Statistica Sinica Preprint No: SS-2020-0207	
Title	Nonparametric Maximum Likelihood Estimation Under a Likelihood Ratio Order
Manuscript ID	SS-2020-0207
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202020.0207
Complete List of Authors	Ted Westling, Kevin J. Downes and Dylan Small
Corresponding Author	Ted Westling
E-mail	twestling@math.umass.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

Statistica Sinica

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

Abstract: Comparing two univariate distributions based on independent samples from them is a fundamental problem in statistics, with applications in a variety of scientific disciplines. In many situations, we might hypothesize that the two distributions are stochastically ordered, meaning that samples from one distribution tend to be larger than those from the other. One type of stochastic order is the likelihood ratio order, in which the ratio of the density functions of the two distributions is monotone nondecreasing. In this article, we derive and study the nonparametric maximum likelihood estimator of the individual distribution functions and the ratio of their densities under the likelihood ratio order. Our work applies to discrete distributions, continuous distributions, and mixed continuous-discrete distributions. We demonstrate convergence in distribution of the estimator in certain cases, and illustrate our results using numerical experiments and an analysis of a biomarker for predicting bacterial infection in children with systemic inflammatory response syndrome.

Key words and phrases: Biomarker evaluation; Density ratio; Monotonicity constraint; Odds ratio; Ordinal dominance curve; Shape-constrained inference.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

1. Introduction

Comparing the distributions of two independent samples is a fundamental problem in statistics. Suppose that $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ are independent real-valued samples with distribution functions F_0 and G_0, respectively. In many situations, we might hypothesize that F_0 and G_0 are *stochastically ordered*, meaning intuitively that samples from F_0 tend to be larger than those from G_0. A particular type of stochastic order that arises in many applications is the *likelihood ratio order*. Specifically, G_0 and F_0 satisfy a likelihood ratio order if the density ratio f_0/g_0 is monotone nondecreasing over the support $\mathcal{G}_0$ of G_0, where $f_0 := dF_0/d\eta$ and $g_0 := dG_0/d\eta$, for some dominating measure η. For this reason, the likelihood ratio order is also called a *density ratio order*.

A likelihood ratio order can arise for a variety of scientific reasons (Beare and Moon, 2015; Roosen and Hennessy, 2004; Dykstra et al., 1995; Yu et al., 2017). In the biomedical sciences and elsewhere, the ratio of two density functions is an object of interest for describing the relative likelihood of a binary status indicator conditional on a covariate. If D is a binary random variable, Z is a scalar random variable, $F_0(z) = P(Z \leq z \mid D = 1)$,

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

$G_0(z) = P(Z \leq z \mid D = 0)$, and $H_0(z) := P(Z \leq z)$, then

$$\frac{f_0(z)}{g_0(z)} = \frac{[dF_0/dH_0](z)}{[dG_0/dH_0](z)} = \frac{P(D = 1 \mid Z = z)/P(D = 1)}{P(D = 0 \mid Z = z)/P(D = 0)} \,. \tag{1.1}$$

Therefore, the density ratio in this context may be interpreted as the relative odds of $D = 1$ given $Z = z$ to the overall odds of $D = 1$. Because the transformation $x \mapsto x/(1 - x)$ is strictly increasing, monotonicity of the density ratio is equivalent to monotonicity of $z \mapsto P(D = 1 \mid Z = z)$. One specific situation in which the representation given in (1.1) is of scientific interest is biomarker evaluation, wherein D represents infection status and Z represents the value of a biomarker. Equation (1.1) implies that the ratio of the densities of biomarker values among infected patients to that among uninfected patients can be interpreted as the odds ratio of infection given a biomarker level relative to the overall odds of infection. Monotonicity of the density ratio corresponds to the assumption that the conditional probability of infection given a biomarker level increases with the biomarker level, which is reasonable if the biomarker can predict the disease.

In this article, we derive the nonparametric maximum likelihood estimators of F_0, G_0, and $\theta_0 = f_0/g_0$ under the likelihood ratio order restriction, and derive certain asymptotic properties of these estimators, including consistency and convergence in distribution. In particular, we use a connection between the estimation of θ_0 and the classical isotonic regression problem

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

with a binary outcome, which both simplifies the derivation of large-sample results and suggests that existing inference methods for the isotonic regression problem can be used to perform inference for θ_0 as well. Our results generalize those of Dykstra et al. (1995), who derived the maximum likelihood estimator of F_0 and G_0 under a likelihood ratio order in the special case where F_0 and G_0 are discrete distributions. We illustrate our results using numerical experiments and an analysis of a biomarker for predicting bacterial infection in children with systemic inflammatory response syndrome.

Recently, Yu et al. (2017) estimated a monotone density ratio and the individual density functions by maximizing a smoothed likelihood function, and demonstrated certain asymptotic properties of their estimator. Yu et al. (2017) considered maximizing a smoothed likelihood rather than maximizing the likelihood directly because the maximum likelihood estimator of the individual densities does not exist. In contrast, we use a definition of the likelihood ratio ordered model based on convexity of the ordinal dominance curve to show that a well-defined nonparametric maximum likelihood estimator of the monotone density ratio function and the individual distribution functions (rather than the density functions) does exist. Furthermore, unlike the smoothed estimator, the derivation of the maximum

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

likelihood estimator does not require the selection of a bandwidth or any other tuning parameter, and does not rely on the existence of Lebesgue density functions.

Additional relevant references include Lehmann and Rojo (1992) and Shaked and Shanthikumar (2007), which contain additional examples and details on stochastic orders, Carolan and Tebbs (2005) and Beare and Moon (2015), who studied tests of the likelihood ratio order, and Rojo and Samaniego (1991), Rojo and Samaniego (1993), Mukerjee (1996), Arcones and Samaniego (2000), Davidov and Herman (2012), and Tang et al. (2017), who considered testing and estimation under other stochastic orders.

2. Likelihood ratio orders

We observe two independent real-valued samples $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ with distribution functions F_0 and G_0, respectively. We define $\mathcal{F}_0$ as the support of F_0, and $\mathcal{G}_0$ as the support of G_0. We denote $n := n_1 + n_2$, and F_n and G_n as the empirical distribution functions of $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$, respectively. We define $x_1 < \cdots < x_{m_1}$ as the unique values of $X_1, \ldots, X_{n_1}$, $y_1 < \cdots < y_{m_2}$ as the unique values of $Y_1, \ldots, Y_{n_2}$, and $z_1 < z_2 < \cdots < z_m$ as the unique values of $(X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2})$. Throughout, we assume that n is fixed, but that n_1 is drawn from a

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

Binomial(n, π_0) distribution, for some $\pi_0 \in (0, 1)$.

