7. Supplement to "Multiscale change-point inference"

In this supplement we collect the proofs of the main assertions in the paper together with some auxiliary lemmas. We further give more general versions of some results in the paper.

7.1. **Large deviation and power estimates.** We begin by recalling some large deviation results for exponential families. By $D(\theta||\tilde{\theta})$ we will denote the *Kullback-Leibler divergence* of F_θ and $F_{\tilde{\theta}}$, i.e.

$$D(\theta||\tilde{\theta}) = \int_{\mathbb{R}} f_\theta(x) \log \frac{f_\theta(x)}{f_{\tilde{\theta}}(x)} \,\mathrm{d}\nu(x) = \psi(\tilde{\theta}) - \psi(\theta) - (\tilde{\theta} - \theta) m(\theta). \tag{43}$$

With the techniques used in (**?**, Thm.7.1) it is readily seen that for a sequence of independent and F_θ-distributed r.v. $Y_1, \ldots, Y_n$ one has that

$$\mathbf{P}\left(\overline{Y} - m(\theta) \geq \eta\right) \leq e^{n(D(\theta||\theta+\varepsilon) - \eta\varepsilon)} \tag{44}$$

for all $\varepsilon > 0$ such that $\theta + \varepsilon \in \Theta$. The following restatement of inequality (44) turns out to be very useful.

Lemma 7.1. Let $Y = (Y_1, \ldots, Y_n)$ be independent random variables such that $Y_i \sim F_\theta$ and assume that $\delta > 0$ is such that $\theta + \delta \in \Theta$. Then,

$$\mathbf{P}(m^{-1}(\overline{Y}) \geq \theta + \delta) \leq e^{-nD(\theta+\delta||\theta)}.$$

Proof. First observe that according to (44)

$$\begin{aligned}\mathbf{P}(m^{-1}(\overline{Y}) \geq \theta + \delta) &= \mathbf{P}(\overline{Y} - m(\theta) \geq m(\theta+\delta) - m(\theta)) \\ &\leq \exp(n(D(\theta||\theta+\delta) - (m(\theta+\delta) - m(\theta))\delta)).\end{aligned}$$

Now it follows from (43) that

$$\begin{aligned}D(\theta||\theta+\delta) - (m(\theta+\delta) - m(\theta))\delta &= \psi(\theta+\delta) - \psi(\theta) - m(\theta+\delta)\delta \\ &= -(\psi(\theta) - \psi(\theta+\delta) - (\theta - (\theta+\delta))m(\theta+\delta)) \\ &= -D(\theta+\delta||\theta).\end{aligned}$$

□

From (44) we further derive a basic power estimate for the likelihood ratio statistic (4).

Lemma 7.2. Let $Y = (Y_1, \ldots, Y_n)$ be independent random variables such that $Y_i \sim F_\theta$ and assume that $\delta \in \mathbb{R}$ is such that $\theta + \delta \in \Theta$. Then,

$$\mathbf{P}\left(T_1^n(Y, \theta+\delta) \geq q\right) \geq 1 - \exp\left(n \inf_{\varepsilon \in [0,\delta]} \left[D(\theta||\theta+\varepsilon) - \frac{\varepsilon}{\delta} D(\theta||\theta+\delta) + \frac{\varepsilon q}{n\delta}\right]\right).$$

Proof. For

$$J(\overline{Y},\theta) = \phi(\overline{Y}) - \big(\overline{Y}\theta - \psi(\theta)\big)$$

we obtain

$$J(\overline{Y},\theta+\delta) = J(\overline{Y},\theta) - \delta\overline{Y} - \psi(\theta) + \psi(\theta+\delta). \tag{45}$$

Thus, we have

$$\begin{aligned}\Pi(q,n,\delta) &:= \mathbf{P}\big(T_1^n(Y,\theta+\delta) \geq q\big)\\ &= \mathbf{P}\left(J(\overline{Y},\theta+\delta) \geq \frac{q}{n}\right)\\ &= \mathbf{P}\left(J(\overline{Y},\theta) - \delta\overline{Y} \geq \frac{q}{n} - \psi(\theta+\delta) + \psi(\theta)\right)\\ &\geq \mathbf{P}\left(-\delta\overline{Y} \geq \frac{q}{n} - \psi(\theta+\delta) + \psi(\theta)\right),\end{aligned}$$

where in the last inequality holds since $J(x,\theta) \geq 0$ for all $x \in \mathbb{R}$ and $\theta \in \Theta$. Now, let us first assume that $\delta > 0$. Then by (43) we find

$$\mathbf{P}\left(-\delta\overline{Y} \geq \frac{q}{n} - \psi(\theta+\delta) + \psi(\theta)\right) = \mathbf{P}\left(\overline{Y} - m(\theta) \leq -\frac{q}{\delta n} + \frac{D(\theta||\theta+\delta)}{\delta}\right). \tag{46}$$

Combining this with the large deviation inequality (44) yields

$$\Pi(q,n,\delta) \geq 1 - \exp\left(n(D(\theta||\theta+\varepsilon) - \frac{\varepsilon}{\delta}D(\theta||\theta+\delta)) + \frac{\varepsilon q}{\delta}\right),$$

for all $0 \leq \varepsilon \leq \delta$. The case when $\delta < 0$ follows analogously. □

For Gaussian observations the estimate can be made explicit.

