Discussiones Mathematicae
Algebra and Stochastic Methods 18(1998) 155–165

ON LOCALLY UNIFORM EXPANSIONS OF REGULAR FUNCTIONALS

Tadeusz Bednarski

Department of Mathematics, Pedagogical University
pl. Słowianski 6, 65–069 Zielona Góra, Poland
and Institute of Mathematics of the Polish Academy of Sciences

AND

Brenton R. Clarke

Mathematics and Statistics, Division of Science Murdoch University
Murdoch, W.A. 6150, Australia

Abstract

The aim of this paper is to show that the concept of Fréchet differentiability of von Mises statistical functionals appears naturally in statistical inference if we impose, on "aproximately" correct parametric models, regularity conditions similar to those typically used for the parametric inference. It is shown that a functional which is regular at a rich family of smooth parametric models containing F and for which the asymptotic limits satisfy a natural continuity property over shrinking Cramér von Mises neighbourhoods of F, is Fréchet differentiable at F for the Cramér von Mises norm.

Keywords: Fréchet differentiability, local asymptotic theory, uniform expansions, robustness.

1991 Mathematics Subject Classification: Primary 62G35; Secondary 62F10.

1. Introduction

Von Mises (1947) was the first to take a systematic approach to the study of statistics via writing estimators as functions of the empirical distribution function, say F_n. The weak notion of the Gâteaux derivative of that paper was replaced by the stronger Fréchet derivative with respect to the

Kolmogorov or sup norm in Kallianpur and Rao (1955). Yet the Fréchet derivative has not been accepted unanimously in the statistical literature as a very useful approach to the study of asymptotic distributions of statistics. Since most common maximum likelihood estimators of parameters from parametric models have unbounded score functions in the observation space variable, these estimators are not Fréchet differentiable in the above sense (cf. Clarke (1983)). Moreover, some of the frequently used estimators constructed from quantiles, such as the median for example, are not Fréchet differentiable with respect to the sup norm, cf. Fernholz (1983).

Generally, a functional which is Fréchet differentiable with respect to a norm yields an estimator which is resistant to outliers modelled by distributions which are in neighbourhoods given by the norm. The central measures of these "small" neighbourhoods would usually form a parametric model along which the functional is Fisher consistent and for which the functional in question is constructed. Norms that induce functionals that are generally recognized as robust are usually defined for cumulative distribution functions. The most commonly used are the supremum norm and the Cramér von Mises norm (Boos and Serfling (1980), Hampel et al. (1986), Huber (1981), Millar(1981), Pollard(1980)). In fact, to a statistician who realises the approximate character of his model, he has a very clear intuitive meaning of closeness of distribution functions.

An additional advantage of a Fréchet differentiable functional with respect to one of the norms is that one has a straightforward path to the description of the asymptotic distribution of the estimator (cf. Huber (1981, p. 39), Boos and Serfling (1980, Lemma 1.1), see also Dvoretzky, Kiefer and Wolfowitz (1956)) over a large nonparametric families of distributions. The approach is to express $T(F_n)$ via a linear expansion. It is indicated in Bednarski, Clarke and Kolkiewicz (1991) that the expansion is valid under violations of the model (mild in terms of the sup distance), which may represent gross errors or asymmetry of the underlying distribution for example. Hence, reliable assessments of the variance of such estimators can easily be constructed. To be more explicit, suppose that the functional T is Fréchet differentiable at a distribution F defined on R^k, with the norm $\|\cdot\|$, with a functional derivative T'_F that is expressible as an integral in the form $T'_F(G-F) = \int \psi_F(x)\, d(G-F)(x)$. Then

$$(1.1) \qquad T(G) - T(F) = \int \psi_F(x) dG(x) + o(\|G-F\|),$$

as $\|G-F\| \to 0$. Hence it is easily seen, for G_n satisfying $\|G_n - F\| \le c/\sqrt{n}$

and F_n the empirical distribution function based on the sample drawn from G_n, that $\sqrt{n}[T(F_n) - T(G_n)]$ is asymptotically normal $N(0, \sigma^2)$, where $\sigma^2 = E_F\psi^2$ and that $\int \psi^2 dF_n$ converges to σ^2 (uniformly for distributions G_n in $\sqrt{n}$ shrinking neighbourhoods). From the practical point of view the last statement is very important because it enables the construction of more reliable tests and confidence intervals concerning parameters of the model. Note that since the derivative T'_F has to be continuous with respect to the norm, the function ψ_F has to be bounded. These facts are exploited in describing robust estimating functionals in some quite sophisticated models in recent examples which include work on the Cox model in Bednarski (1993), finite mixtures of normal distributions in Clarke and Heathcote (1994), and variance components estimation in Bednarski and Zontek (1995).

