# A REMARK ON THE r-th MEAN DIFFERENTIABILITY

**GEORGE D. STAMATELOS**

Department of Mathematics
University of Patras - Greece

ABSTRACT. This paper is concerned with the r-th mean differentiability. In the
mathematical developments regarding the asymptotic expansion and the asymptotic
distribution of the likelihood function, there arises the question whether the
assumptions made on the model imply differentiability in the r'-th mean of the under-
lying random functions, for integer values r'<r. The present paper provides an
answer to this question and also gives the explicit form of the derivatives in the
r'-th mean involved.

1.  INTRODUCTION AND SUMMARY.

For $n \geq 0$, integer, let $X_0, X_1, \ldots, X_n$ be the first $n+1$ r.v´.s from a
stochastic process, defined on the probability space $(X, A, P_\theta)$, whose pro-
bability law depends on a k-dimensional parameter $\theta \in \Theta$, an open subset of
$\mathbb{R}^k$, $k \geq 1$ and satisfies certain regularity conditions. Then, for $\theta, \theta^* \in \Theta$,
let

$$q_j(\theta;\theta^*) = q_j(X_j;\theta,\theta^*) = \frac{dP_{n,\theta^*}}{dP_{n,\theta}}, \quad n \geq 0$$

be specified versions of the Radon-Nikodym derivatives involved, where $X_j = (X_0, X_1, \ldots, X_n)$ and $P_{n,\theta}$ is the restriction of $P_\theta$ to the σ-field $A_n = \sigma(X_0, X_1, \ldots, X_n)$ induced
by the r.v´.s $X_0, X_1, \ldots, X_n$. Also, for any two parameter points $\theta$ and $\theta^*$, let
$q_j(X_j|X_{j-1};\theta,\theta^*)$ be the quotient of the probability density functions of the
random vectors $X_j = (X_0, X_1, \ldots, X_j)$ and $X_{j-1}$. Then, as is well known, for large
sample statistical inference, based on the likelihood function, the unde-
rlying conditions are set primarily on the quantities, $q_j(X_j|X_{j-1};\theta,\theta^*)$,
$j=1,2,\ldots n$. Typically these conditions include pointwise differentiabi-
lity with respect to the "moving" parameter $\theta^*$ of up to third order. It
is also known, however, that these conditions fail to be satisfied, for

instance, in such a simple and interesting case as that of the double exponential distribution.

In relatively recent asymptotic statistical work, the above menti-oned type of conditions have been replaced by conditions refering to the differentiability in the r-th mean of the quantities

$$\varphi_j(\theta,\theta^*;r) = \left[ q_j(\underset{\sim}{X}_j | \underset{\sim}{X}_{j-1} ; \theta,\theta^*) \right]^{1/r} , \quad j \geq 1, \text{ for some integer } r \geq 1. \quad (1.1)$$

These conditions do not suffer from the kind of inadequacies mentioned above and, in addition, are of probabilistic rather than of analytical nature. (See, for example, LeCam [1,2], Johnson and Roussas [3-5], Rous-sas [6], Lind and Roussas [7-8] Akritas [9] and, in particular, Akritas and Roussas [10].)

2.  ASSUMPTIONS AND MAIN RESULT.

In this section, the assumptions are formulated under which the main result of this note holds true. To this end, let $\Theta$ be an open subset of $\mathbb{R}^k$, $k \geq 1$, and for each $\theta \in \Theta$, let $X_0, X_1, \ldots$ be r.v´.s defined on the pro-bability space $(X, A, P_\theta)$ and taking values in $(S, S)$; here S is a Borel su-bset of a Euclidean space and $S$ is the $\sigma$-field of Borel subsets of S. These r.v´.s come from a certain class of stochastic processes which sa-tisfy suitable conditions to be explicitly mentioned below. Let $A_n$ be the $\sigma$-field induced by the first n+1 r.v´.s, $A_n = \sigma(X_0, X_1, \ldots, X_n)$, and let $P_{n,\theta}$ be the restriction of $P_\theta$ to $A_n$; that is,

$$P_{n,\theta} = P_\theta | A_n, \quad A_n = \sigma(X_0, X_1, \ldots, X_n). \quad (2.1)$$

It will be assumed in the following that, for each $\theta, \theta^* \in \Theta$, $P_{n,\theta} \approx P_{n,\theta^*}$ for all $n \geq 0$. Thus, the quantities $\omega_j(\theta,\theta^*;r)$, $j \geq 1$, are well defined by (1.1). Then we have

Assumptions

(A1) For each $n \geq 0$, the probability measures $\{P_{n,\theta}; \theta \in \Theta\}$, defined by (2.1), are mutually absolutely continuous.

