Statistical Science
1989, Vol. 4, No. 2, 164–183

Inference from Grouped Continuous Data: A Review

Daniel F. Heitjan

Abstract. "Grouped" data are defined to be the result of observing continuous variables only up to the nearest interval, rectangle or triangle. This paper traces the development of statistical methods for grouped data, focusing on the major results and their interpretations. It emphasizes the impact of likelihood and Bayesian ideas on the analysis of grouped data, particularly as they influence current work.

Key words and phrases: Asymptotic expansion, heaping, incomplete data, interval censoring, rounding, Sheppard's corrections.

1. INTRODUCTION

Statisticians find it expedient to distinguish between "continuous" and "discrete" variables. In the real world, however, continuous variables are observed and recorded in finite precision. In a fundamental sense, all continuous variables are eventually rounded or coarsened, i.e., grouped.

Grouped data arise in a number of ways. Ideally, the level of grouping coarseness in a data set is the result of a deliberate compromise between the desire to know a continuous variable and the cost of knowing it. In certain situations, for example when confidentiality must be preserved, data are collected at a fine level of precision and then coarsened. Most often, however, grouping results from the unconscious selection of a level of accuracy by the investigator gathering the data. Whatever its source, grouping is ubiquitous, and so throughout this century statistical science has studied its ramifications for data analysis. My purpose is to review this research area.

By grouped data I mean data whose true values are known only up to subsets of the sample space, usually rectangles, triangles or unions of such shapes. This definition includes the important special case of rounding, i.e., the substitution of interval midpoints for true data values. It also includes interval censoring; an example of this is disease relapse time that is only known to lie between a pair of consecutive follow-ups, the times of which may themselves be subject to randomness. A third special case is the actual "grouping" or coarsening of effectively continuous variables into categories during collection or publication. An example of this is the collection of interval income data in sample surveys.

The literature on grouped data spans the modern era of statistics. Thanks to its enduring interest, this body of work has already been reviewed in two encyclopedia articles: Gjeddebaek (1968) discussed some major research areas, primarily from a likelihood perspective, and later Haitovsky (1982) provided a comprehensive summary of results. With this article I intend a more integrated treatment of the methods of grouped data inference, with special emphasis on the central problems, techniques and results of the major research areas and on current research trends.

2. GENERAL STATEMENT AND NOTATION

The common feature of the problems I discuss is that the data values themselves are not known exactly, but can at best be identified with a set of possible values. So for example when Fisher (1936) reports that a flower's petal is 1.4 cm long, we interpret this to mean that the length was actually between 1.35 and 1.45 cm. If the same petal's width is given as 0.2 cm, we surmise that the length and width of that petal actually lie somewhere in the rectangle $(1.35, 1.45) \times (0.15, 0.25)$. If another flower's petal has the same nominal length and width, it does not mean that the two flowers are exactly identical, but that their dimensions were found to lie in the same length/width rectangle. Considerations of this kind lead to the following general statement of the grouped data problem.

Suppose that the random variable $\mathbf{X}$ is distributed according to a density f in a sample space χ, which is partitioned into a collection of disjoint measurable

Daniel F. Heitjan is Assistant Professor, Center for Biostatistics and Epidemiology, Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033.

sets $\{S_i\}$, $\chi = \cup S_i$. Suppose that instead of observing $\mathbf{X}$ exactly, one receives the datum $\mathbf{Y} = \mathbf{Y}(\mathbf{X})$, where $\mathbf{Y}$ is a function that conveys the identity of the subset S_i of χ into which $\mathbf{X}$ has fallen. In the modern language of missing data, $\mathbf{x}$, the unobserved value of $\mathbf{X}$, is the complete data, although $\mathbf{y}$, the observed value of $\mathbf{Y}$, is the incomplete data. I use the name $X(\mathbf{y})$ to refer to the set in which $\mathbf{y}$ has been observed to fall; in symbols

$$X(\mathbf{y}) = \{\mathbf{x}: \mathbf{Y}(\mathbf{x}) = \mathbf{y}\}. \tag{1}$$

The above definition covers a wide variety of missing data situations. I have in mind in particular the situation where the S_i's represent intervals, rectangles, strips, triangles or unions of these kinds of objects, and will refer to $\mathbf{y}$ in such situations as grouped data. The question of interest is how to draw inferences from $\mathbf{y}$ to an unknown parameter $\theta \in \Omega$ that governs the distribution of $\mathbf{X}$.

For samples of size n, the sample space χ is typically the n-fold product of the sample space for a single observation. The component spaces, however, need not always be partitioned in the same way. For example, some data may be rounded to the nearest 0.1 cm and others to the nearest 0.01 cm, the component spaces for the latter points having a finer partition. In the case of interval censoring, each component may have its own partition, and the choice of partition itself may be random. If this source of randomness is of interest, for example if interval pattern is thought to be related stochastically to the value of $\mathbf{x}$, it may be necessary to incorporate random variables for partition choice into the model as well. (For an example of this kind of modeling see Section 7.)

In many grouped data situations, and especially in rounding, the data are presented as the centers of their grouping sets. I refer to the random variable constructed from the grouped data in this way as $\mathbf{Y}^* = \mathbf{Y}^*(\mathbf{Y}(\mathbf{X}))$, and to realizations of it as $\mathbf{y}^*$; note that $\mathbf{y}^*$ usually lies in χ. It is common practice to construct a statistic from rounded data by applying the usual complete data statistic $\mathbf{U}(\cdot)$ to $\mathbf{y}^*$. I therefore distinguish between the complete data statistic $\mathbf{U} = \mathbf{U}(\mathbf{X})$ and the corresponding rounded data statistic $\mathbf{U}^* = \mathbf{U}(\mathbf{Y}^*)$. The random variable $\mathbf{U}^*$ is known in the literature as the simple statistic; thus, one refers to the simple mean $\bar{y}^*$ and simple variance s^{*2}.

3. PRACTICAL RESULTS FOR ROUNDED DATA

It is natural to ask how well standard continuous data methods perform when applied to rounded data, i.e., how well can $\mathbf{U}^*$ act as a substitute for $\mathbf{U}$? In particular, one might inquire about the frequency properties of $\mathbf{U}^*$, or how well it summarizes posterior uncertainty about θ. Not surprisingly, there has been much research into these questions, and answers in some elementary cases are quite well understood. I present the main results useful in practice now, leaving detailed justifications for later sections.

Univariate Data: Mean and Variance

The literature, as cited below, on the simple estimates points to two main conclusions. First, the simple mean $\bar{y}^*$ is a good estimate of $\mu = EX$, and $s^*/\sqrt{n}$ is a good summary of the uncertainty about μ in $\bar{y}^*$. Second, s^{*2} is not a good estimate of $\sigma^2 = \text{Var } X$, but can be improved by subtracting $h^2/12$, where h is the width of the rounding interval for X. A standard error for this estimate can also be constructed from simple moment estimates by elementary corrections. These statements are intended to be interpreted loosely in either a Bayesian or frequentist sense. For the Bayesian, I consider an estimate to be good if it is a reasonable approximation to the posterior mean of the parameter it estimates, using some data-dominated prior. Similarly, I consider a good Bayesian summary of uncertainty to be one that approximates well the posterior standard deviation of the parameter in question. For the frequentist, I consider an estimator to be good if it is at least approximately unbiased, and a summary of uncertainty to be good if it approximates well the sampling standard deviation of the estimate.

There are a number of qualifications on the assertions that the mean and corrected variance are good estimates, and the Bayesian qualifications are rather different from the frequentist ones. There are two common threads in the analysis. First, both the Bayesian and frequentist approximations improve as the rounding interval width h decreases to zero. Second, both kinds of approximation are poor for both large and small samples, although they can be satisfactory for a wide range of moderate values of n.

From the sampling theory perspective (Kendall, 1938), the results stated above hold if

i) the density of $\mathbf{X}$ and several of its derivatives go smoothly down to zero at both terminals of its range, or, if $\mathbf{X}$ has infinite range, at some finite points spanning almost all of its probability,

ii) n is not too small or too large (say, $5 < n < 100$), and

iii) h is not too large (say, $h < 1.6\sigma$).

If these conditions are met, then $\bar{y}^* - \mu$ and $s^{*2} - h^2/12 - \sigma^2$ have sampling means that are small compared to h^3. Although these biases vanish as $h \to 0$, arbitrarily precise estimation for any value of h is not possible, because as $n \to \infty$ the bias comes to dominate the sampling noise. So for example the normal mean μ may be well estimated by $\bar{y}^*$ in moderate samples even with $h = \sigma$ (Kendall, 1938), but by $n = 100$ bias

comes to dominate sampling error (Gjeddebaek, 1957). The simple mean can be a good estimate of μ even if the conditions above are violated, but the variance is more sensitive to assumptions about the tails. For example, the expectation of $\bar{y}^*$ from rounded uniform data is μ, but the simple variance in this case underestimates σ^2 by $h^2/12$. It is also worth noting that the bias in $\bar{y}^*$ depends upon the location of the center of the distribution of X relative to the rounding grid. For discussion of related problems see Holland (1975) and Preece (1981).

Eisenhart (1947) has shown that the Student t with $n - 1$ degrees of freedom is a good approximate sampling distribution for $t^* = \sqrt{n}(\bar{y}^* - \mu)/s^*$. This approximation holds if, say, $h < \sigma$ and $n \geq 5$ (but again, not too large), the chief problem in small samples being discreteness in the distribution of t^*. Similarly, tests and confidence intervals about σ^2 can be based on the approximate χ^2 distribution (with $n - 1$ degrees of freedom) of $(n - 1)s^{*2}/(\sigma^2 + h^2/12)$. There are notable problems with these reference distributions, however. For example, for any sample size there is a positive probability that $s^* = 0$; if $h = \sigma$, $\mu = 0$ and $n = 3$, this probability is about 8.5%. Thus, the sampling distribution of t^* has positive mass at ∞, although the probability of this event decreases rapidly with n. (The discreteness problem, incidentally, was noticed by Gossett in his pioneering paper on t [Student, 1908]). Confidence intervals based on the χ^2 distribution cited above give roughly the correct coverage properties but can include negative values. This can be seen by noting that formulas for these intervals based on the method of pivots simply replace the usual complete data sample variance s^2 with s^{*2}, and reduce the upper and lower bounds by $h^2/12$.