We let $\mathcal{D}$ be the space of distribution functions on $\mathbb{R}$, that is, all nondecreasing, cádlàg functions H such that $\lim_{x\to-\infty} H(x) = 0$ and $\lim_{x\to\infty} H(x) = 1$. For any nondecreasing function $h : \mathbb{R} \to \mathbb{R}$, we define its *generalized-inverse* h^- pointwise as $h^-(u) := \inf\{x : h(x) \geq u\}$. When $h \in \mathcal{D}$, h^- is called the *quantile function* of h. For any interval $I \subseteq \mathbb{R}$ and any function $h : I \to \mathbb{R}$, we define the *greatest convex minorant* (GCM) of h on I, denoted $\mathrm{GCM}_I(h) : I \to \overline{\mathbb{R}}$, for $\overline{\mathbb{R}}$ the extended real line, as the pointwise supremum of all convex functions on I bounded above by h. The least concave majorant operator is defined analogously. We say a function H is *convex over a set* $\mathcal{S} \subseteq \mathbb{R}$ if for every $x, y \in \mathcal{S}$ and $\lambda \in [0, 1]$ such that $\lambda x + (1 - \lambda)y \in \mathcal{S}$, $H(\lambda x + (1 - \lambda)y) \leq \lambda H(x) + (1 - \lambda)H(y)$. We also define ∂_- as the left derivative operator for a left differentiable function and $\mathrm{Im}(h) := \{h(x) : x \in \mathcal{S}\}$ as the image of a function h defined on a domain $\mathcal{S}$.

The unrestricted nonparametric model for the pair (F, G) of distribution functions of the observed data is $\mathcal{M}_{NP} := \mathcal{D}^2$. As mentioned in the introduction, the likelihood ratio order can be defined as the ratio of the density functions f_0 and g_0 of F_0 and G_0, respectively, with respect to some dominating measure η being nondecreasing. By varying the domi-

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

nating measure, both discrete and continuous distributions can be handled this way. However, as noted by Yu et al. (2017), this definition does not lend itself to the derivation of a maximum likelihood estimator, because the likelihood defined through the densities can be made arbitrarily large. Instead, other authors have defined the likelihood ratio order as convexity of the *ordinal dominance curve*, defined as $t \mapsto R_{F,G}(t) := F \circ G^{-}(t)$ for $t \in [0,1]$ (Bamber, 1975; Hsieh and Turnbull, 1996). Lehmann and Rojo (1992) demonstrated the equivalence of this definition to that using the density functions in the special case that F and G are strictly increasing and continuous on their supports, which were assumed to be intervals. As an alternative, Shaked and Shanthikumar (2007) defined the likelihood ratio order as $F(A)G(B) \leq F(B)G(A)$ for all measurable sets $A, B \subseteq \mathbb{R}$ with $A \leq B$, where $F(A) := \int_A dF$ (with some abuse of notation) and $A \leq B$ means that $a \leq b$ for all $a \in A$ and $b \in B$.

In Theorem 1 below, we consolidate and generalize existing results connecting these different definitions of the likelihood ratio order.

Theorem 1. *If $F \ll G$ and $\nu := dF/dG$ is continuous on the support $\mathcal{G}$ of G, then (1) the following are equivalent: $R_{F,G}$ is convex on $\mathrm{Im}(G)$, ν is nondecreasing on $\mathcal{G}$, and $F(A)G(B) \leq F(B)G(A)$ for all measurable sets $A \leq B$; and (2) if ν is nondecreasing on $\mathcal{G}$, then $\nu(x) = \partial_{-}\mathrm{GCM}_{[0,1]}(R_{F,G}) \circ$*

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

$G(x)$ *for all* $x \in \mathcal{G}$.

To the best of our knowledge, Theorem 1 is the most general result to date connecting the three definitions of the likelihood ratio ordered model. Note that the three definitions may not be equivalent when F is not dominated by G or when ν is not continuous. For instance, in the proof of Theorem 1 part (1), we use only the assumption that ν is continuous on $\mathcal{G}$ to show that $R_{F,G}$ is convex on $\text{Im}(G)$ implies that ν is nondecreasing. However, we show that if ν is nondecreasing, then $F(A)G(B) \leq F(B)G(A)$ for all $A \leq B$ (the definition used in Shaked and Shanthikumar, 2007), regardless of whether ν is continuous. Additionally, we show that $F(A)G(B) \leq F(B)G(A)$ for all $A = (a_1, b_1] \leq B = (b_1, b_2]$ implies that $R_{F,G}$ is convex on $\text{Im}(G)$, regardless of whether $F \ll G$ or ν is continuous. However, to show the converse, we use $F \ll G$. For a simple counterexample when F is not dominated by G, consider $F(\{a\}) = 1$ and $G(\{b\}) = 1$, where $a < b$. Then $R_{F,G}(u) = I(u > 0)$ for $u \in [0, 1]$, which is convex on $\text{Im}(G) = \{0, 1\}$, but $1 = F(A)G(B) > F(B)G(A) = 0$ for $A = \{a\} \leq \{b\} = B$. Finally, when $F \ll G$, but dF/dG is not continuous on $\mathcal{G}$, it is unclear whether $R_{F,G}$ being convex on $\text{Im}(G)$ implies that $F(A)G(B) \leq F(B)G(A)$ for *all* measurable sets $A \leq B$, or even all such Borel sets.

Throughout the remainder of the article, we say $(F, G) \in \mathcal{M}_{NP}$ satisfy

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

a likelihood ratio order, and write $G \leq_{LR} F$ if $R_{F,G}$ is convex on $\mathrm{Im}(G)$. We then define the likelihood ratio ordered model $\mathcal{M}_{LR}$ as all $(F, G) \in \mathcal{M}_{NP}$ such that $G \leq_{LR} F$. For any $(F, G) \in \mathcal{M}_{NP}$, we further define $\theta : \mathcal{M}_{NP} \to \boldsymbol{\Theta}$ as $\theta_{F,G} := \partial_- \mathrm{GCM}_{[0,1]}(R_{F,G}) \circ G$, where $\boldsymbol{\Theta}$ is defined as the set of nonnegative, nondecreasing functions on $\mathbb{R}$. Note that this definition allows for the possibility that F is not dominated by G, but by Theorem 1, for all $(F, G) \in \mathcal{M}_{LR}$ such that $F \ll G$ and dF/dG is continuous on $\mathcal{G}$, $\theta_{F,G} = dF/dG$ on $\mathcal{G}$. We define $\theta_0 := \theta_{F_0,G_0}$.

In the context of the likelihood ratio order, many existing works either assume that F_0 and G_0 are discrete (e.g., Dykstra et al., 1995), or that F_0 and G_0 are continuous (e.g., Lehmann and Rojo, 1992; Yu et al., 2017). In the discrete setting, if F_0 and G_0 are discrete distributions with common support and mass functions ΔF_0 and ΔG_0, respectively, such that $(F_0, G_0) \in \mathcal{M}_{LR}$, then $\theta_0 = \Delta F_0 / \Delta G_0$ on $\mathcal{G}_0$. Alternatively, if F_0 and G_0 both possess Lebesgue density functions f_0 and g_0, respectively, and $(F_0, G_0) \in \mathcal{M}_{LR}$, then $\theta_0 = f_0/g_0$ on $\mathcal{G}_0$. However, for the purpose of deriving a maximum likelihood estimator, we show that these two cases do not need to be treated separately. Furthermore, in some applied settings, F_0 and G_0 are neither discrete nor continuous, but rather a mixture of discrete and continuous components, and we derive results that apply in these

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

situations as well. For instance, exposures that are bounded below may have positive mass at their lower boundary, and be continuous thereafter. Many biomarkers exhibit this property. Similarly, some measurements are "clumpy," exhibiting positive mass at integers or other "round" numbers owing to the measurement process, but also possessing positive Lebesgue density between such points. In all cases, θ_0 has a meaningful interpretation as the ratio of the conditional odds of a sample being from the distribution F_0 to the unconditional odds of a sample being from F_0.

