Interval Censoring and Longitudinal Survey Data

by

Norberto Pantoja Galicia

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 2007

Author's declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Being able to explore a relationship between two life events is of great interest to scientists from different disciplines. Some issues of particular concern are, for example, the connection between smoking cessation and pregnancy (Thompson and Pantoja-Galicia 2003), the interrelation between entry into marriage for individuals in a consensual union and first pregnancy (Blossfeld and Mills 2003), and the association between job loss and divorce (Charles and Stephens 2004, Huang 2003 and Yeung and Hofferth 1998).

Establishing causation in observational studies is seldom possible. Nevertheless, if one of two events tends to precede the other closely in time, a causal interpretation of an association between these events can be more plausible. The role of longitudinal surveys is crucial, then, since they allow sequences of events for individuals to be observed. Thompson and Pantoja-Galicia (2003) discuss in this context several notions of temporal association and ordering, and propose an approach to investigate a possible relationship between two lifetime events.

In longitudinal surveys individuals might be asked questions of particular interest about two specific lifetime events. Therefore the joint distribution might be advantageous for answering questions of particular importance. In follow-up studies, however, it is possible that interval censored data may arise due to several reasons. For example, actual dates of events might not have been recorded, or are missing, for a subset of (or all) the sampled population, and can be established only to within specified intervals.

Along with the notions of temporal association and ordering, Thompson and Pantoja-Galicia (2003) also discuss the concept of one type of event "triggering" another. In addition they outline the construction of tests for these temporal relationships.

The aim of this thesis is to implement some of these notions using interval censored data from longitudinal complex surveys. Therefore, we present some proposed tools that may be used for this purpose.

This dissertation is divided in five chapters, the first chapter presents a notion of a temporal relationship along with a formal nonparametric test. The mechanisms of right censoring, interval censoring and left truncation are also overviewed. Issues on complex surveys designs are discussed at the end of this chapter.

For the remaining chapters of the thesis, we note that the corresponding formal nonparametric test requires estimation of a joint density, therefore in the second chapter a nonparametric approach for bivariate density estimation with interval censored survey data is provided. The third chapter is devoted to model shorter term triggering using complex survey bivariate data. The semiparametric models in Chapter 3 consider both noncensoring and interval censoring situations. The fourth chapter presents some applications using data from the National Population Health Survey and the Survey of Labour and Income Dynamics from Statistics Canada. An overall discussion is included in the fifth chapter and topics for future research are also addressed in this last chapter.

Acknowledgements

First of all, I thank God for His love and for sustaining me throughout this journey and our Blessed Mother for her loving intercession and for being the cause of our joy.

I would like to thank my supervisor, Mary Thompson, for her kindness, generosity and most valuable support at all times and especially during the difficult moments. Without her guidance this dissertation could not have been realized.

I would also like to thank the members of my committee, Jock MacKay, for sharing his time, knowledge and advice with me, and Richard Cook, for his kind suggestions, help and advice. Thank you to my external examiners, David Bellhouse and John Goyder, who helped to make my defense a learning experience.

I am grateful to all the faculty members, fellow graduate students and staff members from Waterloo (and abroad) whom I have befriended over the years. Thanks to Milorad Kovacevic for his encouragement and for making my visit at Statistics Canada a pleasant learning experience.

I also thank the Consejo Nacional de Ciencia y Tecnología (Conacyt) for the financial support, the National Program of Complex Data Structures, the Mathematics of Information Technology and Complex Systems and Statistics Canada for their funding and support.

My profound gratitude goes out to my dearest parents Haydee and Norberto, my sisters Monica and Ivonne and my brother Israel, for their endless love, prayers and encouragement.

Also, thanks to my parents-in-law, Salvador and Catalina, to my family-in-law, to Samm, Celza, Gloria, sister Pat, the MC and the GVO communities and all my friends for their prayers and support.

Finally, I thank my beloved wife, Alicia, for her love, patience, help, time, understanding, sacrifices, prayers and encouragement. For being my precious motivation and especially for Ana Maria. Ana Maria, thank you for your presence in our lives, for your laughter, and for bringing such brightness to the joy of my life.

To Jesus, the Virgin Mary and St. Joseph

To Alicia and Ana Maria

To my parents

Contents

List of Figures

Chapter 1

Introduction

Being able to explore a relationship between two life events is of great interest to scientists from different disciplines. Some issues of particular concern are, for example, the connection between smoking cessation and pregnancy (Thompson and Pantoja-Galicia 2003), the interrelation between entry into marriage for individuals in a consensual union and first pregnancy (Blossfeld and Mills 2003), and the association between job loss and divorce (Charles and Stephens 2004, Huang 2003 and Yeung and Hofferth 1998).

Establishing causation in observational studies is seldom possible. Nevertheless, if one of two events tends to precede the other closely in time, a causal interpretation of an association between these events can be more plausible. The role of longitudinal surveys is crucial, then, since they allow sequences of events for individuals to be observed. Thompson and Pantoja-Galicia (2003) discuss in this context several notions of temporal association and ordering, and propose an approach to investigate a possible relationship between two lifetime events.

In longitudinal surveys, individuals might be asked questions of particular interest about two specific lifetime events. Therefore the joint distribution might be advantageous for answering questions of particular importance. In follow-up studies, however, it is possible that interval censored data may arise due to several reasons. For example, actual dates of events might not have been recorded, or are missing, for a subset of (or all) the sampled population, and can be established only to within specified intervals.

Along with the notions of temporal association and ordering, Thompson and Pantoja-Galicia (2003) also discuss the concept of one type of event "triggering" another. In addition, they outline the construction of tests for these temporal relationships.

The aim of this thesis is to implement some of these notions using interval censored data from longitudinal complex surveys. Therefore, we present some proposed tools that may be used for this purpose.

This dissertation is divided in five chapters, this being the introductory one. Sections 1.1 and 1.2 present a notion of a temporal relationship along with a formal nonparametric test. An overview of the mechanisms of right censoring, interval censoring and left truncation is given in Section 1.3. Issues on complex surveys designs are discussed in Section 1.4.

For the remaining chapters of the thesis, we note that the corresponding formal non-parametric test requires estimation of a joint density; therefore in the second chapter a nonparametric approach for bivariate density estimation with interval censored survey data is provided. The third chapter is devoted to modeling shorter term *triggering* using complex survey bivariate data. The semiparametric models in Chapter 3 consider both noncensored and interval censored situations. The fourth chapter presents some applications using data from the National Population Health Survey and the Survey of Labour and Income Dynamics from Statistics Canada. An overall discussion is included in the fifth chapter and topics for future research are also addressed in this last chapter.

1.1 Close Precursor

Let E_1 and E_2 be two types of lifetime events. Let T_1 be the time to occurrence of event E_1, and T_2 be the time to occurrence of event E_2 considering a specified time origin. Knowledge of the exact times of occurrence of each event would provide the appropriate elements to model a temporal relationship through their joint intensities.

From Thompson and Pantoja-Galicia (2003), a local association of T_1 and T_2 is implied by the following concept:

T_1 is a *close precursor* of T_2 if for some positive numbers δ and $\kappa(t_1)$ we have

$$\frac{\mathcal{F}_2(t_1 + \kappa(t_1) \mid T_1 = t_1)}{\mathcal{F}_2(t_1 \mid T_1 = t_1)} < \frac{\mathcal{F}_2(t_1 + \kappa(t_1))}{\mathcal{F}_2(t_1)} - \delta. \tag{1.1}$$

for all t_1 in a specified interval (a, b), with $a, b \in \mathbb{R}$. Here $\mathcal{F}_i(t) = Pr(T_i > t)$ for $i = 1, 2$. In other words, T_1 is a *close precursor* of T_2 if the occurrence of the first event E_1 at T_1 decreases the probability of having to wait longer than $\kappa(t_1)$ to observe the occurrence of the second event E_2, and this happens with some uniformity in (a, b). The decrease is seen relative to the analogous probability if T_1, T_2 are independent.

The interval length $\kappa(t_1)$ may be thought of as the duration of an effect and would come from subject-matter considerations. We have allowed it to depend in general on t_1, anticipating that the effect of T_1 on the hazard of T_2 might not have constant duration.

Although we have given the definition of close precursor in terms of survivor functions, it can be expressed approximately in terms of hazard functions, as follows:

T_1 is a *close precursor* of T_2 if for suitably chosen $\kappa(s)$ and $\delta > 0$,

$$h_2(u \mid T_1 = s) > h_2(u) + \delta,$$

for $u \in (s, s + \kappa(s))$.

Then δ is seen to correspond to an "additive" lower bound to a short term change in the hazard function.

In either formulation, the motivation is that the more closely T_2 tends to follow T_1, the greater the plausibility for a causal connection might be.

Since (1.1) reflects approximately a short term raising of the hazard function for T_2, it is not difficult to formulate an analogue for point process intensities, giving us an alternative way of modelling events less tied to a time origin. Blossfeld and Mills (2003) use interdependent point processes to model interrelated family events, namely entry into marriage (for individuals in a consensual union) and first pregnancy/childbirth. See also Lawless (2003a) for some discussion of intensity models.

1.2 Nonparametric Test for Close Precursor

A formal test for a *close precursor* relationship between T_1 and T_2 as indicated in Section 1.1 (T_1 a close precursor of T_2) is given by the following:

For suitable $\kappa(t_1)$, let

$$Q = \int \left(\frac{\hat{\mathcal{F}}_2(t_1 + \kappa(t_1) \mid T_1 = t_1)}{\hat{\mathcal{F}}_2(t_1 \mid T_1 = t_1)} - \frac{\hat{\mathcal{F}}_2(t_1 + \kappa(t_1))}{\hat{\mathcal{F}}_2(t_1)} \right) d\hat{\mathcal{F}}_1(t_1), \tag{1.2}$$

where the domain of integration is the interval (a, b).

Under the null hypothesis of independence of T_1 and T_2, the mean of Q will be close to 0. Thus in order to test the null hypothesis, the value of Q may be compared with twice its estimated standard error $se(Q)$. A discussion about the estimation of this standard error is presented in Chapter 2 Section 2.4.

Note that the difference within (1.2) approximates the difference between the hazard function conditional on $T_1 = t_1$ and the unconditional hazard.

To compute Q, we first would obtain an estimate of the joint density of (T_1, T_2).Then, we would obtain numerically the corresponding marginal probability density functions and consequently the respective survivor functions (conditional and unconditional versions).

1.3 Censoring, time origin and left truncation

A particular characteristic of data which involves time until occurrence of an event is the presence of censoring, which occurs when the value of the response variable is not observed. There exist different types of censoring for duration time data. Features such as right censoring, interval censoring and left truncation may arise when analyzing time to event data and therefore are briefly introduced in this section.

1.3.1 Right censoring

Let $X_1, X_2, \ldots, X_n$ be random variables, where X_i denotes the time until an event E occurs for the i^{th} individual. Instead of having observed values for the time to occurrence of every individual, right censoring is present when the reported time x_i is either an observed time or a censoring time. In other words $x_1, x_2, \ldots, x_n$ are the observed times to either occurrence of E or censoring. A status indicator variable can be defined to distinguish between these two possibilities. Let $\delta_i = 1$ if $X_i = x_i$ and 0 if $X_i > x_i$. Therefore, $\delta_i = 1$ indicates that x_i is an observed time to occurrence of event E and $\delta_i = 0$ indicates that x_i is a censoring time.

In the following subsections, we assume for simplicity that the duration times X_i are independent and identically distributed.

Type 1 Censoring

Suppose each individual has a specified fixed potential censoring time $C_i > 0$ such that X_i is observed if $X_i \leq C_i$; otherwise the only information we have is that $X_i > C_i$. This type of censoring occurs in studies conducted in a specified time period, such as clinical trials with a predetermined date for terminating follow up.

Type 2 Censoring

Consider n individuals who start on a study all at the same time knowing that the study will be concluded as soon as r of them experience the occurrence of event E. The number of complete observations that will be recorded is predetermined in this case. If r is established in advance as the number of complete observations to be measured then only the r smallest times $x_{(1)}, x_{(1)}, \ldots, x_{(r)}$ are observed from a random sample of n. In this scenario r is an integer between 1 and n, and the observed value of the censoring time $C_i = C = X_{(r)}$ is a random variable.

Note that analysis of data obtained from this type of study design can be performed taking advantage of the theory of order statistics. On the other hand, the fact of not knowing the total time of the duration of the study at its initiation point puts this design at a practical disadvantage.

Independent Random Censoring

For this type of censoring, let X and C represent the time to event and censoring time respectively, with survivor functions $\mathcal{F}(x)$ and $\mathcal{G}(c)$ accordingly. Also assume the C_i values are random variables that are independent of each other and of the response measurements X_i for $i = 1, \ldots, n$. In addition, assume that $\mathcal{G}(c)$ does not depend on any of the parameters of $\mathcal{F}(x)$.

When subjects are removed from a study because of events such as accidental death, death due to a cause unrelated to the occurrence of the event of interest E, migration, patient withdrawal, etc., this type of censoring scheme may fit the process under study. However, it is important to be aware of situations where the censoring process is connected to the duration time process and in consequence, this censoring scheme would not apply. Such circumstances are present in many situations.

For the previous censoring mechanisms, if the X_i's have probability density function $f(x)$

and survivor function $\mathcal{F}(x)$, the observed likelihood function reduces to the following form (see Lawless 2003b, Chapter 2):

$$\mathcal{L} = \prod_{i=1}^{n} f(x_i)^{\delta i} \mathcal{F}(x_i+)^{1-\delta i}.$$

1.3.2 Interval censoring

The general setting for this type of censoring is as follows. The values of the duration time X may not be directly observed, so let us assume a partition $\mathcal{A} = \{a_0, a_1, \ldots, a_m\}$ of the real line. These could be previously specified times or monitoring times, for example. The observed data are of the form $I = (A, B)$, where $A = sup\{a_j : a_j \leq X\}$ and $B = inf\{a_j : a_j \geq X\}$. In other words, $I = (A, B)$ contains the unobserved duration time X. If event E has not occurred by time a_m then we have the censoring case discussed in section 1.3.1 with right censoring time $A = a_m$ for X, and $B = \infty$.

Having n individuals, this scheme can be generalized to the following partition $\mathcal{A}_i = \{a_{i0}, a_{i1}, \ldots, a_{im_i}\}$ for $i = 1, \ldots, n$ and hence $X_i \in I_i = (A_i, B_i)$, where $A_i = sup\{a_{ij} : a_{ij} \leq X_i\}$ and $B_i = inf\{a_{ij} : a_{ij} \geq X_i\}$.

Let $F_i(t)$ denote the distribution function for T_i. The likelihood function from a sample of n independent individuals under this observation scheme is the following (see Lawless 2003):

$$\mathcal{L} = \prod_{i=1}^{n} \big(F_i(B_i) - F_i(A_i)\big), \tag{1.3}$$

assuming that $F_i(0) = 0$.

Some special cases of this general interval censoring scheme are:

Grouped data

The observation times are the same for all individuals. That is, $a_{ij} = a_j$.

Current status data

Suppose individual i is examined once, at time C_i, and at this point it is determined whether event E has occurred (i.e., $X_i \leq C_i$) or not (i.e., $X_i > C_i$). So the interval for the individual is either $(0, C_i]$ or (C_i, ∞).

1.3.3 Time origin

According to Matthews (2003), three basic requirements define duration time measurements:

1. An unambiguous origin for the measurement of "time"

2. An agreed scale of measurement

3. A precise definition of "response", or occurrence of the event of interest.

The time origin need not be the same calendar time for each study subject, but should be precisely defined for each subject. All subjects should be as comparable as possible at the origin.

If observation does not commence at the origin, special treatment of the data is required. For example, left truncation may affect the distributions of the observed times.

1.3.4 Left truncation

If the current duration time at the moment of selection is not $x = 0$ but some value $u > 0$ then we say X_i is left truncated at u_i. By way of illustration, if u denotes the calendar time of the selection determining condition, and X is measured from 0, only subjects for which $X \geq u$ can be observed. This is because subjects with $X < u$ are automatically excluded from the study population. Hence we say the study data are left truncated at the value u.

Lawless (2003b) reviews this topic in detail as well as the censoring mechanisms described throughout this section.

1.4 Complex Survey Data

1.4.1 Probability sampling designs

There exist different types of probability sampling designs. Some examples are: simple random sampling, stratified sampling, cluster sampling, multi-stage sampling, etc. The literature describing them is extensive. Some references are Lohr (1999) and Thompson (1997).

Due to complexities of the sampling design such as without replacement sampling, stratification, clustering or multi-stage sampling, complex survey data may violate the assumption of independent and identically distributed (i.i.d.) data.

For the purpose of this work, in the following section we present the settings for a stratified two-stage sampling design as described in Buskirk and Lohr (2005).

1.4.2 Stratified two-stage sampling design

The finite population is assumed to be divided into L strata. Stratum l has N_l primary sampling units (psu's); we sample n_l of these psu's. Let $N = \sum_{l=1}^{L} N_l$ and $n = \sum_{l=1}^{L} n_l$ be the total number of (psu's) in the population and sample, respectively.

Cluster samples are taken independently from each stratum; the inclusion probabilities are $\pi_i^{(l)} = P_D(\text{(psu) } i \text{ from stratum } l \text{ is included in the sample})$, with $\sum_{i=1}^{N_l} \pi_i^{(l)} = n_l$. The subscript D indicates the probability distribution induced by the design. The joint inclusion probabilities are $\pi_{ij}^{(l)} = P_D(\text{psu's } i \text{ and } j \text{ from stratum } l \text{ are included in the sample})$.

At the secondary sampling unit (ssu) level, (psu) i of stratum l has Q_{li} secondary sampling units (ssu's); $\pi_{m|i}^{(l)}$ is the conditional probability that (ssu) m of (psu) i is included in the sample, given that (psu) i is included. The $\pi_{m|i}^{(l)}$ satisfies $\sum_{m=1}^{Q_{li}} \pi_{m|i}^{(l)} = q_{li}$, where q_{li} is the number of (ssu's) sampled from (psu) i of stratum l. We have $Q_l = \sum_{i=1}^{N_l} Q_{li}, Q = \sum_{l=1}^{L} Q_l$, and $W_l = Q_l/Q$. Thus, Q is the total number of observation units in the population, and

W_l is the stratum weight for stratum l.

1.4.3 Survey Weights

Motivation

If we consider a simple random sample s, each unit in the finite population has the same probability of being sampled. The proportion of the population that is selected is the sample size divided by the population size. For example, let us suppose we have a population of size $N(= 1000)$ and we want to obtain a simple random sample of size $n(= 250)$. Let y_i denote the i^{th} response variable in the sample s, with $i \in s$. In studying a human population, if our interest is to estimate the population total $Y = \sum_{i=1}^{N} y_i$, we would say that each individual in the sample represents $\frac{N}{n}(= 4)$ individuals in the population and therefore we simply assign a sampling weight of $\frac{N}{n}(= 4)$ to each one of the sampled individuals. In other words, any sampled member's response is taken to represent $\frac{N}{n}(= 4)$ identical responses in the population. Let us denote w_i to be the sampling weight of individual i in the sample. In our example, $w_i = 4$, for $i \in s$, and then the estimator of the total Y is $\hat{Y} = \sum_{i \in s} y_i w_i = 4 \sum_{i \in s} y_i$. Due to the simple random sampling, all individuals have the same sampling weight.

In complex surveys, the sample is typically obtained using an unequal probability of selection scheme. Consequently, unequal weights, $w_i, i = 1, \ldots, n$, are assigned to individuals in the sample, with $w_i = \pi_i^{-1}$ where π_i is the probability for individual i to be included in the sample s. Consequently, the i^{th} individual in the sample represents w_i individuals in the population. Therefore, unbiased estimation of the population sum $Y = \sum_{i=1}^{N} y_i$ may be obtained by the weighted sum $\hat{Y} = \sum_{i \in s} y_i w_i$. In general, the basic estimation method is to replace population sums by weighted sample sums.

In practice, there may also be weight adjustments for *nonresponse*. In addition, auxiliary information can be used to adjust the weights so that the survey estimates are consistent with known population totals. *Poststratification*, for example, is employed to adjust the survey weights using a particular variable (e.g. sex) which is appropriate for stratification but was not used at the design stage because the corresponding information was not available (or because updated and more reliable information became available after the selection of the sample). Furthermore, a method called *calibration* is used when the weighted sample totals must agree with reference totals for more than one variable (Statistics Canada; 2003).

In summary, we can say that a survey weight, representing a certain number of individuals (or units) in a finite population, is usually attached to each individual (or unit) in the sample to account for various factors such as unequal probability of selection, nonresponse, poststratification and calibration.

The following section depicts the process of computing longitudinal weights.

1.4.4 Longitudinal weighting

In longitudinal surveys, the target population associated with the longitudinal weight is the population at the time of the panel selection. An example of the multi-step process conducted by a statistical agency to derive weights (at a cycle after the first) for a national longitudinal survey is presented in Naud (2004). A summary of this process is presented next.

First, a classification of all individuals takes place according to whether they are respondents, non-respondents or out of scope (for example, individuals who are deceased or outside the country where the national survey is conducted). A nonzero longitudinal weight is assigned to respondents and out-of-scope individuals, while a weight of zero is given to non-respondents. Non-response adjustment is the next step. For this purpose a non-response model is developed, and the weights of respondents are adjusted so that they represent non-respondents as well. Out-of-scope individuals keep their initial weight, thereby representing the portion of the target population that was present at the time of the panel selection and subsequently left the country or died. Calibration is performed afterwards to ensure that certain totals computed with the weights match the population totals derived from other sources. They apply to the longitudinal target population, which is the population at the time the panel was selected. Completion of this process gives the final longitudinal weights for the panel and those weights are produced for each reference year.

A discussion regarding survey weights is resumed in Section 2.3.

1.4.5 Bootstrap Weights

For the purpose of design-based variance estimation, a number of complex surveys conducted by statistical agencies, such as Statistics Canada, generate bootstrap weights to accompany their data.

A motivation for using the bootstrap method is the possibility of estimating the variance of an estimated parameter by using a large number of somewhat different subsamples (also called replicates) from the original sample. Each replicate is then used to estimate the parameter and the variability among the resulting estimates is used to estimate the variance of the "full sample" estimate. This basic idea is followed by other resampling methods which may differ in the way the replicates are built.

In a simple description, bootstrap replicates are generated by randomly choosing, with replacement, a sample of primary sampling units (psu's) within each stratum and adjusting the original sampling weights of the units in the selected (psu) to reflect the probability of selection into the subsample. If a unit does not appear in the bootstrap replicate, its bootstrap weight variable is set to zero. This process of selecting samples and reweighting is repeated R times to arrive at R bootstrap samples (or replicates), R bootstrap weight variables and consequently R bootstrap estimates.

As it was pointed out in Section 1.4.3, the sampling weight, which reflects the probability of selection of a unit, can be thought of as the number of units in the survey population represented by the sampled unit. This sampling weight is used to estimate a parameter of interest. On the other hand, the bootstrap weight is used for the purpose of estimating the sampling error associated with such parameter of interest. Further discussion on this topic is resumed in Section 2.4. Like the sampling weight, a bootstrap weight could be thought of as the number of individuals in the survey population represented by a unit in the reduced (bootstrap) sample.

Rao and Wu (1988), Rao, Wu and Yue (1992) and Yung (1997) are some references regarding bootstrap variance estimation for complex survey data.

Chapter 2

Nonparametric Density Estimation

The formal nonparametric test of Section 1.2 requires estimation of a joint distribution, therefore in this chapter a nonparametric approach for bivariate density estimation is provided using kernel and local likelihood density estimation techniques. This approach takes into account the interval censored and complex survey data.

Estimation of univariate and multivariate density functions, in the case of independent and identically distributed random variables, is presented for example by Silverman (1986), Scott (1992), Wand and Jones (1995) and Simonoff (1996) with material on kernel density estimation. Turnbull (1976), Gentleman and Geyer (1994) and Li, Watkins and Yu (1997) have proposed nonparametric estimators for the distribution function with univariate interval censored data. Density estimation for univariate interval censored data has been covered by Duchesne and Stafford (2001) and Braun, Duchesne and Stafford (2005). In the context of complex surveys research, density estimation is examined by Bellhouse and Stafford (1999), Breunig (2001), Bellhouse, Goia and Stafford (2003), and Buskirk and Lohr (2005).

In Sections 2.1 and 2.2, using the methods proposed by Duchesne and Stafford (2001) and Braun, Duchesne and Stafford (2005) as a starting point, we present an extension of their procedures to the bivariate case to obtain a simple kernel density estimate as well as local likelihood density estimates. The estimation methods consider the interval censored nature of the data. Since we deal with survey data that have been collected using a complex design, in Section 2.3 we make adaptations to the methodology to account for some of these complexities. An important element of the nonparametric test in (1.2) is the standard error of the statistic Q which is assessed in Section 2.4. Bootstrap replicates of the survey weights are used for this purpose. Asymptotic results from related problems in the literature dealing with interval censoring are included in Section 2.6. Convergence results are discussed in Section 2.7 and the corresponding theory and framework to establish asymptotic properties of our estimators is presented in Section 2.8.

2.1 Kernel Density Estimation

2.1.1 Univariate case

For independent and identically distributed non-censored data $Y_1, Y_2, \ldots, Y_n$, the standard kernel density estimate is given by the expression

$$\hat{f}_{nc}(y) = \frac{1}{n} \sum_{i=1}^{n} K_h(Y_i - y), \tag{2.1}$$

where nc stands for noncensored, and $K_h(u) = h^{-1} K(h^{-1}u)$ is a kernel function with bandwidth h. Note that in (2.1) $\hat{f}_{nc}(y) = \hat{f}_{nc}(y; h)$, i.e. our notation suppresses the dependence on the bandwidth h.

Duchesne and Stafford (2001) propose a natural approach to kernel density estimation with randomly interval censored data $X_i \in I_i = (A_i, B_i)$ and $A_i, B_i \in \mathbb{R}$ $i = 1, ..., n$. They note that

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} E\left[K_h(X_i - x) \middle| X_i \in I_i \right]$$

has the same expectation as $\hat{f}_{nc}(x)$, where the conditional expectation is computed with respect to the distribution for the true value X_i over its corresponding interval I_i.

To estimate the required density, they propose to compute the following expression

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} E_{\hat{f}}\left[K_h(X-x)\Big|X_i \in I_i\right],$$

which may be solved using the iteration

$$\hat{f}_j(x) = \frac{1}{n}\sum_{i=1}^{n} E_{\hat{f}_{j-1}}\left[K_h(X_i-x)\Big|X_i \in I_i\right]. \tag{2.2}$$

The expectation in (2.2) is with respect to the conditional density

$$\hat{f}_{j-1|I_i}(u) = \delta_i(u)\hat{f}_{j-1}(u)/\int_{I_i}\hat{f}_{j-1}(s)ds$$

over I_i, where $\delta_i(u) = 1$ if $u \in I_i$ and 0 otherwise.

