now the stochastic part $s_\varepsilon(h, u)$ of the process. Let $u > 0$. We have

$$s_\varepsilon(h, u) = \frac{1}{\varepsilon} \int_0^\tau (f_1(\theta_1 + \varepsilon h_1, t) - f_1(\theta_1, t)) dW(t) + \frac{1}{\varepsilon} \int_{\tau+\varepsilon^2 u}^1 (f_2(\theta_2 + \varepsilon h_2, t) - f_2(\theta_2, t)) dW(t) \\ + \frac{1}{\varepsilon} \int_\tau^{\tau+\varepsilon^2 u} (f_1(\theta_1 + \varepsilon h_1, t) - f_2(\theta_2, t)) dW(t). \quad (8)$$

Following [10] from Condition F[c] we can obtain the weak convergence of first two terms of (8) to

$$h_1 Z_1 I_1 \stackrel{d}{=} h_1 \int_0^\tau \frac{\partial f_1}{\partial \theta}(\theta_1, t) dW(t) \quad \text{and} \quad h_2 Z_2 I_2 \stackrel{d}{=} h_2 \int_\tau^1 \frac{\partial f_2}{\partial \theta}(\theta_2, t) dW(t),$$

respectively, where I_1 and I_2 are defined in the statement of the lemma and Z_1 and Z_2 are independent $\mathcal{N}(0, 1)$.

Indeed, consider the first term of (8). It can be written as the sum

$$V_\varepsilon(h_1) + \int_0^\tau h_1 \frac{\partial f_1}{\partial \theta}(\theta_1, t) dW(t),$$

where

$$V_\varepsilon(h_1) = \frac{1}{\varepsilon} \int_0^\tau \left[(f_1(\theta_1 + \varepsilon h_1, t) - f_1(\theta_1, t)) - h_1 \varepsilon \frac{\partial f_1}{\partial \theta}(\theta_1, t) \right] dW(t).$$

We have $\mathbf{E}V_\varepsilon(h_1) = 0$. From Condition F[c] it follows that for some constant $C > 0$ and any $h'_1, h''_1 \in H$

$$\mathbf{E}(V_\varepsilon(h'_1) - V_\varepsilon(h''_1))^2 \leq C(h'_1 - h''_1)^2 \int_0^\tau \left(\frac{\partial f_1}{\partial \theta}(\theta_1, t) \right)^2 dt.$$

Consequently, from Theorem 1.A.19 of Prokhorov in [10], p. 372 we obtain that for any $\delta > 0$

$$\lim_{\varepsilon \rightarrow 0} \mathbf{P}_{\theta, \tau} \left\{ \sup_{h_1 \in H_1} |V_\varepsilon(h_1)| > \delta \right\} = 0.$$

Next, from the properties of the stochastic integral we have

$$\int_0^\tau h_1 \frac{\partial f_1}{\partial \theta}(\theta_1, t) dW(t) \stackrel{d}{=} h_1 Z_1 \left(\int_0^\tau \left[\frac{\partial f_1}{\partial \theta}(\theta_1, t) \right]^2 dt \right)^{1/2}.$$

Hence uniformly over $h \in H$ the distribution of the first term in (8) converges to the distribution of $h_1 Z_1 I_1$. Similarly, we can show the weak convergence of the second term of (8) to $h_2 Z_2 I_2$ uniformly over H .

Next, the last term in (8) can be written as

$$\frac{1}{\varepsilon} \int_\tau^{\tau+\varepsilon^2 u} (f_1(\theta_1 + \varepsilon h_1, t) - f_2(\theta_2, t)) dW(t) = \frac{1}{\varepsilon} \int_\tau^{\tau+\varepsilon^2 u} (f_1(\theta_1 + \varepsilon h_1, t) - f_1(\theta_1, \tau)) dW(t) \\ + \frac{1}{\varepsilon} \int_\tau^{\tau+\varepsilon^2 u} (f_1(\theta_1, \tau) - f_2(\theta_2 + \varepsilon h_2, \tau)) dW(t) \\ + \frac{1}{\varepsilon} \int_\tau^{\tau+\varepsilon^2 u} (f_2(\theta_2 + \varepsilon h_2, \tau) - f_2(\theta_2, t)) dW(t).$$

It can be shown, similarly to the proof in [10] that the first and the third terms converge to zero in probability uniformly over $H \times U$. For the second term we have

$$\frac{1}{\varepsilon} \int_\tau^{\tau+\varepsilon^2 u} (f_1(\theta_1, \tau) - f_2(\theta_2 + \varepsilon h_2, \tau)) dW(t) = (f_1(\theta_1, \tau) - f_2(\theta_2 + \varepsilon h_2, \tau)) \frac{1}{\varepsilon} [W(\tau + \varepsilon^2 u) - W(\tau)] \\ \stackrel{d}{=} (f_1(\theta_1, \tau) - f_2(\theta_2 + \varepsilon h_2, \tau)) W_1(u) \xrightarrow{P} \Delta W_1(u).$$

Note that the Wiener process $W_1(u) \stackrel{d}{=} \varepsilon^{-1}[W(\tau + \varepsilon^2 u) - W(\tau)]$ is independent of Z_1 and Z_2 , since three summands in (8) are independent. Thus, combining the estimates for the stochastic and non-stochastic terms, we obtain the convergence of the distribution of $\log Z_{\theta, \tau}^\varepsilon$ for $u > 0$ to the distribution of

$$\begin{aligned} \log Z_{\theta, \tau}^0(h, u) &= \Delta W_1(u) + Z_1 h_1 I_1 + Z_2 h_2 I_2 - \frac{1}{2} \Delta^2 |u| - \frac{1}{2} (h_1^2 I_1^2 + h_2^2 I_2^2) \\ &= \frac{1}{2} (Z_1^2 + Z_2^2) - \frac{1}{2} I_1^2 \left(h_1 - \frac{Z_1}{I_1} \right)^2 - \frac{1}{2} I_2^2 \left(h_2 - \frac{Z_2}{I_2} \right)^2 + \Delta \left(B(u) - \frac{1}{2} \Delta |u| \right) \end{aligned}$$

uniformly over $H \times U$. The similar analysis for $u \leq 0$ yields the statement of the lemma. □

Remark 1. This result can be generalized to the case of multiple change-points.

3. RELATIVE EFFICIENCY OF TWO ESTIMATES OF $\mathcal{L}[f]$

First, recall the problem of estimating the point of discontinuity τ of the discontinuous signal f from observations (1) that was studied by Ibragimov and Hasminskii [10]. For quadratic loss function, they compared asymptotic efficiencies of maximum likelihood and generalized Bayesian estimators of τ . It turned out that asymptotically the ratio of quadratic risks of Bayesian estimate and the MLE of τ does not depend on the function f with the discontinuity point τ and that Bayesian estimator of τ is more efficient than the MLE of τ .

