A SYSTEMATIC APPROACH TO GRAPHICAL METHODS IN BIOMETRY

M. TARTER and S. RAMAN
UNIVERSITY OF CALIFORNIA, BERKELEY

1. Introduction

This paper deals with certain new graphical techniques which may be of value in exploratory biometry. In two senses, emphasis is placed upon the systematization of graphical procedures. One, a new theoretical result is obtained which gives conditions under which nonparametric histogram procedures of Parzen [11], Rosenblatt [13], Watson and Leadbetter [25], as well as others, can be treated as a special case of Fourier series methods of Cencov [2], Tarter and Kronmal [8], [19], [22], and Watson [24]. Two, by utilizing alternative weighted Fourier series, most hitherto considered graphical procedures such as the histogram, scatter diagram, and cumulative polygon are placed within a single computational framework. This systematization is shown to provide a researcher with both a comprehensive as well as a statistically and computationally efficient approach to graphical data analysis.

In Section 6 of this paper, an example of the graphical display of biomedical data is presented. The bivariate case, for example, generalizations of the scatter diagram, is considered in detail and the biomedical variable pair, bone age and chronological age, is used to demonstrate the application of this new graphical procedure.

Before proceeding to the sections of this paper that deal with the systematization and exemplification of graphical methods in biometry, it may be worthwhile to offer a brief explanation concerning what we consider to be the particular relevance of the *new* graphical procedures to biometry. By way of contrast, the following quotation ([1], p. 1), provides a clear exposition of the purpose underlying what might be called the *old* graphical procedures:

"Time after time it happens that some ignorant or presumptuous member of a committee or a board of directors will upset the carefully-thought-out plan of a man who knows the facts, simply because the man with the facts cannot present his facts readily enough to overcome the opposition. It is often with impotent exasperation that a person having the knowledge sees some fallacious

Research supported in part by small grant to faculty—Dr. M. Tarter General Research Support Grant 5 S01 RR5441-10.

conclusion accepted, or some wrong policy adopted, just because known facts cannot be marshalled and presented in such manner as to be effective."

This quotation clearly indicates that the primary function the older graphical methods were usually designed to fulfill was the summarization of information once this information had been obtained.

It is the contention of the authors that while the older methods emphasized the *expository*, the newer graphical methods place an emphasis on the *exploratory*. While it is certainly important for the biometrician or biostatistician to be able to *present* his "conclusions," the process of *reaching* these conclusions would now seem to be an equally important application of graphics.

Unquestionably, the major impetus leading to exploratory graphics has been the availability of high speed digital computation. In particular, recent developments involving the transmission of digitized graphical information over voice grade phone lines may substantially increase the potential for graphical presentation. Unlike the now fairly common IBM 2250 graphics configuration, which usually requires cathode ray tube (CRT) display units to be located within one hundred feet of a large central computer, the new IMLAC and other terminals will make interactive graphics economical for those with access only to a small computer or to a distant time shared computer system.

Before providing details of several new biostatistical graphic methods, one additional comment should be made. In a context where a new and substantially more powerful means of implementation becomes available, it may be of value to thoroughly reconsider traditional methods and the modes of thinking which engendered them and were engendered by them. The histogram, fractile diagram, scatter diagram, and other graphical tools were devised to meet certain specific goals and to cope with a narrow range of practical limitations. Today it is rarely necessary to group continuous data as a preliminary to the computation of sample moments (and then apply Sheppard's corrections). The construction of the traditional histogram based on the division of the range of the random variable into class intervals may be similarly reconsidered.

In the next section, the recent evolution of the histogram will be described. Fortunately, in this situation the methods which in our opinion are the most suitable for biostatistical graphics will be shown to include histogram procedures as a special case. However, it may not always be true that an elaboration of a conventional procedure is the most suitable alternative when matched with new means of implementation. For example, in Section 4, where an alternative to the fractile diagram is considered briefly, and in Section 6 which primarily concerns alternatives to the scatter diagram, it appears that the traditional graphical methods may be supplanted or at least supplemented by substantially different procedures.

2. Nonparametric and series graphical procedures

In this section, the symbolic or mathematical substructure of certain new graphical procedures will be presented. Following an heuristic introduction to the generalized histogram, a nonparametric procedure, introduced in [13] and investigated in [11], [25], [22], and others, it will be shown that for most practical purposes generalized histogram type nonparametric procedures can be treated as special cases of the series procedures, investigated in [2], [8], [22], and [24], as well as Schwartz [14]. Although this identity between nonparametric and series procedures was previously mentioned in [8], it was not explicitly presented, primarily because of the lack of a practical method for implementing this result. However, new procedures devised by Tarter and Fellner [18] have recently been found to make it practical and desirable to consider nonparametric procedures from a series point of view.

The process of constructing a typical "old style" histogram can be considered as consisting of two separate steps. In Step 1 (see Figure 1), the domain of

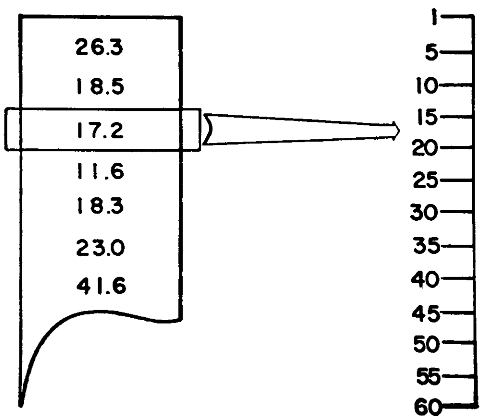


FIGURE 1

Step 1.

interest is divided into equally spaced class intervals and it is ascertained into which interval each data point is to be allocated. In Step 2 (see Figure 2), for each data point X_j, $j = 1, \cdots, n$, a *rectangular* block, with width equal to the class interval length h and with height equal to $(nh)^{-1}$ is added to the pile of blocks already piled within the class interval which contains X_j. It is apparent that Step 1, which is the geometrical analog of moment computation using a frequency table of grouped data, is necessary only if hand rather than machine calculation is used. It is more sensible and efficient to center the jth data point at the exact value assumed by X_j. The resulting irregularly packed pile of blocks can be easily "compressed" numerically by even the smallest digital computer (see Figure 3).

Once one revises Step 1, it is a simple matter to generalize Step 2 and consider alternatives to the rectangular shape of the blocks or counters that compile the contribution of the individual data points to the final graph of the density estimator $\hat{f}$ (see Figure 4).

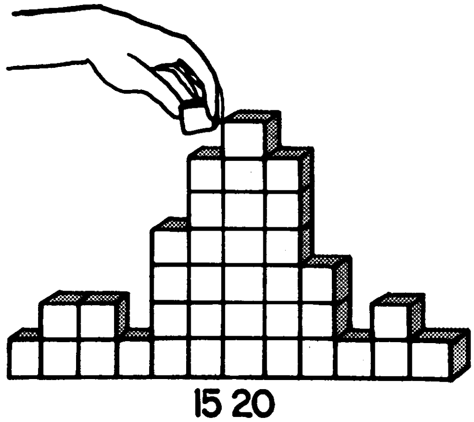


FIGURE 2
Step 2.

