NONPARAMETRIC MAXIMUM LIKELIHOOD ESTIMATION UNDER A LIKELIHOOD RATIO ORDER

Ted Westling, Kevin J. Downes, and Dylan S. Small

University of Massachusetts Amherst, Children's Hospital of Philadelphia, University of Pennsylvania

Supplementary Material

This supplementary material contains an example of the use of Theorem 2 (Section S1), proofs of all theorems (Section S2), additional simulation results (Sections S3, S4, and S5), and additional data analysis results (Section S6).

S1 Example of the use of Theorem 2

We first illustrate the use of Theorem 2 and Corollary 1 using hypothetical data. Suppose that $(Y_1, \ldots, Y_6) = (0, 0, 1, 3, 3, 6)$ and $(X_1, \ldots, X_4) = (-1, 2, 3, 3)$. We first derive F_n^*. The points $\{(H_n(y_k), F_n(y_k)) : k = 0, \ldots, m_2\}$ are given by $\{(0, 0)$, $(0.3, 0.25)$, $(0.4, 0.25)$, $(0.9, 1)$, $(1, 1)\}$, and its GCM is given by $\{(0, 0)$, $(0.3, 3/16)$, $(0.4, 1/4)$, $(0.9, 7/8)$, $(1, 1)\}$. This is displayed in the upper left panel of Figure 1. The values of the GCM imply that $F_n^*(0) = 3/16$, $F_n^*(1) = 1/4$, $F_n^*(3) = 7/8$, and $F_n^*(6) = 1$. We then have that $F_n^*(-1) = F_n^*(-\infty) + [F_n^*(0) - F_n^*(-\infty)]\frac{F_n(-1)-F_n(-1-)}{F_n(0)-F_n(-\infty)} =$

$[3/16]\frac{1/4}{1/4} = 3/16$ and $F_n^*(2) = F_n^*(1) + [F_n^*(3) - F_n^*(1)]\frac{F_n(2) - F_n(2-)}{F_n(3) - F_n(1)} = 1/4 + [5/8]\frac{1/4}{3/4} = 11/24$. The estimators F_n and F_n^* are compared in the bottom left panel of Figure 1.

We next derive G_n^*. The points $\{(H_n(y_k), G_n(y_k)) : k = 0, \ldots, m_2\}$ are given by $\{(0,0)$, $(0.3, 1/3)$, $(0.4, 1/2)$, $(0.9, 5/6)$, $(1,1)\}$, and its LCM is given by $\{(0,0)$, $(0.3, 3/8)$, $(0.4, 1/2)$, $(0.9, 11/12)$, $(1,1)\}$. This is displayed in the center left panel of Figure 1. The values of the LCM imply that $G_n^*(0) = 3/8$, $G_n^*(1) = 1/2$, $G_n^*(3) = 11/12$, and $G_n^*(6) = 1$. The estimators G_n and G_n^* are compared in the bottom left panel of Figure 1.

Finally, we derive θ_n^*. The empirical ordinal dominance curve is given by the points $\{(0,0)$, $(1/3, 1/4)$, $(1/2, 1/4)$, $(5/6, 1)$, $(1,1)\}$, and the vertices of its GCM are given by $\{(0,0,)$, $(1/2, 1/4)$, $(1,1)\}$. This is displayed in the bottom left panel of Figure 1. The left-hand slopes of the GCM are $1/2$ on the interval $(0, 1/2]$ and $3/2$ on the interval $(1/2, 1]$, which implies that $\theta_n^*(z) = 1/2$ for $z \in (-\infty, 1]$ and $\theta_n^*(z) = 3/2$ for $z \in (1, \infty)$. This is displayed in the bottom right panel of Figure 1.

We note that the maximum likelihood estimators $\hat{F}_n$ of F_0 and $\hat{G}_n$ of G_0 derived in Dykstra et al. (1995) for the fully discrete case are different than F_n^* and G_n^*. In particular, both $\hat{F}_n$ and $\hat{G}_n$ have jumps at all the unique values of the data $\{-1, 0, 1, 2, 3, 6\}$ with values $\hat{F}_n(-1) = 1/16$,

$\hat{F}_n(0) = 3/16$, $\hat{F}_n(1) = 1/4$, $\hat{F}_n(2) = 3/8$, and $\hat{F}_n(3) = 7/8$, and $\hat{F}_n(6) = 1$; and $\hat{G}_n(-1) = 1/8$, $\hat{G}_n(0) = 3/8$, $\hat{G}_n(1) = 1/2$, $\hat{G}_n(2) = 7/12$, $\hat{G}_n(3) = 11/12$, and $\hat{G}_n(6) = 1$. However, the maximum likelihood estimator $\hat{\theta}_n(z) = \Delta\hat{F}_n(z)/\Delta\hat{G}_n(z)$ is equal to $\theta_n^*(z)$ for each $z \in \{-1, 0, 1, 2, 3, 6\}$.

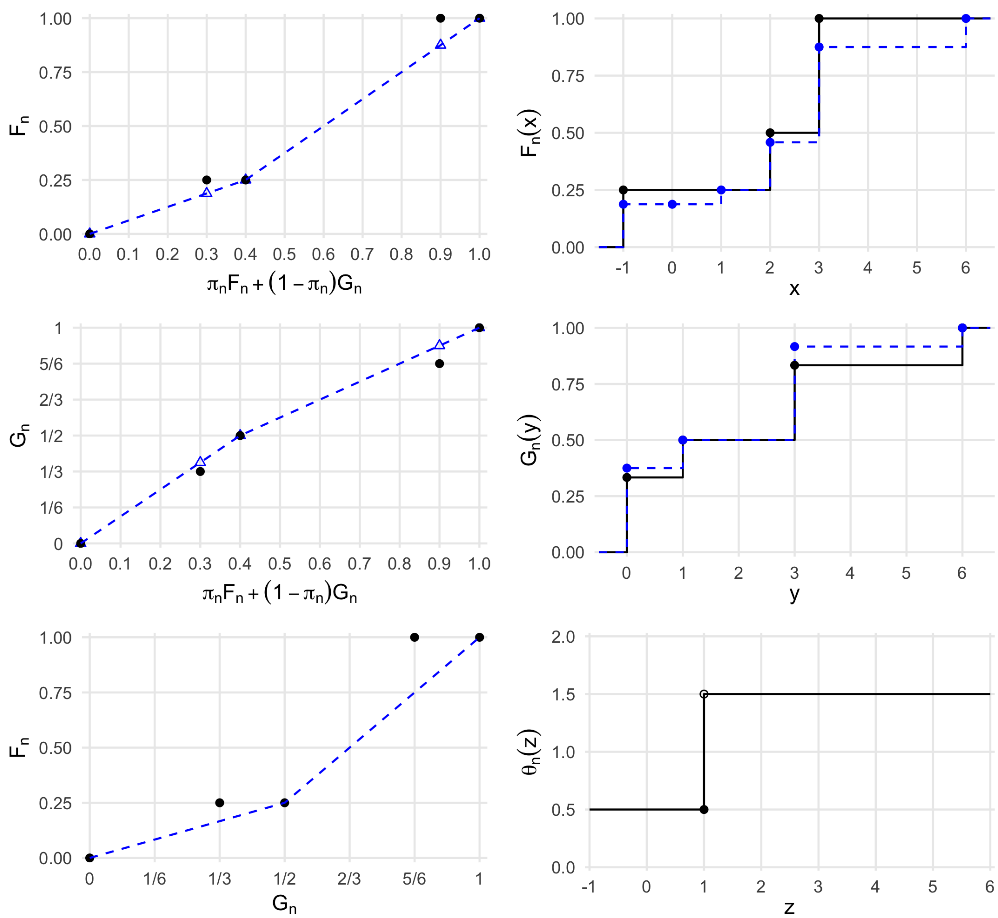

Figure 1: Example of the process of constructing the maximum likelihood estimator for $(Y_1, \ldots, Y_6) = (0, 0, 1, 3, 3, 6)$ and $(X_1, \ldots, X_4) = (-1, 2, 3, 3)$. The graph of F_n versus $\pi_n F_n + (1 - \pi_n) G_n$ evaluated at $z_1, \ldots, z_m$ and its GCM are shown in the upper left. The resulting MLE F_n^* and F_n are shown in the upper right, and the graph of G_n versus $\pi_n F_n + (1 - \pi_n) G_n$ evaluated at $z_1, \ldots, z_m$ and its LCM are shown in the center left, and the resulting MLE G_n^* and G_n are shown in the middle right. The ODC diagram of F_n versus G_n and its GCM are shown in the bottom left, and the resulting MLE θ_n^* is shown in the bottom right.

S2 Proof of Theorems

Proof of Theorem 1. We first suppose that $F \ll G$ and ν is non-decreasing on $\mathcal{G}$, and we show that $F(A)G(B) \leq F(B)G(A)$ for all measurable $A \leq B$. Recall that $F(A) = \int_A dF$ when A is a set, and $A \leq B$ means that $a \leq b$ for all $a \in A$ and $b \in B$. Since $F \ll G$, we have that $F(x) = \int_{-\infty}^{x} \nu(u)\, dG(u)$ for all x. We then have by Fubini's Theorem that

$$\begin{aligned} F(A)G(B) &= \int_{x\in A} dF(x) \int_{y\in B} dG(y) = \int_{x\in A} \nu(x)\, dG(x) \int_{y\in B} dG(y) \\ &= \int_{(x,y)\in A\times B} \nu(x)\, d(G\times G)(x,y). \end{aligned}$$

Now since ν is non-decreasing and $x \leq y$ for all $x \in A$ and $y \in B$, we have

$$\int_{(x,y)\in A\times B} \nu(x)\, d(G\times G)(x,y) \leq \int_{(x,y)\in A\times B} \nu(y)\, d(G\times G)(x,y).$$

Finally, applying Fubini's Theorem again yields

$$\int_{(x,y)\in A\times B} \nu(y)\, d(G\times G)(x,y) = \int_{x\in A} dG(x) \int_{y\in B} \nu(y)\, dG(y) = G(A)F(B).$$

Next, we suppose that $F(A)G(B) \leq F(B)G(A)$ for all measurable $A \leq B$, and we show that $R_{F,G}$ is convex on $Im(G)$. Let $t, u, v \in Im(G)$, where $t < v$ and $u = \lambda t + (1-\lambda)v$ for $\lambda \in (0,1)$. We then let $A = (G^{-}(t), G^{-}(u)]$

and $B = (G^{-}(u), G^{-}(v)]$, which are both Borel sets satisfying $A \leq B$ since G^{-} is necessarily non-decreasing. We then have $F(A) = F(G^{-}(u)) - F(G^{-}(t)) = R_{F,G}(u) - R_{F,G}(t)$ and similarly $F(B) = R_{F,G}(v) - R_{F,G}(t)$. In addition, since $G(G^{-}(z)) = z$ for any $z \in Im(G)$, we also have $G(A) = G(G^{-}(u)) - G(G^{-}(t)) = u - t = (1-\lambda)(v-t)$ and similarly $G(B) = v - u = \lambda(v-t)$. We then have by assumption that

$$\begin{aligned}[R_{F,G}(u) - R_{F,G}(t)][\lambda(v-t)] &= F(A)G(B) \leq F(B)G(A) \\ &= [(1-\lambda)(v-t)][R_{F,G}(v) - R_{F,G}(t)].\end{aligned}$$

Therefore, $\lambda\left[R_{F,G}(u) - R_{F,G}(t)\right] \leq (1-\lambda)\left[R_{F,G}(v) - R_{F,G}(u)\right]$, which implies that $R_{F,G}(u) \leq \lambda R_{F,G}(t) + (1-\lambda)R_{F,G}(v)$, which shows that $R_{F,G}$ is convex on $Im(G)$.