Lemma 7.3. Let $Y_1,\ldots,Y_n$ be i.i.d. random variables such that $Y_1 \sim \mathcal{N}(0,1)$ and let $x_+ = \max(0,x)$ for $x \in \mathbb{R}$. Then,

$$\mathbf{P}\left(T_1^n(Y,\delta) \geq q\right) \geq 1 - \exp\left(-\frac{1}{8}\left(\sqrt{n}\delta - \sqrt{2q}\right)_+^2\right). \tag{47}$$

Proof. Since $D(\theta||\theta+\varepsilon) = \varepsilon^2/2$ we find that

$$\inf_{\varepsilon\in[0,\delta]} n\left[D(\theta||\theta+\varepsilon) - \frac{\varepsilon}{\delta}D(\theta||\theta+\delta) + \frac{\varepsilon q}{n\delta}\right] = -\frac{1}{2}\left(\frac{\delta\sqrt{n}}{2} - \frac{q}{\delta\sqrt{n}}\right)^2 \leq -\frac{1}{8}\left(\sqrt{n}\delta - \sqrt{2q}\right)^2,$$

if $\sqrt{n}\delta \geq \sqrt{2q}$. □

7.2. **Proof of Theorem 2.1.** Throughout this section we will assume that $Y = (Y_1,\ldots,Y_n)$ are independent and identically distributed random variables with $Y_1 \sim F_\theta$ and $\theta \in \Theta$. Without loss of generality we will assume that $m(\theta) = \dot{\psi}(\theta) = 0$ and $v(\theta) = \ddot{\psi}(\theta) = 1$. Moreover, assume that $(c_n)_{n\in\mathbb{N}}$ satisfies (13) and introduce $\mathcal{I}(c_n) = \{(i,j) : j-i+1 \geq c_n n\}$.

We start with some approximation results for the extreme value statistic of the partial sums $\overline{Y}_i^j$.

Lemma 7.4. There exist i.i.d standard normally distributed r.v. $Z_1, \ldots, Z_n$ on the same probability space as $Y_1, \ldots, Y_n$ such that

$$\lim_{n\to\infty} \sqrt{\log n} \max_{(i,j)\in\mathcal{I}(c_n)} \left(\sqrt{j-i+1} \left| \left|\overline{Y}_i^j\right| - \left|\overline{Z}_i^j\right| \right| \right) = 0 \quad \text{a.s.}$$

Proof. We define the partial sums $S_0^Y = 0$ and $S_l^Y = Y_1 + \ldots + Y_l$ and observe that $(j-i+1)\left|\overline{Y}_i^j\right| = \left|S_j^Y - S_{i-1}^Y\right|$. Analogously we define S_l^Z. Now let (i,j) such that $j-i+1 \geq nc_n$ and observe that

$$\left| \frac{\left|S_j^Y - S_{i-1}^Y\right|}{\sqrt{j-i+1}} - \frac{\left|S_j^Z - S_{i-1}^Z\right|}{\sqrt{j-i+1}} \right| \leq \frac{\left|S_j^Y - S_j^Z\right|}{\sqrt{nc_n}} + \frac{\left|S_i^Y - S_i^Z\right|}{\sqrt{nc_n}} \leq 2 \max_{0\leq l\leq n} \frac{\left|S_l^Y - S_l^Z\right|}{\sqrt{nc_n}}.$$

It follows from the KMT inequality (**?**, Thm. 1) and (13) that

$$\sqrt{\log n} \max_{0\leq l\leq n} \frac{\left|S_l^Y - S_l^Z\right|}{\sqrt{nc_n}} = \mathrm{o}(1) \quad \text{a.s.}$$

□

Lemma 7.5.

$$\max_{(i,j)\in\mathcal{I}(c_n)} \left| \sqrt{2T_i^j(Y,\theta)} - \sqrt{j-i+1}\left|\overline{Y}_i^j\right| \right| = \mathrm{o}_{\mathbf{P}}(1)$$

Proof. Set $\xi = m^{-1}$ and note that ξ is strictly increasing. Since Θ is open, there exists for each given $\delta' > 0$ a $\delta > 0$ such that $\xi(B_\delta(0)) \subset B_{\delta'}(\theta) \subset \Theta$. Next define the random variable

$$L_n = \max_{1\leq i<j\leq n} \left|\overline{Y}_i^j\right| \sqrt{j-i+1}.$$

Then it follows from Shao's Theorem (**?**) that $L_n/\sqrt{\log n}$ converges a.s. to some finite constant and we hence find that

$$\max_{(i,j)\in\mathcal{I}(c_n)} \left|\overline{Y}_i^j\right| \leq \sqrt{\frac{\log n}{nc_n}} \frac{L_n}{\sqrt{\log n}} \to 0 \qquad \text{a.s.}$$

Thus, for each $\varepsilon > 0$ there exists an index $n_0 = n_0(\varepsilon) \in \mathbb{N}$ such that for all $n \geq n_0$

$$\mathbf{P}\left(\max_{(i,j)\in\mathcal{I}(c_n)} \left|\overline{Y}_i^j\right| \geq \delta \right) \leq \varepsilon.$$

In other words, $\xi(\overline{Y}_i^j) \in B_\delta(\theta)$ uniformly over $\mathcal{I}(c_n)$ with probability not less than $1-\varepsilon$. Consequently, $\phi(\overline{Y}_i^j) = \max_{\theta\in\Theta} \theta\overline{Y}_i^j - \psi(\theta) = \xi(\overline{Y}_i^j)\overline{Y}_i^j - \psi(\xi(\overline{Y}_i^j))$ which in turn implies that

$$J(\overline{Y}_i^j, \theta) = \phi(\overline{Y}_i^j) - \theta\overline{Y}_i^j + \psi(\theta) = (\xi(\overline{Y}_i^j) - \theta)\overline{Y}_i^j - (\psi(\xi(\overline{Y}_i^j)) - \psi(\theta)).$$

Taylor expansion of ψ around θ gives (recall that $\dot\psi(\theta) = 0$ and $\ddot\psi(\theta) = 1$)

$$\psi(\xi(\overline{Y}_i^j)) - \psi(\theta) = \frac{1}{2}(\xi(\overline{Y}_i^j) - \theta)^2 + \frac{1}{6}\dddot{\psi}(\tilde\theta)(\xi(\overline{Y}_i^j) - \theta)^3$$

for some $\tilde\theta \in B_\varepsilon(\theta)$. This implies

$$J(\overline{Y}_i^j, \theta) = (\xi(\overline{Y}_i^j) - \theta)(\overline{Y}_i^j) - \frac{1}{2}(\xi(\overline{Y}_i^j) - \theta)^2 - \frac{1}{6}\dddot{\psi}(\tilde\theta)(\xi(\overline{Y}_i^j) - \theta)^3.$$

