Biometrika (2019), **xx**, x, *pp*. 1–21
Printed in Great Britain

Supplementary Material for The Essential Histogram

BY HOUSEN LI, AXEL MUNK, HANNES SIELING
Institute for Mathematical Stochastics, Georg-August-Universität Göttingen, Goldschmidstr. 7 37077 Göttingen, Germany
hli1@uni-goettingen.de munk@math.uni-goettingen.de hsielin@uni-goettingen.de

AND GUENTHER WALTHER
Department of Statistics, Stanford University, Sequoia Hall, 390 Serra Mall 94305 Stanford, California, U.S.A.
walther@stat.stanford.edu

SUMMARY

This supplementary material complements the main text. §S1 provides further optimality results on the multiscale confidence set. All the technical proofs are collected in §S2. §S3 gives the computation details, and some additional simulation results are provided in §S4.

S1. OPTIMALITY OF THE CONFIDENCE REGION $C_n(\alpha)$

The first part of Theorem 2 and Theorem S1 show that $C_n(\alpha)$ in (1) is an optimal confidence region for continuous F with respect to the distance d_p in (7) for arbitrary p: The first part of Theorem 2 shows that with probability converging to one, $C_n(\alpha)$ will exclude H with $d_p(F, H) \geq (1 + \epsilon_n)\big\{2\log(e/p)/n\big\}^{1/2}$, where $\epsilon_n \downarrow 0$ sufficiently slowly. In the case of small p, Theorem S1 shows that if $1 + \epsilon_n$ is replaced by $1 - \epsilon_n$, then no test can distinguish F and H with nontrivial power. In the case of larger p, i.e. when p stays bounded away from zero, the condition of the first part of Theorem 2 becomes $d_p(F, H) \geq n^{-1/2}B_n$ with $B_n \to \infty$. On the other hand, a contiguity argument as in the proof of Theorem 4.1(c) in Dümbgen & Walther (2008) shows that for any test to have asymptotic power 1 against a sequence H^n requires $d_p(F, H^n) = n^{-1/2}B_n$ with $B_n \to \infty$.

THEOREM S1. *Let $\phi_n(\boldsymbol{X})$, with $\boldsymbol{X} = (X_1, \ldots, X_n)$, be any test with level $\alpha \in (0, 1)$ under $H_0:\ X_1, \ldots, X_n$ are independent samples from a continuous distribution function F. If $(\log n)^2/n \leq p_n \to 0$ and $\epsilon_n \in (0, 1)$ such that $\epsilon_n(\log e/p_n)^{1/2} \to \infty$, then*

$$\inf_{H:\ d_{p_n}(F,H) \geq (1-\epsilon_n)\left(\frac{2}{n}\log\frac{e}{p_n}\right)^{1/2}} \mathrm{E}_H \phi_n(\boldsymbol{X}) \;=\; \alpha + o(1)\,.$$

*Remark S*1. The price for simultaneously considering all $H \in C_n(\alpha)$ in the second part of Theorem 2, as opposed to a fixed sequence $H = H_n$ in the first part, is a doubling of the distance $d_{p_n}(F, H)$: For a fixed sequence of intervals $I = I_n$, the standardized distance between $F(I)$ and $F_n(I)$ becomes negligible compared to the radius $\big\{2\log e/F_n(I)\big\}^{1/2}$ of the confidence ball around $F_n(I)$. But if one needs to consider all intervals simultaneously, then for the worst-case interval I the standardized distance between $F_n(I)$ and $F(I)$ is also about $\big\{2\log e/F_n(I)\big\}^{1/2}$.

© 2018 Biometrika Trust

The proof of Theorem 2 shows that the first part holds even for smaller intervals, namely for $p_n \in [2\log(n)/n, 1/2)$, provided that also $F(I) > H(I)$. If $F(I) < H(I)$, then (S4) requires a different bound. For example, if $p_n = k\log(n)/n$, $k \geq 2$, then (S4) requires

$$d_{p_n}(F, H) \ > \ (1+\epsilon_n)(2^{1/2} + k^{-1/2}) \left(\frac{1}{n}\log\frac{e}{p_n}\right)^{1/2}, \tag{S1}$$

and it is not clear whether this result can be improved. Note that Theorem S1 does not provide a lower bound for scales of order $\log(n)/n$.

The construction of $C_n(\alpha)$ via the log likelihood ratio statistic $\mathrm{logLR}_n\{H(I), F_n(I)\}$ rather than, say, the standardized binomial statistic $n^{1/2}|H(I) - F_n(I)|\big[H(I)\{1-H(I)\}\big]^{-1/2}$ is crucial for these optimality results: While the tail of $n^{1/2}|F(I) - F_n(I)|\big[F(I)\{1-F(I)\}\big]^{-1/2}$ is close to sub-Gaussian, it does vary with $F(I)$ and becomes increasingly heavy as $F(I)$ decreases to 0, see Shorack & Wellner (1986, Chapter 11.1). It is thus not clear how to construct a penalty that is effective in combining the evidence on the various scales p_n. For example, if $F(I) = k\log(n)/n$ for some fixed $k > 0$, then the penalty $\big(2\log\cdots\big)^{1/2}$ in the definition of $C_n(\alpha)$ would not be sufficiently large for the standardized binomial statistic and therefore the optimality result (S5) would not hold, at least in the case $F(I) > H(I)$.

S2. PROOFS

To ease understanding, we recommend interested readers to read also Rivera & Walther (2013). In all the proofs, we will make it explicit where the continuity assumption on F is used. Whenever the assertions, more precisely, the first part in Theorem 1, and Theorems 2, 3 and 5, also hold for general, possibly discontinuous F, we will detail how such an extension is possible. Recall that $\ell\{F_n(I)\}$ is defined in (2). We will make use of the following

LEMMA S1. *$H \in C_n(\alpha)$ implies for all $I \in \mathcal{J}$:*

$$n^{1/2}\frac{|H(I) - F_n(I)|}{\max\{H(I), F_n(I)\}} \leq c_n(I)\,, \tag{S2a}$$

$$n^{1/2}\frac{|H(I) - F_n(I)|}{M^{1/2}} \leq c_n(I) \qquad \textit{if } \max\{F_n(I), H(I)\} \leq \frac{1}{2}\,, \tag{S2b}$$

$$n^{1/2}\frac{|H(I) - F_n(I)|}{m^{1/2}} \leq c_n(I) + \frac{c_n^2(I)}{2(mn)^{1/2}}\,, \tag{S2c}$$

where $c_n(I) = \ell\{F_n(I)\} + \kappa_n(\alpha)$, $M = \max\big[F_n(I)\{1-F_n(I)\}, H(I)\{1-H(I)\}\big]$, *and* $m = \min\big[F_n(I)\{1-F_n(I)\}, H(I)\{1-H(I)\}\big]$.

Likewise, Lemma S1 hold for all $H \in \tilde{C}_n(\alpha)$ and $I \in \mathcal{J}$ where H' is constant.

Proof. We note that Taylor's theorem implies for $H \in C_n(\alpha)$

$$\begin{aligned}
\frac{c_n^2(I)}{2n} &\geq \frac{\mathrm{logLR}_n\{H(I), F_n(I)\}}{n} \\
&= \frac{\{H(I) - F_n(I)\}^2}{2\xi(1-\xi)} \quad \text{for some } \xi \text{ between } H(I) \text{ and } F_n(I) \\
&\geq \begin{cases} \{H(I) - F_n(I)\}^2/\big[2\max\{H(I), F_n(I)\}\big] & \\ \{H(I) - F_n(I)\}^2/(2M) & \text{if } 2\max\{H(I), F_n(I)\} \leq 1 \end{cases}
\end{aligned} \tag{S3}$$

since $\xi(1-\xi)$ is increasing for $\xi \in (0, 1/2]$, proving (S2a) and (S2b). The assertion (S2c) follows by applying (K.5) in Dümbgen & Wellner (2014) to (S3). $\square$

Proof of Theorem 1. Note that the law of $d_p(F, F_n)$ for an arbitrary F is bounded above by the law of $d_p(F, F_n)$ for a continuous F, and that the latter does not depend on F. Thus, we may assume $X_1, \ldots, X_n$ are independent samples from $F = U[0, 1]$. The statement of the theorem is closely related to the modulus of continuity of the uniform empirical process, see Shorack & Wellner (1986, Chapter 14.2). Unfortunately, the available results appear not strong enough to cover the case where $B_n \to \infty$ slowly. Therefore we employ the Hungarian construction together with elementary calculations and recent results about Brownian motion.

By Shorack & Wellner (1986, Chapter 12.3), there exists a sequence W_n of Brownian motions on the same probability space such that

$$\limsup_n \frac{n^{1/2}}{\log n} \sup_{I \subset [0,1]} \left| n^{1/2} \Big\{ F_n(I) - |I| \Big\} - B_n(I) \right| \le M < \infty \; a.s.,$$

where $B_n(t) = W_n(t) - tW_n(1)$, a standard Brownian bridge. Writing

$$\begin{aligned} R_n = & \sup_{I \subset [0,1]:\, |I| \in [\frac{\log^2 n}{n}, \frac{1}{2})} \left| \frac{n^{1/2}\{F_n(I) - |I|\} - W_n(I)}{\big[|I|\{1 - |I|\}\big]^{1/2}} \right| \\ \le & \sup_{I \subset [0,1]} \frac{(2n)^{1/2}}{\log n} \left| n^{1/2}\{F_n(I) - |I|\} - B_n(I) \right| + |W_n(1)| , \end{aligned}$$

we obtain $R_n = O_p(1)$ and

$$\left| n^{1/2}\, d_p(F, F_n) - \sup_{I : |I| = p} \frac{|W_n(I)|}{\big\{p(1-p)\big\}^{1/2}} \right| \le R_n \;\text{ for all } p \in \Big[\frac{\log^2 n}{n}, \frac{1}{2}\Big) .$$

The first statement of the theorem now follows from Dümbgen & Spokoiny (2001, Theorem 2.1, also Section 6.1) together with $(1-p)^{-1/2} \le 1 + 2p$ and the fact that $p(2 \log e/p)^{1/2}$ stays bounded in p.

For the second claim we note that

$$\sup_{I : |I| = p_n} \frac{|W_n(I)|}{\big\{p_n(1-p_n)\big\}^{1/2}} \;\ge\; \max_{i = 1, \ldots, m_n} N_i ,$$

where the $N_i = \big[W_n(ip_n) - W_n\{(i-1)p_n\}\big] p_n^{-1/2}$ are independent, and identically distributed from $\mathcal{N}(0, 1)$ and $m_n = \lfloor 1/p_n \rfloor$. Without loss of generality, we may assume that B_n is such that

$\lambda_n = (2\log e/p_n)^{1/2} - B_n/2 \to \infty$. Mill's ratio gives

$$\begin{aligned}\mathrm{pr}\Big(\max_{i=1,\dots,m_n} N_i \le \lambda_n\Big) & \\ &\le \Big\{1 - \frac{1}{4\lambda_n}\exp(-\lambda_n^2/2)\Big\}^{m_n} \\ &\le \exp\Big(\frac{-m_n \frac{p_n}{e}\exp\Big[B_n\Big\{\big(2\log e/p_n\big)^{1/2} - B_n/4\Big\}/2\Big]}{4\lambda_n}\Big) \\ &\le \exp\Big[-\frac{\exp\Big\{B_n(2\log e/p_n)^{1/2}/4\Big\}}{5e(2\log e/p_n)^{1/2}}\Big] = o(1)\,.\end{aligned}$$

Theorem 7.1(b) in Dümbgen (2003) suggests that the second statement of the theorem holds for any collection of estimators $\{H_n(I)\}_I$, so it is not possible to improve on the performance of F_n. We will not engage in the lengthy technical work required to establish that claim. □

Proof of Theorem 2. To avoid lengthy technical work we will prove the theorem using $\mathcal{J} = \{$ all intervals in $\boldsymbol{R}\}$ in the definition of $C_n(\alpha)$. The technical work in Rivera & Walther (2013) shows that the approximating set of intervals used in §2 is fine enough so that the optimality results continue to hold with that approximating set.