3. Estimation under a likelihood ratio order

3.1 Maximum likelihood estimator

The pair (F_0, G_0) determines the joint distribution of the observed data. Defining the nonparametric likelihood of the observed data as $L_n(F, G) := \prod_{i=1}^{n_1} [F(X_i) - F(X_i-)] \prod_{j=1}^{n_2} [G(Y_j) - G(Y_j-)]$, the nonparametric (that is, in the model $\mathcal{M}_{NP}$) maximum likelihood estimator of (F_0, G_0) is (F_n, G_n), for F_n and G_n the empirical distribution functions of $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$, respectively. This suggests taking as an estimator of θ_0 the plug-in estimator $\theta_n := \theta_{F_n,G_n} = \partial_- \mathrm{GCM}_{[0,1]}(F_n \circ G_n^-) \circ G_n$. The function $F_n \circ G_n^-$ is known as the *empirical ordinal dominance curve*, and its properties were studied by Hsieh and Turnbull (1996).

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

In this section, we demonstrate, among other results, that θ_n is the maximum likelihood estimator of θ_0 in the likelihood ratio ordered model $\mathcal{M}_{LR}$. A maximum likelihood estimator of (F_0, G_0) in $\mathcal{M}_{LR}$ is defined as $(F_n^*, G_n^*) \in \text{argmax}_{(F,G)\in\mathcal{M}_{LR}} L_n(F, G)$, and a maximum likelihood estimator of θ_0 is defined as $\theta_n^* := \theta_{F_n^*, G_n^*}$.

We define $H_n(z) := \pi_n F_n(z) + (1-\pi_n) G_n(z)$ as the empirical distribution of the combined sample $X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2}$, and $h_k := H_n(y_k)$ for $k = 1, \ldots, m_2$. Our first result characterizes (F_n^*, G_n^*).

Theorem 2. *Let A_k^* be the value at h_k of the GCM over $[0, h_{m_2}]$ of $\{(h_k, F_n(y_k)) : k = 0, \ldots, m_2\}$, and let B_k^* be the value at h_k of the LCM over $[0, h_{m_2}]$ of $\{(h_k, G_n(y_k)) : k = 0, \ldots, m_2\}$. Then G_n^* is a right-continuous step function with jumps at $y_1, \ldots, y_{m_2}$ with $G_n^*(y_k) = B_k^*$, and F_n^* is given by a right-continuous step function with jumps at $z_1, \ldots, z_m$, where $F_n^*(y_k) = A_k^*$, and for any x_i such that $y_{j-1} < x_i \le y_j$, where $y_0 := -\infty$, the mass of F_n^* at x_i is given by*

$$F_n^*(x_i) - F_n^*(x_i-) = [F_n^*(y_j) - F_n^*(y_{j-1})] \, \frac{F_n(x_i) - F_n(x_i-)}{F_n(y_j) - F_n(y_{j-1})} \, .$$

For any x_i such that $y_{m_2} < x_i$, the mass of F_n^* at x_i is given by

$$F_n^*(x_i) - F_n^*(x_i-) = [1 - F_n^*(y_{m_2})] \, \frac{F_n(x_i) - F_n(x_i-)}{1 - F_n(y_{m_2})} \, .$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

Note that $F_n^*(y_k) = \mathrm{GCM}_{[0,h_{m_2}]}(F_n \circ H_n^-)(H_n(y_k))$ *and*

$$G_n^*(y_k) = \mathrm{LCM}_{[0,h_{m_2}]}(G_n \circ H_n^-)(H_n(y_k)).$$

A proof of Theorem 2 (and all other theorems) is provided in the Supplementary Material. Note that F_n^* necessarily has jumps at all x_i and at all y_j such that $y_j \geq x_1$, and G_n^* has jumps at all y_j. Note too that if there are j such that no $x_i \in (y_j, y_{j+1}]$, but $F_n^*(y_j) > F_n^*(y_{j-1})$, then there are infinitely many maximizers F_n^*, because any F_n^* that assigns mass $F_n^*(y_j) - F_n^*(y_{j-1})$ to the interval $(y_j, y_{j+1}]$ yields the same likelihood and satisfies the constraints. In these cases, for the sake of uniqueness, we put mass $F_n^*(y_j) - F_n^*(y_{j-1})$ at the point y_{j+1}.

Theorem 2 implies the following result characterizing θ_n^*.

Corollary 1. *The points* $\{(G_n^*(y_k), F_n^*(y_k)) : k = 1, \ldots, m_2\}$ *lie on the GCM over* $[0, 1]$ *of the empirical ordinal dominance curve*

$$\{(G_n(y_j), F_n(y_j)) : k = 0, \ldots, m_2\},$$

where $y_0 := -\infty$. *Specifically, if* $\{(h_{j_k}, F_n(y_{j_k})) : k = 0, \ldots, K\}$ *are the vertices of the GCM of* $\{(h_k, F_n(y_k)) : k = 0, \ldots, m_2\}$, *then* $(G_n(y_{j_k}), F_n(y_{j_k})) : k = 0, \ldots, K\}$ *are the vertices of the GCM of the empirical ordinal dominance curve. Therefore,* $\theta_n^* := \theta_{F_n^*, G_n^*}$ *is equal to* $\theta_n := \theta_{F_n, G_n}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

Theorem 2 bears a resemblance to, but differs from, Theorem 2.1 of Dykstra et al. (1995), which characterizes the maximum likelihood estimator under a likelihood ratio order in the discrete case. Here, we perform the maximization over all pairs of univariate distribution functions (F, G) such that $R_{F,G} = F \circ G^{-}$ is convex on the support of G. In contrast, Theorem 2.1 of Dykstra et al. (1995) performed the maximization over (F, G) with support contained in $\{z_1, \ldots, z_m\}$ and such that $[\Delta F(z_j)]/[\Delta G(z_j)]$ is nondecreasing. The first set is strictly larger than the second, which results in possibly different maximum likelihood estimators. In particular, our maximum likelihood estimator G_n^* is supported only on $y_1, \ldots, y_{m_2}$, whereas the maximum likelihood estimator of G_0 derived by Dykstra et al. (1995) may have support on x_j that are not equal to any $y_1, \ldots, y_{m_2}$. This difference makes sense in the context of our respective problem formulations: Dykstra et al. (1995) assumed that the supports of F_0 and G_0 are subsets of $\{z_1, \ldots, z_m\}$, whereas we do not assume the supports are known *a priori*. In the Supplementary Material, we illustrate the use of Theorem 2 using hypothetical data in which our maximum likelihood estimators F_n^* and G_n^* differ from those of Dykstra et al. (1995).

