[8] R. E. Rink and R. R. Mohler, "Completely controllable bilinear systems," *SIAM J. Contr.*, vol. 6, pp. 477–486, 1968.
[9] H. J. Sussmann, "Semigroup representations, bilinear approximation of input/output maps and generalized inputs," in *Mathmatics Systems*, Marchesini and Mitter, Eds. New York: Springer, 1976.
[10] R. R. Mohler and P. A. Frick, "Bilinear demographic control process," *Int. J. Policy Anal. Inf. Syst.*, 1979, to be published.
[11] K. J. Aström, "On a first-order stochastic differential equation," *Int. J. Cont.*, vol. 1, pp. 301–326, 1965.
[12] R. S. Liptser and A. N. Shiryayev, *Statistics of Random Process I-General Theory*. New York, Springer, 1977.
[13] R. Rishel, "Necessary and sufficient dynamic programming conditions of a continuous time stochastic optimal control," *SIAM J. Contr.*, vol. 8, pp. 559–571, 1970.
[14] M. H. A. Davis and P. Varaiya, "Dynamic programming conditions for partially observable stochastic systems," *SIAM J. Contr.*, vol. 11, pp. 226–267, 1973.
[15] M. H. A. Davis, "On the existence of optimal policies in stochastic control," *SIAM J. Contr.*, vol. 11, pp. 587–594, 1973.
[16] R. S. Lipster and A. N. Shiryayev, *Statistics of Random Processes II-Applications*, New York, Springer, 1978.
[17] R. R. Mohler and W. J. Kolodziej, "An optimal control problem using incomplete data," to be published.
[18] A. V. Balakrishnan, "Modeling and identification theory: a flight control application," in *Theory and Applications of Variable Structure Systems*, R. R. Mohler and A. Ruberti, Eds. New York, Academic, 1972.
[19] W. M. Wonham, "Random differential equations in control theory," in *Probabilistic Methods in Applied Mathematics*, A. T. Bharucha-Reid, Ed., vol. 2, pp. 131–212, 1974.
[20] ——, "Optimal stationary control of linear systems with state-dependent noise," *SIAM J. Contr.*, vol. 5, pp. 486–500, 1967.
[21] ——, "On a matrix Ricatti equation of stochastic control," *SIAM J. Contr.*, vol. 6, pp. 681–697, 1968.
[22] D. Kleinman, "Numerical solution of the state dependent noise problem," *IEEE Trans. Automat. Contr.*, vol. AC-21, pp. 419–420, 1976.
[23] P. McLane, "Optimal stochastic control of linear systems with state- and control-dependent disturbances," *IEEE Trans. Automat. Contr.*, vol. AC-16, pp. 793–798, 1971.
[24] ——, "Linear optimal stochastic control using instantaneous output feedback," *Int. J. Contr.*, vol. 13, pp. 383–396, 1971.
[25] D. Kleinman, "Optimal stationary control of linear systems with control-dependent noise," *IEEE Trans. Automat. Contr.*, vol. AC-14, pp. 673–677, 1969.
[26] U. Haussmann, "Optimal stationary control with state and control dependent noise," *SIAM J. Contr.*, vol. 9, pp. 184–198, 1971.
[27] M. Zakai, "A Lyapunov criterion for the existence of stationary probability distributions for systems perturbed by noise," *SIAM J. Contr.*, vol. 7, pp. 390–397, 1969.
[28] A. Mahalanabis and R. Rana, "On optimal stationary control of systems with state dependent noise," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 718–719, 1975.
[29] J. Speyer, "A non-linear control law for a stochastic infinite time problem," *IEEE Trans. Automat. Contr.*, vol. AC-21, pp. 560–564, 1976.
[30] S. Ariaratnam and P. Graefe, "Linear systems with stochastic coefficients," *Int. J. Contr.*, vol. 1, pp. 239–250, 1965.
[31] J. Bismut, "Linear-quadratic optimal stochastic control with random coefficients," *SIAM J. Contr.*, vol. 14, pp. 419–444, 1976.
[32] J. Bismut, "On optimal control of linear stochastic equations with a linear-quadratic criterion," *SIAM J. Contr.*, vol. 15, pp. 1–14, 1977.
[33] P. Graefe and P. McLane, "Optimal regulation of a class of linear stochastic systems relative to quadratic criteria," *Int. J. Contr.*, vol. 5, pp. 135–143, 1967.
[34] J. T. Lo and A. Willsky, "Stochastic control of rotational processes with one degree of freedom," *SIAM J. Contr.*, vol. 13, pp. 886–898, 1975.
[35] J. Luh and M. Maguiraga, "Differential games with state-dependent and control-dependent noises." *IEEE Trans. Automat. Contr.*, vol. AC-15, pp. 205–209, 1970.
[36] D. Kleinman, "On the stability of linear stochastic systems," *IEEE Trans. Automat. Contr.*, vol. AC-14, pp. 429–430, 1976.
[37] P. McLane, "Asymptotic stability of linear autonomous systems with state-dependent noise." *IEEE Trans. Automat. Contr.*, vol. AC-14, pp. 754–755, 1969.
[38] M. Hayase, S. Niwa, and I. Sugiura, "Stability of linear time-varying systems with state dependent noise," *IEEE Trans. Automat. Contr.*, vol. AC-21, pp. 175–176, 1976.
[39] U. Haussmann, "Stability of linear systems with control dependent noise," *SIAM J. Contr.*, vol. 11, pp. 382–394, 1973.
[40] U. Haussmann, "On the existence of moments of stationary linear systems with multiplicative noise," *SIAM J. Contr.*, vol. 12, pp. 99–105, 1974.
[41] D. Aeyels and J. Willems, "Moment stability of linear stochastic systems with solvable Lie algebras," *IEEE Trans. Automat. Contr.*, vol. AC-21, p. 285, 1976.
[42] G. Blankenship, "Stability of linear differential equations with random coefficients," *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 834–838, 1977.
[43] J. Willems, "Feedback stabilizability for stochastic systems with state-control dependent noise," *Automatica*, vol. 12, pp. 277–283, 1976.
[44] S. Marcus, D. Martin, and A. Willsky, "On the stochastic stability of linear systems containing coloured multiplicative noise," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 711–713, 1975.
[45] R. W. Brockett, "Parametrically stochastic linear differential equations," *Math. Progr. Study*, vol. 5, pp. 8–21, 1976.
[46] R. W. Brockett, "Stochastic bilinear models," in *Proc. Joint Automat. Contr. Conf.*, San Francisco, pp. 1420–1422, 1979.
[47] D. Q. Mayne and R. W. Brockett, Eds., "Lie algebras and Lie groups in control theory," *Geometric Methods in System Theory*, pp. 43–82, 1973.
[48] P. McLane, "Linear optimal estimation with state-dependent disturbances and time-correlated measurement noise," *IEEE Trans. Automat. Contr.*, vol. AC-16, pp. 198–200, 1971.
[49] ——, "Optimal linear filtering for linear systems with state-dependent noise," *Int. J. Contr.*, vol. 10, pp. 41–51, 1969.
[50] T. Nakamizo, "Moment equivalent system of linear stochastic systems with application to state estimation," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 193–202, 1974.
[51] S. Marcus and A. Willsky, "Algebraic structure and finite dimensional nonlinear estimation," *SIAM J. Math. Anal.*, vol. 9, pp. 312–327, 1978.
[52] J. T. Lo and A. Willsky, "Estimation for rotational processes with one degree of freedom," *IEEE Trans. Automat. Contr.*, vol. 20, pp. 10–21, 1975.
[53] C. Bruni, G. DiPillo, and G. Koch, "Mathematical models and identification of bilinear systems," in *Theory and Applications of Variable Structure Systems*, R. R. Mohler and A. Ruberti, Eds. New York, Academic, 1972, pp. 137–152.

