Statistics & Probability Letters 8 (1989) 411–418
North-Holland

October 1989

IMPUTATION OF MISSING VALUES USING DENSITY ESTIMATION

D.M. TITTERINGTON
University of Glasgow, Glasgow, UK

J. SEDRANSK
University of Iowa, Iowa City, IA 52242, USA

Received June 1988
Revised November 1988

Abstract: The paper proposes the use of density estimation, based on respondents' data, for imputing values for nonrespondents. Multiple imputations are catered for and the procedure is assessed on the basis of estimation of the underlying density itself. Extension to Rubin's version of multiple imputations is described.

Keywords: kernel method, multiple imputations, smoothed bootstrap.

1. Introduction

A common feature of the initial treatment of survey data from very large sample surveys is the imputation of values for the missing data. One of the principal reasons for imputation is to provide a complete data set for the purpose of future, secondary data analyses. Since survey data are usually analyzed by people who are unfamiliar with the missing data process, replacing missing with imputed values will often be beneficial. Whereas the secondary data analyst cannot make an informed decision about how to treat the missing data, the organization that collected the data frequently has knowledge about the missing data process. They can use this knowledge to provide reasonable estimates of the missing values. Kalton and Kasprzyk (1982, Section 1) provide an excellent summary of both the advantages and disadvantages accruing from the use of imputed values. Also, see Little (1988, Section 2).

It is, of course, important that imputations be made to safeguard secondary analysts from gross errors of inference. The most obvious illustration of how things can go wrong is to consider a survey in which a single variable, Y, is of interest. Suppose a random sample of size N is chosen and that n members of the sample respond, providing data $D_0 = (y_1, \ldots, y_n)$. An imputation procedure would then "fill in" values $\{\hat{y}_j\colon\ j = n+1, \ldots, N\}$, say, for the nonrespondents' scores. It is likely that most secondary data analysts will make inferences as if the completed data set were equivalent to N bona fide observations. Thus, it is important to employ imputation procedures that will minimize errors of inference.

To illustrate potential problems associated with alternative imputation methods, consider *mean imputation* and *random imputation*.

(M) *Mean imputation*: For $j = n+1, \ldots, N$, take $\hat{y}_j = \bar{y} = n^{-1}\sum_{i=1}^{n} y_i$.

(R) *Random imputation*: For $j = n+1, \ldots, N$, independently, take $\hat{y}_j = y_i$ with probability n^{-1}, $i = 1, \ldots, n$.

If point estimation of the population mean is of interest, then it is natural to use the sample mean of the completed data as the estimator. Under certain natural conditions, (M) or (R) would both give unbiased estimators. If, however, estimation of the population variance by the sample variance of the completed

data is of interest, then the concentration of the imputed values at $\bar{y}$ leads to gross bias with method (M). Method (R) is much more reliable, as detailed by Titterington and Sedransk (1984). (Other discussions of the problems associated with mean imputation appear in Kim and Curry, 1977; Schieber, 1978; Sande, 1982; and Kalton and Kasprzyk, 1982.)

This superiority of method (R) stimulates a modification, (KR), based on use of a density estimate of the underlying density, f, of Y. Whereas (R) is an *unsmoothed*, nonparametric procedure, (KR), described in Section 2, is the natural smoothed analogue.

2. Imputation using density estimation

The method, to be denoted by (KR), has the following concise description.

(KR) For each $j = n+1, \ldots, N$, independently, impute y_j by $\hat{y}_j = y_i + hz_j$; with probability n^{-1}, $i = 1, \ldots, n$, where the $\{z_j\}$ are independently and identically distributed with density function $K(\cdot)$.