Finally, we suppose that $F \ll G$, $R := R_{F,G}$ is convex on $Im(G)$, and ν is continuous on $\mathcal{G}$, and we show that ν is nondecreasing on $\mathcal{G}$. This is the most difficult of the three implications. The basic argument amounts to using convexity of R to compare the slopes of chords or sequences of chords, and to relate these slopes to values of ν. Let $x, y \in \mathcal{G}$ with $x < y$. Suppose that we can find sequences $\{z_j\}_{j\geq 1}$ and $\{w_j\}_{j\geq 1}$ such that $s_j := [R(G(x)) - R(G(z_j))]/[G(x) - G(z_j)]$ converges to $\nu(x)$, $t_j := [R(G(y)) - R(G(w_j))]/[G(y) - G(w_j)]$ converges to $\nu(y)$, and $z_j \leq w_j$ for all j large enough. Then, by convexity of R, $s_j \leq t_j$ for all j large enough, which implies that $\nu(x) \leq \nu(y)$. The exact form of $\{z_j\}_{j\geq 1}$ and $\{w_j\}_{j\geq 1}$ depends on how G looks near x and y. In particular, there are three cases for y: (1) $G(y) > G(y-)$ and there exists $p \in [x, y)$ such that $G(y-) = G(p)$; (2) $G(y) > G(y-)$ but there is no $p \in [x, y)$ such that $G(y-) = G(p)$; and (3) $G(y) = G(y-)$. We begin by specifying $\{w_j\}_{j\geq 1}$ in each case.

In case (1), we take $w_j = p$ for all j. Since $F \ll G$, we must have $F(G^{-}(G(p))) = F(y-)$, so that $t_j = \nu(y)$ for all j. In case (2), it must be that $G^{-}(G(y-)) = y$. In this case, there exists $\{w_j\}_{j\geq 1}$ increasing to y such that $w_j \in (x, y) \cap \mathcal{G}$ for each j, $G(w_j)$ increases to $G(y-)$ and $F(w_j)$ increases to $F(y-)$. We then have that $R(G(w_j))$ increases to $F(G^{-}(G(y-))-) = F(y-)$, so that t_j increases to $[F(y) - F(y-)]/[G(y) - G(y-)] = \nu(y)$. In case (3), we first note that $F(G^{-}(G(y))) = F(y)$ since $F \ll G$. Additionally, since $y \in \mathcal{G}$, there exist $\{w_j\}_{j\geq 1}$ in $\mathcal{G}$ with $G^{-}(G(w_j)) = w_j$ for each j that either (a) increases to y and $G(w_j) < G(y)$ for each j, or (b) decreases to y and $G(w_j) > G(y)$ for each j. In either case, we have

$$t_j = \frac{\int_{w_j}^{y} \nu(u)\, dG(u)}{G(y) - G(w_j)} = \nu(y) + \frac{\int_{w_j}^{y} [\nu(u) - \nu(y)]\, dG(u)}{G(y) - G(w_j)} \ .$$

For any $\varepsilon > 0$, by continuity of ν over $\mathcal{G}$, we can find m such that $j \geq m$ implies $|\nu(u) - \nu(y)| < \varepsilon$ for all $u \in [w_j, y] \cap \mathcal{G}$. If (a) holds and t_j is bounded above, we then have $\int_{w_j}^{y} |\nu(u) - \nu(y)| \, dG(u) \leq \varepsilon[G(y) - G(w_j)]$ for all $j \geq m$, so that then $\lim_{j\to\infty} t_j = \nu(y)$. If t_j is not bounded above then $\nu(y) = +\infty$, so that $\nu(x) \leq \nu(y)$ trivially. If (b) holds then t_j is bounded below by zero, so by a similar calculation $\lim_{j\to\infty} t_j = \nu(y)$.

The three cases for x are similar: (1) $G(x) > G(x-)$ and there exists $q \in [-\infty, x)$ such that $G(x-) = G(q)$; (2) $G(x) > G(x-)$ but there is no such q; and (3) $G(x) = G(x-)$. In case (1), we take $z_j = q$ for all j. Since $F \ll G$, we must have $F(G^{-}(G(q))) = F(x-)$, so that $s_j = \nu(y)$ for all j. In case (2), it must be that $G^{-}(G(x-)) = x$, and again there exists an increasing sequence $\{z_j\}_{j\geq 1}$ increasing to x such that $z_j \in (-\infty, x) \cap \mathcal{G}$ for each j, $G(z_j)$ increases to $G(x-)$ and $F(z_j)$ increases to $F(x-)$. We then have that $R(G(z_j))$ increases to $F(x-)$, so that s_j increases to $\nu(x)$. In case (3), $F(G^{-}(G(x))) = F(x)$, and since $x \in \mathcal{G}$, there exists $\{z_j\}_{j\geq 1}$ in $\mathcal{G}$ with $G^{-}(G(z_j)) = z_j$ for each j that either (a) increases to x and $G(z_j) < G(x)$ for each j, or (b) decreases to x and $G(z_j) > G(x)$ for each j. If (a) holds and s_j is bounded above, then s_j converges to $\nu(x)$ by continuity of ν as before. If s_j is not bounded above then s_j converges to $\nu(x) = +\infty$. If (b) holds then s_j is bounded below by zero, so again $\lim_{j\to\infty} s_j = \nu(x)$.

Of the nine pairings of cases for y and cases for x, the only situation in which it is not immediately clear that $z_j \leq w_j$ for all j large enough is that

z_j decreases to x (case 3b) and $w_j = p$ for all j (case 1). However, we note that $x = p$ if and only if $G(x) = G(y-)$, which would imply that case (3b) cannot hold for x. Therefore, if z_j decreases to x and $w_j = p$, then $p > x$, so that $z_j < w_j$ for all j large enough. This completes the argument.

Finally, we address statement (2) of the result: we suppose that $F \ll G$ and ν is continuous and non-decreasing on $\mathcal{G}$, and we show that $\theta_{F,G} = \nu$ on $\mathcal{G}$. By (1), R is convex on $Im(G)$. First, we claim that $GCM_{[0,1]}(R) = H$, where $H : [0,1] \to [0,1]$ takes the following form. For any $u \in Im(G)$, $H(u) := R(u)$. If $u \notin Im(G)$, then there exists $x \in \mathbb{R}$ and $\lambda \in [0,1)$ such that $u = \lambda G(x-) + (1-\lambda)G(x)$. We then define $H(u) := \lambda R(G(x-)-) + (1-\lambda)R(G(x))$. Thus, H is the linear interpolation of $R|_{Im(G)}$ to $[0,1]$. In order to show that H indeed equals $GCM_{[0,1]}(R)$, we need to show that (a) H is convex, (b) $H \leq R$, and (c) $H \geq \bar{H}$ for any other convex minorant of R.

For (a), we let $u, v \in [0,1]$ and $p = \lambda u + (1-\lambda)v$ for $\lambda \in (0,1)$. There then exist $u_1 \leq u_2 \leq p_1 \leq p_2 \leq v_1 \leq v_2$ which are all elements of $Im(G)$ and $\lambda_1, \lambda_2, \lambda_3 \in [0,1]$ such that $u = \lambda_1 u_1 + (1-\lambda_1)u_2$, $v = \lambda_2 v_1 + (1-\lambda_2)v_2$, and $p = \lambda_3 p_1 + (1-\lambda_3)p_2$, and furthermore $H(u) = \lambda_1 R(u_1-) + (1-\lambda_1)R(u_2)$, $H(v) = \lambda_2 R(v_1-) + (1-\lambda_2)R(v_2)$, and $H(p) = \lambda_3 R(p_1-) + (1-\lambda_3)R(p_2)$. The remainder of the argument is best seen with a diagram. Let U be the point $(u, H(u))$, U_1 be the point $(u_1, H(u_1))$, and so on. By convexity of R, the line segment $\overline{P_1P_2}$ lies below or on the line segment $\overline{U_2V_1}$, which lies below or on $\overline{UV_1}$, which lies below or on $\overline{UV}$. Therefore, $(p, H(p))$, which falls on $\overline{P_1P_2}$, is no greater than $(p, \lambda H(u) + (1-\lambda)H(p))$, which falls on $\overline{UV}$.

For (b), by definition, $H(u) = R(u)$ for any $u \in Im(G)$. If $u \notin Im(G)$, then $u = \lambda G(x-) + (1-\lambda)G(x)$, and hence $G^{-}(u) = G^{-}(G(x)) = x$. As a result, $R(u) = R(G(x)) > H(u) = \lambda R(G(x-)-) + (1-\lambda)R(G(x))$.

We have now shown that H is a convex minorant of R. For (c), if $\bar{H}$ is another convex minorant of R, then clearly $H(u) \geq \bar{H}(u)$ for all $u \in Im(G)$. If $u \notin Im(G)$, then $u = \lambda G(x-) + (1-\lambda)G(x)$. If $G(x-) \in Im(G)$, then $\bar{H}(u) \leq \lambda H(G(x-)) + (1-\lambda)H(G(x)) \leq \lambda R(G(x-)) + (1-\lambda)R(G(x)) = H(u)$. If $G(x-) \notin Im(G)$, then there must be an $\varepsilon > 0$ such that $z \in Im(G)$ for all $z \in (G(x-) - \varepsilon, G(x-))$, so that $\bar{H}(u) \leq \lambda(z)R(z-) + (1 - \lambda(z))R(G(x))$ for each $z \in (G(x-) - \varepsilon, G(x-))$, where $\lambda(z) \in (0,1)$ and $\lambda(z) \to \lambda$ as $z \to G(x-)$. Taking the limit as $z \to G(x-)$, we have that $\bar{H}(u) \leq \lambda R(G(x-)-) + (1-\lambda)R(G(x)) = H(u)$.

