Statistica Sinica **26** (2016), 1061-1086
doi:http://dx.doi.org/10.5705/ss.202014.0088

MAXIMUM LIKELIHOOD ESTIMATION OF A UNIMODAL PROBABILITY MASS FUNCTION

Fadoua Balabdaoui and Hanna Jankowski

Université Paris Dauphine and York University

Abstract: We develop an estimation procedure for a discrete probability mass function (pmf) with unknown support. We derive its maximum likelihood estimator under the mild and natural shape-constraint of unimodality. Shape-constrained estimation is a powerful and robust technique that additionally provides smoothing of the empirical distribution yielding gains in efficiency. We show that our unimodal estimator is consistent when the model is specified, and that it converges to the best projection of the true pmf on the unimodal class under model misspecification. We derive the limiting distribution of the the estimator, and use this to build asymptotic confidence bands for the unknown pmf when the latter is unimodal. We illustrate our approach using time-to-onset data of the Ebola virus during the 1976 outbreak in the former republic of Zaire.

Key words and phrases: maximum likelihood estimation, probability mass function estimation, shape constrained estimation, unimodal.

1. Introduction

Discrete or discretized data show up in many practical instances, see Harlan et al. (2014); Chowell et al. (2013, 2009); Laskowski et al. (2011); Breman and Johnson (2014). If computing the empirical distribution requires no assumptions on the unknown law, gains in efficiency can be made by imposing additional constraints. One such constraint is unimodality, a natural and mild assumption in many statistical applications.

Nonparametric estimation of a unimodal density has been treated in many research papers. When the mode is known, the problem boils down to fitting the well-known Grenander estimator Grenander (1956). However, as noted by Birgé (1997), it is unrealistic in practice to assume that the location of the mode is known. The main consequence of not making such an assumption is that the maximum likelihood estimator (MLE) fails to exist. To address this, several estimators have been proposed Wegman (1968, 1969); Prakasa Rao (1969); Wegman (1970a,b); Reiss (1973, 1976) that feature additional constraints. More recent work appears in Birgé (1997), where the proposed estimator is chosen among all possible unimodal Grenander estimators as the one with cumulative distribution function closest to the empirical distribution. Durot et al. (2013) consider

the estimation of a discrete convex distribution, using the least squares criterion. Recently, Dümbgen and Rufibach (2009) and Cule, Samworth and Stewart (2010) proposed to use the maximum likelihood estimator of a log-concave density in lieu of the unimodal assumption, partially due to the inherent problems faced when estimating a unimodal density using maximum likelihood. Existence of the unimodal MLE when the data are discrete is guaranteed, even when the mode is unknown. While uniqueness is not always true, this problem is rather marginal, as a rule for selecting from among the *finite* options is immediate, making our estimator fully automatic and easy to compute. Furthermore, if the pmf is not unimodal, the MLE is still consistent, in the sense that it approaches the best unimodal pmf among a finite number of choices. Further details of this behavior are provided in Section 4.

In the recent work of Balabdaoui et al. (2013), the discrete MLE under the constraint of log-concavity was studied. One important consequence of this work is that we can evaluate the loss when data exhibit unimodality but at the same time log-concavity is not a valid assumption. The unimodal MLE seems to be a more natural estimator to consider when additional features of the true distribution besides unimodality are lacking or hard to obtain. On the other hand, one expects the log-concave MLE to be more efficient than the unimodal one in case log-concavity is a correct assumption. This is studied via simulations in Section 3. Although restricted to discrete distributions, our results may be interesting to those studying the continuous setting as well.

The manuscript is organized as follows. In Section 2, we provide the technical details required to define and compute the MLE of a discrete unimodal distribution. In our set-up, the support is assumed to be unknown, and is estimated empirically from the data. In Section 3, we consider the finite sample size behavior of our estimator via simulations. We compare here our estimator with the discrete log-concave MLE, but also assess the loss of efficiency when the support is unknown and must be estimated from the data. Sections 4 and 5 establish consistency and global asymptotic theory for the estimator. One of our key contributions is the application of these to develop global confidence bands for a unimodal pmf, see Section 6. We illustrate the estimator on a data set for the 1976 Ebola outbreak in Zaire; see Section 7. The data show a drastic difference in the time from infection to onset of symptoms depending on the type of infection: whether the individual was infected from person-to-person contact or from injection with an unsterilized needle. R (R Core Team (2014)) code for this analysis (along with all simulations) is available online at `www.math.yorku.ca/~hkj/Research/`. All proofs and additional details are left to the Appendices of the online supplementary material.

2. Maximum Likelihood Estimation

2.1. Discrete unimodal distributions

We consider estimation of a unimodal pmf of a discrete real-valued random variable. We denote the support of such a pmf as $S = \{s_i\}_{i \in K}$, where K is a subset of $\mathbb{Z}$. Without loss of generality, we take $s_i \in \mathbb{R}$ for all $i \in K$, and we assume that $s_i < s_{i+1}$.

We say that a pmf p is unimodal if there exists an integer m such that

$$\begin{aligned} p(s_i) &\geq p(s_{i+1}), \quad \text{for all } \; i \geq m, \quad \text{and} \\ p(s_{i-1}) &\leq p(s_i), \qquad \text{for all } \; i \leq m. \end{aligned} \tag{2.1}$$

The element s_m is thus a mode of the pmf p, but is not necessarily unique. In general, we can define the *modal region*, denoted here by $\mathcal{M}$, as

$$\mathcal{M} = \{s_\kappa \in S : \; p \text{ satisfies (2.1) at } m = \kappa\}. \tag{2.2}$$

$\mathcal{M}$ is necessarily a finite set and we have that $p(s) = p(s')$ for all $s, s' \in \mathcal{M}$. Next, let $\mathcal{U}^1(S)$ denote the space of unimodal pmfs with the same fixed support S. For the purpose of estimating such a p, it is most convenient to decompose the space of unimodal pmfs as

$$\mathcal{U}^1(S) = \bigcup_{\kappa \in K} \mathcal{U}^1|_\kappa(S), \tag{2.3}$$

where $\mathcal{U}^1|_\kappa(S)$ is the space of pmfs which are increasing on $\{s_i : i \leq \kappa - 1\}$ and decreasing on $\{s_i : i \geq \kappa\}$. Here a pmf in $\mathcal{U}^1|_\kappa(S)$ is unimodal either at $s_{\kappa-1}$ or s_κ depending on the order of its values at these points. It may seem, at first, that it would be more natural to decompose $\mathcal{U}^1(S)$ into the spaces of pmfs that are unimodal at κ. However, it turned out that the decomposition (2.3) is much more convenient. In addition, the MLE will always "decide" between these two possibilities by choosing the one that yields the largest value of the likelihood. If $\kappa = \min K$, then $\mathcal{U}^1|_\kappa(S)$ is simply the space of non-increasing pmfs on S. Notably, each space $\mathcal{U}^1|_\kappa(S)$ is convex, whereas $\mathcal{U}^1(S)$ is not.

Known as Khintchine's Theorem, a density with respect to Lebesgue measure is unimodal if and only if it can be written as a mixture of uniform densities, see for example Olshen and Savage (1970). Hence, it is expected that such a representation exists also in the discrete setting.

Proposition 1. *A pmf p satisfies $p \in \mathcal{U}^1|_\kappa(S)$ if and only if*

$$p(s_i) = \sum_{j \geq 0} \frac{1_{i \in \{\kappa, \ldots, \kappa + j\}}}{j+1} q(s_j) + \sum_{j \leq -1} \frac{1_{i \in \{\kappa + j, \ldots, \kappa - 1\}}}{|j|} q(s_j), \tag{2.4}$$

for some pmf q with support S.

A proof of Proposition 1 can be found in the Appendix. Using (2.3), a unimodal $p \in \mathcal{U}^1(S)$ admits a representation (2.4) for some $\kappa \in K$.

Remark 1. Suppose that K is finite, and write $K = \{0, 1, \ldots, k\}$. Then

$$\mathcal{U}^1|_0(S) \subset \mathcal{U}^1|_1(S).$$

2.1.1. Relationship with unimodal densities

Given a probability mass function p with support on $S = \mathbb{Z}$, one can define a density function f on $\mathbb{R}$ by

$$f(x) = p(z) \quad \text{for } x \in (z-1, z],\ z \in \mathbb{Z}.$$

The mass function p is unimodal iff the (piecewise constant) density f is unimodal.

Given a general unimodal density with support on $\mathbb{R}$, one can also define a unimodal pmf p via

$$p(z) = \int_{z-1}^{z} f(x)dx, \quad z \in \mathbb{Z}.$$

Here, the choice of the discretization on $[z-1, z)$ is arbitrary. Indeed, any choice of $a \in \mathbb{R}$, with $[z+a-1, z+a)$ is possible. One could also consider intervals of length other than one, as long as the length is fixed.

In this sense, discrete distributions provide a useful way to analyze data that have been "discretized" in such a manner. One such example is considered in Section 7. This relationship with unimodal densities is particularly noteworthy, since, although the MLE of a unimodal density does not exist, the MLE of its discretized version does.

2.1.2. When the true support is unknown

The discussion above relating unimodal densities and pmfs implies that one natural assumption on the support S is that it is a connected subset of $a+\delta\mathbb{Z}$, for some $a \in \mathbb{R}$ and $\delta > 0$. However, we believe that in certain instances additional generality may be required. For this reason, the only assumption we make about the support S is that it is an ordered subset of $\mathbb{R}$. This provides flexibility to our approach: unimodality of a pmf is preserved under scalar transformations and under removal of elements of the support.

We therefore do not assume that the true support is known a priori. Instead, we estimate both the unimodal pmf and its support from the collected observations. Were the true support known a priori, more efficiency would be gained by using this information in the estimation procedure. Some simulations studying this appear in Section 3.1.

