Bivariate Density Estimation With an Application to Survival Analysis

Charles KOOPERBERG

A procedure for estimating a bivariate density based on data that may be censored is described. After the data are transformed to the unit square, the bivariate density is estimated using linear splines and their tensor products. The combined procedure yields an estimate of the bivariate density on the original scale, which may provide insight about the dependence structure. The procedure can also be used to estimate densities that are known to be symmetric and to test for independence.

Key Words: Censoring; Correlated data; Splines; Test of independence.

1. INTRODUCTION

In this article we develop a procedure for estimating a bivariate density based on possibly censored data. The procedure consists of two stages. First, the univariate marginal distributions are estimated using hazard estimation with flexible tails (HEFT) (Kooperberg, Stone, and Truong 1995a) or logspline density estimation (LOGSPLINE) (Kooperberg and Stone 1992). Then, based on these estimates, the data are transformed to the unit square to have approximately uniform marginal distributions, after which the bivariate density is estimated using linear splines and their tensor products. The combined procedure, referred to as bivariate logspline density estimation, yields an estimate of the bivariate density on the original scale, which may provide insight about the dependence structure.

In a univariate setting, smooth estimates of density and hazard functions based on splines are useful for exploratory data analysis (Kooperberg and Stone 1992; Kooperberg, Stone, and Truong 1995a). Multivariate density estimation procedures can provide similar tools for correlated data. There are several univariate density estimation procedures that can deal with censored data (e.g., Kooperberg and Stone 1992; Koo, Kooperberg, and Park in press). However, while a number of procedures for estimating multivariate densities have been proposed (e.g., Koo 1996; Scott 1992), we are not aware of any bivariate density estimation procedure for dealing with censored data that has explicitly been studied.

Charles Kooperberg is Associate Member, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109-1024 (E-mail: clk@fhcrc.org).

Journal of Computational and Graphical Statistics, Volume 7, Number 3, Pages 322–341

Bivariate censored data arise, for example, in twin studies, where the age when one of the twins get a disease may be correlated with the age when his or her twin sibling gets the disease. Obviously, one or both of the twins may not get the disease at all; thus, different types of censoring are possible. The dependence between the survival times of the two twins may give us information about genetic or environmental influence on the disease.

Although much research in multivariate survival analysis has focused on methods for inference about the marginal survival times, there has also been substantial interest in studying dependence structures. In particular, a popular topic of research in multivariate survival analysis is nonparametric estimation of the survival function (e.g., Dabrowska 1988; Prentice and Cai 1992; Pruitt 1991). Although the bivariate survival function contains information about the dependence structure (see Dabrowska, Zhang, and Duffy 1995), this structure may be hard to visualize because of the discreteness of the estimates of the survival function that have been developed to date.

In Section 2 we describe logspline density estimates for bivariate survival data. In particular, in Section 2.3 we discuss how they can be used in estimating symmetric densities and in tests for independence of the components. Section 4 contains a small simulation study and the analysis of two real data sets. We end the article with a few remarks.

2. BIVARIATE DENSITY ESTIMATION FOR CENSORED DATA

In this section we describe a logspline density estimation procedure for bivariate data in which one or both components may be right censored. For reasons that will be discussed in detail in Section 2.2, the data are pretransformed from $[0, \infty)^2$ to $[0, 1]^2$ in such a manner that the marginal distributions are approximately uniform, after which the bivariate density of the transformed data is estimated. Thus, bivariate logspline models on the unit square and the associated model selection are first introduced. Then transformations from the domain of the data to the unit square and back again and the advantages of such transformations are discussed in the context of bivariate density estimation. Finally, some extensions and implications of the estimation methodology are considered.

2.1 LOGSPLINE DENSITY ESTIMATION ON THE UNIT SQUARE

Let $\mathbf{T} = (T_1, T_2)$ be a pair of random variables that takes values in $\mathcal{U} = [0, 1]^2$. In bivariate logspline density estimation the logdensity of $\mathbf{T}$ is modeled in a linear space G that is adaptively chosen from a family $\mathcal{G}$ of allowable spaces. The basis functions $B_1(\mathbf{t}), \ldots, B_p(\mathbf{t})$, $\mathbf{t} = (t_1, t_2) \in \mathcal{U}$, are said to span an allowable p-dimensional space G if each of the basis functions $B_j(\mathbf{t})$, $1 \leq j \leq p$, has one the following forms:

1. $B_j(\mathbf{t}) = t_l$, $l \in \{1, 2\}$;
2. $B_j(\mathbf{t}) = t_1 t_2$, where t_1 and t_2 are among the basis functions of G;
3. $B_j(\mathbf{t}) = (a_l - t_l)_+$, $l \in \{1, 2\}$, where $0 < a_l < 1$ and t_l is among the basis functions of G (here $x_+ = x$ if $x > 0$ and $x_+ = 0$ otherwise);

4. $B_j(\mathbf{t}) = (a_l - t_l)_+ t_{3-l}$, $l \in \{1,2\}$, where $(a_l - t_l)_+$, t_{3-l} and $t_1 t_2$ are among the basis functions of G;
5. $B_j(\mathbf{t}) = (a_1 - t_1)_+ (a_2 - t_2)_+$, where $(a_1 - t_1)_+$ and $(a_2 - t_2)_+$ are among the basis functions of G.

Basis functions of types 1 and 3 are said to depend on one component, and the other basis functions are referred to as tensor product basis functions; a_l is referred to as a knot in t_l. A function g in G is referred to as a bivariate linear spline.

Given $\boldsymbol{\beta} \in \mathbf{R}^p$, set

$$f(\mathbf{t};\boldsymbol{\beta}) = \exp(\beta_1 B_1(\mathbf{t}) + \cdots + \beta_p B_p(\mathbf{t}) - C(\boldsymbol{\beta})), \qquad \mathbf{t} \in \mathcal{U}, \tag{2.1}$$

where

$$C(\boldsymbol{\beta}) = \log\left(\int_{\mathcal{U}} \exp(\beta_1 B_1(\mathbf{t}) + \cdots + \beta_p B_p(\mathbf{t}))d\mathbf{t}\right). \tag{2.2}$$

Then $f(\cdot;\boldsymbol{\beta})$ is a continuous positive density on $\mathcal{U}$ for $\boldsymbol{\beta} \in \mathbf{R}^p$. For $\boldsymbol{\beta} \in \mathbf{R}^p$ and $0 \le t_1 \le 1$, $\log f(\mathbf{t};\boldsymbol{\beta})$ is a linear function in t_2; for $\boldsymbol{\beta} \in \mathbf{R}^p$ and $0 \le t_2 \le 1$, $\log f(\mathbf{t};\boldsymbol{\beta})$ is a linear function in t_1. Two random variables having $f(\cdot;\boldsymbol{\beta})$ as their joint density are independent if and only if none of the basis functions depend on both t_1 and t_2.

