Electronic Journal of Statistics
Vol. 7 (2013) 364–380
ISSN: 1935-7524
DOI: 10.1214/13-EJS777

Global rates of convergence of the MLE for multivariate interval censoring

Fuchang Gao*

Department of Mathematics
University of Idaho
Moscow, Idaho 83844-1103
e-mail: fuchang@uidaho.edu

and

Jon A. Wellner†

Department of Statistics, Box 354322
University of Washington
Seattle, WA 98195-4322
e-mail: jaw@stat.washington.edu
url: http://www.stat.washington.edu/jaw/

Abstract: We establish global rates of convergence of the Maximum Likelihood Estimator (MLE) of a multivariate distribution function on $\mathbb{R}^d$ in the case of (one type of) "interval censored" data. The main finding is that the rate of convergence of the MLE in the Hellinger metric is no worse than $n^{-1/3}(\log n)^{\gamma}$ for $\gamma = (5d-4)/6$.

AMS 2000 subject classifications: Primary 62G07, 62H12; secondary 62G05, 62G20.
Keywords and phrases: Empirical processes, global rate, Hellinger metric, interval censoring, multivariate, multivariate monotone functions.

Received October 2012.

1. Introduction and overview

Our main goal in this paper is to study global rates of convergence of the Maximum Likelihood Estimator (MLE) in one simple model for multivariate interval-censored data. In section 3 we will show that under some reasonable conditions the MLE converges in a Hellinger metric to the true distribution function on $\mathbb{R}^d$ at a rate no worse than $n^{-1/3}(\log n)^{\gamma_d}$ for $\gamma_d = (5d-4)/6$ for all $d \geq 2$. Thus the rate of convergence is only worse than the known rate of $n^{-1/3}$ for the case $d = 1$ by a factor involving a power of $\log n$ growing linearly with the dimension. These new rate results rely heavily on recent bracketing entropy bounds for $d-$dimensional distribution functions obtained by Gao (2012).

We begin in Section 2 with a review of interval censoring problems and known results in the case $d = 1$. We introduce the multivariate interval censoring model

*Supported in part by a grant from the Simons Foundation (#246211).
†Supported in part by NSF Grant DMS-1104832 and NI-AID grant 2R01 AI291968-04.

of interest here in Section 3, and obtain a rate of convergence for this model for $d \geq 2$ in Theorem 3.1. Most of the proofs are given in Section 4, with the exception being a key corollary of Gao (2012), the statement and proof of which are given in the Appendix (Section 6). Finally, in Section 5 we introduce several related models and further problems.

2. Interval censoring (or current status data) on $\mathbb{R}$

Let $Y \sim F_0$ on $\mathbb{R}^+$, and let $T \sim G_0$ on $\mathbb{R}^+$ be independent of Y. Suppose that we observe $X_1, \ldots, X_n$ i.i.d. as $X = (\Delta, T)$ where $\Delta = 1_{[Y \leq T]}$. Here Y is often the time until some event of interest and T is an observation time. The goal is to estimate F_0 nonparametrically based on observation of the X_i's.

To calculate the likelihood, we first calculate the distribution of X for a general distribution function F: note that the conditional distribution of Δ conditional on T is Bernoulli:

$$(\Delta|T) \sim \text{Bernoulli}(p(T))$$

where $p(T) = F(T)$. If G_0 has density g_0 with respect to some measure μ on $\mathbb{R}^+$, then $X = (\Delta, T)$ has density

$$p_{F,g_0}(\delta, t) = F(t)^{\delta}(1 - F(t))^{1-\delta} g_0(t), \quad \delta \in \{0, 1\}, \quad t \in \mathbb{R}^+,$$

with respect to the dominating measure (counting measure on $\{0, 1\}$) $\times\, \mu$.

The nonparametric Maximum Likelihood Estimator (MLE) $\hat{F}_n$ of F_0 in this interval censoring model was first obtained by Ayer et al. (1955). It is simply described as follows: let $T_{(1)} \leq \cdots \leq T_{(n)}$ denote the order statistics corresponding to $T_1, \ldots, T_n$ and let $\Delta_{(1)}, \ldots, \Delta_{(n)}$ denote the corresponding Δ's. Then the part of the log-likelihood of $X_1, \ldots, X_n$ depending on F is given by

$$\begin{aligned} l_n(F) &= \sum_{i=1}^{n} \{\Delta_{(i)} \log F(T_{(i)}) + (1 - \Delta_{(i)}) \log(1 - F(T_{(i)}))\} \\ &\equiv \sum_{i=1}^{n} \{\Delta_{(i)} \log F_i + (1 - \Delta_{(i)}) \log(1 - F_i)\} \end{aligned} \tag{2.1}$$

where

$$0 \leq F_1 \leq \cdots \leq F_n \leq 1. \tag{2.2}$$

It turns out that the maximizer $\hat{F}_n$ of (2.1) subject to (2.2) can be described as follows: let H^* be the (greatest) convex minorant of the points $\{(i, \sum_{j \leq i} \Delta_{(j)}) : i \in \{1, \ldots, n\}\}$:

$$H^*(t) = \sup \left\{ \begin{array}{c} H(t) : \quad H(i) \leq \sum_{j \leq i} \Delta_{(j)} \quad \text{for each} \quad 0 \leq i \leq n \\ H(0) = 0, \text{ and } H \text{ is convex} \end{array} \right\}.$$

Let $\hat{F}_i$ denote the left-derivative of H^* at $T_{(i)}$. Then $(\hat{F}_1, \ldots, \hat{F}_n)$ is the unique vector maximizing (2.1) subject to (2.2), and we therefore take the MLE $\hat{F}_n$ of F to be

$$\hat{F}_n(t) = \sum_{i=0}^{n} \hat{F}_i 1_{[T_{(i)}, T_{(i+1)})}(t)$$

with the conventions $T_{(0)} \equiv 0$ and $T_{(n+1)} \equiv \infty$. See Ayer et al. (1955) or Groeneboom and Wellner (1992), pages 38-43, for details.

Groeneboom (1987) initiated the study of $\hat{F}_n$ and proved the following limiting distribution result at a fixed point t_0.

Theorem 2.1 (Groeneboom, 1987)**.** *Consider the current status model on $\mathbb{R}^+$. Suppose that $0 < F_0(t_0), G_0(t_0) < 1$ and suppose that F and G are differentiable at t_0 with strictly positive derivatives $f_0(t_0)$ and $g_0(t_0)$ respectively. Then*

$$n^{1/3}(\hat{F}_n(t_0) - F_0(t_0)) \to_d c(F_0, G_0)\mathbb{Z}$$

where

$$c(F_0, G_0) = 2\left(\frac{F_0(t_0)(1 - F_0(t_0))f_0(t_0)}{2g_0(t_0)}\right)^{1/3}$$

and

$$\mathbb{Z} = argmin\{W(t) + t^2\}$$

where W is a standard two-sided Brownian motion starting from 0.