As was mentioned above, if Δ is the jump size at the point τ , then $V(\Delta^2 t)$ is a limiting likelihood ratio process for estimating τ . Denote the MLE and Bayesian estimate of τ by $\hat{\tau}_{\text{mle}}^\varepsilon$ and $\hat{\tau}_{\text{b}}^\varepsilon$, respectively. Let u be a local parameter for τ with the normalization ε^2 . Define

$$\hat{\tau}_{\text{mle}} = \tau + \varepsilon^2 \hat{u}_{\text{mle}}, \quad \hat{\tau}_{\text{b}} = \tau + \varepsilon^2 \hat{u}_{\text{b}},$$

where \hat{u}_{mle} is the point at which the limiting likelihood attains its maximum and \hat{u}_{b} is the generalized Bayesian estimate of u that corresponds to the limiting likelihood. Namely, we have

$$\hat{u}_{\text{mle}} = \arg \max_{t \in \mathbb{R}} V(\Delta^2 t) = \frac{1}{\Delta^2} \arg \max_{t \in \mathbb{R}} V(t), \quad \hat{u}_{\text{b}} = \frac{\int_{\mathbb{R}} t V(\Delta^2 t) dt}{\int_{\mathbb{R}} V(\Delta^2 t) dt} = \frac{1}{\Delta^2} \frac{\int_{\mathbb{R}} t V(t) dt}{\int_{\mathbb{R}} V(t) dt}. \quad (9)$$

Then the asymptotic relative efficiency of $\hat{\tau}_{\text{mle}}^\varepsilon$ and $\hat{\tau}_{\text{b}}^\varepsilon$ coincides with the relative efficiency of the estimates \hat{u}_{mle} and \hat{u}_{b} ,

$$\lim_{\varepsilon \rightarrow 0} \frac{\mathbf{E}_\tau (\hat{\tau}_{\text{b}}^\varepsilon - \tau)^2}{\mathbf{E}_\tau (\hat{\tau}_{\text{mle}}^\varepsilon - \tau)^2} = \lim_{\varepsilon \rightarrow 0} \frac{\mathbf{E}_\tau (\hat{\tau}_{\text{b}} - \tau)^2}{\mathbf{E}_\tau (\hat{\tau}_{\text{mle}} - \tau)^2} = \frac{\mathbf{E} \hat{u}_{\text{b}}^2}{\mathbf{E} \hat{u}_{\text{mle}}^2} \equiv \kappa_0. \quad (10)$$

Ibragimov and Hasminski [10] showed that $\mathbf{E} \hat{u}_{\text{mle}}^2 = 26/\Delta^4$, but they stated that $\mathbf{E} \hat{u}_{\text{b}}^2$ is hard to evaluate explicitly. Using computational methods, they obtained the following approximate value $\Delta^4 \mathbf{E} \hat{u}_{\text{b}}^2 = 19.5 \pm 0.5$, and the efficiency $\kappa_0 \approx 0.73 \pm 0.03$. Later Rubin and Song [13] obtained the exact value of $\mathbf{E} \hat{u}_{\text{b}}^2$ and the asymptotic relative efficiency of two estimates,

$$\kappa_0 = \frac{8}{13} \zeta(3) \approx 0.7397,$$

which appears to be very close to the approximate value found in [10]. Here ζ is Riemann's zeta function defined as $\zeta(s) = \sum_{n=1}^\infty n^{-s}$.

We are interested in estimating the smooth functional $\mathcal{L}[f^\tau(\theta, \cdot)]$. We will compare Bayesian and maximum likelihood estimates of $\mathcal{L}[f^\tau(\theta, \cdot)]$. In fact, the problem is reduced to estimating the parameters $\theta = (\theta_1, \theta_2)$ and τ .

We can write an MLE $\hat{\mathcal{L}}_{\text{mle}}^\varepsilon$ of $\mathcal{L}[f^\tau(\theta, \cdot)]$ in terms of the local parameters h and u as

$$\hat{\mathcal{L}}_{\text{mle}}^\varepsilon = \mathcal{L}[f^{\tau + \varepsilon^2 \hat{u}^\varepsilon}(\theta + \varepsilon \hat{h}^\varepsilon, \cdot)],$$

where

$$(\hat{h}^\varepsilon, \hat{u}^\varepsilon) = \arg \max_{(h, u) \in \Theta_\varepsilon \times T_\varepsilon} Z_{\theta, \tau}^\varepsilon(h, u).$$

Define an MLE that corresponds to the limiting likelihood obtained in Lemma 1,

$$\widehat{\mathcal{L}}_{\text{mle}}^0 = \mathcal{L}[f^{\tau+\varepsilon^2\widehat{u}}(\theta + \varepsilon\widehat{h}, \cdot)], \quad (11)$$

where $(\widehat{h}, \widehat{u}) \equiv (\widehat{h}_1, \widehat{h}_2, \widehat{u})$ is the point at which the limiting log-likelihood process (6) attains its maximum. More precisely,

$$\widehat{h}_i = \arg \max_{h_i \in \mathbb{R}} \left\{ \frac{1}{2} Z_i^2 - \frac{1}{2} I_i^2 \left(h_i - \frac{Z_i}{I_i} \right)^2 \right\}, \quad i = 1, 2, \quad (12)$$

$$\widehat{u} = \arg \max_{u \in \mathbb{R}} V(\Delta^2 u) \equiv \arg \max_{u \in \mathbb{R}} \left\{ |\Delta| \left(B(u) - \frac{|\Delta u|}{2} \right) \right\}. \quad (13)$$

Obviously, $\mathbf{E}\widehat{u} = 0$ and $\mathbf{E}\widehat{h}_i = 0$, $\mathbf{E}\widehat{h}_i^2 = 1/I_i^2$, $i = 1, 2$. We also know that $\mathbf{E}\widehat{u}^2 = 26/\Delta^4$.

Let $\widehat{\mathcal{L}}_{\text{b}}^\varepsilon$ be a Bayesian estimate of $\mathcal{L}[f^\tau(\theta, \cdot)]$ that is defined by

$$\widehat{\mathcal{L}}_{\text{b}}^\varepsilon = \arg \min_A \int_{T_\varepsilon} \int_{\Theta_\varepsilon} (A - \mathcal{L}[f^{\tau+\varepsilon^2 u}(\theta + \varepsilon h, \cdot)])^2 Z_{\theta, \tau}^\varepsilon(h, u) dh_1 dh_2 du.$$

Define the generalized Bayesian estimate corresponding to the limiting likelihood process as

$$\widehat{\mathcal{L}}_{\text{b}}^0 = \arg \min_A \int_{\mathbb{R}} \int_{\mathbb{R}^2} (A - \mathcal{L}[f^{\tau+\varepsilon^2 u}(\theta + \varepsilon h, \cdot)])^2 Z_{\theta, \tau}^0(h, u) dh_1 dh_2 du. \quad (14)$$

Following the approach developed in [10] we can prove the following result.

Lemma 2. *Let Conditions F and L (L') be satisfied for all $\theta \in \Theta$, $\tau \in T$. Then uniformly over $\Theta \times T$*

$$\lim_{\varepsilon \rightarrow 0} \frac{\mathbf{E}_{\theta, \tau}(\widehat{\mathcal{L}}_{\text{mle}}^\varepsilon - \mathcal{L}[f^\tau(\theta, \cdot)])^2}{\mathbf{E}_{\theta, \tau}(\widehat{\mathcal{L}}_{\text{mle}}^0 - \mathcal{L}[f^\tau(\theta, \cdot)])^2} = 1 \quad (15)$$

and

$$\lim_{\varepsilon \rightarrow 0} \frac{\mathbf{E}_{\theta, \tau}(\widehat{\mathcal{L}}_{\text{b}}^\varepsilon - \mathcal{L}[f^\tau(\theta, \cdot)])^2}{\mathbf{E}_{\theta, \tau}(\widehat{\mathcal{L}}_{\text{b}}^0 - \mathcal{L}[f^\tau(\theta, \cdot)])^2} = 1. \quad (16)$$

Proof. The proof follows from Theorems 1.10.1 and 1.10.2 of [10] and the continuity conditions F and L (L'). We do not go into details here and just give a brief overview of the conditions of the two theorems. Lemma 1 guarantees the convergence of the likelihood ratio on compact sets such that Condition 2 of Theorems 1.10.1 and 1.10.2 is satisfied.

