The Annals of Mathematical Statistics
1972, Vol. 43, No. 3, 1030-1034

A REMARK ON QUADRATIC MEAN DIFFERENTIABILITY[1]

BY BRUCE LIND AND GEORGE ROUSSAS

University of Wisconsin, Madison

1. Introduction. In recent years, some authors (see, for example, LeCam [5], [6], Roussas [7], [8], Johnson and Roussas [2], [3], [4]) have replaced the usual assumption of pointwise differentiability of certain random functions depending on a parameter θ by differentiability in quadratic mean (q.m.) for the purpose of obtaining distributional theory and/or making statistical inferences. When the parameter θ is real valued, one may conveniently employ well-known theorems (Vitali's theorem, for example) for checking differentiability in q.m. This process, however, does not avail itself equally well when the underlying parameter θ is multi-dimensional.

The purpose of the present note is to reduce, in essence, the multi-dimensional parameter case to the one-dimensional situation. This reduction is made precise in Theorem 2.1.

In the remaining part of this section, we introduce the necessary terminology and notation, and in the following Section, we formulate and prove first an auxiliary result (Lemma 2.1) and then the main result of this note (Theorem 2.1). Also an additional theorem (Theorem 2.2), relating pointwise differentiability and differentiability in q.m., is stated and proved in the same section. More results along these lines will be presented elsewhere.

Let $(\Omega, \mathscr{A}, P)$ be a probability space and let $f(\theta) = f(\omega; \theta)$ be a real-valued random function defined on $\Omega \times \Theta$, where Θ is an open subset of R^k. Let

$$L_2(\Omega) = L_2(\Omega, P) = \{\text{rv's } X \text{ on } (\Omega, \mathscr{A}, P);\ \mathscr{E}X^2 < \infty\}\,,$$

and for X, $Y \in L_2(\Omega)$, define the inner product $\langle X, Y\rangle$ as follows

$$\langle X, Y\rangle = \mathscr{E}(XY)\,.$$

Denote by $||\cdot||_2$ the L_2-norm induced by the inner product $\langle\cdot,\cdot\rangle$; i.e., for $X \in L_2(\Omega)$,

$$||X||_2 = (\langle X, X\rangle)^{\frac{1}{2}}\,.$$

DEFINITION 1.1. The random function $f(\theta)$ is said to be differentiable in q.m. at θ when P is employed, if there exists a k-dimensional vector of random functions, $\dot{f}(\theta)$, such that

$$|h|^{-1}||f(\theta + h) - f(\theta) - h'\dot{f}(\theta)||_2 \to 0 \qquad \text{as} \quad 0 \neq |h| \to 0\,,$$

where $|\cdot|$ denotes the usual Euclidean norm of the vector h; $\dot{f}(\theta)$ is the q.m. derivative of $f(\theta)$ at θ. Here "′" denotes transpose and $h'\dot{f}(\theta)$ is the inner product of the indicated vectors.

Received August 2, 1971; revised October 12, 1971.

[1] This research was supported by the National Science Foundation, Grant GP-20036.

2. The results. We now formulate the main result of this note, namely,

THEOREM 2.1. *For each $\theta \in \Theta$, we assume that the partial derivatives in* q.m. *of the random function $f(\theta)$ exist and are continuous (in θ) in the L_2-norm, $\|\cdot\|_2$. Then $\dot{f}(\theta)$, the derivative in* q.m. *of $f(\theta)$, exists and is equal to the vector of partial derivatives in* q.m. *That is, for each $\theta \in \Theta$ and $h \neq 0$,*

$$\| \, |h|^{-1}[f(\theta + h) - f(\theta) - \textstyle\sum_{j=1}^{k} h_j \dot{f}_j(\theta)]\|_2 \to 0 \qquad \text{as} \quad |h| \to 0 \,, \tag{2.1}$$

where $\dot{f}_j(\theta), j = 1, \cdots, k$ are the partial q.m. *derivatives of $f(\theta)$ at θ and $h = (h_1, \cdots, h_k)'$.*

To facilitate the proof of the theorem, we first establish the following result.

