The Annals of Mathematical Statistics
1970, Vol. 41, No. 2, 718–722

ON MULTIVARIATE DENSITY ESTIMATES BASED ON ORTHOGONAL EXPANSIONS[1]

BY MICHAEL TARTER AND RICHARD KRONMAL

University of California at Irvine and University of Washington

0. Introduction and summary. The topics of orthogonality and Fourier series occupy a central position in analysis. Nevertheless, there is surprisingly little statistical literature, with the exception of that of time series and regression, which involves Fourier analysis. In the last decade however, several papers have appeared which deal with the estimation of orthogonal expansions of distribution densities and cumulatives. Čencov [1] and Van Ryzin [6] considered general properties of orthogonal expansion based density estimators and the latter applied these properties to obtain classification procedures. Schwartz [3] and the authors [2] and [4] investigated respectively the Hermite and Trigonometric special cases. The authors also obtained certain general results which apply not only to estimators of the population density but also to estimators of the population cumulative [2], [4] and [5]. In this paper several results derived for the univariate case are extended to the multivariate case. Also a new relationship is obtained which involves general Fourier expansions and estimators.

Although there is some reason for calling the Gram–Charlier estimation of distribution densities a Fourier method, one fundamental aspect of Fourier methods is not shared by Gram–Charlier estimation. Gram–Charlier techniques make no use of Parseval's Formula or related error relationships of Fourier analysis. The ease with which the mean integrated square error (MISE) is evaluated, when Fourier methods are applied, accounts for most of the recent interest in this area.

Section 1 of this paper deals with an investigation of two general MISE relationships for multivariate estimates of Fourier expansions. The relationship given in Theorem 2 is particularly simple and yet includes the four MISE's which are involved in the estimation problem.

In Section 2 the choice of orthogonal functions is restricted to the trigonometric polynomials. It is shown that the MISE of multidimensional trigonometric polynomial estimators are related in a simple way to the Fourier coefficients of the distribution which is being estimated. This result is of considerable utility since it yields a rule for deciding which terms should be included in the estimate of the multivariate density.

1. Basic theorems—arbitrary orthogonal expansions. In this section two basic properties of estimates of orthogonal expansions are discussed. The first property was investigated independently by Čencov [1], Schwartz [3], and the authors [2],

Received May 19, 1969; revised October 24, 1969.

[1] This research was supported by a General Research Support Grant from the Public Health Service.

[4] and [5] but is reintroduced here because of its central position relative to the other results of this paper. Although the second property is extremely simple and has immediate application, it seems to be set forth for the first time.

The following notation, given by Zygmund [8] page 300, will be used in this section. The symbols j and k represent p-tuples consisting of positive integers, negative integers or zero's. The sets $\mathscr{M}$ and $\mathscr{N}$ and their complements $\bar{\mathscr{M}}$ and $\bar{\mathscr{N}}$ are composed of such p-tuples. The symbol $\mathscr{X}$ is used to represent a $p \times 1$ vector of real variables and B_k and $\hat{B}_k$ represent complex valued scalars. The scalar $\hat{B}_k$ is a statistic formed from n i.i.d. random variates. The complex valued functions $\psi_k(\mathscr{X})$ are defined to be orthonormal over the set D in p-dimensional Euclidian space E^p, i.e.,

$$\int_D \psi_k(\mathscr{X})\overline{\psi_j(\mathscr{X})}\, d\mathscr{X} = \delta_{kj}. \tag{1}$$

Finally the MISE (mean integrated square error) of the complex valued random functions $G_1(\mathscr{X})$ and $G_2(\mathscr{X})$ is represented by $J(G_1, G_2)$ and defined as

$$J(G_1, G_2) = E\int_D [G_1(\mathscr{X}) - G_2(\mathscr{X})][\overline{G_1(\mathscr{X})} - \overline{G_2(\mathscr{X})}]\, d\mathscr{X}. \tag{2}$$

Theorem 1 concerns the reexpression of (2) for $G_1 = F_{\mathscr{N}}$ and $G_2 = \hat{F}_{\mathscr{M}}$ where

$$F_{\mathscr{N}}(\mathscr{X}) \equiv H(\mathscr{X}) + \textstyle\sum_{k \in \mathscr{N}} B_k \psi_k(\mathscr{X}) \tag{3}$$

$$\hat{F}_{\mathscr{M}}(\mathscr{X}) \equiv H(\mathscr{X}) + \textstyle\sum_{k \in \mathscr{M}} \hat{B}_k \psi_k(\mathscr{X}) \tag{4}$$

and $\mathscr{M} \subseteq \mathscr{N}$. Further, the series defining $F_{\mathscr{N}}$ and $\hat{F}_{\mathscr{M}}$ will be assumed uniformly convergent.

THEOREM 1.

$$J(F_{\mathscr{N}}, \hat{F}_{\mathscr{M}}) = \textstyle\sum_{k \in \mathscr{M}} E[(\hat{B}_k - B_k)(\bar{\hat{B}}_k - \bar{B}_k)] + \sum_{k \in (\mathscr{N} \cap \bar{\mathscr{M}})} B_k \bar{B}_k. \tag{5}$$

The proof of Theorem 1 follows trivially from orthogonality property (1).