We now have that $\theta_{F,G}(x) = (\partial_{-}H)(G(x))$, so it remains to show that $(\partial_{-}H)(G(x)) = \nu(x)$ for all $x \in \mathcal{G}$. First, if $G(x) > G(x-)$, then $H(u) = \lambda R(G(x-)-) + (1-\lambda)R(G(x)) = \lambda F(x-) + (1-\lambda)F(x)$ for all $u = \lambda G(x-) +$

$(1-\lambda)G(x)$ for $\lambda \in (0,1)$. Therefore, $(\partial_- H)(u) = [F(x) - F(x-)]/[G(x) - G(x-)] = \nu(x)$ for all such u, so that $(\partial_- H)(G(x)) = \nu(x)$. If instead $x \in \mathcal{G}$ and $G(x) = G(x-)$ then $H(G(x)) = R(G(x))$, and it is straightforward to see from the definition of R that $(\partial_- R)(G(x)) = \nu(x)$. □

Proof of Theorem 2. We first note that $L_n(F, G) = 0$ for any G such that $G(Y_j) = G(Y_j-)$ for any $j \in \{1, \ldots, n_2\}$. As a result, we may restrict our attention to G such that $G(Y_j) > G(Y_j-)$ for all j, which implies that G^- has support at each $G(Y_j)$. For any such G, we define $\bar{G} := G \circ L$, where $L(y) := \max\{Y_j : Y_j \leq y\}$. We then have $\bar{G}(Y_j) - \bar{G}(Y_j-) \geq G(Y_j) - G(Y_j-)$ for each j. Furthermore, the support of $\bar{G}^-$ is $\{G(Y_j) : j = 1, \ldots, n_2\}$ is contained in the support of G, $\bar{G}(Y_j) = G(Y_j)$ for each j, and $F \circ G^-$ is by assumption convex on the support of G^-. Therefore, $F \circ \bar{G}^-$ is convex on the support of $\bar{G}^-$, so that $(F, \bar{G}) \in \mathcal{M}_0$ and $L_n(F, \bar{G}) \geq L_n(F, G)$. Hence, we may further restrict our attention to G which are discrete with jumps at $Y_1, \ldots, Y_{n_2}$. By a similar argument, we can restrict our attention to F which are discrete with jumps at $X_1, \ldots, X_{n_1}$ *or* $Y_1, \ldots, Y_{n_2}$.

We define $y_0 := -\infty$, and $u_j := G(y_j)$, so that the support of G^- for any discrete G with jumps at $Y_1, \ldots, Y_{n_2}$ is $\{u_j : j = 0, \ldots, m_2\}$, and $G^-(u_j) = y_j$. Defining $g_j := u_j - u_{j-1}$ and s_j the number of Y_k such that $Y_k = y_j$, we have $\prod_{j=1}^{n_2} [G(Y_j) - G(Y_j-)] = \prod_{j=1}^{m_2} g_j^{s_j}$. We then define

$f_j := F(y_j) - F(y_j-)$ for each j, and we note that $(F, G) \in \mathcal{M}_0$ if and only if $f_1/g_1 \leq f_2/g_2 \leq \cdots \leq f_{m_2}/g_{m_2}$. Suppose that the values $f_1, \ldots, f_{m_2}$ are fixed in such a way as to satisfy these constraints. We denote by $\mathcal{I}_j := \{k : x_k \in (y_{j-1}, y_j]\}$ for $j = 1, \ldots, m_2 + 1$, where $y_{m_2+1} := +\infty$, and by r_i the number of X_k such that $X_k = x_i$. Noting that $\mathcal{I}_1, \ldots, \mathcal{I}_{m_2+1}$ are disjoint with union $\{1, \ldots, m_1\}$, we then have

$$\prod_{i=1}^{n_1} [F(X_i) - F(X_i-)] = \prod_{j=1}^{m_2+1} \prod_{k \in \mathcal{I}_j} [F(x_k) - F(x_k-)]^{r_k} \ .$$

Additionally, for each $j \in \{1, \ldots, m_2 + 1\}$, we must have that

$$\sum_{k \in \mathcal{I}_j} [F(x_k) - F(x_k-)] = f_j.$$

Therefore, maximizing $L_n(F, G)$ with respect to F with $f_1, \ldots, f_{m_2+1}$ fixed amounts to maximizing

$$\prod_{k \in \mathcal{I}_j} [F(x_k) - F(x_k-)]^{r_k} \text{ subject to } \sum_{k \in \mathcal{I}_j} [F(x_k) - F(x_k-)] = f_j$$

for each j. This implies that a maximizer F_n^* must satisfy

$$F_n^*(x_k) - F_n^*(x_k-) = f_j \frac{r_k}{\sum_{l \in \mathcal{I}_j} r_l}$$

for each $x_k \in \mathcal{I}_j$. Therefore, $\prod_{k \in \mathcal{I}_j} [F_n^*(x_k) - F_n^*(x_k-)]^{r_k}$ is proportional to $\prod_{k \in \mathcal{I}_j} f_j^{r_k} = f_j^{R_j}$ for $R_j := \sum_{k \in \mathcal{I}_j} r_k$, which is the number of X_i in the interval $(y_{j-1}, y_j]$.

We note that if there are j such that no $x_k \in (y_j, y_{j+1}]$ but $f_j > 0$, then there are infinitely many maximizers because any F_n^* that assigns mass f_j to the interval $(y_{j-1}, y_j]$ yields the same likelihood and satisfies the constraints. In these cases, for the sake of uniqueness we will put mass f_j at the point y_j.

We have at this point reduced the problem to maximizing

$$\left\{ \prod_{k=1}^{m_2+1} f_k^{R_k} \right\} \left\{ \prod_{k=1}^{m_2} g_k^{s_k} \right\} = \left\{ \prod_{k=1}^{m_2} f_k^{R_k} g_k^{s_k} \right\} f_{m_2+1}^{R_{m_2+1}}$$

subject to $f_1/g_1 \leq f_2/g_2 \leq \cdots \leq f_{m_2}/g_{m_2}$ and $\sum_{k=1}^{m_2} g_k = \sum_{k=1}^{m_2+1} f_k = 1$. Letting $\bar{f}_k := f_k/(1 - f_{m_2+1})$ for $k \leq m_2$, this is equivalent to maximizing

$$\bar{L}_n(\bar{f}_1, \ldots, \bar{f}_{m_2}, f_{m_2+1}, g_1, \ldots, g_{m_2}) := \left\{ \prod_{k=1}^{m_2} \bar{f}_k^{R_k} g_k^{s_k} \right\} (1 - f_{m_2+1})^{n_1 - R_{m_2+1}} f_{m_2+1}^{R_{m_2+1}}$$

subject to $\bar{f}_1/g_1 \leq \bar{f}_2/g_2 \leq \cdots \leq \bar{f}_{m_2}/g_{m_2}$ and $\sum_{k=1}^{m_2} g_k = \sum_{k=1}^{m_2} \bar{f}_k = 1$. The term involving f_{m_2+1} is maximized for $f_{m_2+1}^* = R_{m_2+1}/n_1 = 1 - F_n(y_{m_2})$.

From this point we take a similar approach to that in Dykstra et al.

(1995). We define $\bar{n}_1 := \sum_{k=1}^{m_2} R_k = F_n(y_{m_2})n_1$, $\sigma_k := \bar{n}_1 \bar{f}_k + n_2 g_k$ and $\rho_k := \bar{n}_1 \bar{f}_k / \sigma_k$, so that $\bar{f}_k = \rho_k \sigma_k / \bar{n}_1$ and $g_k = (1-\rho_k)\sigma_k / n_2$. Optimizing $\bar{L}_n$ with with respect to $\bar{f}_1, \dots, \bar{f}_{m_2}$ and $g_1, \dots, g_{m_2}$ such that $\sum_{k=1}^{m_2} \bar{f}_k = \sum_{k=1}^{m_2} g_k = 1$ and $\bar{f}_1/g_1 \leq \bar{f}_2/g_2 \leq \cdots \leq \bar{f}_{m_2}/g_{m_2}$ is equivalent to optimizing

$$\begin{aligned}\bar{L}_n(\boldsymbol{\rho}, \boldsymbol{\sigma}) &= \prod_{k=1}^{m_2} \left[\rho_k \sigma_k / \bar{n}_1\right]^{R_k} \left[(1-\rho_k)\sigma_k / n_2\right]^{s_k} \\ &= \bar{n}_1^{-\bar{n}_1} n_2^{-n_2} \prod_{k=1}^{m_2} \rho_k^{R_k} (1-\rho_k)^{s_k} \prod_{k=1}^{m_2} \sigma_k^{R_k + s_k}\end{aligned}$$

such that $\sum_{k=1}^{m_2} \rho_k \sigma_k = \bar{n}_1$, $\sum_{k=1}^{m_2} \sigma_k = \bar{n}_1 + n_2$, and $\rho_1 \leq \cdots \leq \rho_{m_2}$, where $\boldsymbol{\rho} := (\rho_1, \dots, \rho_{m_2})$ and $\boldsymbol{\sigma} := (\sigma_1, \dots, \sigma_{m_2})$.

Now, $\prod_{k=1}^{m_2} \sigma_k^{R_k+s_k}$ such that $\sum_{k=1}^{m_2} \sigma_k = \bar{n}_1 + n_2$ is maximized for $\sigma_k^* = R_k + s_k$. Next, maximizing $\prod_{i=1}^{m_2} \rho_k^{R_k}(1-\rho_k)^{s_k}$ with respect to $\rho_1 \leq \cdots \leq \rho_{m_2}$ is equivalent to maximizing

$$\sum_{k=1}^{m_2} \left[R_k \log \rho_k + s_k \log(1-\rho_k)\right] = \sum_{k=1}^{m_2} w_k \left[t_k \log \rho_k + (1-t_k)\log(1-\rho_k)\right]$$

for $w_k := R_k + s_k \geq 1$ and $t_k := R_k / w_k$. By Theorem 2.1 and Exercise 2.21 of Groeneboom and Jongbloed (2014), the maximizer $(\rho_1^*, \dots, \rho_{m_2}^*)$ of this expression over all $\rho_1 \leq \cdots \leq \rho_{m_2}$ is given by the weighted isotonic regression of $t_1, \dots, t_{m_2}$ with weights $w_1, \dots, w_{m_2}$. By Lemma 2.1 of Groeneboom and Jongbloed (2014), ρ_k^* is equal to the left derivative of the GCM of the set of points

$$\{(0,0)\} \cup \left\{ \left(\sum_{j=1}^{k} w_k, \sum_{j=1}^{k} t_j w_j \right) : k = 1, \ldots, m_2 \right\}$$
$$= \{(n_1 F_n(y_k) + n_2 G_n(y_k), n_1 F_n(y_k)) : k = 0, \ldots, m_2\}$$

evaluated at $n_1 F_n(y_k) + n_2 G_n(y_k)$. We note that $\sum_{k=1}^{m_2} w_k \rho_k^* = \sum_{k=1}^{m_2} \sigma_k^* \rho_k^* = n_1 F(y_{m_2}) = \bar{n}_1$. Therefore, we have that $L_n(\boldsymbol{\rho}, \boldsymbol{\sigma}) \leq L_n(\boldsymbol{\rho}^*, \boldsymbol{\sigma}^*)$ for all $\boldsymbol{\rho}$ such that $\rho_1 \leq \cdots \leq \rho_{m_2}$ and $\boldsymbol{\sigma}$ such that $\sum_{k=1}^{m_2} \sigma_k = \bar{n}_1 + n_2$. Since $\boldsymbol{\rho}^*$ and $\boldsymbol{\sigma}^*$ also satisfy $\sum_{k=1}^{m_2} \sigma_k^* \rho_k^* = \bar{n}_1$, this implies that $(\boldsymbol{\rho}^*, \boldsymbol{\sigma}^*)$ is an optimizer of $\bar{L}_n$ over the set of stated constraints.