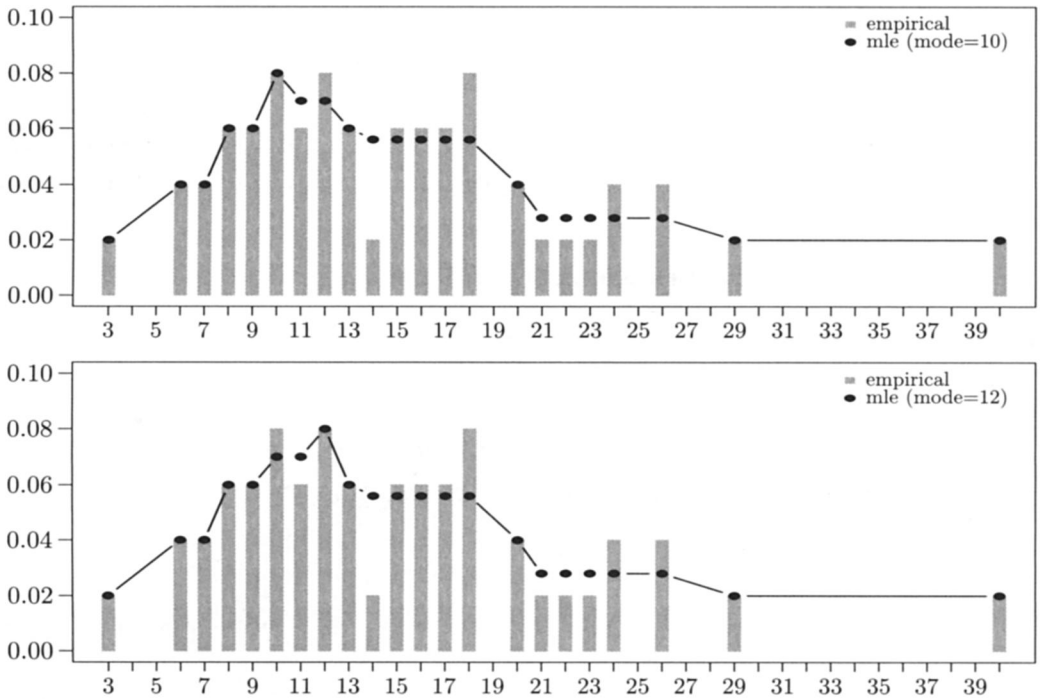

Figure 1. The same empirical observations (shown in grey) yield two different solutions maximizing the likelihood.

Our consistency and asymptotic results apply to both versions of the MLE; support known or unknown. All results are proved and stated for the unknown support version; they are only simplified when the support is known.

2.2. The unimodal maximum likelihood estimator

Let $X_1, \ldots, X_n$ be independent observations from a discrete pmf p_0. Let $\overline{p}_n(z) = n^{-1} \sum_{j=1}^{n} \mathbb{I}_{\{z\}}(X_i)$ and $\mathbb{F}_n(z) = n^{-1} \sum_{j=1}^{n} \mathbb{I}_{\{(-\infty,z]\}}(X_i)$ denote their empirical pmf and empirical cumulative distribution function (cdf), respectively. Let S_n denote the observed support of $\overline{p}_n$, $S_n = \{z_0, \ldots, z_{J-1}\}$ is the set of distinct values in the sample $\{X_1, \ldots, X_n\}$. We assume that $z_0 < z_1 < \ldots < z_{J-1}$.

2.2.1. Definition

As the support S is unknown we define the maximum likelihood estimator (MLE) as

$$\widehat{p}_n = \underset{p \in \mathcal{U}^1(S_n)}{\operatorname{argmax}} \; L_n(p),$$

where the log-likelihood is given by

$$L_n(p) = \int \log p(z)\, d\mathbb{F}_n(z) = \sum_{j=0}^{J-1} \log(p(z_j))\, \overline{p}_n(z_j).$$

This maximization is done in two steps: we maximize L_n over the space $\mathcal{U}^1|_\kappa(S_n)$ for each κ, and we find $\widehat{\kappa}_n$ and the corresponding estimator at which the overall maximum is attained.

2.2.2. The shape operators iso, anti, and uni

It is convenient to define several shape operators. For any $z \in \mathbb{R}^d$, we let $z_{s:t}$ be the sub-vector $(z_s, \ldots, z_t)$, $1 \le s \le t \le d$. Consider the sets of constrained vectors

$$\mathcal{I}_d = \{u = (u_1, \ldots, u_d) \in \mathbb{R}^d : u_1 \le \cdots \le u_d\},$$
$$\mathcal{D}_d = \{w = (w_1, \ldots, w_d) \in \mathbb{R}^d : w_1 \ge \cdots \ge w_d\},$$

and, for $\kappa \in \{1, \ldots, d\}$, let

$$\mathcal{U}_d|_\kappa = \{z = (z_1, \ldots, z_d) \in \mathbb{R}^d : z_{1:(\kappa-1)} \in \mathcal{I}_{\kappa-1} \text{ and } z_{\kappa:d} \in \mathcal{D}_{d-\kappa+1}\},$$

and

$$\mathcal{U}_d = \bigcup_{\kappa=1}^{d} \mathcal{U}_d|_\kappa.$$

We denote the ℓ_2 distance by $\|v - u\|_2^2 = \sum_{j=1}^d (v_j - u_j)^2$.

We define the operators $\text{iso} : \mathbb{R}^d \to \mathcal{I}_d$ and $\text{anti} : \mathbb{R}^d \to \mathcal{D}_d$ as

$$\text{iso}[v] = \operatorname*{argmin}_{u \in \mathcal{I}_d} \|v - u\|_2,$$
$$\text{anti}[v] = \operatorname*{argmin}_{w \in \mathcal{D}_d} \|w - u\|_2.$$

Here, $\text{iso}[v]$ and $\text{anti}[v] = -\text{iso}[-v]$ are the least squares projections of v on the spaces $\mathcal{I}_d$ and $\mathcal{D}_d$ respectively; cf., Barlow et al. (1972); Sen and Meyer (2013). The operator anti is the same as the gren operator discussed in Jankowski and Wellner (2009) and Jankowski (2014).

For $\kappa \in \{1, \ldots, d\}$, define the operators $\text{uni}_\kappa : \mathbb{R}^d \to \mathcal{U}_d|_\kappa$ and $\text{uni} : \mathbb{R}^d \to \mathcal{U}_d$ as

$$\text{uni}_\kappa[v] = (\text{iso}[v_{1:(\kappa-1)}], \text{anti}[v_{\kappa:d}]) = \operatorname*{argmin}_{u \in \mathcal{U}_d|_\kappa} \|v - u\|_2,$$
$$\text{uni}[v] = \operatorname*{argmin}_{u \in \mathcal{U}_d} \|v - u\|_2.$$

As before, we have

$$\text{uni}[v] = \text{uni}_{\kappa=\widetilde{\kappa}}[v], \quad \text{where } \widetilde{\kappa} \in \operatorname*{argmin}_\kappa \|v - \text{uni}_\kappa[v]\|_2.$$

The operators iso and anti are unique, but, the operator uni may yield more than one solution, much like the operator yielding the MLE. Properties of these operators are discussed in detail in Appendix C.3 of the supplementary material.

2.2.3. Existence and characterization of the MLE

Proposition 2. *The restricted MLE $\widehat{p}_n|_\kappa$ exists, is unique, and is characterized by*

$$\widehat{p}_n|_\kappa = \text{uni}_\kappa[\overline{p}_n].$$

The (unrestricted) unimodal MLE $\widehat{p}_n$ exists, but need not be unique. For $\{\widehat{\kappa}_n\} = \text{argmax}_{1\le\kappa\le J-1} L_n(\widehat{p}_n|_\kappa)$, the (finite) collection of solutions to the maximization problem, $\{\widehat{p}_n\}$, is characterized as

$$\{\widehat{p}_n\} = \{\widehat{p}_n|_\kappa; \kappa \in \{\widehat{\kappa}_n\}\}. \tag{2.5}$$

The MLE is not defined in terms of the operator uni though the operator does show up in its limiting distribution.

Remark 2. The size of the set $\{\widehat{p}_n\}$ may be greater than one (see, for example, Figure 1). If non-unique, we take the MLE to be the maximizer with the smallest mode. If $\widehat{\kappa}_n$ is the smallest integer κ such that

$$L_n(\widehat{p}_n|_\kappa) = \max_{1\le l\le J-1} L_n(\widehat{p}_n|_l),$$

then $\widehat{p}_n = \widehat{p}_n|_{\widehat{\kappa}_n}$. Note the slight abuse of notation: we denote $\widehat{\kappa}_n$ as the smallest element of $\{\widehat{\kappa}_n\}$. In order to find $\widehat{\kappa}_n$ we can search over $1 \le \kappa \le J-1$ using Remark 1.

To compute $\widehat{p}_n$. we first find the restricted MLE $\widehat{p}_n|_\kappa$ as the right slopes of the greatest convex minorant of $\{(0,0),(z_j,\mathbb{F}_n(z_j)),\ 0 \le j \le \kappa-1\}$ and the left slopes of the least concave majorant of $\{(0,0),(z_j,\mathbb{F}_n(z_j)),\ \kappa \le j \le J-1\}$. The MLE $\widehat{p}_n$ is then taken to be the $\widehat{p}_n|_\kappa$ that maximizes the overall likelihood for the smallest integer κ. Proposition 2 follows immediately from the more general result of Theorem 2 and Lemma C.3 in Appendix C of the online supplementary material.

3. Finite Sample Performance of the MLE

Here we compare three maximum likelihood estimators for small and medium samples sizes. They are as follows:

(1) the MLE under no assumption on the pmf; i.e., the empirical MLE,
(2) the MLE assuming the pmf is unimodal ($\widehat{p}_n$ as defined in this work),
(3) the log-concave MLE assuming the pmf is log-concave. Theoretical and computational aspects of this estimator have been studied in Balabdaoui et al. (2013).

