ABSTRACT

Title of Dissertation: NONPARAMETRIC ESTIMATION OF A DISTRIBUTION FUNCTION IN BIASED SAMPLING MODELS

Jian-Lun Xu, Doctor of Philosophy, 1993

Dissertation directed by: Grace L. Yang
Professor
Department of Mathematics

The nonparametric maximum likelihood estimator (NPMLE) of a distribution function F in biased sampling models have been studied by Cox (1969), Vardi (1982, 1985), and Gill, Vardi, and Wellner (1988). Their approaches are based on the assumption that the observations are drawn from biased distributions of F and biasing functions do not depend on F. These assumptions have been used in Patil and Rao (1978).

This thesis extends the biased sampling model by making the biasing functions depend on the distribution function F in a variety of ways. With this extension, many of the existing models, including the ranked-set sampling model and the nomination sampling model, become special cases of the biased sampling model. The statistical inference about F becomes to a large extent the study of

the biasing function. We develop conditions under which the generalized model is identifiable. Under these conditions, an estimator of the underlying distribution F is proposed and its strong consistency and asymptotic normality are established.

In certain situation, estimation of F in a biased sampling model is in fact a problem of estimating a monotone decreasing density. Several density estimators are studied. They include the nonparametric maximum likelihood estimator, a kernel estimator, and a modified histogram type estimator. The strong consistency, the asymptotic normality, and the bounds on average error for the estimators are studied in detail.

In summary, this thesis is a generalizations of the estimation results available for the ordinary s-biased sampling model, the ranked-set sampling model, the nomination sampling model, and a monotone decreasing density.

NONPARAMETRIC ESTIMATION OF A DISTRIBUTION

FUNCTION IN BIASED SAMPLING MODELS

by

Jian-Lun Xu

Dissertation submitted to the Faculty of the Graduate School
of The University of Maryland in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
1993

Advisory Committee :

Professor Grace L. Yang Chairman/Advisor
Professor Eric V. Slud
Associate Professor Paul J. Smith
Assistant Professor Issa Fakhre-Zakeri
Associate Professor Estelle Russek-Cohen

DEDICATION

To my family

ACKNOWLEDGMENTS

I would like to express my sincere appreciation and thanks to my advisor, Professor Grace L. Yang for her most excellent guidance, advice, discussions and continued encouragement through the course of this work, and for giving me the opportunity to work in the United States Department of Agriculture. My deepest gratitude is also extended to Professor Paul J. Smith for his useful advice, discussions and encouragement. I also thank Professor Eric V. Slud and Professor Issa Fakhre-Zakeri for their helpful suggestions and discussions on the writing of this dissertation. I would also like to thank Professor Estelle Russek-Cohen for acting as reader of this thesis.

I would like to thank my family for their unlimited support and constant encouragement during my Ph.D study in this country. Without their considerateness and understanding, this work would not have been possible.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

1.1 Literature Review

Let F be the unknown cumulative distribution function (cdf) of a real-valued random variable (rv) X defined on some probability space. We study the nonparametric estimation of F based on random samples that contain biased or constrained observations. Biased or constrained data occur naturally in reliability studies and survival analysis. The following are two frequently encountered situations:

(i) Biased sampling model. Suppose that $w_1(y), \ldots, w_s(y)$ are known non-negative weights, called biasing functions, and F is an unknown cdf. The biased cdf G_i corresponding to F is defined by

$$G_i(x) = \frac{1}{W_i} \int_{-\infty}^{x} w_i(y) dF(y), \quad x \in R^1, \quad i = 1, \ldots, s, \tag{1.1.1}$$

with

$$0 < W_i = \int_{-\infty}^{\infty} w_i(y) dF(y) \overset{def}{\equiv} [F](w_i) < \infty, \quad i = 1, \ldots, s, \tag{1.1.2}$$

where $[F](w_i) \overset{def}{\equiv} \int_{-\infty}^{\infty} w_i(y) dF(y)$ is a functional. *This notation will be used frequently in this thesis for simplicity.* Inference about F is based on s samples of n_i iid random variables, $X_{i1}, \ldots, X_{in_i}$ from the distribution G_i for $i = 1, \ldots, s$. The model (1.1.1) is called an s-biased sampling model. This model has been studied by Vardi (1985) who constructed the nonparametric maximum likelihood estimator (NPMLE) $\hat{F}^{(n)}$ of F. Gill, Vardi, and Wellner (1988) (to be abbreviated by G-V-W (1988)) have obtained the strong consistency and asymptotic normality of estimator $\hat{F}^{(n)}$. A special case of model (1.1.1) ($s = 2$, $w_1(y) \equiv 1$, $w_2(y) \equiv y$, $y \in [0, \infty)$) has been studied by Vardi (1982).

(ii) Density estimation under a constraint. There are certain situations in which the observation X is distorted by a random scale change to ZX, where Z has the uniform distribution over $(0, 1)$ and is independent of X. We assume the underlying random variable X is nonnegative. Let $Y \stackrel{d}{=} ZX$ and G be its cdf. Here "$\stackrel{d}{=}$" means equal in distribution. Then Y has density

$$g(y) = \frac{dG(y)}{dy} = \int_{u \geq y} \frac{1}{u} dF(u), \qquad y \geq 0, \tag{1.1.3}$$

where F is the cdf of X. The problem is to use the data $Y_1, \ldots, Y_n$ from density $g(y)$ to estimate the distribution F. It is easy to check that

$$F(t) = G(t) - tg(t), \quad f(t) = -tg'(t)$$

for $t \in [0, \infty)$, where g is subject to the constraint

$$g'(t) \leq 0, \quad t \in [0, \infty). \tag{1.1.4}$$

In this case, nonparametric estimation of F is equivalent to estimating G and its density g. It is well known that the NPMLE of G under the constraint $g'(t) \leq 0$ is the least concave majorant of the empirical distribution function of $Y_1, \ldots, Y_n$. There are many papers discussing this problem, including Grenander (1956), Prakasa-Rao (1969), Groeneboom (1985), Devroye (1987), Birgé (1987a, b, 1989), and Datta (1992). Recently, Vardi (1989) has generalized this problem from the one sample to the two-sample case in which $(Y_1, \ldots, Y_n)$ is a random sample from G and $(X_1, \ldots, X_m)$ is another random sample from F, and obtained $\hat{F}^{(n)}$, the NPMLE of F. Asymptotic properties of the estimator $\hat{F}^{(n)}$ have been studied in Vardi and Zhang (1992).

In this thesis, we extend the s-biased sampling model (1.1.1) by making the weight functions $w_i(y)$ depend on the population distribution function F in a variety of ways, and we consider the biased sampling model in a substantially more general setting. Also we investigate density estimation under more general constraints than (1.1.4). These generalizations are motivated by some important practical applications to be presented in the next section.

1.2 Some Applications Leading to New Models

Example 1.2.1. The perfect ranked-set-sampling model. The construction of the perfect ranked-set-sampling model (RSS) consists of n cycles. In each cycle the experimenter selects s random samples, each of size s, where s is a predetermined number. For each sample of s observations in the ith cycle, the rth smallest observation, $X_{[r]i}$, is determined and the rest of the $s-1$ observations are discarded. At the conclusion of n cycles, only ns out of the total of ns^2 observations are retained. The retained ns values are certain order statistics $\{X_{[r]i}; r = 1, \ldots, s; i = 1, \ldots, n\}$. Let F be population cdf of the basic rv X. It is easy to see that the cdf of $X_{[r]i}$ is given by

$$G_r(x) = \frac{1}{W_r} \int_{-\infty}^{x} [F(y)]^{r-1}[1-F(y)]^{s-r} dF(y), \quad x \in R^1 \tag{1.2.1}$$

where

$$W_r = \frac{(r-1)!(s-r)!}{s!} = \frac{1}{s\binom{s-1}{r-1}}$$

for $r = 1, \ldots, s$. Comparing with (1.1.5), we see that $G_r(x)$ is a biased distribution function of F with weight function $w_r(y) = [F(y)]^{r-1}[1 - F(y)]^{s-r}$ depending on the population distribution F.

The application of the RSS procedure dates back to 1952. McIntyre (1952) used the procedure to estimate mean pasture yields. Measuring yields of pasture plots requires mowing and weighing the hay, a time-consuming process. But an experienced eye can rank fairly accurately a small number of plots without measurement. The RSS procedure has been applied mostly in agriculture (e.g., Cobby, Ridout, Bassett, and Large 1985; Halls and Dell 1966). Other related studied model can be found in Dell and Clutter (1972), Takahasi and Wakimoto (1968), and Stokes (1977, 1980). Stokes and Sager (1988) carried out an interesting study in which F is estimated with data from s populations $G_1, \ldots, G_s$.

Example 1.2.2. Nomination-sampling model. In the nomination sampling model, we observe not X, but the extreme value of X_i's, either

$$Y = \max(X_1, \ldots, X_k) \tag{1.2.2}$$

or

$$Z = \min(X_1, \ldots, X_k), \tag{1.2.3}$$

where X_i's are iid with cdf F, and k is a fixed integer. Then the cdfs of Y and

Z are respectively

$$(1.2.4) \qquad G_Y(x) = [F(x)]^k = \int_{-\infty}^{x} k[F(y)]^{k-1} dF(y),$$

$$(1.2.5) \qquad G_Z(x) = 1 - [1 - F(x)]^k = \int_{-\infty}^{x} k[1 - F(y)]^{k-1} dF(y).$$

We see that the cdfs G_Y and G_Z are also biased distributions of F with weight functions $w(y) = [F(y)]^{k-1}$ and $w(y) = [1 - F(y)]^{k-1}$, respectively. Again, the weight functions depend on F. For applications, Willemain (1980) used the $Y-$data to estimate the median of F. Boyles and Samaniego (1986) generalized model (1.2.2) to the case in which k is a rv, and studied the NPMLE of F and its properties.

Example 1.2.3. (Gill, Vardi, and Wellner 1988). If we have control over the choice of the number of samples s, the sample fractions $\lambda_{ni} = n_i/n, \quad i = 1, \ldots, s$, and the known weight functions $w_i(y)$, we may want to choose them to obtain an optimal estimator, such as a minimum variance estimator. For example, if we take $s = 1$ and want to estimate the population mean $\mu_F = \int_{-\infty}^{\infty} y dF(y)$ from the biased sampling distribution, then the optimal choice for the weight function is $w(y) = |y - \mu_F|$. This example shows that $w(y)$ depends on the population distribution F through μ_F.

Example 1.2.4. Backward sampling of ages (He and Yang 1993). Let the target population $\mathcal{P}$ for sampling be all those individuals who were born in a specified time interval $[0, T]$ and whose death times $(d = \tau + X)$ are at least as large as T. Here τ is the birth time and X is the lifetime of an individual. Let $B(t)$ be the cdf of birth times τ of all those individuals born in $[0, T]$. A variety of sampling methods can be designed to obtain data for estimating the cdf F of the lifetime X. One problem addressed by He and Yang (1993) is to use the age data to estimate F. The data consists of the age of n individuals taken from $\mathcal{P}$. The age, Y, of an individual taken from $\mathcal{P}$ has the cdf

$$\begin{aligned} G(y) &= P(Y \le y) = P(T - \tau \le y | d > T) \\ &= \frac{1}{c} \int_{T-y}^{T} \overline{F}(T-u) dB(u), \end{aligned}$$

for $0 \le y \le T$, where

$$c = P(X + \tau > T).$$

Evidently, F and B cannot be determined by the knowledge of G alone. Suppose that B is known. Then we can identify F for $x \in [0, T]$. In human populations, this is a reasonable assumption since birth records are available. Under the assumptions that $B(t)$ possesses a strictly positive pdf on $[0, T]$, $\beta_F = \inf\{x : F(x) = 1\} > T$, and that X and τ are independent, we have

$$\text{(1.2.6)} \qquad \overline{F}(x) = P[X > x] = \frac{b(T)}{b(T-x)} \frac{g(x)}{g(0)}, \quad x \in [0, T]$$

where g denotes the pdf of G.

It is interesting to note that in this case estimating $\overline{F}$ is equivalent to estimating the density $g(x)$ under the constraint that $g(x)/b(T-x)$ is a monotone decreasing function of $x \in [0, T]$. This is similar to (1.1.8), but the constraint is that $g(x)/b(T-x)$, ratio of two densities $g(x)$ and $b(T-x)$, is monotone decreasing. More examples will be given in Chapter V.

Motivated by these applications, we proceed to formulate theoretically five biased sampling models that generalize models (1.1.1), (1.1.3), and the constraint (1.1.4).

1.3 New Models and Problems to Be Studied

This thesis treats the following five classes of general s-biased sampling models.

Model I.

$$(1.3.1) \qquad G_i(x) = \frac{1}{B_i}\int_{-\infty}^{x} w_i(y)h(F(y))dF(y), \quad x \in R^1, \quad i = 1, \ldots, s,$$

where

$$\begin{aligned} 0 < B_i &\equiv \int_{-\infty}^{\infty} w_i(y)h(F(y))dF(y) \\ &= \int_{-\infty}^{\infty} w_i(y)dH(F(y)) \\ &\stackrel{def}{=} [H \circ F](w_i) < \infty, \quad i = 1, \ldots, s, \end{aligned}$$

and $h(z)$ is a known, nonnegative and integrable function on $[0,1]$ such that

$$H(x) - H(0) = \int_0^x h(z)dz,$$

and $w_1(y), \ldots, w_s(y)$ are known nonnegative measurable biasing functions.

This model generalizes models (1.1.1), (1.2.4) and (1.2.5). By setting $h(z) \equiv C(\text{constant}) > 0$ for $0 \le z \le 1$, model (1.1.1) is clearly a special case of (1.3.1). Likewise setting $s = 1$, $w_1(y) \equiv C(\text{constant}) > 0$, and $h(z) = z^{k-1}$ for $0 \le z \le 1$, we obtain the nomination sampling model (1.2.4); and setting $s = 1$, $w_1(y) \equiv C(\text{constant}) > 0$, and $h(z) = (1-z)^{k-1}$ for $0 \le z \le 1$, model (1.3.1) becomes the nomination sampling model (1.2.5).

Model II.

$$
(1.3.2) \qquad
\begin{aligned}
G_i(x) &= \frac{1}{W_i}\int_{-\infty}^{x} w_i(F(y))dF(y), \quad x \in R^1, \quad i = 1, \ldots, s, \\
0 < W_i &\equiv \int_{-\infty}^{\infty} w_i(F(y))dF(y) < \infty, \quad i = 1, \ldots, s,
\end{aligned}
$$

where $w_1(z), \ldots, w_s(z)$ are known nonnegative measurable biasing functions defined on $[0,1]$.

The perfect ranked-set sampling model (1.2.1) and the nomination sampling model (1.2.4) are special cases of model (1.3.2) by setting respectively in (1.3.2) $w_i(z) = z^{i-1}(1-z)^{s-i}$, $1 \le i \le s$, and $s = 1$, $w_1(z) = z^{k-1}$ for $0 \le z \le 1$.

Model III. This model is described by the joint distribution of two random variables X and K,

$$
(1.3.3) \qquad
\begin{aligned}
G(x,k) &= P[X \le x, K = k] \\
&= p(k)\int_{-\infty}^{x} \frac{w_k(y)dF(y)}{W_k}, \quad x \in R^1; \\
0 < W_k &= \int_{-\infty}^{\infty} w_k(x)dF(y) < \infty, \\
p(k) &= P(K = k), \quad k \in \mathcal{K},
\end{aligned}
$$

where $\{w_k(x) : k \in \mathcal{K}\}$ are assumed known, the probabilities $\{p(k) : k \in \mathcal{K}\}$ and cdf F are unknown. The observed values are independent pairs $\{(X_i, K_i)\}_{i=1}^{n}$.

In particular, if $P(K = 1) \equiv 1$, then model (1.3.3) becomes the 1-biased sampling model. In fact, model (1.3.3) considers the number of people who contribute the data as a random variable.

Model IV. Density estimation under the constraint:

$$
(1.3.4) \qquad \left[\frac{f(x)}{w(x)}\right]' \le 0, \qquad x \in (0, M)
$$

where $M \leq \infty$, and $w(x)$ is a known and strictly positive weight function.

When $w(x) = C > 0$ for $x \in (0, \infty)$, the constraint specializes to the well known problem of estimating a monotone decreasing pdf f; when $w(x) = b(x-T)$, $M = T$, (1.3.4) becomes the problem described in (1.2.6).

Similar to (1.3.4) we have the following model:

Model V. Density estimation under the constraint

$$\left[\frac{f(x)}{w(x)}\right]' \geq 0, \qquad x \in (0, \beta_F) \tag{1.3.5}$$

where $\beta_F = \inf\{x : F(x) = 1\}$, and $w(x)$ is a known and strictly positive weight function.

Special cases of this model, including the relevation transform (see Chapter VI) of two cdfs, the deconvolution model, and the mixture model will be studied in Chapter VI.

1.4 Research Summary

The dissertation is organized as follows:

In Chapter II, we discuss the estimation of cdf F in model (1.3.1). We first address the identifiability of model (1.3.1) under appropriate assumptions. We show that the model is identifiable by solving the equation (1.3.1) for F. Based on the solution of model (1.3.1), we propose a natural estimator of F in Section 2.2. Strong consistency and asymptotic normality of the estimator are presented in Section 2.3. Proofs of the large sample properties are given in Section 2.4. Some examples are considered in Section 2.5.

In Chapter III, we discuss model (1.3.2). The identifiability of the model and the estimation of F will be addressed in Section 3.2. Some properties of the estimator will be studied in Section 3.3.

In Chapter IV, the identifiability of model (1.3.3) and estimation of F are discussed. In particular, we show that our estimator is in fact the NPMLE of F under model (1.3.3).

In Chapter V, several nonparametric estimators will be considered for model (1.3.4). Specifically, we give five examples which motivate our study of model (1.3.4) in Section 5.1. The NPMLE of f and its properties are studied in Section 5.2. The kernel estimators of f and their properties are investigated in Section 5.3. The modified histogram type estimator of f and its properties are considered in Section 5.4.

In Chapter VI, a discussion of model (1.3.5) is presented, organized parallel to Chapter V.

Finally, the Appendix contains a list of notation and three definitions used in this thesis.

CHAPTER II

ESTIMATING THE DISTRIBUTION FUNCTION IN BIASED SAMPLING MODEL I

2.1 Introduction

In this chapter, we study the statistical inference of the general s-biased sampling model I defined by (1.3.1)

$$
(2.1.1) \qquad
\begin{aligned}
G_i(x) &= \int_{-\infty}^{x} \frac{w_i(y)h(F(y))dF(y)}{B_i}, \quad x \in R^1, \quad i = 1, \ldots, s, \\
0 < B_i &= \int_{-\infty}^{\infty} w_i(x)h(F(y))dF(y) < \infty, \quad i = 1, \ldots, s
\end{aligned}
$$

in Chapter I, where h is a known and nonnegative function on $[0,1]$ such that

$$H(x) - H(0) = \int_0^x h(z)dz \leq \infty,$$

$w_1, \ldots, w_s$ are known and nonnegative functions, and F is an (unknown) cdf. As in the biased sampling model (1.1.1) (see Vardi 1985), it is assumed that the observations are not from F, but from distribution $G_i, \quad i = 1, \ldots, s$. The s independent samples are denoted by

$$(2.1.2) \qquad X_{i1}, \ldots, X_{in_i} \qquad \text{iid from} \quad G_i.$$

We want to use all of the $n = n_1 + \ldots + n_s$ observations to estimate the underlying cdf F efficiently and nonparametrically, and to find a bias-corrected estimator which corrects for the biasing involved in the distributions G_i, $1 \leq i \leq s$.

This chapter is organized as follows: We first consider identifiability and estimator of F for model (2.1.1). The identifiability is proved by solving for

F in equation (2.1.1). Based on the solution of (2.1.1), we propose a natural estimator of F in Section 2.2. In Section 2.3 we state the large sample properties of strong consistency and asymptotic normality of the estimator. The proofs of these results are given in Section 2.4. Examples are discussed in Section 2.5.

2.2 Identifiability and Estimation of F

For nonparametric inference, it is necessary to show that model (2.1.1) is identifiable. That is, we need to show that the system of equations (2.1.1) has a unique solution for F in terms of $G_1, \ldots, G_s$. We make the following three assumptions. The first two are introduced by G-V-W (1988).

Let $\mathcal{X}^+ = \bigcup_{i=1}^s \{x : w_i(x) > 0\}$ and let $\mathcal{X} = (-\infty, \infty)$ be the sample space of the random variables $\{X_{ij}\}$. Meanwhile, we assume that the biasing functions $\{w_i(x)\}_{i=1}^s$ are distinct to avoid trivia. Then

Assumption 1. $\mathcal{X}^+ = \mathcal{X}$.

Remark 2.2.1. If this assumption fails, we must replace the cdf $F(x)$ by the conditional distribution function $F^+(x) = P(X \leq x | \mathcal{X}^+)$.

The next assumption requires the concept of a connected graph. We say the graph $\mathcal{F}$ with points $\{w_1(.), \ldots, w_s(.)\}$ is connected by a path if for each pair (i, j) there exist indices $l_1, \ldots, l_k \in \{1, \ldots, s\}$ such that

$$i \equiv l_1 \leftrightarrow l_2 \leftrightarrow \cdots \leftrightarrow l_k \equiv j, \quad \text{for} \quad 1 \leq i, j \leq s,$$

where $l_1 \leftrightarrow l_2$ means that

$$\int_{-\infty}^{\infty} I_{[w_{l_1}(x)>0]} I_{[w_{l_2}(x)>0]} dH[F(x)] > 0, \quad \text{for} \quad 1 \leq l_1, l_2 \leq s.$$

Assumption 2. The graph $\mathcal{F}$ with points $\{w_1(.), \ldots, w_s(.)\}$ is connected.

Under Assumptions 1 and 2 G-V-W (1988) showed that the ordinary s-biased sampling model ($h(z) = C$) is identifiable. However, we give an example to show that Assumptions 1 and 2 are not sufficient for the more general model (2.1.1) to be identifiable. Some restrictions on the functions h or H are necessary.

Example 2.2.1. Suppose that $s = 1$, $w(y)$ is a positive constant, and $h(z) = I_{[0,\frac{1}{2}]}(z)$ for $z \in [0,1]$. Then

$$\begin{aligned} G(x) &= \frac{\int_{-\infty}^{x} I_{[0 \le F(y) \le \frac{1}{2}]} dF(y)}{\int_{-\infty}^{\infty} I_{[0 \le F(y) \le \frac{1}{2}]} dF(y)} \\ &= 2 \int_0^{F(x)} I_{[0 \le u \le \frac{1}{2}]} du \\ &= 2 \min(F(x), 1/2) = \min(2F(x), 1) \\ &= \begin{cases} 1, & \text{if } F(x) > \frac{1}{2}; \\ 2F(x), & \text{if } F(x) \le \frac{1}{2}. \end{cases} \end{aligned}$$

Thus G cannot determine the unknown F when $F(x) > \frac{1}{2}$.

This example shows that additional assumption is needed.

Assumption 3. The inverse function $H^{-1}(z)$ of $H(z)$ from $[H(0), H(1)]$ to $[0,1]$ exists, where $H(z) - H(0) = \int_0^z h(y)dy \le \infty$ for any $z \in [0,1]$.