We now have that $f_k^* = (R_k + s_k)(\rho_k^*/n_1)$ and $g_k^* = (R_k + s_k)(1 - \rho_k^*)/n_2$. Since $w_k = R_k + s_k$, this implies that $F_n^*(y_k) = \bar{A}_k/n_1$ and $G_n^*(y_k) = [n_2 G_n(y_k) + n_1 F_n(y_k) - \bar{A}_k]/n_2$, where $\bar{A}_k$ is the value of the GCM of the set of points defined above at $n_1 F_n(y_k) + n_2 G_n(y_k)$. We note that $\bar{A}_k/n_1 = A_k^*$, for A_k^* the value of the GCM of

$$\{(\pi_n F_n(y_k) + (1 - \pi_n) G_n(y_k), F_n(y_k)) : k = 0, \ldots, m_2\}$$

evaluated at $\pi_n F_n(y_k) + (1 - \pi_n) G_n(y_k)$. Additionally, $[n_2 G_n(y_k) + n_1 F_n(y_k) -$

$\bar{A}_k]/n_2 = B_k^*$ for B_k^* the value of the LCM of

$$\{(\pi_n F_n(y_k) + (1-\pi_n)G_n(y_k), G_n(y_k)) : k = 0, \ldots, m_2\}$$

at $\pi_n F_n(y_k) + (1-\pi_n)G_n(y_k)$. □

Proof of Corollary 1. From the proof of Theorem 2, we have that $F_n^*(y_k) = A_k^*$ and $G_n^*(y_k) = G_n(y_k) + \frac{\pi_n}{1-\pi_n}[F_n(y_k) - A_k^*]$. Let $j_0', \ldots, j_K'$ denote the indices of the vertices of the GCM of

$$\{(\pi_n F_n(y_k) + (1-\pi_n)G_n(y_k), F_n(y_k)) : k = 0, \ldots, m_2\}.$$

Then $F_n^*(y_{j_k}) = F_n(y_{j_k})$ for each $k = 0, \ldots, K$ and $G_n^*(y_{j_k}) = G_n(y_{j_k})$. It is also straightforward to see that $\{(h_k, A_k) : k = 0, \ldots, m_2\}$ is a convex minorant of $\{(h_k, F_n(y_k)) : k = 0, \ldots, m_2\}$ if and only if $\{(G_n(y_k), A_k) : k = 0, \ldots, m_2\}$ is a convex minorant of $\{(G_n(y_k), F_n(y_k)) : k = 0, \ldots, m_2\}$. Therefore, $\{(F_n(y_{j_k}), G_n(y_{j_k})) : k = 0, \ldots, K\}$ form the vertices of the GCM of $\{(G_n(y_k), F_n(y_k)) : k = 0, \ldots, m_2\}$. □

Proof of Theorem 3. We note that $G_n(y_j) > 0$ for each j with probability tending to one. Then, since the support $\mathcal{G}$ of G_0 is finite, with probability tending to one the empirical ODC is a left-continuous step function with vertices at $(0,0), (G_n(y_1), F_n(y_1)), \ldots, (G_n(y_{m_2}), F_n(y_{m_2})$, where we note that $G_n(y_{m_2}) = 1$ almost surley. We define

$$\delta := \min\left\{\frac{F_0(y_{j+1}) - F_0(y_j)}{\Delta G_0(y_{j+1})} - \frac{F_0(y_j) - F_0(y_{j-1})}{\Delta G_0(y_j)} : j = 1, \ldots, m_2 - 1\right\},$$

which is positive by assumption. We then have

$$\begin{aligned}
&\frac{F_n(y_{j+1}) - F_n(y_j)}{\Delta G_n(y_{j+1})} - \frac{F_n(y_j) - F_n(y_{j-1})}{\Delta G_n(y_j)} \\
&\quad = \frac{F_0(y_{j+1}) - F_0(y_j)}{\Delta G_0(y_{j+1})} - \frac{F_0(y_j) - F_0(y_{j-1})}{\Delta G_0(y_j)} \\
&\qquad + \frac{[F_n(y_{j+1}) - F_0(y_{j+1})] - [F_n(y_j) - F_0(y_j)]}{\Delta G_0(y_{j+1})} \\
&\qquad + [F_n(y_{j+1}) - F_n(y_j)]\left[\frac{1}{\Delta G_n(y_{j+1})} - \frac{1}{\Delta G_0(y_{j+1})}\right] \\
&\qquad - \frac{[F_n(y_j) - F_0(y_j)] - [F_n(y_{j-1}) - F_0(y_{j-1})]}{\Delta G_0(y_j)} \\
&\qquad - [F_n(y_j) - F_n(y_{j-1})]\left[\frac{1}{\Delta G_n(y_j)} - \frac{1}{\Delta G_0(y_j)}\right].
\end{aligned}$$

Now since F_n is uniformly consistent for F_0 and G_n is uniformly consistent for G_0, and $\Delta G_0(y_j) > 0$ for each j, the second through fifth lines above are $o_P(1)$ uniformly over j. Therefore,

$$\begin{aligned}
\min&\left\{\frac{F_n(y_{j+1}) - F_n(y_j)}{\Delta G_n(y_{j+1})} - \frac{F_n(y_j) - F_n(y_{j-1})}{\Delta G_n(y_j)} : j = 1, \ldots, m_2 - 1\right\} \\
&\qquad \geq \delta - o_P(1),
\end{aligned}$$

which implies that

$$\frac{F_n(y_{j+1}) - F_n(y_j)}{\Delta G_n(y_{j+1})} \geq \frac{F_n(y_j) - F_n(y_{j-1})}{\Delta G_n(y_j)}$$

for all $j = 1, \ldots, m_2 - 1$ with probability tending to one. Therefore, with probability tending to one, the diagram of points $(0,0), (G_n(y_1), F_n(y_1))$, $\ldots, (G_n(y_{m_2}), F_n(y_{m_2}))$ is convex. By Corollary 1, $(G_n^*(y_k), F_n^*(y_k))$ lie on the GCM of $(0,0), (G_n(y_1), F_n(y_1)), \ldots, (G_n(y_{m_2}), F_n(y_{m_2}))$. But these points being convex means that they are equal to their GCM, so that with probability tending to one $G_n^*(y_j) = G_n(y_j)$ and $F_n^*(y_j) = F_n(y_j)$ for each j. We can then see by Theorem 2 that $\Delta F_n^*(x_i) = \Delta F_n(x_i)$ with probability tending to one as well, so that $F_n^* = F_n$ with probability tending to one. We then have with probability tending to one that

$$\theta_n^*(y_j) = \left[\partial_- GCM_{[0,1]}(R_{F_n,G_n})\right] \circ G_n(y_j) = \frac{F_n(y_j) - F_n(y_{j-1})}{\Delta G_n(y_j)}$$

for each $j = 1, \ldots, m_2$. since the GCM of R_{F_n,G_n} is with probability tending to one piecewise linear with knots at the y_j and $\theta_n^* = \partial_- GCM_{[0,1]}(R_{F_n,G_n}) \circ G_n$, we then have that with probability tending to one that θ_n^* is a left-continuous step function with jumps at the y_j. Also, since $G_n(z) = 0$ for $z < y_1$ and $R_{F_n,G_n}(u) = 0$ for all $u \leq 0$, $\theta_n^*(z) = 0$ for $z < y_1$.

We now have

$$n^{1/2}[\theta_n^*(y_j) - \theta_0(y_j)] = n^{1/2}\left[\frac{F_n(y_j) - F_n(y_{j-1})}{G_n(y_j) - G_n(y_{j-1})} - \frac{F_0(y_j) - F_0(y_{j-1})}{G_0(y_j) - G_0(y_{j-1})}\right].$$

Using the notation introduced in Section 3.2, this can be written as

$$n^{1/2}\left\{\frac{\left[\frac{1}{n}\sum_{i=1}^n W_{i1}\right] / \left[\frac{1}{n}\sum_{i=1}^n W_{i2}\right]}{\left[\frac{1}{n}\sum_{i=1}^n W_{i3}\right] / \left[\frac{1}{n}\sum_{i=1}^n W_{i4}\right]} - \frac{E_0(W_{i1})/E_0(W_{i2})}{E_0(W_{i3})/E_0(W_{i4})}\right\}$$
$$= n^{1/2}\left\{g\left(\frac{1}{n}\sum_{i=1}^n W_i\right) - g\left(E_0(W_i)\right)\right\},$$

where $W_i = (W_i1, \ldots, W_{i4})^T$ for $W_{i1} = I(D_i = 1, y_{j-1} < Z_i \leq y_j)$, $W_{i2} = I(D_i = 1)$, $W_{i3} = I(D_i = 0, y_{j-1} < Z_i \leq y_j)$, and $W_{i4} = I(D_i = 0)$, and $g(w_1, w_2, w_3, w_4) = \frac{w_1/w_2}{w_3/w_4}$. By the Central Limit Theorem,

$$\sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^n W_i - E_0(W_i)\right\} \xrightarrow{\text{d}} N_4(0, V_0),$$

where the (j,k) element of the covariance matrix V_0 equals $E_0(W_jW_k) - E_0(W_j)E_0(W_k)$. Applying the delta method to the function g yields (after some algebra) $n^{1/2}[\theta_n^*(y_j) - \theta_0(y_j)] \xrightarrow{\text{d}} N\left(0, \sigma_0^2(y_j)\right)$, where $\sigma_0^2(y_j)$ equals

$$\theta_0(y_j)\frac{\pi_0[F_0(y_j) - F_0(y_{j-1})] + (1-\pi_0)\Delta G_0(y_j) - [F_0(y_j) - F_0(y_{j-1})]\Delta G_0(y_j)}{\pi_0(1-\pi_0)[\Delta G_0(y_j)]^2}.$$

□

Proof of Theorem 4. We note that $f_0(z)/g_0(z) \leq f_0(z')/g_0(z')$ for all $z < z'$ in $[a, b]$ implies that

$$\pi_0^2 f_0(z) f_0(z') + \pi_0(1-\pi_0) f_0(z) g_0(z') \leq \pi_0^2 f_0(z) f_0(z') + \pi_0(1-\pi_0) f_0(z') g_0(z),$$

which implies that $z \mapsto \pi_0 f_0(z)/[\pi_0 f_0(z) + (1-\pi_0)g_0(z)]$ is non-decreasing on $[a, b]$. Therefore,

$$\begin{aligned}(G_0 \circ H_0^{-1})' &= \frac{g_0 \circ H_0^{-1}}{\pi_0 f_0 \circ H_0^{-1} + (1-\pi_0) g_0 \circ H_0^{-1}} \\ &= \frac{1}{1-\pi_0}\left[1 - \frac{\pi_0 f_0 \circ H_0^{-1}}{\pi_0 f_0 \circ H_0^{-1} + (1-\pi_0) g_0 \circ H_0^{-1}}\right]\end{aligned}$$

is non-increasing on $H_0([a, b]) = [0, 1]$. Hence, $G_0 \circ H_0^{-1}$ is concave on $[0, 1]$, so

$$LCM_{[0,1]}(G_0 \circ H_0^{-1}) \circ H_0 = G_0 \circ H_0^{-1} \circ H_0 = G_0.$$

We now note that since $G_n \circ H_n^-(u) \geq G_n(y_{m_2}) = 1$ for any $u \geq h_{m_2}$, $LCM_{[0,h_{m_2}]}(G_n \circ H_n^-) = LCM_{[0,1]}(G_n \circ H_n^-)$. Furthermore, since G_n^* only jumps at $y_1, \ldots, y_{m_2}$, we have

$$G_n^*(y) = LCM_{[0,1]}(G_n \circ \tilde{H}_n^-) \circ \tilde{H}_n(y)$$

for any $y \in \mathbb{R}$, where $\tilde{H}_n := \pi_n F_n \circ L_n + (1-\pi_n) G_n$ for $L_n(z) := G_n^- \circ G_n(z) =$

$\max\{Y_j : Y_j \le z\}$.