The distribution of $\mathbb{Z}$ has been studied in detail by Groeneboom (1989) and computed by Groeneboom and Wellner (2001). Balabdaoui and Wellner (2012) show that the density $f_{\mathbb{Z}}$ of $\mathbb{Z}$ is log-concave.

van de Geer (1993) (see also van de Geer (2000)) obtained the following global rate result for $p_{\hat{F}_n}$. Recall that the Hellinger distance $h(p, q)$ between two densities with respect to a dominating measure μ is given by

$$h^2(p, q) = \frac{1}{2}\int \{\sqrt{p} - \sqrt{q}\}^2 d\mu.$$

Proposition 2.2 (van de Geer, 1993)**.** $h(p_{\hat{F}_n}, p_{F_0}) = O_p(n^{-1/3})$.

Now for any distribution functions F and F_0 the (squared) Hellinger distance $h^2(p_F, p_{F_0})$ for the current status model is given by

$$\begin{aligned} h^2(p_F, p_{F_0}) &= \frac{1}{2}\left\{\int (\sqrt{F} - \sqrt{F_0})^2 dG_0 + \int (\sqrt{1-F} - \sqrt{1-F_0})^2 dG_0\right\} \\ &= \frac{1}{2}\int \frac{\{(\sqrt{F} - \sqrt{F_0})(\sqrt{F} + \sqrt{F_0})\}^2}{(\sqrt{F} + \sqrt{F_0})^2} dG_0 \\ &\quad + \frac{1}{2}\int \frac{\{(\sqrt{1-F} - \sqrt{1-F_0})(\sqrt{1-F} + \sqrt{1-F_0})\}^2}{(\sqrt{1-F} + \sqrt{1-F_0})^2} dG_0 \end{aligned}$$

$$\geq \frac{1}{8}\int (F-F_0)^2 dG_0 + \frac{1}{8}\int ((1-F)-(1-F_0))^2 dG_0$$
$$= \frac{1}{4}\int (F-F_0)^2 dG_0, \tag{2.3}$$

and hence Proposition 2.2 yields

$$\int_0^\infty (\hat{F}_n(z)-F_0(z))^2 dG_0(z) = O_p(n^{-2/3}), \tag{2.4}$$

or $\|\hat{F}_n - F_0\|_{L_2(G_0)} = O_p(n^{-1/3})$.

For generalizations of these and other asymptotic results for the current status model to more complicated interval censoring schemes for real-valued random variables Y, see e.g. Groeneboom and Wellner (1992), van de Geer (1993), Groeneboom (1996), van de Geer (2000), Schick and Yu (2000), and Groeneboom, Maathuis and Wellner (2008a,b).

Our main focus in this paper, however, concerns one simple generalization of the interval censoring model for $\mathbb{R}$ introduced above to interval censoring in $\mathbb{R}^d$. We now turn to this generalization.

3. Multivariate interval censoring: multivariate current status data

Let $\underline{Y} = (Y_1,\dots,Y_d) \sim F_0$ on $\mathbb{R}^{+d} \equiv [0,\infty)^d$, and let $\underline{T} = (T_1,\dots,T_d) \sim G_0$ on $\mathbb{R}^{+d}$ be independent of $\underline{Y}$. We assume that G_0 has density g_0 with respect to some dominating measure μ on $\mathbb{R}^d$. Suppose we observe $\underline{X}_1,\dots,\underline{X}_n$ i.i.d. as $\underline{X} = (\underline{\Delta}, \underline{T})$ where $\underline{\Delta} = (\Delta_1,\dots,\Delta_d)$ is given by $\Delta_j = 1_{[Y_j \leq T_j]}$, $j = 1,\dots,d$. Equivalently, with a slight abuse of notation, $\underline{X} = (\underline{\Gamma}, \underline{T})$ where $\underline{\Gamma} = (\Gamma_1,\dots,\Gamma_{2^d})$ is a vector of length 2^d consisting of 0's and 1's and with at most one 1 which indicates which of the 2^d orthants of $\mathbb{R}^{+d}$ determined by $\underline{T}$ the random vector $\underline{Y}$ belongs. More explicitly, define $K \equiv 1 + \sum_{j=1}^d (1-\Delta_j)2^{j-1}$. Then set $\Gamma_k \equiv 1\{k = K\}$ for $k = 1,\dots,2^d$, so that $\Gamma_K = 1$ and $\Gamma_l = 0$ for $l \in \{1,\dots,2^d\} \setminus \{K\}$. Much as for univariate current status data, $\underline{Y}$ represents a vector of times to events, $\underline{T}$ is a vector of observation times, and the goal is nonparametric estimation of the joint distribution function F_0 of $\underline{Y}$ based on observation of the $\underline{X}_i$'s. See Dunson and Dinse (2002), Jewell (2007), Wang (2009), and Lin and Wang (2011) for examples of settings in which data of this type arises.

To calculate the likelihood, we first calculate the distribution of $\underline{X}$ for a general distribution function F: note that the conditional distribution of $\underline{\Gamma}$ conditional on $\underline{T}$ is Multinomial:

$$(\underline{\Gamma}|\underline{T}) \sim \text{Mult}_{2^d}(1, \underline{p}(\underline{T}; F))$$

where $\underline{p}(\underline{T};F) = (p_1(\underline{T};F),\dots,p_{2^d}(\underline{T};F))$ and the probabilities $p_j(\underline{t};F)$, $j = 1,\dots,2^d$, $\underline{t} \in \mathbb{R}^{+d}$ are determined by the F measures of the corresponding sets.

Then our model $\mathcal{P}$ for multivariate current status data is the collection of all densities with respect to the dominating measure (counting measure on $\{0,1\}^{2^d}$) $\times\mu$ given by

$$\prod_{j=1}^{2^d} p_j(\underline{t}; F)^{\gamma_j} g_0(\underline{t})$$

for some distribution function F on $\mathbb{R}^{+d}$ where $\underline{t} \in \mathbb{R}^{+d}$ and $\gamma_j \in \{0,1\}$ with $\sum_{j=1}^{2^d} \gamma_j = 1$.

Now the part of the log-likelihood that depends on F is given by

$$l_n(F) = \sum_{i=1}^{n} \sum_{j=1}^{2^d} \Gamma_{i,j} \log p_j(\underline{T}_i; F),$$

and again the MLE $\hat{F}_n$ of the true distribution function F_0 is given by

$$\hat{F}_n = \text{argmax}\{l_n(F) : \ F \text{ is a distribution function on } \mathbb{R}^{+d}\}. \tag{3.1}$$

For example, when $d = 2$, we can write $\Gamma_1 = \Delta_1\Delta_2$, $\Gamma_2 = (1-\Delta_1)\Delta_2$, $\Gamma_3 = \Delta_1(1-\Delta_2)$, and $\Gamma_4 = (1-\Delta_1)(1-\Delta_2)$, and then

$$\begin{aligned} p_1(\underline{T}; F) &= F(T_1, T_2), \\ p_2(\underline{T}; F) &= F(\infty, T_2) - F(T_1, T_2), \\ p_3(\underline{T}; F) &= F(T_1, \infty) - F(T_1, T_2), \\ p_4(\underline{T}; F) &= 1 - F(T_1, \infty) - F(\infty, T_2) + F(T_1, T_2). \end{aligned}$$

