Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

R.D. Gill, J.A. Wellner

Large sample theory of empirical distributions in biased sampling models

Department of Mathematical Statistics Report MS-R8603 May

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

Large Sample Theory of Empirical Distributions in Biased Sampling Models

Richard Gill
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

Jon A. Wellner
Department of Statistics,
University of Washington,
Seattle Wa. 98195.

Vardi (1985) introduced an s - sample model for biased sampling, gave conditions which guarantee the existence and uniqueness of the nonparametric maximum likelihood estimator $\mathbb{G}_n$ of the common underlying distribution G, and discussed numerical methods for calculating the estimator.

Here we examine the large sample behaviour of the NPMLE $\mathbb{G}_n$, including results on uniform consistency of $\mathbb{G}_n$, convergence of $\sqrt{n}(\mathbb{G}_n - G)$ to a Gaussian process, and asymptotic efficiency of $\mathbb{G}_n$ as an estimator of G. The proofs are based upon recent results for empirical processes indexed by sets and functions, properties of irreducible M-matrices, and the homotopy invariance theorem.

A final section discusses examples and applications to stratified sampling, 'choice-based' sampling in econometrics, and 'case-control' studies in biostatistics.

AMS 1980 Subject Classifications: Primary 62G05, 60F05. Secondary 62G30, 60G44.

Key Words and Phrases: Selection bias models, Vardi's estimator, nonparametric Maximum Likelihood.

1. Introduction: biased sampling models and Vardi's nonparametric MLE.

We begin by describing the biased sampling model studied by Vardi (1985a, 1985b) and a useful iid version of the model.

The s - *sample model* for *biased sampling* introduced by Vardi (1985a) is as follows: Let G be an unknown distribution on $(\mathbf{X},\mathbf{B})$, and let $w_1, \cdots, w_s$ be given non-negative weight functions. Suppose that $X_{i1}, \cdots, X_{in_i}$ are iid $\mathbf{X}$ - valued rv's with distribution F_i given by

$$F_i(A) = \frac{\int_A w_i \, dG}{\int w_i \, dG} \equiv \frac{G(1_A w_i)}{G(w_i)} \tag{1.1}$$

for $A \in \mathbf{B}$ and $i = 1, \cdots, s$. In this model we let $n \equiv n_1 + \cdots + n_s$ and typically assume that $\lambda_{ni} \equiv n_i / n \to \lambda_i > 0$ as $\min(n_i) \to \infty$.

A convenient *iid - model* for biased sampling can be defined in terms of the same weight functions $w_1, \cdots, w_s$, and distribution G by assuming that (X_j, I_j), $j = 1, \cdots, n$ are iid with distribution

$$P(X \in A, \ I = i) = \lambda_i F_i(A) = \lambda_i \frac{\int_A w_i dG}{\int w_i dG} \tag{1.2}$$

for $A \in \mathbf{B}$ and $i = 1, \cdots, s$. In this version of the model the sample sizes $n_i \equiv \sum_{j=1}^{n} 1_{[I_j = i]}$

Report MS-R8603
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

are random and we have $\lambda_{ni} \equiv n_i / n \to_p \lambda_i$ as $n \to \infty$.

In both cases the problem is to estimate G. Vardi (1985a) studied the nonparametric maximum likelihood estimate of G, gave conditions for its existence and uniqueness, and proposed an algorithm for calculating it. Vardi (1985b) also provided a heuristic outline of the large sample theory of his estimator.

In this paper we give a thorough treatment of the asymptotic distribution theory of Vardi's estimator, including results about its efficiency using the methods developed in Begun, Hall, Huang, and Wellner (1983). Consistency is treated in section 2, lower bounds for estimation are discussed in section 3, and asymptotic normality is established in section 4. In the final section, section 5, we discuss examples, applications, and connections between biased sampling models and work in econometrics on 'choice - based' sampling by Cosslett (1981), and in biostatistics on 'case-control' studies by Breslow and Day (1980), Prentice and Pyke (1979), and others.

Now we introduce Vardi's (1985a) nonparametric maximum likelihood estimator of G and some notation which will be used throughout the paper.

Let $\underline{F} \equiv (F_1, \cdots, F_s)^T$ denote the vector of distributions defined by (1.1), and, with $\lambda_{ni} \equiv n_i / n$, $i = 1,\ldots,s$, set

$$F \equiv \underline{\lambda}^T \underline{F} \qquad \text{and} \qquad F_n \equiv \underline{\lambda}_n^T \underline{F}. \tag{1.3}$$

Note that F_n is random in the iid model (1.2). For any vector $\underline{u} \in R^s$, let $\underline{\underline{u}}$ denote the $s \times s$ diagonal matrix with entries u_i on the diagonal, and set

$$\underline{\tilde{w}}(\underline{u}) \equiv \underline{\underline{u}}^{-1} \underline{w}, \qquad r(\underline{u}) \equiv (\underline{\lambda}^T \underline{\tilde{w}}(\underline{u}))^{-1}, \qquad r_n(\underline{u}) \equiv (\underline{\lambda}_n^T \underline{\tilde{w}}(\underline{u}))^{-1} \tag{1.4}$$

where $\underline{w} = (w_1, \cdots, w_s)^T$ is the vector of biasing functions. We also set

$$\underline{W} \equiv \int \underline{w}\, dG \equiv G(\underline{w}) \tag{1.5}$$

and write

$$\underline{\tilde{w}} \equiv \underline{\tilde{w}}(\underline{W}) = \underline{\underline{W}}^{-1} \underline{w}, \qquad r \equiv (\underline{\lambda}^T \underline{\tilde{w}})^{-1}, \qquad r_n \equiv (\underline{\lambda}_n^T \underline{\tilde{w}})^{-1}. \tag{1.6}$$

Now F and F_n are related to G by

$$dF = (\underline{\lambda}^T \underline{\tilde{w}}) dG = r^{-1} dG \tag{1.7}$$

and

$$dF_n = (\underline{\lambda}_n^T \underline{\tilde{w}}) dG = r_n^{-1} dG \tag{1.8}$$

so that, letting $\mathbf{X}^+ \equiv \{x \in \mathbf{X}: r(x)^{-1} > 0\}$ and $\mathbf{X}_n^+ \equiv \{x \in \mathbf{X}: r_n(x)^{-1} > 0\}$, we have

$$G(A) = \int_A r(x) dF(x) \qquad \text{for } A \in \mathbf{B} \cap \mathbf{X}^+ \tag{1.9}$$

and

$$G(A) = \int_A r_n(x) dF_n(x) \qquad \text{for } A \in \mathbf{B} \cap \mathbf{X}_n^+. \tag{1.10}$$

It is clear from (1.9) that G can only be estimated on the set $\mathbf{X}^+$. We therefore define

$$G^+(A) \equiv \frac{G(A \cap \mathbf{X}^+)}{G(\mathbf{X}^+)} = \frac{F(1_{A \cap \mathbf{X}^+} r)}{F(1_{\mathbf{X}^+} r)} \qquad \text{for } A \in \mathbf{B}. \tag{1.11}$$

If support$(G) \subset \mathbf{X}^+$, then $F(r) = G(\mathbf{X}^+) = 1$ so that $G^+ = G$, but in general $G(\mathbf{X}^+) < 1$ and $G^+ \neq G$. Also note that the distribution of the data depends on G through G^+ only, and hence the most we can hope to estimate is G^+ and $\underline{W}^+ \equiv \int \underline{w}\, dG^+ \neq \underline{W}$. Therefore, to ease the notational burden, we henceforth **drop** the plus sign and write $\overline{G}$ for G^+ and $\underline{W}$ for $\underline{W}^+$ even when they are **not**, in fact, equal throughout the remainder of this section and sections 2 - 4. The distinction will be made clearly, and the $+$ sign introduced as needed, in the treatment of the examples in section 5.

Our convention throughout will be to write

$$\int h\, dF \equiv \int_{X^+} h\, dF$$

or, in a notation which we will use frequently,

$$F(h) \equiv F(h\, 1_{X^+})\,. \tag{1.12}$$

Note that the right side of (1.11) is a homogeneous function of degree 0 in the W_i 's: using the above convention and (1.4) and (1.6)

$$\frac{F(1_A r(\underline{W}))}{F(r(\underline{W}))} = \frac{F(1_A r(c\underline{W}))}{F(r(c\underline{W}))} \qquad \text{for any } c > 0\,. \tag{1.13}$$

Thus it follows in particular that, with $c = 1/W_s$,

$$G(A) = \frac{F(1_A\, r(\underline{V}))}{F(r(\underline{V}))} \tag{1.14}$$

where $V \equiv W / W_s$.

Now let $\mathbb{F}_n$ denote the empirical measure of all the X_i's:

$$\mathbb{F}_n \equiv \begin{cases} n^{-1} \sum_{j=1}^{n} \delta_{X_j} & \text{in the iid model} \\ n^{-1} \sum_{i=1}^{s} \sum_{j=1}^{n_i} \delta_{X_{ij}} & \text{in the } s \text{ - sample model.} \end{cases} \tag{1.15}$$

Let $\mathbb{F}_{ni}$ denote the empirical measure for 'sample i':

$$\mathbb{F}_{ni} \equiv \begin{cases} n_i^{-1} \sum_{j=1}^{n} \delta_{X_j} 1_{[I_j = i]} & \text{in the iid model} \\ n_i^{-1} \sum_{j=1}^{n_i} \delta_{X_{ij}} & \text{in the } s \text{ - sample model} \end{cases} \tag{1.16}$$

where δ_x is the measure with mass 1 at x: $\delta_x(A) = 1_A(x)$ for $A \in \mathbf{B}$, and define

$$\begin{aligned} \underline{H}(\underline{u}) &\equiv \int \underline{r}(\underline{u})\underline{\tilde{w}}(\underline{u})\, dF = F((\underline{r}\,\underline{\tilde{w}})(\underline{u})) \\ \underline{H}_n(\underline{u}) &\equiv \int \underline{r}_n(\underline{u})\underline{\tilde{w}}(\underline{u})\, dF_n = F_n((\underline{r}_n\,\underline{\tilde{w}})(\underline{u})) \\ \underline{\mathbb{H}}_n(\underline{u}) &\equiv \int \underline{r}_n(\underline{u})\underline{\tilde{w}}(\underline{u})\, d\mathbb{F}_n = \mathbb{F}_n((\underline{r}_n\,\underline{\tilde{w}})(\underline{u})) \end{aligned} \tag{1.17}$$

In view of (1.7) or (1.9) and (1.8) or (1.10) it follows that

$$\underline{H}(\underline{W}) = \underline{1} \qquad \text{and} \qquad \underline{H}_n(\underline{W}) = 1. \tag{1.18}$$

and, in fact,

$$\underline{H}(c\underline{W}) = \underline{1} \qquad \text{for any } c > 0\,. \tag{1.19}$$

In view of (1.11) and (1.18), estimators $\mathbb{G}_n$ and $\underline{\mathbb{W}}_n$ of G and $\underline{W} = G(\underline{w})$ may be defined as the solution (provided it exists) of the equations

$$\begin{cases} \underline{\mathbb{H}}_n(\underline{\mathbb{W}}_n) = \underline{1} \\ \mathbb{G}_n(h) = \dfrac{\mathbb{F}_n(h\, r_n(\underline{\mathbb{W}}_n))}{\mathbb{F}_n(r_n(\underline{\mathbb{W}}_n))} & \text{for all } h \\ \underline{\mathbb{W}}_n = \mathbb{G}_n(\underline{w})\,. \end{cases} \tag{1.20}$$

Alternatively, in view of (1.14) and (1.19) we can first estimate $V \equiv \underline{W} / W_s$ as the solution, if it exists, of

$$\mathbb{H}_{ni}(\underline{\mathbb{V}}_n) = 1, \qquad i = 1, \cdots, s-1 \tag{1.21}$$

and then estimate G by

$$\mathbb{G}_n(h) \equiv \frac{\mathbb{F}_n(h\, r_n(\mathbb{V}_n))}{\mathbb{F}_n(r_n(\underline{\mathbb{V}}_n))} . \tag{1.22}$$

The latter approach is the one taken by Vardi (1985a, 1985b).