Using the notation of Section 3.2, we can write

$$\pi_n F_n(x) = \frac{n_1}{n}\frac{1}{n_1}\sum_{i=1}^{n} D_i I(Z_i \le x) = \mathbb{P}_n \omega_x$$

for $\omega_x(d,x) := dI(z \le x)$, and similarly $(1-\pi_n)G_n(y) = \mathbb{P}_n\eta_y$ for $\eta_y(d,z) := (1-d)I(z \le y)$. We also have $P_0\omega_x = \pi_0 F_0(x)$ and $P_0\eta_y = (1-\pi_0)G_0(y)$. By standard empirical process theory, we therefore have that

$$\{n^{1/2}[\pi_n F_n(x) - \pi_0 F_0(x)] : x \in \mathbb{R}\} = \{n^{1/2}(\mathbb{P}_n - P_0)\omega_x : x \in \mathbb{R}\}$$

and

$$\{n^{1/2}[(1-\pi_n)G_n(y) - (1-\pi_0)G_0(y)] : y \in \mathbb{R}\} = \{n^{1/2}(\mathbb{P}_n - P_0)\eta_y : y \in \mathbb{R}\}$$

converge weakly (jointly) as processes indexed by $\ell^\infty(\mathbb{R})$ to

$$(\mathbb{G}_1 \circ [\pi_0 F_0], \mathbb{G}_2 \circ [(1-\pi_0)G_0])$$

for $\mathbb{G}_1$ and $\mathbb{G}_2$ independent Brownian bridge processes. The two processes are independent because the covariance between the processes is easily seen to be zero. Since the density of G_0 is bounded strictly away from zero on

$[a,b]$, $n^{1/2}([(1-\pi_n)G_n]^{-} - [(1-\pi_0)G_0]^{-1})$ converges weakly in $\ell^\infty(0,1)$ to $-\mathbb{G}_2/[(1-\pi_0)g_0] \circ [\pi_0 G_0]^{-1}$ by Lemma 3.9.23 of van der Vaart and Wellner, 1996. Hence, by Hadamard differentiability of the composition map (see Lemma 3.9.27 of van der Vaart and Wellner, 1996), the functional delta method yields

$$n^{1/2}(G_n^{-} \circ G_n - Id) = n^{1/2}([(1-\pi_n)G_n]^{-} \circ [\pi_n G_n] - Id)$$

converges weakly in $\ell^\infty[a,b]$ to

$$\begin{aligned}&-(\mathbb{G}_2 \circ [(1-\pi_0)G_0])/[(1-\pi_0)g_0 \circ [(1-\pi_0)G_0]^{-1} \circ [(1-\pi_0)G_0]) \\ &\quad + (\mathbb{G}_2 \circ [(1-\pi_0)G_0])/([(1-\pi_0)g_0 \circ [(1-\pi_0)G_0]^{-1} \circ [(1-\pi_0)G_0]) = 0,\end{aligned}$$

so that $\sup_{z\in[a,b]} |L_n(z) - z| = o_P(n^{-1/2})$. Hence, $n^{1/2}(\pi_n F_n \circ L_n - \pi_0 F_0)$ converges weakly to $\mathbb{G}_1 \circ [\pi_0 F_0]$ in $\ell^\infty[a,b]$, and so $n^{1/2}(\tilde{H}_n - H_0)$ converges weakly to $\mathbb{G}_1 \circ [\pi_0 F_0] + \mathbb{G}_2 \circ [(1-\pi_0)G_0]$ in $\ell^\infty[a,b]$. Since G_0 and F_0 are both continuously differentiable on $[a,b]$, so is H_0, and since the derivative of G_0 is bounded away from zero, so is the derivative of H_0. Therefore, using Lemma 3.9.23 of van der Vaart and Wellner, 1996 again, $n^{1/2}(\tilde{H}_n^{-} - H_0^{-})$ converges weakly in $\ell^\infty(0,1-\varepsilon)$ to $(\mathbb{G}_1 \circ [\pi_0 F_0] + \mathbb{G}_2 \circ [(1-\pi_0)G_0]) \circ H_0^{-1}/(h_0 \circ H_0^{-1})$ for any $\varepsilon > 0$, where $h_0 := H_0' = \pi_0 f_0 + (1-\pi_0) g_0$. Then, using the functional delta method for composition again, we have that $n^{1/2}[G_n \circ \tilde{H}_n^- - G_0 \circ H_0^{-1}]$ converges weakly to

$$\left[1 - \frac{g_0 \circ H_0^{-1}}{h_0 \circ H_0^{-1}}\right] \mathbb{G}_2 \circ [(1-\pi_0)G_0] \circ H_0^{-1} - \frac{g_0 \circ H_0^{-1}}{h_0 \circ H_0^{-1}} \mathbb{G}_1 \circ (\pi_0 F_0) \circ H_0^{-1},$$

which we define as $\mathbb{G}_3$. Now by Proposition 2.1 of Beare and Fang (2017),

$$n^{1/2}[LCM_{[0,1]}(G_n \circ \tilde{H}_n^-) - LCM_{[0,1]}(G_0 \circ H_0^{-1})]$$

converges weakly to $LCM'_{[0,1],G_0 \circ H_0^{-1}}(\mathbb{G}_3)$. Using Hadamard differentiability of composition once more, we have that

$$n^{1/2}[G_n^* - G_0] = n^{1/2}[LCM_{[0,1]}(G_n \circ \tilde{H}_n^-) \circ \tilde{H}_n - LCM_{[0,1]}(G_0 \circ H_0^{-1}) \circ H_0]$$

converges weakly to

$$LCM'_{[0,1],G_0 \circ H_0^{-1}}(\mathbb{G}_3) \circ H_0 + \frac{g_0}{h_0}\left[\mathbb{G}_1 \circ (\pi_0 F_0) + \mathbb{G}_2 \circ [(1-\pi_0)G_0]\right]$$

If f_0/g_0 is strictly increasing on $[a, b]$, then $G_0 \circ H_0^{-1}$ is strictly concave on $[0, 1]$, in which case $LCM'_{[0,1],G_0 \circ H_0^{-1}}$ is the identity operator by Proposition 2.2 of Beare and Fang (2017). Hence, in this case $n^{1/2}[G_n^* - G_0]$

converges weakly to

$$\begin{aligned}
&\mathbb{G}_3 \circ H_0 + \frac{g_0}{h_0}\left[\mathbb{G}_1 \circ (\pi_0 F_0) + \mathbb{G}_2 \circ [(1-\pi_0)G_0]\right] \\
&\quad = \left[1 - \frac{g_0}{h_0}\right] \mathbb{G}_2 \circ [(1-\pi_0)G_0] - \frac{g_0}{h_0}\mathbb{G}_1 \circ (\pi_0 F_0) \\
&\qquad + \frac{g_0}{h_0}\left[\mathbb{G}_1 \circ (\pi_0 F_0) + \mathbb{G}_2 \circ [(1-\pi_0)G_0]\right] \\
&\quad = \mathbb{G}_2 \circ [(1-\pi_0)G_0],
\end{aligned}$$

which, as noted above, is the same limit distribution as $n^{1/2}[G_n - G_0]$. Furthermore, we have

$$\begin{aligned}
n^{1/2}\|G_n^* - G_n\|_\infty &\leq n^{1/2}\|[LCM_{[0,1]}(G_n \circ \tilde{H}_n^-) \circ \tilde{H}_n - G_n \circ \tilde{H}_n^- \circ \tilde{H}_n\|_\infty \\
&\quad + n^{1/2}\|G_n \circ \tilde{H}_n^- \circ \tilde{H}_n - G_n\|_\infty.
\end{aligned}$$

When f_0/g_0 is strictly increasing so that $LCM'_{[0,1],G_0 \circ H_0^{-1}}$ is the identity, the functional delta method (e.g. Theorem 3.9.4 of van der Vaart and Wellner, 1996) implies that

$$\begin{aligned}
&n^{1/2}\|[LCM_{[0,1]}(G_n \circ \tilde{H}_n^-) \circ \tilde{H}_n - G_n \circ \tilde{H}_n^- \circ \tilde{H}_n\|_\infty \\
&\quad \leq n^{1/2}\|[LCM_{[0,1]}(G_n \circ \tilde{H}_n^-) - G_n \circ \tilde{H}_n^-\|_\infty = o_P(1).
\end{aligned}$$

Similarly, since as shown above, $n^{1/2}(\tilde{H}_n^- \circ \tilde{H}_n - Id)$ converges weakly in $[a, b]$

to 0, $n^{1/2}\|G_n \circ \tilde{H}_n^- \circ \tilde{H}_n - G_n\|_\infty = o_P(1)$. Therefore, $n^{1/2}\|G_n^* - G_n\|_\infty = o_P(1)$ if f_0/g_0 is strictly increasing on $[a, b]$.

Now we turn attention to F_n^*. By Theorem 2, for each $y \in \{Y_1, \ldots, Y_n\}$, we know that $F_n^*(y_k) = GCM_{[0,h_{m_2}]}(F_n \circ H_n^-) \circ H_n(y_k)$. We can extend the GCM operation to entirety of $[0, 1]$, so that $F_n^*(y_k) = GCM_{[0,1]}(F_n \circ H_n^-) \circ H_n(y_k)$, because the slope of the secant of $F_n \circ H_n^-$ from h_{m_2} to $H_n(x_j)$ for any $x_j > y_{m_2}$ is $[F_n(x_j) - F_n(y_{m_2})]/[H_n(x_j) - h_{m_2}] = 1/\pi_n$, while the slope of the secant from any other z in the support of H_n is $[F_n(y_{m_2}) - F_n(z)]/[H_n(y_{m_2}) - H_n(z)] \leq [F_n(y_{m_2}) - F_n(z)]/[\pi_n\{F_n(y_{m_2}) - F_n(z)\}] = 1/\pi_n$. Therefore, performing the GCM over $[0, 1]$ rather than $[0, h_{m_2}]$ cannot change the value of the GCM for any $u \leq h_{m_2}$.

