Supplementary material for “Maximum likelihood estimation of a unimodal probability mass function”

Fadoua Balabdaoui and Hanna Jankowski

October 13, 2015

Abstract

In this supplement we present some additional proofs, and discuss further the assumptions of the main manuscript.

Appendices

A Some further empirical coverage results

Table 1: Empirical coverage probabilities for the proposed confidence bands with $\alpha = 0.05$.

	β	R	$n = 100$	$n = 1000$	$n = 5000$
mixture of uniforms	0	S_n	**0.972**	**0.963**	**0.959**
		$\mathrm{supp}(p_0)$	0	**0.953**	**0.959**
	0.25	S_n	**0.991**	**0.971**	**0.970**
		$\mathrm{supp}(p_0)$	0	**0.961**	**0.970**
	0.5	S_n	**0.959**	**0.953**	**0.991**
		$\mathrm{supp}(p_0)$	0	**0.944**	**0.991**
double logarithmic $p = 0.9$	0	S_n	**0.956**	**0.949**	**0.949**
		$p_0(R) \approx 0.95$	0.001	0.922	**0.949**
	0.25	S_n	0.970	**0.950**	**0.948**
		$p_0(R) \approx 0.95$	0.001	0.922	**0.948**
	0.5	S_n	0.980	0.989	0.989
		$p_0(R) \approx 0.95$	0.001	**0.959**	0.990

Let $R \subseteq S_0$. The results in Table 1 give the empirical coverage *on the set* R as indicated in the third column. That is, we report the proportion of times that

$$\widehat{c}_{n,l}(s) \quad \leq \quad p_0(s) \quad \leq \quad \widehat{c}_{n,l}(s), \quad \text{for all } s \in R$$

was observed. The confidence bands are optimized for $R = \text{supp}(\widehat{p}_n)$, and this allows us to compare the behavior for other choices of R.

The results of Table 1 clearly show the cost of only defining the confidence bands on $\text{supp}(\widehat{p}_n) = S_n$ in the simulations. For larger sample sizes, this cost decreases. However, for small sample sizes, the undercoverage is drastically big, simply because S_n does not cover the set R yet. This issue aside, we find that the confidence bands perform rather well. In the double logarithmic setting for $\beta < 0.5$, we expect to obtain asymptotically correct coverage bands, and hence, empirical coverage probabilities not statistically different from 0.95 are shown in bold. In all uniform mixture scenarios, we expect an asymptotically conservative result; that is, the asymptotic coverage should is expected to be greater than 0.95. In the table, empirical coverage probabilities not statistically smaller than 0.95 are shown in bold (for the mixture of uniforms case).

B Proofs and technical details

B.1 Note on finding the MLE

In several proofs, we make use of the following idea (a well known practice in shape-constrained estimation problems):

To compute $\widehat{p}_n|_\kappa$, we first relax consideration over pmfs to positive sequences, $\mathcal{U}|_\kappa(S_n)$, by changing the criterion function L_n to

$$\Phi_n(p) \quad = \quad L_n(p) - \sum_{j=0}^{J-1} p(z_j) \quad = \quad L_n(p) - \sum_{z \in S_n} p(z). \tag{B-1}$$

This is possible because the two maximization problems are equivalent. To see this, note that if p is a positive sequence with support S_n maximizes Φ_n, then for all $c \in \mathbb{R}$ with $|c|$ very small

$$0 = \lim_{c \to 0} \frac{\Phi_n(p + cp) - \Phi_n(p)}{c} = 1 - \sum_{z \in S_n} p(z)$$

implying that p is necessarily a pmf. In the sequel, we denote by $\mathcal{U}(S_n)$ the space of positive unimodal sequences with support S_n.

B.2 Proofs from Section 4

Proof of Theorem 4.1. We first recall that

$$\rho(p|p_0) \quad \geq \quad 0, \tag{B-2}$$

with equality if and only if $p = p_0$ (P_0 a.s.). This result is often referred to as Gibbs' inequality. We now proceed progressively in steps. We first assume that $|S_0| \geq 2$.

1. Let Q denote the cdf of any discrete distribution on $\mathbb{N}$ and $\widehat{Q}$ denote its least concave majorant (on $\mathbb{N}$). Let $\widehat{q}$ denote the pmf associated with $\widehat{Q}$. We first claim that $\widehat{q}$ is such that

$$\int \log \frac{\widehat{q}}{p} dQ \quad \geq \quad 0, \qquad \text{for all decreasing pmf } p.$$

   This follows from the results of Patilea (1997, 2001) for decreasing densities as follows: Let $F_0(z) = Q(z-1)$ denote a cdf on $\mathbb{R}_+$, and let $\widehat{F_0}$ denote its least concave majorant on $\mathbb{R}_+$ (LCM), with associated pdf $\widehat{f_0}$. Then, from Patilea (1997, 2001) it follows that $\widehat{f_0}$ satisfies

$$\int \log \frac{\widehat{f_0}}{f} dF_0 \quad \geq \quad 0,$$

   for all decreasing densities f, and hence also for all decreasing densities with the form $f(x) = \int_0^\infty \theta^{-1} 1_{[0,\theta]}(x) d\mu(\theta)$, where μ is discrete with mass only at $\mathbb{Z}_+$. In other words, any f which is piecewise constant, with points of jump occurring possibly only at $\mathbb{Z}_+$. For such densities f, let $p(z) = \int_z^{z+1} f(x)dx = f(z+1), z \in \mathbb{N}$. In addition, note that $\widehat{q}(z) = \int_z^{z+1} \widehat{f_0}(x)dx = \widehat{f_0}(z+1)$, from the definition of $\widehat{Q}$ and $\widehat{F_0}$. Then we have that

$$\begin{aligned} \int \log \frac{\widehat{q}}{p} dQ &= \sum_{z \geq 0} \log \frac{\widehat{q}(z)}{p(z)} q(z) \\ &= \sum_{z \geq 0} \log \frac{\widehat{f_0}(z+1)}{f(z+1)} \{F_0(z+1) - F_0(z)\} \\ &= \int \log \frac{\widehat{f_0}}{f} dF_0 \quad \geq \quad 0, \end{aligned}$$

   and the result follows.

2. Next, let $\alpha = \sum_{i \geq \kappa} p_0(s_i)$, $q_1(i) = \alpha^{-1} p_0(s_{i+\kappa})$, and $q_2(i) = (1-\alpha)^{-1} p_0(s_{\kappa-1-i})$. Both q_1 and q_2 are pmfs on $\mathbb{N}$ and we apply step one above to find their $\widehat{q_1}, \widehat{q_2}$. Define $\widehat{p_0}|_\kappa(s_i) = \alpha \widehat{q_1}(i-\kappa), i \geq \kappa$ and $\widehat{p_0}|_\kappa(s_i) = (1-\alpha)\widehat{q_2}(\kappa-1-i)$

for $i \le \kappa - 1$. Then clearly $\widehat{p_0}|_\kappa \in \mathcal{U}^1|_\kappa(S_0)$. Furthermore, for any $p \in \mathcal{U}^1|_k(S_0)$

$$\begin{aligned}
\int \log \frac{\widehat{p_0}|_\kappa}{p} dP_0 &= \sum_{i \le \kappa-1} \log \frac{\widehat{p_0}|_\kappa(s_i)}{p(s_i)} p_0(s_i) + \sum_{i \ge \kappa} \log \frac{\widehat{p_0}|_\kappa(s_i)}{p(s_i)} p_0(s_i) \\
&= \sum_{i \le \kappa-1} \log \frac{(1-\alpha)\widehat{q_2}(\kappa-1-i)}{p(s_i)} p_0(s_i) \\
&\quad + \sum_{i \ge \kappa} \log \frac{\alpha \widehat{q_1}(i-\kappa)}{p(s_i)} p_0(s_i) \\
&= (1-\alpha) \sum_{i \le \kappa-1} \log \frac{(1-\alpha)\widehat{q_2}(\kappa-1-i)}{p(s_i)} \frac{p_0(s_i)}{1-\alpha} \\
&\quad + \alpha \sum_{i \ge \kappa} \log \frac{\alpha \widehat{q_1}(i-\kappa)}{p(s_i)} \frac{p_0(s_i)}{\alpha} \\
&= (1-\alpha) \sum_{z \ge 0} \log \frac{\widehat{q_2}(z)}{p_2(z)} q_2(z) + \alpha \sum_{z \ge 0} \log \frac{\widehat{q_1}(z)}{p_1(z)} q_1(z),
\end{aligned}$$

where $p_2(z) = (1-\alpha)^{-1} p(s_{\kappa-1-z})$ and $p_1(z) = \alpha^{-1} p(s_{\kappa+z})$. Now, let c_1 denote the constant such that $c_1 \sum_{z \ge 0} p_1(z) = 1$, and let $\widetilde{p}_1 = c_1 p_1$ (and similarly for p_2). Let $\beta = \sum_{z \ge 0} p(s_{\kappa+z})$. Then, we have that the above is equal to

$$\begin{aligned}
&(1-\alpha) \sum_{z \ge 0} \log \frac{\widehat{q_2}(z)}{p_2(z)} q_2(z) + \alpha \sum_{z \ge 0} \log \frac{\widehat{q_1}(z)}{p_1(z)} q_1(z) \\
&= (1-\alpha) \log c_2 + \alpha \log c_1 + (1-\alpha) \sum_{z \ge 0} \log \frac{\widehat{q_2}}{\widetilde{p}_2(z)} q_2(z) \\
&\quad + \alpha \sum_{z \ge 0} \log \frac{\widehat{q_1}}{\widetilde{p}_1(z)} q_1(z) \\
&\ge (1-\alpha) \log c_2 + \alpha \log c_1 = (1-\alpha) \log \frac{(1-\alpha)}{(1-\beta)} + \alpha \log \frac{\alpha}{\beta} \ge 0,
\end{aligned}$$

where the last inequality follows from the Gibbs' inequality in (B - 2) applied to two Bernoulli distributions with success probabilities α and β respectively. It follows that

$$\int \log \frac{\widehat{p_0}|_\kappa}{p} dP_0 \ge 0$$

for any $p \in \mathcal{U}^1|_\kappa(S_0)$. We have therefore proved existence of $\widehat{p_0}|_\kappa$.

3. Finally, we prove that $\widehat{p_0}|_\kappa$ as defined above is the unique solution to (4.11) in the two cases stated in the proposition.

   • Suppose that $p_0 \in \mathcal{U}^1|_\kappa(S_0)$. Then, by Gibbs' inequality, we have that

   $$\int \log \frac{p_0}{p} dP_0 \quad \ge \quad 0, \quad \forall p \in \mathcal{U}^1|_\kappa(S_0).$$

Suppose then that $\widehat{p}_0$ is another candidate for the KL projection, as above. Then we would have that

$$\int \log \frac{p_0}{\widehat{p}_0} dP_0 \quad \geq \quad 0 \quad \text{and also} \quad \int \log \frac{\widehat{p}_0}{p_0} dP_0 \quad \geq \quad 0.$$

But this implies that

$$\int \log \frac{p_0}{\widehat{p}_0} dP_0 \quad = \quad 0,$$

and (again by Gibbs' inequality) it follows that $\widehat{p}_0 = p_0$, P_0 a.s..