Consider n randomly selected pairs $\mathbf{T}_i = (T_{1i}, T_{2i})$, $1 \le i \le n$. It is assumed that $\mathbf{T}_i$ has density $f(\cdot)$ on $\mathcal{U}$ and that the marginal densities of T_{1i} and T_{2i} are uniform on $[0,1]$. For $1 \le i \le n$ and $l \in \{1,2\}$ let C_{li} be the censoring time for the lth component of the ith pair ($C_{li} = 1$ if this component is uncensored). Set $Y_{li} = \min(T_{li}, C_{li})$ and $\delta_{li} = \text{ind}(T_{li} \le C_{li})$. Also set $\mathbf{Y}_i = (Y_{1i}, Y_{2i})$, $\mathbf{C}_i = (C_{1i}, C_{2i})$, and $\boldsymbol{\delta}_i = (\delta_{1i}, \delta_{2i})$. It is assumed that $\mathbf{T}$ and $\mathbf{C}$ are independent. The random variable $\mathbf{Y}$ is said to be uncensored if $\boldsymbol{\delta} = (1,1)$, censored in the first component if $\boldsymbol{\delta} = (0,1)$, censored in the second component if $\boldsymbol{\delta} = (1,0)$, and doubly censored if $\boldsymbol{\delta} = (0,0)$. The log-likelihood corresponding to $\mathbf{Y}_i = \mathbf{y}_i$, $\boldsymbol{\delta}_i$ and $\boldsymbol{\beta}$ is given by

$$\phi(\mathbf{y}_i, \boldsymbol{\delta}_i, \boldsymbol{\beta}) = f(\mathbf{y}_i;\boldsymbol{\beta}) \qquad \text{if} \quad \boldsymbol{\delta}_i = (1,1),$$

$$\phi(\mathbf{y}_i, \boldsymbol{\delta}_i, \boldsymbol{\beta}) = \int_{y_{1i}}^1 f(u_1, y_{2i};\boldsymbol{\beta})du_1 \qquad \text{if} \quad \boldsymbol{\delta}_i = (0,1),$$

$$\phi(\mathbf{y}_i, \boldsymbol{\delta}_i, \boldsymbol{\beta}) = \int_{y_{2i}}^1 f(y_{1i}, u_2;\boldsymbol{\beta})du_2 \qquad \text{if} \quad \boldsymbol{\delta}_i = (1,0),$$

and

$$\phi(\mathbf{y}_i, \boldsymbol{\delta}_i, \boldsymbol{\beta}) = \int_{y_{1i}}^1 \int_{y_{2i}}^1 f(u_1, u_2;\boldsymbol{\beta})du_2 du_1 \qquad \text{if} \quad \boldsymbol{\delta}_i = (0,0).$$

The log-likelihood function for the observed data is given by

$$\ell(\boldsymbol{\beta}) = \sum_i \log \phi(\mathbf{y}_i, \boldsymbol{\delta}_i, \boldsymbol{\beta}), \quad \boldsymbol{\beta} \in \mathcal{B}.$$

Straightforward manipulation of these expressions yields formulas for the score function and Hessian, which are similar to those in Kooperberg and Stone (1992). The Hessian is guaranteed to be negative definite only when all the data are uncensored, but the possible lack of negative definiteness of the Hessian does not appear to be a problem in practice. The log-likelihood function can be maximized to obtain the maximum likelihood estimate of $\boldsymbol{\beta}$ using, for example, a quasi-Newton algorithm.

The choice of G from the family $\mathcal{G}$ of allowable spaces is carried out using stepwise deletion of basis functions. In particular, let $K_1 = \max\{0, \lfloor 8(\sum_i \delta_{1i})^{.1} - 10 \rfloor\}$ and let the potential knot a_{1k} in t_1 be the empirical $k/(K_1+1)$th quantile of $\{Y_{1i} : \delta_{1i} = 1\}$. The potential knots a_{2k} in t_2 are selected in a similar manner. Initially a saturated model is fit; that is, a model with basis functions t_1, t_2, t_1t_2, $t_1(a_{2k_2} - t_2)_+$, $(a_{1k_1} - t_1)_+t_2$, and $(a_{1k_1} - t_1)_+(a_{2k_2} - t_2)_+$, $k_1 = 1, \ldots, K_1$ and $k_2 = 1, \ldots, K_2$. (However, basis functions of the form $(a_{1k_1} - t_1)_+(a_{2k_2} - t_2)_+$ for which fewer than three uncensored observations are within each of the four quadrants defined by the lines $t_1 = a_{1k_1}$ and $t_2 = a_{2k_2}$ and basis functions that may otherwise cause singularities are not included in this model.) The somewhat complicated form of the formula for K_l is partly motivated by results about the L_2 convergence rates for nonadaptive versions of polynomial spline routines (Stone, Hansen, Kooperberg, and Truong 1997); practical experience leads us to choose the constants 8, 10, and .1.

After the initial model is fit the p-dimensional allowable space G is successively replaced by a $(p-1)$-dimensional allowable subspace G_0 until no basis functions are left, so that the "model" $f(\mathbf{t}) = 1$ is finally considered. An arbitrary basis function depending on one component is removed if the remaining subspace is allowable; otherwise a tensor product basis function is removed, with that candidate space being chosen that decreases the quadratic approximation to the log-likelihood function the least. (The usual Wald statistic for testing that the coefficient of a basis function equals 0 is twice the decrease of the quadratic approximation to the log-likelihood function (Stone, Hansen, Kooperberg, and Truong 1997).) Because it is assumed that the marginal distributions of T_1 and T_2 are uniform, the coefficients of the basis functions that depend on one component are typically insignificant unless the corresponding basis function also appears as part of a tensor product basis function.

During this stepwise deletion procedure, a sequence of models indexed by ν is obtained, with the νth model having p_ν tensor product basis functions. The generalized Akaike information criterion (AIC) can be used to select a model from this sequence. Let $\hat{\ell}_\nu$ denote the log-likelihood for the νth model, and let $\text{AIC}_{a,\nu} = -2\widehat{\ell_\nu} + ap_\nu$ be the Akaike information criterion with penalty parameter a for this model. Among the models from which no basis functions of the form t_1, t_2, $(a_{1k_1} - t_1)_+$ or $(a_{2k_2} - t_2)_+$ can be removed, we select the model corresponding to the value of ν that minimizes $\text{AIC}_{a,\nu}$. In light of practical experience, we generally recommend choosing $a = \log n$ as in the Bayesian information criterion (BIC).

Because the coefficients of the basis functions that depend on one component are typically insignificant, the log-likelihood is not penalized for inclusion of these basis functions in the model (to do so would effectively require the tensor product basis functions to compensate for several parameters).

2.2 TRANSFORMING DATA TO THE UNIT SQUARE

It is possible to apply a procedure similar to the one described in the previous subsection to data on $(-\infty,\infty)^2$ or $[0,\infty)^2$. For a variety of reasons, however, it is better first to transform the data to the unit square. Since these reasons are the same for densities on $(-\infty,\infty)^2$ and on $[0,\infty)^2$, assume that we wish to estimate a bivariate density on $[0,\infty)^2$ using a logspline model (2.1) and based directly on the untransformed data. Then, for a given set of basis functions $B_1,\ldots,B_p$ only vectors $\boldsymbol{\beta}\in\mathbb{R}^p$ for which the normalizing constant

$$C(\boldsymbol{\beta})=\int_0^\infty\int_0^\infty \exp(\beta_1B_1(\mathbf{t})+\cdots+\beta_pB_p(\mathbf{t}))dt_2dt_1$$

given by (2.2) is finite or, equivalently,

$$\lim_{t_l\to\infty}\beta_1B_1(\mathbf{t})+\cdots+\beta_pB_p(\mathbf{t})=-\infty,\quad t_{3-l}\in[0,\infty)\quad\text{and}\quad l\in\{1,2\},\qquad(2.3)$$

yield densities on $[0,\infty)^2$. These conditions result in a number of linear constraints on the coefficients. The decrease of the quadratic approximation to the log-likelihood function is no longer the usual Wald statistic since the constraints (2.3) have to be handled; however, it is conveniently computed using a quadratic optimization package such as LSSOL (Gill, Murray, Saunders, and Wright 1986a).

Although this approach appears straightforward, its problems become more apparent when we consider the quadrant of the plane where $t_1>a_{1K_1}$ and $t_2>a_{2K_2}$. In each rectangle on which the logdensity is a polynomial, it is of the form $b_0+b_1t_1+b_2t_2+b_3t_1t_2$. Therefore, to satisfy (2.3) for $t_1>a_{1K_1}$ or for $t_2>a_{2K_2}$ it is required that $b_3\le 0$. However, only tensor product basis functions can model dependence between T_1 and T_2. A positive dependence is associated with a positive coefficient of t_1t_2. Thus, a positive dependence for values of both random variables larger than the largest knot cannot be modeled.

