



Available at

www.ElsevierMathematics.com

POWERED BY SCIENCE @ DIRECT®

Journal of Statistical Planning and
Inference 128 (2005) 165–190

journal of
statistical planning
and inference

www.elsevier.com/locate/jspi

Asymptotic properties of kernel density estimation with complex survey data

Trent D. Buskirk^{a,1}, Sharon L. Lohr^{b,*,2}

^aBehavioral Research Center, American Cancer Society, 1599 Clifton Road, Atlanta, GA 30329, USA

^bDepartment of Mathematics and Statistics, Arizona State University, Tempe, AZ 85287-1804, USA

Received 4 July 2002; accepted 15 September 2003

Abstract

Kernel density estimation has been used with great success with data that may be assumed to be generated from independent and identically distributed (iid) random variables. The methods and theoretical results for iid data, however, do not directly apply to data from stratified multistage samples. We present finite-sample and asymptotic properties of a modified density estimator introduced in Buskirk (Proceedings of the Survey Research Methods Section, American Statistical Association (1998), pp. 799–801) and Bellhouse and Stafford (Statist. Sin. 9 (1999) 407–424); this estimator incorporates both the sampling weights and the kernel weights. We present regularity conditions which lead the sample estimator to be consistent and asymptotically normal under various modes of inference used with sample survey data. We also introduce a superpopulation structure for model-based inference that allows the population model to reflect naturally occurring clustering. The estimator, and confidence bands derived from the sampling design, are illustrated using data from the US National Crime Victimization Survey and the US National Health and Nutrition Examination Survey.

© 2003 Elsevier B.V. All rights reserved.

MSC: primary 62D05; secondary 62G07; 62G20

Keywords: Smoothing; Sampling weights; Nonparametric density estimation; Quantile estimation; Superpopulation models

* Corresponding author. Tel.: +1-408-965-4440; fax: +1-480-965-8119.

E-mail address: sharon.lohr@asu.edu (S.L. Lohr).

¹ Research supported in part by a grant from the Gallup Organization.

² Research supported by the US Bureau of Justice Statistics, and by the National Science Foundation under Grant No. 0105852.

1. Introduction

Nonparametric kernel density estimation is commonly used to display the shape of a data set without relying on a parametric model. Rosenblatt (1956) and Parzen (1962) provided early results on kernel density estimation; since then, much research has been done in the area. Wand and Jones (1995) summarized some of the work done through the mid-1990s.

In most previous work it is assumed that Y_1, \dots, Y_n are independent and identically distributed (iid) continuous random variables with common density f . Using a kernel function K and a positive bandwidth b , $f(x)$ is estimated by

$$\hat{f}_{\text{iid}}(x; b) = (bn)^{-1} \sum_{i=1}^n K[(x - Y_i)/b]. \quad (1)$$

Density estimates are of interest in survey samples for many of the same reasons they are of interest in the iid setting: they provide a snapshot of the shape of the data and a means for comparing different populations or subgroups. In the National Crime Victimization Survey (NCVS), for example, it is of interest to estimate the density of the variable “victim age” for different types of victimizations. Current Population Survey (CPS) data can be used to estimate the density of household income for two-person households. Data from the National Health and Nutrition Examination Surveys (NHANES) can provide estimates of the density of cholesterol in the population of 40–49-year-old African-American men. In each of these examples, a smoothed density estimate provides more information than a series of point estimates for means and quantiles.

The iid assumptions used for model (1), however, are generally not valid for data from a complex survey. Stratification may reflect a violation of the identically distributed assumption, and clustering may violate the independence assumption. To adapt density estimation to the survey setting, Buskirk (1998, 1999) and Bellhouse and Stafford (1999) independently proposed incorporating the survey weights into a density estimator of the form in (1). Buskirk (1998, 1999) concentrated on a direct analogue of (1); Bellhouse and Stafford (1999) considered in addition the use of binned survey data in density estimation. Although density estimation has been considered for data from complex surveys, asymptotic properties of the estimators have not been thoroughly explored. In this paper we establish sufficient conditions for a kernel density estimator, defined in Section 2, to be consistent and asymptotically normal.

The asymptotic properties established in this paper provide a foundation for inference about a density when estimated using complex survey data. To derive the properties, however, we need to specify frameworks for asymptotic inference with samples from finite populations. Three frameworks are considered: model-based, design-based, and a combination of the design-based and model-based frameworks.

In the traditional (model-based) approach to inference, the function in (1) estimates an underlying continuous density function $f(x)$ that is assumed appropriate for a population of interest. It is assumed, for example, that the observed values of cholesterol level from the sample are the realized values of random variables having density $f(x)$. The joint probability distribution of Y_1, \dots, Y_n provides the basis for inference about

$f(x)$, and it is assumed that the sampling design is noninformative—that is, the probabilities of inclusion only depend on the Y_i 's through concomitant variables in the model.

In the design-based approach to survey sampling, however, inferences depend on the probability distribution induced by the sampling design and not on the probability distribution from an underlying model. Inferences drawn using a design-based approach typically refer to a particular finite population of interest and usually ignore any parametric structure in the corresponding superpopulation. Quantities of interest in survey samples are often characteristics of the finite population such as means or totals—in NHANES, for example, one may be interested in the proportion of 40–49-year-old men who have cholesterol levels greater than 200 in the year 2001. The distribution of cholesterol levels in the finite population is discrete, however. Thus, an underlying theoretical density function assumed to generate the finite population—a superpopulation density—will often be the quantity of interest. There is thus an assumed two-stage process: first, the elements in the finite population are assumed to be realizations of random variables with a joint probability distribution. Then the sample is selected, using a probability sampling design, from the elements in the finite population.

We are thus interested in the asymptotic properties of density estimation in three settings: with design-based inference, with model-based inference that ignores the sampling design, and under a combined approach that assumes the two-stage process described above. Hartley and Sielken (1975) presented a superpopulation framework within which survey design-based inference could be imbedded. Pfeffermann (1993) formalized the combined inference framework in which inference could be made to the superpopulation via inference through the design. With this combined approach emphasis is typically given to estimating parameters associated with the superpopulation model using design-based consistent estimators of finite population quantities; the finite population quantities, in turn, are consistent estimators of the superpopulation parameters under the proposed model. Graubard and Korn (2002) studied variance estimation using a combined inference approach. Rao (1999) discussed the importance of estimators that are consistent under the sampling design—they provide robustness against model misspecification that can occur in a pure model-based approach.

Pfeffermann (1993) and Krieger and Pfeffermann (1997) assumed that finite population values were generated by iid random variables. This assumption works for determining properties of estimated means and totals; for density estimation, however, it is incompatible with a more realistic superpopulation structure in which different strata or clusters have different underlying densities. To derive properties of the estimator under model-based and combined approaches to inference, we propose an expanded superpopulation model for densities in Section 3. The new superpopulation model allows different densities in different strata, and incorporates dependence for clustering in the population.

Nonparametric smoothing has been used in other problems in survey sampling. Korn et al. (1997) introduced a method for smoothing the empirical cumulative distribution function. Korn and Graubard (1998a) suggested nonparametric smoothing as a way to display bivariate relations from survey data, and Cowling et al. (1996) proposed smoothing methods for spatial survey data. Breidt and Opsomer (2000) used

nonparametric smoothing with auxiliary information for regression-type estimators of population totals, and Bellhouse and Stafford (2001) developed estimators for nonparametric regression functions with survey data. Many of the issues and results discussed in this paper apply to those settings as well.

The organization of the paper is as follows. The estimator and some finite population properties are given in Section 2. Consistency and asymptotic normality under the various frameworks of inference are addressed in Section 3. Section 4 provides a discussion of bandwidth issues, and Section 5 treats estimation near a boundary. Section 6 presents applications to the NCVS and NHANES.

2. The sample weighted kernel density estimator

We define the sample weighted kernel density estimator in the context of a stratified two-stage sampling design. Extensions to more than two stages of sampling within each stratum are readily made.

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

Cluster samples are taken independently from each stratum; the inclusion probabilities are

$$\pi_i^{(h)} = P_D(\text{psu } i \text{ from stratum } h \text{ is included in the sample}),$$

with $\sum_{i=1}^{N_h} \pi_i^{(h)} = n_h$. Throughout, the subscript D indicates the probability distribution induced by the design. The joint inclusion probabilities are

$$\pi_{ij}^{(h)} = P_D(\text{psu's } i \text{ and } j \text{ from stratum } h \text{ are included in the sample}).$$

At the secondary sampling unit (ssu) level, psu i of stratum h has Q_{hi} ssu’s; $\pi_{l|i}^{(h)}$ is the conditional probability that ssu l of psu i is included in the sample, given that psu i is included. The $\pi_{l|i}^{(h)}$ satisfy $\sum_{l=1}^{Q_{hi}} \pi_{l|i}^{(h)} = q_{hi}$, where q_{hi} is the number of ssu’s sampled from psu i of stratum h . We have $Q_h = \sum_{i=1}^{N_h} Q_{hi}$, $Q = \sum_{h=1}^L Q_h$, and $W_h = Q_h/Q$. Thus, Q is the total number of observation units in the population, and W_h is the stratum weight for stratum h .

Let $\mathcal{U} = \{1, \dots, Q\}$ denote the index set for the finite population of Q observation units. In addition, let \mathcal{S}_h denote the sample of psu’s selected from stratum h , and \mathcal{S}_{hi} denote the sample of ssu’s selected from psu i of stratum h . The probability sample of size $q = \sum_{h=1}^L \sum_{i \in \mathcal{S}_h} q_{hi}$ that is taken from \mathcal{U} according to the probability distribution D is denoted by \mathcal{S} .

The quantity y_{hik} is observed on unit (hik) . The sampling weight for observation (hik) is $w_{hik} = [\pi_i^{(h)} \pi_{k|i}^{(h)}]^{-1}$. The weights are used in sampling practice to give unbiased estimates of population quantities; w_{hik} is the number of population observation units represented by sampled observation unit (hik) . The Horvitz–Thompson (Horvitz and Thompson, 1952) estimator of the finite population total $\sum_{h=1}^L \sum_{i=1}^{N_h} \sum_{k=1}^{Q_{hi}} y_{hik}$ is $\sum_{(hik) \in \mathcal{S}} w_{hik} y_{hik}$; $\hat{Q} = \sum_{(hik) \in \mathcal{S}} w_{hik}$ estimates Q , the total number of population observation units.