Next, by using the technique developed in [10] it can be shown that for $\theta \in \Theta$, $\tau \in T$, $h', h'' \in \Theta_\varepsilon$, $u', u'' \in T_\varepsilon$, and $\varepsilon < \varepsilon_0$

$$\begin{aligned} \mathbf{E}_{\theta, \tau} | [Z_{\theta, \tau}^\varepsilon(h', u')]^{1/4} - [Z_{\theta, \tau}^\varepsilon(h'', u'')]^{1/4} |^4 &\leq \frac{3}{64} \frac{1}{\varepsilon^2} \| f^{\tau+\varepsilon^2 u'}(\theta + \varepsilon h', \cdot) - f^{\tau+\varepsilon^2 u''}(\theta + \varepsilon h'', \cdot) \|_{\mathbf{L}_2[0,1]}^4 \\ &\leq \frac{3}{64} (C_1 \|h' - h''\|^2 + C_2 |u' - u''|)^2. \end{aligned}$$

It follows that for $\varepsilon < \varepsilon_0$

$$\mathbf{P}_{\theta, \tau} \left\{ Z_{\theta, \tau}^\varepsilon(h, u) > \exp \left(-\frac{1}{8} (C_1 \|h\|^2 + C_2 |\Delta||u|) \right) \right\} \leq \exp \left(-\frac{1}{16} (C_1 \|h\|^2 + C_2 |\Delta||u|) \right).$$

These two relations form Condition 1 of Theorems 1.10.1 and 1.10.2 that is necessary for consistency of the estimates.

Finally, Condition 3 of the theorems is on the unique maximum of the limiting likelihood process (6) in case of an MLE of \mathcal{L} and on the unique minimum in (14). From the results in [10] on the properties of the process $V(u)$ and concavity of $\log \widehat{Z}_{\theta, \tau}^0(h, u)$ in h it follows that these maximal (minimal) values are unique with probability 1. \square

Lemma 3. *Let $\widehat{\mathcal{L}}_{\text{mle}}^\varepsilon$ be an MLE and $\widehat{\mathcal{L}}_{\text{b}}^\varepsilon$ be a generalized Bayesian estimate of $\mathcal{L}[f^\tau(\theta, \cdot)]$, respectively. Assume that Conditions L and F are satisfied. Then, asymptotically as $\varepsilon \rightarrow 0$, the asymptotic quadratic risks of both estimators have the same first order term,*

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \mathbf{E}_{\theta, \tau} \varepsilon^{-2} (\widehat{\mathcal{L}}_{\text{mle}}^\varepsilon - \mathcal{L}[f^\tau(\theta, \cdot)])^2 &= \lim_{\varepsilon \rightarrow 0} \mathbf{E}_{\theta, \tau} \varepsilon^{-2} (\widehat{\mathcal{L}}_{\text{b}}^\varepsilon - \mathcal{L}[f^\tau(\theta, \cdot)])^2 \\ &= \frac{1}{I_1^2} \left(\frac{\partial \mathcal{L}}{\partial \theta_1} [f^\tau(\theta, \cdot)] \right)^2 + \frac{1}{I_2^2} \left(\frac{\partial \mathcal{L}}{\partial \theta_2} [f^\tau(\theta, \cdot)] \right)^2. \end{aligned}$$

Proof. From Lemma 2 it follows that the risks (15) and (16) of $\widehat{\mathcal{L}}_{\text{mle}}^\varepsilon$ and $\widehat{\mathcal{L}}_{\text{b}}^\varepsilon$ have the same asymptotic behavior as the risks of $\widehat{\mathcal{L}}_{\text{mle}}^0$ and $\widehat{\mathcal{L}}_{\text{b}}^0$. Thus we have to calculate the risks of the limiting estimates $\widehat{\mathcal{L}}_{\text{mle}}^0$ and $\widehat{\mathcal{L}}_{\text{b}}^0$ defined in (11) and (14), respectively.

Recall that $\theta = (\theta_1, \theta_2)$ and τ are fixed and the increments $h = (h_1, h_2)$ and u of θ and τ belong to some compact sets, $h \in H \subset \Theta_\varepsilon$, $u \in U \subset T_\varepsilon$.

Let us first show that the following Taylor series expansion holds true

$$\mathcal{L}[f^{\tau+\varepsilon^2 u}(\theta + \varepsilon h, \cdot)] = \mathcal{L}[f^\tau(\theta, \cdot)] + \frac{\partial \mathcal{L}}{\partial \theta_1} [f^\tau(\theta, \cdot)] \varepsilon h_1 + \frac{\partial \mathcal{L}}{\partial \theta_2} [f^\tau(\theta, \cdot)] \varepsilon h_2 + o(\varepsilon), \quad \varepsilon \rightarrow 0. \quad (17)$$

Since \mathcal{L} is Fréchet differentiable, for $f(t) \equiv f^\tau(\theta, t) \in \mathbf{L}_2[0, 1]$ there exists a linear mapping $\Lambda_{\theta, \tau}: \mathbf{L}_2[0, 1] \rightarrow \mathbb{R}$ such that

$$\mathcal{L}[f^{\tau+\varepsilon^2 u}(\theta + \varepsilon h, \cdot)] - \mathcal{L}[f^\tau(\theta, \cdot)] = \Lambda_{\theta, \tau}[\Delta f] + r[\Delta f] \|\Delta f\|_{\mathbf{L}_2[0, 1]}, \quad (18)$$

where

$$\Delta f(t) = f^{\tau+\varepsilon^2 u}(\theta + \varepsilon h, t) - f^\tau(\theta, t), \quad u \in U, \quad h \in H,$$

and $\lim_{\|\Delta f\|_{\mathbf{L}_2[0, 1]} \rightarrow 0} |r[\Delta f]| = 0$ for the remainder term $r: \mathbf{L}_2[0, 1] \rightarrow \mathbb{R}$. Set

$$\Delta f(t) = (f^{\tau+\varepsilon^2 u}(\theta + \varepsilon h, t) - f^\tau(\theta + \varepsilon h, t)) + (f^\tau(\theta + \varepsilon h, t) - f^\tau(\theta, t)) \equiv \Delta_\theta f(t) + \Delta_\tau f(t).$$

Since H and U are compact sets and f is continuous in t everywhere except $t = \tau$ (see Conditions F[a-b]), we have

$$\Delta_\theta f(t) = f^{\tau+\varepsilon^2 u}(\theta + \varepsilon h, t) - f^\tau(\theta + \varepsilon h, t) = r_0^\varepsilon(\theta, \tau, h, u, t),$$

where

$$\lim_{\varepsilon \rightarrow 0} \sup_{h \in H, u \in U} \|r_0^\varepsilon(\theta, \tau, h, u, \cdot)\|_{\mathbf{L}_2[0, 1]} = 0.$$

Indeed, we have for $t \neq \tau$

$$\Delta_\theta f(t) = (f_1(\theta_1 + \varepsilon h_1, t) - f_2(\theta_2 + \varepsilon h_2, t)) \mathbf{1}\{|t - \tau| < \varepsilon^2 u\} (\mathbf{1}\{u > 0\} - \mathbf{1}\{u < 0\}),$$

where the functions f_1 and f_2 are continuous on $[0, 1]$.