An alternative to direct estimation of the joint density of T_1 and T_2 is first to estimate the marginal densities and then to transform the data using the estimated marginal distribution functions. In our approach we use hazard estimation with flexible tails (HEFT) (Kooperberg, Stone, and Truong 1995a) or (univariate) logspline density estimation (LOGSPLINE) (Kooperberg and Stone 1992) to estimate the marginal densities. Both methodologies use cubic splines. LOGSPLINE models the logdensity and HEFT the log-hazard function. HEFT can be used only when the range of the data is $[0,\infty)$.

Let $\hat f_1$ and $\hat f_2$ be estimates of the marginal densities of T_1 and T_2 respectively, and let $\widehat F_1$ and $\widehat F_2$ be the corresponding estimated distribution functions. Set $Y_{li}^*=\widehat F_l(Y_{li})$, $l\in\{1,2\}$, and apply the procedure described in the previous subsection to $\mathbf{Y}_i^*=\{Y_{1i}^*,Y_{2i}^*\}$ and $\boldsymbol{\delta}_i$, $i=1,\ldots,n$, to obtain an estimate $\hat f^*$ of the density of $\mathbf{T}^*=(\widehat F_1(T_1),\widehat F_2(T_2))$ on $\mathcal{U}$. The logspline estimate of the bivariate density of $\mathbf{T}$ is then given by

$$\hat f(\mathbf{t})=\hat f^*(\widehat F_1(t_1),\widehat F_2(t_2))\hat f_1(t_1)\hat f_2(t_2),\qquad 0\le t_1,t_2<\infty.\qquad(2.4)$$

Logspline estimates of the distribution function $F(\mathbf{t})$ and survival function $S(\mathbf{t})=$

$P(\mathbf{T} \geq \mathbf{t}) = P(T_1 \geq t_1, T_2 \geq t_2)$ are given by

$$\widehat{F}(\mathbf{t}) = \int_0^{\widehat{F}(t_1)} \int_0^{\widehat{F}(t_2)} \hat{f}(u_1, u_2) du_1 du_2,$$

and

$$\widehat{S}(\mathbf{t}) = \int_{\widehat{F}(t_1)}^1 \int_{\widehat{F}(t_2)}^1 \hat{f}(u_1, u_2) du_1 du_2.$$

Another advantage of transforming the data to the unit square is that the number of knots in the final model is typically much smaller in the estimation of the bivariate density on the unit square (often zero or one in each component) than it would be if the density were estimated directly. Moreover, since the estimates of the marginal densities are smooth, the final estimate of the bivariate density is smoother than it would be if it were estimated directly.

From a theoretical viewpoint we lose nothing by first transforming the data. In Stone (1994) and Kooperberg, Stone, and Truong (1995b) it is established that the L_2 convergence rate for nonadaptive versions of logspline density estimation and hazard estimation with flexible tails is $n^{-p/(2p+d)}$, where p is the degree of the spline and d is the dimension of the domain of the function that is estimated provided a number of mild conditions on f are satisfied. Thus, the convergence rate for nonadaptive versions of the univariate procedures LOGSPLINE and HEFT is $n^{-3/7}$ since $d = 1$ and $p = 3$ (cubic splines), while the rate for bivariate logspline density estimation is $n^{-1/4}$ since $d = 2$ and $p = 1$ (linear splines). (The degree of smoothness required of f to get the indicated rate of convergence is larger for the procedures with $p = 3$ than for the procedure with $p = 1$.) Thus, the convergence rate for the univariate procedures is so much faster that in practice the convergence properties are determined by the bivariate procedure, whose rate is independent of a possible transformation.

2.3 MORE ABOUT BIVARIATE LOGSPLINE DENSITY ESTIMATION

2.3.1 Symmetric Densities

The framework for bivariate density estimation described in this article is particularly well suited to applications in which it is known in advance that the density is symmetric in its arguments—that is, $f(t_1, t_2) = f(t_2, t_1)$ for all $\mathbf{t}$. In such situations HEFT or LOGSPLINE can be used to estimate one marginal density based on the combined data for the two components, while symmetrized versions of the basis functions can be used to estimate the density on the unit square. Thus, when the density is symmetric, $t_1 + t_2$, $t_1 t_2$, $(a - t_1)_+ + (a - t_2)_+$, $(a - t_1)_+ t_2 + t_1(a - t_2)_+$, $(a - t_1)_+(a - t_2)_+$, and $(a_1 - t_1)_+(a_2 - t_2)_+ + (a_2 - t_1)_+(a_1 - t_2)_+$ with $a_1 \neq a_2$ are used as basis functions. (Note that in this context the basis functions $t_1 + t_2$ and $(a - t_1)_+ + (a - t_2)_+$ play the role of the basis functions that depend on one component in Section 2.1: they do not count in the AIC criterion, and they are removed whenever possible.) The twin study in Section 4.2 is an example in which the density is known a priori to be symmetric.

2.3.2 Integration

Because linear splines are used to model the logdensity on the unit square, all bivariate integrals can be reduced analytically to univariate integrals. The resulting univariate integrals are closely related to the exponential integral (Abramowitz and Stegun 1965), for which fast algorithms exist. This is particularly important when some data may be censored, since integrals then have to be computed for every censored observation. If cubic splines were used to model the density, it would no longer be possible to reduce the bivariate integrals to univariate ones. (Note that if cubic splines were used to model the density, it would be even more critical first to transform the data to the unit square to avoid complications because of tail constraints.)

2.3.3 Numerical Details

After the data have been transformed to the unit square, we use NPSOL (Gill, Murray, Saunders, and Wright 1986b) to find the maximum likelihood estimates. Although the log-likelihood function is not necessarily concave when some data are censored, we have not experienced any numerical difficulties when the knots are positioned such that there are no rectangles $(t_{1i}, t_{1i+1}) \times (t_{2j}, t_{2j+1})$ without a couple of uncensored observations. This experience is in agreement with the experience of Kooperberg and Stone (1992) in the context of univariate logspline density estimation. The real and simulated data sets in Section 4 all took less than 10 seconds of CPU time on the Sparc ULTRA workstation that we used. A crucial aspect of getting fast code is the efficient organization of the computation of the score function and the Hessian involving the integrals discussed previously.

2.3.4 Testing for Independence

There are several ways that model selection can be used to develop tests for independence. One possibility is to fit a model

$$f(t_1, t_2) = \exp(b_1 t_1 + b_2 t_2 + b_3 t_1 t_2 - C(\boldsymbol{\beta})), \qquad \mathbf{t} \in \mathcal{U}, \tag{2.5}$$

without knots to the density of the transformed data and to use a likelihood ratio test of the hypothesis $H_0 : b_3 = 0$. Alternatively, we can forgo the formal testing procedure and see if the AIC-based model selection procedure keeps any interaction terms in the model.

Tests for independence for uncensored data are well established. Tests for independence in the context of censored data are often based on ranks (see, for example, Oakes 1982). Commenges and Andersen (1995) and Gray (1995) gave procedures that can be used for tests of independence in the context of the proportional hazards and frailty models.