Again, Taylor expansion of $\xi = m^{-1}$ around 0 shows

$$\xi(\overline{Y}_i^j) - \theta = \overline{Y}_i^j - \frac{\dddot{\psi}(\tilde\theta)}{2(v(\tilde\theta))^2}(\overline{Y}_i^j)^2$$

for some $\tilde\theta \in B_{\delta'}(\theta)$. This finally proves that

$$2T_i^j(Y,\theta) = (j-i+1)J(\overline{Y}_i^j, \theta) = (j-i+1)(\overline{Y}_i^j)^2 + (j-i+1)f_n(\overline{Y}_i^j)$$

where f_n is such that $\left|f_n(\overline{Y}_i^j)\right| \leq C^2 \cdot (\overline{Y}_i^j)^3$ for a constant $C = C(\delta') > 0$ (independent of ε, i and j) and for all $n \geq n_0$. It thus holds with probability not less than $1 - \varepsilon$ that

$$\begin{aligned}
\max_{(i,j)\in\mathcal{I}(c_n)} \left|\sqrt{2T_i^j(Y,\theta^*)} - \sqrt{j-i+1}\left|\overline{Y}_i^j\right|\right| \leq & C \max_{(i,j)\in\mathcal{I}(c_n)} \left|(j-i+1)\left(\overline{Y}_i^j\right)^3\right|^{1/2} \\
= & C \max_{(i,j)\in\mathcal{I}(c_n)} \left|\frac{\sum_{l=i}^j Y_l}{\sqrt{j-i+1}}(j-i+1)^{-1/6}\right|^{3/2} \\
\leq & C\left(\frac{L_n}{\sqrt{\log n}}\right)^{3/2} \sqrt[4]{\frac{\log^3 n}{nc_n}}.
\end{aligned}$$

From Shao's Theorem it follows that the last term vanishes almost surely as $n \to \infty$. □

Combination of Lemma 7.4 and 7.5 yields

Proposition 7.6. *There exist i.i.d standard normally distributed r.v. $Z_1, \ldots, Z_n$ on the same probability space as $Y_1, \ldots, Y_n$ such that*

$$\max_{(i,j)\in\mathcal{I}(c_n)} \left|\sqrt{2T_i^j(Y,\theta)} - \sqrt{j-i+1}\left|\bar{Z}_i^j\right|\right| = o_{\mathbf{P}}(1).$$

Lemma 7.7. For $n \in \mathbb{N}$, define the continuous functionals $h, h_n : \mathcal{C}([0,1]) \to \mathbb{R}$ by

$$\begin{aligned}
h(x,c) &= \sup_{\substack{0\leq s<t\leq 1 \\ t-s\geq c}} \left(\frac{|x(t)-x(s)|}{\sqrt{t-s}} - \sqrt{2\log\frac{e}{t-s}}\right) \quad \text{and} \\
h_n(x,c) &= \max_{\substack{1\leq i<j\leq n \\ (j-i+1)/n\geq c}} \left(\frac{|x(j/n)-x(i/n)|}{\sqrt{(j-i+1)/n}} - \sqrt{2\log\frac{en}{j-i+1}}\right),
\end{aligned}$$

respectively. Moreover assume that $\{x_n\}_{n\in\mathbb{N}} \subset \mathcal{C}([0,1])$ is such that $x_n \to x$ for some $x \in \mathcal{C}([0,1])$. Then $h_n(x_n, c) \to h(x, c)$.

Proof. Let $\delta > 0$. Then there exists an index $n_0 \in \mathbb{N}$ such that $|x_n(t) - x(t)| \leq \delta$ for all $n \geq n_0$ and $t \in [0,1]$. Thus, it follows directly from the definition that $h_n(x) = h_n(x_n) + \mathcal{O}(\delta)$ for $n \geq n_0$. Since $u \mapsto \sqrt{2\log e/u}$ is uniformly continuous on $[c,1]$ we consequently have that $h_n(x) \to h(x)$ as $n \to \infty$ and the assertion follows. □

Before we proceed, recall the definition of M in (15). Moreover, we introduce for $0 < c \leq 1$ the statistic

$$M(c) := \sup_{\substack{0\leq s<t\leq 1\\ t-s>c}} \left(\frac{|B(t)-B(s)|}{\sqrt{t-s}} - \sqrt{2\log\frac{e}{t-s}} \right). \tag{48}$$

From (**?**, Thm. 6.1) (and the subsequent Remark 1) it can be seen that $M(c)$ converges weakly to M as $c \to 0^+$.

Proposition 7.8. *Let $c > 0$ and define*

$$T_n^c(Y,\theta) = \max_{(i,j)\in\mathcal{I}(c)} \left(\sqrt{2T_i^j(Y,\theta)} - \sqrt{2\log\frac{en}{j-i+1}} \right).$$

Then $\lim_{c\to 0^+}\lim_{n\to\infty} T_n^c(Y,\theta) = M$, weakly.

Proof. Set $S_0 = 0$ and $S_n = Y_1 + \ldots + Y_n$ and let $\{X_n(t)\}_{t\geq 0}$ be the process that is linear on the intervals $[i/n, (i+1)/n]$ with values $X_n(i/n) = S_i/\sqrt{n}$. We obtain from Donsker's Theorem that $X_n \xrightarrow{\mathcal{D}} B$. Now, recall the definition of h and h_n in Lemma 7.7 and observe that

$$h_n(X_n, c) = \max_{(i,j)\in\mathcal{I}(c)} \left(\sqrt{j-i+1}\,|\overline{Y}_i^j| - \sqrt{2\log\frac{en}{j-i+1}} \right).$$

It hence follows from Lemma 7.5 that

$$|T_n^c(Y,\theta) - h_n(X_n,c)| \leq \max_{(i,j)\in\mathcal{I}(c)} \left| \sqrt{2T_i^j(Y,\theta)} - \sqrt{j-i+1}\,|\overline{Y}_i^j| \right| = o_{\mathbf{P}}(1). \tag{49}$$