If $h(z)$ is strictly positive on $[0,1]$, then Assumption 3 holds. On the other hand, without loss of generality, we can assume that $H(0) = 0$, $H(1) = 1$ under Assumption 3. In fact, we can make a transformation $H^*(x) = [H(x) - H(0)]/[H(1) - H(0)]$ so that $H^*(0) = 0$ and $H^*(1) = 1$ always hold.

Theorem 2.2.1. Under Assumptions 1, 2 and 3, model (2.1.1) is identifiable.

Proof. By Assumption 3, model (2.1.1) can be written as

$$G_i(x) = \int_{-\infty}^{x} \frac{w_i(y)h(F(y))dF(y)}{B_i} = \int_{-\infty}^{x} \frac{w_i(y)}{B_i} dH(F(y))$$
$$B_i = \int_{-\infty}^{\infty} w_i(y)dH(F(y)),$$

for $i = 1, \ldots, s$. Denote the sample fractions of $G_1, \ldots, G_s$ by

$$\lambda_{ni} = \frac{n_i}{n}, \quad 1 \leq i \leq s,$$

We introduce an average distribution $\overline{G}_n(x)$ of s cdfs $G_1(x), \ldots, G_s(x)$ as follows:

$$\overline{G}_n(x) = \sum_{i=1}^{s} \lambda_{ni} G_i(x) = \int_{-\infty}^{x} \left[\sum_{i=1}^{s} \frac{\lambda_{ni} w_i(y)}{B_i} \right] dH[F(y)]. \tag{2.2.1}$$

It follows from Assumption 1 that the reciprocal of the integrand is finite and

$$\begin{aligned} H[F(x)] &= \int_{-\infty}^{x} \left[\sum_{i=1}^{s} \frac{\lambda_{ni} w_i(y)}{B_i} \right]^{-1} d\overline{G}_n(y), \\ 1 &= \int_{-\infty}^{\infty} \left[\sum_{i=1}^{s} \frac{\lambda_{ni} w_i(y)}{B_i} \right]^{-1} d\overline{G}_n(y). \end{aligned} \tag{2.2.2}$$

By Assumption 3, we see that H^{-1} exists and thus

$$\begin{aligned} F(x) &= H^{-1} \left\{ \frac{\int_{-\infty}^{x} \left[\sum_{i=1}^{s} \frac{\lambda_{ni} w_i(y)}{B_i} \right]^{-1} d\overline{G}_n(y)}{\int_{-\infty}^{\infty} \left[\sum_{i=1}^{s} \frac{\lambda_{ni} w_i(y)}{B_i} \right]^{-1} d\overline{G}_n(y)} \right\} \\ &= H^{-1}[T_n(x)]. \end{aligned} \tag{2.2.3}$$

Now let us normalize B_i with respect to B_s by defining $V_i = B_i/B_s$ for $i = 1, \ldots, s$. Thus $V_s \equiv 1$. The ratio of integrals in H^{-1} can be expressed in terms of the V_i as

$$T_n(x) = \frac{\int_{-\infty}^{x} \left[\sum_{i=1}^{s} \frac{\lambda_{ni} w_i(y)}{V_i} \right]^{-1} d\overline{G}_n(y)}{\int_{-\infty}^{\infty} \left[\sum_{i=1}^{s} \frac{\lambda_{ni} w_i(y)}{V_i} \right]^{-1} d\overline{G}_n(y)}. \tag{2.2.4}$$

Now proving the model is identifiable is to show for all λ_{ni}, $i = 1, \ldots, s$, $\lambda_{ni} > 0$, and $\sum_{i=1}^{s} \lambda_{ni} = 1$ that $V_1, \ldots, V_{s-1}$, V_s can be uniquely determined as functions of $G_1, \ldots, G_s$. We shall use the technique of Proposition 1.1 in G-V-W (1988). To make the thesis self-contained, a detailed proof will be provided.

Since $V_s \equiv 1$, we only need to consider the system of $s-1$ equations that define B_i or V_i,

$$(2.2.5) \qquad L_{ni}(V_1, \ldots, V_{s-1}, 1) = 1, \qquad i = 1, \ldots, s-1,$$

where

$$(2.2.6) \qquad \begin{aligned} L_{ni}(V_1, \ldots, V_s) &= \frac{1}{V_i} \int_{-\infty}^{\infty} w_i(y) dH[F(y)] \\ &= \frac{1}{V_i} \int_{-\infty}^{\infty} w_i(y) \left[\sum_{j=1}^{s} \frac{\lambda_{nj} w_j(y)}{V_j} \right]^{-1} d\overline{G}_n(y). \end{aligned}$$

We shall reparametrize the arguments in the equations L_{ni} as follows:

Put

$$e^{z_j} \equiv \lambda_{nj}/V_j, \qquad j = 1, \ldots, s,$$

$$\underline{z}^T = (z_1, \ldots, z_s).$$

From now on, the bar underlying the symbol denotes a vector, and "T" denotes the transpose of a column vector. Set

$$K_{nj}(\underline{z}) \equiv \lambda_{nj} L_{nj}(\lambda_{n1} e^{-z_1}, \ldots, \lambda_{ns} e^{-z_s}) - \lambda_{nj},$$

$$D_n(\underline{z}) = \int_{-\infty}^{\infty} \log \left[\sum_{i=1}^{s} e^{z_i} w_i(y) \right] d\overline{G}_n(y) - \sum_{i=1}^{s} \lambda_{ni} z_i.$$

Then the system of equations in (2.2.5) becomes

$$(2.2.7) \qquad K_{ni}(z_1, \ldots, z_{s-1}, \log(\lambda_{ns})) = 0, \qquad i = 1, \ldots, s-1$$

and $\underline{K}_n^T = (K_{n1}, \ldots, K_{ns})$ is the gradient of $D_n(\underline{z})$. We show that under Assumption 2 of a connected graph, $D_n(\underline{z})$ is a strictly convex function of $z_1, \ldots, z_s$. In fact,

$$\begin{aligned} \underline{a}^T D_n^{(2)}(\underline{z})\underline{a} &= \int_{-\infty}^{\infty} \left\{ \sum_{i=1}^{s} a_i^2 p_i(y) - \left[\sum_{i=1}^{s} a_i p_i(y) \right]^2 \right\} d\overline{G}_n(y) \\ &= \int_{-\infty}^{\infty} \mathrm{Var}_y(a_I) d\overline{G}_n(y) \\ &\geq 0 \end{aligned} \tag{2.2.8}$$

for all scalars $\underline{a}^T = (a_1, \ldots, a_s) \in R^s$, where $\{p_i(y)\}_{i=1}^s$ are probability distributions of a rv a_I defined by

$$P(a_I = a_i) = p_i(y) = \frac{e^{z_i} w_i(y)}{\sum_{j=1}^{s} e^{z_j} w_j(y)} > 0, \quad y \in R^1, \qquad i = 1, \ldots, s.$$

Thus to show that system (2.2.7) has a unique solution, it suffices to show that the upper left $(s-1) \times (s-1)$ submatrix of $D_n^{(2)}(\underline{z})$ is positive definite or has rank $s-1$. To do this we argue that if Assumption 2 holds, then a strict inequality holds in (2.2.8) for all $\underline{a} \neq c\sqcap_s$ for $c \neq 0$, where $\sqcap_s^T = (1, \ldots, 1)$ is an $s \times 1$ vector with all components 1.

Suppose instead that equality holds in (2.2.8). Then

$$\mathrm{Var}_y(a_I) = 0, \qquad a.e. \quad (\overline{G}_n) \quad y, \tag{2.2.9}$$

that is on the set $A = \{y : \mathrm{Var}_y(a_I) = 0\}$, $\overline{G}_n(A) = 1$. If so, we show $a_i = a_j$ for $1 \leq i, j \leq s$ with $i \neq j$. Under Assumption 2, for each pair (i, j) there exist indices $i = l_1, l_2, \ldots, l_k = j$ such that

$$\int_{-\infty}^{\infty} I_{[w_{l_{m-1}}(x)>0]} I_{[w_{l_m}(x)>0]} dH[F(x)] > 0 \tag{2.2.10}$$

for $m = 2, \ldots, k$. Let

$$A_m = \{y : w_{l_{m-1}}(y) w_{l_m}(y) > 0\}, \qquad m = 2, \ldots, k.$$

In view of (2.2.2), (2.2.10) implies $H_F(A_m) \equiv [H \circ F](A_m) > 0$, $m = 2, \ldots, k$, and hence $\overline{G}_n(A_m) > 0$. Denote the intersection of A and A_m by A_m^*, $m = 2, \ldots, k$. Choose $y \in A_2^*$. Then $\mathrm{Var}_y(a_I) = 0$ and $p_{l_1}(y), p_{l_2}(y) > 0$ would force $a_i = a_{l_1} = a_{l_2}$ since $\underline{a}$ must be constant for the coordinates with $p_l(y) > 0$. Similarly, choose $y \in A_3^*$, we can conclude $a_{l_2} = a_{l_3}$. Continuing this process yields

$$a_i = a_{l_1} = \cdots = a_{l_k} = a_j.$$

Since the same argument holds for any pair i and j, it follows that (2.2.9) implies that $\underline{a} = c\sqcap_s$ for some c. From this we conclude that $D_n^{(2)}(\underline{z})$ has rank $s - 1$ and its upper left $(s-1) \times (s-1)$ submatrix is nonsingular for all $\underline{z}$. Thus the solution $V_i = B_i / B_s$ of (2.2.5) is unique. The existence of solution of equation (2.2.5) is given in Theorem 2 of Vardi (1985). We shall not prove it here. This completes the proof. ∎

Clearly, from the identifiability result and (2.2.3) we can easily estimate F if the B_i (or V_i) are known. We can use the empirical cdf $\mathbf{G}_n$ of all the observations from s samples to estimate the average df $\overline{G}_n$ of (2.2.1), where

$$\mathbf{G}_n(x) = \frac{1}{n} \sum_{i=1}^{s} \sum_{j=1}^{n_i} I_{(-\infty, x]}(X_{ij}) = \sum_{i=1}^{s} \lambda_{ni} \mathbf{G}_{ni}(x), \tag{2.2.11}$$

and

$$\mathbf{G}_{ni}(x) = \frac{1}{n_i} \sum_{j=1}^{n_i} I_{(-\infty, x]}(X_{ij}), \quad i = 1, \ldots, s, \tag{2.2.12}$$

for $x \in R^1$. Then replacing $\overline{G}_n$ on the right side of (2.2.3) by the empirical cdf $\mathbf{G}_n$ yields a nondecreasing function $\hat{F}_n^0$, i.e.,

$$\hat{F}_n^0(x) = H^{-1}[\hat{T}_n^0(x)], \tag{2.2.13}$$

where

$$\begin{aligned}\hat{T}_n^0(x) &= \frac{\int_{-\infty}^{x}\left[\sum_{i=1}^{s}\frac{\lambda_{ni}w_i(y)}{B_i}\right]^{-1} d\mathbf{G}_n(y)}{\int_{-\infty}^{\infty}\left[\sum_{i=1}^{s}\frac{\lambda_{ni}w_i(y)}{B_i}\right]^{-1} d\mathbf{G}_n(y)} \\ &= \frac{\int_{-\infty}^{x}\left[\sum_{i=1}^{s}\frac{\lambda_{ni}w_i(y)}{V_i}\right]^{-1} d\mathbf{G}_n(y)}{\int_{-\infty}^{\infty}\left[\sum_{i=1}^{s}\frac{\lambda_{ni}w_i(y)}{V_i}\right]^{-1} d\mathbf{G}_n(y)}.\end{aligned} \tag{2.2.14}$$

However, the B_i (and hence V_i) are unknown. We need to address the estimation of the $B_i = \int w_i(y)dH[F(y)]$ first. Once the B_i are properly estimated, we substitute them for the unknown B_i in (2.2.14) to obtain an estimate of F. In order to assume that the resulting estimate indeed has the properties of a distribution function, the following method of substitution is used. We substitute the empirical cdf $\mathbf{G}_n$ for $\overline{G}_n$ in (2.2.6) to obtain the following equations $(1 \le i \le s)$

$$\begin{aligned}\mathbf{L}_{ni}(V_1, \ldots, V_s) &= \frac{1}{V_i}\int_{-\infty}^{\infty} w_i(y)\left[\sum_{j=1}^{s}\frac{\lambda_{nj}w_j(y)}{V_j}\right]^{-1} d\mathbf{G}_n(y) \\ &= 1.\end{aligned} \tag{2.2.15}$$

Let $\mathbf{V}_{n1}, \ldots, \mathbf{V}_{n,s-1}, \mathbf{V}_{ns}(\equiv 1)$ denote the solution of (2.2.15). Replacing $\overline{G}_n$ and $V_1, \ldots, V_s(\equiv 1)$ on the right side of (2.2.3) by the empirical df $\mathbf{G}_n$ and $\mathbf{V}_{n1}, \ldots, \mathbf{V}_{ns}$ yields a nondecreasing function $\hat{F}_n$ which will be used as an estimator of F. That is,

$$\hat{F}_n(x) = \hat{F}_n^0(x; \mathbf{V}_{n1}, \ldots, \mathbf{V}_{ns})$$

$$= H^{-1}\left\{\frac{\int_{-\infty}^{x}\left[\sum_{i=1}^{s}\frac{\lambda_{ni}w_i(y)}{\mathbf{V}_{ni}}\right]^{-1}d\mathbf{G}_n(y)}{\int_{-\infty}^{\infty}\left[\sum_{i=1}^{s}\frac{\lambda_{ni}w_i(y)}{\mathbf{V}_{ni}}\right]^{-1}d\mathbf{G}_n(y)}\right\} \tag{2.2.16}$$

$$= H^{-1}[\hat{T}_n(x)].$$

Finally, with $\hat{F}_n$ we can estimate

$$\mathbf{B}_{ni} = \mathbf{V}_{ni}\mathbf{B}_{ns}, \qquad i = 1, \ldots, s-1, \tag{2.2.17}$$

$$\mathbf{B}_{ns} = \frac{1}{\int_{-\infty}^{\infty}\left[\sum_{i=1}^{s}\frac{\lambda_{ni}w_i(y)}{\mathbf{V}_{ni}}\right]^{-1}d\mathbf{G}_n(y)}. \tag{2.2.18}$$

Remark 2.2.2. The equations (2.2.5), (2.2.15), (2.2.17) and (2.2.18), are the same as (1.12), (1.14), (1.18) and (1.19) of G-V-W (1988).

It is important to note that the solution $\mathbf{V}_{n1}, \ldots, \mathbf{V}_{n,s-1}, \mathbf{V}_{ns}$ of (2.2.15) is not unique in general. To get a unique solution, an assumption similar to Assumption 2 is introduced.

Assumption 4. The graph $\mathcal{F}$ with points $\{w_k(.) : 1 \leq k \leq s\}$ is strongly connected by a path in the sense that for any pair of (i, j) there exist $l_1, \ldots, l_k \in \{1, \ldots, s\}$ such that

$$i \equiv l_1 \rightleftharpoons l_2 \rightleftharpoons \cdots \rightleftharpoons l_k \equiv j,$$

where $l_1 \rightleftharpoons l_2$ if and only if

$$\int_{-\infty}^{\infty} I_{[w_{l_1}(x)>0]} I_{[w_{l_2}(x)>0]} d\mathbf{G}_n(x) = \frac{1}{n}\sum_{i=1}^{s}\sum_{t=1}^{n_i} I_{[w_{l_1}(X_{it})w_{l_2}(X_{it})>0]} > 0$$

for $1 \leq l_1, l_2 \leq s$, where $\mathbf{G}_n$ is defined by (2.2.11).

Under Assumptions 1, 3, and 4, the solution $\mathbf{V}_{n1}, \ldots, \mathbf{V}_{n,s-1}, \mathbf{V}_{ns} \equiv 1$ of (2.2.15) is unique. The proof, which is similar to that of Theorem 2.2.1, will be omitted.

2.3 Asymptotic Properties of the Estimator $\hat{F}_n$

In this section, we state two main large sample results which are the strong consistency and asymptotic normality of the estimators $\underline{\mathbf{V}}_n^T = (\mathbf{V}_{n1}, \ldots, \mathbf{V}_{ns})$, $\underline{\mathbf{B}}_n^T = (\mathbf{B}_{n1}, \ldots, \mathbf{B}_{ns})$, and $\hat{F}_n$. We begin with the following lemma adapted from Dudley and Philipp (1983) for our purpose.

Lemma 2.3.1. If $\lambda_{ni} \to \lambda_i > 0$, $i = 1, \ldots, s$, then there exist a special construction of

$$\mathbf{X}_n^*(g) \equiv \sqrt{n}(\mathbf{G}_n(g) - \overline{G}_n(g)) = \sum_{i=1}^{s} \sqrt{\lambda_{ni}}\sqrt{n_i}(\mathbf{G}_{ni}(g) - G_i(g))$$

and a Gaussian process $\mathbf{X}^*$ on a common probability space $(\Omega, \mathcal{F}, P)$ satisfying

$$||\mathbf{X}_n^* - \mathbf{X}^*||_{\mathcal{G}} \equiv \sup_{g \in \mathcal{G}} |\mathbf{X}_n^*(g) - \mathbf{X}^*(g)| \xrightarrow{P} 0$$

as $n \to \infty$, where $\mathcal{G}$ is a Donsker class for all cdfs G_i, $i = 1, \ldots, s$, (see Definition A.1.3 in Appendix) and $\mathbf{X}^*$ is a mean zero Gaussian process with the following covariance function

$$\begin{aligned}
\mathrm{Cov}(\mathbf{X}^*(q_1), \mathbf{X}^*(q_2)) &= \sum_{i=1}^{s} \lambda_i \{[G_i](q_1 q_2) - [G_i](q_1)[G_i](q_2)\}; \\
&= [H \circ F](r^{-1} q_1 q_2) - [H \circ F](q_1 \underline{\widetilde{w}}^T) \Delta [H \circ F](q_2 \underline{\widetilde{w}}),
\end{aligned}$$

where $q_1, q_2 \in \mathcal{G}$ and

$$r(x) \equiv \left[\sum_{i=1}^{s} \lambda_i \frac{w_i(x)}{B_i}\right]^{-1} = \left[\sum_{i=1}^{s} \lambda_i \widetilde{w}_i(x)\right]^{-1};$$

$$\underline{\widetilde{w}}^T(x) = (\widetilde{w}_1(x), \ldots, \widetilde{w}_s(x)), \quad \Delta \equiv \mathrm{diag}(\lambda_1, \ldots, \lambda_s) > 0.$$

Before stating the main results, let us define a new measure

$$H_F \equiv H \circ F,$$

which was introduced in the proof of Theorem 2.2.1. Defining $b_i(x) = aw_i(x)$, $1 \le i \le s$, we can write model (2.1.1) in terms of H_F as

$$G_i(x) = \int_{-\infty}^{x} b_i(x) dH_F(x)/B_i, \quad i = 1, \ldots, s,$$

for $x \in R^1$. Hence H_F is equivalent to G in G-V-W (1988). All results for G in G-V-W (1988) hold for our new measure H_F.

The first theorem establishes the consistency of the estimates $\underline{\mathbf{V}}_n$ and $\underline{\mathbf{B}}_n$ for $\underline{V}^T = (V_1, \ldots, V_s)$ and $\underline{B}^T = (B_1, \ldots, B_s)$ given by (2.2.15), (2.2.17), and (2.2.18).

Theorem 2.3.1. (Strong consistency of $\underline{\mathbf{V}}_n$ and $\underline{\mathbf{B}}_n$). Suppose that Assumptions 1, 2, 3 and 4 hold, and

$$0 < B_i = \int_{-\infty}^{\infty} w_i(y) h(F(y)) dF(y) < \infty$$

for $i = 1, \ldots, s$. Then equations (2.2.15) have (with probability 1 as $n \to \infty$) a unique solution $\underline{\mathbf{V}}_n$ which satisfies

(2.3.1) $$\underline{\mathbf{V}}_n \xrightarrow{a.s.} \underline{V} = \underline{B}/B_s \quad \text{as} \quad n \to \infty,$$

and

(2.3.2) $$\underline{\mathbf{B}}_n \xrightarrow{a.s.} \underline{B} \quad \text{as} \quad n \to \infty.$$

The second theorem proves the consistency of both $\hat{F}_n$ and $\hat{F}_n^0$ [see (2.2.13) where B_i are known] as estimators of F uniformly over $\mathcal{Q}(F)$, which is the class of functions defined by

$$\mathcal{Q}(F) = \{q_e(x)I_C(x) : C \in \mathcal{C}\}, \tag{2.3.3}$$

where $q_e(x)$ is a fixed nonnegative function, $\mathcal{C}$ is a Vapnik-Čhervonenkis class of subsets of the sample space $\mathcal{X}$, and

$$[F](q_e) \equiv \int_{-\infty}^{\infty} q_e(x)dF(x) < \infty, \tag{2.3.4}$$

(see Definition A.1.1 in Appendix).

Theorem 2.3.2. (Strong consistency of $\hat{F}_n$ and $\hat{F}_n^0$). Suppose that Assumptions 1, 2, 3 and 4 hold, and

$$0 < B_i = \int_{-\infty}^{\infty} w_i(y)h(F(y))dF(y) < \infty$$

for $i = 1, \ldots, s$. Assume that the function $h(x)$ is bounded in the sense that there exist two positive constants m and M such that

$$0 < m \leq h(x) \leq M \quad \text{for} \quad 0 \leq x \leq 1. \tag{2.3.5}$$

Then

$$||\hat{F}_n - F||_{\mathcal{Q}(F)} \equiv \sup\{|[\hat{F}_n](q) - [F](q)| : q \in \mathcal{Q}(F)\} \xrightarrow{a.s.} 0 \tag{2.3.6}$$

as $n \to \infty$. Furthermore, if only Assumptions 1, 3 and 4 hold,

$$||\hat{F}_n^0 - F||_{\mathcal{Q}(F)} \equiv \sup\{|[\hat{F}_n^0](q) - [F](q)| : q \in \mathcal{Q}(F)\} \xrightarrow{a.s.} 0 \tag{2.3.7}$$

as $n \to \infty$.

The following theorems assert the give asymptotic normality of $\underline{\mathbf{V}}_n$, $\underline{\mathbf{B}}_n$ and $\hat{F}_n$.