To prove the first part we will show that for arbitrary $p_n \in [2\log(n)/n, 1/2)$, $H = H_n$ and $I = I_n$ with $F(I) = p_n$ and

$$n^{1/2}\frac{|H(I) - p_n|}{\big\{p_n(1-p_n)\big\}^{1/2}} > \\ (1+\epsilon_n)\Big[2^{1/2} + \Big\{\frac{\log(e/p_n)}{np_n(1-p_n)}\Big\}^{1/2}\mathbb{1}\{F(I) < H(I)\}\Big]\Big(\log\frac{e}{p_n}\Big)^{1/2}, \tag{S4}$$

we have

$$\mathrm{pr}_F\{H \in C_n(\alpha)\} \;\to\; 0 \quad \text{uniformly in } F\,. \tag{S5}$$

The first claim follows since for $p_n \in (\log^2(n)/n, 1/2)$ we have

$$(1+\epsilon_n)\Big\{\frac{\log e/p_n}{np_n(1-p_n)}\Big\}^{1/2} \;\le\; (1+\epsilon_n)(\log n)^{-1/2} \;=\; o(\epsilon_n)\,,$$

since $\epsilon_n \gg (\log e/p_n)^{-1/2} \ge (\log n)^{-1/2}$.

On the other hand, if $p_n = k\log(n)/n$ for some $k \ge 2$, then

$$\Big\{\frac{\log e/p_n}{np_n(1-p_n)}\Big\}^{1/2} \;=\; k^{-1/2} + o\Big\{(\log n)^{-1/2}\Big\}\,,$$

yielding (S1) in the case where $F(I) < H(I)$.

To prove (S5) set $c_n = (1+\epsilon_n)(2\log e/p_n)^{1/2}$. Then the inequality (S4) reads

$$n^{1/2}\frac{|H(I) - p_n|}{\big\{p_n(1-p_n)\big\}^{1/2}} > c_n + \frac{c_n^2}{2(1+\epsilon_n)\big\{np_n(1-p_n)\big\}^{1/2}}\,\mathbb{1}\{F(I) < H(I)\}\,. \tag{S6}$$

We have $b_n = \min\{B_n, (\log n)^{1/2}\} \to \infty$ by the assumption of the theorem, and we define the event $\mathcal{A}_n = \{n^{1/2}|F_n(I) - F(I)| \le \big\{b_n F(I)\{1 - F(I)\}\big\}^{1/2}\}$. We will show that on the event

$\mathcal{A}_n$, (S6) implies for sufficiently large n

$$n^{1/2} \frac{|H(I) - F_n(I)|}{\big[F_n(I)\{1 - F_n(I)\}\big]^{1/2}} \geq \tilde{c}_n + \frac{\tilde{c}_n^2}{2\big[nF_n(I)\{1 - F_n(I)\}\big]^{1/2}} \mathbb{1}\{F_n(I) < H(I)\}\,, \quad \text{(S7)}$$

uniformly in H and I, where $\tilde{c}_n = \ell\{F_n(I)\} + b_n^{1/2}/4$. Hence Lemma S1 (S2b) and (S2c) give

$$\begin{aligned}
&\mathrm{pr}_F\{H \in C_n(\alpha)\} \\
&\leq \mathrm{pr}_F\left(\frac{|H(I) - F_n(I)|}{\big[F_n(I)\{1 - F_n(I)\}\big]^{1/2}} \leq \tilde{c}_n + \frac{\tilde{c}_n^2}{2\big[nF_n(I)\{1 - F_n(I)\}\big]^{1/2}} \mathbb{1}\{F_n(I) < H(I)\} \right) \\
&\qquad \text{once } {b_n}^{1/2}/4 > \kappa_n(\alpha) \\
&\leq \mathrm{pr}_F(\mathcal{A}_n^c) \quad \text{eventually} \\
&\leq \frac{1}{b_n}\,,
\end{aligned}$$

by Chebychev's inequality, and the above conclusions are uniform in F, H and I. Thus (S5) follows since $\kappa_n(\alpha) = O(1)$ by Rivera & Walther (2013, Proposition 1) if F is continuous. The above argument holds for general, possibly discontinuous, F as well; In that case, we will use $\kappa_n^*(\alpha)$ instead of $\kappa_n(\alpha)$, which is also $O(1)$ by Lemma S2.

It remains to prove (S7). On $\mathcal{A}_n$ we have

$$\left| \frac{F_n(I)}{F(I)} - 1 \right| \;\leq\; \left[\frac{b_n\{1 - F(I)\}}{nF(I)} \right]^{1/2} \;\leq\; \left(\frac{b_n}{\log n} \right)^{1/2}, \quad \text{(S8)}$$

since $F(I) \geq 2\log(n)/n$, and the same bound applies to $\{1 - F_n(I)\}/\{1 - F(I)\}$ since $F(I) \leq 1/2$. Hence

$$\left| \left[\frac{F_n(I)\{1 - F_n(I)\}}{p_n(1 - p_n)} \right]^{1/2} - 1 \right| \;\leq\; \left(\frac{b_n}{\log n} \right)^{1/2}.$$

So on $\mathcal{A}_n$:

$$\begin{aligned}
&n^{1/2} \frac{|H(I) - F_n(I)|}{\big[F_n(I)\{1 - F_n(I)\}\big]^{1/2}} \\
&\geq n^{1/2} \frac{|H(I) - F(I)| - [F(I)\{1 - F(I)\}b_n/n]^{1/2}}{\big\{p_n(1 - p_n)\big\}^{1/2}\{1 + \big(b_n/\log n\big)^{1/2}\}} \\
&\geq n^{1/2} \frac{|H(I) - F(I)|}{\{p_n(1 - p_n)\}^{1/2}} \Big\{1 - \big(\frac{b_n}{\log n}\big)^{1/2}\Big\} - {b_n}^{1/2} \\
&\geq \left[c_n + \frac{c_n^2 \,\mathbb{1}\{F(I) < H(I)\}}{2(1 + \epsilon_n)\{np_n(1 - p_n)\}^{1/2}} \right] \Big\{1 - \big(\frac{b_n}{\log n}\big)^{1/2}\Big\} - {b_n}^{1/2}\,,
\end{aligned} \quad \text{(S9)}$$

by (S6). Next for sufficiently large n

$$\begin{aligned}
2\log\frac{e}{p_n} \;&\geq\; 2\log\frac{e}{F_n(I)\{1 - F_n(I)\}} + 2\log\frac{F_n(I)}{2F(I)} \\
&\geq\; 2\log\frac{e}{F_n(I)\{1 - F_n(I)\}} - 4\log 2 \\
&=\; \ell\{F_n(I)\}^2 - 4\log 2\,,
\end{aligned}$$

since $F_n(I) \in (p_n/2,\ 1/2)$ eventually. Hence

$$\begin{aligned} c_n &= \bigl(2\log \frac{e}{p_n}\bigr)^{1/2} + 2^{1/2}B_n \\ &\geq \tilde{c}_n + b_n/4 \quad \text{eventually,} \end{aligned}$$

and

$$\begin{aligned} \frac{c_n^2}{1+\epsilon_n} &= 2\log\frac{e}{p_n} + 2B_n\bigl(\log\frac{e}{p_n}\bigr)^{1/2} \\ &\geq\ 2\log\frac{e}{p_n} + \frac{B_n}{2}\bigl(\log\frac{e}{p_n}\bigr)^{1/2} + \frac{3}{2}B_n \ \text{ eventually} \\ &\geq\ \tilde{c}_n^2 \ \text{ eventually.} \end{aligned}$$

Finally, (S8) yields $\{np_n(1-p_n)\}^{-1/2} \geq [nF_n(I)\{1-F_n(I)\}]^{-1/2}\{1-\bigl(b_n/\log n\bigr)^{1/2}\}$ eventually. Thus (S9) gives

$$\begin{aligned} & n^{1/2}\frac{|H(I)-F_n(I)|}{[F_n(I)\{1-F_n(I)\}]^{1/2}} \\ &\geq\ \left(\tilde{c}_n + \frac{b_n}{4} + \frac{\tilde{c}_n^2\, \mathbb{1}\{F(I)<H(I)\}}{2[nF_n(I)\{1-F_n(I)\}]^{1/2}}\right)\Bigl\{1-\bigl(\frac{b_n}{\log n}\bigr)^{1/2}\Bigr\} - b_n^{1/2} \quad \text{eventually} \\ &\geq\ \tilde{c}_n + \frac{\tilde{c}_n^2\, \mathbb{1}\{F_n(I)<H(I)\}}{2[nF_n(I)\{1-F_n(I)\}]^{1/2}} \quad \text{eventually,} \end{aligned}$$

since

$$\begin{aligned} \left(\tilde{c}_n + \frac{\tilde{c}_n^2}{[nF_n(I)\{1-F_n(I)\}]^{1/2}}\right)\bigl(\frac{b_n}{\log n}\bigr)^{1/2} &\leq\ \Bigl\{\bigl(2\log n\bigr)^{1/2} + b_n^{1/2} + \frac{6\log n + b_n}{\bigl(\log n\bigr)^{1/2}}\Bigr\}\bigl(\frac{b_n}{\log n}\bigr)^{1/2} \\ &=\ \{2^{1/2}+6+o(1)\}b_n^{1/2}, \end{aligned}$$

and because $\mathcal{A}_n$ and (S6) imply that $F_n(I) < H(I)$ if and only if $F(I) < H(I)$. (S7) is proved.

To prove the second part we will consider the event

$$\mathcal{B}_n = \left\{\frac{n^{1/2}|F_n(I)-F(I)|}{[nF(I)\{1-F(I)\}]^{1/2}} \leq \Bigl\{2\log\frac{e}{F(I)}\Bigr\}^{1/2} + b_n^{1/2} \text{ for all } I \text{ with } \frac{\log^2 n}{n} \leq F(I) \leq \frac{1}{2}\right\}$$

in lieu of $\mathcal{A}_n$. Then $\mathrm{pr}_F(\mathcal{B}_n) \to 1$ uniformly in F by Theorem 1. Now we proceed analogously as in the proof of the first part: Suppose that there exist $H, p_n \in (\log^2(n)/n, 1/2)$ and I with $F(I) = p_n$ satisfying

$$n^{1/2}\frac{|H(I)-p_n|}{\{p_n(1-p_n)\}^{1/2}}\ >\ 2c_n \tag{S10}$$

where $c_n = (1+\epsilon_n)\bigl\{2\log(e/p_n)\bigr\}^{1/2}$ as in the first part. We will show that on the event $\mathcal{B}_n$, (S10) implies, for n large enough,

$$n^{1/2}\frac{|H(I)-F_n(I)|}{[nF_n(I)\{1-F_n(I)\}]^{1/2}} > \tilde{c}_n + \frac{\tilde{c}_n^2}{2[nF_n(I)\{1-F_n(I)\}]^{1/2}}\mathbb{1}\{F_n(I)<H(I)\}\,, \tag{S11}$$

uniformly in H and I, where $\tilde{c}_n = \ell\{F_n(I)\} + b_n^{1/2}/4$ as in the first part. Hence

$$
\begin{aligned}
&\mathrm{pr}_F\Big\{(S10) \text{ holds for some } H \in C_n(\alpha), p_n \in \Big(\frac{\log^2 n}{n}, \frac{1}{2}\Big) \text{ and } I \text{ with } F(I) = p_n\Big\} \\
&\quad \le \mathrm{pr}_F\Bigg(n^{1/2}\frac{|H(I) - F_n(I)|}{[F_n(I)\{1 - F_n(I)\}]^{1/2}} > \tilde{c}_n + \frac{\tilde{c}_n^2}{2[nF_n(I)\{1 - F_n(I)\}]^{1/2}}\mathbb{1}\{F_n(I) < H(I)\} \\
&\qquad\qquad \text{for some } H \in C_n(\alpha), p_n \in \Big(\frac{\log^2 n}{n}, \frac{1}{2}\Big)\Bigg) + \mathrm{pr}_F(\mathcal{B}_n^c) \ \text{ eventually} \\
&\quad = \mathrm{pr}_F(\mathcal{B}_n^c) \quad \text{once } b_n^{1/2}/4 > \kappa_n(\alpha) \quad \text{by Lemma S1 (S2b) and (S2c)} \\
&\quad = o(1) \quad \text{uniformly in } F,
\end{aligned}
$$

proving the second part. As in the first part, we use $\kappa_n^*(\alpha)$ instead of $\kappa_n(\alpha)$ for general F.