Sequential Estimates of Probability Densities by Orthogonal Series and Their Application in Pattern Classification

LESZEK RUTKOWSKI

Abstract—**Recursive estimates of probability densities based on orthogonal series are proposed. Pattern classifications procedures derived from these estimates are presented and their asymptotic properties are investigated. For a Haar orthogonal system our procedures are density-free Bayes risk consistent.**

I. Introduction

Let $X_1, \cdots, X_n$ be a sequence of independent observations of a random variable X with Lebesgue density function f on the set A, where A is a subset of the real line. We assume that f has the representation

$$f(x) \sim \sum_{j=0}^{\infty} a_j g_j(x), \tag{1}$$

where

$$a_j = Eg_j(X) \tag{2}$$

and $g_j(\cdot)$, $j=0,1,2,\cdots$, is a complete orthonormal system defined on A. As the estimator of a density we can take

$$\hat{f}_n(x) = \sum_{j=0}^{N(n)} a_{jn} g_j(x), \tag{3}$$

where

$$a_{jn} = \frac{1}{n} \sum_{i=1}^{n} g_j(X_i) \tag{4}$$

Manuscript received May 29, 1980; revised September 22, 1980.
The author is with the Department of Electrical Engineering, Technical University of Czestochowa, Zawadzkiego 17, 42-200 Czestochowa, Poland.

