Supplement to 'Weighted likelihood test for a change in one component of a parametric mixture'

Dominique Abgrall*

April 30, 2021

Abstract

This supplement contains additional results, the proofs of the results from the main paper and further illustrations of the numerical applications.

Keywords: Change-point detection, parametric mixture, weak limit theorems for dependent sequences, weighted likelihood quotient test, applications to insurance data.

Contents

S1 Notations

The numbering of sections, results and equations of this supplement begin with 'S' while the sections, assumptions, results and equations of the main paper do not. For example (1) refers to the equation that defines the Fisher information matrix in Section 2.1 of the main paper while (S1) refers to the equation below that defines the Skorokhod metric in this supplement. With this specification, we omit to indicate whether the reference is in the main paper or in this supplement.

We denote by $D_\theta(.)$, $D^2_\theta(.)$ and $D^3_\theta(.)$ respectively the vector, matrix and hypermatrix differential operators in $\theta \in \mathbb{R}^d$. For $\theta, \tilde{\theta} \in \mathbb{R}^d$, we denote by $[\theta, \tilde{\theta}]$ the segment $[\theta, \tilde{\theta}] := \{\lambda\theta + (1-\lambda)\tilde{\theta},\ \lambda \in [0,1]\}$.

$gl_d(\mathbb{R})$ denotes the set of matrices of size $d \times d$ with real coefficients and $GL_d(\mathbb{R})$ the set of invertible $d \times d$-matrices with real coefficients.
For a given matrix M, its i-th line is denoted by $M_{i,.}$ and its j-th column is denoted by $M_{.,j}$. The same logic is used for hypermatrices: for a given $J \in \mathbb{R}^{d\times d\times d}$ and $1 \leq i \leq d$, we denote by $J_{i,.,.} := (J_{i,j,k,})_{1\leq j,k\leq d}$ the

*Laboratoire de Mathématiques de Bretagne Atlantique, Université de Bretagne Occidentale, dominique.abgrall@gmail.com.

$d \times d$-matrix obtained from J.
For a given matrix M, we denote $\left(M^{-1}\right)^T$ by $M^{-1\,T}$.

For $d_1, d_2 > 0$, we endow the space $F = \mathbb{R}^{d_1} \times gl_{d_2}(\mathbb{R})$ with the norm $\|.\|_2$ defined for the pair $x = (y, Z) \in \mathbb{R}^{d_1} \times gl_{d_2}(\mathbb{R})$ by $\|x\|_2^2 := \sum_{i=1}^{d_1} y_i^2 + \sum_{1\leq i,j\leq d_2} Z_{i,j}^2$. The norm used for y in $\mathbb{R}^{d_1}$ is the Euclidean norm. The norm used for $Z \in gl_{d_2}(\mathbb{R})$ is the *entrywise* 2-norm, also known as the Frobenius norm.

The space of càd-làg functions, defined on some interval $E \subseteq [0,1]$ with values in F, is denoted by $\mathbb{D}(E,F)$ and referred as the Skorokhod metric space with the Skorokhod metric $d_{\mathbb{D}(E,F)}(.,.)$ defined for ζ_1 and ζ_2 in $\mathbb{D}(E,F)$ by

$$d_{\mathbb{D}(E,F)}(\zeta_1,\zeta_2) := \inf_{\tau\in\Gamma_E} \max\left\{\sup_{s\in E}|\tau(s)-s|, \sup_{s\in E}\|\zeta_1(s)-\zeta_2\circ\tau(s)\|_2\right\} \tag{S1}$$

with Γ_E the set of continuous and strictly increasing bijections from E to itself. For some arguments, we also consider the norm $\|.\|_2$ on $\mathbb{D}(E,F)$ defined for $\zeta \in \mathbb{D}(E,F)$ by $\|\zeta\|_2 := \sup_{s\in E}\|\zeta(s)\|_2$.
Refer to Section 12 in Billingsley (1999) for a detailed construction of the Skorokhod topology and the space $\mathbb{D}(E,F)$.

If Σ^2 is a covariance matrix, then it is positive semi-definite, and Σ will denote the unique positive semi-definite square root of Σ^2.

A glossary of notations is given in Section S6.

S2 The estimators $\hat\theta_{0,s}$ and $\hat\theta_{s,1}$

In this section, we give some detailed properties concerning the estimators $\hat\theta_{0,s}$ and $\hat\theta_{s,1}$ defined in Section 2.1.

From Assumption 2.4, we already know that the estimator $\hat\theta$ converges almost surely to $\boldsymbol{\theta}$ when $n \to \infty$. With the following result inspired from Proposition 3.3 in Dehling et al. (2014), we can extend this convergence property to $\hat\theta_{0,s}$, $s \in [\bar s, 1]$, and to $\hat\theta_{s,1}$, $s \in [\bar s, 1-\bar s]$.

Lemma S2.1. *If a sequence $(u_n)_{n\geq 1} \subset \mathbb{R}^d$ converges to some finite limit u, then the sequence $u_{\lfloor sn\rfloor}$ converges to u, uniformly in $s \in [\bar s, 1]$.*

Proof. Fix $\epsilon > 0$. Let N such that for all $n \geq N$, $|u_n - u| \leq \epsilon$ and set $N' := \left\lfloor \frac{N}{\bar s}\right\rfloor + 1$. Then, for any $n \geq N'$, $[\bar s, 1] \subset [\frac{N}{n}, 1]$, thus, for any $n \geq N'$, $\lfloor ns\rfloor \geq N$ and, by the choice of N, $\left|u_{\lfloor sn\rfloor} - u\right| \leq \epsilon$. The result follows. □

Corollary S2.2. *If a sequence $(u_n)_{n\geq 1} \subset \mathbb{R}^d$ converges to some finite limit u, then the sequence $u_{\lfloor sn\rfloor}$ converges to u, uniformly in $s \in [\bar s, 1-\bar s]$.*

Reasoning ω by ω, this result implies directly that the almost sure convergence of $\hat\theta_{0,s}$ and $\hat\theta_{s,1}$ is ω-wise uniform in s. This will represent a key property for the main result.

Proposition S2.3. *Under H_0 and Assumptions 2.1-2.4, the estimator $\hat\theta_{0,s}$ (resp. $\hat\theta_{s,1}$) converges almost surely to $\boldsymbol{\theta}$, uniformly in $s \in [\bar s, 1]$ (resp. in $s \in [\bar s, 1-\bar s]$).*

For n large enough, it is possible to obtain an explicit form for $\hat\theta_{0,s}$. Indeed, the estimator $\hat\theta$ is a sequence of solutions of the likelihood equations $D_\theta L(X,\theta) = 0$. Therefore, we can follow the ideas from the proof of the usual limit theorems for maximum likelihood estimators (see e.g. Theorem 5.1 in Lehmann and Casella (1998), Section 6.5).

Corollary S2.4. *Under H_0 and Assumptions 2.1-2.4, almost surely, the estimator $\hat\theta$ exists for n large enough. Moreover, for almost all $\omega \in \Omega$, we can find some $N(\omega) \geq 1$ that does not depend on $s \in [\bar s, 1-\bar s]$ such that, for all $n \geq N(\omega)$, the three estimators $\hat\theta$, $\hat\theta_{0,s}$ and $\hat\theta_{s,1}$ are respectively the unique solutions of the likelihood equations*

$$D_\theta L(X,\theta) = 0,\ \ D_\theta L((X_1,\ldots,X_{\lfloor sn\rfloor}),\theta) = 0,\ \ \textit{and}\ D_\theta L((X_{\lfloor sn\rfloor+1},\ldots,X_n),\theta) = 0. \tag{S2}$$

Proof. The existence of $\hat{\theta}$ follows from the almost sure convergence. Indeed, by assumption, Θ is an open convex subset of $\mathbb{R}^d$ in which θ_0 belongs. Then, we obtain that, for n large enough, $\hat{\theta}$ also belongs to this open convex set. The proof for the three estimators is a direct application of Proposition S2.3. $\square$

Remark S2.5. *It is clear that the number N in Corollary S2.4 depends on ω. However, since, in this subsection, we always work ω by ω on the set of full probability where the three estimators converge, this will not pose any problem.*

In the sequel, the expression "for n large enough" will always implicitly imply that $\hat{\theta}$ belongs to Θ and solves (S2). In particular, due to the regularity Assumption 2.2, the following Taylor expansion is well defined as soon as $\hat{\theta}$ belongs to Θ: for $1 \leq j \leq d$,

$$D_\theta L(X,\hat{\theta})_j = D_\theta L(X,\boldsymbol{\theta})_j + \sum_{k=1}^{d} D^2_\theta L(X,\boldsymbol{\theta})_{j,k}\left(\hat{\theta}_k - \boldsymbol{\theta}_k\right) + \frac{1}{2}\sum_{k=1}^{d}\sum_{l=1}^{d}\left(\hat{\theta}_l - \boldsymbol{\theta}_l\right) D^3_\theta L(X,\theta')_{j,k,l}\left(\hat{\theta}_k - \boldsymbol{\theta}_k\right) \quad \text{(S3)}$$

for some θ' on the segment $[\hat{\theta},\boldsymbol{\theta}] \subset \mathbb{R}^d$.

Set

$$\hat{A} := -\frac{1}{n}\sum_{i=1}^{n}\left(D^2_\theta(\log f)(X_i,\boldsymbol{\theta}) \; + \; \frac{1}{2}\sum_{l=1}^{d}\left(\hat{\theta}_l - \boldsymbol{\theta}_l\right) D^3_\theta(\log f)(X_i,\theta')_{l,.,.}\right). \quad \text{(S4)}$$

Because of (S2), the left hand side of (S3) vanishes. Thus, replacing $L(.,\boldsymbol{\theta})$ by its explicit expression, we get the equality between the two vectors

$$\hat{A}\left(\hat{\theta} - \boldsymbol{\theta}\right) = \frac{1}{n}\sum_{i=1}^{n} D_\theta(\log f)(X_i,\boldsymbol{\theta}).$$

The limit of $\hat{A}$, when n tends to infinity, is given by the next Proposition. We follow the standard proof of the usual limit theorems for maximum likelihood estimators (e.g. Theorem 5.1 in Lehmann and Casella (1998)) and extend it to the almost sure convergence.

Proposition S2.6. *Under H_0 and Assumptions 2.2-2.4, the matrix $\hat{A}$ converges almost surely to the Fisher Information Matrix $\boldsymbol{I}$.*

Proof. With Assumption 2.2, the result follows by essentially the same logic as in the proof of the convergence in probability (see e.g. Theorem 3.10, in Lehmann and Casella (1998), Section 6.3). $\square$

Corollary S2.7. *Almost surely, the inverse matrix $\hat{A}^{-1}$ exists for large n, and converges to the inverse Fisher Information Matrix $\boldsymbol{I}^{-1}$ as $n \to \infty$.*

Proof. It follows from Proposition S2.6 that $\det(\hat{A}) \xrightarrow[n\to\infty]{a.s.} \det(\boldsymbol{I})$. Now recall that, by Assumption 2.3, $\boldsymbol{I}$ is definite positive and, in particular $\det(\boldsymbol{I}) > 0$. It follows that, for n large enough, $\det(\hat{A}) \neq 0$ and $\hat{A}^{-1}$ exists. The result follows. $\square$

In the same way as above, for any $s \in [\bar{s}, 1]$, there exists some point $\theta'_{0,s}$ on the segment $\left[\hat{\theta}_{0,s},\boldsymbol{\theta}\right]$, such that the matrix

$$\hat{A}_{0,s} := -\frac{1}{\lfloor sn \rfloor}\sum_{i=1}^{\lfloor sn \rfloor}\left(D^2_\theta(\log f)(X_i,\boldsymbol{\theta}) + \frac{1}{2}\sum_{l=1}^{d}\left(\hat{\theta}_{0,s;l} - \boldsymbol{\theta}_l\right) D^3_\theta(\log f)(X_i,\theta'_{0,s})_{l,.,.}\right) \quad \text{(S5)}$$

satisfies

$$\hat{A}_{0,s}(\hat{\theta}_{0,s} - \boldsymbol{\theta}) = \frac{1}{\lfloor sn \rfloor}\sum_{i=1}^{\lfloor sn \rfloor} D_\theta(\log f)(X_i,\boldsymbol{\theta}). \quad \text{(S6)}$$

And, for any $s \in [\bar{s}, 1-\bar{s}]$, there exists some point $\theta'_{s,1}$ on the segment $\left[\hat{\theta}_{s,1}, \boldsymbol{\theta}\right]$, such that the matrix

$$\hat{A}_{s,1} := -\frac{1}{n - \lfloor sn \rfloor} \sum_{i=\lfloor sn \rfloor+1}^{n} \left(D_\theta^2 (\log f)(X_i, \boldsymbol{\theta}) + \frac{1}{2} \sum_{l=1}^{d} \left(\hat{\theta}_{s,1;l} - \boldsymbol{\theta}_l \right) D_\theta^3 (\log f)(X_i, \theta'_{s,1})_{l,.,.} \right)$$

satisfies

$$\hat{A}_{s,1}(\hat{\theta}_{s,1} - \boldsymbol{\theta}) = \frac{1}{n - \lfloor sn \rfloor} \sum_{i=\lfloor sn \rfloor+1}^{n} D_\theta (\log f)(X_i, \boldsymbol{\theta}).$$

To sum up, the following lemma provides an explicit expression for $\hat{\theta}_{0,s}$ and $\hat{\theta}_{s,1}$ and the convergence of $\hat{A}_{0,s}$ and $\hat{A}_{s,1}$.