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

3.2 Representation as a transformation of isotonic regression

Dykstra et al. (1995) and Carolan and Tebbs (2005) provided representations of the maximum likelihood estimators of F_0 and G_0 in terms of an isotonic regression in the discrete and continuous cases, respectively. Here, we show that θ_n^* can also be represented as a transformation of an isotonic regression, which aids in deriving its asymptotic properties. We let $D_1, \ldots, D_n$ be independent Bernoulli random variables with common probability π_0 and such that $n_1 = \sum_{i=1}^n D_i$. Letting $j_1, \ldots, j_{n_1}$ be the indices such that $D_{j_i} = 1$ for each i, we then define $Z_{j_i} := X_i$ for each $i = 1, \ldots, n_1$. Similarly, letting $k_1, \ldots, k_{n_2}$ be the indices such that $D_{k_i} = 0$ for each i, we define $Z_{k_i} := Y_i$ for each $i = 1, \ldots, n_2$. Defining the data unit $\mathbf{O}_i := (Z_i, D_i)$, observing the independent samples $X_1, \ldots, X_{n_1}$ from F_0 and $Y_1, \ldots, Y_{n_2}$ from G_0 is then equivalent to observing independent observations $\mathbf{O}_1, \ldots, \mathbf{O}_n$ from P_0, where P_0 satisfies

$$P_0(Z \leq z, D = d) = d\pi_0 F_0(z) + (1-d)(1-\pi_0)G_0(z) \ .$$

Thus, $Z_1, \ldots, Z_n$ represent the pooled values of $X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2}$, and each D_i represents an indicator that Z_i corresponds to a sample from F_0. Furthermore, $F_0(z) = P_0(Z \leq z \mid D = 1)$, $G_0(z) = P_0(Z \leq z \mid D = 0)$, and $\pi_0 := P_0(D = 1)$. Estimating θ_0 given the independent samples $X_1, \ldots, X_{n_1}$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

and $Y_1, \ldots, Y_{n_2}$ is therefore equivalent to estimating θ_0 given independent observations $\mathbf{O}_1, \ldots, \mathbf{O}_n$ from P_0, where $n_1 := \sum_{i=1}^{n} D_i$.

The benefit to the above reframing of the problem is that θ_0, F_0, and G_0 can then be written as transformations of P_0. First, we have that $\theta_0(z) = T(\mu_0(z))/T(\pi_0)$, where $\mu_0(z) := P_0(D = 1 \mid Z = z)$ and $T : [0, 1) \to \mathbb{R}^+$ is the odds transformation, defined as $T(\mu) := \mu/(1 - \mu)$. Because T is strictly increasing, θ_0 is monotone if and only if μ_0 is monotone. Because the maximum likelihood estimator of μ_0 under the assumption that μ_0 is nondecreasing is given by the isotonic regression μ_n^* of $D_1, \ldots, D_n$ on $Z_1, \ldots, Z_n$, and the maximum likelihood estimator of π_0 is given by π_n, the maximum likelihood estimator of $\theta_0(z)$ is then given by $T(\mu_n^*(z))/T(\pi_n)$. It is straightforward to see that this form of the maximum likelihood estimator is equivalent to the forms given above. In the next section, we use this form of θ_n^* to derive its asymptotic properties and to construct asymptotic confidence intervals.

4. Asymptotic results

4.1 Discrete distributions

We first consider the situation where G_0 has finite support $\mathcal{G}_0$ and θ_0 is strictly increasing on $\mathcal{G}_0$. The next result demonstrates that in this case,

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

F_n^* and G_n^* are asymptotically equivalent to F_n and G_n, respectively, and θ_n^* is asymptotically equivalent to the ratio of the empirical masses on the support of G_0.

Theorem 3 (Discrete distributions)**.** *Suppose that the support $\mathcal{G}$ of G_0 is a finite set $\{y_1 < y_2 < \cdots < y_{m_2}\}$ and that $[F_0(y_j) - F_0(y_{j-1})]/\Delta G_0(y_j) < [F_0(y_{j+1}) - F_0(y_j)]/\Delta G_0(y_{j+1})$ for $j = 1, \ldots, m_2 - 1$, where $y_0 := -\infty$. Then, $F_n^* = F_n$ and $G_n^* = G_n$ with probability tending to one, so that with probability tending to one, θ_n^* is a left-continuous step function with jumps at $y_1, \ldots, y_{m_2-1}$ and $\theta_n^*(y_j) = [F_n(y_j) - F_n(y_{j-1})]/\Delta G_n(y_j)$ and $\theta_n^*(z) = 0$ for $z < y_1$. As a result, $n^{1/2}[\theta_n^*(y_j) - \theta_0(y_j)] \xrightarrow{\mathrm{d}} N(0, \sigma_0^2(y_j))$ for*

$$\sigma_0^2(y_j) := \theta_0(y_j)\frac{\pi_0 F_{0,j} + (1-\pi_0)\Delta G_0(y_j) - F_{0,j}\Delta G_0(y_j)}{\pi_0(1-\pi_0)[\Delta G_0(y_j)]^2},$$

where $F_{0,j} := F_0(y_j) - F_0(y_{j-1})$.

Note that the above result does not require that F_0 be discrete as well, or that it be dominated by G_0. If F_0 is dominated by G_0, then $\theta_0 = \Delta F_0/\Delta G_0$ corresponds to the ratio of the mass functions.

4.2 Continuous distributions

Here, we address the situation where F_0 and G_0 are both absolutely continuous on $\mathcal{G}_0$, and θ_0, which now corresponds to the ratio f_0/g_0 of the density

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

functions, is strictly increasing. We first consider the large-sample behavior of F_n^* and G_n^*.

Theorem 4. *Suppose that G_0 is supported on a bounded interval $[a, b] \subset \mathbb{R}$, that F_0 and G_0 possess continuous density functions f_0 and g_0, respectively, on $[a, b]$ such that f_0/g_0 is strictly increasing on $[a, b]$, and $g_0(z) \geq \kappa > 0$ on $[a, b]$. Then, $\|G_n^* - G_n\|_\infty = o_P(n^{-1/2})$ and $\|F_n^* - F_n\|_\infty = o_P(n^{-1/2})$.*

Theorem 4 demonstrates that when θ_0 is strictly increasing, the maximum likelihood estimators of the individual distribution functions are asymptotically equivalent to the empirical distribution functions at the rate $n^{-1/2}$, and hence possess the same limit distributions as the empirical distribution functions. This result is proved using the functional delta method and the results of Beare and Fang (2017), who demonstrated that the LCM operation is a directionally Hadamard differentiable mapping at any concave function.

We now turn to large-sample results for θ_n^* at points z where both F_0 and G_0 possess Lebesgue density functions f_0 and g_0, respectively. First, consistency of μ_n^* implies consistency of θ_n^*.