We shall assume that $\int uK(u)\,du = 0$ and $\int u^2K(u)\,du < \infty$. The method has more than one interpretation. Adopting a "superpopulation" approach to finite population inference, one regards the values of Y for the *entire finite population* as a sample from an underlying superpopulation density, f. In a parametric superpopulation approach with *no missing data*, one uses the sample data to estimate the parameters indexing f. Finally, to estimate a quantity such as the finite population total one adds to the sum of the sample values, predictions (using the estimated f) for the *nonsampled values*. (See Royall, 1983, for references.) Method (KR) uses a nonparametric version of this approach to predict the *missing values*, $\{y_j: j = n+1, \ldots, N\}$. That is, the $\{\hat{y}_j\}$ are a random sample from the kernel-based estimate of the underlying superpopulation density, f, of Y, obtained from the respondents' data; i.e.,

$$\hat{f}_0(y) = (nh)^{-1} \sum_{j=1}^{n} K\{(y - y_i)/h\}.$$

An alternative interpretation is to regard the imputations from method (R) as a bootstrap sample of size $N - n$ from the respondents' data, and method (KR) as a corresponding smoothed bootstrap sample; see Efron (1979), Silverman (1986, Section 6.4).

In the technical part of the paper we shall investigate the value of this imputation method in particular senses. Before that, we generalize the method to allow multiple imputations for each missing value. This technique (Rubin, 1987, p. 16) has proved to be very useful in providing more appropriate representations of uncertainty when inferences are based upon a "data set" consisting of D_0 *and* imputed values $\{\hat{y}_j: j = n+1, \ldots, N\}$. Rubin also discusses other advantages of using multiple rather than single imputation. For a more extensive discussion of multiple imputations, see Rubin (1978, 1987).

To be specific, we allow r independent imputations per missing value and denote them by $\hat{y}_{jk}$, $j = n+1, \ldots, N$, $k = 1, \ldots, r$, where, for each j and k,

$$\hat{y}_{jk} = y_i + hz_{jk}, \text{ w.p. } 1/n, \quad i = 1, \ldots, n,$$

where the z_{jk} are as before. From now on we refer to this generalization as method (KR).

All values, "real" and imputed, can then be used to define a final density estimate of the underlying superpopulation density of Y, in the form

$$\hat{f}(y) = (Nh)^{-1}\left[\sum_{i=1}^{n} K\{(y - y_i)/h\} + r^{-1} \sum_{j=n+1}^{N} \sum_{i=1}^{r} K\{(y - \hat{y}_{jk})/h\}\right]. \tag{1}$$

There is, of course, a generalization of method (R) to the multiple-imputation case. Here, for $j = n+1, \ldots, N$ and $k = 1, \ldots, r$, independent imputations $\hat{y}_{jk}$ are generated from the discrete uniform

distribution on $\{y_1, \ldots, y_n\}$; substitution of these $\hat{y}_{jk}$ instead of the y_{jk} in (1) leads to a density estimate, $\hat{\hat{f}}(\cdot)$, based on method (R).

Note that, averaging $\hat{f}(y)$ over the $\{z_{jk}\}$, we obtain

$$\tilde{f}(y) = (Nh)^{-1}\left[\sum_{i=1}^{n} K\{(y-y_i)/h\} + (N-n)\int K\{(y-x)/h\}\hat{f}_0(x)\,dx\right]. \tag{2}$$

Apart from the use of $\hat{f}_0(\cdot)$ instead of the true $f(\cdot)$ in the right side of (2), this corresponds to the "ideal" imputation of the density of the missing value in each case. If the convolution integration can be carried out, (2) can be evaluated explicitly (e.g., if $K(\cdot)$ is chosen to be Normal).

3. Effectiveness of method (KR)

We now assess the use of $\hat{f}(\cdot)$, $\hat{\hat{f}}(\cdot)$ and $\tilde{f}(\cdot)$ as estimators of $f(\cdot)$. We assume that the sampling mechanisms and missing-data mechanisms are ignorable (Rubin, 1976; Little, 1982, 1983).

We can obtain, for a given imputation method, a measure of closeness of the resulting density estimate to the true density by calculating the integrated mean square error

$$\int E\{\hat{f}(y) - f(y)\}^2\,dy = \delta(\hat{f}, f),$$

say.