Thus

$$P_F(\underline{\Gamma} = \underline{\gamma} | \underline{T}) = \prod_{j=1}^{4} p_j(\underline{T}; F)^{\gamma_j}, \ \text{ for } \underline{\gamma} = (\gamma_1, \gamma_2, \gamma_3, \gamma_4), \ \ \gamma_j \in \{0,1\}, \ \sum_{j=1}^{4} \gamma_j = 1.$$

Note that

$$p_j(\underline{t}; F) = \int_{[0,\infty)^2} 1_{C_j(\underline{t})}(\underline{y}) dF(\underline{y}), \qquad j = 1, \ldots, 4 \tag{3.2}$$

where

$$\begin{aligned} C_1(\underline{t}) &= [0, t_1] \times [0, t_2], \\ C_2(\underline{t}) &= [0, t_1] \times (t_2, \infty), \\ C_3(\underline{t}) &= (t_1, \infty) \times [0, t_2], \\ C_4(\underline{t}) &= (t_1, \infty) \times (t_2, \infty). \end{aligned}$$

Characterizations and computation of the MLE (3.1), mostly for the case $d = 2$ have been treated in Song (2001), Gentleman and Vandal (2002), and

Maathuis (2005, 2006). Consistency of the MLE for more general interval censoring models has been established by Yu, Yu and Wong (2006). For an interesting application see Betensky and Finkelstein (1999). This example and other examples of multivariate interval censored data are treated in Sun (2006) and Deng and Fang (2009). For a comparison of the MLE with alternative estimators in the case $d = 2$, see Groeneboom (2012a).

An analogue of Groeneboom's Theorem 2.1 has not been established in the multivariate case. Song (2001) established an asymptotic minimax lower bound for pointwise convergence when $d = 2$: if F_0 and G_0 have positive continuous densities at $\underline{t}_0$, then no estimator has a local minimax rate for estimation of $F_0(\underline{t}_0)$ faster than $n^{-1/3}$. By making use of additional smoothness hypotheses, Groeneboom (2012a) has constructed estimators which achieve the pointwise $n^{-1/3}$ rate, but it is not yet known if the MLE achieves this.

Our main goal here is to prove the following theorem concerning the global rate of convergence of the MLE $\hat{F}_n$.

Theorem 3.1. *Consider the multivariate current status model. Suppose that F_0 has supp$(F_0) \subset [0, M]^d$ and that F_0 has density f_0 which satisfies*

$$c_1^{-1} \le f_0(\underline{y}) \le c_1 \quad \textit{for all} \quad \underline{y} \in [0, M]^d \tag{3.3}$$

where $0 < c_1 < \infty$. Suppose that G_0 has density g_0 which satisfies

$$c_2^{-1} \le g_0(\underline{y}) \le c_2 \quad \textit{for all} \quad \underline{y} \in [0, M]^d. \tag{3.4}$$

Then the MLE $\widehat{p}_n \equiv p_{\widehat{F}_n}$ of $p_0 \equiv p_{F_0}$ satisfies

$$h(\widehat{p}_n, p_0) = O_p\left(\frac{(\log n)^\gamma}{n^{1/3}}\right)$$

for $\gamma \equiv \gamma_d \equiv (5d - 4)/6$.

Since the inequality (2.3) continues to hold in $\mathbb{R}^d$ for $d \ge 2$ (with $1/4$ replaced by $1/8$ on the right side), we obtain the following corollary:

Corollary 3.2. *Under the conditions of Theorem 3.1 it follows that*

$$\int_{\mathbb{R}^{+d}} (\widehat{F}_n(z) - F_0(z))^2 dG_0(z) = O_p(n^{-2/3}(\log n)^\beta)$$

for $\beta \equiv \beta_d = 2\gamma_d = (5d - 4)/3$.

4. Proofs

Here we give the proof of Theorem 3.1. The main tool is a method developed by van de Geer (2000). We will use the following lemma in combination with Theorem 7.6 of van de Geer (2000) or Theorem 3.4.1 of van der Vaart and Wellner (1996) (Section 3.4.2, pages 330-331). Without loss of generality we can take $M = 1$ where M is the upper bound of the support of F (see Theorem 3.1).

Let $\mathcal{P}$ be a collection of probability densities p on a sample space $\mathcal{X}$ with respect to a dominating measure μ. Define

$$\mathcal{G}^{(conv)} \equiv \left\{ \frac{2p}{p+p_0} : \ p \in \mathcal{P} \right\}, \tag{4.1}$$

$$\sigma(\delta) \equiv \sup\{\sigma \geq 0 : \ \int_{\{p_0 \leq \sigma\}} p_0 d\mu \leq \delta^2\} \qquad \text{for} \quad \delta > 0, \tag{4.2}$$

$$\mathcal{G}_\sigma^{(conv)} \equiv \left\{ \frac{2p}{p+p_0} 1_{[p_0 > \sigma]} : \ p \in \mathcal{P} \right\}, \qquad \text{for} \quad \sigma > 0. \tag{4.3}$$

The following general result relating the bracketing entropies $\log N_{[\,]}(\cdot, \mathcal{G}^{(conv)}, L_2(P_0))$, $\log N_{[\,]}(\cdot, \mathcal{G}^{(conv)}_{\sigma(\epsilon)}, L_2(P_0))$, $\log N_{[\,]}(\cdot, \mathcal{P}, L_2(Q_{\sigma(\epsilon)}))$, and $\log N_{[\,]}(\cdot, \mathcal{P}, L_2(\tilde{Q}_{\sigma(\epsilon)}))$ is due to van de Geer (2000).

Lemma 4.1 (van de Geer, 2000)**.** *For every* $\epsilon > 0$

$$\begin{aligned} \log N_{[\,]}(3\epsilon, \mathcal{G}^{(conv)}, L_2(P_0)) &\leq \log N_{[\,]}(\epsilon, \mathcal{G}^{(conv)}_{\sigma(\epsilon)}, L_2(P_0)) & (4.4)\\ &\leq \log N_{[\,]}(\epsilon/2, \mathcal{P}, L_2(Q_{\sigma(\epsilon)})) & (4.5)\\ &= \log N_{[\,]}\left(\frac{\epsilon/2}{\sqrt{Q_{\sigma(\epsilon)}(\mathcal{X})}}, \mathcal{P}, L_2(\tilde{Q}_{\sigma(\epsilon)}) \right) & (4.6) \end{aligned}$$

where $dQ_\sigma \equiv p_0^{-1} 1_{[p_0 > \sigma]} d\mu$ *and* $\tilde{Q}_\sigma \equiv Q_\sigma / Q_\sigma(\mathcal{X})$.