LEMMA 2.1. *Let $F(t) = F(\omega, t)$, $t \in T$, an open subset of R, be a real-valued random function defined on $\Omega \times T$ which is differentiable in* q.m. *on T with* q.m. *derivative $\dot{F}(t)$. Then for every $g \in L_2(\Omega)$, independent of t, we have*

$$\frac{d}{dt}\langle F(t), g\rangle = \langle \dot{F}(t), g\rangle \,.$$

PROOF. For $0 \neq l \in R$ such that t, $t + l \in T$, one has

$$\begin{aligned} |l^{-1}[\langle F(t+l), g\rangle - \langle F(t), g\rangle] - \langle \dot{F}(t), g\rangle| & \\ = |l^{-1}\langle F(t+l) - F(t), g\rangle - \langle \dot{F}(t), g\rangle| & \\ = |\langle l^{-1}[F(t+l) - F(t)] - \dot{F}(t), g\rangle| & \\ \leqq \|l^{-1}[F(t+l) - F(t)] - \dot{F}(t)\|_2\|g\|_2 \to 0 & \qquad \text{as} \quad l \to 0 \,. \end{aligned}$$

The proof of the lemma is complete. Of course, the lemma implies that $\langle F(t), g\rangle$ is continuous on T.

PROOF OF THEOREM 2.1. For the proof of the theorem, it will be convenient to introduce the following notation.

$$\begin{aligned} f^*(\theta, h, j) &= f(\theta_1, \cdots, \theta_j, \theta_{j+1} + h_{j+1}, \cdots, \theta_k + h_k) \\ f^*(\theta, h, 0) &= f(\theta + h) \\ f^*(\theta, h, k) &= f(\theta) \\ \dot{f}_1{}^*(\theta, h, j) &= \dot{f}_1(\theta_1, \cdots, \theta_j, \theta_{j+1} + h_{j+1}, \cdots, \theta_k + h_k) \,. \end{aligned}$$

Consider the following identity

$$\begin{aligned} f(\theta + h) - f(\theta) &= f^*(\theta, h, 0) - f^*(\theta, h, k) \\ &= [f^*(\theta, h, 0) - f^*(\theta, h, 1)] + [f^*(\theta, h, 1) - f^*(\theta, h, 2)] \\ &\quad + \cdots + [f^*(\theta, h, k-1) - f^*(\theta, h, k)] \,. \end{aligned} \tag{2.2}$$

Then instead of (2.1), it suffices to establish the following equivalent formulation. Namely, for every $\varepsilon > 0$, there exists a $\delta = \delta(\varepsilon) > 0$ such that $|h| < \delta$ implies

$$\|[f(\theta + h) - f(\theta)] - \textstyle\sum_{j=1}^{k} h_j \dot{f}_j(\theta)\| < k^2\varepsilon|h| \,. \tag{2.3}$$

Inserting (2.2) in (2.3), the assertion may be stated as follows: $|h| < \delta$ implies

$$\begin{aligned} &\|\{[f^*(\theta, h, 0) - f^*(\theta, h, 1)] - h_1\dot{f}_1(\theta)\} \\ (2.4) \qquad &\quad + \{[f^*(\theta, h, 1) - f^*(\theta, h, 2)] - h_2\dot{f}_2(\theta)\} + \cdots \\ &\quad + \{[f^*(\theta, h, k-1) - f^*(\theta, h, k)] - h_k\dot{f}_k(\theta)\}\|_2 < k^2\varepsilon\,|h|\,. \end{aligned}$$

We first show that there exists a δ such that $|h| < \delta$ implies

$$(2.5) \qquad \|[f^*(\theta, h, 0) - f^*(\theta, h, 1)] - h_1\dot{f}_1(\theta)\|_2 < k\varepsilon\,|h|\,.$$

Let $g \in L_2(\Omega)$ be independent of θ, but otherwise arbitrarily chosen. Then by our assumptions and Lemma 1.1, we have

$$\frac{\partial}{\partial\theta_1}\langle f^*(\theta, h, 1), g\rangle = \langle \dot{f}_1^*(\theta, h, 1), g\rangle\,.$$

By the same lemma, we have that $\langle f(t, \theta_2 + h_2, \cdots, \theta_k + h_k), g\rangle$ as a function of t, is continuous in the closed interval with end-points $\theta_1, \theta_1 + h_1$. Thus the mean value theorem (for real-valued functions of a real variable) applies to $\langle f(t, \theta_2 + h_2, \cdots, \theta_k + h_k), g\rangle$, regarded as a function of t, and gives

$$(2.6) \qquad \langle f^*(\theta, h, 0) - f^*(\theta, h, 1), g\rangle = h_1\langle \dot{f}_1(t_1, \theta_2 + h_2, \cdots, \theta_k + h_k), g\rangle\,,$$

where t_1 lies between θ_1 and $\theta_1 + h_1$ (and, in general, also depends on $\theta_2, \cdots, \theta_k$, $h_2, \cdots, h_k$ and g). But