If every unit in the finite population were observed, then a density estimator corresponding to the iid estimator in (1) would be

$$\hat{f}_Q(x; b) = (bQ)^{-1} \sum_{h=1}^L \sum_{i=1}^{N_h} \sum_{k=1}^{Q_{hi}} K[(x - y_{hik})/b]. \tag{2}$$

Buskirk (1998, 1999) and Bellhouse and Stafford (1999) proposed estimating $\hat{f}_Q(x; b)$ by using the sampling weights. Define the sample weighted kernel density estimator (SWKDE) by

$$\hat{f}_G(x; b) = (b\hat{Q})^{-1} \sum_{(hik) \in \mathcal{G}} w_{hik} K[(x - y_{hik})/b]. \tag{3}$$

In a simple random sample, where $L = 1$, $\hat{Q} = Q = N$, and $w_{hik} = N/n$, (3) gives the same estimate of the density as (1).

The following assumptions for the kernel function K will be referenced throughout the paper.

- (K1) $K(x) \geq 0$ and $K(x) = K(-x)$ for all x .
- (K2) $\int_{-\infty}^{\infty} K(x) dx = 1$.
- (K3) $\int_{-\infty}^{\infty} x^4 K(x) dx < \infty$.
- (K4) There exists a constant m such that $K(x) \leq m$ for all x .

The following theorem gives two useful properties of \hat{f}_G . The proof of the theorem is straightforward and hence is omitted.

Theorem 1. Assume that (K1) and (K2) hold, and that the bandwidth b is the same for all x . Then

- (i) $\hat{f}_G(x; b)$ is a probability density function in x .
- (ii) $\int_{-\infty}^{\infty} x \hat{f}_G(x; b) dx = \hat{Q}^{-1} \sum_{(hik) \in \mathcal{G}} w_{hik} y_{hik}$.

Thus, if X has density \hat{f}_G and the bandwidth is constant, the expected value of X is the usual Hájek-type estimator of the finite population mean.

The estimator $\hat{f}_G(x; b)$ in (3) is a ratio of Horvitz–Thompson (1952) estimators of population totals. Let

$$K_b(u) = b^{-1} K[u/b] \tag{4}$$

and define

$$u_{hik}(b) = K_b(x - y_{hik}). \tag{5}$$

Then

$$\hat{U}_{hi}(b) = \sum_{k \in \mathcal{G}_{hi}} u_{hik}(b) / \pi_k^{(h)}$$

is a design-unbiased estimator of the population total in psu i of stratum h , $U_{hi}(b) = \sum_{k=1}^{Q_{hi}} u_{hik}(b)$. Also

$$\hat{U}(b) = \sum_{h=1}^L \sum_{i \in \mathcal{S}_h} \hat{U}_{hi}(b) / \pi_i^{(h)} = \sum_{(hik) \in \mathcal{S}} w_{hik} u_{hik}(b) = \hat{Q} \hat{f}_{\mathcal{G}}(x; b)$$

is the Horvitz–Thompson estimator of the finite population total $U(b) = \sum_{h=1}^L U_h(b) = \sum_{h=1}^L \sum_{i=1}^{N_h} \sum_{k=1}^{Q_{hi}} u_{hik}(b)$ and $\hat{U}(b)$ is thus design-unbiased for $U(b)$. As noted in Bellhouse and Stafford (1999), then, $\hat{f}_{\mathcal{G}}(x; b) = \hat{U}(b) / \hat{Q}$ is approximately design-unbiased for $\hat{f}_{\mathcal{U}}(x; b) = U(b) / Q$, for each x .

The design-based variance of $\hat{f}_{\mathcal{G}}(x; b)$, denoted $V_D[\hat{f}_{\mathcal{G}}(x; b)]$, may be calculated by standard survey sampling techniques described in Lohr (1999, Chapter 6). The Sen–Yates–Grundy form of the design-based variance of the population total $\hat{U}(b)$ is

$$V_D[\hat{U}(b)] = \sum_{h=1}^L \left\{ \sum_{i=1}^{N_h} \sum_{j>i}^{N_h} [\pi_i^{(h)} \pi_j^{(h)} - \pi_{ij}^{(h)}] \left[\frac{U_{hi}(b)}{\pi_i^{(h)}} - \frac{U_{hj}(b)}{\pi_j^{(h)}} \right]^2 + \sum_{i=1}^{N_h} \frac{V_D[\hat{U}_{hi}(b)]}{\pi_i^{(h)}} \right\}. \tag{6}$$

Using linearization,

$$V_D[\hat{f}_{\mathcal{G}}(x; b)] \approx V_D\{Q^{-1}[\hat{U}(b) - \hat{f}_{\mathcal{U}}(x; b)\hat{Q}]\}. \tag{7}$$

Quantiles may also be estimated using $\hat{f}_{\mathcal{G}}(x; b)$. Let θ_p represent the 100 p th percentile. Denote the ordered sample values by $y_{1:q}, \dots, y_{q:q}$ and let $w_{1:q}, \dots, w_{q:q}$ be the weights corresponding to $y_{1:q}, \dots, y_{q:q}$. Then θ_p may be estimated by $\hat{\theta}_p$ satisfying $\int_{-\infty}^{\hat{\theta}_p} \hat{f}_{\mathcal{G}}(x) dx = p$. As a consequence, $\hat{\theta}_p$ is between $y_{j:q}$ and $y_{j+1:q}$, where $\sum_{i \leq j} w_{i:q} / \hat{Q} \leq p$ and $\sum_{i > j} w_{i:q} / \hat{Q} \leq 1 - p$. This differs from many of the other methods proposed for estimating quantiles (see, for example, Woodruff, 1952; Sedransk and Sedransk, 1979; Kuk, 1988) in that we use the smoothed density to interpolate between $y_{j:q}$ and $y_{j+1:q}$. Francisco and Fuller (1991) used a step function to estimate the cumulative distribution function, and estimate θ_p by $y_{j+1:q}$. Under their regularity conditions, $\hat{\theta}_p$ has the same asymptotic properties as the estimators in Francisco and Fuller (1991). However, we expect that the smoothed estimate will behave better in small samples when the underlying superpopulation density is continuous; the small-sample performance is currently being investigated.

Chambers and Dunstan (1986), Rao et al. (1990), Chambers et al. (1992), and Rao (1994) all use auxiliary information in estimating the cumulative distribution function, so their estimates of quantiles can be expected to be somewhat more efficient if the auxiliary variables are correlated with y . We note that auxiliary information can be incorporated into the smoothed estimate of the quantile by calibrating the sampling weights.

Likewise, the weights used in $\hat{f}_{\mathcal{G}}(x; b)$ may also be modified to incorporate calibration and nonresponse adjustments. Since the quantity of interest is a density, methods discussed in Deville and Särndal (1992) may be used to ensure that the calibrated weights are nonnegative.

3. Asymptotic properties and inference

We now examine consistency and asymptotic normality of $\hat{f}_{\mathcal{G}}(x; b)$ under different sampling designs and superpopulation models. For design-based inference, we use the setup of Isaki and Fuller (1982), with a sequence of nested finite populations $\{\mathcal{U}(t)\}$ where $\mathcal{U}(i) \subset \mathcal{U}(i+1)$. The corresponding sample from $\mathcal{U}(t)$ is denoted $\mathcal{S}(t)$; the samples from successive superpopulations are not necessarily nested. Population $\mathcal{U}(t)$ has $L(t)$ strata and a total of $N(t)$ psu's and $Q(t)$ ssu's; similarly the sample $\mathcal{S}(t)$ contains a total of $n(t)$ psu's and $q(t)$ ssu's. We examine properties of the estimator as $t \rightarrow \infty$.

In the design-based setting, writing $\hat{f}_{\mathcal{G}}(x; b) = \hat{U}(b)/\hat{Q}$ allows calculation of design-based means and variances for fixed sample and population sizes. Asymptotic results, however, depend on the sample size, the sampling design, and the bandwidth b , which is assumed to converge to 0 as $n(t) \rightarrow \infty$. In a simple random sample, for example, $\hat{f}_{\mathcal{G}}(x; b)$ is the sample average of the $u_{hik}(b)$ for units in the sample. The observations $u_{hik}(b)$, though, are a function of the bandwidth b which is converging to 0; consequently, standard results on consistency and central limit theorems for finite population sampling which treat the observations as fixed quantities (see, for example, Krewski and Rao, 1981) do not directly apply. We write $b = b(t)$ to treat the convergence of b to 0. We then specify conditions for the rate of convergence of $b(t)$ in order to obtain consistency in the different frameworks for inference. In the following, the index t is suppressed unless needed for clarity.

For model-based inference in the sample survey setting, assume $Y_{111}, \dots, Y_{L(t), N_L(t), Q_L, N_L(t)}$ are distributed according to some joint probability distribution and that y_{hik} is a realization of Y_{hik} that gives the measurement in the t th finite population. Probabilities in the model-based setting are denoted by the subscript M . Since interest is often in a theoretical underlying density, $f(x)$, rather than in the finite population quantity, $\hat{f}_{\mathcal{U}}(x; b)$, we also examine asymptotic properties under the combined framework discussed in Pfeiffermann (1993) and denoted by the subscript C . The main interest in this framework is in estimating the superpopulation density f , but an estimator that is design-based consistent for the corresponding finite population quantity will be consistent under the combined distribution if the model holds for the finite population. As stated in Section 1, this approach can provide protection against model misspecification.

In this section, in addition to deriving asymptotic properties of $\hat{f}_{\mathcal{G}}(x; b)$, we also consider consistency and asymptotic normality of $\hat{f}_{\mathcal{U}}(x; b)$ under an assumed model. Model-based inference using $\hat{f}_{\mathcal{U}}(x; b)$ is of interest when the entire finite population is observed and underlying densities vary among strata and clusters, or when the model assumptions are believed to hold for a data set regardless of sampling design.

3.1. Consistency in stratified multistage sampling

Different regularity conditions for pointwise mean squared error (MSE) consistency are needed in the three frameworks for inference. Krieger and Pfeffermann (1997) discussed the notion of consistency in these frameworks. In design-based inference, since $\hat{f}_{\mathcal{G}}(x; b)$ is approximately design-unbiased for the corresponding finite population quantity $\hat{f}_{\mathcal{U}}(x; b)$, we need only show that $V_D[\hat{f}_{\mathcal{G}}(x; b)] \rightarrow 0$. Model-based MSE consistency is achieved if $MSE_M[\hat{f}_{\mathcal{G}}(x; b)] \rightarrow 0$, where the subscript M indicates that the expectation is taken using the joint probability distribution in the model; in this case, the sampling design is irrelevant and the finite population quantity $\hat{f}_{\mathcal{U}}(x; b)$ is one realization of a random process. Finally, $\hat{f}_{\mathcal{G}}(x; b)$ is pointwise MSE consistent for $f(x)$ under the combined distribution if

$$E_M\{E_D[\hat{f}_{\mathcal{G}}(x; b) - f(x)]^2 \mid \mathbf{Y}_t\} \rightarrow 0 \quad \text{as } t \rightarrow \infty,$$

where \mathbf{Y}_t is the vector of the $Q(t)$ random variables generated in finite population $\mathcal{U}(t)$.