For Bayesian inference, the main assertions above hold if h is small and n is not too large, and if

i) $\mathbf{X}$ is normal, or
ii) $\mathbf{X}$ is regular (see Section 5) and n is also not too small.

The basis of the assertions is that, when these conditions are satisfied, $\bar{y}^*$ and $s^{*2} - h^2/12$ approximately maximize the correct, incomplete data likelihood of μ and σ^2 (Lindley, 1950; Tallis, 1967; Don, 1981; Dempster and Rubin, 1983). Standard errors for the mean and variance can be derived by inverting the normal information matrix for μ and σ^2 (Fisher, 1922; Tallis, 1967). Therefore, $\bar{y}^*$ and $s^{*2} - h^2/12$ may be thought of either as approximate maximum likelihood estimates (MLEs) or approximate posterior modes of μ and σ^2, and their standard errors may be interpreted as either large n sampling standard deviations of the MLE or as standard deviations of a data-dominated posterior distribution. An explicitly Bayesian approach has not been applied to the case of small grouped samples; thus, it is not known to what extent Student t and χ^{-2} posteriors are appropriate for μ and σ^2.

Multivariate Rounded Data and Regression

As in the univariate case, a single set of formulas for multivariate rounded data are justified on both frequentist and Bayesian grounds. Specifically, the mean vector $\boldsymbol{\mu} = E\mathbf{X}$ may be estimated by the simple mean vector $\bar{\mathbf{y}}^*$, with associated dispersion matrix S^{*2}/n, where S^{*2} is the uncorrected variance-covariance matrix. The variance-covariance matrix Σ can be estimated by subtracting from each diagonal element of S^{*2} the quantity $h^2/12$ for the h corresponding to the widths of the rounding intervals in that direction. Off-diagonal elements require no correction. Standard errors for the elements of Σ and cross-covariances of Σ and $\boldsymbol{\mu}$ terms are more complex; I refer the interested reader to the paper by Tallis (1967).

The theory of estimation from grouped multivariate data has not been studied as intensively as that for univariate data, although in what is known there are many similarities. Conditions for small bias in the mean and corrected variance are, as before, small rounding intervals and high order contact (Wold, 1934). The corrections also give approximate MLEs for the mean and variance if, as before, the grouping widths are small, n is not too large, and either $\mathbf{X}$ is normal or $\mathbf{X}$ is regular and n is not too small (Dempster and Rubin, 1983).

The results of the preceding paragraphs lay the foundation for the treatment of rounding errors in regression. If the model satisfies either set of conditions stated above one can account for rounding by simply using a corrected sums of squares and cross-products matrix as input to a regression routine. Regression coefficients computed in this way lie close to a mode of the likelihood and are roughly unbiased. If grouping prevails even in the "independent" variables, it is essential to specify, at least vaguely, their marginal distribution, because if the grouped predictors are not regular, simple corrections may not be appropriate. For example, the variance adjustment for the grouped uniform (which is not regular) involves adding $h^2/12$ rather than subtracting it. For more discussion of these ideas see Dempster and Rubin (1983) and Beaton, Rubin and Barone (1976).

Other aspects of the sampling theory of multivariate rounded data have been discussed by Baten (1931) and Haitovsky (1973, Chapter 6). In addition to the references cited above, likelihood theory leading to simple corrections in normal models has been treated by Fryer and Pethybridge (1972), Pethybridge (1973, 1975) and Indrayan and Rustagi (1979). Some other

papers of interest are by Berkson (1950), Durbin (1954) and Cochran (1968), who considered grouping in the context of errors in variables.

Two Kinds of Asymptotics

The mean and corrected variance estimators discussed in this section have sampling properties that most would label quite poor. They are biased under most models, have annoying singularities in small samples and are inconsistent. Because bias decreases with interval width, however, small h asymptotics are favorable. To the extent that simple moment corrections lead to approximate MLEs, one can be reasonably sure that their frequency properties are good, but even so approximation error (due not only to grouping but to model misspecification) will come to dominate sampling error for sufficiently large n. For the Bayesian, adjustments provide information on the approximate maximum of the likelihood and the curvature there that can be valuable in summarizing beliefs about parameters, although such summaries are also potentially misleading if n or h is large.

The limitations of the simple approximation methods are troubling. This has led investigators to study theoretical and numerical aspects of an approach based on maximizing the incomplete data likelihood rather than a substitute complete data likelihood. The fruits of this research are reviewed in Section 6. Before embarking on a detailed analysis of the results of this section, I will bring them into clearer focus by considering an example.

An Example: Fisher's Iris Data

The Fisher (1936) iris data has not been viewed as grouped data, but as I suggested in Section 2, it might well be. Figure 1 is a display of petal length and width of the 50 specimens of *Iris setosa*. The rectangles are sets in the partition of the bivariate sample space, and the numbers are the counts in those sets. So for example eight flowers fell in the cell (1.35, 1.45] × (0.15, 0.25], and were therefore reported as (1.4, 0.2).

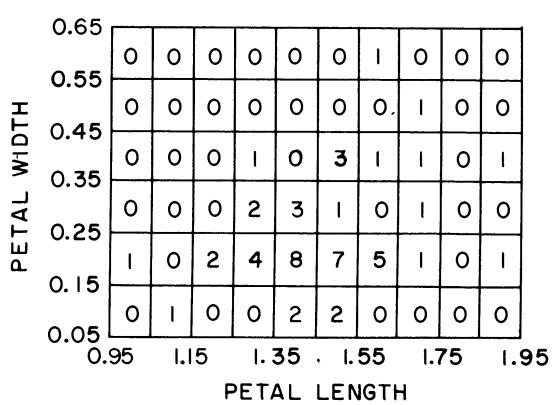

FIG. 1. *Data on petal length and petal width of fifty specimens of Iris setosa (Fisher, 1936). The number of flowers in each rectangle is shown at the center of the rectangle.*

A standard analysis of these data might assume that length and width follow a bivariate normal distribution, whose parameters one could estimate by maximum likelihood. From the standpoint of grouping, at least three levels of precision are possible in such an analysis. The first is the usual analysis: ignore the grouping and estimate the parameters by simple, uncorrected sample moments. A second, more precise analysis consists of correcting the sample moments by subtracting $h^2/12 = (0.1)^2/12$ from each variance. The correlation estimate is the uncorrected covariance divided by the product of the square roots of the corrected variances. A third, most precise analysis consists of maximizing the correct, grouped data likelihood, i.e., the product over units of the integrals of a bivariate normal density over the rounding rectangles. The first two analyses are trivial to apply; I have implemented the third in FORTRAN using a Newton-Raphson algorithm.

Results of the three analyses appear in Table 1. Standard errors for the simple estimates follow from the usual sampling variance formulas, whereas standard errors for the corrected estimates are based on the corrected moment formulas of Tallis (1967), and standard errors for the MLE are the square roots of the diagonal of the inverse negative Hessian of the grouped data log-likelihood. The salient feature in this table is the relative insensitivity of estimates and standard errors to the level of approximation used in the analysis. Means change not at all, and variances and the correlation only slightly, across the rows of the table. In these data, where the groups are of uniform size and h is no larger than σ in each direction, corrections give a fine approximation to the grouped data MLE.

In recent years it has become popular to transform these data before analysis; Gnanadesikan (1977), for instance, took logarithms. Figure 2 presents the data again, this time with rectangles stretched to their size on the log scale; the counts are centered at the transformed values of the raw scale rectangle centers. Although this nonaffine transformation presumably leads to greater homogeneity and a closer approxi-

TABLE 1
Iris setosa data on petal dimensions, raw scale parameter estimates

Parameter	Simple Estimate	(SE)	Corrected Estimate	(SE)	MLE Estimate	(SE)
Mean (length)	1.462	(0.024)	1.462	(0.024)	1.462	(0.024)
Mean (width)	0.246	(0.015)	0.246	(0.015)	0.246	(0.015)
Var (length)	0.0296	(0.0059)	0.0287	(0.0059)	0.0287	(0.0059)
Var (width)	0.0109	(0.0022)	0.0100	(0.0022)	0.0101	(0.0022)
Correlation	0.332	(0.126)	0.350	(0.133)	0.350	(0.132)

mation to normality, it does so by forfeiting the convenient structure of rounded data and therefore the opportunity for corrections.

Results of the simple and ML analyses of the log scale data are presented in Table 2. For petal length the differences between the simple and MLE are similar to what was found in the raw scale analysis: no change in the mean and a small decrease in the variance. Petal width is grouped more coarsely and unevenly, however. Consequently both the mean and variance estimates are more dramatically different in the ML analysis: the mean changes by 1/6 of a standard error and the variance by 2/3 of a standard error. The difference between the simple variance of log petal width and the ML variance is $h'^2/12$ for $h' = 0.51$ (i.e., $h' = \sqrt{12[0.164 - 0.142]}$), which is approximately equal to the log-scale length of the modal petal width interval. This is a common finding; even when the correction formulas are not strictly appropriate, they can give rough guidance on the size and direction of the needed adjustments.

The sensitivity analyses presented here may seem unconvincing in that the observed effects are fairly small. In larger samples, however, the magnitude of the corrections stays the same although the standard error decreases, so that eventually some account must be taken of grouping. As the example shows, the effects of grouping in this normal model are first felt in the variances. Although the variance itself is rarely

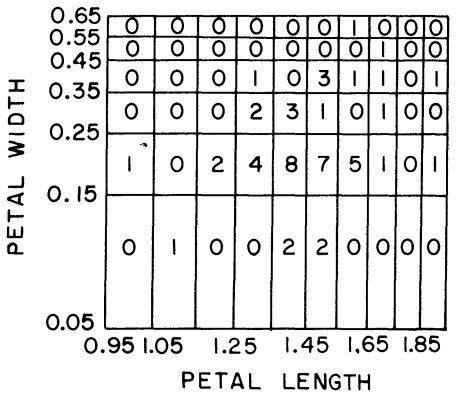

FIG. 2. *The data of Figure 1 reexpressed on the log scale. The number of flowers in each rectangle is shown at the transformed value of the raw scale rectangle center.*

TABLE 2
Iris setosa data on petal dimensions, log scale parameter estimates

Parameter	Simple Estimate (SE)	MLE Estimate (SE)
Mean (log length)	0.373 (0.017)	0.373 (0.017)
Mean (log width)	−1.485 (0.057)	−1.477 (0.057)
Var (log length)	0.0143 (0.0029)	0.0138 (0.0029)
Var (log width)	0.164 (0.033)	0.142 (0.032)
Correlation	0.309 (0.128)	0.346 (0.136)

of interest, errors in its estimation can distort inferences on more useful quantities such as correlations and regression coefficients.