In our notation, Vardi's (1985a) condition for existence of a unique nonparametric maximum likelihood estimator (i.e. a solution of (1.20) or (1.21)) is:

$$n \sum_{i \in B} \int_{[x:\, w_i(x) > 0]} d\mathbb{F}_n(x) > \sum_{i \in B} n_i \qquad \text{for every nonempty proper subset } B \text{ of } \{1,...,s\} . \tag{1.23}$$

Vardi (1985a) shows that (1.23) has several equivalent forms, one of which is in terms of a directed graph M on s vertices defined as follows: a directed edge connects vertex i to vertex i', $i \to i'$, if and only if $\int w_i \, d\mathbb{F}_{ni'} > 0$. Then we say that M is *strongly connected* if for any two vertices x and x' there exists a directed path from x to x' and a directed path from x' to x. Vardi (1985a) shows that (1.23) is equivalent to:

$$\text{the graph } M \text{ is strongly connected} . \tag{1.24}$$

Note that M is strongly connected if and only if the matrix with elements $\mathbb{F}_{ni'}(w_i) = \int w_i \, d\mathbb{F}_{ni'}$ is irreducible; see Berman and Plemmons (1979), pages 27 - 30.

THEOREM 1.1. (VARDI). *If (1.23) or equivalently (1.24) holds, then the equations (1.21) have a unique solution and $\mathbb{G}_n$ given in (1.22) is the nonparametric maximum likelihood estimate of G. Equivalently, if (1.23) or (1.24) holds, then the system of equations (1.20) has a unique solution $\mathbb{G}_n$, $\mathbb{W}_n$ which is the nonparametric maximum likelihood estimate of G, W. Conversely, if (1.23) or (1.24) fail, then the equations (1.20) do not have a unique solution $\mathbb{G}_n$, $\underline{\mathbb{W}}_n$.*

2. CONSISTENCY OF $\mathbb{G}_n$.

Our first task is to establish the consistency of the nonparametric maximum likelihood estimator $\mathbb{G}_n$ of G introduced in section 1 under both the iid and s - sample models. Suppose that

$$G(w_i^2) < \infty \qquad \text{for } i = 1, \cdots, s . \tag{2.1}$$

Then it follows from the strong law of large numbers (and the assumption $n_i / n \to \lambda_i$ for $i = 1, \cdots, s$ in the s - sample case) that for any $i,j = 1, \cdots, s$

$$\mathbb{F}_{nj}(w_i) \to_{a.s.} F_j(w_i) = \frac{1}{W_j} G(w_i w_j) \qquad \text{as } n \to \infty . \tag{2.2}$$

Thus we can define a graph M^* with s vertices and an edge from i to j, $i \leftrightarrow j$, if and only if

$$G(w_i w_j) > 0 . \tag{2.3}$$

In terms of the graph M^*, an asymptotic version of the condition (1.23) is:

$$M^* \quad \text{is a connected graph} . \tag{2.4}$$

The main result of this section can now be stated as:

THEOREM 2.1. (CONSISTENCY OF $\mathbb{G}_n$). *Suppose that (2.1) and (2.4) hold, and that* $\mathcal{H}$ *is a collection of functions of the form* $\mathcal{H} = \{h_e 1_C: C \in \mathcal{C}\}$ *where the envelope function* h_e *satisfies* $G(h_e) = F(rh_e) < \infty$ *and* $\mathcal{C}$ *is a Vapnik - Chervonenkis class of subsets of* $\mathbf{X}$. *Then*

$$\|\mathbb{G}_n - G\| \equiv \sup\{|\mathbb{G}_n(h) - G(h)|: h \in \mathcal{H}\} \to_{a.s.} 0 \qquad \text{as } n \to \infty . \quad (2.5)$$

COROLLARY 2.1. *If* $\mathbf{X} = R^d$ *and* $\mathcal{H}$ *is the class of all indicator functions of lower left orthants, or of all rectangles, or of all balls, or the class of all half - spaces, then (2.1) and (2.4) imply that (2.5) holds.*

REMARK 2.1. Several variants of theorem 2.1 are possible, depending on which basic Glivenko - Cantelli theorem is used. For some other possible formulations, see e.g. Dudley (1985) theorems 6.1.5 and 11.1.6 or Pollard (1984).

REMARK 2.2. What can be said if (2.4) fails? If $r \leqslant s$ is the number of components of the graph M^*, let $C_1, \cdots, C_r$ be a partition of the indices $\{1, \cdots, s\}$ corresponding to the r components, and define $\mathbf{X}_j \equiv \{x \in \mathbf{X}: \sum_{i \in C_j} \lambda_i w_i(x) > 0$ for some $i \in C_j\}$. (Thus when $r = 1$, $\mathbf{X}_1 = \mathbf{X}^+$.) Then we can estimate $G_j \equiv G(\cdot 1_{\mathbf{X}_j}) / G(1_{\mathbf{X}_j})$ for each j : note that Vardi's estimator works asymptotically for each component of the graph separately, that M is reducible, rank(M) $= s - r$, and hence, after a relabeling of the samples, M is a block - diagonal matrix with blocks M_{jj} which are irreducible and rank$(M_{jj}) = \dim(M_{jj}) - 1$.

An important part of our argument, which we use repeatedly in the following sections, concerns the derivative matrix $\nabla \underline{H}^T$ of the vector of functions $\underline{H}(\underline{u})$ defined in (1.17). By straightforward calculation (justified by the monotone convergence theorem)

$$(\nabla H)(\underline{u}) \equiv (\nabla \underline{H}^T(\underline{u}))^T$$

$$= - \left\{ \underline{\underline{H(u)\lambda^{-1}}} - F(r^2(u)\underline{\tilde{w}(u)}\underline{\tilde{w}^T(u)}) \right\} \underline{\underline{\lambda u^{-1}}} \quad (2.6)$$

$$= - \left\{ \underline{\underline{\lambda^{-1}}} - F(r^2 \underline{\tilde{w}}\,\underline{\tilde{w}^T}) \right\} \underline{\underline{\lambda u^{-1}}} \qquad \text{when } \underline{u} = c\underline{W} \text{ so } \underline{H}(\underline{u}) = \underline{1}. \quad (2.7)$$

$$\equiv - M \underline{\underline{\lambda u^{-1}}}$$

PROPOSITION 2.1. *If the graph* M^* *is connected, then the matrix*

$$M \equiv \underline{\underline{\lambda^{-1}}} - F(r^2 \underline{\tilde{w}}\,\underline{\tilde{w}^T})$$

defined in (2.7) has rank s-1. In fact, every principal proper submatrix of M *is nonsingular, and the same holds for* $(\nabla H)(\underline{u})$. *In particular, if the ith row and column are deleted, a matrix of full rank (s-1) results.*

To prove the preceding assertions, it will be convenient to work with $\underline{\mathbb{V}}_n \equiv (\mathbb{V}_{n1}, \cdots, \mathbb{V}_{ns-1}, 1)^T$ defined by (1.21) and $\underline{V} \equiv (V_1, \cdots, V_{s-1}, 1)^T = \underline{W} / W_s$.

PROPOSITION 2.2. *If the graph* M^* *is connected, then the unique solution* $\underline{\mathbb{V}}_n$ *of (1.21) satisfies*

$$\underline{\mathbb{V}}_n \to_{a.s.} \underline{V} \equiv \underline{W} / W_s \qquad \text{as } n \to \infty \quad (2.8)$$

where $\underline{V}$ *is the unique solution of*

$$H_i(\underline{V}) = 1, \qquad i = 1, \cdots, s-1 . \quad (2.9)$$

This proposition will be proved by means of the following lemmas.

LEMMA 2.1. $(\nabla H)(\underline{u})\cdot\underline{u} = \underline{0}$ *for all* $u \in R^{+s}$.

PROOF. Since each H_i is homogeneous of degree zero, $\underline{H}(c\,\underline{u}) = \underline{H}(\underline{u})$ for all $c > 0$. Hence

$$\underline{0} = \frac{\partial}{\partial c}\underline{H}(c\,\underline{u}) = (\nabla H)(c\,\underline{u})\cdot\underline{u} \qquad \text{for all} \quad c > 0 \tag{a}$$

and in particular for $c = 1$. □

LEMMA 2.2. $(-(\nabla H)(\underline{u}))_{ij} \leqslant 0$ *for* $i \neq j$, *and hence the upper left* $(s-1)\times(s-1)$ *submatrix of* $-(\nabla H)$ *is in the collection* $Z^{(s-1)\times(s-1)}$ *of Berman and Plemmons (1979) page 132.*

PROOF. Note that $F(r^2(\underline{u})\tilde{w}_i(\underline{u})\tilde{w}_j(\underline{u}) \geqslant 0$ in (2.6). □

PROOF OF PROPOSITION 2.1. By Berman and Plemmons (1979) exercise 6.4.14 page 155, and the facts that $M \in Z^{s\times s}$, $\lambda >> 0$, and $M\lambda = 0$, M is a singular M - matrix with 'property c'. Since the graph M^* is connected, M is also irreducible; see theorem 2.7, Berman and Plemmons (1979), page 30. Hence by theorem 6.4.16(4), Berman and Plemmons (1979), every principal proper submatrix of M is nonsingular. The same argument applies to $(\nabla H)(\underline{u})$ starting from the observation that $(\nabla H)(\underline{u})\cdot\underline{u} = 0$. □

LEMMA 2.3. *If* M^* *is connected (i.e. (2.4) holds), then every solution of (2.9) is an isolated solution.*

PROOF This follows from Proposition 2.1 and the implicit function theorem; see e.g. Apostol (1957) or Ortega and Rheinboldt (1970). □

LEMMA 2.4. (VARDI) *If* M^* *is connected, then with probability one for* $n \geqslant$ *some* N_ω *there is a unique solution* $\underline{\mathbb{V}}_n$ *of the equations (1.21).*

PROOF. This follows from Vardi (1985a), theorems 1 and 2. □

LEMMA 2.5. *Let* K *be a compact set in* R^{+s} *with all coordinates bounded away from* 0. *Then*

$$\max_{1\leqslant i\leqslant s}\ \sup_{\underline{u}\in K} |\mathbb{H}_{ni}(\underline{u}) - H_i(\underline{u})| \rightarrow_{a.s.} 0 \qquad \text{as} \quad n\rightarrow\infty . \tag{2.10}$$