From Condition F[c] it follows that for $t \neq \tau$

$$\begin{aligned} \Delta_\tau f(t) &= f^\tau(\theta_1 + \varepsilon h_1, \theta_2 + \varepsilon h_2, t) - f^\tau(\theta_1, \theta_2, t) \\ &= \frac{\partial f^\tau(\theta, t)}{\partial \theta_1} \varepsilon h_1 + \frac{\partial f^\tau(\theta, t)}{\partial \theta_2} \varepsilon h_2 + r_1(\varepsilon h_1, t) \varepsilon |h_1| \mathbf{1}\{t < \tau\} + r_2(\varepsilon h_2, t) \varepsilon |h_2| \mathbf{1}\{t > \tau\}, \end{aligned}$$

where $\lim_{\varepsilon \rightarrow 0} \sup_{h_i \in H_i} \|r_i(\varepsilon h_i, \cdot)\|_{\mathbf{L}_2[0, 1]} = 0$. Combining two formulas for $\Delta_\tau f(t)$ and $\Delta_\theta f(t)$ we obtain

$$\Delta f(t) = \frac{\partial f^\tau(\theta_1, t)}{\partial \theta_1} \varepsilon h_1 + \frac{\partial f^\tau(\theta_2, t)}{\partial \theta_2} \varepsilon h_2 + r^\varepsilon(\theta, \tau, h, u, t),$$

where

$$\sup_{h \in H, u \in U} \|r^\varepsilon(\theta, \tau, h, u, \cdot)\|_{\mathbf{L}_2[0, 1]} \rightarrow 0, \quad \varepsilon \rightarrow 0.$$

Thus

$$\|\Delta f\|_{\mathbf{L}_2[0,1]} \leq \|\Delta_\theta f\|_{\mathbf{L}_2[0,1]} + \|\Delta_\tau f\|_{\mathbf{L}_2[0,1]} \rightarrow 0 \quad \text{as } \varepsilon \rightarrow 0$$

uniformly over h and u and, consequently, $|r[\Delta f]| = o(\varepsilon)$ in (18).

Substituting the obtained expansion in (18) gives the desired formula (17), where the linear mappings are defined as $\frac{\partial \mathcal{L}}{\partial \theta_i}[f^\tau(\theta, \cdot)] = \frac{\partial f^\tau(\theta, \cdot)}{\partial \theta_i} \circ \Lambda_{\theta, \tau}: \Theta_i \rightarrow \mathbb{R}$.

First, we will calculate the asymptotic mean-square error of $\widehat{\mathcal{L}}_{\text{mle}}^0$ defined in (11). From formula (17) and independence of \widehat{h}_1 and \widehat{h}_2 we obtain

$$\begin{aligned} \mathbf{E}_{\theta, \tau}(\widehat{\mathcal{L}}_{\text{mle}}^0 - \mathcal{L}[f^\tau(\theta, \cdot)])^2 &= \varepsilon^2 \left(\frac{\partial \mathcal{L}}{\partial \theta_1}[f^\tau(\theta, \cdot)] \right)^2 \mathbf{E}_{\theta, \tau} \widehat{h}_1^2 + \varepsilon^2 \left(\frac{\partial \mathcal{L}}{\partial \theta_2}[f^\tau(\theta, \cdot)] \right)^2 \mathbf{E}_{\theta, \tau} \widehat{h}_2^2 + o(\varepsilon^2) \\ &= \frac{\varepsilon^2}{I_1^2} \left(\frac{\partial \mathcal{L}}{\partial \theta_1}[f^\tau(\theta, \cdot)] \right)^2 + \frac{\varepsilon^2}{I_2^2} \left(\frac{\partial \mathcal{L}}{\partial \theta_2}[f^\tau(\theta, \cdot)] \right)^2 + o(\varepsilon^2), \end{aligned}$$

since $\mathbf{E}_{\theta, \tau} \widehat{h}_i^2 = 1/I_i^2$ and $\mathbf{E}_{\theta, \tau} \widehat{h}_i = 0$.

Now we calculate the risk of the Bayesian estimate $\widehat{\mathcal{L}}_{\text{b}}^0$ defined in (14). Using the Taylor series expansion (17) and formula (6) we have

$$\begin{aligned} \widehat{\mathcal{L}}_{\text{b}}^0 &= \frac{\int_{\mathbb{R}} \int_{\mathbb{R}^2} \mathcal{L}[f^{\tau+\varepsilon^2 u}(\theta + \varepsilon h, \cdot)] \widetilde{Z}_{\theta, \tau}^0(h, u) dh_1 dh_2 du}{\int_{\mathbb{R}} \int_{\mathbb{R}^2} \widetilde{Z}_{\theta, \tau}^0(h, u) dh_1 dh_2 du} \\ &= \mathcal{L}[f^\tau(\theta, \cdot)] + \varepsilon \frac{\partial \mathcal{L}}{\partial \theta_1}[f^\tau(\theta, \cdot)] \frac{Z_1}{I_1} + \varepsilon \frac{\partial \mathcal{L}}{\partial \theta_2}[f^\tau(\theta, \cdot)] \frac{Z_2}{I_2} + o(\varepsilon), \end{aligned}$$

where Z_i 's are independent $\mathcal{N}(0, 1)$. Calculating the mean square risk gives exactly the same asymptotic behavior as the one for the risk of $\widehat{\mathcal{L}}_{\text{mle}}^0$. \square

Remark 2. If Condition L' is satisfied and \mathcal{L} is twice differentiable w.r.t. θ_i , we can calculate the second order terms of the asymptotic risks applying the facts that

$$\mathbf{E} \widehat{u}_{\text{b}} = \mathbf{E} \widehat{u}_{\text{mle}} = 0, \quad \mathbf{E} \widehat{u}_{\text{mle}}^2 = 26/\Delta^4, \quad \mathbf{E} \widehat{u}_{\text{b}}^2 = 16\zeta(3)/\Delta^4.$$