For $\theta, \theta^*$ in $\Theta$, define $\omega_j(\theta,\theta^*;r)$ by (1.1) Then

(A2) For each $\theta \in \Theta$, the random functions $\varphi_j(\theta,.;r)$ are differentiab-le $[P_\theta]$ in the r-th mean at $\theta$, uniformly in $j \geq 1$, for some integer $r \geq 1$.

Let $\dot{\varphi}_j(\theta;r)$, $j \geq 1$, be the r-th mean derivatives involved. Then

(A3) For each $\theta \in \Theta$ and each $h \in \mathbb{R}^k$,

$$E_\theta | h´\dot{\varphi}_j(\theta;r) |^r \leq M_r(\theta,h) \quad (<\infty), \quad j \geq 1.$$

We may now formulate the main result of this paper, namely,

Theorem 2.1.  Let the number $r_k$ be defined by

$$r_k = \frac{r}{r-k} , \quad k = r-1, \ldots, \frac{2r}{3}, \frac{r}{2}. \quad (2.2)$$

and let the random functions $\varphi_j(\theta, .; r_k)$, $j \geq 1$, be defined by (1.1) with $r$ being replaced by $r_k$. Then, for each $\theta \epsilon \Theta$ and under assumptions (A1)-(A3), these random functions are differentiable in the $r_k$-th mean $[P_\theta]$ at $\theta$, uniformly in $j \geq 1$. The $r_k$-th mean derivative $\dot{\varphi}_j(\theta; r_k)$ is given by

$$\dot{\varphi}_j(\theta; r_k) = (r-k)\dot{\varphi}_j(\theta; r) \qquad (2.3)$$

and

$$E_\theta |h \acute{\dot{\varphi}}_j(\theta; r_k)|^{r_k} \leq M_{r_k}(\theta, h) \ (<\infty), \ j \geq 1. \qquad (2.4)$$

## 3. PROOF OF MAIN RESULT.

In the course of the proof of Theorem 2.1, the following auxiliary result will be needed which is formulated here as a lemma. Namely,

Lemma 3.1  For $n$, $j \geq 1$ integers, let $X_j$ and $X_{nj}$ be r.v´.s defined on the probability space $(\Omega, F, P)$ and suppose that, for some $r > 0$,

$$X_{nj} \xrightarrow{(r)} X_j, \text{ uniformly in } j \geq 1.$$

Then

$$E|X_{nj}|^r \longrightarrow E|X_j|^r, \text{ uniformly in } j \geq 1.$$

Proof.  For the two cases $0 < r \leq 1$ and $r \geq 1$, use the $c_r$-inequality and the Minkowski inequality, respectively (see, for example, Loève [11] pages 1955-1956), in order to obtain

$$|E^{1/r}|X_{nj}|^r - E^{1/r}|X_j|^r| \leq E|X_{nj} - X_j|^r + E^{1/r}|X_{nj} - X_j|^r.$$

Since the right hand side above tends to 0, uniformly in $j \geq 1$, so does the left hand side. The proof is completed.

We may now proceed with the proof of the main result, namely,

Proof of Theorem 2.1.  To show that, uniformly on bounded sets of $h$ and as $(0<)\lambda \to 0$,

$$E_\theta |\frac{1}{\lambda}[\varphi_j(\theta, \theta + \lambda h; r_k) - 1] - h \acute{\dot{\varphi}}_j(\theta; r_k)|^{r_k} \longrightarrow 0, \text{ uniformly in } j \geq 1, \qquad (3.1)$$

and that relation (2.4) holds.

Suppose for a moment that (2.3) holds true. Then  the  fact  that $r_k \leq r$ implies that

$$E_\theta |h \acute{\dot{\varphi}}_j(\theta, r_k)|^{r_k} \leq (r-k)^r E_\theta^{r_k/r} |h \acute{\dot{\varphi}}_j(\theta; r)|^r.$$

Thus, (2.4) is satisfied by means of assumptions (A3).