If we assume that $(N-n) \sim \mathrm{Bi}(N, \theta)$, then

$$\delta(\hat{f}, f) = \tfrac{1}{4}h^4H^2 + \hat{G}N^{-1}h^{-1} + o(h^4 + N^{-1}h^{-1}) \tag{3}$$

and

$$\delta(\tilde{f}, f) = \tfrac{1}{4}h^4H^2 + \tilde{G}N^{-1}h^{-1} + o(h^4 + N^{-1}h^{-1}) \tag{4}$$

where $\hat{G} > \tilde{G}$. Furthermore,

$$\delta(\hat{\hat{f}}, f) = \tfrac{1}{4}h^4\hat{\hat{H}}^2 + \hat{\hat{G}}N^{-1}h^{-1} + o(h^4 + N^{-1}h^{-1}). \tag{5}$$

These results are special cases of those of Titterington (1983). The main steps and the values of $\hat{G}$, $\tilde{G}$, $\hat{\hat{G}}$, H and $\hat{\hat{H}}$ are given in Appendix 1. The optimal choice of the smoothing parameter h is obtained by minimizing the dominant terms in (3), (4) or (5). For instance,

$$(\hat{h})^5 = \hat{G}/(NH^2),$$

giving, for the dominant term,

$$\delta(\hat{f}, f) \alpha (\hat{G}^2 H N^{-2})^{2/5}.$$

To indicate how well the various methods perform, values of the dominant terms in $\delta(f^*, f)$ were calculated for various θ, and for

(i) $f^* = \hat{f}$ for various r, and optimal $\hat{h}$;
(ii) $f^* = \tilde{f}$ for optimal $\tilde{h}$;
(iii) $f^* = \hat{f}_0$ for optimal $\hat{h}$;
(iv) $f^* = \hat{\hat{f}}$ for various r and optimal $\hat{\hat{h}}$;
(v) $f^* = \hat{f}$ with one imputation per missing value ($r = 1$) and h optimal for a complete data set of size N; that is (Scott and Factor, 1981), $h^5 = I_2/(NH^2)$ where $I_2 = \int K^2(u)\,du$.

Table 1

Method	r	θ				
		0.1	0.3	0.5	0.7	0.9
(i)	1	1.03	1.09	1.10	1.09	1.04
	2	1.00	0.99	0.99	0.98	0.99
	5	0.98	0.93	0.91	0.92	0.96
	10	0.97	0.92	0.89	0.90	0.95
(ii)	∞	0.96	0.90	0.86	0.87	0.94
(iv)	1	1.07	1.16	1.20	1.16	1.07
	2	1.04	1.08	1.10	1.08	1.04
	5	1.01	1.03	1.04	1.03	1.01
	10	1.01	1.02	1.02	1.02	1.01
(v)		1.04	1.09	1.11	1.09	1.08

Table 2

Density estimate	Associated moments conditioned on D_0	
	Mean	Variance
$\hat{f}_0$	$\bar{y}_n$	$s_y^2 + h^2$
$\tilde{f}$	$\bar{y}_n$	$s_y^2 + h^2 + h^2\{1-(n/N)\}$
$\hat{f}$	$\bar{y}_n$	$s_y^2 + h^2 + h^2\{1-(n/N)\}$
$\hat{f}_m$	$\bar{y}_n$	$(n/N)s_y^2 + h^2$

A normal kernel was used. Table 1 gives the values for $\delta(f^*, f)$ with, for each θ, $\delta(\hat{f}_0, f)$ normalized to 1. We would hope, therefore, that all values would be close to 1. The main features are that r need not be large to do well, and that method (v) is also satisfactory. This latter finding is particularly helpful because it means that, if the density estimate is being constructed as part of a secondary analysis, the whole data set can indeed be treated "as if" it were a sample of size N, without generating misleading results.

Although nonnormal data would be expected, and although, in practice, a data based choice of h will be necessary (h depends on the true $f(\cdot)$; see Appendix 1 and Scott and Factor, 1981), it is anticipated that the features mentioned above will persist in more general cases. Note that the relative values of $\delta(f^*, f)$, as f^* varies, are independent of f.