4. SAMPLING BASIS OF SHEPPARD'S CORRECTIONS

Asymptotic Expansions

The formulas described in Section 3 are usually attributed to Sheppard (1898) (who, incidentally, advocated a Bayesian interpretation). In those early days, statistical thinking was dominated by the method of moments, wherein it was held that the best way to learn about a distribution was to discover its moments. Because, whatever else may be said about them, the $\mathbf{U}^*$ moment statistics are unbiased estimates of their expectations, it was natural to seek formulas relating the moments of $\mathbf{Y}^*$ to those of $\mathbf{X}$. For arbitrary grouping schemes, no general results exist, but for rounded data, the observation that $E\mathbf{U}^* \to E\mathbf{U}$ as $h \to 0$ suggested that expansions asymptotic in h might provide some guidance. Such reasoning has led to a number of arguments for Sheppard's corrections.

The key mathematical tool in this area is the Euler-Maclaurin theorem (e.g., Stoer and Bulirsch, 1980). Suppose that the finite interval $[a, b]$ is partitioned into N subintervals each of length h, $h = (b - a)/N$. The integral of a function g on this range may be approximated by the trapezoidal sum

$$T_{g(x)}(h) = [h/2][g(a) + 2g(a + h) + 2g(a + 2h) + \cdots + 2g(b - h) + g(b)].$$

If g has four continuous derivatives on $[a, b]$, Euler-Maclaurin asserts the following representation for the dependence of the error in this approximation on h:

$$(2)\quad \int_a^b g(t)\,dt = T_{g(x)}(h) + c_2h^2(g'(a) - g'(b)) + c_4h^4g^{(4)}(\xi),$$

for some $\xi \in (a, b)$, where $c_l = B_l/l!$, and B_l is the lth Bernoulli number.

Suppose now that the continuous variable X is distributed according to density f on the finite interval $[a, b]$, but is observed as the rounded variable Y^*; specifically, Y^* takes values $a + h(2i - 1)/2$, $i = 1, \cdots, N$. The kth moment of Y^* is

$$\nu_k = E[Y^{*k}] = \sum_{i=1}^{N} \left[a + \frac{(2i - 1)h}{2}\right]^k \int_{a+(i-1)h}^{a+ih} f(t)\,dt,$$

and the kth moment of X is

$$\mu_k = E[X^k] = \int_a^b t^k f(t)\,dt.$$

Sheppard's corrections are intended to relate μ_k to ν_k.

It is enlightening to explore the Sheppard corrections for the first two moments. If f has four continuous derivatives on $[a, b]$, applying (2) to μ_1 yields

$$\begin{aligned}\mu_1 &= \int tf(t)\,dt \\ &= T_{xf(x)}(h) + [h^2/12] \\ &\quad \cdot [af'(a) + f(a) - bf'(b) - f(b)] + R_1(h)h^4,\end{aligned}$$

where $R_1(h)$ is bounded in h. Similarly, by applying (2) with $N = 1$ on each of the subintervals and summing (assuming also a fifth continuous derivative), one obtains

$$\begin{aligned}\nu_1 = T_{xf(x)}(h) + (h^2/12)(2[f(a) - f(b)] + af'(a) \\ - bf'(b)) + R_2(h)h^4,\end{aligned}$$

where $R_2(h)$ is another bounded function. The difference is then

$$\mu_1 - \nu_1 = [h^2/12][f(b) - f(a)] + o(h^3) \quad \text{as } h \to 0, \tag{3}$$

whose h^2 term drops out when $f(b) = f(a)$. Similar arguments for second moments lead to

$$\begin{aligned}\mu_2 - \nu_2 &= -[h^2/4]T_{f(x)}(h) - [h^2/6]T_{xf'(x)}(h) \\ &\quad - [h^2/3][af(a) - bf(b)] + o(h^3) \\ &= -h^2/12 + [h^2/6][bf(b) - af(a)] + h^4Q(h),\end{aligned}$$

where $Q(h)$ is bounded, which under equality of $bf(b)$ and $af(a)$ reduces to

$$\mu_2 - \nu_2 = -h^2/12 + o(h^3), \tag{4}$$

the classic Sheppard's correction for the variance.

A more general statement of the Sheppard formulas is

$$\nu_k = \sum_{m=0}^{[k/2]} (h/2)^{2m}\binom{k}{2m}\mu_{k-2m} \Big/ (2m + 1) + R, \tag{5}$$

where $[\cdot]$ is the greatest integer function and R is a remainder term. Kendall (1938) has stated conditions sufficient for R to be negligible. Specifically, if there exists an integer j such that

i) $2[k/2] < j$,

ii) f and its first j derivatives vanish at a and b, and

iii) $F^{(j+1)}(x) = d^{j+1}/dx^{j+1}(x^k \int_{-h/2}^{h/2} f(x + u)\,du)$ is not large in $[a, b]$,

then $R = o(h^{j-1})$. As the examples of equations (3) and (4) suggest, Kendall's conditions, although they are sufficient for all moments, are not also necessary for all moments; more careful analyses for any given k can lead to more precise conditions at a and b. In addition, moments of variables with infinite range (the normal, for instance) can be approximated by truncating at points that include most of the mass. If the truncated distribution satisfies the conditions, at least approximately, then corrections may be safely applied. Wold (1934) derived the multivariate corrections discussed in Section 3 by similar arguments.

The deriving of Sheppard's corrections was a popular topic in the first half of this century, and there was considerable controversy over which set of sufficient conditions was most realistic and correct. Sheppard's (1898) derivations used both Bayesian arguments and asymptotic series. Pearson (1902) clarified the asymptotic series argument, and in later papers (Pairman and Pearson, 1919; Sandon, 1924; Pearse, 1928; Martin, 1934), he and others discussed ways of applying the series when the remainder terms are not negligible and can be estimated from the data. Fisher (1922) noted that the moments of $\mathbf{Y}^*$ are periodic functions in the location of the distribution with respect to the rounding grid, and so derived Sheppard's corrections as the aperiodic terms in a trigonometric series. Langdon and Ore (1930) and Kullback (1935) discussed Sheppard's corrections for the cumulants. Elderton (1933) proposed basing moment corrections for J-shaped distributions on corrections appropriate for the exponential. Less research on moment corrections per se has appeared since the definitive work of Kendall (1938), although Hartley (1950) discussed computational issues, and Chao (1975) derived Sheppard's corrections using only elementary calculus.

Average Moments

Assuming that the rounding lattice itself is random also leads to Sheppard's corrections. Suppose that each sample is drawn using a different rounding grid, where the center of the rounding grid is uniform on $(-h/2, h/2)$. Then the average over samples of the kth moment of $\mathbf{Y}^*$ is

$$\begin{aligned}E[\nu_k] &= h^{-1}\int_{-h/2}^{h/2} \sum_{-\infty}^{\infty} y_i^k \int_{-h/2}^{h/2} f(y_i + u)\,du\,dy_i \\ &= h^{-1}\int_{-\infty}^{\infty} y^k \int_{-h/2}^{h/2} f(y + u)\,du\,dy.\end{aligned}$$

The latter equation reduces to (5) with $R = 0$, and so shows that the average of the kth moment of $\mathbf{Y}^*$ may be related to the moments of $\mathbf{X}$ *via* Sheppard's corrections.

This result, although interesting, does little to supplement the asymptotic series approach. A commonly encountered situation in which the assumptions for the asymptotic series are not applicable is the case of skewed densities defined on $(0, \infty)$. Data from such distributions are typically rounded to the nearest

multiple of an integer, so that the random rounding grid assumption also is not valid. Furthermore, the random grid argument applies only to moments about zero. Moments about the sample mean (including s^{*2}) require further conditions before Sheppard's adjustments can be used.

Abernethy (1933) pioneered the average moments argument, and Kendall (1938) contributed a more elegant proof. Related work may be found in Carver (1936), Craig (1936, 1941), Cornish and Fisher (1937) and Pierce (1940, 1943).

5. LIKELIHOOD BASIS OF SHEPPARD'S CORRECTIONS

Approximating the Likelihood

For likelihood and Bayesian data analysis, the data contribute to the inference through the likelihood function. In incomplete data problems such as grouped data, the log-likelihood is of the form

$$(6) \qquad L(\theta) = \ln\left\{\int_{X(\mathbf{y})} f(\mathbf{x} \mid \theta)\, d\mathbf{x} \Big/ \int_{X(\mathbf{y})} d\mathbf{x}\right\}.$$

In the grouped data case, $X(\mathbf{y})$ refers to the set of possible $\mathbf{x}$ values that would appear, when grouped, as $\mathbf{y}$. (Note that the denominator on the right of (6) does not depend on θ, and so is irrelevant and can be ignored in applications.) For example, in the case of grouped normal data, the (unscaled) contribution to the log-likelihood from a data point known only to fall in the interval $X(\mathbf{y}) = (a, b)$ is $L(\mu, \sigma) = \ln[\Phi((b - \mu)/\sigma) - \Phi((a - \mu)/\sigma)]$, where Φ is the standard normal integral. For rounded data, compare this with the log-likelihood more commonly used, i.e., the complete data log-likelihood arising from $\mathbf{y}^*$,

$$(7) \qquad L^*(\theta) = \ln f(\mathbf{y}^* \mid \theta).$$

In the example just alluded to, if $y^* = (a + b)/2$ is the midpoint of the interval in which the point is known to lie, this substitute log-likelihood is $L^*(\mu, \sigma) = -0.5 \ln 2\pi - \ln \sigma - 0.5((y^* - \mu)/\sigma)^2$.