At this point, it is advisable to switch from graphical to symbolic expression. Let $f(x)$ represent the joint probability density function of the p dimensional random variate column vectors $\{X_j\}, j = 1, \cdots, n$, where the X_j will be assumed to be independent and identically distributed. Let k represent an arbitrary column vector of integers or p-tuples and let N represent the set of all k. For a fixed sample size n, a generalized histogram $\hat{f}$, that is, nonparametric estimator of f [13], can be expressed as

$$\hat{f}(x) = \frac{1}{n} \sum_{j=1}^{n} \delta_n(x - X_j). \tag{1}$$

If Step 1 but not Step 2 is revised so that the rectangular shape of each counter is retained but the jth counter is centered at X_j, then

$$\delta_n(z) = \frac{I_H(z)}{h^p}, \tag{2}$$

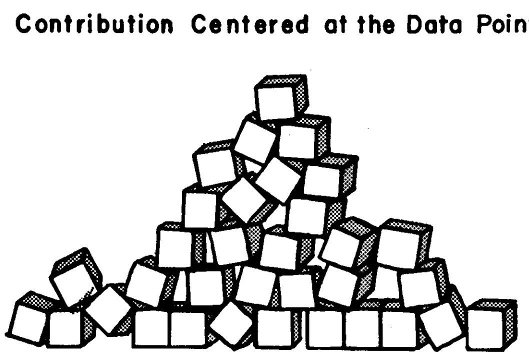
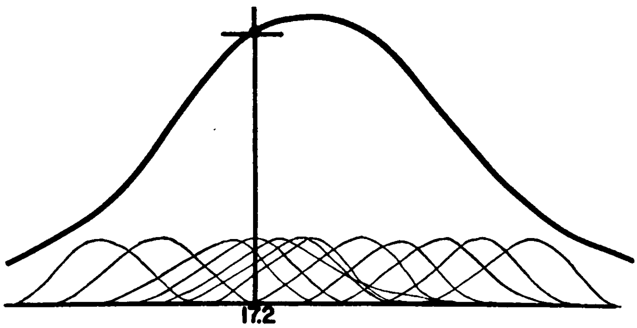


FIGURE 3
Improving Step 1.



Contribution Distributed Over the Entire Range

FIGURE 4

Improving Step 2.

where H is a p dimensional rectangular parallelopiped within space E^p with diagonal corners $\pm(\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_p)$, where $\tilde{x}_k = h/2$ for all k and I_H represents the indicator function of H.

We will now show that a broad class of nonparametric estimators can be expressed as series estimators. The latter were introduced independently in [2], [8], [22], as well as [24]. This result is somewhat surprising since at least two authors, Schwartz [14] and Wegman [26], have stressed comparisons between nonparametric and series estimators and, hence, given the impression that there are fundamental mathematical and statistical differences between them. Theorem 1 tends to indicate that for almost all purposes, nonparametric and series estimators can be treated as being different forms of the same estimator. Thus, in our opinion, the choice between the two alternatives should be made solely on the basis of computational criteria.

THEOREM 1. *Assume that we are interested in the estimation of the multivariate density f over a finite subregion of the p dimensional Euclidean space. (Without loss of generality we will define the support of f to be the hypercube*

$$U \equiv \{X: -\tfrac{1}{4} < X_s \leqq \tfrac{1}{4},\ s = 1, \cdots, p\}, \tag{3}$$

where X_s is the sth *component of X.) Assume that the p dimensional nonparametric kernel, as defined in expression* (1) *has a uniformly convergent Fourier expansion on the hypercube*

$$V = \{X: -\tfrac{1}{2} < X_s < \tfrac{1}{2},\ s = 1, \cdots, p\} \tag{4}$$

of the form

$$\delta_n(X) \equiv \sum_{k \in N} b_k \exp\{2\pi i k'(X - R)\}, \tag{5}$$

where $R' = (-\tfrac{1}{2}, \cdots, -\tfrac{1}{2})$. Then at every point X of the support region of f, the "nonparametric" estimator $\hat{f}$ defined by expression (1) *is identical to the "series" estimator*

$$\sum_{k \in N} \hat{B}_k b_k \exp \{2\pi i k'(X - R)\}, \tag{6}$$

where

$$\hat{B}_k = n^{-1} \sum_{j=1}^{n} \exp \{-2\pi i k'(X_j - R)\}. \tag{7}$$

The proof of Theorem 1 is obtained through simple algebraic substitution and interchange of the order of summation. It is, of course, not necessary to expand δ about the point R. However, expansion about R helps to establish the identity between expression (7) and the definition of $\hat{B}_k$ given in [22]. (In the remaining sections of this paper we will use the earlier definition, that is, omit R and assume V to be the usual unit hypercube.) It might also be noted that if the function $\tilde{f}$ is defined as

$$\tilde{f}(X) = \begin{cases} 1/n & \text{if } X = X_j, \ j = 1, \cdots, n, \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

then the Fourier series associated with $\tilde{f}$ is

$$\tilde{f}(X) \sim \sum_{k \in N} \hat{B}_k \exp \{2\pi i k'(X - R)\} \tag{9}$$

and

$$\hat{f}(X) = \int_V \tilde{f}(Z)\, \delta_n(X - Z)\, dZ, \tag{10}$$

where the integral is taken in the Lebesgue-Stieltjes sense. Hence, $\hat{f}$ can be considered to be the convolution of $\tilde{f}$ with δ_n.

Now define the Fourier coefficient of the density f as

$$B_k = \int_V \exp \{-2\pi i k'(X - R)\} f(X)\, dX \tag{11}$$

and the usual goodness of fit criterion [22], [25], mean integrated square error (MISE), as

$$J = E \int_V |f(X) - \hat{f}(X)|^2\, dX. \tag{12}$$

By simple algebraic manipulation as in [24] and [18], one finds that the MISE of nonparametric-series estimator $\hat{f}$ equals

$$J(f, \hat{f}) = \sum_{k \in N} \{n^{-1}|b_k|^2(1 - |B_k|^2) + |1 - b_k|^2|B_k|^2\}. \tag{13}$$

Consequently, for f and δ as defined in the statement of Theorem 1, the problem of optimal choice of the "best" kernel δ is identical to the problem of choice of "best" weights, $b_k = b_k^{\text{opt}}$, in expression (6). This problem has been considered in [24] as well as [18] with the result that

$$b_k^{\text{opt}} = \frac{n|B_k|^2}{1 + (n-1)|B_k|^2}. \tag{14}$$

The estimation of b_k^{opt} is considered separately by Fellner and Tarter in [18], where an estimator

$$\hat{b}_k^{\text{opt}} = \frac{n}{n-1} - \frac{1}{(n-1)|\hat{B}_k|^2} \tag{15}$$

is derived as an analog to the inclusion rule for the choice of $b_k = 0$ or 1 investigated in [22].