It remains to prove that on $\mathcal{B}_n$, (S10) implies (S11): On $\mathcal{B}_n$ we have

$$
\Big|\frac{F_n(I)}{F(I)} - 1\Big| \ \le\ \frac{[2\log\{e/F(I)\}]^{1/2} + b_n^{1/2}}{\log n} \ \le\ \frac{2}{(\log n)^{1/2}} \quad \text{eventually,}
$$

since $F(I) \ge \log^2(n)/n$, and the same bound holds for $\{1 - F_n(I)\}/\{1 - F(I)\}$, hence

$$
\left|\left[\frac{F_n(I)\{1 - F_n(I)\}}{p_n(1 - p_n)}\right]^{1/2} - 1\right| \ \le\ \frac{2}{(\log n)^{1/2}}.
$$

Thus on $\mathcal{B}_n$

$$
\begin{aligned}
& n^{1/2}\frac{|H(I) - F_n(I)|}{[F_n(I)\{1 - F_n(I)\}]^{1/2}} \\
&\ge n^{1/2}\frac{|H(I) - F(I)| - \Big[\{2\log(e/p_n)\}^{1/2} + b_n^{1/2}\Big]\{p_n(1 - p_n)/n\}^{1/2}}{\{p_n(1 - p_n)\}^{1/2}\{1 + 2(\log n)^{-1/2}\}} \\
&\ge n^{1/2}\frac{|H(I) - p_n|}{\{p_n(1 - p_n)\}^{1/2}}\Big\{1 - \frac{2}{(\log n)^{1/2}}\Big\} - \Big[\{2\log(e/p_n)\}^{1/2} + b_n^{1/2}\Big] \\
&\ge c_n + \frac{1}{2}B_n \quad \text{by (S10),}
\end{aligned}
\tag{S12}
$$

since $\big(\log e/p_n\big)^{1/2} \le \big(\log n\big)^{1/2}$. As in the proof of the first part, one finds $c_n \ge \tilde{c}_n + b_n/4$ eventually. Moreover, on $\mathcal{B}_n$ we have $F_n(I) \ge F(I)\{1 - 2(\log n)^{-1/2}\} \ge \log^2(n)/(2n)$, hence

$$
\frac{\tilde{c}_n^2}{[nF_n(I)\{1 - F_n(I)\}]^{1/2}} \ \le\ \frac{6\log n + b_n}{2^{-1}\log n} \ \le\ 13 \quad \text{eventually,}
$$

and so (S12) is no smaller than

$$
\tilde{c}_n + \frac{\tilde{c}_n^2}{2[nF_n(I)\{1 - F_n(I)\}]^{1/2}}\,\mathbb{1}\{F_n(I) < H(I)\} \quad \text{eventually,}
$$

completing the proof. □

Proof of Proposition 1. Let F have density

$$
f(x) = \frac{3}{2}\mathbb{1}\big(x \in [0, \tfrac{1}{2})\big) + \frac{1}{2}\mathbb{1}\big(x \in [\tfrac{1}{2}, 1]\big).
$$

If k_n is odd, then the $(k_n+1)/2$th bin is $\big[1/2-1/(2k_n), 1/2+1/(2k_n)\big)$. Denote the height of the histogram, i.e. the slope of H_n, on that bin by h. Set

$$I = \begin{cases} \big[1/2-1/(6k_n), 1/2\big) & \text{if } h \le 3/4, \\ \big[1/2, 1/2+1/(2k_n)\big) & \text{if } h > 3/4. \end{cases}$$

Then $p_n = F(I) = 1/(4k_n)$ and $|F(I)-H_n(I)| \ge 1/(8k_n) = p_n/2$, hence

$$n^{1/2}\, d_{p_n}(F, H_n) \;\ge\; n^{1/2}\frac{|F(I)-H_n(I)|}{[F(I)\{1-F(I)\}]^{1/2}} \;\ge\; \frac{1}{2}\big(np_n\big)^{1/2}.$$

Proof of Theorem 3. By construction of $C_n(\alpha)$ we have $\mathrm{pr}\{F\in C_n(\alpha)\} \ge 1-\alpha$, which also holds true for general F, if we set $\kappa_n^*(\alpha)$ as the threshold, see Lemma S2. Hence Lemma S1 (S2c) gives

$$\mathrm{pr}\Big\{\big|\bar f(I) - |I|^{-1}F_n(I)\big| \le \frac{1}{2}r_n(I) \text{ for all } I\in\mathcal{J}\Big\} \ge 1-\alpha.$$

Furthermore, Lemma S1 (S2c) yields $\big|\bar h(I) - |I|^{-1}F_n(I)\big| \le r_n(I)/2$ for all $I\in\mathcal{J}$ and every $H\in\tilde C_n(\alpha)$ whose density h is constant on I. □

Proof of Theorem 4. Let $\epsilon_n\in(0,1)$ and $p_n\in\big(2n^{-1}\log^2 n, 1/2\big)$ be arbitrary sequences with $\epsilon_n\big(\log e/p_n\big)^{1/2}\to\infty$. In particular, $p_n\to 0$. We proceed as in the proof of Theorem S1 and will construct $m_n = \lfloor 1/2p_n\rfloor^2$ densities $f_{njk}\in\mathcal{I}_n(1-\epsilon_n)$ such that

$$\lim_{n\to\infty}\mathrm{E}\Big|\frac{1}{m_n}\sum_{j,k=1}^{m_n^{1/2}} L_{njk} - 1\Big| \;=\; 0, \tag{S13}$$

where $L_{njk} = \prod_{i=1}^n f_{njk}(X_i)$ with X_i independent $U(0,1)$. Unfortunately, the truncation argument used in the proof of Theorem S1 will not go through as the covariances of the L_{njk} are not small enough. Instead, we follow an idea in the proof of Theorem 3.1(a) in Dümbgen & Spokoiny (2001) and write $f_{njk}(x) = g_{nj}(x)h_{nk}(x)$, where for $j,k=1,\ldots,m_n^{1/2}$:

$$\begin{aligned} g_{nj}(x) &= 1_{[0,1]}(x) - c_n 1_{I_{nj}}(x) + a_n 1_{[0,1/2)\setminus I_{nj}}(x), \\ h_{nk}(x) &= 1_{[0,1]}(x) + c_n 1_{J_{nk}}(x) - a_n 1_{[1/2,1)\setminus J_{nk}}(x), \end{aligned}$$

and $c_n = (1-\epsilon_n)\{(1-p_n)/np_n\}^{1/2}\{2\log(2e/p_n)\}^{1/2}$, $a_n = c_np_n/(1/2-p_n)$, $I_{nj} = [(j-1)p_n, jp_n)$, $J_{nk} = [1/2+(k-1)p_n, 1/2+kp_n)$. One readily checks that f_{njk}, g_{nj} and h_{nk} are densities. For each pair (j,k):

$$\begin{aligned} \bar f_{njk}(J_{nk}) - \bar f_{njk}(I_{nj}) \;&=\; 2c_n \;=\; 2(1-\epsilon_n)\left\{\frac{p_n(1-p_n)}{n}\right\}^{1/2}\frac{\{2\log(2e/p_n)\}^{1/2}}{p_n} \\ &\ge (1-\epsilon_n)\bigg(\left[\frac{F_{njk}(J_{nk})\{1-F_{njk}(J_{nk})\}}{n}\right]^{1/2}\frac{\big[2\log\big\{e/F_{njk}(J_{nk})\big\}\big]^{1/2}}{|J_{nk}|} \\ &\quad + \left[\frac{F_{njk}(I_{nj})\{1-F_{njk}(I_{nj})\}}{n}\right]^{1/2}\frac{\big[2\log\big\{e/F_{njk}(I_{nj})\big\}\big]^{1/2}}{|I_{nj}|}\bigg), \end{aligned}$$

since $F_{njk}(J_{nk}) = (1+c_n)p_n$, $F_{njk}(I_{nj}) = (1-c_n)p_n \ge p_n/2$, and the function $g(t) = \{t(1-t)\}^{1/2}$ is concave for $t\in(0,1)$. Thus $f_{njik}\in\mathcal{I}_n(1-\epsilon_n)$. While $F_{njk}(J_{nk})$

is larger than p_n, one can easily bound it by, say, $2p_n$, and changing the definition of $\mathcal{I}_n(c)$ to this end will not affect the conclusions of the theorem. Moreover, $p_n \geq 2\log^2(n)/n$ implies $np_n \geq 2\big\{\log(e/p_n)\big\}^2$, hence $\Delta_\infty = \sup_{x\in[0,1]}|g_{nj}(x)-1| = c_n \leq \{2\log(2e/p_n)/(np_n)\}^{1/2} \leq \big\{\log(e/p_n)\big\}^{-1/2} \leq (\frac{1}{2}\log m_n)^{-1/2}$, while $\Delta_2^2 = \int_0^1 (g_{nj}-1)^2 = p_n c_n^2 + (1/2-p_n)a_n^2 \leq 2(1+2p_n)(1-\epsilon_n)^2\log(2e/p_n)/n \leq 2(1-\epsilon_n/2)\log(2e/p_n)/n$ since $p_n \leq \epsilon_n/4$. Therefore we obtain as in the proof of Theorem S1:

$$\lim_{n\to\infty} \mathrm{E}\Big|\frac{1}{m_n^{1/2}}\sum_{j=1}^{m_n^{1/2}} L_{g_{nj}} - 1\Big| \;=\; 0, \tag{S14}$$

where $L_{g_{nj}} = \prod_{i=1}^n g_{nj}(X_i)$, and the same result holds for $L_{h_{nk}}$. Conditional on $N = \#\{i : X_i < 1/2\}$, $\sum_j L_{g_{nj}}$ and $\sum_k L_{h_{nk}}$ are independent and the $\{X_i : X_i < 1/2\}$ are independent $U(0,1/2)$, hence (S14) gives

$$\mathrm{E}\Big(\Big|\frac{1}{m_n^{1/2}}\sum_{j=1}^{m_n^{1/2}} L_{g_{nj}} - 1\Big| \mid N\Big)\,\mathbb{1}\Big(N \in (\frac{n}{4},\frac{3n}{4})\Big) \;\to\; 0 \quad \text{a.s.},$$

and likewise for the average of the $L_{h_{nk}}$. Thus

$$\begin{aligned}
&\mathrm{E}\Big|\frac{1}{m_n}\sum_{j,k=1}^{m_n^{1/2}}(L_{g_{nj}}-1)(L_{h_{nk}}-1)\Big| \\
&\leq \mathrm{E}\left\{\mathrm{E}\Big(\Big|\frac{1}{m_n^{1/2}}\sum_{j=1}^{m_n^{1/2}} L_{g_{nj}} - 1\Big| \mid N\Big)\mathrm{E}\Big(\Big|\frac{1}{m_n^{1/2}}\sum_{k=1}^{m_n^{1/2}} L_{h_{nk}} - 1\Big| \mid N\Big)\mathbb{1}\Big(N \in (\frac{n}{4},\frac{3n}{4})\Big)\right\} \\
&\qquad + 2^2\mathrm{pr}\Big(N \notin (\frac{n}{4},\frac{3n}{4})\Big) \\
&\to 0\,,
\end{aligned} \tag{S15}$$