Proposition S2.8. *Under H_0 and Assumptions 2.1-2.4, almost surely, for n large enough,*

- *for all $s \in [\bar{s}, 1]$, the matrix $\hat{A}_{0,s}$ is invertible and*

$$\hat{\theta}_{0,s} - \boldsymbol{\theta} = \hat{A}_{0,s}^{-1} \frac{1}{\lfloor sn \rfloor} \sum_{i=1}^{\lfloor sn \rfloor} D_\theta (\log f)(X_i, \boldsymbol{\theta}), \qquad \text{(S7)}$$

  with $\mathbb{E}_{H_0}[D_\theta(\log f)(X_1, \boldsymbol{\theta})] = 0$, and where $\hat{A}_{0,s}^{-1}$ converges almost surely to $\boldsymbol{I}^{-1}$, uniformly in $s \in [\bar{s}, 1]$,

- *for all $s \in [\bar{s}, 1-\bar{s}]$ the matrix $\hat{A}_{s,1}$ is invertible and*

$$\hat{\theta}_{s,1} - \boldsymbol{\theta} = \hat{A}_{s,1}^{-1} \frac{1}{n - \lfloor sn \rfloor} \sum_{i=\lfloor sn \rfloor+1}^{n} D_\theta (\log f)(X_i, \boldsymbol{\theta})$$

  with $\mathbb{E}_{H_0}[D_\theta(\log f)(X_1, \boldsymbol{\theta})] = 0$, and where $\hat{A}_{s,1}^{-1}$ converges almost surely to $\boldsymbol{I}^{-1}$, uniformly in $s \in [\bar{s}, 1-\bar{s}]$.

Proof of Propositions 3.2 and S2.8. For n large enough, $\hat{A}_{0,s}^{-1}$ is well defined and Equation (S7) follows directly from (S6). Assumption 2.2 guarantees that the expectation $\mathbb{E}_{H_0}[D_\theta(\log f)(X_1, \boldsymbol{\theta})]$ vanishes. Finally we use Corollary S2.7 and Lemma S2.1 to obtain the almost sure convergence of $\hat{A}_{0,s}^{-1}$ to $\boldsymbol{I}^{-1}$, uniformly in $s \in [\bar{s}, 1]$. The proof of the analogue result for $\hat{\theta}_{s,1}$ and $\hat{A}_{s,1}$ is the same. □

The explicit expression obtained for $\hat{\theta}_{0,s} - \boldsymbol{\theta}$ and $\hat{\theta}_{s,1} - \boldsymbol{\theta}$ already points out the direction of the next steps: since $\hat{A}_{0,s}^{-1}$ and $\hat{A}_{s,1}^{-1}$ converge almost surely to $\boldsymbol{I}^{-1}$, uniformly in s, and with $\mathbb{E}_{H_0}[D_\theta(\log f)(X_1, \boldsymbol{\theta})] = 0$, we will be able to establish Donsker-type result for $\hat{\theta}_{0,s} - \boldsymbol{\theta}$ and $\hat{\theta}_{s,1} - \boldsymbol{\theta}$. This can be used to derive a Donsker-type result for $Q_{s,n}^1$ and $Q_{s,n}^2$.

We will need the following variant of Glivenko-Cantelli's Theorem that exploits the almost sure convergence of $\hat{\theta}_{0,s}$ and $\hat{\theta}_{s,1}$ to $\boldsymbol{\theta}$.

Lemma S2.9. *Consider an application $h : (x, \theta) \in \mathcal{X} \times \Theta \mapsto h(x, \theta) \in \mathbb{R}$ and a convex subset $\mathcal{O}$ of Θ, such that $\boldsymbol{\theta}$ is in the interior of $\mathcal{O}$ and*

1. *for almost all $x \in \mathcal{X}$, the application $\theta \mapsto h(x, \theta)$ is continuous on $\mathcal{O}$,*
2. *we can find some application $\mathcal{X} \ni x \mapsto \kappa_3(x)$, such that, for all θ in $\mathcal{O}$, $|h(x, \theta)| \le \kappa_3(x)$ and $\mathbb{E}_{H_0}[|\kappa_3(X_1)|] < \infty$.*

Then, under H_0 and Assumptions 2.1-2.4, one has $\mathbb{E}_{H_0}[|h(X_1, \boldsymbol{\theta})|] < \infty$ and

- *for $\theta'_{0,s} \in [\hat{\theta}_{0,s}, \boldsymbol{\theta}]$, $\frac{1}{\lfloor sn \rfloor} \sum_{i=1}^{\lfloor sn \rfloor} h(X_i, \theta'_{0,s}) \xrightarrow[n\to\infty]{a.s.} \mathbb{E}_{H_0}[h(X_1, \boldsymbol{\theta})]$, uniformly in $s \in [\bar{s}, 1]$,*
- *for $\theta'_{s,1} \in [\hat{\theta}_{s,1}, \boldsymbol{\theta}]$, $\frac{1}{n - \lfloor sn \rfloor} \sum_{i=\lfloor sn \rfloor+1}^{n} h(X_i, \theta'_{s,1}) \xrightarrow[n\to\infty]{a.s.} \mathbb{E}_{H_0}[h(X_1, \boldsymbol{\theta})]$, uniformly in $s \in [\bar{s}, 1-\bar{s}]$.*

Proof. We only show the case $\theta'_{0,s} \in [\hat{\theta}_{0,s}, \boldsymbol{\theta}]$, $s \in [\bar{s}, 1]$. By the second condition of the lemma, $\boldsymbol{\theta} \in \mathcal{O}$ implies that $|h(x, \boldsymbol{\theta})| \leq \kappa_3(x)$ for all $x \in \mathcal{X}$, thus $\mathbb{E}_{H_0}[|h(X_1, \boldsymbol{\theta})|] \leq \mathbb{E}_{H_0}[|\kappa_3(X_1)|] < \infty$. Let us fix some $\epsilon > 0$ small enough so that, with $B(\boldsymbol{\theta}, \epsilon)$ the closed ball centered in $\boldsymbol{\theta}$ with radius ϵ, $B(\boldsymbol{\theta}, \epsilon) \cap \Theta$ is strictly contained in the subset $\mathcal{O}$. This is possible since, from the first condition of the lemma, $\mathcal{O}$ is a convex subset of Θ such that $\boldsymbol{\theta}$ is in the interior of $\mathcal{O}$. From Proposition S2.3, $\hat{\theta}_{0,s} \xrightarrow[n\to\infty]{a.s.} \boldsymbol{\theta}$, uniformly in $s \in [\bar{s}, 1]$. Therefore, almost surely, we can find some $N \geq 1$ such that for all $n \geq N$ and for all $s \in [\bar{s}, 1]$, $\hat{\theta}_{0,s} \in B(\boldsymbol{\theta}, \epsilon) \cap \Theta$. Since $\theta'_{0,s}$ is a point on the segment $[\hat{\theta}_{0,s}, \boldsymbol{\theta}]$, it also belongs to $B(\boldsymbol{\theta}, \epsilon) \cap \Theta$. It follows that

$$
\begin{aligned}
&\left| \frac{1}{\lfloor sn \rfloor} \sum_{i=1}^{\lfloor sn \rfloor} h(X_i, \theta'_{0,s}) - \mathbb{E}_{H_0}\left[h(X_1, \boldsymbol{\theta})\right] \right| \\
&\leq \sup_{\theta \in B(\boldsymbol{\theta}, \epsilon) \cap \Theta} \left| \frac{1}{\lfloor sn \rfloor} \sum_{i=1}^{\lfloor sn \rfloor} h(X_i, \theta) - \mathbb{E}_{H_0}\left[h(X_1, \theta)\right] \right| + \left| \mathbb{E}_{H_0}\left[h(X_1, \theta'_{0,s})\right] - \mathbb{E}_{H_0}\left[h(X_1, \boldsymbol{\theta})\right] \right|. \qquad \text{(S8)}
\end{aligned}
$$

With conditions 1. and 2., thanks to the dominated convergence theorem, the application $\theta \mapsto \mathbb{E}_{H_0}[h(X_1, \theta)]$ is continuous on $\mathcal{O}$. Since $\boldsymbol{\theta} \in \mathcal{O}$, $\hat{\theta}_{0,s} \xrightarrow[n\to\infty]{a.s.} \boldsymbol{\theta}$, uniformly in $s \in [\bar{s}, 1]$ and, for all $n \geq 1$, $\theta'_{0,s} \in [\hat{\theta}_{0,s}, \boldsymbol{\theta}]$, we have also $\theta'_{0,s} \xrightarrow[n\to\infty]{a.s.} \boldsymbol{\theta}$. Therefore the second term of the right hand side of (S8) converges almost surely to 0, uniformly in $s \in [\bar{s}, 1]$, when $n \to \infty$.

To conclude the proof, we show that the first term also converges almost surely to 0, uniformly in $s \in [\bar{s}, 1]$, when $n \to \infty$. For all fixed $\theta \in B(\boldsymbol{\theta}, \epsilon) \cap \Theta$, the following convergence is an application of the Strong Law of Large Numbers:

$$
Y_n(\theta) := \frac{1}{n} \sum_{i=1}^{n} h(X_i, \theta) - \mathbb{E}_{H_0}\left[h(X_1, \theta)\right] \xrightarrow[n\to\infty]{a.s.} 0.
$$

We deduce from assumptions 1. and 2. that $\theta \to Y_n(\theta)$ is continuous. Since $B(\boldsymbol{\theta}, \epsilon) \cap \Theta$ is compact, we get the convergence of the supremum $\sup_{\theta \in B(\boldsymbol{\theta}, \epsilon) \cap \Theta} |Y_n(\theta)| \xrightarrow[n\to\infty]{a.s.} 0$. And finally we can conclude by Lemma S2.1. $\square$

This Lemma concludes the collection of properties that are required for the estimators $\hat{\theta}_{0,s}$ and $\hat{\theta}_{s,1}$.