Theorem 2.3.3. Suppose that Assumptions 1, 2, 3 and 4 hold. Then

$$\sqrt{n}(\hat{F}_n(t) - F(t)) \xrightarrow{d} Z^*(t), \quad t \in R^1 \tag{2.3.8}$$

where $Z^*(t)$ is the mean zero Gaussian process with the following covariance function

$$K(t_1, t_2) = \mathrm{Cov}(Z^*(t_1), Z^*(t_2)) = \frac{1}{h(F(t_1))h(F(t_2))} c(t_1, t_2),$$

where

$$\begin{aligned} c(t_1, t_2) &= \mathrm{Cov}(Z(I_{t_1}), Z(I_{t_2})) \\ &= [H_F]\Big(r[I_{t_1} - H_F(t_1)][I_{t_2} - H_F(t_2)]\Big) \\ &+ [H_F]\Big([I_{t_1} - H_F(t_1)]r\underline{\widetilde{w}}^T\Big)M^-[H_F]\Big([I_{t_2} - H_F(t_2)]r\underline{\widetilde{w}}\Big), \end{aligned} \tag{2.3.9}$$

$Z(I_x)$ is the limiting process of $\sqrt{n}(\hat{T}_n(x) - T(x))$, M^- is the $\{1,2\}$-generalized inverse of M in the sense that M^- satisfies $M^-MM^- = M^-$, $MM^-M = M$, and

$$\begin{aligned} I_x \equiv I_{(-\infty, x]}(z), &\quad M = \Delta^{-1} - A; \\ A \equiv [\overline{G}](r^2 \underline{\widetilde{w}}\underline{\widetilde{w}}^T) &\equiv \int_{-\infty}^{\infty} r^2(x)\underline{\widetilde{w}}(x)\underline{\widetilde{w}}^T(x) d\overline{G}(x); \\ \overline{G}(x) = \sum_{i=1}^{s} \lambda_i G_i(x), &\quad x \in R^1, \end{aligned} \tag{2.3.10}$$

and r and $\underline{\widetilde{w}}$ are defined in Lemma 2.3.1.

Theorem 2.3.4. (Asymptotic normality of $\underline{\mathbf{B}}_n$) If Assumptions 1, 2, 3 and 4 hold, then

$$\sqrt{n}[\underline{\mathbf{B}}_n - \underline{B}] \xrightarrow{d} K^{-}\mathbf{X}^*(r\underline{\widetilde{w}}) + \underline{B}Z_\alpha, \tag{2.3.11}$$

where

$$K = M\Delta B^{-1};$$

$$\underline{B}^T = (B_1, \ldots, B_s), \quad B \equiv \operatorname{diag}(B_1, \ldots, B_s) > 0;$$

$$Z_\alpha = -\mathbf{X}^*(r) - [H_F](r\underline{\widetilde{w}}^T)M^{-}\mathbf{X}^*(r\underline{\widetilde{w}}),$$

defined in G-V-W (1988).

Theorem 2.3.5. (Asymptotic normality of $\underline{\mathbf{V}}_n$) If Assumptions 1, 2, 3 and 4 hold, then

$$\begin{aligned} J^T\sqrt{n}[\underline{\mathbf{V}}_n - \underline{V}] &\xrightarrow{d} B_s^{-1}K_*^{-1}J^T\mathbf{X}^*(r\underline{\widetilde{w}}) \\ &\sim N_{s-1}(\underline{0}, \Sigma) \end{aligned} \tag{2.3.12}$$

where

$$\Sigma = K_*^{-1}C_*[K_*^{-1}]^T/B_s^2,$$

$$K_* = J^TKJ, \quad J^T = (I_{s-1}, \underline{0})_{s\times(s-1)}, \quad C_* = J^TCJ;$$

$$C = \operatorname{Cov}[\mathbf{X}^*(r\underline{\widetilde{w}}), \mathbf{X}^*(r\underline{\widetilde{w}})] = A - A\Delta A.$$

2.4 Proofs of the Asymptotic Results

By Remark 2.2.2. it is unnecessary to prove the asymptotic normality given in Theorems 2.3.4 and 2.3.5 since the proofs are the same as those of Propositions 2.2 and 2.3 of G-V-W (1988) except for a change of notation from W_s to B_s. We shall therefore prove the consistency only. Since $\hat{F}_n$ is a function of $\hat{T}_n$ (cf. (2.2.16)), an estimator of the ordinary s-biased sampling model, the result of G-V-W (1988) can be applied again.

Proof of Theorem 2.3.1.

The proof that $\underline{\mathbf{V}}_n \xrightarrow{a.s.} \underline{V}$ is the same as the first part of the proof of Proposition 2.1 of G-V-W (1988). So (2.3.1) holds. Under Assumption 1, by (2.2.2), (2.3.1) and the fact that $\mathbf{G}_n(x) \xrightarrow{a.s.} \overline{G}(x)$, $\mathbf{B}_{ns}$ given by (2.2.18) converges a.s. to

$$\frac{1}{\int_{-\infty}^{\infty}\left[\sum_{i=1}^{s}\frac{\lambda_i w_i(y)}{V_i}\right]^{-1} d\overline{G}(y)} = \frac{B_s}{\int_{-\infty}^{\infty}\left[\sum_{i=1}^{s}\frac{\lambda_i w_i(y)}{B_i}\right]^{-1} d\overline{G}(y)} = B_s.$$

Hence

$$\mathbf{B}_{ni} = \mathbf{V}_{ni}\mathbf{B}_{ns} \xrightarrow{a.s.} V_i B_s = B_i \quad \text{as} \quad n \to \infty$$

for $i = 1, \ldots, s$, so (2.3.2) holds. ∎

Proof of Theorem 2.3.3.

The quantity $\hat{T}_n(x) = [\hat{T}_n](I_x)$ appearing in (2.2.16) is a distribution function. Let

$$T(x) = \int_{-\infty}^{x}\left[\sum_{i=1}^{s}\lambda_i \frac{w_i(y)}{B_i}\right]^{-1} d\overline{G}(y) = [T](I_x)$$

for any $x \in R^1$. Applying Theorem 2.2 of G-V-W (1988), we have

$$\sqrt{n}\{[\hat{T}_n](I_x) - [T](I_x)\} \xrightarrow{d} Z(I_x),$$

where $Z(I_x)$ is a mean zero Gaussian process with the covariance function $c(t_1, t_2)$ given by (2.3.9). On the other hand, by (2.2.3) and (2.2.16)

$$\hat{F}_n(x) = H^{-1}(\hat{T}_n(x)),$$

$$F(x) = H^{-1}(T(x)).$$

Applying the δ-method we have

$$\begin{aligned}\sqrt{n}(\hat{F}_n(x) - F(x)) &\xrightarrow{d} \frac{1}{h(F(x))} Z(I_x) \\ &\equiv Z^*(x).\end{aligned}$$

That is, the limiting process of $\sqrt{n}(\hat{F}_n(x) - F(x))$ is a mean zero Gaussian process with covariance function

$$\begin{aligned} K(t_1, t_2) &= \text{Cov}(Z^*(t_1), Z^*(t_2)) \\ &= \frac{1}{h(F(t_1))h(F(t_2))} \text{Cov}(Z(I_{t_1}), Z(I_{t_2})) \end{aligned}$$

for any $t_1, t_2 \in R^1$. So the proof is complete. ∎

Proof of Theorem 2.3.2.

From Theorem 2.1 of G-V-W (1988), we know that

$$\begin{aligned} ||\hat{T}_n - T||_{\mathcal{Q}(H_F)} &= \sup_{q^* \in \mathcal{Q}(H_F)} |[\hat{T}_n](q^*) - [T](q^*)| \\ &\xrightarrow{a.s.} 0 \quad \text{as} \quad n \to \infty. \end{aligned} \tag{2.4.1}$$

Moreover, (2.2.16), (2.4.1) and the continuity of H^{-1} imply that

$$\hat{F}_n(x) = H^{-1}(\hat{T}_n(x)) \xrightarrow{a.s.} H^{-1}(T(x)) = F(x).$$

Thus, for a fixed function $q \in \mathcal{Q}(F)$, we have $\frac{q}{h \circ F} \in \mathcal{Q}(H_F)$ and

$$\begin{aligned} |[\hat{F}_n](q) - [F](q)| &= \left| \int_{-\infty}^{\infty} \frac{q(x)}{h(\hat{F}_n(x))} d\hat{T}_n(x) - \int_{-\infty}^{\infty} \frac{q(x)}{h(F(x))} dT(x) \right| \\ &\leq \left| \int_{-\infty}^{\infty} q(x) \left[\frac{1}{h(\hat{F}_n(x))} - \frac{1}{h(F(x))} \right] d\hat{T}_n(x) \right| \\ &+ \left| \int_{-\infty}^{\infty} \frac{q(x)}{h(F(x))} d[\hat{T}_n(x) - T(x)] \right| \\ &= \left\| \frac{h(F(x))}{h(\hat{F}_n(x))} - 1 \right\|_{\infty} [\hat{T}_n]\left(\frac{q}{h \circ F} \right) \\ &+ |[\hat{T}_n](\frac{q}{h \circ F}) - [T](\frac{q}{h \circ F})| \\ &\xrightarrow{a.s.} 0 \end{aligned}$$

as $n \to \infty$. The second equality holds because $h(F(x))/h(\hat{F}_n(x))$ is bounded by

(2.3.5). Therefore

$$\sup_{q\in\mathcal{Q}(F)} |[\hat{F}_n](q) - [F](q)| \le \left\| \frac{h(F(x))}{h(\hat{F}_n(x))} - 1 \right\|_\infty \sup_{q\in\mathcal{Q}(F)} [\hat{T}_n]\left(\frac{q}{h\circ F}\right)$$
$$+ \sup_{q\in\mathcal{Q}(F)} |[\hat{T}_n](\frac{q}{h\circ F}) - [T](\frac{q}{h\circ F})|$$
$$\xrightarrow{a.s.} 0$$

as $n \to \infty$. That is, (2.3.6) holds. The proof of (2.3.7) is similar and will be omitted. But it is worth noting that we do not need (2.3.1) for the proof. Instead we use

$$||\hat{T}_n^0 - T||_{\mathcal{Q}(F)} = \sup_{q\in\mathcal{Q}(F)} |[\hat{T}_n^0](q) - [T](q)| \xrightarrow{a.s.} 0$$

as $n \to \infty$ since $\hat{F}_n^0$ depends only on the B_i and not on $\mathbf{B}_{ni}$. ∎

2.5 Some Examples

We give some examples of model (2.1.1).

Example 2.5.1. (Ordinary s-biased sampling model). Let $h(z) = C > 0$ for $z \in [0,1]$. Then (2.2.2) reduces to the solution of the ordinary s-biased sampling model. Thus, the results of Section 2.3 are generalizations of G-V-W (1988).

Example 2.5.2. (Nomination sampling model for the maximum (1.2.2)). Let $s = 1$, $w_1(x) \equiv C > 0$ for $x \in R^1$ and $h(z) = [p + (q-p)z]^{k-1}$ for $z \in [0,1]$, $0 \le p < q \le 1$, $p + q = 1$. Then (2.2.3) becomes

$$F(x) = \frac{[(q^k - p^k)G(x) + p^k]^{1/k} - p}{q - p}$$

while (2.2.16) becomes

$$\hat{F}_n(x) = \frac{[(q^k - p^k)\mathbf{G}_n(x) + p^k]^{1/k} - p}{q - p}.$$

By (2.3.8) we obtain that $\sqrt{n}[\hat{F}_n(x) - F(x)]$ converges in law to a Gaussian process with mean zero and covariance function

$$K(s,t) = \frac{a^2}{h(F(s))h(F(t))} F(s)[1 - F(t)], \quad s \le t,$$

where

$$a = \frac{q^k - p^k}{k[q-p]}.$$

When $q = 1$, we obtain the results for nomination sampling model (1.2.2). That is,

$$\sqrt{n}[\hat{F}_n(x) - F(x)] \xrightarrow{d} Z(x);$$

$$\begin{aligned} K(s,t) &= \mathrm{Cov}(Z(s), Z(t)) \\ &= \left[\frac{1}{k}\right]^2 \frac{1}{[F(s)F(t)]^{k-1}} F(s)[1 - F(t)], \quad s \le t. \end{aligned}$$

Similar results can be obtained for model (1.2.4).

Example 2.5.3. (Nomination sampling model for the minimum (1.2.3)). Let $s = 1$, $w_1(x) \equiv C > 0$ for $x \in R^1$ and $h(z) = [1 - qz]^{k-1}$ for $z \in [0,1]$, $0 \le q \le 1$. Then (2.2.3) becomes

$$\begin{aligned} F(x) &= \frac{1 - [1 - akqG(x)]^{1/k}}{q}, \\ a &= \frac{1 - (1-q)^k}{kq}. \end{aligned}$$

Hence,

$$\hat{F}_n(x) = \frac{1 - [1 - akq\mathbf{G}_n(x)]^{1/k}}{q}.$$

By (2.3.8) we obtain that $\sqrt{n}[\hat{F}_n(x) - F(x)]$ converges in law to a Gaussian process with mean zero and covariance function

$$K(s,t) = \frac{a^2}{h(F(s))h(F(t))} F(s)[1 - F(t)], \quad s \le t.$$

When $q = 1$, we obtain the results for the nomination sampling model (1.2.2). That is,

$$\sqrt{n}[\hat{F}_n(x) - F(x)] \xrightarrow{d} Z(x);$$

$$\begin{aligned} K(s,t) &= \mathrm{Cov}(Z(s), Z(t)) \\ &= \left[\frac{1}{k}\right]^2 \frac{1}{[\overline{F}(s)\overline{F}(t)]^{k-1}} F(s)[1 - F(t)], \quad s \leq t. \end{aligned}$$

CHAPTER III

ESTIMATING THE DISTRIBUTION FUNCTION IN BIASED SAMPLING MODEL II

3.1 Introduction

In this chapter, we study statistical inference of the class of general s-biased sampling models defined by (1.3.2)

$$
\begin{aligned}
G_i(x) &= \frac{1}{W_i}\int_{-\infty}^{x} w_i(F(y))dF(y), \quad x \in R^1, \quad i = 1, \dots, s, \\
0 < W_i &= \int_{-\infty}^{\infty} w_i(F(y))dF(y) < \infty, \quad i = 1, \dots, s
\end{aligned}
\tag{3.1.1}
$$

in Chapter I, where $w_1, \dots, w_s$ are known nonnegative and measurable biasing functions, and F is an (unknown) absolutely continuous cdf. As in Chapter II, iid observations from F are not available. Instead we can only observe s independent samples:

$$
X_{i1}, \dots, X_{in_i} \quad \text{iid from} \quad G_i, \qquad i = 1, \dots, s,
\tag{3.1.2}
$$

We wish to use all of the $n = n_1 + \dots + n_s$ observations to estimate efficiently the underlying cdf F. This estimator corrects for the biasing involved in the distributions G_i, $1 \le i \le s$.

The perfect ranked-set-sampling model (1.2.1) and nomination sampling models (1.2.4) and (1.2.5) are special cases of (3.1.1).

This chapter is organized as follows: We first investigate the identifiability of model (3.1.1). It is followed by the construction of an estimator for F. This is given in Section 3.2. The strong consistency and asymptotic normality of the estimator are established. While the technical treatment is similar to that of Chapter II, however, we are unable to provide a unified presentation for both chapters.

3.2 Identifiability and Estimator of F

The model (3.1.1) is identifiable if given G_i, $i = 1, \ldots, s$, F can be uniquely determined by (3.1.1). Obviously, the answer depends on the biasing functions $w_1(z), \ldots, w_s(z)$. Example 2.2.1 shows that we have to impose conditions on the biasing functions $w_1(z), \ldots, w_s(z)$ to assure the identifiability of the model. Hence, the following assumptions are made throughout this chapter:

Let $\mathcal{Y}^+ = \bigcup_{i=1}^{s} \{y \in [0,1] : w_i(y) > 0\}$. Then

Assumption 1. $\mathcal{Y}^+ = [0,1]$.

As in Assumption 1 of Section 2.2, we must replace $F(x)$ by $F^+(x) = P(X \leq x | \mathcal{X}_+)$ if the above assumption fails, where $F(\mathcal{X}_+) = \mathcal{Y}^+$.

Let $\lambda_{ni} = n_i/n$, $1 \leq i \leq s$ be the sampling fractions of $G_1, \ldots, G_s$. If $\lim_{n\to\infty} \lambda_{ni} = \lambda_i > 0$ exist for $i = 1, \ldots, s$, then we write $\sum_{i=1}^{s} \frac{\lambda_i w_i(u)}{W_i} = h(u)$ and set

$$H(u) - H(0) = \int_0^u h(z)dz, \quad 0 \leq u \leq 1.$$

Similar to Chapter II, without loss of generality, we can assume that $H(1) \equiv 1$ and $H(0) \equiv 0$.

Assumption 2. For $i = 1, \dots, s$,

$$\lim_{n \to \infty} \lambda_{ni} = \lambda_i > 0,$$

exist.

Theorem 3.2.1. Under Assumptions 1 and 2, model (3.1.1) is identifiable.

Proof. Write (3.1.1) as

$$G_i(x) = \int_{-\infty}^{x} \frac{w_i(F(y))}{W_i} dF(y) = \int_0^{F(x)} \frac{w_i(u)}{W_i} du,$$
$$W_i = \int_{-\infty}^{\infty} w_i(F(y)) dF(y),$$

for $i = 1, \dots, s$. Then the average distribution of $G_1(x), \dots, G_s(x)$ is given by

$$\overline{G}(x) = \sum_{i=1}^{s} \lambda_i G_i(x) = \int_0^{F(x)} \left[\sum_{i=1}^{s} \frac{\lambda_i w_i(u)}{W_i} \right] du = H(F(x)). \tag{3.2.1}$$

Thus, it follows from Assumption 1 that (cf. (2.2.3))

$$F(x) = H^{-1}(\overline{G}(x)). \tag{3.2.2}$$

Although (3.2.2) shows that the solution for F involves $W_1, \dots, W_s$, they are in fact independent of F, since

$$W_i = \int_{-\infty}^{\infty} w_i(F(y)) dF(y) = \int_0^1 w_i(u) du \tag{3.2.3}$$

for $i = 1, \dots, s$. Furthermore H by definition is also independent of F. Thus, the solution (3.2.2) for F is what we are seeking.

Let us further illustrate the solution by examples 1.2.1 and 1.2.2.

Example 3.2.1. (Example 1.2.1 continued). Take $\lambda_{ni} = 1/s$, $1 \le i \le s$. Since

$$h(u) = \sum_{i=1}^{s} \frac{\lambda_i w_i(u)}{W_i} = \sum_{i=1}^{s} \binom{s-1}{i-1} u^{i-1} (1-u)^{s-i} \equiv 1,$$

then

$$H(u) - H(0) = \int_0^u dz = u,$$

$$H^{-1}(u) = u.$$

Then

$$F(x) = H^{-1}(\overline{G}(x)) = \overline{G}_n(x) = \frac{1}{s}\sum_{i=1}^{s} G_i(x).$$

This relation between F and $\{G_1, \dots, G_s\}$ can be used to estimate F. A heuristic estimator introduced by Stokes and Sager (1988) is in fact the estimator $\hat{F}_n(x) = (1/s)\sum_{i=1}^{s} \mathbf{G}_{ni}(x)$, where $\mathbf{G}_{ni}(x)$ is the empirical distribution of $X_{i1}, \dots, X_{in_i}$, $1 \le i \le s$.

Example 3.2.2. (Example 1.2.2 continued). We have

$$H(u) = \int_0^u kz^{k-1} dz = u^k$$

and

$$F(x) = H^{-1}(\overline{G}(x)) = [G_Y(x)]^{\frac{1}{k}}.$$

Similarly, when $w(z) = (1-z)^{k-1}$, we have

$$F(x) = 1 - [1 - G_Z(x)]^{\frac{1}{k}}.$$

Clearly (3.2.2) suggests an estimator $\hat{F}_n$ for F. Namely, take

$$\hat{F}_n(x) = H^{-1}[\mathbf{G}_n(x)], \tag{3.2.4}$$

where

$$\mathbf{G}_n(x) = \sum_{i=1}^{s} \lambda_{ni}\mathbf{G}_{ni}(x);$$

$$\mathbf{G}_{ni} = \frac{1}{n_i}\sum_{j=1}^{n_i} I_{(-\infty,x]}(X_{ij}), \quad 1 \le i \le s.$$

3.3 Asymptotic Properties of the Estimator $\hat{F}_n$

Our main task in this section is to establish consistency and asymptotic normality of the estimator $\hat{F}_n$ of F. But first note

Theorem 3.3.1. $EH[\hat{F}_n(x)] = H[F(x)]$ for any $x \in R^1$.

This result follows directly from (3.1.1) and (3.2.4). When H is a linear function, we conclude that $\hat{F}_n(x)$ is an unbiased estimator of $F(x)$. This is the case for the perfect ranked-set-sampling model. In general, $\hat{F}_n(x)$ is of course biased for F. However, we can show that $\hat{F}_n$ is an asymptotically unbiased estimator of F by the Lebesgue Dominated Convergence Theorem and the following Theorem 3.3.2.

Theorem 3.3.2. (Strong consistency of $\hat{F}_n$) For any $x \in R^1$,

$$\hat{F}_n(x) \xrightarrow{a.s.} F(x)$$

as $n \to \infty$.

This result also follows directly from (3.1.1) and (3.2.4), and the Glivenko theorem for the empirical distribution $\mathbf{G}_n$.

Theorem 3.3.3. (Asymptotic normality of $\hat{F}_n$)

$$\sqrt{n}[\hat{F}_n(x) - F(x)] \xrightarrow{d} Z(x), \quad x \in R^1,$$

as $n \to \infty$, where $Z(x)$ is a Gaussian process with zero mean and the following covariance function

$$\begin{aligned}
K(t_1,t_2) &= \mathrm{Cov}(Z(t_1),Z(t_2)) \\
&= \frac{[\underline{\lambda}^{\frac{1}{2}}]^T \mathrm{Cov}(\underline{Z}(t_1),\underline{Z}(t_2))[\underline{\lambda}^{\frac{1}{2}}]}{h(H^{-1}(\overline{G}(t_1)))h(H^{-1}(\overline{G}(t_2)))} \\
&= \frac{\sum_{i=1}^{s}\lambda_i c_i(t_1,t_2)}{h(H^{-1}(\overline{G}(t_1)))h(H^{-1}(\overline{G}(t_2)))}, \\
c_i(t_1,t_2) &= \mathrm{Cov}(Z_i(t_1),Z_i(t_2)) \\
&= G_i(\min(t_1,t_2)) - G_i(t_1)G_i(t_2),
\end{aligned}$$

for any $t_1, t_2 \in R^1$.

Proof. It is well known that

$$\sqrt{n_i}[\mathbf{G}_{ni}(x) - G_i(x)] \xrightarrow{d} Z_i(x)$$

as $n_i \to \infty$ for $i = 1,\ldots,s$, where Z_i is a Brownian bridge with zero mean and the covariance function

$$c_i(t_1,t_2) = \mathrm{Cov}(Z_i(t_1),Z_i(t_2)) = G_i(\min(t_1,t_2)) - G_i(t_1)G_i(t_2).$$

Let $\underline{Z}^T = (Z_1,\ldots,Z_s)$, and $\underline{\lambda}^{\frac{1}{2}} = (\lambda_1^{\frac{1}{2}},\ldots,\lambda_s^{\frac{1}{2}})^T$. By the $\delta-$method, we have

$$\sqrt{n}[\hat{F}_n(x) - F(x)] \xrightarrow{d} Z(x) = \frac{\underline{Z}^T(x)\underline{\lambda}^{\frac{1}{2}}}{h[H^{-1}(\overline{G}(x))]}.$$

It remains only to compute the covariance function of $Z(x)$. In fact,

$$\begin{aligned}
K(t_1,t_2) &= \mathrm{Cov}[Z(t_1),Z(t_2)] \\
&= \mathrm{Cov}\left[\frac{\underline{Z}^T(t_1)\underline{\lambda}^{\frac{1}{2}}}{h(H^{-1}(\overline{G}(t_1)))}, \frac{\underline{Z}^T(t_2)\underline{\lambda}^{\frac{1}{2}}}{h(H^{-1}(\overline{G}(t_2)))}\right] \\
&= \frac{[\underline{\lambda}^{\frac{1}{2}}]^T \mathrm{Cov}(\underline{Z}(t_1),\underline{Z}(t_2))[\underline{\lambda}^{\frac{1}{2}}]}{h(H^{-1}(\overline{G}(t_1)))h(H^{-1}(\overline{G}(t_2)))}.
\end{aligned}$$

By noting that the following covariance function is a diagonal matrix,

$$\mathrm{Cov}(\underline{Z}(t_1),\underline{Z}(t_2)) = \mathrm{diag}(c_1(t_1,t_2),\ldots,c_s(t_1,t_2)),$$

the result follows immediately.