3. Computational considerations

It is not usually feasible to directly apply the series estimator derived in Theorem 1. For biomedical and other applications, graphical procedures must be considered from computational as well as statistical points of view. While optimum weights (14) are estimable from given data, it is impractical to compute more than a moderate number of these coefficients. On the other hand, the estimators considered in [22] result in a very simple inclusion rule, namely, include the coefficient B_k if and only if

$$|B_k|^2 > \frac{1}{n+1}, \tag{16}$$

which is the dichotomous analog of weight (14) or, in the usual situation where B_k is unknown, use

$$|\hat{B}_k|^2 > \frac{2}{n+1}, \tag{17}$$

which is the dichotomous analog of weight (15).

Furthermore, since the asymptotically optimal kernel is the Dirac δ function (see [24]), the above procedure approaches optimality, that is, the MISE approaches zero as n approaches infinity. Investigations with real data by Raman [12] indicate that satisfactory results may be obtained by using the above technique with sample size as low as 45 pairs of observations to estimate certain bivariate distributions.

In contradistinction to series methods, the use of nonparametric estimates of form (1) for graphical purposes requires that the entire file of data points be either stored or reentered into the computer in order to graph $\hat{f}$ at each specific value of x. Since the estimator $\hat{B}_k$ and the inclusion rule (17) are symmetric functions of the observations, there is no need to allocate memory space in the computer for the storage of data. On heuristic considerations, we propose as a stopping rule the termination of the search for a subsequence of coefficients (in the bivariate case) as soon as two consecutive coefficients become negligible in a horizontal and vertical scan of the array of coefficients.

Since the inclusion rule is the same for B_k and B_{-k}, it becomes necessary to compute and store the real and imaginary parts of only half the total number of coefficients. This again results in saving of computer memory and time.

To further optimize the running time, we test the coefficients by the inclusion and stopping rule after reading each set of 100 data points. Since the observations are assumed independent, these intermediate estimates are unbiased. However, to be conservative we perform the test with the final sample size substituted for n in the right side of inequality (17).

In a typical example with 1000 pairs of observations from a 50 per cent mixture of Gaussian distributions, the computation of the Fourier coefficients took about nine minutes on an IBM 1130. The computational advantage of the method will be evident by noting that in the above situation, the first 100 observations took three minutes, the second 100 took one and a half minutes, the third 100 took one minute, the fourth and the rest took half a minute each. Moreover, the economy of summarizing the characteristic density by the series technique can be appreciated by noting that the number of coefficients needed in the above situation was only 23. Studies performed by Kronmal and Tarter [8] have shown that the number of parameters required in the univariate case is less than fifteen for most distributions, sample sizes, and intervals of estimation.

Besides the advantage of series forms that is related to the condensation of statistical information into a relatively few estimators, one might note a second closely related computational property possessed by most series procedures and, in particular, by expression (6). The statistics $\hat{B}_k$ are symmetric functions of the observations and, hence, can be computed iteratively.

The utility of the iterative computation of $\hat{B}_k$ is best illustrated by the application of univariate modifications of expression (6) to the problem of estimating a cumulative distribution function. Consider that the process of graphing the sample cumulative is almost identical to the process of ranking the data points. This has led to investigations [8] and [21] which deal with the use of series estimators as replacements for the sample cumulative.

In the next section, a general result is derived for which a special case, related to estimation of the cumulative distribution function, has been considered by Kronmal and Tarter [8]. It has been shown ([8], Section 7) that the same subset of indices M, minimizes the MISE of estimator $\hat{F}_M$ that minimizes the MISE of density estimator $\hat{f}_M$. Specifically, in the notation of this paper, letting $p = 1$, and $\bar{x}$ represent the sample mean, and defining density estimator $\hat{f}_M$ and cumulative estimator $\hat{F}_M$, respectively, as

$$\hat{f}_M(x) = 1 + \sum_{k \in M} \hat{B}_k \exp\{2\pi ikx\} \tag{18}$$

and

$$\hat{F}_{M'}(x) = (\tfrac{1}{2} + x - \bar{x}) + \sum_{k \in M'} (2\pi ik)^{-1}\hat{B}_k \exp\{2\pi ikx\}, \tag{19}$$

where $0 \notin M$ or M', then if goodness of fit is measured in terms of MISE, the set M should equal the set M'. Note that one can treat the term

$$\sum_{k \in M'} (2\pi ik)^{-1}\hat{B}_k \exp\{2\pi ikx\} \tag{20}$$

of expression (19) as a special case of the general series estimator defined by expression (6), where

$$b_k = \begin{cases} (2\pi ik)^{-1} & \text{if } k \in M', \\ 0 & \text{otherwise.} \end{cases} \tag{21}$$

4. Weighted series estimators of functions derived from the density

In Sections 4 and 5, the choice of specific predetermined sequences of b_k will be considered from the following two points of view. One, the researcher may wish to obtain an estimate of a distribution *density* which possesses certain desirable mathematical or statistical properties. (Weights chosen for this purpose will be considered in Section 5.) For example, one may wish to constrain a density estimator to be nonnegative. Two, the target of the estimation process, rather than the density itself, may be a function derived from the density. For example, as previously discussed, one may wish to estimate and graph the cumulative distribution function. In this section, we will consider approaches to the latter class of problems which may be desirable from both a computational and statistical point of view.

It may be of value to distinguish between the previously described two classes of problems by examining the MISE criteria associated with estimates of the *density* as opposed to the MISE between a statistical construct *derived* from a density and alternative estimates of this construct. If this latter construct can be expressed as a weighted sum of the coefficients of the Fourier series expansion of the density, then the following result applies.

THEOREM 2. *Define*

$$g(x) \equiv \sum_{k\in N} b_k B_k \exp\{2\pi i k' x\} \tag{22}$$

and

$$\hat{g}(x) \equiv \sum_{k\in M} b_k \hat{B}_k \exp\{2\pi i k' x\}, \tag{23}$$

where $\{b_k\}$, $k \in N$, *is a preselected sequence of complex valued constants,* N, B_k *and* $\hat{B}_k$ *are as defined in Section 2, and* $M \subseteq N$. *Then the same set* M *minimizes the MISE* $J(g, \hat{g})$ *for all sequences* $\{b_k\}$.