S3 Proofs

Lemma S3.1. *Set*

$$
\boldsymbol{\xi} := (0, 0, \boldsymbol{I}) \in (\mathbb{R}^d)^2 \times GL_d(\mathbb{R}). \qquad \text{(S9)}
$$

Under H_0, for all $s \in [\bar{s}, 1]$, $\mathbb{E}_{H_0}[\hat{\xi}_{0,s}] = \boldsymbol{\xi}$, the sequence of random variables $\hat{\xi}_{0,s}$ converges a.s. to $\boldsymbol{\xi}$, uniformly in $s \in [\bar{s}, 1]$, and the process $\sqrt{n}\left(\hat{\xi}_{0,s} - \boldsymbol{\xi}\right)_{s \in [\bar{s},1]}$ converges weakly in the Skorokhod metric space of càd-làg paths $\mathbb{D}_{[\bar{s},1]} := \mathbb{D}([\bar{s}, 1], (\mathbb{R}^d)^2 \times gl_d(\mathbb{R}))$, as follows

$$
\sqrt{n}\left(\hat{\xi}_{0,s} - \boldsymbol{\xi}\right)_{s \in [\bar{s},1]} \xrightarrow[n\to\infty]{\mathcal{D}} \left(\frac{1}{s} \Sigma W_s\right)_{s \in [\bar{s},1]},
$$

where $W := (W_s)_{s \in [0,1]}$ is a standard $2d + d^2$-dimensional Brownian motion and ΣW_s is reorganized as a triple in $(\mathbb{R}^d)^2 \times gl_d(\mathbb{R})$.

Proof of Lemmas 3.3 and S3.1. Under H_0, the random vector $\hat{\xi}_{0,s}$ is the sum of independent identically distributed random variables. In addition, we already know that

- ⋄ from Assumptions 2.2 and 2.3, $\mathbb{E}_{H_0}\left[D_\theta(\log f)(X_1, \boldsymbol{\theta})\right] = 0$ and $\mathbb{E}_{H_0}\left[-D^2_\theta(\log f)(X_1, \boldsymbol{\theta})\right] = \boldsymbol{I}$,
- ⋄ by the expression for $\boldsymbol{u}$ in (6), $\mathbb{E}_{H_0}\left[D_\theta(w \log f_1)(X_1, \boldsymbol{\theta})\right] = \boldsymbol{u}$.

Therefore $\mathbb{E}_{H_0}[\hat{\xi}_{0,s}] = \boldsymbol{\xi}$, and the uniform a.s. convergence of the random variables $\hat{\xi}_{0,s}$ to $\boldsymbol{\xi}$ is a direct consequence of Lemma S2.9. The second part of the lemma follows then from Donsker's Theorem[1] and Slutsky's Theorem. $\square$

Lemma S3.2. *Almost surely, for large n (depending on ω), the variable $\hat{A}_{0,s}$ can be written as*

$$\hat{A}_{0,s} = \hat{I}_{0,s} - \frac{1}{2}\sum_{l=1}^{d}\left(\hat{\iota}_{0,s}^{T}\left(\hat{A}_{0,s}^{-1\,T}\right)_{.,l}\right)(\hat{J}_{0,s})_{.,.,l}, \tag{S10}$$

for all $s \in [\bar{s}, 1]$, where $\hat{J}_{0,s}$ is the hypermatrix defined by

$$\hat{J}_{0,s} := \frac{1}{\lfloor sn \rfloor}\sum_{i=1}^{\lfloor sn \rfloor} D_\theta^3(\log f)(X_i, \theta'_{0,s}).$$

In addition, under H_0, almost surely, $\hat{J}_{0,s}$ converges to the hypermatrix $\boldsymbol{J}$, uniformly in $s \in [\bar{s}, 1]$.

Proof of Lemmas 3.4 and S3.2. Recall that the explicit expression of $\hat{A}_{0,s}$ in (S5) depends itself on $\hat{\theta}_{0,s} - \boldsymbol{\theta}$, which, by Proposition S2.8, almost surely, can once more be replaced by

$$\hat{A}_{0,s}^{-1}\frac{1}{\lfloor sn \rfloor}\sum_{i=1}^{\lfloor sn \rfloor} D_\theta(\log f)(X_i, \boldsymbol{\theta})$$

for n large enough. This gives:

$$\begin{aligned}\hat{A}_{0,s} = &\left(-\frac{1}{\lfloor sn \rfloor}\sum_{i=1}^{\lfloor sn \rfloor} D_\theta^2(\log f)(X_i, \boldsymbol{\theta})\right)\\ &-\frac{1}{2}\sum_{l=1}^{d}\left(\left(\frac{1}{\lfloor sn \rfloor}\sum_{i=1}^{\lfloor sn \rfloor} D_\theta(\log f)(X_i, \boldsymbol{\theta})^T\right)\left(\hat{A}_{0,s}^{-1\,T}\right)_{.,l}\right)\left(\frac{1}{\lfloor sn \rfloor}\sum_{i=1}^{\lfloor sn \rfloor} D_\theta^3(\log f)(X_i, \theta'_{0,s})_{l,.,.}\right).\end{aligned}$$

The result given in (S10) follows. With Assumption 2.2, the convergence of $\hat{J}_{0,s}$ is a direct application of Lemma S2.9, taking the parameter set Θ as $\mathcal{O}$. $\square$

Theorem S3.3. *Under H_0, the process $\sqrt{n}(\hat{\xi}'_{0,s} - \boldsymbol{\xi}')_{s\in[\bar{s},1]}$ converges weakly to $(\frac{1}{s}\boldsymbol{g}(\Sigma W_s))_{s\in[\bar{s},1]}$ in $\mathbb{D}_{[\bar{s},1]}$, where $\boldsymbol{g}$ is the linear map defined for $(\iota, u, I) \in (\mathbb{R}^d)^2 \times gl_d(\mathbb{R})$ by*

$$\boldsymbol{g}(\iota, u, I) := \left(\iota, u, -\boldsymbol{I}^{-1}\left(I - \frac{1}{2}\sum_{l=1}^{d}\left(\iota^T\left(\boldsymbol{I}^{-1\,T}\right)_{.,l}\right)\boldsymbol{J}_{.,.,l}\right)\boldsymbol{I}^{-1}\right), \tag{S11}$$

and ΣW_s is reorganized as a triple in $(\mathbb{R}^d)^2 \times gl_d(\mathbb{R})$.

Before giving the proof of Theorems 3.5 and S3.3, we start by recalling the Functional Delta Method in normed spaces (van der Vaart (1998)).

Theorem S3.4 (Functional Delta Method in normed spaces, Theorem 20.8 in van der Vaart (1998))**.** *For D and E two normed linear spaces, consider a map Φ from a subset D_Φ of D to E, that is Hadamard differentiable at $\theta \in D_\Phi$ with differential denoted by $\Phi'_\theta(.)$. Consider also a sequence of random maps X_n with values in D_Φ and a sequence of numbers a_n which tends to infinity as $n \to \infty$. If, as $n \to \infty$, the sequence $a_n(X_n - \theta)$ converges weakly to some random map X, then the sequence $a_n(\Phi(X_n) - \Phi(\theta))$ converges weakly to the random map $\Phi'_\theta(X)$.*

Proof. The proof is a direct application of the Continuous Mapping Theorem with $f_n(\zeta) := a_n(\Phi(\theta + a_n^{-1}\zeta) - \Phi(\theta))$ where the maps f_n are defined on the sets $\{\zeta : \theta + a_n^{-1}\zeta \in D_\Phi\}$. The Hadamard differentiability ensures that the conditions of the Continuous Mapping Theorem[2] hold. $\square$

[1] See e.g. Theorem 14.1 in Billingsley (1999).
[2] See e.g. Theorem 18.11 in van der Vaart (1998).

The space $\mathbb{C}([0,1],\mathbb{R}^d)$ of continuous functions $[0,1]\to\mathbb{R}^d$ is a linear normed space with the norm $\|.\|_2$ defined in Section S1, and therefore falls in the scope of Theorem S3.4. However the Skorokhod metric space $\mathbb{D}([0,1],\mathbb{R}^d)$ is not a normed space. We give here a Corollary of the result from van der Vaart (1998) for càd-làg processes.

Corollary S3.5 (Functional Delta Method in the Skorokhod metric space)**.** *For $0<d_1,d_2<\infty$, consider a map $\Phi : D_\Phi \subseteq \mathbb{D}([0,1],\mathbb{R}^{d_1}) \to \mathbb{D}([0,1],\mathbb{R}^{d_2})$. Consider also a sequence of random maps X_n with values in D_Φ and a sequence of numbers a_n which tends to infinity as $n\to\infty$. If, as $n\to\infty$,*

- ⋄ *the sequence $a_n(X_n-\theta)$ converges weakly to some random map X,*
- ⋄ *we can find some linear map $\Phi'_\theta(.)$ from $\mathbb{D}([0,1],\mathbb{R}^{d_1})$ to $\mathbb{D}([0,1],\mathbb{R}^{d_2})$ such that for every sequence $\zeta_n\in\{z:\theta+a_n^{-1}z\in D_\Phi\}$ for which we can find a subsequence $\zeta_{n'}$ that converges in $\mathbb{D}([0,1],\mathbb{R}^{d_1})$ to ζ, the sequence $a_{n'}(\Phi(\theta+a_{n'}^{-1}\zeta_{n'})-\Phi(\theta))$ converges in $\mathbb{D}([0,1],\mathbb{R}^{d_2})$ to $\Phi'_\theta(\zeta)$,*

then the sequence $a_n(\Phi(X_n)-\Phi(\theta))$ converges weakly to the random map $\Phi'_\theta(X)$.

Proof. As for Theorem S3.4, the proof is an application of the Continuous Mapping Theorem where the Hadamard differentiability is replaced by the second condition. □

With this result, we can give below the proof of Theorems 3.5 and S3.3.

Proof of Theorems 3.5 and S3.3. (i) Recall that, by Lemma S3.1, $\sqrt{n}\left(\hat{\xi}_{0,s}-\boldsymbol{\xi}\right)_{s\in[\bar{s},1]}$ converges weakly to $(\frac{1}{s}\Sigma W_s)_{s\in[\bar{s},1]}$ in $\mathbb{D}_{[\bar{s},1]}$, while, by Proposition S2.8 and Lemma S3.2, the couple $(\hat{A}_{0,s}^{-1},\hat{J}_{0,s})$ converges a.s. to $(\boldsymbol{I}^{-1},\boldsymbol{J})$ uniformly in $s\in[\bar{s},1]$. It follows by Slutsky's Theorem that the random process $(\sqrt{n}(\hat{\xi}_{0,s}-\boldsymbol{\xi}),\hat{A}_{0,s}^{-1},\hat{J}_{0,s})_{s\in[\bar{s},1]}$ converges weakly to $(\frac{1}{s}\Sigma W_s,\boldsymbol{I}^{-1},\boldsymbol{J})_{s\in[\bar{s},1]}$ in $\mathbb{D}([\bar{s},1],(\mathbb{R}^d)^2\times gl_d(\mathbb{R})\times GL_d(\mathbb{R})\times\mathbb{R}^{d\times d\times d})$.

(ii) Using Lemma S3.2, almost surely for n large enough (i.e. n depends on ω but not on s), we can write

$$\hat{\xi}'_{0,s}=\varphi\circ g(\hat{\xi}_{0,s};\hat{A}_{0,s}^{-1},\hat{J}_{0,s}),\ s\in[\bar{s},1],$$

where, for all $(\iota,u,I;A,J)\in(\mathbb{R}^d)^2\times(gl_d(\mathbb{R}))^2\times\mathbb{R}^{d\times d\times d}$, we define

$$g(\iota,u,I;A,J):=\left(\iota,u,I-\frac{1}{2}\sum_{l=1}^{d}\left(\iota^T(A^T)_{.,l}\right)J_{.,.,l}\right),\tag{S12}$$

and, for $(\iota,u,I)\in(\mathbb{R}^d)^2\times GL_d(\mathbb{R})$,

$$\varphi(\iota,u,I):=(\iota,u,I^{-1}).$$

Remark that $g(0,u,I;A,J)=(0,u,I)$ for all $(u,I;A,J)\in\mathbb{R}^d\times(gl_d(\mathbb{R}))^2\times\mathbb{R}^{d\times d\times d}$. In particular,

$$\boldsymbol{\xi}=g(\boldsymbol{\xi};\hat{A}_{0,s}^{-1},\hat{J}_{0,s})$$

because $\boldsymbol{\xi}=(0,0,\boldsymbol{I})$ by definition in (S9). Since $g(\iota,u,I;A,J)$ is linear in (ι,u,I), then the following equality holds for each $s\in[\bar{s},1]$:

$$\sqrt{n}(g(\hat{\xi}_{0,s};\hat{A}_{0,s}^{-1},\hat{J}_{0,s})-\boldsymbol{\xi})=g(\sqrt{n}(\hat{\xi}_{0,s}-\boldsymbol{\xi}),\hat{A}_{0,s}^{-1},\hat{J}_{0,s}).$$

By (i) and the Continuous Mapping Theorem[3], the process

$$\sqrt{n}(g(\hat{\xi}_{0,s};\hat{A}_{0,s}^{-1},\hat{J}_{0,s})-\boldsymbol{\xi})_{s\in[\bar{s},1]}$$

converges weakly to $g(\frac{1}{s}\Sigma W_s;\boldsymbol{I}^{-1},\boldsymbol{J})_{s\in[\bar{s},1]}$.