2.3.5 Odds Ratios

In many epidemiological studies the odds ratio is the preferred measure of dependence between two random variables. A local version of this ratio can be defined as

$$\theta(\mathbf{t}) = \frac{f(\mathbf{t})S(\mathbf{t})}{\int_{t_1}^{\infty} f(v_1, t_2)dv_1 \int_{t_2}^{\infty} f(t_1, v_2)dv_2}$$

(Clayton 1978; Oakes 1989). The relation with the odds ratio is clarified by rewriting θ as a ratio of conditional hazard rates:

$$\theta(\mathbf{t}) = \frac{\lambda_{T_1}(t_1|T_2 = t_2)}{\lambda_{T_1}(t_1|T_2 > t_2)} = \frac{\lambda_{T_2}(t_2|T_1 = t_1)}{\lambda_{T_2}(t_2|T_1 > t_1)}.$$

In particular, when $\theta(\mathbf{t}) > 1$ there is a local positive association, and when $\theta(\mathbf{t}) < 1$ there is a local negative association. In certain circumstances the local odds ratio, together with the marginal distributions, determines the bivariate distribution of $\mathbf{T}$ (Oakes 1989). In the context of bivariate logspline density estimation it is straightforward to compute the estimate $\hat{\theta}(\mathbf{t})$ of the local odds ratio corresponding to the density estimate $\hat{f}(\mathbf{t})$. Note that the local odds ratio is invariant under monotone transformations of the individual coordinates; thus, in the notation of (2.4),

$$\hat{\theta}(\mathbf{t}) = \hat{\theta}^*(\widehat{F}_1(t_1), \widehat{F}_2(t_2)), \qquad 0 \le t_1, t_2 < \infty.$$

It is easy to establish that if f_2^* is a continuous, positive density function on $[0, 1]^2$, then $\lim_{t_1 \uparrow 1} \theta^*(t_1, t_2) = 1$ for $0 \le t_2 < 1$ and $\lim_{t_2 \uparrow 1} \theta^*(t_1, t_2) = 1$ for $0 \le t_1 < 1$. In particular, for bivariate logspline models, the local odds ratio is 1 if either $t_1 = 1$ or $t_2 = 1$ on the transformed scale or, equivalently, if t_1 or t_2 approaches infinity on the original scale.

3. RELATED WORK

3.1 BIVARIATE DENSITY ESTIMATION USING SPLINES

In a recent paper Koo (1996) introduced tensor logspline density estimation (TELDE). This procedure, like the one described in the present paper, is a bivariate density estimation methodology based on tensor product splines and a stepwise deletion algorithm. However, there are a number of differences between TELDE and bivariate logspline density estimation: (1) TELDE cannot deal with censored data; (2) since in TELDE the data are not transformed to a bounded region the tail behavior is much harder to control (Koo (1996) did not fit tails to his density, but rather restricted his estimates to the rectangle shown in the figures in his paper; Ja-Yong Koo, private communication); and (3) TELDE employs cubic splines and therefore usually produces smoother estimates than bivariate logspline density estimation. Although in principle it would be possible to employ cubic splines in bivariate logspline density estimation, the integrals can no longer be reduced to univariate integrals, which is an important consideration when some data are censored (see Sec. 2.3).

3.2 BIVARIATE SURVIVAL ANALYSIS

In univariate survival analysis the distribution of the survival times is uniquely determined by its hazard function $\lambda(t) = f(t)/S(t)$. For bivariate survival analysis, however, such a unique representation in terms of hazard functions does not exist. Although there are several consistent estimators of the bivariate distribution function, efficient estimation has proven to be a difficult problem (van der Laan 1996). The two representations of the bivariate survival distribution that are most commonly used are due to Dabrowska (1988) and Prentice and Cai (1992). Andersen, Borgan, Gill, and Keiding (1993) discussed both representations and their connections in detail.

Dabrowska (1988) modeled the joint survival function in terms of two conditional hazard functions and a "double" hazard function, representing the instantaneous rate of "double failures" at time $\mathbf{t}$ given that both components were alive until that time. Cumulative versions of each of these three hazard functions can be estimated using a Kaplan–Meier type estimate. For further properties of this estimator see Dabrowska (1989), Gill, van der Laan, and Wellner (1995), and Pruitt (1991). Prentice and Cai (1992) modeled the bivariate survival function in terms of the two marginal cumulative hazard functions and a covariance rate function, which measures the covariance between two martingales defined in terms of counting processes.

Both representations have attractive properties, such as the direct link to the Kaplan–Meier estimate (Dabrowska representation) or the relation to the estimation of dependence parameters and the possibility of combining the bivariate model with a marginal proportional hazards model (Prentice and Cai representation). However, neither representation is well suited for estimation with a polynomial spline model since there are no explicit conditions under which the double hazard function (Dabrowska representation) or the covariance rate function (Prentice and Cai representation) yields a valid joint survival function. As a result, the optimization problem has complicated constraints which are hard to handle; also, the corresponding log-likelihood is nonconcave even if none of the data are censored.

Neither representation yields a direct estimate of the bivariate density. While it is possible to obtain an estimate of the bivariate density as the derivative of the convolution of any estimate of the bivariate distribution function with a kernel, we know of no methodology other than that presented in this article for directly estimating a bivariate density when some data are censored.

4. EXAMPLES

4.1 SIMULATED EXAMPLES

Clayton (1978) proposed the family of bivariate distributions

$$S(\mathbf{t};\theta) = \begin{cases} [S_1(t_1)^{1-\theta} + S_2(t_2)^{1-\theta} - 1]^{-(\theta-1)^{-1}}, & t_1, t_2 \geq 0, \quad \theta > .5, \quad \theta \neq 1, \\ S_1(t_1)S_2(t_2) & t_1, t_2 \geq 0, \quad \theta = 1, \end{cases} \tag{4.1}$$

where S_1 and S_2 are the marginal survival functions of $T_1 \geq 0$ and $T_2 \geq 0$. When $\mathbf{T}$ has such a distribution, the local odds ratio $\theta(\mathbf{t})$ equals θ for all $\mathbf{t}$, independently of the

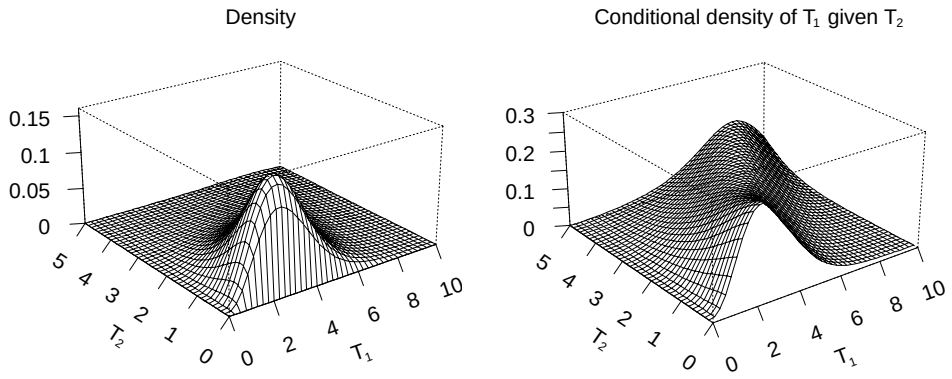

Figure 1. True Bivariate Density From Which Data Were Generated for the Simulated Example (correlation is .6).

marginal distributions. The random variables T_1 and T_2 are positively correlated if $\theta > 1$, negatively correlated if $\theta < 1$, and independent if $\theta = 1$.

In the following examples $\mathbf{T}$ is generated from (4.1) with T_1 having the gamma distribution with shape parameter 4 and mean 4 and T_2 having the gamma distribution with shape parameter 1.5 and mean 1.5. In Figure 1 we show the density of $\mathbf{T}$ and the conditional density of T_1 given T_2 when $\theta = 2$ (left), so that $\text{cor}(T_1, T_2) = .605$. Perspective plots of the bivariate density tend to look very similar for a wide range of values of θ; plots of the conditional density, like the one on the right side, show more dependence on θ.

In the simulation studies we looked at the performance of the bivariate logspline density estimate for a variety of values of θ with both uncensored data and data censored by independent exponential random variables. For the censored data, C_1 was generated from the exponential distribution with mean 8 and C_2 was generated from the exponential distribution with mean 3, yielding about 38% censoring for T_1 and 35% censoring for T_2. We considered sample sizes of $n = 100$ and $n = 500$ pairs.