In the remaining part of the proof of this theorem, all convergences will be taken as above, that is, for $(0<)\lambda \to 0$ and uniformly on boun-

ded sets of h. Relation (3.1) holds true for $r_k = r$ by **assumption** (A2).
Assume it to hold true for some $r_k = 1, 2, \ldots, r-2$ and establish it for
$r_k - 1 = r_{k-\delta}$ where $\delta$ is defined by the relation

$$\delta = \frac{(r-k)^2}{k}.$$

To this end, consider the relation

$$\varphi_j(\theta, \theta + \lambda h; r_k - 1) = \varphi_j(\theta, \theta + \lambda h; r_{k-\delta}) = \varphi_j(\theta, \theta + \lambda h; r_k)\varphi_j(\theta, \theta + \lambda h; \frac{r}{\delta}) \qquad (3.2)$$

which follows from (1.1). Then by means of (3.2) and (2.3), one has the
following identity

$$\frac{1}{\lambda}\left[\varphi_j(\theta, \theta + \lambda h; r_k - 1) - 1\right] - h\dot\varphi_j(\theta; r_k - 1)$$

$$= \frac{1}{\lambda}\left[\varphi_j(\theta, \theta + \lambda h; r_{k-\delta}) - 1\right] - h\dot\varphi_j(\theta; r_{k-\delta})$$

$$= \left\{\frac{1}{\lambda}\left[\varphi_j(\theta, \theta + \lambda h; r_k) - 1\right] - h\dot\varphi_j(\theta; r_k)\right\}$$

$$+ \left\{\frac{1}{\lambda}\left[\dot\varphi_j(\theta, \theta + \lambda h; \frac{r}{\delta}) - 1\right] - h\dot\varphi_j(\theta; \frac{r}{\delta})\right\}$$

$$+ \left\{\frac{1}{\lambda}\left[\dot\varphi_j(\theta, \theta + \lambda h; r_k) - 1\right]\left[\varphi_j(\theta, \theta + \lambda h; \frac{r}{\delta}) - 1\right]\right\}. \qquad (3.3)$$

Hence the $c_r$-inequality gives

$$E_\theta\left|\frac{1}{\lambda}\left[\varphi_j(\theta, \theta + \lambda h; r_k - 1) - 1\right] - h\dot\varphi_j(\theta; r_k - 1)\right|^{r_k - 1}$$

$$\leq c^2_{r_k - 1} E_\theta\left|\frac{1}{\lambda}\left[\varphi_j(\theta, \theta + \lambda h; r_k) - 1\right] - h\dot\varphi_j(\theta; r_k)\right|^{r_k - 1}$$

$$+ c^2_{r_k - 1} E_\theta\left|\frac{1}{\lambda}\left[\varphi_j(\theta, \theta + \lambda h; \frac{r}{\delta}) - 1\right] - h\dot\varphi_j(\theta; \frac{r}{\delta})\right|^{r_k - 1}$$

$$+ c^2_{r_k - 1} E_\theta\left|\frac{1}{\lambda}\left[\varphi_j(\theta, \theta + \lambda h; r_k) - 1\right]\left[\dot\varphi_j(\theta, \theta + \lambda h; \frac{r}{\delta}) - 1\right]\right|^{r_k - 1}$$

$$= I_1(\lambda, j) + I_2(\lambda, j) + I_3(\lambda, j), \qquad (3.4)$$

where the quantities $I_i(\lambda, j)$ stand for the i-th term, i=1,2,3 on the
right hand side of relation (3.4). Since for all values of k for which
$r_k$ is integer $r_k - 1 < r_k < r$,

$$I_1(\lambda, j) \leq c^2_{r_k - 1} E_\theta^{(r_k - 1)/r_k}\left|\frac{1}{\lambda}\left[\varphi_j(\theta, \theta + \lambda h; r_k) - 1\right] - h\dot\varphi_j(\theta; r_k)\right|^{r_k} \to 0, \qquad (3.5)$$

uniformly in $j \geq 1$, by induction hypothesis.

Next,

$$I_2(\lambda,j) \leq c^2_{r_k-1} \, E_\theta^{\delta(r_k-1)/r} \, |\frac{1}{\lambda}\left[\varphi_j(\theta,\theta+\lambda h;\frac{r}{\delta})-1\right] - h\dot{\varphi}_j(\theta;\frac{r}{\delta})|^{\frac{r}{\delta}} \to 0, \quad (3.6)$$

uniformly in $j \geq 1$, by hypothesis (A2).