- Suppose that $p_0 \notin \mathcal{U}^1|_\kappa(S_0)$ with $\sum_{j \neq 0} \log |j| p_0(s_j) < \infty$. Then, by Proposition 4.3, we have that $\sup_{p \in \mathcal{U}^1|_\kappa(S_0)} \int \log p \, dP_0 \in (-\infty, 0]$. Hence, (4.11) is equivalent to

$$\widehat{p}_0 = \text{argmax}_{p \in \mathcal{U}^1|_\kappa(S_0)} \int \log p \, dP_0.$$

By the strict concavity of log, we have that $\log(\alpha a + (1-\alpha)b) \geq \alpha \log a + (1-\alpha) \log b$, with equality iff $a = b$. Suppose that $\widehat{p}_1$ and $\widehat{p}_2$ are two different pmfs at which the cross entropy achieves its maximum. Then, by convexity of $\mathcal{U}^1|_\kappa(S_0)$, $\widehat{p}_0 = \alpha \widehat{p}_1 + (1-\alpha)\widehat{p}_2$ is also in $\mathcal{U}^1|_\kappa(S_0)$, and hence

$$\begin{aligned} \int \log \widehat{p}_0 \, dP_0 &= \int \log \{\alpha \widehat{p}_1 + (1-\alpha)\widehat{p}_2\} \, dP_0 \\ &> \alpha \int \log \widehat{p}_1 dP_0 + (1-\alpha) \int \log \widehat{p}_2 dP_0 \\ &= \text{argmax}_{p \in \mathcal{U}^1|_\kappa(S_0)} \int \log p \, dP_0 \end{aligned}$$

which yields a contradiction. Therefore, we must have that $\widehat{p}_1 = \widehat{p}_2$, P_0-almost surely. But this implies that on the set $\{s \in S_0 : p_0(s) > 0\}$, $\widehat{p}_1$ and $\widehat{p}_2$ must both be equal to the slope of the greatest convex minorant (GCM) of the cumulative sum of $p_0(s_i)$ to the left of $s_{\kappa-1}$ and the slope of its LCM to the right of s_κ. Since the latter has the same support as p_0, we conclude that uniqueness has to hold everywhere.

Lastly, suppose that $|S_0| = 1$. Then $p_0 \in \mathcal{U}^1|_\kappa(S_0)$ must be unimodal with $S_0 = \{s_\kappa\}$, and $\widehat{p}_0 = p_0$. The same proof as the first part of point three above applies. □

Lemma B.1. *Suppose that $\sum_{j \neq 0} \log |j| \, p_0(s_j) < \infty$. Then for each $\kappa \in \mathbb{Z}$ such that $s_\kappa \in S_0$, there exists a $q \in \mathcal{U}^1|_\kappa(S_0)$ such that $\int \log q \, dP_0(x) \in (-\infty, 0]$.*

Proof. Fix $\kappa \in \mathbb{Z}$ with $\kappa \neq 0$. Then

$$\begin{aligned}\sum_{j\neq\kappa} \log|j-\kappa|\, p_0(s_j) &= \sum_{j\notin\{\kappa,0\}} \log|j|\,\frac{|j-\kappa|}{|j|}\, p_0(s_j) + \log|\kappa|\, p_0(s_0) \\ &= \sum_{j\neq 0,\kappa} \log|j|\, p_0(s_j) + \sum_{j\notin\{\kappa,0\}} \log\frac{|j-\kappa|}{|j|} p_0(s_j) + \log|\kappa|\, p_0(s_0) \\ &\leq \sum_{j\neq 0} \log|j|\, p_0(s_j) + \sum_{j\notin\{\kappa,0\}} \log\frac{|j-\kappa|}{|j|} p_0(s_j) + \log|\kappa|.\end{aligned}$$

Now, since $\lim_{|j|\to\infty} \log(|j-\kappa|/|j|) = 0$, and by the assumption of the lemma, all three terms above are finite, and it follows that $\sum_{j\neq\kappa} \log|j-\kappa|\, p_0(s_j) < \infty$ for all $\kappa \in \mathbb{Z}$.

Define a pmf q with support S_0 as

$$q(s_j) \quad \propto \quad \begin{cases} \frac{1}{|j-\kappa|\log^2|j-\kappa|} & j \neq \kappa-1, \kappa, \kappa+1, \\ \frac{1}{2\log^2 2} & j = \kappa-1, \kappa, \kappa+1, \end{cases} \tag{B-3}$$

for $s_j \in S_0$. Since $\int_2^\infty (x\log^2 x)^{-1} dx = 1/(\log 2)$, there exists a normalizing constant for $q \in \mathcal{U}^1|_\kappa(S_0)$. It remains to calculate its entropy. That is,

$$\sum_j \log q(s_j) p_0(s_j) \;=\; D - \sum_{|j-\kappa|\geq 2} \log|j-\kappa| p_0(s_j) - 2\sum_{|j-\kappa|\geq 2} \log\log|j-\kappa| p_0(s_j),$$

where D is some finite constant. The second term is also finite by the first part of this proof. For the last term we have that

$$0 \;\leq\; \sum_{|j-\kappa|\geq e} \log\log|j-\kappa| p_0(s_j) \;\leq\; \sum_{|j-\kappa|\geq e} \log|j-\kappa| p_0(s_j),$$

and hence this term is also finite. The result follows. □

Proof of Theorem 4.2. The first point can be shown using Gibbs' inequality as done above in the proof of Theorem 4.1. To prove the second point, we first note that by Proposition 4.3, under the assumptions of the theorem, (4.9) and (4.10) are equivalent. Therefore, to prove that (4.9) holds, it is sufficient to show that (4.10) holds, for some $\widehat{p}_0$. By Lemma B.1, for each $\kappa \in \mathbb{Z}$ there exists a $q \in \mathcal{U}^1|_\kappa(S_0)$ such that $\int \log q dP_0 > -\infty$. Therefore, each $\int \log \widehat{p}_0|_\kappa(x)\, dP_0(x) > -\infty$ (although this bound is not uniform in κ). Next, by Lemma C.1, we have that

$$\begin{aligned}\int \log \widehat{p}_0|_\kappa(x)\, dP_0(x) &= \sum_j \log \widehat{p}_0|_\kappa(s_j) p_0(s_j) \\ &\leq -\sum_{j\neq\kappa} \log|j-\kappa|\, p_0(s_j) \\ &\leq -\log|\kappa - m|\, p_0(s_m),\end{aligned}$$

for some fixed m such that $s_m \in S_0$. Letting $\kappa \to \pm\infty$, it follows that the maximum cannot be attained for large values of $|\kappa|$, and hence the supremum of $\int \log \widehat{p}_0|_\kappa(x)\, dP_0(x)$ can be found by considering a finite collection of values of κ. This proves existence of a maximizer $\widehat{p}_0 \in \mathcal{U}^1(S_0)$ (and also that $\{\widehat{p}_0\}$ is a finite set). ☐

Proof of Proposition 4.3. We first show that if $\sum_{j\neq 0} \log|j| dP_0 = \infty$, then $\int \log p dP_0 = -\infty$, for any unimodal p. This follows since, if p is unimodal, then $p \in \mathcal{U}^1|_\kappa(S_0)$ for some $\kappa \in \mathbb{Z}$. Hence, by Lemma C.1, we have that

$$\begin{aligned}\int \log p\, dP_0 &\leq \sum_j \log \min\left(1, |j-\kappa|^{-1}\right) p_0(s_j)\\ &= -\sum_{j\neq\kappa} \log|j-\kappa|\, p_0(s_j).\end{aligned}$$

Now, if $\kappa = 0$, then $\int \log p\, dP_0 \leq -\sum_{j\neq\kappa} \log|j-\kappa|\, p_0(s_j) = \sum_{j\neq 0} \log|j|\, p_0(s_j)$. If $\kappa \neq 0$, then

$$\begin{aligned}-\sum_{j\neq\kappa} \log|j-\kappa|\, p_0(s_j) &= -\sum_{j\notin\{\kappa,0\}} \log|j| \frac{|j-\kappa|}{|j|} p_0(s_j) - \log|\kappa|\, p_0(s_0)\\ &= -\sum_{j\neq 0} \log|j|\, p_0(s_j) - \sum_{j\notin\{\kappa,0\}} \log \frac{|j-\kappa|}{|j|} p_0(s_j) - \log|\kappa|\, p_0(s_0) + \log|\kappa| p_0(s_\kappa)\\ &\leq -\sum_{j\neq 0} \log|j|\, p_0(s_j) - \sum_{j\notin\{\kappa,0\}} \log \frac{|j-\kappa|}{|j|} p_0(s_j) + \log|\kappa|.\end{aligned}$$

Since $\lim_{|j|\to\infty} \log\left(|j-\kappa|/|j|\right) = 0$, there exists an integer $J > 0$ such that for all $|j| > J$

$$\log \frac{|j-\kappa|}{|j|} \in \left[-\frac{1}{2}, \frac{1}{2}\right].$$

Then,

$$\begin{aligned}&\sum_{j\notin\{\kappa,0\}} \log \frac{|j-\kappa|}{|j|} p_0(s_j)\\ &\quad= \sum_{j\notin\{\kappa,0\}, |j|\leq J} \log \frac{|j-\kappa|}{|j|} p_0(s_j) + \sum_{j\notin\{\kappa,0\}, |j|>J} \log \frac{|j-\kappa|}{|j|} p_0(s_j)\\ &\quad\geq \sum_{j\notin\{\kappa,0\}, |j|\leq J} \log \frac{|j-\kappa|}{|j|} p_0(s_j) - \frac{1}{2} \sum_{j\notin\{\kappa,0\}, |j|>J} p_0(s_j)\\ &\quad\geq \sum_{j\notin\{\kappa,0\}, |j|\leq J} \log \frac{|j-\kappa|}{|j|} p_0(s_j) - \frac{1}{2} = -C,\end{aligned}$$

for some finite constant C. Therefore,

$$\int \log p\, dP_0 \leq -\sum_{j\neq 0} \log|j| p_0(s_j) + C + \log|\kappa|,$$

and the first part of the claim follows (noting that since $\kappa \in \mathbb{Z}$ and $\kappa \neq 0$, then $\log|\kappa| < \infty$). The second part of the claim follows immediately from Lemma B.1. □

B.2.1 Proof of Theorem 4.4

We start by showing the following lemma.