Table 1 summarizes the simulations with 100 pairs and with 500 pairs. For both the censored and the uncensored data we report how often, out of 200 simulations, the logspline procedure estimated a model in which T_1 and T_2 were independent (using AIC with the default parameter of $\log n$ for model selection) and how often the test $H_0 : b_3 = 0$ was not rejected at the 5% level (see Sec. 2.3).

We see from Table 1 that the test for independence based on the bivariate logspline model has good power, even when θ is small. Actually, the formal likelihood ratio test appears to perform almost the same as the model selection procedure using BIC.

As a measure of accuracy of the estimate of the bivariate density after back-transformation, we computed the integrated squared error (ISE) between the true underlying density and the estimated density, $\int_0^\infty \int_0^\infty [\hat{f}(\mathbf{t}) - f(\mathbf{t}; \theta = 2)]^2 dt_1 dt_2$ for each of the simulated densities with $\theta = 2$. By comparison, the integrated squared difference between the underlying density with $\theta = 2$ and the underlying density with $\theta = 1$ (i.e., when T_1 and T_2 are independent) is .00556. The results are summarized in Table 2. From

Table 1. Results of 200 Simulations

		n=100				n=500			
		Uncensored		Censored		Uncensored		Censored	
θ	cor($F_1(T_1)$, $F_2(T_2)$)	indep. models	$b_3 = 0$ not rejected	indep. models	$b_3 = 0$ not rejected	indep. models	$b_3 = 0$ not rejected	indep. models	$b_3 = 0$ not rejected
1.0	.000	188	192	180	185	198	189	200	193
1.2	.138	140	160	163	166	46	28	123	117
1.4	.247	39	44	106	102	0	0	9	7
1.6	.338	8	8	57	58	0	0	0	0
1.8	.415	1	1	28	23	0	0	0	0
2.0	.478	0	0	6	4	0	0	0	0

this table we see that the density estimates are quite accurate when $n = 500$, but they are substantially more variable when $n = 100$. The amount of censoring in this example seems to have relatively little influence on the rate of change of the ISE as n changes: for both the censored and the uncensored data, the average ISE is reduced by about a factor of 3–4 when the sample size is increased from 100 to 500.

In Figure 2 we show the estimates of the density and conditional density for the simulation that was at the median ISE among the simulations with $n = 500$ and censored data. In the right side of this figure for large values of T_2, it is easy to recognize the location of a knot in t_1. Actually, for this data set 77 observations were doubly censored; for 103 observations T_1 was censored but not T_2; for 105 observations T_2 was censored but not T_1; and 215 observations were uncensored. The ISE was .00175. After examining this and many similar figures, we are convinced that the bivariate logspline density estimation procedure does a good job in estimating bivariate densities, even when some of the data are censored.

In many practical situations the censoring distribution will essentially be supported on a bounded subset of the failure distribution; for example, because of the limited period of follow-up in clinical research or because the majority of people never get a certain disease (we could say they die before they get the disease). This is the case for the Australia twin data example in the next section. In such situations the interest is typically in the density restricted to the region where data are observed. While the bivariate logspline density estimate provides an estimate on $[0, \infty)^2$, the procedure really extrapolates outside the region where data are observed. In particular, the univariate logspline and HEFT procedures extrapolate the marginal densities using exponential and Weibull-like tails, respectively.

When the data are limited to a region $[0, c_1] \times [0, c_2]$ there are two possible approaches to bivariate density estimation:

Table 2. Integrated Squared Error (ISE) with $\theta = 2$

	$n = 100$		$n = 500$	
	uncensored	censored	uncensored	censored
Average ISE	.00499	.00651	.00133	.00196
SD ISE	.00334	.00453	.00062	.00097

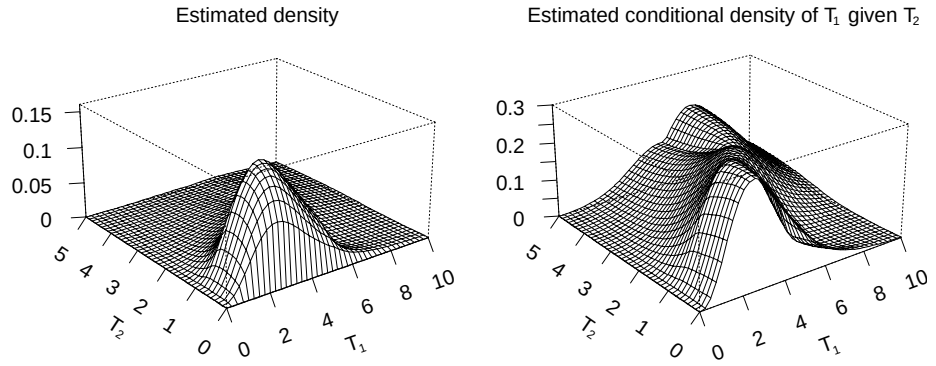

Figure 2. Bivariate Logspline Density Estimate and the Corresponding Conditional Density of T_1 given T_2; n =500, 36% Censoring.

- we can use only the data for which both times are within this region, and estimate the conditional density of $\mathbf{T}$ given that $t_1 \leq c_1$ and $t_2 \leq c_2$; or
- we can use all the data to estimate the density of T on $[0, \infty)^2$, but only consider the estimate on $[0, c_1] \times [0, c_2]$.

We would expect the second approach, which is the one used by the bivariate logspline procedure, to be more accurate, unless the dependence is extreme, since it makes use of the data for which one of the two components is uncensored to estimate the marginal density, which would conceivably help in estimating the marginal densities more accurately. We carried out a small simulation study to confirm this.

For the computations reported in Table 3 we generated $\mathbf{T}$ from the same distributions as before, but now the censoring times are fixed at $C_1 = 4.5$ and $C_2 = 1.7$ for each observation. This results in about 34% censoring of T_1 and 33% censoring of T_2; for the bivariate data between 56% ($\theta = 0$) and 47% ($\theta = 2$) were censored. In Table 3 we report the average ISE over 200 simulations on the rectangle $[0, 4.5] \times [0, 1.7]$ for the logspline procedure using only the uncensored data to estimate the density function (the conditional density function, rescaled by the inverse of the fraction of data that were uncensored) and for the logspline procedure using all the data. As a comparison we also give the ISE over the relevant rectangle based upon the completely uncensored data. As can be seen from this table, it is indeed advantageous to include the singly censored

Table 3. Integrated Squared Error (ISE) on [0,4.5]×[0,1.7]

	n = 100			*n = 500*		
θ	*using only uncensored data*	*using all data*	*without censoring*	*using only uncensored data*	*using all data*	*without censoring*
1.0	.00262	.00087	.00062	.00068	.00016	.00010
1.2	.00279	.00116	.00097	.00071	.00023	.00020
1.4	.00326	.00137	.00125	.00077	.00033	.00026
1.6	.00344	.00149	.00146	.00088	.00049	.00033
1.8	.00360	.00189	.00169	.00095	.00064	.00037
2.0	.00378	.00224	.00191	.00121	.00081	.00041