(iii) By Lemma S3.1, the sequence of random variables $\hat{\xi}_{0,s}$ converges a.s. to $\boldsymbol{\xi}$, uniformly in $s\in[\bar{s},1]$. With (i), the triple $(\hat{\xi}_{0,s},\hat{A}_{0,s}^{-1},\hat{J}_{0,s})$ also converges a.s. to $(\boldsymbol{\xi},\boldsymbol{I}^{-1},\boldsymbol{J})$, uniformly in $s\in[\bar{s},1]$. Then, again by the

[3] See e.g. Theorem 18.11 in van der Vaart (1998).

Continuous Mapping Theorem, $g(\hat{\xi}_{0,s}; \hat{A}_{0,s}^{-1}, \hat{J}_{0,s})$ converges a.s. to $\boldsymbol{\xi}$, uniformly in $s \in [\bar{s}, 1]$.

(iv) Remark that, from Proposition S2.8 and Lemma S3.2, $g(\hat{\xi}_{0,s}; \hat{A}_{0,s}^{-1}, \hat{J}_{0,s})$ is well defined when $\hat{A}_{0,s}$ is invertible. That is the case almost surely for n large enough, uniformly in $s \in [\bar{s}, 1]$: i.e. n depends on ω but not on s.

From Assumption 2.3, $\boldsymbol{I}$ is positive definite with finite components. So $\boldsymbol{I}^{-1}$ is also positive definite and $0 < \left\|\boldsymbol{I}^{-1}\right\|_2^{-1} < \infty$. Fix some $0 < r < \left\|\boldsymbol{I}^{-1}\right\|_2^{-1}$ such that the closed ball $B(\boldsymbol{\xi}, r)$ centered in $\boldsymbol{\xi}$ with radius r is included in $(\mathbb{R}^d)^2 \times GL_d(\mathbb{R})$. With (iii), we see that, almost surely, the following holds for n large enough, uniformly in $s \in [\bar{s}, 1]$:

$$\hat{A}_{0,s} \text{ is invertible and } \left\| g(\hat{\xi}_{0,s}; \hat{A}_{0,s}^{-1}, \hat{J}_{0,s}) - \boldsymbol{\xi} \right\|_2 < r. \tag{S13}$$

Let $\hat{\xi}''_{0,\cdot}$ denote the process on $[\bar{s}, 1]$ defined for all $\omega \in \Omega$, all $n \geq 1$ and all $s \in [\bar{s}, 1]$ as follows:

$$\hat{\xi}''_{0,s}(\omega) := \begin{cases} g(\hat{\xi}_{0,s}; \hat{A}_{0,s}^{-1}, \hat{J}_{0,s})(\omega) & \text{if (S13) holds,} \\ \boldsymbol{\xi} & \text{otherwise.} \end{cases}$$

Then, almost surely, $\sqrt{n}(\hat{\xi}''_{0,s} - g(\hat{\xi}_{0,s}; \hat{A}_{0,s}^{-1}, \hat{J}_{0,s}))$ is equal to 0 for n large enough, uniformly in $s \in [\bar{s}, 1]$. We denote by $\boldsymbol{\xi}_\cdot$ the constant process such that $\boldsymbol{\xi}_s = \boldsymbol{\xi}$ for all $s \in [\bar{s}, 1]$. Therefore, by the conclusion from (ii) and Slutsky's Theorem and the Continuous Mapping Theorem,

$$\sqrt{n}(\hat{\xi}''_{0,\cdot} - \boldsymbol{\xi}_\cdot) = \sqrt{n}(g(\hat{\xi}_{0,s}; \hat{A}_{0,s}^{-1}, \hat{J}_{0,s})_{s\in[\bar{s},1]} - \boldsymbol{\xi}_\cdot) + \sqrt{n}(\hat{\xi}''_{0,\cdot} - g(\hat{\xi}_{0,s}; \hat{A}_{0,s}^{-1}, \hat{J}_{0,s})_{s\in[\bar{s},1]})$$

converges weakly to $g(\frac{1}{s}\Sigma W_s; \boldsymbol{I}^{-1}, \boldsymbol{J})_{s\in[\bar{s},1]}$.

(v) Let us denote by Φ the function from $\mathbb{D}([\bar{s}, 1], (\mathbb{R}^d)^2 \times GL_d(\mathbb{R})) \subset \mathbb{D}_{[\bar{s},1]}$ onto itself defined by:

$$\Phi(\zeta)_s := \varphi(\zeta_s), \; s \in [\bar{s}, 1]. \tag{S14}$$

It follows from the definition of $\hat{\xi}'_{0,s}$ in Section 3.1 that

$$\sqrt{n}(\hat{\xi}'_{0,s} - \boldsymbol{\xi}')_{s\in[\bar{s},1]} = \sqrt{n}(\Phi(\hat{\xi}''_{0,\cdot}) - \Phi(\boldsymbol{\xi}_\cdot)) + \sqrt{n}(\Phi(g(\hat{\xi}_{0,s}; \hat{A}_{0,s}^{-1}, \hat{J}_{0,s})_{s\in[\bar{s},1]}) - \Phi(\hat{\xi}''_{0,\cdot})). \tag{S15}$$

With (iii) and (iv), almost surely, $\sqrt{n}(\varphi(g(\hat{\xi}_{0,s}; \hat{A}_{0,s}^{-1}, \hat{J}_{0,s})) - \varphi(\hat{\xi}''_{0,s}))$ is equal to 0 for n large enough, uniformly in $s \in [\bar{s}, 1]$. Once more, by Slutsky's Theorem and the Continuous Mapping Theorem, both processes $\sqrt{n}(\hat{\xi}'_{0,s} - \boldsymbol{\xi}')_{s\in[\bar{s},1]}$ and $\sqrt{n}(\Phi(\hat{\xi}''_{0,\cdot}) - \Phi(\boldsymbol{\xi}_\cdot))$ have the same limit distribution.

The remainder of the proof is based on the functional delta method in the Skorokhod metric space given in Corollary S3.5. This result, applied to the map Φ, would conclude the proof and provide the limit distribution of $\sqrt{n}(\Phi(\hat{\xi}''_{0,\cdot}) - \Phi(\boldsymbol{\xi}_\cdot))$.

The first condition of Corollary S3.5 holds by (iv) since $\sqrt{n}(\hat{\xi}''_{0,\cdot} - \boldsymbol{\xi}_\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} g(\frac{1}{s}\Sigma W_s; \boldsymbol{I}^{-1}, \boldsymbol{J})_{s\in[\bar{s},1]}$.

To conclude the proof, it now sufficient to show that the second condition of Corollary S3.5 also holds. For that purpose, we start by noticing that, by (iv), for all $n \geq 1$, the process $\sqrt{n}(\hat{\xi}''_{0,\cdot} - \boldsymbol{\xi}_\cdot)$ is in the closed ball

$$B(0, r\sqrt{n}) := \{\zeta \in \mathbb{D}_{[\bar{s},1]}, \|\zeta\|_2 \leq r\sqrt{n}\},$$

where 0 is the null function on $[\bar{s}, 1]$. Let us consider the sequence of applications γ_n defined for ζ_n in $B(0, r\sqrt{n}) \subset \mathbb{D}_{[\bar{s},1]}$ by

$$\gamma_n(\zeta_n) := \sqrt{n}\left(\Phi\left(\boldsymbol{\xi}_\cdot + \frac{1}{\sqrt{n}}\zeta_n\right) - \Phi(\boldsymbol{\xi}_\cdot)\right).$$

Further denote the differential of Φ at $\boldsymbol{\xi}_\cdot$ by $D\Phi$. The differential is a function from $\mathbb{D}_{[\bar{s},1]}$ onto itself, defined for $\zeta = (\zeta_1, \zeta_2, \zeta_3)$ in $\mathbb{D}_{[\bar{s},1]}$ by[4]

$$D\Phi(\zeta)_s := (\zeta_1(s), \zeta_2(s), -\boldsymbol{I}^{-1}\zeta_3(s)\boldsymbol{I}^{-1}), \; s \in [\bar{s}, 1].$$

[4]We extend here the well known differential of the inversion of matrices given in Lemma 2.5.5 in Abraham et al. (1988).

It is then sufficient to show that the convergence of every sequence $\zeta_n \in B(0, r\sqrt{n})$ to $\zeta \in \mathbb{D}_{[\bar{s},1]}$ implies the convergence of $\gamma_n(\zeta_n)$ to $D\Phi(\zeta)$.

Let us consider some sequence $\zeta_n = (\zeta_{1,n}, \zeta_{2,n}, \zeta_{3,n}) \in B(0, r\sqrt{n})$ and some path $\zeta = (\zeta_1, \zeta_2, \zeta_3) \in \mathbb{D}_{[\bar{s},1]}$ such that $d_{\mathbb{D}_{[\bar{s},1]}}(\zeta_n, \zeta) \to 0$ as $n \to \infty$. From the definition of the Skorokhod metric in (S1), this means[5] that there exists some sequence of strictly increasing bijections τ_n^* from $[\bar{s}, 1]$ onto itself such that, as $n \to \infty$,

$$\sup_{s\in[\bar{s},1]} |\tau_n^*(s) - s| \to 0 \quad \text{and} \quad \sup_{s\in[\bar{s},1]} \|\zeta_n(\tau_n^*(s)) - \zeta(s)\|_2 \to 0. \tag{S16}$$

To conclude the proof, we need only to show that $\sup_s \|\gamma_n(\zeta_n(\tau_n^*(.))_s) - D\Phi(\zeta)_s\|_2 \to 0$. First, we remark that

$$\gamma_n(\zeta_n(\tau_n^*(s))_{s\in[\bar{s},1]}) = \left(\zeta_{1,n}(\tau_n^*(s)), \zeta_{2,n}(\tau_n^*(s)), \sqrt{n}\left(\left(\boldsymbol{I} + \frac{\zeta_{3,n}(\tau_n^*(s))}{\sqrt{n}}\right)^{-1} - \boldsymbol{I}^{-1}\right)\right)_{s\in[\bar{s},1]}$$

and $D\Phi(\zeta) = (\zeta_1(s), \zeta_2(s), -\boldsymbol{I}^{-1}\zeta_3(s)\boldsymbol{I}^{-1})_{s\in[\bar{s},1]}$. By (S16) and the definition of $\|.\|_2$ in Section S1, it is sufficient to show that $\sup_s \|\zeta_{3,n}(\tau_n^*(s)) - \zeta_3(s)\|_2 \to 0$ implies

$$\sup_s \left\| \sqrt{n}\left(\left(\boldsymbol{I} + \frac{\zeta_{3,n}(\tau_n^*(s))}{\sqrt{n}}\right)^{-1} - \boldsymbol{I}^{-1}\right) + \boldsymbol{I}^{-1}\zeta_3(s)\boldsymbol{I}^{-1} \right\|_2 \to 0.$$