PROOF. Note that the functions

$$h_i(x, \underline{u}) \equiv r_n(\underline{u})\tilde{w}_i(\underline{u})(x), \qquad i = 1, \cdots, s, \quad \underline{u} \in K \tag{a}$$

are uniformly bounded (by $\max_{1\leqslant i\leqslant s-1}(1/\lambda_i)$). Since K is compact and since each h_i is a monotone decreasing function of each coordinate of u, the supremum in (2.10) may easily be reduced to a maximum over a finite set of u 's in K and then (2.10) follows by the strong law of large numbers. Alternatively, (2.10) follows from a Glivenko - Cantelli theorem for $\mathbb{F}_n$ indexed by the collection of functions in (a): see e.g. Pollard (1982), (1984), or Dudley (1985). □

PROOF OF PROPOSITION 2.2. If (2.4) holds, then by lemma 2.3, $\underline{V} = \underline{W}/W_s$ is an isolated solution of (2.9). Let K be a small compact set containing V in its relative interior K^0 in $R^{+(s-1)}$ and no other solution. Since $\underline{H}(v_1, \cdots, v_{s-1}, 1)$, is continuous,

$$\inf_{\underline{v}\in\partial K}\{\max_{1\leqslant i\leqslant s-1}|H_i(\underline{v}) - 1|\} \equiv \epsilon > 0 . \tag{a}$$

Furthermore, for $0 < \delta < \epsilon$,

$$\max_{1 \leq i \leq s-1} \sup_{\underline{v} \in K} |\mathbb{H}_{ni}(\underline{v}) - H_i(\underline{v})| < \frac{\delta}{7} \quad \text{for } n \geq \text{some } N_\omega \tag{b}$$

by lemma 2.5. Hence by the homotopy invariance theorem (or 6.1.6 page 152) of Ortega and Rheinboldt (1970), for $n \geq N_\omega$

$$\deg(\underline{\mathbb{H}}_n, K^0, \underline{1}) = \deg(\underline{H}, K^0, \underline{1}) \tag{c}$$

where the right side is ± 1 since the upper left $(s-1)\times(s-1)$ submatrix of ∇H^T is nonsingular everywhere by proposition and since H has only one solution of (2.9) in K^0. Thus the unique solution $\hat{\underline{V}}$ of (1.21) guaranteed by lemma 2.4 is also in K^0. Since K can be chosen to be arbitrarily small, this implies (2.8).

Now this same argument applies to any other solution $\tilde{\underline{V}}$ of (2.9), since any other such solution is isolated by lemma 2.3. Hence $\underline{V} = \underline{V}$; i.e. (2.9) has only one solution, namely $\underline{V}$. □

PROOF OF THEOREM 2.1. For a fixed function h we write

$$\begin{aligned}
&|\mathbb{F}_n(hr_n(\underline{\mathbb{V}}_n)) - F(hr)| \\
&\quad \leq |\mathbb{F}_n(h(r_n(\underline{\mathbb{V}}_n) - r))| + |\mathbb{F}_n(hr) - F(hr)| \\
&\quad \leq \|\frac{r_n(\underline{\mathbb{V}}_n)}{r} - 1\|_\infty |\mathbb{F}_n(hr)| + |\mathbb{F}_n(hr) - F(hr)| \\
&\quad \to_{a.s.} 0 \quad \text{as } n \to \infty
\end{aligned} \tag{a}$$

for any fixed function h with $F(hr) = G(h) < \infty$ by the strong law of large numbers, and by continuity and boundedness of $r(\underline{u})/r$ (as a function of x) together with proposition 2.2. Thus by Pollard's Glivenko - Cantelli theorem (see Dudley (1985), theorems 11.1.2 and 11.1.6), for $\mathcal{H}$ of the form hypothesized,

$$\begin{aligned}
&\sup_{h \in \mathcal{H}} |\mathbb{F}_n(hr_n(\underline{\mathbb{V}}_n)) - F(hr)| \\
&\quad \leq \|\frac{r_n(\underline{\mathbb{V}}_n)}{r} - 1\|_\infty \sup_{h \in \mathcal{H}} |\mathbb{F}_n(hr)| + \sup_{h \in \mathcal{H}} |\mathbb{F}_n(hr) - F(hr)| \\
&\quad \to_{a.s.} 0 \cdot \sup_{h \in \mathcal{H}} |F(hr)| + 0 = 0 .
\end{aligned} \tag{b}$$

Thus, since

$$\begin{aligned}
|\mathbb{G}_n(h) - G(h)| &= |\frac{\mathbb{F}_n(hr_n(\underline{\mathbb{V}}_n))}{\mathbb{F}_n(r_n(\underline{\mathbb{V}}_n))} - \frac{F(hr)}{F(r)}| \\
&\leq \frac{|\mathbb{F}_n(hr_n(\underline{\mathbb{V}}_n)) - F(hr)|}{\mathbb{F}_n(r_n(\underline{\mathbb{V}}_n))} + \frac{|F(hr)|}{F(r)} \frac{|\mathbb{F}_n(r_n(\underline{\mathbb{V}}_n)) - F(r)|}{\mathbb{F}_n(r_n(\underline{\mathbb{V}}_n))},
\end{aligned}$$

(2.5) follows from (a) with $h = 1$ and (b). □

Before ending this section, we record some useful facts concerning generalized inverses of the matrix M which will be used repeatedly in sections 2 - 5.

PROPOSITION 2.3. *If the graph* M^* *is connected, then the matrix* M *has a* $\{1,2\}$ *- inverse* M^-. *Thus* M^- *satisfies both*

$$M M^- M = M \tag{2.11}$$

and

$$M^- M M^- = M^- . \tag{2.12}$$

Any such {1,2} - inverse M^- also satisfies:

$$\underline{y} = M\underline{x} \quad \text{implies} \quad \underline{x} = M^-\underline{y} + c\underline{\lambda} \quad \text{for some} \quad c \tag{2.13}$$

where $\underline{\lambda}$ is the unique eigenvector of M with 0 eigenvalue. If, in addition, M^- is the {1,2,3,4} or Moore - Penrose generalized inverse of M, then

$$M M^- = 1 - \underline{\theta}\underline{\theta}^T \tag{2.14}$$

where $\theta \equiv \underline{\lambda} / |\underline{\lambda}|$.

PROOF. By proposition 2.2, deleting any row and column, in particular the last row and column, from M yields an $(s-1)\times(s-1)$ matrix M_{11} of rank $s-1$. Thus a {1,2} inverse M^- of M is given by

$$M^- = \begin{bmatrix} M_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} ; \tag{2.15}$$

see e.g. Seber (1977) page 76 and also Berman and Plemmons (1979) page 117.

Since M is symmetric, another way to get a {1,2} - inverse M^- satisfying (2.11) and (2.12) is to use the decomposition $M = P D P^T$ where the columns of P are the normalized eigenvectors of M and D is the diagonal matrix of the eigenvalues $d_1 \geqslant d_2 \geqslant \cdots \geqslant d_{s-1} > d_s = 0$ of M. Then

$$M^- = P D^- P^T \tag{2.16}$$

where D^- is the diagonal matrix with $D_{ii}^- = 1/d_i$ for $i = 1, \cdots, s-1$ and $D_{ss}^- = 0$, is in fact the Moore - Penrose (or {1,2,3,4}) generalized inverse of M^- ; see Berman and Plemmons (1979) page 117.

Now for any {1,2} - inverse M^- (an inverse satisfying both (2.11) and (2.12)), $M^- M$ is the projector on $R(M^-)$ along $N(M)$; see Berman and Plemmons (1979) page 118. Thus if $\underline{x} = \underline{x}_0 + c\underline{\lambda}$ and $\underline{y} = M\underline{x}$, it follows that $M^-\underline{y} = M^- M\underline{x} = \underline{x}_0 = \underline{x} - c\underline{\lambda}$ and hence (2.13) holds. The proof of (2.14) proceeds by direct computation using $M = P D P^T$ and (2.16). □

3. A LOWER BOUND FOR ESTIMATION OF G.

Our goal here will be to derive a convolution theorem for regular estimates of G. Our computations will follow the approach of Begun, Hall, Huang, and Wellner (1983). An alternative derivation could be based on the results of Millar (1985).

We suppose that support(G) $\subset$ support($\lambda^T w$) so that $G = G^+$ and $F(r) = 1$, that $r^{-1} = \underline{\lambda}^T \underline{\tilde{w}}$ and r are bounded, that G has density g with respect to ν, and write

$$f_i = \frac{w_i g}{W_i} \equiv \tilde{w}_i g , \tag{3.1}$$

with

$$W_i = \int w_i g \, d\nu = G(w_i) = \langle w_i, 1 \rangle_G$$

$$r = (\sum_{i=1}^{s} \lambda_i \tilde{w}_i)^{-1} = (\underline{\lambda}^T \underline{\tilde{w}})^{-1} , \tag{3.2}$$

$$f = \sum_{i=1}^{s} \lambda_i f_i = r^{-1} g . \tag{3.3}$$

It is easily verified that each f_i is Hellinger differentiable with respect to g with derivative

$$A_i\beta = f_i^{1/2}\left(\frac{\beta}{g^{1/2}} - \int \tilde{w}_i \beta g^{1/2}\,d\nu\right), \qquad i = 1, \cdots, s \tag{3.4}$$

so that

$$A_i^T A_i \beta = \frac{f_i\beta}{g} - \frac{f_i}{g^{1/2}} \int \beta f_i g^{-1/2}\,d\nu, \qquad i = 1, \cdots, s \tag{3.5}$$

and hence

$$\begin{aligned} U\beta &\equiv \sum_{i=1}^{s} \lambda_i A_i^T A_i \beta \\ &= \sum_{i=1}^{s} \lambda_i \tilde{w}_i \beta - \sum_{i=1}^{s} \lambda_i \tilde{w}_i g^{1/2} <\tilde{w}_i g^{1/2}, \beta>_\nu \\ &= r^{-1}\beta - \underline{\tilde{w}}^T \underline{\underline{\lambda}} g^{1/2} <\underline{\tilde{w}} g^{1/2}, \beta>_\nu . \end{aligned} \tag{3.6}$$

The operator U maps $L^2(\nu)$ to $L^2(\nu)$ (since r^{-1} is bounded) and is precisely the operator $A^T A$ of Begun, Hall, Huang, and Wellner (1983) in the iid case. It is often convenient to work instead with $\tilde{U}$ mapping $L^2(G)$ to $L^2(G)$ defined by

$$\tilde{U}\tilde{\beta} \equiv g^{-1/2} U(g^{1/2}\tilde{\beta}), \qquad \text{for } \tilde{\beta} \in L^2(G).$$

Thus

$$\tilde{U}\tilde{\beta} = r^{-1}\tilde{\beta} - \underline{\tilde{w}}^T \underline{\underline{\lambda}} <\underline{\tilde{w}}, \tilde{\beta}>_G \tag{3.7}$$

and we have

$$\begin{aligned} <\underline{\tilde{w}}, r\tilde{U}\tilde{\beta}>_G &= (I - <\underline{\tilde{w}}, r\underline{\tilde{w}}^T>_G \underline{\underline{\lambda}}) <\underline{\tilde{w}}, \tilde{\beta}>_G \\ &= (\underline{\underline{\lambda}}^{-1} - <\underline{\tilde{w}}, r\underline{\tilde{w}}^T>_G) \underline{\underline{\lambda}} <\underline{\tilde{w}}, \tilde{\beta}>_G \\ &= M \underline{\underline{\lambda}} <\underline{\tilde{w}}, \tilde{\beta}>_G \end{aligned} \tag{3.8}$$

where

$$M \equiv \underline{\underline{\lambda}}^{-1} - <\underline{\tilde{w}}, r\underline{\tilde{w}}^T>_G . \tag{3.9}$$