Finally to the term $I_3(\lambda,j)$ apply the Hölder inequality with

$$s = \frac{r}{k} \quad \text{and} \quad t = \frac{r}{r-n}$$

in order to get

$$I_3(\lambda,j) \leq c_{r_k-1} \, E_\theta^{k/r} \, |\frac{1}{\lambda}\varphi_j(\theta,\theta+\lambda h;r_k)-1|^{[(r-1)/:]\cdot[r/k]}.$$

$$E_\theta^{(r-k)/r} |\varphi_j(\theta,\theta+\lambda h;\frac{r}{\delta})-1|^{(r_k-1)\cdot r/(r-k)}$$

$$= c_{r_k-1} \, E_\theta^{k/r} \, |\frac{1}{\lambda}\left[\varphi_j(\theta,\theta+\lambda h;r_k)-1\right]|^{r_k} \cdot E_\theta^{(r-k)/r}|\varphi_j(\theta,\theta+\lambda h;\frac{r}{\delta})-1|^{r/\delta}. \quad (3.7)$$

By the induction hypothesis and Lemma 3.1,

$$E_\theta|\frac{1}{\lambda}\varphi_j(\theta,\theta+\lambda h;r_k)-1|^{r_k} \to E_\theta|h\dot{\varphi}_j(\theta;r_k|^{r_k}, \quad (3.8)$$

uniformly in $j \geq 1$.

Also, by assumptions (A2), (A3) and Lemma 3.1,

$$E_\theta|\varphi_j(\theta,\theta+\lambda h;\frac{r}{\delta})-1|^{r/\delta} \to 0, \text{ uniformly in } j \geq 1. \quad (3.9)$$

Relations (3.7)-(3.9) imply that $I_3(\lambda,j) \to 0$, uniformly in $j \geq 1$.

Hence, from (3.5), (3.6) and (3.9), relation (3.4) gives

$$E_\theta|\frac{1}{\lambda}\left[\varphi_j(\theta,\theta+\lambda h;r_k-1)-1\right] - h\dot{\varphi}_j(\theta;r_k-1)|^{r_k-1} \to 0,$$

uniformly in $j \geq 1$, which completes the proof of the theorem.

ACKNOWLEDGEMENT. The author wishes to express his sincere thanks to referee for his useful comments.

## REFERENCES

1. LECAM, L. Locally asymptotically normal families of distributions. Univ. California Publ. Statist. 3, (1960) 37-98.

2. LECAM, L. On the assumptions used to prove asymptotic normality of maximum likelihood estimates. Ann. Math. Statist. 41, (1970) 802-828.

3. JOHNSON, R. A. and ROUSSAS, G.G. Asymptotically most powerful tests in Markov processes. Ann. Math. Statist. 40, (1969) 1207-1215.

4.   JOHNSON, R.A. and ROUSSAS, G.G. Asymptotically tests in Markov pro-
     cesses. Ann. Math. Statist. 41 (1970) 918-938.

5.   JOHNSON, R. A. and ROUSSAS, G.G. Applications of contiguity to mul-
     tiparameter hypothesis testing. Proc. 6th. Berkeley Symp. of Math.
     Statist. and Prob. 1 (1972) 195-226.

6.   ROUSSAS, G.G. Contiguity of Probability Measures: Some Applications
     in Statistics. Cambridge University Press, 1972.

7.   LIND, B. and ROUSSAS, G.G. A remark on quadratic mean differentiabi-
     lity. Ann. Math. Statist. 43, (1972) 1030-1034.

8.   LIND, B. and ROUSSAS, G.G. Cramér-type conditions and quadratic mean
     differentiability. Ann. Inst. Statist. Math. 29, (1977) Part A,
     189-201.

9.   AKRITAS, M.G. Contiguity of probability measures associated with con-
     tinuous time stochastic processes. Ph. D. dissertation, Univ. of Wis-
     consin, 1978.

10.  AKRITAS, M.G. and ROUSSAS, G.G. (1978). A characterization of conti-
     guity by sample size and parameter rates. Symposia Mathematica, In-
     stituto Nazionale di Alta Mathematica, Vol. 25, (1980) 11. 155-170.

11.  LOÈVE, M. Probability Theory (3rd ed.) Van Nostrand, New York, 1963.