An additional comparison is provided by Table 2, where the mean and variance of the random variables associated with the various density estimates are stated, without proof. The moments are those conditioned on $\{y_1, \ldots, y_n\}$, so that, in some cases, some averaging has been undertaken. The densities considered in Table 2 are $\hat{f}_0$, $\hat{f}$, $\tilde{f}$ and $\hat{f}_m$, the latter obtained by the method of mean imputation (M). The unsatisfactory variance associated with $\hat{f}_m$ is clearly seen.

4. Extension to Rubin's multiple imputations

In this section we briefly indicate how some of the calculations of Section 3 can be extended to deal with a variation of the multiple imputations procedure. The modified form of (KR) is as follows.

(KR′) For each $k = 1, \ldots, r$, independently, create, from the observed data D_0, a sample D_k of values $\{y_{ik}: i = 1, \ldots, n\}$. Then, for each $j = n+1, \ldots, N$, independently,

$$\hat{y}_{jk} = y_{ik} + hz_{jk}, \text{ w.p. } n^{-1}, \quad i = 1, \ldots, n,$$

where the z_{jk} are as in Section 2.

The idea of creating a different set of $\{y_{ik}\}$ for each of the r sets of imputations is to reflect the uncertainty arising from the fact that $y_1, \ldots, y_n$ are merely a *sample* from f. This has important consequences for the calculation of reliable confidence intervals for population totals (Rubin and Schenker, 1986). In this paper we merely indicate the effect this modification has on the estimation of f, in terms of the dominant terms in the integrated mean square error.

The two density estimates obtained from the procedure are

$$\tilde{f}(y) = (Nh)^{-1}\left[\sum_{i=1}^{n} K\{(y-y_i)/h\} + r^{-1}\sum_{j=n+1}^{N}\sum_{k=1}^{r} K\{(y-\hat{y}_{jk})/h\}\right]$$

and

$$f^*(y) = (Nh)^{-1}\left[\sum_{i=1}^{n} K\{(y-y_i)/h\} + r^{-1}(N-n)\sum_{k=1}^{r}\int K\{(y-x)/h\}\hat{f}_k(x)\,dx\right],$$

where

$$\hat{f}_k(x) = (nh)^{-1}\sum_{i=1}^{n} K\{(x-y_{ik})/h\}.$$

Note that the formula for $\tilde{f}$ is identical to that of $\hat{f}$, although its statistical properties are different. There is more than one way to generate D_k: the n values might be drawn as a random sample from the $y_1, \ldots, y_n$ with replacement or as a random sample from $\hat{f}_0$, the latter corresponding to a smoothed bootstrap sample. Appendix 2 concentrates on the former procedure, showing that the biases of $\tilde{f}(x)$ and $f^*(x)$ are the same as the common bias of $\hat{f}(x)$ and $\tilde{f}(x)$. The difference comes in the variance term, reflecting the extra source of variation imposed by the method. To be specific, the integrated mean squared errors are related by

$$\delta(\tilde{f}, f) - \delta(\hat{f}, f) = \delta(f^*, f) - \delta(\tilde{f}, f) = (Nh)^{-1}\theta^2 I_4(1-\theta)^{-1}r^{-1} + o(N^{-1}h^{-1}), \qquad (6)$$

where I_4 is defined at the beginning of Appendix 1. Comparison of (6) and (3) indicates that $\hat{f}$ is preferable to $\tilde{f}$ on this basis, although the latter has important positive attributes as discussed by Rubin and Schenker (1986).

5. Discussion

The univariate case is clearly a very simple situation. In the multivariate version it is natural to use data available on covariates: Titterington and Mill (1983) use techniques similar to those of Section 3 to look at the bivariate case, whereas Sedransk and Titterington (1980) investigate analogues of the (M) and (R) procedures for bivariate normal data with missing values on one component.