Because τ_n^* is a bijection from $[\bar{s}, 1]$ onto itself and $\zeta_n = (\zeta_{1,n}, \zeta_{2,n}, \zeta_{3,n})$ is in $B(0, r\sqrt{n})$, we obtain that, for all $s \in [\bar{s}, 1]$ and $n \geq 1$, the random variable $\frac{\zeta_{3,n}(\tau_n^*(s))}{\sqrt{n}}$ is in the closed ball $B(0, r)$. In addition, we chose r such that $\|\boldsymbol{I}^{-1}\|_2 < 1/r$. Because the Frobenius norm is submultiplicative[6], it follows that

$$\left\| -\boldsymbol{I}^{-1}\frac{\zeta_{3,n}(\tau_n^*(s))}{\sqrt{n}} \right\|_2 < 1.$$

Therefore, using Theorem 4.16 in Dudley and Norvaiša (2011), we can expand the term $\left(\boldsymbol{I} + \frac{\zeta_{3,n}(\tau_n^*(s))}{\sqrt{n}}\right)^{-1}$ as a Neumann series. We obtain

$$\left(\boldsymbol{I} + \frac{\zeta_{3,n}(\tau_n^*(s))}{\sqrt{n}}\right)^{-1} = \left(Id_d + \boldsymbol{I}^{-1}\frac{\zeta_{3,n}(\tau_n^*(s))}{\sqrt{n}}\right)^{-1}\boldsymbol{I}^{-1} = \left(\sum_{k\geq 0}(-1)^k\boldsymbol{I}^{-k}\left(\frac{\zeta_{3,n}(\tau_n^*(s))}{\sqrt{n}}\right)^k\right)\boldsymbol{I}^{-1}.$$

For all $s \in [\bar{s}, 1]$ and $n \geq 1$,

$$\begin{aligned}&\sqrt{n}\left(\left(\boldsymbol{I} + \frac{\zeta_{3,n}(\tau_n^*(s))}{\sqrt{n}}\right)^{-1} - \boldsymbol{I}^{-1}\right) + \boldsymbol{I}^{-1}\zeta_3(s)\boldsymbol{I}^{-1}\\ &= \sqrt{n}\left(\sum_{k\geq 2}(-1)^k\boldsymbol{I}^{-k}\left(\frac{\zeta_{3,n}(\tau_n^*(s))}{\sqrt{n}}\right)^k\right)\boldsymbol{I}^{-1} - \boldsymbol{I}^{-1}\Big(\zeta_{3,n}(\tau_n^*(s)) - \zeta_3(s)\Big)\boldsymbol{I}^{-1}.\end{aligned}$$

The result follows from the fact that $\sup_s \|\zeta_{3,n}(\tau_n^*(s)) - \zeta_3(s)\|_2 \to 0$ and

$$\begin{aligned}&\sup_s \left\| \sqrt{n}\left(\sum_{k\geq 2}(-1)^k\boldsymbol{I}^{-k}\left(\frac{\zeta_{3,n}(\tau_n^*(s))}{\sqrt{n}}\right)^k\right)\boldsymbol{I}^{-1} \right\|_2\\ &\leq \frac{1}{\sqrt{n}}\left\|\boldsymbol{I}^{-1}\right\|_2^3 \sup_s \|\zeta_{3,n}(\tau_n^*(s))\|_2^2 \sum_{k\geq 0}\left(\frac{\left\|\boldsymbol{I}^{-1}\right\|_2}{\sqrt{n}}\sup_s \|\zeta_{3,n}(\tau_n^*(s))\|_2\right)^k.\end{aligned} \tag{S17}$$

We already know that the constant $\left\|\boldsymbol{I}^{-1}\right\|_2$ is finite. From Lemma 12.1 in Billingsley (1999), $\|\zeta_3\|_2$ is a finite constant since ζ is a càd-làg process with finite values in $\mathbb{D}_{[\bar{s},1]}$. It follows that $\sup_s \|\zeta_{3,n}(\tau_n^*(s))\|_2 \leq \|\zeta_3\|_2 + \sup_s \|\zeta_{3,n}(\tau_n^*(s)) - \zeta_3(s)\|_2$ converges to $\|\zeta_3\|_2$ and, for n large enough, the series above can be dominated by a convergent geometric series. Then (S17) converges to 0 and the result follows. □

[5] See Section 12 in Billingsley (1999) for more details on the Skorokhod topology.

[6] This property is a consequence of the Cauchy-Schwarz inequality. See e.g. Trefethen and Bau (1997), p.23.

Remark S3.6. *The map* $z \in \mathbb{R}^{2d+d^2} \mapsto \boldsymbol{q}(z)$ *is a quadratic form.*

Proof of Remarks 3.6 and S3.6. Remark that, for any A and J, the map $(\iota, u, I) \mapsto q(\iota, u, I; A, J)$ is a quadratic form but not a norm. Its unique associated symmetric bilinear form[7] is given by $\big((\iota, u, I), (\iota', u', I')\big) \mapsto \frac{1}{2}\big(q(\iota+\iota', u+u', I+I'; A, J) - q(\iota, u, I; A, J) - q(\iota, u, I; A, J)\big)$. From Theorem S3.3, we already know that the map $\boldsymbol{g}$ is linear. It follows that the map

$$\begin{aligned}(z, z') \mapsto & \frac{1}{2}\left(\boldsymbol{q}(z+z') - \boldsymbol{q}(z) - \boldsymbol{q}(z')\right) \\ = & \frac{1}{2}\left(q(\boldsymbol{g}(\Sigma(z+z')); \boldsymbol{I}^{-1}, \boldsymbol{U}) - q(\boldsymbol{g}(\Sigma z); \boldsymbol{I}^{-1}, \boldsymbol{U}) - q(\boldsymbol{g}(\Sigma z'); \boldsymbol{I}^{-1}, \boldsymbol{U})\right)\end{aligned}$$

is symmetric bilinear. The result follows. □

Theorem S3.7. *Under* H_0*, the process* Q_n^1 *converges weakly as* $n \to \infty$ *to the process* $(\frac{1}{s}\boldsymbol{q}(W_s))_{s\in[\bar{s},1]}$ *in* $\mathbb{D}([\bar{s}, 1], \mathbb{R})$*.*

Proof of Theorems 3.7 and S3.7. For all $s \in [\bar{s}, 1]$ and n large enough, from (10), we can reorganize the variable $Q_{s,n}^1$ as follows

$$\begin{aligned}Q_{s,n}^1 = & \frac{\lfloor sn \rfloor}{n}\left(\sqrt{n}(\hat{u}_{0,s} - \boldsymbol{u})^T\, \hat{A}_{0,s}^{-1}\, \sqrt{n}\hat{\iota}_{0,s} + \sqrt{n}\hat{\iota}_{0,s}^T\ \ {A_{0,s}^{-1}}^T\, \hat{U}_{0,s}\hat{A}_{0,s}^{-1}\, \sqrt{n}\hat{\iota}_{0,s} + \boldsymbol{u}^T\sqrt{n}\left(\hat{A}^{-1} - \boldsymbol{I}^{-1}\right)\, \sqrt{n}\hat{\iota}_{0,s}\right) \\ = & \frac{\lfloor sn \rfloor}{n} q(\sqrt{n}(\hat{\xi}_{0,s}' - \boldsymbol{\xi}'); \hat{A}_{0,s}^{-1}, \hat{U}_{0,s}),\end{aligned} \tag{S18}$$

with

$$\hat{U}_{0,s} := \frac{1}{\lfloor sn \rfloor}\sum_{i=1}^{\lfloor sn \rfloor} D_\theta^2(w \log f_1)(X_i, \theta_{0,s}')$$

and $\hat{\xi}_{0,s}' = (\hat{\iota}_{0,s}, \hat{u}_{0,s} - \boldsymbol{u}, \hat{A}_{0,s}^{-1})$ from its definition in Section 3.1.
We know from Theorem S3.3 that the process $(\sqrt{n}(\hat{\xi}_{0,s}' - \boldsymbol{\xi}'))_{s\in[\bar{s},1]}$ converges weakly to the process $(\frac{1}{s}\boldsymbol{g}(\Sigma W_s))_{s\in[\bar{s},1]}$. Recall that, by Proposition S2.8, $\hat{A}_{0,s}^{-1}$ converges a.s. to $\boldsymbol{I}^{-1}$, uniformly in $s \in [\bar{s}, 1]$. Further, by Assumption 2.5, Lemma S2.9 can be applied to $h(x, \theta) = D_\theta^2(w \log f_1)(x, \theta)$, taking the set Θ' as $\mathcal{O}$. Therefore $\hat{U}_{0,s}$ converges a.s. to $\boldsymbol{U}$, uniformly in $s \in [\bar{s}, 1]$. It follows by Slutsky's Theorem, that the process

$$\left(\sqrt{n}(\hat{\xi}_{0,s}' - \boldsymbol{\xi}'), \hat{A}_{0,s}^{-1}, \hat{U}_{0,s}\right)_{s\in[\bar{s},1]}$$

converges weakly to the process $(\frac{1}{s}\boldsymbol{g}(\Sigma W_s), \boldsymbol{I}^{-1}, \boldsymbol{U})_{s\in[\bar{s},1]}$. The map q being continuous, the result follows by the Continuous Mapping Theorem. □

Theorem S3.8. *Under* H_0 *and Assumptions 2.1-2.5, the test statistic*

$$S_n \xrightarrow[n\to\infty]{\mathcal{D}} \sup_{s\in[\bar{s},1-\bar{s}]} \frac{\boldsymbol{q}(W_s - sW_1)}{s(1-s)}$$

where $(W_s)_{s\in[0,1]}$ *is a standard* $2d+d^2$*-dimensional Brownian motion and the application* $\boldsymbol{q}$ *is defined in* (14)*.*

Before giving the proof of Theorems 3.8 and S3.8, we start by a result for quadratic forms.

Lemma S3.9. *Fix* $d \geq 1$*. If* $x \in \mathbb{R}^d \mapsto q(x)$ *is a quadratic form, then, for any* $x, y \in \mathbb{R}^d$ *and any real* $s \neq 0$*, the following equality holds:*

$$\frac{1}{s}q\left(x\right) + \frac{1}{1-s}q\left(y-x\right) - q(y) = \frac{q(x-sy)}{s(1-s)}.$$

[7] See for instance Section 41 in O'Meara (2000).

Proof. The unique symmetric bilinear form[8] associated to q is the application

$$(x,y) \in \mathbb{R}^d \times \mathbb{R}^d \mapsto b_q(x,y) = \frac{1}{2}\big(q(x+y) - q(x) - q(y)\big).$$

By definition of b_q, we have that $q(x) = b_q(x,x)$ and $q(x+y) = q(x) + q(y) + 2\, b_q(x,y)$. Further, for any real s and any $x \in \mathbb{R}^d$, $q(sx) = s^2 q(x)$. It follows that

$$\begin{aligned}
&\frac{1}{s}q\,(x) + \frac{1}{1-s}q\,(y-x) - q(y) \\
=&\frac{(1-s)q(x) + s(q(x) + q(y) - 2\, b_q(x,y)) - s(1-s)q(y)}{s(1-s)} \\
=&\frac{q(x) - 2s\, b_q(x,y) + s^2 q(y)}{s(1-s)} = \frac{q(x) + 2\, b_q(x,-sy) + q(-sy)}{s(1-s)}.
\end{aligned}$$

The result follows. □

We can now give below the proof of Theorems 3.8 and S3.8.