2.1.2 Bivariate case

In the presence of complete (non-censored) data $\mathbf{Y}_i = (Y_{i,1}, Y_{i,2})$, $i = 1, \ldots, n$, the bivariate kernel density estimator with kernel $K_{\mathbf{h}}(\mathbf{y})$, $\mathbf{y} = (y_1, y_2)$ and bandwidth $\mathbf{h} = (h_1, h_2)$ is given by

$$\hat{f}_{nc}(\mathbf{y}) = \frac{1}{n}\sum_{i=1}^{n} K_{\mathbf{h}}(\mathbf{Y}_i - \mathbf{y}).$$

In the context of interval censored data, $\mathbf{X}_i = (X_{i,1}, X_{i,2})$ lies within the 2-dimensional interval $\mathbf{I}_i = (A_{i,1}, B_{i,1}) \times (A_{i,2}, B_{i,2})$, and $A_{i,1}, B_{i,1}, A_{i,2}, B_{i,2} \in \mathbb{R}$. Therefore, for $\mathbf{x} = (x_1, x_2)$, a generalization of the univariate approach proposed by Braun, Duchesne and Stafford (2005) to the bivariate scenario gives the following estimator:

$$\hat{f}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} E_{\hat{f}}\left[K_{\mathbf{h}}(\mathbf{X}_i-\mathbf{x})\Big|\mathbf{X}_i \in \mathbf{I}_i\right], \tag{2.3}$$

which involves the conditional expectation of the kernel given that $\mathbf{X}_i$ lies within $\mathbf{I}_i$ (the information we know about $\mathbf{X}_i$). Here, the conditional expectation is with respect to the density $\hat{f}$.

Then, in terms of iterated conditional expectation, a solution to (2.3) should be

$$\hat{f}_j(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} E_{\hat{f}_{j-1}}\left[K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{x})\Big|\mathbf{X}_i \in \mathbf{I}_i\right]. \tag{2.4}$$

The expectation in (2.4) is with respect to the conditional density

$$\hat{f}_{j-1|\mathbf{I}_i}(\mathbf{u}) = \delta_i(\mathbf{u})\hat{f}_{j-1}(\mathbf{u})/\int_{\mathbf{I}_i}\hat{f}_{j-1}(\mathbf{s})\mathbf{ds}$$

over $\mathbf{I}_i$, where $\delta_i(\mathbf{u}) = 1$ if $\mathbf{u} \in \mathbf{I}_i$ and 0 otherwise.

Equations (2.4) and (2.2) imply that the conditional expectation with respect to $\hat{f}_{j-1}$ is employed to obtain $\hat{f}_j$. Note that, in both cases, we need to have an initial estimate $\hat{f}_0$ of the density.

2.1.3 Conditional Expectation

In (2.4), computation of the corresponding conditional expectation for each interval censored observation $\mathbf{X}_i$ is needed. Therefore we extend to the bivariate case the importance sampling scheme used by Duchesne and Stafford (2001). Let us define

$$\mu_{j-1|\mathbf{I}}(\mathbf{x}) = E_{\hat{f}_{j-1}}\left[K_{\mathbf{h}}(\mathbf{X} - \mathbf{x})\Big|\mathbf{X} \in \mathbf{I}\right]. \tag{2.5}$$

So, we may estimate (2.5) by using

$$E_{\hat{f}}\left[K_{\mathbf{h}}(\mathbf{X} - \mathbf{x})\Big|\mathbf{X} \in \mathbf{I}\right] = E_g\left[K_{\mathbf{h}}(\mathbf{X} - \mathbf{x})w(\mathbf{X})\right], \tag{2.6}$$

where g is a suitable distribution over the interval $\mathbf{I}$, and $w(\mathbf{X}) = \hat{f}_{j-1|\mathbf{I}}(\mathbf{X})/g(\mathbf{X})$ is the importance sampling weight.

The desired conditional expectation, (2.5), may be approximated by the following expression

$$\hat{\mu}_{j-1|\mathbf{I}}(\mathbf{x}) = \sum_{b=1}^{B} \left[K_{\mathbf{h}}(\mathbf{X}_b^u - \mathbf{x}) w_b^u \right]. \tag{2.7}$$

Here we let $g(\mathbf{X})$ be a bivariate uniform density and therefore the $\mathbf{X}_b^u$'s are generated over the interval $\mathbf{I}$ using a bivariate uniform sampling scheme (U sampling) derived from the orthogonal array-based Latin hypercubes described by Tang (1993) and $w_b^u = w(\mathbf{X}_b^u) / \sum_{k=1}^{B} w(\mathbf{X}_k^u)$, with $b = 1, \dots, B$. Tang's procedure establishes that B has to be a perfect square, i.e. $B \in \{1^2, 2^2, 3^2, \dots\}$. Figure 2.1 is reproduced from Tang (1993) to illustrate the case when $B = 4$. It also shows the difference between a four-point Orthogonal Array-Based Latin Hypercube design and a four-point Latin Hypercube design.

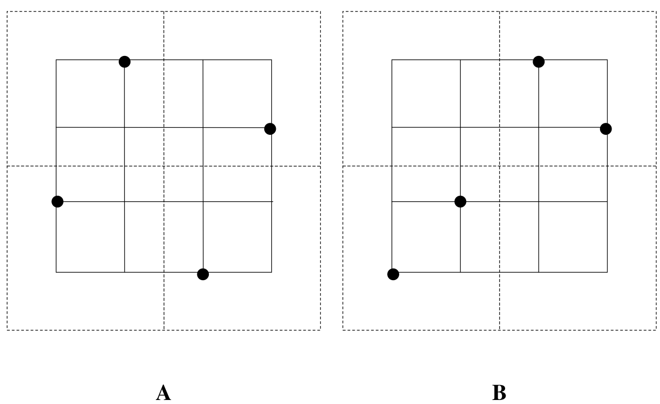

Figure 2.1: Orthogonal Array Based Latin Hypercube design (A) and Latin Hypercube design (B).

At the j^{th} step, an estimate of (2.4) may be obtained by

$$\hat{f}_j(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} \hat{\mu}_{j-1|\mathbf{I}_i}(\mathbf{x}). \tag{2.8}$$

2.2 Local likelihood density estimation

Kernel density estimation may present increased bias at and near a boundary. Wand and Jones (1995) present a discussion on this issue. One way to overcome this is by using a local likelihood approach, which we present next.

2.2.1 Univariate case

Local likelihood density estimation was introduced by Hjort and Jones (1996) and Loader (1996). In the presence of univariate non-censored data $Y_1, Y_2, ..., Y_n \in \mathbb{R}$, they defined equivalent local log-likelihood functions for density estimation. These are based on the concept of having an approximating parametric family, $f(y) = f(y; a) = f(y; a_0, a_1, \ldots, a_p)$, which may be locally estimated at y by maximizing the local log likelihood function.

Thus $a = (a_0, a_1, \ldots, a_p)$ is chosen to maximize the local log likelihood given by

$$L = \frac{1}{n}\sum_{i=1}^{n} K_h(Y_i - y)\log f(Y_i) - \int K_h(t-y)f(t)dt. \tag{2.9}$$

where $K_h(u) = \frac{1}{h}K(\frac{u}{h})$ is a kernel function with bandwidth h.

Maximization of (2.9) amounts to solving $\partial L/\partial a = 0$, i.e.

$$\frac{1}{n}\sum_{i=1}^{n} K_h(Y_i - y)A(y, Y_i, a) = \int K_h(t-y)A(y,t,a)f(t)dt, \tag{2.10}$$

with

$$A(y,t,a) = \left[\frac{\partial}{\partial a_1} logf(t,a), \ldots, \frac{\partial}{\partial a_p} logf(t,a)\right]^T.$$

Furthermore, Loader (1996) supposes that $\log f(t)$ can be approximated by a low-degree polynomial around y. That is

$$\log f(t) \approx P(t-y) = \sum_{i=0}^{p} a_i (t-y)^i;$$

therefore the local log likelihood (2.9) is

$$L \approx \frac{1}{n}\sum_{i=1}^{n} K_h(Y_i - y)P(Y_i - y) - \int K_h(t-y)\exp\bigl(P(t-y)\bigr)dt,$$

and the maximizing equation becomes

$$\frac{1}{n}\sum_{i=1}^{n} K_h(Y_i - y)A(y, Y_i, a) = \int K_h(t-y)A(y,t,a)\exp\bigl(P(t-y)\bigr)dt.$$

If $\tilde{a} = (\tilde{a}_0, \tilde{a}_1, \ldots, \tilde{a}_p)$ is a solution that maximizes (2.9), then the density estimate of $f(y)$ using this procedure is given by $\tilde{f}(y) = e^{\tilde{a}_0}$.

Braun, Duchesne and Stafford (2005) take this concept to the context of univariate interval censored data $X_1, X_2, ..., X_n \in \mathbb{R}$, proposing the following local log likelihood function

$$L_{ic} = \frac{1}{n}\sum_{i=1}^{n} E\left[K_h(X_i - x)\log\{f(X_i)\}\middle| X_i \in I_i\right] - \int K_h(t-x)f(t)dt,$$

where ic stands for interval censored. Using the polynomial approximation, $\partial L_{ic}/\partial a = 0$ leads to a system of local likelihood equations for the coefficients $(a_0, a_1, \ldots, a_p)$:

$$\frac{1}{n}\sum_{i=1}^{n} E\left[K(X_i - x)A(x, X_i, a)\middle| X_i \in I_i\right] - \int K_h(t-x)A(x,t,a)\exp\{P(t-x)\}dt = 0.$$

Solving leads to a local EM algorithm.

2.2.2 Bivariate case

We generalize the local log likelihood function to the bivariate scenario as follows

$$\begin{aligned} L_{ic} = \frac{1}{n}\sum E&\left[K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{x})P(\mathbf{X}_i - \mathbf{x})\middle|\mathbf{X}_i \in \mathbf{I}_i\right] \\ &- \int K_{\mathbf{h}}(\mathbf{t}-\mathbf{x})\exp\bigl(P(\mathbf{t}-\mathbf{x})\bigr)\mathbf{dt}, \end{aligned} \tag{2.11}$$

with the assumption that $\log f(\mathbf{t})$ can be approximated locally by

$$\log f(\mathbf{t}) \approx P(\mathbf{t}-\mathbf{x}) = a_0 + a_1(t_1 - x_1) + a_2(t_2 - x_2). \tag{2.12}$$

Then, maximization of (2.11) with respect to a_0, a_1 and a_2 amounts to solving the three equations

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} E\left[K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{x})A(\mathbf{x}, \mathbf{X}_i, a)\middle|\mathbf{X}_i \in \mathbf{I}_i\right] \\ - \int K_{\mathbf{h}}(\mathbf{t}-\mathbf{x})A(\mathbf{x}, \mathbf{t}, a)e^{P(\mathbf{t}-\mathbf{x})}\mathbf{dt} = 0,\end{aligned} \tag{2.13}$$

where $a = (a_0, a_1, a_2)$ and the corresponding score function $A(\mathbf{x}, \mathbf{t}, a) = (1, t_1 - x_1, t_2 - x_2)^T$.

Solving this system of log likelihood equations for the coefficients of (2.12) leads to a local EM algorithm as described in Braun, Duchesne and Stafford (2005).

Let us suppose that the logarithm of the density is locally constant, i.e. $\log f(\mathbf{t}) = a_0$. From equation (2.13), solving the system of one local likelihood equation with one unknown coefficient a_0 results in the following estimator for $f(\mathbf{x})$:

$$\frac{1}{n}\sum_{i=1}^{n} E\left[K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{x})\middle|\mathbf{X}_i \in \mathbf{I}_i\right] = \int K_{\mathbf{h}}(\mathbf{t}-\mathbf{x})e^{\tilde{a}_0}\mathbf{dt} = e^{\tilde{a}_0} = \tilde{f}(\mathbf{x}).$$

This corresponds to the kernel density estimate in (2.4) which is estimated by (2.8).

If the polynomial approximation is taken as in (2.12) and the product normal kernel (see Appendix A) is employed, we have to solve the following system of three local likelihood equations with three unknown coefficients a_0, a_1, and a_2:

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} E\left[K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{x})(1, X_{i,1} - x_1, X_{i,2} - x_2)^T\middle|\mathbf{X}_i \in \mathbf{I}_i\right] \\ = \int K_{\mathbf{h}}(\mathbf{t}-\mathbf{x})(1, t_1 - x_1, t_2 - x_2)^T e^{a_0 + a_1(t_1 - x_1) + a_2(t_2 - x_2)}\mathbf{dt}.\end{aligned} \tag{2.14}$$

If the solutions are $\tilde{a}_0, \tilde{a}_1$ and $\tilde{a}_2$, the local likelihood density estimate of $f(\mathbf{x})$ is given by $\tilde{f}(\mathbf{x}) = e^{\tilde{a}_0}$. The first equation in (2.14) leads to

$$\frac{1}{n}\sum_{i=1}^{n} E\left[K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{x})\Big|\mathbf{X}_i \in \mathbf{I}_i\right] = \int K_{\mathbf{h}}(\mathbf{t} - \mathbf{x})e^{\tilde{a}_0 + \tilde{a}_1(t_1 - x_1) + \tilde{a}_2(t_2 - x_2)}\mathbf{dt}.$$

Using the product gaussian kernel, this yields

$$\frac{1}{n}\sum_{i=1}^{n} E\left[K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{x})\Big|\mathbf{X}_i \in \mathbf{I}_i\right] = e^{\tilde{a}_0}\int K_{h_1}(t_1 - x_1)e^{\tilde{a}_1(t_1 - x_1)}dt_1 \int K_{h_2}(t_2 - x_2)e^{\tilde{a}_2(t_2 - x_2)}dt_2$$

or

$$\frac{1}{n}\sum_{i=1}^{n} E\left[K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{x})\Big|\mathbf{X}_i \in \mathbf{I}_i\right] = e^{\tilde{a}_0}e^{\frac{1}{2}(h_1\tilde{a}_1)^2}e^{\frac{1}{2}(h_2\tilde{a}_2)^2}$$

or

$$\frac{1}{n}\sum_{i=1}^{n} E\left[K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{x})\Big|\mathbf{X}_i \in \mathbf{I}_i\right] = e^{\tilde{a}_0}m(h_1\tilde{a}_1)m(h_2\tilde{a}_2), \tag{2.15}$$

where $m(\cdot)$ is the moment generating function of $y_k = t_k - x_k$, with $y_k \sim N(0, h_k^2), k = 1, 2$. This implies that

$$e^{\tilde{a}_0} = \frac{1}{n}\sum_{i=1}^{n} E\left[K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{x})\Big|\mathbf{X}_i \in \mathbf{I}_i\right]\left(m(h_1\tilde{a}_1)m(h_2\tilde{a}_2)\right)^{-1},$$

or

$$\tilde{f}(\mathbf{x}) = \hat{f}(\mathbf{x})e^{-\frac{1}{2}[(h_1\tilde{a}_1)^2 + (h_2\tilde{a}_2)^2]}, \tag{2.16}$$

where $\hat{f}(\mathbf{x})$ is obtained as in (2.8). To obtain $\tilde{a}_1$ and $\tilde{a}_2$, we proceed to solve the second and third equations from (2.14). The second equation in (2.14) gives

$$\frac{1}{n}\sum_{i=1}^{n} E\left[K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{x})(X_{i,1} - x_1)\Big|\mathbf{X}_i \in \mathbf{I}_i\right]$$
$$= \int K_{\mathbf{h}}(\mathbf{t} - \mathbf{x})(t_1 - x_1)e^{\tilde{a}_0 + \tilde{a}_1(t_1 - x_1) + \tilde{a}_2(t_2 - x_2)}\mathbf{dt}.$$

This implies that

$$\frac{1}{n}\sum_{i=1}^{n} E\left[K_{h_1}(X_{i,1}-x_1)(X_{i,1}-x_1)K_{h_2}(X_{i,2}-x_2)\Big|\mathbf{X}_i \in \mathbf{I}_i\right]$$
$$= e^{\tilde{a}_0}\int K_{h_1}(t_1-x_1)(t_1-x_1)e^{\tilde{a}_1(t_1-x_1)}dt_1 \int K_{h_2}(t_2-x_2)e^{\tilde{a}_2(t_2-x_2)}dt_2$$

or

$$\frac{1}{n}\frac{\partial}{\partial x_1}\sum_{i=1}^{n} E\left[K_{h_1}(X_{i,1}-x_1)h_1^2 K_{h_2}(X_{i,2}-x_2)\Big|\mathbf{X}_i \in \mathbf{I}_i\right]$$
$$= e^{\tilde{a}_0}h_1 m'(h_1\tilde{a}_1)m(h_2\tilde{a}_2)$$

or

$$h_1^2\frac{\partial}{\partial x_1}\hat{f}(x_1,x_2) = e^{\tilde{a}_0}h_1^2\tilde{a}_1 m(h_1\tilde{a}_1)m(h_2\tilde{a}_2). \tag{2.17}$$

In the same way, the third equation in (2.14) leads to

$$h_2^2\frac{\partial}{\partial x_2}\hat{f}(x_1,x_2) = e^{\tilde{a}_0}h_2^2\tilde{a}_2 m(h_2\tilde{a}_2)m(h_1\tilde{a}_1). \tag{2.18}$$

Dividing (2.17) and (2.18) correspondingly by (2.15) we obtain the expressions to be used in (2.16).

$$\frac{\frac{\partial}{\partial x_1}\hat{f}(x_1,x_2)}{\hat{f}(x_1,x_2)} = \tilde{a}_1,$$

$$\frac{\frac{\partial}{\partial x_2}\hat{f}(x_1,x_2)}{\hat{f}(x_1,x_2)} = \tilde{a}_2.$$

Therefore, in terms of iterated conditional expectation, the explicit expression for linear adjustments to the kernel density estimate is as follows:

$$\tilde{f}_j = \hat{f}_j(\mathbf{x})\exp\left[-\frac{1}{2}\sum_{k=1}^{2} h_i^2\left(\frac{\partial}{\partial x_k}\hat{f}_j(x_1,x_2)/\hat{f}_j(\mathbf{x})\right)^2\right], \tag{2.19}$$

which is parallel to the result of Hjort and Jones (1996) with the difference of having

$$\frac{\partial}{\partial x_k}\hat{f}_j(x_1, x_2) = \frac{1}{n}\sum_{i=1}^{n} E_{\hat{f}_{j-1}}\left[\frac{\partial}{\partial x_k}K_{\mathbf{h}}(\mathbf{X}-\mathbf{x})\Big|\mathbf{X}\in\mathbf{I}_i\right]. \tag{2.20}$$

Let us define

$$\mu^k_{j-1|\mathbf{I}}(\mathbf{x}) = E_{\hat{f}_{j-1}}\left[\frac{\partial}{\partial x_k}K_{\mathbf{h}}(\mathbf{X}-\mathbf{x})\Big|\mathbf{X}\in\mathbf{I}\right]. \tag{2.21}$$

In the same way as in Section 2.1, (2.21) may be approximated by the following expression

$$\hat{\mu}^k_{j-1|\mathbf{I}}(\mathbf{x}) = \sum_{b=1}^{B}\left(\frac{\partial}{\partial x_k}K_{\mathbf{h}}(\mathbf{X}^u_b-\mathbf{x})w^u_b\right),$$

where each w^u_b and $\mathbf{X}^u_b$ are obtained as described in Section 2.1. Since an estimate of the conditional expectation has been obtained using an importance sampling scheme, (2.20) can be approximated by

$$\frac{1}{n}\sum_{i=1}^{n}\hat{\mu}^k_{j-1|\mathbf{I}_i}(\mathbf{x}). \tag{2.22}$$

2.3 Survey weights

The estimates in Sections 2.1 and 2.2 do not consider the complexities of the survey design. The purpose of this section is to take into account some of these complexities by incorporating the survey weights into these estimates.

Let w^l_i be the longitudinal weight derived for individual i in the survey sample. This weight is broadly interpretable as the number of subjects represented by subject i in the population at the time of recruitment. Section 1.4.3 indicates that the survey weights are constructed to compensate for nonresponse, selection bias, stratification and postratification. An important feature of the longitudinal weights is that they add up to the size of the population from which the longitudinal sample was selected.

Let w^*_i be the standarized weight for individual i on the survey sample so that $\sum_{i\in S} w^*_i =$

1, where S corresponds to the longitudinal sample. By replacement of the population totals in (2.8) by weighted totals, a kernel density estimate which accounts for some of the complexities of the survey design is given in the following form

$$\hat{f}_j^w(\mathbf{x}) = \sum_{i \in S} \hat{\mu}_{j-1|\mathbf{I}_i}(\mathbf{x}) w_i^*. \tag{2.23}$$

For the estimate (2.22), we propose the following weighted expression

$$\hat{f}_j^{k,w}(\mathbf{x}) = \sum_{i \in S} \hat{\mu}_{j-1|\mathbf{I}_i}^k(\mathbf{x}) w_i^*. \tag{2.24}$$

Consequently, the corresponding weighted estimate for local linear adjustments to the kernel density estimate in (2.19) is given by

$$\tilde{f}_j^w = \hat{f}_j^w(\mathbf{x}) \exp\left[-\frac{1}{2}\sum_{k=1}^{2} h_i^2 \left(\hat{f}_j^{k,w}(\mathbf{x})/\hat{f}_j^w(\mathbf{x})\right)^2\right]. \tag{2.25}$$

The test statistic Q in (1.2) may be computed from (2.25) as outlined in Section 1.2.

2.4 Standard Error

To test the null hypothesis in Section 1.2, it is necessary to obtain an estimate of the standard error of Q in (1.2). The method we use for estimation of the design-based variance of the non-linear statistic (1.2) is the survey bootstrap as outlined in Rao, Wu and Yue (1992). Statistics Canada, the statistical agency whose data will be analysed in Chapter 4, produces a large number (500 and more) of bootstrap replicates of the survey weights for most of its national longitudinal surveys. These bootstrap weights allow for calculation of correct design-based variance estimators (Rao and Wu, 1988; Yung, 1997).

Let $w_i^{(r)}$ be the normalized bootstrap weight of the r^{th} replicate for individual i, such that $\sum_{i \in S} w_i^{(r)} = 1$. If we employ R of these bootstrap weight replicates, the required standard error can be assessed as follows:

For each set r of replicates:

1. Obtain (2.25) using $w_i^{(r)}$ (instead of w_i^*) for $i \in S$. Let $\tilde{f}_j^{w(r)}$ be the corresponding estimate.

2. Calculate (1.2) using $\tilde{f}_j^{w(r)}$ and call it Q_r^*.

Finally, compute $var^*(Q) = \frac{1}{R-1} \sum_{r=1}^{R} (Q_r^* - \bar{Q}^*)^2$, where $\bar{Q}^* = R^{-1} \sum_{r=1}^{R} Q_r^*$ and obtain $se(Q) = \sqrt[2]{var^*(Q)}$.

2.5 Estimation of density with a cusp or discontinuity

In this section, we work out the local likelihood density estimation method for a joint density with a cusp or a discontinuity in a certain region. Figures 2.2, 2.3 and 2.3 depict some examples of such densities in the univariate case. By way of illustration, we refer first to a univariate and non-censored case. We develop a local likelihood estimator suitable for a density which is believed to have a value greater than zero at and near a known boundary as in the case of Figure 2.3. Then we generalize the idea to the bivariate case for estimating a joint density with a discontinuity along the line $y_1 = y_2$ and indicate how the method can be applied for interval censored and complex survey data.

The idea is to estimate the density on each side of the discontinuity point or line using observations only from that side.

2.5.1 Univariate Case

Estimation at and near a Boundary

Let us suppose first that 0 is a boundary point and that near $y(> 0)$ $\log f(t)$ may be locally approximated by a constant

$$\log f(t) \approx a_{0+}. \tag{2.26}$$

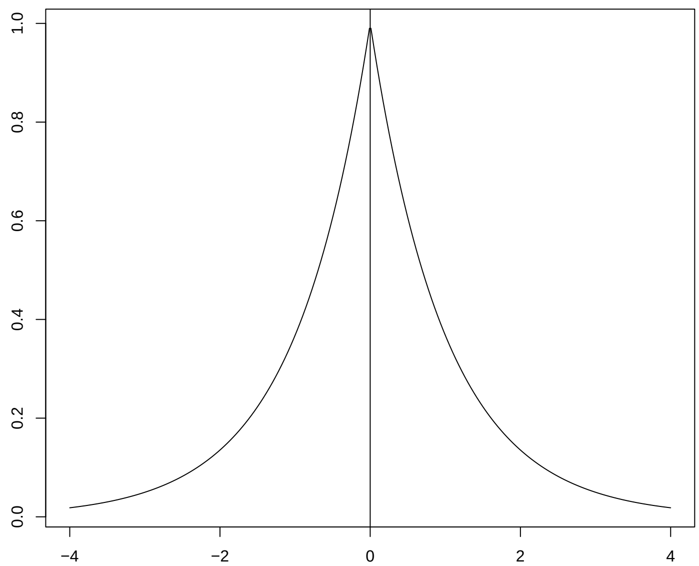

Figure 2.2: Cusp at zero.

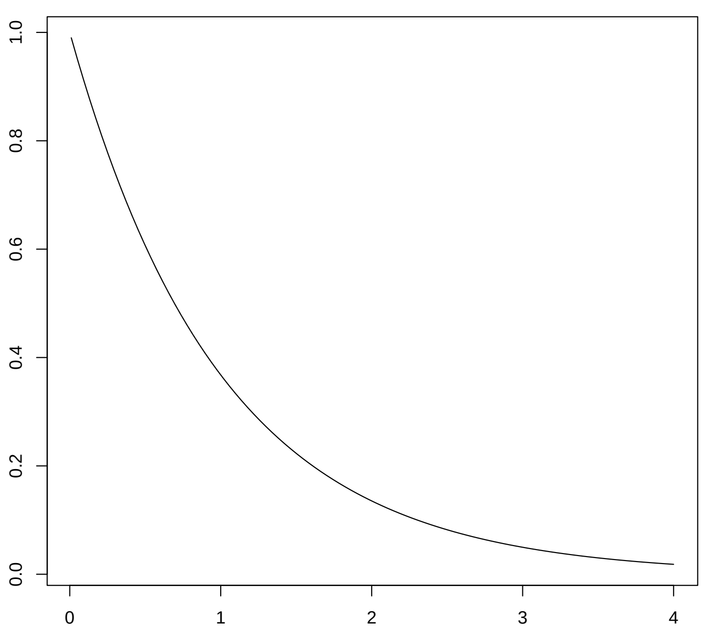

Figure 2.3: Discontinuity at zero.