Fixing $\mathbf{x}$ and letting the group width approach zero, for every $\theta \in \Omega$, $|L(\theta) - L^*(\theta)| \to 0$. It therefore makes sense to consider the distorting effect of grouping as a function of h, and to explore the small group asymptotic behavior of the likelihood and its functionals. This approach, pioneered by Fisher (1922), has been the thrust of much grouped data research in the latter part of this century.

Suppose that $\mathbf{X}$ is of dimension k, rounded into rectangles of length h_j in direction j, and set $\mathbf{h} = \{h_j\}$. Consider the likelihood arising from a single grouped observation $\mathbf{y}^*$. If the density f of $\mathbf{X}$ has enough continuous derivatives, the (unscaled) likelihood

$$l(\theta) = \int_{X(\mathbf{y}^*)} f(x \mid \theta)\, d\mathbf{x}$$

may be approximated by the integral over $X(\mathbf{y}^*)$ of a Taylor expansion of f in $\mathbf{x}$ about $\mathbf{y}^*$. Setting

$$f^* = f(\mathbf{y}^* \mid \theta),$$

$$f^*_{jj} = \partial^2/\partial x_j^2 f(\mathbf{x} \mid \theta)\,|_{\mathbf{x}=\mathbf{y}^*}$$

yields the approximation

$$(8) \qquad l(\theta) = f^* + \sum_{j=1}^{k} h_j^2 f^*_{jj} \Big/ 24 + o(\|\mathbf{h}\|^2)$$

as $\|\mathbf{h}\| \to 0$. Similarly, for the log-likelihood L

$$(9) \qquad L(\theta) = \ln f^* + \sum_{j=1}^{k} h_j^2 (f^*_{jj}/f^*) \Big/ 24 + o(\|\mathbf{h}\|^2).$$

Formula (9) serves as a simple numerical approximation to the log-likelihood and as an analytic platform for computing further derivatives.

Correction Formulas

Tocher (1949) and Lindley (1950) independently proposed using (9) as the basis for a Newton-Raphson iteration to improve the simple version of the MLE. Starting from $\hat{\theta}^*$, a maximum of the substitute likelihood $L^*(\theta)$, the result of one Newton step, correct to terms of order $\|\mathbf{h}\|^2$, is $\hat{\theta} = \hat{\theta}^* - \mathbf{I}^{*-1}\mathbf{B}$, where

$$\mathbf{B} = \left\{\partial/\partial\theta_l \left(\sum_{j=1}^{k} h_j^2 (f^*_{jj}/f^*) \Big/ 24\right)\right\}_{\theta=\hat{\theta}^*},$$

and $\mathbf{I}^*$ is the substitute observed information matrix computed at $\hat{\theta}^*$. When there are n data points, $\mathbf{B}$ and $\mathbf{I}^*$ represent sums of these terms over the units. If n is large and h small, both $\mathbf{I}^*$ and $\mathbf{B}$ may be replaced by their complete data expectations at $\hat{\theta}^*$. In the case of rounded multivariate normal data, this formula reduces to the Sheppard corrections discussed in Section 3.

Another approach, this one due to Dempster and Rubin (1983), leads to Sheppard's corrections somewhat more generally. If estimation of a multivariate mean and variance or a set of regression coefficients is desired, and the intervals are small, a single step of an EM algorithm (Dempster, Laird and Rubin, 1977) can be used in lieu of a Newton-Raphson step. The E step here reduces to computing the expected sums of squares and cross-products matrix for the data, conditional on the fact that the observations are constrained to lie in the rectangles about the $\mathbf{y}^*$ data points.

Letting x_j be the jth component of the complete-data random vector, $f^* = f(\mathbf{y}^* \mid \hat{\theta}^*)$, and

$$f_j^* = \partial/\partial x_j f(\mathbf{x} \mid \hat{\theta}^*) \mid_{\mathbf{x}=\mathbf{y}^*},$$

one arrives at the approximations

$$\begin{aligned} &E[X_j \mid \mathbf{y}^*, \hat{\theta}^*] = y_j^* + h_j^2(f_j^*/f^*)/12, \\ &E[X_j^2 \mid \mathbf{y}^*, \hat{\theta}^*] = y_j^{*2} + h_j^2(1 + 2(f_j^*/f^*)y_j^*)/12, \\ &E[X_jX_l \mid \mathbf{y}^*, \hat{\theta}^*] \\ &\quad = y_j^*y_l^* + h_j^2(f_l^*/f^*)y_l^*/12 + h_l^2(f_j^*/f^*)y_j^*/12. \end{aligned} \tag{10}$$

These formulas reduce to Sheppard's corrections under two conditions. First, if $\mathbf{X}$ is multinormal, they reduce exactly regardless of n. Second, if f is regular, in the sense that

$$\partial/\partial v_j \int f(\mathbf{x} - \mathbf{v})\, d\mathbf{x} = \int \partial/\partial v_j f(\mathbf{x} - \mathbf{v})\, d\mathbf{x}$$

and

$$\partial/\partial v_j \int x_j f(\mathbf{x} - \mathbf{v})\, d\mathbf{x} = \int \partial/\partial v_j x_j f(\mathbf{x} - \mathbf{v})\, d\mathbf{x},$$

for $j = 1, \cdots, k$, the equations (10) reduce to the Sheppard corrections in large samples.

"Regularity" is a curious property, and its implications have not yet been fully studied. For example, gamma densities (i.e., densities proportional to $x^{\lambda-1} \exp[-\alpha x]$) are regular iff $\lambda > 1$. Thus, although many highly skewed distributions are regular, the exponential, with $\lambda = 1$, is not. Another nonregular distribution is the uniform, where the moment-based correction is to add $h^2/12$. This can be seen using the EM arguments or more directly, as in Beaton, Rubin and Barone (1976).

The likelihood expansion (9) is due to Fisher (1922). Lindley (1950) rederived it and is credited with developing the correction formulas, although the idea appeared earlier in a paper by Tocher (1949), who attributed it to Bliss. Tallis (1967) established that Lindley's formula implied the Sheppard corrections for multinormal data; he also derived asymptotic standard errors. A more elegant derivation is due to Don (1981). The idea of using EM instead of Newton-Raphson, and the consequent discovery of the broader applicability of Sheppard's corrections, are due to Dempster and Rubin (1983).

6. BEYOND MOMENT CORRECTIONS

Theoretical Developments

Compared with many other kinds of data, the frequency theory of grouped data analysis is poorly developed. The main problem is that the sampling distributions of many potential estimators are complex and unattractive, so that the usual sort of decision-theoretic analysis has not been undertaken. Thus for a long time the bulwark of the sampling theory of grouped data was the Sheppard moment corrections, whose sampling properties are, as I have indicated, poor.

On the other hand, maximum likelihood estimates have been shown in the normal and exponential cases to be consistent and asymptotically efficient (Kulldorff, 1961). Except for these distributions, the asymptotic theory has not been carefully examined. This has not deterred practice, however, where it has become more common to see examples of grouped data ML estimation.

The problems with discreteness mentioned in the context of moment corrections affect the sampling distribution of the MLE as well. For instance, suppose a variable $\mathbf{X}$ is distributed normally with mean μ and unit variance, and that instead of observing $\mathbf{X}$, one observes only its sign. If in n independent trials one records n_1 negatives and n_2 positives, the MLE of μ is the solution of $n_1/n = \Phi(-\mu)$. Thus, if $n_1 = 0$, as frequently happens for small n or large μ, the MLE is at positive infinity, even though the log-likelihood is concave. This problem has typically been handled by considering the sampling properties of the MLE conditional on its existence.

The example just cited is an extreme case, and in fairness it should be noted that likelihood-based tests and confidence intervals might still give sensible answers when the MLE does not exist. In real applications such radical behavior is rare. More commonly one observes that, if the grouping is coarse, the MLE exists but is difficult to locate because of near ridges in the likelihood. Heitjan (1987) has recorded this phenomenon for the grouped bivariate normal.

Unimodality

Although the mode of a grouped data log-likelihood can be hard to find, in some common grouped data models any mode that exists is known to be unique. Burridge (1981a, 1982) has shown that log-concave densities (including the normal and the logistic) lead to likelihoods that are also log-concave in appropriately chosen parameterizations. Thus, if such a parameterization exists, and one can demonstrate a mode (in any parameterization), the mode must be unique. Well chosen parameters can have other tangible benefits, including more rapid convergence of iterative algorithms and better normal approximations.

The most powerful results in this area are based upon theorems from convex analysis that have only

recently come to the attention of statisticians (Prekopa, 1973; Brascamp and Lieb, 1975; Pratt, 1981). Yet there are tantalizing gaps in the theory. For example, grouped data from a univariate normal with unknown μ and σ lead to a log-concave likelihood in an appropriate parameterization; $1/\sigma$ and μ/σ will do. The methods of proof used for this result, however, do not readily generalize to all higher dimensional problems. For instance, it is not known whether there exists a concavity parameterization for the log-likelihood arising from rectangularly grouped bivariate normal data with unspecified mean and variance-covariance matrix. For the special case of bivariate normal data where each item is either fully observed or completely missing, multimodal likelihoods exist (see Murray, 1977), and so there can be no concavity parameterization. However, this still leaves open the possibility that with some restrictions on the grouping a concavity parameterization might be available.

Parametric Maximum Likelihood from Arbitrarily Grouped Data

Approximation methods based on small h asymptotics are useful and interesting, but fine rounding is just one special case of grouped data. As in the iris data example, it is often expedient to fit a model to nonlinear transformations of the same set of data; however, rounding, and so the relevance of adjustments, are preserved only under affine transformation. It is also possible to have finely rounded data, but so much of it as to arouse concern that the error in the adjustments is larger than the standard error of the parameter. For these situations the approach of calculating the likelihood as precisely as is feasible and iterating to an MLE may be the only practical alternative. In this section, I summarize studies that have been conducted along these lines. The interested reader is referred to these works.

For grouped univariate normal data, the work of Gjeddebaek (1949, 1956, 1957, 1959a, 1959b, 1961) is foremost. Other early contributors were Stevens (1948) and Yoneda and Uchiyama (1956). Tallis and Young (1962) and Heitjan (1987) considered arbitrarily grouped bivariate normal data; at this time this is the highest dimensional normal model that can be routinely fit. Deken (1983) discussed higher dimensional multivariate data, but his algorithm is appropriate only for very small grouping rectangles, in which case corrections can be a viable alternative. Kulldorff (1961) has fit grouped data exponential models. Aigner and Goldberger (1970) covered the Pareto, and Flygare, Austin and Buckwalter (1985) have treated the Weibull. Boardman (1973) considered the compound exponential, a special case of bivariate grouping in which the grouping sets are unions of rectangles and triangles. Most recently, Pettitt (1985) and Beckman and Johnson (1987) have fit the t distribution, including a degrees of freedom parameter, the latter by Newton's method, the former by EM.