Thus, by (2.13), (3.8) can be inverted to yield

$$\underline{\underline{\lambda}} <\underline{\tilde{w}}, \tilde{\beta}>_G = M^- <\underline{\tilde{w}}, r\tilde{U}\tilde{\beta}>_G - \alpha \underline{\lambda} \tag{3.10}$$

where the constant $\alpha = \alpha(\tilde{\beta})$ is still to be determined. Substitution of (3.10) into (3.7) yields

$$r\tilde{U}\tilde{\beta} \equiv \tilde{\beta} - r\underline{\tilde{w}}^T M^- <\underline{\tilde{w}}, r\tilde{U}\tilde{\beta}>_G - \alpha \tag{3.11}$$

which implies

$$\begin{aligned} \tilde{U}^{-1}\tilde{\beta} &= r\tilde{\beta} + r\underline{\tilde{w}}^T M^- <\underline{\tilde{w}}, r\tilde{\beta}>_G + \alpha \\ &\equiv \tilde{V}\tilde{\beta} + \alpha \end{aligned} \tag{3.12}$$

where

$$\tilde{V}\tilde{\beta} = r\tilde{\beta} + r\underline{\tilde{w}}^T M^- <\underline{\tilde{w}}, r\tilde{\beta}>_G . \tag{3.13}$$

Now we want $\tilde{U}^{-1}\tilde{\beta} \perp 1$ in $L^2(G)$; therefore

$$\alpha = \alpha(\tilde{\beta}) = - <\tilde{V}\tilde{\beta}, 1>_G$$

and

$$\tilde{U}^{-1}\tilde{\beta} = \tilde{V}\tilde{\beta} - <\tilde{V}\tilde{\beta},1>_G$$
$$= \tilde{V}\tilde{\beta} - G(\tilde{V}\tilde{\beta}). \tag{3.14}$$

Thus it follows that

$$K(h,\tilde{h}) \equiv <g^{1/2}(h - G(h)), U^{-1}g^{1/2}(\tilde{h} - G(\tilde{h}))>_\nu$$
$$= <h - G(h), \tilde{U}^{-1}(\tilde{h} - G(\tilde{h}))>_G$$
$$= <h - G(h), \tilde{V}(\tilde{h} - G(\tilde{h})) - G(\tilde{V}(\tilde{h} - G(\tilde{h})))>_G$$
$$= <h - G(h), \tilde{V}(\tilde{h} - G(\tilde{h}))>_G \tag{3.15}$$

where $\tilde{V}$ is given in (3.13) and M is given in (3.9).

We say that $\hat{\mathbb{G}}_n$ is a *regular estimator* of G if, under $P_n = P_{g_n}$,

$$\sqrt{n}(\hat{\mathbb{G}}_n - G_n) \Rightarrow \mathbb{S} \qquad \text{independent of } \beta \tag{3.16}$$

whenever $\sqrt{n}(g_n^{1/2} - g^{1/2}) \to \beta$ in $L_2(\nu)$. Then the above calculations in combination with theorem 4.1 of Begun, Hall, Huang, and Wellner (1983) yield:

THEOREM 3.1. *Suppose that M^* is connected, that $F(r) = 1$ so that $G^+ = G$, and that both r^{-1} and r are bounded. Then the limit process $\mathbb{S}$ for any regular estimator of G in the biased sampling model (1.1) can be represented as*

$$\mathbb{S} = \mathbb{Z} + \mathbb{W} \tag{3.17}$$

where $\mathbb{W}$ is independent of the mean 0 Gaussian process $\mathbb{Z}$ with covariance function given by (3.15).

A local asymptotic minimax lower bound can also be stated on the basis of the preceeding calculations, but we forego this here.

4. ASYMPTOTIC NORMALITY OF THE NONPARAMETRIC MLE $\mathbb{G}_n$.

We now study the asymptotic behavior of the process

$$\mathbb{Z}_n \equiv \sqrt{n}(\mathbb{G}_n - G) \tag{4.1}$$

regarded as a process indexed by functions $h \in L_2(G)$. We do this in both the iid case and the s - sample case simultaneously (getting the same result in each case). To accomplish this, we define

$$\mathbb{X}_n^* \equiv \sqrt{n}(\mathbb{F}_n - F_n) \tag{4.2}$$

$\lambda_{ni} \equiv n_i / n,\ i = 1, \cdots, s$ and

$$F_n \equiv \sum_{i=1}^{s} \lambda_{ni}F_i \neq \sum_{i=1}^{s} \lambda_i F_i \equiv F. \tag{4.3}$$

as in (1.3). Here the n_i's are random sample sizes in the iid sampling model (and hence the λ_{ni}'s are random too), and deterministic sample sizes in the s - sample model, and we have

$$\begin{aligned} \lambda_{ni} &\to_p \lambda_i && \text{as } n \to \infty && \text{in the iid model} \\ \lambda_{ni} &\to \lambda_i && \text{as } n \to \infty && \text{in the } s \text{ - sample model} \end{aligned} \tag{4.4}$$

by the weak law of large numbers in the first case, and by assumption in the second case. Centering by F_n rather than F amounts to conditioning on the λ_{ni}'s in the iid case. In either case we can write

$$\mathbb{X}_n^* = \sum_{i=1}^{s} \sqrt{\lambda_{ni}} \sqrt{n_i} (\mathbb{F}_{ni} - F_i) . \tag{4.5}$$

Then, if $\mathcal{F}$ is a *Donsker class for* F (or if $\mathcal{F}$ is a Donsker class for all F_i, $i = 1, \cdots, s$ in the s - sample case), for a special construction as in Dudley and Philipp (1983),

$$\|\mathbb{X}_n^* - \mathbb{X}^*\|_{\mathcal{F}} \equiv \sup_{f \in \mathcal{F}} |\mathbb{X}_n^*(f) - \mathbb{X}^*(f)| \to_p 0 \quad \text{as} \quad n \to \infty \tag{4.6}$$

where $\mathbb{X}^*$ is (a sequence of) mean 0 Gaussian processes with covariance function

$$\begin{aligned} \mathrm{Cov}(\mathbb{X}^*(h), \mathbb{X}^*(\tilde{h})) &= \sum_{i=1}^{s} \lambda_i \{F_i(h\tilde{h}) - F_i(h)F_i(\tilde{h})\} \\ &= F(h\tilde{h}) - \underline{F}(h)^T \underline{\underline{\lambda}} \underline{F}(\tilde{h}) \qquad (4.7) \\ &= G(r^{-1}h\tilde{h}) - G(h\underline{\tilde{w}}^T) \underline{\underline{\lambda}} \underline{G}(\tilde{h}\underline{\tilde{w}}) . \qquad (4.8) \end{aligned}$$

This follows in the iid case by a random sample size version of the classical theorems which can easily be deduced from the results of Dudley and Philipp (1983), and in the s - sample case by straightforward calculation, also using the Dudley - Philipp results or other central limit theorems for the empirical process in the one - sample case such as Pollard (1982) or Ossiander (1985).

Note that in the iid case $\mathbb{X}_n^*$ and $\mathbb{X}^*$ are related to the more familiar empirical process $\mathbb{X}_n \equiv \sqrt{n}(\mathbb{F}_n - F)$ and its limit $\mathbb{X}$ by

$$\mathbb{X}_n = \mathbb{X}_n^* + \sqrt{n}(\underline{\lambda}_n - \underline{\lambda})^T \underline{F} \tag{4.9}$$

and

$$\mathbb{X} = \mathbb{X}^* + \underline{\mathbb{Z}}_\lambda \underline{F} \tag{4.9'}$$

where $\mathbb{X}_n^*$ and $\sqrt{n}(\underline{\lambda}_n - \underline{\lambda}) \to_d \underline{\mathbb{Z}}_\lambda \cong N_s(\mathbf{0}, \underline{\underline{\lambda}} - \underline{\lambda}\underline{\lambda}^T)$ (by the multivariate CLT for a multinomial random vector) are asymptotically independent by a straightforward covariance calculation. Thus

$$\begin{aligned} \mathrm{Cov}\left[\mathbb{X}(h), \mathbb{X}(\tilde{h})\right] &= \mathrm{Cov}(\mathbb{X}_n^*(h), \mathbb{X}^*(\tilde{h})) + \underline{F}^T(h)(\underline{\underline{\lambda}} - \underline{\lambda}\underline{\lambda}^T)\underline{F}(\tilde{h}) \\ &= F(h\tilde{h}) - F(h)F(\tilde{h}) . \end{aligned}$$

The first step in studying $\mathbb{Z}_n$ is to establish asymptotic normality of $\sqrt{n}(\underline{\mathbb{W}}^n - \underline{W})$ where $\underline{\mathbb{W}}_n$ results from solving (1.20).

THEOREM 4.1. (CLT FOR $\underline{\mathbb{W}}_n$). *If* M^* *is connected, and* $\underline{\mathbb{W}}_n$ *is the solution of (1.20), then, for the same special construction as in (4.6),*

$$\begin{aligned} \sqrt{n}(\underline{\mathbb{W}}_n - \underline{W}) &\to_p K^- \mathbb{X}^*(r\underline{\tilde{w}}) + \mathbb{Z}_\alpha \underline{W} \qquad (4.10) \\ &= K^- \mathbb{X}^*(r\underline{\tilde{w}}) \qquad (4.11) \\ &\quad + \left[-\mathbb{X}^*(r / F(r)) + \{(1 - F(r))\underline{1}^T - G(r\underline{\tilde{w}}^T)\}M^- \mathbb{X}^*(r\underline{\tilde{w}})\right] \underline{W} / F(r) \\ &= K^- \mathbb{X}^*(r\underline{\tilde{w}}) - \mathbb{X}^*(r + r\underline{\tilde{w}}^T M^- G(r\underline{\tilde{w}}))\underline{W} \quad \text{if} \quad F(r) = 1 \qquad (4.12) \\ &= \mathbb{X}^*(\{K^- \underline{\tilde{w}} - (1 + \underline{\tilde{w}}^T M^- G(r\underline{\tilde{w}}))\underline{W}\} r) \end{aligned}$$

where both

$$K \equiv \underline{\underline{M\lambda W^{-1}}} \quad \text{and} \quad K^- = \underline{\underline{W\lambda^{-1}M^-}} \tag{4.13}$$

have rank $s-1$. *Recall that* $M = \underline{\underline{\lambda}}^{-1} - F(r^2 \underline{\tilde{w}}\,\underline{\tilde{w}}^T)$, *and that* M^- *is a {1,2} - generalized inverse as defined in proposition 2.3.*