PROOF. From Theorem 1 of [22], one finds that

$$J(g, \hat{g}) = \sum_{k\in M} \operatorname{Var}(b_k\hat{B}_k) + \sum_{k\in(N\cap \bar{M})} b_k^2|B_k|^2. \tag{24}$$

Therefore, the error increment ΔJ_{k_0} due to adding a term $k_0 \in M$ (as defined in Corollary 2 [22]) equals

$$b_k^2((\operatorname{Var}\hat{B}_k) - |B_k|^2). \tag{25}$$

Theorem 2 follows from the observation that for all nonzero b_k, the sign of expression (25) is identical to the sign of

$$(\operatorname{Var}\hat{B}_k) - |B_k|^2. \tag{26}$$

It is easily seen that estimates of density derivatives can be obtained, for which Theorem 2 applies. Also, consider the problem of estimating the function

$$g(x) = f(x) + C\,\frac{\partial f(x)}{\partial x_s}, \tag{27}$$

which may (at least in the univariate case) be of value in empirical Bayes investigations (here x_s is the sth component of the vector x). To estimate $g(x)$, one might choose estimator $\hat{g}(x)$ of expression (23) with

$$b_k = 1 + 2\pi i C k_s, \tag{28}$$

where k_s is the sth component of p-tuple k. If MISE is chosen as the measure of fit of $\hat{g}$ to g, then Theorem 2 implies that M can be determined by means of inclusion rule (17). Naturally, it is also possible to modify "Fellner weights" (15) to obtain a more statistically efficient estimator of expression (27). The decision of whether to use an inclusion rule or a weighting procedure should, of course, take into account computational as well as statistical properties of the alternative procedures.

A function derived from the density, which differs substantially from the integrals or derivatives of f, will be considered in the remainder of this section. Define $g^{(\lambda)}(x)$ as a univariate special case of expression (22) with

$$b_k = \exp \{2(\pi k \lambda)^2\}. \tag{29}$$

Consider the following special case of the Fourier coefficients B_k of density f,

$$B_k = \sum_{s=1}^{c} p_s \exp \{2\pi k \mu_s - 2(\pi k \sigma_s)^2\}. \tag{30}$$

If all values of μ_s and σ_s are sufficiently small, then f closely approximates a superposition of c Gaussian densities with component means equal to $\{\mu_s\}$, $s = 1, \cdots, c$, component standard deviations equal to $\{\sigma_s\}$, $s = 1, \cdots, c$, and mixing parameters $\{p_s\}$, $s = 1, \cdots, c$. Moreover, assuming $\lambda < \sigma_s$ for all s, the function $g^{(\lambda)}(x)$ closely approximates a superposition of Gaussian densities which differs from $f(x)$ only in that the set of component variances equals

$$\{(\sigma_s^2 - \lambda^2)\}, \qquad s = 1, \cdots, c. \tag{31}$$

Thus, if $\hat{B}_k$ is obtained from independent and identically distributed data arising from a superposition of c Gaussian densities, one can estimate $g^{(\lambda)}(x)$ by substituting the special case of b_k given by expression (29) into expression (23) and then determining the set M by means of inclusion rule (17).

It may be of interest to mention that a very slight modification of the method of decomposing superpositions described above can be shown to be identical to a particular form of nonparametric density estimation procedure considered in [11] (see p. 1068) and elsewhere. From expression (5), one finds that

$$b_k = \int_V \delta(x) \exp \{-2\pi i k'(x - R)\}\, dx. \tag{32}$$

If one considers the univariate Gaussian kernel δ (see [11], p. 1068), one finds that b_k of expression (32), that is, a constant times the Gaussian characteristic function evaluated at $t = -2\pi k$, is identical to expression (29) with $\lambda = \sigma i$ (where σ is the standard deviation of the Gaussian kernel). Hence, the method for decomposing superpositions of distribution functions described in this section and considered from other points of view in [3], [4], [7], [9], [10], and [15] is closely connected to the nonparametric method for estimating densities based upon Gaussian kernels. In fact, as M approaches N the specific series estimator $\hat{g}$ with b_k given in expression (29), which is used to decompose superpositions of Gaussian densities, approaches the specific nonparametric estimator (1) with δ_n set equal to a complex Gaussian function with $\mu = 0$ and $\sigma = \lambda i$. Interestingly, this particular choice of δ *reduces* the variances of superimposed Gaussian components while a choice of δ with a real positive variance can be shown from expression (10) to cause the variance of the density estimate to be greater than that of the density which is estimated.

Although functions of form g of expression (22) are of interest and of practical value in biomedical investigations, the display of composite functions, especially transgenerations of f and F are probably of primary biomedical utility (at least in the univariate case). Consider, for example, the ratio of $f(x)$ to the survival curve $1 - F(x)$, that is, the hazard function or, in some applications, the age specific death rate.

It would not be appropriate to give an extensive survey of composite graphical functions here, and hence, the remaining sections of the paper will deal with a discussion and examples of the use of $\hat{g}$ with various choices of b_k for the purpose of estimating a multivariate *density*. However, before leaving the topic of composite functions and statistical constructs derived from the density, two general comments seem appropriate.

One, the inclusion rule given by Theorem 2 applies to derived functions g of expression (20). Composite functions such as the hazard, fractile [6], confidence band [16], transformation selection [20], [17] functions may make use of combinations of derived functions $\hat{g}$. However, there is no guarantee that the particular choices of $\hat{g}$ which are optimal (even in the weak sense of inclusion rule (16)) for the purpose of estimating various choices of g *singly* will, when combined to form a composite estimate, be optimal.

Two, the computational convenience of the various estimators, which can be put into form $\hat{g}$, makes it feasible to try new and more complex composite functions. For example, Tarter and Kowalski [20] have found graphs of the function

$$\frac{\phi\Phi^{-1}\hat{F}(x)}{\hat{f}(x)} \tag{33}$$

(where ϕ and Φ represent the standard Gaussian and $\hat{f}$ and $\hat{F}$ the estimated unknown density and cumulative) to be superior in many instances to the usual fractile diagram for the purpose of selecting a transformation of data to normality.

5. Series density estimators utilizing a predetermined sequence of weights

In this section a hybrid form of density estimator will be taken up. Consider the case where one is interested in density estimates that are constrained to satisfy certain mathematical properties, for example, be nonnegative. Alternatively, suppose that one wishes to find a computationally convenient estimator that corresponds as closely as possible to a conventional statistical form, for example, a histogram that utilizes rectangular blocks. We will consider in this section a hybrid or compromise technique that tends to retain certain of the above prespecified mathematical properties while it approaches the statistical and computational efficiency of the series estimators introduced in Section 2.

Consider a specific nonparametric estimator, or equivalently, a series estimator $\hat{g}$ with coefficients b_k chosen to satisfy some mathematical constraint. For example, let $\hat{f}$ be defined by expression (1) with δ_n given by expression (2). Here

$$b_k = \frac{\sin \pi k'h}{\pi k'h}. \tag{34}$$

Alternatively, if one wishes to estimate a density f with an estimator $\hat{f}$ that is constrained to be nonnegative one might, in the univariate case, choose coefficients b_k to be the Fejer weights

$$b_k = \begin{cases} \left(1 - \dfrac{k}{m+1}\right) & \text{if } |k| \leqq m, \\ 0 & \text{elsewhere,} \end{cases} \tag{35}$$

where m is some predetermined constant. Fejer forms of series estimators are considered in the univariate case in [8] and in the multivariate case in [16].

This section deals with density estimators $\hat{g}$ of form (23) with predetermined sequences of weights b_k, for example, as given in expressions (34) or (35). Theorem 3 concerns the choice of an appropriate inclusion rule (choice of set M) for such density estimators.