We now define $F_n^\ell(y) := GCM_{[0,1]}(F_n \circ H_n^-) \circ H_n \circ L_n$, where $L_n = G_n^- \circ G_n$ as above, so that $\bar{F}_n$ is the right-continuous step function with jumps at $y_1, \ldots, y_{m_2}$ and agreeing with F_n^* at these points. We similarly define $F_n^u := GCM_{[0,1]}(F_n \circ H_n) \circ H_n \circ R_n$, where $R_n := G_n^- \circ \bar{G}_n$ for $\bar{G}_n(y) := \frac{1}{n}\sum_{i=1}^n I(Y_i < y) + 1/n$. Since the Y_j's are unique with probability one, $\bar{G}_n$ is a left-continuous version of G_n that agrees at $y_1, \ldots, y_{m_2}$, and $\bar{G}_n \geq G_n$. Therefore, since any MLE F_n^* is a proper CDF, we have $F_n^\ell \leq F_n^* \leq F_n^u$. One can show that $\|F_n^\ell - F_n\|_\infty = o_P(n^{-1/2})$ and $\|F_n^u - F_n\|_\infty = o_P(n^{-1/2})$ when f_0/g_0 is strictly increasing using the same argument as that used above for showing that $\|G_n^* - G_n\| = o_P(n^{-1/2})$. We then have $\|F_n^* - F_n\|_\infty = o_P(n^{-1/2})$ as well. □

Proof of Theorem 5. The conditions of Theorem 1 of Westling and Carone (2020) are satisfied by the uniform consistency of empirical distribution functions. □

Proof of Theorem 6. This result follows by the delta method, as discussed in the text. □

S3 Additional simulations: discrete case

We now present results from a numerical study of the properties of the maximum likelihood estimator in the case where both F_0 and G_0 are fully discrete. We set F_0 and G_0 as the distribution functions of Poisson random variables with rates 6 and 4, respectively, and we set π_0 to 0.4. We simulated 1000 datasets each for $n \in \{500, 1000, 5000, 10000\}$ and estimated the maximum likelihood estimator θ_n^*, the empirical mass ratio function, defined as the ratio of the empirical mass functions of $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$, and the sample splitting estimators with $m \in \{5, 10, 20\}$ (Banerjee et al., 2019). We computed Wald-type confidence intervals (constructed around $\log \theta_n^*$ and exponentiated) using the asymptotic variance provided in Section 4.1 of the main text, likelihood ratio-based confidence intervals, and confidence intervals around the sample splitting estimators as outlined in Section 5 of the main text.

The left panel of Figure 2 displays the distribution of $\theta_n^*(z) - \theta_0(z)$ for $z \in \{0, 1, \ldots, 6\}$, and demonstrates that θ_n^* is approximately unbiased in large samples. The right panel of Figure 2 displays the ratio of the empirical standard deviation of $n^{1/2}[\theta_n^*(z) - \theta_0(z)]$ to the standard deviation based on the asymptotic theory, and demonstrates that the empirical standard deviation of $\theta_n^*(z)$ approaches the standard deviation defined by the limit

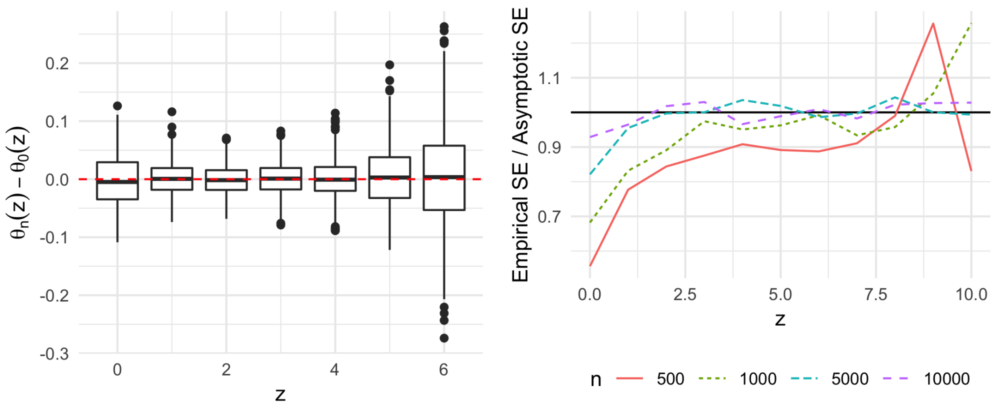

Figure 2: Left: boxplots of $\theta_n^*(z) - \theta_0(z)$ with $n = 10K$ in the fully discrete case. Right: empirical standard errors of $n^{1/2}[\theta_n^*(z) - \theta_0(z)]$ divided by the limit theory-based counterparts for $z \in \{0, 1, \ldots, 10\}$.

theory as the sample size grows, and that $\theta_n^*(z)$ is more efficient than the limit theory suggests in smaller samples for small values of z.

Figure 3 displays the ratio of the mean squared errors of the empirical and sample splitting estimators to that of the maximum likelihood estimator. For the empirical estimator, this ratio approaches one as sample size grows, which agrees with our theoretical result suggesting that the two estimators are asymptotically equivalent. However, in small samples, the maximum likelihood estimator has strictly smaller mean squared error than the empirical estimator. The mean squared errors of the sample splitting estimators also approach that of the maximum likelihood estimator as the sample size grows, which is concurrent with existing theory for $n^{-1/2}$-rate asymptotics.

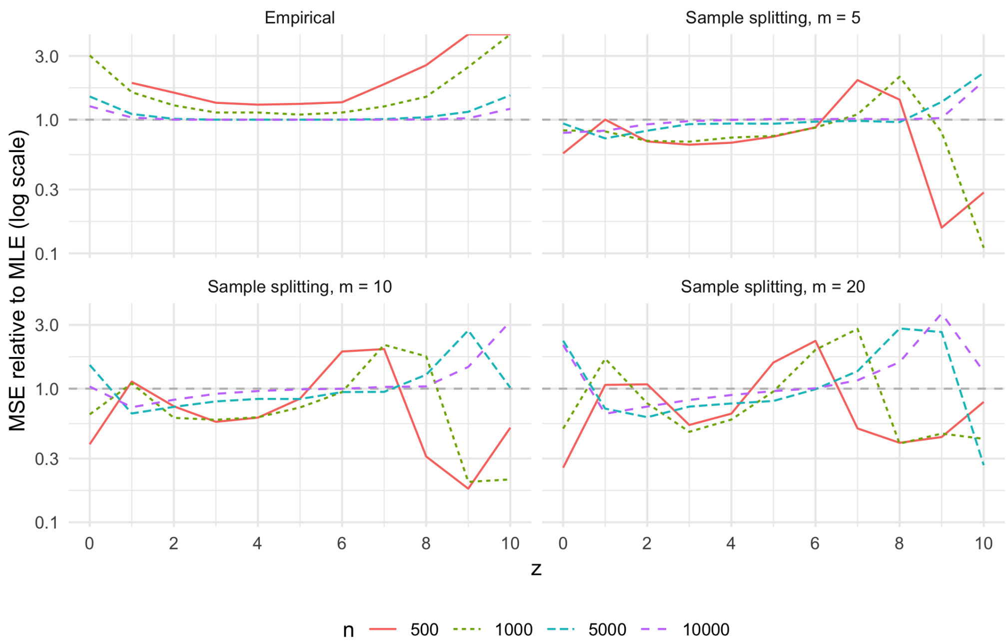

Figure 3: Relative mean squared errors of the empirical estimator and the sample splitting estimators to the maximum likelihood estimator for $z \in \{0, 1, \ldots, 10\}$ and various sample sizes n in the fully discrete case. The maximum likelihood has better mean squared error for y-values greater than one, and the other estimator has better mean squared error for y-values less than one.

Figure 4 shows the empirical coverage of 95% confidence intervals for $\theta_0(z)$ constructed using Wald-type confidence intervals with a plug-in standard error according to the results presented in Section 4.1 of the main text, the inverted likelihood ratio test approach of Banerjee and Wellner (2001), and the sample splitting approach of Banerjee et al. (2019) described in the main text. We note that the likelihood ratio approach does not provide intervals at the end point $z = 0$. The plug-in method is conservative in small samples, but its coverage approaches 95% for $z \neq 0$ as n grows. The

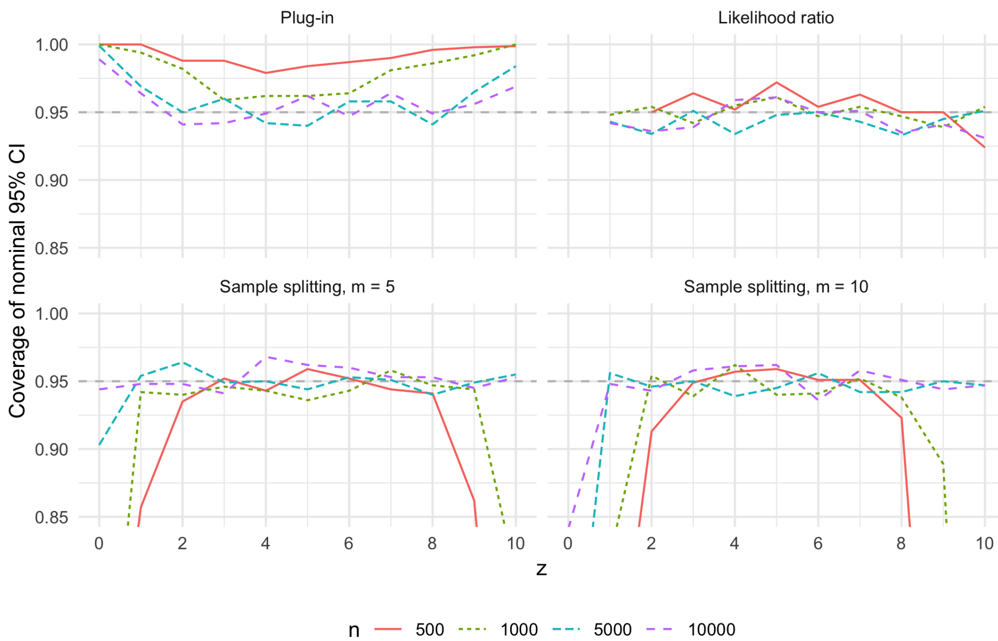

Figure 4: Coverage of 95% CIs in the fully discrete case for $z \in \{0, 1, \ldots, 10\}$, various sample sizes n, and four methods: the plug-in method centered around the log of the maximum likelihood estimator (upper left), the inverted likelihood ratio tests (upper right), and the sample splitting method with $m = 5$ (lower left) and $m = 10$ (lower right). Note that the likelihood ratio method does not provide intervals at the endpoints.

likelihood ratio method provides excellent coverage at all sample sizes. The sample splitting method has good coverage in large enough sample sizes.