3.1.1. Design-based consistency

We first consider consistency under the probability distribution induced by the sampling design. In Section 2, $f_{\mathcal{G}}(x; b)$ was shown to be approximately unbiased for $\hat{f}_{\mathcal{U}}(x; b)$ under the design. We use the following assumptions (C1)–(C5), adapted from Isaki and Fuller (1982) and Korn and Graubard (1998b), to show MSE consistency of $\hat{f}_{\mathcal{G}}(x; b)$ in the design-based framework.

- (C1) $n_h(t) \geq 1$ for all h .
- (C2) There exists a constant B such that $Q_{hi}(t) < B$ for all h, i , and t .
- (C3) There exists a constant $\delta > 0$ such that $\delta < \pi_i^{(h)}(t)$ and $\delta < \pi_{k|i}^{(h)}(t)$ for all h, i, k, t .
- (C4) There exist sequences $\{\alpha_h(t)\}$, $h = 1, \dots, L(t)$, such that $\pi_i^{(h)}(t)\pi_j^{(h)}(t) - \pi_{ij}^{(h)}(t) \leq \alpha_h(t)\pi_i^{(h)}(t)\pi_j^{(h)}(t)$ and $\max_{1 \leq h \leq L(t)} N_h(t)\alpha_h(t) = O(1)$.
- (C5) $N(t)b^2(t) \rightarrow \infty$ and $b(t) \rightarrow 0$ as $t \rightarrow \infty$.

Note that in this section we assume that the same bandwidth is used for both $\hat{f}_{\mathcal{G}}(x; b)$ and for $\hat{f}_{\mathcal{U}}(x; b)$; we discuss the possibility of using different bandwidths in Section 4.

Theorem 2. *Suppose that conditions (C1)–(C5) hold in stratified two-stage sampling and that the kernel function satisfies (K1)–(K4). Then $V_D[\hat{f}_{\mathcal{G}}(x; b(t))] \rightarrow 0$ as $t \rightarrow \infty$, uniformly in x .*

Theorem 2 is proven in the appendix.

Assumptions (C1)–(C5) are not unduly restrictive; they would be satisfied under many reasonable sampling designs. In the special case of stratified random sampling, $Q_{hi}(t) = 1$, $\pi_i^{(h)} = \pi_j^{(h)} = n_h/N_h$, and $\pi_{ij}^{(h)} = n_h(n_h - 1)/[N_h(N_h - 1)]$, so $\pi_i^{(h)}(t)\pi_j^{(h)}(t) - \pi_{ij}^{(h)}(t) \leq \alpha_h(t)\pi_i^{(h)}(t)\pi_j^{(h)}(t)$ if and only if $\alpha_h(t) \geq (N_h - n_h)/[n_h(N_h - 1)]$. Condition (C4) is thus equivalent in stratified random sampling to the condition

$\max N_h(t)/n_h(t) = O(1)$; it ensures that all strata are represented in the sample as $t \rightarrow \infty$. Likewise, condition (C3) prevents occurrences such as almost all of the observations being taken from one stratum. Conditions (C4) and (C5) together imply that $n(t)b^2(t) \rightarrow \infty$.

With conditions (C1)–(C5), design-based consistency is achieved either when the sample sizes within each stratum increase with t or when the number of strata $L(t)$ increases with t . If different strata have different underlying densities, the requirements that each stratum be represented asymptotically in the sample ensure that we will not miss part of the overall density because of the sampling design.

Cluster sampling presents different challenges for density estimation than does stratification. In stratified sampling, each stratum is guaranteed to be represented in the sample; if the underlying densities f_h differ among the strata, each density will be represented by at least one data point. The danger in cluster sampling, if clusters have different underlying densities, is that a cluster making an important contribution to the overall density f may not be sampled; this situation may lead to poor estimation of f . Conditions (C3) and (C4) ensure that no one psu within a stratum will dominate the density estimation within that stratum.

Condition (C5) is stronger than the usual condition for pointwise MSE consistency in the iid model; namely, that $n(t)b(t) \rightarrow \infty$. The condition $n(t)b(t) \rightarrow \infty$ is also the one used for the binned estimators in Bellhouse and Stafford (1999). The stronger condition (C5) is needed here because pointwise MSE consistency is required for every possible finite population. Indeed, Theorem 2 proved uniform consistency, but the same conditions are needed for pointwise consistency in the design-based framework. The following counterexample, using a simple random sample, shows that $n(t)b(t) \rightarrow \infty$ is not sufficient for design-based pointwise MSE consistency.

Example 1. In this example, we construct sequences of finite populations $\mathcal{U}(t)$ and samples $\mathcal{S}(t)$ that satisfy (C1)–(C4), and show that condition (C5) is necessary to have $V_D[\hat{f}_{\mathcal{S}(t)}(x; b(t))] \rightarrow 0$ as $t \rightarrow \infty$ for these sequences. Let $g(k) = \max \{j \in \mathbb{Z} : 2j(j-1) < k\}$ and define the sequence y_1, y_2, \dots by $y_k = (-1)^k [2g(k) - 1]^{-1}$ if $k \leq 2g^2(k)$ and $y_k = (-1)^k [2g(k)]^{-1}$ if $k > 2g^2(k)$. Let $N(t) = t^2$, $n(t) = t^2/2$, and $b(t) = t^{-1}$. Let $\mathcal{U}(t)$ be the first $N(t)$ elements of $\{y_i\}_{i \geq 1}$ and let $\mathcal{S}(t)$ be a simple random sample of $n(t)$ elements in $\mathcal{U}(t)$. Then $n(t)b^{2-\varepsilon}(t) \rightarrow \infty$ for any $0 < \varepsilon \leq 1$ and assumptions (C1)–(C4) are met, but $n(t)b^2(t) \rightarrow 1/2$ as $t \rightarrow \infty$. Take the kernel to be uniform, $K(z) = (1/2)I_{[0,1]}(|z|)$. With $x = 0$, $U_{1i} = (t/2)I_{[0,1]}(ty_i)$. By construction, if t is even, then exactly half of the elements of $\mathcal{U}(t)$ are greater than $b(t)$ in absolute value. If t is odd, then exactly $[N(t) - 1]/2$ elements of $\mathcal{U}(t)$ are greater than $b(t)$ in absolute value. Thus, using (6), $V_D[\hat{f}_{\mathcal{S}(t)}(0; b(t))] = \frac{1}{16}t^2/(t^2 - 1) + o(1)$. The asymptotic variance of $\hat{f}_{\mathcal{S}(t)}(0; b(t))$ is $1/16$.

3.1.2. Consistency under model assumptions

In contrast to design-based inference, the condition $n(t)b(t) \rightarrow \infty$ is sufficient for consistency in model-based inference, or inference under the combined distribution induced by the model and design. Often, in inference under the combined distribution, the random variables generating the finite population are assumed to be iid (as in

Isaki and Fuller, 1982, and Bellhouse and Stafford, 1999). This assumption causes difficulty for density estimation, however, because it implies that there are no clustering effects in the superpopulation. It seems reasonable that units in the same cluster would be considered dependent in the superpopulation. We introduce here a model for the densities that supports the concept of cluster sampling.

For a superpopulation model that accords with the motivation for a stratified cluster design, assume that the finite population is generated by random variables $\{Y_{hik}\}$. Assume that any random variables Y_{hik} and Y_{hil} within the same psu are positively correlated, and that the random variables generating psu i in stratum h are independent of those generating any other psu. Specifically, let (Y_{hik}, Y_{hil}) have joint density g_h , with the property that

$$\int g_h(x, u) du = \int g_h(u, x) du = f_h(x) \tag{8}$$

and assume that Y_{hik} and Y_{rpl} are independent if $(hi) \neq (rp)$. Property (8) is satisfied, for example, if $Y_{hik} = \mu_h + \alpha_{hi} + \varepsilon_{hik}$, where $\alpha_{hi} \sim N(0, \sigma_{\alpha h}^2)$, $\varepsilon_{hik} \sim N(0, \sigma_{\varepsilon h}^2)$, and all α_{hi} and ε_{hik} are independent, i.e., the random variables generating the finite population satisfy the conditions for the one-way random effects model with normal errors within each stratum. Then the common joint density for any pair of random variables in the

same psu is bivariate normal with mean $[\mu_h, \mu_h]^T$ and covariance matrix $\begin{bmatrix} \tau_h^2 & \sigma_{\alpha h}^2 \\ \sigma_{\alpha h}^2 & \tau_h^2 \end{bmatrix}$,

where $\tau_h^2 = \sigma_{\alpha h}^2 + \sigma_{\varepsilon h}^2$. The common marginal density in stratum h is $N(\mu_h, \tau_h^2)$. With this random effects model, the random variables within each psu satisfy the conditions of a single-parameter version of the model proposed by Isaki and Fuller (1982).

We employ the following regularity conditions for the superpopulation model.

- (M1) (Y_{hik}, Y_{hil}) have joint density g_h , where g_h satisfies (8). The marginal density of Y_{hik} , f_h , has continuous second derivative f_h'' that is square integrable and monotone in both $(-\infty, -M)$ and (M, ∞) for some M . The variables Y_{hik} and Y_{rpl} are independent if $(hi) \neq (rp)$.
- (M2) $f(x) = \sum_{h=1}^L W_h f_h(x)$.
- (M3) $\sup_x \max_{1 \leq h \leq L(t)} f_h(x) = G(t) = O(1)$.
- (M4) $\sup_x \max_{1 \leq h \leq L(t)} |f_h''(x)| = D(t) = O(b^{-1}(t))$.
- (M5) $N(t)b(t) \rightarrow \infty$ and $b(t) \rightarrow 0$ as $t \rightarrow \infty$.

Assumption (M2) states that the overall density is a mixture of the stratum densities, weighted by $W_h = Q_h/Q$. Stratification with random sampling accords with a special case of this model, in which Y_{hi} has density f_h and all random variables are independent.

Note that for model-based approach to be useful for density estimation, we must assume that the sampled units follow the model in (M1). If the sample selection probabilities depend on the values of the response after conditioning on explanatory variables in the model, there may be sample selection bias due to the informative design. In this case a more complicated model would need to be adopted in order to incorporate the information in the sampling design. See Pfeffermann et al. (1998) for more discussion of informative designs in model-based inference. We assume throughout the

remainder of the paper that the sample design is noninformative: the sample inclusion probabilities within each stratum are independent of the values of the response.

Under the superpopulation model in (M1)–(M4), $\hat{f}_{\mathcal{M}}(x; b)$ is a random quantity. Let $*$ denote the convolution operator. Define $R(\phi) = \int \phi^2(u) du$ for any function ϕ and let $\mu_2(K) = \int u^2 K(u) du$. The following results are given in Wand and Jones (1995, Chapter 2).