THEOREM 3. *Let δ be any prespecified kernel whose Fourier expansion (assumed, as in Theorem 1, to converge in V) generates the weight sequence $\{b_k\}$. Then the weighted series estimator $\hat{g}$ of form* (21), *chosen so that an index $k_0 \in M$ if and only if*

$$n^{-1}|b_{k_0}|^2 < |B_{k_0}|^2(1 + n^{-1}|b_{k_0}|^2 - |1 - b_{k_0}|^2), \tag{36}$$

has at least as small a MISE as the nonparametric estimator obtained by substituting δ into expression (1).

The proof of Theorem 3 follows directly from MISE expression (13). Note that if we assume that δ is a symmetric kernel, then the b_k are all real and the above inequality reduces to

$$b_k < \frac{2n|B_k|^2}{1 + (n-1)|B_k|^2}, \tag{37}$$

which is equivalent to

$$b_k < 2b_k^{\text{opt}} \tag{38}$$

(see expression (14)).

An alternative interpretation of inclusion rule (37) can be obtained as follows. Define $\delta^{\text{opt}}(z)$ by using the values of b_k^{opt} given in expression (14) as the coefficients of the Fourier expansion of $\delta^{\text{opt}}(z)$ (see expression (5)). Consider that by

Parseval's theorem the integrated square error ISE between δ and δ^{opt}, that is, $\int (\delta(z) - \delta^{opt}(z))^2 \, dz$, equals $\sum_{k \in N} (b_k - b_k^{opt})^2$. Hence, to minimize the ISE one should include index k in the set M if and only if $(b_k - b_k^{opt})^2 < (b_k^{opt})^2$, that is, if $b_k < 2b_k^{opt}$, which is identical to inequality (38).

It is, of course, necessary to check whether estimate $\hat{g}$, to which the above inclusion rule is applied, retains the mathematical properties of the nonparametric estimators based upon δ. However, empirical studies with prespecified sequences of b_k, for example, Fejer weights (35), have tended to show that guaranteed possession of "mathematical" property, for example, nonnegativity, can usually be purchased only with an unacceptable increase in MISE. Thus, the hybrid procedures considered in this section may offer a reasonable compromise in certain applications.

While the choice of prespecified Fejer weight sequence (35) in conjunction with inclusion rule (38) appears to be a useful graphical method, we are not at all certain of the value of prespecified sequences of b_k, as given by expressions (34) and (29), for the purpose of density estimation. Like most procedures which arise from nonparametric estimator (1), the effective use of weights (34) and (29) depends on the estimation of at least one parameter, for weight (34) the class interval length h and for weight (29) the kernel standard deviation λ. Although it is, of course, possible to estimate the parameters of "prespecified" b_k by fitting b_k, considered as a function of h or λ, to b_k^{opt}, this seems to be a roundabout way of handling the estimation problem. Hence, in the example which will be considered in the next section, estimation is implemented from the series point of view, using the computational procedures described in Section 3.

6. Biomedical application

In this section, we consider a specific application of the computational methods outlined in Section 3, for estimating a bivariate density.

The basic data used in these calculations were measurements of chronological age and bone age of children included in the Child Development Studies (Kaiser Foundation, Oakland, California). The children included were a 10 per cent sample, stratified by sex, race, and height (classified as tall, medium, short). For purposes of illustration, we have included here the data for white males of medium height. The particular bone selected is hamate and the bone ages are determined by matching the X-ray picture of the child with the standard radiological atlas prepared by Gruelich and Pyle [5]. The values read in are close to within three months of the actual bone age since the graduation of the atlas is in intervals of three months.

The program computes the bivariate probability density nonparametrically from the data, using the technique of Fourier approximation of multivariate densities (see [22]). The x variable is the chronological age in months and the y variable is the bone age in months. The lower and upper limits are obtained by calculating the maximum and minimum values and choosing the closest number in tenths. Table I gives the Fourier coefficients calculated from the data for the upper half plane. The values for the lower half plane can be obtained as conjugates of the values on the upper half plane.

TABLE I

HAMATE BONE AGE STUDY, GROUP II: WHITE MALES, MEDIUM HEIGHT
X lower = 30.0, X upper = 130.0, Y lower = 20.0,
Y upper = 130.0. Number of observations = 98.

		Fourier coefficients (upper half plane)	
X coord	Y coord	Real	Imaginary
−3	1	−0.16353922E-04	0.79695746E-05
−3	2	−0.11921363E-04	−0.73126666E-05
−2	0	−0.26822795E-04	0.13461094E-04
−2	1	−0.16394707E-04	−0.22885004E-04
−2	2	0.40507031E-04	−0.24095239E-04
−1	0	−0.17124028E-04	−0.31696021E-04
−1	1	0.75077390E-04	−0.18168164E-04
−1	2	−0.22633103E-05	0.23782737E-04
0	0	0.90909103E-04	0.00000000E 00
0	1	−0.12383168E-04	0.32439042E-04
0	2	−0.18815233E-04	−0.34065936E-06
0	3	−0.13230197E-04	−0.13690030E-04
1	0	−0.17124028E-04	0.31696021E-04
1	1	−0.27838021E-04	−0.83455852E-05
1	2	−0.52113328E-05	−0.19308896E-04
2	0	−0.26822795E-04	−0.13461094E-04
2	1	0.41055191E-05	−0.16014244E-04

Table II shows the bivariate probability density calculated from the Fourier coefficients for the grid of points within the specified limits. The corresponding scatter diagram for these limits are shown in Table III. The limits chosen for the display of scatter diagram and bivariate density are not the same as the ones used for the calculation of the Fourier coefficients, which accounts for the discrepancy in the number of observations shown on the scatter diagram. Since total probability density is always unity, the appropriate actual density height in Table II can be obtained by dividing the number shown by the scale factor given in the title.

For ease of visualization, the table was converted to a contour diagram which is shown in Table IV. In the preparation of this table, numbers to be displayed were truncated to tenths. The table displays only those numbers starting with even second digits which are greater than or equal to 20.

By abstracting the essential features of the scatter diagram, the contour chart clearly exhibits distributional features of the data. For example, Table IV indicates the possible decomposition of the data into a bivariate normal distribution and a degenerate uniform distribution.

TABLE II

BIVARIATE PROBABILITY DENSITY (TIMES 110,000)
HAMATE BONE AGE STUDY, GROUP II: WHITE MALES, MEDIUM HEIGHT
Number of observations = 98.