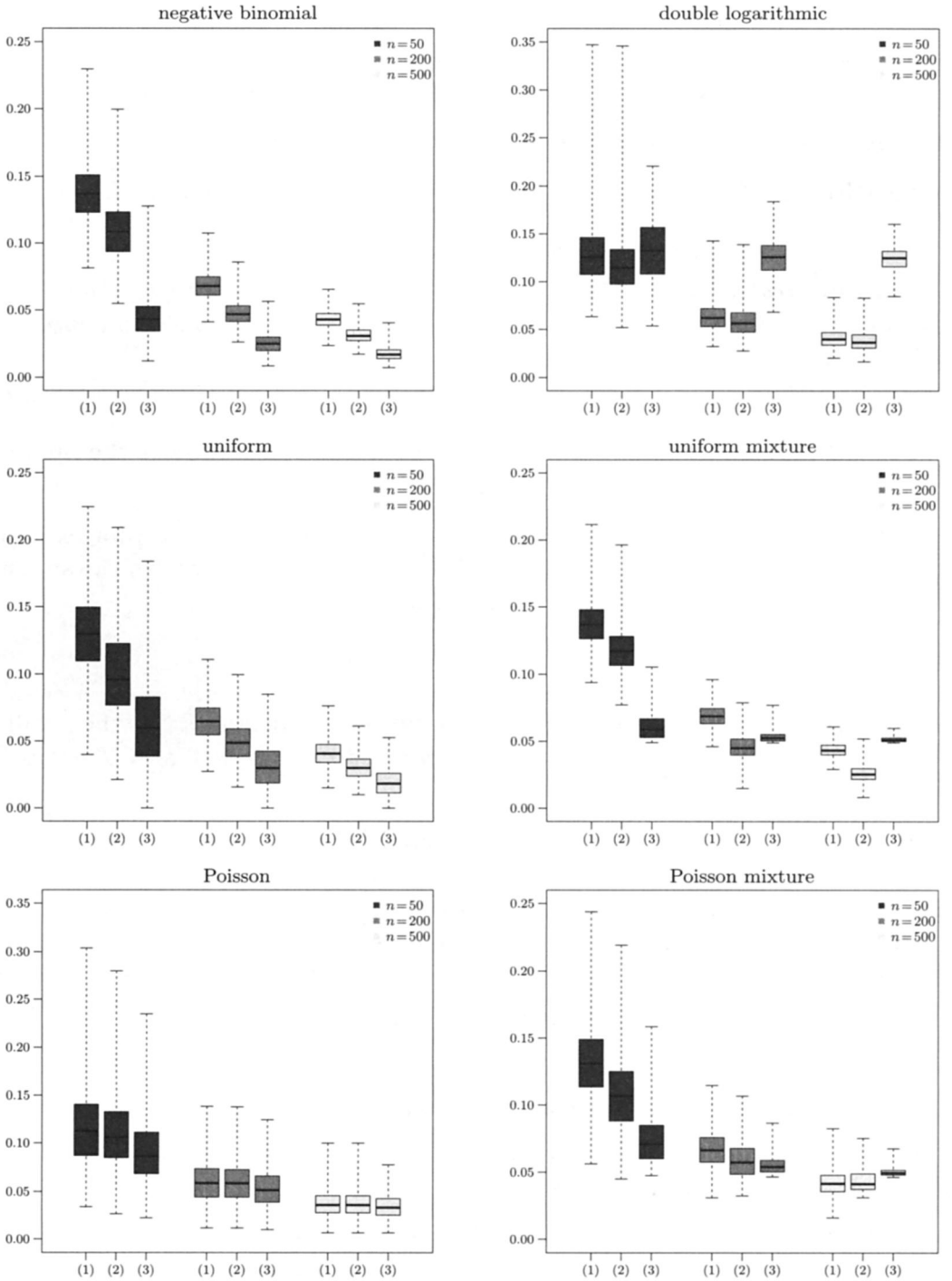

Figure 2. Boxplots of the ℓ_2 distance of the estimated pmf from the true pmf under each of three estimators: the empirical MLE (1), the unimodal MLE (2), the log-concave MLE (3). Each boxplot is the result of B =1,000 simulations. Properties of these distributions are summarized in Table 1.

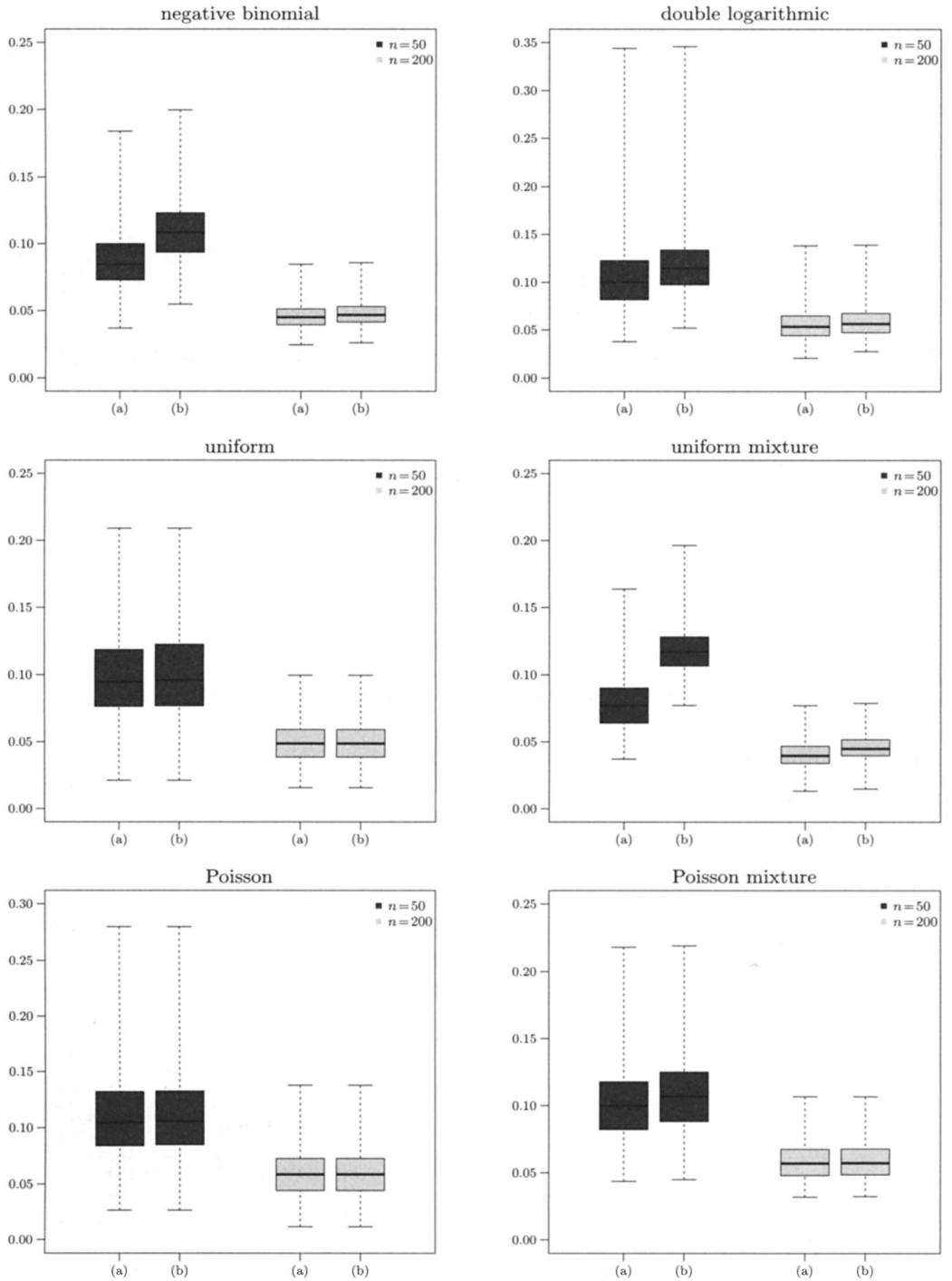

Figure 3. Boxplots for the six distributions of Figure 2 of the ℓ_2 distance of the estimated pmf from the true pmf for the unimodal MLE when the support is known (a) and unknown (b). Each boxplot is the result of B =1,000 simulations.

In our simulations, we considered six distributions; as follows:

- The negative binomial distribution with parameters $r = 6, p = 0.3$; it is both strictly unimodal and strictly log-concave.
- The double logarithmic distribution with $S = \mathbb{Z}$, and

$$p(z) = \begin{cases} \frac{p^{|z|}}{2|z|(p-\log(1-p))}, & z \leq -1, \\ \frac{p}{p-\log(1-p)}, & z = 0, \\ \frac{p^z}{2z(p-\log(1-p))}, & z \geq 1. \end{cases} \tag{3.1}$$

  This distribution is strictly unimodal but not log-concave. In the simulations, we took $p = 0.9$.
- The uniform pmf with $S = \{0, \ldots, 9\}$; it is neither strictly unimodal nor strictly log-concave.
- The mixture of uniform distributions with support on $\{0, \ldots, 49\}$, with pmf given by taking $S = \mathbb{Z}, \kappa = 0$ and

$$q(z) = \begin{cases} \frac{1}{3}, & z = 9, 39, 49, \\ 0, & \text{otherwise.} \end{cases} \tag{3.2}$$

  in decomposition (2.4). This distribution is unimodal, though not strictly, and is not log-concave.
- The Poisson with rate $\lambda = 2$, a strictly log-concave and unimodal distribution.
- A mixture of Poisson distributions: Let p_λ denote the Poisson pmf with rate λ, we consider ed the mixture $(1/4) \cdot p_1(\cdot) + (1/8) \cdot p_3(\cdot) + (5/8) \cdot p_8(\cdot)$. This distribution is (strictly) bimodal, and is therefore neither unimodal nor log-concave.

Properties of the six distributions are summarized in Table 1 for convenience.

In Figure 2, we can see that the unimodal MLE performs better than the empirical MLE for all six distributions. It is, however, outperformed by the log-concave MLE for the distributions that are log-concave, although they appear to

Table 1. Properties of distributions considered.

	unimodal	log-concave	finite support
negative binomial	yes (strict)	yes (strict)	no
double logarithmic	yes (strict)	no	no
uniform	yes	yes	yes
uniform mixture	yes	no	yes
Poisson	yes (strict)	yes (strict)	no
Poisson mixture	no	no	no

have comparable errors in the case of the Poisson distribution with rate $\lambda = 2$. The unimodal MLE outperforms the log-concave MLE when the distribution is not log-concave, at least for sample sizes that are "large enough". Our simulations show that this sample size is related to how far away the true pmf is from the set of log-concave distributions. In Figure 2, the ℓ_2 distance to the corresponding log-concave Kullback-Leibler projection (cf., Balabdaoui et al. (2013)) is approximatively 0.363 for the double logarithmic and 0.050 for the uniform mixture. Overall, we expect that when log-concavity fails to hold, the unimodal MLE is the better estimator for larger sample sizes. Moreover, this behavior also holds for smaller sample sizes for pmfs that are further from the log-concave class. The bimodal Poisson mixture model is the only example in which neither the log-concave nor unimodal classes are correct. Notably, although the empirical pmf is the only well-specified MLE in this case, it outperforms the other two estimators only for the largest sample size.