by bounded convergence, since $\mathrm{E}\big(\big|L_{g_{nj}}-1\big| \mid N\big) \leq \mathrm{E}\big(L_{g_{nj}} \mid N\big) + 1 = 2$. Hence

$$\mathrm{E}\Big|\frac{1}{m_n}\sum_{j,k=1}^{m_n^{1/2}} L_{njk} - 1\Big| = \mathrm{E}\Big|\frac{1}{m_n}\sum_{j,k=1}^{m_n^{1/2}}\Big\{(L_{g_{nj}}-1)(L_{h_{nk}}-1) + (L_{g_{nj}}-1) + (L_{h_{nk}}-1)\Big\}\Big|$$

converges to zero by (S14) and (S15), proving (S13). □

Proof of Theorem 5. For $k \in \{0,1\}$, let $F \in \mathcal{I}_n(1+2k+\epsilon_n)$, so there exist $I_1 < I_2$ for which (9) holds with $c = 1+2k+\epsilon_n$ and $F(I_i) \geq \log^2(n)/n$, $i = 1,2$. It is readily checked that these lower bounds on $F(I_i)$ imply on the event

$$\mathcal{A}_n = \Big\{n^{1/2}\frac{|F_n(I_i)-F(I_i)|}{[F(I_i)\{1-F(I_i)\}]^{1/2}} \leq \frac{\epsilon_n}{2}\{2\log\frac{e}{F(I_i)}\}^{1/2} \text{ for } i = 1,2\Big\},$$

it holds, for large n enough,

$$(1+\frac{\epsilon_n}{6})\left[\frac{F(I_i)\{1-F(I_i)\}}{n}\right]^{1/2}\frac{[2\log\{e/F(I_i)\}]^{1/2}}{|I_i|} \;\geq\; \frac{1}{2}r_n(I_i) \quad i = 1,2, \tag{S16}$$

where $r_n(I)$ is defined in Theorem 3. For general F, we replace $\kappa_n(\alpha)$ by $\kappa_n^*(\alpha)$.

Let $H \in \tilde{C}_n(\alpha)$ with h being constant on I_1 and I_2, and suppose

$$\bar{h}(I_1) - \bar{h}(I_2) \;\geq\; -k\Big\{r_n(I_1) + r_n(I_2)\Big\} \tag{S17}$$

holds. Then on $\mathcal{A}_n$:

$$\begin{aligned}
&\frac{H(I_1) - F_n(I_1)}{|I_1|} + \frac{F_n(I_2) - H(I_2)}{|I_2|} \\
&\geq \bar{h}(I_1) - \bar{h}(I_2) + \bar{f}(I_2) - \bar{f}(I_1) - \frac{\epsilon_n}{2}\sum_{i=1}^{2}\left[\frac{F(I_i)\{1-F(I_i)\}}{n}\right]^{1/2}\frac{[2\log\{e/F(I_i)\}]^{1/2}}{|I_i|} \\
&> -k\{r_n(I_1) + r_n(I_2)\} + (1 + 2k + \frac{\epsilon_n}{2})\sum_{i=1}^{2}\left[\frac{F(I_i)\{1-F(I_i)\}}{n}\right]^{1/2}\frac{[2\log\{e/F(I_i)\}]^{1/2}}{|I_i|} \\
&\qquad \text{by (S17) and (9)} \\
&\geq \frac{1}{2}\{r_n(I_1) + r_n(I_2)\} \;\; \text{eventually} \;\; \text{by (S16),}
\end{aligned}$$

which gives the contradiction $h \notin \tilde{C}_n(\alpha)$ by Lemma S1 (S2c). Therefore (S17) can only hold on $\mathcal{A}_n^c$. As in the proof of Theorem 2 we assumed $\mathcal{J} = \{$ all real intervals $\}$ and refer to Rivera & Walther (2013) for the technical work that can be used to show that the conclusion also obtains with the approximating set used in §2.

If h is nonincreasing on $\mathrm{conv}(I_1 \cup I_2)$, then (S17) holds with $k = 0$ and so for $F \in \mathcal{I}_n(1 + \epsilon_n)$:

$$\begin{aligned}
&\mathrm{pr}_F\Big\{\text{there exists } H \in \tilde{C}_n(\alpha) \text{ whose density is constant on } I_1 \\
&\qquad\qquad \text{and } I_2 \text{ and nonincreasing on } \mathrm{conv}(I_1 \cup I_2)\Big\} \\
&\leq \mathrm{pr}_F(\mathcal{A}_n^c) \;\leq\; \frac{8}{2\epsilon_n^2\log(e/p_n)} \;\to 0 \;\; \text{uniformly in } F\,,
\end{aligned}$$

by Chebychev's inequality, proving the first part. The second part follows analogously since (8) together with $\bar{f}(I_2) \leq \bar{f}(I_1)$ implies that (S17) holds with $k = 1$. □

Proof of Theorem 6. It follows immediately from Proposition S1 below and the fact that $\kappa_n(\alpha) \leq \{2\log(C/\alpha)\}^{1/2}$ for some constant C, see Rivera & Walther (2013). □

PROPOSITION S1. *Under the same notation as Theorem* 6*, it controls overestimating the number of bins uniformly over all F's*

$$\mathrm{pr}_F\Big\{N_{\mathtt{bin}}(H) > N_{\mathtt{bin}}(F)\Big\} \leq \alpha;$$

and it controls underestimating the number of bins, for $n \geq 16\log n/(\lambda\underline{\theta})$,

$$\mathrm{pr}_F\Big\{N_{\mathtt{bin}}(H) < N_{\mathtt{bin}}(F)\Big\} \leq 4K\left[2\exp\left(-\frac{1}{128}n\lambda^2\underline{\theta}^2\right) + \exp\left\{-\frac{1}{72}n\lambda^2\left(\frac{\Delta}{2} - \frac{12\delta_n}{\lambda n^{1/2}}\right)_+^2\right\}\right],$$

with $\delta_n = \{2\log(256e)/(\lambda^2\underline{\theta}^2)\}^{1/2} + \kappa_n(\alpha)$; Moreover, it controls the number of modes and troughs, for $n \geq 32\log n/(\lambda\underline{\theta})$, with $\tilde{\delta}_n = \{2\log(1024e)(\lambda^2\underline{\theta}^2)\}^{1/2} + \kappa_n(\alpha)$,

$$\begin{aligned}
&\mathrm{pr}_F\Big(h \textit{ and } f \textit{ have the same number of modes and troughs}\Big)\\
&\qquad \geq 1-\alpha-12K\left[2\exp\left(-\frac{1}{512}n\lambda^2\underline{\theta}^2\right)+\exp\left\{-\frac{1}{288}n\lambda^2\left(\frac{\Delta}{2}-\frac{24\tilde{\delta}_n}{\lambda n^{1/2}}\right)_+^2\right\}\right].
\end{aligned}$$

*Remark S*2. Note that the assertions in Proposition S1 also hold for sequences of $f = f_n$ with $\underline{\theta} = \underline{\theta}_n$, $\lambda = \lambda_n$, $\Delta = \Delta_n$, $K = K_n$, and $\alpha = \alpha_n$.

*Proof of Proposition S*1. For the first part, by the definition of the essential histogram in (5), we have

$$\mathrm{pr}_F\Big\{N_{\mathtt{bin}}(H) \leq N_{\mathtt{bin}}(F)\Big\} \geq \mathrm{pr}_F\Big\{F \in \tilde{C}_n(\alpha)\Big\} \geq 1-\alpha.$$

For the second and the third parts, we use arguments similar to Frick et al. (2014, Theorem 7.10), but with notable differences due to the use of the reduced system $\mathcal{J}$. We will frequently use the following inequality, which comes as an application of the Hoeffding's inequality,

$$\mathrm{pr}_F\Big\{|F(I)-F_n(I)| \geq x\Big\} \leq 2\exp(-2nx^2). \tag{S18}$$

The detail is as follows: For the second part, let m_k be the mid-point of I_k, $I_k^- = I_k \cap (-\infty, m_k]$, and $I_k^+ = I_k \cap (m_k, \infty)$. For a fixed k, we have

$$\begin{aligned}
&\mathrm{pr}_F\Big(h \text{ is constant on } (m_{k-1}, m_k]\Big)\\
=&\mathrm{pr}_F\Big(h \equiv c \geq \frac{c_{k-1}+c_k}{2} \text{ on } (m_{k-1}, m_k] \text{ for some constant } c\Big)\\
&\quad+\mathrm{pr}_F\Big(h \equiv c < \frac{c_{k-1}+c_k}{2} \text{ on } (m_{k-1}, m_k] \text{ for some constant } c\Big)\\
\leq&\mathrm{pr}_F\Big\{\big|\bar{h}(I)-\bar{f}(I)\big| \geq \frac{\Delta}{2},\, h \text{ is constant on } I \equiv I_{k-1}^+\Big\}\\
&\quad+\mathrm{pr}_F\Big\{\big|\bar{h}(I)-\bar{f}(I)\big| \geq \frac{\Delta}{2},\, h \text{ is constant on } I \equiv I_k^-\Big\}.
\end{aligned}$$

By symmetry we only need to consider the first term in the r.h.s. of the above equation, where $I = I_{k-1}^+$. By the construction of $\mathcal{J}$ in (4), it holds that for any I with $F_n(I) \geq 6\log(n)/n$ there is an interval $J \subseteq I$ and $J \in \mathcal{J}$ such that $F_n(J) \geq F_n(I)/3$. Conditioned on $|F_n(I) - F(I)| \leq \lambda\underline{\theta}/16$ and $|F_n(J) - F(J)| \leq \lambda\underline{\theta}/16$, we have for $n \geq 16\log n/(\lambda\underline{\theta})$

$$F(J) \;\geq\; F_n(J) - \frac{1}{16}\lambda\underline{\theta} \;\geq\; \frac{1}{3}F_n(I) - \frac{1}{16}\lambda\underline{\theta} \;\geq\; \frac{1}{3}F(I) - \frac{1}{12}\lambda\underline{\theta} \;\geq\; \frac{1}{6}F(I),$$

which implies $|J| \geq |I|/6 \geq \lambda/12$. Thus, for $n \geq 16\log n/(\lambda\underline{\theta})$

$$\begin{aligned}
&\mathrm{pr}_F\Big\{\big|\bar{h}(I)-\bar{f}(I)\big| \geq \frac{\Delta}{2},\, h \text{ is constant on } I \equiv I_{k-1}^+\Big\}\\
\leq&\mathrm{pr}_F\Big\{|F_n(I)-F(I)| \geq \frac{1}{16}\lambda\underline{\theta}\Big\} + \mathrm{pr}_F\Big\{|F_n(J)-F(J)| \geq \frac{1}{16}\lambda\underline{\theta}\Big\}\\
&\quad+\mathrm{pr}_F\Big\{\big|\bar{h}(J)-\bar{f}(J)\big| \geq \frac{\Delta}{2},\, h \text{ is constant on } J,\, |J| \geq \frac{1}{12}\lambda, |F_n(J)-F(J)| \leq \frac{1}{16}\lambda\underline{\theta}\Big\}\\
\leq& 2\exp\left(-\frac{1}{128}n\lambda^2\underline{\theta}^2\right) + 2\exp\left(-\frac{1}{128}n\lambda^2\underline{\theta}^2\right)
\end{aligned}$$

$$+\mathrm{pr}_F\Big\{\big|\bar h(J)-\bar f(J)\big|\ge\frac{\Delta}{2},\ h\text{ is constant on }J,\ |J|\ge\frac{1}{12}\lambda,\ |F_n(J)-F(J)|\le\frac{1}{16}\lambda\underline{\theta}\Big\}$$
by (S18)
$$\le 4\exp\left(-\frac{1}{128}n\lambda^2\underline{\theta}^2\right)+\mathrm{pr}_F\Big\{\big|\bar h(J)-\bar f(J)\big|\ge\frac{\Delta}{2}-\frac{12\delta_n}{\lambda n^{1/2}}\Big\}\qquad\text{by Lemma S1 (S2a)}$$
$$\le 4\exp\left(-\frac{1}{128}n\lambda^2\underline{\theta}^2\right)+2\exp\left\{-\frac{1}{72}n\lambda^2\left(\frac{\Delta}{2}-\frac{12\delta_n}{\lambda n^{1/2}}\right)_+^2\right\}\qquad\text{by (S18).}$$