Lemma B.2. *Suppose that* $\sum_{i\neq 0} \log|i|\, p_0(s_i) < \infty$. *Let* $\widehat{\mathcal{M}}_n$ *be the modal region of* $\widehat{p}_n$. *Then, we can find* $M > 0$ *sufficiently large, such that with probability one there exists an integer* $n_0 > 0$ *such that*

$$\sup_{n\geq n_0} \max_{\kappa\in\widehat{\mathcal{M}}_n} |\kappa| \leq M + 1.$$

Proof. Fix $\varepsilon_1 \in (0, p_0(s_0)/4)$, and define the event that $A_n^c = \{\sup|\mathbb{F}_n - F_0| \leq \varepsilon_1\}$. By the Dvoretzky-Kiefer-Wolfowitz inequality the probability of A_n is at most $2e^{-2n\varepsilon_1^2}$. Applying Lemma C.1, we have that

$$\begin{aligned} \int \log \widehat{p}_n|_\kappa d\mathbb{F}_n \quad &\leq \quad -\sum_{i\neq\kappa} \log|i-\kappa|\overline{p}_n(s_i) \\ &\leq \quad -\log|\kappa|\overline{p}_n(s_0) \quad \leq \quad -\log|\kappa|(p_0(s_0) - 2\varepsilon_1) \\ &\leq \quad -\log|\kappa|\, p_0(s_0)/2 \quad \leq \quad -\log M\, p_0(s_0)/2, \end{aligned}$$

if $|\kappa| > M$. Let B_n denote the event that $\int \log \widehat{p}_n|_\kappa d\mathbb{F}_n > -\log M\, p_0(s_0)/2$, whenever $|\kappa| > M$. By the above, we have that $B_n \subset A_n$. Since $P(A_n)$ is summable, the Borel-Cantelli lemma implies that $P(B_n \text{ i.o.}) = 0$. Thus, we have shown that, with probability one, there exists an integer n_1 such that for all $n \geq n_1$

$$\int \log \widehat{p}_n|_\kappa d\mathbb{F}_n \quad \leq \quad -\log M\, p_0(s_0)/2, \quad \forall |\kappa| > M.$$

Without loss of generality, we can assume that $S_0 = \{s_i, i \in K\}$ with $K = \mathbb{Z}$. Next, define q as in (B-3), and note that here we have

$$\sum_i |\log q(s_i)| p_0(s_i) < \infty,$$

using similar arguments to those used in the proof of Proposition 4.3. Recall that the pmf $q \in \mathcal{U}^1|_{\kappa=0}(S_0)$. Fix $\varepsilon_2 > 0$. By the strong law of large numbers, we can find with probability one an integer n_2 such that for all $n \geq n_2$

$$\int \log q\, d\mathbb{F}_n \quad \geq \quad \int \log q\, dF_0 - \varepsilon_2.$$

Since $\int \log q\, dF_0 \in \mathbb{R}$, we can furthermore choose ε_2 and M so that

$$\int \log q\, d\mathbb{F}_n \quad > \quad -\log M\, p_0(s_0)/2.$$

Thus, it follows that with probability one, there exists an n_0 (in fact, $n_0 = \max\{n_1, n_2\}$), such that

$$\int \log q \, d\mathbb{F}_n \quad > \quad \int \log \widehat{p}_n|_\kappa d\mathbb{F}_n$$

for all $|\kappa| > M$. But this implies that $\widehat{p}_n|_\kappa$ cannot be equal to the MLE $\widehat{p}_n$ when $|\kappa| > M$, proving the result. □

Proof of Theorem 4.4. We want to show that $\widehat{p}_n \to \widehat{p}_0$. First, we recall that pointwise convergence and convergence in $\ell_k, 1 \le k \le \infty$ and Hellinger distance h are all equivalent for sequences of pmfs. This follows for example from Lemma C.2 in the on-line supporting material of Balabdaoui et al. (2013). We also recall that a collection of probability measures is tight if, for all $\varepsilon > 0$, there exists a compact set $K = K(\varepsilon)$, such that for all measures μ in the collection, we have $\mu(K^c) < \varepsilon$. Let $\widehat{P}_n$ denote the measure induced by $\widehat{p}_n$. We first claim that $\{\widehat{P}_n\}_{n\ge1}$ is tight with probability one. Fix $\varepsilon > 0$. Then, by the Glivenko-Cantelli theorem, we can find with probability one an integer $n_1 > 0$ such that for all $n \ge n_1$, $\sup_{s\in S_0} |\mathbb{F}_n(s) - F_0(s)| < \varepsilon/6$. Also, by definition of the cdf, there exists a constant $M_0 > 0$ such that for all $M \ge M_0$,

$$1 - F_0(M) + F_0(-M-1) < \varepsilon/6.$$

Note that this implies that we have with probability one

$$\begin{aligned}
1 - \mathbb{F}_n(M) &+ \mathbb{F}_n(-M-1) \\
&= 1 - F_0(M) + F_0(-M-1) \\
&\quad + \{F_0(M) - \mathbb{F}_n(M)\} + \{\mathbb{F}_n(-M-1) - F_0(-M-1)\} \\
&< \varepsilon/2,
\end{aligned}$$

for all $n \ge n_1$ and all $M \ge M_0$.

Next, let $\widehat{\kappa}_n$ be such that $\widehat{p}_n \in \mathcal{U}^1|_{\widehat{\kappa}_n}(S_0)$. Then, by the result of Lemma B.2, with probability one, there exist $M > M_0$ and an integer $n_2 > 0$ such that for all $n \ge n_2$, $\sup_{n\ge n_2} |\widehat{\kappa}_n| \le M$. On this event, we have that

$$\begin{aligned}
\widehat{P}_n\left([-M, M]^c\right) &= \sum_{z\ge M+1} \widehat{p}_n(z) + \sum_{z\le -M-1} \widehat{p}_n(z) \\
&\le \sum_{z\ge M+1} \overline{p}_n(z) + \sum_{z\le -M-1} \overline{p}_n(z) \\
&= 1 - \mathbb{F}_n(M) + \mathbb{F}_n(-M-1) < \varepsilon/2
\end{aligned}$$

where the inequality in the second line follows from Proposition C.2. We have therefore shown that there exists a sufficiently large $n_0 = \max\{n_1, n_2\}$ such that $\{\widehat{P}_n\}_{n\ge n_0}$ is tight. Since any finite collection of distributions is also tight, it follows that $\{\widehat{P}_n\}_{n\ge1}$ is tight, with probability one.

Since $\{\widehat{P}_n\}$ is tight, it is also sequentially compact. Thus, let $\{\widehat{P}_{n_k}\}$ denote a weakly convergent subsequence, which, for convenience, we continue to denote as $\{\widehat{P}_n\}$. The Portmanteau theorem then implies that the associated pmf $\widehat{p}_n(s_i)$ converges for all $s_i \in S_0$ (since $(s-\delta, s+\delta)$ are continuity sets for appropriate choice of δ), and we let $\widetilde{p}$ denote the limiting pmf. To complete the proof, we need only show that $\widetilde{p}$ is an element of $\{\widehat{p}_0\}$. Note that convergence in the set metric then follows because $\{\widehat{p}_0\}$ is necessarily a finite set.

Now, since we maximize the criterion function $\int \log p \, d\mathbb{F}_n - \sum_{z \in S_n} p(z)$ (B-1) over positive and unimodal sequences and since $\sum_{z \in S_n} \widehat{p}_n(z) = 1$, we can write

$$\begin{aligned} \sum \log \widehat{p}_0(z_j)\, \overline{p}_n(z_j) - \sum \widehat{p}_0(z_j) &\leq \sum \log \widehat{p}_n(z_j)\, \overline{p}_n(z_j) - 1 \\ &\leq \sum \log(b + \widehat{p}_n(z_j))\, \overline{p}_n(z_j) - 1, \end{aligned}$$

for $b > 0$. Re-arranging the terms above, this yields

$$\begin{aligned} 0 &\leq \sum \log(b + \widehat{p}_n(z_j))\, \overline{p}_n(z_j) - \sum \log \widehat{p}_0(z_j)\, \overline{p}_n(z_j) + \sum \widehat{p}_0(z_j) - 1 \\ &\leq \sum \log(b + \widehat{p}_n(z_j))\, \overline{p}_n(z_j) - \sum \log \widehat{p}_0(z_j)\, \overline{p}_n(z_j), \end{aligned}$$

where the last inequality follows since $\sum \widehat{p}_0(z_j) \leq 1$. Finally, because $\overline{p}_n$ puts all of its mass only on the z_j, we can re-write the latter as

$$0 \leq \sum \log(b + \widehat{p}_n(s_j))\, \overline{p}_n(s_j) - \sum \log \widehat{p}_0(s_j)\, \overline{p}_n(s_j).$$

On the other hand, we have that

$$\begin{aligned} &\sum \log(b + \widehat{p}_n(s_j))\, \overline{p}_n(s_j) - \sum \log \widehat{p}_0(s_j)\, \overline{p}_n(s_j) \\ &= \sum \log(b + \widehat{p}_n(s_j))\, (\overline{p}_n(s_j) - p_0(s_j)) + \sum \log \widehat{p}_0(s_j)\, (p_0(s_j) - \overline{p}_n(s_j)) \\ &\quad + \sum \log\left(\frac{b + \widehat{p}_n(s_j)}{b + \widehat{p}_0(s_j)}\right) p_0(s_j) + \sum \log\left(\frac{b + \widehat{p}_0(s_j)}{\widehat{p}_0(s_j)}\right) p_0(s_j). \end{aligned} \tag{B-4}$$

Next we get rid of the first two terms on the right-hand side. First, using summation by parts,

$$\begin{aligned} &\sum \log(b + \widehat{p}_n(s_j))\, (\overline{p}_n(s_j) - p_0(s_j)) \\ &\qquad = \sum (\mathbb{F}_n(s_j) - F_0(s_j)) \left[\log(b + \widehat{p}_n(s_j)) - \log(b + \widehat{p}_n(s_{j-1}))\right]. \end{aligned}$$

Now, we know that $\widehat{p}_n = \widehat{p}_n|_\kappa$ for some κ. Then,

$$\begin{aligned} &\left|\sum \log(b + \widehat{p}_n(s_j))\, (\overline{p}_n(s_j) - p_0(s_j))\right| \\ &\leq \sup |\mathbb{F}_n(s_j) - F(s_j)| \left\{ \sum_{j \leq \kappa - 1} \left[\log(b + \widehat{p}_n(s_j)) - \log(b + \widehat{p}_n(s_{j-1}))\right] \right. \\ &\qquad + \left|\log(b + \widehat{p}_n(s_\kappa)) - \log(b + \widehat{p}_n(s_{\kappa-1}))\right| \\ &\qquad \left. + \sum_{j \geq \kappa + 1} \left[\log(b + \widehat{p}_n(s_j)) - \log(b + \widehat{p}_n(s_{j-1}))\right] \right\} \\ &\leq 4 \left|\log(b + \max_j \widehat{p}_n(s_j))\right| \sup |\mathbb{F}_n(s_j) - F_0(s_j)| \\ &\leq 4 \max\{\log(1+b), |\log(b)|\} \sup |\mathbb{F}_n(s_j) - F_0(s_j)|, \end{aligned}$$

which converges to zero. The law of large numbers shows that the second term also converges to zero. This follows because $\sup_{p\in\mathcal{U}^1(S_0)}\int \log p\, dP_0 > -\infty$, which implies that $\sum|\log\widehat{p}_0(s_j)|p_0(s_j) < \infty$. Therefore, rearranging (B-4), we find that

$$\limsup_n \sum \log\left(\frac{b+\widehat{p}_0(s_j)}{b+\widehat{p}_n(s_j)}\right) p_0(s_j) \quad\leq\quad \sum \log\left(\frac{b+\widehat{p}_0(s_j)}{\widehat{p}_0(s_j)}\right) p_0(s_j).$$