Lemma 1. Assume conditions (K1)–(K4) and (M1). Then for any $x \in \mathfrak{X}$,

$$E_M[K_b(x - Y_{hik})] = (K_b * f_h)(x), \tag{9}$$

$$(K_b * f_h)(x) = f_h(x) + \frac{1}{2} b^2 f_h''(x) \mu_2(K) + o(b^2), \tag{10}$$

$$(K_b^2 * f_h)(x) = b^{-1} f_h(x) R(K) + o(b^{-1}). \tag{11}$$

Under the model, the population quantity $\hat{f}_{\mathcal{M}}(x; b)$ is a biased estimator of the underlying function $f(x)$; from Lemma 1, the model-based bias under conditions (M1) and (M2) is

$$\begin{aligned} E_M[\hat{f}_{\mathcal{M}}(x; b) - f(x)] &= \sum_{h=1}^L W_h (K_b * f_h)(x) - f(x) \\ &= \frac{1}{2} b^2 f''(x) \mu_2(K) + o(b^2). \end{aligned} \tag{12}$$

This coincides with the bias for the estimator (1).

Standard results from density estimation do not apply for calculating the model-based variance of $\hat{f}_{\mathcal{M}}(x; b)$ and $\hat{f}_{\mathcal{G}}(x; b)$ because of the assumed dependence within clusters. Instead, we use a bivariate convolution, defined as

$$\Psi_h(x; b) = \iint K_b(x - u) K_b(x - v) g_h(u, v) du dv. \tag{13}$$

Note that $\Psi_h(x; b) = E_M[K_b(x - Y_{hik}) K_b(x - Y_{hil})]$ for any $i \leq N_h$ and any $k \neq l$. Under condition (M1), then,

$$\begin{aligned} V_M[\hat{f}_{\mathcal{M}}(x; b)] &= \frac{1}{Q^2} \sum_{h=1}^L \sum_{i=1}^{N_h} V_M \left[\sum_{j=1}^{Q_{hi}} K_b(x - Y_{hij}) \right] \\ &= \frac{1}{Q^2} \sum_{h=1}^L \sum_{i=1}^{N_h} \{ Q_{hi} [(K_b^2 * f_h)(x) - (K_b * f_h)^2(x)] \\ &\quad + Q_{hi}(Q_{hi} - 1) [\Psi_h(x; b) - (K_b * f_h)^2(x)] \}. \end{aligned} \tag{14}$$

If the stratum densities and their absolute second derivatives are uniformly bounded, in the sense of conditions (M3) and (M4), then $\hat{f}_{\mathcal{G}}(x; b(t))$ is MSE-consistent under the model and under the combined distribution when $nb \rightarrow \infty$, provided that certain conditions on the designs are met. The conditions needed for consistency are given in Theorems 3 and 4.

Theorem 3. Assume conditions (C1), (C2), and (M1)–(M5) hold, and that the kernel function satisfies (K1)–(K4). Then $MSE_M[\hat{f}_h(x; b(t))] \rightarrow 0$ as $t \rightarrow \infty$. If, in addition, (C3) holds, and

$$\sum_{i \in \mathcal{S}_h} \sum_{k \in \mathcal{S}_{hi}} w_{hik} = Q_h + o(Q_h) \tag{15}$$

for each h , then $MSE_M[\hat{f}_{\mathcal{S}}(x; b(t)) | \mathcal{S}] \rightarrow 0$ as $t \rightarrow \infty$.

Theorem 4. Assume conditions (C1)–(C4) and (M1)–(M5) hold and that the kernel function satisfies (K1)–(K4). Then $E_M\{E_D[\hat{f}_{\mathcal{S}}(x; b) - f(x)]^2 | \mathbf{Y}_t\} \rightarrow 0$ as $t \rightarrow \infty$.

Note that in Theorem 3, the condition in (15) is necessary since the expectation operator averages over populations generated by the model, not over possible samples. Without (15), the estimator $\hat{f}_{\mathcal{S}}(x; b(t))$ could be badly biased under the model because the sum of the weights in the particular sample \mathcal{S} could be far from Q_h . An alternate estimator, using actual values of Q_h instead of the sampling weights, would be better in the model-based setting; such an estimator would avoid restrictions on the sampling design.

Why are different rates needed for consistency under the design-based and the model-based or combined inference? The model assumptions in the latter two modes of inference allow a slower increase in the sample size. In design-based inference, one must have consistency under every possible finite population and sample, no matter how implausible. In model-based inference, the boundedness requirement for the individual stratum densities and their absolute second derivatives ensures that on average, misbehaving populations will not occur.

Although strict conditions on g_h are not needed for consistency, the expressions for the variances of $\hat{f}_h(x; b(t))$ and $\hat{f}_{\mathcal{S}}(x; b(t))$ simplify if we require the joint density to be smooth and bounded. The conditions in the following theorem, which is proven in the appendix, guarantee that the within-psu correlations do not approach 1 as $t \rightarrow \infty$.

Theorem 5. For all h , suppose g_h is absolutely continuous with respect to Lebesgue measure and that g_h and its absolute second partial derivatives are uniformly bounded and continuous in a neighborhood of (x, x) . Then, under conditions (M1)–(M5) and (K1)–(K4),

$$V_M[\hat{f}_h(x; b(t))] = \frac{1}{Q_h b} f(x)R(K) + o[(Q_h b)^{-1}]. \tag{16}$$

If, in addition, (C1)–(C4) hold, then

$$V_M[\hat{f}_{\mathcal{S}}(x; b(t))] = \frac{1}{Q^2 b} \sum_{(hik) \in \mathcal{S}} w_{hik}^2 [f_h(x)R(K) + o(1)] \tag{17}$$

and

$$V_C[\hat{U}(b)] = \sum_{h=1}^L \sum_{i=1}^{N_h} \frac{Q_{hi}}{b\pi_i^{(h)}} [f_h(x)R(K) + o(1)]. \tag{18}$$

The asymptotic variance of $\hat{f}_{\mathcal{M}}(x; b(t))$ under the model is the same as if all observations were independent. The asymptotics are driven by the increase in the number of psu's as $t \rightarrow \infty$. Because the psu sizes in the population are bounded, as $t \rightarrow \infty$ it becomes less and less likely that two observations from the same psu would appear in the bandwidth window, so the dependence has negligible effect asymptotically.

3.2. Asymptotic normality in stratified multistage sampling

3.2.1. Design-based inference

It is a common practice to sample the psu's without replacement with inclusion probabilities proportional to size. The dependence induced by sampling without replacement, however, greatly complicates central limit results, as documented by Sen (1988). To avoid the dependence induced by sampling without replacement, following Krewski and Rao (1981) we show design-based asymptotic normality for an estimator where the psu's are sampled with replacement.

We use the following setup for sampling with replacement. As before, there are $L(t)$ strata. From stratum h , $n_h(t)$ psu's are sampled with replacement; on draw j (for $j = 1, \dots, n_h(t)$), psu (hi) is sampled with probability p_{hi} , where $\sum_{i=1}^{N_h(t)} p_{hi} = 1$. Define the random variable Z_{hjk} to be 1 if psu k is selected on draw j and 0 otherwise. Since sampling is done with replacement, Z_{hjk} and $Z_{h'j'l}$ are independent when $(hj) \neq (h'j')$. Let

$$X_{hj}(b) = \sum_{k=1}^{N_h} Z_{hjk} \frac{\hat{U}_{hk}(b)}{n_h p_{hk}}. \tag{19}$$

The X_{hj} 's are independent with $E_D[X_{hj}(b)] = U_h(b)/n_h$. Define $X(b, t) = \sum_{h=1}^{L(t)} \sum_{j=1}^{n_h(t)} X_{hj}(b)$. Let the sample estimator of the density be

$$\hat{f}_R(x; b) = \hat{Q}^{-1} X(b, t) \tag{20}$$

for \hat{Q} a design-consistent estimator of Q ; then $\hat{f}_R(x; b)$ is approximately design-unbiased for $\hat{f}_{\mathcal{M}}(x; b)$ since $E_D[Q^{-1} X(b, t)] = \hat{f}_{\mathcal{M}}(x; b)$.

Also define $\sigma^2(t) = V_D[X(b, t)]$ and, when all $n_h \geq 2$, define

$$\hat{\sigma}^2(t) = \sum_{h=1}^L \sum_{j=1}^{n_h} \frac{n_h}{n_h - 1} \left(X_{hj}(b) - \sum_{l=1}^{n_h} X_{hl}(b)/n_h \right)^2. \tag{21}$$

By (2.10) of Krewski and Rao (1981), $E_D[\hat{\sigma}^2(t)] = \sigma^2(t)$. The following limiting distribution results, proven in the appendix, then hold.

Theorem 6. Assume conditions (K1)–(K4), (C1), (C2), and (C5) hold. Suppose there exists a constant $\delta > 0$ such that $\delta < N_h(t)p_{hi}$ and $\delta < \pi_k^{(h)}|_i$ for all h, i , and k . Further suppose that $\max_h N_h(t)/n_h(t)$ is bounded and that $\lim_{t \rightarrow \infty} b(t)\sigma(t) = \infty$. Then, conditionally on \mathbf{Y} ,

$$[X(b, t) - Q\hat{f}_{\mathcal{M}}(x; b)]/\sigma(t) \rightarrow_{\mathcal{D}} N(0, 1)$$

as $t \rightarrow \infty$. Furthermore, if $n_h \geq 2$ for $h = 1, \dots, L(t)$, then

$$\frac{b(t)}{n(t)} [\hat{\sigma}^2(t) - \sigma^2(t)] | \mathbf{Y} \rightarrow_p 0 \tag{22}$$

as $t \rightarrow \infty$.

Asymptotic normality can be demonstrated for without replacement sampling for the special situation where $n_h(t) \rightarrow \infty$ and $N_h(t) - n_h(t) \rightarrow \infty$ for every stratum h . This is done by applying Lemma 2.1 of Hájek (1960), which shows that with-replacement results apply to the without-replacement situation.

Using Theorem 6, an approximate 95% confidence interval for $\hat{f}_u(x; b)$ is given by $\hat{f}_R(x; b) \pm 1.96\hat{\sigma}(t)/\hat{Q}$. If sampling were done without replacement, substituting $\hat{f}_u(x; b)$ for $\hat{f}_R(x; b)$ in the confidence interval while retaining the with-replacement variance estimate $\hat{\sigma}^2(t)$ would generally result in a slightly conservative design-based confidence interval.

The variance $\sigma^2(t)$ can also be estimated using the jackknife method. In this case, since $\hat{U}(b)$ is an estimated population total, the jackknife estimate of $\sigma^2(t)$ reduces to $\hat{\sigma}^2(t)$.

3.2.2. Asymptotic normality under model assumptions

Under the model assumptions in (M1)–(M5), $\hat{f}_u(x; b)$ is asymptotically normal; in addition, $\hat{f}_R(x; b)$ is asymptotically normal under the combined distribution. The following theorem is proven in the appendix.