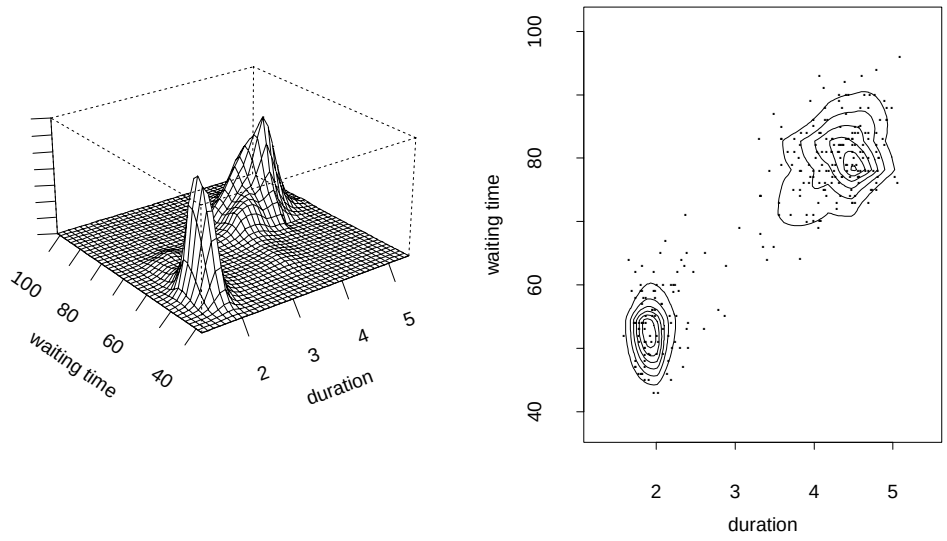

Figure 3. Density Estimate for the Geyser Data.

observations even if there is a fairly strong dependence.

4.2 REAL EXAMPLES

4.2.1 Geyser Data

Although bivariate logspline density estimation was developed with survival analysis in mind, it is interesting to see how it performs on data without censoring when the underlying density is bimodal. The geyser data set consists of 274 measurements of the duration in minutes of an eruption of the Old Faithful geyser and the waiting time until the next eruption of this geyser. This data set has been widely used in the literature. There are several versions of the data set; here we use the one published in Härdle (1990), which was also used by Koo (1996). Figure 3 contains a perspective plot and a contour plot of the bivariate logspline estimate (the data are displayed in the contour plot).

For the bivariate logspline estimate, both marginal distributions were first estimated using LOGSPLINE. After the data were transformed to the unit square, we computed the bivariate logspline estimate using four knots in each variable, resulting in a 35-dimensional starting model. The stepwise deletion procedure removed 13 interaction terms. None of the knots were completely removed.

When we compare this figure to the estimates shown in Härdle (1990) and Koo (1996) we note that the bivariate logspline procedure provides reasonable estimates. In particular, keeping in mind that while the bivariate part of the procedure is based only on linear splines, the estimate is smoother than expected. It is our experience that the bivariate logspline density estimation procedure usually gives reasonable estimates, even when the density has several modes. However, because of the piecewise linear nature of

the estimates and the relatively small number of knots that are being used, the estimates may sometimes be somewhat rough, for example, in comparison to those presented in Koo (1996). The power of the bivariate logspline density estimation procedure becomes more evident when we consider censored data.

4.2.2 The Australian Twin Study (Duffy, Martin and Mathews 1990)

The Australian twin study was a retrospective cohort study for analyzing the correlations between monozygotic twins (MZ) and between dizygotic twins (DZ) for various diseases. Such twin studies allow one to separate effects of shared environment and genetic components. While it is generally assumed that twins share the same environment, monozygotic twins (MZ) are genetically identical, while dizygotic twins (DZ) have half their genome in common. In the Australian twin study, between 1980 and 1982 a questionnaire was mailed to all registered 5,968 twins over the age of 18 asking them for their history of diseases and operations. A total of 3,808 complete pairs returned the questionnaire.

Appendicitis is an inflammation of the vermiform appendix and is usually acute. It occurs in approximately 10–15% of the population, mostly adolescents and young adults. The causes of appendicitis are not fully understood, According to a leading theory, the initial event is obstruction of the lumen by factors such as foreign bodies, intestinal parasites, tumors, or lymphoid follicular enlargement due to a viral infection. However, obstructive elements have been identified in only 30–40% of the removed inflamed appendixes. Additional factors may contribute to the occurrence of appendicitis, such as genetic predisposition, perhaps interacting with diet or other environmental factors.

Here we investigate the age of appendectomy for twin female pairs. There were 1,218 MZ and 735 DZ such pairs. Of these pairs 770 MZ twins (63%) and 464 DZ twins (63%) were doubly censored; that is, neither sibling had undergone appendectomy; in 304 MZ twins (25%) and 208 DZ twins (29%) one sibling had undergone appendectomy; in 144 MZ twins (12%) and 63 DZ twins (9%) both siblings had undergone appendectomy. While only about 30% of the people will eventually get an appendectomy, the use of the logspline procedure requires us to make the usual survival analysis assumption that eventually everybody would have gotten an appendectomy, but that most people die before this happens. (According to the HEFT fit shown in Fig. 4, after approximately 29,000 years 90% of the people would have gotten an appendectomy.) As such, the interesting part of the density will be on the region $[0, 60]^2$, since very few people get appendectomies after age 60. However, this region represents only about 10% of the probability mass of the density.

HEFT was used to estimate the marginal density of the age to appendectomy for the MZ twins and for the DZ twins. Since it is reasonable to assume that the marginal distribution of the age to appendectomy is the same for the MZ and DZ twins, data for both types of twins were combined to get one preliminary estimate of the marginal density, which is shown as the solid line in Figure 4. Because we prefer to detect details using the bivariate estimation on the unit square, we oversmoothed this estimates somewhat by doubling the default penalty parameter.

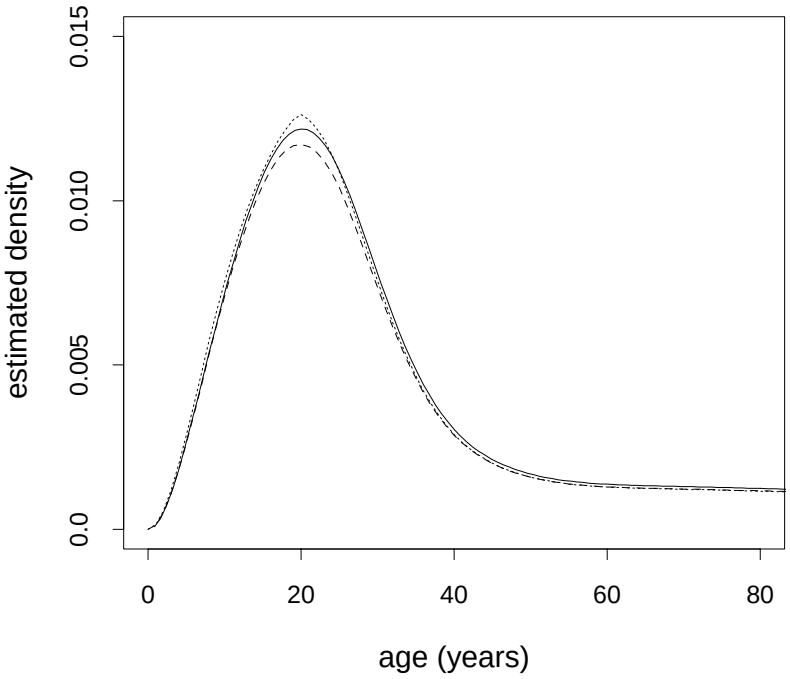

Figure 4. Marginal Density Estimates for the Twin Data. Solid: all data, using HEFT; dotted: MZ, combined procedure; dashed: DZ, combined procedure.

In this situation the bivariate density estimate is required to be symmetric in its arguments, that is, $\hat{f}(t_1, t_2) = \hat{f}(t_2, t_1)$. For the MZ twins, we initially fit a model with four knots and 19 basis functions; for the DZ twins we fit a model with three knots and 11 basis functions. After the stepwise deletion procedure, two basis functions, neither involving knots, were left for the DZ twins and four for the MZ procedure. The fits are summarized in Table 4.