0018-9472/80/1200-0918$00.75

and $N(n)$ is a sequence of integers such that

$$N(n) \to \infty \tag{5}$$

as $n \to \infty$. Density estimates (3) were introduced by Čencov [3] and studied by Schwartz [11], Kronmal and Tarter [9], and Bosq [2] among others.

In this correspondence the following estimator of a probability density function is proposed,

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=0}^{N(i)} g_j(X_i) g_j(x). \tag{6}$$

Observe that estimator (6) may be expressed as

$$\hat{f}_n(x) = \frac{n-1}{n} \hat{f}_{n-1}(x) + \frac{1}{n} \sum_{j=0}^{N(n)} g_j(X_n) g_j(x). \tag{7}$$

A great advantage of the definition (7) over (3) is that $\hat{f}_n$ can be computed by making use of the current observation X_n and the preceding estimator $\hat{f}_{n-1}$. Thus the unknown probability density function is estimated sequentially. It should be noted that estimator (7) is analogous to a recursive version of a Parzen–Rosenblatt kernel estimate, introduced by Wolverton and Wagner [13], [14],

$$\hat{p}_n(x) = \frac{n-1}{n} \hat{p}_{n-1}(x) + \frac{1}{nh_n} K((x - X_n)/h_n), \tag{8}$$

where the function K, the so-called kernel, and the sequence h_n satisfy suitable conditions. Recently several interesting results related to the pointwise and the integral convergence of estimator (8) have been obtained by Devroye [5]. For surveys on commonly used nonparametric probability density estimators, the reader is referred to Cover [4] and Wegman [12].

It is well known that nonparametric density estimates can be applied to many engineering problems including pattern classification, cluster analysis, and reliability theory. As an example of these applications we present pattern classification procedures with class conditional density estimates of type (6). Theorem 1 gives conditions for weak or strong consistency of class density estimates whereas Theorem 2 establishes Bayes risk consistency of pattern classification procedures. For a Haar orthogonal system our procedures are Bayes risk consistent under no restrictions imposed on class conditional densities. In [8] and [10] this problem was also treated by orthogonal series but authors supposed that class conditional densities are square integrable.

In this correspondence we assume that

$$|g_j(x)| \leqslant G_j \tag{9}$$

for all $x \in A$ where G_j is a sequence of numbers. The condition (9) is more general than

$$|g_j(x)| \leqslant \text{const} \tag{9a}$$

for all $x \in A$ which was assumed in [2], [3], [9], and [11]. It is worth recalling that Legendre and Haar orthogonal systems do not satisfy (9a) at all.

II. Pattern Classification Procedures

In this section only two category classification problems will be treated but the results can be extended to the multicategory problem. Let (T, X) be a pair of random variables; $P(T=k) = p_k$, $k = 1, 2$, X is A-valued, where A is a subset of the real line. Let f_k be a conditional density of X given $T = k$. When $f_k, p_k, k = 1, 2$, are known, a Bayes decision function,

$$D(x) = p_1 f_1(x) - p_2 f_2(x),$$

classifies every $x \in A$ as coming from class 1 if $D(x) \geqslant 0$ and from class 2 otherwise. We assume that f_k, p_k, $k = 1, 2$, are unknown and have a learning sequence,

$$(T_1, X_1), \cdots, (T_n, X_n),$$

i.e., a sample of independent observations of (T, X) that we wish to estimate the function D. Let n_k be the number of observations from the class k. Partition observations $(X_1, \cdots, X_n)$ into subsequences $(X_1^1, \cdots, X_{n_1}^1)$, $(X_1^2, \cdots, X_{n_2}^2)$. As estimates of conditional densities we propose

$$\hat{f}_k(x) = \frac{1}{n_k} \sum_{i=1}^{n_k} \sum_{j=0}^{N(i)} g_j(X_i^k) g_j(x). \tag{10}$$

The following theorem establishes weak and strong pointwise consistency of estimates (10).