The same bound holds for $\mathrm{pr}_F\big\{\big|\bar h(I)-\bar f(I)\big|\ge\Delta/2,\ h\text{ is constant on }I\equiv I_k^-\big\}$ due to symmetry. Therefore, we have for $n\ge 16\log n/(\lambda\underline{\theta})$

$$\mathrm{pr}_F\Big\{N_{\mathtt{bin}}(H)<N_{\mathtt{bin}}(F)\Big\}\le\mathrm{pr}_F\Big\{h\text{ is constant on }(m_{k-1},m_k]\text{ for some }k\Big\}$$
$$\le 4K\left[2\exp\left(-\frac{1}{128}n\lambda^2\underline{\theta}^2\right)+\exp\left\{-\frac{1}{72}n\lambda^2\left(\frac{\Delta}{2}-\frac{12\delta_n}{\lambda n^{1/2}}\right)_+^2\right\}\right].$$

For the third part, we further divide I_k^+ into two subintervals $I_{k,1}^+$ and $I_{k,2}^+$ of equal lengths, and I_k^- into $I_{k,1}^-$ and $I_{k,2}^-$ of equal lengths. For any fixed k, it holds that

$$\mathrm{pr}_F\Big(h\text{ has exactly one jump, but has a different trend from }f\text{ on }(m_{k-1},m_k]\Big)$$
$$\le\mathrm{pr}_F\Big\{\big|\bar h(I)-\bar f(I)\big|\ge\frac{\Delta}{2},\text{ and }h\text{ is constant on }I\equiv I_{k,1}^+\Big\}$$
$$+\mathrm{pr}_F\Big\{\big|\bar h(I)-\bar f(I)\big|\ge\frac{\Delta}{2},\text{ and }h\text{ is constant on }I\equiv I_{k,2}^+\Big\}$$
$$+\mathrm{pr}_F\Big\{\big|\bar h(I)-\bar f(I)\big|\ge\frac{\Delta}{2},\text{ and }h\text{ is constant on }I\equiv I_{k,1}^-\Big\}$$
$$+\mathrm{pr}_F\Big\{\big|\bar h(I)-\bar f(I)\big|\ge\frac{\Delta}{2},\text{ and }h\text{ is constant on }I\equiv I_{k,2}^-\Big\}.$$

Each term above can be bounded in a similar way as in the second part, which leads to

$$\mathrm{pr}_F\Big(h\text{ has exactly one jump, but has a different trend from }f\text{ on }(m_{k-1},m_k]\Big)$$
$$\le 16\exp\left(-\frac{1}{512}n\lambda^2\underline{\theta}^2\right)+8\exp\left\{-\frac{1}{288}n\lambda^2\left(\frac{\Delta}{2}-\frac{24\tilde\delta_n}{\lambda n^{1/2}}\right)_+^2\right\}\qquad\text{for }n\ge\frac{32\log n}{\lambda\underline{\theta}}.$$

It follows from (i) and (ii) that for for $n\ge 16\log n/(\lambda\underline{\theta})$

$$\mathrm{pr}_F\Big(h\text{ has exactly one jump in each }(m_{k-1},m_k]\Big)\ge 1-\alpha$$
$$-4K\left[2\exp\left(-\frac{1}{128}n\lambda^2\underline{\theta}^2\right)+\exp\left\{-\frac{1}{72}n\lambda^2\left(\frac{\Delta}{2}-\frac{12\delta_n}{\lambda n^{1/2}}\right)_+^2\right\}\right].$$

Thus, for $n\ge 32\log n/(\lambda\underline{\theta})$

$$\mathrm{pr}_F\Big(h\text{ and }f\text{ have the same number of modes and troughs}\Big)$$
$$\ge\mathrm{pr}_F\Big(h\text{ has exactly one jump, and has the same trend as }f,\text{ on each }(m_{k-1},m_k]\Big)$$

$$\geq 1-\alpha-12K\left[2\exp\left(-\frac{1}{512}n\lambda^2\underline{\theta}^2\right)+\exp\left\{-\frac{1}{288}n\lambda^2\left(\frac{\Delta}{2}-\frac{24\tilde{\delta}_n}{\lambda n^{1/2}}\right)_+^2\right\}\right].$$

□

*Proof of Theorem S*1. Using the probability integral transformation we may assume $F=U[0,1]$. For $j=1,\ldots,m_n=\lfloor 1/p_n\rfloor$ define the densities $f_{nj}(x)=(1+c_n)\mathbb{1}_{I_{nj}}(x)+(1-a_n)\mathbb{1}_{[0,1]\setminus I_{nj}}(x)$, where $I_{nj}=[(j-1)p_n,jp_n)$, $a_n=c_np_n/(1-p_n)$ and $c_n=\{(1-p_n)/np_n\}^{1/2}(1-\epsilon_n)\{2\log(e/p_n)\}^{1/2}$. Then $d_{p_n}(F,F_{nj})=|F_{nj}(I_{nj})-F(I_{nj})|\{p_n(1-p_n)\}^{-1/2}=(1-\epsilon_n)\{2\log(e/p_n)/n\}^{1/2}$. The claim of the theorem will follow as in the proof of Theorem 4.1(b) in Dümbgen & Walther (2008) once we show that

$$\lim_{n\to\infty}\mathrm{E}_F\Big|\frac{1}{m_n}\sum_{j=1}^{m_n}L_{nj}-1\Big| \;=\; 0, \tag{S19}$$

where $L_{nj}=\prod_{i=1}^n f_{nj}(X_i)$. Since the sets $\{f_{nj}\neq 1\}$ are not disjoint, Lemma 7.4 of Dümbgen & Walther (2008) is not applicable to prove (S19) and we have to account for the covariances of the L_{nj}. For $i\neq j$ we obtain $\int_0^1 f_{ni}(x)f_{nj}(x)dx=1-\big\{p_nc_n/(1-p_n)\big\}^2$, hence $\mathrm{E}_F(L_{ni}L_{nj})=\Big[1-\big\{p_nc_n/(1-p_n)\big\}^2\Big]^n<1$. Using $\mathrm{E}_FL_{nj}=1$ and Hölder's inequality gives for any $\epsilon>0$

$$\begin{aligned}
&\mathrm{E}_F\Big|\frac{1}{m_n}\sum_{j=1}^{m_n}L_{nj}-1\Big|\\
&\leq\left[\mathrm{var}_F\Big\{\frac{1}{m_n}\sum_{j=1}^{m_n}L_{nj}\mathbb{1}(L_{nj}\leq\epsilon m_n)\Big\}\right]^{1/2}+\frac{2}{m_n}\sum_{j=1}^{m_n}\mathrm{E}_F\big\{L_{nj}\mathbb{1}(L_{nj}>\epsilon m_n)\big\}\\
&\leq\left[\frac{1}{m_n}\mathrm{E}_F\Big\{L_{n1}^2\mathbb{1}(L_{n1}\leq\epsilon m_n)\Big\}+\frac{1}{m_n^2}\sum_{i<j}\mathrm{cov}\Big\{L_{ni}\mathbb{1}(L_{ni}\leq\epsilon m_n),L_{nj}\mathbb{1}(L_{nj}\leq\epsilon m_n)\Big\}\right]^{1/2}\\
&\qquad+2\mathrm{E}_F\big\{L_{n1}\mathbb{1}(L_{n1}>\epsilon m_n)\big\}\\
&\leq\left[\frac{1}{m_n}\mathrm{E}_F(\epsilon m_nL_{n1})+\mathrm{E}_FL_{n1}L_{n2}-\Big\{\mathrm{E}_FL_{n1}\mathbb{1}(L_{n1}\leq\epsilon m_n)\Big\}^2\right]^{1/2}+2\mathrm{E}_F\big\{L_{n1}\mathbb{1}(L_{n1}>\epsilon m_n)\big\}\\
&\leq\left[\epsilon+1-\Big\{1-\mathrm{E}_FL_{n1}\mathbb{1}(L_{n1}>\epsilon m_n)\Big\}^2\right]^{1/2}+2\mathrm{E}_F\big\{L_{n1}\mathbb{1}(L_{n1}>\epsilon m_n)\big\}.
\end{aligned}$$

Thus (S19) follows by showing

$$\lim_{n\to\infty}\mathrm{E}_F\big\{L_{n1}\mathbb{1}(L_{n1}>\epsilon m_n)\big\} \;=\; 0\,. \tag{S20}$$

To this end, observe that $p_n\geq\log^2 n/n$ implies $np_n\geq(\log e/p_n)^2$, hence

$$\Delta_\infty=\sup_{x\in[0,1]}|f_{nj}(x)-1|=c_n\leq\left\{\frac{2\log(e/p_n)}{np_n}\right\}^{1/2}\leq\left\{\frac{2}{\log(e/p_n)}\right\}^{1/2}\leq\left(\frac{2}{\log m_n}\right)^{1/2}.$$

Further, $\Delta_2^2 = \int_0^1 \{f_{nj}(x) - 1\}^2 dx = p_n c_n^2 + (1-p_n)a_n^2 = 2(1-\epsilon_n)^2 \log(e/p_n)/n$, hence

$$\begin{aligned}
\big(\log m_n\big)^{1/2}\Big(1 - \frac{n\Delta_2^2}{2\log m_n}\Big) &\geq \big(\log m_n\big)^{1/2}\Big\{1 - (1-\epsilon_n)\frac{\log(e/p_n)}{\log\lfloor 1/p_n\rfloor}\Big\} \\
&\geq \big(\log m_n\big)^{1/2}\Big\{1 - (1-\epsilon_n)(1 - \frac{3}{\log p_n})\Big\} \\
&= \big(\log m_n\big)^{1/2}\epsilon_n\{1 + o(1)\} + o(1) \;\to\; \infty\,,
\end{aligned}$$

by the assumption on ϵ_n. It is shown in the proof of Lemma 7.4 in Dümbgen & Walther (2008) that (S20) follows from these properties of Δ_∞ and Δ_2^2. □

S3. COMPUTATIONAL DETAILS

S3.1. *Computation of the threshold $\kappa_n(\alpha)$*

Let F be an arbitrary distribution function, i.e., right continuous with limits from left. We define its standard left continuous inverse as $F^{-1}(t) = \inf\{x : F(x) \geq t\}$. Let $Z_1, \ldots, Z_n$ be independent and identically distributed uniform random variables on $[0,1]$, and denote their common distribution function by U. Note that $X_i = F^{-1}(Z_i)$ are independent, and identically distributed according to F. By U_n and F_n we denote the empirical distribution functions of $Z_1, \ldots, Z_n$ and $X_1, \ldots, X_n$, respectively. For an interval $I = (j,k]$, we define $Z_I = (Z_{(j)}, Z_{(k)}]$ and similarly $X_I = (X_{(j)}, X_{(k)}]$. Then, T_n in (3) can be written as

$$T_n \;=\; \max_{I \in \mathcal{J}_0 \cap (\mathcal{D}\times\mathcal{D})}\Big([2\mathrm{logLR}_n\big\{F(X_I), F_n(X_I)\big\}]^{1/2} - \ell\big\{F_n(X_I)\big\}\Big)$$

where $\mathcal{D} = \{i \,:\, X_{(i)} \neq X_{(i+1)}\}$ as in (4), and $\mathcal{J}_0$ is a collection of intervals

$$\mathcal{J}_0 = \bigcup_{l=2}^{l_{max}} \mathcal{J}_0(l), \quad \text{where } l_{max} = \Big\lfloor \log_2 \tfrac{n}{\log n} \Big\rfloor \quad \text{and}$$

$$\mathcal{J}_0(l) = \Big\{(j,k] : j,k \in \{1 + id_l, i \in \mathbb{N}_0\} \text{ and } m_l < k - j \leq 2m_l\Big\},$$

$$\text{where } m_l = n2^{-l},\; d_l = \Big\lceil \tfrac{m_l}{6l^{1/2}} \Big\rceil.$$

In the case that F is continuous, it holds that $F(X_I) = U(Z_I)$ and $F_n(X_I) = U_n(Z_I)$ for any $I = (j,k]$. Then T_n can be further written as

$$T_n \;=\; \max_{I \in \mathcal{J}_0}\Big([2\mathrm{logLR}_n\big\{U(Z_I), U_n(Z_I)\big\}]^{1/2} - \ell\big\{U_n(Z_I)\big\}\Big), \tag{S21}$$

and is thus independent of F. This makes it possible to estimate the distribution, as well as the quantile $\kappa_n(\alpha)$, of T_n via Monte Carlo simulations.