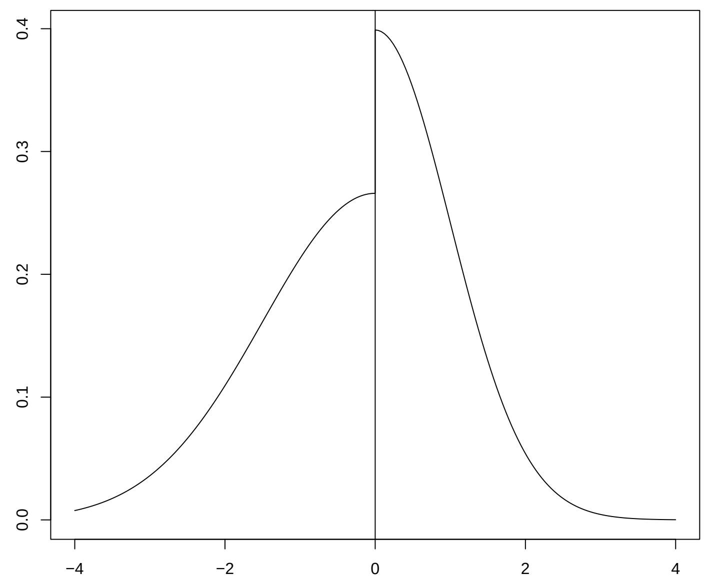

Figure 2.4: Discontinuity at zero.

Maximization of (2.9) amounts to solving $\partial L/\partial a = 0$. An analogue using only positive observations leads to the following equation for $y > 0$:

$$\frac{1}{n}\sum 1(Y_i \geq 0)K_h(Y_i - y) = \int_0^\infty K_h(t-y)e^{a_{0+}}dt. \tag{2.27}$$

Let us denote the left hand-side of (2.27) by $B(y,h)$. If $\tilde{a}_{0+}$ is a solution to (2.27), then we have that

$$B(y,h) = e^{\tilde{a}_{0+}}\int_0^\infty K_h(t-y)dt.$$

Assuming the gaussian kernel we have

$$B(y,h) = e^{\tilde{a}_{0+}}\frac{1}{\sqrt{2\pi}h}\int_0^\infty e^{-\frac{1}{2}\left(\frac{t-y}{h}\right)^2}dt$$

$$= e^{\tilde{a}_{0+}}\frac{1}{\sqrt{2\pi}}\int_{-y/h}^\infty e^{-\frac{1}{2}s^2}ds$$

where $s = \frac{t-y}{h}$ and $ds = \frac{dt}{h}$. Therefore

$$B(y,h) = e^{\tilde{a}_{0+}} \Phi\left(\frac{y}{h}\right), \tag{2.28}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function.

Consequently a density estimate when $\log f$ is locally approximated by a constant is given by

$$e^{\tilde{a}_{0+}} = \left(\frac{B(y,h)}{\Phi\left(\frac{y}{h}\right)}\right), \tag{2.29}$$

where $\tilde{a}_{0+}$ depends on y and h.

Now, let us assume that $\log f(t)$ can be locally approximated by a polynomial (t near y, $t > 0$) up to the linear term:

$$\log f(t) \approx P(t-y) = a_{0+} + a_{1+}(t-y). \tag{2.30}$$

Maximization of (2.9) which amounts to solving $\partial L / \partial a = 0$ yields the following system of 2 local likelihood equations with 2 unknown coefficients a_{0+} and a_{1+} :

$$\frac{1}{n} \sum 1(Y_i \geq 0) K_h(Y_i - y)(1, Y_i - y)^T = \int_0^\infty K_h(t-y)(1, t-y)^T e^{a_{0+} + a_{1+}(t-y)} dt. \tag{2.31}$$

The first equation leads to

$$\frac{1}{n} \sum 1(Y_i \geq 0) K_h(Y_i - y) = \int_0^\infty K_h(t-y) e^{a_{0+} + a_{1+}(t-y)} dt. \tag{2.32}$$

Let us again denote by $B(y,h)$ the left hand-side of (2.32). Using the normal kernel we have that

$$B(y,h) = e^{\tilde{a}_{0+}} \frac{1}{\sqrt{2\pi} h} \int_0^\infty e^{\tilde{a}_{1+}(t-y)} e^{-\frac{1}{2}\left(\frac{t-y}{h}\right)^2} dt$$

or

$$B(y,h) = e^{\tilde{a}_{0+}} \frac{1}{\sqrt{2\pi}} \int_{-y/h}^{\infty} e^{\tilde{a}_{1+}sh} e^{-\frac{1}{2}s^2} ds$$

where $s = \frac{t-y}{h}$ and $ds = \frac{dt}{h}$. Let $\lambda = a_{1+}h$. Then,

$$B(y,h) = e^{\tilde{a}_{0+}} \frac{1}{\sqrt{2\pi}} \int_{-y/h}^{\infty} e^{\tilde{\lambda}s} e^{-\frac{1}{2}s^2} ds$$

$$= e^{\tilde{a}_{0+}} e^{\frac{1}{2}\tilde{\lambda}^2} \frac{1}{\sqrt{2\pi}} \int_{-y/h}^{\infty} e^{-\frac{1}{2}(s-\tilde{\lambda})^2} ds$$

Let $r = s - \lambda$ and $dr = ds$. Then,

$$B(y,h) = e^{\tilde{a}_{0+}} e^{\frac{1}{2}\tilde{\lambda}^2} \frac{1}{\sqrt{2\pi}} \int_{-\frac{y}{h}-\tilde{\lambda}}^{\infty} e^{-\frac{1}{2}r^2} dr$$

$$= e^{\tilde{a}_{0+}} e^{\frac{1}{2}\tilde{\lambda}^2} \Phi\left(\frac{y}{h} + \tilde{\lambda}\right).$$

If we denote

$$m_+(\lambda) = \frac{1}{\sqrt{2\pi}} \int_{-y/h}^{\infty} e^{\lambda s} e^{-\frac{1}{2}s^2} ds = e^{\frac{1}{2}\lambda^2} \Phi\left(\frac{y}{h} + \lambda\right),$$

then we can also say that

$$B(y,h) = e^{\tilde{a}_{0+}} m_+(\tilde{\lambda}). \tag{2.33}$$

The second expression in (2.31) gives

$$\frac{1}{n} \sum 1(Y_i \geq 0) K_h(Y_i - y)(Y_i - y) = \int_0^{\infty} K_h(t-y)(t-y) e^{\tilde{a}_{0+} + \tilde{a}_{1+}(t-y)} dt. \tag{2.34}$$

Correspondingly, we denote the left hand-side of (2.34) by $C(y,h)$. Using the normal kernel we obtain

$$C(y,h) = e^{\tilde{a}_{0+}} \frac{1}{\sqrt{2\pi}h} \int_0^{\infty} e^{-\frac{1}{2}\left(\frac{t-y}{h}\right)^2} (t-y) e^{\tilde{a}_{1+}(t-y)} dt.$$

Let $s = \frac{t-y}{h}$ and $ds = \frac{dt}{h}$. Then

$$C(y,h) = e^{\tilde{a}_{0+}} \frac{1}{\sqrt{2\pi}} \int_{-y/h}^{\infty} e^{-\frac{1}{2}s^2} she^{\tilde{a}_{1+}sh} ds$$

or

$$C(y,h) = e^{\tilde{a}_{0+}} h \frac{1}{\sqrt{2\pi}} \int_{-y/h}^{\infty} e^{-\frac{1}{2}s^2} se^{\tilde{\lambda}s} ds$$

where $\tilde{\lambda} = \tilde{a}_{1+}h$.

Note that

$$m_+^{'}(\lambda) = \frac{d}{d\lambda} m_+(\lambda) = \frac{1}{\sqrt{2\pi}} \int_{-y/h}^{\infty} e^{-\frac{1}{2}s^2} se^{\lambda s} ds.$$

Then,

$$C(y,h) = e^{\tilde{a}_{0+}} hm_+^{'}(\tilde{\lambda}). \tag{2.35}$$

From (2.33) and (2.35) we respectively have that

$$e^{\tilde{a}_{0+}} = \frac{B(y,h)}{m_+(\tilde{\lambda})} \tag{2.36}$$

and

$$e^{\tilde{a}_{0+}} = \frac{C(y,h)}{hm_+^{'}(\tilde{\lambda})}. \tag{2.37}$$

Therefore

$$\frac{C(y,h)}{hm_+^{'}(\tilde{\lambda})} = \frac{B(y,h)}{m_+(\tilde{\lambda})}, \tag{2.38}$$

which implies that

$$\frac{C(y,h)}{hB(y,h)} = \frac{m_+^{'}(\tilde{\lambda})}{m_+(\tilde{\lambda})}. \tag{2.39}$$

On the other hand note that integration by parts leads to

$$e^{\lambda s} e^{-\frac{1}{2}s^2}|_{-y/h}^{\infty} = \lambda \int_{-y/h}^{\infty} e^{\lambda s} e^{-\frac{1}{2}s^2} ds - \int_{-y/h}^{\infty} e^{\lambda s} se^{-\frac{1}{2}s^2} ds \tag{2.40}$$

or

$$0 - e^{-\lambda(y/h)} e^{-\frac{1}{2}(y/h)^2} = \lambda\sqrt{2\pi} m_+(\lambda) - \sqrt{2\pi} m_+^{'}(\lambda).$$

Thus

$$m_+^{'}(\lambda) = \frac{1}{\sqrt{2\pi}} \left(e^{-\left(\lambda(y/h) + \frac{1}{2}(y/h)^2\right)} + \lambda\sqrt{2\pi} m_+(\lambda) \right)$$

or

$$m_+^{'}(\lambda) = e^{\frac{1}{2}\lambda^2} \phi\left(\frac{y}{h} + \lambda\right) + \lambda m_+(\lambda), \tag{2.41}$$

where $\phi(\cdot)$ is the standard normal probability density function.

Let us denote $\frac{C(y,h)}{B(y,h)}$ by $\alpha(y,h)$. Then, using (2.39) and (2.41) we have that

$$\frac{\alpha(y,h)}{h} = \frac{e^{\frac{1}{2}\tilde{\lambda}^2} \phi(\frac{y}{h} + \tilde{\lambda}) + \tilde{\lambda} m_+(\tilde{\lambda})}{m_+(\tilde{\lambda})}. \tag{2.42}$$

Solving for $\tilde{\lambda}$ leads to

$$\tilde{\lambda} = \frac{\alpha(y,h)}{h} - \frac{e^{\frac{1}{2}\tilde{\lambda}^2} \phi(\frac{y}{h} + \tilde{\lambda})}{m_+(\tilde{\lambda})} \tag{2.43}$$

or

$$\tilde{\lambda} = \frac{\alpha(y,h)}{h} - \frac{e^{\frac{1}{2}\tilde{\lambda}^2} \phi(\frac{y}{h} + \tilde{\lambda})}{e^{\frac{1}{2}\tilde{\lambda}^2} \Phi(\frac{y}{h} + \tilde{\lambda})} \tag{2.44}$$

and consequently

$$\tilde{\lambda} = \frac{\alpha(y,h)}{h} - \frac{\phi(\frac{y}{h} + \tilde{\lambda})}{\Phi(\frac{y}{h} + \tilde{\lambda})}. \tag{2.45}$$

We can solve (2.45) for $\tilde{\lambda}$ and use the following expression, derived from (2.33), to obtain $e^{\tilde{a}_{0+}}$, the density estimate at y:

$$e^{\tilde{a}_{0+}} = \frac{B(y,h)}{m_+(\tilde{\lambda})}. \tag{2.46}$$

Note that if $\tilde{\lambda}$ is close to 0, (2.46) is close to (2.29) which was derived by assuming that the density is locally constant to the right of the boundary.

Example

Suppose we have a random sample $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$, which is distributed according to an exponential probability density function with intensity 1. If we estimate the p.d.f. of $\mathbf{Y}$ using a standard kernel density estimator, we observe bias at and near the boundary as shown in Figure 2.5. This picture also depicts the estimate obtained by using the adapted local likelihood method, which shows very little bias.

In Figure 2.5, the sample size is $n = 1000$. In addition, the kernel density estimate shown in Figure 2.5 uses a gaussian kernel with bandwidth $= 0.1754$. The choice of this smoothing parameter was according to *Silverman's rule of thumb* which is presented in Silverman (1986, page 48, equation 3.31). The smoothing parameter ($= 1$) for the adapted local likelihood method was selected by using a least squares cross validation method (Silverman; 1986, page 49).

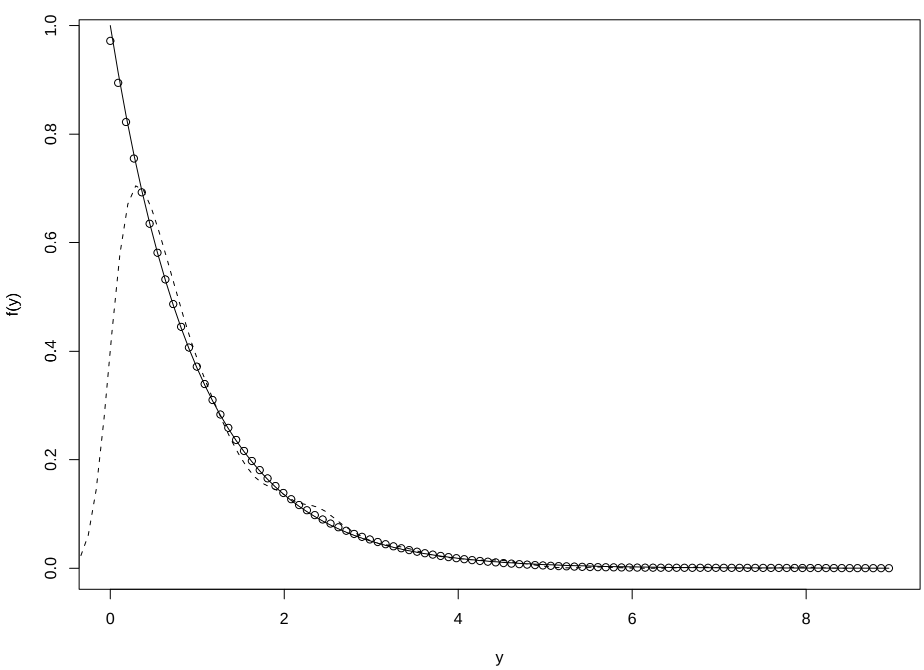

Figure 2.5: Solid line: f(y) = exp(-y). Dotted line: kernel density estimate. Circles: adapted local likelihood estimate.

Besides the bias at and near zero, we observe that the kernel density estimate in Figure 2.5 shows an anomaly near $y = 2$, which is not present using the adapted local likelihood density estimation approach.

2.5.2 Bivariate Case

We now proceed to generalize the previous idea to the bivariate case for estimating a joint density with a discontinuity along the line $y_1 = y_2$. (Later this will be applied to event times T_1 and T_2). We also indicate how the method can be applied for interval censored and complex survey data.

Let us now consider the following transformation. Let

$$v_1 = (y_1 + y_2)/\sqrt{2} \tag{2.47}$$

and

$$v_2 = (y_2 - y_1)/\sqrt{2}. \tag{2.48}$$

Assuming that $\log f(\mathbf{t})$ may be locally approximated by a constant a_{0+} for positive v_2, the corresponding estimating equation is given by

$$\frac{1}{n}\sum 1(V_{i,2} \geq 0)K_{\mathbf{h}}(\mathbf{V}_i - \mathbf{v}) = \int_{-\infty}^{\infty}\int_{0}^{\infty} K_{\mathbf{h}}(\mathbf{t} - \mathbf{v})e^{a_{0+}}\mathbf{dt}. \tag{2.49}$$

Let us denote the left hand-side of (2.49) by $\mathbf{B}(\mathbf{v}, \mathbf{h})$. If $\tilde{a}_{0+}$ is a solution to (2.49), then assuming the product gaussian kernel, it can be shown that

$$e^{\tilde{a}_{0+}} = \left(\frac{\mathbf{B}(\mathbf{v}, \mathbf{h})}{\Phi\left(\frac{v_2}{h_2}\right)}\right). \tag{2.50}$$

Now, consider the assumption that $\log f(\mathbf{t})$ can be approximated for example by

$$\log f(\mathbf{t}) \approx P(\mathbf{t}-\mathbf{v}) = a_{0+} + a_{1+}(t_1 - v_1) + a_{2+}(t_2 - v_2). \tag{2.51}$$

In this case, solving $\partial L_{biv}/\partial a = 0$ is equivalent to solving the following system in 3 equations with 3 unknown coefficients a_{0+}, a_{1+} and a_{2+}:

$$\begin{aligned} \frac{1}{n}\sum 1(V_{i,2} \geq 0) K_{\mathbf{h}}(\mathbf{V}_i - \mathbf{v})(1, V_{i,1} - v_1, V_{i,2} - v_2)^T = \\ \int_{-\infty}^{\infty}\int_{0}^{\infty} K_{\mathbf{h}}(\mathbf{t}-\mathbf{v})(1, t_1 - v_1, t_2 - v_2)^T e^{a_{0+}+a_{1+}(t_1-v_1)+a_{2+}(t_2-v_2)} \mathbf{dt}. \end{aligned} \tag{2.52}$$

The first equation in (2.52) leads to

$$\begin{aligned} \frac{1}{n}\sum 1(V_{i,2} \geq 0) K_{\mathbf{h}}(\mathbf{V}_i - \mathbf{v}) = \\ \int_{-\infty}^{\infty}\int_{0}^{\infty} K_{\mathbf{h}}(\mathbf{t}-\mathbf{v}) e^{a_{0+}+a_{1+}(t_1-v_1)+a_{2+}(t_2-v_2)} \mathbf{dt}. \end{aligned} \tag{2.53}$$

Let us denote by $\mathbf{B}(\mathbf{v},\mathbf{h})$ the left hand-side of (2.53). Using the product normal kernel (with respect to the new coordinates) it can be shown that

$$\mathbf{B}(\mathbf{v},\mathbf{h}) = e^{\tilde{a}_{0+}} m(\tilde{\lambda}_1) m_+(\tilde{\lambda}_2), \tag{2.54}$$

where $\lambda_1 = a_{1+}h_1$ and $m(\lambda_1)$ is the moment generating function of W_1, where W_1 follows a normal distribution with mean 0 and variance h_1^2. Equation (2.54) implies that

$$e^{\tilde{a}_{0+}} = \frac{\mathbf{B}(\mathbf{v},\mathbf{h})}{m(\tilde{\lambda}_1) m_+(\tilde{\lambda}_2)}. \tag{2.55}$$

The second equation in (2.52) gives

$$\begin{aligned}\frac{1}{n}\sum 1(V_{i,2}\geq 0)K_{\mathbf{h}}(\mathbf{V}_i-\mathbf{v})(V_{i,1}-v_1) &=\\ \int_{-\infty}^{\infty}\int_{0}^{\infty} K_{\mathbf{h}}(\mathbf{t}-\mathbf{v})(t_1-v_1)e^{\tilde{a}_{0+}+\tilde{a}_{1+}(t_1-v_1)+\tilde{a}_{2+}(t_2-v_2)}\mathbf{dt}.\end{aligned} \tag{2.56}$$

Correspondingly, we denote the left hand-side of (2.56) by $\mathbf{C}(\mathbf{v},\mathbf{h})$. Using the product normal kernel it can be shown that

$$\mathbf{C}(\mathbf{v},\mathbf{h}) = e^{\tilde{a}_{0+}}h_1 m^{'}(\tilde{\lambda}_1)m_+(\tilde{\lambda}_2). \tag{2.57}$$

From (2.57) we have that

$$e^{\tilde{a}_{0+}} = \frac{\mathbf{C}(\mathbf{v},\mathbf{h})}{h_1 m^{'}(\tilde{\lambda}_1)m_+(\tilde{\lambda}_2)}. \tag{2.58}$$

Therefore from (2.57) and (2.54) we have that

$$\frac{\mathbf{C}(\mathbf{v},\mathbf{h})}{h_1\mathbf{B}(\mathbf{v},\mathbf{h})} = \frac{m^{'}(\tilde{\lambda}_1)}{m(\tilde{\lambda}_1)} = \tilde{\lambda}_1. \tag{2.59}$$

The third equation in (2.52) yields

$$\begin{aligned}\frac{1}{n}\sum 1(V_{i,2}\geq 0)K_{\mathbf{h}}(\mathbf{V}_i-\mathbf{v})(V_{i,2}-v_2) &=\\ \int_{-\infty}^{\infty}\int_{0}^{\infty} K_{\mathbf{h}}(\mathbf{t}-\mathbf{v})(t_2-v_2)e^{\tilde{a}_{0+}+\tilde{a}_{1+}(t_1-v_1)+\tilde{a}_{2+}(t_2-v_2)}\mathbf{dt}.\end{aligned} \tag{2.60}$$

We denote the left hand-side of (2.60) by $\mathbf{D}(\mathbf{v},\mathbf{h})$. It can be shown by using the product gaussian kernel that

$$\mathbf{D}(\mathbf{v},\mathbf{h}) = e^{\tilde{a}_{0+}}h_2 m_+^{'}(\tilde{\lambda}_2)m(\tilde{\lambda}_1). \tag{2.61}$$

Therefore

$$\frac{\mathbf{D}(\mathbf{v},\mathbf{h})}{h_2\mathbf{B}(\mathbf{v},\mathbf{h})} = \frac{m_+^{'}(\tilde{\lambda}_2)}{m_+(\tilde{\lambda}_2)} = \frac{e^{\frac{1}{2}\tilde{\lambda}_2^2}\phi(\frac{v_2}{h_2}+\tilde{\lambda}_2)+\tilde{\lambda}_2 m_+(\tilde{\lambda}_2)}{m_+(\tilde{\lambda}_2)}. \tag{2.62}$$

Solving for $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ leads to

$$\tilde{\lambda}_1 = \frac{\mathbf{C}(\mathbf{v},\mathbf{h})}{h_1\mathbf{B}(\mathbf{v},\mathbf{h})}. \tag{2.63}$$

and

$$\tilde{\lambda}_2 = \frac{\mathbf{D}(\mathbf{v},\mathbf{h})}{h_2\mathbf{B}(\mathbf{v},\mathbf{h})} - \frac{\phi(\frac{v_2}{h_2}+\tilde{\lambda}_2)}{\Phi(\frac{v_2}{h_2}+\tilde{\lambda}_2)}. \tag{2.64}$$

Solve (2.63) and (2.64) for $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ and use (2.55) to obtain the density estimate.

Interval censored and complex survey data

Recall the following notation from (2.49), (2.56) and (2.60):

$$\mathbf{B}(\mathbf{v},\mathbf{h}) = \frac{1}{n}\sum 1(V_{i,2}\geq 0)K_{\mathbf{h}}(\mathbf{V}_i-\mathbf{v}),$$

$$\mathbf{C}(\mathbf{v},\mathbf{h}) = \frac{1}{n}\sum 1(V_{i,2}\geq 0)K_{\mathbf{h}}(\mathbf{V}_i-\mathbf{v})(V_{i,1}-v_1)$$

and

$$\mathbf{D}(\mathbf{v},\mathbf{h}) = \frac{1}{n}\sum 1(V_{i,2}\geq 0)K_{\mathbf{h}}(\mathbf{V}_i-\mathbf{v})(V_{i,2}-v_2).$$

In the presence of interval censored data, we follow the ideas of Sections 2.1.2 and 2.1.3. Therefore $\mathbf{B}(\mathbf{v},\mathbf{h}), \mathbf{C}(\mathbf{v},\mathbf{h})$ and $\mathbf{D}(\mathbf{v},\mathbf{h})$ may be substituted by

$$\frac{1}{n}\sum_{i=1}^{n} E_{\hat{f}}\left[1(V_{i,2}\geq 0)K_{\mathbf{h}}(\mathbf{V}_i-\mathbf{v})\middle|\mathbf{V}_i\in\mathbf{I}'_i\right], \tag{2.65}$$

$$\frac{1}{n}\sum_{i=1}^{n} E_{\hat{f}}\left[1(V_{i,2}\geq 0)K_{\mathbf{h}}(\mathbf{V}_i-\mathbf{v})(V_{i,1}-v_1)\middle|\mathbf{V}_i\in\mathbf{I}'_i\right], \tag{2.66}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} E_{\hat{f}}\left[1(V_{i,2}\geq 0)K_{\mathbf{h}}(\mathbf{V}_i-\mathbf{v})(V_{i,2}-v_2)\middle|\mathbf{V}_i\in\mathbf{I}'_i\right], \tag{2.67}$$

where $\mathbf{I}'_i$ is determined by the transformations in (2.47) and (2.48); in the (v_1, v_2) space, $\mathbf{I}'_i$ is the rotation of a rectangle aligned with the coordinates in the (y_1, y_2) space.

Correspondingly, using an importance sampling scheme, the conditional expectations in (2.65), (2.66) and (2.67) may be approximated by the following expressions

$$\hat{\mu}_{\mathbf{B}|\mathbf{I}}(\mathbf{v}) = \sum_{r=1}^{R} 1(V^u_{r,2} \geq 0) K_{\mathbf{h}}(\mathbf{V}^u_r - \mathbf{v}) w^u_r. \tag{2.68}$$

$$\hat{\mu}_{\mathbf{C}|\mathbf{I}}(\mathbf{v}) = \sum_{r=1}^{R} 1(V^u_{r,2} \geq 0) K_{\mathbf{h}}(\mathbf{V}^u_r - \mathbf{v})(V_{r,1} - v_1) w^u_r. \tag{2.69}$$

$$\hat{\mu}_{\mathbf{D}|\mathbf{I}}(\mathbf{v}) = \sum_{r=1}^{R} 1(V^u_{r,2} \geq 0) K_{\mathbf{h}}(\mathbf{V}^u_r - \mathbf{v})(V_{r,2} - v_2) w^u_r. \tag{2.70}$$

where $\mathbf{V}^u_r$ is generated over the interval $\mathbf{I}'$ using the previously mentioned bivariate U sampling and $w^u_r = w(\mathbf{V}^u_r)/\sum_{k=1}^{R} w(\mathbf{V}^u_k)$, with $r = 1, \ldots, R$ and $R \in \{1^2, 2^2, 3^2, \ldots\}$.

Taking into account some of the complexities of the design, the weighted versions of (2.65), (2.66) and (2.67) are given by

$$\mathbf{B}^w(\mathbf{v}, \mathbf{h}) = \sum_{i \in S} \hat{\mu}_{\mathbf{B}|\mathbf{I}_i}(\mathbf{v}) w^*_i \tag{2.71}$$

$$\mathbf{C}^w(\mathbf{v}, \mathbf{h}) = \sum_{i \in S} \hat{\mu}_{\mathbf{C}|\mathbf{I}_i}(\mathbf{v}) w^*_i \tag{2.72}$$

and

$$\mathbf{D}^w(\mathbf{v}, \mathbf{h}) = \sum_{i \in S} \hat{\mu}_{\mathbf{D}|\mathbf{I}_i}(\mathbf{v}) w^*_i \tag{2.73}$$

where w^*_i are the survey weights of Section 2.3.

Therefore analogue expressions to (2.55), (2.63) and (2.64) for the interval censored and complex survey data case may be given by

$$e^{\tilde{a}_{0+}} = \frac{\mathbf{B}^w(\mathbf{v}, \mathbf{h})}{m(\tilde{\lambda}_1) m_+(\tilde{\lambda}_2)} \tag{2.74}$$

$$\tilde{\lambda}_1 = \frac{\mathbf{C}^w(\mathbf{v}, \mathbf{h})}{h_1 \mathbf{B}^w(\mathbf{v}, \mathbf{h})} \tag{2.75}$$

and

$$\tilde{\lambda}_2 = \frac{\mathbf{D}^w(\mathbf{v}, \mathbf{h})}{h_2 \mathbf{B}^w(\mathbf{v}, \mathbf{h})} - \frac{\phi(\frac{v_2}{h_2} + \tilde{\lambda}_2)}{\Phi(\frac{v_2}{h_2} + \tilde{\lambda}_2)}. \tag{2.76}$$

Solve (2.75) and (2.76) for $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ and use (2.74) to obtain the density estimate.