Theorem 5 (Consistency)**.** *If f_0 is continuous at x, g_0 is continuous at x, and $g_0(x) > 0$, then $\theta_n^*(x) \xrightarrow{\mathrm{P}} \theta_0(x)$. If f_0 and g_0 are uniformly continuous on $\mathcal{G}_0$, then $\sup_{x \in I} |\theta_n^*(x) - \theta_0(x)| \xrightarrow{\mathrm{P}} 0$ for any strict sub-interval $I \subsetneq \mathcal{G}_0$.*

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

Recall that, at any z such that $h_0 = \pi_0 f_0 + (1-\pi_0) g_0$ is positive and continuous in a neighborhood of z, $\mu_0(z) \in (0,1)$, and μ_0 is continuously differentiable in a neighborhood of z, it holds that

$$n^{1/3}\left[\mu_n^*(z) - \mu_0(z)\right] \xrightarrow{\mathrm{d}} \left\{4\mu_0'(z)\mu_0(z)[1-\mu_0(z)]h_0(z)^{-1}\right\}^{1/3} W \ , \qquad (4.1)$$

where W follows *Chernoff's distribution*, defined as the point of maximum of $Z(u) - u^2$ for Z a two-sided standard Brownian motion originating from zero (Brunk, 1970; Groeneboom and Jongbloed, 2014). We can then use the delta method to see that

$$n^{1/3}\left[\theta_n^*(z) - \theta_0(z)\right] \xrightarrow{\mathrm{d}} T(\pi_0)T'(\mu_0)\left\{4\mu_0'(z)\mu_0(z)[1-\mu_0(z)]h_0(z)^{-1}\right\}^{1/3} W \ .$$

The scale parameter in the above limit distribution is equal to $[4\kappa_0(z)\theta_0'(z)]^{1/3}$ for

$$\kappa_0(z) := \theta_0(z)\frac{\pi_0 f_0(z) + (1-\pi_0)g_0(z)}{\pi_0(1-\pi_0)g_0(z)^2} \ .$$

This yields the following result.

Theorem 6 (Pointwise convergence in distribution)**.** *Suppose that, in a neighborhood of z, θ_0 is continuously differentiable with $\theta_0'(z) > 0$, and f_0 and g_0 are positive and continuous. Then,*

$$n^{1/3}[\theta_n^*(z) - \theta_0(z)] \xrightarrow{\mathrm{d}} [4\kappa_0(z)\theta_0'(z)]^{1/3} W.$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

Theorem 6 reflects certain common tradeoffs of the monotonicity constraint. Theorem 6 indicates that the nonsmoothed estimator converges pointwise at the rate $n^{-1/3}$. In contrast, the smoothed estimator proposed by Yu et al. (2017) converges at the faster rate $n^{-2/5}$, albeit under stronger smoothness assumptions. While Yu et al. (2017) did not propose a method for conducting inference, smoothed estimators typically possess an asymptotic bias that complicates the task of performing valid inference. In contrast, the limit distribution in Theorem 6 has mean zero, which we can use to construct asymptotically valid confidence intervals. Defining $\tau_n(z)$ as an estimator of $\tau_0(z) := \kappa_0(z)\theta_0'(z)$ and q_α as the $1-\alpha/2$ quantile of W, a $100(1-\alpha)\%$ Wald-type confidence interval for $\theta_0(z)$ is given by $\theta_n^*(z) \pm [4\tau_n(z)/n]^{1/3} q_{1-\alpha/2}$. If $\tau_n(z) \xrightarrow{\mathrm{P}} \tau_0(z)$, then this interval has asymptotic coverage of $100(1-\alpha)\%$. The quantiles of W were computed by Groeneboom and Wellner (2001), and in particular $q_{0.975} \approx 0.9982$.

In practice, we recommend using an alternative method to construct confidence intervals for $\theta_0(z)$. We recommend first constructing confidence intervals for $\mu_0(z)$ using either of two existing methods, then transforming these intervals into intervals for $\theta_0(z)$. Specifically, if $[\ell_n(z), u_n(z)]$ represents a $100(1-\alpha)\%$ confidence interval for $\mu_0(z)$, then we take $[T(\ell_n(z))/T(\pi_n)$, $T(u_n(z))/T(\pi_n)]$ as a $100(1-\alpha)\%$ confidence interval for $\theta_0(z)$. Two ex-

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

isting method used to construct $[\ell_n(z), u_n(z)]$ are Wald-type intervals with plug-in estimation of the nuisance parameters and intervals based on likelihood ratio tests. The former intervals are analogous to the Wald-type interval, but based on the limit distribution for $n^{1/3}[\mu_n^*(z) - \mu_0(z)]$ given in (4.1). Alternatively, we can form confidence intervals by inverting likelihood ratio tests, proposed first by Banerjee and Wellner (2001) and studied further by, among others, Banerjee (2007) and Groeneboom and Jongbloed (2015), based on the limiting distribution of twice the log of the ratio of the likelihoods of the maximum likelihood estimator and a suitably constrained maximum likelihood estimator. Because this limiting distribution is pivotal, meaning it does not depend on any unknown features of the true distribution, this approach does not require estimating any unknown nuisance parameters. We therefore expect this method to have better finite-sample properties than intervals based on plug-in estimation of the nuisance parameters.

5. Numerical studies

In the Supplementary Material, we present the results of two simulation studies in the cases where F_0 and G_0 are fully discrete and fully continuous. In short, these studies confirm the validity of our large-sample theory,

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

and demonstrate that the maximum likelihood estimator and various proposed methods of conducting inference perform well in both cases. Here, we present the results of a numerical study illustrating the behavior of θ_n^* when F_0 and G_0 are mixed discrete-continuous distributions. Note that our asymptotic results do not address the behavior of θ_n^* at mass points in mixed discrete-continuous distributions; to the best of our knowledge, no such results yet exist for monotone estimators. We use this numerical study to explore this important case.

We simulated Y as a mixed discrete-continuous random variable with probability 1/9 each of being 0, 0.5, and 1, and probability 2/3 of being from the uniform distribution on $[0, 1]$. We simulated X as a mixed discrete-continuous random variable with probabilities 1/18, 1/9, and 3/18 of being 0, 0.5, and 1, respectively, and probability 2/3 of being from the density function $x \mapsto I_{[0,1]}(x)(0.5 + x)$. We then have that $\theta_0(x) = 0.5 + x$ for $x \in [0, 1]$. We set $\pi_0 := 0.4$. For each combined sample size $n \in \{500, 1K, 5K, 10K\}$, we simulated 1000 data sets, and in each data set we computed the maximum likelihood estimator, maximum smoothed likelihood estimator of Yu et al. (2017), and nonmonotone estimator based on kernel density estimates for each $z \in \{0, 0.05, \ldots, 0.95, 1\}$. We constructed confidence intervals at each z using the transformed plug-in and

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

likelihood ratio-based methods described in Section 4.2. To estimate the scale parameter in the limit distribution of $\mu_n^*(z)$, as defined in equation 4.1, we used the plug-in estimator $\mu_n^*(z)$ for $\mu_0(z)$, and estimated $\mu_0'(z)/h_0(z) = (\mu_0 \circ H_0^{-1})' \circ H_0(z)$ using the derivative of a local linear smoother of $\mu_n^* \circ H_n^-$ evaluated at $H_n(z)$.