Bone age in months	Age in months																									
	35					52					69					86					103					120
130																							8	11	12	9
126																					8	15	20	21	18	11
123																			5	15	25	32	35	32	25	14
119																		9	21	34	45	51	50	43	32	18
115																	11	25	40	55	65	68	64	53	37	21
112																10	24	41	59	73	82	82	74	59	40	22
108															6	19	37	56	74	88	93	89	77	60	40	21
104															11	27	47	67	84	95	97	90	74	55	35	17
101															16	34	54	73	88	95	93	82	65	45	27	12
97														6	20	38	57	74	85	88	82	69	51	33	17	5
93														9	24	41	57	70	76	75	66	52	36	20	7	0
90														14	27	42	54	62	64	59	49	35	21	8	0	0
86												5	11	20	31	41	49	52	50	43	32	19	8		0	0
82								5	6	7	9	13	18	25	33	39	42	41	36	28	17	8			0	0
79						5	8	11	14	16	18	21	25	30	33	36	35	31	24	15	7				0	0
75					5	9	14	18	21	23	26	28	30	31	32	30	26	21	13	6					0	0
71					7	13	18	23	26	29	30	31	31	30	27	23	18	11	5						0	0
68					9	15	21	26	29	31	31	30	28	25	20	15	10	0	0						0	5
64					10	16	22	26	28	29	28	25	22	17	12	7	0	0	0						5	5
60				5	10	16	21	24	25	24	22	18	13	8	0	0	0	0	0						0	0
57				6	11	16	20	21	21	18	14	9	5	0	0	0	0	0	0						0	0
53				9	13	17	18	18	16	12	7			0	0	0	0	0	0						0	0
49		5	9	13	16	18	18	16	11	6				0	0	0	0	0	0						0	0
46	8	11	14	17	19	20	18	14	8					0	0	0	0	0	5						0	0
42	13	16	19	22	22	21	18	12	6					0	0	0	7	7	6						0	0
38	17	20	23	24	24	22	17	11						0	5	8	9	8	5						0	0
35	19	21	24	25	24	21	15	9						0	8	10	9	6							0	5
31	17	19	21	22	21	18	12	6						6	9	9	6								0	7
27	12	14	16	17	16	13	8							6	7	5									0	7
24	6	8	10	11	10	7																			7	8
20	0	0	0	0	0	0																			0	0

TABLE III

SCATTER DIAGRAM
HAMATE BONE AGE STUDY, GROUP II: WHITE MALES, MEDIUM HEIGHT
Number of observations = 96.

Bone age in months	Age in months																									
	35					52					69					86					103					120
130																										
126																										
123																							1			
119																						1	2			
115																							1	2		
112																			4	1			3			
108																			2	3		1	1			
104																										
101																2	1	1	1		2	3	1			
97																1		2		4		1	2			
93																1			2	1		1	1			
90															1	1	1									
86															1	1	1	1								
82															1		1	1		1						
79													2			1				2						
75													1				1									
71				1					1	1							1									
68								1				1						1								
64						1		1	2	1	1		1													
60																										
57								1																		
53						1			1																	
49																										
46	1								1																	
42																										
38	2			1	1	1																				
35	1			1																						
31	1		1																							
27																										
24						1																				
20																										

TABLE IV

Contours of Probability Density (Times 110,000)
Hamate Bone Age Study, Group II: White Males, Medium Height
Number of observations = 98. Values below 20 have been suppressed.

Bone age in months	Age in months																									
	35					52					69					86					103					120
130																										
126																							20	20		
123																					20				20	
119																			20		40			40		
115																		20	40		60	60	60			20
112																	20	40			80	80			40	20
108																				80		80		60	40	20
104																20	40	60	80							
101																			80			80	60	40	20	
97															20				80	80	80	60				
93															20	40					60			20		
90															20	40		60	60		40		20			
86														20		40	40			40						
82														20			40	40		20						
79												20	20						20							
75									20	20	20	20					20	20								
71								20	20	20					20	20										
68							20	20	20				20	20	20											
64							20	20	20	20	20	20	20													
60							20	20	20	20	20															
57							20	20	20																	
53																										
49																										
46						20																				
42				20	20	20																				
38		20	20	20	20	20																				
35		20	20	20	20	20																				
31			20	20	20																					
27																										
24																										
20																										

TABLE V

CONDITIONAL PROBABILITY DENSITY
HAMATE BONE AGE STUDY, GROUP II: WHITE MALES, MEDIUM HEIGHT
Number of observations = 98.

Bone age in months	Age in months																									
	35					52					69					86					103					120
130																							1	2	3	4
126																					1	2	3	4	5	6
123																				1	3	4	5	7	8	7
119																		1	2	4	5	7	8	9	10	9
115																	2	3	5	7	8	9	10	11	12	10
112																2	4	6	7	9	10	11	12	13	13	11
108															2	4	6	8	9	11	12	12	13	13	13	11
104															3	6	8	10	11	12	12	12	12	12	11	9
101															4	7	9	11	11	12	12	11	11	10	8	6
97														2	6	8	10	11	11	11	10	9	8	7	5	2
93														4	7	9	10	10	10	9	8	7	6	4	2	
90														6	8	9	9	9	8	7	6	5	3	1		
86												2	6	8	9	9	8	7	6	5	4	2	1			
82								2	3	3	5	7	10	11	10	8	7	6	4	3	2	1				
79						2	3	4	6	8	9	11	13	12	10	8	6	4	3	2						
75					2	3	5	7	9	11	13	15	16	13	9	6	4	3	1							
71					3	5	7	9	12	14	16	17	16	12	8	5	3	1								
68					3	6	8	10	13	15	16	16	15	10	6	3	1									
64					4	6	8	10	13	14	14	13	11	7	3	1										
60				2	4	6	8	10	11	12	11	9	7	3												
57				3	4	6	7	8	9	9	7	5	2													
53				4	5	6	7	7	7	6	3															
49		4	6	7	6	7	7	6	5	3																
46	8	9	10	9	8	7	7	5	3																	
42	14	14	14	11	9	8	7	5	2																	
38	18	17	16	13	10	8	6	4																		
35	20	18	17	13	10	8	6	3																		
31	18	16	15	12	8	7	4	2																		
27	13	12	12	9	6	5	3																			
24	6	7	7	6	4	3																				
20																										

Table V shows the empirical conditional probability distribution $\hat{f}(Y|X)$ (obtained by dividing the bivariate probability density by the marginal density). Also shown, in Table VI, are the estimated regression $E(Y|X)$, the standard deviation, mode, median, and the two quartiles of $\hat{f}(Y|X)$.