2.6 Interval Censoring Asymptotic Theory

2.6.1 Univariate Case

In the context of interval censored data, nonparametric estimators for a univariate cumulative distribution function have been proposed widely. Recall from Section 1.3.2 the general setting for interval censored data. The value of the duration times, T, may not be directly observed, but there exist observed inspection or monitoring times $\{a_0, a_1, \ldots, a_m\}$ such that $T \in I = (A, B)$, where $A = \sup\{a_j : a_j \leq T\}$ (the last inspection time prior to occurrence of event E) and $B = \inf\{a_j : a_j \geq T\}$ (the first monitoring time after occurrence of E). Having n individuals, the set of monitoring times generalizes to $\{a_{i0}, a_{i1}, \ldots, a_{im_i}\}$ for $i = 1, \ldots, n$ and hence $T_i \in I_i = (A_i, B_i)$, where $A_i = \sup\{a_{ij} : a_{ij} \leq T_i\}$ and $B_i = \inf\{a_{ij} : a_{ij} \geq T_i\}$.

Suppose that the inspection times and the duration times are independent, in order to ensure that the censoring is noninformative. In addition, assume that no event time point occurs with positive probability among the inspection times, so that it is ensured that the occurrences of E cannot coincide with the monitoring times. Moreover, suppose T arises from the cumulative distribution $F(t)$. Then the likelihood conditional upon the observed intervals is:

$$\mathcal{L} = \prod_{i=1}^{n} \left(F(B_i) - F(A_i) \right). \tag{2.77}$$

Let $\mathcal{I} = \{[A_i, B_i], i = 1, \ldots, n\}$. Peto (1973) and Turnbull (1976) defined a new set of disjoint intervals, $\mathcal{J} = \{[c_j, d_j], j = 1, \ldots, J\}$, of which the elements are the intersections of the observed intervals in $\mathcal{I}$. Specifically, the left endpoints of the disjoint set of intervals in $\mathcal{J}$ lie in the set $\mathcal{A} = \{A_i, i = 1, \ldots, n\}$, the right endpoints lie in the set $\mathcal{B} = \{B_i, i = 1, \ldots, n\}$, and the intervals in $\mathcal{J}$ contain no other members of $\mathcal{A}$ and $\mathcal{B}$ except at their left and right endpoints respectively. Also, $c_1 \leq d_1 < c_2 \leq \cdots < c_J \leq d_J$. Turnbull (1976) proved that any cumulative distribution function that has mass outside of the union of $[c_j, d_j]$ for $j = 1, \ldots, J$ cannot be a maximum likelihood estimate of the true cumulative distribution function.

The likelihood can be rewritten as follows:

$$\mathcal{L} = \prod_{i=1}^{n} \Big(\sum_{j=1}^{J} \alpha_j^i p_j \Big), \tag{2.78}$$

where $p_j = F(d_j) - F(c_j)$ and $\alpha_j^i = 1$ if $[c_j, d_j]$ is a subset of $[A_i, B_i]$ and 0 otherwise. Therefore, the log likelihood is given by

$$l = \sum_{i=1}^{n} log \Big(\sum_{j=1}^{J} \alpha_j^i p_j \Big). \tag{2.79}$$

Let $\mathbf{p} = (p_1, \ldots, p_J)$. To find the nonparametric maximum likelihood estimate of $\mathbf{p}$ Gentleman and Geyer (1994) maximized l with respect to p subject to the constraints

$$1 - \sum_{j=1}^{J} p_j = 0, \tag{2.80}$$

and

$$p_j \geq 0 (j = 1, \ldots, J). \tag{2.81}$$

They noted that for a concave programming problem with linear constraints, the Kuhn-Tucker conditions (see Appendix A) are necessary and sufficient for optimality. In addition, their resulting estimator convergences to a unique maximum if the log likelihood is strictly concave, i.e. the Hessian H associated with the likelihood is strictly negative definite. They let A denote the $n \times J$ matrix with elements α_j^i; then $H = A'DA$, where D is the diagonal matrix with elements $-1/(\sum_{j=1}^{J} \alpha_j^i p_j)$. Hence, H will be of full rank and the maximum likelihood estimate will be unique if $\text{rank}(A) = J$ (Gentleman and Geyer; 1994, section 2.3). They also indicated that there may be situations in which the likelihood is concave, but not strictly concave, and the maximum likelihood estimator is unique nevertheless. Therefore, Gentleman and Geyer (1994, Theorem 1) provided a sufficient condition for uniqueness.

Gentleman and Geyer (1994, p. 621) stated that if the inspection time process samples $[0, \infty)$ densely as the number of individuals increases, their corresponding estimator of the cumulative distribution function converges strongly to the true distribution.

Yu, Schick, Li and Wong (1998) proved that if, on the other hand, the censoring vector takes on finitely many values as the number of individuals increases, then under additional assumptions their maximum likelihood estimator is asymptotically normally distributed.

It is important to mention that considering interval censoring, the convergence rate of the nonparametric maximum likelihood estimator of the cumulative distribution function is $n^{1/3}$ in general. See Groeneboom and Wellner (1992), Geskus and Groeneboom (1997), Geskus and Groeneboom (1999). However, considering partly interval censored data in which the exact values of some duration times are observed in addition to interval censored observations, Huang (1999) showed that the corresponding convergence rate can be $n^{1/2}$.

Other Algorithms

Turnbull (1976) proposed a simple self-consistent estimator which is not necessarily the maximum likelihood estimator.

Let $\mathbf{p} = (p_1, \ldots, p_J)$, where $\sum p_j = 1$ and $p_j \geq 0$. A self-consistent estimator of $\mathbf{p}$ is defined to be any solution of the simultaneous equations (Turnbull; 1976):

$$p_j = \pi_j(p_1, \ldots, p_J)$$

for $j = 1, \ldots, J$, where $\pi_j(\mathbf{p})$ is the expected proportion of observations in $[c_j, d_j]$, given $\mathcal{A}$ and $\mathcal{B}$. That is,

$$\pi_j(\mathbf{p}) = \frac{1}{n}\sum_{i=1}^{n} \mu_j^i(\mathbf{p})$$

with

$$\mu_j^i(\mathbf{p}) = \frac{\alpha_j^i p_j}{\sum_{k=1}^{J} \alpha_k^i p_k}.$$

Turnbull's nonparametric estimator of F, which is a self-consistent estimator of F, can be obtained by the following iterative procedure.

1. Obtain initial estimates by setting $p_j^0 = 1/J, 1 \leq j \leq J$.

2. Compute $\mu_j^i(\mathbf{p}^0)$ for $1 \leq i \leq n$ and $1 \leq j \leq J$.

3. Set $p_j^1 = \pi_j(\mathbf{p}^0), 1 \leq j \leq J$.

4. Return to step 2 with $\mathbf{p}^1$ replacing $\mathbf{p}^0$, and so on.

This algorithm is easy to implement but is known to have slow convergence. The algorithm converges monotonically to an estimate of the weight vector $\mathbf{p}$ (at least for $\mathbf{p}^0$ close enough to $\hat{\mathbf{p}}$). Another issue with this algorithm is that there can be self-consistency points other than the maximum likelihood estimate (see Gentleman and Geyer (1994) for an example).

Turnbull's nonparametric estimate $\hat{F}(x)$ of the cumulative distribution function $F(x)$ is equal to:

$$\begin{array}{lll} 0 & \text{if } x < c_1, & \\ \hat{p}_1 + \hat{p}_2 + \cdots + \hat{p}_j & \text{if } d_j < x < c_{j+1} & (1 \leq j \leq J-1), \\ 1 & \text{if } x > d_J, & \end{array}$$

and is not defined for $x \in [c_j, d_j]$ for $1 \leq j \leq J$. It is noted that the plot of $\hat{F}$ presents a series of $m+1$ horizontal lines of increasing heights with gaps in between. The location of the probability mass p_j associated with $[c_j, d_j]$ is left unspecified because we know only the amount of weight on the intervals $[c_j, d_j]$ but not the way the weights vary within these intervals.

Designed to fill these gaps and due to the non-uniqueness of Turnbull's estimator over innermost intervals, members of $\mathcal{J}$, Li, Watkins and Yu (1997) proposed an EM algorithm which coincides with Turnbull's where it is uniquely defined (outside of the union of $[c_j, d_j]$ for $j = 1, \ldots, J$), but converges over the innermost intervals to a value that depends on the starting point of the algorithm. The algorithm involves computing the conditional expectation of F_n, the empirical distribution function, at each step, that is

$$\tilde{F}_k(x) = E_{k-1}\{F_n(x) \mid I_i, i = 1, \ldots, n\}. \tag{2.82}$$

where E_{k-1} is expectation with respect to the distribution $\tilde{F}_{k-1}(x)$. Braun, Duchesne and Stafford (2005) proved that for a vanishing bandwidth h, their algorithm coincides with that of Li, Watkins and Yu (1997) and with Turnbull's. They provide a theorem that shows that at step j, (2.82) may be obtained as a limit of (2.2) as the bandwidth h of the kernel shrinks to zero. And this happens at every step j. So, their proposed algorithm modifies the usual self-consistency algorithms by introducing kernel smoothing at each step of the iteration (Braun et al.; 2005). Moreover, the estimator obtained from (2.2) has an attractive shape, is defined in regions of interest and is uniquely determined. However, the convergence of their algorithm to a unique fixed point will depend critically on h. The convergence will be based on having a sufficiently large bandwidth h (see Theorem 2 in Braun, Duchesne and Stafford (2005) for sufficient conditions to assure existence of a fixed point). We discuss convergence in more detail in Section 2.7.

Braun, Duchesne and Stafford (2005) pointed out that smoothing permits all data to influence the estimate at any location. Hence probability massed on the innermost intervals is smoothed repeatedly over the entire region and the extent to which this occurs depends on the size of the bandwidth h. The difficulty with Turnbull's algorithm and Li, Watkins and Yu's, is that $h = 0$ and no smoothing takes place. Braun et al (2005) also mentioned that their estimate may present the same difficulties as Turnbull's as h goes to 0, namely, convergence to a wrong fixed point, i.e. a fixed point other than the maximum likelihood estimate. An example of this situation is given in Gentleman and Geyer (1994).

The consistency and asymptotic normality of the estimator of Braun et al (2005) appear to be open questions. When no interval censoring is present, a gaussian kernel smoothed estimate is consistent if the density function is well behaved and the bandwidth $h \to 0$ and $nh \to \infty$ as $n \to \infty$. These conditions imply that, while the size of the smoothing parameter h must decrease as the sample size n increases, h must not converge to zero as rapidly as n^{-1} (Silverman; 1986, p.71). In addition, the density estimator is asymptotically distributed as a normal random variable if the density function is twice continuously differentiable and the bandwidth h is proportional to $n^{-1/5}$ (Härdle; 1990, p.62).

2.6.2 Bivariate Case

For the bivariate case, consider the estimation of a joint distribution function F_0 of a bivariate random vector $\mathbf{X}_i = (X_{i,1}, X_{i,2})$ which is subject to interval censoring. That is, $\mathbf{X}_i$ lies within the 2-dimensional interval $\mathbf{I}_i = (A_{i,1}, B_{i,1}) \times (A_{i,2}, B_{i,2})$.

Wong and Yu (1999) proposed nonparametric estimation of the distribution function based on multivariate interval censored data. Their method generalized the concept of Turnbull's estimator to the multivariate case. They generalized the concept of innermost intervals (they called them maximal intersections) to the multivariate case and proposed a likelihood function equivalent to (2.79). Therefore, they also use Turnbull's self-consistent algorithm to obtain an estimate. In addition, their estimator may not be unique (they provide an example of this situation).

Wong and Yu (1999) established consistency and asymptotic normality of their estimator. Essentially, strong consistency is guaranteed if F_0 is continuous and $\mathcal{G}_*$ is dense in $[0,\infty)^2$, where $\mathcal{G}_*$ is the set which contains the grid points generated by the multi-dimensional intervals ($\mathbf{I}_i = (A_{i,1}, B_{i,1}) \times (A_{i,2}, B_{i,2})$ in the bivariate case). Asymptotic normality of the estimator is obtained under an alternative assumption that $\mathcal{G}_*$ contains finitely many elements. Note that these conditions are in tune with the univariate results previously mentioned.

Betensky and Finkelstein (1999) extended the approach of Gentleman and Geyer (1994) to the bivariate case. Hence, they also used the Kuhn-Tucker conditions as necessary and sufficient for optimality. Their resulting estimator, as in the univariate case, is unique if the log likelihood is strictly concave. Therefore their resulting estimator converges to a unique maximum if the hessian associated with the likelihood is strictly negative definite. They also use Gentleman and Geyer (1994, Theorem 1) to provide a sufficient condition for uniqueness of their estimate.

In the bivariate case also, we may conjecture that a kernel smoothed density in the interval censored case would be consistent and asymptotically normal if the conditions of Wong and Yu (1999) are satisfied and if the conditions of (Silverman; 1986, p.71) and (Härdle; 1990, p.62) are satisfied assuming that $h = h_1 = h_2$.

2.7 Convergence

At this point we would like to mention the following properties of the estimator proposed by Braun, Duchesne and Stafford (2005) in the univariate i.i.d. case.

- Local EM algorithms are implemented by computing conditional expectations using numerical integration (setting out an equal-spaced mesh for that purpose).

- Convergence to a unique estimate in the local constant case is proven (the fixed point of their implementation does not depend on $\hat{f}_0$).

- Convergence of the fixed point iteration for sufficiently large bandwidths is guaranteed.

- Their results are for kernels with compact support but can be extended to kernels with noncompact support such as the gaussian kernel.

Braun, Duchesne and Stafford (2005) consider local EM algorithms (constant, linear and quadratic) in an attempt to solve

$$\mathbf{f} = G(\mathbf{f})$$

Note that in our case, the local EM algorithms are implemented by computing conditional expectations using numerical integration by means of a Monte Carlo approach (importance sampling). The attractiveness of our method relies on the implementation of a bivariate uniform sampling scheme derived from the orthogonal array-based latin hypercubes described by Tang (1993), also called U sampling. Below we assume for simplicity that the set of sample points once chosen is fixed.

In this Section we generalize the convergence results of Braun, Duchesne and Stafford (2005) to a bivariate scenario for the local constant case. First, we generalize the approach of Braun, Duchesne and Stafford (2005) to compute the required conditional expectations by using numerical integration over an equal spaced bivariate mesh. Second, we prove convergence of the fixed point iteration using our method, which computes the desired conditional expectations by means of the importance sampling approach. Third, the convergence of the fixed point iteration when the survey weights are incorporated is established.

The contraction mapping theorem (Ortega; 1972) is employed to prove convergence results.

Definition 1 *A function G from $D \subset \mathbb{R}^m$ into $\mathbb{R}^m$ has a fixed point at $\mathbf{p} \in D$ if $\mathbf{p} = G(\mathbf{p})$.*

Theorem 1 *Let D be a closed convex set. Suppose G is a continuous function from $D \subset \mathbb{R}^m$ into $\mathbb{R}^m$ with the property that $G(\mathbf{p}) \in D$ whenever $\mathbf{p} \in D$. Then G has a fixed point in D.*

Suppose in addition, that G is continuously differentiable on the convex set D and that

$$\| G'(\mathbf{p}) \|_\infty \leq \alpha < 1 \tag{2.83}$$

whenever $\mathbf{p} \in D$. Then the sequence $\{\mathbf{p}_j\}$ defined by an arbitrarily selected $\mathbf{p}_0$ in D and generated by

$$\mathbf{p}_j = G(\mathbf{p}_{j-1}),$$

converges to a unique fixed point $\mathbf{p}_{FP} \in D$.

In (2.83), $\| \cdot \|_\infty$ denotes the infinity-norm (i.e. the maximum row-sum of the absolute values of the matrix entries) and $G'(\mathbf{p})$ corresponds to the $m \times m$ Jacobian.

The framework of Braun, Duchesne and Stafford (2005) is extended to the bivariate scenario in the following manner. Let $\mathcal{M}_k$ be a univariate mesh consisting of m_k equidistant points $\{x_k^i\}_{i=1}^{m_k}$, with $x_k^i - x_k^{i-1} = \Delta_k$, for $k = 1, 2$. Let $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2$ be a bivariate mesh consisting of points $\{\mathbf{x}^r = (x_1, x_2)^r\}_{r=1}^m$, with $m = m_1 \times m_2$. Define $f^r = f(\mathbf{x}^r)$ with $\mathbf{f} = (f^1, f^2, \ldots, f^m)^T$. Therefore for $r = 1, \ldots, m$, the fixed point equation (2.3) is equivalent to

$$f^r = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\sum_{v:\mathbf{x}^v \in I_i} K_{\mathbf{h}}(\mathbf{x}^v - \mathbf{x}^r) f^v \Delta_1 \Delta_2}{\sum_{v:\mathbf{x}^v \in I_i} f^v \Delta_1 \Delta_2} \right) \tag{2.84}$$

or

$$f^r = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\sum_{v:\mathbf{x}^v \in I_i} K_{\mathbf{h}}(\mathbf{x}^v - \mathbf{x}^r) f^v}{\sum_{v:\mathbf{x}^v \in I_i} f^v} \right). \tag{2.85}$$

Note that this numerical integration method corresponds to a bivariate trapezoidal quadrature rule (ignoring correction at interval endpoints).

A generalization of the convergence theorem of Braun, Duchesne and Stafford (2005, p.51-52) to the bivariate case is given by the following.

If we denote G_2 as the mapping with r^{th} component as given by the right-hand side of (2.85), Theorem 1 may be used to prove convergence of the fixed point iteration (2.3) for fine enough grids and large enough bandwidths h_1, h_2. We consider the case $h = h_1 = h_2$.

Theorem 2 *Assume that $K(\mathbf{u})$ is a continuous probability density function with compact support in $\mathbb{R}^2$, and such that $K(\mathbf{u})$ is of the form $K_1(u_1)K_2(u_2)$, where $K_1(u_1)$ and $K_2(u_2)$ are non-negative and symmetric about 0. Let δ be a small positive number and let*

$$D_h^2 = \left\{ (f^1, \ldots, f^m) : 0 \leq f^r \leq \sup_{\mathbf{u}} K_{\mathbf{h}}(\mathbf{u}) = K_{\mathbf{h}}(\mathbf{0}) \, and \sum_{r:\mathbf{x}^r \in \mathbf{I}_i} f^r \geq K_{\mathbf{h}}(\mathbf{0})(1+\delta) \, for \, i = 1, \ldots, n \right\}.$$

There exists a combination of mesh size m and bandwidth h (depending on $\mathbf{I}_1, \cdots, \mathbf{I}_n$) such that $G_2(\mathbf{f})$ has a unique fixed point $\mathbf{f}_{FP2}$ in D_h^2, and for any $\mathbf{f}_0 \in D_h^2$, the corresponding fixed point iteration converges to $\mathbf{f}_{FP2}$.

The proof is analogous to that in Braun, Duchesne and Stafford (2005) for the univariate case.

Proof of Theorem 2. For any fixed $h > 0$, it can be shown that D_h^2 is a closed and convex subset of $\mathbb{R}^m$. We now proceed to show that the image of D_h^2 under the continuous mapping given by G_2 lies in D_h^2. Suppose that $\mathbf{f} \in D_h^2$, the nonnegativity of $G_2^r(\mathbf{f})$ for $r = 1, \ldots, m$ is due to the nonnegativity of the kernel and of $\mathbf{f}$. Also

$$G_2^r(\mathbf{f}) \leq \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\sum_{v:\mathbf{x}^v \in \mathbf{I}_i} \sup_{\mathbf{u}} K_{\mathbf{h}}(\mathbf{u}) f^v}{\sum_{v:\mathbf{x}^v \in \mathbf{I}_i} f^v} \right) \leq \sup_{\mathbf{u}} K_{\mathbf{h}}(\mathbf{u}) = K_{\mathbf{h}}(\mathbf{0}).$$

Moreover, it can be noted that for a grid sufficiently fine (depending on h and $\mathbf{I}_1, \cdots, \mathbf{I}_n$) that

$$\sum_{r:\mathbf{x}^r \in \mathbf{I}_i} K_{\mathbf{h}}(\mathbf{x}^v - \mathbf{x}^r) \geq n K_{\mathbf{h}}(\mathbf{0})(1+\delta)$$

for each $\mathbf{x}^v \in \mathbf{I}_i$, $i = 1, \cdots, n$, and thus we have

$$\sum_{r:\mathbf{x}^r \in \mathbf{I}_i} G_2^r(\mathbf{f}) \geq \frac{1}{n} \frac{\sum_{v:\mathbf{x}^v \in \mathbf{I}_i} f^v \sum_{r:\mathbf{x}^r \in \mathbf{I}_i} K_{\mathbf{h}}(\mathbf{x}^v - \mathbf{x}^r)}{\sum_{v:\mathbf{x}^v \in \mathbf{I}_i} f^v} \geq \frac{\sum_{v:\mathbf{x}^v \in \mathbf{I}_i} f^v K_{\mathbf{h}}(\mathbf{0})(1+\delta)}{\sum_{v:\mathbf{x}^v \in \mathbf{I}_i} f^v} = K_{\mathbf{h}}(\mathbf{0})(1+\delta).$$

Up to this point the existence of a fixed point has been guaranteed. To prove the convergence of (2.3) to a unique fixed point, the Jacobian condition is checked in the following manner.

Differentiating with respect to $\mathbf{f}$ at points in the mesh gives

$$\frac{\partial G_2^r(\mathbf{f})}{\partial f^j} = \frac{1}{n}\sum_{i=1}^{n} 1(\mathbf{x}^j \in \mathbf{I}_i)\left\{\frac{\sum_{v:\mathbf{x}^v\in\mathbf{I}_i} f^v\left[K_{\mathbf{h}}(\mathbf{x}^j - \mathbf{x}^r) - K_{\mathbf{h}}(\mathbf{x}^v - \mathbf{x}^r)\right]}{(\sum_{v:\mathbf{x}^v\in\mathbf{I}_i} f^v)^2}\right\}.$$

It follows that

$$\frac{\partial G_2^r(\mathbf{f})}{\partial f^j} \leq \frac{1}{n}\sum_{i=1}^{n} 1(\mathbf{x}^j \in \mathbf{I}_i)\left\{\frac{\sum_{v:\mathbf{x}^v\in\mathbf{I}_i} f^v K_{\mathbf{h}}(\mathbf{0})}{(\sum_{v:\mathbf{x}^v\in\mathbf{I}_i} f^v)^2}\right\} \quad (2.86)$$

$$\leq \frac{1}{1+\delta} \quad (2.87)$$

It can then be shown that $\alpha = \parallel G_2'(\mathbf{f}) \parallel < 1$, for all $\mathbf{f} \in D_h^2$.

As the bandwidth $h \to 0$, the mesh must be correspondingly finer and $\{f^r\}$ must stay within D_h^2, for the argument to remain valid.

Recall that $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2$ is a bivariate mesh consisting of points $\{\mathbf{x}^r = (x_1, x_2)^r\}_{r=1}^m$, with $m = m_1 \times m_2$, and $f^r = f(\mathbf{x}^r)$ with $\mathbf{f} = (f^1, f^2, \ldots, f^m)^T$.

The unweighted version of our method implements a bivariate uniform sampling (U sampling) scheme derived from the orthogonal array-based latin hypercubes of Tang (1993). In this case the iteration equation

$$\hat{f}_j(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\sum_{b=1}^{B} K_{\mathbf{h}}(\mathbf{X}_{i,b}^u - \mathbf{x})\hat{f}_{j-1}(\mathbf{X}_{i,b}^u)}{\sum_{k=1}^{B} \hat{f}_{j-1}(\mathbf{X}_{i,k}^u)}\right)$$

is intended to solve the fixed point equation

$$f^r = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\sum_{b_i=1}^{B_i} K_{\mathbf{h}}(\mathbf{x}^{b_i} - \mathbf{x}^r)f^{b_i}}{\sum_{b_i=1}^{B_i} f^{b_i}}\right), \quad (2.88)$$

for $r = 1, \ldots, m$, where B_i is the number of points $\mathbf{x}^{b_i}$ of the uniform sample over the interval $\mathbf{I}_i$. We assume for simplicity that the points of the uniform sample are also points of the bivariate mesh.

If we denote by G_3 the mapping with r^{th} component as given by the right-hand side of (2.88), Theorem 1 may be used to prove convergence of the fixed point iteration (2.3).

Theorem 3 *Assume that $K(\mathbf{u})$ is a continuous probability density function with compact support in $\mathbb{R}^2$, and such that $K(\mathbf{u})$ is of the form $K_1(u_1)K_2(u_2)$, where $K_1(u_1)$ and $K_2(u_2)$ are non-negative and symmetric about 0. Let δ be a small positive number and let*

$$D_h^3 = \left\{ (f^1, \ldots, f^m) : 0 \le f^r \le \sup_{\mathbf{u}} K_{\mathbf{h}}(\mathbf{u}) = K_{\mathbf{h}}(\mathbf{0}), \sum_{\mathbf{x}^{b_i} \in \mathbf{I}_i} f^{b_i} \ge K_{\mathbf{h}}(\mathbf{0})(1+\delta) \ for \ i = 1, \ldots, n \right\}.$$

There exists a combination of number of points B_i, $i = 1, \cdots, n$, and bandwidth h (depending on $\mathbf{I}_1, \cdots, \mathbf{I}_n$) such that $G_3(\mathbf{f})$ has a unique fixed point $\mathbf{f}_{FP3}$ in D_h^3, and for any $\mathbf{f}_0 \in D_h^3$, the corresponding fixed point iteration converges to $\mathbf{f}_{FP3}$. Here m is the size of the mesh.

The proof is similar to the proof of Theorem 2.