Proof. We first show that (4.4) holds. Suppose that $\{[g_{L,j}, g_{U,j}], \ j = 1, \ldots, m\}$ are ϵ-brackets with respect to $L_2(P_0)$ for $\mathcal{G}^{(conv)}_{\sigma(\epsilon)}$ with

$$\mathcal{G}^{(conv)}_{\sigma(\epsilon)} \subset \bigcup_{j=1}^m [g_{L,j}, g_{U,j}], \qquad m = N_{[\,]}(\epsilon, \mathcal{G}^{(conv)}_{\sigma(\epsilon)}, L_2(P_0)).$$

Then for $g \in \mathcal{G}^{(conv)}$, let $g_\sigma \equiv g 1_{[p_0 > \sigma]}$ be the corresponding element of $\mathcal{G}^{(conv)}_{\sigma(\epsilon)}$. Suppose that $g_\sigma \in [g_{L,j}, g_{U,j}]$ for some $j \in \{1, \ldots, m\}$. Then

$$g = g 1_{[p_0 \leq \sigma]} + g_\sigma \ \begin{cases} \leq g 1_{[p_0 \leq \sigma]} + g_{U,j} \equiv \tilde{g}_{U,j} \\ \geq 0 + g_{L,j} \equiv \tilde{g}_{L,j}, \end{cases}$$

where, by the triangle inequality, $0 \leq g \leq 2$ for all $g \in \mathcal{G}^{(conv)}$, and the definition of $\sigma(\epsilon)$, it follows that

$$\left\| \tilde{g}_{U,j} - \tilde{g}_{L,j} \right\|_{P_0,2} \leq \left\| g_{U,j} - g_{L,j} \right\|_{P_0,2} + 2\epsilon \leq 3\epsilon.$$

Thus $\{[\tilde{g}_{L,j}, \tilde{g}_{U,j}] : \ j \in \{1, \ldots, m\}\}$ is a collection of $3\epsilon-$brackets for $\mathcal{G}^{(conv)}$ with respect to $L_2(P_0)$ and hence (4.4) holds.

Now we show that (4.5) holds. Suppose that $\{[p_{L,j}, p_{U,j}] : \ j = 1, \ldots, m\}$ is a set of $\epsilon/2-$brackets with respect to $L_2(Q_\sigma)$ for $\mathcal{P}$ with

$$\mathcal{P} \subset \bigcup_{j=1}^m [p_{L,j}, p_{U,j}] \qquad \text{and} \qquad m = N_{[\,]}(\epsilon/2, \mathcal{P}, L_2(Q_{\sigma(\epsilon)})).$$

Suppose $p \in [p_{L,j}, p_{U,j}]$ for some j. Then, since

$$\frac{2p}{p+p_0} 1_{[p_0>\sigma]} \quad \begin{cases} \leq \frac{2p_{U,j}}{p_{U,j}+p_0} 1_{[p_0>\sigma]} \equiv g_{U,j}, \\ \\ \geq \frac{2p_{L,j}}{p_{U,j}+p_0} 1_{[p_0>\sigma]} \equiv g_{L,j} \end{cases}$$

where

$$\begin{aligned} & |g_{U,j} - g_{L,j}| \\ &= \left| \frac{2p_{U,j}}{p_{U,j}+p_0} 1_{[p_0>\sigma]} - \frac{2p_{L,j}}{p_{U,j}+p_0} 1_{[p_0>\sigma]} \right| \\ &= \frac{2(p_{U,j}-p_{L,j})}{p_{L,j}+p_0} 1_{[p_0>\sigma]} \leq \frac{2|p_{U,j}-p_{L,j}|}{p_0} 1_{[p_0>\sigma]}. \end{aligned}$$

Thus

$$\|g_{U,j} - g_{L,j}\|_{P_0,2} \leq 2 \|p_{U,j} - p_{L,j}\|_{Q_\sigma,2} \leq \epsilon,$$

and hence $\{[g_{L,j}, g_{U,j}] : \ j = 1, \ldots, m\}$ is a set of ϵ-brackets with respect to $L_2(P_0)$ for $\mathcal{G}_\sigma^{(conv)}$. This shows that (4.5) holds.

It remains only to show that (4.6) holds. But this is easy since $\|g\|^2_{Q_\sigma,2} = \|g\|^2_{\tilde{Q}_\sigma,2} \cdot Q_\sigma(\mathcal{X})$.

This lemma is based on van de Geer (2000), pages 101 and 103. Note that our constants differ slightly from those of van de Geer. □

Lemma 4.2. *Suppose that F_0 has density f_0 which satisfies, for some $0 < c_1 < \infty$,*

$$\frac{1}{c_1} \leq f_0(\underline{y}) \leq c_1 \qquad \text{for all } \ \underline{y} \in [0,1]^d. \tag{4.7}$$

Then p_0 (which we can identify with the vector $p_0(\cdot, F_0)$) satisfies

$$p_{0,1}(\underline{t}; F_0) \begin{cases} \leq c_1 \prod_{j=1}^d t_j \\ \geq c_1^{-1} \prod_{j=1}^d t_j, \end{cases} \qquad \text{for all } \ \underline{t} \in [0,1]^d,$$

$$\vdots$$

$$p_{0,2^d}(\underline{t}; F_0) \begin{cases} \leq c_1 \prod_{j=1}^d (1-t_j) \\ \geq c_1^{-1} \prod_{j=1}^d (1-t_j), \end{cases} \qquad \text{for all } \ \underline{t} \in [0,1]^d.$$

Proof. This follows immediately from the general d version of (3.2) and the assumption on f_0. □

These inequalities can also be written in the following compact form: For $k = 1 + \sum_{j=1}^d (1-\delta_j) 2^{j-1}$ with $\delta_j \in \{0,1\}$,

$$p_{0,k}(\underline{t}; F_0) \begin{cases} \leq c_1 \prod_{j=1}^d t_j^{\delta_j} (1-t_j)^{1-\delta_j} \\ \geq c_1^{-1} \prod_{j=1}^d t_j^{\delta_j} (1-t_j)^{1-\delta_j}, \end{cases} \qquad \text{for all } \ t \in [0,1]^d.$$

Lemma 4.3. *Suppose that the assumption of Lemma 4.2 holds. Suppose, moreover, that G_0 has density g_0 which satisfies*

$$\frac{1}{c_2} \le g_0(\underline{y}) \le c_2 \qquad \textit{for all} \;\; \underline{y} \in [0,1]^d. \tag{4.8}$$

Then

$$\int_{[p_0 \le \sigma]} p_0 d\mu \le 2^d (c_1 c_2)^2 \sigma.$$

Furthermore, with $\sigma(\delta) \equiv \delta^2/(2^d(c_1c_2)^2)$ we have

$$\int_{[p_0 \le \sigma(\delta)]} p_0 d\mu \le \delta^2.$$

Proof. The first inequality follows easily from Lemma 4.2: note that

$$\begin{aligned} \int_{[p_0 \le \sigma]} p_0 d\mu &= \sum_{k=1}^{2^d} \int_{[p_k(\underline{t}, F_0) \le \sigma]} p_k(\underline{t}, F_0) g_0(\underline{t}) \underline{dt} \\ &\le 2^d \int_{[F_0(\underline{t}) g_0(\underline{t}) \le \sigma]} F_0(\underline{t}) g_0(\underline{t}) \underline{dt} \\ &\le 2^d c_1 c_2 \int_{[c_1^{-1} c_2^{-1} \prod_{j=1}^d t_j \le \sigma]} \prod_{j=1}^d t_j \, \underline{dt} \le 2^d (c_1 c_2)^2 \sigma. \end{aligned}$$