Theorem 7. Assume conditions (K1)–(K4), (M1)–(M5), (C1) and (C2) hold. Also assume that g_h and its second partial derivatives are continuous and uniformly bounded in a neighborhood of (x, x) . Then, as $t \rightarrow \infty$,

$$\frac{\hat{f}_u(x; b) - E_M[\hat{f}_u(x; b)]}{\sqrt{V_M[\hat{f}_u(x; b)]}} \rightarrow_D N(0, 1). \tag{23}$$

If, in addition, the conditions of Theorem 6 hold, and if $n(t)/Q(t) \rightarrow \lambda$ and $n(t)b(t)\sigma^2(t)/Q^2(t) \rightarrow_{a.s.} \gamma$ as $t \rightarrow \infty$, where $0 \leq \lambda < 1$ and $0 < \gamma < \infty$, then

$$\frac{\hat{f}_R(x; b) - E_M[\hat{f}_u(x; b)]}{\sqrt{V_C[\hat{f}_R(x; b)]}} \rightarrow_D N(0, 1). \tag{24}$$

Theorem 7 can be used to construct confidence intervals for $E_M[\hat{f}_u(x; b)]$ under the combined mode of inference. Since $V_C[X(b, t)] = E_M[\sigma^2(t)] + V_M E_D[X(b, t)]$, an approximately unbiased estimator of $V_C[\hat{f}_R(x; b)]$ is given by

$$\hat{V}_C[\hat{f}_R(x; b)] = \hat{\sigma}^2/\hat{Q}^2 + R(K)\hat{f}_R(x; b)/(\hat{Q}b), \tag{25}$$

where the second term estimates $V_M[\hat{f}_u(x; b)]$ given in (16). Then an approximate 95% confidence interval for $E_M[\hat{f}_u(x; b)]$ can be calculated as $\hat{f}_R(x; b) \pm 1.96\sqrt{\hat{V}_C[\hat{f}_R(x; b)]}$. If sampling fractions n_h/N_h are small, the second term in (25) will be small relative

to the first term, so that $\hat{V}_C[\hat{f}_R(x; b)] \approx \hat{V}_D[\hat{f}_R(x; b)]$. As with design-based confidence intervals, $\hat{f}_R(x; b)$ can be replaced by the without-replacement estimate $\hat{f}_{\mathcal{G}}(x; b)$ to give a slightly conservative confidence interval under the combined distribution.

The confidence intervals in this section are for $E_M[\hat{f}_{\mathcal{U}}(x; b)]$. Following standard practice, they can be adapted to form confidence intervals for $f(x)$ either by under-smoothing so that the asymptotic bias is negligible, or by shifting the interval using an estimate of the bias.

4. Bandwidths

One important aspect of density estimation is choice of the bandwidth b . When b is too large, the density is oversmoothed and the model-based bias of $\hat{f}_{\mathcal{G}}(x; b)$ is large; when b is too small, the bias is small but the variance of $\hat{f}_{\mathcal{G}}(x; b)$ is large. In order for $\hat{f}_{\mathcal{G}}(x; b)$ to be a density, the same bandwidth should be used for each value of x ; however, local kernel density estimators are sometimes used in which the optimal bandwidth can vary with x . In this section we first explore bandwidth issues from a design-based perspective. We then derive two sets of optimal bandwidth sequences to be used with $\hat{f}_{\mathcal{G}}(x; b)$ and $\hat{f}_{\mathcal{U}}(x; b)$ under the model-based and combined approaches to inference.

4.1. Bandwidths under design-based inference

In Sections 2 and 3 we used the same bandwidth sequence $b(t)$ in $\hat{f}_{\mathcal{G}}(x; b)$, $\hat{f}_R(x; b)$, and $\hat{f}_{\mathcal{U}}(x; b)$. Using the same bandwidth ensures that in the design-based setting, $\hat{f}_{\mathcal{G}}(x; b)$ and $\hat{f}_R(x; b)$ are both consistent and approximately unbiased for $\hat{f}_{\mathcal{U}}(x; b)$. However, it raises the question of whether $\hat{f}_{\mathcal{U}}(x; b)$ is the appropriate quantity to be estimated: typically, the population size is much larger than the sample size, so it might be more appropriate to estimate $\hat{f}_{\mathcal{U}}(x; b_{\mathcal{U}})$ instead, where $b_{\mathcal{U}}$ is smaller than the bandwidth b used in $\hat{f}_{\mathcal{G}}(x; b)$ or $\hat{f}_R(x; b)$.

If one bandwidth sequence $\{b_{\mathcal{U}}\}$ is used for the finite population quantity, and a different bandwidth sequence $\{b_{\mathcal{G}}\}$ is used for the estimator from the sample, then it is possible that $\hat{f}_{\mathcal{G}}(x; b_{\mathcal{G}})$ may not be consistent for $\hat{f}_{\mathcal{U}}(x; b_{\mathcal{U}})$. The following example illustrates that the pointwise design-based bias may be infinite when the bandwidths for sample and population differ.

Example 2. Define the sequence y_1, y_2, \dots by $y_k = k^{-1}(-1)^k$ and let $K(z) = (1/2)I_{[0,1]}(|z|)$. Let $N(t) = t^2$, $n(t) = t^2/2$, $b_{\mathcal{U}}(t) = t^{-1}$ and $b_{\mathcal{G}}(t) = 2t^{-1}$. Let $\mathcal{U}(t)$ be the first $N(t)$ elements of $\{y_i\}_{i \geq 1}$ and let $\mathcal{S}(t)$ be a simple random sample of $n(t)$ elements in $\mathcal{U}(t)$. Under this construction there are precisely $(t - 1)$ elements of $\mathcal{U}(t)$ outside of $[-b_{\mathcal{U}}, b_{\mathcal{U}}]$ and $[(t + t \bmod 2)/2 - 1]$ elements outside of $[-b_{\mathcal{G}}, b_{\mathcal{G}}]$. Then

$$E_D[\hat{f}_{\mathcal{G}}(0; b_{\mathcal{G}})] - \hat{f}_{\mathcal{U}}(0; b_{\mathcal{U}}) = \frac{1}{2t} \sum_{i=1}^{t^2} K(ty_i/2) - \frac{1}{t} \sum_{i=1}^{t^2} K(ty_i) = -\frac{t}{2} + O(1).$$

The design-based bias at 0 thus tends to $-\infty$ as $t \rightarrow \infty$.

In the design-based setting, then, it is best to use the same bandwidth $b = b_{\mathcal{U}} = b_{\mathcal{G}}$ for both $\hat{f}_{\mathcal{G}}$ and $\hat{f}_{\mathcal{U}}$. This does not tell what the common bandwidth should be, however; we need to adopt a model-based or combined inference framework to select a bandwidth.

4.2. Bandwidths under the combined distribution

In the combined approach to inference, an underlying theoretical density function f , rather than $\hat{f}_{\mathcal{U}}(x; b)$, is the quantity of interest. In this approach, then, it is acceptable to consider the same bandwidth for $\hat{f}_{\mathcal{U}}$ and $\hat{f}_{\mathcal{G}}$ since $\hat{f}_{\mathcal{U}}$ is used only in proving properties of $\hat{f}_{\mathcal{G}}$ and is not considered to be of interest in itself. The model-based bias of $\hat{f}_{\mathcal{G}}$ is used in determining the mean squared error. Problems such as that in Example 2 will not occur because of the averaging over possible finite populations under the model.

Throughout this section, we assume that the conditions in Theorem 5 hold. Using the combined distribution for the sample quantity and Eq. (18), the asymptotic mean squared error (AMSE) of $\hat{f}_{\mathcal{G}}(x; b)$ is

$$\text{AMSE}_C[\hat{f}_{\mathcal{G}}(x; b)] = \left[\frac{1}{2} b^2 f''(x) \mu_2(K) \right]^2 + \sum_{h=1}^L \sum_{i=1}^{N_h} \frac{Q_{hi}}{Q^2 b \pi_i^{(h)}} f_h(x) R(K),$$

where μ_2 and $R(K)$ were defined preceding Lemma 1. Thus the locally optimal bandwidth sequence for $\hat{f}_{\mathcal{G}}(x; b)$ is given by

$$b^{\text{opt}}(\mathcal{G}, x) = \left(\frac{R(K) \sum_{h=1}^L \sum_{i=1}^{N_h} Q_{hi} f_h(x) / \pi_i^{(h)}}{Q^2 [f''(x)]^2 \mu_2^2(K)} \right)^{1/5}. \tag{26}$$

By imposing integrability assumptions on the second-order derivatives of the underlying densities f_h , we can use the asymptotic mean integrated squared error (AMISE) to derive globally optimal bandwidth sequences. Here, $\text{AMISE}(\hat{f}) = \int \text{AMSE}[\hat{f}(x; b)] dx$. Then

$$\text{AMISE}_C[\hat{f}_{\mathcal{G}}(\cdot; b)] = \frac{1}{4} b^4 \mu_2^2(K) R(f'') + \sum_{h=1}^L \sum_{i=1}^{N_h} \frac{Q_{hi}}{Q^2 b \pi_i^{(h)}} R(K)$$

and the globally optimal bandwidth for $\hat{f}_{\mathcal{G}}$ is

$$b^{\text{opt}}(\mathcal{G}, \cdot) = \left(\sum_{h=1}^L \sum_{i=1}^{N_h} \frac{Q_{hi}}{Q \pi_i^{(h)}} \right)^{1/5} \left(\frac{R(K)}{QR(f'') \mu_2^2(K)} \right)^{1/5}. \tag{27}$$

As would be expected, the locally optimal bandwidth is more sensitive to the pointwise behavior of both f and its second derivative than is the global bandwidth. Note that if the entire finite population is included in the sample, i.e., $\hat{f}_{\mathcal{G}} = \hat{f}_{\mathcal{U}}$, then the first factor in (27) is one, so that the optimal global bandwidth for $\hat{f}_{\mathcal{U}}$ is the same as in the iid setting.

The optimal bandwidths discussed in this section are theoretical quantities that depend on the unknown underlying density function and its second derivative. In practice, a plug-in or cross-validation method (see Wand and Jones, 1995) can be used to select a bandwidth for the iid case and then the bandwidth can be adjusted for $\hat{f}_{\mathcal{G}}$ as in (27).

Note that the expression in (27) simplifies under certain sampling designs. If a probability proportional to size design is used with $\pi_i^{(h)} = n_h Q_{hi} / Q_h$ and if $n_h / N_h = n / N$ for all h , then $b^{\text{opt}}(\mathcal{S}, \cdot) = \{[NR(K)]/[nQR(f'')\mu_2^2(K)]\}^{1/5}$. Consequently, under stratified random sampling with proportional allocation as well as certain other designs, the optimal global bandwidth is the same as in the iid case.