Proof of Theorem 3. For any fixed $h > 0$, it can be shown that D_h^3 is a closed and convex subset of $\mathbb{R}^m$. We now proceed to show that the image of D_h^3 under the continuous mapping given by G_3 lies in D_h^3. Suppose that $\mathbf{f} \in D_h^3$. The nonnegativity of $G_3^r(\mathbf{f})$ for $r = 1, \ldots, m$ is due to the nonnegativity of the kernel and of $\mathbf{f}$. Also

$$G_3^r(\mathbf{f}) \le \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\sum_{b_i=1}^{B_i} \sup_{\mathbf{u}} K_{\mathbf{h}}(\mathbf{u}) f^{b_i}}{\sum_{b_i=1}^{B_i} f^{b_i}} \right) \le \sup_{\mathbf{u}} K_{\mathbf{h}}(\mathbf{u}) = K_{\mathbf{h}}(\mathbf{0}).$$

Moreover, it can be noted that for a sufficiently large number B_i (depending on h and $\mathbf{I}_1, \cdots, \mathbf{I}_n$) that

$$\sum_{b_i=1}^{B_i} K_{\mathbf{h}}(\mathbf{x}^{b_i} - \mathbf{x}^r) \ge n K_{\mathbf{h}}(\mathbf{0})(1+\delta),$$

for each $\mathbf{x}^{b_i} \in \mathbf{I}_i$, $i = 1, \ldots, n$, and thus we have

$$\sum_{r:\mathbf{x}^r \in \mathbf{I}_i} G_3^r(\mathbf{f}) \ge \frac{1}{n} \frac{\sum_{b_i=1}^{B_i} f^{b_i} \sum_{r:\mathbf{x}^r \in \mathbf{I}_i} K_{\mathbf{h}}(\mathbf{x}^{b_i} - \mathbf{x}^r)}{\sum_{b_i=1}^{B_i} f^{b_i}} \ge \frac{\sum_{b_i=1}^{B_i} f^{b_i} K_{\mathbf{h}}(\mathbf{0})}{\sum_{b_i=1}^{B_i} f^{b_i}} = K_{\mathbf{h}}(\mathbf{0})$$

Therefore, (2.88) has a fixed point in D_h^3. To prove the convergence of (2.88) to a unique fixed point, we consider the following.

Differentiating with respect to $\mathbf{f}$ at points of the importance sample gives

$$\frac{\partial G_3^r(\mathbf{f})}{\partial f^b} = \frac{1}{n}\sum_{i=1}^{n} 1(\mathbf{x}^b \in \mathbf{I}_i)\left\{\frac{\sum_{b_i=1}^{B_i} f^{b_i}\left[K_{\mathbf{h}}(\mathbf{x}^b - \mathbf{x}^r) - K_{\mathbf{h}}(\mathbf{x}^{b_i} - \mathbf{x}^r)\right]}{(\sum_{b_i=1}^{B_i} f^{b_i})^2}\right\}.$$

It follows that

$$\begin{aligned}
\frac{\partial G_3^r(\mathbf{f})}{\partial f^b} &\leq \frac{1}{n}\sum_{i=1}^{n} 1(\mathbf{x}^b \in \mathbf{I}_i)\left\{\frac{\sum_{b_i=1}^{B_i} f^{b_i} K_{\mathbf{h}}(\mathbf{0})}{(\sum_{b_i=1}^{B_i} f^{b_i})^2}\right\} \\
&\leq \frac{1}{n}\sum_{i=1}^{n} 1(\mathbf{x}^b \in \mathbf{I}_i)\frac{K_{\mathbf{h}}(\mathbf{0})}{\sum_{b_i=1}^{B_i} f^{b_i}} \\
&\leq \frac{1}{(1+\delta)}.
\end{aligned}$$

It can then be shown that $\alpha' = \| G_3'(\mathbf{f}) \| < 1$, for all $\mathbf{f} \in D_h^3$.

As the bandwidth $h \to 0$, the number of uniform sampling points B_i must increase and $\{f^r\}$ must stay within D_h^3, for the argument to remain valid.

Incorporation of the survey weights in both equations (2.85) and (2.88) is implemented in following manner

$$f^{r,w} = \sum_{i=1}^{n}\left(\frac{\sum_{v:\mathbf{x}^v \in \mathbf{I}_i} K_{\mathbf{h}}(\mathbf{x}^v - \mathbf{x}^r) f^{v,w}}{\sum_{v:\mathbf{x}^v \in \mathbf{I}_i} f^{v,w}}\right) w_i^* \tag{2.89}$$

and

$$f^{r,w} = \sum_{i=1}^{n}\left(\frac{\sum_{b_i=1}^{B_i} K_{\mathbf{h}}(\mathbf{x}^{b_i} - \mathbf{x}^r) f^{b_i,w}}{\sum_{b_i=1}^{B_i} f^{b_i,w}}\right) w_i^*, \tag{2.90}$$

for $r = 1, \ldots, m$, where $w_i^* = w_i/(\sum_{i\in S} w_i)$. Recall that B_i is the number of $\mathbf{x}^{b_i}$'s generated over the interval I_i using a bivariate U sampling.

Let us denote by $G_{w,1}$ the mapping with r^{th} component as given by the right-hand side of (2.89), and by $G_{w,2}$ the mapping with r^{th} component as given by the right-hand side of

(2.90).

Theorem 4 *Assume that $K(\mathbf{u})$ is a continuous probability density function with compact support in $\mathbb{R}^2$, and such that $K(\mathbf{u})$ is of the form $K_1(u_1)K_2(u_2)$, where $K_1(u_1)$ and $K_2(u_2)$ are non-negative and symmetric about 0. Let δ be a small positive number and let*

$$D_h^{w,1} = \left\{ (f^{1,w}, \ldots, f^{m,w}) : 0 \le f^{r,w} \le \sup_{\mathbf{u}} K_{\mathbf{h}}(\mathbf{u}) = K_{\mathbf{h}}(\mathbf{0})\, and \sum_{r:\mathbf{x}^r \in \mathbf{I}_i} f^{r,w} \ge K_{\mathbf{h}}(\mathbf{0})(1+\delta)\, for\, i = 1, \ldots, n \right\}.$$

There exists a combination of grid size m and bandwidth h (depending on $\mathbf{I}_1, \cdots, \mathbf{I}_n$) such that $G_{w,1}(\mathbf{f}^w)$ has a unique fixed point $\mathbf{f}^w_{FP1}$ in $D_h^{w,1}$, and for any $\mathbf{f}_0 \in D_h^{w,1}$, the corresponding fixed point iteration converges to $\mathbf{f}^w_{FP1}$.

Theorem 5 *Assume that $K(\mathbf{u})$ is a continuous probability density function with compact support in $\mathbb{R}^2$, and such that $K(\mathbf{u})$ is of the form $K_1(u_1)K_2(u_2)$, where $K_1(u_1)$ and $K_2(u_2)$ are non-negative and symmetric about 0. Let δ be a small positive number and let*

$$D_h^{w,2} = \left\{ (f^{1,w}, \ldots, f^{m,w}) : 0 \le f^{r,w} \le \sup_{\mathbf{u}} K_{\mathbf{h}}(\mathbf{u}) = K_{\mathbf{h}}(\mathbf{0}),\ \sum_{\mathbf{x}^{b_i} \in \mathbf{I}_i} f^{b_i,w} \ge K_{\mathbf{h}}(\mathbf{0})(1+\delta),\ i = 1, \ldots, n \right\}.$$

There exists a combination of number of points B_i, $i = 1, \cdots, n$, and bandwidth h (depending on $\mathbf{I}_1, \cdots, \mathbf{I}_n$) such that $G_{w,2}(\mathbf{f}^w)$ has a unique fixed point $\mathbf{f}^w_{FP2}$ in $D_h^{w,2}$, and for any $\mathbf{f}_0 \in D_h^{w,2}$, the corresponding fixed point iteration converges to $\mathbf{f}^w_{FP2}$. Proofs of theorems 4 and 5 are similar to those of theorems 2 and 3 and are omitted here.

2.8 Asymptotics for Sample Survey Data

In Chapter 1, Section 1.2, we presented the test statistic Q for close precursor. We need to justify the asymptotic distribution assumed for this test statistic. Therefore the aim of this section is to outline a corresponding theory for this purpose. Since Q is a functional of a bivariate density estimate, the focus will be on the asymptotic properties of the density estimate itself.

As it has been indicated in Section 1.4, the independent and identically distributed assumptions used for model (2.1) are generally not valid for complex survey data. Stratification can reflect a violation of the identically distributed assumption, while clustering can violate the independence assumption. Considering univariate non-censored data from stratified multistage samples, Buskirk and Lohr (2005) provided asymptotic properties of a kernel density estimator that incorporates the sampling weights. They presented regularity conditions which lead the sample estimator to be consistent and asymptotically normal under various modes of inference used with sample survey data.

The settings used by Buskirk and Lohr (2005) for the stratified two-stage sampling design are those specified in Section 1.4.2.

For design-based inference they used the following set-up from Isaki and Fuller (1982). Let $\{\mathcal{U}(t)\}$ be a sequence of nested finite populations where $\mathcal{U}(i)$ is a subset of $\mathcal{U}(i+1)$. Let $s(t)$ denote the corresponding sample from $\mathcal{U}(t)$. Population $\mathcal{U}(t)$ has $L(t)$ strata and a total of $N(t)$ primary sampling units (psu's) and $Q(t)$ secondary sampling units (ssu's); similarly the sample $s(t)$ contains a total of $n(t)$ psu's and $q(t)$ ssu's. They examined properties of the estimator as $t \to \infty$.

Accordingly, from Section 1.4.2 we have that the corresponding inclusion probabilities are $\pi_i^{(l)}(t) = P_D(\text{psu } i \text{ from stratum } l \text{ is included in the sample})$, with $\sum_{i=1}^{N_l} \pi_i^{(l)}(t) = n_l(t)$. Recall that the subscript D indicates the probability distribution induced by the design. The joint inclusion probabilities are $\pi_{ij}^{(l)}(t) = P_D(\text{psu's } i \text{ and } j \text{ from stratum } l \text{ are included in the sample})$. At the secondary sampling unit (ssu) level, (psu) i of stratum l has $Q_{li}(t)$ secondary sampling units (ssu's); $\pi_{m|i}^{(l)}(t)$ is the conditional probability that (ssu) m of (psu) i is included in the sample, given that (psu) i is included.

Buskirk and Lohr (2005) pointed out that if each unit in the finite population were observed, then a density estimator corresponding to the i.i.d. estimator in (2.1) would be

$$\hat{f}_{\mathcal{U}}(y;h) = \frac{1}{hQ}\sum_{l=1}^{L}\sum_{i=1}^{N_l}\sum_{k=1}^{Q_{li}} K\left(\frac{y - Y_{lik}}{h}\right). \tag{2.91}$$

The sample weighted kernel density estimator of Buskirk and Lohr (2005) is given by

$$\hat{f}_s(y;h) = \frac{1}{h\hat{Q}}\sum_{(lik)\in s} w_{lik} K\left(\frac{y - Y_{lik}}{h}\right), \tag{2.92}$$

where $\hat{Q} = \sum_{(lik)\in s} w_{lik}$. Note that Bellhouse and Stafford (1999) also used sampling weights to estimate $\hat{f}_{\mathcal{U}}(y;h)$.

Buskirk and Lohr (2005) used the following assumptions for the kernel function K:

(1.k) $K(u) \geq 0$ for all u and K is symmetric about zero.

(2.k) $\int K(u)du = 1$.

(3.k) $\int u^4 K(u)du < \infty$.

(4.k) There exists a constant m such that $K(u) \leq m$ for all u.

In stratified two-stage sampling, the conditions for consistency of the kernel density estimator given by Buskirk and Lohr (2005) are the following:

(1.c) $n_l(t) \geq 1$ for all l, where l indexes the strata.

(2.c) There exists a constant B such that $Q_{li}(t) < B$ for all l, i, and t.

(3.c) There exists a constant $\delta > 0$ such that $\delta < \pi_i^{(l)}(t)$ and $\delta < \pi_{k|i}^{(l)}(t)$ for all l, i, k, t.

(4.c) There exist sequences $\{\alpha_l(t)\}, l = 1, \ldots, L(t)$, such that

$$\pi_i^{(l)}(t)\pi_j^{(l)}(t) - \pi_{ij}^{(l)}(t) \leq \alpha_l(t)\pi_i^{(l)}(t)\pi_j^{(l)}(t)$$

and $\max_{1\leq l\leq L(t)} N_l(t)\alpha_l(t) = O(1)$.

(5.c) The bandwidth satisfies $N(t)h^2(t) \to \infty$ and $h(t) \to 0$ as $t \to \infty$.

Buskirk and Lohr (2005) first showed that $\hat{f}_s(y;h)$ is approximately unbiased for $\hat{f}_{\mathcal{U}}(y;h)$ under the design. Then Buskirk and Lohr (2005, Theorem 2) proved design consistency of the sample estimator of the finite population estimate under their conditions (1.c) to (5.c) and (1.k) to (4.k). They did this by bounding the design mean square error (MSE) of the sample estimate by a constant divided by $N(t)h^2(t)$. We note the fact that $h(t) \to 0$ but in such a way that $N(t)h^2(t) \to \infty$ as $t \to \infty$.

We reproduce Buskirk and Lohr (2005, Theorem 2) as follows.

Theorem 6 *(Buskirk and Lohr; 2005, Theorem 2). Suppose that conditions* (1.c) *to* (5.c) *hold in stratified two-stage sampling and that the kernel function satisfies* (1.k) *to* (4.k). *Then* $V_D(\hat{f}_s(y;h)) \to 0$ *as* $t \to \infty$*, uniformly in* y.

Here $V_D(\hat{f}_s(y;h))$ refers to the design based variance of $\hat{f}_s(y;h)$.

Furthermore, Buskirk and Lohr (2005, Theorem 6) proved design based asymptotic normality of the sample estimate as an estimator of the finite population estimate under basically the same previous conditions except that (3.c) and (4.c) were replaced by assumptions that guaranteed asymptotic normality of a weighted sample sum (see theorem below). They used the following set-up for sampling with replacement. As before, there are $L(t)$ strata. From stratum l, $n_l(t)$ psu's are sampled with replacement; on draw j (for $j = 1, \ldots, n_l(t)$), psu (li) is sampled with probability p_{li}, where $\sum_{i=1}^{N_l(t)} p_{li} = 1$. They defined the random variable Z_{lji} to be 1 if psu i is selected on draw j and 0 otherwise. Since sampling is done with replacement, Z_{lji} and $Z_{l'j'i'}$ are independent when $(lj) \neq (l'j')$.

Then, they let

$$Z_{lj}(y,h) = \sum_{i=1}^{N_l} Z_{lji} \frac{\sum_{k \in s_{li}} K_h(y - Y_{lik})}{n_l p_{li}}. \tag{2.93}$$

Their Z_{lj}'s are independent with

$$E_D[Z_{lj}(y,h)] = \frac{1}{n_l}\sum_{i=1}^{N_l}\sum_{k=1}^{Q_{li}} K_h\left(y - Y_{lik}\right) \tag{2.94}$$

Then, they defined $Z(y,h,t) = \sum_{l=1}^{L(t)}\sum_{j=1}^{n_l(t)} Z_{lj}(y,h)$, and let the sample estimator of the density be

$$\hat{f}_R(y;h) = \frac{1}{\hat{Q}} Z(y,h,t) \tag{2.95}$$

for $\hat{Q}$ a design consistent estimator of Q; then $\hat{f}_R(y;h)$ is approximately design unbiased for $\hat{f}_{\mathcal{U}}(y;h)$ since $E_D[Q^{-1}Z(y,h,t)] = \hat{f}_{\mathcal{U}}(y;h)$. Also they defined $\sigma^2(t) = V_D[Z(y,h,t)]$. Their theorem for asymptotic normality in this context is presented next.

Theorem 7 *(Buskirk and Lohr; 2005, Theorem 6). Assume conditions* $(1.k)$ *to* $(4.k)$, $(1.c)$, $(2.c)$ *and* $(5.c)$ *hold. Suppose there exists a constant* δ *such that* $\delta < N_l(t)p_{li}$ *and* $\delta < \pi_{k|i}^{(l)}(t)$ *for all* l, i *and* k. *Further suppose that* $\max_l N_l(t)/n_l(t)$ *is bounded and that* $\lim_{t\to\infty} h(t)\sigma(t) = \infty$. *Then, the distribution of*

$$\frac{\left(Z(y,h,t) - Q\hat{f}_{\mathcal{U}}(y;h)\right)}{\sigma(t)},$$

conditional on $\mathbf{Y}_t = (Y_{111}, \ldots, Y_{L(t),N_L(t),Q_{L(t),N_L(t)}})$, *approaches* $N(0,1)$ *as* $t \to \infty$.

In the proofs of both previous theorems, the bandwidth may be considered the same for the sample estimate and the population estimate. Moreover, their conditions can be satisfied if $h(t)$ remains fixed (i.e. $h(t) \to 0$ may not be regarded as necessary), as long as $N(t) \to \infty$ and the sample size $n(t) \to \infty$ in such a way that the assumption of $\max_l N_l(t)/n_l(t)$ being bounded holds.

For model based inference (i.e. assuming that $Y_{111}, \ldots, Y_{L(t),N_L(t),Q_{L(t),N_L(t)}}$ are distributed according to some joint probability distribution and that y_{lik} is a realization of Y_{lik} that gives the measurement in the t^{th} finite population), and model-design inference, Buskirk and Lohr

(2005) used conditions (1.r) to (5.r) in the survey sample setting. The corresponding regularity conditions for f (the density of the superpopulation) are the following.

(1.r) For any ssu labels k and j, and psu label i within stratum l, (Y_{lik}, Y_{lij}) have joint density g_l, where g_l satisfies $\int g_l(x,u)du = \int g_l(u,x)du = f_l(x)$. The marginal density of Y_{lik}, f_l, has continuous second derivative f_l'' that is square integrable and monotone in both $(-\infty, -M)$ and (M, ∞) for some M. The variables Y_{lik} and Y_{rpj} are independent if $(li) \neq (rp)$.

(2.r) $f(x) = \sum_{l=1}^{L} W_l f_l(x)$.

(3.r) $\sup_x \max_{1 \leq l \leq L(t)} f_l(x) = G(t) = O(1)$.

(4.r) $\sup_x \max_{1 \leq l \leq L(t)} \mid f_l''(x) \mid = D(t) = O(h^{-1}(t))$.

(5.r) $N(t)h(t) \to \infty$ and $h(t) \to 0$ as $t \to \infty$.

Note that (5.r) is implied by condition (5.c).

In Buskirk and Lohr (2005, Theorem 4), consistency in the combined model-design sense is shown. In this theorem they did not require condition (5.c), but the condition in (5.r) that the bandwidth $h(t) \to 0$ was crucial. Explicitly, that theorem stated the following.

Theorem 8 *(Buskirk and Lohr; 2005, Theorem 4). Assume that conditions* (1.c) *to* (4.c) *and* (1.r) *to* (5.r) *hold and that the kernel function satisfies* (1.k) *to* (4.k). *Then*

$$E_M \left\{ E_D[\hat{f}_s(y; h(t)) - f(y)]^2 | \mathbf{Y}_t \right\} \to 0$$

as $t \to \infty$.

Buskirk and Lohr (2005, Theorem 7), referring to the paper of Bleuer and Kratina (2000), established asymptotic normality of the finite population estimate and the sample estimate as an estimator of the model expectation of the finite population estimator. In this theorem they replaced conditions $(3.c)$ and $(4.c)$ by a further assumption on the superpopulation joint distributions of $\mathbf{Y}$. In their proof we noted that the bandwidth did not have to go to 0.

Having all these previous results, under further regularity conditions it is reasonable to indicate that a test statistic which is a smooth functional of the density estimator is also consistent and asymptotically normal. See Geskus and Groeneboom (1997) and Geskus and Groeneboom (1999) for some examples.

Generalization of these results to the bivariate kernel density estimation case is straightforward. In the case of local density estimation, the theory that generalizes these results is also straightforward since the idea involves replacing a single estimating equation with a system of two or three similar estimating equations.

Consequently, once this apparatus is available we may conjecture that the test statistic Q presented in Section 1.2, which is a smooth functional of the density estimator, is also consistent and asymptotically normal.

However, when interval censoring is introduced, we need to add conditions to ensure the following:

- Convergence of both the finite population and sample algorithms.
- Convergence of the finite population density estimate $\hat{f}_{\mathcal{U}}(\mathbf{x})$ to the true density f (i.e. the density of the superpopulation).
- Asymptotic normality under the model assumptions in $(1.r) - (5.r)$ of the finite population density estimate $\hat{f}_{\mathcal{U}}(\mathbf{x})$.

These are conditions on:

- The superpopulation density f.
- The superpopulation structure.
- The density of the endpoints of the censoring intervals.

- The kernel function.
- The mesh $\mathcal{M}$ (and the number of U sampled points).
- The bandwidth, so that $h_i(t) \to 0$ as $t \to \infty, i = 1, 2$.

We assume that these conditions may be satisfied, so that if we could establish design-based consistency and normality of the sample estimator $\hat{f}_s(\mathbf{x})$ as an estimate of the finite population estimator $\hat{f}_{\mathcal{U}}(\mathbf{x})$, the apparatus for sample-based inference about $\hat{f}_{\mathcal{U}}(\mathbf{x})$ and f would be complete.

Now, let us consider the finite population estimator $\hat{f}_{\mathcal{U}}(\mathbf{x})$. It satisfies the self-consistency equation in Section 2.6, and it is obtained by iterations (see sections 2.1 and 2.2). In addition, let us assume the following.

- The density f of the superpopulation satisfies conditions $(1.r) - (5.r)$.
- The conditions of Buskirk and Lohr (2005) on the superpopulation hold.
- The interval endpoints become dense as $t \to \infty$.

For the kernel function K we assume the following.

$(1.k')$ K is of the form $K_1(u_1)K_2(u_2)$, where $K_i(u_i)$ is a non-negative probability density function and symmetric about 0, $i = 1, 2$.

$(2.k')$ $\int u_i^4 K_i(u_i) du_i < \infty$ for $i = 1, 2$.

$(3.k')$ There exists a constant m such that $K(\mathbf{u}) \leq m$ for all $\mathbf{u}$.

We may conjecture that if the above conditions hold there will be an analogue of the first statements in Buskirk and Lohr (2005, Theorems 3 and 7) which do not involve the sampling design. The analogue we are referring to is as follows.

Theorem 9 *Assume conditions (1.k')-(3.k'), (1.r)-(5.r), (1.c) and (2.c) hold. Then*

$$MSE_M[\hat{f}_{\mathcal{U}}(\mathbf{x})] \to 0$$

as $t \to \infty$*. Furthermore,*

$$\frac{\hat{f}_{\mathcal{U}}(\mathbf{x}) - E_M[\hat{f}_{\mathcal{U}}(\mathbf{x})]}{\sqrt{V_M[\hat{f}_{\mathcal{U}}(\mathbf{x})]}} \to_{\mathcal{D}} N(0,1).$$

as $t \to \infty$*.*

We will assume these conclusions as background conditions.

Recall from Sections 2.1.2 and 2.1.3 that at the population level

$$\begin{aligned}
\hat{f}_{j,\mathcal{U}}(\mathbf{x}) &= \frac{1}{Q}\sum_{i=1}^{Q} E_{\hat{f}_{j-1,\mathcal{U}}}\left[K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{x})\middle|\mathbf{X}_i \in \mathbf{I}_i\right] \\
&= \frac{1}{Q}\sum_{i=1}^{Q} E_{g_{j-1}}\left[K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{x})w(\mathbf{X}_i)\right]
\end{aligned}$$

where g is a suitable distribution over the interval $\mathbf{I}_i$, and $w(\mathbf{X}) = \hat{f}_{j-1,\mathcal{U}|\mathbf{I}}(\mathbf{X})/g(\mathbf{X})$ is the importance sampling weight. Thus, $\hat{f}_{j,\mathcal{U}}(\mathbf{x})$ may be approximated by

$$\frac{1}{Q}\sum_{i=1}^{Q}\left(\frac{\sum_{b=1}^{B} K_{\mathbf{h}}(\mathbf{X}_{i,b}^u - \mathbf{x})w(\mathbf{X}_{i,b}^u)}{\sum_{k=1}^{B} w(\mathbf{X}_{i,k}^u)}\right) \tag{2.96}$$

$$= \frac{1}{Q}\sum_{i=1}^{Q}\left(\frac{\sum_{b=1}^{B} K_{\mathbf{h}}(\mathbf{X}_{i,b}^u - \mathbf{x})\hat{f}_{j-1,\mathcal{U}}(\mathbf{X}_{i,b}^u)}{\sum_{k=1}^{B} \hat{f}_{j-1,\mathcal{U}}(\mathbf{X}_{i,k}^u)}\right) \tag{2.97}$$

Here we let $g(\mathbf{X})$ be a bivariate uniform distribution density and therefore $\mathbf{X}_{i,b}^u$ is generated over the interval $\mathbf{I}_i$ using a bivariate uniform sampling scheme derived from the orthogonal array-based Latin hypercubes described by Tang (1993) for $b = 1, \ldots, B$.

Consider the survey weights to be the inverse inclusion probability weights w_i, $i \in s$.

Incorporation of these weights yields a corresponding sample iteration

$$\hat{f}_j^{w^s}(\mathbf{x}) = \sum_{i \in s} \left(\frac{\sum_{b=1}^{B} K_{\mathbf{h}}(\mathbf{X}_{i,b}^{u} - \mathbf{x}) \hat{f}_{j-1}^{w^s}(\mathbf{X}_{i,b}^{u})}{\sum_{k=1}^{B} \hat{f}_{j-1}^{w^s}(\mathbf{X}_{i,k}^{u})} \right) \frac{w_i}{\sum_{l \in s} w_l}.$$

Alternatively, this may be expressed as

$$\hat{f}_{j,s}(\mathbf{x}) = \sum_{i \in s} \left(\frac{\sum_{b=1}^{B} K_{\mathbf{h}}(\mathbf{X}_{i,b}^{u} - \mathbf{x}) \hat{f}_{j-1,s}(\mathbf{X}_{i,b}^{u})}{\sum_{k=1}^{B} \hat{f}_{j-1,s}(\mathbf{X}_{i,k}^{u})} \right) \frac{w_i}{\sum_{l \in s} w_l}. \tag{2.98}$$

Thus,

$$\hat{f}_{0,s}(\mathbf{x}) = \sum_{i \in s} \left(\frac{1}{B} \sum_{b=1}^{B} K_{\mathbf{h}}(\mathbf{X}_{i,b}^{u_0} - \mathbf{x}) \right) \frac{w_i}{\sum_{l \in s} w_l}. \tag{2.99}$$

Taking a sufficiently fine grid we may obtain $\hat{f}_{0,s}(\mathbf{X}_{i,b}^{u_1})$, for $b = 1, \ldots, B$. Then,

$$\hat{f}_{1,s}(\mathbf{x}) = \sum_{i \in s} \left(\frac{\sum_{b=1}^{B} K_{\mathbf{h}}(\mathbf{X}_{i,b}^{u_1} - \mathbf{x}) \hat{f}_{0,s}(\mathbf{X}_{i,b}^{u_1})}{\sum_{k=1}^{B} \hat{f}_{0,s}(\mathbf{X}_{i,k}^{u_1})} \right) \frac{w_i}{\sum_{l \in s} w_l}. \tag{2.100}$$

Once again we take a sufficiently fine grid to obtain $\hat{f}_{1,s}(\mathbf{X}_{i,b}^{u_1})$ for $b = 1, \ldots, B$, so we have

$$\hat{f}_{2,s}(\mathbf{x}) = \sum_{i \in s} \left(\frac{\sum_{b=1}^{B} K_{\mathbf{h}}(\mathbf{X}_{i,b}^{u_2} - \mathbf{x}) \hat{f}_{1,s}(\mathbf{X}_{i,b}^{u_2})}{\sum_{k=1}^{B} \hat{f}_{1,s}(\mathbf{X}_{i,k}^{u_2})} \right) \frac{w_i}{\sum_{l \in s} w_l}.$$

In order to consider mean square error consistency under the probability distribution induced by the sampling design, we then need to show that the variance of the estimator under the sampling design goes to 0.