Choice of Algorithms

Because the usual definition of grouped data subsumes standard kinds of censoring as well, good algorithms for grouped data likelihood calculations can be quite generally useful. Available methods are variants of either EM, Newton-Raphson or Fisher scoring. Newton-Raphson appears to be the swiftest, followed by scoring and EM (Schader and Schmid, 1984). Burridge (1981a) has suggested using concavity parameterizations to speed up Newton-Raphson algorithms. EM algorithms, although they have at best a linear rate of convergence, are guaranteed never to decrease the likelihood and so are quite robust, a virtue not to be taken lightly in the face of the vagaries of real data. They are also easier to program and less costly per iteration than the other methods.

Computer programs in the literature are by Swan (normal by Newton-Raphson, 1969), Benn and Sidebottom (scale and location families by scoring, 1976) and Wolynetz (normal linear model by quasi-EM, 1979a, b). Stirling (1984) has recommended the use of iteratively reweighted least squares for the linear part (i.e., means and regression coefficients) in grouped data models. This would permit fitting these models in GLIM, but difficulties in estimation of the nuisance parameters may render this approach impractical.

7. RECENT TRENDS IN GROUPED DATA MODELING

Models in Survival Analysis

Survival times are inherently continuous, and for the most part statisticians model them that way. Yet times cannot often be observed very exactly. For example, in cancer clinical trials, patients may be reexamined only at, say, 3-month intervals, and often the follow-up times themselves are random. If the end point of interest is time until detectable tumor recurrence, this quantity is known only up to the approximate 3-month gap between the last negative and the first positive follow-up. Data of this kind have come to be known as "interval censored," a name that emphasizes the connection with the common problem of right censoring. Interval and right censoring are not special cases of grouping in a narrow sense of the word, because the partition of the sample space (see Section 2) in censoring need not be fixed. However, if the censoring time distribution or the distribution of censoring interval limits is not related to survival time,

the relevant part of the likelihood is effectively a grouped data likelihood, and so these problems can be treated within the scope of this review

Historically, the robust frequency properties of rank-based methods of inference have made them popular in survival analysis. When data are grouped these methods lead to problems, however, because the sampling distributions of rank-based statistics can be severely distorted by ties (Lehmann, 1975); the coarser is the grouping the more abundant will ties be. Thus it is not surprising that discussions of grouping in survival analysis most often appear in the context of concern about ties.

An englightening example is the effect of grouping on fitting the proportional hazards model with non-time varying covariates. This model relates survivorship X to covariates $\mathbf{Z}$ through the hazard function $\lambda(x \mid \mathbf{z})$. Specifically, the hazard for an individual with covariate $\mathbf{z}$ is assumed to be proportional to a baseline hazard λ_0 as

$$\lambda(x \mid \mathbf{z}, \boldsymbol{\theta}) = \lambda_0(x)\exp(\mathbf{z}\boldsymbol{\theta}). \tag{11}$$

Cox (1972) proposed basing inferences about $\boldsymbol{\theta}$ on a distribution that, in the case of continuous data, is equivalent to the distribution of the ranks of X (Kalbfleisch and Prentice, 1973). This distribution, as it turns out, has no dependence on the nuisance function λ_0. When continuous data are observed, the likelihood based on this distribution (the so-called marginal likelihood) is a product of n factors, one for each of the order statistics of x. If $\mathbf{z}_{(i)}$ is the row vector covariate for the ith failure, and $R(i)$ is the set of indices of failures that occur at or after the ith failure (the risk set of i) then the corresponding factor in the likelihood is

$$l_i(\boldsymbol{\theta}) = \exp(\mathbf{z}_{(i)}\boldsymbol{\theta}) \Big/ \sum_{k \in R(i)} \exp(\mathbf{z}_{(k)}\boldsymbol{\theta}). \tag{12}$$

This likelihood can also be justified as a partial likelihood (Kalbfleisch and Prentice, 1980).

Generalization of the marginal likelihood to the case of grouped continuous data is in principle straightforward but in practice quite difficult. The correct grouped data likelihood is the sum of the complete data likelihoods for each of the possible complete data rank vectors consistent with the observed, tied pattern of ranks. If there are k distinct failure times, where r_j failed at the jth time, $j = 1, \cdots, k$, there are $r_1! \cdot r_2! \cdot \cdots \cdot r_k!$ such rank vectors. The likelihood simplifies to a product of k factors, the jth factor consisting of a sum of $r_j!$ terms. Even without right censoring, which introduces complicated terms involving λ_0, the grouped data marginal likelihood is an unwieldy expression. Approximations to the grouped data marginal likelihood are available, however, including those proposed by Peto (1972), Breslow (1974) and Efron (1977).

Another approach to constructing grouped data likelihoods is the method of partial likelihood (Cox, 1972). A problem with this method is that, in order to get a decent likelihood, one must assume a non-proportional hazards model. Although the grouped data model comes to agree with the proportional hazards model (11) as group size approaches zero (see Thompson, 1977), for any positive group size the partial likelihood estimates are inconsistent for $\boldsymbol{\theta}$ in (11) (Kalbfleisch and Prentice, 1973). The partial likelihood model is also computationally difficult, although Howard (1972) and Gail, Lubin and Rubinstein (1981) have shown how to simplify the calculations.

Prentice and Gloeckler (1978) proposed a third approach to this problem; their idea was to handle the nuisance function λ_0 by reducing it to a finite number of dimensions and then estimating it. They supposed that the time line is divided into intervals $S_j = (a_{j-1}, a_j), j = 1, \cdots, k$, and defined

$$\alpha_j = \exp\left\{-\int_{S_j} \lambda_0(u)\, du\right\}.$$

Then the probability that an individual with covariate $\mathbf{z}$ survives until set S_i and then fails in it is

$$(1 - \alpha_i^{\exp[\mathbf{z}\boldsymbol{\theta}]}) \prod_{j=1}^{i-1} \alpha_j^{\exp[\mathbf{z}\boldsymbol{\theta}]}.$$

Thus, the distribution of failures can be expressed in terms of the finite-dimensional vector of α terms and $\boldsymbol{\theta}$, which can be estimated simultaneously and to which standard asymptotics apply. This method is effective if the data set is large, the number of groups is small and the same grouping sets apply to all subjects. Another method appropriate in this situation is due to Breslow (1974), who assumed the hazard rate to be constant between observed failures. He estimated the hazard function and $\boldsymbol{\theta}$ simultaneously.

An alternative analysis that may prove valuable for grouped and interval censored data was proposed by Lindley (1972). He argued that any hazard function may be changed, by suitable transformation of the time metameter, into a constant hazard, i.e., an exponential model. Lindley's proposal is to fit this model using transformations of the time variable suggested by a range of potential hazard functions, thereby assessing the sensitivity of the inference on $\boldsymbol{\theta}$ to assumptions about the hazard. The idea holds great promise for grouped data analysis because the exponential is one of the easiest to use and best understood of grouped data models (Kulldorff, 1961).

In addition to the work discussed above, there has been progress on other problems in the analysis of grouped survival data, to the point that this area would benefit from a review of its own. Moreau, LeMinor, Myquel and Lellouch (1985) and O'Neill (1985) have discussed testing in the two-sample problem, extending censored data methods to grouped data. Self and Grossman (1986) considered tests based on the marginal distribution of ranks in grouped samples. Turnbull (1974, 1976) has extended the product limit method of estimating a survival function to the case of data that can be censored on the right or left, grouped and truncated. Finkelstein and Wolfe (1985) have incorporated Turnbull's algorithm into a method for relating grouped survival data to covariates.

Models Explicitly for Grouping

Although it has been known since Fisher's time how one should fit models to coarsely grouped data (i.e., what likelihood one ought to compute), until recently this was thought to be an ideal unattainable in all but the simplest cases. Improvements in computing technology, however, have brought many grouped data models within the grasp of the applied statistician. These developments have made possible the assumption of more complex models for grouping behavior, which in turn have led to greater insights into how grouping arises in real data, its importance in inference and how best to handle it. I will illustrate these points with two examples.

Hasselblad, Stead and Galke (1980) presented data on blood lead levels, year, race and age from a large sample of New York children. Lead level is known to be approximately lognormally distributed in homogeneous populations, and so the authors wished to fit linear models relating the logarithm of lead level to the other variables. Analysis was complicated, however, by the fact that the lead data had been recorded in coarse groups; it was not uncommon for half of the children in a given age/race/year bin to fall in the lowest lead level category. Furthermore, the grouping interval pattern had been changed midway through the study, and estimated means and SD's based on interval midpoints were sharply different under the two classification systems. Correction methods were not helpful because the groupings were coarse and uneven on both the raw and log scales.

The authors' solution was to fit the lognormal linear model by maximizing the grouped data log-likelihood as in (6). They found the results of this analysis, unlike the analysis based on interval midpoints, to be insensitive to the grouping scheme used. Estimates of the log-scale SD were reduced when the grouping was taken into account and were more consistent between groups. In light of the large sample size (over 100,000) and the observed sensitivity of their results, they concluded that it was essential to take the grouping seriously.

Heitjan and Reboussin (1987) considered a model for grouped data in which the grouping intervals themselves are unobserved random quantities. They studied a sample of lengths (in months) of post-partum amenorrhea from Guatemalan mothers. A problem with this data set is the appearance of heaps at quarter and full years in the histogram of amenorrhea lengths. A possible explanation for these heaps is that at least some of the mothers rounded their amenorrhea lengths to the nearest quarter or full year, rather than the nearest month. Furthermore, there were hints that the rounding was coarser at the greater amenorrhea lengths, suggesting that the longer the amenorrhea time the more likely it was to be rounded coarsely. There has been some research on rounding of this kind, which is known as heaping (see Ewbank [1981] and references therein), principally in the context of age measurement. However, much of it is concerned with simple smoothing techniques and is not designed to yield insight into patterns of heaping.