Since $X_n \xrightarrow{\mathcal{D}} B$, Lemma 7.7 and (Billingsley, 1968, Thm. 5.5) imply that $h_n(X_n, c) \xrightarrow{\mathcal{D};} h(B, c)$. Theorem 4.1 in (Billingsley, 1968) and (49) thus imply that $T_n^c(Y,\theta) \xrightarrow{\mathcal{D}} h(B,c) = M(c)$ as $n \to \infty$ for all $c > 0$. Thus, the assertion finally follows, since $M(c) \to M$ weakly as $c \to 0^+$ □

Theorem 7.9. *Let $Y = (Y_1, \ldots, Y_n)$ be independent and identically distributed random variables with distribution F_θ, $\theta \in \Theta$. Moreover, assume that $\{c_n\}_{n\in\mathbb{N}}$ is a sequence of positive numbers such that $n^{-1}\log^3 n/c_n \to 0$ and set*

$$T_n(Y,\theta,c_n) = \max_{(i,j)\in\mathcal{I}(c_n)} \left(\sqrt{2T_i^j(Y,\theta)} - \sqrt{2\log\frac{en}{j-i+1}} \right).$$

Then, $T_n(Y,\theta,c_n) \to M$ weakly as $n\to\infty$.

Proof. First observe that according to Proposition 7.6 we have for all $t>0$ that

$$\begin{aligned}\mathbf{P}\left(T_n(Y,\theta;c_n)\le t\right) &= \mathbf{P}\left(\max_{(i,j)\in\mathcal{I}(c_n)}\left(\sqrt{j-i+1}\left|\overline{Z}_i^j\right| - \sqrt{2\log\frac{en}{j-i+1}}\right)\le t\right) + \mathrm{o}(1)\\ &\ge \mathbf{P}\left(\sup_{0\le s<t\le 1}\left(\frac{|B(t)-B(s)|}{\sqrt{t-s}} - \sqrt{2\log\frac{e}{t-s}}\right)\le t\right)+\mathrm{o}(1)\end{aligned}$$

This shows that for all $t>0$

$$\liminf_{n\to\infty}\mathbf{P}(T_n(Y,\theta,c_n)\le t)\ge \mathbf{P}(M\le t)$$

Now let $c>0$ be fixed and assume w.l.o.g. $c_n<c$ for all $n\in\mathbb{N}$. With T_n^c as defined in Proposition 7.8 we conversely find

$$\limsup_{n\to\infty}\mathbf{P}(T_n(Y,\theta,c_n)\le t)\le \limsup_{n\to\infty}\mathbf{P}(T_n^c(Y,\theta,c_n)\le t) = \mathbf{P}(M(c)\le t).$$

Hence the assertion follows from Proposition 7.8 after letting $c\to 0^+$ and the fact that $M>0$ a.s. $\square$

Proof of Theorem 2.1. Let $T_n(Y,\vartheta;c_n)$ be defined as in (14). From Theorem 7.9 it then follows that

$$T_n(Y,\vartheta;c_n)\xrightarrow{\mathcal{D}}\max_{0\le k\le K}\sup_{\tau_k\le s<t\le\tau_{k+1}}\left(\frac{|B(t)-B(s)|}{\sqrt{t-s}} - \sqrt{2\log\frac{e}{t-s}}\right).$$

Clearly the limiting statistic on the right hand side is stochastically bounded from above by M. Conversely, we observe by the scaling property of the Brownian motion that

$$\begin{aligned}&\sup_{\tau_k\le s<t\le\tau_{k+1}}\left(\frac{|B(t)-B(s)|}{\sqrt{t-s}} - \sqrt{2\log\frac{e}{t-s}}\right)\\ &\overset{\mathcal{D}}{=}\sup_{0\le s<t\le 1}\left(\frac{|B(t)-B(s)|}{\sqrt{t-s}} - \sqrt{2\log\frac{e}{t-s}+2\log\frac{1}{\tau_{k+1}-\tau_k}}\right)\overset{\mathcal{D}}{\ge} M-\sqrt{2\log\frac{1}{\tau_{k+1}-\tau_k}}.\end{aligned}$$

$\square$

7.3. **A general exponential inequality.** In this section we give a general exponential inequality for the probability that SMUCE underestimates the number of change-points. To this end, we will make use of the functions

$$\kappa_1^{\pm}(v,w,x,y) = \inf_{\substack{v\le\theta\le w\\ \theta\pm x\in[v,w]}}\sup_{\varepsilon\in[0,x]}\left[\frac{\varepsilon}{x}\left(D(\theta||\theta\pm x)-y\right)-D(\theta||\theta\pm\varepsilon)\right], \tag{50}$$

$$\kappa_2^{\pm}(v,w,x) = \inf_{\substack{v\le\theta\le w\\ \theta\pm x\in[v,w]}} D(\theta\pm x||\theta). \tag{51}$$

Theorem 7.10. *Let $q \in \mathbb{R}$ and $\hat{K}(q)$ be defined as in* (18)*. Moreover, assume that $\kappa_1^{\pm}$ and $\kappa_2^{\pm}$ are defined as in* (50) *and* (51)*, respectively and set*

$$\kappa_1 = \min\left\{\kappa_1^+\left(\underline{\theta}, \overline{\theta}, \frac{\Delta}{2}, \frac{\left(q+\sqrt{2\log\frac{2e}{\lambda}}\right)^2}{n\lambda}\right), \kappa_1^-\left(\underline{\theta}, \overline{\theta}, \frac{\Delta}{2}, \frac{\left(q+\sqrt{2\log\frac{2e}{\lambda}}\right)^2}{n\lambda}\right)\right\} \quad \textit{and}$$

$$\kappa_2 = \min\left\{\kappa_2^+\left(\underline{\theta}, \overline{\theta}, \frac{\Delta}{2}\right), \kappa_2^-\left(\underline{\theta}, \overline{\theta}, \frac{\Delta}{2}\right)\right\}.$$