Therefore, in the same manner as Buskirk and Lohr (2005), suppose that conditions (1.c) to (4.c) hold in stratified two-stage sampling and that the kernel function satisfies (1.k') to (3.k'). Suppose also the following condition:

(5.c') The bandwidth $h_i(t)$ satisfies $N(t)h_i^2(t) \to \infty$ and $h_i(t) \to 0$ as $t \to \infty$, for $i = 1, 2$. For convenience, let us take $h_1(t) \equiv h_2(t) \equiv h(t)$.

Let us note that under the assumptions (2.c) and (3.c) of Buskirk and Lohr (2005), saying that $A = O\left((Q(t)h^2(t))^{-1/2}\right)$ is nearly the same as saying that $A = O\left((N(t)h^2(t))^{-1/2}\right)$. We may thus define $R(t) = (Q(t)h^2(t))^{-1/2}$.

Let us note that in (2.98) the characteristic of interest (for $j \geq 1$) for the i^{th} subject actually depends on the rest of the sample. Here, our consistency result refers then to convergence in probability.

At the sample level, equation (2.98) for $j = 0$ yields equation (2.99). Replacing weighted sums in (2.99) by population sums yields the corresponding equation at the population level given by

$$\hat{f}_{0,\mathcal{U}}(\mathbf{x}) = \sum_{i \in \mathcal{U}(t)} \left(\frac{1}{B} \sum_{b=1}^{B} K_{\mathbf{h}}(\mathbf{X}_{i,b}^{u_0} - \mathbf{x}) \right) \frac{1}{\sum_{i \in \mathcal{U}(t)} 1}. \tag{2.101}$$

Assuming $(4.k)$ (note this becomes $(3.k')$ in our case), $(2.c), (3.c)$ and $(4.c)$, Buskirk and Lohr (2005) showed that the sample estimator $\hat{f}_{0,s}(\mathbf{x})$ is approximately unbiased for its population counterpart $\hat{f}_{0,\mathcal{U}}(\mathbf{x})$, and that the design variance of the sample estimator $V_D(\hat{f}_{0,s}(\mathbf{x}))$ is bounded by $c_0 R^2(t)$, uniformly in $\mathbf{x}$, with $c_0 \in \mathbb{R}$.

Therefore the maximum over any finite set of points $\{\mathbf{x}\}$ of the difference between the sample estimate $\hat{f}_{0,s}(\mathbf{x})$ and the population estimate $\hat{f}_{0,\mathcal{U}}(\mathbf{x})$ is $O_p(R(t))$.

Let us consider the corresponding finite set of points $\{\mathbf{x}\}$ to be a mesh $\mathcal{M}$ which also includes the points of the uniform sample, as in the convergence argument previous to Theorem 3 in Section 2.7.

Now let us consider equation (2.100), that is the sample estimate in (2.98) at the first iteration (i.e. $j = 1$), and also consider the corresponding population equation (2.97) at the first iteration. Alternatively, the above referred equations may be respectively expressed as follows.

$$\hat{f}_{1,s}(\mathbf{x}) = \frac{\sum_s w_i Y_i}{\sum_s w_l} \tag{2.102}$$

and

$$\hat{f}_{1,\mathcal{U}}(\mathbf{x}) = \frac{\sum_{\mathcal{U}} Z_i}{\sum_{\mathcal{U}} 1}, \tag{2.103}$$

where

$$Y_i = \frac{\sum_{b=1}^{B} K_{\mathbf{h}}(\mathbf{X}_{i,b}^{u_1} - \mathbf{x})\hat{f}_{0,s}(\mathbf{X}_{i,b}^{u_1})}{\sum_{k=1}^{B} \hat{f}_{0,s}(\mathbf{X}_{i,k}^{u_1})}$$

and

$$Z_i = \frac{\sum_{b=1}^{B} K_{\mathbf{h}}(\mathbf{X}_{i,b}^{u_1} - \mathbf{x})\hat{f}_{0,\mathcal{U}}(\mathbf{X}_{i,b}^{u_1})}{\sum_{k=1}^{B} \hat{f}_{0,\mathcal{U}}(\mathbf{X}_{i,k}^{u_1})}.$$

We want to bound the difference between $\hat{f}_{1,s}(\mathbf{x})$ and $\hat{f}_{1,\mathcal{U}}(\mathbf{x})$. First, we break up the difference between the right hand side in (2.102) and the right hand side in (2.103) into the following two parts:

$$\frac{\sum_s w_i(Y_i - Z_i)}{\sum_s w_l} \tag{2.104}$$

$$\frac{\sum_s w_i Z_i}{\sum_s w_l} - \frac{\sum_{\mathcal{U}} Z_i}{\sum_{\mathcal{U}} 1}. \tag{2.105}$$

Since the Z_i's come from the algorithm at the population level, they are not dependent on the sample, and they have the same bound as on the kernel. Therefore, the proof of Buskirk and Lohr (2005, Theorem 2) again applies directly to show that the design variance of (2.105) is bounded by $cR^2(t)$, with $c \in \mathbb{R}$.

Now, we may say that the difference between Y_i and Z_i must be bounded by the maximum difference between the two densities normalized times the integral of the kernel over the interval I_i, that is

$$\max_b \left(\frac{\hat{f}_{0,s}(\mathbf{X}_{i,b}^{u_1})}{\sum_{k=1}^{B} \hat{f}_{0,s}(\mathbf{X}_{i,k}^{u_1})} - \frac{\hat{f}_{0,\mathcal{U}}(\mathbf{X}_{i,b}^{u_1})}{\sum_{k=1}^{B} \hat{f}_{0,\mathcal{U}}(\mathbf{X}_{i,k}^{u_1})} \right) \int_{\mathbf{I}_i} K_{\mathbf{h}}(\mathbf{y} - \mathbf{x})\mathbf{dy}$$

with $\int_{\mathbf{I}_i} K_{\mathbf{h}}(\mathbf{y} - \mathbf{x})\mathbf{dy} < 1$. So essentially we have a bound on the magnitude of a weighted sample ratio or mean where the maximum of the characteristic $(Y_i - Z_i)$ of the numerator in (2.104) is bounded in probability by $c'R(t)$, with $c' \in \mathbb{R}$. The same bound $c'R(t)$ would apply to the whole sample average (2.104).

Thus we can conclude that the difference between $\hat{f}_{1,s}(\mathbf{x})$ and $\hat{f}_{1,\mathcal{U}}(\mathbf{x})$ is bounded in probability by $c_1 R(t)$, uniformly in x, with $c_1 \in \mathbb{R}$.

Proceeding by induction, we can say the same after any finite number of iterations J. The constant c_j will tend to increase with each iteration j. We realize that the convergence of the algorithm would be governed by the evolution of the population and the intervals as well as the bandwidth. Therefore we want to re-emphasize that completion of the arguments to prove consistency would have to consider these elements.

Let us note that in practice, between 4 and 7 iterations are required. This is the case considering settings similar to those employed in Sections 4.4, 4.8.1 and 4.8.2. When we use a finer mesh and increase the number of uniform sampled points within the intervals, the number of required iterations decreases significantly (less than 3 for example).

For the asymptotic normality argument we note that the asymptotic normality of (2.105), normalized, is a consequence of the argument of Buskirk and Lohr (2005, Theorem 6). The asymptotic normality of

$$\frac{\sum_s w_i(Y_i - Z_i)}{\sum_s w_l} + \frac{\sum_s w_i Z_i}{\sum_s w_l} - \frac{\sum_{\mathcal{U}} Z_i}{\sum_{\mathcal{U}} 1}$$

normalized would be clear if the Y_i's in the first term were not dependent on the rest of the sample s. Since it also comes from a ratio, a way to proceed is clear in principle, because we may iterate the linearization of (2.104) and, as in the convergence argument, this could be carried out up to a fixed number of iterations J (which in practice is a small number).

Chapter 3

Semiparametric Models

The present chapter introduces *triggering* models from Thompson and Pantoja-Galicia (2003). A distinction between long term and short term *triggering* is presented in Section 3.1. In Section 3.2 we derive likelihood functions that can be used to estimate the parameters of these models considering interval censored times. Subsequently, in Section 3.3 we mention the work on multi-state analysis of bivariate interval censored event times developed by Cook, Zeng and Lee (2007).

Here E_1 and E_2 also denote two lifetime events. T_1 denotes the time to occurrence of event E_1, and T_2 denotes the time to occurrence of event E_2 considering a specified time origin.

3.1 Triggering Models

Thompson and Pantoja-Galicia (2003) establish that the concept of *triggering* is a special case of temporal ordering, where a causal model is explicit. They provide the following conditional formulation for triggering.

3.1.1 Long term triggering

T_1 *triggers* T_2 if the occurrence of the end of duration T_1 increases the hazard function of T_2 immediately after T_1.

In a simple example, suppose that if T_1 were infinity (i.e. E_1 never occurred), T_2 would have survivor function $\mathcal{F}_{02}(t_1)$ and hazard function $\lambda_{02}(u)$. However, if $T_1 = t_1$ (i.e. E_1 does occur at time T_1), then T_2 has hazard function $\lambda_{02}(u)$ before t_1 and $e^{\beta}\lambda_{02}(u)$ after t_1, where $\beta > 0$. This kind is a long term triggering property.

3.1.2 Short term triggering

The "local" formulation in terms of a short term scale change in the hazard of T_2 is as follows.

In a local sense, event E_1 *triggers* event E_2 if it increases (for a time) the conditional intensity of E_2, given recent history.

For example, let $\mathcal{H}(s-)$ denote the history of the joint process before time s. Then E_1 triggers E_2 if for some $\kappa(s)$, we have

$$\lambda_2(u \mid \mathcal{H}(s-) \,\text{and}\, E_1 \,\text{at}\, s) = e^{\beta}\lambda_2(u \mid \mathcal{H}(s-)), \tag{3.1}$$

for $s < u < s + \kappa(s)$ $(\beta > 0)$.

Note that here triggering induces a multiplicative change in the hazard function, whereas the nonparametric expression of close precursor in Chapter 1 section 1.1 suggested an additive change.

Suppose $\lambda_1(t)$ and $\lambda_{02}(\tau)$ are respectively the hazard functions for T_1 and for T_2 when T_1 is ∞ (i.e. E_1 never occurs). Suppose that $\lambda_{12}(u \mid t; \beta)$ is the hazard function for T_2 at u, given $T_1 = t$, so that $\lambda_{12}(u \mid t; \beta)$ could be $e^{\beta}\lambda_{02}(u)$ for u between t and $t + \kappa(t)$, and $\lambda_{02}(u)$ for other u.

An example of this situation would be to have, for example, $\lambda_1(t) = \lambda_1$, $\lambda_{02}(u) = \lambda_{02}$ and

$\lambda_{12}(u) = e^{\beta}\lambda_{02}$ for $t \le u < t + \kappa(t)$ and λ_{02} for other u.

3.2 Likelihood Functions

In longitudinal surveys conducted at widely spaced time points, it is possible for the endpoint of T_1 or T_2 to be interval censored. Let us suppose that $T_1 \in (T_{10}, T_{11})$ and $T_2 \in (T_{20}, T_{21})$, and that the period between interviews is $[0, a_1]$ where a_1 is a general endpoint.

The basic idea of the model in Section 3.1.2 is that the occurrence of E_1 at time T_1 changes the hazard for T_2 for some time. If we have this simple model, we can deal with this situation simply by calculating the appropriate likelihood function and using it to estimate the corresponding parameters including β.

For simplicity, let us consider first the case when T_1 is not interval censored. For subject i, taking $t_{i,20}$ to be a_1 if T_2 is unobserved, the likelihood function L looks like the following:

$$L = \prod_i \mathcal{L}(t_{i,1}, t_{i,20}, t_{i,21}) \tag{3.2}$$

where $\mathcal{L}(t_1, t_{20}, t_{21}) =:$

$$\begin{aligned}
L_1 &= \mathcal{A}(a_1) \\
&\quad if \quad a_1 \le t_1, t_{20} \\
L_2 &= \mathcal{A}(t_1)\lambda_1(t_1)dt_1\mathcal{B}_{12}(a_1 \mid t_1; \beta) \\
&\quad if \quad 0 < t_1 < a_1 = t_{20} \\
L_3 &= \int_{t_{20}}^{t_{21}} \mathcal{A}(\tau)\lambda_{02}(\tau)\mathcal{B}_{01}(a_1 \mid \tau)d\tau \\
&\quad if \quad 0 \le t_{20} < t_{21} \le a_1 < t_1 \\
L_4 &= \int_{t_{20}}^{t_1} \mathcal{A}(\tau)\lambda_{02}(\tau)\mathcal{B}_{01}(t_1 \mid \tau)\lambda_1(t_1)d\tau dt_1 + \mathcal{A}(t_1)\lambda_1(t_1)dt_1\left[1 - \mathcal{B}_{12}(t_{21} \mid t_1; \beta)\right] \\
&\quad if \quad 0 \le t_{20} < t_1 < t_{21} \le a_1
\end{aligned}$$

$$
\begin{aligned}
L_5 &= \mathcal{A}(t_1)\lambda_1(t_1)dt_1[\mathcal{B}_{12}(t_{20} \mid t_1;\beta) - \mathcal{B}_{12}(t_{21} \mid t_1;\beta)] \\
& \quad if \quad 0 < t_1 < t_{20} < t_{21} \leq a_1 \\
L_6 &= \int_{t_{20}}^{t_{21}} \mathcal{A}(\tau)\lambda_{02}(\tau)\mathcal{B}_{01}(t_1 \mid \tau)\lambda_1(t_1)d\tau dt_1 \\
& \quad if \quad 0 < t_{20} < t_{21} \leq t_1 < a_1
\end{aligned}
$$

with

$$
\begin{aligned}
\mathcal{A}(t) &= exp\{-\int_0^t [\lambda_1(u) + \lambda_{02}(u)]du\}, \\
\mathcal{B}_{01}(\tau \mid t) &= exp\{-\int_t^\tau \lambda_1(u)du\}, \\
\mathcal{B}_{12}(\tau \mid t;\beta) &= exp\{-\int_t^\tau \lambda_{12}(u \mid t;\beta)du\}.
\end{aligned}
$$

Note that in complete data likelihood, L_4 is not needed, L_1 and L_2 remain the same (using the appropriate notation) and L_3, L_5 and L_6 become as indicated in the following segment. Therefore, in this case the likelihood function is equal to

$$
L = \prod_i \mathcal{L}(t_{i,1}, t_{i,2}) \tag{3.3}
$$

where $\mathcal{L}(t_1, t_2) =$:

$$
\begin{aligned}
L_1 &= \mathcal{A}(a_1) \\
& \quad if \quad a_1 \leq t_1, t_2 \\
L_2 &= \mathcal{A}(t_1)\lambda_1(t_1)dt_1\mathcal{B}_{12}(a_1 \mid t_1;\beta) \\
& \quad if \quad 0 < t_1 < a_1 \leq t_2 \\
L_3 &= \mathcal{A}(t_2)\lambda_{02}(t_2)dt_2\mathcal{B}_{01}(a_1 \mid t_2) \\
& \quad if \quad 0 < t_2 < a_1 \leq t_1
\end{aligned}
$$

$$
\begin{aligned}
L_5 &= \mathcal{A}(t_1)\lambda_1(t_1)dt_1\mathcal{B}_{12}(t_2 \mid t_1;\beta)\lambda_{12}(t_2)dt_2 \\
&\quad if \quad 0 < t_1 < t_2 < a_1 \\
L_6 &= \mathcal{A}(t_2)\lambda_{02}(t_2)dt_2\mathcal{B}_{01}(t_1 \mid t_2)\lambda_1(t_1)dt_1 \\
&\quad if \quad 0 < t_2 < t_1 < a_1.
\end{aligned}
$$

Now, let us consider the case where both times T_1 and T_2 are interval censored, i.e. $T_1 \in (T_{10}, T_{11})$ and $T_2 \in (T_{20}, T_{21})$. Let $[0, a_{i,1}]$ be the period between interviews for subject i. Let $t_{i,10} = a_{i,1}$ if T_1 is unobserved for subject i, and similarly $t_{i,20} = a_{i,1}$ if T_2 is unobserved for subject i. Accordingly, we can form the following likelihood function:

$$
L = \prod_i \mathcal{L}(t_{i,10}, t_{i,11}, t_{i,20}, t_{i,21}) \tag{3.4}
$$

where $\mathcal{L}(t_{10}, t_{11}, t_{20}, t_{21}) =$:

$$
\begin{aligned}
L_1 &= \mathcal{A}(a_1) \\
&\quad if \quad a_1 \le t_{10}, t_{20} \\
L_2 &= \mathcal{A}(t_{10}) \int_{t_{10}}^{t_{11}} \mathcal{A}(s)\lambda_1(s)\mathcal{B}_{12}(a_1 \mid s;\beta)ds \\
&\quad if \quad 0 < t_{10} < t_{11} < a_1 = t_{20} \\
L_3 &= \int_{t_{20}}^{t_{21}} \mathcal{A}(\tau)\lambda_{02}(\tau)\mathcal{B}_{01}(a_1 \mid \tau)d\tau \\
&\quad if \quad 0 \le t_{20} < t_{21} \le a_1 < t_{10}
\end{aligned}
$$

$$
\begin{aligned}
L_4 &= \int_{t_{20}}^{t_{10}} \mathcal{A}(\tau)\lambda_{02}(\tau)\mathcal{B}_{01}(t_{10} \mid \tau)d\tau\{\int_{t_{10}}^{t_{11}} \mathcal{B}_{01}(s \mid t_{10})\lambda_1(s)ds\} \\
&\quad + \int_{t_{10}}^{t_{11}} \mathcal{A}(s)\lambda_{02}(s)\{\int_{s}^{t_{11}} \mathcal{B}_{01}(u \mid s)\lambda_1(u)du\}ds + \int_{t_{10}}^{t_{11}} \mathcal{A}(s)\lambda_1(s)[1-\mathcal{B}_{12}(t_{21} \mid s;\beta)]ds \\
&\quad if \quad 0 \le t_{20} < t_{10} < t_{11} < t_{21} \le a_1 \\
L_5 &= \int_{t_{10}}^{t_{11}} \mathcal{A}(s)\lambda_1(s)ds[\mathcal{B}_{12}(t_{20} \mid t_{11};\beta) - \mathcal{B}_{12}(t_{21} \mid t_{11};\beta)] \\
&\quad if \quad 0 < t_{10} < t_{11} < t_{20} < t_{21} \le a_1 \\
L_6 &= \int_{t_{20}}^{t_{21}} \mathcal{A}(\tau)\lambda_{02}(\tau)\mathcal{B}_{01}(t_{10} \mid \tau)\{\int_{t_{10}}^{t_{11}} \mathcal{B}_{01}(s \mid t_{10})\lambda_1(s)ds\}d\tau \\
&\quad if \quad 0 < t_{20} < t_{21} \le t_{10} < t_{11} < a_1 \\
L_{4'} &= \int_{t_{10}}^{t_{20}} \mathcal{A}(s)\lambda_1(s)\mathcal{B}_{12}(t_{20} \mid s;\beta)ds[1-\mathcal{B}_{12}(t_{21} \mid t_{20};\beta)] \\
&\quad + \int_{t_{20}}^{t_{21}} \mathcal{A}(s)\lambda_1(s)[1-\mathcal{B}_{12}(t_{21} \mid s;\beta)]ds + \int_{t_{20}}^{t_{21}} \mathcal{A}(s)\lambda_{02}(s)\{\int_{s}^{t_{11}} \mathcal{B}_{01}(u \mid s)\lambda_1(u)du\}ds \\
&\quad if \quad 0 < t_{10} < t_{20} < t_{21} < t_{11} \le a_1 \\
L_{5'} &= \int_{t_{10}}^{t_{20}} \mathcal{A}(s)\lambda_1(s)\mathcal{B}_{12}(t_{20} \mid s;\beta)ds[1-\mathcal{B}_{12}(t_{21} \mid t_{20};\beta)] \\
&\quad + \int_{t_{20}}^{t_{11}} \mathcal{A}(s)\lambda_1(s)[\mathcal{B}_{12}(t_{21} \mid s;\beta)]ds + \int_{t_{20}}^{t_{11}} \mathcal{A}(s)\lambda_{02}(s)\{\int_{s}^{t_{11}} \mathcal{B}_{01}(u \mid s)\lambda_1(u)du\}ds \\
&\quad if \quad 0 < t_{10} < t_{20} < t_{11} < t_{21} \le a_1 \\
L_{6'} &= \int_{t_{20}}^{t_{10}} \mathcal{A}(\tau)\lambda_{02}(\tau)\mathcal{B}_{01}(t_{10} \mid \tau)d\tau\{\int_{t_{10}}^{t_{11}} \mathcal{B}_{01}(s \mid t_{10})\lambda_1(s)ds\} \\
&\quad + \int_{t_{10}}^{t_{21}} \mathcal{A}(s)\lambda_{02}(s)\{\int_{s}^{t_{11}} \mathcal{B}_{01}(u \mid s)\lambda_1(u)du\}ds + \int_{t_{10}}^{t_{11}} \mathcal{A}(s)\lambda_1(s)[1-\mathcal{B}_{12}(t_{21} \mid s;\beta)]ds \\
&\quad if \quad 0 < t_{20} < t_{10} \le t_{21} < t_{11} < a_1.
\end{aligned}
$$

As it is pointed out in Thompson and Pantoja-Galicia (2003), more realistic models would allow the hazards λ_1, λ_{02} and λ_{12} to depend on appropriate covariates.

3.2.1 Estimation

If we assume the exponential distribution as the underlying distribution of our data, we would be able to obtain a closed form for many parts of the expressions in Section 3.2, which can be supplemented by numerical integration.

The approach in Section 3.2 can become very complicated with a complex survey design. In principle, this would not be hard to deal with as long as we have a single stage design, so that individuals can be regarded as independent. In that case, the log likelihood at the population level is a population sum, and the sample analogue is a weighted sample sum. For example, considering (3.2), (3.3) and (3.4), we respectively have the following log likelihood functions at the population level

$$\sum_{i=1}^{N} \log \mathcal{L}(t_{i,1}, t_{i,20}, t_{i,21}) \tag{3.5}$$

$$\sum_{i=1}^{N} \log \mathcal{L}(t_{i,1}, t_{i,2}) \tag{3.6}$$

$$\sum_{i=1}^{N} \log \mathcal{L}(t_{i,10}, t_{i,11}, t_{i,20}, t_{i,21}) \tag{3.7}$$

where N is the number of units in the finite population. Correspondingly, at the sample level we have

$$\sum_{i \in s} w_i \log \mathcal{L}(t_{i,1}, t_{i,20}, t_{i,21}) \tag{3.8}$$

$$\sum_{i \in s} w_i \log \mathcal{L}(t_{i,1}, t_{i,2}) \tag{3.9}$$

$$\sum_{i \in s} w_i \log \mathcal{L}(t_{i,10}, t_{i,11}, t_{i,20}, t_{i,21}) \tag{3.10}$$

where s refers to the sample and w_i to the survey weight for subject i.

3.3 Multi-state analysis

It is important to mention the work developed by Cook, Zeng and Lee (2007) who present a multi-state analysis of bivariate interval censored event times. They propose an analysis based on a four-state stochastic model, where they define the states as follows in terms of the events E_1 and E_2 having occurred or not:

State 0: No occurrence of E_1 nor E_2.

State 1: Occurrence of E_1, but not yet E_2.

State 2: Occurrence of E_2, but not yet E_1.

State 3: Occurrence of E_1 and E_2.

With associated intensities $\lambda_{jk}(t \mid H(t))$ for the transitions from state j to state k, they restrict consideration to flexible Markov models with piecewise constant transition intensities λ_{jkl}, where l is the index associated with the l^{th} break point. The break points are fixed. Cook, Zeng and Lee (2007) derive the corresponding complete data log likelihood function. And since interval censored data is a type of incomplete data, an EM algorithm can be worked out for this scenario.

In our case, we also use the idea of piecewise constant hazards, but in our situation the break point is not fixed (the break points would be T_1 and $T_1 + \kappa$ if T_1 occurs before T_2). Cook, Zeng and Lee (2007) express how one type of event alters the risk of another, by allowing different intensities for the transitions from state j to state k. Our case involves a more local influence, where we have the intensity for T_2 affected in a certain way in the next short interval of time.

To estimate the parameters we have worked out the required integration (see Chapter 4) instead of following an EM algorithm. So a question would be whether the EM algorithm would be easier.

We would like to mention that in our situation, the main purpose is to model a shorter term triggering using complex survey data. The semiparametric model in principle allows us to make use of right censored as well as interval censored lifetimes.

Chapter 4

Applications

The purpose of this chapter is to illustrate the application of the methodologies proposed in Chapters 2 and 3 using data from two Statistics Canada's national longitudinal surveys. Our first application involves time to pregnancy and time to smoking cessation using data from the National Population Health Survey (NPHS). The second application considers time to job loss and time to divorce (or separation) using data from the Survey of Labour and Income Dynamics (SLID).

The following section provides a brief introduction to the survey that supplied the data for our first application.

4.1 The National Population Health Survey

The National Population Health Survey had been intended to collect both cross-sectional and longitudinal information related to the health of the population in Canada. The first cycle of data collection took place in 1994-1995. Subsequent cycles have been collected every second year thereafter and this process is planned to continue for up to 20 years. It is important to note that beginning with cycle 4 in 2000-2001, the survey became purely longitudinal.

The target population of the National Population Health Survey consists of three parts:

the household, the institutional and the northern component respectively. Our focus will be on the first component which includes household residents from every province, except persons living on Indian Reserves, Canadian Forces Bases and some remote areas in Quebec and Ontario (Statistics Canada; 1996).

During the first three cycles of the NPHS, in every household, a knowledgeable person provided information related to demographic, socio-economic and limited health issues about each household member. This cross-sectional part is contained in the *general component* of the NPHS questionnaire. On the other hand, a randomly selected individual was asked to answer more in-depth and detailed health questions. This information is catalogued in the *health component* of the questionnaire. Every second year and for longitudinal purposes, only the randomly selected person is followed up.

The sample design used outside Quebec is based on the Labour Force Survey of Statistics Canada. In the province of Quebec, the NPHS utilised the design of a health survey organized by Santé Québec: the 1992-93 Enquête sociale et de santé (ESS).

The NPHS considered a stratified two-stage design. Hence, homogeneous strata were created in the first stage and independent samples of clusters were drawn from each stratum. During the second stage, dwelling lists were prepared for each cluster and then households were selected from these lists. The sampling design for the initial cycle selected one individual at random from each of about 17,000 households across the country. Since only one member in each sampled household responds to the in-depth health questions, the probability of being selected as a respondent is inversely related to the number of people in the household.

For a detailed description of the design and other important features of this survey Larry and Catlin (1995) and Swain, Catlin and Beaudet (1999) are useful references.

This survey yields important information for the purpose of investigating a relationship between pregnancy and smoking cessation in the context that has been discussed in previous chapters.