The second inequality follows from the first inequality of the lemma. □

Lemma 4.4. *If the hypotheses of Lemmas 4.2 and 4.3 hold, then the measure Q_σ defined by $dQ_\sigma \equiv (1/p_0) 1\{p_0 > \sigma\} d\mu$ has total mass $Q_\sigma(\mathcal{X})$ given by*

$$\begin{aligned} \int dQ_\sigma &= \int_{\{p_0 > \sigma\}} \frac{1}{p_0} d\mu \\ &= \sum_{j=1}^{2^d} \int_{\{\underline{t}:\; p_{0,j}(\underline{t}) g_0(\underline{t}) > \sigma\}} \frac{1}{p_{0,j}(\underline{t}) g_0(\underline{t})} \underline{dt} \\ &\le 2^d \int_{\{\underline{t} \in [0,1]^d :\; \prod_{j=1}^d t_j > \sigma/(c_1 c_2)\}} \frac{c_1 c_2}{\prod_{j=1}^d t_j} \underline{dt} && (4.9) \\ &= \frac{2^d c_1 c_2}{d!} (\log(c_1 c_2/\sigma))^d. && (4.10) \end{aligned}$$

Proof. This follows from Lemma 4.2, followed by an explicit calculation. In particular, the equality in (4.10) follows from

$$\begin{aligned} \int_{[\prod_{j=1}^d t_j > b]} \frac{1}{\prod_{j=1}^d t_j} \underline{dt} &= \int_{[\sum_1^d x_j \le \log(1/b)]} \underline{dx} \quad \text{by the change of variables } t_j = e^{-x_j}, \\ &= \frac{1}{d!} (\log(1/b))^d \qquad \text{for} \;\; 0 < b \le 1 \end{aligned}$$

where the second equality follows by induction: it holds easily for $d = 1$ (and $d = 2$); and then an easy calculation shows that it holds for d if it holds for $d - 1$. $\square$

Lemma 4.5. *If the hypotheses of Lemmas 4.2 and 4.3 hold, and $d \geq 2$, then*

$$\log N_{[\,]}(\epsilon, \mathcal{G}^{(conv)}, L_2(P_0)) \leq K \frac{[\log(1/\epsilon)]^{5d/2-2}}{\epsilon}$$

for all $0 < \epsilon <$ some ϵ_0 and some constant $K < \infty$.

Proof. This follows by combining the results of Lemmas 4.3 and 4.4 with Lemma 4.1, and then using Corollary 6.2 of the bracketing entropy bound of Gao (2012) and stated here as Theorem 6.1. Here is the explicit calculation:

$$\begin{aligned}
&\log N_{[\,]}(6\epsilon, \mathcal{G}^{(conv)}, L_2(P)) \\
&\leq \log N_{[\,]}\left(\frac{\epsilon}{\sqrt{Q_{\sigma(\epsilon)}(\mathcal{X})}}, \mathcal{P}, L_2(\tilde{Q}_{\sigma(\epsilon)})\right) \qquad \text{by Lemma 4.1} \\
&\leq \log N_{[\,]}\left(\frac{\epsilon}{\sqrt{\frac{2^d c_1 c_2}{d!}[\log((c_1c_2)^3 \cdot 2^d/(\epsilon^2)]^d}}, \mathcal{P}, L_2(\tilde{Q}_{\sigma(\epsilon)})\right) \text{ by Lemmas 4.3 and 4.4} \\
&\leq \log N_{[\,]}\left(\frac{V\epsilon}{[\log(1/\epsilon)]^{d/2}}, \mathcal{P}, L_2(\tilde{Q}_{\sigma(\epsilon)})\right) \qquad \text{for } \ V = V_d(c_1, c_2) \\
&\leq K \frac{[\log(1/\epsilon)]^{d/2}}{V\epsilon}\left[\log\left(\frac{(\log(1/\epsilon))^{d/2}}{V\epsilon}\right)\right]^{2(d-1)} \qquad \text{by Corollary 6.2(b)} \\
&\leq \tilde{K}\frac{[\log(1/\epsilon)]^{5d/2-2}}{\epsilon}
\end{aligned}$$

for ϵ sufficiently small. $\square$

Proof. (Theorem 3.1) This follows from Lemma 4.5 and Theorem 7.6 of van de Geer (2000) or Theorem 3.4.1 of van der Vaart and Wellner (1996) together with the arguments given in Section 3.4.2. By Lemma 4.5 the bracketing entropy integrals

$$\begin{aligned}
J_{[\,]}(\delta, \mathcal{G}^{(conv)}, L_2(P_0)) &\equiv \int_0^\delta \sqrt{1 + \log N_{[\,]}(\epsilon, \mathcal{G}^{(conv)}, L_2(P_0))}\, d\epsilon \\
&\lesssim \int_0^\delta \epsilon^{-1/2}\left\{\log(1/\epsilon)\right\}^{3\gamma_d/2}\ d\epsilon
\end{aligned}$$

where the bound on the right side behaves asymptotically as a constant times $2\delta^{1/2}(\log(1/\delta))^{3\gamma_d/2}$ with $3\gamma_d \equiv 5d/2 - 2$, and hence (using the notation of Theorem 3.4.1 of van der Vaart and Wellner (1996)), we can take $\phi_n(\delta) = K2\delta^{1/2}(\log(1/\delta))^{3\gamma_d/2}$. Thus with $r_n \equiv n^{1/3}/(\log n)^\beta$ with $\beta = \gamma_d$ we find that $r_n^2\phi_n(1/r_n) \sim K\sqrt{n}$ and hence the claimed order of convergence holds. $\square$

5. Some related models and further problems

There are several related models in which we expect to see the same basic phenomenon as established here, namely a global convergence rate of the form $n^{-1/3}(\log n)^{\gamma}$ in all dimensions $d \geq 2$ with only the power γ of the log term depending on d. Three such models are:
(a) the "in-out model" for interval censoring in $\mathbb{R}^d$;
(b) the "case 2" multivariate interval censoring models studied by Deng and Fang (2009); and
(c) the scale mixture of uniforms model for decreasing densities in $\mathbb{R}^{+d}$.
Here we briefly sketch why we expect the same phenomenon to hold in these three cases, even though we do not yet know pointwise convergence rates in any of these cases.

5.1. The "in-out model" for interval censoring in $\mathbb{R}^d$

The "in-out model" for interval censoring in $\mathbb{R}^d$ was explored in the case $d = 2$ by Song (2001). In this model $\underline{Y} \sim F$ on $\mathbb{R}^2$, R is a random rectangle in $\mathbb{R}^2$ independent of $\underline{Y}$ (say $[\underline{U}, \underline{V}] = \{\underline{x} = (x_1, x_2) \in \mathbb{R}^2 : \ U_1 \leq x_1 \leq V_1, \ U_2 \leq x_2 \leq V_2\}$ where $\underline{U}$ and $\underline{V}$ are random vectors in $\mathbb{R}^2$ with $\underline{U} \leq \underline{V}$ coordinatewise). We observe only $(1_R(\underline{Y}), R)$, and the goal is to estimate the unknown distribution function F.