TABLE VI

ESTIMATED MODE, QUARTILES, MEDIAN, REGRESSION, CONDITIONAL STANDARD DEVIATION, VARIANCE, AND CORRECTION

X	Mode	$Q(1)$	$Q(3)$	Median	$E(Y\|X)$	S.D.	$V(Y\|X)$	Correction
35.0	34.6	31.7	36.8	34.3	35.0	6.1	38.20	0
38.4	34.6	31.6	42.9	33.6	35.7	6.9	47.79	2
41.8	34.6	31.5	42.2	33.3	36.1	7.2	52.19	5
45.2	34.5	30.0	44.2	37.8	43.4	20.7	432.55	1
48.6	38.3	34.7	55.1	47.4	49.3	21.7	473.22	0
52.0	38.3	39.6	59.0	50.6	49.3	14.8	220.58	2
55.4	64.0	41.9	62.2	51.2	53.9	14.2	202.41	1
58.8	63.9	49.3	66.2	60.2	58.8	13.0	171.01	0
62.2	67.7	56.5	71.1	63.4	63.8	9.9	99.70	−1
65.6	67.6	60.3	69.7	68.5	66.5	8.3	69.52	0
69.0	67.5	65.4	75.8	66.9	68.2	7.6	58.18	0
72.4	71.4	63.3	73.9	72.2	70.4	7.4	55.68	1
75.8	71.2	68.7	79.1	70.3	72.3	7.5	57.16	3
79.2	75.0	66.7	81.5	74.3	74.2	14.2	203.40	4
82.6	78.6	70.7	88.1	83.3	78.9	18.4	339.89	3
86.0	89.6	73.2	96.8	84.7	84.3	18.6	346.65	1
89.4	97.0	82.7	100.1	94.5	89.4	17.3	301.94	0
92.8	96.9	85.2	104.2	97.9	94.1	15.1	230.55	−1
96.2	100.6	89.2	109.1	102.4	97.4	13.8	190.81	−1
99.6	104.4	93.2	106.7	99.7	101.1	10.6	114.47	−1
103.0	104.3	98.0	111.9	104.8	103.5	10.4	109.38	0
106.4	104.2	95.7	110.1	102.8	105.6	10.1	102.50	0
109.8	108.0	100.8	115.3	108.1	107.8	9.9	99.85	1
113.2	107.9	105.9	113.6	106.2	109.9	9.6	93.53	3
116.6	111.7	104.9	119.6	112.3	108.7	17.5	307.17	7
120.0	111.6	101.4	119.7	113.3	98.2	32.1	1032.62	21

Biologically, for the average normal child the bone age should be the same as the chronological age. In practice, however, there are sources of error due to observer bias. Further, the atlas on which the assessments are based was calibrated 40 years ago and, hence, the possibility of a secular trend on the osteological maturation of California's children cannot be ruled out. Consequently, a correction at each age to within three months can be obtained from the difference of the chronological age and the regression estimate shown in the last column of Table VI. After applying the correction to the original data, we then recomputed the bivariate distribution. The corresponding results are shown in Tables VII, VIII, and IX. It can be seen from Table X that the second order corrections are now negligible, at least at those levels where there are observations. Further, the contour chart after the correction shows a sharper segregation of the com-

TABLE VII

BIVARIATE PROBABILITY DENSITY (TIMES 110,000)
HAMATE BONE AGE STUDY, GROUP II: WHITE MALES, MEDIUM HEIGHT
Number of observations = 98.

Bone age in months	Age in months																									
	35					52					69					86					103					120
130																							8	10	10	6
126																					10	17	21	20	16	9
123																			8	18	28	34	36	32	24	13
119																		11	24	37	48	54	52	45	32	17
115																	12	26	43	58	68	71	67	55	38	21
112																9	24	42	60	76	85	85	76	61	42	23
108																16	35	55	75	89	95	92	79	61	41	23
104															7	23	44	65	84	95	98	91	76	56	36	19
101															10	28	50	70	86	94	92	82	65	46	27	13
97															14	33	53	70	82	85	80	68	51	33	17	6
93															18	36	53	66	73	72	64	50	34	19	7	
90														10	23	38	51	59	61	56	46	32	19	7		
86													9	18	29	39	47	50	48	40	29	17	6			
82							5	7	7	8	9	13	18	26	33	40	43	41	35	26	15	5				
79						7	10	13	15	17	19	22	27	32	36	38	37	32	23	14	5					
75					6	11	15	19	22	25	27	30	33	36	36	34	30	22	14	5						
71					8	13	19	23	27	30	32	34	35	35	33	28	22	14	6							
68					9	15	20	25	28	31	33	34	33	31	26	20	13	6								
64					9	14	20	24	27	29	30	29	26	23	18	12	5	0								
60					8	13	18	21	23	24	23	21	17	13	8	0	0	0								
57					9	13	16	18	18	17	15	11	8		0	0	0	0								
53				7	10	13	15	15	13	10	7				0	0	0	0								
49		5	8	11	14	15	15	13	9	5					0	0	0	0								
46	9	12	14	17	18	18	16	12	7						0	0	0	0								
42	14	17	20	22	22	21	17	11	5						0	0	5	6								
38	18	21	24	25	25	22	17	11							0	7	8	7								
35	19	23	25	26	25	22	16	9							7	9	9	6								
31	17	20	23	24	23	19	14	7							8	8	6									
27	12	15	18	19	18	15	9								7	5										
24	5	8	12	13	13	9	5																			
20																										

TABLE VIII

SCATTER DIAGRAM
HAMATE BONE AGE STUDY, GROUP II: WHITE MALES, MEDIUM HEIGHT
Number of observations = 96.

Bone age in months	Age in months																									
	35					52					69					86					103					120
130																										
126																										
•123																							1			
119																						1	2			
115																							1			
112																			4	1			2	2		
108																			2	3		1	2			
104																			1		2					
101																	1	1		4		3	1			
97																2		2				1	2			
93																1			2	1		1	1			
90																1	1	1								
86															1	1	1	1		1						
82															1	1	1			2						
79															1											
75									1	1				2		1	1									
71														1			1	1								
68				1				1	1																	
64								1	1	1	1	1														
60						1																				
57								1	1					1												
53																										
49						1			1																	
46		1																								
42																										
38		2				1																				
35		1			1		1																			
31		1			1																					
27				1																						
24							1																			
20																										

TABLE IX

CONTOURS OF PROBABILITY DENSITY (TIMES 110,000)
HAMATE BONE AGE STUDY, GROUP II: WHITE MALES, MEDIUM HEIGHT
Number of observations = 98. Values below 20 have been suppressed.

Bone age in months	Age in months																									
	35					52					69					86					103					120
130																										
126																							20	20		
123																					20				20	
119																			20		40			40		
115																		20	40		60		60			20
112																	20	40	60		80	80		60	40	20
108																				80				60	40	20
104																20	40	60	80							
101																20			80			80	60	40	20	
97																			80	80	80	60				
93																		60			60					
90															20				60		40					
86															20		40		40	40	20					
82														20		40	40	40		20						
79												20	20						20							
75									20	20	20							20								
71								20	20							20	20									
68							20	20	20						20	20										
64							20	20	20	20		20	20	20												
60								20	20	20	20	20														
57																										
53																										
49																										
46																										
42			20	20	20	20																				
38		20	20	20	20	20																				
35		20	20	20	20	20																				
31		20	20	20	20																					
27																										
24																										
20																										