The approach adopted by Heitjan and Reboussin was to assume that heaping type behavior (nearest year, nearest quarter year, nearest month) has an ordered categories probit regression on the true length of amenorrhea. In particular, they let heaping type X_g be a normally distributed random variable with mean μ_g and SD σ_g, where $X_g \geq 1$ corresponded to nearest year rounding, $0 \leq X_g < 1$ to nearest quarter year rounding, and $X_g < 0$ to nearest month rounding. To relate heaping to amenorrhea length X_a, they assumed that $\mu_g = \theta_0 + \theta_1 X_a$. The set of possible true values for amenorrhea length and grouping type consistent with a reported length of, say, 12 months is the following region in the (X_a, X_g) plane:

$$(6, 18) \times [1, \infty) \cup (10.5, 13.5) \times [0, 1) \cup (11.5, 12.5) \times (-\infty, 0).$$

This region reflects the fact that a woman reporting an amenorrhea length of 12 months may be a full year reporter or a quarter year reporter or a nearest month reporter. An analogous region was built up for each mother's reported time. The model specification was completed by assuming a marginal distribution for amenorrhea times; the authors tried various transformations to normality.

The authors fit the model using a Newton-Raphson algorithm for the grouped bivariate normal. Grouping coarseness was positively related to amenorrhea length, as expected, and the slope coefficient was significant. The SD of the back-transformed amenorrhea time was affected by taking the grouping into account, and, depending on the nonlinearity of the

transformation, so was the mean. The authors validated their model by comparing it with an analysis of data in which both grouped (retrospective) and ungrouped (prospective) amenorrhea lengths were available.

Besides these, relatively few examples of the serious modeling of grouped data have appeared. Wachter and Trussell's (1982) study of historical data on British Navy recruits fit models to the height variable that assumed grouping to the nearest inch. The model in Heitjan and Reboussin (1987) is equivalent to a model used in the analysis of heaped age data in Heitjan (1985) and Heitjan and Rubin (1986).

Discrete Data Modeling Tools

The grouped continuous model (McCullagh, 1980) is popular in the analysis of ordered categorical data. Because this model assumes that the observed discrete data have arisen by grouping an underlying continuous variate (often the normal or logistic), there are many connections with grouped data methods. The main difference in current practice is that for ordered categories models one usually assumes an ad hoc shape for the underlying distribution, e.g., a unit normal or logistic, then estimates the cutpoints. In grouped data analysis, on the other hand, the cutpoints are known; inference concerns instead the parameters of the underlying distribution.

As an example of the equivalence between grouped data modeling per se and the grouped continuous model for discrete ordinal data, consider a classification into three ordered categories. If F is the standard distribution function of the underlying variate, assumed to belong to a location-scale family, and θ_1 and θ_2 represent the unknown cutpoints, the probabilities of the three categories can be written

$$F(\theta_1),\ F(\theta_2) - F(\theta_1),\ 1 - F(\theta_2).$$

A grouped data model using known cutpoints t_1 and t_2 and assuming location μ and scale σ would result in probabilities

$$F((t_1 - \mu)/\sigma),\quad F((t_2 - \mu)/\sigma) - F((t_1 - \mu)/\sigma),$$
$$1 - F((t_2 - \mu)/\sigma).$$

The parameters of the two models are interchangeable, because

$$\sigma = (t_1 - t_2)/(\theta_1 - \theta_2)$$

and

$$\mu = (\theta_1 t_2 - t_1\theta_2)/(\theta_1 - \theta_2).$$

Equivalence between a grouped data model and an ordered categories model is possible if the number of parameters required to index the underlying distribution is one less than the number of groups. Equations relating the two sets of parameters are not in general easily solved.

8. RESEARCH FRONTIERS

Bayesian Ideas

There has not yet appeared an explicitly Bayesian analysis of an elementary grouped data problem. Yet, as in other areas of statistics, Bayesian ideas are being used to make progress on important problems in grouped data.

One such idea is the notion of multiple imputation, proposed by Rubin (1978, 1987). Recalling the notation of Section 2 and suppressing subscripts where possible, the posterior distribution of parameter $\boldsymbol{\theta}$ based on grouped data $\mathbf{y}$ can be written

$$\begin{aligned} f(\boldsymbol{\theta} \mid \mathbf{y}) &= \int_{X(\mathbf{y})} f(\boldsymbol{\theta} \mid \mathbf{x}, \mathbf{y}) f(\mathbf{x} \mid \mathbf{y})\, d\mathbf{x} \\ &= \int_{X(\mathbf{y})} f(\boldsymbol{\theta} \mid \mathbf{x}) f(\mathbf{x} \mid \mathbf{y})\, d\mathbf{x}. \end{aligned} \tag{13}$$

This means that one way to construct summary inferences for $\boldsymbol{\theta}$ is to average complete data inferences $f(\boldsymbol{\theta} \mid \mathbf{x})$ with respect to the predictive distribution of the complete data given the grouped data, i.e., $f(\mathbf{x} \mid \mathbf{y})$.

Equation (13) also suggests some practical measures one might take when faced with severe grouping or rounding. For example, for a given set of grouped data there may be a natural set of extreme complete data sets, i.e., those located at corners of $X(\mathbf{y})$. By comparing complete data inferences computed at the extreme data sets, it is possible to determine a range of potential sensitivity for final inferences. In the univariate case, extreme data sets for estimating the mean would be those obtained by fixing all the data to be at the left or right edges of their grouping intervals. Extreme data sets for a variance parameter would be those obtained by moving half of the data to the right edges of their intervals and simultaneously moving the other half to the left edges of their intervals. In complex problems this can be a useful exploratory device (Heitjan, 1985).

As another practical application of (13), suppose, as is often the case, that $f(\mathbf{x} \mid \mathbf{y})$, the predictive distribution of the complete data, is not readily available. Equation (13) suggests substituting some density that is easily simulated for the true predictive distribution. By drawing complete data sets from this distribution and averaging the resulting complete data inferences, one can get an approximate posterior that may convey valuable information on the effect of the grouping. For example, Beaton, Rubin and Barone (1976) studied sensitivity of the regression estimates in

the Longley (1967) data by drawing complete data independently and uniformly from their seven-dimensional rounding sets. Although subsequent evidence (Dempster and Rubin, 1983) suggested that the independent, uniform predictive distribution is a poor substitute for $f(\mathbf{x} \mid \mathbf{y})$ for regular $\mathbf{X}$, the investigation did show that errors as the result of numerical inaccuracy in regression routines were a less important source of bias than errors as the result of rounding.

Another means of computing grouped data posteriors is due to Tanner and Wong (1987a, b), who call their method "data augmentation." They rewrote (13) as

$$(14) \qquad f(\boldsymbol{\theta} \mid \mathbf{y}) = \int_{\Omega} K(\boldsymbol{\theta}, \boldsymbol{\theta}') f(\boldsymbol{\theta}' \mid \mathbf{y})\, d\boldsymbol{\theta}',$$

where

$$(15) \qquad K(\boldsymbol{\theta}, \boldsymbol{\theta}') = \int_{X(\mathbf{y})} f(\boldsymbol{\theta} \mid \mathbf{x}, \mathbf{y}) f(\mathbf{x} \mid \mathbf{y}, \boldsymbol{\theta}')\, d\mathbf{x}.$$

Thus, the posterior distribution $f(\boldsymbol{\theta} \mid \mathbf{y})$ is a solution of the integral equation (14). Typically, $f(\boldsymbol{\theta} \mid \mathbf{x}, \mathbf{y})$ is easy to construct, and $f(\mathbf{x} \mid \mathbf{y}, \boldsymbol{\theta})$ is easily simulated. Therefore a way of solving (14) is to select a starting guess $f_{(0)}(\boldsymbol{\theta} \mid \mathbf{y})$ and compute, by Monte Carlo techniques,

$$f_{(1)}(\boldsymbol{\theta} \mid \mathbf{y}) = \int_{X(\mathbf{y})} f(\boldsymbol{\theta} \mid \mathbf{x}, \mathbf{y}) \cdot \left[\int_{\Omega} f(\mathbf{x} \mid \mathbf{y}, \boldsymbol{\theta}') f_{(0)}(\boldsymbol{\theta}' \mid \mathbf{y})\, d\boldsymbol{\theta}' \right] d\mathbf{x},$$

iterating this process until the sequence of working posteriors $f_{(i)}$ converges. Tanner and Wong (1987b) have stated precise conditions under which convergence to the true posterior is guaranteed.

As an example of a way in which data augmentation might be applied, consider Bayesian estimation of a normal mean μ from coarsely grouped data when the variance is known to be one. A starting guess for $f_{(0)}(\mu \mid \mathbf{y})$ might be the normal distribution with mean $\bar{y}^*$ and variance n^{-1}. One data augmentation procedure for this problem would consist of the following five steps:

i) Draw m values of μ, $\mu_j, j = 1, \cdots, m$, from $f_{(0)}(\mu \mid \mathbf{y})$.

ii) For $j = 1, \cdots, m$, draw a data set $\mathbf{x}_{0j}$ by drawing values independently from the normal distribution (with mean μ_j and variance one) conditional on the intervals observed.

iii) Compute $\bar{x}_{0_j}$ to be the mean of the jth imputed data set $\mathbf{x}_{0j}$; then $f(\mu \mid \mathbf{x}_{0j}, \mathbf{y})$ is the normal with mean $\bar{x}_{0j}$ and variance n^{-1}.

iv) Compute the overall augmented data mean $\bar{x}_0 = m^{-1} \sum \bar{x}_{0j}$ and a variance estimate $v_0 = n^{-1} + (m-1)^{-1} \sum (\bar{x}_{0j} - \bar{x}_0)^2$.

v) Set $f_1(\mu \mid \mathbf{y})$ to be the average of the complete-data posteriors computed in step (iii), i.e., a normal with mean $\bar{x}_0$ and variance v_0.

Iteration of these steps will result in convergence to a normal approximation to the posterior of μ. A more realistic approach might use t distributions for the series of posteriors $f_{(i)}$, whose parameters could be found by matching the posterior mean, variance and higher moments.