Proof of Theorems 3.8 and S3.8. Recall that from (S15) and (S18), for $s \in [\bar{s}, 1]$, $Q^1_{s,n}$ can be written as follows for $s \in [\bar{s}, 1]$:

$$Q^1_{s,n} = \frac{\lfloor sn \rfloor}{n} q\Big(\sqrt{n}\Big(\Phi(g(\hat{\xi}_{0,t}; \hat{A}^{-1}_{0,t}, \hat{J}_{0,t})_{t\in[\bar{s},1]}) - \Phi(\boldsymbol{\xi}_{\cdot})\Big)_s; \hat{A}^{-1}_{0,s}, \hat{U}_{0,s}\Big).$$

We remark from (7) that $Q^1_{s,n}$ and $Q^2_{s,n}$ have a similar structure and differ only from the fact that $Q^1_{s,n}$ depends from the sample $(X_1, \ldots, X_{\lfloor sn \rfloor})$ and the estimator $\hat{\theta}_{0,s}$, while $Q^2_{s,n}$ depends from the sample $(X_{\lfloor sn \rfloor + 1}, \ldots, X_n)$ and the estimator $\hat{\theta}_{1,s}$. With the definition of $\hat{\xi}_{0,s}$ in (9), we can write $Q^2_{s,n}$ for $s \in [\bar{s}, 1-\bar{s}]$ as follows:

$$Q^2_{s,n} = \frac{n - \lfloor sn \rfloor}{n} q\left(\sqrt{n}\left(\tilde{\Phi}\left(g\left(\frac{n}{n - \lfloor tn \rfloor}\hat{\xi}_{0,1} - \frac{\lfloor tn \rfloor}{n - \lfloor tn \rfloor}\hat{\xi}_{0,t}; \hat{A}^{-1}_{t,1}, \hat{J}_{t,1}\right)_{t\in[\bar{s},1-\bar{s}]}\right) - \tilde{\Phi}((\boldsymbol{\xi})_{s\in[\bar{s},1-\bar{s}]})\right)_s; \hat{A}^{-1}_{s,1}, \hat{U}_{s,1}\right)$$

where $\tilde{\Phi}$ is the map from the set of càd-làg paths $\mathbb{D}([\bar{s}, 1-\bar{s}], (\mathbb{R}^d)^2 \times GL_d(\mathbb{R}))$ onto itself that coincide with Φ on $[\bar{s}, 1-\bar{s}]$, i.e. for $s \in [\bar{s}, 1-\bar{s}]$ and $x \in \mathbb{D}([\bar{s}, 1-\bar{s}], (\mathbb{R}^d)^2 \times GL_d(\mathbb{R}))$, $\tilde{\Phi}(x)_s := \varphi(x_s)$. In addition, the random variable $Q^1_{1,n}$ can be written as:

$$Q^1_{1,n} = q\Big(\sqrt{n}\Big(\varphi(g(\hat{\xi}_{0,1}; \hat{A}^{-1}, \hat{J}_{0,1})) - \varphi(\boldsymbol{\xi})\Big); \hat{A}^{-1}, \hat{U}_{0,1}\Big).$$

From the three equations above, the process $(Q^1_{s,n}, Q^2_{s,n}, -Q^1_{1,n})_{s\in[\bar{s},1-\bar{s}]}$ can be seen as a function of the triple process

$$\left(\hat{\xi}_{0,s}, \frac{n}{n - \lfloor sn \rfloor}\hat{\xi}_{0,1} - \frac{\lfloor sn \rfloor}{n - \lfloor sn \rfloor}\hat{\xi}_{0,s}, \hat{\xi}_{0,1}\right)_{s\in[\bar{s},1-\bar{s}]}. \tag{S19}$$

Recall that, by Lemma S3.1, the process $\sqrt{n}\left(\hat{\xi}_{0,s} - \boldsymbol{\xi}\right)_{s\in[\bar{s},1]}$ converges weakly to $(\frac{1}{s}\Sigma W_s)_{s\in[\bar{s},1]}$ in $\mathbb{D}_{[\bar{s},1]}$. Then a similar central limit result holds for the triple defined in (S19) and, by a succession of composition of the applications g, Φ and q along with arguments based on Slutsky's Theorem and the Continuous Mapping Theorem, the result obtained for Q^1_n is extended to the process $(Q^1_{s,n}, Q^2_{s,n}, -Q^1_{1,n})_{s\in[\bar{s},1-\bar{s}]}$. To show this, we reuse the arguments of Theorem S3.3. The functional delta method on the Skorokhod metric space is still

[8] See for instance Section 41 in O'Meara (2000).

applicable and the process

$$\Bigg(\sqrt{n}(\Phi(g(\hat{\xi}_{0,t};\hat{A}_{0,t}^{-1},\hat{J}_{0,t})_{t\in[\bar{s},1]})-\Phi(\boldsymbol{\xi}_{\cdot}))_s,$$

$$\sqrt{n}\left(\tilde{\Phi}\left(g\left(\frac{n}{n-\lfloor tn\rfloor}\hat{\xi}_{0,1}-\frac{\lfloor tn\rfloor}{n-\lfloor tn\rfloor}\hat{\xi}_{0,t};\hat{A}_{t,1}^{-1},\hat{J}_{t,1}\right)_{t\in[\bar{s},1-\bar{s}]}\right)-\tilde{\Phi}((\boldsymbol{\xi})_{t\in[\bar{s},1-\bar{s}]})\right)_s$$

$$\sqrt{n}(\varphi(g(\hat{\xi}_{0,1};\hat{A}^{-1},\hat{J}_{0,1}))-\varphi(\boldsymbol{\xi}))\Bigg)_{s\in[\bar{s},1-\bar{s}]}$$

converges weakly to the process

$$\left(\boldsymbol{g}\left(\frac{1}{s}\Sigma W_s\right),\boldsymbol{g}\left(\frac{1}{1-s}\Sigma\left(W_1-W_s\right)\right),\boldsymbol{g}(\Sigma W_1)\right)_{s\in[\bar{s},1-\bar{s}]}$$

in the Skorokhod metric space of càd-làg functions on $[\bar{s},1-\bar{s}]$ with values in $(\mathbb{R}^d)^2\times GL_d(\mathbb{R})$. Then, with a simple extension of the arguments of Theorem S3.7, the triple $\left(Q_{s,n}^1,Q_{s,n}^2,-Q_{1,n}^1\right)_{s\in[\bar{s},1-\bar{s}]}$ converges as follows:

$$\left(Q_{s,n}^1,Q_{s,n}^2,-Q_{1,n}^1\right)_{s\in[\bar{s},1-\bar{s}]}\xrightarrow[n\to\infty]{\mathcal{D}}\left(\frac{1}{s}\boldsymbol{q}\left(W_s\right),\frac{1}{1-s}\boldsymbol{q}\left(W_1-W_s\right),-\boldsymbol{q}(W_1)\right)_{s\in[\bar{s},1-\bar{s}]}.$$

From Remark S3.6, $\boldsymbol{q}$ is a quadratic form and, from Lemma S3.9,

$$\frac{1}{s}\boldsymbol{q}\left(W_s\right)+\frac{1}{1-s}\boldsymbol{q}\left(W_1-W_s\right)-\boldsymbol{q}(W_1)=\frac{\boldsymbol{q}(W_s-sW_1)}{s(1-s)},\ s\in[\bar{s},1-\bar{s}].$$

The result follows from a last application of the Continuous Mapping Theorem to the application that sums the elements of the triple above and takes the supremum over $[\bar{s},1-\bar{s}]$. □

Theorem S3.10. *Under H_0 and Assumptions 2.1-2.5, the test statistic*

$$S_n^*\xrightarrow[n\to\infty]{\mathcal{D}}\sup_{s\in[\bar{s},1-\bar{s}]}\frac{\boldsymbol{q}^*(W_s-sW_1)}{s(1-s)}$$

where $(W_s)_{s\in[0,1]}$ is a standard $3d+d^2$-dimensional Brownian motion and the application $\boldsymbol{q}^*$ is defined in (17).

Proof of Theorems 4.1 and S3.10. The proof follows the same logic as the proof of Theorem S3.8. The arguments are based on a functional delta method (Corollary S3.5) and multiple applications of Slutsky's Theorem and the Continuous Mapping Theorem. □

Proposition S3.11. *Under Assumption 2.1 and with the parameter set Θ defined above, the validity conditions of Theorems S3.8 and S3.10 hold for a finite Gaussian mixture.*

Proof of Proposition 5.2 and S3.11. The result is obtained as soon as we show that Assumptions 2.2-2.5 are valid. First of all, the conditions of the model introduced in Section 2.1 and Assumptions 2.2 and 2.3 are standard prerequisites for limit results of likelihood based estimators (McLachlan and Peel (2000)). In particular, from Example 6.10 in Lehmann and Casella (1998), one sees that the assumptions of Theorem 5.1 in Lehmann and Casella (1998) hold for identifiable Gaussian mixtures. With condition 1. above, it follows that Assumptions 2.2 and 2.3 hold.
With condition 2. above, Theorem 3.3 in Hathaway (1985) ensures that Assumption 2.4 is valid, i.e. the estimator $\hat{\theta}$ is strongly consistent.
Since the parameter set Θ is a convex open subset of $\mathbb{R}^d$ that contains the true parameter $\boldsymbol{\theta}$, it is possible to find a bounded convex set $\Theta'\subset\Theta$ such that $\boldsymbol{\theta}$ is in the interior of Θ'. We show that Assumption 2.5 is valid for this set Θ'. First we recall that, $\mathbb{E}_{H_0}[|X_1|^k]$ is finite for all $k\geq 0$. Thus it is sufficient to show that we

can find some function $\theta \mapsto \kappa(\theta)$ with positive values, not depending on x and continuous on Θ', such that, for all $1 \leq i, j \leq d$, θ in Θ' and $x \in \mathbb{R}$,

$$\left|D^2_\theta(w \log f_1)(x, \theta)_{i,j}\right| \leq \sum_{k=0}^{6} \kappa(\theta)|x|^k. \tag{S20}$$

On the one hand, since $x \in \mathbb{R} \mapsto f_1(x, (\mu_1, \sigma_1))$ is the density function of a Gaussian random variable, $\log f_1(x, (\mu_1, \sigma_1))$ can be written as a second-order polynomial of x. Its coefficients are infinitely differentiable functions of (μ_1, σ_1) on $\mathbb{R} \times \mathbb{R}^+_*$. It follows that the absolute values of the first and second order partial derivatives of $f_1(x, (\mu_1, \sigma_1))$ can be bounded by a second-order polynomial as on the left side of (S20). On the other hand, by definition, the weight function $w(x, \theta)$ takes its values in $[0, 1]$ for all x and all θ. In order to conclude the proof, we need only to bound the absolute value of the first and second partial derivatives of $w(x, \theta)$.
From the definition of w in (2), its first order partial derivatives can be written as

$$\frac{\partial}{\partial \theta_i} w(x, \theta) = \frac{\frac{\partial}{\partial \theta_i} \tilde{f}_1(x, \theta)}{f(x, \theta)} - w(x, \theta) \sum_{k=1}^{m} \frac{\frac{\partial}{\partial \theta_i} \tilde{f}_k(x, \theta)}{f(x, \theta)}, \ 1 \leq i \leq d$$

where $\tilde{f}_k : (x, \theta) \mapsto p_k f_k(x, \lambda_k)$. We can show that

$$\frac{\frac{\partial}{\partial \theta_i} \tilde{f}_k(x, \theta)}{f(x, \theta)} = \frac{p_k f_k(x, \lambda_k)}{f(x, \theta)} \kappa_{i,k}(x, \theta),$$

where $\kappa_{i,k}(x, \theta)$ is a second-order polynomial of x with coefficients that are infinitely differentiable functions of θ on Θ. Moreover we recognize that $p_k f_k(x, \lambda_k)/f(x, \theta)$ is another weight function that takes its values in $[0, 1]$ for all x and all θ. Thus $|\frac{\partial}{\partial \theta_i} w(x, \theta)|$ can be bounded by a second-order polynomial as on the left side of (S20). With similar arguments, the absolute value of the second partial derivatives of $w(x, \theta)$ can be bounded by a polynomial of degree four as in the left side of (S20). The result follows. □

S4 The constant $\boldsymbol{u}$

The constant $\boldsymbol{u} = \mathbb{E}_{H_0}[D_\theta(w \log f_1)(X_1, \boldsymbol{\theta})]$, defined in (6), plays a central role in the proof of Theorems S3.8 and S3.10 since most of the technical difficulties in Section 3 emerge only when $\boldsymbol{u} \neq 0$. This is the case in general, as illustrated in the following numerical example.
With the notations of Section 5, we consider a numerical simulation for an univariate Gaussian mixture with 3 components defined by

$$\boldsymbol{\theta} := \big((1/3, 1/3), (-1.00, 0.25), (0.00, 0.25), (1.00, 0.25)\big).$$

For 10^5 simulations, the Monte-Carlo approximation of $\mathbb{E}_{H_0}\left[\frac{\partial}{\partial \mu_2}(w \log f_1)(X_1, \boldsymbol{\theta})\right]$ converges to a non null limit (Figure S1). This illustrates that $\boldsymbol{u}$ is not null in general.