3.1. Comparison of known versus unknown support

Some efficiency may be lost by assuming that the support is unknown, and we briefly consider the question of "how much?" via simulations. When the support is known, the MLE is defined as

$$\underset{p \in \mathcal{U}^1(S)}{\operatorname{argmax}} \; L_n(p),$$

unlike in the definition of $\widehat{p}_n$, where S is replaced by its estimate S_n. In order to avoid existence issues, the class $\mathcal{U}^1(S)$ should be viewed as the set of probability mass functions p with support contained in S. Our simulations show that although some difference is seen for small sample sizes, the cost is not great, and the difference disappears with increased sample size. As mentioned previously, our consistency and asymptotic results developed later apply to both versions of the MLE.

Figure 4 gives an example of the two estimators, with known and unknown support, for a sample from the negative binomial distribution for $n = 50$. Both unimodal MLE approaches provide considerable "smoothing" to the empirical pmf. However, when the support is unknown, the MLE only places mass on S_n, whereas the MLE with known support places mass on the entire range $\{X_{(1)}, \ldots, X_{(n)}\}$. This is clearly seen in Figure 4. The potential loss of efficiency most likely occurs in the tails of the true distribution, and this is particularly true for distributions with a fatter tail.

We compared the two approaches via simulations, the results of which are shown in Figure 3. The distributions considered are those described on page 1070. The loss is small for the uniform distribution that has support on ten points, and

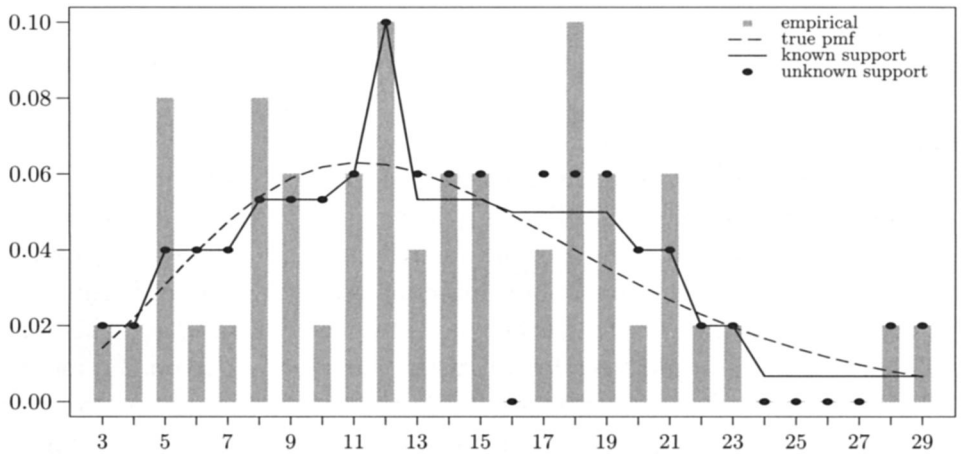

Figure 4. Example comparing the unimodal MLE when the support is known vs. unknown. The true distribution is the negative binomial with sample size $n = 50$.

also for both Poisson distributions, where the tails converge to zero quickly. For the other distributions, with slower rate of decay in the tails, some efficiency is lost for the small sample size ($n = 50$). However, the loss appears almost negligible for the medium sample size ($n = 200$).

Remark 3. When $S = S_n$, the known and unknown support MLEs are the same. When $|S|$ is finite, the probability that this does not happen for a given n decreases exponentially with n. Furthermore, with probability one, there exists an n_0, such that for all $n \geq n_0$, $S = S_n$.

4. The Kullback-Leibler Projection and Consistency of the Unimodal MLE

Let p_0 denote a fixed probability mass function on S_0 with distribution function P_0 and let

$$\rho(p|p_0) = \int \log \frac{p_0}{p}\, dP_0,$$

denote the Kullback-Leibler (KL) divergence. In this section, we seek the KL projection $\widehat{p}_0 \in \mathcal{U}^1(S_0)$ of a given pmf p_0. The KL projection has been considered extensively for the log-concave shape constraint for densities on $\mathbb{R}^d$ in Cule and Samworth (2010) and Dümbgen, Samworth, and Schuhmacher (2011), and for probability mass functions in Balabdaoui et al. (2013). As in Cule and Samworth (2010) and Cule, Samworth and Stewart (2010) and Balabdaoui et al. (2013), we take

$$\widehat{p}_0 = \operatorname*{argmin}_{p \in \mathcal{U}^1(S_0)} \int_{S_0} \log \frac{p_0}{p}\, dP_0 \;=\; \operatorname*{argmin}_{p \in \mathcal{U}^1(S_0)} \rho(p|p_0), \tag{4.1}$$

the element of $\mathcal{U}^1(S_0)$ closest to the unknown pmf p_0 in the sense of Kullback-Leibler divergence. From a practical point of view, this allows us to view the shape constrained estimator as the closest approximation within a class of distributions.

Alternatively, Patilea (2001) uses the definition

$$\int \log \frac{\widehat{p}_0}{p} dP_0 \geq 0, \qquad \text{for all } p \in \mathcal{U}^1(S_0), \tag{4.2}$$

and refers to the pmf $\widehat{p}_0$ satisfying (4.2) as the pseudo-true pmf. If the integrals involved are finite, one can re-arrange (4.2) into (4.1) and vice versa. In particular, if $\inf_{q \in \mathcal{U}^1(S_0)} \rho(q|p_0) = \rho(\widehat{p}_0|p_0) < \infty$ for some $\widehat{p}_0$ then (4.1) is equivalent to (4.2), since then

$$\begin{aligned} 0 \leq \int \log \frac{p_0}{\widehat{p}_0} dP_0 &\leq \int \log \frac{p_0}{p} dP_0 = \int \log \frac{p_0}{\widehat{p}_0} \frac{\widehat{p}_0}{p} dP_0 \\ &= \int \log \frac{p_0}{\widehat{p}_0} dP_0 + \int \log \frac{\widehat{p}_0}{p} dP_0. \end{aligned}$$

Alternatively, as in Dümbgen, Samworth, and Schuhmacher (2011), one could also consider

$$\int \log \widehat{p}_0 \, dP_0 \geq \int \log p \, dP_0, \qquad \text{for all } p \in \mathcal{U}^1(S_0), \tag{4.3}$$

which is akin to maximizing the likelihood. If p_0 admits a finite entropy, $\int \log p_0 \, dP_0 > -\infty$, then (4.3) is equivalent to (4.1). Furthermore, (4.2) is equivalent to (4.3) whenever $\sup_{p \in \mathcal{U}^1(S_0)} \int \log p dP_0 > -\infty$ and is attained.

In what follows, we work with the formulation of Patilea (2001) in (4.2), although we continue to refer to it as the KL projection. Recall that $\mathcal{U}^1|_\kappa(S_0)$ is the space of unimodal pmfs with support S_0 and mode at either $s_{\kappa-1}$ or s_κ.

Theorem 1. *Let p_0 be a discrete pmf with support S_0. Let $\widehat{P}_0|_\kappa$ denote the greatest convex majorant of the cumulative sum of $p_0(s_i), i \leq \kappa - 1$ and the least concave minorant of the cumulative sum of $p_0(s_i), i \geq \kappa$, and let $\widehat{p}_0|_\kappa$ denote the pmf corresponding to $\widehat{P}_0|_\kappa$. Then*

$$\int \log \frac{\widehat{p}_0|_\kappa}{p} dP_0 \geq 0, \qquad \textit{for all } p \in \mathcal{U}^1|_\kappa(S_0). \tag{4.4}$$

Furthermore, when $p_0 \in \mathcal{U}^1|_\kappa(S_0)$, or when $\sum_{j \neq 0} \log |j| p_0(s_j) < \infty$, $q = \widehat{p}_0|_\kappa$ is the unique pmf which satisfies $\int \log(q/p) dP_0 \geq 0$ for all $p \in \mathcal{U}^1|_\kappa(S_0)$.

Theorem 2. *Let p_0 be a discrete pmf with support S_0.*

1. *If* $p_0 \in \mathcal{U}^1(S_0)$, *then* $\widehat{p}_0 = p_0$ *is the unique unimodal pmf satisfying*

$$\int \log \frac{\widehat{p}_0}{p} dP_0 \geq 0 \qquad \text{for all } p \in \mathcal{U}^1(S_0).$$

2. *If* $p_0 \notin \mathcal{U}^1(S_0)$ *and* $\sum_{j\neq 0} \log|j| p_0(s_j) < \infty$, *then there exists a* $\widehat{p}_0 \in \mathcal{U}^1(S_0)$ *such that*

$$\int \log \frac{\widehat{p}_0}{p} dP_0 \geq 0 \qquad \text{for all } p \in \mathcal{U}^1(S_0).$$

When $\widehat{p}_0$ *is not unique, we denote by* $\{\widehat{p}_0\}$ *the (finite) collection of all such projections.*

Thus when the model is well-specified, the KL projection of p_0 is unique and is the true pmf itself under no additional assumptions. However, if the model is misspecified, there may exist several different KL projections; these make-up the (finite) set $\{\widehat{p}_0\}$. Examples of such non-uniqueness are given in Figure 5. We believe that this lack of uniqueness is due to the fact that the space of unimodal densities is not convex. Although the condition $\sum_{j\neq 0} \log|j| p_0(s_j) < \infty$ may seem a bit unnatural at first, one can express it in a more transparent form.