We have

$$\begin{aligned} \mathbf{E}_{\theta, \tau}(\widehat{\mathcal{L}}_{\text{mle}}^\varepsilon - \mathcal{L}[f^\tau(\theta, \cdot)])^2 &= \varepsilon^2 \sum_{i=1,2} \frac{1}{I_i^2} \left(\frac{\partial \mathcal{L}}{\partial \theta_i}[f^\tau(\theta, \cdot)] \right)^2 + \varepsilon^4 \frac{26}{\Delta^4} \left(\frac{\partial \mathcal{L}}{\partial \tau}[f^\tau(\theta, \cdot)] \right)^2 \\ &\quad + \varepsilon^4 \sum_{i=1,2} \frac{3}{I_i^4} \left(\frac{\partial^2 \mathcal{L}}{\partial \theta_i^2}[f^\tau(\theta, \cdot)] \right)^2 + o(\varepsilon^4), \\ \mathbf{E}_{\theta, \tau}(\widehat{\mathcal{L}}_{\text{b}}^\varepsilon - \mathcal{L}[f^\tau(\theta, \cdot)])^2 &= \varepsilon^2 \sum_{i=1,2} \frac{1}{I_i^2} \left(\frac{\partial \mathcal{L}}{\partial \theta_i}[f^\tau(\theta, \cdot)] \right)^2 + \varepsilon^4 \frac{16\zeta(3)}{\Delta^4} \left(\frac{\partial \mathcal{L}}{\partial \tau}[f^\tau(\theta, \cdot)] \right)^2 \\ &\quad + \varepsilon^4 \sum_{i=1,2} \frac{I_i^4 + 2I_i^2 + 3}{I_i^4} \left(\frac{\partial^2 \mathcal{L}}{\partial \theta_i^2}[f^\tau(\theta, \cdot)] \right)^2 + o(\varepsilon^4). \end{aligned}$$

If Condition L' is satisfied and $\mathcal{L}[f^\tau(\theta, \cdot)]$ is a function of τ only, $\mathcal{L}[f^\tau(\theta, \cdot)] \equiv g(\tau)$, we obtain the result (10) of Ibragimov and Hasminski as a corollary:

$$\lim_{\varepsilon \rightarrow 0} \frac{\mathbf{E}_{\theta, \tau}(\widehat{\mathcal{L}}_{\text{b}}^\varepsilon - \mathcal{L}[f^\tau(\theta, \cdot)])^2}{\mathbf{E}_{\theta, \tau}(\widehat{\mathcal{L}}_{\text{mle}}^\varepsilon - \mathcal{L}[f^\tau(\theta, \cdot)])^2} = \lim_{\varepsilon \rightarrow 0} \frac{\mathbf{E}_{\theta, \tau}(\widehat{\mathcal{L}}_{\text{b}}^\varepsilon - g(\tau))^2}{\mathbf{E}_{\theta, \tau}(\widehat{\mathcal{L}}_{\text{mle}}^\varepsilon - g(\tau))^2} = \frac{8}{13} \zeta(3).$$

4. ESTIMATION IN THE SEQUENCE MODEL

In this section we give explicit estimates for the problem of estimating a smooth functional in the equivalent sequence model. We assume that a very simple signal is observed, which is constant up to some moment of time τ and equals zero afterwards.

We observe the vector $X = (X_1, \dots, X_n)$, where X_i 's are Gaussian random variables with distribution $\mathbf{P}_{\theta_i, \tau}$ defined on the probability space $(\mathcal{X}, \mathcal{B}_i, \mathbf{P}_{\theta_i, \tau})$,

$$X_i = \theta_i + \varepsilon \xi_i, \quad i = 1, \dots, n. \tag{19}$$

The signal is constant up to some moment of time τ , $\theta_i = \theta \mathbf{1}(i - \tau \leq 0)$. The parameters $\tau \in \{1, \dots, n\}$ and θ are unknown, ξ_i 's are i.i.d. $\mathcal{N}(0, 1)$, $\varepsilon > 0$ is known. The goal is to estimate a smooth function $L(\theta, \tau)$ of the signal.

4.1. Overview

The change-point problem for the change in mean in Bayesian set-up was considered by Chernoff and Zacks [6]. They studied the change in mean for a sequence of Gaussian r.v.'s and obtained a Bayes estimate of the difference in means before and after the change. This estimate was obtained under the assumption that the current mean and the jump size have normal prior distributions. The change-point was also assumed to be random with some arbitrary discrete prior. In their paper, Chernoff and Zacks compare Bayesian estimates of the current mean and the minimum variance linear unbiased estimates (MVLUE) when the signal-to-noise ratio is greater than 2. In particular, they state that if there is exactly one change in the observations, the Bayes estimator is very efficient if the change takes place at the beginning of the sequence and it loses its efficiency when the change takes place very close to the last observation. In the case of at most one change Bayesian procedures are always better than MVLUE. Rubin [12] considered the change-point problem in the context of estimating discontinuities in multivariate densities. He discussed Bayesian and maximum likelihood approach to the problem and mentioned that the Bayes procedure with the uniform prior distribution on the unknown parameter τ gives more efficient estimates than the maximum likelihood approach.

A maximum likelihood estimate (MLE) of a change-point was first obtained by Hinkley [9] in the problem of estimating a moment of the change in mean of Gaussian data under the assumption that the jump size is small. Later an asymptotic distribution of the MLE of a change-point for the case of close normal means was derived by Bhattacharya and Brockwell [3]. Their result was generalized by Bhattacharya [2] to the case of a small jump size in a multidimensional parameter. Following [3] and [2] Ferger [8] proposed a class of estimates for a change-point based on U -statistics for the case of small disorders in the distribution. Later Brodskii and Darhovskii [4] studied the asymptotic behavior of an estimate for the change-point in Gaussian sequence with unknown mean without the assumption that the difference between the means (the jump size) tends to zero. They proposed a family of estimates for the change-point based on the Kolmogorov–Smirnov statistics which includes an MLE. An asymptotic distribution of these estimates was derived and the corresponding testing problem was considered.

Let ν^n be a σ -finite measure on the σ -algebra $\mathcal{B} = \mathcal{B}_1 \times \dots \times \mathcal{B}_n$ and $\mathbf{P}_{\theta, \tau}^n = \mathbf{P}_{\theta_1, \tau} \times \dots \times \mathbf{P}_{\theta_n, \tau}$. Then the joint density of X (likelihood) is given by

$$\frac{d\mathbf{P}_{\theta, \tau}^n}{d\nu^n}(X) \equiv p_n^\varepsilon(X; \theta, \tau) = (2\pi\varepsilon^2)^{-n/2} \exp \left\{ -\frac{1}{2\varepsilon^2} \left(\sum_{i=1}^{\tau} (X_i - \theta)^2 + \sum_{i=\tau+1}^n X_i^2 \right) \right\}. \tag{20}$$

Assuming that $\varepsilon = 1$ in model (19), Bhattacharya and Brockwell [3] derived a limiting process for the likelihood ratio under the conditions that the parameter θ is small, $\theta = \delta\nu_n^{-1}$, where $\nu_n \rightarrow \infty$ slower than $n^{1/2}$ and the length of the observed sequence $n \rightarrow \infty$. According to their result, as $n \rightarrow \infty$, the following weak convergence holds with respect to uniform convergence on compact sets,

$$\log \frac{d\mathbf{P}_{\theta+n^{-1/2}h, \tau+\nu_n^2u}}{d\mathbf{P}_{\theta, \tau}} \xrightarrow{w} \frac{Z^2}{2} - \frac{1}{2}\lambda \left(h - \frac{Z}{\sqrt{\lambda}} \right)^2 + |\delta| \left(B(u) - \frac{1}{2}|\delta||u| \right), \quad n \rightarrow \infty,$$

where Z is $\mathcal{N}(0, 1)$, $B(u)$ is a two-sided Wiener process (5) independent of Z , and $\lambda = \lim_{n \rightarrow \infty} \tau/n$.