Consider now that F is a general distribution function. For $I = (j,k] \in \mathcal{J}_0 \cap (\mathcal{D}\times\mathcal{D})$, it still holds that $F_n(X_I) = U_n(Z_I) = (k-j)/n$, but $F(X_I) = U(Z_I)$ may no longer be valid. However, it is possible to find a universal distribution which stochastically dominates the distribution of T_n. More precisely, we introduce the notation $I_+ = (j, k+1]$ and $I_- = (j+1, k]$, for any $I = (j,k]$, and define

$$T_n^* \;=\; \max_{I\in\mathcal{J}_0}\Big\{\Big(2\max\big[\mathrm{logLR}_n\big\{U(Z_{I_+}), U_n(Z_I)\big\}, \mathrm{logLR}_n\big\{U(Z_{I_-}), U_n(Z_I)\big\}\big]\Big)^{1/2}$$

$$- \ell\big(U_n(Z_I)\big)\Big\}, \qquad \text{(S22)}$$

and $\kappa_n^*(\alpha)$ as its $(1-\alpha)$-quantile, which can be determined by Monte Carlo simulations.

LEMMA S2. *It holds that* $\kappa_n(\alpha) \le \kappa_n^*(\alpha)$ *and* $\sup_n \kappa_n^*(\alpha) < \infty$, *for every* $\alpha \in (0,1)$.

Proof. We claim that $i \in \mathcal{D}$ implies $Z_{(i)} \le F(X_{(i)}) < Z_{(i+1)}$. In fact, since $F\{F^{-1}(t)\} \ge t$ for all t, we have $Z_{(i)} \le F\{F^{-1}(Z_{(i)})\} = F(X_{(i)})$. If $F(X_{(i)}) \ge Z_{(i+1)}$, then, by the definition of F^{-1}, we have $X_{(i)} \ge F^{-1}(Z_{(i+1)}) = X_{(i+1)}$, which contradicts with $i \in \mathcal{D}$. Thus, $F(X_{(i)}) < Z_{(i+1)}$.

For an arbitrary interval $I = (j,k] \in \mathcal{J}_0 \cap (\mathcal{D} \times \mathcal{D})$, it then holds that

$$U(Z_{I_-}) = Z_{(k)} - Z_{(j+1)} < F(X_I) = F(X_{(k)}) - F(X_{(j)}) < Z_{(k+1)} - Z_{(j)} = U(Z_{I_+}).$$

As $\log\mathrm{LR}_n\{x, F_n(X_I)\}$ is decreasing for $x \le F_n(X_I)$ and increasing for $x \ge F_n(X_I)$, we have

$$\log\mathrm{LR}_n\{F(X_I), F_n(X_I)\} \le \max\left[\log\mathrm{LR}_n\{U(Z_{I_+}), U_n(Z_I)\},\ \log\mathrm{LR}_n\{U(Z_{I_-}), U_n(Z_I)\}\right].$$

This proves the first part, i.e., $\kappa_n(\alpha) \le \kappa_n^*(\alpha)$.

The second part, namely $\kappa_n^*(\alpha) = O(1)$, can be proven in a similar way as Rivera & Walther (2013, Theorem 1), by noticing that $|U_n(Z_{I_+}) - U_n(Z_I)| = |U_n(Z_{I_-}) - U_n(Z_I)| = 1/n$. $\square$

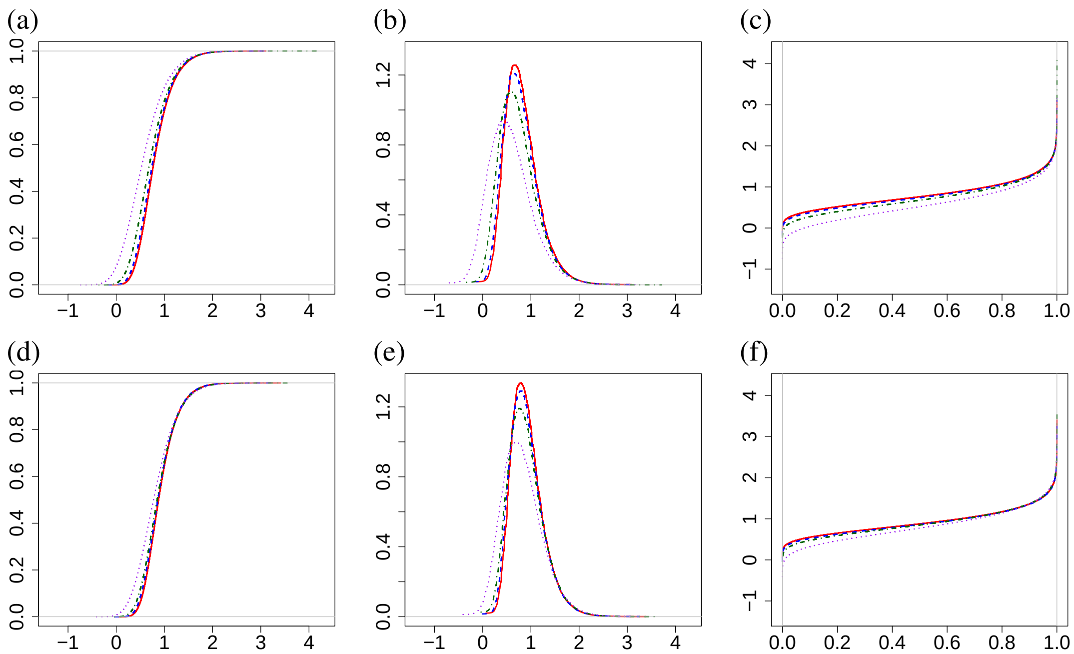

Fig. S1: The (a) distribution function, (b) density and (c) quantile function of T_n in (S21), equivalently, T_n in (3) under any continuous F; and the (d) distribution function, (e) density and (f) quantile function of T_n^* in (S22); for sample size $n = 10^3$ in dotted, $n = 10^4$ in dot-dash, $n = 5 \times 10^4$ in dashed and $n = 10^5$ in solid. For each sample size n, the distribution function, density, and quantile function are estimated over 10^5 random repetitions.

Lemma S2 ensures that all the theoretical results about the confidence set $C_n(\alpha)$ and the essential histogram remain valid if we use $\kappa_n^*(\alpha)$ in place of $\kappa_n(\alpha)$ as the threshold in (1), cf. §S2. Therefore, in practice, we choose $\kappa_n^*(\alpha)$ as the threshold whenever there are tied observations; otherwise we treat F as continuous, and use $\kappa_n(\alpha)$, determined by setting F as uniform, as the threshold. This is also the default choice in our R-pakcage essHist. As mentioned, $\kappa_n(\alpha)$

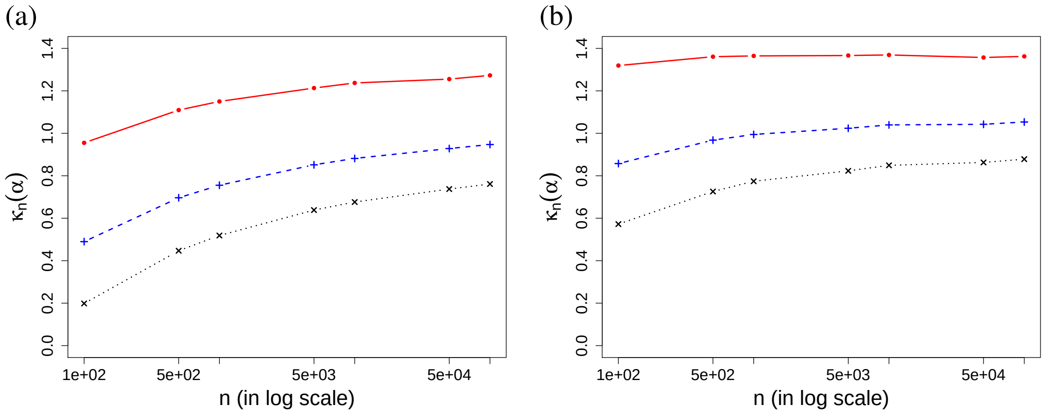

Fig. S2: Quantiles: (a) $\kappa_n(\alpha)$ of T_n in (S21), equivalently, T_n in (3) under any continuous F; (b) $\kappa_n^*(\alpha)$ of T_n^* in (S22); for $\alpha = 0.1$ in solid, $\alpha = 0.3$ in dashed and $\alpha = 0.5$ in dotted. For each sample size n, the quantiles are estimated over 10^5 random repetitions.

and $\kappa_n^*(\alpha)$ are estimated via Monte-Carlo simulations, which are needed only once for a given sample size, which are automatically recorded for later usage in our R-package essHist, while the computation time could be much longer than the pruned dynamic program in Algorithm 1 (see §S3.2) for large sample sizes, say $n \geq 10,000$. By simulation, we observe that the distributions of T_n and T_n^* seem to converge to a limit, and that such a convergence is fairly good when $n \geq 10,000$, see Figs. S1 and S2. This is theoretically underpinned by the tightness of T_n and T_n^*, see Rivera & Walther (2013, Proposition 1) and Lemma S2. For the sake of computational speed, we use the values of $\kappa_n(\alpha)$ and $\kappa_n^*(\alpha)$ with $n = 10,000$ as default values for sample sizes $n \geq 10,000$ in our R-package essHist. Through extensive simulation studies, we find that the performance of the essential histogram is hardly affected by such a default rule, while the computation gets a significant speedup.

S3.2. *Dynamic programming*

Now we consider the computation of the relaxed essential histogram defined in (5). As argued in §3.2, it can be computed by an accelerated dynamic programming algorithm. The idea of designing such an algorithm follows from Frick et al. (2014), but the constraint that the solution should be a histogram introduces an additional difficulty.