In later sections the question of the choice of terms to be included in subset $\mathscr{M}$ will be investigated. For this purpose Theorem 2, which involves an identity between the four MISE's; $J(F_{\mathscr{N}}, \hat{F}_{\mathscr{M}})$, $J(\hat{F}_{\mathscr{N}}, \hat{F}_{\mathscr{M}})$, $J(F_{\mathscr{N}}, F_{\mathscr{M}})$ and $J(\hat{F}_{\mathscr{N}}, F_{\mathscr{N}})$, will be of importance.

THEOREM 2. *If* $E(\hat{B}_k) = B_k$ *then*

$$J(F_{\mathscr{N}}, \hat{F}_{\mathscr{M}}) + J(\hat{F}_{\mathscr{N}}, \hat{F}_{\mathscr{M}}) = 2J(F_{\mathscr{N}}, F_{\mathscr{M}}) + J(\hat{F}_{\mathscr{N}}, F_{\mathscr{N}}). \tag{6}$$

PROOF.

$$\begin{aligned} J(\hat{F}_{\mathscr{N}}, \hat{F}_{\mathscr{M}}) &= E\textstyle\sum_{k \in (\mathscr{N} \cap \bar{\mathscr{M}})} \hat{B}_k \bar{\hat{B}}_k \\ &= \textstyle\sum_{k \in (\mathscr{N} \cap \bar{\mathscr{M}})} E[(\hat{B}_k - B_k)(\bar{\hat{B}}_k - \bar{B}_k)] + \sum_{k \in (\mathscr{N} \cap \bar{\mathscr{M}})} B_k \bar{B}_k. \end{aligned} \tag{7}$$

$$J(F_{\mathscr{N}}, F_{\mathscr{M}}) = \textstyle\sum_{k \in (\mathscr{N} \cap \bar{\mathscr{M}})} B_k \bar{B}_k. \tag{8}$$

$$J(\hat{F}_{\mathscr{N}}, F_{\mathscr{N}}) = \textstyle\sum_{k \in \mathscr{M}} E[(\hat{B}_k - B_k)(\bar{\hat{B}}_k - \bar{B}_k)] + \sum_{k \in (\mathscr{N} \cap \bar{\mathscr{M}})} E[(\hat{B}_k - B_k)(\bar{\hat{B}}_k - \bar{B}_k)]. \tag{9}$$

The theorem follows from the rearrangement of the terms of expressions (5), (7), (8) and (9).

Although Theorems 1 and 2 are valid for any orthonormal expansion when $\mathscr{N}$ is finite and for all of the commonly used orthonormal systems $\{\psi_k\}$ when $\mathscr{N}$ is denumerable, primary emphasis will be placed in the remainder of this paper upon the trigonometric functions

$$\psi_k(x) = e^{2\pi i k' x} \tag{10}$$

2. Basic theorems—trigonometric system. In the remainder of this paper the n $p \times 1$ vectors of *real*-valued random variates X_s ; $s = 1, \cdots, n$; will be defined to be i.i.d. with distribution density $f(x)$ where $f \in L_1 \cap L_2$. Also, B_k will be defined as the Fourier coefficient

$$B_k = \int_{\mathscr{R}} f(x)\, e^{-2\pi i k' x}\, dx \tag{11}$$

where $\mathscr{R}$ represents the p-dimensional unit hypercube in p-dimensional Euclidean space E^p. The support of f will be assumed to be a subset of $\mathscr{R}$ and $\hat{B}_k$ will be defined as the following estimate of B_k

$$\hat{B}_k = n^{-1} \textstyle\sum_{s=1}^{n} e^{-2\pi i k' X_s}. \tag{12}$$

It is important to note the obvious fact that

$$E\hat{B}_k = B_k. \tag{13}$$

However, not only is the first moment of the statistic $\hat{B}_k$ expressible in terms of the Fourier coefficients of f, as would be true if any orthogonal ψ_k were substituted into expressions (11) and (12), but the second ordinary moments as well as the product moments of $\hat{B}_k$ are likewise easily expressible by equation (14).

THEOREM 3. *For all j and k*

$$E[(\hat{B}_k - B_k)(\bar{\hat{B}}_j - \bar{B}_j)] = n^{-1}[B_{k-j} - B_k B_{-j}] \tag{14}$$

which in turn equals

$$n^{-1}[B_{k-j} - B_k \bar{B}_j]. \tag{15}$$

PROOF.

$$E[(\hat{B}_k - B_k)(\bar{\hat{B}}_j - \bar{B}_j)] = E\hat{B}_k \bar{\hat{B}}_j - B_k \bar{B}_j. \tag{16}$$

$$\begin{aligned} E\hat{B}_k \bar{\hat{B}}_j &= E(n^{-2})\{\textstyle\sum_{s=1}^{n} e^{-2\pi i[(k-j)' X_s]} + \sum_{s \neq t} e^{-2\pi i (k' X_s - j' X_t)}\} \\ &= n^{-1}\{E\, e^{-2\pi i (k-j)' x} + (n-1)(E\, e^{-2\pi i k' x})(E\, e^{2\pi i j' x})\}. \end{aligned} \tag{17}$$

Expression (14) follows from the substitution of expression (11) and (17) into (16) and expression (15) results from the identity $\bar{B}_j = B_{-j}$ which is apparent from expression (11).