Proposition 3. *Let* p_0 *be a discrete pmf with support* S_0. *Then*

$$\sum_{j\neq 0} \log|j|\ p_0(s_j) < \infty \quad \text{if and only if} \quad \sup_{p\in\mathcal{U}^1(S_0)} \int \log p\ dP_0 \in (-\infty, 0].$$

Under this condition, (4.2) is equivalent to (4.3). If we assume in addition that

$$0 < \delta_1 \quad \leq \quad \inf(s_{j+1} - s_j) \leq \sup(s_{j+1} - s_j) \quad \leq \delta_2 < \infty,$$

then one can show that the condition $\sum_{j\neq 0} \log|j| p_0(s_j) < \infty$ is equivalent to $\int \log|x - a| dP_0(x) \in \mathbb{R}$ for some $a \notin S_0$. Therefore, this condition gives a bound on the speed of decay of p_0. Also, it is weaker than the assumption of having a finite mean required by Cule and Samworth (2010) and Dümbgen, Samworth, and Schuhmacher (2011). Our assumption is also weaker than that made by Patilea (2001, Corollary 5.6), although the latter is a condition for deriving rates of convergence. In our setting, Patilea's assumption boils down to existence of an $\epsilon > 0$ such that

$$\int \widehat{p_0}^{\epsilon} dP_0 < \infty,$$

where P_0 is the cumulative distribution function of p_0. Since $\log(x) \leq x^{\epsilon}/\epsilon$ for $x \in (0, \infty)$,

$$\int \log \frac{1}{\widehat{p}_0} dP_0 \leq \frac{1}{\epsilon} \int \widehat{p_0}^{\epsilon} dP_0 < \infty,$$

implying our condition in Proposition 3 (since then $\int \log \widehat{p}_0 dP_0 > -\infty$).

4.1. Consistency

For two pmfs p and q defined on S, the ℓ_k and Hellinger distances between p and q are respectively,

$$\ell_k(p,q) = \begin{cases} \left(\sum_{x\in\mathcal{S}} |p(x)-q(x)|^k\right)^{1/k}, & \text{if } 1 \le k < \infty, \\ \sup_{x\in\mathcal{S}} |p(x)-q(x)|, & \text{if } k = \infty, \end{cases}$$

$$h(p,q) = \frac{1}{2}\sum_{k\in\mathcal{S}} \left(\sqrt{p(x)} - \sqrt{q(x)}\right)^2.$$

In the following, we establish almost sure consistency of the unimodal MLE under a mild condition on the true pmf p_0. Fix a discrete pmf p_0 with support S_0, and assume that we observe i.i.d. data $X_1, \ldots, X_n \sim p_0$. Here, we do not necessarily assume that p_0 is itself unimodal. Let $\widehat{p}_n$ denote again the unimodal MLE based on the sample $(X_1, \ldots, X_n)$. In the well-specified model, the KL projection $\widehat{p}_0$ in the sense of (4.2) is p_0 itself; when the model is misspecified and p_0 satisfies $\sum_j \log|j| p_0(s_j) < \infty$, then the KL projection $\widehat{p}_0$ exists in the sense of (4.3) but may not be unique, with $\{\widehat{p}_0\}$ the set of all such KL projections.

Theorem 3. *Suppose that* $\sum_{i\neq 0} \log|i| p_0(s_i) < \infty$, *and let* $d \equiv \ell_k$ *or* h. *Then*

$$d(\widehat{p}_n, \{\widehat{p}_0\}) \equiv \inf_{\widehat{q}\in\{\widehat{p}_0\}} d(\widehat{p}_n, \widehat{q}) \to 0$$

almost surely. If p_0 *is unimodal, then* $d(\widehat{p}_n, p_0) \to 0$ *almost surely.*

Remark 4. Pointwise convergence and convergence in $\ell_k, 1 \le k \le \infty$ and Hellinger distance h are all equivalent for probability mass functions. This follows for example from Lemma C.2 in the online supporting material of Balabdaoui et al. (2013).

The fact that $\{\widehat{p}_0\}$ is not necessarily a singleton means that the MLE does not necessarily converge to a particular element of $\{\widehat{p}_0\}$. Rather, our proof shows instead that the MLE is sequentially compact: there exists an element $\widehat{q} \in \{\widehat{p}_0\}$ and a subsequence n_k such that $d(\widehat{p}_{n_k}, \widehat{q}) \to 0$. We illustrate this behaviour via an example. Let $S_0 = \{-2, -1, 0, 1, 2\}$ and take

$$p_0(s_i) = \begin{cases} \frac{1}{6}, & s_i = -2, 0, 2, \\ \frac{1}{4}, & s_i = -1, 1. \end{cases} \tag{4.5}$$

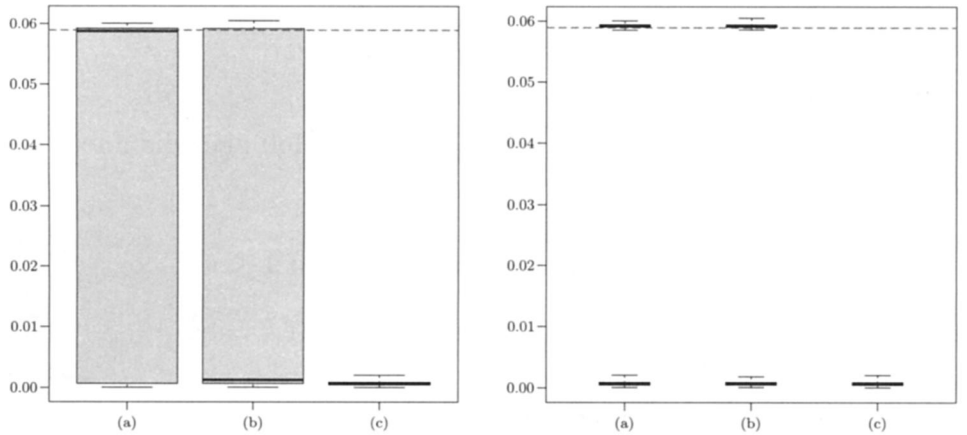

Figure 5. Convergence of $\widehat{p}_n$ to $\{\widehat{p}_0\}$ for p_0 as in (4.5). The boxplots show the $d = \ell_2$ distance for B =1,000 Monte Carlo samples with a sample size of n =1,000,000. The three columns give (a) $d(\widehat{p}_n, \widehat{p}_0^1)$, (b) $d(\widehat{p}_n, \widehat{p}_0^2)$, and (c) $d(\widehat{p}_n, \{\widehat{p}_0\})$. The plot on the right differs from the plot on the left in that, on the right, in (a) and (b) the boxplots have been split into large/small values to show the bimodal nature of the data. For reference, the dashed horizontal line gives $d(\widehat{p}_0^1, \widehat{p}_0^2)$.

Here $\{\widehat{p}_0\}$ has two elements, $\widehat{p}_0^1$ and $\widehat{p}_0^2$, say. Straightforward calculations show that

$$\widehat{p}_0^1(s_i) = \begin{cases} \frac{1}{6}, & s_i = -2, 2, \\ \frac{1}{4}, & s_i = -1, \\ \frac{5}{24}, & s_i = 0, 1, \end{cases}$$

with mode at -1 and $\widehat{p}_0^2(s_i) = \widehat{p}_0^1(-s_i)$ (with mode at 1). Simulations for a very large sample size are shown in Figure 5, where the convergence in set distance is clearly visible.

On the other hand, if $|\{\widehat{p}_0\}| = 1$, then the unimodal MLE converges to the unique element of $\{\widehat{p}_0\}$. For the restricted MLE $\widehat{p}_n|_\kappa$, a similar result holds. A proof follows by using, for example, Marshall's lemma as in Patilea (2001, Lemma 5.5, p.114), without any restrictions on p_0.

Let $\widehat{\mathcal{M}}_n$ be the modal region of the unimodal MLE $\widehat{p}_n$, cf., (2.2). An immediate corollary of the preceding theorem is the following statement about convergence of $\widehat{\mathcal{M}}_n$. For simplicity, we assume that the KL projection is unique. One can state the following with some additional generality, albeit in a less clear manner.

Corollary 1. *Assume* $\sum_{i \neq 0} \log |i| p_0(s_i) < \infty$ *and that* $|\{\widehat{p}_0\}| = 1$, *and let* $\mathcal{M}$ *denote the modal region of* p_0. *Then with probability one, there exists a sufficiently large integer* n_0 *such that for all* $n \geq n_0$, $\widehat{\mathcal{M}}_n \subset \mathcal{M}$.

Thus, if $|\mathcal{M}| = 1$, then, with probability one, there exists a sufficiently large n_0, such that the mode of the MLE coincides with the true mode. If $|\mathcal{M}| > 1$, this is no longer true, and all we can say is that eventually the estimated mode will be in $\mathcal{M}$.

Corollary 2. *Assume* $\sum_{i\neq 0} \log|i| p_0(s_i) < \infty$ *and that* $|\{\widehat{p}_0\}| = 1$. *Let* $\widehat{F}_n$ *and* $\widehat{F}_0$ *denote the cdfs of* $\widehat{p}_n$ *and* $\widehat{p}_0$ *respectively. Then*

$$\lim_{n\to\infty} \sup_{s\in S_0} |\widehat{F}_n(s) - \widehat{F}_0(s)| = 0$$

almost surely.

5. Global Asymptotics

The results for the asymptotic behaviour of the unimodal MLE share many similarities with those given in Jankowski and Wellner (2009) for the Grenander estimator of a decreasing pmf on $\mathbb{N}$. Our main interest here is to derive the weak limit of the estimator when p_0 is unimodal. We do not consider the misspecified setting. One could, however, mimic the work in Jankowski (2014) to obtain the asymptotic distributions in this case under some further restrictions on p_0. Despite the similarity mentioned earlier with the monotone problem, some technical details need special attention since the mode of the true pmf is unknown, and we do not assume that the true support is known.