$$\begin{aligned} &\langle \dot{f}_1(t_1, \theta_2 + h_2, \cdots, \theta_k + h_k), g\rangle \\ (2.7) \quad &= \langle \dot{f}_1(\theta) + [\dot{f}_1(t_1, \theta_2 + h_2, \cdots, \theta_k + h_k) - \dot{f}_1^*(\theta, h, 1)] \\ &\quad + [\dot{f}_1^*(\theta, h, 1) - \dot{f}_1^*(\theta, h, 2)] + \cdots + [\dot{f}_1^*(\theta, h, k-1) - \dot{f}_1(\theta)], g\rangle\,. \end{aligned}$$

By means of (2.6) and (2.7), one then has

$$\begin{aligned} &\langle[f^*(\theta, h, 0) - f^*(\theta, h, 1)] - h_1\dot{f}_1(\theta), g\rangle \\ (2.8) \qquad &\quad = h_1\langle[\dot{f}_1(t_1, \theta_2 + h_2, \cdots, \theta_k + h_k) - \dot{f}_1^*(\theta, h, 1)], g\rangle \\ &\quad\quad + h_1\langle[\dot{f}_1^*(\theta, h, 1) - \dot{f}_1^*(\theta, h, 2)], g\rangle \\ &\quad\quad + \cdots + h_1\langle[\dot{f}_1^*(\theta, h, k-1) - \dot{f}_1(\theta)], g\rangle\,. \end{aligned}$$

One of our assumptions is that $\dot{f}_1(\theta)$ is continuous in $\|\cdot\|_2$-norm. Thus, since t_1 lies between θ_1 and $\theta_1 + h_1$, it follows that for δ_1 sufficiently small and h such that $|h| < \delta_1$, one has

$$\begin{aligned} &\|\dot{f}_1(t_1, \theta_2 + h_2, \cdots, \theta_k + h_k) - \dot{f}_1^*(\theta, h, 1)\|_2 < \varepsilon \\ (2.9) \qquad &\|\dot{f}_1^*(\theta, h, 1) - \dot{f}_1^*(\theta, h, 2)\|_2 < \varepsilon \\ &\cdots\cdots\cdots\cdots\cdots \\ &\|\dot{f}_1^*(\theta, h, k-1) - \dot{f}_1(\theta)\|_2 < \varepsilon\,. \end{aligned}$$

Taking absolute values on both sides of (2.8) and utilizing (2.9), we obtain

$$(2.10) \quad |\langle[f^*(\theta, h, 0) - f^*(\theta, h, 1)] - h_1\dot{f}_1(\theta), g\rangle| \leqq k\,\varepsilon\,|h_1|\,\|g\|_2 \leqq k\varepsilon\,|h|\,\|g\|_2\,,$$

the last inequality following since $|h_1| \leqq |h|$. If we now take the supremum of both sides of (2.10) over $g \in L_2(\Omega)$, such that $||g||_2 = 1$, the left-hand side will be equal to

$$||[f^*(\theta, h, 0) - f^*(\theta, h, 1)] - h_1 \dot{f}_1(\theta)||_2 .$$

(See, for example, Rudin [9] pages 128–130). Thus, we get

$$||[f^*(\theta, h, 0) - f^*(\theta, h, 1)] - h_1 \dot{f}_1(\theta)||_2 < k\varepsilon \, |h| ,$$

which is (2.5).

Inequalities analogous to (2.5) can be established for the other terms on the left-hand side of (2.4). Applying the triangle inequality for norms to (2.4) gives the desired result.

The theorem just proved changes the problem from a search for a vector of functions and the taking of limits as a vector variable approaches zero in norm to one of finding the partial derivatives in q.m., which involves only a single function and a real argument, and then checking to see that they are continuous in $||\cdot||_2$-norm.

In the final theorem of this section, we present another result, Theorem 2.2, which also deals with the case $k > 1$. This theorem gives another method of finding the q.m. derivative by relating it to the pointwise partial derivatives.