The assumption of an ignorable missing-data mechanism is clearly restrictive, but is shared with much of the recent work on imputation methodology; see Lanke (1983), Herzog and Rubin (1983) and Rubin and Schenker (1986). These authors derive imputation procedures with the aim of constructing confidence intervals with predictable coverages. Unfortunately, that particular aim is often incompatible with other possible goals of the secondary analyst. In fact, it has to be admitted that no imputation procedure is simultaneously reliable for all possible such goals.

Appendix 1. Derivation of (3), (4) and (5)

We shall require the technical conditions that $f''(x)$ exists and is square integrable and that the kernel function satisfies the following properties:

$$I_1 = \int u^2 K(u)\,du < \infty; \qquad \int uK(u)\,du = 0; \qquad I_2 = \int K^2(u)\,du < \infty;$$

$$I_3 = \iint K(u)K(u+v)K(v)\,du\,dv < \infty; \qquad I_4 = \int \left\{ \int K(u)K(u+v)\,du \right\}^2 dv < \infty.$$

In (1), we write the two components as

$$\hat{f}(y) = T_0(y) + U_1(y).$$

Similarly, in (2), we write

$$\tilde{f}(y) = T_0(y) + T_1(y).$$

Let $D_0 = (y_1, \ldots, y_n)$ and $D_1 = \{\hat{y}_{jk}: j = n+1, \ldots, N,\ k = 1, \ldots, r\}$. We assume that, given N, $N - n \sim \mathrm{Bi}(N, \theta)$. Our calculations are similar to those of Rosenblatt (1956).

First, note that

$$E_{D_1}\hat{f}(y) = \tilde{f}(y) \quad \text{and} \quad \mathrm{var}_{D_1}\hat{f}(y) = (N-n)N^{-2}r^{-1}h^{-2}\ \mathrm{var}\ K((y-x)/h),$$

where the variance on the right side is based on the density $\hat{f}_0(x)$. Substitution and Taylor expansion gives

$$E_n\ \mathrm{var}_{D_1}\hat{f}(y) = \theta(Nrh)^{-1}I_2 f(y) + o(N^{-1}h^{-1}). \tag{A.1}$$

Now consider the variance of $\tilde{f}(y)$. Since we shall eventually be considering $h = O(N^{-1/5})$ it will turn out (cf. Titterington and Mill, 1983) that, for the dominant term, we need consider only

$$E_n\ \mathrm{var}_{D_0}\tilde{f}(y) = E_n\ \mathrm{var}_{D_0}[T_0(y) + T_1(y)]. \tag{A.2}$$

Taylor expansion and expectation give

$$E_n\ \mathrm{var}_{D_0}T_0(y) = (1-\theta)I_2 f(y)(Nh)^{-1} + o(N^{-1}h^{-1}), \tag{A.3}$$

$$E_n\ \mathrm{cov}_{D_0}(T_0(y), T_1(y)) = \theta I_3 f(y)(Nh)^{-1} + o(N^{-1}h^{-1}), \tag{A.4}$$

$$E_n\ \mathrm{var}_{D_0}T_1(y) = \theta^2(1-\theta)^{-1}I_4 f(y)(Nh)^{-1} + o(N^{-1}h^{-1}). \tag{A.5}$$

Combination of (A.1)–(A.5) gives var $\tilde{f}(y)$ and var $\hat{f}(y)$. For the common bias of $\tilde{f}(y)$ and $\hat{f}(y)$ note that

$$\tilde{f}(y) = N^{-1}\left\{ n\hat{f}_0(y) + h^{-1}(N-n)\int K((y-x)/h)\hat{f}_0(x)\,dx \right\}.$$

Substitution of the kernel-based formula for $\hat{f}_0(x)$, expansion and averaging give

$$E\tilde{f}(y) = E\hat{f}(y) = f(y) + \tfrac{1}{2}(1+\theta)h^2 I_1 f''(y) + o(h^2).$$