Remark 3. Some general results on the behavior of the log-likelihood ratio and of the maximum likelihood estimate of the change point τ can be found in Section 1.6 of [7]. In particular, Theorems 1.6.2 and 1.6.3 of [7] state that the asymptotic distribution as $n \rightarrow \infty$ of the likelihood ratio in the situation of a decreasing size of the change in means obtained in [3] differs from the one in the situation of a fixed change in mean.

In our case the number of observations n is fixed and the change in mean (the jump size at τ) θ is fixed. Let $\theta \in \Theta$, where Θ is an open subset of \mathbb{R} and $\tau = [n\alpha]$, where $\alpha \in A = (n^{-1}, 1)$. Define the sets $\Theta_\varepsilon = \varepsilon^{-1}(\Theta - \theta)$ and $A_\varepsilon = \varepsilon^{-2}(A - \alpha)$. The following lemma that is given without proof describes the asymptotic behavior of the likelihood ratio as $\varepsilon \rightarrow 0$.

Lemma 4. *Let $X = (X_1, \dots, X_n)$ be given by (19) and H, V be some compact subsets of Θ_ε and A_ε , respectively. Then the following weak convergence holds uniformly over $(h, v) \in H \times V$ as $\varepsilon \rightarrow 0$,*

$$\log \frac{d\mathbf{P}_{\theta+\varepsilon h, \tau+\varepsilon^2 n v}}{d\mathbf{P}_{\theta, \tau}}(X) \xrightarrow{w} \frac{Z^2}{2} - \frac{\tau}{2} \left(h - \frac{Z}{\sqrt{\tau}} \right)^2 + |\theta| \sqrt{n} \left(B(v) - \frac{1}{2} \sqrt{n} |\theta| |v| \right), \quad \varepsilon \rightarrow 0, \quad (21)$$

where Z is $\mathcal{N}(0, 1)$ independent of the two-sided Brownian motion $B(v)$.

4.2. MLE and Bayesian Estimate of $L(\theta, \tau)$

Let us find an MLE of $L(\theta, \tau)$. The log-likelihood $\log p_n^\varepsilon(X; \theta, \tau)$ satisfies

$$\begin{aligned} \log p_n^\varepsilon(X; \theta, \tau) + \frac{n}{2} \log(2\pi\varepsilon^2) &= - \sum_{i=1}^{\tau} \frac{(X_i - \theta)^2}{2\varepsilon^2} - \sum_{i=\tau+1}^n \frac{X_i^2}{2\varepsilon^2} \\ &= - \frac{1}{2\varepsilon^2} \sum_{i=1}^n X_i^2 + \frac{\theta}{\varepsilon^2} \sum_{i=1}^{\tau} X_i - \frac{\theta^2 \tau}{2\varepsilon^2}. \end{aligned}$$

First, we maximize the log-likelihood with respect to θ and replace θ by its conditional MLE $\bar{X}_\tau = \frac{1}{\tau} \sum_{i=1}^{\tau} X_i$. Next, maximizing the obtained log-likelihood with respect to τ we obtain an MLE of the change-point τ ,

$$\hat{\tau}_{\text{mle}}^\varepsilon = \arg \max_{1 \leq k \leq n} \left\{ \frac{1}{2\varepsilon^2 k} \left(\sum_{i=1}^k X_i \right)^2 \right\} = \arg \max_{1 \leq k \leq n} U_k, \quad (22)$$

where

$$U_k = \frac{1}{2\varepsilon^2 k} \left(\sum_{i=1}^k X_i \right)^2. \quad (23)$$

The estimate (22) of τ for unknown change in mean θ of normal distribution was first obtained by Hinkley in [9]. An asymptotic distribution as $\varepsilon \rightarrow 0$ of this estimate was derived by Brodskii and Darkhovskii in [4].

Finally, an MLE of $L(\theta, \tau)$ is given by

$$\hat{L}_{\text{mle}}^\varepsilon = L(\hat{\theta}_{\text{mle}}, \hat{\tau}_{\text{mle}}) = L(\bar{X}_{\hat{\tau}_{\text{mle}}}, \hat{\tau}_{\text{mle}}).$$

For example, if $L(\theta, \tau) = \sum_{i=1}^{\tau} \theta_i = \theta\tau$, then $\hat{L}_{\text{mle}}^\varepsilon = \sum_{i=1}^{\hat{\tau}_{\text{mle}}} X_i$.

Remark 4. Note that using (19) we can write

$$\hat{\tau}_{\text{mle}} = \arg \max_{1 \leq k \leq n} \left\{ \left| \frac{\varepsilon}{\sqrt{k}} \sum_{i=1}^k \xi_i + \theta \sqrt{k} \mathbf{1}\{k \leq \tau\} + \frac{\theta\tau}{\sqrt{k}} \mathbf{1}\{k > \tau\} \right| \right\}.$$

Thus, to find a non-asymptotic risk of \hat{L}_{mle} we need to calculate the joint distribution of $\hat{\tau}_{\text{mle}}$ and $\sum_{i=1}^{\hat{\tau}_{\text{mle}}} \xi_i$. This problem is similar to calculation of the joint distribution of

$$\hat{\tau} = \arg \max_{t > 0} \left| \theta \sqrt{t} \mathbf{1}\{t \leq \tau\} + \frac{\theta\tau}{\sqrt{t}} \mathbf{1}\{t > \tau\} + \varepsilon \frac{W(t)}{\sqrt{t}} \right| \quad \text{and} \quad W(\hat{\tau}).$$

Let us now find a Bayesian estimate of $L(\theta, \tau)$. Assume that θ has a non-informative prior distribution $\mathcal{N}(0, \sigma^2)$, where $\sigma^2 \rightarrow \infty$ and τ is uniformly distributed on the set $\{1, \dots, n\}$. Then the generalized posterior density of (θ, τ) is given by

$$\pi(\theta, \tau|X) = \frac{1}{\sqrt{2\pi\varepsilon^2}} \frac{\exp\left(-\frac{\tau}{2\varepsilon^2}(\theta - \bar{X}_\tau)^2 + U_\tau\right)}{\sum_{k=1}^n e^{U_k}/\sqrt{k}}, \tag{24}$$

where U_k is defined in (23). Chernoff and Zacks [6] used Bayesian approach with normal priors on θ and uniform prior on τ to obtain an estimate of the mean of the observations after the change. In [11] under the same assumptions the posterior distribution (24) in the change-point problem for normal observations was calculated.