THEOREM 2.2. *For each $\theta \in \Theta$, assume that the pointwise partial derivatives of the random function $f(\theta)$ exist and are continuous, belong to $L_2(\Omega)$ and satisfy the following condition: For each $\theta_0 \in \Theta$, there exists a neighborhood containing θ_0 in which*

$$(2.11) \qquad |\hat{f}_j(\omega; \theta) - \hat{f}_j(\omega; \theta_0)| < H_{\theta_0}^{(j)}(\omega) \quad \text{a.s.} \quad [P]$$

for each θ in the neighborhood, where $H_{\theta_0}^{(j)} \in L_2(\Omega)$, $j = 1, \cdots, k$ and $\hat{f}_j$ denotes the pointwise partial derivative of $f(\theta)$ with respect to θ_j. Under the above assumptions, $\dot{f}(\theta)$ exists and is equal to the vector of pointwise partial derivatives.

PROOF. We first remark that the uniformity condition above, clearly, implies continuity in the $||\cdot||_2$-norm.

Next, for $t \neq 0$, we have

$$(2.12) \qquad || \, |h|^{-1}[f(\theta + h) - f(\theta) - h'(\hat{f}_1(\theta), \cdots, \hat{f}_k(\theta))']||_2 = |h|^{-1}||f(\theta + h) - f(\theta) - \textstyle\sum_{j=1}^k h_j \hat{f}_j(\theta)||_2 .$$

Now, by the mean value theorem for functions of several variables (see, for example, Franklin [1] page 336), we have, under the assumptions of the theorem,

$$(2.13) \qquad f(\theta + h) - f(\theta) = \textstyle\sum_{j=1}^k \hat{f}_j(\theta + \alpha(\omega)h)h_j \quad \text{a.s.} \quad [P]$$

for some $0 < \alpha(\omega) < 1$.

Using this in (2.12), we obtain

$$(2.14) \qquad |h|^{-1}||f(\theta + h) - f(\theta) - \textstyle\sum_{j=1} \hat{f}_j(\theta)h_j||_2 = |h|^{-1}||\textstyle\sum_{j=1}^k h_j[\hat{f}_j(\theta) - \hat{f}_j(\theta + \alpha(\omega)h)]||_2 = |h|^{-1}||(h, \hat{f}(\theta) - \hat{f}(\theta + \alpha(\omega)h))||_2 ,$$

where $(\cdot, \cdot)$ denotes the usual inner product on R^k and $\hat{f}(\theta)$ denotes the vector of pointwise partial derivatives of $f(\theta)$. By the Cauchy-Schwarz inequality, applied to the Euclidean inner product and norm, the last expression on the right-hand side of (2.14) is bounded by

$$\begin{aligned} |h|^{-1} \| |h| \, |\hat{f}(\theta) - \hat{f}(\theta + \alpha(\omega)h)| \, \|_2 & \\ \leqq \| \textstyle\sum_{j=1}^{k} |\hat{f}_j(\theta) - \hat{f}_j(\theta + \alpha(\omega)h)| \, \|_2 & \\ \leqq \textstyle\sum_{j=1}^{k} \|\hat{f}_j(\theta) - \hat{f}_j(\theta + \alpha(\omega)h)\|_2 \, . & \end{aligned}$$

By the Dominated convergence theorem, the uniformity condition, along with the continuity of $f_j, j = 1, \cdots, k$, implies that each individual term on the right hand side of the last expression above approaches zero as $|h| \to 0$. The proof is thus complete.

The use of the theorems in this paper will be demonstrated in a forthcoming paper which will be devoted to the calculation of some asymptotically optimal tests for certain failure distributions.

REFERENCES

[1] Franklin, P. (1964). *A Treatise on Advanced Calculus*. Dover Publications, Inc. New York.

[2] Johnson, R. A. and Roussas, G. G. (1969). Asymptotically most powerful tests in Markov processes. *Ann. Math. Statist.* **40** 1207–1215.

[3] Johnson, R. A. and Roussas, G. G. (1970). Asymptotically optimal tests in Markov processes. *Ann. Math. Statist.* **41** 918–938.

[4] Johnson, R. A. and Roussas, G. G. (1971). Applications of contiguity to multiparameter hypotheses testing. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* (to appear).

[5] LeCam, L. (1960). Locally asymptotically normal families of distributions. *Univ. California Publ. Statist.* **3** 37–98.

[6] LeCam, L. (1966). Likelihood functions for large numbers of independent observations. *Research Papers in Statistics*, ed. F. N. David. Wiley, New York.

[7] Roussas, G. G. (1965). Asymptotic inference in Markov processes. *Ann. Math. Statist.* **36** 978–992.

[8] Roussas, G. G. (1968). Some applications of the asymptotic distribution of likelihood functions to the asymptotic efficiency of estimates. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **10** 252–260.

[9] Rudin, W. (1966). *Real and Complex Analysis*. McGraw-Hill, New York.