Now, letting $b\to 0$, we have by Fatou's lemma that

$$\limsup_{b\to 0}\limsup_n \sum \log\left(\frac{b+\widehat{p}_0(s_j)}{b+\widehat{p}_n(s_j)}\right) p_0(s_j) \quad\leq\quad 0.$$

Next, we take the limits on the right-hand side. First, by the dominated convergence theorem

$$\limsup_n \sum \log\left(\frac{b+\widehat{p}_0(s_j)}{b+\widehat{p}_n(s_j)}\right) p_0(s_j) \quad=\quad \sum \log\left(\frac{b+\widehat{p}_0(s_j)}{b+\widetilde{p}(s_j)}\right) p_0(s_j),$$

since $|\log((b+\widehat{p}_0)/(b+\widehat{p}_n))| \leq 2\max\{\log(b+1), |\log b|\}$. Next, we want to show that

$$\lim_{b\downarrow 0} \sum \log\left(\frac{b+\widehat{p}_0(s_j)}{b+\widetilde{p}(s_j)}\right) p_0(s_j) \quad=\quad \sum \log\left(\frac{\widehat{p}_0(s_j)}{\widetilde{p}(s_j)}\right) p_0(s_j). \qquad \text{(B-5)}$$

To do this, consider both pieces separately. First, $\log(b+\widehat{p}_0(z))$ is decreasing in b and bounded above by $\log 2$, and hence by the monotone convergence theorem we have that

$$\lim_b \sum \log\left(b+\widehat{p}_0(s_j)\right) p_0(s_j) \quad=\quad \sum \log\widehat{p}_0(s_j)\, p_0(s_j).$$

Similarly $-\log(b+\widetilde{p}(s_j))$ is increasing as b decreases, and bounded below by $-\log 2$. Therefore also,

$$\lim_b \sum \log\left(b+\widetilde{p}(s_j)\right) p_0(s_j) \quad=\quad \sum \log\widetilde{p}(s_j)\, p_0(s_j).$$

Note that $\int \log p\, dP_0$ is always finite for any unimodal p, and therefore we may subtract the last two lines above to yield (B-5). We have thus shown that

$$\sum \log\left(\frac{\widehat{p}_0(s_j)}{\widetilde{p}(s_j)}\right) p_0(s_j) \quad\leq\quad 0.$$

Rearranging, this gives

$$\sup_{p\in\mathcal{U}^1(S_0)}\int \log p\, dP_0 \quad=\quad \sum \log\widehat{p}_0(s_i)\, p_0(s_i) \quad\leq\quad \sum \log\widetilde{p}(s_i)\, p_0(s_i),$$

and hence $\widetilde{p} \in \{\widehat{p}_0\}$. □

Recall the definition of knots in (5.14) and the preceding paragraph.

Lemma B.3. *Suppose that* $\sum_{j\neq 0}\log|j|p_0(s_j) < \infty$ *and* $|\{\widehat{p}_0\}| = 1$. *Let* $\tau \in \mathcal{T}$ *be a knot point of* p_0. *Then, almost surely, there exists an* n_0 *such that for all* $n \geq n_0$ *we have that* τ *is also a knot of* $\widehat{p}_n$.

Proof. Without loss of generality, assume that $\tau = s_{k_0}$ and that $\widehat{p}_0(s_{k_0}) > \widehat{p}_0(s_{k_0-1})$. Then, from Theorem 4.4, we know that $\sup|\widehat{p}_0(s_j) - \widehat{p}_n(s_j)| < \varepsilon$, where $\varepsilon < (\widehat{p}_0(s_{k_0}) - \widehat{p}_0(s_{k_0-1}))/2$, for all sufficiently large n. Therefore,

$$\widehat{p}_n(s_{k_0}) \geq \widehat{p}_0(s_{k_0}) - \varepsilon \quad > \widehat{p}_0(s_{k_0-1}) + \varepsilon \geq \widehat{p}_n(s_{k_0-1}),$$

and the result follows. □

Proof of Corollary 4.6. Write $\mathcal{M} = \{s_{k_1}, \ldots, s_{k_2}\}$, and note that, by definition, we have that $\widehat{p}_0(s_{k_1}) > \widehat{p}_0(s_{k_1-1})$ and $\widehat{p}_0(s_{k_2}) > \widehat{p}_0(s_{k_2+1})$. From Lemma B.3 and the ℓ_1 consistency results of Theorem 4.4, it follows that, with probability one, there exists an n_0 such that for all $n \geq n_0$,

$$\begin{array}{rcccl} \widehat{p}_n(s_j) & \leq & \widehat{p}_n(s_{k_1-1}) & < & \widehat{p}_n(s_{k_1}), \quad j \leq k_1 - 1, \\ \widehat{p}_n(s_{k_2}) & > & \widehat{p}_n(s_{k_2+1}) & \geq & \widehat{p}_n(s_j), \quad j \geq k_2 + 1. \end{array}$$

This, of course, implies that the mode of $\widehat{p}_n$ must be in $\mathcal{M} = \{s_{k_1}, \ldots, s_{k_2}\}$. □

Proof of Corollary 4.7. This is an immediate consequence of Theorem 4.4 and the inequality

$$|\widehat{F}_n(s_i) - \widehat{F}_0(s_i)| \quad \leq \quad \sum_j |\widehat{p}_n(s_j) - \widehat{p}_0(s_j)|.$$

□

B.3 Proof of Theorem 5.1

Let $\{\mathbb{W}_n(s), s \in S_n\} \equiv \{\sqrt{n}(\overline{p}_n(s) - p_0(s)), s \in S_n\}$, denote the empirical white noise process.

Proposition B.4. *Let* $C = \left\{\cup_{j=1}^k I_j\right\} \cup \left\{\cup_{j=1}^k D_j\right\}$ *with* k *finite and* I_j, D_j *defined as in (5.13). Then, with probability one, there exists an integer* $n_0 > 0$ *such that for* $n \geq n_0$

$$\sqrt{n}(\widehat{p}_n - p_0)(s) = \varphi[\mathbb{W}_n](s), \ \textit{for all } s \in C.$$

Proof. By the strong law of large numbers, with probability one, we can find an integer $n_1 > 0$ such that for all $n \geq n_1$, $C \subset S_n$. Next, by Corollary 4.6, with probability one, we can find $n_2 \geq n_1$ such that for $n \geq n_2$ we have that $\widehat{\mathcal{M}}_n \subset \mathcal{M}$. This means that the $\widehat{p}_n$ is found as the minimizer in $\mathcal{U}^1|_\kappa(S_0)$

where $\kappa \in \mathcal{M}$. By Lemma B.3, again with probability one, we can find an $n_3 \geq n_2$ such that the knots $\tau_i^I, \tau_i^D, i = 0, \ldots, k$ are also knots of $\widehat{p}_n$ for all $n \geq n_3$ (recall the definitions of the knots from (5.14) and the preceding paragraph). Therefore, by Lemma C.4, for all $n \geq n_3$ we have that for $1 \leq j \leq k$

$$\begin{aligned}
\widehat{p}_n(s) &= \text{iso}[(\overline{p}_n)_{I_j}](s), s \in I_j, \\
\widehat{p}_n(s) &= \text{anti}[(\overline{p}_n)_{D_j}](s), s \in D_j.
\end{aligned}$$

That is, we have that $\widehat{p}_n(s) = \varphi[\overline{p}_n](s), s \in C$, for $n \geq n_3$. Since p_0 is constant on each I_j, D_j by definition, this implies that

$$\sqrt{n}(\widehat{p}_n - p_0)(s) = \varphi[\mathbb{W}_n](s), \text{ for all } s \in C,$$

see Lemma C.5. □

Lemma B.5. *Let $\mathbb{V}$ be a mean-zero Gaussian vector of dimension $d > 0$ with variance-covariance matrix Σ given by $cov(\mathbb{V}_i, \mathbb{V}_j) = d^{-1}\delta_{i=j} - d^{-2}$. Then* uni$[\mathbb{V}]$ *is unique with probability one.*

Proof. Suppose that $\widehat{\mathbb{V}}_1$ and $\widehat{\mathbb{V}}_2$ are two different solutions for the minimization problem. Our goal will be to show that $P(\widehat{\mathbb{V}}_1 \neq \widehat{\mathbb{V}}_2) = 0$. Since any minimizer of uni$(\mathbb{V})$ can be re-written as local averages of the original vector $\mathbb{V}$, it follows that we can find $d \times d$ matrices $\widehat{A}_1$ and $\widehat{A}_2$ such that $\widehat{\mathbb{V}}_1 = \widehat{A}_1\mathbb{V}$ and $\widehat{\mathbb{V}}_2 = \widehat{A}_2\mathbb{V}$, where $\widehat{A}_i, i = 1, 2$ can be written as

$$\widehat{A}_i = \begin{bmatrix} \widehat{A}_i^1 & 0 & 0 & \ldots & 0 \\ 0 & \widehat{A}_i^2 & 0 & \ldots & 0 \\ & & \cdots & & \\ 0 & 0 & 0 & 0 & \widehat{A}_i^{m_i} \end{bmatrix},$$

with $\widehat{A}_i^j$, $1 \leq j \leq m_i$, given by the $l_j \times l_j$ matrix

$$\widehat{A}_i^j = \frac{1}{l_j}\begin{bmatrix} 1 & 1 & \ldots & 1 \\ \vdots & \vdots & \ldots & \vdots \\ 1 & 1 & \ldots & 1 \end{bmatrix}.$$

Also, note that if $\widehat{\mathbb{V}}_i = \widehat{A}_i\mathbb{V}$ then

$$||\widehat{\mathbb{V}}_i - \mathbb{V}||_2^2 = \mathbb{V}^T(I - \widehat{A}_i)\mathbb{V}.$$

Finally, let $\mathcal{A}$ denote the set of all possible matrices $\widehat{A}_i$, and note that $|\mathcal{A}|$ is finite. Hence,

$$\begin{aligned}
P\left(\widehat{\mathbb{V}}_1 \neq \widehat{\mathbb{V}}_2\right) &= P\left(\widehat{\mathbb{V}}_1 \neq \widehat{\mathbb{V}}_2, \mathbb{V}^T(I - \widehat{A}_1)\mathbb{V} = \mathbb{V}^T(I - \widehat{A}_2)\mathbb{V}\right) \\
&\leq \sum_{B_1, B_2, \in \mathcal{A}, B_1 \neq B_2} P\left(\mathbb{V}^T(I - B_1)\mathbb{V} = \mathbb{V}^T(I - B_2)\mathbb{V}\right) \\
&= \sum_{B_1, B_2, \in \mathcal{A}, B_1 \neq B_2} P\left(\mathbb{V}^T(B_1 - B_2)\mathbb{V} = 0\right).
\end{aligned}$$