Song (2001) (page 86) produced a local asymptotic minimax lower bound for estimation of F at a fixed $\underline{t}_0 \in \mathbb{R}^2$. Under the assumption that F has a positive density f at $\underline{t}_0$, Song (2001) showed that any estimator of $F(\underline{t}_0)$ can have a local-minimax convergence rate which is at best $n^{-1/3}$. Groeneboom (2012a) has shown that this rate can be achieved by estimators involving smoothing methods. Based on the results for current status data in $\mathbb{R}^d$ obtained in Theorem 3.1 and the entropy results for the class of distribution functions on $\mathbb{R}^d$, we conjecture that the global Hellinger rate of convergence of the MLE $\hat{F}_n(\underline{t}_0)$ will be $n^{-1/3}(\log n)^{\nu}$ for all $d \geq 2$ where $\nu = \nu_d$.

5.2. "Case 2" multivariate interval censoring models in $\mathbb{R}^d$

Recall that "case 2" interval censored data on $\mathbb{R}$ is as follows: suppose that $\underline{Y} \sim F_0$ on $\mathbb{R}^+$, the pair of observation times (U, V) with $U \leq V$ determines a random interval $(U, V]$, and we observe $\underline{X} = (\underline{\Delta}, U, V) = (\Delta_1, \Delta_2, \Delta_3, U, V)$ where $\Delta_1 = 1\{Y \leq U\}$, $\Delta_2 = 1\{U < Y \leq V\}$, and $\Delta_3 = 1\{V < Y\}$. Nonparametric estimation of F_0 based on $\underline{X}_1, \ldots, \underline{X}_n)$ i.i.d. as $\underline{X}$ has been discussed by a number of authors, including Groeneboom and Wellner (1992), Geskus and Groeneboom (1999), and Groeneboom (1996). Deng and Fang (2009) studied generalizations of this model to $\mathbb{R}^d$, and obtained rates of convergence of the MLE with respect to the Hellinger metric given by $n^{-(1+d)/(2(1+2d)}(\log n)^{d^2/(2(2d+1)}$ in the case most comparable to the multivariate interval censoring model studied here. While this rate reduces when $d = 1$ to the known rate $n^{-1/3}(\log n)^{1/6}$, it is slower than $n^{-1/3}(\log n)^\nu$ for some ν when $d > 1$ due to the use of entropy bounds involving convex hulls (see Deng and Fang (2009), Proposition A.1, page 66) which are not necessarily sharp. We expect that rates of the form $n^{-1/3}(\log n)^\nu$ with $\nu > 0$ are possible in these models as well.

5.3. Scale mixtures of uniform densities on $\mathbb{R}^{+d}$

Pavlides (2008) and Pavlides and Wellner (2012) studied the family of scale mixtures of uniform densities of the following form:

$$f_G(\underline{x}) = \int_{\mathbb{R}^{+d}} \frac{1}{\prod_{j=1}^d y_j} 1_{(0,\underline{y}]}(\underline{x}) dG(\underline{y}) \equiv \int_{\mathbb{R}^{+d}} \frac{1}{|\underline{y}|} 1_{(0,\underline{y}]}(\underline{x}) dG(\underline{y}) \tag{5.1}$$

for some distribution function G on $(0,\infty)^d$. (Note that we have used the notation $\prod_{j=1}^d y_j = |\underline{y}|$ for $\underline{y} = (y_1, \ldots, y_d) \in \mathbb{R}^{+d}$.) It is not difficult to see that such densities are decreasing in each coordinate and that they also satisfy

$$(\Delta_d f_G)(\underline{u}, \underline{v}] = (-1)^d \int_{(\underline{u},\underline{v}]} |\underline{y}|^{-1} 1_{(\underline{y},\underline{v}]} dG(\underline{y}) \ge 0$$

for all $\underline{u}, \underline{v} \in \mathbb{R}^{+d}$ with $\underline{u} \le \underline{v}$; here Δ_d denotes the $d-$dimensional difference operator. This is the same key property of distribution functions which results in (bracketing) entropies which depend on dimension only through a logarithmic term. The difference here is that the density functions f_G need not be bounded, and even if the true density f_0 is in this class and satisfies $f_0(\underline{0}) < \infty$, then we do not yet know the behavior of the MLE $\hat{f}_n$ at zero. In fact we conjecture that: (a) If $f_0(\underline{0}) < \infty$ and f_0 is a scale mixture of uniform densities on rectangles as in (5.1), then $\hat{f}_n(\underline{0}) = O_p((\log n)^\beta)$ for some $\beta = \beta_d > 0$. (b) Under the same hypothesis as in (a) and the hypothesis that f_0 has support contained in a compact set, the MLE converges with respect to the Hellinger distance with a rate that is no worse than $n^{-1/3}(\log n)^\xi$ where $\xi = \xi_d$. Again Pavlides (2008) and Pavlides and Wellner (2012) establish asymptotic minimax lower bounds for estimation of $f_0(\underline{x}_0)$ proving that no estimator can have a (local minimax) rate of convergence faster than $n^{-1/3}$ in all dimensions. This is in sharp contrast to the class of block-decreasing densities on $\mathbb{R}^{+d}$ studied by Pavlides (2012) and by Biau and Devroye (2003): Pavlides (2012) shows that the local asymptotic minimax rate for estimation of $f_0(x_0)$ is no faster than $n^{-1/(d+2)}$, while Biau and Devroye (2003) show that there exist (histogram type) estimators $\tilde{f}_n$ which satisfy $E_{f_0}\|\tilde{f}_n - f_0\|_1 = O(n^{-1/(d+2)})$.

6. Appendix

We begin by summarizing the results of Gao (2012). For a (probability) measure μ on $[0,1]^d$, let $F \equiv F_\mu$ denote the corresponding distribution function given by

$$F(\underline{x}) = F_\mu(\underline{x}) = \mu([0,\underline{x}]) = \mu([0,x_1] \times \cdots \times [0,x_d])$$

for all $\underline{x} = (x_1, \ldots, x_d) \in [0,1]^d$. Let $\mathcal{F}_d$ denote the collection of all distribution functions on $[0,1]^d$; i.e.

$$\mathcal{F}_d = \{F : \ F \ \text{ is a distribution function on } \ [0,1]^d\}.$$

For example, if λ_d denotes Lebesgue measure on $[0,1]^d$, then the corresponding distribution function is $F(\underline{x}) = F_{\lambda_d}(\underline{x}) = \prod_{j=1}^d x_j$.

Theorem 6.1 (Gao, 2012)**.** *For $d \geq 2$ and $1 \leq p < \infty$*

$$\log N_{[\,]}(\epsilon, \mathcal{F}_d, L_p(\lambda_d)) \lesssim \epsilon^{-1} \left(\log(1/\epsilon)\right)^{2(d-1)}$$

for all $0 < \epsilon \leq 1$.