Example 3.3.1. (Example 3.2.1 continued) By Theorem 3.3.3, we have

$$\hat{F}_n(x) = \frac{1}{s}\sum_{i=1}^{s} \mathbf{G}_{ni}(x),$$

$$\sqrt{n}[\hat{F}_n(x) - F(x)] \xrightarrow{d} Z(x)$$

for any $x \in R^1$, where Z is a Gaussian process with mean zero and the covariance function

$$K(t_1,t_2) = \frac{1}{s^2}\sum_{i=1}^{s}[G_i(\min(t_1,t_2)) - G_i(t_1)G_i(t_2)],$$

for $t_1, t_2 \in R^1$. In the case $t_1 = t_2 = x$, we have

$$\frac{\sqrt{n}[\hat{F}_n(x) - F(x)]}{\sqrt{K(x,x)}} \xrightarrow{d} N(0,1)$$

which is given by Stokes and Sager (1988).

Example 3.3.2. (Example 3.2.2 continued) Similar to Example 3.3.1, we have

$$\hat{F}_n(x) = [\mathbf{G}_{ns}(x)]^{1/s},$$

$$\sqrt{n}[\hat{F}_n(x) - F(x)] \xrightarrow{d} Z(x)$$

for $x \in R^1$, where Z has the Gaussian process with mean zero and the covariance function mean zero Gaussian process with covariance function

$$K(t_1,t_2) = \frac{F^s(\min(t_1,t_2)) - F^s(t_1)F^s(t_2)}{s[F(t_1)F(t_2)]^{s-1}},$$

for any $t_1, t_2 \in R^1$.

The following theorem strengthens Theorem 3.3.2.

Theorem 3.3.4. (Strong consistency uniformly over $Q(F)$). Suppose that there exist two positive real numbers M and m such that

$$0 < m \leq h(z) \leq M \quad \text{for all} \quad z \in [0,1].$$

Then

$$\sup_{q\in\mathcal{Q}(F)} ||[\hat{F}_n](q) - [F](q)|| \xrightarrow{a.s.} 0$$

as $n \to \infty$, where $\mathcal{Q}(F)$ is defined by (2.3.3).

Proof. This proof is similar to that of Theorem 2.3.3. For a fixed function $q(x)$ we write

$$\begin{aligned}
|[\hat{F}_n](q) - [F](q)| &= \left|[\mathbf{G}_n]\left(\frac{q}{h\circ\hat{F}_n}\right) - [\overline{G}]\left(\frac{q}{h\circ F}\right)\right| \\
&\le \left|[\mathbf{G}_n]\left(q\left[\frac{1}{h\circ\hat{F}_n} - \frac{1}{h\circ F}\right]\right)\right| \\
&\quad + \left|[\mathbf{G}_n]\left(\frac{q}{h\circ F}\right) - [\overline{G}]\left(\frac{q}{h\circ F}\right)\right| \\
&\le \left\|\frac{h\circ F}{h\circ\hat{F}_n} - 1\right\|_\infty [\mathbf{G}_n]\left(\frac{q}{h\circ F}\right) \\
&\quad + \left|[\mathbf{G}_n]\left(\frac{q}{h\circ F}\right) - [\overline{G}]\left(\frac{q}{h\circ F}\right)\right| \\
&\xrightarrow{a.s.} 0
\end{aligned}$$

as $n \to \infty$ with $[\overline{G}](q/h\circ F) = [F](q) < \infty$. The second inequality holds because $h(u)/h$ is bounded together with $\hat{F}_n(x) \xrightarrow{a.s.} F(x)$ by Theorem 3.3.2. Thus,

$$\begin{aligned}
\sup_{q\in\mathcal{Q}(F)} |[\hat{F}_n](q) - [F](q)| &\le \left\|\frac{h\circ F}{h\circ\hat{F}_n} - 1\right\|_\infty [\mathbf{G}_n]\left(\frac{q_e}{h\circ F}\right) \\
&\quad + \sup_{q\in\mathcal{Q}(F)} \left|[\mathbf{G}_n]\left(\frac{q}{h\circ F}\right) - [\overline{G}]\left(\frac{q}{h\circ F}\right)\right| \\
&\xrightarrow{a.s.} 0.
\end{aligned}$$

The result is proved. ∎

Remark 3.3.1. The condition that h is bounded above and below is equivalent to the condition that the function H satisfies a Lipschitz condition.

CHAPTER IV

ESTIMATING THE DISTRIBUTION FUNCTION IN BIASED SAMPLING MODEL III

4.1 Introduction

In this chapter, we study the statistical inference of the general biased sampling model III defined by (1.3.3)

$$\begin{aligned} & G(x,k) = P[X \le x, K = k] = p(k) \int_{-\infty}^{x} \frac{w_k(y)dF(y)}{W_k}, \\ (4.1.1) \qquad & 0 < W_k = \int_{-\infty}^{\infty} w_k(x)dF(y) < \infty, \\ & p(k) = P(K = k), \quad k \in \mathcal{K}, \end{aligned}$$

in Chapter I, where $\{w_k : k \in \mathcal{K}\}$ are known and nonnegative measurable functions, $\{p(k) : k \in \mathcal{K}\}$ and F are unknown. The observed data for estimating F are n independent observations $(X_1, K_1), \ldots, (X_n, K_n)$, where $(X, K) \sim G(x, k)$.

The random variable K in model (4.1.1) can be considered as the "*label*" of individuals who contribute the data set. Model (4.1.1) is different from the ordinary biased sampling model when $\mathcal{K} = \{1, \ldots, s\}$ and $s \ge 1$. For model (4.1.1) we take the sample of n independent random vectors $(X_1, K_1), \ldots, (X_n, K_n)$ from a mixture of s populations while we sample independently from each of s populations in s-biased sampling model. We may have some choices of the design elements of the biasing sampling (choice of s, sample fractions and the biasing functions $w_i(x)$). We, however, do not have any choice since the design is fixed and given in the s-biased sampling model. On the other hand, the number of observationss from each biased distribution G_i is predetermined for s-biased sampling while we do not know this information in advance for model (4.1.1). Hence we may often be interested in model (4.1.1)

This chapter is organized as follows: We first consider identifiability for model (4.1.1). Based on the solution of (4.1.1), we propose a natural estimator of F in Section 4.2 and prove the strong consistency and asymptotic normality of the estimator in Section 4.3. Finally, we show that the estimator proposed in Section 4.2 is in fact the NPMLE of F.

4.2 Identifiability of the Model and an Estimator of F

The following three assumptions are needed for establishing identifiability. The first two are similar to Assumptions 1 and 2 Section 2.2.

Assumption 1. $\mathcal{X}^+ = \bigcup_{i \in \mathcal{K}} \{x : w_i(x) > 0\} = \mathcal{X}$, where $\mathcal{X}$ is the sample space of the random variable X.

As before, if this assumption fails, we must replace the cdf F by the conditional distribution $F^+(x) = P(X \leq x | \mathcal{X}^+)$. The following example shows that the assumptions used in G-V-W (1988) are not enough to produce an identifiable model even in their stratified sampling model.

Example 4.2.1. Stratified sampling. Suppose that $\mathcal{K}$ contains at least three elements, say $\{1, 2, 3\} \subset \mathcal{K}$. Let $\{D_k : k \in \mathcal{K}\}$ be a measurable partition of the sample space $\mathcal{X} = R^1$. If the weight functions are the indicators of the sets D_k, $w_k(x) = I_{D_k}(x)$ for $k \in \mathcal{K}$, then

$$G(x, k) = \frac{F((-\infty, x] \cup D_k)}{F(D_k)} p(k)$$

is just the product of $P(K = k)$ and the conditional distribution given the event D_k. Note here, without introducing further notation, we have for simplicity used F for both the probability distribution and the cdf. It is clear that estimation of F itself not possible without the knowledge of the probabilities $F(D_k)$ and $p(k)$. Even if $\{p(k) > 0 : k \in \mathcal{K}\}$ is known, F is still not estimable.

Assumption 2. The graph $\mathcal{F}$ with points $\{w_k(.) : k \in \mathcal{K}\}$ is connected by a path. That is, for any pair of (i, j) there exist $l_1, \ldots, l_k \in \mathcal{K}$ such that

$$i \equiv l_1 \leftrightarrow l_2 \leftrightarrow \cdots \leftrightarrow l_k \equiv j,$$

where $l_1 \leftrightarrow l_2$ if and only if

$$\int_{-\infty}^{\infty} I_{[w_{l_1}(x)>0]} I_{[w_{l_2}(x)>0]} dF(x) > 0, \quad (l_1, l_2 \in \mathcal{K}).$$

Additionally, we assume that the set $\mathcal{K}$ contains only a finite number of elements for conveience.

Assumption 3. $\mathcal{K} = \{1, \ldots, s\}$, where $1 \leq s < \infty$.

Theorem 4.2.1. Under the Assumptions 1, 2 and 3, the model (4.1.1) is identifiable.

Proof. By Assumption 3

$$G(x) = \sum_{k \in \mathcal{K}} G(x, k) = \int_{-\infty}^{x} \sum_{k=1}^{s} \frac{w_k(y) p(k)}{W_k} dF(y). \tag{4.2.1}$$

Applying Assumptions 1 and 3, we compute

$$
\begin{aligned}
F(x) &= \int_{-\infty}^{x} \left[\sum_{k=1}^{s} \frac{w_k(y)p(k)}{W_k} \right]^{-1} dG(y) \\
&= \frac{\int_{-\infty}^{x} \left[\sum_{k=1}^{s} \frac{w_k(y)p(k)}{W_k} \right]^{-1} dG(y)}{\int_{-\infty}^{\infty} \left[\sum_{k=1}^{s} \frac{w_k(y)p(k)}{W_k} \right]^{-1} dG(y)} \\
&= \frac{\int_{-\infty}^{x} \left[\sum_{k=1}^{s} \frac{w_k(y)p(k)}{V_k} \right]^{-1} dG(y)}{\int_{-\infty}^{\infty} \left[\sum_{k=1}^{s} \frac{w_k(y)p(k)}{V_k} \right]^{-1} dG(y)}
\end{aligned}
\tag{4.2.2}
$$

where $V_i = W_i/W_s, \quad i = 1, \ldots, s$. The rest is to show that $(W_1, \ldots, W_s)$ or $(V_1, \ldots, V_s)$ can be uniquely determined as functions of $G(x, k)$, $1 \le k \le s$. The procedures are the same as in the proof of Theorem 2.2.1 with a slight change of notation, namely λ_{ni}, H_F, B_i and $\overline{G}_n$ are changed to $p(i)$, F, W_i and G, respectively. So we omit the rest of the proof. ∎

As in Chapter II, $V_1, \ldots, V_{s-1}, V_s = 1$ are solutions of the system of $s-1$ equations

$$
L_i(V_1, \ldots, V_{s-1}, 1) = 1, \qquad i = 1, \ldots, s-1, \tag{4.2.3}
$$

where

$$
\begin{aligned}
L_i(V_1, \ldots, V_s) &= \frac{1}{V_i} \int_{-\infty}^{\infty} w_i(y) dF(y) \\
&= \frac{1}{V_i} \int_{-\infty}^{\infty} w_i(y) \left[\sum_{j=1}^{s} \frac{w_j(y)p(j)}{V_j} \right]^{-1} dG(y).
\end{aligned}
\tag{4.2.4}
$$

Now we use the equation (4.2.2) to construct an estimator for F. We write

$$
\begin{aligned}
\mathbf{G}_n(x) &= \frac{1}{n} \sum_{i=1}^{s} \sum_{j=1}^{n} I_{(X_j \le x, K_j = i)} = \frac{1}{n} \sum_{i=1}^{n} I_{(X_i \le x)}; \\
\hat{p}_n(k) &= \frac{1}{n} \sum_{i=1}^{n} I_{[K_i = k]}.
\end{aligned}
\tag{4.2.5}
$$

Replacing G and $p(k)$ on the right side of (4.2.2) by the empirical cdf $\mathbf{G}_n$ and $\hat{p}_n(k)$ yields a nondecreasing function $\hat{F}_n^0$,

$$\hat{F}_n^0(x) = \int_{-\infty}^{x} \left[\sum_{k=1}^{s} \frac{w_k(y)\hat{p}_n(k)}{W_k} \right]^{-1} d\mathbf{G}_n(y) \tag{4.2.6}$$

$$= \frac{\int_{-\infty}^{x} \left[\sum_{k=1}^{s} \frac{w_k(y)\hat{p}_n(k)}{W_k} \right]^{-1} d\mathbf{G}_n(y)}{\int_{-\infty}^{\infty} \left[\sum_{k=1}^{s} \frac{w_k(y)\hat{p}_n(k)}{W_k} \right]^{-1} d\mathbf{G}_n(y)}.$$

The next step of construction is to replace the unknown W_i by appropriate estimates. Thus we need to estimate $W_i = \int w_i(y)dF(y)$ first. For this purpose, we replace G and $p(k)$ on the right side of (4.2.4) by the empirical cdf $\mathbf{G}_n$ and $\hat{p}_n(j)$, to obtain the following equations $(1 \le j \le s)$

$$\mathbf{L}_{ni}(V_1, \ldots, V_s) = \frac{1}{V_i} \int_{-\infty}^{\infty} w_i(y) \left[\sum_{j=1}^{s} \frac{w_j(y)\hat{p}_n(j)}{V_j} \right]^{-1} d\mathbf{G}_n(y) \tag{4.2.7}$$

$$= 1.$$

Solving the system (4.2.7) gives a solution $\mathbf{V}_{n1}, \ldots, \mathbf{V}_{n,s-1}, \mathbf{V}_{ns}(= 1)$. Finally, substituting G, $p(k)$, and $V_1, \ldots, V_s$ on the right side of (4.2.2) by the empirical cdf $\mathbf{G}_n$, $\hat{p}_n(k)$, and $\mathbf{V}_{n1}, \ldots, \mathbf{V}_{ns}$ yields an estimator $\hat{F}_n$,

$$\hat{F}_n(x) = \hat{F}_n^0(x; \mathbf{V}_{n1}, \ldots, \mathbf{V}_{ns})$$

$$= \frac{\int_{-\infty}^{x} \left[\sum_{i=1}^{s} \frac{w_i(y)\hat{p}_n(i)}{\mathbf{V}_{ni}} \right]^{-1} d\mathbf{G}_n(y)}{\int_{-\infty}^{\infty} \left[\sum_{i=1}^{s} \frac{\hat{p}_n(i)w_i(y)}{\mathbf{V}_{ni}} \right]^{-1} d\mathbf{G}_n(y)}. \tag{4.2.8}$$

The estimator $\hat{F}_n$ fulfills the requirements of a cdf. We can estimate W_i by

$$\mathbf{W}_{ni} = \mathbf{V}_{ni}\mathbf{W}_{ns}, \qquad i = 1, \ldots, s-1, \tag{4.2.9}$$

$$\mathbf{W}_{ns} = \frac{1}{\int_{-\infty}^{\infty} \left[\sum_{i=1}^{s} \frac{\hat{p}_n(i)w_i(y)}{\mathbf{V}_{ni}} \right]^{-1} d\mathbf{G}_n(y)}. \tag{4.2.10}$$

Remark 4.2.2. In general, the solution $\mathbf{V}_{n1}, \ldots, \mathbf{V}_{n,s-1}, \mathbf{V}_{ns}(=1)$ of (4.2.8) is not unique. For uniqueness of the solution, we impose the following two assumptions. The first is similar to Assumption 4 in Section 2.2.

Assumption 4. The graph $\mathcal{F}$ with points $\{w_1(.), \ldots, w_s(.)\}$ is strongly connected by a path. That is, for any pair of (i, j) there exist $l_1, \ldots, l_k \in \{1, \ldots, s\}$ such that

$$i \equiv l_1 \rightleftharpoons l_2 \rightleftharpoons \cdots \rightleftharpoons l_k \equiv j,$$

where $l_1 \rightleftharpoons l_2$ if and only if

$$\int_{-\infty}^{\infty} I_{[w_{l_1}(x)>0]} I_{[w_{l_2}(x)>0]} d\mathbf{G}_n(x) = \frac{1}{n} \sum_{t=1}^{n} I_{[w_{l_1}(X_t) w_{l_2}(X_t)>0]} > 0$$

for $1 \leq l_1, l_2 \leq s$.

Assumption 5. The estimates $\hat{p}_n(k)$ defined by (4.2.5) are strictly positive for all $k = 1, \ldots, s$.

Assumption 5 requires that the data set must contain at least one observation from each data contributor.

Under Assumptions 4 and 5, the solution $\mathbf{V}_{n1}, \ldots, \mathbf{V}_{n,s-1}, \mathbf{V}_{ns}(=1)$ of (4.2.7) is unique. The proof, which is similar to that of Theorem 4.2.1, will be omitted.

4.3 Asymptotic Properties of the Estimator $\hat{F}_n$

The first theorem gives consistency of the $\underline{\mathbf{V}}_n^T = (\mathbf{V}_{n1}, \ldots, \mathbf{V}_{n,s-1}, 1)$ and $\underline{\mathbf{W}}_n^T = (\mathbf{W}_{n1}, \ldots, \mathbf{W}_{ns})$ given by (4.2.7), (4.2.9), (4.2.10) and (4.2.8), respectively.

Theorem 4.3.1. (Strong consistency of $\underline{\mathbf{V}}_n$ and $\underline{\mathbf{W}}_n$). Suppose that Assumptions 1-5 hold, and

$$0 < W_i = \int_{-\infty}^{\infty} w_i(y) dF(y) < \infty$$

for $i = 1, \ldots, s$. Then equations (4.2.7) have (with probability 1 as $n \to \infty$) the unique solution $\underline{\mathbf{V}}_n$ which satisfies

$$\underline{\mathbf{V}}_n \xrightarrow{a.s.} \underline{V} = \underline{W}/W_s \qquad \text{as} \quad n \to \infty. \tag{4.3.1}$$

Furthermore,

$$\underline{\mathbf{W}}_n \xrightarrow{a.s.} \underline{W} \qquad \text{as} \quad n \to \infty. \tag{4.3.2}$$

The proof of $\underline{\mathbf{V}}_n \xrightarrow{a.s.} \underline{V}$ is similar to the first part of the proof of Proposition 2.1 of G-V-W (1988) if we note the only difference is we have $\{\hat{p}_n(i)\}$ instead of $\{\lambda_{ni}\}$. Hence we omit this part of the proof.

Now applying (4.3.1), (4.2.2) and Assumption 1, $\mathbf{W}_{ns}$ given by (4.2.10) converges a.s. to

$$\frac{1}{\int_{-\infty}^{\infty} \left[\sum_{i=1}^{s} \frac{p(i) w_i(y)}{V_i} \right]^{-1} dG(y)} = \frac{W_s}{\int_{-\infty}^{\infty} \left[\sum_{i=1}^{s} \frac{p(i) w_i(y)}{W_i} \right]^{-1} dG(y)} = W_s.$$

Hence

$$\mathbf{W}_{ni} = \mathbf{V}_{ni} \mathbf{W}_{ns} \xrightarrow{a.s.} V_i W_s = W_i \quad \text{as} \quad n \to \infty$$

for $i = 1, \ldots, s$, so (4.3.2) holds. ∎

The second theorem asserts the uniform strong consistency of $\hat{F}_n$ and $\hat{F}_n^0$.

Theorem 4.3.2. (Strong consistency of $\hat{F}_n$ and $\hat{F}_n^0$). Suppose that Assumptions 1-5 hold, and

$$0 < W_i = \int_{-\infty}^{\infty} w_i(y) dF(y) < \infty$$

for $i = 1, \ldots, s$. Then

$$\text{(4.3.3)} \qquad ||\hat{F}_n - F||_{\mathcal{Q}(H_F)} \equiv \sup\{|[\hat{F}_n](q) - [F](q)| : q \in \mathcal{Q}(F)\} \xrightarrow{a.s.} 0$$

as $n \to \infty$. Furthermore, under Assumptions 1,3, 4 and 5,

$$\text{(4.3.4)} \qquad ||\hat{F}_n^0 - F||_{\mathcal{Q}(F)} \equiv \sup\{|[\hat{F}_n^0](q) - [F](q)| : q \in \mathcal{Q}(F)\} \xrightarrow{a.s.} 0$$

as $n \to \infty$.