In many problems, computation of the grouped data log-likelihood and its derivatives either analytically, numerically or by some mixture of the two is feasible, and this information can be combined with a prior distribution to produce the posterior of $\boldsymbol{\theta}$. Sometimes several approximations based on matching derivatives to the parameters of familiar distributions are available. Laird and Louis (1982) have formalized these ideas in the context of missing data problems, although not specifically for grouped data.

Research Opportunities

In this paper I have indicated some of the major unsolved questions in grouped data theory and practice, including i) regularity conditions and the likelihood basis for Sheppard's corrections (Dempster and Rubin, 1983); ii) frequentist asymptotics in grouped data likelihood analysis (Kulldorff, 1961); iii) concavity of grouped data log-likelihoods (Burridge, 1981a, 1982); and iv) Bayesian analysis of grouped survivorship data (Lindley, 1972). Two other promising areas are v) the recovery of model diagnostic information from grouped samples (Pettitt, 1985); and vi) analysis of heaping and other nonignorable grouping phenomena (Heitjan, 1985; Heitjan and Rubin, 1986; Heitjan and Reboussin, 1987). Thanks to the incompleteness of the general theory, almost any other area that I have mentioned would benefit from further research.

9. CONCLUSIONS

Statisticians have long known that their models do not perfectly mirror reality. In the past most attention has been fixed on such problems as what to do with data that are more long tailed or prone to outliers than standard models imply. The motivation for this sort of analysis was to develop procedures that would give reliable inferences in the absence of strong knowledge of the process generating the data. By contrast, the study of grouping is aimed at ensuring comparability of results from different data sets. The goal is to liberate model-based inferences from the distorting effects of grouping.

When intervals are small enough, grouping makes little difference, and simple estimates are essentially valid. This is the case, for example, in the grouped data studied by Beckman and Johnson (1987). However, in the studies of Beaton, Rubin and Barone (1976), Hasselblad, Stead and Galke (1980), Heitjan and Rubin (1986) and Heitjan and Reboussin (1987), it was essential to account for grouping. Large, unevenly spaced intervals, uncertain group widths, highly multicollinear data and large samples are indicators of a need for careful attention to grouping.

Classical frequency statistics has made relatively little progress on grouping, probably because it has been considered either too difficult or insufficiently important to merit much serious attention. The major results are that moment-based statistics are inconsistent, but that maximum likelihood, at least in the normal and exponential cases, is consistent and efficient. Recently the Cox model and the various likelihoods used to fit it have been examined. Although progress in this area too has been hindered by analytical and computational difficulties, several satisfactory alternatives have been proposed and are finding their way into practice.

By contrast, Bayesian solutions to grouping problems do not require the development of special theory, although the specific models one uses are usually dependent on the available computing software and hardware. A Bayesian's approach might simply be to fit candidates from a range of potential models by computing and summarizing the correct likelihoods, combining the answers so obtained with prior information. For grouped data, this approach has only recently become practical and has not yet been attempted on a significant scale. Extrapolating from the recent past, however, it appears that applied Bayesian analysis for these problems will become easier at a rate that will soon be considered astounding. Statistically motivated numerical tools like imputation, data augmentation and importance sampling should contribute to progress on this objective.

ACKNOWLEDGMENTS

Support for this research was provided by the Warner-Lambert Research Institute, by the National Science Foundation under Grants MCS-80-02217 and MCS-83-01459 and by the National Cancer Institute under United States Public Health Service Grant CA-16042.

REFERENCES

ABERNETHY, J. R. (1933). On the elimination of systematic errors due to grouping. *Ann. Math. Statist.* **4** 263–277.

AIGNER, D. J. and GOLDBERGER, A. S. (1970). Estimation of Pareto's law from grouped observations. *J. Amer. Statist. Assoc.* **65** 712–723.

BATEN, W. D. (1931). Correction for the moments of a frequency distribution in two variables. *Ann. Math. Statist.* **2** 309–319.

BEATON, A. E., RUBIN, D. B. and BARONE, J. L. (1976). The acceptability of regression solutions: Another look at computational accuracy. *J. Amer. Statist. Assoc.* **71** 158–168.

BECKMAN, R. J. and JOHNSON, M. E. (1987). Fitting the Student t distribution to grouped data, with application to a particle scattering experiment. *Technometrics* **29** 17–22.

BENN, R. T. and SIDEBOTTOM, S. (1976). Algorithm AS 95. Maximum likelihood estimation of location and scale parameters from grouped data. *Appl. Statist.* **25** 88–93.

BERKSON, J. (1950). Are there two regressions? *J. Amer. Statist. Assoc.* **45** 164–180.

BOARDMAN, T. J. (1973). Estimation in compound exponential failure models—when the data are grouped. *Technometrics* **15** 271–277.

BRASCAMP, H. J. and LIEB, E. H. (1976). On extensions of the Brunn-Minkowski and Prekopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *J. Funct. Anal.* **22** 366–389.

BRESLOW, N. E. (1974). Covariance analysis of censored survival data. *Biometrics* **30** 89–99.

BURRIDGE, J. (1981a). A note on maximum likelihood estimation for regression models using grouped data. *J. Roy. Statist. Soc. Ser. B* **43** 41–45.

BURRIDGE, J. (1982). Some unimodality properties of likelihoods derived from grouped data. *Biometrika* **69** 145–151.

CARVER, H. C. (1936). The fundamental nature and proof of Sheppard's adjustments. *Ann. Math. Statist.* **7** 154–163.

CHAO, S. T. (1975). A note on Sheppard's corrections for grouping. *Tamkang J. Math.* **6** 55–60.

COCHRAN, W. G. (1968). Errors of measurement in statistics. *Technometrics* **10** 637–666.

CORNISH, E. A. and FISHER, R. A. (1937). Moments and cumulants in the specification of distributions. *Rev. Internat. Statist. Inst.* **5** 307–320.

COX, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34** 187–220.

CRAIG, C. C. (1936). Sheppard's corrections for a discrete variable. *Ann. Math. Statist.* **7** 55–61.

CRAIG, C. C. (1941). A note on Sheppard's corrections. *Ann. Math. Statist.* **12** 339–345.

DEKEN, J. G. (1983). Approximating conditional moments of the multivariate normal distribution. *SIAM J. Sci. Statist. Comput.* **4** 720–732.

DEMPSTER, A. P., LAIRD, N. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. Ser. B* **39** 1–38.

DEMPSTER, A. P. and RUBIN, D. B. (1983). Rounding error in regression: The appropriateness of Sheppard's corrections. *J. Roy. Statist. Soc. Ser. B* **45** 51–59.

DON, F. J. H. (1981). A note on Sheppard's corrections for grouping and maximum likelihood estimation. *J. Multivariate Anal.* **11** 452–458.

DURBIN, J. (1954). Errors in variables. *Rev. Internat. Statist. Inst.* **22** 23–32.

EFRON, B. (1977). The efficiency of Cox's likelihood function for censored data. *J. Amer. Statist. Assoc.* **72** 557–565.

EISENHART, C. (1947). Effect of rounding or grouping data. In *Selected Techniques of Statistical Analysis* (C. Eisenhart, M. W. Hastay and W. A. Wallis, eds.). McGraw-Hill, New York.

ELDERTON, W. P. (1933). Adjustments for the moments of J-shaped curves. *Biometrika* **25** 179–180.

EWBANK, D. C. (1981). *Age Misreporting and Age-Selective Underenumeration: Sources, Patterns, and Consequences for Demographic Analysis.* National Academy Press, Washington.

FINKELSTEIN, D. M. and WOLFE, R. A. (1985). A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics* **41** 933–945.

FISHER, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London Ser. A* **222** 309–368.

FISHER, R. A. (1936). The use of multiple measurements in taxonomic problems. *Ann. Eugenics* **8** 179–188.

FLYGARE, M. E., AUSTIN, J. A. and BUCKWALTER, R. M. (1985). Maximum likelihood estimation for the 2-parameter Weibull distribution based on interval-data. *IEEE Trans. Reliability* **34** 57–59.

FRYER, J. G. and PETHYBRIDGE, R. J. (1972). Maximum likelihood estimation of a linear regression function with grouped data. *Appl. Statist.* **21** 142–154.

GAIL, M. H., LUBIN, J. H. and RUBINSTEIN, L. V. (1981). Likelihood calculations for matched case-control studies and survival studies with tied death times. *Biometrika* **68** 703–707.

GJEDDEBAEK, N. F. (1949). Contributions to the study of grouped observations. Application of the method of maximum likelihood in case of normally distributed observations. *Skandinavisk Aktuarietidskrift* **32** 135–159.

GJEDDEBAEK, N. F. (1956). Contributions to the study of grouped observations. II. Loss of information caused by grouping of normally distributed observations. *Skandinavisk Aktuarietidskrift* **39** 154–159.

GJEDDEBAEK, N. F. (1957). Contributions to the study of grouped observations. III. The distribution of estimates of the mean. *Skandinavisk Aktuarietidskrift* **40** 20–25.

GJEDDEBAEK, N. F. (1959a). Contributions to the study of grouped observations. IV. Some comments on simple estimates. *Biometrics* **15** 433–439.

GJEDDEBAEK, N. F. (1959b). Contributions to the study of grouped observations. V. Three-class grouping of normal observations. *Skandinavisk Aktuarietidskrift* **42** 194–207.

GJEDDEBAEK, N. F. (1961). Contributions to the study of grouped observations. VI. *Skandinavisk Aktuarietidskrift* **44** 55–73.

GJEDDEBAEK, N. F. (1968). Statistical analysis: Grouped observations. In *International Encyclopedia of the Social Sciences* (D. R. Sills, ed.) **15**. Macmillan and the Free Press, New York.

GNANADESIKAN, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations*. Wiley, New York.

HAITOVSKY, Y. (1973). *Regression Estimation from Grouped Observations*. Griffin, London.

HAITOVSKY, Y. (1982). Grouped data. *Encyclopedia of Statistical Sciences* **3** 527–536. Wiley, New York.

HARTLEY, H. O. (1950). A simplified form of Sheppard's correction formulae. *Biometrika* **37** 145–148.

HASSELBLAD, V., STEAD, A. G. and GALKE, W. (1980). Analysis of coarsely grouped data from the lognormal distribution. *J. Amer. Statist. Assoc.* **75** 771–778.