Our goal here is to use this result to control bracketing numbers for $\mathcal{F}_d$ with respect to two other measures C_d and $R_{d,\sigma}$ defined as follows. Let C_d denote the finite measure on $[0,1]^d$ with density with respect to λ_d given by

$$c_d(\underline{u}) = \frac{d!}{d^d} \prod_{j=1}^d \frac{1}{u_j^{1-1/d}} \cdot 1\left\{\sum_{j=1}^d u_j^{1/d} > d-1\right\}.$$

For fixed $\sigma > 0$, let $R_{d,\sigma}$ denote the (probability) measure on $(0,1]^d$ with density with respect to λ_d given by

$$r_{d,\sigma}(\underline{t}) = \frac{d!}{(\log(1/\sigma))^d} \frac{1}{\prod_{j=1}^d t_j} 1\left\{\prod_{j=1}^d t_j > \sigma\right\}.$$

Corollary 6.2. *(a) For each $d \geq 2$ it follows that for $\epsilon \leq \epsilon_0(d)$*

$$\log N_{[\,]}(2^{d/2}\epsilon, \mathcal{F}_d, L_2(C_d)) \lesssim \epsilon^{-1} \left(\log(1/\epsilon)\right)^{2(d-1)}.$$

(b) For each $d \geq 2$ and $\sigma \leq \sigma_0(d)$ it follows that for $\epsilon \leq \epsilon_0(d)/2$

$$\log N_{[\,]}(2^{d/2+1}\epsilon, \mathcal{F}_d, L_2(R_{d,\sigma})) \lesssim \epsilon^{-1} \left(\log(1/\epsilon)\right)^{2(d-1)}.$$

Proof. We first prove (a). We set $p \equiv p_d = 2r_d \equiv 2r$ where $r \equiv r_d = 2d-1$ and $s = (d-1/2)/(d-1)$ satisfy $r^{-1} + s^{-1} = 1$. Let $\{[g_j, h_j], j = 1, \ldots, m\}$ be a collection of $\epsilon-$brackets for $\mathcal{F}_d$ with respect to $L_p(\lambda_d)$. (Thus for $d = 2$, $r = 3$, $s = 3/2$, and $p = 6$, while for $d = 4$, $r = 7$, $s = (13/2)/3 = 13/6$, and $p = 14$.) By Theorem A.1 we know that $m \lesssim \epsilon^{-1}(\log(1/\epsilon))^{2(d-1)}$. Now we bound the size of the brackets $[g_j, h_j]$ with respect to C_d. Using Hölder's inequality with $1/r + 1/s = 1$ as chosen above we find that

$$\begin{aligned}\int_{[0,1]^d} (h_j - g_j)^2 c_d(u) du &\leq \left(\int_{[0,1]^d} |h_j - g_j|^{2r} \underline{du}\right)^{1/r} \cdot \left(\int_{[0,1]^d} c_d(\underline{u})^s \underline{du}\right)^{1/s} \\ &\leq (\epsilon^p)^{1/r} \cdot 2^{d/s} \leq 2^d \epsilon^2. \end{aligned} \quad (6.1)$$

Here are some details of the computation leading to (6.1):

$$
\begin{aligned}
\int_{[0,1]^d} c_d(\underline{u})^s \underline{du} &= \int_{[0,1]^d} \left(\frac{d!}{d^d}\right)^s \prod_{j=1}^d \frac{1}{u_j^{(d-1/2)/d}} \cdot 1\left\{\sum_{j=1}^d u_j^{1/d} > d-1\right\} \underline{du} \\
&= \left(\frac{d!}{d^d}\right)^s \cdot (2d)^d \int_{[0,1]^d} 1\left\{\sum_{j=1}^d x_j^2 > d-1\right\} \underline{dx} \\
&\leq \left(\frac{d!}{d^d}\right)^s \cdot (2d)^d \cdot \int_{[0,1]^d} 1\left\{\sum_{j=1}^d x_j > d-1\right\} \underline{dx} \\
&\leq \left(\frac{d!}{d^d}\right)^s \cdot (2d)^d \cdot \int_{[0,1]^d} 1\left\{\sum_{j=1}^d t_j < 1\right\} \underline{dt} \\
&= 2^d \left(\frac{d!}{d^d}\right)^{s-1} \leq 2^d.
\end{aligned}
$$

To prove (b) we introduce monotone transformations $t_j(u_j)$ and their inverses $u_j(t_j)$ which relate c_d and $r_{d,\sigma}$: we set

$$
\begin{aligned}
u_j(t_j) &\equiv \left(\frac{\log(t_j/\sigma)}{\log(1/\sigma)}\right)^d, \\
t_j(u_j) &\equiv \sigma \exp(u_j^{1/d} \log(1/\sigma))
\end{aligned}
$$

for $j = 1, \ldots, m$. These all depend on $\sigma > 0$, but this dependence is suppressed in the notation.

For the same brackets $[g_j, h_j]$ used in the proof of (a), we define new brackets $[\tilde{g}_j, \tilde{h}_j]$ for $j = 1, \ldots, m$ by

$$
\begin{aligned}
\tilde{g}_j(\underline{t}) &\equiv \tilde{g}_{j,\sigma}(\underline{t}) = g_j(u(\underline{t})) = g_j(u_1(t_1), \ldots, u_d(t_d)), \\
\tilde{h}_j(\underline{t}) &\equiv \tilde{h}_{j,\sigma}(\underline{t}) = h_j(u(\underline{t})) = h_j(u_1(t_1), \ldots, u_d(t_d))).
\end{aligned}
$$

Then it follows easily by direct calculation using

$$
\begin{aligned}
\prod_{j=1}^d t_j &= \sigma^d \exp\left(\log(1/\sigma) \sum_{j=1}^d u_j^{1/d}\right), \\
\underline{dt} &= \prod_{j=1}^d \left\{\sigma \exp(\log(1/\sigma) u_j^{1/d}) \cdot d^{-1} u_j^{1/d-1} \cdot \log(1/\sigma)(du_j)\right\} \\
&= \frac{\sigma^d (\log(1/\sigma))^d}{d^d} \prod_{j=1}^d t_j \cdot \prod_{j=1}^d u_j^{-(1-1/d)} \cdot \underline{du}
\end{aligned}
$$

$$\left\{\prod_{j=1}^{d} t_j > \sigma\right\} = \left\{\exp\left(\log(1/\sigma)\sum_{j=1}^{d} u_j^{1/d}\right) > \sigma^{-(d-1)}\right\}$$
$$= \left\{\log(1/\sigma)\sum_{j=1}^{d} u_j^{1/d} > (d-1)\log(1/\sigma)\right\}$$
$$= \left\{\sum_{j=1}^{d} u_j^{1/d} > d-1\right\},$$

that

$$\int_{[0,1]^d} (\tilde{h}_j(t) - \tilde{g}_j(t))^2 r_{d,\sigma}(t)\underline{dt} = \int_{[0,1]^d} (h_j(u) - g_j(u))^2 c_d(u)\underline{du}.$$

Thus for $\sigma \le \sigma_0(d)$ we have

$$\|\tilde{h}_j - \tilde{g}_j\|_{L_2(R_{d,\sigma})} \le 2^{d/2+1}\epsilon$$

by the arguments in (a). Hence the brackets $[\tilde{g}_j, \tilde{h}_j]$ yield a collection of $2^{d/2+1}\epsilon-$ brackets for $\mathcal{F}_d$ with respect to $L_2(R_{d,\sigma})$, and this implies that (b) holds. □

Acknowledgements

We owe thanks to the referees for a number of helpful suggestions and for pointing out the work of Yu, Yu and Wong (2006) and Deng and Fang (2009).