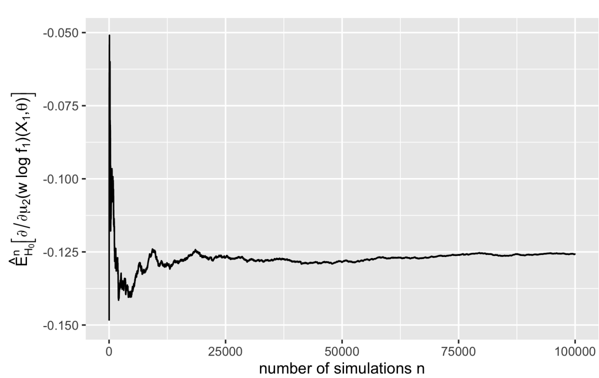
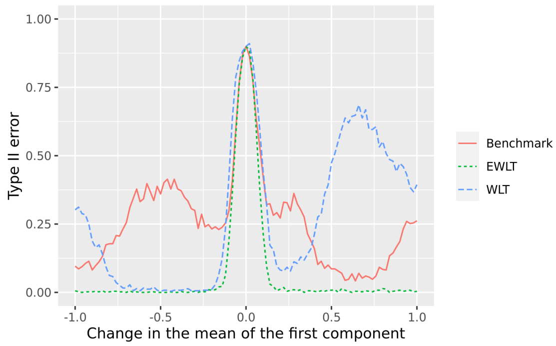
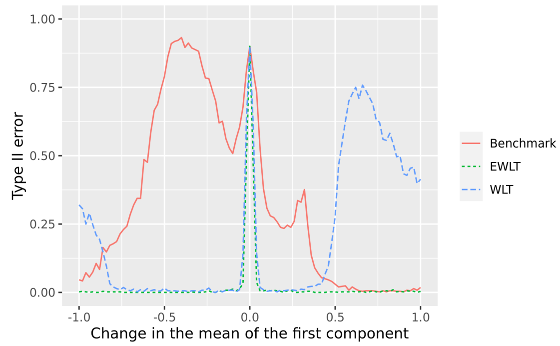


Figure S1: Convergence of the Monte-Carlo approximation of $\mathbb{E}_{H_0}\left[\frac{\partial}{\partial \mu_2}(w \log f_1)(X_1, \boldsymbol{\theta})\right]$.

S5 Additional numerical results

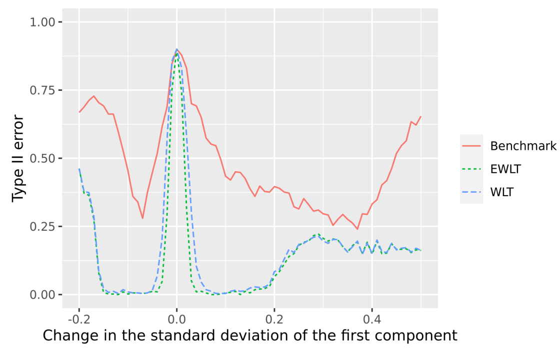


(a) Change in the mean (1k obs.)



(b) Change in the mean (10k obs.)

(c) Change in the standard deviation (1k obs.)

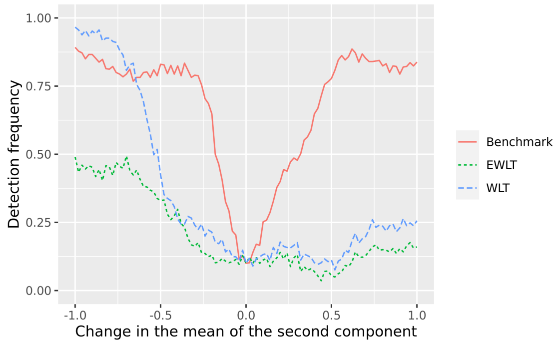


(d) Change in the standard deviation (10k obs.)

(e) Change in the weight parameter (1k obs.)

(f) Change in the weight parameter (10k obs.)

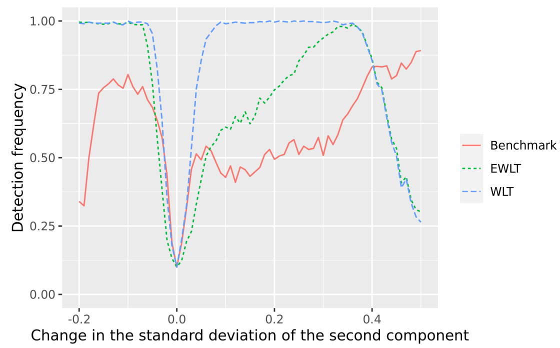
Figure S2: Type II error for a change in the first component



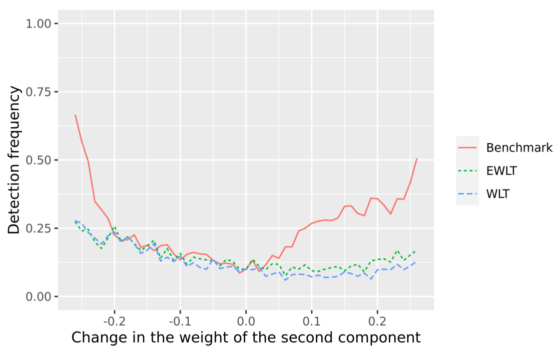
(a) Change in the mean (1k obs.)

(b) Change in the mean (10k obs.)

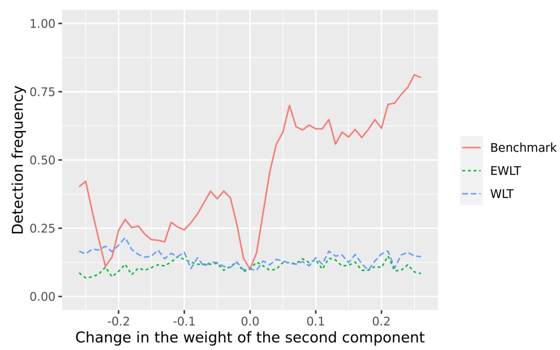
(c) Change in the standard deviation (1k obs.)



(d) Change in the standard deviation (10k obs.)

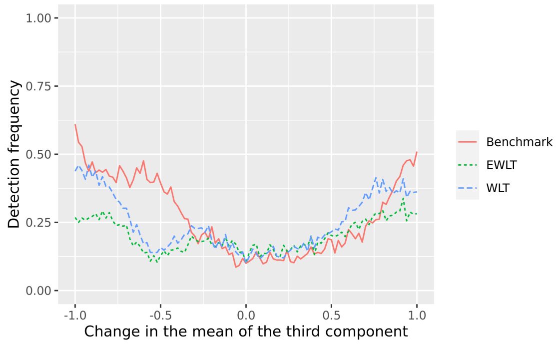


(e) Change in the weight parameter (1k obs.)



(f) Change in the weight parameter (10k obs.)

Figure S3: Detection frequency for a change in the second component



(a) Change in the mean (1k obs.)

(b) Change in the mean (10k obs.)

(c) Change in the standard deviation (1k obs.)

(d) Change in the standard deviation (10k obs.)

(e) Change in the weight parameter (1k obs.)

(f) Change in the weight parameter (10k obs.)

Figure S4: Detection frequency for a change in the third component

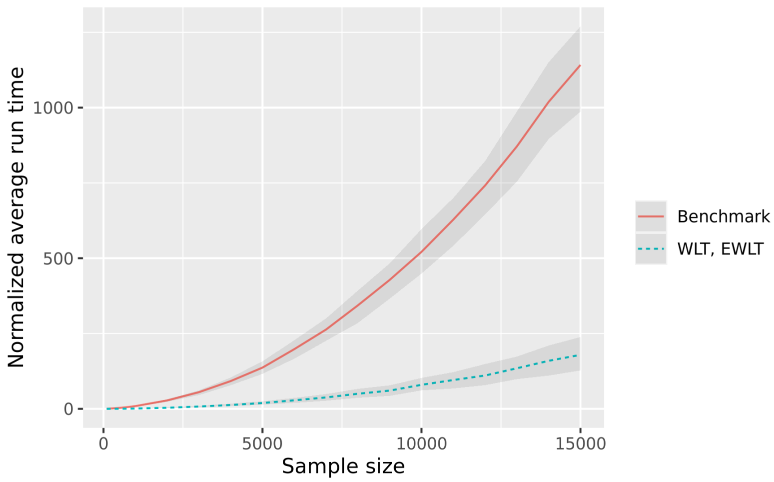


Figure S5: Run time of the benchmark and the WL and EWL tests for an increasing sample size.