Theorem 1: Assume that (5) and (9) hold. $E(\hat{f}_k(x) - f_k(x))^2 \overset{n}{\to} 0$ if

$$\frac{1}{n^2} \sum_{i=1}^{n} \left(\sum_{j=0}^{N(i)} G_j^2 \right)^2 \overset{n}{\to} 0, \tag{11}$$

$\hat{f}_k(x) \overset{n}{\to} f_k(x)$ with probability one if

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \left(\sum_{j=0}^{N(n)} G_j^2 \right)^2 < \infty, \tag{12}$$

at every point $x \in A$ at which

$$\sum_{j=0}^{N(n)} b_j^k g_j(x) \overset{n}{\to} f_k(x), \tag{13}$$

where

$$b_j^k = \int_A f_k(x) g_j(x)\, dx.$$

Proof: The proof of (11) is analogous to that in [15]. The proof of (12) is based on the strong law of large numbers. Details are left to the reader.

Let $\hat{p}_k = n_k / n$ be an estimate of p_k. Consider an empirical decision function.

$$\hat{D}(x) = \hat{p}_1 \hat{f}_1(x) - \hat{p}_2 \hat{f}_2(x) \tag{14}$$

classifying every $x \in A$ as coming from class 1 if $\hat{D}(x) \geqslant 0$ and from class 2 otherwise. By theorem 1 and theorem 2 in [7] one obtains the following theorem.

Theorem 2: Pattern classification procedures based on (14) are weakly (strongly) Bayes risk consistent if (13) holds almost everywhere (with respect to the Lebesgue measure on A).

Now we are interested in the Haar orthogonal system. For this system condition (13) holds almost everywhere without any restrictions imposed on f_1 and f_2 (see [1, theorem 1.6.1]). Consequently pattern classification procedures based on (14) are density-free Bayes risk consistent. From the construction of the Haar orthogonal system (see [1]) it follows that $G_j = \text{const}\, j^{1/2}$. Therefore conditions (11) and (12) take the form

$$\frac{1}{n^2} \sum_{i=1}^{n} N^4(i) \overset{n}{\to} 0, \qquad \sum_{n=1}^{\infty} \frac{N^4(n)}{n^2} < \infty.$$

For $N(n)$ of type n^t they are satisfied if $0 < t < 1/4$.

III. Concluding Remarks

The result in Theorem 2 is density-free for a Haar orthogonal system. Similar results for discrimination rules with Parzen–Rosenblatt kernel estimates have been obtained by Devroye and Wagner [6].

One can extend Theorems 1 and 2 to the multivariate case without much difficulty; treating g_j as the multivariate orthogonal system conclusions of these theorems hold with exactly the same assumptions.

Finally we note that procedures (14) may be alternatively expressed as

$$\hat{D}(x)=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=0}^{N(i)}(t_{1i}-t_{2i})g_j(X_i)g_j(x),$$

where

$$t_{ki}=\begin{cases}1, & \text{if } T_i=k\\ 0, & \text{otherwise.}\end{cases}$$

ACKNOWLEDGMENT

The author thanks the referee for his helpful suggestions.

REFERENCES

[1] G. Alexits, *Convergence Problems of Orthogonal Series.* Budapest, Hungary: Akademiai Kiado, 1961.
[2] D. Bosq, "Sur l'estimation d'une densite multivariee par une serie de fonctions orthogonales," *C. R. Acad. Sci. Paris*, vol. 268, pp. 555–557, 1969.
[3] N. N. Čencov, "Evaluation of an unknown distribution density from observations," *Sov. Math.*, vol. 3, pp. 1559–1562, 1962.
[4] T. Cover, "A hierarchy of probability density function estimates," in *Frontiers of Pattern Recognition*, S. Watanabe, Ed. New York: Academic, pp. 83–98, 1972.
[5] L. P. Devroye, "On the pointwise and the integral convergence of recursive kernel estimates of probability densities," *Utilitas Math.*, to be published.
[6] L. P. Devroye and T. J. Wagner, "On the L1 convergence of kernel estimators of regression functions with applications in discrimination," *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete.*, to be published.
[7] W. Greblicki, "Asymptotically optimal pattern recognition procedures with density estimates," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 250–251, Mar. 1978.
[8] ——, "Pattern recognition procedures with nonparametric density estimates," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-8, pp. 809–812, Nov. 1978.
[9] R. Kronmal and M. Tarter, "The estimation of probability densities and cumulatives by Fourier series methods," *J. Amer. Statist. Ass.*, vol. 63, pp. 925–952, 1968.
[10] J. Van Ryzin, "Bayes risk consistency of classification procedures using density estimates," *Sankhya*, Series A, vol. 28, pp. 16–170, 1966.
[11] S. C. Schwartz, "Estimation of probability density by an orthogonal series," *Ann. Math. Statist.*, vol. 38, pp. 1261–1265, 1967.
[12] E. J. Wegman, "Nonparametric probability density estimation; A summary of available methods," *Technometrics*, vol. 14, pp. 533–546, 1972.
[13] C. T. Wolverton and T. J. Wagner, "Asymptotically optimal discriminant functions for pattern recognition," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 258–265, Mar. 1969.
[14] C. T. Wolverton and T. J. Wagner, "Recursive estimates of probability densities," *IEEE Trans. Syst. Sci. Cybern.*, vol. SSC-5, p. 307, 1969.
[15] H. Yamato, "Sequential estimation of a continuous probability density function and mode," *Bull. Math. Statis*, vol. 14, pp. 1112, 1971.