We then obtain formulas (3) and (4) of Section 3 with

$$H^2 = (1+\theta)^2 \int \{f''(x)\}^2\,dx \cdot I_1^2,$$

$$\tilde{G} = (1-\theta)I_2 + 2\theta I_3 + \theta^2(1-\theta)^{-1}I_4,$$

$$\hat{G} = \tilde{G} + \theta I_2/r.$$

For use in the numerical illustration in Section 3 we have, if $K(\cdot)$ is the Standard Normal kernel, $I_1 = 1$, $I_2 = (2\sqrt{\pi})^{-1}$, $I_3 = (4\sqrt{\pi})^{-1}$ and $I_4 = (2\sqrt{2\pi})^{-1}$.

The analysis for $\hat{f}(\cdot)$ is similar but simpler. The corresponding G-value is

$$\hat{\hat{G}} = \{(1-\theta)^{-1} + \theta/r\} I_2,$$

and the H-value is different from the above:

$$\hat{\hat{H}}^2 = \int \{f''(x)\}^2 \, dx \cdot I_1^2.$$

Appendix 2. Derivation of (6)

Clearly,

$$E_{D_1}(\tilde{f}(y) \mid \{D_k\}, D_0, n) = f^*(y)$$

and

$$\operatorname{var}_{D_1}(\tilde{f}(y) \mid \{D_k\}, D_0, n) = (N-n)N^{-2}r^{-2}h^{-2} \sum_{k=1}^{r} \operatorname{var}_{y_{1k}} K\{(y - y_{1k})/h\},$$

from which

$$E \operatorname{var}_{D_1}(\tilde{f}(y) \mid \{D_k\}, D_0, n) = \theta(Nrh)^{-1} I_2 f(y) + o(N^{-1}h^{-1}); \tag{A.6}$$

cf. (A.1).

Next, it is easy to see that

$$E_{\{D_k\}}(f^*(y) \mid D_0, n) = \tilde{f}(y), \tag{A.7}$$

so that the biases of $\tilde{f}(y)$ and $f^*(y)$ are the same as the common bias of $\hat{f}(y)$ and $\tilde{f}(y)$. Furthermore, it follows from (A.6) and (A.7) that

$$\operatorname{var} \tilde{f}(y) - \operatorname{var} \hat{f}(y) = \operatorname{var} f^*(y) - \operatorname{var} \tilde{f}(y) = E \operatorname{var}_{\{D_k\}}(f^*(y) \mid D_0, n).$$

Now

$$\begin{aligned} \operatorname{var}_{\{D_k\}}(f^*(y) \mid D_0, n) &= \left(\frac{N-n}{Nhrn}\right)^2 \operatorname{var}_{\{y_{jk}\}} \left[\sum_{k=1}^{r} \sum_{j=1}^{n} \int K\{(y-x)/h\} h^{-1} K\{(x - y_{jk})/h\} \, dx \right] \\ &= \left(\frac{N-n}{Nhrn}\right)^2 nr \left\{ \frac{n-1}{n^2} \sum_{j=1}^{n} J_j^2 - \frac{1}{n^2} \sum_{\substack{j=1 \\ j \neq l}}^{n} \sum_{l=1}^{n} J_j J_l \right\}, \end{aligned} \tag{A.8}$$

where $J_j = \int K\{(y-x)/h\} h^{-1} K\{(x - y_j)/h\} \, dx$, $j = 1, \ldots, n$, Now

$$E_{D_0} J_j = O(h) \quad \text{and} \quad E_{D_0} J_j = h I_4 f(y) + o(h).$$

Thus, when (A.8) is averaged over D_0, the second term is $O(h^2)$ and of smaller order than the first, giving

$$E_{D_0} \operatorname{var}_{\{D_k\}}(f^*(y) \mid D_0, n) = \left(\frac{N-n}{Nhrn}\right)^2 nrh I_4 f(y)(1 + o(1))$$

and

$$E_n E_{D_0} \operatorname{var}_{\{D_k\}}(f^*(y) \mid D_0, n) = (Nh)^{-1}\theta^2 I_4(1-\theta)^{-1} r^{-1} f(y)(1-o(1)).$$

The relationships among the integrated mean squared errors then follow as in (6).