With $\mathcal{M}$ the modal region of p_0 as defined in (2.2), write

$$D = \Big\{ s_i : s_i \notin \mathcal{M} \text{ and } p_0(s_i) \geq p_0(s_{i+1}) \Big\}, \text{ and}$$
$$I = \Big\{ s_i : s_i \notin \mathcal{M} \text{ and } p_0(s_{i-1}) \leq p_0(s_i) \Big\}$$

as the decreasing and increasing regions of S_0 respectively. We write $\mathcal{M} = \{\tau_0^I, \ldots, \tau_0^D\}$ (where $\tau_0^I \leq \tau_0^D$), and let $\{\tau_i^D\}_{i\geq 1}$ enumerate the points in D such that $p_0(s_i) > p_0(s_{i+1})$, where $\tau_i^D < \tau_{i+1}^D$. Similarly, let $\{\tau_i^I\}_{i\geq 1}$ enumerate the points in I such that $p_0(s_{i-1}) < p_0(s_i)$, where $\tau_{i+1}^I < \tau_i^I$. We write $D_j = \{s \in S_0, \tau_{j-1}^D < s \leq \tau_j^D\}$ for $j \geq 1$, and $I_j = \{s \in S_0, \tau_j^I \leq s < \tau_{j-1}^I\}$ for $j \geq 1$. Notice that each of these regions is necessarily finite, and that p_0 is constant on each subset I_j, D_j and $\mathcal{M}$. We therefore have that

$$I = \cup\, I_j,\ D = \cup\, D_j, \text{ and } \mathcal{S}_0 = I \cup \mathcal{M} \cup D. \tag{5.1}$$

We also denote the collection of knots as

$$\mathcal{T} = \{\tau_j^I, j \geq 1\} \cup \{\tau_0^D, \tau_0^I\} \cup \{\tau_j^D, j \geq 1\}. \tag{5.2}$$

Our definition of a knot, as well as the collection of knots, depends on the underlying pmf p_0. Let q be an element of $\ell_2(S_0)$ and, for a subset $C \subset S_0$, write the vector $q_C = \{q(s_j), s_j \in C\}$ to denote the sequence q restricted to C. Define φ :

$$\varphi[q](s) = \begin{cases} \text{iso}[q_{I_j}](s), & s \in I_j, \\ \text{uni}[q_{\mathcal{M}}](s), & s \in \mathcal{M}, \\ \text{anti}[q_{D_j}](s), & s \in D_j. \end{cases} \tag{5.3}$$

The definition of φ technically depends on p_0, although we omit this dependence in the notation. In addition, φ satisfies $\varphi[p_0] = p_0$.

Theorem 4. *Suppose that p_0 is unimodal and that $\sum_{i \neq 0} \log |i| p_0(s_i) < \infty$. Let $\mathbb{W}$ be the centered Gaussian process defined on S_0 such that $cov(\mathbb{W}(s_i), \mathbb{W}(s_j)) = p_0(s_i)\delta_{i,j} - p_0(s_i)p_0(s_j)$. Then $\sqrt{n}(\widehat{p}_n - p_0) \Rightarrow \varphi[\mathbb{W}]$, in $\ell_k(S_0)$, where $2 \leq k \leq \infty$.*

Thus if s is such that $s \in C$ where $C = I_j, \mathcal{M}$, or D_j and $|C| = 1$, then $\sqrt{n}(\widehat{p}_n(s) - p_0(s)) \Rightarrow \mathbb{W}(s)$, since in such cases $\varphi[q](s) = q(s)$; in regions where p_0 is strictly unimodal, the asymptotics of $\widehat{p}_n$ are the same as those of $\overline{p}_n$. Similar observations have been made in Jankowski and Wellner (2009) for the Grenander estimator and Balabdaoui et al. (2013) for the log-concave MLE. Here $\ell_2(S_0)$ is the smallest space, of those considered above, where one can prove the asymptotics; convergence in smaller spaces such as $\ell_1(S_0)$ cannot be considered without additional assumptions on p_0. We refer to Jankowski and Wellner (2009) for additional details.

The next result follows from the definition of φ, as well as Jankowski and Wellner (2009, Theorem 2.1).

Proposition 4. *For $2 \leq k \leq \infty$, we have that $\|\varphi[\mathbb{W}]\|_k \leq \|\mathbb{W}\|_k$.*

It is also possible to develop a Marshall's lemma type result in our setting. The (asymptotically negligible) error term not seen in the usual type of result here is due to estimation of the support in our approach.

Proposition 5 (Marshall's Lemma)**.** *Suppose that $\sum_{s \in S_0} p_0^{1/2}(s) < \infty$ and that the true pmf p_0 is unimodal with associated cumulative distribution function F_0. Then, with probability one, there exists an n_0 such that for all $n \geq n_0$*

$$\sup_{s \in S_0} |\widehat{F}_n(s) - F_0(s)| \leq \sup_{s \in S_0} |\mathbb{F}_n(s) - F_0(s)| + o_p(n^{-1/2}).$$

6. Global Confidence Bands for p_0

We consider the calculation of confidence bands for the true pmf p_0, which we assume to be unimodal. To this end, let $q_{0,\alpha}$ be such that $P(\|\mathbb{W}\|_\infty > q_{0,\alpha}) = \alpha$.

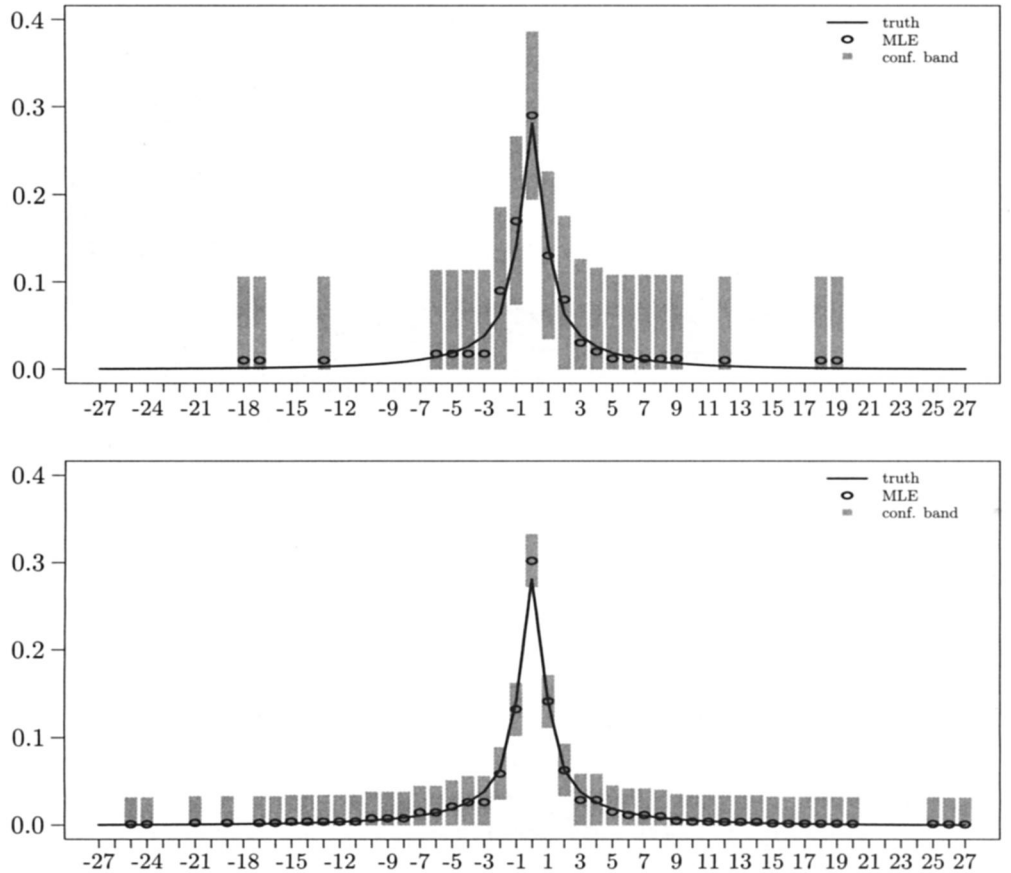

Figure 6. 95% constant-width ($\beta = 0$) confidence bands for the true pmf when sampling from the double logarithmic distribution $p = 0.9$. The sample size is $n = 100$ on the top and n =1,000 on the bottom.

Then, it follows that

$$\lim_n P\left(\sqrt{n}\|\widehat{p}_n - p_0\|_\infty \le q_{0,\alpha}\right) \ge 1 - \alpha,$$

since $\sqrt{n}\|\widehat{p}_n - p_0\|_\infty \Rightarrow \|\varphi[\mathbb{W}]\|_\infty \le \|\mathbb{W}\|_\infty$. If p_0 is strictly monotone then $\varphi[\mathbb{W}] = \mathbb{W}$, then the last inequality becomes an equality, resulting in an asymptotically *exact* confidence band.

To estimate $q_{0,\alpha}$, we use $\widehat{p}_n$ in place of p_0. In Proposition B.7 of the online supplementary material, we show that this yields an almost surely consistent method of estimating $q_{0,\alpha}$. We estimate each quantile using Monte Carlo simulations. Thus, if $\widehat{q}_{0,\alpha}$ denotes a Monte Carlo estimate of the quantile of $\|\mathbb{W}\|_\infty$, an asymptotically correct conservative confidence band is given by

$$\left\{\left[\left(\widehat{p}_n(s_i) - \frac{\widehat{q}_{0,\alpha}}{\sqrt{n}}\right) \vee 0,\ \widehat{p}_n(s_i) + \frac{\widehat{q}_{0,\alpha}}{\sqrt{n}}\right], s_i \in \operatorname{supp}(\widehat{p}_n)\right\}, \tag{6.1}$$

where $\text{supp}(\widehat{p}_n)$ denotes the support of $\widehat{p}_n$. When the support of p_0 is estimated from the data, $\text{supp}(\widehat{p}_n) = S_n$, the support of the empirical distribution.

In Figure 6, we show an example of confidence bands thus constructed, when the true pmf is the double logarithmic distribution with $p = 0.9$. We found the constant width of the confidence bands, particularly for the smaller sample size, somewhat visually jarring. For this reason, we created confidence bands which are visually more appealing in that they do not have uniform width. Define, for $\beta \geq 0$,

$$\widehat{\mathbb{W}}_n^{\beta}(s) = \begin{cases} \frac{\sqrt{n}(\widehat{p}_n - p_0)(s)}{\widehat{p}_n^{\beta}(s)}, & s \in \text{supp}(\widehat{p}_n), \\ 0, & s \notin \text{supp}(\widehat{p}_n). \end{cases}$$

If $\beta = 0$, then $\widehat{\mathbb{W}}_n^{\beta} = \sqrt{n}(\widehat{p}_n - p_0)$, and we are in the situation of constant-width confidence bands.