Similar relationships involving non-trigonometric orthogonal expansions have been investigated by the authors and G. Anderson. These will be presented separately. In the univariate case a formula obtained by Watson [7] does yield a simplified expression for the covariance of two coefficients of Hermite expansions. However, expression (14) presently appears to be unique. The covariance or variance is a simple function of the Fourier coefficients for no other orthogonal family which has been investigated.

Letting $j=k$ in expression (15) and substituting the result into expression (5) yields the following corollary.

COROLLARY 1.

$$J(F_{\mathscr{N}},\hat{F}_{\mathscr{M}}) = n^{-1}\textstyle\sum_{k\in\mathscr{M}}[1-B_k\bar{B}_k]+\sum_{k\in(\mathscr{N}\cap\bar{\mathscr{M}})}B_k\bar{B}_k. \tag{18}$$

If the set $\mathscr{M}^+$ is defined as $\mathscr{M}\cup\{k_0\}$, $k_0\notin\mathscr{M}$, we get:

COROLLARY 2. *Define the error increment* (ΔJ_{k_0}) *due to adding a term* k_0 *as*

$$J(F_{\mathscr{N}},\hat{F}_{\mathscr{M}^+})-J(F_{\mathscr{N}},\hat{F}_{\mathscr{M}}), \quad \text{then} \quad \Delta J_{k_0}=\{n^{-1}-n^{-1}(n+1)B_{k_0}\bar{B}_{k_0}\}. \tag{19}$$

Note that the sign of ΔJ_{k_0}, equals the sign of $\{(n+1)^{-1}-B_{k_0}\bar{B}_{k_0}\}$ or, in other words, the inclusion of the k_0-th term to form estimate $\hat{F}_{\mathscr{M}^+}$, as opposed to the estimate $\hat{F}_{\mathscr{M}}$ is indicated whenever

$$B_{k_0}\bar{B}_{k_0} > (n+1)^{-1}. \tag{20}$$

As suggested by Čencov [1], "it is then possible to select n, 'here k_0,' according to results of observations, restricting oneself in '(4)' only to such terms whose coefficients '$\hat{B}_{k_0}$' are essentially greater 'than' their experimental mean square error." However the substitution of the estimate $\hat{B}_k$ for B_k in (20) will result in a biased rule that would select too many terms. It will be shown using Corollary 3 that the appropriate inclusion rule would indicate the use of the term k_0 whenever

$$\hat{B}_{k_0}\bar{\hat{B}}_{k_0} > 2(n+1)^{-1} \tag{21}$$

rather than $\hat{B}_{k_0}\bar{\hat{B}}_{k_0} > (n+1)^{-1}$.

COROLLARY 3. Let

$$\widehat{\Delta J}_{k_0}=\{n^{-1}-n^{-1}(n+1)[n(n-1)^{-1}\hat{B}_{k_0}\bar{\hat{B}}_{k_0}-(n-1)^{-1}]\}. \tag{22}$$

Then

$$E(\widehat{\Delta J}_{k_0}) = \Delta J_{k_0}. \tag{23}$$

Corollary 3 follows from expression (17). Now from (22) we see that the sign of $\widehat{\Delta J}_{k_0}$, the unbiased estimator of ΔJ_{k_0}, equals the sign of $\{2(n+1)^{-1}-\hat{B}_{k_0}\bar{\hat{B}}_{k_0}\}$ i.e., we use expression (21) for our inclusion rule.

3. Conclusion. Some computational properties and applications of expressions (12) and stopping rule (21) are discussed by the authors in [2], [4] and [5] for the univariate case. In [2] and [4] the estimator given in this section is shown to compare favorably with univariate estimators of the population density proposed by other authors. In a forthcoming paper it will be shown that the results of Section 1 can be applied to the estimation of the univariate cumulative. Further, these results may also be useful in defining nonparametric classification rules.

Acknowledgment. The authors thank the referee for his helpful suggestions.

REFERENCES

[1] ČENCOV, N. N. (1962). Evaluation of an unknown distribution density from observations. *Soviet Math.* **3** 1559–1562.

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

[3] SCHWARTZ, S. C. (1967). Estimation of probability density by an orthogonal series. *Ann. Math. Statist.* **38** 1261–1265.

[4] TARTER, M. and KRONMAL, R.A. (1967). A description of new computer methods for estimating the population density. *Proceedings of the ACM* **22** 511–519.

[5] TARTER, M. and KRONMAL, R. A. (1968). Estimation of the cumulative by Fourier series methods and application to the insertion problem. *Proceedings of the ACM* **23** 491–497.

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

[7] WATSON, G. N. (1938). A note on the polynomials of Hermite and Laguerre. *J. London Math. Soc.* **13** 29–32.

[8] ZYGMUND, A. (1959). *Trigonometric Series*, **2**. Cambridge, London.