In practice, the quantities Q_{hi} may be unknown by a secondary data analyst. We believe a reasonable approach to bandwidth selection in many large-scale surveys (which often are designed so that the sampling weights are approximately equal) is to start with a bandwidth selected as though the data were iid with sample size q ; the initial bandwidth may need to be expanded if measurements from the same psu are highly positively correlated because in small samples the variance term under the combined distribution may be larger than the asymptotic quantity given in Theorem 5.

If the assumptions in Theorem 5 are met and the optimal asymptotic bandwidth is used, the AMISE under the combined distribution has order $O(q^{-4/5})$. Since a sample size of 100 is often considered to be adequate in the iid setting, we suggest that an overall sample size of 100–200 will generally give a reasonably accurate estimate of the density with data from a complex survey.

5. Density estimation near boundaries

For the asymptotic results in Section 3 and the bandwidth discussions in Section 4 we have assumed that no boundary effects arise. Measurements in surveys, however, often take only positive values. In this section we consider density estimation near a boundary. Without loss of generality suppose that f has support $[0, \infty)$. Also suppose that the kernel K has support $[-1, 1]$. In this setting the design-based results are still valid. In Theorem 2, for example, $\hat{f}_{\mathcal{S}}(x, b)$ is design-based consistent for $\hat{f}_{\mathcal{M}}(x, b)$ for all x . However, $\hat{f}_{\mathcal{M}}(x, b)$ may be positive for $x < 0$ and may be model-biased for estimating $f(x)$ for x near 0.

One method for reducing the model-based bias for x near 0 is to use the boundary kernel introduced in Gasser and Müller (1979): Let $p = x/b$ and define

$$K^*(u, p) = \frac{a_2(p) - a_1(p)u}{a_0(p)a_2(p) - a_1^2(p)} K(u), \tag{28}$$

where $a_\ell(p) = \int_{-1}^p z^\ell K(z) dz$. Note that for $p \geq 1$, $K^*(u, p) = K(u)$. The density estimators $\hat{f}_{\mathcal{M}}^*(x, b)$ and $\hat{f}_{\mathcal{S}}^*(x, b)$ are defined by (2) and (3), respectively, with $K[(x - y_{hik})/b]$ replaced by $K^*[(x - y_{hik})/b, x/b]$ for $x \leq b$.

We study asymptotic properties for a sequence of points $x = x(t) = pb(t)$ for fixed $p \in (0, 1)$ as $t \rightarrow \infty$. First note that under the conditions of Theorem 2,

$$E_D[\{\hat{f}_{\mathcal{S}}^*(x(t); b(t)) - \hat{f}_{\mathcal{M}}^*(x(t); b(t))\}^2] \rightarrow 0 \tag{29}$$

as $t \rightarrow \infty$, so the estimator is design-consistent near the boundary. The proof of (29) follows that of Theorem 2, noting that for all $p \in (0, 1)$ $a_0(p)a_2(p) - a_1^2(p) \geq a_0(0)a_2(0) - a_1^2(0) > 0$, so $K^*(u, p) \leq cK(u)$ for a constant c .

For properties under the model-based and combined approaches, replace condition (M1) by (M1*):

(M1*) (Y_{hik}, Y_{hil}) have joint density g_h , where g_h satisfies (8). The marginal density of Y_{hik} , f_h , has second derivative f''_h that is square integrable and continuous on $(0, \infty)$ and monotone in (M, ∞) for some M . The variables Y_{hik} and Y_{rpl} are independent if $(hi) \neq (rp)$.

Then under conditions (M1*), (M2), and (K1)–(K4), the model-based bias near the boundary has order $O(b^2)$: for $x = pb$ and $p < 1$,

$$E_M[\hat{f}_{\mathcal{Q}}^*(x(t); b)] = f(x) + \frac{b^2}{2} f''(0+) \int_{-1}^p u^2 K^*(u, p) du + o[b^2], \tag{30}$$

where $f''(0+) = \lim_{x \downarrow 0} f''(x)$. Also, under the conditions in Theorem 5 with (M1) replaced by (M1*),

$$V_M[\hat{f}_{\mathcal{Q}}^*(x(t); b(t))] = \frac{1}{Qb} f(0+) \int_{-1}^p [K^*(u, p)]^2 du + o[(Qb)^{-1}] \tag{31}$$

and

$$V_C[\hat{Q}\hat{f}_{\mathcal{G}}^*(x(t); b(t))] = \sum_{h=1}^L \sum_{i=1}^{N_h} \frac{Q_{hi}}{b\pi_i^{(h)}} \left[f_h(0+) \int_{-1}^p [K^*(u, p)]^2 du + o(1) \right]. \tag{32}$$

Eqs. (31) and (32) are proven similarly to the corresponding results in Theorem 5, using results in Jones (1993). Neither $\hat{f}_{\mathcal{Q}}^*(x; b)$ nor $\hat{f}_{\mathcal{G}}^*(x; b)$ is guaranteed to be a density, but the estimated functions can be normalized so that they integrate to 1.

6. Applications

We present two examples of smoothed density estimates from complex surveys, illustrating the confidence bands calculated using the results of this paper. Since the sampling fractions are small in both surveys, we estimate the variance under the combined distribution by $\hat{\sigma}^2/\hat{Q}$, where $\hat{\sigma}^2$ is defined in (21). The jackknife is used to calculate $\hat{\sigma}^2$.

The US National Crime Victimization Survey (NCVS) is an ongoing stratified multistage sample with a rotating panel design; although it is designed to be approximately self-weighting, nonresponse and ratio adjustments vary the final weights. Fig. 1 shows the estimated density function for ages of female victims of sexual assault, using the 1994 NCVS (US Department of Justice, 1998). The bandwidth value $b=5$ was chosen subjectively. Since the NCVS interviews only persons aged 12 and over, the boundary kernel in (28) was used for women under age 17; for the boundary region, we used the bandwidth value $[2b - (x - 12)]$ discussed in Gasser and Müller (1979) and Jones (1993). In the 1994 NCVS, there were 237 female victims of sexual assault, so the design-based confidence bands calculated using the jackknife are relatively wide. The density estimate shows a feature of the data that is not readily discernible from summaries using means and quantiles. The estimated mode for age of sexual assault

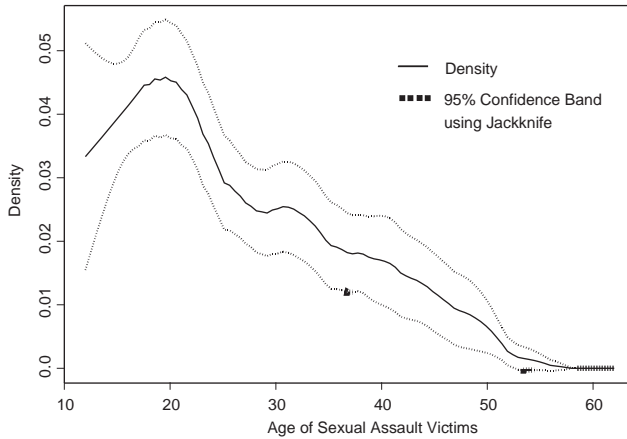


Fig. 1. Density estimate for the ages of female victims of sexual assault, using Epanechnikov kernel function and a 5-year bandwidth.

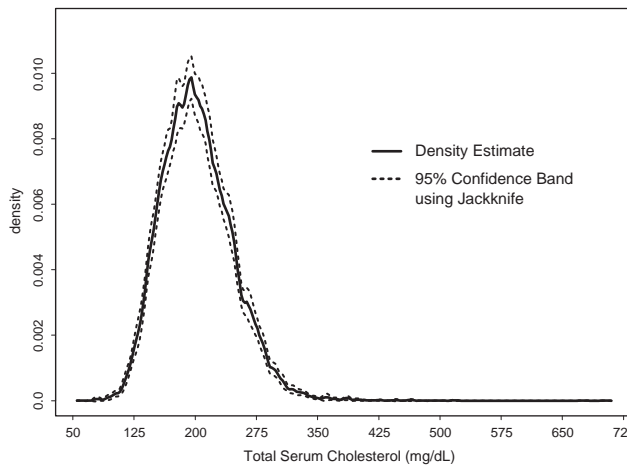


Fig. 2. Density estimate for the total serum cholesterol levels of US adults, aged 17 or older, using Epanechnikov kernel function and a 6.5-unit bandwidth.

victims is close to age 20; however, another minor peak appears to occur around the age of 32. Although the confidence bands are wide around the secondary peak, the same phenomenon occurs in NCVS data sets from other years.

Fig. 2 estimates the density of total serum cholesterol using data from the US National Health and Nutrition Examination Survey III ($q = 16764$), with data provided by the US Department of Health and Human Services (1996). We used a Gaussian approximation to the density to estimate $R(f'')$ by $(3/8)\pi^{-1/2}s^{-5} = 1.3 \times 10^{-9}$, where $s=44$ estimates the standard deviation of the population. For the Epanechnikov kernel used, $R(K) = 3/5$ and $\mu_2(K) = 1/5$, so a preliminary bandwidth,

using results in Section 4 and acting as though the sample were iid, was estimated as $((3/5)/(16764R(f'')/25))^{1/5} = 14.4$. We reduced this to 6.5 for the figure so that we would not miss features of the density by possible oversmoothing. It is readily seen that the distribution is slightly skewed to the right. Because of the large sample size, the jackknife confidence bands are tight around the density estimate. The jackknife, however, gives confidence intervals around $\hat{f}_{\mathcal{S}}(x; b)$ as an estimator of $\hat{f}_{\mathcal{M}}(x; b)$. The confidence intervals thus include only the variance and do not incorporate the model-based bias in $\hat{f}_{\mathcal{M}}(x; b)$.

7. Discussion

In this paper, we have presented sufficient conditions needed for consistency and asymptotic normality under three frameworks for inference in survey sampling: design-based, model-based, and a combined approach. These theoretical results justify inference about density functions in the survey sampling context. We illustrated the theoretical results by estimating the density and constructing confidence bands for data from two large US surveys. We also introduced a new superpopulation model that is more appropriate for stratified multistage samples than the one usually used in which random variables generating the finite population are assumed to be iid. The superpopulation model developed here may also be used for model-based or combined inference in settings other than density estimation.

We believe that each of the three settings for inference may be appropriate in certain contexts. Design-based inference is appropriate if the population quantity of interest is in fact $f_{\mathcal{M}}$. Model-based inference is appropriate if the entire finite population is measured, or if one has reason to believe that the model adopted explains all the salient features of the data.

In most survey sampling settings, however, we believe that the combined distribution will be most appropriate for inference about densities. For the NCVS example, the combined approach assumes an underlying density for the ages of female victims of sexual assault in 1994; this would likely be the quantity of interest for social scientists. The combined distribution also incorporates features of the design, however, so that the design-based variance is employed and the resulting estimator is more robust to model misspecification.