TABLE X

ESTIMATED MODE, QUARTILES, MEDIAN, REGRESSION, CONDITIONAL STANDARD DEVIATION, VARIANCE, AND CORRECTION

X	Mode	$Q(1)$	$Q(3)$	Median	$E(Y\|X)$	S.D.	$V(Y\|X)$	Correction
35.0	34.6	31.4	36.6	34.0	35.2	6.1	37.78	0
38.4	34.6	31.6	42.9	33.6	35.7	6.9	47.71	2
41.8	34.6	31.5	41.8	33.2	39.2	18.9	359.02	2
45.2	34.6	31.2	40.2	39.6	40.8	20.3	414.60	4
48.6	34.5	36.1	58.1	42.3	46.6	20.3	415.92	1
52.0	38.3	33.7	65.7	44.9	48.7	15.6	244.81	3
55.4	67.7	36.5	67.9	51.2	53.8	15.9	253.14	1
58.8	67.6	49.3	72.3	59.0	59.4	13.7	187.86	0
62.2	67.6	55.3	70.3	62.3	64.7	9.9	98.65	−2
65.6	67.6	59.6	76.7	68.0	67.0	8.2	67.57	−1
69.0	67.5	65.3	75.8	66.9	68.3	7.5	57.15	0
72.4	71.4	63.7	74.7	72.9	69.6	7.0	49.84	2
75.8	71.3	62.1	80.1	71.2	71.4	7.5	56.62	4
79.2	75.0	68.1	77.3	76.4	72.1	12.4	154.25	7
82.6	74.9	72.7	84.8	79.1	76.7	16.4	271.27	5
86.0	82.3	75.2	98.4	86.9	82.9	18.2	334.86	3
89.4	93.3	76.9	100.4	88.1	88.6	17.4	304.33	0
92.8	96.9	85.9	104.0	98.0	93.9	15.4	237.68	−1
96.2	100.6	88.7	108.5	101.9	98.2	13.2	175.87	−2
99.6	104.3	92.6	106.2	99.2	101.6	10.7	114.89	−2
103.0	104.3	97.4	111.4	104.2	104.1	10.4	108.39	−1
106.4	108.0	102.4	117.0	109.6	106.1	10.0	100.14	0
109.8	107.9	100.3	115.1	107.7	108.2	9.8	97.14	1
113.2	107.9	105.7	113.6	113.4	109.9	9.5	90.45	3
116.6	111.7	105.3	120.2	112.7	108.3	16.7	280.88	8
120.0	111.7	100.6	120.7	106.3	101.0	27.3	748.79	18

ponents. Also, it will be noted that the quartiles after correction give a more reasonable range between the $P_{0.25}$ and $P_{0.75}$ quartiles.

◇ ◇ ◇ ◇ ◇

The authors would like to thank Professor J. Yerushalmy for making available the data utilized in Section 6 and Dr. W. Fellner and Dr. R. Brand for suggestions concerning statistical aspects of this paper.

REFERENCES

[1] W. C. BRINTON, *Graphic Methods for Presenting Facts*, New York, The Engineering Magazine Company, 1917.

[2] N. N. CENCOV, "Evaluation of an unknown distribution density from observations," *Soviet Math. Dokl.*, Vol. 3 (1962), pp. 1559–1562.

[3] G. DOETSCH, "Zerlegung einer Function in Gauss'sche Fehlerkurven," *Math. Z.*, Vol. 41 (1936), pp. 283–318.

[4] J. GREGOR, "An algorithm for the decomposition of a distribution into Gaussian components," *Biometrics*, Vol. 25 (1969), pp. 79–93.

[5] W. W. GRUELICH and S. I. PYLE, *Radiographic Atlas of Skeletal Development of the Hand and Wrist*, Stanford, Stanford University Press, 1966 (2nd ed.).

[6] A. HALD, *Statistical Theory with Engineering Applications*, New York, Wiley, 1952.

[7] R. A. KRONMAL, "The estimation of probability densities," unpublished Ph.D. thesis, Division of Biostatistics, University of California, Los Angeles, 1964.

[8] R. A. KRONMAL and M. TARTER, "The estimation of probability densities and cumulatives by Fourier series methods," *J. Amer. Statist. Assoc.*, Vol. 63 (1968), pp. 925–952.

[9] P. MEDGYESSY, "The decomposition of compound probability distributions," *Hungar. Acad. Sci. Inst. Appl. Math.*, Vol. 2 (1953), pp. 165–177. (In Hungarian.)

[10] ———, *Decomposition of Superpositions of Distribution Functions*, Budapest, Pub. House of the Hungarian Academy of Sciences, 1961.

[11] E. PARZEN, "On estimation of a probability density function and mode," *Ann. Math. Statist.*, Vol. 33 (1962), pp. 1065–1076.

[12] S. RAMAN, "Contribution to the theory of Fourier estimation of multivariate probability density functions with application to data on bone age determinations," unpublished Ph.D. thesis, Department of Biostatistics, University of California, Berkeley, 1971.

[13] M. ROSENBLATT, "Remarks on some nonparametric estimates of a density function," *Ann. Math. Statist.*, Vol. 27 (1956), pp. 832–837.

[14] S. C. SCHWARTZ, "Estimation of probability density by an orthogonal series," *Ann. Math. Statist.*, Vol. 38 (1967), pp. 1961–1965.

[15] D. F. STANAT, "Nonsupervised pattern recognition through the decomposition of probability functions," Technical Report, University of Michigan, Sensory Intelligence Laboratory, 1966.

[16] M. TARTER, "Variance and covariance formulas for evaluations of estimated orthogonal expansions," to appear.

[17] ———, "Inverse cumulative approximation and applications," *Biometrika*, Vol. 55 (1968), pp. 29–42.

[18] M. TARTER and W. FELLNER, "Some new results concerning density estimates based on Fourier series," *Proceedings of the Fifth Conference on the Interface Between Statistics and Computation*, North Hollywood, Western Periodicals, 1972.

[19] M. TARTER, R. HOLCOMB, and R. A. KRONMAL, "A description of new computer methods for estimating the population density," *Proc. Assoc. for Computing Machinery*, Vol. 22 (1967), pp. 511–519.

[20] M. TARTER and C. J. KOWALSKI, "A new test for, and class of transformations to, normality," *Technometrics*, Vol. 50 (1972), in press.

[21] M. TARTER and R. A. KRONMAL, "Estimation of the cumulative by Fourier series methods and application to the insertion problem," *Proc. Assoc. for Computing Machinery*, Vol. 23 (1968), pp. 491–497.

[22] ———, "On multivariate density estimates based on orthogonal expansions," *Ann. Math. Statist.*, Vol. 41 (1970), pp. 718–722.

[23] J. VAN RYZIN, "Bayes risk consistency of classification procedures using density estimation," *Sankhyā Ser. A*, Vol. 28 (1966), pp. 261–270.

[24] G. S. WATSON, "Density estimation by orthogonal series," *Ann. Math. Statist.*, Vol. 40 (1969), pp. 1496–1498.

[25] G. S. WATSON and M. R. LEADBETTER, "On the estimation of the probability density, I," *Ann. Math. Statist.*, Vol. 34 (1963), pp. 480–491.

[26] E. J. WEGMAN, "Nonparametric probability density estimation," University of North Carolina at Chapel Hill, Institute of Statistics Mimeo Series, No. 638, 1969.