Let $S = \Sigma^{1/2}$ so that we can write $\mathbb{V} = SZ$ for $Z \sim \mathcal{N}_d(0, I)$. The matrix $S^T(B_1 - B_2)S$ is Hermitian, and therefore admits a spectral decomposition, which we write as $\Gamma\Lambda\Gamma^T$, where Γ is an orthogonal matrix and $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p, -\lambda_{p+1}, \dots, -\lambda_d)$ with $\lambda_i \geq 0,\ 1 \leq i \leq d$. Note that since $B_1 \neq B_2$, there exists at least one index $i \in \{1, \dots, d\}$ such that $\lambda_i \neq 0$. It is also important to note that only $B \in \mathcal{A}$ with $m = 1$ yields $B\mathbb{V} = 0$. Finally, let $U = \Gamma^T Z$. Note that $U \sim \mathcal{N}(0, I)$. Then, we can write

$$\begin{aligned} P\left(\mathbb{V}^T(B_1 - B_2)\mathbb{V} = 0\right) &= P(Z^T\Gamma\Lambda\Gamma^T Z = 0) \\ &= P(U^T\Lambda U = 0) \\ &= P(\lambda_1 U_1^2 + \ldots + \lambda_p U_p^2 = \lambda_{p+1}U_{p+1}^2 + \ldots + \lambda_d U_d^2). \end{aligned}$$

Notice that in the last line at least one of the quantities on the left or right hand side is not equal to zero and that in such case it is a continuous random variable (in fact, each has a gamma distribution). Also, notice that the left hand side is independent of the right hand side. This shows that $P\left(\mathbb{V}^T(B_1 - B_2)\mathbb{V} = 0\right) = 0$, and the result follows. □

Proof of Theorem 5.1. The proof is divided into several main steps. We first address a slight technicality: the MLE $\widehat{p}_n$ is defined on S_n, while p_0 is defined on S_0. The results we prove here all "live" in the space of ℓ_k sequences defined on S_0. To make our results concrete, we therefore embed all sequences on S_n into sequences on S_0 by setting them equal to zero for $s \notin S_n$.

Below, we present the proof for $\ell_k(S_0)$ with $k = 2$ only. Convergence for $3 \leq k \leq \infty$ follows immediately, because $||q||_k \leq ||q||_2$ for $k \geq 2$ and $q \in \ell_2$ and hence $||\cdot||_k$ is a continuous mapping on $\ell_2(S_0)$ for $k \geq 2$.

1. We first show that $\mathbb{W}_n$ converges in $\ell_2(S_0)$ to the limit $\mathbb{W}$. This is essentially a well known result (cf. Jankowski and Wellner (2009, Theorem 3.1)), noting that for $s \notin S_n$, $\mathbb{W}_n(s) = -\sqrt{n}p_0(s) = \sqrt{n}(\overline{p}_n(s) - p_0(s))$ is still well-defined, since for $s \notin S_n, \overline{p}_n(s) = 0$.

2. We will next show that $\sqrt{n}(\widehat{p}_n - p_0) \Rightarrow \varphi[\mathbb{W}]$ in $\ell_2(S_0 \setminus \mathcal{M})$. That is, we consider the sequence *only* on the set $S_0 \setminus \mathcal{M}$. This result is proved in two sub-steps:

   (a) We first show that φ is continuous in $\ell_2(S_0 \setminus \mathcal{M})$. This, together with step one above implies that $\varphi[\mathbb{W}_n] \Rightarrow \varphi[\mathbb{W}]$ in $\ell_2(S_0 \setminus \mathcal{M})$.

   (b) The next step is to show that $||\sqrt{n}(\widehat{p}_n - p_0) - \varphi[\mathbb{W}_n]||_2^2 \xrightarrow{p} 0$ (where the ℓ_2 norm is calculated only on the support $S_0 \setminus \mathcal{M}$). In fact, we prove slightly stronger convergence (in expectation).

3. Finally, we will tackle convergence on the set $\mathcal{M}$. This follows essentially from the argmax continuous mapping theorem. Note that since $|\mathcal{M}|$ is necessarily finite, we also have convergence in $\ell_2(\mathcal{M})$ of the process on the set $\mathcal{M}$.

4. To put the two results together, note that the convergence in steps two and three can also be stated as joint convergence (and not *just* convergence of marginals). This holds because of the joint convergence of $\mathbb{W}_n$ in step one. From here the full result follows.

We now fill in the details in steps 2 and 3 above. To prove 2(a), consider a converging sequence $q_n \to q$ in $\ell_2(S_0 \setminus \mathcal{M})$ and fix $\varepsilon > 0$. Then we can find an integer n_0 and $K > 0$ large enough such that

$$\sup_{n \geq n_0} \sum_{|i|>K} q_n^2(s_i) < \varepsilon/6, \text{ and } \sum_{|i|>K} q^2(s_i) < \varepsilon/6.$$

Now, let $K_1 \leq -K$ and $K_2 \geq K$ be such that $s_{K_1}, s_{K_2} \in \mathcal{T}$. We then have that

$$\begin{aligned}\sum_{s_i \notin \mathcal{M}} (\varphi[q_n](s_i) - \varphi[q](s_i))^2 \quad &\leq \sum_{i \in [K_1, K_2], s_i \notin \mathcal{M}} (\varphi[q_n](s_i) - \varphi[q](s_i))^2 \\ &+ 2 \sum_{i \notin [K_1, K_2]} \varphi[q_n]^2(s_i) + 2 \sum_{i \notin [K_1, K_2]} \varphi[q]^2(s_i).\end{aligned}$$

Now, by Lemma C.5 (choosing $p = q = 0$) we have that

$$\sum_{i \notin [K_1, K_2]} \varphi[q_n]^2(s_i) \leq \sum_{i \notin [K_1, K_2]} q_n^2(s_i),$$

and similarly for q_n replaced with q. Also, by continuity of the operators iso and anti (Proposition C.6), we can choose an $n_1 \geq n_0$ such that for all $n \geq n_1$

$$\sum_{i \in [K_1, K_2], s_i \notin \mathcal{M}} (\varphi[q_n](s_i) - \varphi[q](s_i))^2 < \varepsilon/3.$$

It follows that for all $n \geq n_1$, we have that

$$\begin{aligned}\sum_{s_i \notin \mathcal{M}} (\varphi[q_n](s_i) - \varphi[q](s_i))^2 \quad &\leq \quad \varepsilon/3 + 2 \sum_{|i| \geq K} q_n^2(s_i) + 2 \sum_{|i| \geq K} q^2(s_i) \\ &\leq \quad \varepsilon/3 + 4\varepsilon/6 = \varepsilon.\end{aligned}$$

This shows that φ is continuous in $\ell_2(S_0 \setminus \mathcal{M})$.

To prove 2(b), we fix $\varepsilon > 0$ and pick K large enough so that $\sum_{|i|>K} p_0(s_i) < \varepsilon$. Now, let $K_1 \leq -K$ and $K_2 \geq K$ be such that $s_{K_1}, s_{K_2} \in \mathcal{T}$. Let $\widehat{\mathbb{W}}_n(s) = \sqrt{n}(\widehat{p}_n - p_0)(s)$. Then

$$\begin{aligned}\sum_{s_i \notin \mathcal{M}} (\widehat{\mathbb{W}}_n - \varphi[\mathbb{W}_n])^2(s_i) \quad &\leq \sum_{i \in [K_1, K_2], s_i \notin \mathcal{M}} (\widehat{\mathbb{W}}_n(s_i) - \varphi[\mathbb{W}_n](s_i))^2 \\ &+ 2 \sum_{i \notin [K_1, K_2], s_i \in S_n} \widehat{\mathbb{W}}_n^2(s_i) + 2 \sum_{i \notin [K_1, K_2], s_i \in S_n} \mathbb{W}_n^2(s_i) \\ &+ 4 \sum_{i \notin [K_1, K_2], s_i \notin S_n} \mathbb{W}_n^2(s_i).\end{aligned}$$

Now, for n large enough, by Proposition B.4 we have that $\widehat{\mathbb{W}}_n(s_i) = \varphi[\mathbb{W}_n](s_i)$ for all $i \in [K_1, K_2], s_i \notin \mathcal{M}$. Also, $\sum_{i \notin [K_1,K_2], s_i \in S_n} \widehat{\mathbb{W}}_n^2(s_i) \le \sum_{i \notin [K_1,K_2], s_i \in S_n} \mathbb{W}_n^2(s_i)$ by Lemma C.5. We therefore have that for n sufficiently large

$$||\widehat{\mathbb{W}}_n - \varphi[\mathbb{W}_n]||_2^2 \quad \le \quad 4 \sum_{|i|>K} \mathbb{W}_n^2(s_i).$$

$$\begin{aligned} 0 \le \overline{\lim}\, E\left[||\widehat{\mathbb{W}}_n - \varphi[\mathbb{W}_n]||_2^2\right] \quad &\le \quad E\left[\overline{\lim}\, ||\widehat{\mathbb{W}}_n - \varphi[\mathbb{W}_n]||_2^2\right] \\ &\le \quad 4E\left[\overline{\lim} \sum_{|i|>K} \mathbb{W}_n^2(s_i)\right] \le 4E\left[\sum_{|i|>K} \mathbb{W}^2(s_i)\right] \\ &= \quad \sum_{|i|>K} p_0(s_i) < \varepsilon. \end{aligned}$$

Since ε was arbitrary, this proves the result.