Acknowledgements

The authors are grateful to J.N.K. Rao and the referees for their many helpful suggestions. They also thank David Bellhouse for making his papers available to them prior to publication.

Appendix. Proofs

Proof of Theorem 2. We suppress the index t for notational ease. Condition (K4) implies that $u_{hik}(b) \leq b^{-1}m$ for all h, i, k and consequently, by assumption (C2), that

$$U_{hi}(b) \leq b^{-1}Q_{hi}m \leq b^{-1}Bm.$$

Thus for each h , (C3) and (C4) imply that

$$\begin{aligned} & \sum_{i=1}^{N_h} \sum_{j>i}^{N_h} [\pi_i^{(h)} \pi_j^{(h)} - \pi_{ij}^{(h)}] \left[\frac{U_{hi}(b)}{\pi_i^{(h)}} - \frac{U_{hj}(b)}{\pi_j^{(h)}} \right]^2 \\ & \leq \frac{\alpha_h}{2} \sum_{i=1}^{N_h} \sum_{j=1}^{N_h} \pi_i^{(h)} \pi_j^{(h)} \left[\left(\frac{U_{hi}(b)}{\pi_i^{(h)}} \right)^2 + \left(\frac{U_{hj}(b)}{\pi_j^{(h)}} \right)^2 \right] \\ & \leq \frac{\alpha_h B m^2}{2b^2} \sum_{i=1}^{N_h} \sum_{j=1}^{N_h} \left[\frac{\pi_j^{(h)} Q_{hi}}{\pi_i^{(h)}} + \frac{\pi_i^{(h)} Q_{hj}}{\pi_j^{(h)}} \right] \\ & = \frac{\alpha_h B m^2}{2b^2} \sum_{i=1}^{N_h} \frac{2n_h Q_{hi}}{\pi_i^{(h)}} \leq \frac{\alpha_h B m^2 Q_h n_h}{b^2 \delta}. \end{aligned}$$

Similarly,

$$\begin{aligned} V_D[\hat{U}_{hi}] &= \frac{1}{2} \sum_{k=1}^{Q_{hi}} \sum_{l \neq k}^{Q_{hi}} (\pi_{k|i}^{(h)} \pi_{l|i}^{(h)} - \pi_{kl|i}^{(h)}) \left[\frac{u_{hik}}{\pi_{k|i}^{(h)}} - \frac{u_{hil}}{\pi_{l|i}^{(h)}} \right]^2 \\ & \leq \frac{m^2}{2b^2} \sum_{k=1}^{Q_{hi}} \sum_{l=1}^{Q_{hi}} \left[\frac{\pi_{k|i}^{(h)}}{\pi_{l|i}^{(h)}} + \frac{\pi_{l|i}^{(h)}}{\pi_{k|i}^{(h)}} \right] \leq \frac{B Q_{hi} m^2}{b^2 \delta}. \end{aligned}$$

Thus, using the expression in (6),

$$\begin{aligned} Q^{-2} V_D[\hat{U}(b)] & \leq Q^{-2} \sum_{h=1}^{L(t)} \left\{ \frac{\alpha_h B m^2 Q_h n_h}{b^2 \delta} + \frac{Q_h B m^2}{b^2 \delta^2} \right\} \\ & \leq \frac{B m^2}{Q b^2 \delta^2} \sum_{h=1}^{L(t)} W_h [\alpha_h n_h + 1]. \end{aligned}$$

This expression converges to 0 as $Nb^2 \rightarrow \infty$. The variance $V_D[\hat{f}_{\mathcal{G}}(x; b)]$ is shown to converge to 0 uniformly in x by using standard linearization arguments for the variance of a ratio, since $\hat{f}_{\mathcal{U}}(x; b) \leq b^{-1} m$, $Q^{-2} \hat{f}_{\mathcal{U}}^2(x; b) V_D[\hat{Q}]$ and $Q^{-2} \hat{f}_{\mathcal{U}}(x; b) \text{Cov}_D[\hat{U}(b), \hat{Q}]$ both converge to 0 as $Nb^2 \rightarrow \infty$. \square

Proof of Theorem 3. Note that by the Cauchy–Schwarz inequality, $\Psi_h(x) \leq (K_b^2 * f_h)(x)$. Thus, using (14),

$$V_M[\hat{f}_U(x; b)] \leq \frac{1}{Q^2} \sum_{h=1}^L \sum_{i=1}^{N_h} Q_{hi}^2 (K_b^2 * f_h)(x).$$

Using (11), (M3), and (M4), we have for any $x \in \mathfrak{R}$ that

$$V_M[\hat{f}_U(x; b)] \leq \frac{1}{Q^2} \sum_{h=1}^L \sum_{i=1}^{N_h} Q_{hi}^2 \frac{1}{b(t)} [f_h(x) R(K) + o(1)]$$

$$\begin{aligned} &\leq \frac{1}{Q^2} \sum_{h=1}^L \sum_{i=1}^{N_h} Q_{hi}^2 \frac{G(t)R(K)}{b(t)} [1 + o(1)] \\ &\leq \frac{B}{Q} \frac{G(t)R(K)}{b(t)} [1 + o(1)]. \end{aligned}$$

This quantity converges to 0 by (M5). Likewise, the bias of $\hat{f}_U(x; b)$ tends to 0 by (M4).

Similarly,

$$V_M[\hat{f}_{\mathcal{S}}(x; b)] \leq \frac{1}{\hat{Q}^2} \sum_{h=1}^L \sum_{i \in \mathcal{S}_h} \left(\sum_{k \in \mathcal{S}_{hi}} w_{hik} \right)^2 (K_b^2 * f_h)(x).$$

Because $\sum_{i \in \mathcal{S}_h} \sum_{k \in \mathcal{S}_{hi}} w_{hik} = Q_h + o(Q_h)$ for each h , $\hat{Q} = \sum_{h=1}^L [Q_h + o(Q_h)] = Q + o(Q)$ and

$$V_M[\hat{f}_{\mathcal{S}}(x; b)] \leq \frac{B}{Q^2 + o(Q^2)} \frac{1}{\delta^2} \sum_{h=1}^L [Q_h + o(Q_h)] \frac{G(t)R(K)}{b(t)} [1 + o(1)],$$

which tends to 0 as $Qb \rightarrow \infty$. \square

Proof of Theorem 4. Since $\hat{f}_{\mathcal{S}}(x; b)$ is approximately design-unbiased for $\hat{f}_{\mathcal{U}}(x; b)$, and since $E_M[\hat{f}_{\mathcal{U}}(x; b)] - f(x) \rightarrow 0$ by Theorem 3, we need only examine the asymptotic behavior of $E_M[V_D(\hat{U}(b) | \mathbf{Y})]$. Viewing the components of (6) separately, the model assumptions in (M1) imply that

$$\begin{aligned} &E_M \left[\left(\frac{U_{hi}(b)}{\pi_i^{(h)}} - \frac{U_{hj}(b)}{\pi_j^{(h)}} \right)^2 \right] \\ &= \sum_{k=1}^{Q_{hi}} \left(\frac{1}{\pi_i^{(h)}} \right)^2 (K_b^2 * f_h)(x) + \sum_{k=1}^{Q_{hi}} \sum_{l \neq k}^{Q_{hi}} \left(\frac{1}{\pi_i^{(h)}} \right)^2 \Psi_h(x; b) \\ &\quad + \sum_{k=1}^{Q_{hj}} \left(\frac{1}{\pi_j^{(h)}} \right)^2 (K_b^2 * f_h)(x) + \sum_{k=1}^{Q_{hj}} \sum_{l \neq k}^{Q_{hj}} \left(\frac{1}{\pi_j^{(h)}} \right)^2 \Psi_h(x; b) \\ &\quad - 2 \frac{Q_{hi}Q_{hj}}{\pi_i^{(h)}\pi_j^{(h)}} (K_b * f_h)^2(x) \leq \left[\left(\frac{Q_{hi}}{\pi_i^{(h)}} \right)^2 + \left(\frac{Q_{hj}}{\pi_j^{(h)}} \right)^2 \right] (K_b^2 * f_h)(x). \end{aligned}$$

Similarly,

$$E_M[V_D(\hat{U}_{hi}(b))] \leq \frac{1}{\delta} Q_{hi} (K_b^2 * f_h)(x).$$

So, using the above results with (C3) and (C4),

$$E_M [E_D (\{\hat{U}(b) - U(b)\}^2 | \mathbf{Y})]$$

$$\begin{aligned}
 &= \frac{1}{2} \sum_{h=1}^L E_M \left[\sum_{i=1}^{N_h} \sum_{j \neq i}^{N_h} [\pi_i^{(h)} \pi_j^{(h)} - \pi_{ij}^{(h)}] \left(\frac{U_{hi}(b)}{\pi_i^{(h)}} - \frac{U_{hj}(b)}{\pi_j^{(h)}} \right)^2 + \sum_{i=1}^{N_h} \frac{V_D[\hat{U}_{hi}(b)]}{\pi_i^{(h)}} \right] \\
 &\leq \frac{1}{2} \sum_{h=1}^L \sum_{i=1}^{N_h} \left\{ \sum_{j=1}^{N_h} \alpha_h \left[\frac{Q_{hi}^2 \pi_j^{(h)}}{\pi_i^{(h)}} + \frac{Q_{hj}^2 \pi_i^{(h)}}{\pi_j^{(h)}} \right] + \frac{Q_{hi}}{\delta^2} \right\} (K_b^2 * f_h)(x) \\
 &= \frac{1}{2} \sum_{h=1}^L \left\{ 2 \sum_{i=1}^{N_h} \alpha_h \frac{Q_{hi}^2 n_h}{\pi_i^{(h)}} + \frac{Q_h}{\delta^2} \right\} (K_b^2 * f_h)(x) \\
 &\leq \sum_{h=1}^L [B \alpha_h n_h + 1] \frac{Q_h}{\delta^2} (K_b^2 * f_h)(x) \\
 &= \sum_{h=1}^L [B \alpha_h n_h + 1] \frac{Q_h}{\delta^2 b(t)} [f_h(x) R(K) + o(b(t))].
 \end{aligned}$$

Thus $Q^{-2} E_M [E_D(\{\hat{U}(b) - U(b)\}^2 | \mathbf{Y})] \rightarrow 0$ as $t \rightarrow \infty$ by (C4) and (M3). \square