4.2 Pregnancy and Smoking Cessation

The longitudinal nature of the NPHS makes possible to observe changes in the responses of the surveyed people across cycles. We pay special interest to women's responses to questions related to pregnancy and smoking cessation. The following questions are taken from the NPHS questionnaire.

- Have you ever smoked cigarettes at all?
- At the present time do you smoke cigarettes daily, occasionally or not at all?
- At what age did you stop smoking cigarettes daily?
- Are you currently pregnant?
- Have you given birth since last interview?

Therefore, at every cycle we are able to collect information such as: date of birth and gender of every participant, date of interview, whether a female respondent is pregnant, and the smoking status for each respondent (daily or occasional smoker and non-smoker). Also, each former daily smoker states at what age she stopped smoking cigarettes daily.

The presence of a household (cross-sectional) file in the NPHS plays an important role at providing information about age and date of birth of every member in the household from the first to the third cycle.

Considering the time origin to be the date of the interview at cycle n, let T_1 denote the time until a pregnancy begins, and T_2 the time until a smoking cessation occurs which lasts until the next interview at cycle $n + 1$ to be reported. The way we determine these times, which are interval censored for each subject, is revealed in subsection 4.3. The unit of measurement for time is years.

4.2.1 Subsample

Responses at two successive cycles, n and $n+1$, are needed to determine the intervals for the beginning of the pregnancy and the smoking cessation.

Our analysis is based in the following subsample of longitudinal respondents. For cycles $1, 2$ and 3, we select individuals who at the moment of the interview in the n^{th} cycle are between the ages of 15 and 49, are regular daily smokers and are not pregnant . The second inclusion criteria is that among these subjects, we select those who at the next cycle, $n+1$, report being pregnant or having given birth since the last interview; and having abandoned cigarettes. This subsample is depicted in figure 4.1.

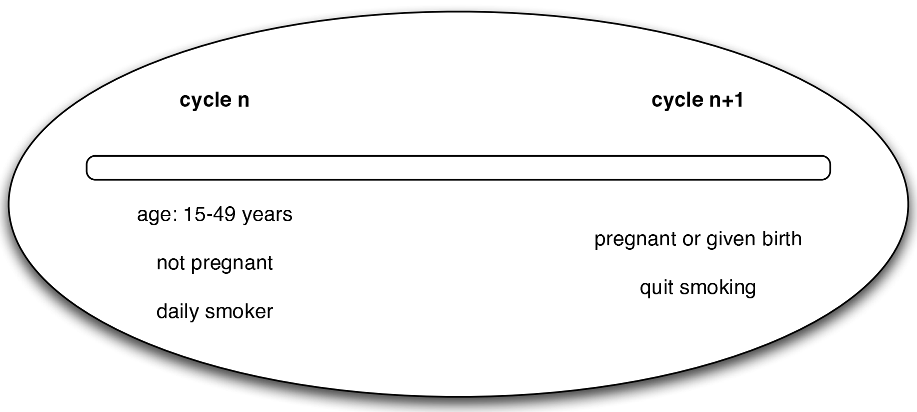

Figure 4.1: Sampled population of the first application using NPHS data.

4.3 Interval Censoring in the NPHS

4.3.1 Time to Pregnancy

We have previously mentioned that a household file has been helpful to obtain the date of birth for every household member living with the longitudinal respondent. Therefore, inferring the approximate dates of pregnancy for the longitudinal individual is possible by simply subtracting 9 months to the date of birth of the appropriate child. More specifically, if we consider 280 days (or 40 weeks) to be the average duration of a pregnancy, plus or minus 14 days (or 2 weeks) (Taylor et al.; 2003, p. 67), the date of becoming pregnant belongs to an interval which is 28 days in length. More specifically, let

- D_n be the date of the interview at cycle n (time origin);
- B be the date of birth of the appropriate child;
- P be the inferred date of pregnancy ($B - 280$ days);
- $P_L = P - 14$ days.
- $P_R = P + 14$ days (according to (Taylor et al.; 2003, p. 67)).

Therefore, $T_1 \in I_1 = (P_L - D_n, P_R - D_n)$. Figure 4.2 illustrates this description.

According to Sections 4.2 and 4.2.1, note that the time origin is not at the same date for every respondent. For some respondents it is D_1 , for others it is D_2, and D_3 for some others.

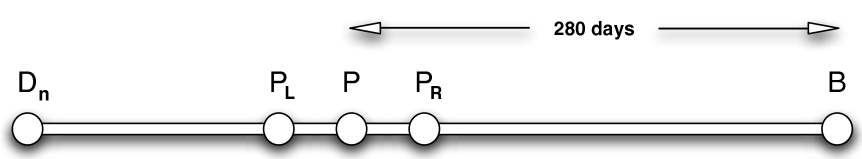

Figure 4.2: Time to Pregnancy

Due to the strictly longitudinal nature of the NPHS starting with cycle 4, the household (also called cross-sectional) file for the year 2000-2001 does not exist. Therefore, for the corresponding selected individuals, their respective time to becoming pregnant, T_1, would lie in a wider interval I_1. By way of illustration, consider the time unit for T_1 in months, therefore an individual that in cycle 4 reported to have given birth since her last interview, is assigned the interval $I_1 = (0, 15)$. This indicates that she could have become pregnant at any time between the day of her interview in cycle 3, i.e. D_3, and 15 months later. For a respondent that reported to be pregnant at the time of the interview in cycle 4, the interval $I_1 = (15, 24)$ is assigned to her, indicating that she could have started her pregnancy any time in the past 9 months.

4.3.2 Time to Smoking Cessation

At every cycle, subjects who became former daily smokers were asked the age at which they quit smoking. The approximate duration between interviews is of two years. If smoking cessation is reported, T_2 is thus to be observed between two end points T_{20} and T_{21}. T_{20} is either 0, which corresponds to the date of the interview in cycle n (D_n), or the duration of time between the date of the interview in cycle n and the corresponding birthday. On the other side, T_{21} is either the time to a birthday or 2 (obtained using the date of the interview in cycle $n + 1$).

By way of an example, and recalling the time units to be in years, let us consider the case of a longitudinal individual who at cycle n is 24 years old and who has her interview in cycle $n + 1$ exactly 2 years after the interview in cycle n. Let B_{25} denote the duration of time between the date of her interview at cycle n and the date of her twenty-fifth birthday (equivalently for B_{26}). Depending on the reported age at which she quit smoking, T_2 belongs to the interval I_2 which may be $(0, B_{25})$ if the reported age of quitting is 24; (B_{25}, B_{26}) if the age of smoking cessation is reported to be 25; or $(B_{26}, 2)$ if the subject responded 26 as her age of abandoning cigarettes.

- $(0, B_{25})$ if the reported age of quitting is 24;
- (B_{25}, B_{26}) if the age of smoking cessation is reported to be 25;
- $(B_{26}, 2)$ if the subject responded 26 as her age of abandoning cigarettes.

Figure 4.3 depicts this example.

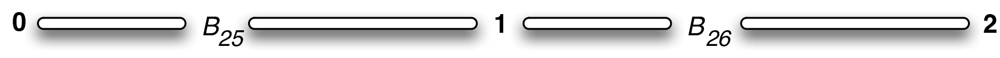

Figure 4.3: An example of the determination of the time to smoking cessation

We note that the records of each respondent at the 4^{th} cycle present the age at which the individual stopped smoking as *not stated*. Therefore, the length of the time interval I_2 for each of these individuals will be of about two years, which is the period of time between the interviews at cycle 3 and 4.

4.4 Results

Recall that to compute the test statistic Q presented in (1.2), we first obtain an estimate, within the observation window, of the joint density of (T_1, T_2) using (2.25). Figure 4.5 depicts the contour plot of such estimated joint density of (T_1, T_2) and shows the expected ordering of T_1 and T_2. Subsequently, we obtain numerically the corresponding marginal probability density functions and consequently the respective survivor functions (conditional and unconditional versions).

The size of the sample satisfying the conditions described in Section 4.2.1 is of 57 individuals. These 57 respondents represent about $68,000$ people of our total target population. For every subject the corresponding intervals I_1 and I_2 have been determined according to the description in Section 4.3. To obtain (2.25) and figure 4.5, the following settings were also employed. We let $g(\mathbf{X})$ in (2.6) be a bivariate uniform density and considered $B = 7^2$ to be an appropriate choice for the value B required in (2.7). Therefore at every iteration j that estimates $\hat{f}_j$, we generated 7^2 bivariate uniform random values within each rectangle $\mathbf{I}_i$. This value of B was also employed for the results in Sections 4.8.1 and 4.8.2.

The two modes on Figure (4.5) appear because the data subset being used contains two groups of respondents. Group 1 are those who had children and quit smoking but did not

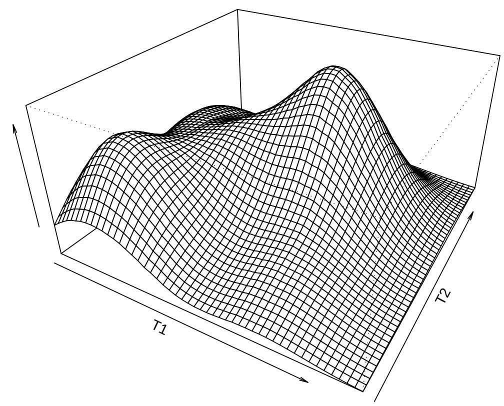

Figure 4.4: Estimated joint density of T_1 and T_2.

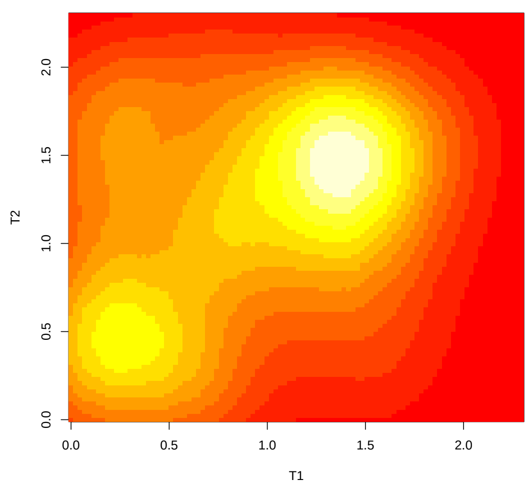

Figure 4.5: Contour plot of the estimated joint density of T_1 and T_2. Light-colored areas indicate high density.

relapse between cycle n and $n+1$. Group 2 represent those quitters who have become pregnant and are still pregnant at cycle $n+1$. A feature of the NPHS data means that some of the intervals for these two groups are fixed. The visual representation given by the estimated joint density is a useful tool for seeing the effect of this additional structure.

Regarding the test for close precursor as discussed in Section 1.2, we let $\kappa(t_1)$ have a constant value of about 2.5 months, assuming that smoking cessation might occur with increased probability within the first trimester of pregnancy.

For each t_1, we have that $Q = 0.029$ with $se(Q) = (0.004)$. After comparing Q with twice its estimated standard error, we reject the null hypothesis, which is effectively that the mean of Q is 0. Therefore, we can argue that there is evidence that T_1 is a close precursor of T_2, i.e. the occurrence of the pregnancy at time T_1 decreases the probability of having to wait longer than 2.5 months for the smoking cessation to occur.

The estimated standard error was obtained as discussed in Section 2.4, by using Statistics Canada's bootstrap weights replicates. The size of our domain is small, however, it is widely dispersed. Therefore the chances of getting many people from a primary sampling unit (cluster) are small. We examined Statistics Canada's bootstrap weights replicates for the domain and found that the distribution of the weights that are equal to zero is fairly uniform. In addition, the corresponding coefficients of variation of the bootstrap weights replicates are fairly uniform as well. A similar situation applies to the respective results of the SLID application.

R code was written so that estimation of the joint density could be implemented according to the methodology of Chapter 2. The corresponding figures and computation of the test statistic were also obtained from this code.

To check the assumed null distribution for Q, we have performed simulations using models and parameter settings appropriate to the NPHS application (time to pregnancy and time to smoking cessation) and using our density estimation method. These show a mean for Q under independence that is very close to 0 and thus validate the significance test of this section.

4.4.1 An illustration of estimation of a density with a discontinuity

Using the approach presented in Section 2.5.2 yields the following density estimate of (T_1, T_2) in the case of the NPHS example. In Figure 4.6, the density is estimated separately using the local likelihood method for positive and negative $V_2 = (T_2 - T_1)/\sqrt{2}$.

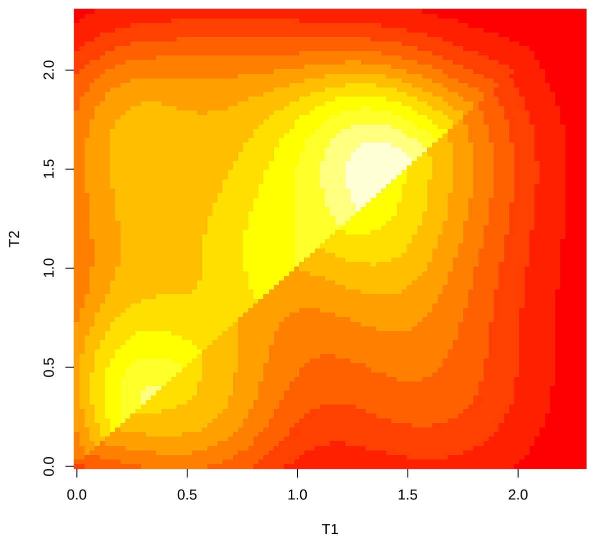

Figure 4.6: Contour plot of the estimated joint density with a discontinuity at $T_1 = T_2$. Light-colored areas indicate high density. Time units: years.

4.5 The Survey of Labour and Income Dynamics

The Survey of Labour and Income Dynamics (SLID) is a longitudinal survey composed of panels of six years in length. The purpose of this survey is to track the experiences of individuals in the labour market, their income and changes in family life. A longitudinal panel of all persons belonging to the sample of households derived from the Labour Force Survey is formed and kept for six years. Longitudinal respondents aged 16 years or older are contacted twice a year. At the first interview, in January, they answer questions about their labour activities. In May, information about their income is obtained either through a second interview or by access with permission to the respondent's tax return.

The longitudinal sample of the SLID is a stratified multistage sample, with strata defined within provinces, and, generally, with two primary sampling units (psu 's) selected from each stratum with a probability proportional to size. A sample of households is chosen within each (psu). So, the dwelling place is the last-stage sampling unit. The first panel started in 1993 and consisted of about 15,000 households, which account for approximately 40,000 people (31,000 persons who are over 16 years of age). The second panel of approximately the same size was selected in 1996, the third in 1999, etc. In this work we analyzed data from the first and second panels.

Further details regarding design as well as other important issues of the Survey of Labour and Income Dynamics can be found at Statistics Canada (1997) and Statistics Canada (2005).

4.6 Job loss and divorce

According to reports from Marienthal, given in a classic book in unemployment research (Jahoda, Lazarsfeld and Zeisel; 1933, p. 86), improvements in the relationship between husband and wife as a result of unemployment are definitely exceptional. As it is a subject

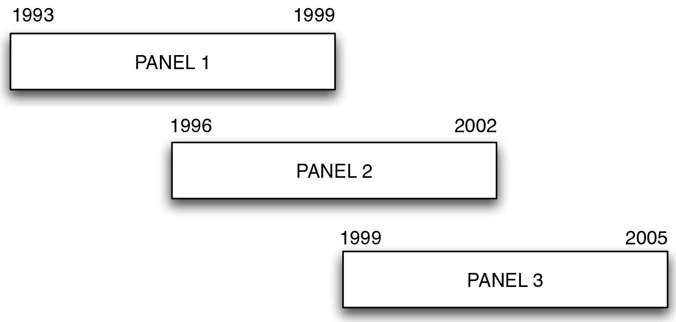

Figure 4.7: First three panels of SLID

of interest for social scientists, the topic of job loss and divorce has been examined more recently by Yeung and Hofferth (1998), Huang (2003) and Charles and Stephens (2004).

Considering the time origin to be the date of the first interview of the respondent (day zero), which takes place at the beginning of the life of the panel (some time in January of 1993 for panel 1), let T_1 denote the time to the termination of the job of the subject, and let T_2 be the time to either separation or divorce (whichever comes first, as the result of the termination of the marriage or common-law relationship of the individual). The unit of measurement for time is years.

A vector of all the dates of changes of marital status for each respondent, along with associated type of change, can be obtained from a panel of the SLID. In the same manner, a vector of job history with dates of changes in employment status can be retrieved. With this information, an appropriate data set can be generated for our own subsample, which is described in the next section.

An in-depth description regarding the extraction and structure of the set of variables to be used for this and the previous application is given by Pantoja-Galicia and Thompson (2006). A summary description has been reproduced in Appendix B.

4.6.1 Subsample

We take into account subjects from the second panel of the SLID with the following characteristics: employed and married (or in a common-law union) at day zero; with only one marriage (or common-law relationship) and one job during the life of the panel; and with termination of the job and dissolution of the marital relationship occurring during the life of the panel. We restrict further to cases where the job ended due to involuntary reasons. We condition on observing both events during the time window from January 1993 to April 1999 (for the first panel). In other words, let

- D_U be the date at which the union started. The union can be either marriage or a common-law relationship. Consider whichever started first.
- D_J be the date at which the job started.
- D_I be the date of the first interview.
- D_T be the date of the termination of the panel.
- D_{E1} be the date of the occurrence of E_1 (end of job).
- D_{E2} be the date of the occurrence of E_2 (end of the relationship).

Then, according to the description of our subsample, these dates are restricted to

$$D_U \leq D_I, D_J \leq D_I.$$

Also

$$D_I < D_{E1} \leq D_{E2} < D_T$$

or

$$D_I < D_{E2} \leq D_{E1} < D_T.$$

4.7 Interval Censoring in the SLID

Memory plays an important role in survey responding. Whenever a date of an event is reported, there exists the potential for dating errors. *Forward telescoping* is a type of memory error which involves reporting the occurrence of events more recently than they actually happened. The events are seen as closer in time than they really are, according to the interview's vantage point. As stated by Tourangeau, Rips and Rasinski (2000), this phenomenon has been studied by survey methodologists and cognitive psychologists since Neter and Waksberg (1964) first documented it. In the opposite direction, *Backward telescoping* is another possibility for a dating error. Tourangeau, Rips and Rasinski (2000) present, in the fourth chapter, a vast review of the literature documenting these sort of memory errors. From the same source in page 11, we quote:

> Reporting errors due to incorrect dating seem to arise through several distinct mechanisms. People may make incorrect inferences about timing based on the accessibility (or other properties) of the memory, incorrectly guess a date within an uncertain range, and round vague temporal information to prototypical values (such as 30 days).

These issues might be reflected in survey data as measurement error. In reports like those from Huttenlocher, Hedges and Bradburn (1990) it is also shown that respondents round estimates to times that are stand-ins for calendar units, that is seven days or thirty days. This background justifies trusting the reported dates of the events to be within a certain period of time instead of a specific day. This leads us to have times to occurrence which are interval censored for the events of our interest.

By way of illustration, let T_J and T_D be the reported times to occurrence of job termination and divorce (or separation) respectively. Therefore, we may say that $T_1 \in I_1 = (T_J - \delta_1, T_J + \delta_2)$ and $T_2 \in I_2 = (T_D - \delta_3, T_D + \delta_4)$, where δ_k is a particular period of time, for $k = 1, \ldots, 4$. For instance, if the trusted period of time is of 30 days then $\delta_k =$ (15 days) for all k. Note that $T_J = (D_{E1} - D_I)$ and $T_D = (D_{E2} - D_I)$.

4.8 Results

4.8.1 First Panel

Using data from the first panel of the SLID we found that the size of the sample population satisfying the conditions described in Section 4.6.1 is of 70 individuals, who represent about 49,000 people of the total target population. The corresponding intervals I_1 and I_2 have been determined for every subject in our sample according to Section 4.7.

Figure 4.8 shows an estimate of the joint density of (T_1, T_2) using (2.25). Recall that T_1 is the time until the end of job (on the horizontal axis) and T_2 is the time until separation or divorce (on the vertical axis). This picture shows the expected ordering.

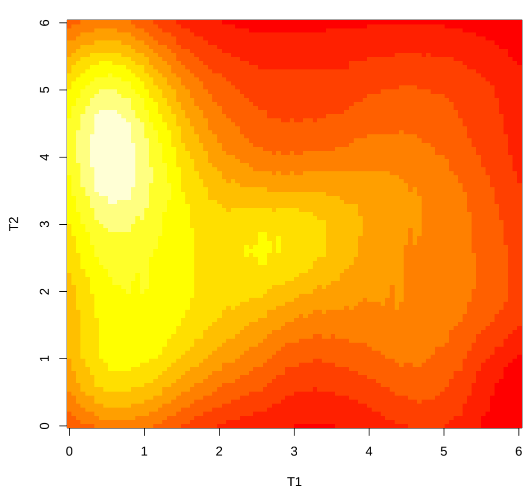

Figure 4.8: Contour plot of the estimated joint density of T_1 and T_2. Time in years. Light-colored areas indicate high density.

Concerning the test for close precursor from section 1.2, if we let $\kappa(t_1)$ to have a constant value of 6 months, for every t_1, the test statistic for close precursor Q will have a value of 0.034 with a standard error of 0.0025. The comparison of Q with twice its estimated standard error, gives us evidence to reject the null hypothesis from Section 1.2. Therefore, there exists evidence to argue that T_1 is a close precursor of T_2, i.e. losing a job at time T_1 decreases the probability of having to wait longer than 6 months to observe a separation or divorce.

4.8.2 Second Panel

Figure 4.9 shows an estimate of the joint density of (T_1, T_2) obtained according to (2.25). In compliance with the conditions established in Section 4.6.1, 53 individuals were selected. These represent about $40,000$ people of the total target population. In accordance with Section (4.7), the intervals I_1 and I_2 were determined for each subject.

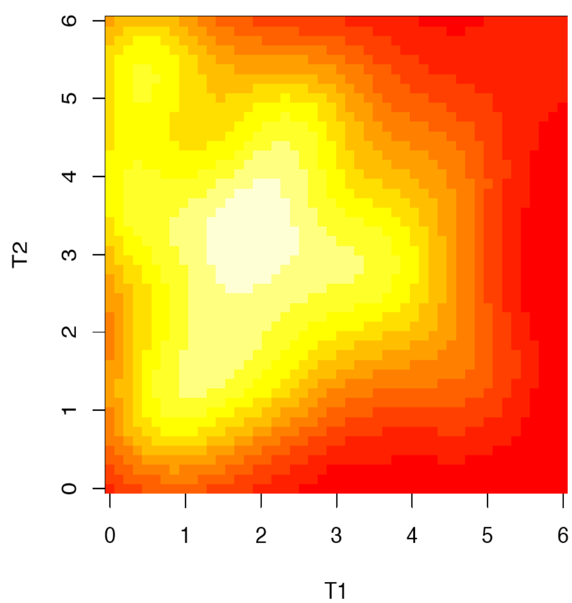

Figure 4.9: Contour plot of the estimated joint density of T_1 and T_2.

If we let $\kappa(t_1)$ have a constant value of about 6.5 months, for every t_1, the test statistic for close precursor Q from (1.2) results in a value of 0.0304 with a standard error of 0.0008. The comparison of Q with twice its estimated standard error gives evidence to reject the null hypothesis from Section 1.2. Therefore, there is evidence that T_1 is a close precursor of

T_2, i.e. losing a job at time T_1 decreases the probability of having to wait longer than 6.5 months to observe a separation or divorce.

We have selected 6.5 months as being a censoring half-interval plus one twelfth of the length of a 6 year panel, and thus representing a short term in comparison with the panel length.

4.8.3 Results from a semiparametric model

In Chapter 3, we presented short term triggering models and likelihood functions for the case when T_1 is not interval censored and $T_2 \in (T_{20}, T_{21})$.

Assume

$$\begin{aligned}
\lambda_1(t) &= \lambda_1, \\
\lambda_{02}(u) &= \lambda_{02} \\
\lambda_{12}(t) &= e^{\beta}\lambda_{02},\ t \le u < t + \kappa(t).
\end{aligned}$$

In principle, we could maximize the log likelihood function (3.8) to obtain estimates of λ_1, λ_{02} and β. Note that the log likelihood function (3.8) includes the corresponding terms L_k, $k = 1, \ldots, 6$, depending on whether the data satisfies the following conditions:

$$\begin{array}{lll}
L_1 & if & a_1 \le t_1, t_{20} \\
L_2 & if & 0 < t_1 < a_1 = t_{20} \\
L_3 & if & 0 \le t_{20} < t_{21} \le a_1 < t_1 \\
L_4 & if & 0 \le t_{20} < t_1 < t_{21} \le a_1 \\
L_5 & if & 0 < t_1 < t_{20} < t_{21} \le a_1 \\
L_6 & if & 0 < t_{20} < t_{21} \le t_1 < a_1
\end{array}$$

In the following implementation, however, we consider data for which both events E_1 and E_2 are observed to occur within a time window, i.e. the same data that were used in the previous applications. Due to this conditioning, the corresponding log likelihood function is modified accordingly and includes only the terms L'_4, L'_5 and L'_6, where

$$
\begin{aligned}
L'_4 &= \frac{L_4}{\int_{t_{20}}^{t_{21}} L_4 dt_1 + \int_{t_0}^{t_{20}} L_5 dt_1 + \int_{t_{21}}^{a_1} L_6 dt_1} && if \quad 0 \leq t_{20} < t_1 < t_{21} \leq a_1 \\
L'_5 &= \frac{L_5}{\int_{t_{20}}^{t_{21}} L_4 dt_1 + \int_{t_0}^{t_{20}} L_5 dt_1 + \int_{t_{21}}^{a_1} L_6 dt_1} && if \quad 0 < t_1 < t_{20} < t_{21} \leq a_1 \\
L'_6 &= \frac{L_6}{\int_{t_{20}}^{t_{21}} L_4 dt_1 + \int_{t_0}^{t_{20}} L_5 dt_1 + \int_{t_{21}}^{a_1} L_6 dt_1} && if \quad 0 < t_{20} < t_{21} \leq t_1 < a_1
\end{aligned}
$$

Consequently, considering data from the second panel of SLID, the corresponding estimates and standard errors are given by

$$
\begin{aligned}
\hat{\lambda}_1 &= 0.6821\ (0.4609) \\
\hat{\lambda}_{02} &= 0.2736\ (0.4084) \\
\hat{\beta} &= 1.4831\ (0.7437).
\end{aligned}
$$

4.9 Remarks

The density estimates from Sections 4.4, 4.8.1 and 4.8.2 were obtained for the data with (T_1, T_2) interval censored. The null hypothesis we are testing is that T_1 and T_2 are independent, given that both are within the observation window. This is a weaker null hypothesis than global independence of T_1 and T_2, and is more appropriate in our context because the data are extracted for joint density estimation conditional on being in the observation window.

It can be noted that the selection criteria from Section 4.2.1 make it possible to select a respondent more than once over the four cycles of the NPHS. In fact, however, we did not have respondents with more than one contribution to the data.