In addition to the properties of the estimators listed above, we also investigated the properties of the general sample-splitting procedure proposed by Banerjee et al. (2019). Given a generic monotone estimator γ_n of a monotone function γ_0 such that $n^{1/3}[\gamma_n(z) - \gamma_0(z)] \xrightarrow{\mathrm{d}} G$ for G a mean-zero distribution with finite variance, Banerjee et al. (2019) proposed randomly splitting the sample into m subsets of roughly equal size, computing monotone estimates $\gamma_{n,1}, \ldots, \gamma_{n,m}$ in each subset, then defining $\bar{\gamma}_{n,m}(z) := \frac{1}{m}\sum_{j=1}^m \gamma_{n,j}(z)$. They demonstrated that if $m > 1$ is fixed, then under mild conditions, $\bar{\gamma}_{n,m}(z)$ has a strictly better asymptotic mean squared error than $\gamma_n(z)$, and that for moderate m, $\bar{\gamma}_{n,m}(z) \pm \sigma_{n,m}(z) t_{1-\alpha/2,m-1}/\sqrt{m}$ forms an asymptotic $100(1-\alpha)\%$ confidence interval for $\gamma_0(z)$, where $\sigma_{n,m}^2(z) := \frac{1}{m-1}\sum_{j=1}^m [\gamma_{n,j}(z) - \bar{\gamma}_{n,m}(z)]^2$, and $t_{1-\alpha/2,m-1}$ is the $100(1-\alpha/2)$ quantile of the t-distribution with $m-1$ degrees of freedom. Therefore, $\bar{\gamma}_{n,m}(z)$ is preferable to $\gamma_n(z)$ for two reasons: it has a better asymptotic mean squared error, and asymptotically valid pointwise confidence intervals for γ_0 based

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

on $\bar{\gamma}_{n,m}$ can be formed without estimating any nuisance parameters. They also studied the asymptotic properties of $\bar{\gamma}_{n,m_n}(z)$ when m_n grows with n. In our simulation study, we considered the estimator $\bar{\theta}_{n,m}$ defined as $\bar{\theta}_{n,m}(z) := \frac{1}{m}\sum_{j=1}^{m}\theta^*_{n,j}(z)$, where $\theta^*_{n,j}$ is the maximum likelihood estimator in the jth subset, and the corresponding confidence intervals defined above. We considered only the situation where $m \in \{5, 10\}$ is fixed with the sample size.

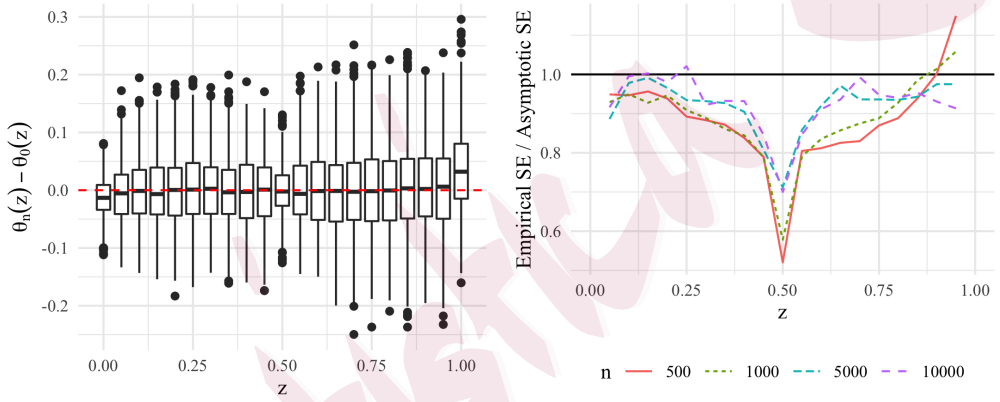

Figure 1: Left: Box plots of $\theta^*_n(z) - \theta_0(z)$ with $n = 10K$. Right: Empirical standard errors of $r_n[\theta^*_n(z) - \theta_0(z)]$ divided by the limit theory-based counterparts for $z \in (0, 1)$, where $r_n = n^{1/2}$ for $z = 0.5$, and $r_n = n^{1/3}$ otherwise.

We now turn to the results of the simulation study. The left panel of Figure 1 displays the distribution of $\theta^*_n(z) - \theta_0(z)$ for $z \in [0, 1]$ and $n = 10K$.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

These distributions are approximately centered around 0 for $z \in (0, 1)$, but not for $z \in \{0, 1\}$. Hence, despite the positive mass at the boundaries, the maximum likelihood estimator does not appear to be consistent at the boundaries. This is a common problem among monotonicity-constrained estimators, and various correction procedures have been proposed and could be considered in this context (see, e.g., Woodroofe and Sun, 1993; Kulikov and Lopuhaä, 2006).

The right panel of Figure 1 displays the ratio of the standard deviation of $r_n[\theta_n^*(z) - \theta_0(z)]$ to the standard deviation of the asymptotic distributions derived in Section 4 for $z \neq 0, 1$. For $z = 0.5$, $r_n = n^{1/2}$ and the asymptotic distribution is that of the fully discrete case presented in Section 4.1, though the results presented in that section do not apply here because of the mixed discrete-continuous nature of F_0 and G_0. Otherwise, $r_n = n^{1/3}$ and the asymptotic distribution is that of the continuous case presented in Section 4.2. We see that, for $z \neq 0.5$, the empirical standard error approaches the asymptotic standard deviation as n grows. However, for $z = 0.5$, the empirical standard error is converging to a limit that is strictly smaller than the asymptotic standard deviation. This suggests that, at points that have both positive mass and positive density in a neighborhood of the point, the maximum likelihood estimator gains efficiency from the positive den-

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

sity. In addition, points of continuity near the mass point also experience finite-sample efficiency gains.

Figure 2 shows the ratios of the mean squared errors of the maximum smoothed likelihood estimator, the kernel density-based estimator, and the sample splitting estimators to that of the maximum likelihood estimator. The maximum smoothed likelihood estimator is slightly more efficient than the maximum likelihood estimator at continuity points, but is less efficient around mass points. Furthermore, the relative performance of the maximum likelihood estimator at positive mass points increases as the sample size grows. The kernel density estimator is, in general, less efficient than the maximum likelihood estimator, especially near mass points, and the discrepancy grows with the sample size.

For large enough n, the sample splitting estimator is more efficient than the maximum likelihood estimator at all points at which the latter is consistent. The relative improvement of $\bar{\theta}_{n,m}$ grows with the number of splits m, as does the sample size n required for $\bar{\theta}_{n,m}$ to outperform θ_n^*.

Figure 3 shows the empirical coverage of 95% confidence intervals for $\theta_0(z)$ constructed using the plug-in method described in Section 4.2, the inverted likelihood ratio test approach of Banerjee and Wellner (2001), and the sample splitting approach of Banerjee et al. (2019), described above.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