For notation simplicity, we consider the case that there are no ties in the data, otherwise it is sufficient to treat only distinct observations as the candidate locations of breakpoints, and the computation complexity will essentially depend on the number of distinct samples. Since the data can be sorted in $O(n \log n)$ computation time, we simply assume $X_i = X_{(i)}$, for $i = 1, \ldots, n$. For every interval I in $\mathcal{J}$, defined in (4), the corresponding local constraint in $\tilde{C}_n(\alpha)$ in (6) leads to simply an interval, namely,

$$[l_I, u_I] = \Big\{\mu \mid \Big[2\mathrm{logLR}_n\{\mu\,|I|\,, F_n(I)\}\Big]^{1/2} - \ell\{F_n(I)\} \leq \kappa_n(\alpha)\Big\}. \tag{S23}$$

Here l_I and u_I are roots of a smooth nonlinear equation, which can be computed in $O(1)$ time by standard algorithms, such as quasi-Newton methods (see e.g. Nocedal & Wright, 2006). For ease of exposition, we introduce the notation, for every $I \in \mathcal{J}$ and $\mu \in \boldsymbol{R}$

$$T_{n,I}(\mu) = \sup_{J \subset I,\, J \in \mathcal{J}} \Big(\Big[2\mathrm{logLR}_n\{\mu\,|J|\,, F_n(J)\}\Big]^{1/2} - \ell\{F_n(I)\}\Big).$$

We consider the multiscale constraint and entropy minimization, which is needed in case of multiple solutions, simultaneously, and thus define, for $j < i$

$$c_{(j,i]} = \begin{cases} -(i-j)\log \frac{i-j}{n(X_{(i)}-X_{(j)})} & \text{if } T_{n,(X_{(j)},X_{(i)}]}\{\frac{i-j}{n(X_{(i)}-X_{(j)})}\} \le \kappa_n(\alpha) \\ \infty & \text{otherwise} \end{cases}$$

Let $K[i]$ be the number of blocks of the essential histogram $\hat{h}_i$ being applied to restricted data $X_{(1)}, \ldots, X_{(i)}$, and $L[i]$ be the index of the leftmost sample in the last block of $\hat{h}_i$. Then it holds

$$\begin{aligned} K[0] &= -1, \\ K[i] &= \min\left\{K[j]+1 \mid c_{(j,i]} < \infty, j = 0,\ldots,i-1\right\}; \qquad \text{(S24)} \\ L[0] &= 1, \\ L[i] &= \operatorname*{arg\,min}_{j=0,\ldots,i-1} \left\{c_{(j,i]} \mid K[j]+1 = K[i]\right\}. \qquad \text{(S25)} \end{aligned}$$

The above relation is often referred to as *Bellman equation* (Bellman, 2010), which is the crucial part of a dynamic programing algorithm. Note that the histogram is completely determined by the segmentation, i.e. breakpoints. Thus, the essential histogram can be easily determined from vector $L = \{L[1], \ldots, L[n]\}$ in $O(K[n])$, less than $O(n)$, computation time. Now we are ready to present the whole algorithm for computing the essential histogram in Algorithm 1.

Algorithm 1. Pruned dynamic algorithm for the essential histogram

```
   Data: X_(1), ..., X_(n)
   Result: the essential histogram
1  k = 1, u = ∞, l = −∞, A_0 = {0}, B = {1, ..., n}, and K, L vectors of length n;
2  while i ∈ B and B ≠ ∅ do
3  |   u = min[u, min{u_I | I ∈ J, max I = i}] with u_I defined in (S23);
4  |   l = max[v, max{l_I | I ∈ J, max I = i}] with l_I defined in (S23);
5  |   Compute K[i] and L[i] by (S24) and (S25) over A_{k−1};
6  |   if l > u then
7  |   |   A_k = {i : K[i] = k}, and B = B \ A_k;
8  |   |   u = ∞, l = −∞, and k = k + 1;
9  |   end
10 end
11 Compute the histogram from L;
```

Unlike standard dynamic programs, Algorithm 1 incorporates two important pruning steps: One is in line 5, where the search space is $\mathcal{A}_{k-1}$ instead of $\{0, \ldots, i-1\}$; The other is in line 6, which prohibits further search towards right if no constant signal is admitted to the multiscale constraint, that is, no further right beyond $X_{(r_k)}$ with

$$r_k = \left\{ i \mid T_{n,(X_{(j)},X_{(i)}]}(\mu) \le \kappa(\alpha) \quad \text{for some } i \in \mathcal{A}_{k-1} \text{ and } \mu \in \boldsymbol{R} \right\}. \qquad \text{(S26)}$$

Further pruning is possible, for instance, by introducing a similar stopping rule as r_k in (S26) on the reverse order of the data $\{X_{(n)}, \ldots, X_{(1)}\}$ (see Pein et al., 2017). We refer to our R-package essHist for further technical details.

The computation complexity of Algorithm 1 is bounded from above by

$$O\Big\{\log n \sum_{k=1}^{K[n]} \#\mathcal{A}_{k-1}(r_k - \min \mathcal{A}_{k-1})\Big\} \leq O\Big\{n \log n \max_{k=1,\ldots,K[n]} (r_k - \min \mathcal{A}_{k-1})\Big\}$$

In particular, if $K[n] = 1$, it implies that the computation complexity is $O(n \log n)$. In most cases, $\max_{k=1,\ldots,K[n]}(r_k - \min \mathcal{A}_{k-1})$ stays bounded, which thus leads to a nearly linear computation complexity $O(n \log n)$. However, in very rare cases, $\max_{k=1,\ldots,K[n]}(r_k - \min \mathcal{A}_{k-1})$ can be of order n, which gives the worst case complexity $O(n^2 \log n)$. Clearly, the memory complexity of Algorithm 1 is always linear, i.e., $O(n)$.

Furthermore, we point out that Algorithm 1 applies to the multiscale constraint with arbitrary system of intervals besides $\mathcal{J}$.

S4. ADDITIONAL SIMULATION RESULTS

This section collects all quantitative comparison results for the essential histogram, the Davies & Kovac estimator, and the classical histograms, being a companion of §6.1.

While density estimation is not the primary purpose of the essential histogram, we now consider estimation errors measured by L^2-loss, and Kolmogorov loss. The former gives the mean integrated squared error, namely,

$$\mathrm{E}\left\{\int \left|f(x) - \hat{f}(x)\right|^2 dx\right\} \qquad \text{for true density } f \text{ and its estimator } \hat{f};$$

The latter leads to

$$\mathrm{E} \sup_x \left|\int_{-\infty}^{x} \{f(t) - \hat{f}(t)\} dt\right| \qquad \text{for true density } f \text{ and its estimator } \hat{f}.$$

In practice, the expectations are approximated by averages over independent repetitions.

From data exploration and shape recovery perspective, we introduce evaluation measures via numbers of modes/troughs, and skewness. For a histogram density $h = \sum_{k=0}^{K} c_k 1_{I_k}$ with $c_k \neq c_{k+1}$, we define the number of modes, i.e. maxima, as $\#\big\{k : c_k > \max(c_{k-1}, c_{k+1})\big\}$ and the number of troughs, i.e. minima, as $\#\{k : c_k < \min(c_{k-1}, c_{k+1})\}$. The number of extrema is then defined as the total number of modes and troughs. Note that the number of modes and troughs also capture the number of increases and decreases. From a slightly different viewpoint, the difference between the skewness of the estimator and that of the truth reflects how well the shape of the truth is recovered by the estimator.

Table S1 is for the uniform density example; Table S2 for the monotone density example; Tables S3 and S4 for the histogram density example; Tables S5, S6 and S7 for the claw density example; Tables S8, S9, S10 and S11 for the harp density example; and Tables S12 and S13 for the heavy tail example, in §6.1. Within each table, the best value along each column, i.e. among different methods, is marked in bold.

REFERENCES

BELLMAN, R. (2010). *Dynamic programming*. Princeton University Press, Princeton, NJ.
DÜMBGEN, L. (2003). Optimal confidence bands for shape-restricted curves. *Bernoulli* **9**, 423–449.
DÜMBGEN, L. & SPOKOINY, V. G. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.* **29**, 124–152.
DÜMBGEN, L. & WALTHER, G. (2008). Multiscale inference about a density. *Ann. Statist.* **36**, 1758–1785.

Table S1: Number of modes, i.e. false positives, on the uniform density, in Fig. 3, averaged over 500 repetitions

Methods		Number of observations				
		$n = 100$	$n = 300$	$n = 500$	$n = 700$	$n = 900$
Essential histogram	$\alpha = 0.1$	**0.000**	**0.002**	**0.000**	**0.000**	**0.000**
	$\alpha = 0.2$	0.004	0.006	0.004	0.006	0.008
	$\alpha = 0.3$	0.006	0.010	0.016	0.012	0.016
	$\alpha = 0.5$	0.030	0.046	0.054	0.048	0.072
	$\alpha = 0.7$	0.108	0.148	0.162	0.188	0.178
	$\alpha = 0.9$	0.424	0.442	0.518	0.548	0.560
Davies & Kovac		0.332	0.052	0.244	0.062	0.038
Equal bin width	Sturges' rule	5.126	5.220	5.194	5.148	5.262
	Scott's rule	1.938	3.344	3.914	4.798	5.300
Equal block area	Scott's rule	2.060	3.378	4.062	4.732	5.326

Table S2: Number of modes, i.e. false positives, on the exponential density, in Fig. 4, averaged over 500 repetitions

Methods		Number of observations				
		$n = 100$	$n = 300$	$n = 500$	$n = 700$	$n = 900$
Essential histogram	$\alpha = 0.1$	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
	$\alpha = 0.2$	0.006	**0.000**	**0.000**	**0.000**	0.004
	$\alpha = 0.3$	0.006	0.002	**0.000**	**0.000**	0.006
	$\alpha = 0.5$	0.014	0.012	0.006	0.008	0.012
	$\alpha = 0.7$	0.034	0.034	0.028	0.018	0.028
	$\alpha = 0.9$	0.100	0.066	0.074	0.064	0.076
Davies & Kovac		0.048	**0.000**	0.022	0.004	0.010
Equal bin width	Sturges' rule	0.650	0.610	0.534	0.610	0.566
	Scott's rule	0.328	0.942	1.406	1.674	2.074
Equal block area	Scott's rule	1.158	2.472	3.348	4.044	4.580

Table S3: Frequency of detecting the correct number of extrema on the histogram density, in Fig. 5, in 500 repetitions

Methods		Number of observations				
		$n = 600$	$n = 700$	$n = 800$	$n = 900$	$n = 1000$
Essential histogram	$\alpha = 0.1$	**95.6%**	**98.0%**	**99.2%**	**98.8%**	**98.4%**
	$\alpha = 0.2$	95.2%	97.6%	97.6%	96.4%	96.6%
	$\alpha = 0.3$	94.4%	95.2%	95.0%	94.8%	93.2%
	$\alpha = 0.5$	89.4%	90.4%	88.6 %	89.0%	88.6%
	$\alpha = 0.7$	83.0%	80.8%	79.8%	81.0%	78.6%
	$\alpha = 0.9$	64.6%	64.8%	68.6%	67.2%	60.8%
Davies & Kovac		50.2%	51.8%	50.4%	48.2%	46.2%
Equal bin width	Sturges' rule	33.6%	36.6%	31.6%	29.0%	34.0%
	Scott's rule	30.8%	28.6%	32.2%	29.8%	28.0%
Equal block area	Scott's rule	36.0%	42.2%	44.2%	44.8%	43.6%

DÜMBGEN, L. & WELLNER, J. (2014). Confidence bands for distribution functions: A new look at the law of the iterated logarithm. *arXiv: 1402.2918* .

FRICK, K., MUNK, A. & SIELING, H. (2014). Multiscale change point inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76**, 495–580. With 32 discussions by 47 authors and a rejoinder by the authors.

NOCEDAL, J. & WRIGHT, S. J. (2006). *Numerical optimization*. Springer, New York, 2nd ed.

PEIN, F., SIELING, H. & MUNK, A. (2017). Heterogeneous change point inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79**, 1207–1227.

RIVERA, C. & WALTHER, G. (2013). Optimal detection of a jump in the intensity of a Poisson process or in a density with likelihood ratio statistics. *Scand. J. Stat.* **40**, 752–769.