Finally, we tackle step 3. We will do this by applying the argmax continuous mapping theorem, cf. van der Vaart and Wellner (1996, Theorem 3.2.2, page 286). Let $L_n(p)$ denote again the empirical log-likelihood, and recall that

$$\widehat{p}_n \quad = \quad \mathrm{argmax}_{\kappa \in S_n} L_n(\widehat{p}_n|_\kappa).$$

Now, by Lemma B.3 applied to τ_0^I and τ_0^D, and by Lemma C.4 we can also have that

$$\widehat{p}_n \quad = \quad \mathrm{argmax}_{\kappa \in \mathcal{M}} L_n(\widehat{p}_n|_\kappa),$$

and furthermore, each $\widehat{p}_n|_\kappa(s), s \in \mathcal{M}$ is determined by the LCM/GCM characterization only on $\mathcal{M}$. Let $d = |\mathcal{M}|$ and recall the definition of $\mathcal{U}_d$ from Section 2.2.2 as the space of unimodal vectors of length d. Also, let $\mathcal{U}_d^+ = \{u \in \mathcal{U}_d : u > 0\}$. For $s \in \mathcal{M}$, and for sufficiently large n, we have that

$$\begin{aligned} \sqrt{n}(\widehat{p}_n - p_0) \quad &= \quad \sqrt{n}\left\{\mathrm{argmin}_{p \in \mathcal{U}_d^+}\left[-\sum_{s\in\mathcal{M}} \log\left(\frac{p}{\overline{p}_n}\right)\overline{p}_n + \sum_{s\in\mathcal{M}} p\right] - p_0\right\} \\ &= \quad \mathrm{argmin}_{q \in \sqrt{n}(\mathcal{U}_d^+ - p_0)}\left[-\sum_{s\in\mathcal{M}} \log\left(\frac{p_0 + q/\sqrt{n}}{\overline{p}_n}\right)\overline{p}_n + \sum_{s\in\mathcal{M}} \left(p_0 + q/\sqrt{n}\right)\right] \\ &= \quad \mathrm{argmin}_{q \in \sqrt{n}(\mathcal{U}_d^+ - p_0)}\left[-\sum_{s\in\mathcal{M}} \log\left(\frac{p_0 + \frac{q}{\sqrt{n}}}{\overline{p}_n}\right)\overline{p}_n + \sum_{s\in\mathcal{M}} \frac{q}{\sqrt{n}} - \sum_{s\in\mathcal{M}} \frac{\mathbb{W}_n}{\sqrt{n}}\right], \end{aligned}$$

since $\sum_{s \in \mathcal{M}} p_0$ and $\sum_{s\in\mathcal{M}} \mathbb{W}_n$ are constants on which the minimization does not depend. Now, let

$$
\begin{aligned}
\mathbb{M}_n(q) &= -\sum_{s\in\mathcal{M}} \log\left(\frac{p_0 + q/\sqrt{n}}{\bar{p}_n}\right)\bar{p}_n + \frac{1}{\sqrt{n}}\sum_{s\in\mathcal{M}} q - \frac{1}{\sqrt{n}}\sum_{s\in\mathcal{M}} \mathbb{W}_n \\
&= -\sum_{s\in\mathcal{M}} \log\left(1 + \frac{1}{\sqrt{n}}\frac{q - \mathbb{W}_n}{\bar{p}_n}\right)\bar{p}_n + \frac{1}{\sqrt{n}}\sum_{s\in\mathcal{M}} q - \frac{1}{\sqrt{n}}\sum_{s\in\mathcal{M}} \mathbb{W}_n \\
&\approx -\sum_{s\in\mathcal{M}} \left\{\frac{1}{\sqrt{n}}\frac{q - \mathbb{W}_n}{\bar{p}_n} - \frac{1}{2}\left(\frac{1}{\sqrt{n}}\frac{q - \mathbb{W}_n}{\bar{p}_n}\right)^2\right\}\bar{p}_n + \sum_{s\in\mathcal{M}} \frac{q}{\sqrt{n}} - \sum_{s\in\mathcal{M}} \frac{\mathbb{W}_n}{\sqrt{n}} \\
&= \frac{1}{2n}\sum_{s\in\mathcal{M}} \frac{(q - \mathbb{W}_n)^2}{\bar{p}_n} = \widetilde{\mathbb{M}}_n(q)
\end{aligned}
$$

where

$$
n\widetilde{\mathbb{M}}_n(q) \quad \Rightarrow \quad \frac{1}{2}\sum_{s\in\mathcal{M}} \frac{(q - \mathbb{W})^2}{p_0}.
$$

Finally, since p_0 is constant on $\mathcal{M}$, we would therefore like to conclude that

$$
\sqrt{n}(\widehat{p}_n - p_0) \quad \Rightarrow \quad \operatorname{argmin}_{q\in\mathcal{U}_{|\mathcal{M}|}} \sum_{s\in\mathcal{M}} (q - \mathbb{W})^2 \quad = \quad \operatorname{uni}[\mathbb{W}_{\mathcal{M}}],
$$

where $\mathbb{W}_{\mathcal{M}}$ denotes the vector of random variables $\{\mathbb{W}(s), s \in \mathcal{M}\}$. To do this, we need to check the criteria of the argmax continuous mapping theorem, that is

1. $\sqrt{n}(\widehat{p}_n - p_0)$ is tight ("uniformly tight" in the sense of van der Vaart and Wellner (1996)) since it is equal to $\sqrt{n}(\widehat{p}_n|_{\widehat{\kappa}_n} - p_0)$ for some $\widehat{\kappa}_n \in \mathcal{M}$, and each $\sqrt{n}(\widehat{p}_n|_k - p_0)$ converges, using for example, Marshall's lemma.

2. The requirement that $\mathbb{M}_n(\sqrt{n}(\widehat{p}_n - p_0)) \geq \sup_q \mathbb{M}_n(q)$ is satisfied by definition of the $\widehat{p}_n$.

3. By Lemma B.5, $\sum_{s\in\mathcal{M}}(q - \mathbb{W})^2$ has a unique minimum on $\mathcal{U}_d$, that is, $\operatorname{uni}[\mathbb{W}_{\mathcal{M}}]$ has a unique solution. To see this, recall that on $\mathcal{M}, \mathbb{W}$ is normally distributed with mean zero and covariance given by $\operatorname{cov}(\mathbb{W}(s_i), \mathbb{W}(s_j)) = \theta\delta_{i,j} - \theta^2$, letting $\theta = p_0(s), s \in \mathcal{M}$. Now, define $\mathbb{V}(s) = (\theta|\mathcal{M}|)^{-1/2}(\mathbb{W}(s) - \sum_{s\in\mathcal{M}} \mathbb{W}(s)/|\mathcal{M}|)$, so that $\mathbb{W} = (\theta d)^{1/2}\mathbb{V} + \sum_{s\in\mathcal{M}} \mathbb{W}(s)/d$, using $d = |\mathcal{M}|$. A quick check shows that $\mathbb{V}$ is still normally distributed with mean zero and $\operatorname{cov}(\mathbb{V}(s_i), \mathbb{V}(s_j)) = d^{-1}\delta_{i,j} - d^{-2}$. Applying Lemma C.5, we have that

$$
\operatorname{uni}[\mathbb{W}] \quad = \quad (\theta d)^{1/2}\operatorname{uni}[\mathbb{V}] + \sum_{s\in\mathcal{M}} \mathbb{W}(s)/d.
$$

   By Lemma B.5, $\operatorname{uni}[\mathbb{V}]$ has a unique solution, and therefore, $\operatorname{uni}[\mathbb{W}]$ does also.

4. Note lastly that $\sum_{s\in\mathcal{M}}(q - \mathbb{W})^2$ is a.s. continuous in q.

The result follows. □

B.4 Proof of Proposition 5.3

Lemma B.6. *Suppose that* $\sum_{s\in S_0} p_0^{1/2}(s) < \infty$. *Then*

$$\sqrt{n}\sum_{s\notin S_n} p_0(s) = o_p(1).$$

Proof. Let $P_0(A) = \sum_{s\in A} p_0(s)$ and $\mathbb{P}_n(A) = \sum_{s\in A} \overline{p}_n(s)$. By the Borisov-Durst theorem (Dudley, 1999, Theorem 7.9, page 279) the power set of S_0, 2^{S_0}, is a Donsker class for P_0 if and only of $\sum_{s\in S_0} p_0^{1/2}(s) < \infty$. Now, let $\mathbb{G}$ denote the zero-mean Gaussian random field on 2^{S_0} with covariance

$$E\left[\mathbb{G}(A)\mathbb{G}(B)\right] = P_0(A\cap B) - P_0(A)P_0(B).$$

The Borisov-Durst theorem tells us that

$$\sqrt{n}(\mathbb{P}_n - P_0) \quad\Rightarrow\quad \mathbb{G} \quad \text{in } \ell_\infty(2^{S_0}). \tag{B-6}$$

Since S_0 is countable, we have that

$$\sup_{A\in 2^{S_0}} \left|\mathbb{P}_n(A) - P_0(A)\right| = \frac{1}{2}\,\ell_1(\bar{p}_n, p_0) \to 0,$$

almost surely as $n\to\infty$ since the class 2^{S_0} is also Glivenko-Cantelli. Since by definition, $\mathbb{P}_n(S_n^c) = 0$, the latter implies that

$$\lim_{n\to\infty} P_0(S_n^c) = 0 \tag{B-7}$$

almost surely. Furthermore, using the Skorokhod representation the convergence in (B-6) (see e.g. Theorem 1.10.4 of van der Vaart and Wellner, 1996) implies that we can assume that there exists a common probability space on which $\sqrt{n}(\mathbb{P}_n - P_0)$ and $\mathbb{G}$ are defined such that

$$\sup_{A\in 2^{S_0}} \left|\sqrt{n}(\mathbb{P}_n(A) - P_0(A)) - \mathbb{G}(A)\right| \to 0$$

almost surely. This implies that

$$\lim_{n\to\infty} \left(\sqrt{n}(\mathbb{P}_n(S_n^c) - P_0(S_n^c)) - \mathbb{G}_{P_0}(S_n^c)\right) = 0$$

almost surely. However, $\mathbb{G}(S_n^c) \overset{d}{=} Z\sqrt{P_0(S_n^c)(1-P_0(S_n^c))}$ with $Z\sim\mathcal{N}(0,1)$. Using this along with $\mathbb{P}_n(S_n^c) = 0$ and (B-7) it follows that

$$\lim_{n\to\infty} \sqrt{n}P_0(S_n^c) = 0.$$

We conclude that on the original probability space $\sqrt{n}P_0(S_n^c) \to_d 0$ which is equivalent to

$$\sqrt{n}P_0(S_n^c) = \sqrt{n}\sum_{s\in S_n^c} p_0(s) \overset{p}{\to} 0,$$

because the limit is degenerate. □

Proof of Proposition 5.3. Let us first fix $\kappa \in \mathcal{M}$. From Corollary 4.6 we have that with probability one, there exists a sufficiently large n_0 such that $\widehat{p}_n = \widehat{p}_n|_{\widehat{\kappa}_n}$ with $\widehat{\kappa}_n \in \mathcal{M}$ if $n \geq n_0$. This also implies that $\mathcal{M} \subset S_n$ for all $n \geq n_0$. Consider such an n.