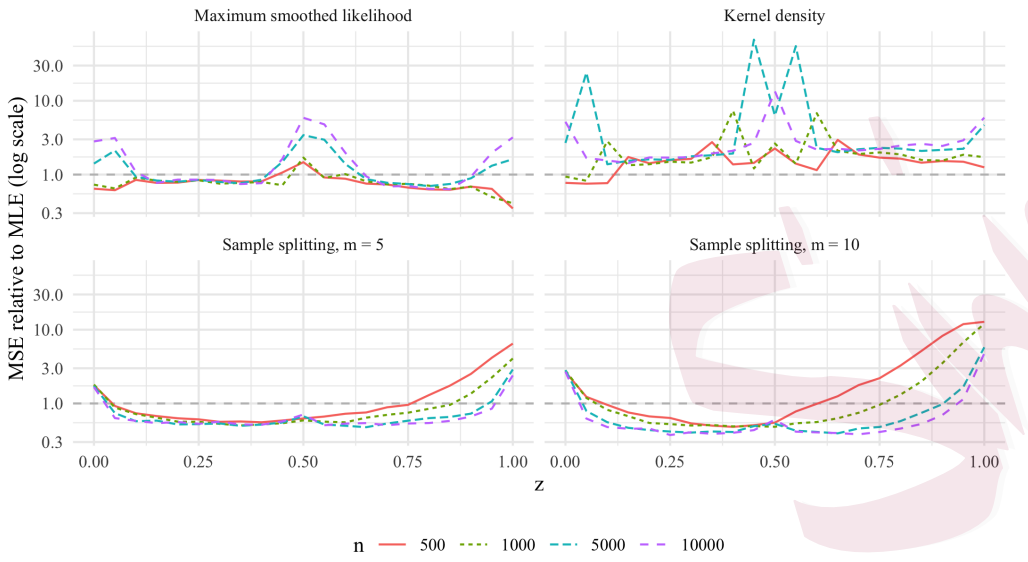

Figure 2: Relative mean squared errors of the maximum smoothed likelihood estimator, kernel density-based estimator, and sample splitting estimators to the maximum likelihood estimator for $z \in [0, 1]$ and various sample sizes n. The maximum likelihood has a better mean squared error for y-values greater than one, and the other estimator has a better mean squared error for y-values less than one.

Note that the likelihood ratio approach does not provide intervals at the end points $z = 0$ or $z = 1$. The plug-in method is conservative in large samples near mass points, but anti-conservative at some points of positive density. This is because the plug-in method is designed to work when the distributions are fully continuous, and estimation of the required nuisance

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

parameters in the limit distribution fails in the presence of mass points. The likelihood ratio method is conservative in smaller samples, but approaches nominal coverage in large samples for points z of absolute continuity. The sample splitting method with $m = 5$ has adequate coverage for all sample sizes, except for z close to the boundaries. The sample splitting method with $m = 10$ (and similarly for $m = 20$, which is not shown) appears to require very large sample sizes to attain adequate coverage over a large range of z. Note that the sample splitting methods were able to achieve good coverage in large samples at both interior absolutely continuous points and interior mass points, without the user specifying which points are which.

6. Analysis of C-reactive protein for predicting bacterial infection

In this section, we use the methods presented herein to assess using the biomarker C-reactive protein (CRP) to determine the presence or absence of bacterial infection in children with systemic inflammatory response syndrome (SIRS). The Optimizing Antibiotic Strategies in Sepsis (OASIS) II study enrolled a prospective observational cohort of children under the age of 19 at the pediatric intensive care unit at The Children's Hospital of Philadelphia from August 2012 to June 2016 (Downes et al., 2018). Patients

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

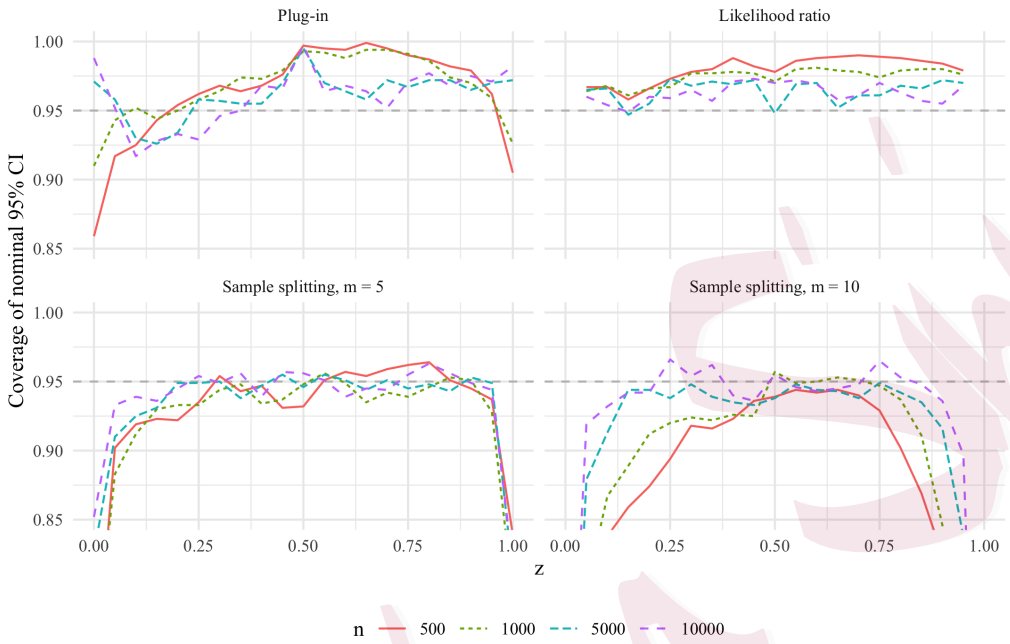

Figure 3: Coverage of 95% CIs for $z \in [0, 1]$, various sample sizes n, and four methods: the plug-in method centered around the maximum likelihood estimator (upper left), the inverted likelihood ratio tests (upper right), and the sample splitting method with $m = 5$ (lower left) and $m = 10$ (lower right). Note that the likelihood ratio method does not provide intervals at the endpoints.

were enrolled in the study if they presented signs of SIRS, were started on a new broad-spectrum antibiotic for suspected bacterial infection, and had blood cultures taken within six hours of SIRS onset. A primary goal of the study was to assess whether CRP, which had previously been found to be

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

predictive of bacterial infection (Downes et al., 2017), could be used to determine when antibiotic therapy could be safely ended. Additional details of the study design and the results of the primary analysis may be found in Downes et al. (2018).

We analyzed all patients in the OASIS II cohort with measured biomarkers and bacterial infection status to assess the odds of bacterial infection as a function of CRP value. Some patients had measurements at multiple episodes; because all such episodes were at least 30 days apart, we treated them as independent of one another. We analyzed a total of $n = 504$ CRP measurements among 443 unique patients, with $n_1 = 202$ bacterial infections among 191 unique patients and $n_2 = 302$ non-infections among 266 unique patients.

Because CRP has previously been found to be predictive of bacterial infection in this patient population, there is scientific reason to believe that the density ratio order holds. We therefore computed the MLE of the density ratio function and corresponding 95% likelihood ratio-based pointwise confidence intervals, and the sample splitting estimator of Banerjee et al. (2019) with $m = 5$ splits and corresponding 95% pointwise confidence intervals.

Figure 4 displays the estimated odds of bacterial infection given a CRP

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

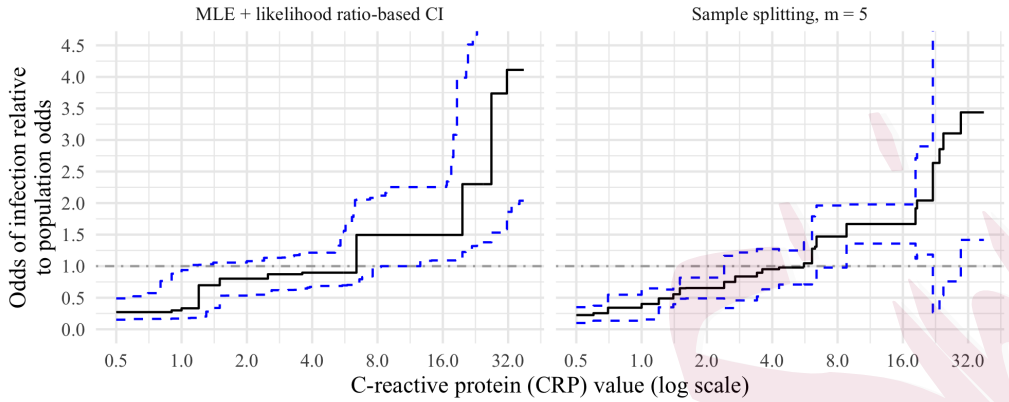

Figure 4: Odds of bacterial infection given C-reactive protein value relative to population odds in children with treating SIRS.

value, relative to the population odds of bacterial infection, and 95% pointwise confidence intervals. We find that values of CRP under one are indicative of roughly quartered odds of infection relative to the population odds of infection, and values of CRP greater than 20 are indicative of roughly doubled odds of infection relative to the population odds. Values of CRP between one and 20 do not clearly indicate that a patient's odds of infection are larger or smaller than the population odds.