Proof. Let

$$\underline{\mathbf{p}}_n^T = (\hat{p}_n(1), \ldots, \hat{p}_n(s)), \quad \underline{p}^T = (p(1), \ldots, p(s));$$
$$r_n(\underline{\mathbf{V}}_n, \underline{\mathbf{p}}_n) = \left[\sum_{i=1}^{s} \frac{w_i(x)\hat{p}_n(i)}{\mathbf{V}_{ni}}\right]^{-1}, \quad r(\underline{V}, \underline{p}) = \left[\sum_{i=1}^{s} \frac{w_i(x)p(i)}{V_i}\right]^{-1}.$$

both r_n and r are functions of x. Then for a fixed function $q(x)$ with $[F](q) < \infty$ we have

$$\begin{aligned}
&|[\mathbf{G}_n](qr_n(\underline{\mathbf{V}}_n, \underline{\mathbf{p}}_n)) - [G](qr(\underline{V}, \underline{p}))| \\
&\leq |[\mathbf{G}_n](q(r_n(\underline{\mathbf{V}}_n, \underline{\mathbf{p}}_n) - r(\underline{V}, \underline{p})))| \\
&+ |[\mathbf{G}_n](qr(\underline{V}, \underline{p})) - [G](qr(\underline{V}, \underline{p}))| \\
&\leq \left|\left|\frac{r_n(\underline{\mathbf{V}}_n, \underline{\mathbf{p}}_n)}{r(\underline{V}, \underline{p})} - 1\right|\right|_\infty |[\mathbf{G}_n](qr(\underline{V}, \underline{p}))| \\
&+ |[\mathbf{G}_n](qr(\underline{V}, \underline{p})) - [G](qr(\underline{V}, \underline{p}))| \\
&\xrightarrow{a.s.} 0 \qquad \text{as} \quad n \to \infty.
\end{aligned}$$

The second inequality holds because $r_n(\underline{\mathbf{V}}_n, \underline{\mathbf{p}}_n)/r(\underline{V}, \underline{p})$ is bounded as a function of x together with $\underline{\mathbf{V}}_n \xrightarrow{a.s.} \underline{V}$ and $\underline{\mathbf{p}}_n \xrightarrow{a.s.} \underline{p}$. Thus,

$$
\begin{aligned}
&\sup_{q\in\mathcal{Q}(F)} |[\mathbf{G}_n](qr_n(\underline{\mathbf{V}}_n, \underline{\mathbf{p}}_n)) - [G](qr(\underline{V}, \underline{p}))| \\
&\leq \left\| \frac{r_n(\underline{\mathbf{V}}_n, \underline{\mathbf{p}}_n)}{r(\underline{V}, \underline{p})} - 1 \right\|_\infty \sup_{q\in\mathcal{Q}(F)} |[\mathbf{G}_n](qr(\underline{V}, \underline{p}))| \\
&+ \sup_{q\in\mathcal{Q}(F)} |[\mathbf{G}_n](qr(\underline{V}, \underline{p})) - [G](qr(\underline{V}, \underline{p}))| \\
&\leq \left\| \frac{r_n(\underline{\mathbf{V}}_n, \underline{\mathbf{p}}_n)}{r(\underline{V}, \underline{p})} - 1 \right\|_\infty [\mathbf{G}_n](q_e r(\underline{V}, \underline{p}))| \\
&+ \sup_{q\in\mathcal{Q}(F)} |[\mathbf{G}_n](qr(\underline{V}, \underline{p})) - [G](qr(\underline{V}, \underline{p}))| \\
&\xrightarrow{a.s.} 0.
\end{aligned}
$$

But

$$
\begin{aligned}
|[\hat{F}_n](q) - [F](q)| &= \left| \frac{[\mathbf{G}_n](qr_n(\underline{\mathbf{V}}_n, \underline{\mathbf{p}}_n))}{[\mathbf{G}_n](r_n(\underline{\mathbf{V}}_n, \underline{\mathbf{p}}))} - \frac{[G](qr(\underline{V}, \underline{p}))}{[G](r(\underline{V}, \underline{p}))} \right| \\
&\leq \frac{|[\mathbf{G}_n](qr_n(\underline{\mathbf{V}}_n, \underline{\mathbf{p}}_n)) - [G](qr(\underline{V}, \underline{p}))|}{[\mathbf{G}_n](r_n(\underline{\mathbf{V}}_n, \underline{\mathbf{p}}_n))} \\
&+ \frac{|[G](qr(\underline{V}, \underline{p}))||[\mathbf{G}_n](r_n(\underline{\mathbf{V}}_n, \underline{\mathbf{p}}_n)) - [G](r(\underline{V}, \underline{p}))|}{[G](r(\underline{V}, \underline{p}))[\mathbf{G}_n](r_n(\underline{\mathbf{V}}_n, \underline{\mathbf{p}}_n))}.
\end{aligned}
$$

So (4.3.3) follows immediately. The proof of (4.3.4) (i.e., consistency of $\hat{F}_n^0$) is similar, but does not use (4.3.1), since $\hat{F}_n^0$ depends only on the W_i and not on the $\mathbf{W}_{ni}$. We therefore omit it. ∎

The following theorem establishes the asymptotic normality of $\hat{F}_n^0$ under the further assumption that $\underline{W}^T = (W_1, \ldots, W_s)$ and $\underline{p}^T = (p(1), \ldots, p(s))$ are known values. The asymptotic normality of general $\hat{F}_n$ is still under investigation.

Theorem 4.3.3. Suppose that Assumptions 1-4 hold. $0 < W_i =$

$[F](w_i) < \infty$, and $p(i), \quad 1 \leq i \leq s$ are known. Then

$$\sqrt{n}[\hat{F}_n^0(x) - F(x)] \xrightarrow{d} N(0, \sigma_x^2),$$

where

$$\sigma_x^2 = \mathrm{Var}\Big[I_{(-\infty,x]}(X)\Big(\sum_{i=1}^{s} \frac{p(i)w_i(X)}{W_i}\Big)^{-1}\Big];$$

$$X \sim G(x) = \int_{-\infty}^{x} \Big(\sum_{i=1}^{s} \frac{p(i)w_i(y)}{W_i}\Big) dF(y).$$

Proof. According to (4.2.8) and assumptions, we have

$$\sqrt{n}[\hat{F}_n(x) - F(x)] = \sqrt{n}\Big[[\mathbf{G}_n]\Big(I_x r(\underline{W}, \underline{p})\Big) - [G]\Big(q r(\underline{W}, \underline{p})\Big)\Big]$$
$$\xrightarrow{d} N(0, \sigma_x^2)$$

by the classical CLT. The proof is complete. ∎

4.4 Maximum Likelihood Property of the Estimator $\hat{F}_n$

A natural question can be raised:

Is $\hat{F}_n$ in (4.2.8) a NPMLE of F under model (4.1.1)?

In this section we are going to give a positive answer to this question. We can prove the following lemma by using the method used in Theorem 2.2.1.

Lemma 4.4.1. Suppose that Assumption 4 holds, $X_{(1)} \leq \cdots \leq X_{(n)}$, and $K_{(i)}$ are the accompanying K's. Let

$$R = \sum_{i=1}^{n} \Big\{ -z_i - \log\Big[\sum_{j=1}^{n} w_{K_{(i)}}(X_{(j)}) e^{-z_j}\Big]\Big\}, \tag{4.4.1}$$

where $z_i > 0$ such that $\sum_{i=1}^{n} e^{-z_i} = 1$. Then R is a strictly concave function of $z_1, \ldots, z_n$.

Put

$$A_{ij} = w_{K_{(i)}}(X_{(j)})e^{-z_j}, \quad i,j = 1,\dots,n,$$

(4.4.2)

$$R_l^{(1)} \equiv \frac{\partial R}{\partial z_l}, \quad l = 1,\dots,n.$$

Then

$$R_l^{(1)} = -1 + \sum_{i=1}^{n} \frac{A_{il}}{\sum_{j=1}^{n} A_{ij}}. \quad l = 1,\dots,n.$$

Same argument implies that the solution of $R_l^{(1)} = 0, l = 1,\dots,n$ exists uniquely.

Theorem 4.4.1. The estimator $\hat{F}_n$ defined by (4.2.8) is the NPMLE of F under model (4.1.1).

Proof. Let $(X_1, K_1),\dots,(X_n, K_n)$ be a random sample from the distribution G as in (1.3.3). Relabel the sample in terms of ordered values of $X-$ that is, as $(X_{(1)}, K_{(1)}),\dots,(X_{(n)}, K_{(n)})$, where $X_{(1)} \leq \cdots \leq X_{(n)}$ and $K_{(i)}$ are the accompanying K's. It is easy to show that the NPMLE in this problem places of its mass on $\{(X_{(i)}, K_{(i)})\}$ and maximizes the probability element

(4.4.3)

$$\prod_{i=1}^{n} \frac{w_{K_{(i)}}(X_{(i)})dF(X_{(i)})}{W_{K_{(i)}}}.$$

Let $dF(X_{(i)}) = p_i$ and $\underline{p}^T = (p_1,\dots,p_n)$. Then it follows that

$$W_{K_{(i)}} = \sum_{j=1}^{n} w_{K_{(i)}}(X_{(j)})p_j.$$

Maximizing (4.4.3) with respect to F is equivalent to maximizing $L(\underline{p})$ with respect to $\underline{p}$:

(4.4.4)

$$L(\underline{p}) = \prod_{i=1}^{n} \frac{p_i}{\sum_{j=1}^{n} w_{K_{(i)}}(X_{(j)})p_j}$$

subject to

$$\sum_{i=1}^{n} p_i = 1, \quad p_i > 0, \quad i = 1,\dots,n.$$

We reparametrize: Let $e^{-z_j} = p_j$, and set

$$R(\underline{z}) = \log L(\underline{p}).$$

We know from Lemma 4.5.1 that the maximizer of the likelihood function (4.4.4) exists uniquely. Now we compute this maximizer.

In fact, simple calculation gives

$$p_j^{-1} = e^{z_j} = \sum_{i=1}^{n} \frac{w_{K_{(i)}}(X_{(j)})}{\sum_{l=1}^{n} A_{il}}. \tag{4.4.5}$$

Hence the NPMLE of F is obtained by

$$F^*(x) = \sum_{X_{(j)} \le x} p_j. \tag{4.4.6}$$

In fact, (4.4.6) and (4.2.8) are equivalent. To see this point, we rewrite (4.2.9) as follows

$$\begin{aligned}
\hat{F}_n(x) &= \frac{\int_{-\infty}^{x} \left[\sum_{i=1}^{s} \frac{w_i(y)\hat{p}_n(i)}{\mathbf{W}_{ni}}\right]^{-1} d\mathbf{G}_n(y)}{\int_{-\infty}^{\infty} \left[\sum_{i=1}^{s} \frac{w_i(y)\hat{p}_n(i)}{\mathbf{W}_{ni}}\right]^{-1} d\mathbf{G}_n(y)} \\
&= \frac{\sum_{X_{(j)} \le x} \left[\sum_{i=1}^{n} \frac{w_{K_{(i)}}(X_{(j)})}{\mathbf{W}_{nK_{(i)}}}\right]^{-1}}{\sum_{j=1}^{n} \left[\sum_{i=1}^{n} \frac{w_{K_{(i)}}(X_{(j)})}{\mathbf{W}_{nK_{(i)}}}\right]^{-1}}.
\end{aligned}$$

We claim that

$$\mathbf{W}_{nK_{(i)}} = \sum_{l=1}^{n} w_{K_{(i)}}(X_{(l)}) p_l. \tag{4.4.7}$$

In fact, we obtain from definition of $\mathbf{W}_{nK_{(i)}}$ that

$$\begin{aligned}
\mathbf{W}_{nK_{(i)}} &= \int_{-\infty}^{\infty} w_{K_{(i)}}(y) \left[\sum_{j=1}^{s} \frac{w_j(y)\hat{p}_n(j)}{\mathbf{W}_{nj}}\right]^{-1} d\mathbf{G}_n(y) \\
&= \sum_{t=1}^{n} w_{K_{(i)}}(X_{(t)}) \left[\sum_{j=1}^{n} \frac{w_{K_{(j)}}(X_{(t)})}{\mathbf{W}_{nK_{(j)}}}\right]^{-1}.
\end{aligned} \tag{4.4.8}$$

It is easy to see from (4.4.5) that (4.4.7) satisfies (4.4.8). This concludes the proof. ∎

CHAPTER V

DENSITY ESTIMATION UNDER CONSTRAINT I

5.1 Introduction

Suppose that X is a lifetime random variable with an unknown pdf f. In this chapter we study the nonparametric estimation of f with a sample of n iid observations $X_1, \ldots, X_n$ drawn from population f subject to constraint (1.3.4)

$$\left[\frac{f(x)}{w(x)}\right]' \leq 0, \quad x \in (0, M),$$

where M may be infinite. The following applications motivate our study.

Example 5.1.1. We see from Example 1.2.4 that the backward sampling plan is a special case of model (1.3.4) by setting $w(x) = b(T - x)$, and $M = T$.

Example 5.1.2. (Estimating a monotone decreasing pdf). As we mentioned in Section 1.3, estimating a monotone decreasing pdf f is a special case of model (1.3.4) which is obtained by setting $w(x) = C > 0$ and $M = \infty$. This is a well-known problem. See e.g., Grenander (1956), who introduced the MLE of f; Prakasa-Rao (1969), who provided a thorough analysis of the pointwise properties of the MLE; Groeneboom (1985), who obtained the exact convergence of the L_1 risk of the MLE; Devroye (1987), who devotes an entire chapter to the various methods used so far to estimate the density f; Birgé (1987a, b, 1989), who established the lower bound for the L_1 minimax risk and obtained a minimax optimal estimator, and Datta (1992), who discussed nonasymptotic bounds for L_1 density estimation using kernels.

Example 5.1.3. (Mixture model). Suppose that $X \stackrel{d}{=} YZ$, where positive random variables X and Y have the pdfs g and f, respectively, and $Z \sim B(\alpha, 1)$ (Beta distribution) independent of Y. Let $X_1, \ldots, X_n$ be an iid sample from the population g. The purpose is to estimate the pdf f. Then it is easy to check that

$$g(x) = \int_{y \geq x} \frac{\alpha x^{\alpha-1}}{y^\alpha} f(y) dy = \alpha x^{\alpha-1} \int_{y \geq x} \frac{1}{y^\alpha} f(y) dy.$$

In order to estimate f, we first need to estimate g under the constraint

$$\left[\frac{g(x)}{\alpha x^{\alpha-1}}\right]' \leq 0.$$

This is a special case of model (1.3.4) with $w(x) = \alpha x^{\alpha-1}$ and $M = \infty$. When $\alpha = 1$, it reduces to Example 5.1.2 with $w(y) \equiv 1$.

Example 5.1.4. (Deconvolution model). Suppose that

$$X \stackrel{d}{=} Y + Z,$$

that f and g are the pdfs of Y and X, respectively, and that $-Z$ has an exponential distribution with parameter 1 and is independent of Y. The problem is to derive the NPMLE of the pdf f (or cdf F) of Y by using iid observations $X_1, \ldots, X_n$ from g. It is easy to see that

$$g(x) = e^x \int_x^\infty e^{-y} dF(y).$$

This implies that

$$\left[\frac{g(x)}{e^x}\right]' = -e^{-x} f(x) \leq 0;$$

$$f(x) = g(x) - g'(x),$$

for $x \in R^1$. We first need to estimate $g(x)$ under the constraint $[g(x)e^{-x}]' \leq 0$ in order to estimate $f(x)$ nonparametrically.

Note that this example is not an exact special case of model (1.3.4) because of the range of x. If we, however, make a simple change for (5.1.3) by $X \stackrel{d}{=} \max(0, Y + Z)$, then this example is a special case of (1.3.4) with $w(x) = e^x$, and $M = \infty$.

Example 5.1.4 is a prototype example of the deconvolution model since Z can have any distribution in a deconvolution model.

The following biological example is from the corpuscle problem which has applications in tumor growth. This example shows that the constraint (1.3.4) sometimes may fail for population distribution, but a reasonable estimator must satisfy this constraint.

Example 5.1.5. (Keiding, Jensen, and Ranek (1972)). From sections of liver biopsies, the distribution of the radii of sections of liver cell nuclei is recorded. Under the assumption that the nuclei are spherical, the problem of inferring the distribution of the radii of the nuclei in the liver from the observed distribution of the radii of sections is an example of the *corpuscle problem* studied long ago by Wicksell (1925, 1926). In its general setting the corpuscle problem concerns a conglomerate consisting of a material A and bodies of another material B distributed therein. We shall assume throughout that the bodies are spheres with a random size distribution. The density of the distribution of the sphere radii is denoted by $f(y)$.

The Wicksell theory considers the case where the surface of a plane section is observed. The density $g(x)$ of the radius of the spherical section is then related to $f(y)$ by the Volterra integral equation of the first kind:

$$g(x) = \frac{1}{\mu} \int_x^\infty \frac{x f(y)}{\sqrt{y^2 - x^2}} dy, \tag{5.1.1}$$

where $\mu = \int_0^\infty x f(x) dx$ is the expected spherical radius. It can be shown that $g(x)/x$ is monotone decreasing if we assume that f is monotone decreasing. Thus (5.1.1) is a special case of model (1.3.4) with $w(x) = x$ and $M = \infty$. In general, $g(x)/x$ need not have the monotone decreasing property. The inversion formula for the model (5.1.1) was not available until Anderssen and Jakeman (1975) provided the solution:

$$f(x) = -\frac{2x\mu}{\pi} \int_x^\infty \frac{1}{\sqrt{y^2 - x^2}} \frac{d}{dy} \left[\frac{g(y)}{y} \right] dy. \tag{5.1.2}$$

Applying (5.1.2), we may choose an estimator $\hat{g}_n$ for g such that $\hat{g}_n(x)/x$ is monotone decreasing to ensure $\hat{f}_n(x) \geq 0$, a necessary requirement for a density estimator.

This chapter is organized as follows: We first consider in Section 5.2 the NPMLE estimator $\hat{f}_n^*$ of f and its large sample properties for model (1.3.4). The construction of $\hat{f}_n^*$ makes use of a a simple transformation of the data set. In Section 5.3 we investigate the kernel estimator of f and its properties. Similar discussion for modified histogram type estimator will be given in Section 5.4.

5.2 The NPMLE and Asymptotic Properties

The NPMLE $\hat{f}_n^*(x)$ of $f(x)$ for model (1.3.4) is defined as a density function such that $\hat{f}_n^*(x)/w(x)$ is a monotone decreasing function and

$$\prod_{i=1}^n \hat{f}_n^*(X_i)$$

is maximal.

We first note that since we consider only one sample from f, the weight function $w(x)$ needs no subscripts.

The construction of $\hat{f}_n^*$ is based on a transformation of the data

$$Y = \int_0^X w(z)dz \equiv W(X) \tag{5.2.1}$$

for any $X \in (0, M)$. Here M may be equal to ∞. If the pdf of X is f, then the pdf of Y is given by

$$r(y) = \frac{f(W^{-1}(y))}{w(W^{-1}(y))}, \qquad y \in (0, W(M)) \tag{5.2.2}$$

where $W(M) = \int_0^\infty w(z)dz \leq \infty$.

It is easy to check that r is monotone decreasing if and only if f/w is monotone decreasing. Thus the problem can be solved by using the transformed sample

$$Y_1, \ldots, Y_n \quad \text{iid} \quad r(y) \quad \text{(monotone decreasing)}, \tag{5.2.3}$$

$$Y_i = \int_0^{X_i} w(z)dz, \qquad i = 1, \ldots, n. \tag{5.2.4}$$

We claim that

$$\hat{f}_n^*(x) = w(x)\hat{r}_n(W(x)), \qquad x \in (0, M), \tag{5.2.5}$$

where $\hat{r}_n(.)$ is Grenander's NPMLE of $r(.)$ based on data $Y_1, \ldots, Y_n$. The following Lemmas 5.2.1 and 5.2.2 and Theorem 5.2.1 give rigorous proof to our claim (5.2.5).

Lemma 5.2.1. The NPMLE $\hat{f}_n^*(x)$ of $f(x)$ is the product of $w(x)$ and a step function with breakpoints (jumps) at the order statistics $X_{(i)} (1 \le i \le n)$.

Proof. Write the log-likelihood of the sample in terms of the order statistics $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ as

$$L(f) = \sum_{i=1}^{n} \log[f(X_i)] = \sum_{i=1}^{n} \log[f(X_{(i)})].$$

Define

$$\begin{aligned} f^*(x) &= \begin{cases} 0 & \text{if } x \le 0; \\ cw(x)\frac{f(X_{(i)})}{w(X_{(i)})} & \text{if } X_{(i-1)} < x \le X_{(i)}; \\ 0 & \text{if } x > X_{(n)}; \end{cases} \\ &= cw(x) \sum_{i=1}^{n} \frac{f(X_{(i)})}{w(X_{(i)})} I_{[X_{(i-1)} < x \le X_{(i)}]}, \end{aligned}$$

where c is a normalizing constant such that $f^*(x)$ is a pdf on $[0, \infty)$. Observe that

$$L(f^*) = n \log(c) + L(f) \ge L(f).$$

This is because

$$\begin{aligned} \frac{1}{c} &= \sum_{i=1}^{n} \frac{f(X_{(i)})}{w(X_{(i)})} \int_{X_{(i-1)}}^{X_{(i)}} w(y)dy \\ &\le \sum_{i=1}^{n} \int_{X_{(i-1)}}^{X_{(i)}} \frac{f(y)}{w(y)} w(y)dy \\ &\quad \text{(by the decreasing property of } f(y)/w(y)) \\ &= \int_{0}^{X_{(n)}} f(y)dy \\ &\le 1. \end{aligned}$$

Thus for every density $f_n(x)$ there exists a density function $f_n^*(x)$ such that $f_n^*(x)$ is the product of $w(x)$ and a step function with breakpoints at the order statistics and $L(f_n^*) \ge L(f_n)$. ∎

Lemma 5.2.1 asserts that the form of the NPMLE of $f(x)$ is the product of $w(x)$ and a data dependent histogram. The next Lemma states that once we have settled on the breakpoints, the NPMLE is completely specified. Its proof is patterned after the proof of Lemma 8.3 of Devroye (1987).

Lemma 5.2.2. Consider a partition $A_1, \ldots, A_k$ of a compact subset A of R^1, and a histogram type density estimate $f_n(x)$ taking the value $w(x)g_i$ on A_i, subject to the normalization $\sum g_i \int_{A_i} w(x)dx = 1$. That is, $f_n(x) = w(x)\sum_{i=1}^{k} g_i I_{A_i}(x)$. Then the maximum over all these histogram type estimates of the likelihood product

$$\prod_{i=1}^{n} f_n(X_i) \tag{5.2.6}$$

is attained for the histogram type estimate with

$$g_i = \frac{\mu_n(A_i)}{\int_{A_i} w(y)dy}, \tag{5.2.7}$$

where μ_n is the empirical measure for the data.

Proof. Put $\Delta_i = \int_{A_i} w(y)dy$ and $C_i = n\mu_n(A_i)$. Observe that for any $g_1, \ldots, g_k$

$$\begin{aligned}\prod_{i=1}^{n} f_n(X_i) &= \prod_{i=1}^{k} [g_i w(X_{(i)})]^{C_i} \\ &= \prod_{i=1}^{k} \left[\frac{g_i n\Delta_i}{C_i}\right]^{C_i} \prod_{i=1}^{k} \left[\frac{C_i w(X_{(i)})}{n\Delta_i}\right]^{C_i} \\ &\le \left[\frac{\sum_{i=1}^{k} g_i n\Delta_i}{\sum_{i=1}^{k} C_i}\right]^{n} \prod_{i=1}^{k} \left[\frac{C_i w(X_{(i)})}{n\Delta_i}\right]^{C_i} \\ &= \prod_{i=1}^{k} \left[\frac{C_i w(X_{(i)})}{n\Delta_i}\right]^{C_i}.\end{aligned}$$

Here we have used the arithmetic-geometric mean inequality. This proves the lemma. ∎

When Lemma 5.2.2 is applied with $A_i = (X_{(i-1)}, X_{(i)}], (1 \le i \le n)$, then it is easily seen that among all densities which are the products of $w(x)$ and all step functions with breakpoints at the order statistics, the likelihood product (5.2.6) is maximized if we take a density which on $(X_{(i-1)}, X_{(i)}]$, takes the value

$$\frac{w(x)}{n \int_{X_{(i-1)}}^{X_{(i)}} w(y)dy}$$

since the empirical measure of each interval is precisely one.

Now let us use $I_1, \ldots, I_k$ to denote a partition of indices $\{1, \ldots, n\}$ determined by the breakpoints of the smallest concave majorant of the empirical distribution function of $\{Y_i\}_{i=1}^n$ defined by (5.2.4).