Proposition 6. *Fix $\beta > 0$ and assume that the support of p_0 is finite. Then*

$$\left|\left|\widehat{\mathbb{W}}_n^{\beta}\right|\right|_{\infty} \Rightarrow \left|\left|\frac{\varphi[\mathbb{W}]}{p_0^{\beta}}\right|\right|_{\infty} \leq \left|\left|\frac{\mathbb{W}}{p_0^{\beta}}\right|\right|_{\infty}.$$

In this case, an asymptotically correct conservative confidence band is given by

$$\left\{\left[\left(\widehat{p}_n(s) - \widehat{p}_n^{\beta}(s)\frac{\widehat{q}_{\beta,\alpha}}{\sqrt{n}}\right) \vee 0,\ \widehat{p}_n(s) + \widehat{p}_n^{\beta}(s)\frac{\widehat{q}_{\beta,\alpha}}{\sqrt{n}}\right], s \in \text{supp}(\widehat{p}_n)\right\},$$

where $\widehat{q}_{\beta,\alpha}$ is an estimate of $q_{\beta,\alpha}$ where $P\left(\left\|p_0^{-\beta}\mathbb{W}\right\|_{\infty} > q_{\beta,\alpha}\right) = \alpha$. Estimation of this quantile can be done using a Monte Carlo approach, as before.

Remark 5. When p_0 has an infinite support, the limiting distribution $p_0^{-\beta}\mathbb{W}$ exists in ℓ_2 provided $\sum_i p_0^{1-2\beta} < \infty$, which adds the restriction that $\beta < 1/2$. We conjecture that this continues to hold for distributions with infinite support with the restriction that $\beta \in [0, 1/2)$, although we do not pursue the proof here. The assumption of finite support may be highly plausible in certain practical situations, whereas the (weaker) assumption of $\sum_i p_0^{1-2\beta} < \infty$, is not as easy to motivate.

In Figure 7 we compare the constant-width confidence bands to the varying width confidence bands (with $\beta = 0.5$) when the true distribution is the mixture of uniforms, whose mixing distribution is given in (3.2). Visually, we find the choice of $\beta = 0.5$ preferable in that the values, where $\widehat{p}_n$ is smaller, express slightly more accuracy, as one would expect. In this particular example, the difference is not great, but is still eye-pleasing. For $\beta = 0$, the width of the

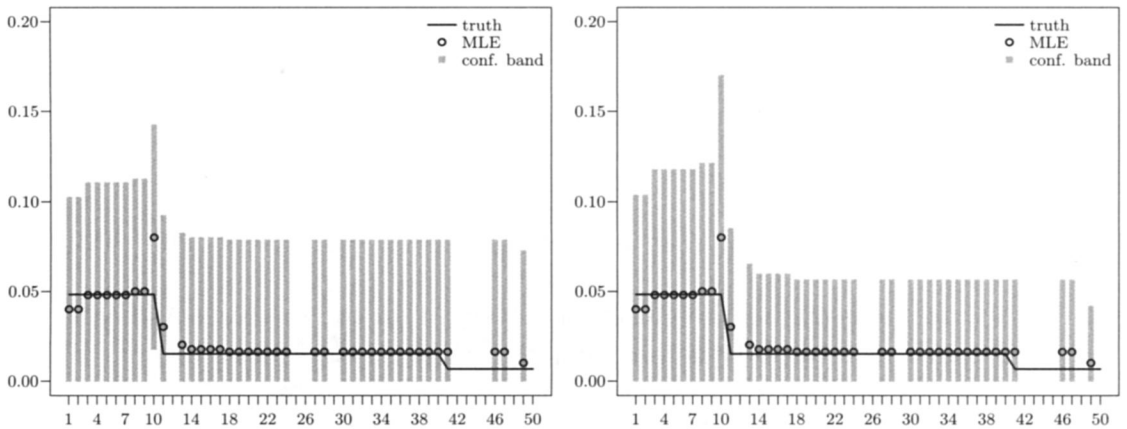

Figure 7. 95% confidence bands for the true pmf when sampling from the mixture of uniforms distribution with mixing distribution given in (3.2). The sample size is $n = 100$ and we chose $\beta = 0$ (left, constant width) and $\beta = 0.5$ (right, varying width).

Table 2. Empirical coverage probabilities for the proposed confidence bands with $\alpha = 0.05$.

	β	$n = 100$	n =1,000	n =5,000
mixture of uniforms	0	0.972	0.963	0.959
	0.25	0.991	0.971	0.970
	0.5	0.959	0.953	0.991
double logarithmic	0	0.956	0.949	0.949
	0.25	0.970	0.950	0.948
	0.5	0.980	0.989	0.989

confidence bands varies from 0.13 to 0.07 (median 0.08), while for $\beta = 0.5$, the width of the confidence bands varies from 0.17 to 0.04 (median 0.06). Note that, although for $\beta = 0$ the confidence bands have a constant width, we have to cut off the lower bound at a maximum value of zero, and hence the bands end up being non-constant in reality. Without this cutoff, the width would be constant at 0.13.

In Table 2, we examine the empirical performance of the proposed confidence bands. We considered two different unimodal distributions: the mixture of uniforms as above, and the double logarithmic with $p = 0.9$ from (3.1). Our simulations spanned various samples sizes and values of β. Note that when $\beta = 0.5$ and the true pmf is double logarithmic, the conditions for convergence are violated (see Proposition 6 and Remark 5), and we include this example for comparison only (seeing as the condition that $\sum_i p_0^{1-2\beta} < \infty$ may be difficult to verify without additional information about p_0).

For $\beta \geq 0$ take

$$\widehat{c}_{n,u}(s) = \widehat{p}_n(s) + \widehat{p}_n^{\beta}(s)\frac{\widehat{q}_{\beta,\alpha}}{\sqrt{n}}, \ s \in \text{supp}(\widehat{p}_n),$$
$$\widehat{c}_{n,l}(s) = 0 \vee \left(\widehat{p}_n(s) - \widehat{p}_n^{\beta}(s)\frac{\widehat{q}_{\beta,\alpha}}{\sqrt{n}}\right), \ s \in \text{supp}(\widehat{p}_n),$$
$$\widehat{c}_{n,u}(s) = \widehat{c}_{n,l}(s) \ = \ 0, \ s \notin \text{supp}(\widehat{p}_n).$$

The results in Table 2 give the empirical coverage *on the set* S_n as indicated in the third column, we report the proportion of times that

$$\widehat{c}_{n,l}(s) \quad \leq p_0(s) \leq \quad \widehat{c}_{n,l}(s), \quad \text{for all } s \in S_n \tag{6.2}$$

was observed.

Overall, we find that the confidence bands perform rather well. For the double logarithmic case $\beta < 0.5$ we expect to obtain asymptotically correct bands, whereas in both uniform mixture scenarios, we expect an asymptotically conservative result. In Appendix A of the online supplementary material, we provide some additional results where we study the cost of defining the bands on $\text{supp}(\widehat{p}_n) = S_n$ in the simulations.

7. Time-to-onset of the Ebola Virus

Breman and Johnson (2014) describe their experiences during the 1976 Ebola virus outbreak in Zaire (currently, the Democratic Republic of the Congo). They show histograms of the time of onset of the disease based on the transmission route: patients became infected either with an unsterilized needle or through person-to-person contact. This data was also published in Breman et al. (1978). Here, we use the histograms in Breman and Johnson (2014) to transcribe the data and perform a brief analysis. Transcribing the histograms resulted in samples sizes of $n = 57$ and $n = 108$, which differs slightly from those presented in Breman and Johnson (2014).

Figure 8 shows the empirical observations and the fitted unimodal MLEs for the two types of transmission routes. The 95% asymptotic global confidence bands are also included, where we used the version with constant width. The time-to-onset is measured in days, and we assume that the support of the true pmf is the natural numbers, or is a connected subset thereof. Thus, we use the version of the likelihood maximization where the support is not estimated from the data. Visually, there is no glaring reason that the assumption of unimodality is not appropriate in these two cases. On the other hand, the fitted MLE provides a slight smoothing to the empirical distribution, which is appealing.

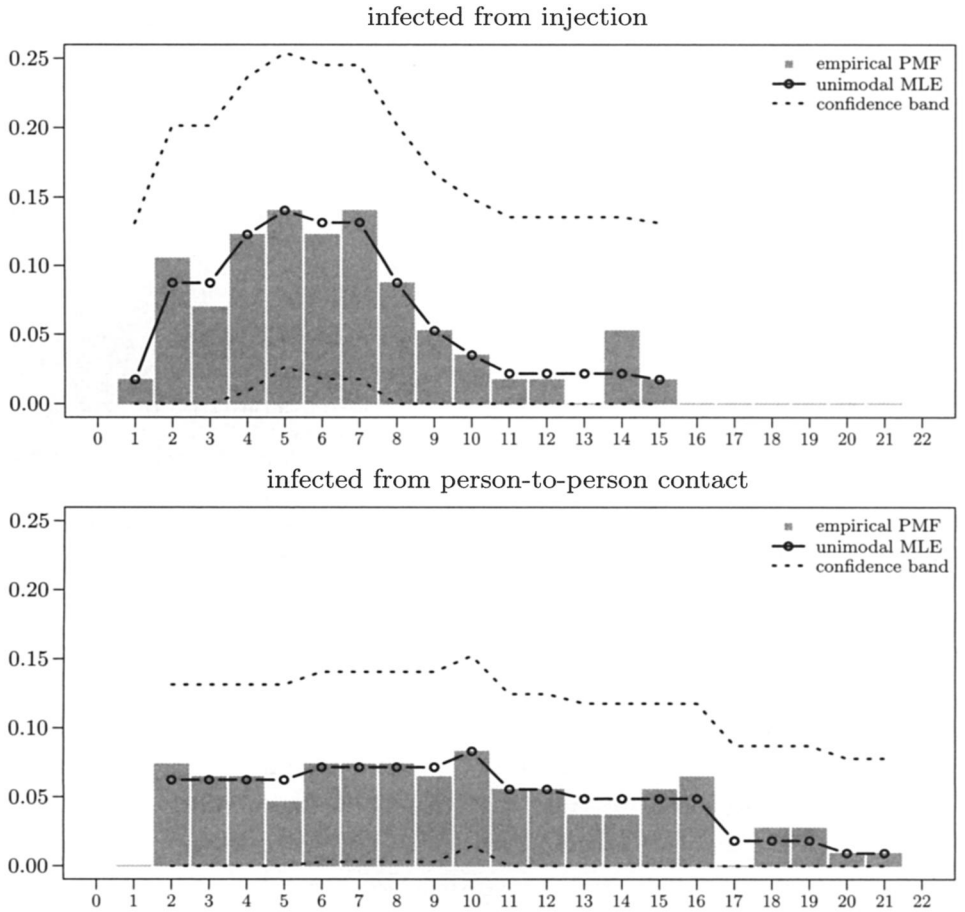

Figure 8. Time to onset of symptoms of the Ebola virus based on transmission type. The sample size is $n = 57$ for those infected from unsterilized needles and $n = 108$ for person to person contact.