7. Conclusion

We have considered nonparametric maximum likelihood inference for the density ratio function and individual distribution functions under the as-

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

sumption that the density ratio is nondecreasing. We applied these methods to analyze the biomarker C-reactive protein for predicting bacterial infection in children with SIRS. The methods apply broadly to biomarker analysis, as well as to other areas of biomedical research.

One of our important contributions is the ability to deal with discrete, continuous, and mixed discrete-continuous distributions. Such distributions arise frequently in applied settings, and particularly in the context of biomarker analysis. Furthermore, we have demonstrated via numerical studies that sample splitting provides good pointwise confidence interval coverage without knowing which values correspond to discrete mass points and which correspond to points of Lebesgue continuity of the underlying densities. This is important, because in practice, analysts may not have such knowledge *a priori*. However, a theoretical treatment of the precise asymptotic behavior of the estimator at mass points remains unknown, and is an interesting topic for future research.

Supplementary Materials

The Supplementary Material contains an example of the use of Theorem 2, proofs of all theorems, additional simulation results, and additional results from the data analysis.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

Acknowledgments

The authors thank Craig Boge for help compiling the OASIS II data. The authors also gratefully acknowledge the constructive comments of the editors and anonymous reviewers and the support of the CDC Epicenters program (KJD), NICHD grant K23HD091365 (KJD), the Center for Pediatric Clinical Effectiveness and the Pediatric IDEAS Research Group of the Children's Hospital of Philadelphia (KJD, TW), and the Department of Pediatrics of the University of Pennsylvania Perelman School of Medicine.

References

Arcones, M. A. and F. J. Samaniego (2000). On the Asymptotic Distribution Theory of a Class of Consistent Estimators of a Distribution Satisfying a Uniform Stochastic Ordering Constraint. *Ann. Stat. 28*(1), 116–150.

Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J. Math. Psychol. 12*(4), 387 – 415.

Banerjee, M. (2007, 07). Likelihood based inference for monotone response models. *Ann. Statist. 35*(3), 931–956.

Banerjee, M., C. Durot, and B. Sen (2019, 04). Divide and conquer in nonstandard problems and the super-efficiency phenomenon. *Ann. Statist. 47*(2), 720–757.

Banerjee, M. and J. A. Wellner (2001). Likelihood ratio tests for monotone functions. *Ann.*

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

Statist. 29(6), 1699–1731.

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

Beare, B. K. and J.-M. Moon (2015). Nonparametric tests of density ratio ordering. *Econom. Theory 31*(3), 471–492.

Brunk, H. D. (1970). Estimation of isotonic regression. In *Nonparametric Techniques in Statistical Inference (Proc. Sympos., Indiana Univ., Bloomington, Ind., 1969)*, London, pp. 177–197. Cambridge Univ. Press.

Carolan, C. A. and J. M. Tebbs (2005, 03). Nonparametric tests for and against likelihood ratio ordering in the two-sample problem. *Biometrika 92*(1), 159–171.

Davidov, O. and A. Herman (2012). Ordinal dominance curve based inference for stochastically ordered distributions. *J. R. Stat. Soc. Ser. B. Stat. Methodol. 74*(5), 825–847.

Downes, K. J., J. C. Fitzgerald, E. Schriver, et al. (2018). Implementation of a Pragmatic Biomarker-Driven Algorithm to Guide Antibiotic Use in the Pediatric Intensive Care Unit: the Optimizing Antibiotic Strategies in Sepsis (OASIS) II Study. *J. Pediatric Infect. Dis. Soc.*. advance online publication.

Downes, K. J., S. L. Weiss, J. S. Gerber, et al. (2017). A Pragmatic Biomarker-Driven Algorithm to Guide Antibiotic Use in the Pediatric Intensive Care Unit: The Optimizing Antibiotic Strategies in Sepsis (OASIS) Study. *J. Pediatric Infect. Dis. Soc. 6*(2), 134–141.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

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

Groeneboom, P. and G. Jongbloed (2014). *Nonparametric estimation under shape constraints*. Cambridge University Press.

Groeneboom, P. and G. Jongbloed (2015, 10). Nonparametric confidence intervals for monotone functions. *Ann. Statist. 43*(5), 2019–2054.

Groeneboom, P. and J. A. Wellner (2001). Computing Chernoff's distribution. *J. Comput. Graph. Statist. 10*(2), 388–400.

Hsieh, F. and B. W. Turnbull (1996, 02). Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *Ann. Statist. 24*(1), 25–40.

Kulikov, V. N. and H. P. Lopuhaä (2006, 04). The behavior of the NPMLE of a decreasing density near the boundaries of the support. *Ann. Statist. 34*(2), 742–768.

Lehmann, E. L. and J. Rojo (1992, 12). Invariant Directional Orderings. *Ann. Statist. 20*(4), 2100–2110.

Mukerjee, H. (1996). Estimation of survival functions under uniform stochastic ordering. *J. Amer. Statist. Assoc. 91*(436), 1684–1689.

Rojo, J. and F. J. Samaniego (1991). On nonparametric maximum likelihood estimation of a distribution uniformly stochastically smaller than a standard. *Statist. Probab. Lett. 11*(3), 267 – 271.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0207

Rojo, J. and F. J. Samaniego (1993). On estimating a survival curve subject to a uniform stochastic ordering constraint. *J. Amer. Statist. Assoc. 88*(422), 566–572.

Roosen, J. and D. A. Hennessy (2004). Testing for the monotone likelihood ratio assumption. *J. Bus. Econ. Stat. 22*(3), 358–366.

Shaked, M. and J. G. Shanthikumar (2007). *Stochastic Orders*. Springer Science & Business Media.

Tang, C.-F., D. Wang, and J. M. Tebbs (2017, 12). Nonparametric goodness-of-fit tests for uniform stochastic ordering. *Ann. Statist. 45*(6), 2565–2589.

Woodroofe, M. and J. Sun (1993). A penalized maximum likelihood estimate of f(0+) when f is non-increasing. *Statist. Sinica 3*(2), 501–515.

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

Department of Mathematics and Statistics, University of Massachusetts Amherst

E-mail: twestling@math.umass.edu

Center for Pediatric Clinical Effectiveness, Children's Hospital of Philadelphia

E-mail: downeskj@email.chop.edu

Department of Statistics, The Wharton School, University of Pennsylvania

E-mail: dsmall@wharton.upenn.edu