PROOF OF THEOREM 4.1. Since

$$\begin{aligned} \underline{1} &= \underline{\mathbb{H}_n}(\underline{\mathbb{W}_n}) = \underline{H_n}(\underline{W}) \\ &= \underline{\mathbb{H}_n}(\underline{W}) + \nabla\underline{\mathbb{H}_n}(\underline{\mathbb{W}_n^*})(\underline{\mathbb{W}_n} - \underline{W}) \end{aligned} \tag{a}$$

where $\underline{\mathbb{W}_n^*}$ lies on the line segment between $\underline{\mathbb{W}_n}$ and $\underline{W}$, it follows that

$$-\nabla\underline{\mathbb{H}_n}(\underline{\mathbb{W}_n^*})\sqrt{n}(\underline{\mathbb{W}_n} - \underline{W}) = \sqrt{n}(\underline{\mathbb{H}_n}(\underline{W}) - \underline{H_n}(\underline{W})) = \mathbb{X}_n^*(\underline{r\tilde{w}}) \tag{b}$$

and hence, letting $n \to \infty$, and writing $\underline{\mathbb{Z}_W} \equiv \lim_n \sqrt{n}(\underline{\mathbb{W}_n} - \underline{W})$,

$$\mathbb{X}^*(\underline{r\tilde{w}}) = -\nabla\underline{H}(\underline{W})\underline{\mathbb{Z}_W} = \underline{\underline{M\lambda W^{-1}}}\,\underline{\mathbb{Z}_W} \tag{c}$$

or, by (2.13) in proposition 2.3

$$\begin{aligned} \underline{\mathbb{Z}_W} &= \underline{\underline{W\lambda^{-1}M^-}}\mathbb{X}^*(\underline{r\tilde{w}}) + \underline{\underline{W\lambda^{-1}}}\mathbb{Z}_\alpha\underline{\lambda} \\ &= K^-\mathbb{X}^*(\underline{r\tilde{w}}) + \mathbb{Z}_\alpha \underline{W}\,. \end{aligned} \tag{d}$$

This completes the proof of (4.10). We postpone identification of $\mathbb{Z}_\alpha$ as given in (4.11) since this step requires the convergence of $\mathbb{Z}_n$ which will be established in theorem 4.2. In proving theorem 4.2 we will use only (4.10), and *not* (4.11). □

Now define a Gaussian process $\mathbb{Z}^*$ by

$$\mathbb{Z}^*(h) = \mathbb{X}^*(\tilde{V}(h - G(h))) \tag{4.14}$$

where, for β with $\beta\sqrt{r} \in L_2(G)$,

$$\begin{aligned} \tilde{V}\beta &\equiv \frac{r}{F(r)}\beta + \underline{r\tilde{w}}^T M^- G(\underline{r\tilde{w}}\beta) \\ &= \tilde{V}\beta \qquad \text{of (3.13) if } F(r) = 1\,. \end{aligned} \tag{4.15}$$

THEOREM 4.2. (CLT FOR $\mathbb{Z}_n$). *Suppose that (2.1) and (2.4) hold (so the graph* M^* *is connected), and that* $\mathcal{H}$ *is a collection of functions containing the constant function* 1 *such that (4.6) holds for the class* $\mathcal{F} \equiv \{hr : h \in \mathcal{H}\}$. *We also assume that* $G(h_e^2 r) = F(h_e^2 r^2) < \infty$ *where* h_e *is an envelope function for* $\mathcal{H}$; $|h| \leq h_e$ *for all* $h \in \mathcal{H}$. *Then,*

$$\|\mathbb{Z}_n - \mathbb{Z}^*\|_{\mathcal{H}} \equiv \sup_{h \in \mathcal{H}} |\mathbb{Z}_n(h) - \mathbb{Z}^*(h)| \to_p 0 \quad \text{as } n \to \infty \tag{4.16}$$

where $\mathbb{Z}$ *is the (sequence of) mean* 0 *Gaussian processes (4.14) with covariance function*

$$\begin{aligned} K^*(h,\tilde{h}) &\equiv \mathrm{Cov}(\mathbb{Z}^*(h), \mathbb{Z}^*(\tilde{h})) \\ &= G((h - G(h))\tilde{V}(\tilde{h} - G(\tilde{h}))) \\ &= G(r^+(h - \bar{h})(\tilde{h} - \bar{\tilde{h}})) + G((h - \bar{h})\underline{r\tilde{w}}^T)M^- G((\tilde{h} - \bar{\tilde{h}})\underline{r\tilde{w}}) \\ &\qquad \text{with } r^+ \equiv r / F(r), \text{ and } \bar{h} \equiv G(h) \end{aligned} \tag{4.17}$$

$$= G((h - G(h))\tilde{V}(\tilde{h} - G(\tilde{h}))) \quad \text{if } F(r) = 1\,. \tag{4.18}$$

where $\tilde{V}$ *is given by (4.15).*

There are many sufficient conditions which imply that (4.6) holds for $\mathcal{F} \equiv \{hr: h \in \mathcal{H}\}$; see e.g. Dudley and Philipp (1983), Pollard (1984), or Ossiander (1985). The important special case of $\mathcal{H} = \{1_C: C \in \mathcal{C}\}$ where $\mathcal{C}$ is a Vapnik - Chervonenkis class of subsets of **X**, which follows as a consequence of Pollard's (1982) theorem, is singled out in the following corollary.

COROLLARY 4.2. *Suppose that* $\mathbf{X} = R^d$ *and* $\mathcal{H}$ *is the collection of all indicator functions of lower left orthants, or of all rectangles, or of all balls, or the class of all half spaces. If* $G(r) = F(r^2) < \infty$ *and (2.1) and (2.4) hold, then* $\mathbb{Z}_n$ *satisfies (4.16).*

PROOF OF THEOREM 4.2. Now for any fixed h with $F(h^2r^2) = G(h^2r) < \infty$

$$\mathbb{Z}_n(h) \equiv \sqrt{n}(\mathbb{G}_n(h) - G(h)) \tag{a}$$

$$= \sqrt{n}\left\{\frac{\mathbb{F}_n(hr_n(\underline{\mathbb{W}}_n)}{\mathbb{F}_n(r_n(\underline{\mathbb{W}}_n))} - \frac{F_n(hr_n(\underline{W}_n)}{F_n(r_n(\underline{W}_n))}\right\}$$

$$= \{\mathbb{X}_n^*(hr_n - r_n\frac{F_n(hr_n)}{F_n(r_n)})$$

$$+ \mathbb{F}_n((h - r_n\frac{F_n(hr_n)}{F_n(r_n)})\sqrt{n}(r_n(\underline{\mathbb{W}}_n) - r_n))\} / \mathbb{F}_n(r_n(\underline{\mathbb{W}}_n))$$

$$= \{\mathbb{X}_n^*(hr_n - r_n\frac{F_n(hr_n)}{F_n(r_n)}) \tag{b}$$

$$+ \mathbb{F}_n((h - r_n\frac{F_n(hr_n)}{F_n(r_n)})r_n^2(\underline{\mathbb{W}}_n^*)(\lambda \underline{\underline{\mathbb{W}}}_n^{*-1}\underline{\tilde{w}}(\underline{\mathbb{W}}_n^*))^T)$$

$$\times \sqrt{n}(\underline{\mathbb{W}}_n - \underline{W})\} / \mathbb{F}_n(r_n(\underline{\mathbb{W}}_n))$$

where $\underline{\mathbb{W}}_n^*$ is on the line segment between $\underline{\mathbb{W}}_n$ and $\underline{W}$ since $\underline{\nabla_u} r_n(\underline{u}) = r_n^2(\underline{u})\lambda \underline{\underline{u}}^{-1}\underline{\tilde{w}}(\underline{u})$

$$\rightarrow_p \{\mathbb{X}^*(hr - r\frac{F(hr)}{F(r)}) \tag{c}$$

$$+ F((h - r\frac{F(hr)}{F(r)})r^2(\lambda \underline{\underline{W}}^{-1}\underline{\tilde{w}})^T)\underline{\mathbb{Z}}_W\} / F(r)$$

by (4.6), (4.4), proposition 2.2, and theorem 4.1

$$= \{\mathbb{X}^*(hr - r\frac{F(hr)}{F(r)})$$

$$+ \{F(hr^2\underline{\tilde{w}}^T) - \frac{F(hr)}{F(r)}F(r^2\underline{\tilde{w}}^T)\}\lambda \underline{\underline{W}}^{-1}\underline{\mathbb{Z}}_W\} / F(r)$$

$$= \{\mathbb{X}^*(hr - r\frac{F(hr)}{F(r)})$$

$$+ \ \{F(hr^2\tilde{\underline{w}}^T) - \frac{F(hr)}{F(r)}F(r^2\tilde{\underline{w}}^T)\}M^- \mathbb{X}^*(r\tilde{\underline{w}})\} / F(r)$$

since $\mathbb{X}^*(r\tilde{\underline{w}}) = M\underline{\underline{\lambda W}}^{-1}\underline{\mathbb{Z}}_W$ implies

$\underline{\underline{\lambda W}}^{-1}\underline{\mathbb{Z}}_W = M^- \mathbb{X}^*(r\tilde{\underline{w}}) + \mathbb{Z}_\alpha \underline{\lambda}$, and

$\{F(hr^2\tilde{\underline{w}}^T) - \frac{F(hr)}{F(r)}F(r^2\tilde{\underline{w}}^T)\}\underline{\lambda} = F(hr) - \frac{F(hr)}{F(r)}F(r) = 0$

$$= \mathbb{X}^* \left[\frac{r}{F(r)}(h - G(h)) + r\tilde{\underline{w}}^T M^- \left\{G(hr\tilde{\underline{w}}) - G(h)G(r\tilde{\underline{w}})\right\}\right]$$

$$= \mathbb{X}^*(\tilde{V}^+(h - G(h))) \equiv \mathbb{Z}^*(h)$$

where $\tilde{V}\beta$ is defined in (4.15).

It remains only to establish that the convergence in (c) holds uniformly in $h \in \mathcal{H}$. By comparison of (b) and (c) and by theorem 4.1, it clearly suffices to show

$$\sup_{h \in \mathcal{H}} |\mathbb{X}_n^*(hr_n - r_n\frac{F_n(hr_n)}{F_n(r_n)}) - \mathbb{X}^*(hr - r\frac{F(hr)}{F(r)})| \to_p 0 \qquad \text{(d)}$$

and

$$\sup_{h \in \mathcal{H}} |\mathbb{F}_n((h - r_n\frac{F_n(hr_n)}{F_n(r_n)})r_n^2(\underline{\mathbb{W}_n^*})(\underline{\underline{\lambda \mathbb{W}_n^*}}^{-1}\underline{\tilde{w}}(\underline{\mathbb{W}_n^*}))^T) \qquad \text{(e)}$$
$$- F((h - r\frac{F(hr)}{F(r)})r^2(\underline{\underline{\lambda W}}^{-1}\underline{\tilde{w}})^T)| \to_p 0 .$$

Now, since $||\mathbb{W}_n^* - W|| \leq ||\mathbb{W}_n - W|| \to_p 0$ by proposition 2.2 and $F(h_e^2 r^2) < \infty$, (e) follows from a Glivenko - Cantelli theorem for $\mathbb{F}_n$ as in the proof of theorem 2.2.