Bednarski (1993) has improved the statement of conditions given in Bednarski, Clarke and Kolkiewicz (1991), which show an asymptotic equivalence between estimators satisfying a uniform expansion and M-estimators that are in fact Fréchet differentiable with respect to the sup norm. The current paper in a way contributes to this line of research. More specifically, we ask about the extent to which the uniform weak convergence of distributions of the functional under regular parametric models defines its strong smoothness properties. Briefly stated, we examine the von Mises functionals T which satisfy

$$\sqrt{n}[T(F_n) - T(G_n)]_{|G_n} \rightarrow_w M, \tag{1.2}$$

where $\rightarrow_w$ denotes weak convergence to a probability law M, for sequences G_n which are members of a rich family of regular parametric models centered at F (to be defined in Section 2) and satisfying $G_n \in \{G \in \mathcal{G} : \|G - F\| \leq c/\sqrt{n}\}$.

We shall also assume that under the same sequences G_n, there is a continuous linear functional L defined on the linear space spanned by the differences of distribution functions, such that

$$\sqrt{n}[T(F_n) - T(F)]_{|G_n} - L(\sqrt{n}(G_n - F)) \rightarrow_w M, \tag{1.3}$$

where $*$ denotes the convolution, while δ_x is the distribution having mass 1 at x.

There are several immediate questions which arise from such a set of assumptions. For instance, what should be the choice of norm; why do we insist on linearity and continuity of the functional L; and why should the limiting family in (1.3) be a shift family ? Uniform convergence, which has clear practical implications for inference, is also justified more generally by robustness considerations. Since the functional T is meant to be a good estimator at the model distribution F, it is reasonable to suppose that the resulting estimator is regular at F for a rich family of parametric models containing F. The above assumptions will be motivated in a detailed way in Chapter 2.

The main objective of the paper is achieved in Chapter 3. Using a Riesz representation theorem we show that the functional L is expressible as an integral with respect to a function ψ_F which is continuous and of bounded variation. This representation is conditional on the choice of the Cramér von Mises norm which reflects continuity with respect to small perturbations in terms of rounding errors as well as gross errors and small mis-specifications of parametric models. While perhaps a more general form of the Riesz representation theorem could yield a similar result for a linear functional L continuous with respect to the sup norm, the authors do not know of any useful statistical functionals which are Fréchet differentiable and which have discontinuous ψ_F. It is pertinant to note here that the median functional does not converge to a single limit law uniformly under $\sqrt{n}$ shrinking neighbourhoods of the Kolmogorov type as it is remarked in Lemma 3.1 of Bednarski, Clarke and Kolkiewicz (1991). The representation theorem then leads to a conclusion that the functional $T(F)$ is Fréchet differentiable and that the distribution M has to be normal. Due to our concern about robustness properties of the functionals we use "small" norms in our considerations. In his remarkable work, Dudley (1992,1994) uses "bigger" norms to show Fréchet differentiability of functionals which asymptotically do not enjoy the simple structure of limiting distributions for the norm we are concerned with; for instance, the variance of the asymptotic distribution may vary with the bias over shrinking neighbourhoods generated by the sup norm for such functionals.

2. DISCUSSION OF ASSUMPTIONS

The starting point of our considerations will be a short review of properties of estimators in the case of smooth parametric families. Suppose $\mathcal{P} = \{P_\theta\}_{\theta \in R^k}$ is a parametric model. Since efficiency properties of estimators are usually studied locally at the model, the typical smoothness assumptions have a local character and are phrased in terms of L_2 differentiability for the square roots of density functions. We say that $\mathcal{P}$ is L_2 differentiable at θ if there exists some function $\Lambda_\theta \in L_k^2(P_\theta)$ such that as $t \to 0$

$$\int \left(\sqrt{p_{\theta+t}} - 1 - \frac{1}{2}t'\Lambda_\theta\right)^2 dP_\theta \;=\; o(|t|^2)\,,$$

where $p_{\theta+t} = dP_{\theta+t}/dP_\theta$.

Estimators $\hat{\theta}_n$ which include good competitors to the maximum likelihood estimator in the smooth case are usually required to satisfy the following asymptotic uniform stability conditions (see for instance Hájek(1972), Le Cam(1986) and Millar(1983)):

$$\text{(2.1)} \qquad \sqrt{n}[\hat{\theta}_n - (\theta + \tau_n/\sqrt{n})]_{\,|P^{\otimes n}_{\theta+\tau_n/\sqrt{n}}} \to_w M\,,$$

where all τ_n belong to a bounded set and M is a probability distribution. These estimators are called regular and their asymptotic properties relative to asymptotically efficient estimators are explained by convolution theorems.