HEITJAN, D. (1985). Analysis of a set of coarsely grouped data. Ph.D. dissertation, Univ. Chicago.

HEITJAN, D. (1987). Maximum likelihood from grouped bivariate normal data. Unpublished manuscript.

HEITJAN, D. F. and REBOUSSIN, D. M. (1987). A model for coarse grouping with application to survey data. *Proc. Survey Res. Methods Sec.* 271–276. Amer. Statist. Assoc., Washington.

HEITJAN, D. F. and RUBIN, D. B. (1986). Inference from coarse data using multiple imputation. *Computer Science and Statistics: Proceedings of the 18th Symposium on the Interface* (T. J. Boardman, ed.) 138–143. Amer. Statist. Assoc., Washington.

HOLLAND, B. S. (1975). Some results on the discretization of continuous probability distributions. *Technometrics* **17** 333–339.

HOWARD, S. (1972). Discussion of "Regression models and life tables" by D. R. Cox. *J. Roy. Statist. Soc. Ser. B* **34** 210–211.

INDRAYAN, A. and RUSTAGI, J. S. (1979). Approximate maximum likelihood estimates in regression models for grouped data. In *Optimizing Methods in Statistics* (J. S. Rustagi, ed.). Academic, New York.

KALBFLEISCH, J. D. and PRENTICE, R. L. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika* **60** 267–278.

KALBFLEISCH, J. D. and PRENTICE, R. L. (1980). *The Statistical Analysis of Failure Time Data*. Wiley, New York.

KENDALL, M. G. (1938). The conditions under which Sheppard's corrections are valid. *J. Roy. Statist. Soc.* **101** 592–605.

KULLBACK, S. (1935). A note on Sheppard's corrections. *Ann. Math. Statist.* **6** 158–159.

KULLDORFF, G. (1961). *Contributions to the Theory of Estimation from Grouped and Partially Grouped Samples*. Almqvist and Wiksell, Stockholm.

LAIRD, N. M. and LOUIS, T. A. (1982). Approximate posterior distributions for incomplete data problems. *J. Roy. Statist. Soc. Ser. B* **44** 190–200.

LANGDON, W. H. and ORE, O. (1930). Semi-invariants and Sheppard's correction. *Ann. of Math. (2)* **31** 230–232.

LEHMANN, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day, San Francisco.

LINDLEY, D. V. (1950). Grouping corrections and maximum likelihood equations. *Proc. Cambridge Philos. Soc.* **46** 106–110.

LINDLEY, D. V. (1972). Discussion of "Regression models and life tables" by D. R. Cox. *J. Roy. Statist. Soc. Ser. B* **34** 208–209.

LONGLEY, J. W. (1967). An appraisal of least squares programs for the electronic computer from the point of view of the user. *J. Amer. Statist. Assoc.* **62** 819–841.

MARTIN, E. S. (1934). On corrections for the moment coefficients of frequency distributions when the start of the frequency is one of the characteristics to be determined. *Biometrika* **26** 12–58.

MCCULLAGH, P. (1980). Regression models for ordinal data (with discussion). *J. Roy. Statist. Soc. Ser. B* **42** 109–142.

MOREAU, T., LEMINOR, M., MYQUEL, P. and LELLOUCH, J. (1985). Estimation of the hazards ratio in two grouped samples. *Biometrics* **41** 245–252.

MURRAY, G. D. (1977). Discussion of "Maximum likelihood from incomplete data via the EM algorithm" by A. P. Dempster, N. Laird and D. B. Rubin. *J. Roy. Statist. Soc. Ser. B* **39** 27–28.

O'NEILL, T. J. (1985). A test of equality of survival distributions for two sample censored grouped survival data. *Biometrical J.* **27** 697–705.

PAIRMAN, E. and PEARSON, K. (1919). On correcting for the moment-coefficients of limited range frequency-distributions when there are finite or infinite ordinates and any slopes at the terminals of the range. *Biometrika* **12** 231–258.

PEARSE, G. E. (1928). On corrections for the moment-coefficients of frequency distributions when there are infinite ordinates at one or both terminals of the range. *Biometrika* **20A** 314–355.

PEARSON, K. (1902). On the systematic fitting of curves to observations and measurements. *Biometrika* **1** 265–303.

PETHYBRIDGE, R. J. (1973). Maximum likelihood estimation of a polynomial regression function with grouped data. *Appl. Statist.* **22** 203–212.

PETHYBRIDGE, R. J. (1975). Maximum likelihood estimation of a linear regression function with grouped data. *Appl. Statist.* **24** 28–41.

PETO, R. (1972). Discussion of "Regression models and life tables" by D. R. Cox. *J. Roy. Statist. Soc. Ser. B* **34** 205–207.

PETTITT, A. N. (1985). Re-weighted least squares estimation with censored and grouped data: An application of the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **47** 253–260.

PIERCE, J. A. (1940). A study of a universe of n finite populations with application to moment-function adjustments for grouped data. *Ann. Math. Statist.* **11** 311–334.

PIERCE, J. A. (1943). Correction formulas for moments of a grouped-distribution of discrete variates. *J. Amer. Statist. Assoc.* **38** 57–62.

PRATT, J. W. (1981). Concavity of the log-likelihood. *J. Amer. Statist. Assoc.* **76** 103–106.

PREECE, D. A. (1981). Distribution of final digits in data. *The Statistician* **30** 31–60.

PREKOPA, A. (1973). On logarithmic concave measures and functions. *Acta Sci. Math. (Szeged)* **34** 335–343.

PRENTICE, R. L. and GLOECKLER, L. A. (1978). Regression analysis of grouped survival data with applications to breast cancer data. *Biometrics* **34** 57–67.

RUBIN, D. B. (1978). Multiple imputation in sample surveys—a phenomenological Bayesian approach to nonresponse. *Proc. Survey Res. Methods Sec.* Amer. Statist. Assoc., Washington.

RUBIN, D. B. (1987). *Multiple Imputation for Nonresponse in Sample Surveys and Censuses.* Wiley, New York.

SANDON, F. (1924). Note on the simplification of the calculation of abruptness coefficients to correct crude moments. *Biometrika* **16** 193–195.

SCHADER, M. and SCHMID, F. (1984). Computation of maximum likelihood estimates for μ and σ from a grouped sample of a normal population. A comparison of algorithms. *Statist. Hefte* **25** 245–258.

SELF, S. G. and GROSSMAN, E. A. (1986). Linear rank tests for interval-censored data with application to PCB levels in adipose tissue of transformer repair workers. *Biometrics* **42** 521–530.

SHEPPARD, W. F. (1898). On the calculation of the most probable values of frequency constants for data arranged according to equidistant divisions of a scale. *Proc. London Math. Soc.* **29** 353–380.

STEVENS, W. L. (1948). Control by gauging (with discussion). *J. Roy. Statist. Soc. Ser. B* **10** 54–98.

STIRLING, W. D. (1984). Iteratively reweighted least squares for models with a linear part. *Appl. Statist.* **33** 7–17.

STOER, J. and BULIRSCH, R. (1980). *Introduction to Numerical Analysis.* Springer, New York.

STUDENT (1908). The probable error of a mean. *Biometrika* **6** 1–25.

SWAN, A. V. (1969). Algorithm AS 16. Maximum likelihood estimation from grouped and censored normal data. *Appl. Statist.* **18** 110–114.

TALLIS, G. M. (1967). Approximate maximum likelihood from grouped data. *Technometrics* **9** 599–606.

TALLIS, G. M. and YOUNG, S. S. (1962). Maximum likelihood estimation of parameters of the normal, log-normal, truncated normal and bivariate normal distributions from grouped data. *Austral. J. Statist.* **4** 49–54.

TANNER, M. A. and WONG, W. H. (1987a). An application of imputation to an estimation problem in grouped lifetime analysis. *Technometrics* **29** 23–32.

TANNER, M. A. and WONG, W. H. (1987b). The calculation of posterior distributions by data augmentation (with discussion). *J. Amer. Statist. Assoc.* **82** 528–540.

THOMPSON, W. A. (1977). On the treatment of grouped observations in life studies. *Biometrics* **33** 463–470.

TOCHER, K. D. (1949). A note on the analysis of grouped probit data. *Biometrika* **36** 9–17.

TURNBULL, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *J. Amer. Statist. Assoc.* **69** 169–173.

TURNBULL, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *J. Roy. Statist. Soc. Ser. B* **38** 290–295.

WACHTER, K. W. and TRUSSELL, J. (1982). Estimating historical heights. *J. Amer. Statist. Assoc.* **77** 279–293.

WOLD, H. (1934). Sheppard's correction formulae in several variables. *Skandinavisk Aktuarietidskrift* **17** 248–255.

WOLYNETZ, M. S. (1979a). Algorithm AS 138. Maximum likelihood estimation from confined and censored samples. *Appl. Statist.* **28** 185–195.

WOLYNETZ, M. S. (1979b). Algorithm AS 139. Maximum likelihood estimation in a linear model from confined and censored normal data. *Appl. Statist.* **28** 195–206.

YONEDA, K. and UCHIYAMA, M. (1956). Some estimations in the case of relatively large class intervals. *Yokohama Math. J.* **4** 99–118.

Comment

James Burridge

Heitjan's paper is a useful and interesting survey of the current state of the art regarding "grouped data." Grouping is, as Heitjan says, "ubiquitous." Yet all of us have been brought up on statistical theory and methods intended to deal with "continuous" data—data that none of us will ever see! Justifications for such a perverse situation are of course that it is usually convenient to treat the data as if they were continuous and, often, that the grouping is fine enough for any necessary corrections to be ignorable. There remains of course the grey area where it is not clear whether or not adjustments ought to be used. It is irritating in practice to have, on occasion, to worry about such things. Perhaps, in the near future hopefully, authors of statistical packages will enable us to analyze grouped data as a matter of routine. Certainly the continuing advances in computer processor power are making it increasingly feasible, if not desirable, to analyze the data that are actually observed. However, much of the conventional elegant theory of mathematical statistics may seem less compelling if we routinely adopt such a view: I wonder, for example, whether many results associated with sufficiency may ultimately be seen as mathematical curiosities or, at best, as approximations.

James Burridge is a Lecturer in the Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, England.