To prove (d), note that the left side is bounded by

$$\sup_{h \in \mathcal{H}} |\mathbb{X}_n^*(hr_n - r_n\frac{F_n(hr_n)}{F_n(r_n)}) - \mathbb{X}^*(hr_n - r_n\frac{F_n(hr_n)}{F_n(r_n)})| \qquad \text{(f)}$$
$$+ \sup_{h \in \mathcal{H}} |\mathbb{X}^*(hr_n - r_n\frac{F_n(hr_n)}{F_n(r_n)}) - \mathbb{X}^*(hr - r\frac{F(hr)}{F(r)})|$$
$$\equiv I + II .$$

Then, since

$$II \leq \sup_{h \in \mathcal{H}} |\mathbb{X}^*(h(r_n - r))| + \sup_{h \in \mathcal{H}} |\frac{F_n(hr_n)}{F_n(r_n)}\mathbb{X}^*(r_n) - \frac{F(hr)}{F(r)}\mathbb{X}^*(r)|$$

where

$$||h(r_n - r)||_{L_2(F)} \leq ||\frac{r_n}{r} - 1||_\infty F(h_e^2 r^2)^{1/2} \to 0 \quad \text{uniformly in } h \in \mathcal{H} \qquad \text{(g)}$$

and

$$|F_n(hr_n) - F(hr)| \leq |F_n(hr_n - hr)| + |(hr)| \quad \text{(h)}$$
$$\leq \|\frac{r_n}{r} - 1\|_\infty F_n(h_e r) + \sum_{i=1}^{s} |\lambda_{ni} - \lambda_i| F_i(h_e r)$$
$$\to 0 \quad \text{uniformly in} \quad h \in \mathcal{H},$$

and both (g) and (h) hold with $h \equiv 1$, $II \to_p 0$ by uniform continuity of $\mathbb{X}^*$. To handle I, note that

$$I \leq \sup_{h \in \mathcal{H}} |\mathbb{X}_n^*(hr_n) - \mathbb{X}^*(hr_n)| + \sup_{h \in \mathcal{H}} \frac{F_n(hr_n)}{F_n(r_n)} |\mathbb{X}_n^*(r_n) - \mathbb{X}^*(r_n)|$$
$$\equiv A + B$$

where

$$A \leq \sup_{h \in \mathcal{H}} |\mathbb{X}_n^*(hr(\frac{r_n}{r} - 1))| + \sup_{h \in \mathcal{H}} |\mathbb{X}_n^*(hr) - \mathbb{X}^*(hr)|$$
$$+ \sup_{h \in \mathcal{H}} |\mathbb{X}^*(hr(1 - \frac{r_n}{r}))|$$
$$\leq \|\frac{r_n}{r} - 1\|_\infty \sup_{h \in \mathcal{H}} |\mathbb{X}_n^*(hr)| + \|\mathbb{X}_n^*(\cdot\, r) - \mathbb{X}^*(\cdot\, r)\|$$
$$+ \|\frac{r_n}{r} - 1\|_\infty \sup_{h \in \mathcal{H}} |\mathbb{X}^*(hr)|$$
$$\to_p 0 \qquad \text{by (4.6).} \quad \text{(i)}$$

Since $1 \in \mathcal{H}$, $B \to_p 0$ by (h) and (i). Thus $I \to_p 0$, and hence (d) holds, which completes the proof of (4.16).

Now we establish the covariance formula (4.17). First suppose that $G(h) = G(\tilde{h}) = 0$. Now from (4.8) and (4.15) it follows that

$$\text{Cov}(\mathbb{Z}^*(h), \mathbb{Z}^*(\tilde{h})) = G(r^{-1}\tilde{V}(h)\tilde{V}^+(\tilde{h})) - G(\underline{\tilde{w}}^T\tilde{V}(h))\underline{\underline{\lambda}}G(\underline{\tilde{w}}\tilde{V}(\tilde{h}))\,. \quad \text{(j)}$$

But

$$G(\underline{\tilde{w}}\tilde{V}(h)) = \langle\underline{\tilde{w}},rh\rangle_G / F(r) + \langle\underline{\tilde{w}},r\underline{\tilde{w}}^T\rangle_G M^- G(hr\underline{\tilde{w}})$$
$$= \left[I + \langle\underline{\tilde{w}},r\underline{\tilde{w}}^T\rangle_G M^-\right]G(hr\underline{\tilde{w}})\,. \quad \text{(k)}$$

Now if M^- is the Moore - Penrose generlized inverse of M, it follows from (2.14) of proposition 2.3 that $MM^- = I - \underline{\theta}\underline{\theta}^T$ where $\underline{\theta}$ is the (only) normalized eigenvector of M with eigenvalue 0. Since $M\underline{\lambda} = \underline{0}$ it follows that

$$MM^- = \left[\underline{\underline{\lambda}}^{-1} - \langle\underline{\tilde{w}},r\underline{\tilde{w}}\rangle_G\right]M^- = I - \underline{\lambda}\,\underline{\lambda}^T / \underline{\lambda}^T\underline{\lambda} \quad \text{(l)}$$

and hence

$$G(\underline{\tilde{w}}\tilde{V}(h)) = (\underline{\underline{\lambda}}^{-1}M^- + \underline{\lambda}\,\underline{\lambda}^T / \underline{\lambda}^T\underline{\lambda})G(hr\underline{\tilde{w}})$$
$$= \underline{\underline{\lambda}}^{-1}M^- G(hr\underline{\tilde{w}}) \quad \text{(m)}$$

since $\underline{\lambda}^T G(hr\underline{\tilde{w}}) = G(hrr^{-1}) = G(h) = 0$. Also,

$$G(r^{-1}\tilde{V}(h)\tilde{V}(\tilde{h})) = G((\frac{1}{F(r)}h + \tilde{\underline{W}}^T M^- G(r\tilde{\underline{w}}))\tilde{V}(\tilde{h})$$

$$= G(h\tilde{V}(\tilde{h})) + G(hr\tilde{\underline{w}}^T)M^- G(\underline{\tilde{w}}\tilde{V}^+(\tilde{h})). \quad \text{(n)}$$

Thus

$$\mathrm{Cov}\left[\mathbb{Z}^*(h), \mathbb{Z}^*(\tilde{h})\right] = G(h\tilde{V}(\tilde{h})) + G(r\tilde{\underline{w}}^T h)M^- G(\underline{\tilde{w}}\tilde{V}^+(\tilde{h}))$$

$$- G(hr\tilde{\underline{w}}^T)M^- \underline{\underline{\lambda}}^{-1}\underline{\underline{\lambda}} G(\underline{\tilde{w}}\tilde{V}(\tilde{h}))$$

$$= G(h\tilde{V}(\tilde{h})). \quad \text{(o)}$$

Now for arbitrary h, since $\mathbb{Z}^*(1) = 0$,

$$\mathbb{Z}^*(h) = \mathbb{Z}^*(h - G(h)) = \mathbb{X}^*(\tilde{V}(h - G(h)))$$

so from (o) it follows that

$$\mathrm{Cov}\left[\mathbb{Z}^*(h), \mathbb{Z}^*(\tilde{h})\right] = \langle h - G(h), \tilde{V}(\tilde{h} - G(\tilde{h}))\rangle_G ,$$

and hence (4.17) holds. □

PROOF OF (4.11) OF THEOREM 4.1: (IDENTIFICATION OF $\mathbb{Z}_\alpha$) From (1.20) we have $\underline{\mathbb{W}}_n = \underline{\mathbb{G}}_n(\underline{w})$, and since $\underline{W} = G(\underline{w})$ by definition, it follows that

$$\sqrt{n}(\underline{\mathbb{W}}_n - \underline{W}) = \mathbb{Z}_n(\underline{w})$$

where $\max_{1 \leq i \leq s} \|w_i r\|_\infty \leq \max_{1 \leq i \leq s}(W_i / \lambda_i) < \infty$. Thus, on the one hand by theorem 4.2

$$\sqrt{n}(\underline{\mathbb{W}}_n - \underline{W}) \to_p \mathbb{Z}^*(\underline{w})$$

$$= \{\mathbb{X}^*(r\underline{w} - r\frac{F(r\underline{w})}{F(r)}) + \left[F(r^2\underline{w}\,\tilde{\underline{w}}^T) - \frac{F(r\underline{w})}{F(r)}F(r^2\tilde{\underline{w}}^T)\right]M^- \mathbb{X}^*(r\underline{w})\} / F(r)$$

$$= \{\mathbb{X}^*(\frac{r}{F(r)}(\underline{w} - \underline{W} / F(r))) + \left[G((\underline{w} - \underline{W} / F(r))r\tilde{\underline{w}}^T)\right]M^- \mathbb{X}^*(r\tilde{\underline{w}})\} / F(r) \quad \text{(a)}$$

while on the other hand, by (4.10)

$$\sqrt{n}(\underline{\mathbb{W}}_n - \underline{W}) \to_p \underline{\underline{W\lambda}}^{-1}M^- X^*(r\tilde{\underline{w}}) + \mathbb{Z}_\alpha \underline{W}. \quad \text{(b)}$$

Thus the expressions on the right sides in (a) and (b) must be equal, and, upon multiplying across by $\underline{\underline{\lambda}}^T \underline{\underline{W}}^{-1}$ and using $\underline{\underline{\lambda}}^T \underline{\underline{W}}^{-1} \underline{W} = 1$, this yields

$$\{\mathbb{X}^*(r(r^{-1} - 1 / F(r))) + G(r(r^{-1} - 1 / F(r))\tilde{\underline{w}}^T)M^- \mathbb{X}^*(r\underline{w})\} / F(r)$$

$$= \underline{1}^T M^- X^*(r\tilde{\underline{w}}) + \mathbb{Z}_\alpha$$

or, since $\mathbb{X}^*(1) = 0$,

$$\{-\mathbb{X}^*(r / F(r)) + \left[\underline{1}^T - G(r\tilde{\underline{w}}^T)\right]M^- \mathbb{X}^*(r\tilde{\underline{w}})\} / F(r) = \underline{1}^T M^- \mathbb{X}^*(r\tilde{\underline{w}}) + \mathbb{Z}_\alpha$$

so that

$$\mathbb{Z}_\alpha = \left[-\mathbb{X}^*(r / F(r)) + \{\underline{1}^T - G(r\tilde{\underline{w}}^T)\}M^- \mathbb{X}^*(r\tilde{\underline{w}})\right] / F(r) - \underline{1}^T M^- \mathbb{X}^*(r\tilde{\underline{w}})$$

$$= \left[-\mathbb{X}^*(r / F(r)) + \{(1 - F(r))\underline{1}^T - G(r\tilde{\underline{w}}^T)\}M^- \mathbb{X}^*(r\tilde{\underline{w}})\right] / F(r)$$

which proves (4.11). When $F(r) = 1$ (4.11) reduces to (4.12). □

We close this section with some remarks on the estimation of the asymptotic variance of the process $\sqrt{n}(\mathbb{G}_n - G)$. From the Glivenko Cantelli theorems similar to those used in section 2 one can derive conditions for the sample analogue of (4.17) (with G and λ replaced throughout -- i.e. also in $\tilde{V}$, M, $\tilde{w}$ and r -- by $\mathbb{G}_n$ and λ_n) to converge in probability to (4.17) as $n \to \infty$. In fact this (co-) variance estimator is also obtained by formal likelihood calculations (inversion of the $(n-1)\times(n-1)$ matrix of second derivatives of the loglikelihood) continuing the derivation of $\mathbb{G}_n$ itself as a maximum likelihood estimator in the model where G is discrete, with mass at the actual observations only.