If $\lambda \geq 2c_n$, then

$$\mathbf{P}\left(\hat{K}(q) < K\right) \leq 2K\left[e^{-\frac{n\lambda\kappa_1}{2}} + e^{-\frac{n\lambda\kappa_2}{2}}\right]. \tag{52}$$

Proof. Let Δ and λ be the smallest jump size and the smallest interval length of the true regression function ϑ, i.e.

$$\Delta = \inf_{1\leq k\leq K}|\theta_k - \theta_{k-1}| \quad \text{and} \quad \lambda = \inf_{0\leq k\leq K}\tau_{k+1} - \tau_k.$$

Now define K disjoint intervals $I_i = (\tau_i - \lambda/2, \tau_i + \lambda/2) \subset [0,1]$. Let $\theta_i^+ = \max\{\theta_{i-1}, \theta_i\}$, $\theta_i^- = \min\{\theta_{i-1}, \theta_i\}$ and split each interval I_i accordingly, i.e. $I_i^+ = \{t \in I_i : \vartheta(t) = \theta_i^+\}$ and $I_i^- = \{t \in I_i : \vartheta(t) = \theta_i^-\}$. Clearly $I_i = I_i^- \cup I_i^+$.

From the definition of the estimator $\hat{K}(q)$ it is clear that

$$\hat{K}(q) < K \quad \Leftrightarrow \quad \exists\hat{\vartheta} \in \mathcal{S}_n[K-1] \text{ such that } T_n(Y, \hat{\vartheta}) \leq q.$$

If $\hat{\vartheta} \in \mathcal{S}_n[K-1]$, then there exists an index $k \in \{1, \ldots, K\}$ such that $\hat{\vartheta}$ is constant on I_k. Let $\Omega_k = \left\{\exists\hat{\theta} \in \Theta : \sqrt{T_{I_k^+}(Y, \hat{\theta})} - \sqrt{\log\frac{en}{\#I_k^+}} \leq \frac{q}{\sqrt{2}} \text{ and } \sqrt{T_{I_k^-}(Y, \hat{\theta})} - \sqrt{\log\frac{en}{\#I_k^-}} \leq \frac{q}{\sqrt{2}}\right\}$ Since the K intervals I_i are disjoint we find

$$\mathbf{P}(\hat{K}(q) < K) \leq \sum_{k=1}^{K}\mathbf{P}\left(\Omega_k\right).$$

If $\hat{\vartheta} \in \mathcal{S}_n[K-1]$ is constant on some I_k with value $\hat{\theta}$, then either $\hat{\theta} \leq \theta_k^+ - \Delta/2$ or $\hat{\theta} \geq \theta_k^- + \Delta/2$, by construction. Set

$$\Omega_k^+ = \left\{\exists\hat{\theta} \leq \theta_k^+ - \Delta/2 : \sqrt{T_{I_k^+}(Y, \hat{\theta})} - \sqrt{\log\frac{en}{\#I_k^+}} \leq \frac{q}{\sqrt{2}}\right\}$$
$$\Omega_k^- = \left\{\exists\hat{\theta} \geq \theta_k^- + \Delta/2 : \sqrt{T_{I_k^-}(Y, \hat{\theta})} - \sqrt{\log\frac{en}{\#I_k^-}} \leq \frac{q}{\sqrt{2}}\right\}$$

and observe that $\mathbf{P}(\Omega_k) \leq \mathbf{P}(\Omega_k^+) + \mathbf{P}(\Omega_k^-)$. We proof an upper bound for $\mathbf{P}(\Omega_k^-)$, the same bound can be obtained for $\mathbf{P}(\Omega_k^+)$ analogously. Recall that $\theta \mapsto T_{I_k^-}(Y, \cdot)$ is convex and has its minimum at $m^{-1}(\overline{Y}_{I_k^-})$. Thus, $T_{I_k^-}(Y, \hat{\theta}) \geq T_{I_k^-}(Y, \theta_k^- + \Delta/2)$ whenever $m^{-1}(\overline{Y}_{I_k^-}) \leq$

$\theta_k^- + \Delta/2$. This yields

$$\begin{aligned}
\mathbf{P}\left(\Omega_k^-\right) &\leq \mathbf{P}\left(\Omega_k^- \cap \left\{m^{-1}(\overline{Y}_{I_k^-}) \leq \theta_k^- + \frac{\Delta}{2}\right\}\right) + \mathbf{P}\left(m^{-1}(\overline{Y}_{I_k^-}) > \theta_k^- + \frac{\Delta}{2}\right) \\
&\leq 1 - \mathbf{P}\left(T_{I_k^-}\left(Y, \theta_k^- + \frac{\Delta}{2}\right) \geq 1/2\left(q + \sqrt{2\log(2e/\lambda)}\right)^2\right) + \mathbf{P}\left(m^{-1}(\overline{Y}_{I_k^-}) > \theta_k^- + \frac{\Delta}{2}\right) \\
&\leq \exp\left(\frac{\lambda n}{2}\inf_{\varepsilon\in[0,\Delta/2]}\left(D(\theta_k^-||\theta_k^- + \varepsilon) - \frac{\varepsilon}{\Delta/2}D(\theta_k^-||\theta_k^- + \Delta/2) + \frac{2\varepsilon\left(q + \sqrt{2\log(2e/\lambda)}\right)^2}{\Delta\lambda n}\right)\right) \\
&\qquad + \exp\left(-\frac{\lambda n}{2}D(\theta_k^- + \Delta/2||\theta_k^-)\right) \\
&\leq \exp\left(-\frac{n\lambda}{2}\kappa_1^+\left(\underline{\theta}, \overline{\theta}, \frac{\Delta}{2}, \frac{\left(q + \sqrt{2\log(2e/\lambda)}\right)^2}{\lambda n}\right)\right) + \exp\left(-\frac{n\lambda}{2}\kappa_2^+\left(\underline{\theta}, \overline{\theta}, \frac{\Delta}{2}\right)\right)
\end{aligned}$$

by Lemma 7.1 and Lemma 7.2. With the definition of the constants κ_j as in the Theorem ($j = 1, 2$) we eventually obtain

$$\mathbf{P}(\hat{K}(q) < K) \leq 2K\left[\exp\left(-\frac{n\lambda\kappa_1}{2}\right) + \exp\left(-\frac{n\lambda\kappa_2}{2}\right)\right].$$

□

The constants $\kappa_i^\pm$ ($i = 1, 2$) basically depend on the exponential family $\mathcal{F}$. Their explicit computation can be rather tedious and has to be done for each exponential family separately (for the Gaussian case see below). Therefore, it is useful to have a lower bound for these constants.