Thus, we obtain the following Bayesian estimate of $L(\theta, \tau)$,

$$\hat{L}_b = \sum_{\tau=1}^n \int_{\mathbb{R}} L(\theta, \tau) \pi(\theta, \tau|X) d\theta = \frac{1}{\sqrt{2\pi\varepsilon^2}} \sum_{\tau=1}^n p_\tau \sqrt{\tau} \int_{\mathbb{R}} L(\theta, \tau) \exp\left(-\frac{\tau}{2\varepsilon^2}(\theta - \bar{X}_\tau)^2\right) d\theta, \tag{25}$$

where

$$p_k = \frac{e^{U_k}}{\sqrt{k}} \left(\sum_{i=1}^n \frac{e^{U_i}}{\sqrt{i}} \right)^{-1}. \tag{26}$$

For example, if $L(\theta, \tau) = \theta\tau = \sum_{i=1}^\tau \theta_i$, then \hat{L}_b is a weighted sum of X_i 's with weights p_k

$$\hat{L}_b^\varepsilon = \sum_{k=1}^n p_k \sum_{i=1}^k X_i \equiv \frac{\sum_{k=1}^n (e^{U_k}/\sqrt{k}) \sum_{i=1}^k X_i}{\sum_{k=1}^n e^{U_k}/\sqrt{k}}.$$

Note that the Bayesian estimate of τ for quadratic loss function is given by

$$\hat{\tau}_b^\varepsilon = \sum_{k=1}^n k p_k \equiv \sum_{k=1}^n \sqrt{k} e^{U_k} \left(\sum_{k=1}^n \frac{e^{U_k}}{\sqrt{k}} \right)^{-1}$$

and the corresponding estimate for θ is

$$\hat{\theta}_b^\varepsilon = \sum_{k=1}^n p_k \bar{X}_k.$$

5. SIMULATION STUDY

We studied the quadratic risks of Bayesian and maximum likelihood estimates of

$$L(\theta, \tau) = \sum_{i=1}^n \theta_i \equiv \theta\tau.$$

10^4 simulations were made for $n = 20$ observations in model (19) with the values of $\theta \in \{0.5, 1, 1.5, 2\}$, for the change-points $\tau = 3, 4, \dots, 17, 18$, and the noise level $\varepsilon = 1$.

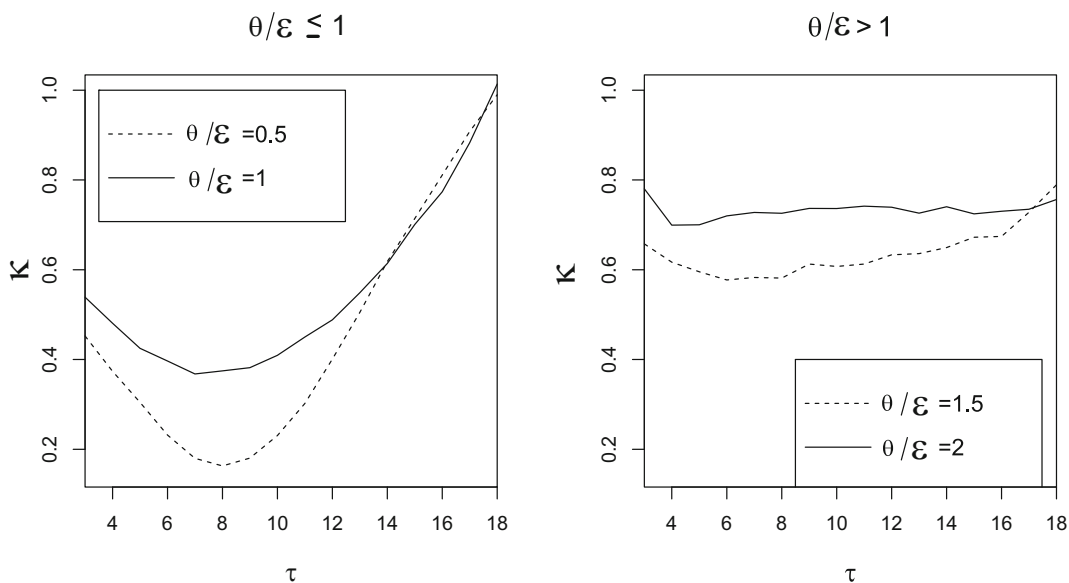
First, introduce the following notation for risk ratios,

$$\kappa(\tau, \theta/\varepsilon) = \frac{\mathbf{E}_{\theta, \tau}(\hat{\tau}_b^\varepsilon - \tau)^2}{\mathbf{E}_{\theta, \tau}(\hat{\tau}_{mle}^\varepsilon - \tau)^2}, \quad \tilde{\kappa}(\tau, \theta/\varepsilon) = \frac{\mathbf{E}_{\theta, \tau}(\hat{L}_b^\varepsilon - L)^2}{\mathbf{E}_{\theta, \tau}(\hat{L}_{mle}^\varepsilon - L)^2}.$$

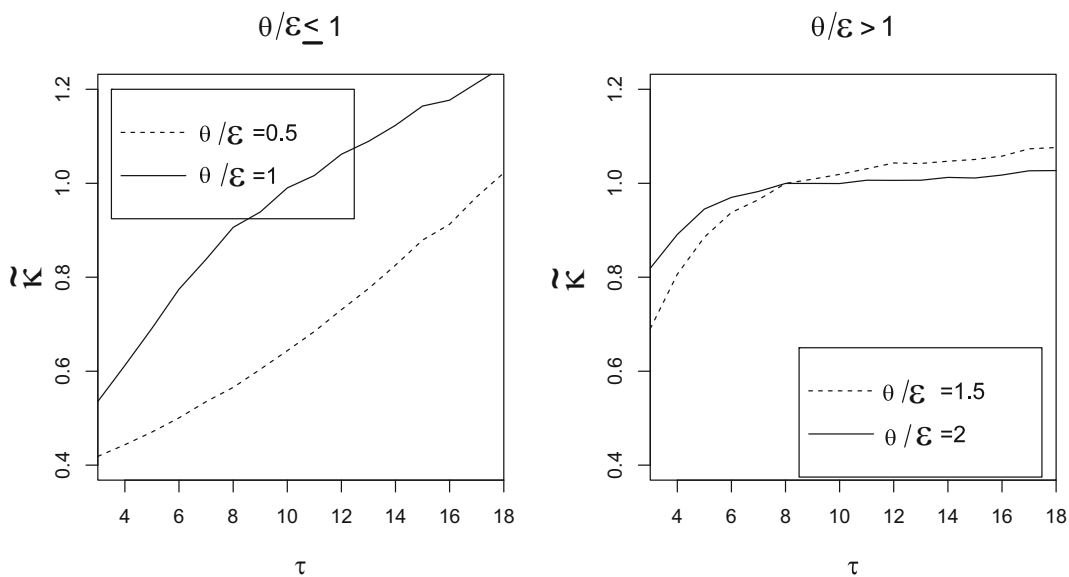
In Fig. 1 the graphs of the empirical risk ratios $\kappa = \kappa(\tau, \theta/\varepsilon)$ and $\tilde{\kappa} = \tilde{\kappa}(\tau, \theta/\varepsilon)$ depending on $\tau = 3, 4, \dots, 17, 18$ are presented for different values of the signal-to-noise ratio (SNR), $\theta/\varepsilon = 0.5, 1, 1.5, 2$. For simplicity we assume that the signal θ is positive.

Recall that asymptotically as $\varepsilon \rightarrow 0$ the relative efficiency of the MLE of τ with respect to the Bayes estimate of τ is about 0.74, $\lim_{\varepsilon \rightarrow 0} \kappa(\tau, \theta/\varepsilon) = \kappa_0 \approx 0.7397$. It means that the MLE of τ is about 17% less efficient than Bayesian estimate $\hat{\tau}_b$ if the SNR θ/ε is large.

The examination of our numerical results leads to the following conclusions.