5. Examples and applications.

Example 5.1. (Length biased sampling, Vardi (1982)). Let $\mathbf{X} = R^+ = [0,\infty)$ and suppose $0 < \mu \equiv \int x\, dG(x) < \infty$. Let $w_1(x) \equiv 1$, so that the first sample is from G itself, and let $w_2(x) \equiv x$, so that the second sample is from the 'length - biased' distribution $F_2(x) = \mu^{-1}\int_0^x y dG(y)$. Then $\mathbf{X}^+ = \mathbf{X}$, $G^+ = G$, and (2.4) holds. Furthermore $\underline{W} = (1,\mu)^T$, and letting $\lambda_1 \equiv \lambda \in (0,1)$, $\lambda_2 = 1-\lambda \equiv \bar{\lambda}$,

$$r(x) = (\lambda + \bar{\lambda}\frac{x}{\mu})^{-1}.$$

Thus we have

$$M = \begin{bmatrix} 1/\lambda - F(r^2) & -F(r^2\tilde{w}_2) \\ -F(r^2\tilde{w}_2) & 1/\bar{\lambda} - F(r^2\tilde{w}_2^2) \end{bmatrix} = \begin{bmatrix} \bar{\lambda}K/\lambda & -K \\ -K & \lambda K/\bar{\lambda} \end{bmatrix},$$

since $M\underline{\lambda} = \underline{0}$, where

$$K \equiv F(r^2\tilde{w}_2) = G(r\tilde{w}_2) = \int_0^\infty \frac{x}{\mu}\frac{1}{\lambda + \bar{\lambda}x/\mu}dG(x). \tag{5.1}$$

Also define

$$K(x) \equiv \int_0^x \tilde{w}_2 r dG = G(1_{[0,x]}\underline{w}_2 r) \tag{5.2}$$

and note that

$$G(1_{[0,x]}r) = \frac{1}{\lambda}(G(x) - \bar{\lambda}K(x)). \tag{5.3}$$

Then by (4.17) with $K(s,t) \equiv K(1_{[0,s]}, 1_{[0,t]})$,

$$K(s,t) = G(r(1_{[0,s]} - G(s))(1_{[0,t]} - G(t)) + G(r(1_{[0,s]} - G(s)))\frac{\lambda}{\bar{\lambda}K}G(r(1_{[0,t]} - G(t)))$$

by using $M^- = \begin{bmatrix} \lambda/\bar{\lambda}K & 0 \\ 0 & 0 \end{bmatrix}$ as the $\{1,2\}$ - inverse of M; recall proposition 2.3.

$$= \frac{1}{\lambda}\left\{G(s\wedge t) - G(s)G(t)\right\} - \frac{\bar{\lambda}}{\lambda}K\left\{\frac{K(s\wedge t)}{K} - \frac{K(s)}{K}\frac{K(t)}{K}\right\}, \tag{5.4}$$

in agreement with equation (3.6) of Vardi (1982).

Moreover $G(r\tilde{w}) = (G(r),K)^T = (1 - \bar{\lambda}/\lambda K, K)^T$, and hence, using $K^- = \underline{\underline{W}}\lambda^{-1}M^-$, and the same version of M^-,

$$\left\{ K^- \underline{\tilde{w}} - (1 + \underline{\tilde{w}}^T M^- G(r\underline{\tilde{w}}))\underline{W} \right\} r$$

$$= \begin{bmatrix} 0 \\ \frac{x}{\lambda K} \end{bmatrix} r - \frac{1}{\lambda} r^{-1} r \begin{bmatrix} 1 \\ \mu \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{x}{\lambda K} \end{bmatrix} r - \frac{1}{\lambda} \begin{bmatrix} 1 \\ \mu \end{bmatrix}$$

and hence the limit rv $\underline{\mathbb{Z}}_W$ of theorem 4.1 is

$$\mathbb{X}^* \begin{bmatrix} 0 \\ \frac{x}{\lambda K} r \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbb{X}^*(\frac{x}{\lambda K} r) \end{bmatrix}$$

where, by (4.8),

$$\mathbb{X}^*(\frac{x}{\lambda K} r) \cong N(0, \frac{1}{\lambda^2 K^2} \left\{ G(x^2 r) - G(xr(1, \frac{x}{\mu})) \underline{\lambda} G(xr(1, \frac{x}{\mu})^T) \right\})$$

$$= N(0, \frac{\mu^2(1 - \bar{K})}{\lambda \bar{\lambda} K}) \tag{5.5}$$

by straightforward calculation, which agrees with (3.3) of Vardi (1982).

EXAMPLE 5.2. ($s = 1$; biasing with only one stratum). For a general sample space $\mathbf{X}$ suppose that $s = 1$ and that $w_1 = w$ satisfies both $G(w) < \infty$ and $W_{-1} \equiv G(w^{-1}) < \infty$. Then condition (2.4) is trivially satisfied, and, letting $W \equiv G(w)$, $r = W/w = \tilde{w}^{-1}$, $F(r) = G(\mathbf{X}^+)$ may be $= 1$ or < 1 depending on w and G. Since $M = 0$ and $r\tilde{w} = 1$, theorem 4.1 yields

$$\sqrt{n}(\mathbb{W}_n - W) \to_d -\mathbb{X}^*(r/F^2(r))W \cong N(0, \frac{W^2}{F(r)^4}(W_{-1}W - 1)). \tag{5.6}$$

Furthermore, the process $\mathbb{Z}^*$ of theorem 4.2 reduces to

$$\mathbb{Z}^*(h) = \mathbb{X}^*(r(h - G^+(h))/F(r)) \tag{5.7}$$

$$= \mathbb{X}^*(r(h - G^+(h))) \qquad \text{when } F(r) = 1$$

with covariance function

$$K(h,\tilde{h}) = G^+(r(h - G^+(h))(\tilde{h} - G^+(\tilde{h})))/F(r) \tag{5.8}$$

$$= G(r(h - G(h))(\tilde{h} - G(\tilde{h}))) \qquad \text{when } F(r) = 1$$

$$= W_{-1}W \left\{ \left\{ \frac{G(\frac{h\tilde{h}}{w})}{W_{-1}} - G(h)\frac{G(\frac{\tilde{h}}{w})}{W_{-1}} \right\} + \left\{ G(h\tilde{h}) - \frac{G(\frac{h}{w})}{W_{-1}} G(\tilde{h}) \right\} \right\}.$$

These formulas agree with the results of Vardi (1985a) section 7(ii) in the case $F(r) = 1$. The further special case $w(x) = x$ (which is also the special case $\lambda = 0$ in example 1) was considered by Cox (1969).

EXAMPLE 5.3. (Truncated sampling or restricted measurement). This is a further special case of example 2. For a general sample space $\mathbf{X}$ suppose that $s = 1$ and that $w_1(x) = 1_C(x)$ where

$C \in \mathbf{B}$, $C \neq \mathbf{X}$, and $G(C) < 1$. Then $\mathbf{X}^+ = C$, $W_1 = G(C) < 1$, $r(x) = G(C)1_C(x)$, and $G^+ = G / G(C)$ is simply the conditional distribution $G(\cdot|C)$. Thus

$$G^+(A) = \frac{G(A \cap C)}{G(C)} = F(A) \qquad \text{for } A \in \mathbf{B} \cap C ,$$

and the estimator $\mathbb{G}_n^+$ of G^+ is simply $\mathbb{F}_n$. Note that $W_1 \equiv G(C)$ is not identifiable in this situation.

EXAMPLE 5.4. ('Choice - based' sampling in econometrics; 'case - control studies' in biostatistics). Suppose that $\mathbf{X} = (Y,Z)$ where Y takes values in $\{1, \cdots, M\}$ and $Z \cong H$ with density h with respect to μ is a covariate vector with values in $\mathbf{Z} \subset$ some R^p. The basic (unbiased or prospective) model G has density

$$g(y,z) = p_\theta(y|z)h(z) \tag{5.9}$$

so that

$$G(\{y\} \times A) = \int_A p_\theta(y|z)dH(z) \tag{5.10}$$

for $y = 1, \cdots, M$ and $A \in \mathbf{B}(R^p)$ where $p_\theta(y\,|z) \equiv P_\theta(Y = y|Z = z)$ is a parametric (finite - dimensional) model. A frequent choice is the logistic regression model

$$p_\theta(y|z) = \frac{exp(\alpha_y + \beta_y^T z)}{\sum_{y'=1}^{M} exp(\alpha_{y'} + \beta_{y'}^T z)} \tag{5.11}$$

with $\theta = (\alpha,\beta) \in R^{(p+1)M}$.

The biased (or retrospective) sampling model F is obtained from G via the weight functions $w_i(x) = w_i(y) = 1_{D_i}(y)$ where $D_i \subset \{1, \cdots, M\}$ for $i = 1, \cdots, s$. This again yields a semiparametric submodel, since only distributions G of the form (5.10) are considered.

One case of particular interest is that of 'pure choice - based sampling' in the terminology of Cosslett (1981). In this sampling scheme, the strata D_i are taken to be just $D_i = \{i\}$, $i = 1, \cdots, s \equiv M$. In this case the graph M^* of section 2 is *not* connected: in fact $G(w_i w_j) = 0$ for all $i \neq j$ and hence there is no unique nonparametric MLE $\mathbb{G}_n$ of G for this sampling scheme. Manski and Lerman (1977) avoid this difficulty of pure choice - based sampling by assuming that the 'aggregate shares' $G(y) \equiv G(\{y\} \times \mathbf{Z}) = \int p_\theta(y\,|z)dH(z)$, $y = 1, \cdots, M$ are known. Note that for this biasing system we can view F as a biased distribution derived from H with new biasing (weight) functions $w_y^*(z;\theta) \equiv p_\theta(y\,|z)$, $y = 1, \cdots, M$, depending on the unknown parameter θ. Then $W_y^* = G(y)$, typically the condition on M^* for these w^*'s will hold, and if θ is known the methods of the preceding sections yield estimates of H together with the asymptotic behavior of the estimates.

This same pure choice - based sampling design is also frequently used in 'case - control studies' in biostatistics where the y 's often denote different disease categories. In the biostatistics applications interest centers on odds ratios which can be estimated from purely choice - based sampling in spite of the fact that G itself cannot be estimated; see e.g. Prentice and Pyke (1979), who examine the case of (5.11), and Breslow and Day (1980). If the 'pure choice - based' design is 'enriched' by taking $s = M+1$, $\lambda_M + 1 = 0$, and choosing $w_{M+1}(x) = 1_{\{1, \cdots, M\}}(y)$, then (2.4) holds and the nonparametric MLE $\mathbb{G}_n$ of G exists (a.s. for $n \geq$ some N_ω) and is unique. See example 5.