Lemma 7.11. Let v be as in (11) and $\kappa_1^\pm$ and $\kappa_2^\pm$ be defined as in (50) and (51), respectively. Then,

$$\kappa_1^\pm(v, w, x, y) \geq \frac{x^2}{8}\frac{\inf_{v\leq t\leq w} v(t)^2}{\sup_{v\leq t\leq w} v(t)} - y \quad \text{and} \quad \kappa_2^\pm(v, w, x) \geq \frac{x^2}{2}\inf_{v\leq t\leq w} v(t).$$

Proof. First observe from (43), that for any $\theta \in \Theta$ and $\varepsilon > 0$ such that $\theta + \varepsilon \in \Theta$ one has $D(\theta||\theta + \varepsilon) = \int_\theta^{\theta+\varepsilon}(\theta + \varepsilon - t)v(t)\,\mathrm{d}t$. Thus if follows that for all $0 \leq \varepsilon \leq x$

$$\begin{aligned}
\frac{\varepsilon}{x}D(\theta||\theta + x) - D(\theta||\theta + \varepsilon) &= \frac{\varepsilon}{x}\int_\theta^{\theta+x}(\theta + x - t)v(t)\,\mathrm{d}t - \int_\theta^{\theta+\varepsilon}(\theta + \varepsilon - t)v(t)\,\mathrm{d}t \\
&\geq \frac{\varepsilon x}{2}\inf_{t\in[\theta,\theta+x]} v(t) - \frac{\varepsilon^2}{2}\sup_{t\in[\theta,\theta+x]} v(t).
\end{aligned}$$

Maximizing over $0 \leq \varepsilon \leq x$ then yields

$$\sup_{\varepsilon\in[0,x]} \frac{\varepsilon}{x} D(\theta||\theta+x) - D(\theta||\theta+\varepsilon) \geq \frac{x^2}{8} \frac{\inf_{t\in[\theta,\theta+x]} v(t)^2}{\sup_{t\in[\theta,\theta+x]} v(t)}.$$

This proves that

$$\kappa_1^+(v,w,x,y) \geq \frac{x^2}{8} \frac{\inf_{v\leq t\leq w} v(t)^2}{\sup_{v\leq t\leq w} v(t)} - y.$$

Likewise, one finds

$$\kappa_2^+(v,w,x) \geq \frac{x^2}{2} \inf_{v\leq t\leq w} v(t).$$

The estimates for κ_1^- and κ_2^- are derived analogously. □

The combination of Theorem 7.10 and the estimates in Lemma 7.11 yield the handy result in Theorem 2.2. For the case of Gaussian observations, the constants $\kappa_i^\pm$ $(i = 1,2)$ can be computed explicitly and in particular κ_1 is strictly larger than the approximations obtained from Lemma 7.11 by setting $v(t) \equiv 1$.

Theorem 7.12. *Let $q \in \mathbb{R}$ and $\hat{K}(q)$ be defined as in* (18) *and assume that $\mathcal{F}$ is the family of Gaussian distributions with fixed variance* 1*. Then,*

$$\mathbf{P}\left(\hat{K}(q) < K\right) \leq 2K \left[\exp\left(-\frac{1}{8}\left(\frac{\Delta\sqrt{\lambda n}}{2\sqrt{2}} - q - \sqrt{2\log\frac{2e}{\lambda}}\right)_+^2\right) + \exp\left(-\frac{\lambda n \Delta^2}{16}\right)\right]$$

Proof. The proof is similar to the proof of Lemma 7.3. From Lemma 7.11 it follows that $\kappa_2^\pm(v,w,x) = \frac{x^2}{2}$ and one computes explicitly that $\kappa_1^\pm(v,w,x,y) = \frac{1}{2}(\frac{x}{2} - \frac{y}{x})^2 \geq \frac{1}{8}(x - \sqrt{2y})^2$ if $x^2 \geq 2y$. The assertion now follows from Theorem 7.10. □

We close this section with the proof of Theorem 2.8 which is very much in the same spirit than the proof of Theorem 7.10 above.

Proof of Theorem 2.8. Let again Δ be the smallest jump of the true signal ϑ and recall that $\vartheta(t) \in [\underline{\theta}, \overline{\theta}]$ for all $t \in [0,1]$. Moreover, define the K disjoint intervals $I_i = (\tau_i - c_n, \tau_i + c_n) \subset [0,1]$ and accordingly I_i^-, I_i^+, θ_i^-, θ_i^+ and $\hat{\vartheta}_i$ as in the proof of Theorem 7.10.
Now assume that $\hat{K} \in \mathbb{N}$ and that $\hat{\vartheta} \in \mathcal{S}_n[\hat{K}]$ is an estimator of ϑ such that $T_n(Y, \hat{\vartheta}) \leq q$ and

$$\max_{0\leq k\leq K} \min_{0\leq l\leq \hat{K}} |\hat{\tau}_l - \tau_k| > c_n.$$

Put differently, there exists an index $i \in \{1,\ldots,K\}$ such that $|\hat{\tau}_l - \tau_i| > c_n$ for all $0 \leq l \leq \hat{K}$ or, in other words, $\hat{\vartheta}$ contains no change-point in the interval I_i. With the very same reasoning as in the proof of Theorem 7.10 we find that

$$\mathbf{P}\left(\exists \hat{K} \in \mathbb{N}, \hat{\vartheta} \in \mathcal{S}_n[\hat{K}] : T_n(Y, \hat{\vartheta}) \leq q \text{ and } \max_{0 \leq k \leq K} \min_{0 \leq l \leq \hat{K}} |\hat{\tau}_l - \tau_k| > c_n\right)$$

$$\leq \sum_{k=1}^{K} \mathbf{P}\left(\exists \hat{\theta} \in \Theta : T_{I_k^+}(Y, \hat{\theta}) \leq \frac{1}{2}\left(q + \sqrt{\log \frac{e}{c_n}}\right)^2 \text{ and } T_{I_k^-}(Y, \hat{\theta}) \leq \frac{1}{2}\left(q + \sqrt{\log \frac{e}{c_n}}\right)^2\right).$$