Variation of Spatial Cues in Node Arrangement

WILLIAM G. BEAZLEY, MEMBER, IEEE, AND JOHN J. ALLAN, III, MEMBER, IEEE

***Abstract*—Directed and undirected graphs are a common message in communication to and from computers in many different tasks. Examples of these messages are finite element grids, signal flow graphs, flow charts, performance evaluation review technique (PERT) charts, and scale drawings of objects. Several different formats these messages might take are examined on a graphics display. Digraphs (directed graphs) are selected for study as the more general case of message. (Undirected graphs are digraphs in which directions of arcs do not matter.) The degree to which spatial cues could be utilized in learning and recalling digraphs was varied by controlling the nature and consistency of node location on a graphics screen. Thirty-eight, mechanical engineering students were required to learn three digraphs under one of four presentation conditions. Twenty-nine completed post-tests either immediately or after one day's delay. Results indicate that digraphs can be learned with significantly fewer learning errors when spatial cues are available to the subject. The increased recall errors with delay time for a regularly ordered node arrangement indicate that node formats where spatial cues are available but arcs may lie on top of one another may be difficult to learn or remember for great lengths of time. Other results are discussed. The results can be useful in guiding the presentation of digraph data to the user for his retention in his continuing work. Future possibilities for more detailed study of the helpful or harmful effects of node position and organization are indicated.**

Manuscript received March 26, 1979; revised March 26, 1980, September 10, 1980.

W. G. Beazley was with the Department of General Engineering, University of Illinois, Urbana, IL 61801. He is now with the Department of Marine Structures and Naval Architecture, Brown and Root Inc., P.O. Box 3, Houston, TX 77001.

J. J. Allan, III, is with Productivity International Incorporated, 5622 Dyer Street, Dallas, TX 75205.

I. INTRODUCTION

Many types of messages are composed of individual items with an indicated relationship. Examples range from words and their position in a list [1] to pictorial items and their relative position in a scene [2]. One way to represent these messages is to use directed or undirected (directed graphs in which the directions do not matter [3]) graphs as a basis, with the general characteristics of the nodes and arcs indicating the particular items and relationships of the message. Paynter [4] has asserted that partitioning systems into networks is a basic design activity and it has been demonstrated that a display based on directed graph (digraph) concepts can be used as a designer–computer interface for assembling components of a system model [5]. Communication of many messages can be viewed, at least in part, as communication of directed graphs.

There are many ways in which the nodes and arcs of the graph can be varied to complete the message when communicated by a spatial media, i.e., a drawing or computer graphic display. They can be labeled directly, as in signal flow graphs and PERT charts. In addition, the appearance of the nodes and arcs can be changed, as in computer program flow charts and in logic diagrams. Finally, they may also communicate by their spatial position on the media as in finite element grids, data plots [6], scale drawings [7], maps [8], and vectors [9]. When computer graphic displays or other *spatial* media are used to communicate directed graph-based messages, user performance is often improved. Rouse [10], [11] has found fault logic displays helpful to fault diagnosis. Corley and Allan [12], using a pipe layout task with a computer graphic display, showed that tablet data entry was superior to keyboard or a keyboard–tablet combination.

While the format of directed graph-*based* messages has been studied [1], [2], [6]–[13], little is known about the effects of different formats on a directed graph alone. It has not been established in the literature that spatial cues play an important role in learning digraphs, though they commonly appear as drawings or pictorial displays. The nodes and arcs are, in practice, organized in different formats depending on the task being performed. The finite element specialist might need a regularly ordered display to find mislocated or missing nodes and arcs. The computer programmer might require nodes to be grouped according to program structure. The PERT chart analyst may require the critical path to occupy the center of the display, and less critical activity arcs on the periphery. In the interactive computer graphics environment, retention of these node placements from use to use may be important in allowing spatial cues in the appearance of the digraph to refresh the memory of the user.

One way of gauging the effectiveness of a message format is to study how well its content can be learned and how long it can be