References

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T. and Silverman, E. (1955). An empirical distribution function for sampling with incomplete information. *Ann. Math. Statist.* **26** 641–647. MR0073895 (17,504f)

Balabdaoui, F. and Wellner, J. A. (2012). Chernoff's density is log-concave. Technical Report No. 512, Department of Statistics, University of Washington. available as arXiv:1207.6614.

Betensky, R. A. and Finkelstein, D. M. (1999). A non-parametric maximum likelihood estimator for bivariate interval-censored data. *Statistics in Medicine* **18** 3089-3010.

Biau, G. and Devroye, L. (2003). On the risk of estimates for block decreasing densities. *J. Multivariate Anal.* **86** 143–165. MR1994726 (2005c:62055)

Deng, D. and Fang, H.-B. (2009). On nonparametric maximum likelihood estimations of multivariate distribution function based on interval-censored data. *Comm. Statist. Theory Methods* **38** 54–74. MR2489672 (2010j:62139)

Dunson, D. B. and Dinse, G. E. (2002). Bayesian models for multivariate current status data with informative censoring. *Biometrics* **58** 79–88. MR1891046

GAO, F. (2012). Bracketing entropy of high dimensional distributions. Technical Report, Department of Mathematics, University of Idaho. "High Dimensional Probability VI", to appear.

GENTLEMAN, R. and VANDAL, A. C. (2002). Nonparametric estimation of the bivariate CDF for arbitrarily censored data. *Canad. J. Statist.* **30** 557–571. MR1964427 (2004b:62090)

GESKUS, R. and GROENEBOOM, P. (1999). Asymptotically optimal estimation of smooth functionals for interval censoring, case 2. *Ann. Statist.* **27** 627–674. MR1714713 (2000j:60044)

GROENEBOOM, P. (1987). Asymptotics for interval censored observations. Technical Report No. 87-18, Department of Mathematics, University of Amsterdam.

GROENEBOOM, P. (1989). Brownian motion with a parabolic drift and Airy functions. *Probab. Theory Related Fields* **81** 79–109. MR981568 (90c:60052)

GROENEBOOM, P. (1996). Lectures on inverse problems. In *Lectures on probability theory and statistics (Saint-Flour, 1994). Lecture Notes in Math.* **1648** 67–164. Springer, Berlin. MR1600884 (99c:62092)

GROENEBOOM, P. (2012a). The bivariate current status model. Technical Report No. ??, Delft Institute of Applied Mathematics, Delft University of Technology. available as arXiv:1209.0542.

GROENEBOOM, P. (2012b). Local minimax lower bounds for the bivariate current status model. Technical Report No. ??, Delft Institute of Applied Mathematics, Delft University of Technology. Personal communication.

GROENEBOOM, P., MAATHUIS, M. H. and WELLNER, J. A. (2008a). Current status data with competing risks: consistency and rates of convergence of the MLE. *Ann. Statist.* **36** 1031–1063.

GROENEBOOM, P., MAATHUIS, M. H. and WELLNER, J. A. (2008b). Current status data with competing risks: limiting distribution of the MLE. *Ann. Statist.* **36** 1064–1089.

GROENEBOOM, P. and WELLNER, J. A. (1992). *Information bounds and nonparametric maximum likelihood estimation. DMV Seminar* **19**. Birkhäuser Verlag, Basel. MR1180321 (94k:62056)

GROENEBOOM, P. and WELLNER, J. A. (2001). Computing Chernoff's distribution. *J. Comput. Graph. Statist.* **10** 388–400. MR1939706

JEWELL, N. P. (2007). Correspondences between regression models for complex binary outcomes and those for structured multivariate survival analyses. In *Advances in statistical modeling and inference. Ser. Biostat.* **3** 45–64. World Sci. Publ., Hackensack, NJ. MR2416109 (2009e:62407)

LIN, X. and WANG, L. (2011). Bayesian proportional odds models for analyzing current status data: univariate, clustered, and multivariate. *Comm. Statist. Simulation Comput.* **40** 1171–1181. MR2818097

MAATHUIS, M. H. (2005). Reduction algorithm for the NPMLE for the distribution function of bivariate interval-censored data. *J. Comput. Graph. Statist.* **14** 352–362. MR2160818

Maathuis, M. H. (2006). *Nonparametric estimation for current status data with competing risks*. ProQuest LLC, Ann Arbor, MI Thesis (Ph.D.)–University of Washington. MR2708977

Pavlides, M. G. (2008). *Nonparametric estimation of multivariate monotone densities*. ProQuest LLC, Ann Arbor, MI Thesis (Ph.D.)–University of Washington. MR2717518

Pavlides, M. G. (2012). Local asymptotic minimax theory for block-decreasing densities. *J. Statist. Plann. Inference* **142** 2322–2329. MR2911847

Pavlides, M. G. and Wellner, J. A. (2012). Nonparametric estimation of multivariate scale mixtures of uniform densities. *J. Multivariate Anal.* **107** 71–89. MR2890434

Schick, A. and Yu, Q. (2000). Consistency of the GMLE with mixed case interval-censored data. *Scand. J. Statist.* **27** 45–55. MR1774042

Song, S. (2001). Estimation with bivariate interval–censored data. PhD thesis, University of Washington, Department of Statistics.

Sun, J. (2006). *The Statistical Analysis of Interval-censored Failure Time Data*. *Statistics for Biology and Health*. Springer, New York. MR2287318 (2007h:62007)

van de Geer, S. (1993). Hellinger-consistency of certain nonparametric maximum likelihood estimators. *Ann. Statist.* **21** 14–44. MR1212164 (94c:62062)

van de Geer, S. A. (2000). *Applications of Empirical Process Theory*. *Cambridge Series in Statistical and Probabilistic Mathematics* **6**. Cambridge University Press, Cambridge. MR1739079 (2001h:62002)

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. *Springer Series in Statistics*. Springer-Verlag, New York. MR1385671 (97g:60035)

Wang, Y.-F. (2009). *Topics on multivariate two-stage current-status data and missing covariates in survival analysis*. ProQuest LLC, Ann Arbor, MI Thesis (Ph.D.)–University of California, Davis. MR2736679

Yu, S., Yu, Q. and Wong, G. Y. C. (2006). Consistency of the generalized MLE of a joint distribution function with multivariate interval-censored data. *J. Multivariate Anal.* **97** 720–732. MR2236498 (2007i:62068)