Since $\kappa \in \mathcal{M}$, we know that $p_0 \in \mathcal{U}^1|_\kappa(S_0)$. From the characterization of the restricted MLE, we know that $\widehat{F}_{n|k}(s), s \in S_n$ (the associate CDF) is found as the least concave majorant of the graph

$$\{(i, \mathbb{F}_n(z_i)), k \leq i \leq m\}$$

where z_i denotes an ordered enumeration of the elements of $S_n = \{z_1, \ldots, z_m\}$ and $k = k(\kappa, S_n)$ is such that $z_k = s_\kappa$ (recalling that $\mathcal{M} \subset S_n$). Next, define the function

$$\overline{F}_0(z_i) = \sum_{j \leq i} p_0(z_j).$$

This depends of course on the observed S_n. Note that by definition this function is concave on $k \leq i \leq m$ and convex on $1 \leq i \leq k-1$. Now, the usual proof of Marshall's lemma applies. That is, let $a = \sup_{i \geq k} |\mathbb{F}_n(z_i) - \overline{F}_0(z_i)|$. Then for all $i \geq k$, we have (1)

$$\widehat{F}_{n|k}(z_i) - \overline{F}_0(z_i) \geq \mathbb{F}_n(z_i) - \overline{F}_0(z_i) \geq -a.$$

On the other hand $\overline{F}_0(z_i) + a$ is a concave majorant of $\mathbb{F}_n(z_i)$ on $i \geq k$, and hence (2) $\overline{F}_0(z_i) + a \geq \widehat{F}_n(z_i)$. Combining the results of (1) and (2) gives that $\sup_{i \geq \kappa} |\widehat{F}_n(z_i) - \overline{F}_0(z_i)| \leq \sup_{i \geq \kappa} |\mathbb{F}_n(z_i) - \overline{F}_0(z_i)|$. Repeating the argument on greatest concave minorants, yields a similar result for $i \leq \kappa - 1$, which combined gives

$$\sup_{z \in S_n} |\widehat{F}_{n|k}(z) - \overline{F}_0(z)| \leq \sup_{z \in S_n} |\mathbb{F}_n(z) - \overline{F}_0(z)|.$$

This result holds for any choice of $\kappa \in \mathcal{M}$. Next, from Corollary 4.6 we have that with probability one, there exists a sufficiently large n_0 such that $\widehat{p}_n = \widehat{p}_n|_{\widehat{\kappa}_n}$ with $\widehat{\kappa}_n \in \mathcal{M}$. Let $\widehat{F}_n$ denote the CDF associated with $\widehat{p}_n$. We then have that

$$\begin{aligned} \sup_{z \in S_n} |\widehat{F}_n(z) - \overline{F}_0(z)| &\leq \sup_{\kappa \in \mathcal{M}} \sup_{z \in S_n} |\widehat{F}_{n|k}(z) - \overline{F}_0(z)| \\ &\leq \sup_{z \in S_n} |\mathbb{F}_n(z) - \overline{F}_0(z)|. \end{aligned}$$

Next, it follows that

$$\begin{aligned} \sup_{s \in S_n} |\widehat{F}_n(s) - \overline{F}_0(s)| &\leq \sup_{s \in S_n} |\mathbb{F}_n(s) - F_0(s)| + \sup_{s \in S_n} |F_0(s_i) - \overline{F}_0(s_i)| \\ &\leq \sup_{s \in S_0} |\mathbb{F}_n(s) - F_0(s)| + \sum_{s \in S_n^c} p_0(s). \end{aligned}$$

On the other hand,

$$\begin{aligned}|\widehat{F}_n(s) - \overline{F}_0(s)| &= |\widehat{F}_n(s) - F_0(s) + F_0(s) - \overline{F}_0(s)| \\ &\geq |\widehat{F}_n(s) - F_0(s)| - |F_0(s) - \overline{F}_0(s)|\end{aligned}$$

This yields

$$\sup_{s\in S_0} |\widehat{F}_n(s) - F_0(s)| \leq \sup_{s\in S_0} |\mathbb{F}_n(s) - F_0(s)| + 2 \sum_{s\in S_n^c} p_0(s).$$

The full result is obtained by applying Lemma B.6. □

B.5 Proofs for Section 6

Proof of Proposition 6.1. Using our assumption of finite support, the result follows immediately from Theorems 4.4 and 5.1 via Slutsky's theorem and the continuity of norms on $\mathbb{R}^d$. Next, recall the definition of φ in (5.15). We have that

$$\frac{\varphi(\mathbb{W})}{p_0^\beta} = \varphi\left(\frac{\mathbb{W}}{p_0^\beta}\right)$$

since p_0 is constant on the intervals $\mathcal{I}_j, \mathcal{M}, \mathcal{D}_j$. The final inequality now follows as in Proposition 5.2. □

Proposition B.7. *Let $\mathbb{W}$ denote a mean zero Gaussian process defined on S_0 such that $cov(\mathbb{W}(s_i), \mathbb{W}(s_j)) = p_0(s_i)\delta_{i,j} - p_0(s_i)p_0(s_j),\ s_i \in S_0$. Let $\widetilde{\mathbb{W}}_n$ denote a mean zero Gaussian process defined on $supp(\widehat{p}_n)$ such that $cov(\widetilde{\mathbb{W}}_n(s_i), \widetilde{\mathbb{W}}_n(s_j)) = \widehat{p}_n(s_i)\delta_{i,j} - \widehat{p}_n(s_i)\widehat{p}_n(s_j),\ s_i \in supp(\widehat{p}_n)$. Let $q_{0,\alpha}$ and $\widetilde{q}_{0,\alpha}$ denote the quantiles such that*

$$P(\|\mathbb{W}\|_\infty > q_{0,\alpha}) = \alpha, \qquad P(\|\widetilde{\mathbb{W}}_n\|_\infty > \widetilde{q}_{0,\alpha}) = \alpha,$$

respectively. Then $\widetilde{q}_{0,\alpha} \to q_{0,\alpha}$ almost surely.

Proof. First, let p_n denote *any fixed* pmf such that p_n converges to p_0 and has the same properties as $\widehat{p}_n$:

(a) p_n converges pointwise to p_0, and

(b) $\lim_{m\to\infty} \lim_n \sum_{|s_i|>m} p_n(s_i) = 0$.

Suppose also that $\widetilde{\mathbb{W}}_n$ is defined as above, except that p_n replaces $\widehat{p}_n$ in the definition (in essence, we remove the randomness associated with this choice). Then one can easily show that $\widetilde{\mathbb{W}}_n$ converges weakly to $\mathbb{W}$ in ℓ_2. This follows from (a) convergence of finite dimensional distributions, which is immediate from convergence of p_n to p_0, and (b) tightness in ℓ_2. To prove tightness, we refer again to Jankowski and Wellner (2009, Lemma 6.2). Note that we have that

1. $E[\|\widetilde{\mathbb{W}}_n\|_2^2] \le 1$ for all n

2. For sufficiently large n, we have that

$$\sum_{|s_i|>m} E[\widetilde{\mathbb{W}}_n^2(s_i)] \quad \le \quad \sum_{|s_i|>m} p_n(s_i),$$

which shows that $\widetilde{\mathbb{W}}_n$ is tight in ℓ_2. The required weak convergence follows. Now, since the ℓ_∞ is continuous in ℓ_2, convergence of the quantiles follows.

Thus, we obtain convergence of the quantiles (as numbers), based on conditions (a) and (b) of p_n. We will now show that these conditions hold almost surely, establishing the full result. Condition (a) follows immediately from Theorem 4.4. To see also that Condition (b) holds, note that from Propositions C.2 and B.4, there exists a sufficiently large n such that with probability one

$$\sum_{|s_i|>m} \widehat{p}_n(s_i) \quad = \quad \widehat{F}_n(-m) + 1 - \widehat{F}_n(m) \quad \le \quad \mathbb{F}_n(-m) + 1 - \mathbb{F}_n(m)$$

for $m \notin \mathcal{M}$. That $\lim_m \lim_n (\mathbb{F}_n(-m) + 1 - \mathbb{F}_n(m)) = 0$ almost surely follows from the properties of the empirical CDF and CDFs in general. □

C Additional Technical Results

C.1 Useful bounds

Lemma C.1. *Any $p \in \mathcal{U}^1|_\kappa(S_0)$ satisfies*

$$p(s_j) \le \min\left\{1, |j-\kappa|^{-1}\right\}.$$

Proof. We have that

$$1 \quad \ge \quad \sum_{i=\kappa}^{j} p(s_i) \quad \ge \quad \sum_{i=\kappa}^{j} p(s_j) \quad = \quad (j-\kappa+1)p(s_j) \quad \ge \quad (j-\kappa)p(s_j).$$

Similarly, we have

$$1 \quad \ge \quad \sum_{i=j}^{\kappa-1} p(s_i) \quad \ge \quad \sum_{i=j}^{\kappa-1} p(s_j) \quad = \quad (\kappa-j)p(s_j).$$

Together, these yield the first inequality. □

Proposition C.2. *The restricted MLE $\widehat{p}_n|_\kappa$ satisfies the inequalities*

$$\begin{aligned} \widehat{F}_n|_\kappa(z) \quad &\ge \quad \mathbb{F}_n(z) \quad z \ge s_\kappa, \\ \widehat{F}_n|_\kappa(z) \quad &\le \quad \mathbb{F}_n(z) \quad z \le s_{\kappa-1}. \end{aligned}$$

Proof. Follows immediately from the GCM/LCM characterization of $\widehat{p}_n|_\kappa$. □

C.2 Proof of Proposition 2.1

Suppose that there exists q such that p satisfies (2.4). It is clear that p is a pmf. We now verify that p is unimodal with mode either at $s_{\kappa-1}$ or s_κ. Let $(\Delta p)(j) = p(s_{j+1}) - p(s_j)$. We calculate

$$(\Delta p)(j) \quad = \quad \begin{cases} -\frac{q(s_{j-\kappa})}{j-\kappa+1} & \leq \ 0 \quad j \geq \kappa, \\ \frac{q(s_{j+1-\kappa})}{|j+1-\kappa|} & \geq 0 \quad j \leq \kappa - 2. \end{cases}$$

Therefore, p is non-decreasing on $\{s_j : j \geq \kappa\}$ and non-increasing on $\{s_j : j \leq \kappa - 1\}$. For $j = \kappa - 1$, we calculate

$$p(s_\kappa) - p(s_{\kappa-1}) = \sum_{i=\kappa}^{\infty} \frac{q(s_i)}{i - \kappa + 1} - \sum_{-\infty}^{i=\kappa-1} \frac{q(s_i)}{|i - \kappa|}$$

which could be either ≥ 0 or < 0. This shows that p is unimodal with mode either at $s_{\kappa-1}$ or s_κ.

Conversely, if p is a pmf which is unimodal with mode either at $s_{\kappa-1}$ or s_κ. Let q be defined as

$$q(s_i) \quad = \quad \begin{cases} -(i-\kappa+1)(\Delta p)(i) & i \geq \kappa, \\ |i-\kappa|(\Delta p)(i-1) & i \leq \kappa - 1. \end{cases}$$

By the property of p, $q \geq 0$. Furthermore, using Fubini's theorem and the fact that p is a pmf, we have that

$$\begin{aligned} \sum_j q(s_j) &= -\sum_{j\geq\kappa}(j-\kappa+1)(\Delta p)(j) + \sum_{j\leq\kappa-1}(\kappa-j)(\Delta p)(j-1) \\ &= -\sum_{i=0}^{\infty}\sum_{j=i+\kappa}^{\infty}(p(s_{j+1}) - p(s_j)) + \sum_{i=0}^{\infty}\sum_{j=-\infty}^{\kappa-1-i}(p(s_j) - p(s_{j-1})) \\ &= \sum_{i=0}^{\infty} p(s_{i+\kappa}) + \sum_{i=0}^{\infty} p(s_{\kappa-1-i}) \\ &= \sum_{i\geq\kappa} p(s_i) + \sum_{i\leq\kappa-1} p(s_i) = 1 \end{aligned}$$

and hence q is a pmf. Finally, q satisfies

$$\begin{aligned} \sum_i (|i|+1)^{-1} q(s_{i+\kappa}) &= \sum_{i\geq\kappa} \frac{q(s_i)}{i-\kappa+1} + \sum_{i\leq\kappa-1} \frac{q(s_i)}{\kappa - i} \\ &= -\sum_{i\geq\kappa}(\Delta p)(i) + \sum_{i\leq\kappa-1}(\Delta p)(i-1) \\ &= p(s_\kappa) + p(s_{\kappa-1}) < \infty \end{aligned}$$

which completes the proof. □

C.3 Properties of the anti, iso, and uni operators

There is a well-known equivalence between the monotonic projection in the sense of least squares and likelihood maximization (e.g. the maximum likelihood and least squares estimators are the same for a decreasing density). As such equivalences are not always readily available in a standard reference on isotonic estimation, for completeness, we state this relationship explicitly in the following lemma. Let $\mathcal{I}_d^+ = \mathcal{I}_d \cap \{u \in \mathbb{R}^d : u_j > 0\}$ and $\mathcal{D}_d^+ = \mathcal{D}_d \cap \{u \in \mathbb{R}^d : u_j > 0\}$.