S6 Glossary of notations

$\hat{A}$	$= -\frac{1}{n}\sum_{i=1}^{n}\left(D_\theta^2(\log f)(X_i,\boldsymbol{\theta}) + \frac{1}{2}\sum_{l=1}^{d}\left(\hat{\theta}_l - \boldsymbol{\theta}_l\right) D_\theta^3(\log f)(X_i,\theta')_{l,.,.}\right)$ defined in (S4).
$\hat{A}_{0,s}$ or $\hat{A}_{s,1}$	$\hat{A}_{s_1,s_2} = -\frac{1}{\lfloor s_2 n\rfloor - \lfloor s_1 n\rfloor}\sum_{i=\lfloor s_1 n\rfloor+1}^{\lfloor s_2 n\rfloor}\Big(D_\theta^2(\log f)(X_i,\boldsymbol{\theta})$ $+\frac{1}{2}\sum_{l=1}^{d}(\hat{\theta}_{s_1,s_2;l} - \boldsymbol{\theta}_l)D_\theta^3(\log f)(X_i,\theta'_{s_1,s_2})_{l,.,.}\Big)$ defined in (S5).
$\boldsymbol{\beta}$	$= \mathbb{E}_{H_0}\left[\log f_1(Y,\boldsymbol{\lambda}_1)\right]$, with Y a r.v. with density $f_1(.,\boldsymbol{\lambda}_1)$. See Section 4.
$c_{s,n}$	$= \sum_{i=1}^{\lfloor sn\rfloor} w(X_i,\hat{\theta}_{0,s}) + \sum_{i=\lfloor sn\rfloor+1}^{n} w(X_i,\hat{\theta}_{s,1})$, defined in Section 4.
d	dimension of the parameter set Θ, see Section 2.1.
$\mathbb{D}_{[\bar{s},1]}$	$= \mathbb{D}([\bar{s},1],(\mathbb{R}^d)^2\times gl_d(\mathbb{R}))$. Skorokhod metric space of càd-làg paths, see Lemma S3.1.
$f_k(.,\lambda_k)$	density function of the k-th component, see Section 2.1.
$g(.,.,.;.,.)$	map defined in (S12) for $(\iota,u,I;A,J)\in(\mathbb{R}^d)^2\times(gl_d(\mathbb{R}))^2\times\mathbb{R}^{d\times d\times d}$ by $(\iota,u,I;A,J)\mapsto\left(\iota,u,I-\frac{1}{2}\sum_{l=1}^{d}(\iota^T(A^T)_{.,l})J_{.,.,l}\right)$.
$\boldsymbol{g}(.,.,.)$	linear map defined in (S11) for $(\iota,u,I)\in(\mathbb{R}^d)^2\times gl_d(\mathbb{R})$ by $(\iota,u,I)\mapsto\left(\iota,u,-\boldsymbol{I}^{-1}\boldsymbol{(}I-\frac{1}{2}\sum_{l=1}^{d}(\iota^T(\boldsymbol{I}^{-1\,T})_{.,l})\boldsymbol{J}_{.,.,l}\boldsymbol{)}\boldsymbol{I}^{-1}\right)$.
$\hat{\iota}_{0,s}$	$= \frac{1}{\lfloor sn\rfloor}\sum_{i=1}^{\lfloor sn\rfloor} D_\theta(\log f)(X_i,\boldsymbol{\theta})$, defined in (9).
$\hat{I}_{0,s}$	$= -\frac{1}{\lfloor sn\rfloor}\sum_{i=1}^{\lfloor sn\rfloor} D_\theta^2(\log f)(X_i,\boldsymbol{\theta})$, defined in (9).
$\boldsymbol{I}$	$= -\mathbb{E}_{H_0}\left[D_\theta^2(\log f)(X_1,\boldsymbol{\theta})\right]$. Fisher information matrix defined in (1).
$\hat{J}_{0,s}$	$= \frac{1}{\lfloor sn\rfloor}\sum_{i=1}^{\lfloor sn\rfloor} D_\theta^3(\log f)(X_i,\theta'_{0,s})$ defined in Lemma S3.2.
$\boldsymbol{J}$	$= \mathbb{E}_{H_0}\left[D_\theta^3(\log f)(X_1,\boldsymbol{\theta})\right]$ defined in (11).
λ_k	density function parameter of the k-th component, see Section 2.1.
$\hat{\lambda}_1$	see $\hat{\theta}$.
$\hat{\lambda}_{0,s,1}$ or $\hat{\lambda}_{s,1,1}$	see $\hat{\theta}_{0,s}$ or $\hat{\theta}_{s,1}$.
Λ_n	$= (\Lambda_{s,n})_{s\in[\bar{s},1-\bar{s}]}$. Detection process of the WL test, defined in (4).
$\Lambda_{s,n}$	$= Q_{s,n}^1 + Q_{s,n}^2 - Q_{1,n}^1$ defined in (4) (see also (7)).
Λ_n^*	$= (\Lambda_{s,n}^*)_{s\in[\bar{s},1-\bar{s}]}$. Detection process of the EWL test, defined in (16).
$\Lambda_{s,n}^*$	$= \frac{c_{1,n}}{c_{s,n}}\Lambda_{s,n} + \left(\frac{c_{1,n}}{c_{s,n}} - 1\right)\sum_{i=1}^{n} w(X_i,\hat{\theta})\log f_1(X_i,\hat{\lambda}_1)$ defined in (16).
m	number of components in the mixture, see Section 2.1.
p_k	weight of the k-th component in the mixture, see Section 2.1.
$\Phi(.)$	map defined in (S14) by $(x_s)_{s\in[\bar{s},1]}\mapsto(\varphi(x_s))_{s\in[\bar{s},1]}$.
$q(.,.,.;.,.)$	map defined in (13) by $(\iota,u,I;A,J)\mapsto u^T A\iota + \iota^T A^T J A\iota + \boldsymbol{u}^T I\iota$.
$\boldsymbol{q}(.)$	map defined in (14) by $z\in\mathbb{R}^{2d+d^2}\mapsto q(\boldsymbol{g}(\Sigma z);\boldsymbol{I}^{-1},\boldsymbol{U})$.
$\boldsymbol{q}^*(.)$	map defined in (17) for $z\in\mathbb{R}^{3d+d^2}$ by $z\mapsto q^*(\boldsymbol{g}^*(\Sigma^* z);\boldsymbol{I}^{-1},\boldsymbol{U},\boldsymbol{V})$.
Q_n^1	$= (Q_{s,n}^1)_{s\in[\bar{s},1]}$ defined in Section 3.
$Q_{s,n}^1$	$= \sum_{i=1}^{\lfloor sn\rfloor}\left(w(X_i,\hat{\theta}_{0,s})\log f_1(X_i,\hat{\lambda}_{0,s,1}) - w(X_i,\boldsymbol{\theta})\log f_1(X_i,\boldsymbol{\lambda}_1)\right)$ $-\boldsymbol{u}^T\boldsymbol{I}^{-1}\sum_{i=1}^{\lfloor sn\rfloor} D_\theta(\log f)(X_i,\boldsymbol{\theta})$ defined in Section 3.
Q_n^2	$= (Q_{s,n}^2)_{s\in[\bar{s},1-\bar{s}]}$ defined in Section 3.
$Q_{s,n}^2$	$= \sum_{i=\lfloor sn\rfloor+1}^{n}\left(w(X_i,\hat{\theta}_{s,1})\log f_1(X_i,\hat{\lambda}_{s,1,1}) - w(X_i,\boldsymbol{\theta})\log f_1(X_i,\boldsymbol{\lambda}_1)\right)$ $-\boldsymbol{u}^T\boldsymbol{I}^{-1}\sum_{i=\lfloor sn\rfloor+1}^{n} D_\theta(\log f)(X_i,\boldsymbol{\theta})$ defined in Section 3.
$\bar{s}$	$\in(0,0.5)$. We assume that the change-point is in $[\bar{s},1-\bar{s}]$ if it exists (Assumption 2.1).
S_n	$= \sup_{s\in[\bar{s},1-\bar{s}]}\Lambda_{s,n}$ defined in (5). Detection statistic of the WL test.
S_n^*	$= \sup_{s\in[\bar{s},1-\bar{s}]}\Lambda_{s,n}^*$ defined in Section 4. Detection statistic of the EWL test.
Σ	unique positive semi-definite square root of the covariance matrix of the i.i.d. terms in the average $\hat{\xi}_{0,1}$, see Lemma S3.1.

$\hat{\theta}$	$= (\hat{p}_1, \ldots, \hat{p}_{m-1}, \hat{\lambda}_1, \ldots, \hat{\lambda}_m)$. One consistent sequence of solutions of the likelihood equations over the sample $X_1, \ldots, X_n$, see Section 2.1.
$\hat{\theta}_{0,s}$ or $\hat{\theta}_{s,1}$	$\hat{\theta}_{s_1,s_2} = (\hat{p}_{s_1,s_2,1}, \ldots, \hat{p}_{s_1,s_2,m-1}, \hat{\lambda}_{s_1,s_2,1}, \ldots, \hat{\lambda}_{s_1,s_2,m})$. One consistent sequence of solutions of the likelihood equations over the sample $X_{\lfloor s_1 n \rfloor+1}, \ldots, X_{\lfloor s_2 n \rfloor}$, see Section 2.1.
$\boldsymbol{\theta}$	true parameter under the null hypothesis, see Section 2.1.
θ'	point on the segment $[\hat{\theta}, \boldsymbol{\theta}]$, see (S3).
$\theta'_{0,s}$ or $\theta'_{s,1}$	θ'_{s_1,s_2} is a point on the segment $[\hat{\theta}_{s_1,s_2}, \boldsymbol{\theta}]$, see (S5).
Θ_0	parameter set for the weights of the mixture components, see Section 2.1.
Θ'	convex subset of Θ for which Assumption 2.5 holds.
$\hat{u}_{0,s}$	$= \frac{1}{\lfloor sn \rfloor} \sum_{i=1}^{\lfloor sn \rfloor} D_\theta(w \log f_1)(X_i, \boldsymbol{\theta})$, defined in (9).
$\boldsymbol{u}$	$= \mathbb{E}_{H_0}\left[D_\theta(w \log f_1)(X_1, \boldsymbol{\theta})\right]$ defined in (6).
$\hat{U}_{0,s}$	$= \frac{1}{\lfloor sn \rfloor} \sum_{i=1}^{\lfloor sn \rfloor} D^2_\theta(w \log f_1)(X_i, \theta'_{0,s})$, see proof of Theorem S3.7.
$\boldsymbol{U}$	$= \mathbb{E}_{H_0}\left[D^2_\theta(w \log f_1)(X_1, \boldsymbol{\theta})\right]$ defined in (15).
$\hat{v}_{0,s}$	$= \frac{1}{\lfloor sn \rfloor} \sum_{i=1}^{\lfloor sn \rfloor} D_\theta w(X_i, \boldsymbol{\theta})$, see Section 4.
$\boldsymbol{v}$	$= \mathbb{E}_{H_0}\left[D_\theta w(X_1, \boldsymbol{\theta})\right]$, see Section 4.
$\boldsymbol{V}$	$= \mathbb{E}_{H_0}\left[D^2_\theta w(X_1, \boldsymbol{\theta})\right]$, see Section 4.
$w(x, \theta)$	$= (p_1 f_1(x, \lambda_1))/(f(x, \theta))$ defined in (2).
$w \log f_1$	application defined in (3) by $(x, \theta) \mapsto w(x, \theta) \log f_1(x, \lambda_1)$.
$\hat{\xi}_{0,s}$	$= (\hat{\imath}_{0,s}, \hat{u}_{0,s} - \boldsymbol{u}, \hat{I}_{0,s})$, defined in (9).
$\boldsymbol{\xi}$	$= (0, 0, \boldsymbol{I})$ defined in (S9).
$\boldsymbol{\xi}_\cdot$	constant process s.t. $\boldsymbol{\xi}_s = \boldsymbol{\xi}$ for all $s \in [\bar{s}, 1]$, see proof of Theorem S3.3.
$\hat{\xi}'_{0,s}$	$= (\hat{\imath}_{0,s}, \hat{u}_{0,s} - \boldsymbol{u}, \hat{A}^{-1}_{0,s})$ defined in Section 3.1.
$\boldsymbol{\xi}'$	$= (0, 0, \boldsymbol{I}^{-1})$, see Theorem S3.3.
$\hat{\xi}''_{0,\cdot}$	$= (\hat{\xi}''_{0,s})_{s \in [\bar{s}, 1]}$, see proof of Theorem S3.3.
$\hat{\xi}''_{0,s}$	$g(\hat{\xi}_{0,s}; \hat{A}^{-1}_{0,s}, \hat{J}_{0,s})$ if (S13) holds, $\boldsymbol{\xi}$ otherwise. See proof of Theorem S3.3.
$\hat{\xi}^*_{0,s}$	$= (\hat{\imath}_{0,s}, \hat{u}_{0,s} - \boldsymbol{u}, \hat{v}_{0,s} - \boldsymbol{v}, \hat{I}_{0,s})$, see Section 4.
$\boldsymbol{\xi}^*$	$= (0, 0, 0, \boldsymbol{I})$, see Section 4.

References

Abraham, R., Marsden, J. E., and Ratiu, T. (1988). *Manifolds, tensor analysis, and applications*, volume 75 of *Applied Mathematical Sciences*. Springer-Verlag, New York, second edition.

Billingsley, P. (1999). *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition. A Wiley-Interscience Publication.

Dehling, H., Franke, B., Kott, T., and Kulperger, R. (2014). Change point testing for the drift parameters of a periodic mean reversion process. *Stat. Inference Stoch. Process.*, 17(1):1–18.

Dudley, R. M. and Norvaiša, R. (2011). *Concrete functional calculus*. Springer Monographs in Mathematics. Springer, New York.

Hathaway, R. J. (1985). A constrained formulation of maximum-likelihood estimation for normal mixture distributions. *Ann. Statist.*, 13(2):795–800.

Lehmann, E. L. and Casella, G. (1998). *Theory of point estimation*. Springer Texts in Statistics. Springer-Verlag, New York, second edition.

McLachlan, G. and Peel, D. (2000). *Finite mixture models*. Wiley Series in Probability and Statistics: Applied Probability and Statistics. Wiley-Interscience, New York.

O'Meara, O. T. (2000). *Introduction to quadratic forms*. Classics in Mathematics. Springer-Verlag, Berlin. Reprint of the 1973 edition.

Trefethen, L. N. and Bau, III, D. (1997). *Numerical linear algebra*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

van der Vaart, A. W. (1998). *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.