Acknowledgement

This work was supported by the U.S. Bureau of the Census, under contract JSA 80-12, and by the U.K. Science and Engineering Research Council, under Research Grant GR/D/06896.

References

Efron, B. (1979), Bootstrap methods: another look at the jackknife, *Ann. Statist.* 7, 1–26.

Herzog, T. and D. Rubin (1983), Using multiple imputation to handle nonresponse in sample surveys, in: W.G. Madow, I. Olkin and D.B. Rubin, eds., *Incomplete Data in Sample Surveys: Theory and Bibliographies* (Academic Press, New York) pp. 209–245.

Kalton, G. and D. Kasprzyk (1982), Imputing for missing survey response, *Proc. Survey Res. Methods Sect. Amer. Statist. Assoc.*, 22–31.

Kim, J.-O. and J. Curry (1977), Treating missing data in multivariate analysis, *Sociol. Methods Res.* **6**.

Lanke, J. (1983), Hot deck imputation techniques that permit standard methods for assessing precision of estimates, *Statist. Rev. (Statist. Tidskrift)* **21**, 105–110.

Little, R.J.A. (1982a), Maximum likelihood inference for multiple regression with missing values: A simulation study, *J. Roy. Statist. Soc. Ser. B* **41**, 76–87.

Little, R.J.A. (1982b), Models for nonresponse in sample surveys, *J. Amer. Statist. Assoc.* **77**, 237–250.

Little, R.J.A. (1983), Superpopulation models for nonresponse, in: W.G. Madow, I. Olkin and D.B. Rubin, eds., *Incomplete Data in Sample Surveys, Vol. 2, Part VI* (Academic Press, New York) pp. 337–413.

Little, R.J.A. (1988), Missing-data adjustments in large surveys, *J. Business Econom. Statist.* **6**, 287–301.

Rosenblatt, M. (1956), Remarks on some nonparametric estimates of a density function, *Ann. Math. Statist.* **27**, 832–837.

Royall, R. (1983), Finite populations, sampling from, in: S. Kotz and N. Johnson, eds., *Encyclopedia of Statistical Sciences, Vol. 3* (Wiley, New York) pp. 96–101.

Rubin, D.B. (1976), Inference and missing data, *Biometrika* **63**, 581–590.

Rubin, D.B. (1978), Multiple imputations in sample surveys—a phenomenological Bayesian approach to nonresponse, *Proc. Survey Res. Methods Sect. Amer. Statist. Assoc.*, 20–28.

Rubin, D.B. (1987), *Multiple Imputation for Nonresponse in Surveys* (Wiley, New York).

Rubin, D.B. and N. Schenker (1986), Multiple imputation for interval estimation from simple random samples with ignorable nonresponse, *J. Amer. Statist. Assoc.* **81**, 366–374.

Sande, I.G. (1982), Imputation in surveys: Coping with reality, *Amer. Statist.* **36**, 145–152.

Schieber, S.J. (1978). A comparison of three alternative techniques for allocating unreported social security income in the survey of the low-income aged and disabled, *Proc. Survey Res. Methods Sect. Amer. Statist. Assoc.*, 212–218.

Scott, D.W. and L.E. Factor (1981), Monte Carlo study of three data-based nonparametric density estimators, *J. Amer. Statist. Assoc.* **76**, 9–15.

Sedransk, J. and D.M. Titterington (1980), Non-response in sample surveys, unpublished report.

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

Titterington, D.M. (1983), Kernel methods for censored, truncated and grouped data, *Comm. Statist.—Theory Methods* **12**, 2151–2167.

Titterington, D.M. and G.M. Mill (1983), Kernel-based density estimates from incomplete data, *J. Roy. Statist. Soc. Ser. B* **45**, 258–266.

Titterington, D.M. and J. Sedransk (1984), Mean imputation and random imputation methods, unpublished report.