Table S4: Number of false bins, i.e., $\max\{N_{\mathtt{bin}}(\hat{F}) - N_{\mathtt{bin}}(F), 0\}$, on the histogram density, in Fig. 5, averaged over 500 repetitions

Methods		Number of observations				
		$n = 600$	$n = 700$	$n = 800$	$n = 900$	$n = 1000$
Essential histogram	$\alpha = 0.1$	**0.02**	**0.03**	**0.02**	**0.02**	**0.04**
	$\alpha = 0.2$	0.06	0.07	0.06	0.08	0.08
	$\alpha = 0.3$	0.11	0.10	0.11	0.13	0.14
	$\alpha = 0.5$	0.23	0.23	0.23	0.29	0.28
	$\alpha = 0.7$	0.40	0.47	0.45	0.49	0.53
	$\alpha = 0.9$	0.88	0.90	0.85	0.99	1.10
Davies & Kovac		7.23	6.89	7.58	7.75	8.73
Equal bin width	Sturges' rule	2.81	2.79	2.83	2.80	2.82
	Scott's rule	0.90	0.83	0.23	0.23	0.62
Equal block area	Scott's rule	1.00	1.13	2.00	2.00	2.15

Table S5: Number of modes on the claw density, in Fig. 6, averaged over 500 repetitions; The truth has 5 modes

Methods		Number of observations				
		$n = 1000$	$n = 1200$	$n = 1500$	$n = 2000$	$n = 3000$
Essential histogram	$\alpha = 0.1$	1.58	1.85	2.46	3.24	4.74
	$\alpha = 0.2$	1.92	2.35	2.99	3.74	4.9
	$\alpha = 0.3$	2.19	2.68	3.34	4.13	4.96
	$\alpha = 0.5$	2.65	3.19	3.91	4.6	4.99
	$\alpha = 0.7$	3.16	3.72	4.38	4.82	**5.00**
	$\alpha = 0.9$	3.84	4.39	4.75	4.96	**5.00**
Davies & Kovac		**4.99**	**5.00**	**5.00**	5.01	5.01
Equal bin width	Sturges' rule	1.23	1.24	1.19	1.11	1.08
	Scott's rule	4.39	4.86	5.58	6.23	6.84
Equal block area	Scott's rule	5.07	5.08	5.15	5.24	5.50

Table S6: Average number of bins, with standard deviation given in the parenthesis, on the claw density, in Fig. 6, over 500 repetitions

Methods		Number of observations				
		$n = 1000$	$n = 1200$	$n = 1500$	$n = 2000$	$n = 3000$
Essential histogram	$\alpha = 0.1$	**7.1** (1.1)	**8.4** (1.1)	**9.8** (1.1)	**11.2** (0.9)	**13.3** (0.8)
	$\alpha = 0.2$	8.1(1.2)	9.3 (1.1)	10.6 (1.0)	11.9 (**0.8**)	13.9 (0.8)
	$\alpha = 0.3$	8.7 (1.2)	9.9 (1.1)	11.2 (1.0)	12.4 (**0.8**)	14.2 (**0.7**)
	$\alpha = 0.5$	9.8 (1.2)	10.8 (1.1)	12.0 (0.9)	13.1 (**0.8**)	14.7 (**0.7**)
	$\alpha = 0.7$	10.7 (1.1)	11.7 (1.0)	12.7 (0.9)	13.7 (**0.8**)	15.1 (**0.7**)
	$\alpha = 0.9$	11.9 (1.0)	12.8 (1.0)	13.6 (0.9)	14.6 (0.9)	15.8 (0.9)
Davies & Kovac		48.6 (4.9)	52.5 (5.6)	57.8 (5.9)	67.7 (5.9)	79.1 (6.6)
Equal bin width	Sturges' rule	12.5 (**0.8**)	12.8 (**0.8**)	13.1 (**0.8**)	13.5 (0.9)	14.0 (0.8)
	Scott's rule	19.3 (1.2)	20.8 (1.2)	22.9 (1.3)	25.6 (1.3)	30.1 (1.4)
Equal block area	Scott's rule	20.5 (1.7)	22.1 (1.7)	24.3 (1.8)	27.4 (2.0)	32.4 (2.3)

SHORACK, G. R. & WELLNER, J. A. (1986). *Empirical processes with applications to statistics*. John Wiley & Sons, Inc., New York.

Table S7: Skewness of the histograms on the claw density, in Fig. 6, averaged over 500 repetitions; The skewness of the truth is 0

Methods		Number of observations				
		$n = 1000$	$n = 1200$	$n = 1500$	$n = 2000$	$n = 3000$
Essential histogram	$\alpha = 0.1$	**0.010**	0.002	0.010	0.007	**0.009**
	$\alpha = 0.2$	0.012	0.007	0.009	0.006	0.010
	$\alpha = 0.3$	0.016	0.005	0.007	**0.005**	0.010
	$\alpha = 0.5$	0.018	0.002	0.006	0.006	**0.009**
	$\alpha = 0.7$	0.017	0.002	0.008	0.007	0.010
	$\alpha = 0.9$	0.019	0.005	0.007	**0.005**	0.010
Davies & Kovac		0.057	0.043	0.039	0.038	0.035
Equal bin width	Sturges' rule	0.076	0.056	0.060	0.053	0.050
	Scott's rule	0.063	0.049	0.049	0.039	0.040
Equal block area	Scott's rule	0.032	**0.001**	**0.004**	0.008	0.014

Table S8: Frequency of detecting the correct number of extrema on the harp density, in Fig. 8, in 500 repetitions

Methods		Number of observations				
		$n = 600$	$n = 800$	$n = 1000$	$n = 1200$	$n = 1500$
Essential histogram	$\alpha = 0.1$	14.6%	64.4%	93.4%	98.8%	**100%**
	$\alpha = 0.2$	35.6%	80.0%	96.6%	99.4%	**100%**
	$\alpha = 0.3$	49.6%	88.6%	97.4%	**99.8%**	**100%**
	$\alpha = 0.5$	69.6%	95.2%	97.8%	**99.8%**	**100%**
	$\alpha = 0.7$	82.6%	97.4%	99.4%	**99.8%**	**100%**
	$\alpha = 0.9$	93.0%	98.4%	99.6%	**99.8%**	**100%**
Davies & Kovac		**99.8%**	**99.6%**	**99.8%**	**99.8%**	**100%**
Equal bin width	Sturges' rule	0.6%	0.0%	0.0%	2.0%	1.4%
	Scott's rule	0.0%	0.0%	0.0%	0.0%	0.0%
Equal block area	Scott's rule	1.0%	24.2%	81.0%	98.8%	**100%**

Table S9: Mean integrated squared error $\times 10^5$ on the harp density, in Fig. 8, averaged over 500 repetitions

Methods		Number of observations				
		$n = 600$	$n = 800$	$n = 1000$	$n = 1200$	$n = 1500$
Essential histogram	$\alpha = 0.1$	3.9	2.8	2.4	2.2	1.9
	$\alpha = 0.2$	3.4	2.6	2.3	2.0	1.8
	$\alpha = 0.3$	3.2	2.5	2.2	1.9	1.7
	$\alpha = 0.5$	2.9	2.4	2.1	1.8	1.6
	$\alpha = 0.7$	2.7	2.3	2.0	1.7	1.4
	$\alpha = 0.9$	2.6	2.2	1.9	1.6	1.3
Davies & Kovac		**1.5**	**0.8**	**0.6**	**0.4**	**0.3**
Equal bin width	Sturges' rule	61.9	63.1	63.7	55.8	57.3
	Scott's rule	35.0	31.7	30.1	31.0	36.0
Equal block area	Scott's rule	20.4	15.6	9.8	8.5	7.7

Table S10: Error measured by Kolmogorov metric on the harp density, in Fig. 8, averaged over 500 repetitions

Methods		Number of observations				
		$n = 600$	$n = 800$	$n = 1000$	$n = 1200$	$n = 1500$
Essential histogram	$\alpha = 0.1$	0.05	**0.04**	**0.03**	**0.03**	**0.03**
	$\alpha = 0.2$	0.05	**0.04**	**0.03**	**0.03**	**0.03**
	$\alpha = 0.3$	0.05	**0.04**	**0.03**	**0.03**	**0.03**
	$\alpha = 0.5$	**0.04**	**0.04**	**0.03**	**0.03**	**0.03**
	$\alpha = 0.7$	**0.04**	**0.04**	**0.03**	**0.03**	**0.03**
	$\alpha = 0.9$	**0.04**	**0.04**	**0.03**	**0.03**	**0.03**
Davies & Kovac		0.05	0.05	0.05	0.04	0.04
Equal bin width	Sturges' rule	0.18	0.18	0.18	0.15	0.16
	Scott's rule	0.09	0.09	0.08	0.08	0.08
Equal block area	Scott's rule	0.06	0.06	0.05	0.05	0.04

Table S11: Skewness of the histograms on the claw density, in Fig. 8, averaged over 500 repetitions; The skewness of the truth is 0.9

Methods		Number of observations				
		$n = 600$	$n = 800$	$n = 1000$	$n = 1200$	$n = 1500$
Essential histogram	$\alpha = 0.1$	0.94	0.92	0.91	0.91	0.91
	$\alpha = 0.2$	0.93	**0.91**	0.91	0.91	0.91
	$\alpha = 0.3$	0.92	**0.91**	0.91	0.91	0.91
	$\alpha = 0.5$	0.91	**0.91**	0.91	0.91	**0.90**
	$\alpha = 0.7$	0.91	**0.91**	0.91	**0.90**	**0.90**
	$\alpha = 0.9$	**0.90**	**0.91**	**0.90**	**0.90**	**0.90**
Davies & Kovac		0.60	0.65	0.69	0.72	0.75
Equal bin width	Sturges' rule	1.08	1.10	1.12	1.09	1.10
	Scott's rule	0.83	0.85	0.86	0.87	**0.90**
Equal block area	Scott's rule	0.98	0.94	0.93	0.96	0.98

Table S12: Frequency of detecting the correct number of extrema on the Cauchy density, in Fig. 9, in 500 repetitions

Methods		Number of observations				
		$n = 100$	$n = 200$	$n = 300$	$n = 400$	$n = 500$
Essential histogram	$\alpha = 0.1$	**100%**	**100%**	**100%**	**100%**	**100%**
	$\alpha = 0.2$	**100%**	**100%**	**100%**	**100%**	**100%**
	$\alpha = 0.3$	**100%**	**100%**	**100%**	**100%**	**100%**
	$\alpha = 0.5$	**100%**	**100%**	**100%**	99.8%	**100%**
	$\alpha = 0.7$	**100%**	**100%**	99.8%	99.8%	**100%**
	$\alpha = 0.9$	99.6%	99.6%	99.0%	99.4%	99.4%
Davies & Kovac		**100%**	**100%**	**100%**	**100%**	**100%**
Equal bin width	Sturges' rule	76.6%	72.8%	76.0%	74.0%	74.0%
	Scott's rule	35.6%	25.6%	21.0%	15.2%	15.4%
Equal block area	Scott's rule	5.4%	0.0%	0.0%	0.0%	0.0%

Table S13: Average number of bins, with standard deviation given in the parenthesis, on the Cauchy density, in Fig. 9, over 500 repetitions

Methods		Number of observations				
		$n = 100$	$n = 200$	$n = 300$	$n = 400$	$n = 500$
Essential histogram	$\alpha = 0.1$	**4.3** (0.58)	5.3 (**0.53**)	6.1 (0.59)	6.9 (**0.47**)	7.3 (**0.51**)
	$\alpha = 0.2$	4.5 (0.56)	5.6 (0.61)	6.4 (0.57)	7.2 (0.52)	7.7 (0.59)
	$\alpha = 0.3$	4.7 (**0.53**)	5.8 (0.61)	6.6 (0.56)	7.4 (0.58)	7.9 (0.63)
	$\alpha = 0.5$	4.9 (0.54)	6.1 (0.63)	6.9 (**0.51**)	7.8 (0.68)	8.4 (0.66)
	$\alpha = 0.7$	5.2 (0.54)	6.5 (0.66)	7.2 (0.58)	8.3 (0.71)	8.8 (0.68)
	$\alpha = 0.9$	5.6 (0.66)	7.1 (0.74)	7.7 (0.70)	9.0 (0.79)	9.4 (0.80)
Davies & Kovac		14.7 (2.59)	22.8 (3.27)	28.2 (3.60)	33.3 (3.85)	37.5 (4.14)
Equal bin width	Sturges' rule	4.4 (1.26)	**4.6** (1.47)	**4.7** (1.50)	**4.7** (1.48)	**4.7** (1.48)
	Scott's rule	6.0 (2.13)	7.7 (2.91)	8.9 (3.60)	10.0 (4.06)	10.8 (4.30)
Equal block area	Scott's rule	14.2 (1.79)	24.7 (3.44)	34.6 (4.75)	44.1 (6.22)	53.0 (7.35)