By replacing $\lambda/2$ in the proof of Theorem 7.10 by c_n and applying Lemma 7.11 the assertion follows. □

7.4. **Proof of Theorems 2.6 and 2.7.**

Proof of Theorem 2.6. W.l.o.g. we shall assume that $\delta_n \geq 0$. The main idea of the proof is as follows: Let $J_n = \operatorname{argmax}\{|J| : J \subset [0,1], J \cap I_n = \emptyset\}$. In order to show that (24) holds, we construct a sequence $\theta_n^* \in \Theta$ such that

$$\sup\nolimits_{\theta \geq \theta_n^*} \mathbf{P}\left(T_{J_n}(Y, \theta) \leq 1/2\left(q_n + \sqrt{2 \log\left(e/|J_n|\right)}\right)^2\right) \to 0 \text{ and} \tag{53}$$

$$\sup\nolimits_{\theta \leq \theta_n^*} \mathbf{P}\left(T_{I_n}(Y, \theta) \leq 1/2\left(q_n + \sqrt{2 \log(e/|I_n)|}\right)^2\right) \to 0. \tag{54}$$

Note that the true signal ϑ_n takes the value $\theta_0 + \delta_n$ on I_n and θ_0 on J_n and it is not restrictive to assume that $\inf_{n \in \mathbb{N}} |J_n| > 0$. We construct $\theta_n^* = \theta_0 + \sqrt{\beta_n/n}$ for a sequence $(\beta_n)_{n \in \mathbb{N}}$ that satisfies $\sqrt{\beta_n}/q_n \to \infty$.
We first consider (53). To this end observe that for all $t \in J_n$ we have $|\theta_n^* - \vartheta_n(t)| \sqrt{|J_n|\, n} = \sqrt{\beta_n |J_n|}$. We further find that

$$\Gamma_{J_n} := \sqrt{\beta_n |J_n|} - q_n - \sqrt{2 \log(e/|J_n|)} = q_n \left(\frac{\sqrt{\beta_n}}{q_n} - 1 - \frac{\sqrt{2 \log(e/|J_n|)}}{q_n}\right) \to \infty.$$

Thus, we can apply (47) and find for all $\theta \geq \theta_n^*$

$$\mathbf{P}\left(T_{J_n}(Y, \theta) \leq 1/2\left(q_n + \sqrt{2 \log(e/|J_n|)}\right)^2\right) \leq \exp\left(-\frac{\Gamma_{J_n}^2}{8}\right) \to 0.$$

Now observe that for $t \in I_n$ we have $|\theta_n^* - \vartheta_n(t)| \sqrt{|I_n|\, n} = \delta_n \sqrt{|I_n|\, n} - \sqrt{\beta_n |I_n|}$. Thus (54) follows from (47) given

$$\Gamma_{I_n} := \delta_n \sqrt{|I_n|\, n} - \sqrt{\beta_n |I_n|} - q_n - \sqrt{2 \log(e/|I_n|)} \to \infty.$$

It hence remains to construct sequences (β_n) for each case (1) and (2) such that the previous condition holds while $\sqrt{\beta_n}/q_n \to \infty$.

We assume first that $\liminf_{n\to\infty} |I_n| > 0$ and define β_n through the equation $\sqrt{\beta_n |I_n|} = c\left(\delta_n\sqrt{|I_n|\, n} - q_n - \sqrt{2\log(e/|I_n|)}\right)$ for some $0 < c < 1$. Then,

$$\frac{\sqrt{\beta_n |I_n|}}{q_n} = c\left(\frac{\delta_n\sqrt{|I_n|\, n}}{q_n} - 1 - \frac{\sqrt{2\log(e/|I_n|)}}{q_n}\right)$$

From the condition in case (1) of the theorem and the fact that $|I_n|$ is bounded away from zero for large n, we find that $\sqrt{\beta_n}/q_n \to \infty$. Further we find $\Gamma_{I_n} = (1-c)\sqrt{\beta_n |I_n|} \to \infty$. Finally we consider the case when $|I_n| \to 0$ and define β_n through the equation $\sqrt{\beta_n |I_n|} = c\varepsilon_n\sqrt{-\log|I_n|}$ for some $0 < c < 1$. From the conditions in case (2) of the theorem and the inequality $\sqrt{x+1} - \sqrt{x} \leq 1/(2\sqrt{x})$, which holds for any $x > 0$, one obtains

$$\begin{aligned}
\Gamma_{I_n} &\geq (\sqrt{2} + \varepsilon_n)\sqrt{-\log|I_n|} - \sqrt{\beta_n |I_n|} - q_n - \sqrt{2\log(e/|I_n|)} \\
&= (\sqrt{2} + (1-c)\varepsilon_n)\sqrt{-\log|I_n|} - q_n - \sqrt{2}\sqrt{1 + \log(1/|I_n|)} \\
&\geq ((1-c)\varepsilon_n)\sqrt{-\log|I_n|} - \frac{1}{\sqrt{-2\log|I_n|}} - q_n.
\end{aligned}$$

This shows that $\Gamma_{I_n} \to \infty$ for a suitable small c, such that $\sup_{n\in\mathbb{N}} q_n/(\epsilon_n\sqrt{\log(1/|I_n|)}) \leq 1 - 2c$. Again from the assumptions in the theorem it follows that $\sqrt{\beta_n}/q_n \to \infty$. $\square$