Proof of Theorem 5. Using Taylor’s theorem as is done in the proof of Lemma 1, it is shown that $\Psi_h(x; b) = g_h(x, x) + o(b)$. Because $Q_{hi} \leq B$, the other terms in (14) are asymptotically negligible relative to the first term. The results in Lemma 1, then, imply (16) and (17). Eq. (18) is shown similarly: Using the expression for the $E_M[V_D(\hat{U}(b) | \mathbf{Y})]$ given in the proof of Theorem 4 together with the relations $\sum_{i=1}^{N_h} \pi_i^{(h)} = n_h$ and $\sum_{j \neq i}^{N_h} \pi_{ij}^{(h)} = (n_h - 1)\pi_i^{(h)}$ from Theorem 6.1 of Lohr (1999), it is shown that

$$E_M V_D[\hat{U}(b)] = \sum_{h=1}^L \sum_{i=1}^{N_h} \frac{(1 - \pi_i^{(h)}) Q_{hi}}{\pi_i^{(h)}} [b^{-1} f_h(x) R(K) + o(b^{-1})].$$

Thus $V_C[\hat{U}(b)] = E_M V_D[\hat{U}(b)] + V_M[U(b)]$, and $V_M[U(b)]$ is derived from (16). \square

Proof of Theorem 6. Since the X_{hi} ’s are independent and $V_D[X(b, t)] = \sigma^2(t) < \infty$, the asymptotic normality requires only verification of Liapunov’s condition that

$$\sum_{h=1}^L \sum_{j=1}^{n_h} E_D[|X_{hj} - E_D X_{hj}|^{2+\eta}] = o(\sigma^{2+\eta}(t))$$

for some $\eta > 0$. Since $u_{hjk} \leq b^{-1}m$ for all h, k , and j , we have that

$$|\hat{U}_{hk}(b)/p_{hk} - U_h(b)| \leq 2(b\delta^2)^{-1} N_h B m$$

for all h and k . Consequently, using the boundedness of N_h/n_h ,

$$\sum_{h=1}^L \sum_{j=1}^{n_h} E_D[|X_{hj} - E_D X_{hj}|^3]$$

$$\begin{aligned} &\leq \sum_{h=1}^L \sum_{j=1}^{n_h} E_D \left[\frac{1}{n_h} (X_{hj} - E_D X_{hj})^2 \max_{h,k} | \hat{U}_{hk}(b) / p_{hk} - U_h(b) | \right] \\ &\leq 2Bm(b\delta^2)^{-1} \sum_{h=1}^L \sum_{j=1}^{n_h} \frac{N_h}{n_h} E_D [(X_{hj} - E_D X_{hj})^2] \\ &\leq b^{-1} C\sigma^2(t), \end{aligned}$$

where C is a constant. Liapunov’s condition with $\eta = 1$ thus holds because $b(t)\sigma(t) \rightarrow \infty$. For the consistency of the variance estimate, we use similar bounds on the second moments of $X_{hj}(b)$ along with the weak law of large numbers. \square

Proof of Theorem 7. Both of these results are proven similarly to the asymptotic normality result in Parzen (1962), again using the Liapunov condition employed in the proof of Theorem 6. To show (23), note that

$$\begin{aligned} &\sum_{h=1}^L \sum_{i=1}^{N_h} E_M \left\{ \left[\frac{1}{Q} \sum_{k=1}^{Q_{hi}} \frac{1}{b} K \left(\frac{x - Y_{hik}}{b} \right) \right]^{2+\eta} \right\} \\ &\leq \frac{1}{(Qb)^{2+\eta}} \sum_{h=1}^L \sum_{i=1}^{N_h} \sum_{k=1}^{Q_{hi}} E_M \left[K^{2+\eta} \left(\frac{x - Y_{hik}}{b} \right) \right] \\ &\leq \left(\frac{B}{Qb} \right)^{1+\eta} [f(x) \int K^{2+\eta}(u) du + o(1)]. \end{aligned}$$

Using the asymptotic model-based variance in (16), then, Liapunov’s condition holds as $Qb \rightarrow \infty$.

The result in (24) is shown by applying the central limit theorem for superpopulation parameter estimators in Bleuer and Kratina (2000). We write

$$\begin{aligned} \sqrt{nb}(\hat{f}_R(x; b) - E_M[\hat{f}_\mathcal{U}(x; b)]) &= \sqrt{nb}(\hat{f}_R(x; b) - \hat{f}_\mathcal{U}(x; b)) \\ &\quad + \sqrt{nb}(\hat{f}_\mathcal{U}(x; b) - E_M[\hat{f}_\mathcal{U}(x; b)]). \end{aligned}$$

The first term in the sum converges to a normal distribution by Theorem 6; the second term converges in distribution to $N(0, \lambda f(x)R(K))$ by (23). Using Lemma 5.1 of Bleuer and Kratina (2000), the necessary conditions for asymptotic independence of the two terms are met because the third absolute moment of $X_{hj}(b)$ is $O(N_h^3 n_h^{-3} b^{-2})$ because of the condition that $N_h(t)p_{hi} > \delta > 0$ and because $E_M[U_{hkl}^3] = O(b^{-2})$. Result (24) follows by Slutsky’s theorem. \square

References

Bellhouse, D., Stafford, J., 1999. Density estimation from complex surveys. *Statist. Sin.* 9, 407–424.
 Bellhouse, D., Stafford, J., 2001. Local polynomial regression in complex surveys. *Survey Methodol.* 27, 197–203.

- Bleuer, S.R., Kratina, I.S., 2000. Some issues in the analysis of complex survey data. Proceedings of the Survey Research Methods Section, American Statistical Association, Washington, DC, pp. 734–739.
- Breidt, F.J., Opsomer, J.D., 2000. Local polynomial regression estimators in survey sampling. *Ann. Statist.* 28, 1026–1053.
- Buskirk, T., 1998. Nonparametric density estimation using complex survey data. Proceedings of the Survey Research Methods Section, American Statistical Association, Washington, DC, pp. 799–801.
- Buskirk, T., 1999. Using nonparametric methods for density estimation with complex survey data. Ph.D. Dissertation, Department of Mathematics, Arizona State University.
- Chambers, R., Dunstan, R., 1986. Estimating distribution functions from survey data. *Biometrika* 73, 597–604.
- Chambers, R., Dorfman, A., Hall, P., 1992. Properties of estimators of the finite population distribution function. *Biometrika* 79, 577–582.
- Cowling, A., Chambers, R., Lindsay, R., Parameswaran, B., 1996. Applications of spatial smoothing to survey data. *Survey Methodol.* 22, 175–183.
- Deville, J.-C., Särndal, C.-E., 1992. Calibration estimators in survey sampling. *J. Amer. Statist. Assoc.* 87, 376–382.
- Francisco, C., Fuller, W., 1991. Quantile estimation with a complex survey design. *Ann. Statist.* 19, 454–469.
- Gasser, T., Müller, H.-G., 1979. Kernel estimation of regression functions. In: Gasser, T., Rosenblatt, M. (Eds.), *Smoothing Techniques for Curve Estimation*. Springer, Heidelberg, pp. 23–68.
- Graubard, B.I., Korn, E.L., 2002. Inference for superpopulation parameters using sample surveys. *Statist. Sci.* 17, 73–96.
- Hájek, J., 1960. Limiting distributions in simple random sampling from a finite population. *Publ. Math. Inst. Hungarian Acad. Sci.* 5, 361–374.
- Hartley, H.O., Sielken Jr., R.L., 1975. A “super-population viewpoint” for finite population sampling. *Biometrics* 31, 411–422.
- Horvitz, D.G., Thompson, D.J., 1952. A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* 47, 663–685.
- Isaki, C.T., Fuller, W.A., 1982. Survey design under the regression superpopulation model. *J. Amer. Statist. Assoc.* 77, 89–96.
- Jones, M.C., 1993. Simple boundary correction for kernel density estimator. *Statist. Comput.* 3, 135–146.
- Korn, E., Graubard, B., 1998a. Scatterplots with survey data. *Amer. Statist.* 52, 58–69.
- Korn, E., Graubard, B., 1998b. Variance estimation for superpopulation parameters. *Statist. Sin.* 8, 1131–1151.
- Korn, E., Midthune, D., Graubard, B., 1997. Estimating interpolated percentiles from grouped data with large samples. *J. Official Statist.* 13, 385–399.
- Krewski, D., Rao, J.N.K., 1981. Inference from stratified samples: properties of the linearization, jackknife and balanced repeated replication methods. *Ann. Statist.* 9, 1010–1019.
- Krieger, A., Pfeffermann, D., 1997. Testing of distribution functions from complex sample surveys. *J. Official Statist.* 13, 123–142.
- Kuk, A., 1988. Estimation of distribution functions and medians under sampling with unequal probabilities. *Biometrika* 75, 97–103.
- Lohr, S., 1999. *Sampling: Design and Analysis*. Duxbury Press, Pacific Grove, CA.
- Parzen, E., 1962. On estimation of a probability function and mode. *Ann. Math. Statist.* 33, 1065–1076.
- Pfeffermann, D., 1993. The role of sampling weights when modeling survey data. *Internat. Statist. Rev.* 61, 317–337.
- Pfeffermann, D., Krieger, A.M., Rinott, Y., 1998. Parametric distributions of complex survey data under informative probability sampling. *Statist. Sin.* 8, 1087–1114.
- Rao, J.N.K., 1994. Estimating totals and distribution functions using auxiliary information at the estimation stage. *J. Official Statist.* 10, 153–165.
- Rao, J.N.K., 1999. Some current trends in sample survey theory and methods. *Sankhya* 61, 1–57.
- Rao, J.N.K., Kovar, J., Mantel, H., 1990. On estimating distribution functions and quantiles from survey data using auxiliary information. *Biometrika* 77, 365–375.

- Rosenblatt, M., 1956. Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* 27, 186–190.
- Sedransk, N., Sedransk, J., 1979. Distinguishing among distributions using data from complex sample designs. *J. Amer. Statist. Assoc.* 74, 754–760.
- Sen, P.K., 1988. Asymptotics in finite population sampling. In: Krishnaiah, P.R., Rao, C.R. (Eds.), *Handbook of Statistics*, Vol. 6. North-Holland, Amsterdam, pp. 291–331.
- U.S. Department of Health and Human Services, 1996. National Center for Health Statistics. Third National Health and Nutrition Examination Survey, 1988–1994, NHANES III Laboratory Data file (CD-ROM). Public use data file documentation number 76200. Centers for Disease Control and Prevention, Hyattsville, MD.
- U.S. Department of Justice, Bureau of Justice Statistics, 1998. National Crime Victimization Survey, 1992–1995 [Computer file]. Conducted by the U.S. Department of Commerce, Bureau of the Census, 4th ICPSR Edition. Interuniversity Consortium for Political and Social Research, Ann Arbor, MI.
- Wand, M.P., Jones, M.C., 1995. *Kernel Smoothing*. Chapman & Hall, New York.
- Woodruff, R., 1952. Confidence intervals for medians and other position measures. *J. Amer. Statist. Assoc.* 47, 635–646.