(a) Graphs of risk ratio $\kappa(\tau, \theta/\varepsilon) = \frac{\mathbf{E}_{\tau}(\widehat{\tau}_b^{\varepsilon} - \tau)^2}{\mathbf{E}_{\tau}(\widehat{\tau}_{mle}^{\varepsilon} - \tau)^2}$.



(b) Graphs of risk ratio $\tilde{\kappa}(\tau, \theta/\varepsilon) = \frac{\mathbf{E}_{\theta, \tau}(\widehat{L}_b^{\varepsilon} - L)^2}{\mathbf{E}_{\theta, \tau}(\widehat{L}_{mle}^{\varepsilon} - L)^2}$.

Fig. 1. Graphs of risk ratios κ and $\tilde{\kappa}$ depending on $\tau \in \{3, 4, \dots, 18\}$ for $N = 20$ observations, $\varepsilon = 1$, and different values of $\theta = 0.5, 1, 1.5, 2$.

(a) **Large SNR, $\theta/\varepsilon > 1$.**

It is clearly seen from Fig. 1 (a) that for large θ/ε the ratio $\kappa(\tau, \theta/\varepsilon)$ is close to its asymptotic theoretical value 0.7397. For $\theta/\varepsilon = 2$ the risk ratio κ fluctuates between 0.72 and 0.77. For $\theta/\varepsilon = 1.5$ we have $0.57 < \kappa < 0.8$. However, this is not the case for the ratio $\tilde{\kappa}(\tau, \theta/\varepsilon)$ of risks of estimating L presented in Fig. 1 (b). In this case the behavior of the risk ratio depends both on θ/ε and τ . For large SNR $\theta/\varepsilon = 1.5, 2$ the relative efficiency is close to 1. It means that asymptotically both estimates of L have very close risks.

(b) **Small SNR, $\theta/\varepsilon \leq 1$.**

For small SNR and moderate or small values of τ , the Bayes estimate \widehat{L}_b has to be preferred to the MLE \widehat{L}_{mle} . For example, if $\theta/\varepsilon = 0.5$ and the change in the data takes place close to the beginning of the sequence $\tau/N \leq 0.4$, then the Bayes estimate \widehat{L}_b of L is almost twice more efficient than the MLE estimate \widehat{L}_{mle} . If τ is large, then the MLE of L has to be chosen instead of the Bayes estimate. At the same time, the Bayes estimate of τ (Fig. 1 (a)) is always more efficient than the MLE in the case of small SNR ($\theta/\varepsilon = 0.5, 1$). Moreover, the smaller SNR is, the better is the behavior of Bayesian estimates compared to the maximum likelihood estimates, both for estimating τ and L .

(c) **Dependence on τ .**

Fig. 1 (b) shows that the Bayesian estimate of L is more efficient if the change takes place close to the beginning of the sequence. For example, for $\theta/\varepsilon = 1$ and $\tau/N < 0.4$ the Bayes estimate is more efficient than MLE, and vice versa, the MLE of L is more efficient for large values of τ , $\tau/N > 0.7$. If the values of τ are moderate, $0.4 \leq \tau/N \leq 0.7$, then the choice of better estimate of L depends on the value of SNR.

Our simulation results are very similar to the results of Sen and Srivastava [14]. They made a comparative study of the Bayes and likelihood ratio tests for the problem of testing the hypothesis of “no change” in Gaussian data. It turned out that in the case of known mean θ in the data the Bayes test is superior for $\tau/N \leq 0.4$, the LRT is superior for $\tau/N \geq 0.75$ and for $0.4 < \tau/N < 0.75$ the Bayes test dominates the LRT for small θ and vice versa.

The risks of MLE and Bayesian estimates of a smooth functional \mathcal{L} have the same first order asymptotic term as $\varepsilon \rightarrow 0$. Thus, from the viewpoint of asymptotic behavior there is no difference between two estimation approaches.

For small values of the signal-to-noise ratio the Bayesian procedure has much better performance than the ML procedure for a large part of values of the change-point τ in case of quadratic losses. We cannot explain this fact theoretically, since the behavior of the risk ratio is only known for large SNR as $\varepsilon \rightarrow 0$.

Simulation studies show that Bayesian procedures work remarkably better than MLE procedures in the case of small signal-to-noise ratio. On the other hand, asymptotically, both procedures show the same performance. We think that due to this fact Bayesian estimates have to be used in non-asymptotic framework. Unfortunately, in non-asymptotic setting, their theoretical risk properties are very difficult to obtain.

ACKNOWLEDGMENTS

The author is grateful to an anonymous referee for constructive comments and suggestions that helped to improve the paper.

REFERENCES

1. P. K. Bhattacharya, “Some Aspects of Change-Point Analysis”, in *IMS Lecture Notes*, Vol. 23: *Change-Point Problems* (1994), pp. 28–55.
2. P. K. Bhattacharya, “Maximum Likelihood Estimation of a Change-Point in the Distribution of Independent Random Variables: General Multiparameter Case”, *J. Multivar. Anal.* **23**, 183–208 (1987).
3. P. K. Bhattacharya and P. J. Brockwell, “The Minimum of an Additive Process with Applications to Signal Estimation and Storage Theory”, *Z. Wahrsch. verw. Gebiete* **37**, 51–75 (1976).
4. B. E. Brodskii and B. S. Darkhovskii, “Asymptotic Analysis of Some Estimates in the a Posteriori ‘Disorder’ Problem”, *Theory Probab. Appl.* **35**, 550–556 (1990).
5. B. E. Brodsky and B. S. Darkhovskiy, *Nonparametric Methods in Change-Point Problems* (Kluwer Acad. Publ., the Netherlands, 1993).
6. H. Chernoff and S. Zacks “Estimating the Current Mean of a Normal Distribution which is Subject to Changes in Time”, *Ann. Math. Statist.* **35**, 999–1028 (1964).
7. M. Csörgő and L. Horváth, *Limit Theorems In Change-Point Analysis* in *Wiley Series in Probab. and Statist.* (Wiley, 1997).

8. D. Ferger, *Change-Point Estimators in Case of Small Disorders*, J. Statist. Plann. Inf. **40** (1994), 33–49.
9. D. V. Hinkley, “Inference About the Change Point in a Sequence of Random Variables”, *Biometrika* **57**, 1–17 (1970).
10. I. A. Ibragimov and R. Z. Hasminski, *Statistical Estimation: Asymptotic Theory* (Springer, New York, 1981).
11. A. F. S. Lee and S. M. Heghinian, “A Shift of the Mean Level in a Sequence of Independent Normal Random Variables — a Bayesian Approach”, *Technometrics* **19**, 503–506 (1977).
12. H. Rubin, “The Estimation of Discontinuities in Multivariate Densities and Related Problems in Stochastic Processes”, in: *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* (Univ. California Press, Berkeley, 1961), Vol. 1, 563–574.
13. H. Rubin and K.-S. Song, “Exact Computation of the Asymptotic Efficiency of Maximum Likelihood Estimators of a Discontinuous Signal in a Gaussian White Noise”, *Ann. Statist.* **23**, 732–739 (1995).
14. A. Sen and M. S. Srivastava, “On Tests for Detecting Change in Mean”, *Ann. Statist.* **3**, 98–108 (1975).
15. A. N. Shiryaev, *Optimal Stopping Rules* (Springer, New York, 1978).