Theorem 5.2.1. The NPMLE $\hat{f}_n^*$ is a density whose distribution function is given by

$$\hat{F}_n^*(x) = \int_0^x w(y)B_n(y)d\hat{F}_n(y) = \hat{R}_n(W(x)) \tag{5.2.8}$$

where $\hat{R}_n$ and $\hat{F}_n$ are the smallest concave majorants of the empirical distribution functions R_n and F_n based on data $Y_1, \ldots, Y_n$ and $X_1, \ldots, X_n$, respectively, and

$$R_n(y) = \frac{1}{n}\sum_{i=1}^n I_{(-\infty,y]}(Y_i);$$

$$B_n(y) = \sum_{i=1}^k \frac{1}{a_i} I_{D_i}(y);$$

$$D_i = \cup_{j \in I_i} A_j = \cup_{j \in I_i} (X_{(j-1)}, X_{(j)}];$$

$$a_i = \frac{\sum_{j \in I_i} \int_{X_{(j-1)}}^{X_{(j)}} w(y)dy}{\sum_{j \in I_i} \int_{X_{(j-1)}}^{X_{(j)}} dy},$$

i.e., a_i is the average of function $w(y)$ over set $D_i, (1 \leq i \leq k)$.

Proof. According to Lemma 5.2.1, it suffices to consider only histogram type estimators with monotone histogram and breakpoints at the order statistics. Consider such a density $g^*(x)$, and let its value be $w(x)g_i$ on the interval $A_i = (X_{(i-1)}, X_{(i)}]$. Let $\Delta_i = \int_{X_{(i-1)}}^{X_{(i)}} w(y)dy$. Consider a partition of $1, \ldots, n$ into intervals of indices, $I_1, \ldots, I_k$, and define

$$\begin{aligned} p_i &= \sum_{j \in I_i} g_j \Delta_j, \\ q_i &= \frac{1}{n} \text{Cardinality of } \quad \cup_{j \in I_i} A_j, \\ h_j &= \left(\frac{q_i}{p_i}\right) g_j, \quad j \in I_i, \end{aligned}$$

where $i = 1, \ldots, k$. Note that the h_j's define another histogram estimator $h^*(x)$ with the following properties:

(I) The integral of $h^*(x)$ is one, since

$$\int_0^M h^*(x)dx = \sum_j \int_0^M h_j w(x)dx = \sum_{i=1}^k \left(\frac{q_i}{p_i}\right) \sum_{j \in I_i} g_j \int_{A_j} w(x)dx = 1.$$

(II) The likelihood evaluated at h^* is larger than at g^*, since

$$\begin{aligned} \prod_{i=1}^n h^*(X_i) &= \prod_{i=1}^n w(X_i) \prod_{i=1}^k \prod_{j \in I_i} h_j \\ &= \prod_{i=1}^n w(X_i) \prod_{i=1}^k \left(\frac{q_i}{p_i}\right)^{nq_i} \prod_{j \in I_i} g_j \\ &\geq \prod_{i=1}^n w(X_i) \prod_{i=1}^k \prod_{j \in I_i} g_j \\ &= \prod_{i=1}^n w(X_i) \prod_{i=1}^n g_i \end{aligned}$$

because $\int_{-\infty}^{\infty} q(x) \log \left[\frac{q(x)}{p(x)}\right] dx \geq 0$ for all densities $q(x)$ and $p(x)$. This improvement is applicable to any histogram type estimator with an arbitrarily selected partition. In particular, we can partition the indices $\{1, \ldots, n\}$ by the break-points of the smallest concave majorant of the R_n with $\{Y_i\}_{i=1}^n$ defined by (5.2.4). For such a partition, we have

$$\sum_{j \in I_i} \int_0^M h_j(x)dx = \frac{q_i}{p_i} \sum_{j \in I_i} g_j \int_{A_j} w(x)dx = q_i.$$

Furthermore, since the g_j are nonincreasing, the h_j are nonincreasing in each I_i, $i = 1, \ldots, k$. As in the proof of Theorem 8.2 of Devroye (1987), we can make a further improvement from $h_1, \ldots, h_n$ to $l_1, \ldots, l_n$, which has the property that they are independent of the original choice of g_i's. This is the desired improvement, and the product of the histogram type density defined by $\{l_1, \ldots, l_n\}$ and $w(x)$ is the NPMLE. Define l_j for $j \in I_i$ by

$$l_j = \frac{\sum_{j \in I_i} \int_0^M h_j(x)dx}{\sum_{j \in I_i} \int_{A_j} w(x)dx} = \frac{q_i}{\sum_{j \in I_i} \Delta_j}.$$

They agree with the NPMLE. It suffices to show that we have a likelihood product improvement for every I_i. To see this, we need to show that

$$\prod_{j \in I_i} h_j \leq \prod_{j \in I_i} l_j$$

for all $i = 1, \ldots, k$. In fact,

$$\left[\prod_{j \in I_i} h_j\right]^{\frac{1}{nq_i}} \leq \frac{1}{nq_i} \sum_{j \in I_i} h_j$$

by the arithmetic-geometric mean inequality, and

$$\begin{aligned} q_i &= \int_0^M w(x)\left[\sum_{j \in I_i} h_j I_{A_j}(x)\right] dx \\ &= \sum_{j \in I_i} h_j \int_0^M w(x) I_{A_j}(x)dx \\ &= \sum_{j \in I_i} h_j \Delta_j \\ &\geq \left[\frac{1}{nq_i} \sum_{j \in I_i} h_j\right]\left[\sum_{j \in I_i} \Delta_j\right] \end{aligned}$$

by association inequality. This concludes the proof of the first equality of (5.2.8). The second equality of (5.2.8) follows by a simple transformation $z = W(y)$ in the integral. ∎

The NPMLE $\hat{f}_n^*$ in (5.2.5) has the following convergence properties:

Theorem 5.2.2. $\int_0^M |\hat{f}_n^*(x) - f(x)|dx \to 0$ almost surely if and only if $f(x)/w(x)$ is monotone decreasing.

Proof. We apply Theorem 8.3 of Devroye (1987) which asserts

$$\int_0^{W(M)} |\hat{r}_n(y) - r(y)|dy \to 0$$

almost surely if and only if r, defined by (5.2.3) is monotone decreasing. Direct computation gives

$$\begin{aligned}
&\int_0^{W(M)} |\hat{r}_n(y) - r(y)|dy \\
&= \int_0^{W(M)} \left| \hat{r}_n(y) - \frac{f(W^{-1}(y))}{w(W^{-1}(y))} \right| dy \\
&= \int_0^{W(M)} |w(W^{-1}(y))\hat{r}_n(y) - f(W^{-1}(y))| \frac{dy}{w(W^{-1}(y))} \\
&= \int_0^M |w(x)\hat{r}_n(W(x)) - f(x)|dx \\
&= \int_0^M |\hat{f}_n^*(x) - f(x)|dx.
\end{aligned}$$

The result follows immediately from the fact r is monotone decreasing if and only if f/w is monotone decreasing. ∎

Theorem 5.2.3. Let $X_1, \ldots, X_n$ be independent observations generated by f such that $f(x)/w(x)$ is a monotone decreasing function on $(0, M)$ which has a nonzero derivative $\left[f/w\right]'$ at a point $t \in (0, M)$. If $\hat{f}_n^*$ is the NPMLE of

f, then

$$n^{1/3}[w(t)]^{-2/3}\left|\frac{1}{2}\left[\frac{f(t)}{w(t)}\right]\left[\frac{f(t)}{w(t)}\right]'\right|^{-1/3}\left[\hat{f}_n^*(t)-f(t)\right]\to 2Z$$

in distribution, where Z is distributed as the location of the maximum of the process $[L(u)-u^2, u\in R^1]$, and L is standard Brownian motion on R^1 originating from zero (i.e., $L(0)=0$).

Proof. Using the notation of Groeneboom (1985) we have

$$P\left[\hat{f}_n^*(t)-f(t)\le xn^{-1/3}[w(t)]^{2/3}\left|\frac{1}{2}\left[\frac{f(t)}{w(t)}\right]\left[\frac{f(t)}{w(t)}\right]'\right|^{1/3}\right]$$

$$=P\left[w(t)\hat{r}_n(W(t))-r(W(t))w(t)\le xn^{-1/3}[w(t)]^{2/3}\left|\frac{1}{2}\left[\frac{f(t)}{w(t)}\right]\left[\frac{f(t)}{w(t)}\right]'\right|^{1/3}\right]$$
$$=P\left[\hat{r}_n(W(t))-r(W(t))\le \frac{x}{w(t)}n^{-1/3}[w(t)]^{2/3}\left|\frac{1}{2}\left[\frac{f(t)}{w(t)}\right]\left[\frac{f(t)}{w(t)}\right]'\right|^{1/3}\right]$$
$$=P\left[n^{1/3}\left|\frac{1}{2}\left[\frac{f(t)}{w(t)}\right]\left[\frac{f(t)}{w(t)}\right]'\frac{1}{w(t)}\right|^{-1/3}\Big(\hat{r}_n(W(t))-r(W(t))\Big)\le x\right]$$
$$\to P(2Z\le x).$$

The weak convergence follows by Theorem 2.1 of Groeneboom (1985). ∎

Theorem 5.2.4. Assume that f/w has two bounded continuous derivatives on $(0,M)$ and that $[f/w]'<0$ on $(0,M)$. Then

$$n^{\frac{1}{3}}E\left[\int_0^M|\hat{f}_n^*(x)-f(x)|dx\right]\to c\int_0^M\left[\frac{1}{2}w(x)f(x)|[f(x)/w(x)]'|\right]^{\frac{1}{3}}dx$$

where $c\approx 0.82$ is a universal constant.

Proof. Applying Theorem 8.4 of Devroye (1987) we have

$$n^{\frac{1}{3}}E\left[\int_0^{W(M)}|\hat{r}_n(y)-r(y)|dy\right]\to c\int_0^{W(M)}\left[\frac{1}{2}r(y)|r'(y)|\right]^{\frac{1}{3}}dy$$

where $c \approx 0.82$ is a universal constant (see Theorem 5.2.5.). The result follows by the transformation $y = W(x) = \int_0^x w(z)dz$ and the following facts:

$$n^{\frac{1}{3}} E\left[\int_0^{W(M)} |\hat{r}_n(y) - r(y)|dy\right] = n^{\frac{1}{3}} E\left[\int_0^{M} |\hat{r}_n(W(x)) - r(W(x))|w(x)dx\right]$$
$$= n^{\frac{1}{3}} E\left[\int_0^{M} |\hat{f}_n^*(x) - f(x)|dx\right];$$
$$c\int_0^{W(M)} \left[\frac{1}{2}r(y)|r'(y)|\right]^{\frac{1}{3}} dy = c\int_0^{M} \left[\frac{1}{2}r(W(x))|r'(W(x))|\right]^{\frac{1}{3}} w(x)dx$$
$$= c\int_0^{M} \left[\frac{1}{2}w(x)f(x)\left|\Big(f(x)/w(x)\Big)'\right|\right]^{\frac{1}{3}} dx.$$

This concludes the proof. ∎

Next we investigate the asymptotic normality of the L_1–norm $||\hat{f}_n^* - f||_1$. In order to do that, let f be a density, concentrated on a bounded interval $[0, B]$ such that $[f(x)/w(x)]' < 0$ for $x \in (0, B)$ and $[f(x)/w(x)]^{(2)}$ is bounded and continuous. Furthermore, let $(L(t), t \in R^1)$ be Brownian motion on R^1, originating from zero, and let the process $(V(a), a \in R^1)$ be defined by

$$V(a) = \sup\{t \in R^1 : L(t) - (t-a)^2 \quad \text{is maximal}\}.$$

Then V is an increasing pure-jump process, generated by the Brownian motion sample paths. Let

$$||\hat{f}_n^* - f||_{1,B} = \int_0^B |\hat{f}_n^*(t) - f(t)|dt.$$

Then we have the following result:

Theorem 5.2.5. (Asymptotic normality) Let B be a positive number. Assume that w is differentiable . Then

$$n^{\frac{1}{6}}\{n^{\frac{1}{3}}||\hat{f}_n^* - f||_{1,B} - C\} \to N(0, \sigma^2) \quad \text{as} \quad n \longrightarrow \infty$$

where

$$C = 2E|V(0)| \int_0^B |\frac{1}{2}\frac{f(x)}{w(x)}\Big(f'(x)w(x) - w'(x)f(x)\Big)|^{\frac{1}{3}} dx,$$

$$\approx 0.82 \int_0^B |\frac{1}{2}\frac{f(x)}{w(x)}\Big(f'(x)w(x) - w'(x)f(x)\Big)|^{\frac{1}{3}} dx,$$

$$\sigma^2 = 8 \int_0^\infty \mathrm{Cov}(|V(0)|, |V(a) - V(0)|) da$$

$$\approx 0.17.$$

Proof. Let

$$||\hat{r}_n - r||_{1,W(B)} = \int_0^{W(B)} |\hat{r}_n(y) - r(y)| dy.$$

From the proof of preceding theorem we have $||\hat{r}_n - r||_{1,W(B)} = ||\hat{f}_n^* - f||_{1,B}$.

Further, compute

$$\int_0^{W(B)} |\frac{1}{2} r(y) r'(y)|^{\frac{1}{3}} dy = \int_0^B \Big|\frac{1}{2} w(x) f(x) \Big(\frac{f(x)}{w(x)}\Big)'\Big|^{\frac{1}{3}} dx$$
$$= \int_0^B |\frac{1}{2}\frac{f(x)}{w(x)}[f'(x)w(x) - w'(x)f(x)]|^{\frac{1}{3}} dx.$$

Theorem is proved by application of Groeneboom (1985) to $\hat{r} - r$. ∎

5.3 Kernel Estimators and Their Properties

In this section we are going to discuss kernel estimators of f and their properties for model (1.3.4). We first study asymmetric kernels and then symmetric kernels. The kernel estimator of f is defined by

$$\hat{f}_n(x) = \frac{1}{nh}\sum_{i=1}^n k\Big(\frac{x - X_i}{h}\Big) = \frac{1}{nh}\sum_{i=1}^n k_h(x - X_i) \tag{5.3.1}$$

where the kernel k is a function on R^1 satisfying $k \geq 0$ and $\int_{-\infty}^\infty k(x)dx = 1$, and $h = h_n$ depending on n is a sequence of positive reals decreasing to 0.

Let E denote the expectation with respect to the joint distribution of $X_1, \ldots, X_n$, and let $C > 0$ be a fixed constant, and let

$$D_C = \left\{ \begin{array}{c} f(x) : f(x) \quad \text{is a density on } [0, \infty) \quad \text{such that} \\ [f(x)/w(x)]' \leq 0, f(0)/w(0) \leq C \end{array} \right\}.$$

Theorem 5.3.1. Let $\hat{f}_n$ be defined by (5.3.1). Assume that the kernel k is left sided, that is, $k(x)I_{(0,\infty)}(x) = 0$. Then for all $f \in D_C$,

$$E \int_0^\infty |\hat{f}_n(x) - f(x)|dx \leq \left(\frac{1}{nh}\right)^{\frac{1}{2}} M_{1h}(f) + Chk_1 M(f),$$

where

$$k_1 = \int_{-\infty}^{\infty} |x| k(x) dx;$$

$$w_{h,j}(x) \equiv \int_{-\infty}^{0} [k(u)]^j w(x - uh) du, \quad j = 1, 2;$$

$$M_{1h}(f) = \int_0^\infty \left[f(x) \frac{w_{h,2}(x)}{w(x)} \right]^{\frac{1}{2}} dx;$$

$$M(f) = \liminf_{h \to 0+} \int_0^\infty \left[\frac{f(0)}{w(0)}\right]^{-1} \left[\frac{f(x)}{w(x)}\right] w_{h,1}(x) dx.$$

Proof. As usual, we split $E \int_0^\infty |\hat{f}_n(x) - f(x)|dx$ into two parts

$$\begin{aligned} E \int_0^\infty |\hat{f}_n(x) - f(x)|dx &\leq E \int_0^\infty |\hat{f}_n(x) - E\hat{f}_n(x)|dx \\ &\quad + \int_0^\infty |E\hat{f}_n(x) - f(x)|dx \\ &\equiv \text{VARIATION} + \text{BIAS}. \end{aligned}$$

Now for fixed x, we define $w_{h,j}(x) \equiv \int_{-\infty}^{0} [k(u)]^j w(x - uh) du$, with $f_{h,j}(x)$ defined similarly. Then for $x > 0$,

$$\begin{aligned} \mathrm{Var}[\hat{f}_n(x)] &\leq \frac{1}{nh^2} \int_0^\infty k^2\left(\frac{x-y}{h}\right) f(y) dy \\ &= \frac{1}{nh} \int_{-\infty}^{0} k^2(u) f(x - uh) du \\ &\leq \frac{1}{nh} \frac{f(x)}{w(x)} \int_{-\infty}^{0} k^2(u) w(x - uh) du \\ &\equiv \frac{1}{nh} \frac{f(x)}{w(x)} w_{h,2}(x). \end{aligned}$$

It follows that

$$\begin{aligned}
\text{VARIATION} &\leq \int_0^\infty \left[\operatorname{Var}(\hat{f}_n(x))\right]^{\frac{1}{2}} dx \\
&\leq \left(\frac{1}{nh}\right)^{\frac{1}{2}} \int_0^\infty \left[f(x)\frac{w_{h,2}(x)}{w(x)}\right]^{\frac{1}{2}} dx \\
&\equiv \left(\frac{1}{nh}\right)^{\frac{1}{2}} M_{1h}(f),
\end{aligned}$$

where

$$M_{1h}(f) = \int_0^\infty \left[f(x)\frac{w_{h,2}(x)}{w(x)}\right]^{\frac{1}{2}} dx.$$

On the other hand, note that $E\hat{f}_n(x) = f_{h,1}(x)$. Therefore, applying Theorem 7.1 of Devroye (1987), we obtain an upper bound for the integrated absolute bias:

$$\text{BIAS} = \int_0^\infty |E\hat{f}_n(x) - f(x)|dx \leq hk_1 D^*(f),$$

where

$$\begin{aligned}
D^*(f) &= \liminf_{h\to 0+} \int_0^\infty |f_{h,1}(x)|dx \\
&\leq \liminf_{h\to 0+} \int_0^\infty \frac{f(x)}{w(x)} w_{h,1}(x)dx \\
&\leq C \liminf_{h\to 0+} \int_0^\infty \left[\frac{f(0)}{w(0)}\right]^{-1} \left[\frac{f(x)}{w(x)}\right] w_{h,1}(x)dx \\
&\equiv CM(f),
\end{aligned}$$

and

$$M(f) = \liminf_{h\to 0+} \int_0^\infty \left[\frac{f(0)}{w(0)}\right]^{-1} \left[\frac{f(x)}{w(x)}\right] w_{h,1}(x)dx.$$

The result is proved. ∎

Corollary 1. Under the condition of Theorem 5.3.1 and assume that $\sup_{h>0} M_{1h}(f) = M_1(f)$ exists, then

$$E\int_0^\infty |\hat{f}_n(x) - f(x)|dx \leq \left(\frac{1}{nh}\right)^{\frac{1}{2}} M_1(f) + Chk_1 M(f),$$

Corollary 2. Under the condition of Corollary 1, the bound in Corollary 1 is minimized for

$$h_n = \left(\frac{M_1(f)}{2\sqrt{n}k_1CM(f)}\right)^{2/3}$$

and for this choice of h_n

$$E\int_0^\infty |\hat{f}_n(x) - f(x)|dx \le \left(\frac{k_1CM(f)}{n}\right)^{\frac{1}{3}} M_1^{\frac{2}{3}}(f)[2^{1/3} + 2^{-2/3}].$$

Corollary 3. For all n and the band width $h > 0$,

$$\sup_{f\in D_C} E\int_0^\infty |\hat{f}_n(x) - f(x)|dx \le \left(\frac{1}{nh}\right)^{1/2} M_{1h} + hk_1CM$$

provided $M_{1h} = \sup_{f\in D_C} M_{1h}(f)$ and $M = \sup_{f\in D_C} M(f)$ are finite. Hence

$$\inf_h \sup_{f\in D_C} E\int_0^\infty |\hat{f}_n(x) - f(x)|dx \le \left(\frac{k_1CM}{n}\right)^{\frac{1}{3}} M_1^{\frac{2}{3}}[2^{1/3} + 2^{-2/3}]$$

if $M_1 = \sup_{h>0} M_{1h}$ exists.

The next theorem considers a symmetric kernel.

Theorem 5.3.2. Suppose that the kernel in (5.3.1) is symmetric about zero. Then for all n and h,

$$E\int_0^\infty |\hat{f}_n(x) - f(x)|dx \le \left(\frac{C}{nh}\right)^{\frac{1}{2}} M_{1h}^*(f) + Chk_1M^*(f),$$

where

$$M_{1h}^*(f) = \int_0^\infty [w_{h,2}^*(x)]^{\frac{1}{2}}dx;$$

$$w_{h,j}^*(x) = \int_{-\infty}^{\frac{x}{h}} k^j(u)w(x - uh)du, \quad j = 1, 2;$$

$$M^*(f) = \liminf_{h\to 0+} \int_0^\infty \left[\frac{f(0)}{w(0)}\right]^{-1} \left[\frac{f(x)}{w(x)}\right] w_{h,1}^*(x)dx.$$

Proof. The proof is similar to that of Theorem 5.3.1. It is produced here for completeness.

$$E\int_0^\infty |\hat{f}_n(x) - f(x)|dx \le E\int_0^\infty |\hat{f}_n(x) - E\hat{f}_n(x)|dx + \int_0^\infty |E\hat{f}_n(x) - f(x)|dx$$
$$\equiv \text{VARIATION} + \text{BIAS}.$$

For fixed x,

$$\begin{aligned}
\operatorname{Var}[\hat{f}_n(x)] &\le \frac{1}{nh^2}\int_0^\infty k^2\left(\frac{x-y}{h}\right) f(y)dy \\
&= \frac{1}{nh}\int_{-\infty}^{\frac{x}{h}} k^2(u) f(x-uh)du \\
&\le \frac{1}{nh}\frac{f(0)}{w(0)}\int_{-\infty}^{\frac{x}{h}} k^2(u) w(x-uh)du \\
&\equiv \frac{1}{nh}\frac{f(0)}{w(0)} w^*_{h,2}(x).
\end{aligned}$$

It follows that

$$\begin{aligned}
\text{VARIATION} &\le \int_0^\infty \left[\operatorname{Var}(\hat{f}_n(x))\right]^{\frac{1}{2}} dx \\
&\le \left(\frac{1}{nh}\right)^{\frac{1}{2}} \int_0^\infty \left[f(0)\frac{w^*_{h,2}(x)}{w(0)}\right]^{\frac{1}{2}} dx \\
&\equiv \left(\frac{C}{nh}\right)^{\frac{1}{2}} M^*_{1h}(f),
\end{aligned}$$

where

$$M^*_{1h}(f) = \int_0^\infty [w^*_{h,2}(x)]^{\frac{1}{2}} dx.$$

On the other hand, note that $E\hat{f}_n(x) = f^*_{h,1}(x)$. Applying Theorem 7.1 of Devroye (1987) we have

$$\text{BIAS} = \int_0^\infty |E\hat{f}_n(x) - f(x)|dx \le hk_1 D^*(f),$$

where

$$\begin{aligned}
D^*(f) &= \liminf_{h\to 0+} \int_0^\infty |f^*_{h,1}(x)|dx \\
&\le \liminf_{h\to 0+} \int_0^\infty \frac{f(x)}{w(x)} w^*_{h,1}(x)dx \\
&\le C \liminf_{h\to 0+} \int_0^\infty \left[\frac{f(0)}{w(0)}\right]^{-1} \left[\frac{f(x)}{w(x)}\right] w^*_{h,1}(x)dx \\
&\equiv CM^*(f),
\end{aligned}$$

and

$$M^*(f) = \liminf_{h\to 0+} \int_0^\infty \left[\frac{f(0)}{w(0)}\right]^{-1} \left[\frac{f(x)}{w(x)}\right] w^*_{h,1}(x)dx.$$

The proof is complete. ∎

5.4 Modified Histogram Type Estimates

Let f be a density defined on a bounded interval $[a, a+L]$ such that f/w is monotone decreasing and $f(a)/w(a) \le C$. Let f_n be an estimator of f constructed from n iid observations from f. We define the risk of f_n at f by

$$R(f_n, f) = E\left[\int_a^{a+L} |f_n(x) - f(x)|dx\right]. \tag{5.4.1}$$

In (5.4.1), we make the change of variable $y = \int_0^x w(z)dz = W(x)$. Then

$$\begin{aligned}
R(f_n, f) &= E\left[\int_{W(a)}^{W(a+L)} \left|f_n(W^{-1}(y)) - f(W^{-1}(y))\right| \frac{dy}{w(W^{-1}(y))}\right] \\
&= E\left[\int_{W(a)}^{W(a+L)} \left|\frac{f_n(W^{-1}(y))}{w(W^{-1}(y))} - \frac{f(W^{-1}(y))}{w(W^{-1}(y))}\right| dy\right] \\
&= E\left[\int_{W(a)}^{W(a+L)} |r_n(y) - r(y)|dy\right].
\end{aligned}$$

Hence, as long as we define $f_n(x) = w(x)r_n(W(x))$, where r_n is Birgé's histogram estimator, then the results of Birgé (1987) also hold for our estimator. One of such results is the following.