Proof of Theorem 2.7. Theorem 7.12 implies $\mathbf{P}(\hat{K}(q_n) < K_n) \leq e^{-\Gamma_{1,n}} + e^{-\Gamma_{2,n}}$ with

$$\Gamma_{1,n} = \frac{1}{8}\left(\frac{\sqrt{n\lambda_n}\Delta_n}{2\sqrt{2}} - q_n - \sqrt{2\log(2e/\lambda_n)}\right)_+^2 - \log K_n \quad \text{and} \quad \Gamma_{2,n} = \frac{n\lambda_n\Delta_n^2}{16} - \log K_n.$$

It is easy to see, that any condition (1) - (3) implies $\Gamma_{2,n} \to \infty$. It remains to check that $\Gamma_{1,n} \to \infty$. Under condition (1) we observe that

$$\frac{\Gamma_{1,n}}{q_n^2} = \frac{1}{8}\left(\frac{\sqrt{n\lambda_n}\Delta_n}{2\sqrt{2}q_n} - \frac{q_n + \sqrt{2\log(2e/\lambda_n)}}{q_n}\right)_+^2 - \frac{\log K_n}{q_n^2} \to \infty.$$

Since q_n is bounded away from zero, the assertion follows. Next, we consider conditions (2) and (3). To this end, assume that $\sqrt{n\lambda_n}\Delta_n \geq (C + \varepsilon_n)\sqrt{\log(1/\lambda_n)}$ for some constant $C > 0$ and a sequence ε_n such that $\varepsilon_n\sqrt{\log(1/\lambda_n)} \to \infty$. We find that

$$\begin{aligned}
\Gamma_{1,n} &\geq \frac{1}{8}\left(\frac{(C+\varepsilon_n)\sqrt{\log\frac{1}{\lambda_n}}}{2\sqrt{2}} - q_n - \sqrt{2\log(2e/\lambda_n)}\right)_+^2 - \log K_n \\
&= \frac{1}{8}\left(\frac{\varepsilon_n\sqrt{\log\frac{1}{\lambda_n}}}{2\sqrt{2}} + \left(\frac{C-4}{2\sqrt{2}}\right)\sqrt{\log\frac{1}{\lambda_n}} - q_n - \frac{1+\log 2}{\sqrt{2\log(1/\lambda_n)}}\right)_+^2 - \log K_n,
\end{aligned}$$

where we have used the inequality $\sqrt{x+1}-\sqrt{x}\leq 1/(2\sqrt{x})$. If $\sup_{n\in\mathbb{N}} K_n<\infty$, then the choice $C=4$ implies $\Gamma_{1,n}\to\infty$. Otherwise, we use the estimate $K_n\leq 1/\lambda_n$ which results in $C=12$ as a sufficient condition for $\Gamma_{1,n}\to\infty$. $\square$

7.5. **Proof of Lemma 3.1.**

Proof. First observe that the definition of $\hat{\vartheta}(q)$ in (6) implies that $q\geq T_n(Y,\hat{\vartheta}(q))$ and hence, by identifying $\hat{\vartheta}(q)$ with the pair $(\hat{\mathcal{P}}(q),\hat{\theta}(q))$, we find

$$
\begin{aligned}
(\hat{K}(q)+1)q \geq (\hat{K}(q)+1)T_n(Y,\hat{\vartheta}(q)) &\geq \sum_{I\in\hat{\mathcal{P}}(q)}\left(\sqrt{2T_I(Y,\hat{\vartheta}(q))}-\sqrt{2\log(e/|I|)}\right)\\
&\geq \sqrt{2}\sqrt{\sum_{I\in\hat{\mathcal{P}}(q)}\big(|I|\,\phi(\bar{Y}_I)\big)-l(Y,\hat{\vartheta}(q)}-n\sqrt{2\log(en)}\\
&\geq \sqrt{2}\sqrt{l(\bar{Y},\hat{\vartheta}(q))-l(Y,m^{-1}(\bar{Y}))}-n\sqrt{2\log(en)}
\end{aligned}
$$

The last inequality follows from the fact that $\phi(\overline{Y}_I)\geq \overline{Y}_I\theta-\psi(\theta)$ for all $\theta\in\Theta$ and all $I\in\hat{\mathcal{P}}(q)$ for the choice $\theta=m^{-1}(\overline{Y})$. Summarizing, we find

$$
\gamma\geq\left((\hat{K}(q)+1)q+n\sqrt{2\log(en)}\right)^2/2+l(Y,m^{-1}(\bar{Y}))\geq l(Y,\hat{\vartheta}(q)).
$$

Now, let $\hat{\vartheta}=(\hat{\mathcal{P}},\hat{\theta})$ be a minimizer of (31). The definition of $\hat{K}(q)$ in (18) implies that $D(\mathcal{P},\theta)=\infty$ if $\#\mathcal{P}<\hat{K}(q)$. Thus we have that $\left|\hat{\mathcal{P}}\right|\geq\hat{K}(q)$. Assume that there exists $k\geq 1$ such that $\#\mathcal{P}=\hat{K}(q)+k$ (for $k=0$ nothing is to show). Since $(\hat{\mathcal{P}},\hat{\theta})$ is a minimizer of (31) and since $D\geq 0$ we find

$$
\begin{aligned}
\gamma(\left|\hat{\mathcal{P}}\right|-1) &\leq D(\hat{\mathcal{P}}(q),\hat{\theta}(q))-D(\hat{\mathcal{P}},\hat{\theta})+\gamma\big(\left|\hat{\mathcal{P}}(q)\right|-1\big)\\
&\leq D(\hat{\mathcal{P}}(q),\hat{\theta}(q))-k\gamma+\gamma\big(\left|\hat{\mathcal{P}}\right|-1\big)\\
&< (1-k)l(Y,\hat{\vartheta}(q))+\gamma\big(\left|\hat{\mathcal{P}}\right|-1\big).
\end{aligned}
$$

This is a contradiction for $l(\hat{\vartheta})$ being non-negative and hence we conclude that $\left|\hat{\mathcal{P}}\right|=\hat{K}(q)$ and that $\hat{\vartheta}=(\hat{\mathcal{P}},\hat{\theta})$ solves (6). $\square$