The fitted densities, back-transformed to $[0, \infty)^2$, are shown in Figure 5. After the bivariate densities are fit on the unit square, the (back-transformed) marginal densities are no longer identical to the estimate we obtained by HEFT. In particular, in the present situation, where both data sets were combined to estimate one marginal density, differences could occur. However, as can be noticed from the new estimates of the marginal densities (shown as the dashed and dotted lines in Fig. 4), there is little difference between the combined estimate of the marginal densities and the separate marginals for the MZ and DZ twins. This is comforting since, as the response patterns of the DZ and

Table 4. Fitted Densities for the Twin Data on the Unit Square

Basis function	*MZ twins*	*DZ twins*
$t_1 + t_2$	-2.15	-1.65
$t_1 t_2$	4.47	3.56
$(.13 - t_1)_+ + (.13 - t_2)_+$	3.03	—
$(.13 - t_1)_+ \, t_2 + t_1 (.13 - t_2)_+$	-13.32	—

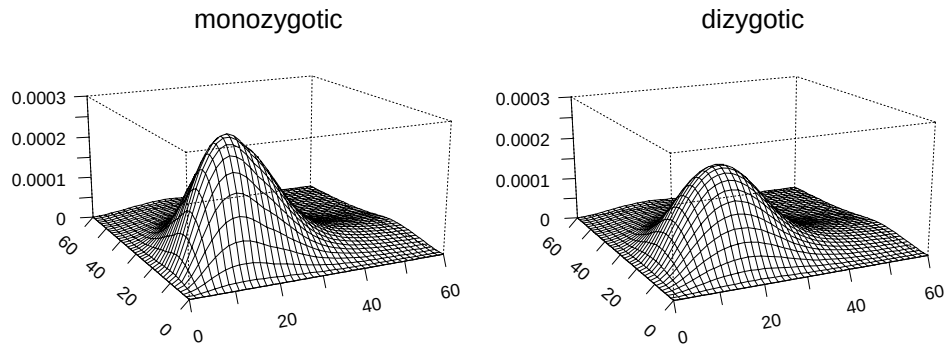

Figure 5. Logspline Estimates of the Bivariate Densities. Left: monozygotic twins; right: dizygotic twins.

MZ twins differed somewhat (Duffy, Martin, and Mathews 1990), there may have been differences between the two groups.

In a twin study, such as the Australian twin study, two of the questions of interest are: are the survival times (age of appendectomy) of the twins correlated? If so, are the correlations in the survival time identical for MZ and DZ twins? A positive answer to the first question would suggest that either environmental or genetic factors play a role in appendicitis. A negative answer to the second question would suggest that genetic factors play a role in appendicitis.

Although the fact that the AIC procedure did not remove the basis function t_1t_2 already suggests that the two survival times are not independent for either of the two types of twins, we can carry out a formal likelihood-ratio test to investigate this further. For the MZ twins the difference in log-likelihood between the models with and without the basis function t_1t_2 is 54.11; for the DZ twins it is 10.21. Both results give strong support to the hypothesis of dependence. Actually, the fitted correlation between the two survival times on the unit square was .42 for the MZ twins and .28 for the DZ twins.

To answer the second question, first a model with the same basis functions $t_1 + t_2$ and t_1t_2 as were used for the DZ twins were fit to the MZ twins and to the combined data for all twins, still using the transformed data. The test for differences between MZ twins and DZ twins now becomes a test for differences in the coefficient of t_1t_2. This test can conveniently be carried out as a likelihood ratio test (with one degree of freedom). The relevant statistics can be found in Table 5. As can be seen, the difference in log-likelihood is 3.40 ($p = .009$) so that we can conclude that the MZ twins have a higher correlation than the DZ twins.

Table 5. Statistics for Testing the Difference Between MZ and DZ Twins

Data	*Coefficient of t_1 t_2*	*Log-likelihood*
MZ twins	6.18	−440.70
DZ twins	3.56	−290.77
		−731.47
All twins	5.25	−734.87

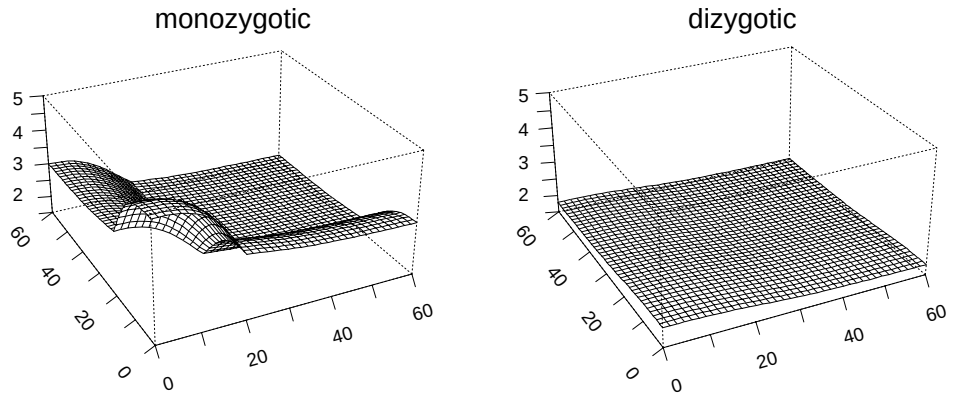

Figure 6. Logspline Estimates of the Local Odds Ratio. Left: monozygotic twins; right: dizygotic twins.

As discussed in Section 2.3 the local odds ratio gives a local measure of dependence. In Figure 6 the estimated local odds ratio for the MZ and the DZ twins are shown. For the MZ twins the estimated local odds ratio varies between 4.9 (at $t_1 = t_2 = 0$) and 1.7 (at $t_1 = t_2 = 60$); for the DZ twins it varies between 2.0 (at $t_1 = t_2 = 0$) and 1.5 (at $t_1 = t_2 = 60$). When examining these plots one should keep in mind that for either $t_1 = \infty$ or $t_2 = \infty$ the odds ratio equals one for logspline models. Note that the local dependence between T_1 and T_2 is everywhere stronger for the MZ twins than for the DZ twins. Because the local odds ratio is larger than 1 everywhere, the risk of getting an appendectomy for a twin increases if her sibling also got an appendectomy. From the left side of Figure 6, we are led to believe that this increase in risk is particularly large when a MZ twin is young and her sibling got the appendectomy at a young age.

To investigate the validity of this belief or, equivalently, that the local odds ratio for the MZ twins is indeed much larger near (0,0) than elsewhere, we analyzed simulated data that was generated from a Clayton model (4.1) having the same marginal distributions and censoring pattern as the MZ twin data. In particular, we generated a new set $\mathbf{T}_i^\circ = (T_{1i}^\circ, T_{2i}^\circ)$, $i = 1, \ldots, 1{,}218$, of survival times from a Clayton model with $\theta = 2.5$ and marginal densities as shown in Figure 4. For the twin data the censoring times C_{1i} and C_{2i} are identical for all pairs; furthermore, analysis of the data suggested that the distribution of of $C_1 - 20$ is reasonably modeled by an exponential distribution with mean 18. We thus generated censoring times $C_{1i}^\circ = C_{2i}^\circ$ accordingly. For each i we set $Y_{1i}^\circ = \min(T_{1i}^\circ, C_{1i}^\circ)$, $Y_{2i}^\circ = \min(T_{2i}^\circ, C_{2i}^\circ)$, $\delta_{1i}^\circ = \text{ind}(T_{1i}^\circ \le C_{1i}^\circ)$ and $\delta_{2i}^\circ = \text{ind}(T_{2i}^\circ \le C_{2i}^\circ)$. This yielded 744 doubly censored, 314 singly censored and 140 uncensored pairs in the simulated data set. We applied the bivariate logspline procedure to $(\mathbf{Y}_i^\circ, \boldsymbol{\delta}_i^\circ)$, $i = 1, \ldots, 1218$, requiring a symmetric density, as we did for the MZ twin data. The fitted bivariate density and the corresponding local odds ratio are shown in Figure 7. (We actually repeated this experiment ten times and got very similar results each time.) While the estimated density in the left side of Figure 7 looks similar to the one in the left side of Figure 5, the local odds ratios are very different. In particular, we note that the local odds ratio in Figure 7 is fairly constant at a level between 2 and 2.5. Since the true odds ratio for the simulated data was indeed constant at 2.5, this adds credence to the conclusion of Figure 6 that