Theorem 5.4.1. Let

$$f_n(x) = w(x) \sum_{i=0}^{p-1} (nl_i)^{-1} N_i I_{A_i}(W(x)),$$

where N_i denotes the number of observations $\{Y_i\}_{i=1}^n$ which belong to A_i, where $\{A_i = [y_i, y_{i+1})\}_{i=0}^{p-1}$ is a partition of $[W(a), W(a+L)]$, and $l_i = y_{i+1} - y_i$ is the length of A_i such that $\{l_i\}_{i=0}^{p-1}$ is an increasing sequence. Then

$$R(f_n, f) \leq 1.89 \left[\frac{S}{n}\right]^{\frac{1}{3}} + 0.2 \left[\frac{S}{n}\right]^{\frac{2}{3}},$$

where $S = \log[1 + C(W(a+L) - W(a))]$.

CHAPTER VI

DENSITY ESTIMATION UNDER CONSTRAINT II

6.1 Introduction

Suppose that X is a lifetime random variable with an unknown pdf f. The problem we discuss in this chapter is similar to that of Chapter V, but the estimation problem is subject to the constraint (1.3.5):

$$\left[\frac{f(x)}{w(x)}\right]' \geq 0, \quad x \in [0, M].$$

Suppose we are given a sample of n iid observations $X_1, \ldots, X_n$ taken from the population density f. Our goal is to construct nonparametric estimators for f under the constraint (1.3.5). There are numerous practical examples to motivate our study:

Example 6.1.1. (Relevation transform in reliability theory). The Stieltjes convolution of two distribution functions F and G with support on the nonnegative axis is denoted by

$$F * G(t) = \int_0^t F(t-u)dG(u).$$

It represents the distribution function of the time to failure of the second of two components when the second component (with lifetime distribution G) is put into service on the failure of the first (with lifetime distribution F). The replacement component is usually assumed to be new on installation. Suppose, however, that we replace the failed component by one of equal age. The survival function $\overline{H}$ of S, the time to system failure (i.e., both components are failed) is the relevation of $\overline{F}$ and $\overline{G}$, the survival functions of the first and second components, respectively, and is denoted by

$$\overline{H}(t) = \overline{F}(t) + \int_0^t \frac{\overline{G}(t)}{\overline{G}(u)} dF(u).$$

If we assume that F and G have pdfs f and g, respectively, and $F(0) = 0$, we can check that

$$\left[\frac{h}{g}\right]' = \frac{f}{\overline{G}} \geq 0,$$

provided $g \neq 0$ and $\overline{G} \geq 0$. This is a special case of (1.3.5) with $w(x) = g(x)$.

The *relevation transform* was introduced by Krakowski (1973). Its applications in reliability analysis can be found in Baxter (1982) and the references therein.

The following two examples are similar to Examples 5.1.4 and 5.1.3, but the distributions of Z are different.

Example 6.1.2. (Deconvolution model). Suppose that

$$X \stackrel{d}{=} Y + Z,$$

that f and g are the pdfs of nonnegative random variables Y and X, respectively, and that Z has an exponential distribution and is independent of Y. The problem is to estimate the pdf (or cdf) of Y by using iid data $X_1, \ldots, X_n$ from g. It is easy to see that

$$g(x) = e^{-x} \int_{-\infty}^{x} e^y f(y) dy.$$

This implies that

$$\left[\frac{g(x)}{e^{-x}}\right]' = e^x f(x) \geq 0.$$

This is a special case of (1.3.5) with $w(x) = e^{-x}$. Of course, this can also be considered as a special case of Example 6.1.1 in which g has the exponential distribution.

Example 6.1.3. (Mixture model). Suppose that $X \stackrel{d}{=} YZ$, where the nonnegative rvs X and Y have pdfs g and f, respectively, and $Z \sim B^{-1}(\alpha, 1)$ (inverse Beta distribution with $\alpha > 1$) is independent of Y. Let $X_1, \ldots, X_n$ be an iid sample from X. The purpose is to estimate g. Then it is easy to check that

$$g(x) = \int_{y \leq x} \frac{(\alpha-1)y^{\alpha-1}}{x^\alpha} f(y)dy = \frac{\alpha-1}{x^\alpha} \int_{y \leq x} y^{\alpha-1} f(y)dy.$$

This implies that $\left[g(x)x^\alpha\right]' \geq 0$. This is also a special case of model (1.3.5) with $w(x) = x^{-\alpha}$.

Since model (1.3.5) is similar to model (1.3.4), all discussions in this Chapter are parallel to Chapter V. Specifically, we first consider $\hat{f}_n^*$, the NPMLE estimator of f, and its large sample properties for model (1.3.5) in Section 6.2. The construction of the estimator $\hat{f}_n^*$ is based on a transformation of the data. In Section 6.3 we investigate kernel estimators of f and their properties, and a similar discussion for modified histogram type estimators will be given in Section 6.4.

6.2 The NPMLE and Its Asymptotic Properties

The NPMLE $\hat{f}_n^*$ of f for model (1.3.5) is a density function such that

$\hat{f}_n^*(x)/w(x)$ is a monotone increasing function of x and

$$\prod_{i=1}^{n} \hat{f}_n^*(X_i)$$

is maximal. From the results of Chapter V, one would expect that $\hat{f}_n^*$ is the product of $w(x)$ and a histogram estimator which is nondecreasing. If so, we need a finite right end point for the histogram. Thus the following assumption is made.

ASSUMPTION: $\beta_F = \inf\{x : F(x) = 1\} < \infty$.

Now we make the following transformation

(6.2.1) $$Y = \int_X^{\beta_F} w(z)dz \equiv W_*(X)$$

for any $X \in (0, \beta_F)$. If the pdf of X is f, then the pdf of Y is given by

(6.2.2) $$l(y) = \frac{f(W_*^{-1}(y))}{w(W_*^{-1}(y))}, \qquad y \in (0, M^*)$$

where $M^* = W_*(0) = \int_0^{\beta_F} w(z)dz \leq \infty$. It is easy to check that l is monotone decreasing if and only if f/w is monotone increasing. As in Chapter V, we introduce the following histogram type estimator

(6.2.3) $$\begin{aligned} f_n(x) &= w(x) \sum_{i=1}^{n} g_i I_{[X_{(i)} \leq x < X_{(i+1)}]} \\ &= [w(x)] \times [\text{histogram estimator}], \end{aligned}$$

where $0 = X_{(0)} \leq X_{(1)} \leq \ldots \leq X_{(n)} \leq X_{(n+1)} = \beta_F[= W_*^{-1}(0)]$ are the order statistics of $X_1, \ldots, X_n$. Now we claim that the NPMLE is

(6.2.4) $$\hat{f}_n^*(x) = w(x)\hat{l}_n(W_*(x)), \quad x \in [0, \beta_F),$$

where $\hat{l}_n$ is Grenander's NPMLE of l based on data $Y_1, \ldots, Y_n$. We are going to prove this claim in Lemma 6.2.1 and Theorem 6.2.1. Their proofs are parallel to those of Lemma 5.2.1 and Theorem 5.2.1.

Lemma 6.2.1. The NPMLE $\hat{f}_n^*$ of f is the product of $w(x)$ and a step function with breakpoints (jumps) at the order statistics $X_{(i)} (1 \le i \le n)$.

Proof. The log-likelihood of the sample with order statistics $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$ is

$$L(f) = \sum_{i=1}^{n} \log[f(X_{(i)})].$$

Define

$$\begin{aligned} f^*(x) &= \begin{cases} 0 & \text{if } x < X_{(1)}; \\ cw(x)\frac{f(X_{(i)})}{w(X_{(i)})} & \text{if } X_{(i)} \le x < X_{(i+1)}; \\ 0 & \text{if } x \ge X_{(n+1)}. \end{cases} \\ &= cw(x) \sum_{i=1}^{n} \frac{f(X_{(i)})}{w(X_{(i)})} I_{[X_{(i)} \le x < X_{(i+1)}]}, \end{aligned}$$

where c is a normalizing constant. Observe that

$$L(f^*) = n \log(c) + L(f) \ge L(f)$$

since

$$\begin{aligned} \frac{1}{c} &= \sum_{i=1}^{n} \frac{f(X_{(i)})}{w(X_{(i)})} \int_{X_{(i)}}^{X_{(i+1)}} w(y)dy \\ &\le \sum_{i=1}^{n} \int_{X_{(i)}}^{X_{(i+1)}} \frac{f(y)}{w(y)} w(y)dy \\ &\quad \text{(by the increasing property of } f(y)/w(y)) \\ &= \int_{X_{(1)}}^{X_{(n+1)}} f(y)dy = \int_{X_{(1)}}^{\beta_F} f(y)dy \\ &\le \int_{0}^{\beta_F} f(y)dy = 1. \end{aligned}$$

Thus for every density $f_n(x)$ there exists a density function $f_n^*(x)$ such that $f_n^*(x)$ is the product of $w(x)$ and a step function with breakpoints at the order statistics for which $L(f_n^*) \geq L(f_n)$. ∎

Lemma 6.2.1 asserts that the form of the NPMLE of f is the product of $w(x)$ and a data dependent histogram. Next we have to use Lemma 5.2.2.

When Lemma 5.2.2 is applied with $A_i = [X_{(i)}, X_{(i+1)}), (1 \leq i \leq n)$, then it is easily seen that among all densities which are the products of $w(x)$ and all step functions with breakpoints at the order statistics, the likelihood product

$$\prod_{i=1}^{n} f_n(X_i),$$

is maximized if we take a density which on $[X_{(i)}, X_{(i+1)})$, takes the value

$$\frac{w(x)}{n \int_{X_{(i)}}^{X_{(i+1)}} w(y)dy}$$

since the empirical measure of each interval is precisely one.

As in Chapter V, we use $I_1, \ldots, I_k$ to denote a partition of indices $\{1, \ldots, n\}$ determined by the breakpoints of the smallest concave majorant of the empirical distribution function of $\{Y_i\}_{i=1}^{n}$ defined by (6.2.1).

Theorem 6.2.1. The NPMLE $\hat{f}_n^*$ is the density whose distribution function is the following

$$\hat{F}_n^*(x) = \begin{cases} 1 - \hat{L}_n(W_*(x)), & \text{if } x \in [0, \beta_F), \\ 1, & \text{if } x \geq \beta_F, \end{cases}$$

where $\hat{L}_n$, is the smallest concave majorant of the empirical distribution functions L_n, based on data $Y_1, \ldots, Y_n$. Here we have used the fact $\hat{L}_n(M^*) = 1$.

The proof is the same as that of Theorem 5.2.1 except $A_i = (X_{(i-1)}, X_{(i)}]$ is replaced by $A_i = [X_{(i)}, X_{(i+1)})$, $\Delta_i = \int_{X_{(i-1)}}^{X_{(i)}} w(x)dx$ by $\Delta_i = \int_{X_{(i)}}^{X_{(i+1)}} w(x)dx$, and $\int_0^\infty$ by $\int_0^{\beta_F}$. So we omit it here.

The NPMLE $\hat{f}_n^*$ in (6.2.4) has the following convergence properties:

Theorem 6.2.2. $\int_0^{\beta_F} |\hat{f}_n^*(x) - f(x)|dx \to 0$ almost surely if and only if f/w is monotone increasing.

Proof. We apply Theorem 8.3 of Devroye (1987) which asserts

$$\int_0^{M^*} |\hat{r}_n(y) - r(y)|dy \to 0$$

almost surely if and only if r is monotone decreasing. However,

$$\begin{aligned}
\int_0^{M^*} |\hat{r}_n(y) - r(y)|dy &= \int_0^{M^*} \left|\hat{r}_n(y) - \frac{f(W_*^{-1}(y))}{w(W_*^{-1}(y))}\right| dy \\
&= \int_0^{M^*} |w(W_*^{-1}(y))\hat{r}_n(y) - f(W_*^{-1}(y))| \frac{dy}{w(W_*^{-1}(y))} \\
&= \int_0^{\beta_F} |w(x)\hat{r}_n(W_*(x)) - f(x)|dx \\
&= \int_0^{\beta_F} |\hat{f}_n^*(x) - f(x)|dx,
\end{aligned}$$

and r is monotone decreasing if and only if f/w is monotone increasing. So the result follows immediately. ∎

Theorem 6.2.3. Let $X_1, \ldots, X_n$ be independent observations generated by f such that $f(x)/w(x)$ is a monotone increasing function on $[0, \beta_F)$ which has a nonzero derivative $\left[f/w\right]'$ at a point $t \in (0, \beta_F)$. If $\hat{f}_n^*$ is the NPMLE of f, then

$$n^{1/3}[w(t)]^{-2/3} \left|\frac{1}{2}\left[\frac{f(t)}{w(t)}\right]\left[\frac{f(t)}{w(t)}\right]'\right|^{-1/3} \left[\hat{f}_n^*(t) - f(t)\right] \to 2Z$$

in distribution, where Z is distributed as the location of the maximum of the process $[L(u)-u^2, u \in R^1]$, and L is standard Brownian motion on R^1 originating from zero (i.e., $L(0)=0$).

Proof. Let $a(t) = f(t)/w(t)$. By using the notation of Groeneboom (1985) we have

$$P(\hat{f}_n^*(t) - f(t) \leq xn^{-1/3}[w(t)]^{2/3}|\frac{1}{2}a(t)a'(t)|^{1/3})$$

$$\begin{aligned}
&= P[w(t)\hat{r}_n(W_*(t)) - r(W_*(t))w(t) \leq xn^{-1/3}[w(t)]^{2/3}|\frac{1}{2}a(t)a'(t)|^{1/3}] \\
&= P[\hat{r}_n(W_*(t)) - r(W_*(t)) \leq \frac{x}{w(x)}n^{-1/3}[w(t)]^{2/3}|\frac{1}{2}a(t)a'(t)|^{1/3}] \\
&= P\left(n^{1/3}|\frac{1}{2}a(t)a'(t)\frac{1}{w(t)}|^{-1/3}[\hat{r}_n(W_*(t)) - r(W_*(t))] \leq x\right) \\
&\to P(2Z \leq x).
\end{aligned}$$

The weak convergence follows by Theorem 2.1 of Groeneboom (1985). ∎

Theorem 6.2.4. Assume that f/w has two bounded continuous derivatives on $(0, \beta_F)$, and that $[f/w]' > 0$ on $(0, \beta_F)$. Then

$$n^{\frac{1}{3}}E\left[\int_0^{\beta_F} |\hat{f}_n^*(x) - f(x)|dx\right] \to c\int_0^{\beta_F}\left[\frac{1}{2}w(x)f(x)|[f(x)/w(x)]'|\right]^{\frac{1}{3}}dx$$

where $c \approx 0.82$ is a universal constant.

Proof. Applying Theorem 8.4 of Devroye (1987) we have

$$n^{\frac{1}{3}}E\left[\int_0^{M^*} |\hat{r}_n(y) - r(y)|dy\right] \to c\int_0^{M^*}\left[\frac{1}{2}r(y)|r'(y)|\right]^{\frac{1}{3}}dy$$

where $c \approx 0.82$ is a universal constant. The result follows from simple transformation $y = \int_x^{\beta_F} w(z)dz = W_*(x)$ and the facts:

$$n^{\frac{1}{3}}E\Big[\int_0^{M^*} |\hat{r}_n(y) - r(y)|dy\Big] = n^{\frac{1}{3}}E\Big[\int_0^{\beta_F} |\hat{r}_n(W_*(x)) - r(W_*(x))|w(x)dx\Big]$$
$$= n^{\frac{1}{3}}E(\int_0^{\beta_F} |\hat{f}_n^*(x) - f(x)|dx);$$
$$c\int_0^{M^*}\Big[\frac{1}{2}r(y)|r'(y)|\Big]^{\frac{1}{3}}dy = c\int_0^{\beta_F}\Big[\frac{1}{2}r(W_*(x))|r'(W_*(x))|\Big]^{\frac{1}{3}}w(x)dx$$
$$= c\int_0^{\beta_F}\Big[\frac{1}{2}w(x)f(x)\Big|\Big[f(x)/w(x)\Big]'\Big|\Big]^{\frac{1}{3}}dx.$$

So the proof is complete. ∎

In the following we discuss the asymptotic normality of the L_1−norm

$$||\hat{f}_n^* - f||_{1,\beta_F} = \int_0^{\beta_F} |\hat{f}_n^*(t) - f(t)|dt.$$

We assume that $[f(x)/w(x)]' > 0$ for $x \in (0, \beta_F)$ and $[f(x)/w(x)]^{(2)}$ is bounded and continuous. Then we have the following result:

Theorem 6.2.5. (Asymptotic normality of the L_1−norm $||\hat{f}_n^* - f||_{1,\beta_F}$) Assume that w is differentiable. Then

$$n^{\frac{1}{6}}\{n^{\frac{1}{3}}||\hat{f}_n^* - f||_{1,\beta_F} - C\} \to N(0, \sigma^2),$$

where

$$C = 2E|V(0)|\int_0^{\beta_F}\Big|\frac{1}{2}\frac{f(x)}{w(x)}(f'(x)w(x) - w'(x)f(x))\Big|^{\frac{1}{3}}dx,$$
$$\approx 0.82\int_0^{\beta_F}\Big|\frac{1}{2}\frac{f(x)}{w(x)}(f'(x)w(x) - w'(x)f(x))\Big|^{\frac{1}{3}}dx,$$
$$\sigma^2 = 8\int_0^{\infty}\mathrm{Cov}(|V(0)|, |V(a) - V(0)|)da$$
$$\approx 0.17.$$

Proof. Applying Theorem 3.1 of Groeneboom (1985) and noting the following facts:

$$||\hat{r}_n - r||_{1,M^*} = ||\hat{f}_n^* - f||_{1,\beta_F};$$

$$\int_0^{M^*} |\frac{1}{2}r(y)r'(y)|^{\frac{1}{3}} dy = \int_0^{\beta_F} \left|\frac{1}{2}w(x)f(x)\left(\frac{f(x)}{w(x)}\right)'\right|^{\frac{1}{3}} dx$$
$$= \int_0^{\beta_F} \left|\frac{1}{2}\frac{f(x)}{w(x)}\left[f'(x)w(x) - w'(x)f(x)\right]\right|^{\frac{1}{3}} dx,$$

the result follows immediately. ∎

6.3 Kernel Estimators and Their Properties

In this section we are going to discuss kernel estimators of f and their properties under condition (6.1.1). As in Chapter V, we first study asymmetric kernels and then symmetric kernels. The kernel estimator of f is defined by

$$\hat{f}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x - X_i}{h}\right) = \frac{1}{nh}\sum_{i=1}^{n} k_h(x - X_i) \tag{6.3.1}$$

where the kernel k and the sequence $\{h = h_n\}$ are defined in Chapter V.

Let E denote the expectation with respect to the joint distribution of $X_1, \ldots, X_n$, and let $C > 0$ be a fixed constant, and let

$$U_C = \left\{ \begin{array}{ll} f(x) : f(x) \quad \text{is a density on}[0, \beta_F) \quad \text{such that} \\ \qquad [f(x)/w(x)]' \geq 0, \quad f(\beta_F)/w(\beta_F) \leq C \end{array} \right\}.$$

Theorem 6.3.1. Let $\hat{f}_n$ be defined by (6.3.1). Assume that the kernel k is left sided, that is, $k(x)I_{(0,\infty)}(x) = 0$. Then for all $f \in U_C$,

$$E\int_0^{\beta_F} |\hat{f}_n(x) - f(x)|dx \leq \left(\frac{C}{nh}\right)^{\frac{1}{2}} N_1(h) + Chk_1N,$$

where

$$k_1 = \int |x| k(x) dx,$$

$$N_1(h) = \int_0^{\beta_F} [w_{h,2}^+(x)]^{\frac{1}{2}} dx,$$

$$N = \liminf_{h \to 0+} \int_0^{\beta_F} w_{h,1}^+(x) dx.$$

The proof is similar to that of Theorem 5.3.1. In order to make the thesis self-contained, we give the proof as follows:

Proof. As usual, we split $E \int_0^{\beta_F} |\hat{f}_n(x) - f(x)| dx$ into two parts

$$\begin{aligned} E \int_0^{\beta_F} |\hat{f}_n(x) - f(x)| dx &\le E \int_0^{\beta_F} |\hat{f}_n(x) - E\hat{f}_n(x)| dx \\ &\quad + \int_0^{\beta_F} |E\hat{f}_n(x) - f(x)| dx \\ &\equiv \text{VARIATION} + \text{BIAS}. \end{aligned}$$

Now for fixed x, we define $w_{h,j}^+(x) = \int_{\frac{x-\beta_F}{h}}^0 [k(u)]^j f(x - uh) du$, $f_{h,j}^+(x)$ with defined similarly for $j = 1, 2$. Then

$$\begin{aligned} \text{Var}[\hat{f}_n(x)] &\le \frac{1}{nh^2} \int_0^{\beta_F} k^2\left(\frac{x-y}{h}\right) f(y) dy \\ &= \frac{1}{nh} \int_{\frac{x-\beta_F}{h}}^0 k^2(u) f(x - uh) du \\ &\le \frac{C}{nh} \int_{\frac{x-\beta_F}{h}}^0 k^2(u) w(x - uh) du \\ &\equiv \frac{C}{nh} w_{h,2}^+(x). \end{aligned}$$

It follows that

$$\begin{aligned} \text{VARIATION} &\le \int_0^{\beta_F} \left[\text{Var}(\hat{f}_n(x)) \right]^{\frac{1}{2}} dx \\ &\le \left(\frac{C}{nh}\right)^{\frac{1}{2}} \int_0^{\beta_F} [w_{h,2}^+(x)]^{\frac{1}{2}} dx \\ &\equiv \left(\frac{C}{nh}\right)^{\frac{1}{2}} N_1(h), \end{aligned}$$

where

$$N_1(h) = \int_0^{\beta_F} [w_{h,2}^+(x)]^{\frac{1}{2}} dx.$$

On the other hand, note that $E\hat{f}_n(x) = f_{h,1}^+(x)$. As in the proof of Theorem 5.3.1, we have

$$\text{BIAS} = \int_0^{\beta_F} |E\hat{f}_n(x) - f(x)| dx \leq hk_1 D^*(f),$$

where

$$\begin{aligned} D^*(f) &= \liminf_{h\to 0+} \int_0^{\beta_F} |f_{h,1}^+(x)| dx \\ &\leq C \liminf_{h\to 0+} \int_0^{\beta_F} w_{h,1}^+(x) dx \\ &\equiv CN, \end{aligned}$$

and

$$N = \liminf_{h\to 0+} \int_0^{\beta_F} w_{h,1}^+(x) dx.$$

The result is proved. ∎

Corollary 1. Under the condition of Theorem 6.3.1 and assume that $N_1 = \sup_{h>0} N_1(h)$ exists, then

$$E\int_0^{\beta_F} |\hat{f}_n(x) - f(x)| dx \leq \left(\frac{C}{nh}\right)^{\frac{1}{2}} N_1 + Chk_1 N,$$

Corollary 2. Under the condition of Corollary 1, the bound in Corollary 1 is minimized for

$$h_n = \left(\frac{N_1}{2\sqrt{n}k_1\sqrt{C}N}\right)^{2/3}$$

and for this choice of h_n

$$E\int_0^{\beta_F} |\hat{f}_n(x) - f(x)| dx \leq \left(\frac{k_1 N}{n}\right)^{\frac{1}{3}} (CN_1)^{\frac{2}{3}} [2^{1/3} + 2^{-2/3}].$$

Corollary 3. For all n,

$$\inf_h \sup_{f \in D_C} E \int_0^{\beta_F} |\hat{f}_n(x) - f(x)| dx \leq \left(\frac{k_1 N}{n}\right)^{\frac{1}{3}} (CN_1)^{\frac{2}{3}} [2^{1/3} + 2^{-2/3}].$$

The next theorem considers a symmetric kernel.