Some important specifications are considered in the selection of the sample for the second application. First, we condition on being at risk for both events at the day of the first interview. This eliminates the possibility of including individuals who become married and/or employed after day 0, for whom different distributions will apply. Second, we observe both events in the time window of the life of the panel. This conditioning is reasonable and gives us a sensible estimate since we are looking at close following and therefore are interested in events that are close in time. Third, we select those individuals whose job is reported to have ended during the life of the panel due to involuntary reasons. This inclusion criteria is intended to avoid as much as possible a potential effect of divorce as a trigger for job loss.

In Section 4.8.3 we employed the approach presented in Chapter 3 to estimate the parameters of interest using data which is conditioned to be within a time window, i.e. data for which both events E_1 and E_2 are observed to occur within a time window $(0, a_1)$. In principle, the log likelihood function from Chapter 3 can be used to estimate the parameters of interest using data which is not conditioned to be within a time window.

4.10 Bandwidth Selection

An important aspect of local likelihood density estimation is the selection of the smoothing parameter. When h is very large, the density estimate is oversmoothed and the bias of the density estimate is large; when h is very small, the bias is small but the variance of the estimate is large.

In some contexts, the selection of the smoothing parameter obtained in a subjective manner is adequate (Silverman; 1986, p.44). Quoting Silverman:

> A natural method for choosing the smoothing parameter is to plot out several curves and choose the estimate that is most in accordance with one's prior ideas

about the density. For many applications this approach will be perfectly satisfactory. Indeed, the process of examining several plots of the data, all smoothed by different amounts, may well give more insight into the data than merely considering a single automatically produced curve.

However, for many purposes a more formal analysis is necessary. Sheather (2004) briefly reviews and classifies some methods for choosing a global value of the window width according to Rules of Thumb, Cross-Validation and Plug-in methods.

In the univariate context of nonparametric density estimation from clustered sample survey data, Breunig (2001) examined an optimal bandwidth selection using a higher-order kernel. A method using likelihood cross-validation was presented by Braun, Duchesne and Stafford (2005) in the context of univariate local likelihood density estimation with interval censored data. Faraway and Jhun (1990), Taylor (1989) and Hall (1990) have presented bootstrap approaches to select the smoothing parameter in kernel density estimation.

We approached the problem in the following manner. Consider the use of a gaussian kernel and that $h = h_1 = h_2$. According to Silverman (1986, p.86-87), assuming the underlying distribution to be the unit bivariate normal density and acting as though the sample were i.i.d., then the smoothing parameter which minimizes the Asymptotic Mean Integrated Square Error (AMISE) is given by

$$h_{opt}^{sil} = 0.96 n^{-1/(d+4)}, \tag{4.1}$$

where d indicates the dimensionality of the data, which in our case is $d = 2$. Considering the NPHS application, a preliminary bandwidth was estimated as 0.48.

Note that although this approach works well if the sample were i.i.d. and the population really followed a unit bivariate normal density, it may oversmooth somewhat if the population is multimodal, as result of the value of the curvature (integrated square second derivative) being larger (Silverman; 1986). Therefore, in Figures 4.4 and 4.5, we reduced the bandwidth

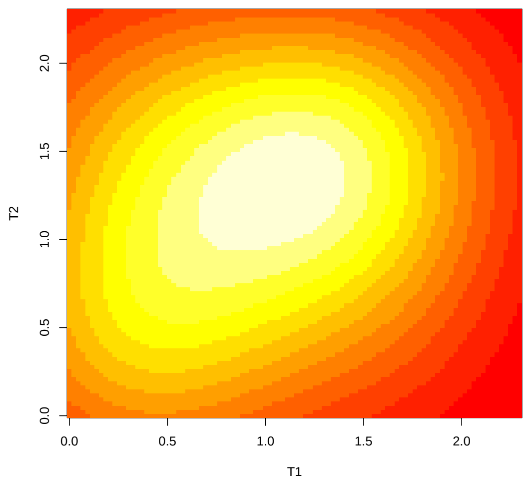

Figure 4.10: Contour plot of the estimated joint density. NPHS example. Bandwidth $h_{opt}^{sil} = 0.48$.

to 0.25 so that we would not miss features of the density by possible oversmoothing (compare Figure 4.5 with Figure 4.10).

We can also think of $h_{opt}^{sil} = 0.48$ as a starting point for subsequent fine tuning of the bandwidth.

An approach to using the bootstrap replicates of the survey weights (see Sections 1.4.5 and 2.4) to estimate the Mean Integrated Square Error (MISE) for a given bandwidth h would be the following.

Let $w_i^{(r)}$ be the normalized bootstrap weight of the r^{th} replicate for individual i, such that $\sum_{i \in S} w_i^{(r)} = 1$. If we employ R of these bootstrap weight replicates, the MISE may be assessed as follows.

For each set r of replicates:

1. Obtain (2.25) using the standardized survey weights w_i^* for $i \in S$. Let $\tilde{f}_j^{w^*}$ be the corresponding estimate.

2. Obtain (2.25) using $w_i^{(r)}$ (instead of w_i^*) for $i \in S$. Let $\tilde{f}_j^{w^{(r)}}$ be the corresponding estimate.

3. Compute

$$BMISE(h) = \frac{1}{R}\sum_{r=1}^{R}\int(\tilde{f}_j^{w^{(r)}}(\mathbf{x};h) - \tilde{f}_j^{w^*}(\mathbf{x};h))^2\mathbf{dx}. \tag{4.2}$$

We may select a bandwidth by minimizing $BMISE(h)$ over h.

However, the estimate of the bias implicit in (4.2), vanishes. The component of bias increases with h and can be of considerable importance. Since the bias term penalizes oversmoothing, a small bandwidth would be favourable (to minimize the bias component). This method considers only an estimate of the variance implicit in (4.2). The variance term penalizes undersmoothing, so that a large bandwidth would be preferred. Consequently, minimization of $BMISE(h)$ over h would suggest to take h as large as possible.

An improvement to this method would then include a penalty function $g(h)$ for oversmoothing in (4.2). Therefore, we may proceed as follows.

For each set r of replicates:

1. Obtain (2.25) using the standardized survey weights w_i^* for $i \in S$. Let $\tilde{f}_j^{w^*}$ be the corresponding estimate.

2. Obtain (2.25) using $w_i^{(r)}$ (instead of w_i^*) for $i \in S$. Let $\tilde{f}_j^{w^{(r)}}$ be the corresponding estimate.

3. Compute

$$PBMISE(h) = \left\{\frac{1}{R}\sum_{r=1}^{R}\int(\tilde{f}_j^{w^{(r)}}(\mathbf{x};h) - \tilde{f}_j^{w^*}(\mathbf{x};h))^2\mathbf{dx}\right\} + g(h) \tag{4.3}$$

Finally, select the bandwidth $\hat{h}_p$ by minimizing $PBMISE(h)$ over h.

Considering the NPHS application and taking $g(h) = h^2$, we obtained a value for $\hat{h}_p$ which is closer to the bandwidth employed in Figures 4.4 and 4.5 and thus further validates our choice of smoothing parameter in the first place.

Chapter 5

Discussion and Future Research

The new aspects of this thesis include an extension of nonparametric density estimation methods for interval censored survey data to the bivariate case. Furthermore, we propose a method for estimation of a density with a cusp or a discontinuity in a certain region. Additional contributions cover the development of a semiparametric approach to deal with interval censoring in two event times. The application of these methodologies to complex surveys, using data from the National Population Health Survey and the Survey of Labour and Income Dynamics, is also a new aspect of this dissertation.

Close precursor

As it has been mentioned in Section 1.1, the quantity $\kappa(t_1)$, which may be thought of as the duration of an effect, would come from subject-matter considerations. Although (1.1) indicates it depends on t_1 (anticipating that the effect of T_1 on the hazard of T_2 might not have constant duration), in neither of our applications was there a clear reason not to take κ to be constant.

In section 1.2 it has been pointed out that the difference within (1.2) approximates the difference between the hazard function conditional on $T_1 = t_1$ and the unconditional hazard (i.e. a local additive change).

A topic for future research would examine the sensitivity to the assumption of the size of κ. We could consider estimating $Q(\kappa)$ for a range of different values of κ and check for which values of κ there is evidence for a close precursor relationship of T_1 and T_2 and for which values of κ there is no evidence to reject the null hypothesis of independence.

In a future stage, more extensive simulation studies may be conducted to check the assumed null distribution for Q. These simulation studies may be implemented using different models and parameter settings which may be appropriate to the NPHS and SLID applications.

In Section 1.1 we also mentioned that (1.1) reflects approximately a short term raising of the hazard function for T_2. In principle, it may be not difficult to formulate an analogue for point process intensities, giving us an alternative way of modelling events less tied to a time origin. This may be pursued in future work. Blossfeld and Mills (2003), for example, used interdependent point processes to model interrelated family events, namely entry into marriage (for individuals in a consensual union) and first pregnancy or childbirth.

Density estimation

Note that in our case, the local EM algorithms of Chapter 2 are implemented by computing conditional expectations using an importance sampling technique. This method relies on the implementation of a bivariate uniform sampling scheme derived from the orthogonal array-based latin hypercubes described by Tang (1993), also called U sampling. This method is appealing, particularly for nonparametric density estimation in higher dimensions.

The estimated joint density has an easily interpreted visual representation. We can also observe whether the plot is consistent with the inference, and whether it is consistent with the idea of triggering by comparing it with the contour plot under the null hypothesis.

Other topics for further research include the following.

Semiparametric Models

In principle the semiparametric models of Chapter 3 allow us to make use of right censored as well as interval censored lifetimes for estimation of the corresponding intensities. The estimating equations may be derived respectively from the likelihood functions in Section 3.2 assuming underlying distributions for the lifetimes. Therefore, such implementation constitutes material for future work.

In Chapter 3, we also indicated that more realistic models would allow the hazards to depend on appropriate covariates. This could be implemented in future work as well.

We have mentioned that the approach in Section 3.2 can become very complicated considering a complex survey design. In principle this would not be hard to deal with as long as we have a single stage design, so that individuals can be regarded as independent. In that case the log likelihood at the population level is a population sum, and the sample analogue is a weighted sample sum. We have the following open questions:

(a) Is this type of model the most appropriate for survey data, particularly when the design has more than a single stage of sampling?

(b) Should some kind of mixture be considered to account for heterogeneity?

Convergence

Regarding local likelihood density estimation, the results in Chapter 2 related to convergence of our algorithms to a unique density estimate refer to the locally constant case. Appropriate conditions can be established to guarantee the existence of a fixed point density estimate in the locally linear case. However, establishing conditions to guarantee the uniqueness of or convergence to a fixed point density estimate requires further research.

A generalization of Theorems 2 and 3 for h_1 not necessarily equal to h_2 may be pursued also.

Asymptotic results considering that the survey weights are not only the inverse inclusion probability weights may be implemented in further work.

Bandwidth selection methods in kernel and local likelihood density estimation

The development of procedures for an optimal choice of the smoothing parameter in kernel density estimation and local likelihood density estimation is a topic for future research. Faraway and Jhun (1990), Taylor (1989) and Hall (1990) have presented bootstrap approaches to select the smoothing parameter in kernel density estimation. Further research on bootstrap procedures to select the smoothing parameter in the context of local likelihood density estimation for multivariate interval censored and complex survey data may be pursued.

Appendix A

The Normal Kernel

The normal (gaussian) kernel with bandwidth h mentioned in Chapter 2 is the following:

$$K_h(u) = \frac{1}{h}K(u/h) = \frac{1}{h\sqrt{2\pi}}e^{-(1/2)(u/h)^2}.$$

The Kuhn-Tucker Conditions

Recall the log likelihood given in (2.79)

$$l(\mathbf{p}) = \sum_{i=1}^{n} log\Big(\sum_{j=1}^{J} \alpha_j^i p_j\Big).$$

Let $\mathbf{p} = (p_1, \ldots, p_J)$. To find the maximum likelihood estimate of $\mathbf{p}$ Gentleman and Geyer (1994) maximize $l(\mathbf{p})$ with respect to $\mathbf{p}$ subject to the constraints

$$1 - \sum_{j=1}^{J} p_j = 0 \tag{A.1}$$

and

$$p_j \geq 0. \tag{A.2}$$

A point $\hat{\mathbf{p}}$ is a maximum likelihood estimate if and only if there exist Lagrange multipliers λ_j for $j = 0, \ldots, J$ such that the Kuhn-Tucker conditions (A.1) - (A.5) hold, for $j = 1, \ldots, J$, where

$$\lambda_j p_j = 0 \tag{A.3}$$

$$\lambda_j \geq 0 \tag{A.4}$$

$$\frac{\partial}{\partial p_j}\left[l(\mathbf{p}) + \sum_{j=1}^{J} p_j(\lambda_j - \lambda_0)\right] = \frac{\partial l}{\partial p_k} + \lambda_j - \lambda_0 = 0 \tag{A.5}$$

Note that $\lambda_0 = n$, which is obtained by multiplying (A.5) by p_j, summing over all j and using the fact that

$$\frac{\partial l}{\partial p_k} = \sum_{i=1}^{n} \frac{\alpha_k^i}{\sum_{j=1}^{J} \alpha_j^i p_j}.$$

Appendix B

NPHS Variables

For every longitudinal respondent a set of useful variables was obtained and according to the description of our subsample, the appropriate data file was created using the software SAS v8.

A list of the variables we have employed for each selected longitudinal individual is as follows:

- Record identifier for the household: REALUKEY.
- Number identifying the person in the household: PERSONID.
- Longitudinal response pattern: LONGPAT.
- Sex: SEX.
- Day, month and Year of birth: DOB, MOB, YOB.

Day, month and year of the interview:

- AM54_BDD, AM54_BMM, AM54_BYY, (cycle 1).
- AM56_BDD, AM56_BMM, AM56_BYY, (cycle 2).
- AM58_BDD, AM58_BMM, AM58_BYY, (cycle 3).

- AM60_BDD, AM60_BMM, AM60_BYY, (cycle 4).

Age of the respondent at the time of the interview:

- DHC4_AGE, DHC6_AGE, DHC8_AGE, DHC0_AGE, (cycles 1 to 4).

Whether the respondent gave birth since last interview:

- GHC6_21, GHC8_21, GHC0_21, (cycles 2 to 4)

Whether the individual is pregnant at the moment of the interview:

- HWC4_1, HWC6_1, HWC8_1, HWC0_1, (cycles 1 to 4).

Type of smoker (daily or occasional):

- SMC4_2, SMC6_2, SMC8_2, SMC0_2, (cycles 1 to 4).

Age at which stopped smoking daily (for the former daily smoker):

- SMC6_8, SMC8_8, SMC0_8, (cycles 2 to 4).

Reason for quitting smoking:

- SMC6_9, SMC8_9, SMC0_9, (cycles 2 to 4).

Our exercise employed demographic variables from the household (cross-sectional) files of 1996-97 and 1998-99 (cycles 2 and 3). These were obtained from the general component of the NPHS. This component is present only in the first three cycles and include:

Household and person identifiers:

- REALUKEY and PERSONID.

Gender of each household member:

- DHC6_SEX, DHC8_SEX, (cycles 2 and 3)

Dates of birth (day, month and year of birth) of every individual in the household:

- DHC6_DOB, DHC6_MOB and DHC6_YOB, (cycle 2).

- DHC8_DOB, DHC8_MOB and DHC8_YOB, (cycle 3).

Age of each person in the household:

- DHC6_AGE, DHC8_AGE, (cycle 2 and 3).

Family identification code of every member in the household:

- DHC6_FID, DHC8_FID, (cycles 2 and 3).

SLID Variables

For every longitudinal respondent a set of useful variables was obtained from each of the first and second panels of the survey.

SLID RETrieval Software Version 2.2 (SLIDRET v2.2) was used to extract a vector of all dates of changes of the marital status (with associated type of change) for each person in Panels 1 and 2. We employed entity number 4 (MARSTAT) for this purpose and the chosen unit of analysis was: Marital status.

The type of analysis was indicated as: Longitudinal, with reference years covering waves 1 to 6 of the panel. Finally, according to the description of our subsample, the appropriate data file was created using the software SAS v8.

For panel 1, the corresponding variables and their description are listed below.

- **personid**: Unique identifier for a person

- **ilgwt26_1993**: Internal longitudinal weight for the person for the reference year which is 1993 in this case. ilgwt26_1994 to ilgwt26_1999 were also part of the output of the SLIDRET query. Applicable to panel one of respondents.

- **stateid**: Unique identifier for one marital state for one person.

- **strdat4**: Date on which the marital state began.

- **enddat4**: Date of termination of the marital state.

- **ended4**: Indicates whether marital state ended.

- **state4**: Marital status of respondent, which includes:

  - Married.

  - Common-law.

  - Separated (persons separated from a common-law relationship are included here).

  - Divorced.

  - Widowed.

  - Single (never married). This is the first state for all persons.

Similarly, for dates of employment (or dates of changes in employment status), a vector of job history was also obtained using SLIDRET v2.2. We have utilized entity 9 (JOB) and the unit of analysis was: Person-Job.

The type of analysis was also indicated as: Longitudinal, with reference years for retrieval from 1993 to 1998.

A description of each variable follows its name:

- **personid**: Unique identifier for a person.

- **ilgwt26_1993** : Internal longitudinal weight for the person for the reference year (1993) in panel 1. Same as above.

- **jobid**: Unique identifier for a job or employment spell with an employer.

- **strdat9** : Start date of job.

- **enddat9**: End date of job.
- **ended9** : Flag to indicate if job had ended by the end of the most current survey reference period of the data file.
- **jobdur9** : Duration of job expressed in months.
- **endtyp9** : Reason why job was ended in processing.
- **typjs9** : Type of job separation (voluntary or involuntary).
- **reaend9** : Reason why work came to an end.

The reasons why the work came to an end that are considered to be involuntary job separations are the following:

- Company moved.
- Company went out of business.
- Layoff/Business slowdown (not caused by seasonal conditions).
- Labour dispute.
- Dismissal by employer.

These involuntary reasons for job separation were included in our subsampling conditions to avoid as much as possible a potential effect of divorce as a trigger for job loss. Some of the voluntary job separations include:

- Caring for own children or elder relative(s).
- School.
- Found new job.
- Move to a new residence.

- Poor pay.

- Not enough hours of work.

- Too many hours of work.

An equivalent list was obtained for panel 2. Further details concerning the variables and methods of extracting data using SLIDRET v2.2 can be found at the SLID microdata user's guide, the SLID electronic data dictionary and the SLIDRET user's manual (see bibliography).

Bibliography

Bellhouse, D. R. and Stafford, J. E. (1999). Density estimation from complex surveys, *Statistica Sinica* **9**: 407–424.

Bellhouse, D. R., Goia, C. M. and Stafford, J. E. (2003). Graphical displays of complex survey data through kernel smoothing, in Chambers, R. and Skinner, C.J. (eds), *Analysis of Survey Data,* pp. 133–150. John Wiley.

Betensky, R. and Finkelstein, D. (1999). A non-parametric maximum likelihood estimator for bivariate interval censored data, *Statistics in Medicine* **18**: 3089–3100.

Bleuer, S. J. and Kratina, I. S. (2000). Some issues in the analysis of complex survey data, *Proceedings of the Survey Research Methods Section, American Statistical Association, Washington DC* pp. 734–739.

Blossfeld, H. and Mills, M. (2003). A causal approach to interrelated family events, *Proceedings of Statistics Canada Symposium 2002: Modelling Survey Data for Social and Economic Research*. Statistics Canada. 11-522-XIE, Ottawa.

Braun, J., Duchesne, T. and Stafford, J. (2005). Local likelihood density estimation for interval censored data, *The Canadian Journal of Statistics* **33**(1): 39–60.

Breunig, R. V. (2001). Density estimation for clustered data, *Econometric Reviews* **20**: 353–367.

Buskirk, T. and Lohr, S. (2005). Asymptotic properties of kernel density estimation with complex survey data, *Journal of Statistical Planning and Inference* **128**: 165–190.

Charles, K. K. and Stephens, M. J. (2004). Job displacement, disability and divorce, *Journal of Labour Economics* **22**: 489–522.

Cook, R. J., Zeng, L. and Lee, K.-A. (2007). A multistate model for bivariate interval censored failure time data, *Technical Report. Department of Statistics and Actuarial Science. University of Waterloo.*

Duchesne, T. and Stafford, J. (2001). A kernel density estimate for interval censored data, *Technical Report No. 0106. University of Toronto.*

Faraway, J. J. and Jhun, M. (1990). Bootstrap choice of bandwidth for density estimation, *Journal of the American Statistical Association* **85**: 1119–1122.

Gentleman, R. and Geyer, C. (1994). Maximum likelihood for interval censored data: Consistency and computation, *Biometrika* **81**: 618–623.

Geskus, R. and Groeneboom, P. (1997). Asymptotically optimal estimation of smooth functionals for interval censoring, part 2, *Statistica Neerlandica* **51**: 201–219.

Geskus, R. and Groeneboom, P. (1999). Asymptotically optimal estimation of smooth functionals for interval censoring, case 2, *The Annals of Statistics* **27**: 627–674.

Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*, Birkhäuser, Boston.

Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems, *Journal of Multivariate Analysis* **32**: 177–203.

Härdle, W. (1990). *Smoothing Techniques with implementation in S.*, Springer.

Hjort, N. L. and Jones, M. C. (1996). Locally parametric nonparametric density estimation, *The Annals of Statistics* **24**: 1619–1647.

Huang, J. (1999). Asymptotic properties of nonparametric estimation based on partly interval-censored data, *Statistica Sinica* **9**: 501–519.

Huang, J. (2003). Unemployment and family behavior in Taiwan, *Journal of Family and Economic Issues* **24**: 27–48.

Huttenlocher, J., Hedges, L. V. and Bradburn, N. M. (1990). Reports of elapsed time: Bounding and rounding processes in estimation, *Journal of Experimental Psychology: Learning, Memory and Cognition* **16**: 196–213.

Isaki, C. and Fuller, W. (1982). Survey design under the regression superpopulation model, *Journal of the American Statistical Association* **77**: 89–96.

Jahoda, M., Lazarsfeld, P. F. and Zeisel, H. (1933). *Marienthal. The Sociography of an Unemployed Community*, Aldine Atherton Inc.

Larry, J.-L. T. and Catlin, G. (1995). Sampling design of the National Population Health Survey, *Health Reports. Statistics Canada. Catalogue 82–003.* **7**(1): 29–38.

Lawless, J. (2003a). Event history analysis and longitudinal surveys, in Chambers, R. and Skinner, C.J. (eds), *Analysis of Survey Data,* pp. 221–243. John Wiley.

Lawless, J. (2003b). *Statistical Models and Methods for Lifetime Data*, second edn, John Wiley and Sons.

Li, L., Watkins, T. and Yu, Q. (1997). An EM algorithm for smoothing the self-consistent estimator of survival functions with interval-censored data, *Scandinavian Journal of Statistics* **24**: 531–542.

Loader, C. R. (1996). Local likelihood density estimation, *The Annals of Statistics* **24**: 1602–1618.

Lohr, S. L. (1999). *Sampling: Design and Analysis*, Duxbury Press.

Naud, J.-F. (2004). Combined-panel longitudinal weighting. Survey of Labour and Income Dynamics 1996-2002, *Income research paper series*. Statistics Canada. Catalogue no. 75F0002MIE No. 008.

Neter, J. and Waksberg, J. (1964). A study of response errors in expenditures data from household interviews, *Journal of the American Statistical Association* **59**: 17–55.

Ortega, J. M. (1972). *Numerical Analysis: A Second Course*, Academic Press, New York.

Pantoja-Galicia, N. and Thompson, M. E. (2006). Examples of testing for temporal order using data from the National Population Health Survey and the Survey of Labour and Income Dynamics, *Technical Project Report. Statistics Canada.*

Peto, R. (1973). Experimental survival curves for interval-censored data, *Applied Statistics* **24**: 86–91.

Rao, J. N. K. and Wu, C. F. J. (1988). Resampling inference with complex survey data, *Journal of the American Statistical Association* **83**: 231–241.

Rao, J. N. K., Wu, C. F. J. and Yue, K. (1992). Some recent work on resampling methods for complex survey data, *Survey Methodology* **18**: 209–217.

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

Sheather, S. J. (2004). Density estimation, *Statistical Science* **19**: 588–597.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman-Hall, London.

Simonoff, J. S. (1996). *Smoothing Methods in Statistics*, Springer.

Statistics Canada (1996). *1994-95 NPHS Public Use Microdata Documentation*.

Statistics Canada (1997). *Survey of Labour and Income Dynamics Microdata User's Guide*. www.statcan.ca/english/freepub/75M0001GIE/free.htm.

Statistics Canada (2003). *Survey methods and practices*. Catalogue 12-587-XPE.

Statistics Canada (2004a). *SLID RETrieval Software "SLIDRET" Version 2.2. User's Manual*.

Statistics Canada (2004b). *Survey of Labour and Income Dynamics electronic data dictionary*. http://www.statcan.ca/english/IPS/Data/75F0026XIB.htm.

Statistics Canada (2005). *Survey of Labour and Income Dynamics (SLID) - A Survey Overview*. www.statcan.ca/english/freepub/75F0011XIE/free.htm.

Swain, L., Catlin, G. and Beaudet, M. P. (1999). The National Population Health Survey–its longitudinal nature, *Health Reports. Statistics Canada. Catalogue 82–003.* **10**(4): 69–82.

Tang, B. (1993). Orthogonal array-based latin hypercubes, *Journal of the American Statistical Association* **88**: 1392–1397.

Taylor, C. C. (1989). Bootstrap choice of the smoothing parameter in kernel density estimation., *Biometrika* **76**: 705–712.

Taylor, R. B., David, A. K. and Fields, S. A. (2003). *Fundamentals of Family Medicine: The Family Medicine Clerkship Textbook*, third edn, Springer.

Thompson, M. E. (1997). *Theory of Sample Surveys*, Chapman and Hall.

Thompson, M. E. and Pantoja-Galicia, N. (2003). Interval censoring of smoking cessation in the National Population Health Survey, *Proceedings of Statistics Canada Symposium 2002:*

Modelling Survey Data for Social and Economic Research. Statistics Canada. 11-522-XIE, Ottawa.

Tourangeau, R., Rips, L. J. and Rasinski, K. (2000). *The Psychology of Survey Response*, Cambridge University Press, New York.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data, *Journal of the Royal Statistical Society. Series B (Methodological* **38**: 290–295.

Wand, M. and Jones, M. (1995). *Kernel Smoothing*, Chapman and Hall.

Wong, G. and Yu, Q. (1999). Generalized mle of a joint distribution function with multivariate interval-censored data, *Journal of Multivariate Analysis* **69**: 155–166.

Yeung, W. J. and Hofferth, S. L. (1998). Family adaptations to income and job loss in the U.S., *Journal of Family and Economic Issues* **19**: 255–283.

Yu, Q., Schick, A., Li, L. and Wong, G. (1998). Asymptotic properties of the gmle of a survival function with case 2 interval-censored data, *Statist. Prob. Lett.* **37**: 223–228.

Yung, W. (1997). Variance estimation for public use files under confidentiality constraints, *Statistics Canada*.