The confidence bands in Figure 8 appear somewhat wide due to the small sample sizes observed in both distributions. The average width for the injection infection was found to be 0.18, and 0.12 for infection from person-to-person contact. As a crude benchmark, the average widths of 95% *pointwise* confidence intervals were calculated for the true pmf $\widehat{p}_n \pm 1.96\ \sqrt{\widehat{p}_n(1-\widehat{p}_n)}$, based on Theorem 4 and under the (untested) assumption that the true pmf is strictly unimodal. Here, the average width for the injection infection was found to be 0.12, and 0.08 for infection from person-to-person contact. These are also wide, but less so than the global confidence bands, as expected.

It is interesting how different the two distributions appear to be. The standard Kolmogorov-Smirnov test does not yield exact p-values in this setting because the data is discretized, and hence we used a permutation test; see Jöckel

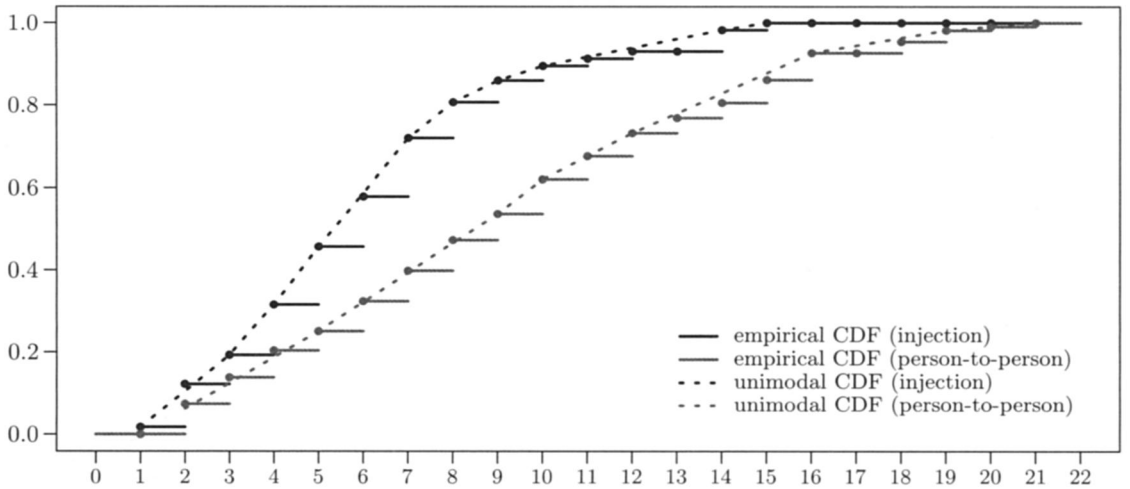

Figure 9. Time to onset of symptoms of the Ebola virus based on transmission type: a comparison of cumulative distribution functions.

(1986). This modified approach yielded a p-value of 0.0014 for the hypothesis that the two distributions are the same. This is in line with what we observe in Figures 8 and 9. R (R Core Team (2014)) code for performing this analysis is available online at `www.math.yorku.ca/~hkj/Research`.

A biological explanation for the difference between the two distributions was provided to us by Jane Heffernan (2014, private communication): "Injection gets the pathogen into the blood stream. Person-to-person contact provides exposure to the mucosa (innate immunity) first, so the pathogens that ultimately make it to the blood will be different in fitness distribution than the injection method. Also, the amount of pathogen ultimately making it to the blood could be smaller compared to the injection method. Both of these variables will affect the incubation period." In the data, we see this difference not only through a mean comparison (the mean time-to-onset is 6.3 days for transmission via injection and 9.4 days for person-to-person infections) but also in the stochastic dominance observed via the fitted and empirical CDFs in Figure 9. The latter suggest that $T_{inj} \leq T_{ptp}$ stochastically, where T_{inj} and T_{ptp} denote the times to onset for injection and person-to-person infections, respectively. Repeating the permutation test for the hypothesis that the two distributions are equal against the alternative that $F_{inj} > F_{ptp}$ yields a p-value of 0.0008.

Supplementary Material

In this supplement we present some additional proofs, and a further discussion of our assumptions.

Acknowledgements

The authors thank Hélène Massam and Jon Wellner for their help with the proof of Lemma B.5 and Lemma B.6 of the online supplementary material. We also thank the anonymous referees for useful comments that helped improve the paper.

References

Balabdaoui, F., Jankowski, H., Rufibach, K. and Pavlides, M. (2013). Asymoptic distribution of the discrete log-concave MLE and some applications. *J. Roy. Statist. Soc. Ser. B* **75**, 769-790.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). *Statistical Inference under Order Restrictions. The Theory and Application of Isotonic Regression.* John Wiley, London-New York-Sydney.

Birgé, L. (1997). Estimation of unimodal densities without smoothness assumptions. *Ann. Statist.* **25**, 970-981.

Breman, J. G. and Johnson, K. M. (2014). Ebola then and now. *New England J. Medicine* **371**, 1663-1666.

Breman, J. G., Piot, P., Johnson, K. M., White, M. W., Mbuyi, M., Sureau, P., Heymann, D. L., van Nieuwenhove, S., McCormick, J. B., Ruppol, J. P., Kintoki, V., Isaacson, M., van der Groen, G., Webb, P. A. and Ngvete, K. (1978). The epidemiology of ebola hemorrhagic fever in zaire, 1976. In *Pattyn SR, ed. Ebola virus hemorrhagic fever*, 85-97, Elsevier, Amsterdam.

Chowell, G., Bertozzi, S. M., Colchero, M. A., Lopez-Gatell, H., Alpuche-Aranda, C., Hernandez, M. and Miller, M. A. (2009). Severe respiratory disease concurrent with the circulation of H1N1 influenza. *New England J. Medicine* **361**, 674-679.

Chowell, G., Fuentes, R., Olea, A., Aguilera, X., Nesse, H. and Hyman, J. (2013). The basic reproduction number R_0 and effectiveness of reactive interventions during dengue epidemics: The 2002 dengue outbreak in Easter Island, Chile. *Math. Biosci. Eng.* **10**, 1455-1474.

Cule, M. and Samworth, R. (2010). Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. *Electronic J. Stat.* **4**, 254-270.

Cule, M., Samworth, R. and Stewart, M. (2010). Maximum likelihood estimation of a multidimensional log-concave density. *J. Roy. Statist. Soc. Ser. B* **72**, 545-607.

Dümbgen, L. and Rufibach, K. (2009). Maximum likelihood estimation of a log-concave density and its distribution function. *Bernoulli* **15**, 40-68.

Dümbgen, L., Samworth, R. and Schuhmacher, D. (2011). Approximation by log-concave distributions with applications to regression. *Ann. Statist.* **39**, 702-730.

Durot, C., Huet, S., Koladjo, F. and Robin, S. (2013). Least-squares estimation of a convex discrete distribution. *Comput. Statist. Data Anal.* **67**, 282-298.

Grenander, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.* **39**, 125-153.

Harlan, S., Chowell, G., Yang, S., Petitti, D., Morales Butler, E., Ruddell, B. and Ruddell, D. M. (2014). Heat-related deaths in hot cities: Estimates of human tolerance to high temperature thresholds. *Int. J. Environ. Res. Public Health* **11**, 3304-3326.

Jankowski, H. (2014). Convergence of linear functionals of the Grenander estimator under misspecification. *Ann. Statist.* **42**, 625-653.

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

Jöckel, K.-H. (1986). Finite sample properties and asymptotic efficiency of Monte Carlo tests. *Ann. Statist.* **14**, 336-347.

Laskowski, M., Mostaco-Guidolin, L., Greer, A., Wu, J. and Moghadas, S. (2011). The impact of demographic variables on disease spread: influenza in remote communities. *Scientific Reports* (*Nature*) **1**, 1-7.

Olshen, R. A. and Savage, L. J. (1970). A generalized unimodality. *J. Appl. Probab.* **7**, 21-34.

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

Prakasa Rao, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31**, 23-36.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org`

Reiss, R.-D. (1973). On the measurability and consistency of maximum likelihood estimates for unimodal densities. *Ann. Statist.* **1**, 888-901.

Reiss, R.-D. (1976). On minimum distance estimators for unimodal densities. *Metrika* **23**, 7-14.

Sen, B. and Meyer, M. C. (2013). Testing against a linear regression model using ideas from shape-restricted estimation. Preprint. ArXiv:1311.6849.

Wegman, E. J. (1968). *On Estimating a Unimodal Density*. Ph.D. Thesis, The University of Iowa.

Wegman, E. J. (1969). A note on estimating a unimodal density. *Ann. Math. Statist.* **40**, 1661-1667.

Wegman, E. J. (1970a). Maximum likelihood estimation of a unimodal density function. *Ann. Math. Statist.* **41**, 457-471.

Wegman, E. J. (1970b). Maximum likelihood estimation of a unimodal density. II. *Ann. Math. Statist.* **41**, 2169-2174.

Université Paris Dauphine, CEREMADE, Room B-617 Place du Marechal de Lattre de Tassigny 75775 PARIS CEDEX 16, France.

ETH, Seminar fuer Statistk, Room HG G12 Raemistrasse 101 8092 Zurich, Switzerland.

E-mail: fadoua@ceremade.dauphine.fr

Department of Mathematics and Statistics, York University, 4700 Keele Street Toronto, ON M3J 1P3, Canada.

E-mail: hkj@mathstat.yorku.ca

(Received October 2014; accepted September 2015)