Theorem 6.3.2. Suppose that the kernel in (6.3.1) is symmetric about zero. Then for all n and h,

$$E \int_0^{\beta_F} |\hat{f}_n(x) - f(x)| dx \leq \left(\frac{1}{nh}\right)^{\frac{1}{2}} N_1^*(h) + Chk_1 N^*,$$

where

$$\begin{aligned} k_1 &= \int |x| k(x) dx, \\ N_1^*(h) &= \int_0^{\beta_F} [w_{h,2}^*(x)]^{\frac{1}{2}} dx, \\ N^* &= \liminf_{h \to 0+} \int_0^{\beta_F} w_{h,1}^*(x) dx, \\ w_{h,j}^*(x) &= \int_{\frac{x-\beta_F}{h}}^{\frac{x}{h}} k^2(u) w(x - uh) du, \quad j = 1, 2. \end{aligned}$$

Proof. Similar to the proof of Theorem 6.3.1,

$$\begin{aligned} E \int_0^{\beta_F} |\hat{f}_n(x) - f(x)| dx &\leq E \int_0^{\beta_F} |\hat{f}_n(x) - E\hat{f}_n(x)| dx \\ &\quad + \int_0^{\beta_F} |E\hat{f}_n(x) - f(x)| dx \\ &\equiv \text{VARIATION} + \text{BIAS}. \end{aligned}$$

Now for fixed x, we define

$$w_{h,j}^*(x) = \int_{\frac{x-\beta_F}{h}}^{\frac{x}{h}} k^j(u) w(x - uh) du, \quad j = 1, 2,$$

$f^*_{h,j}(x)$ with defined similarly. Then

$$\begin{aligned}\mathrm{Var}[\hat{f}_n(x)] &\le \frac{1}{nh^2}\int_0^{\beta_F} k^2\left(\frac{x-y}{h}\right) f(y)dy \\ &= \frac{1}{nh}\int_{\frac{x-\beta_F}{h}}^{\frac{x}{h}} k^2(u) f(x-uh)du \\ &\le \frac{C}{nh}\int_{\frac{x-\beta_F}{h}}^{\frac{x}{h}} k^2(u) w(x-uh)du \\ &\equiv \frac{C}{nh} w^*_{h,2}(x).\end{aligned}$$

It follows that

$$\begin{aligned}\text{VARIATION} &\le \int_0^{\beta_F}\left[\mathrm{Var}(\hat{f}_n(x))\right]^{\frac{1}{2}} dx \\ &\le \left(\frac{C}{nh}\right)^{\frac{1}{2}} \int_0^{\beta_F} [w^*_{h,2}(x)]^{\frac{1}{2}} dx \\ &\equiv \left(\frac{C}{nh}\right)^{\frac{1}{2}} N_1^*(h),\end{aligned}$$

where

$$N_1^*(h) = \int_0^{\beta_F} [w^*_{h,2}(x)]^{\frac{1}{2}} dx.$$

On the other hand, note that $E\hat{f}_n(x) = f^*_{h,1}(x)$. Applying Theorem 7.1 of Devroye (1987) we have

$$\text{BIAS} = \int_0^{\beta_F} |E\hat{f}_n(x) - f(x)| dx \le hk_1 D^*(f),$$

where

$$\begin{aligned}D^*(f) &= \liminf_{h\to 0+} \int_0^{\beta_F} |f^*_{h,1}(x)| dx \\ &\le C \liminf_{h\to 0+} \int_0^{\beta_F} w^*_{h,1}(x) dx \\ &\equiv CN^*,\end{aligned}$$

and

$$N^* = \liminf_{h\to 0+} \int_0^{\beta_F} w^*_{h,1}(x) dx.$$

The proof is complete. ∎

6.4 Modified Histogram Type Estimators

As in Chapter V, let f be a density defined on a bounded interval $[a, a+L]$ such that f/w is monotone increasing and $f(a+L)/w(a+L) \leq C$. Let f_n be an estimator of f constructed from n iid observations from f. The risk of f_n at f is defined by (5.4.1).

In (5.4.1), we make the change of variable $y = \int_x^{\beta_F} w(z)dz = W_*(x)$, where $\beta_F \geq a+L$. Then

$$\begin{aligned}
R(f_n, f) &= E\left[\int_{W_*(a+L)}^{W_*(a)} \left|f_n(W_*^{-1}(y)) - f(W_*^{-1}(y))\right| \frac{dy}{w(W_*^{-1}(y))}\right] \\
&= E\left[\int_{W_*(a+L)}^{W_*(a)} \left|\frac{f_n(W_*^{-1}(y))}{w(W_*^{-1}(y))} - \frac{f(W_*^{-1}(y))}{w(W_*^{-1}(y))}\right| dy\right] \\
&= E\left[\int_{W_*(a+L)}^{W_*(a)} |r_n(y) - r(y)| dy\right].
\end{aligned}$$

Hence, as long as we define $f_n(x) = w(x) r_n(W_*(x))$, where r_n is Birgé's histogram estimator, then the results of Birgé (1987) also hold for our estimator. One of such results is the following.

Theorem 6.4.1. Let

$$f_n(x) = w(x) \sum_{i=0}^{p-1} (nl_i)^{-1} N_i I_{A_i}(W_*(x)),$$

where N_i denotes the number of observations $\{Y_i\}_{i=1}^n$ which belong to A_i, where $\{A_i = [y_i, y_{i+1})\}_{i=0}^{p-1}$ is a partition of $[W_*(a+L), W_*(a)]$, and $l_i = y_{i+1} - y_i$ is the length of A_i such that $\{l_i\}_{i=0}^{p-1}$ is an increasing sequence. Then

$$R(f_n, f) \leq 1.89 \left[\frac{S}{n}\right]^{\frac{1}{3}} + 0.2 \left[\frac{S}{n}\right]^{\frac{2}{3}},$$

where $S = \log[1 + C(W_*(a) - W_*(a+L))]$.

APPENDIX

DEFINITIONS AND NOTATION

A.1 Definitions

The following definitions can be found in Shorack and Wellner (1986, p.828-p.839).

Definition A.1.1. Let $\mathcal{A}$ denote a class of subsets of the sample space $\mathcal{X}$. For any finite subset D of $\mathcal{X}$, let $\#_{\mathcal{A}}(D)$ denote the number of different subsets of D that can be obtained by intersecting D with members of $\mathcal{A}$. For any $r = 0, 1, \ldots,$ define

$$m_{\mathcal{A}}(r) = \max\{\#_{\mathcal{A}}(D) : D \subset \mathcal{X}, \#(D) = r\}$$

Clearly, $m_{\mathcal{A}}(r)$ cannot be larger than 2^r. We call $m_{\mathcal{A}}(r)$ the *growth function* for $\mathcal{A}$. Let

$$v = v_{\mathcal{A}} = \begin{cases} \inf\{r : m_{\mathcal{A}}(r) < 2^r\}; & \\ \infty, & \text{if } m_{\mathcal{A}} = 2^r \text{ for all } r < \infty. \end{cases}$$

Then $v_{\mathcal{A}}$ is called the *Vapnik-Čhervonenkis index number* of $\mathcal{A}$. If $v_{\mathcal{A}} < \infty$, then $\mathcal{A}$ is called a *Vapnik-Čhervonenkis class* or VC class. We always use $\mathcal{C}$ to denote a *Vapnik-Čhervonenkis class.*

The next definition requires the following notation.

Let $\mathcal{L}_2 = \mathcal{L}_2(\mathcal{X}, \mathcal{F}, P)$ be the Hilbert space of (equivalent classes of) $\mathcal{F}-$measurable functions $f : \mathcal{X} \longrightarrow R^1$ such that $\int f^2 dP < \infty$. Let $\mathbf{W}_P$ be a *Gaussian process* indexed by $\mathcal{L}_2$. That is, the rvs $\{\mathbf{W}_P(f) : f \in \mathcal{L}_2\}$ are jointly Gaussian with mean 0 and covariance

$$P(fg) = \int_{\mathcal{X}} fg dP$$

for $f, g \in \mathcal{L}_2$. Let

$$\mathbf{Z}_P(f) \stackrel{d}{=} \mathbf{W}_P(f) - P(f)\mathbf{W}_P(1).$$

Definition A.1.2. A class $\mathcal{B} \subset \mathcal{L}_2$ will be called a $\mathbf{Z}_P$–BUC class if and only if the process $\mathbf{Z}_P(f)(\omega)$ can be chosen so that for all ω the sample functions $f \longrightarrow \mathbf{Z}_P(f)(\omega)$ restricted to $f \in \mathcal{B}$ are bounded and

$$\rho_P(f,g) = \{P[(f-g)^2] - [P(f-g)]^2\}^{\frac{1}{2}}$$

is uniformly continuous with respect to the $\mathcal{L}_2$-norm.

Definition A.1.3. A $\mathbf{Z}_P$–BUC class $\mathcal{B}$ will be called a functional Donsker class for P if and only if there exist processes $Y_j(f), Y \in \mathcal{B}$ where $Y_j \stackrel{d}{=} \mathbf{Z}_P$ are independent copies of $\mathbf{Z}_P$ for which $\mathcal{B}$ is $\mathbf{Z}_P$–BUC for each j such that

$$n^{-\frac{1}{2}} \max_{m \le n} ||\sqrt{m}\mathbf{Z}_m - \sum_{j=1}^{m} Y_j||_{\mathcal{B}} \le M_n,$$

where

$$M_n \xrightarrow{P} 0 \quad \text{as} \quad n \longrightarrow \infty$$

$$\mathbf{Z}_m = \sqrt{m}[\mathbf{P}_m - P]$$

$$\mathbf{P}_m = \frac{1}{m}\sum_{i=1}^{m} \delta_{X_i}, \quad \text{(empirical measure)}$$

and $||.||$ is the supremum norm.

A.2 Special Notation for Chapters II and III

$$\mathcal{Q}(F) \equiv \{q_e(x)I_C : C \in \mathcal{C}, \quad q_e(x) \geq 0 \quad \text{is fixed with} \quad [F](q_e) < \infty\}$$

$$\text{where} \quad x \in R^1, \quad \text{and} \quad [F](q_e) \equiv \int_{-\infty}^{\infty} q_e(x)dF(x)$$

$$\mathcal{C} \equiv \text{Vapnik-Chervonenkis class of subsets of sample space} \quad \mathcal{X}$$

$$Q(H_F) \equiv \left\{ \frac{q}{h \circ F} : q \in \mathcal{Q}(F) \right\}$$

$$\mathcal{X}^+ \equiv \bigcup_{i=1}^{s} \{x : w_i(x) > 0\}$$

$$n \equiv n_1 + \cdots + n_s \quad \text{(total sample size)}$$

$$\lambda_{ni} \equiv n_i/n \quad \text{(sample fractions)} \quad i = 1, \ldots, s$$

$$\lambda_i \equiv \lim_{n \to \infty} \lambda_{ni} > 0 \quad \text{(if the limit exists)}, \quad i = 1, \ldots, s$$

$$\Delta \equiv \text{diagonal matrix with diagonal} \quad (\lambda_1, \ldots, \lambda_s)$$

$$\overline{G}_n(x) = \sum_{i=1}^{s} \lambda_{ni} G_i(x) \quad \text{average distribution of cdfs}$$

$$G_1(x), \ldots, G_s(x) \quad \text{for} \quad x \in R^1$$

$$\overline{G}(x) = \sum_{i=1}^{s} \lambda_i G_i(x) \quad \text{for} \quad x \in R^1$$

$$\mathbf{G}_n(x) = \frac{1}{n} \sum_{i=1}^{s} \sum_{j=1}^{n_i} I_{(-\infty,x]}(X_{ij}), \quad x \in R^1$$

$$\underline{V}^T = (V_1, \ldots, V_s), \quad \text{(solution of equation (2.2.5))}$$

$$\underline{\mathbf{V}}_n^T = (\mathbf{V}_{n1}, \ldots, \mathbf{V}_{ns}), \quad \text{(solution of equation (2.2.15))}$$

$$\underline{W}^T = (W_1, \ldots, W_s), \quad \text{where}$$

$$W_i = \int_{-\infty}^{\infty} w(F(y))dF(y) = \int_0^1 w_i(u)du, \quad 1 \leq i \leq s$$

$$\underline{B}^T = (B_1, \ldots, B_s), \quad \text{where} \quad B_i = V_i B_s, \quad 1 \leq i \leq s$$

$$\underline{\mathbf{B}}_n^T = (\mathbf{B}_{n1}, \ldots, \mathbf{B}_{ns}), \quad \text{where}$$

$$\mathbf{B}_{ni} = \mathbf{V}_{ni}\mathbf{B}_{ns}, \quad i = 1, \ldots, s-1, \quad \text{(see (2.2.17))}$$

$$\mathbf{B}_{ns} = \frac{1}{\int_{-\infty}^{\infty} \left[\sum_{i=1}^{s} \frac{\lambda_{ni} w_i(y)}{\mathbf{V}_{ni}}\right]^{-1} d\mathbf{G}_n(y)}, \quad \text{(see (2.2.18))}$$

$$r_n(\underline{\mathbf{B}}_n) = \left[\sum_{i=1}^{s} \frac{\lambda_{ni} w_i(x)}{\mathbf{B}_{ni}}\right]^{-1}, \quad x \in R^1$$

$$r(\underline{B}) = \left[\sum_{i=1}^{s} \frac{\lambda_i w_i(x)}{B_i}\right]^{-1}, \quad x \in R^1$$

A.3 Special Notation for Chapter IV

$$\underline{\mathbf{p}}_n^T = (\hat{p}_n(1), \ldots, \hat{p}_n(s)), \quad \text{where} \quad \hat{p}_n(i) = \frac{1}{n} \sum_{j=1}^{n} I_{[K_j = i]}, \quad 1 \le i \le s$$

$$\underline{p}^T = (p(1), \ldots, p(s)), \quad \text{where} \quad p(i) = P(K = i), \quad 1 \le i \le s$$

$$\underline{V}^T = (V_1, \ldots, V_s), \quad \text{(solution of equation (4.2.3))}$$

$$\underline{\mathbf{V}}_n^T = (\mathbf{V}_{n1}, \ldots, \mathbf{V}_{ns}), \quad \text{(solution of equation (4.2.7))}$$

$$\underline{W}^T = (W_1, \ldots, W_s), \quad \text{where} \quad W_i = V_i W_s, \quad 1 \le i \le s$$

$$\underline{\mathbf{W}}_n^T = (\mathbf{W}_{n1}, \ldots, \mathbf{W}_{ns}), \quad \text{where}$$

$$\mathbf{W}_{ni} = \mathbf{V}_{ni}\mathbf{W}_{ns}, \quad i = 1, \ldots, s-1, \quad \text{(see (4.2.9))}$$

$$\mathbf{W}_{ns} = \frac{1}{\int_{-\infty}^{\infty} \left[\sum_{i=1}^{s} \frac{\hat{p}_n(i) w_i(y)}{\mathbf{V}_{ni}}\right]^{-1} d\mathbf{G}_n(y)}, \quad \text{(see (4.2.10))}$$

$$\mathbf{G}_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty, x]}(X_i), \quad x \in R^1$$

$$r(\underline{\mathbf{W}}_n, \underline{\mathbf{p}}_n) = \left[\sum_{i=1}^{s} \frac{\hat{p}_n(i) w_i(x)}{\mathbf{W}_{ni}}\right]^{-1}, \quad x \in R^1$$

$$r(\underline{\mathbf{W}}, \underline{p}) = \left[\sum_{i=1}^{s} \frac{p(i) w_i(x)}{W_i}\right]^{-1}, \quad x \in R^1$$

BIBLIOGRAPHY

[1] Anderssen, R. S., and Jakeman, A. J. (1975). Abel type integral equations in stereology. II. Computational methods of solution and the random spheres approximation. *J. of Microscopy* **105**, 135-153.

[2] Baxter, L. A. (1982). Reliability applications of the relevation transform. *Naval Research Logistics Quarterly* **29**, 323-330.

[3] Birgé, L.(1987a). Estimating a density under order restrictions: Nonasymptotic minimax risk. *Ann. Statist.* **15**, 995-1012.

[4] Birgé, L. (1987b). On the risk of histogram estimating decreasing densities. *Ann. Statist.* **15**, 1013-1022.

[5] Birgé, L. (1989). The Grenander estimator: A nonasymptotic approach. *Ann. Statist.* **17**, 1532-1549.

[6] Boyles, R. A., and Samaniego, F. J. (1986). Estimating a distribution function based on nomination sampling. *J. Amer. Statist. Assoc.* **81**, 1039-1045.

[7] Cobby, J. M., Ridout, M. S., Bassett, P. J., and Large, R. V. (1985). An investigation into the use of ranked set sampling on grass and grass-clover swards. *Grass and Forage Science* **40**, 257-263.

[8] Cox, D. R. (1969). Some sampling problems in technology. In *New Developments in Survey Sampling* (N. L. Johnson and H. Smith, Jr., eds), 506-527, Wiley, New York.

[9] Datta, S. (1992). Some nonasymptotic bounds for L_1 density estimation using kernel. *Ann. Statist.* **20**, 1658-1667.

[10] Dell, T. R., and Clutter, J. L. (1972). Ranked set sampling theory with order statistics background. *Biometrics* **28**, 545-555.

[11] Devroye, L. (1987). *A Course in Density Estimation.* Birkhauser, Boston.

[12] Dudley, R. M., and Philipp, W. (1983). Invariance principles for sums of Banach space valued random elements and empirical processes. *Z. Wahrsch. verw. Gebiete* **62**, 509-552.

[13] Feller, W. (1971). *Introduction to Probability Theory and Its Applications*, Vol.2, 2d ed. Wiley, New York.

[14] Gill, Richard D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method (part 1), *Scand J. Statist.* **16**, 97-128.

[15] Gill, Richard D., Vardi, Y., and Wellner, Jon A. (1988). Large sample theory of empirical distributions in biased sampling models. *Ann. Statist.* **16**, 1069-1112.

[16] Grenander, U. (1956). On the theory of mortality measurements. II. *Skand. Aktuar.* **39**, 125-153.

[17] Groeneboom, P. (1985). Estimating a monotone density. *In Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (L. M. Le Cam and R. A. Olshen, eds.) **2**, 539-555. IMS, Hayward, Calif.

[18] Halls, L. K., and Dell, T. R. (1966). Trail of ranked set sampling for forage yields, *Forest Science* **12**, 22-26.

[19] He, S. Y., and Yang, G. L. (1993). Estimating lifetime distribution under different sampling plans. *Statistical Decision Theory and Related Topics, V*, (to appear).

[20] Keiding, N., Jensen, S. T., and Ranek, L. (1972). Maximum likelihood estimation of the size distribution of liver cell nuclei from the observed distribution in a plane section. *Biometrics* **28**, 813-829.

[21] Krakowski, M. (1973). The relevation transform and a generalization of the gamma distribution function. *Revue Francaise d'Automatique, Informatique et Recherche Opérationnelle* **7**, Sér. V-2, 107-120.

[22] McIntyre, G. A. (1952). A method of unbiased selective sampling, using ranked sets. *Australian J. of Agricultural Research* **3**, 385-390.

[23] Miller, R. (1981). *Survival Analysis*, Wiley, New York.

[24] Patil, G. P., and Rao, C. R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics* **34**, 179-189.

[25] Prakasa-Rao, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31**, 23-36.

[26] Shorack, G. R., and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics.* John Wiley, New York.

[27] Stokes, S. L. (1977). Ranked set sampling with concomitant variables. *Comm. in Stat. Part A-Theory and Methods* **6**, 1207-1211.

[28] Stokes, S. L. (1980). Estimation of variance using judgment ordered ranked set samples. *Biometrics* **36**, 35-42.

[29] Stokes, S. L., and Sager, T. W. (1988). Characterization of a ranked-set sample with application to estimating distribution functions. *J. Amer. Statist. Assoc.* **83**, 374-381.

[30] Takahasi, K., and Wakimoto, K. (1968). On unbiased estimates of the

population mean based on the sample stratified by means of ordering. *Ann. Math. Statist.* **20**, 1-31.

[31] Vardi, Y. (1982). Nonparametric estimation in the presence of length bias. *Ann. Statist.* **10**, 616-620.

[32] Vardi, Y. (1985). Empirical distributions in selection bias model. *Ann. Statist.* **13**, 178-203.

[33] Vardi, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation. *Biometrika* **76**, 4, 751-61.

[34] Vardi, Y., and Zhang, C.-H. (1992). Large sample study of empirical distributions in a random-multiplicative censoring model. *Ann. Statist.* **20**, 1022-1039.

[35] Wicksell, S. D. (1925). The corpuscle problem. *Biometrika* **17**, 82-99.

[36] Wicksell, S. D. (1926). The corpuscle problem II. *Biometrika* **18**, 151-172.

[37] Willemain, T. R. (1980). Estimating the population median by nomination sampling. *J. Amer. Statist. Assoc.* **75**, 908-911.