S4 Additional simulations: continuous case

We now present results from a numerical study of the properties of the maximum likelihood estimator in the case where both F_0 and G_0 are fully continuous. We set F_0 and G_0 as the distribution functions of exponential random variables with rates 1 and 2, respectively, and we set π_0 to 0.4. We simulated 1000 datasets each for $n \in \{500, 1000, 5000, 10000\}$ and estimated the maximum likelihood estimator, the maximum smoothed likelihood estimator of Yu et al. (2017), the non-monotone estimator based on kernel density estimates for each $z \in \{0, 0.1, \ldots, 1.9, 2\}$, and the sample splitting estimator with $m \in \{5, 10, 20\}$ (Banerjee et al., 2019). We constructed confidence intervals at each z using the transformed plug-in and likelihood ratio-based methods described in Section 4.2 of the main text.

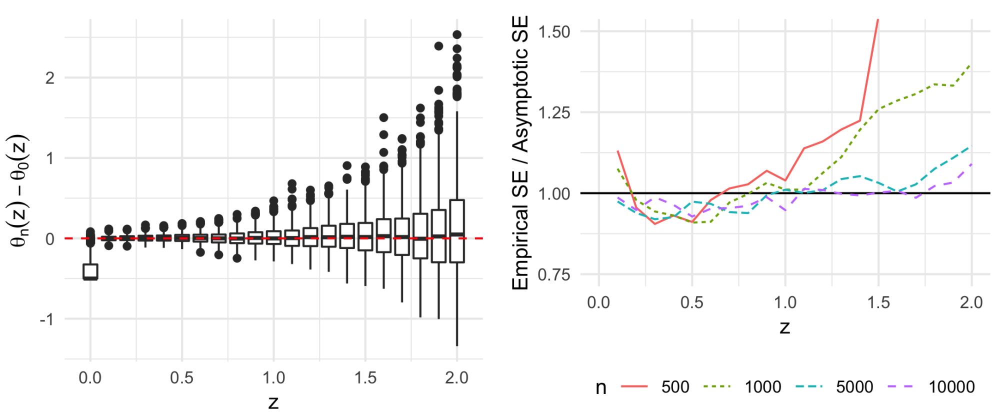

Figure 5: Left: boxplots of $\theta_n^*(z) - \theta_0(z)$ with $n = 10K$ in the fully continuous case. Right: empirical standard errors of $n^{1/2}[\theta_n^*(z) - \theta_0(z)]$ divided by the limit theory-based counterparts for $z \in [0, 2]$.

The left panel of Figure 5 displays the distribution of $\theta_n^*(z) - \theta_0(z)$ for $z \in [0, 2]$, and demonstrates that the sampling distribution of θ_n^* is approximately centered around $\theta_0(z)$ in large samples for $z > 0$. The right panel of Figure 5 displays the ratio of the empirical standard deviation of $n^{1/2}[\theta_n^*(z) - \theta_0(z)]$ to the standard deviation based on the asymptotic theory, and demonstrates that the empirical standard deviation of $\theta_n^*(z)$ approaches the standard deviation defined by the limit theory as the sample size grows.

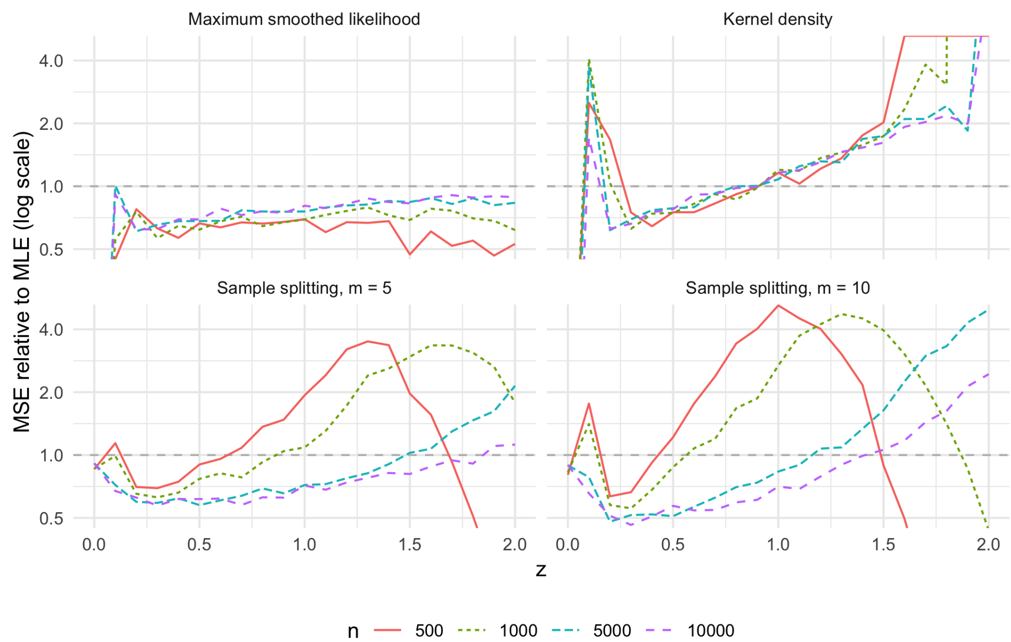

Figure 6: Relative mean squared errors of the maximum smoothed likelihood estimator, the kernel density estimator, and the sample splitting estimators to the maximum likelihood estimator for $z \in [0, 2]$ and various sample sizes n in the fully continuous case. The maximum likelihood has better mean squared error for y-values greater than one, and the other estimator has better mean squared error for y-values less than one.

Figure 6 displays the ratio of the mean squared errors of maximum smoothed likelihood estimator, the kernel density estimator, and the sample splitting estimators to the maximum likelihood estimator. The maximum smoothed likelihood estimator is more efficient than the maximum likelihood estimator. The kernel density estimator is more efficient for some values of z, but less efficient for others. In large enough samples, the sample splitting estimators are more efficient than the maximum likelihood estimator, but in smaller samples, they are less efficient for some values of z. The sample size required for improvement grows with m, as does the gain in asymptotic efficiency.

Finally, Figure 7 shows the empirical coverage of 95% confidence intervals for $\theta_0(z)$ constructed using Wald-type confidence intervals with a plug-in standard error according to the results presented in Section 4.2 of the main text, the inverted likelihood ratio test approach of Banerjee and Wellner (2001), and the sample splitting approach of Banerjee et al. (2019) described in the main text. The plug-in method is conservative in large enough samples due to the difficulty of accurately estimating the derivative of θ_0. The likelihood ratio method provides slightly conservative coverage at all sample sizes. The sample splitting method has excellent coverage for $m = 5$, but requires larger samples to have good coverage for $m = 10$.

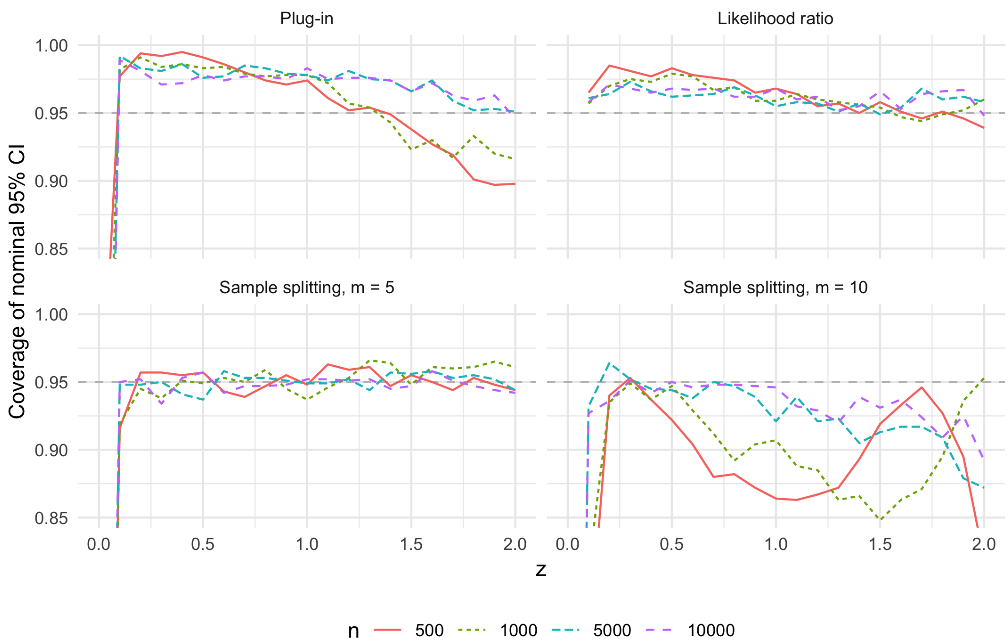

Figure 7: Coverage of 95% CIs in the fully continuous case for $z \in (0, 2]$, various sample sizes n, and four methods: the plug-in method (upper left), the inverted likelihood ratio tests (upper right), and the sample splitting method with $m = 5$ (lower left) and $m = 10$ (lower right).

S5 Additional simulations: flat case with jumps

Here we present results from a numerical study of the properties of the various estimators in the case where F_0 and G_0 are mixed distributions, and $\theta_0 = dF_0/dG_0$ is discontinuous. We set $F_0 := (2/3)F_0^c + (1/3)\delta_0$, where F_0^c is the uniform distribution on $[0, 1]$ and δ_0 is a discrete distribution with mass 1/6 at 0, 1/3 at 1/2, and 1/2 at 1. We set $G_0 := (2/3)F_0^c + (1/3)\gamma_0$, where γ_0 is a discrete distribution with mass 1/3 each at 0, 1/2, and 1. We set π_0 to 0.4.

With these definitions, we have $\theta_0(x) = 1/2$ for $x = 0$, $\theta_0(x) = 1$ for $x \in (0, 1)$, and $\theta_0(x) = 3/2$ for $x = 1$. Hence, θ_0 has jumps at the extremal mass points $x = 0$ and $x = 1$, and is flat between these mass points. Therefore, our large-sample theory does not cover this case for two reasons: because θ_0 is flat in the interior, and because it is discontinuous at the boundaries.

We simulated 1000 datasets each for $n \in \{500, 1000, 5000, 10000\}$ and estimated the maximum likelihood estimator, the maximum smoothed likelihood estimator of Yu et al. (2017), the non-monotone estimator based on kernel density estimates for each $z \in \{0, 0.1, \ldots, 1.9, 2\}$, and the sample splitting estimator with $m \in \{5, 10\}$ (Banerjee et al., 2019). We constructed confidence intervals at each z using the transformed plug-in and likelihood ratio-based methods described in Section 4.2 of the main text. We were unable to use the plug-in method of constructing confidence intervals because it failed in this case due to the difficulty of estimating the derivative of a flat function.

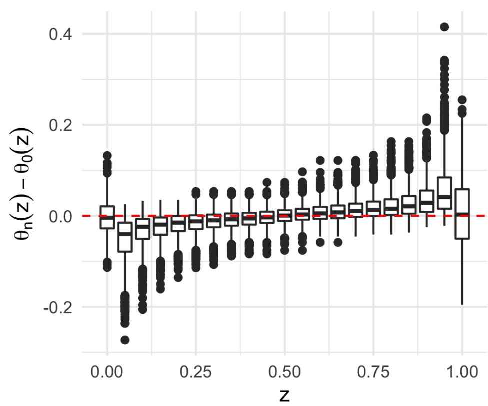

Figure 8: Boxplots of $\theta_n^*(z) - \theta_0(z)$ with $n = 10K$ in the flat case with jumps for $z \in [0, 1]$.