For general D_i's the biased distribution F has density

$$f(y,z,i) = \lambda_i \frac{1_{D_i}(y)p_\theta(y\,|z)h(z)}{\int \sum_{y'=1}^{M} 1_{D_i}(y')p_\theta(y'\,|z')h(z')d\mu(z')} , \tag{5.12}$$

and the condition (2.4) for existence of a unique solution is precisely Cosslett's (1981) Assumption 10:

$$\left\{\bigcup_{i \in B} D_i\right\} \cap \left\{\bigcup_{i \in B^c} D_i\right\} \neq \emptyset \qquad \text{for every proper subset } B \text{ of } \{1,\ldots,s\}\ ; \tag{5.13}$$

see Vardi (1985a) sections 2 and 8.

For known θ, efficient estimates of H and their asymptotic behavior via the preceding sections can be obtained as follows: The marginal distribution of (Z,I) is

$$f(z,i) = \lambda_i \frac{\left\{\sum_{y'=1}^{M} 1_{D_i}(y') p_\theta(y' \mid z)\right\} h(z)}{\int \left\{\sum_{y'=1}^{M} 1_{D_i}(y') p_\theta(y' \mid z')\right\} h(z') d\mu(z')}, \tag{5.14}$$

$$\equiv \lambda_i \frac{w_i^*(z) h(z)}{\int w_i^*(z') dH(z')}$$

where the new biasing functions $w_i^*(z) = w_i^*(z;\theta)$ depend on θ. Thus if θ is known, the methods of Vardi (1985a) and the preceding sections apply to yield efficient estimates of H, which can, in turn, be used to construct efficient estimates of θ. This method is implicit in Cosslett (1981) section 4, and will be discussed in more detail in Bickel, Klaassen, Ritov, and Wellner (1986).

EXAMPLE 5.5. ('Enriched' stratified sampling). Let $\mathbf{X}$ be a general sample sample space and suppose that $D_1, \cdots, D_s$ form a (measurable) partition of $\mathbf{X}$: $D_i \cap D_j = \emptyset$ for $i \neq j$ and $\bigcup_{i=1}^{s-1} D_i = \mathbf{X}$. Let $w_i(x) = 1_{D_i}(x)$ for $i = 1, \cdots, s-1$ and suppose that $w_s(x) = 1$. Thus the stratified sample from $D_1, \cdots, D_{s-1}$ is 'enriched' by sampling from all of $\mathbf{X}$ with sampling fraction $\lambda_s > 0$; this terminology is that of Cosslett (1981).

For this sampling scheme (2.4) holds (assuming without loss that $G(D_i) > 0$, $i = 1, \cdots, s-1$), $F(r) = 1$, and we have

$$W_i = G(D_i), \qquad i = 1, \cdots, s-1$$

$$r(x) = \sum_{i=1}^{s-1} \frac{1}{\lambda_s + \lambda_i / G(D_i)} 1_{D_i}(x)$$

so that the upper left $s-1 \times s-1$ submatrix of $F(r^2 \underline{\tilde{w}}\, \underline{\tilde{w}}^T)$ is diagonal with elements

$$F(r\tilde{w}_i^2) = \frac{1}{\lambda_s G(D_i) + \lambda_i}.$$

Hence the upper left $s-1 \times s-1$ submatrix of M is diagonal with elements

$$M_{ii} = \frac{1}{\lambda_i} - \frac{1}{\lambda_s G(D_i) + \lambda_i} = \frac{\lambda_s G(D_i)}{\lambda_i(\lambda_s G(D_i) + \lambda_i)},$$

for $i = 1, \cdots, s-1$, and a $\{1,2\}$ - inverse M^- of M is given by the diagonal matrix with last row and column containing all zeros and having diagonal elements M_{ii}^{-1}; recall proposition 2.3 and (2.15). Thus, $K^- = \underline{\underline{W}}\,\underline{\underline{\lambda}}^{-1} M^-$ is also diagonal with last row and column all zero and diagonal entries

$$K_{ii}^- = \frac{1}{\lambda_s}\left\{\lambda_s G(D_i) + \lambda_i\right\} \equiv \frac{1}{\lambda_s} a_i \tag{5.15}$$

for $i = 1, \cdots, s-1$. Similarly,

$$G(r\tilde{w}_i) = G(D_i)/a_i\,,$$

and hence $M^- G(r\underline{\tilde{w}}) = \left[\lambda_1/\lambda_s, \cdots, \lambda_{s-1}/\lambda_s, 0\right]$. Therefore

$$(1 + \underline{\tilde{w}}^T M^- G(r\underline{\tilde{w}}))\underline{W}r = (1 + \sum_{j=1}^{s-1} \frac{1_{D_i}}{G(D_i)}\frac{\lambda_i}{\lambda_s}) r\underline{W}$$

$$= \frac{1}{\lambda_s} r^{-1} r\underline{W} = \frac{1}{\lambda_s}\underline{W}$$

which is constant in x, and hence the limiting random vector in theorem 4.1 becomes $K^- \mathbb{X}^*(r\underline{\tilde{w}})$, the last element of which is 0 by the form of K^-, and where the first $s-1$ elements of $\mathbb{X}^*(r\underline{\tilde{w}})$ have an $(s-1)\times(s-1)$ covariance matrix

$$C = \underline{\underline{a}} - (\underline{\underline{a}}, \underline{\underline{a}}G(\underline{D}))\underline{\lambda}(\underline{\underline{a}}, \underline{\underline{a}}G(\underline{D}))^T \tag{5.16}$$

$$= \underline{\underline{a}}\left[\underline{\underline{a}}^{-1} - (\underline{I},\underline{p})\underline{\lambda}(\underline{I},\underline{p})^T\right]\underline{\underline{a}}\,,$$

where $p \equiv G(D) \equiv (G(D_1), \cdots, G(D_s))^T$. Thus by theorem 4.1, and straightforward calculation using (5.15) and (5.16), it follows that

$$\sqrt{n}(\mathbb{G}_n(\underline{D}) - G(\underline{D})) \to_d N_{s-1}(0, \frac{1}{\lambda_s}(G(\underline{\underline{D}}) - G(\underline{D})G(\underline{D})^T))\,.$$

Note that this is just the covariance matrix for the usual (multinomial) estimate of $G(\underline{D})$ from a random sample of size $n\lambda_s$ from all of $\mathbb{X}$. In other words, sampling *within* the strata D_i does not help in estimating the strata probabilities $G(\underline{D})$.

EXAMPLE 5.6. (Stratified or truncated regression). This interesting and rich family of semiparametric submodels of the general biased sampling model begin with ordinary linear regression with unknown error distribution G_0 as the basic (unbiased) model: Suppose that $X = (Y,Z) \cong G$ where $Y = \theta^T Z + \epsilon$ with $\epsilon \cong G_0$ with density g_0 with respect to Lebesgue measure and $Z \cong H$ independent of ϵ with density h with respect to μ. Thus G has density

$$g(y,z) = g_0(y - \theta^T z)h(z)\,.$$

The biased sampling model is typically determined by weight functions $w_i(x) \equiv w_i(y) \equiv 1_{D_i}(y)$, $i = 1, \cdots, s$, where the D_i's are disjoint subintervals of R^1. The case of $s = 1$ and $D_1 = (-\infty, y_0]$, which is also a special case of example 3, has been considered by Bhattacharya, Chernoff, and Yang (1983). Jewell (1985) considers the case $s = 2$ and $D_1 = (-\infty, y_0]$, $D_2 = (y_0, \infty)$ in which condition (2.4) fails, so a unique completely nonparametric estimator of G does not exist in view of Vardi's theorem 1.1. Nevertheless the parameters θ, G_0, and H are identifiable in this model, and for known θ the methods of Vardi (1985a) can be applied iteratively by first regarding G_0 as known and absorbing it into the biasing functions and estimating H, and then by treating H as known and absorbing it into the biasing functions and estimating G_0, and so forth. This type of semiparametric submodel of the biased sampling model will be treated by Bickel, Klaassen, Ritov, and Wellner (1986).

REFERENCES

1. T. M. APOSTOL (1957). *Mathematical Analysis,* Addison-Wesley, Reading.
2. J. M. BEGUN, W. J. HALL, W. HUANG, J. A. WELLNER (1983). Information and asymptotic efficiency in parametric - nonparametric models, *Ann. Statist.* **11**, 432 - 452.
3. A. BERMAN, ROBERT J. PLEMMONS (1979). *Nonnegative matrices in the Mathematical Sciences,* Academic Press, New York.
4. P. K. BHATTACHARYA, H. CHERNOFF, S. S. YANG (1983). Nonparametric estimation of the slope of a truncated regression, *Ann. Statist.* **11**, 505 - 514.
5. P. J. BICKEL, C. A. J. KLAASSEN, Y. RITOV, J. A. WELLNER (1986). *Efficient and Adaptive Inference in Semiparametric Models,* Forthcoming monograph, Johns Hopkins University Press, Baltimore.
6. N. E. BRESLOW, N. E. DAY (1980). *The Analysis of Case-Control Studies,* International Agency for Research on Cancer, Lyon.
7. STEPHEN R. COSSLETT (1981). Maximum likelihood estimator for choice - based samples, *Econometrica* **49**, 1289 - 1316.
8. D. R. COX (1969). *Some sampling problems in technology,* (N. L. Johnson and H. Smith Jr., eds.) 506 - 527, Wiley - Interscience, New York.
9. R. M. DUDLEY, W. PHILIPP (1983). Invariance principles for sums of Banach space valued random elements and empirical processes, *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **62**, 509 - 552.
10. P. GAENSSLER (1983). *Empirical Processes,* IMS Lecture Notes - Monograph Series **3**, IMS, Hayward.
11. N. P. JEWELL (1985). Least squares regression with data arising from stratified samples of the dependent variable, *Biometrika* **72**, 11 - 21.
12. C. F. MANSKI, S. R. LERMAN (1977). The estimation of choice probabilities from choice-based samples, *Econometrica* **45**, 1977 - 1988.
13. C. F. MANSKI, D. MCFADDEN (EDS) (1981). *Structural Analysis of Discrete Data with Applications,* MIT Press, Cambridge, Mass.
14. P. W. MILLAR (1985). Nonparametric applications of an infinite dimensional convolution theorem, *Z. Wahrsch. verw. Gebiete* **68**, 545 - 556.
15. J. M. ORTEGA, W. C. RHEINBOLDT (1970). *Iterative Solution of Nonlinear Equations in Several Variables,* Academic Press, New York.
16. M. OSSIANDER (1985). *A central limit theorem under metric entropy with L 2 bracketing,* Preprint, University of Washington.
17. D. POLLARD (1982). A central limit theorem for empirical processes, *J. Austral. Math. Soc. (Series A)* **33**, 235 - 248.
18. DAVID POLLARD (1984). *Convergence of Stochastic Processes,* Springer - Verlag, New York.
19. R. L. PRENTICE, R. PYKE (1979). Logistic disease incidence models and case-control studies, *Biometrika* **66**, 403 - 411.
20. G. A. F. SEBER (1977). *Linear Regression Analysis,* Wiley,New York.
21. Y. VARDI (1983). Nonparametric estimation in the presence of length bias, *Ann. Statist.* **10**, 616 - 620.
22. Y. VARDI (1985a). Empirical distributions in selection bias models, *Ann. Statist.* **13**, 178 - 203.
23. Y. VARDI (1985b). *Asymptotics for empirical distributions in selection bias models,* Preprint, AT&T Bell Laboratories, Murray Hill, New Jersey.