Lemma C.3. *Suppose that $v \in \mathbb{R}^d$ such that $v_j > 0$ for $j = 1, \dots, d$. Then*

$$\begin{aligned} \mathrm{iso}[v] &= \mathrm{argmax}_{u \in \mathcal{I}_d^+} \Big\{ \sum_{j=1}^d v_j \log(u_j) - \sum_{j=1}^d u_j \Big\}, \\ \mathrm{anti}[v] &= \mathrm{argmax}_{u \in \mathcal{D}_d^+} \Big\{ \sum_{j=1}^d v_j \log(u_j) - \sum_{j=1}^d u_j \Big\}. \end{aligned}$$

Proof. It is known that

$$\mathrm{argmin}_{u \in \mathcal{I}_d} \sum_{j=1}^d (v_j - u_j)^2$$

is equal to the right slope of the GCM of the cumulative sum diagram $\big\{(0,0), (j, \sum_{i=1}^j v_j), j = 1, \dots, d\big\}$. Note that implies in particular that these slopes have to be positive if $v_j > 0$ for all $j \in \{1, \dots, d\}$, and hence

$$\mathrm{argmin}_{u \in \mathcal{I}_d} \sum_{j=1}^d (v_j - u_j)^2 = \mathrm{argmin}_{u \in \mathcal{I}_d^+} \sum_{j=1}^d (v_j - u_j)^2.$$

Now maximizing the criterion $L(u) = \sum_{j=1}^d v_j \log(u_j) - \sum_{j=1}^d u_j$ on $\mathcal{I}_d^+$ admits a unique solution. Let $\{u^s\}_{s \in \mathbb{N}}$ be a maximizing sequence of L. Suppose that there exists $j \in \{1, \dots, d\}$ such that

$$\lim_{s \to \infty} u_j^s = 0 \quad \text{or} \quad \lim_{s \to \infty} u_j^s = \infty.$$

Then, in this case we would have $\lim_{s \to \infty} L(u^s) = -\infty$ contradicting the fact that $\{u^s\}_{s \in \mathbb{N}}$ is a maximizing sequence since it must satisfy $\lim_{s \to \infty} L(u^s) \geq L(v) = \sum_{j=1}^d v_j \log(v_j) - \sum_{j=1}^d v_j > -\infty$. Hence, there exists $K_2 > K_1 > 0$ such that $K_1 \leq u_j \leq K_2$ for $j = 1, \dots, d$. It follows that the maximization is performed on a compact set and existence of the maximum is now guaranteed by continuity of L. Uniqueness follows from strict concavity of L. We denote this unique solution by $\widehat{v}$. Let $j \in \{1, \dots, d\}$. For $\epsilon \in \mathbb{R}$, let

$$\widehat{v}_i^{\epsilon} = \widehat{v}_i + \epsilon \mathbb{I}_{1 \leq i \leq j}, \ 1 \leq i \leq d.$$

Then, for $\epsilon > 0$ small enough, we have $\widehat{v}^\epsilon \in \mathcal{I}_d^+$, and $L(\widehat{v}^\epsilon) \le L(\widehat{v})$. Therefore

$$\lim_{\epsilon \searrow 0} \epsilon^{-1} \left(L(\widehat{v}^\epsilon) - L(\widehat{v}) \right) \le 0.$$

When j is a knot point, that is, $\widehat{v}_{j+1} > \widehat{v}_j$ then it is easy to see that

$$\lim_{\epsilon \to 0} \epsilon^{-1} \left(L(\widehat{v}^\epsilon) - L(\widehat{v}) \right) = 0.$$

This yields

$$\sum_{i=1}^{j} \frac{v_i}{\widehat{v}_i} \begin{cases} \le j, \text{ for all } j \in \{1, \dots, d\} \\ = j, \text{ if } j \text{ is a knot point.} \end{cases} \tag{C-8}$$

Let $B_1, \dots, B_r$ denote a partition of $\{1, \dots, d\}$ such that $\forall l \in B_i$, $u_l = c_i$ some positive constant, for $i = 1, \dots, r$. Let $i_1, i_2, \dots, i_r$ denote the largest integers of $B_1, \dots, B_d$ respectively. Note that $i_r = d$. Then, if follows from (C-8) that

$$\sum_{i=1}^{j} v_i \begin{cases} \le j\widehat{v}_1 = \sum_{i=1}^{j} \widehat{v}_i, \text{ for all } j \in B_1 = \{1, \dots, i_1\} \\ = j\widehat{v}_1 = \sum_{i=1}^{j} \widehat{v}_i, \text{for } j = i_1. \end{cases}$$

The same reasoning can be applied for the other sets $B_i, 2 \le i \le r$ to conclude that

$$\sum_{i=1}^{j} \widehat{v}_i \begin{cases} \ge \sum_{i=1}^{j} v_i, \text{for all } j \in \{1, \dots, d\} \\ = \sum_{i=1}^{j} v_i, \text{if } j \text{ is a jump point.} \end{cases} \tag{C-9}$$

Hence, the solution $\widehat{v}$ is given by the slope of the LCM of the cumulative sum of v. The same reasoning can be applied to the projection on $\mathcal{D}_d^+$, proving the result. □

In the following, we state a result which shows that isotonic/antitonic projections can be transformed into "localized" projections between the knots of the "global" isotonic/antitonic solution. Recall that if $v = (v_1, \dots, v_d) \in \mathbb{R}^d$, then $v_{s:t} = (v_s, \dots, v_t)$ for $1 \le s \le t \le d$.

Lemma C.4. *Let $v = (v_1, \dots, v_d) \in \mathbb{R}^d$ such that $v_j > 0,\ j = 1, \dots, d$. Also let $\widehat{v} = \mathrm{iso}[v]$ and $1 \le s_1 < \ldots < s_r \le d$ the locations of the knot points of $\widehat{v}$, that is*

$$\widehat{v}_1 = \ldots = \widehat{v}_{s_1} < \widehat{v}_{s_1+1} = \ldots \widehat{v}_{s_2} < \ldots < \widehat{v}_{s_r+1} = \ldots = \widehat{v}_d.$$

Then, for $1 \le j < k \le r$

$$\widehat{v}_{(s_j+1):s_k} = \mathrm{iso}\big[v_{(s_j+1):s_k}\big].$$

Proof. The proof follows immediately from the fact that $\widehat{v}_{(s_j+1):s_k}$ is characterized by the same Fenchel conditions as $\mathrm{iso}(v_{(s_j+1):s_k})$. Indeed, we know from the characterization of $\widehat{v} = \mathrm{iso}(v)$ that

$$\sum_{i=1}^{t} \widehat{v}_i \begin{cases} \geq \sum_{i=1}^{t} v_i, \text{for all } t \in \{1, \dots, d\} \\ = \sum_{i=1}^{t} v_i, \text{if } t \text{ is a jump point} \end{cases}$$

and therefore

$$\sum_{i=s_j}^{t} \widehat{v}_i \begin{cases} \geq \sum_{i=s_j}^{t} v_i, \text{for all } t \in \{s_j+1, \dots, s_k\} \\ = \sum_{i=s_j}^{t} v_i, \text{if } t \text{ is a jump point} \end{cases}$$

which give exactly the characterization of the isotonic projection of the subvector $v_{(s_j+1):s_k}$.

□

Lemma C.5. *Suppose that* $v \in \mathbb{R}^d$ *and let* $p \in \mathcal{I}_d, q \in \mathcal{D}_d$. *Also, let* $a > 0, b \in \mathbb{R}$ *denote two fixed constants. Then the following (in)equalities hold*

$$\begin{aligned} \|\mathrm{iso}[v] - p\|_2^2 &\leq \|v - p\|_2^2, & \|\mathrm{anti}[v] - q\|_2^2 &\leq \|v - q\|_2^2 \\ \mathrm{anti}[av + b] &= a\,\mathrm{anti}[v] + b, & \mathrm{iso}[av + b] &= a\,\mathrm{iso}[v] + b, \\ \mathrm{uni}[av + b] &= a\,\mathrm{uni}[v] + b. \end{aligned}$$

Proof. The first two inequalities appear in Robertson et al. (1988, Theorem 1.6.1); cf. Jankowski and Wellner (2009, Lemma 6.1). The three equalities are all proved in a similar manner. For example,

$$\begin{aligned} \mathrm{anti}[v + b] &= \mathrm{argmin}_{u \in \mathcal{D}_d} \|u - (v + b)\|_2^2 \\ &= \mathrm{argmin}_{u \in \mathcal{D}_d} \|(u - b) - v\|_2^2 \\ &= \mathrm{argmin}_{u+b\ \in \mathcal{D}_d} \|u - v\|_2^2 \\ &= \mathrm{argmin}_{u \in \mathcal{D}_d} \|u - v\|_2^2 + b \quad = \quad \mathrm{anti}[v] + b. \end{aligned}$$

□

Continuity of the operators anti and iso follows immediately from Jankowski and Wellner (2009, Lemma 6.1).

Proposition C.6. *Suppose that* $v_n \in \mathbb{R}^d$ *and that* $\lim_{n\to\infty} v_n = v$. *Then*

$$\lim_{n\to\infty} \mathrm{iso}[v_n] = \mathrm{iso}[v], \quad \textit{and} \quad \lim_{n\to\infty} \mathrm{anti}[v_n] = \mathrm{anti}[v].$$

References

Balabdaoui, F., Jankowski, H., Rufibach, K. and Pavlides, M. (2013). Asymoptic distribution of the discrete log-concave mle and some applications. *JRSS, Series B* **75** 769–790.

Dudley, R. M. (1999). *Uniform central limit theorems*, vol. 63 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge.

Jankowski, H. K. and Wellner, J. A. (2009). Estimation of a discrete monotone distribution. *Electron. J. Stat.* **3** 1567–1605.

Patilea, V. (1997). *Convex models, NPMLE and misspecification*. Ph.D. thesis, Institute of Statistics, Univ. catholique de Louvain.

Patilea, V. (2001). Convex models, MLE and misspecification. *Ann. Statist.* **29** 94–123.

Robertson, T., Wright, F. and Dykstra, R. (1988). *Order Restricted Statistical Inference*. John Wiley & Sons Ltd.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer Series in Statistics, Springer-Verlag.