Figure 8 displays the distribution of $\theta_n^*(z) - \theta_0(z)$ for $z \in [0, 1]$. The pattern is quite interesting. For $z \in \{0, 1\}$, the estimator appears to be centered around the truth. This suggests that the estimator may be consistent at mass points even if the function is discontinuous at these points or these points lie on the boundary of the domain. However, for $z \in (0, 0.25)$, the distribution of $\theta_n^*(z)$ is biased downward, and for $z \in (0.75, 1)$, the distribution is biased upward. This is likely due to the discontinuity of θ_0 at 0 and 1: although the estimator is consistent for any $z \in (0, 1)$, in any finite sample the estimator is flat in a region of the discontinuity, which biases the finite-sample distribution of the estimator near these discontinuities. We will see below that this also makes inference in these areas challenging.

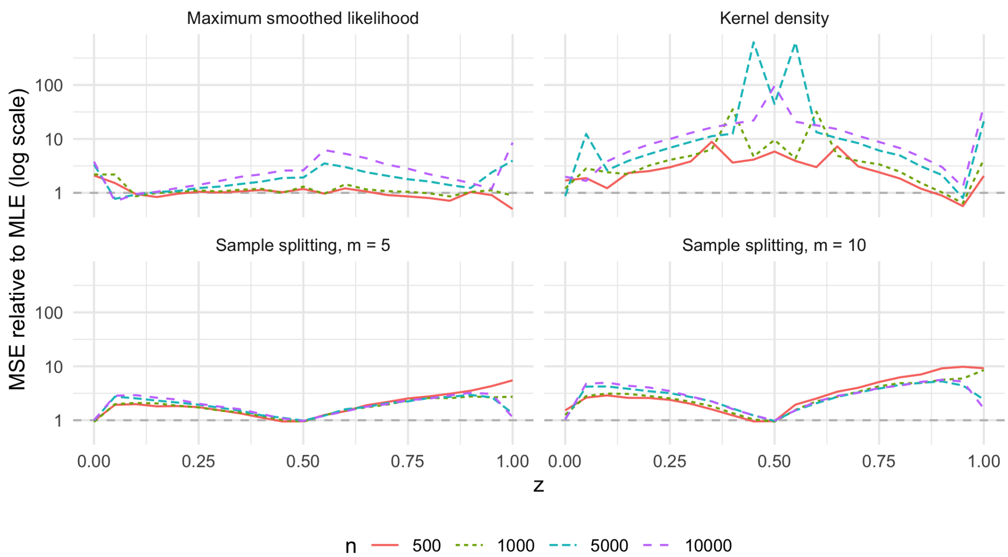

Figure 9: Relative mean squared errors of the maximum smoothed likelihood estimator, the kernel density estimator, and the sample splitting estimators to the maximum likelihood estimator for $z \in [0, 1]$ and various sample sizes n in the flat case with jumps.
.

Figure 9 displays the ratio of the mean squared errors of maximum smoothed likelihood estimator, the kernel density estimator, and the sample splitting estimators to the maximum likelihood estimator. The maximum smoothed likelihood estimator is comparable to the maximum likelihood estimator for $n \in \{500, 1000\}$, but is less efficient for most z in larger samples. This is especially true for $z \in \{0, 1\}$, where the maximum likelihood estimator appears to benefit from the mass points. The kernel density estimator is less efficient, and in large samples much less efficient, for all z except those very close to 0 and 1. Somewhat surprisingly, the sample splitting estimators are less efficient than the maximum likelihood estimator except for z near the mass points. This is likely due to the fact that the sample splitting estimator inherit the bias of the maximum likelihood estimator at a smaller sample size, and the maximum likelihood estimator is biased near the points of discontinuity.

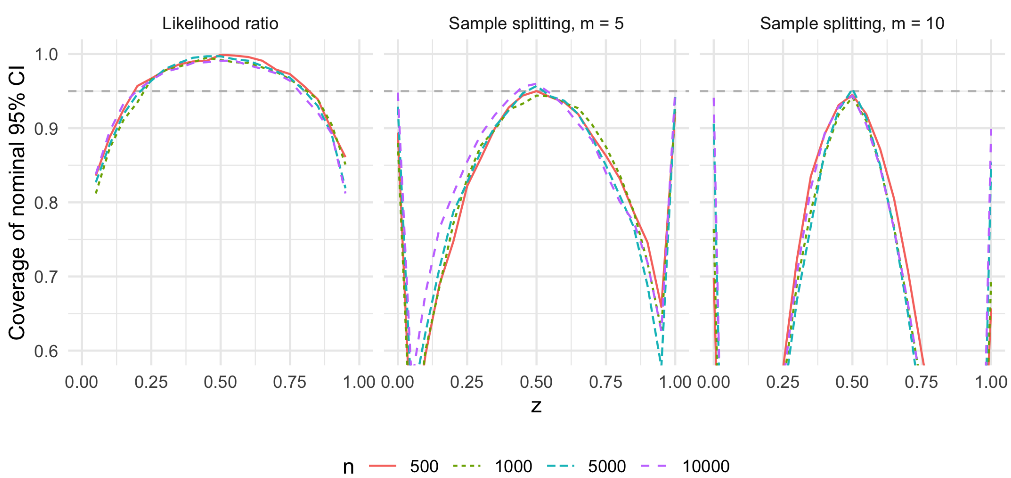

Figure 10: Coverage of 95% CIs in the flat case with jumps for $z \in [0, 1]$, various sample sizes n, and three methods: the inverted likelihood ratio tests (left) and the sample splitting method with $m = 5$ (middle) and $m = 10$ (right).

Finally, Figure 10 shows the empirical coverage of 95% confidence intervals for $\theta_0(z)$ constructed using the inverted likelihood ratio test approach of Banerjee and Wellner (2001) and the sample splitting approach of Banerjee et al. (2019) described in the main text. None of the methods do well near

$z \in \{0, 1\}$ due to the bias of the estimators in these regions. The likelihood ratio method provides conservative coverage at all sample sizes for z near $1/2$, which is because it relies on limit theory that only holds when θ_0 is strictly increasing. The sample splitting method has good coverage for z at the mass points $\{0, 1/2, 1\}$, but poor coverage otherwise.

S6 Additional data analysis results

Figure 11 displays the empirical and likelihood ratio order maximum likelihood cumulative distribution function estimates of C-reactive protein for patients with bacterial infections and those without. Figure 12 displays the empirical and likelihood ratio order maximum likelihood ordinal dominance curve estimates for C-reactive protein.

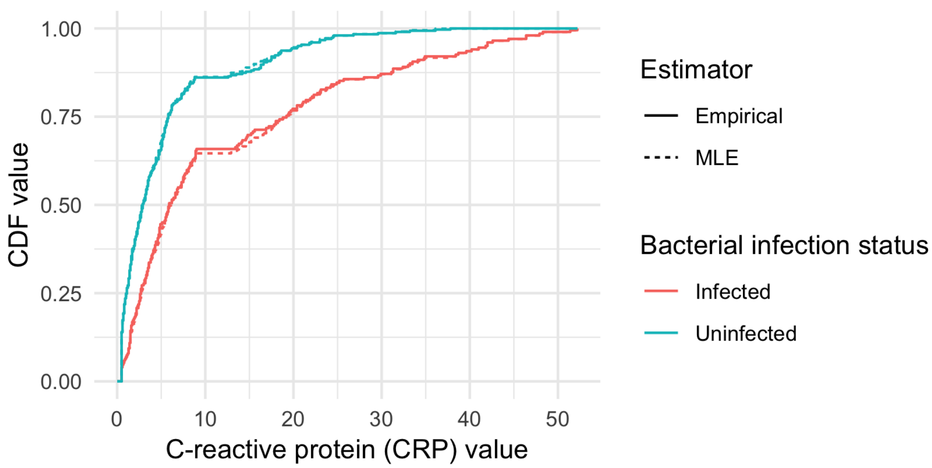

Figure 11: Estimated cumulative distribution functions of C-reactive protein value among patients with bacterial infections and those without. Both the empirical distribution functions and the maximum likelihood estimators under the likelihood ratio order are shown.

Bibliography

Banerjee, M., Durot, C., and Sen, B. (2019). Divide and conquer in nonstandard problems and the super-efficiency phenomenon. *Ann. Statist.*,

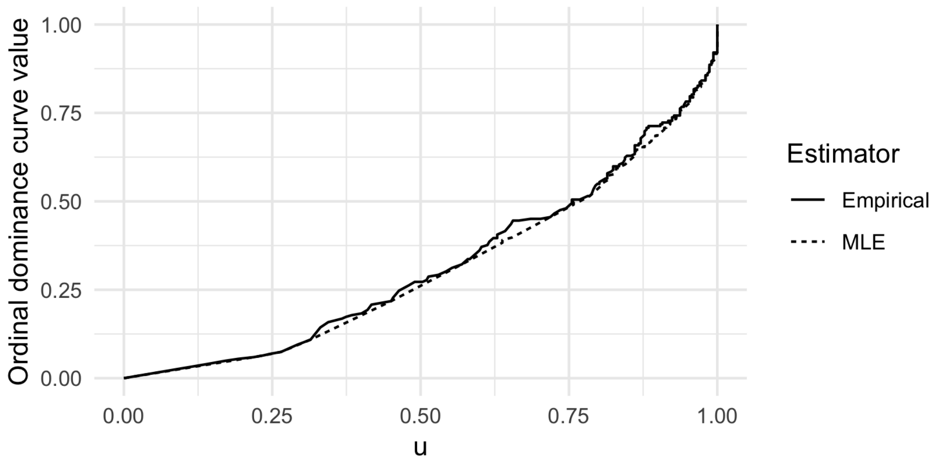

Figure 12: Estimated ordinal dominance curve for C-reactive protein. Both the empirical distribution functions and the maximum likelihood estimators under the likelihood ratio order are shown.

47(2):720–757.

Banerjee, M. and Wellner, J. A. (2001). Likelihood ratio tests for monotone functions. *Ann. Statist.*, 29(6):1699–1731.

Beare, B. K. and Fang, Z. (2017). Weak convergence of the least concave majorant of estimators for a concave distribution function. *Electron. J. Statist.*, 11(2):3841–3870.

Dykstra, R., Kochar, S., and Robertson, T. (1995). Inference for likelihood ratio ordering in the two-sample problem. *J. Amer. Statist. Assoc.*, 90(431):1034–1040.

Groeneboom, P. and Jongbloed, G. (2014). *Nonparametric estimation un-*

der shape constraints. Cambridge University Press.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag New York.

Westling, T. and Carone, M. (2020). A unified study of nonparametric inference for monotone functions. *Ann. Statist.*, 48(2):1001–1024.

Yu, T., Li, P., and Qin, J. (2017). Density estimation in the two-sample problem with likelihood ratio ordering. *Biometrika*, 104(1):141–152.