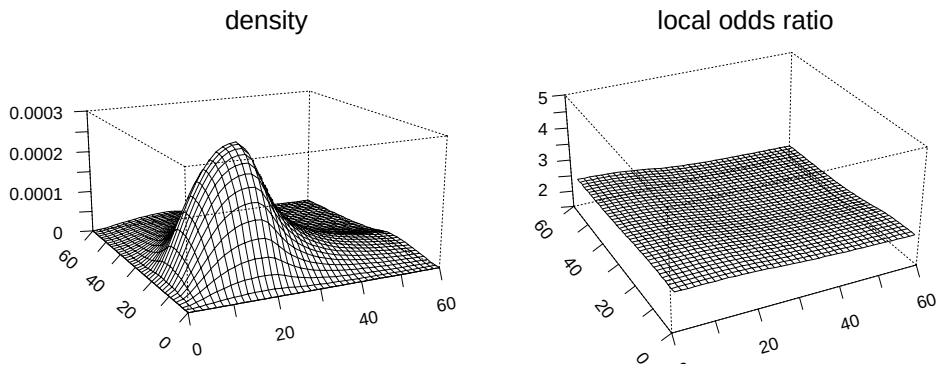

Figure 7. Logspline Estimate for a Simulated Data Set With the Same Marginal Distributions as the Twin Data, a Constant Odds Ratio of 2.5, and Sample Size and Censoring Patterns as for the MZ Twins. Left: density estimate; right: local odds ratio estimate.

the dependence between the two survival times is stronger for low values of T_1 and T_2 for the MZ twins. (Actually, this stronger dependence for low values of T_1 and T_2 was also found in other investigations [Dr. R. L. Prentice, private communication].) A similar simulation for the DZ twins suggested that their odds ratio may very well be constant.

5. DISCUSSION

In this article we have introduced a procedure for estimating bivariate densities based upon data that may be censored. By first estimating the marginal densities in order to transform the data to the unit square, we facilitate the computations, avoid an artifact that would prevent us from estimating a positive dependence for large values of the components, and obtain a final estimate that is smoother than it otherwise would have been.

From the examples in Section 4 we see that our procedure can be used in visualizing dependencies, which can provide further insight in our data. Also, it yields a powerful test of independence.

The present methodology cannot deal with covariates beyond what was done with type-of-twins in the Australian twin study example. However, it should be possible to extend this methodology to include covariates by combining hazard regression (HARE) (Kooperberg, Stone, and Truong 1995a) with a bivariate conditional density estimation procedure. While in principle it would be possible to extend the present methodology to estimate densities in more than two dimensions (the numerical problems would not become harder; in particular, all multivariate integrals would still reduce to univariate ones), we expect that there are few situations in which such a procedure would be useful because of the large amount of data that would be required to obtain a reasonably accurate estimate of such a multivariate density.

ACKNOWLEDGMENTS

This research was supported in part by NSF grant DMS-9403371 and NIH grant CA61937. The author thanks Li Hsu and Charles J. Stone for many helpful discussions. The set of data for the Australian twin study in Section 4.2 was kindly made available by David L. Duffy of the Queensland Institute of Medical Research. This data set was collected by the Australian National Health and Medical Research Council Twin Registry, Professor N. G. Martin of the Queensland Institute of Medical Research, and Professor J. D. Mathews of the Menzies School of Health Research.

[Received July 1996. Revised August 1997.]

REFERENCES

Abramowitz, M., and Stegun, I. A. (1965), *Handbook of Mathematical Functions*, New York: Dover.

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993), *Statistical Models Based on Counting Processes*, New York: Wiley.

Clayton, D. G. (1978), "A Model for Association in Bivariate Life Tables and its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence," *Biometrika*, 65, 141–151.

Commenges, D., and Andersen, P. K. (1995), "Score Test of Homogeneity for Survival Data," *Lifetime Data Analysis*, 1, 145–156.

Dabrowska, D. M. (1988), "Kaplan–Meier Estimate on the Plane," *The Annals of Statistics*, 16, 1475–1489.

Dabrowska, D. M., Zhang, Z., and Duffy, D. L. (1995), "Graphical Comparisons of Bivariate Survival Functions With an Application to Twin Studies," unpublished manuscript.

Duffy, D. L., Martin, N. G., and Matthews, J. D. (1990), "Appendectomy in Australian Twins," *The American Journal of Human Genetics*, 47, 590–592.

Gill, P. E., Murray, W., Saunders, M. A., and Wright, M. H. (1986a), "User's Guide for LSSOL (Version 1.0)," Technical Report No. SOL 86-1, Stanford University, Department of Operations Research.

——— (1986b), "User's Guide for NPSOL (Version 4.0)," Technical Report No. SOL 86-2, Stanford University, Department of Operations Research.

Gill, R. D., van der Laan, M. J., and Wellner, J. A. (1995), "Inefficient Estimators of the Bivariate Survival Function for Three Models," *Annales de l'Institut Henri Poincaré*, 31, 545–597.

Gray, R. J. (1995), "Tests for Variation Over Groups in Survival Data," *Journal of the American Statistical Association*, 90, 198–203.

Härdle, W. (1990), *Smoothing Techniques With Implementation in S*, New York: Springer.

Koo, J.-Y. (1996), "Bivariate B-splines for Tensor Logspline Density Estimation," *Computational Statistics and Data Analysis*, 21, 31–42.

Koo, J.-Y., Kooperberg, C., and Park, J. (in press), "Logspline Density Estimation Under Censoring and Truncation," *Scandinavian Journal of Statistics*.

Kooperberg, C., and Stone, C. J. (1992), "Logspline Density Estimation for Censored Data," *Journal of Computational and Graphical Statistics*, 1, 301–328.

Kooperberg, C., Stone, C. J., and Truong, Y. K. (1995a), "Hazard Regression," *Journal of the American Statistical Association*, 90, 78–94.

——— (1995b), "The L_2 Rate of Convergence of Hazard Regression," *Scandinavian Journal of Statistics*, 22, 143–157.

Oakes, D. (1982), "A Concordance Test for Independence in the Presence of Censoring," *Biometrics*, 38, 451–455.

——— (1989), "Bivariate Survival Models Induced by Frailties," *Journal of the American Statistical Association*, 84, 487–493.

Prentice, R. L., and Cai, J. (1992), "Covariance and Survivor Function Estimation Using Censored Failure Time Data," *Biometrika*, 79, 495–512.

Pruitt, R. C. (1991), "On Negative Mass Assigned by the Bivariate Kaplan–Meier Estimator," *The Annals of Statistics*, 19, 443–453.

Scott, D. W. (1992), *Multivariate Density Estimation*, New York: Wiley.

Stone, C. J. (1994), "The use of Polynomial Splines and Their Tensor Products in Multivariate Function Estimation" (with discussion), *The Annals of Statistics*, 22, 118–184.

Stone, C. J., Hansen, M., Kooperberg, C., and Truong, Y. K. (1997), "Polynomial Splines and Their Tensor Products in Extended Linear Modeling" (with discussion), *The Annals of Statistics*, 25, 1371–1470.

van der Laan, M. J. (1996), "Efficient and Inefficient Estimation in Semiparametric Models," technical report, CWI Amsterdam.