The above notions and properties are basic fundaments of a theory of asymptotic statistics which was largely developed by Le Cam (1969,1986). We also refer the reader to Rieder (1994) where more detailed information along with relevant citations can be found in connection with asymptotically robust testing and estimation.

We are going to consider functionals which in a sense inherit the regularity property for smooth parametric models containing the fixed distribution P_{θ_0}, denoted further either by F or P_0 (we put $\theta_0 = 0$). The family is defined for bounded functions in $L^2(P_0)$. Namely, if a bounded $g \in L^2(P_0)$, then the induced model has the form $dP_{t,g}/dP_0 = 1 + tg$, if for some $t_0 > 0$ $1 + tg \geq 0$ is a density function with respect to P_0 for all $|t| \leq t_0$. The functions g are called tangents (see for instance Rieder (1994)) and further on we shall refer to the family of models as tangent models. As one can see, these models are smooth. It is customary in parametric statistics to assume that the asymptotic distribution of a regular functional T would have a bias that is linear in the local parameter. In the context of the above defined family of smooth models this would mean that

$$\text{(2.2)} \qquad \sqrt{n}T(F_n)_{|P^{\otimes n}_{t/\sqrt{n},g}} \;\to_w\; M * \delta_{L(tg)},$$

where the functional L would be linear in argument. From the point of view of robust inference one would expect the bias term to be bounded under the shrinking Cramér von Mises (henceforth refered to as CvM) neighbourhoods of P_0. For a distribution function G, we define its CvM norm here to be

$$\sqrt{\int G^2(x)\, dF(x)}$$

where $F = P_0$ is the central measure.

As it was previously mentioned, in addition to common requirements on regularity of estimators in the asymptotic parametric setup we consider it justifiable that the property (2.2) is satisfied uniformly over all g_n's such that $(1+g_n)dP_0$ stays in the $1/\sqrt{n}$ CvM neighbourhoods of P_0. This makes much more sense from the inferential point of view when asymptotically approximating distributions are used. The uniform approximation requirement together with the boundedness of bias in a small CvM neighbourhood of P_0 imply the boundedness of the functional L for all measures (given by the tangents) that are in some small neighbourhood. The CvM distance between two tangent models $1+tg_1$, $1+tg_2$ is given by:

$$t\sqrt{\int \left(\int_{-\infty}^{x} (g_1 - g_2)\, dF\right)^2 dF(x)}$$

and it naturally defines a norm on the tangent space, which we shall also call CvM. The boundedness of L implies then the continuity of L with respect to the CvM norm on the linear space of tangents. This can be interpreted in such a way that if there is a negligible change in the underlying distribution in terms of the CvM norm compared to $1/\sqrt{n}$, then the resulting bias has a negligible effect, as always takes place along any basic parametric model P_θ for which the functional was constructed.

The continuity has also a justification stemming from robustness considerations. For instance, suppose that $F = N(0,1)$ is the standard normal law. Then we may find it very reasonable to require from our estimator $T(F_n)$, where F_n is the empirical distribution function, to have the same asymptotic limit under the following sequences of pairs of marginal distributions

$$(1-\epsilon/\sqrt{n})F + \epsilon/\sqrt{n}\delta_x,\ \ (1-\epsilon/\sqrt{n})F + \epsilon/\sqrt{n}\delta_{x+a_n}$$

and

$$(1-\epsilon/\sqrt{n})F + \epsilon/\sqrt{n}N(5, a_n),\ \ (1-\epsilon/\sqrt{n})F + \epsilon/\sqrt{n}N(5+a_n, a_n),$$

where $a_n \to 0$ as $n \to \infty$. All these distributions are indistinguishable under the Cramér von Mises norm goodness of fit test. Athough some of the differences are picked up by the Kolmogorov Smirnov test we now take a functional not too sensitive to local shifts of the distribution as illustrated in the above examples.

3. Fréchet differentiability of regular functionals

Let $\mathcal{G}$ be a space of all probability distributions on R^k. We denote by $\mathcal{H}_0$ the linear space spanned by $\{G-F : G \in \mathcal{G}\}$. Taking into account properties of usable parametric models we assume that either F has a density f with support given by an interval (possibly the whole real line) or F is a discrete distribution. In the first case, it is supposed that f is bounded away from 0 on any compact of the support's interior. For the other, which because of its simplicity will not be considered in detail, we take the CvM neighbourhoods of probabilities with the same support as F. The functional L naturally defines on part of $\mathcal{H}_0$ which is given by the tangent models and moreover it extends to $\mathcal{H}_0$ by the assumed continuity with respect to the norm. This can easily be seen since the space of tangents induces probabilities that are dense in $\mathcal{G}$ for the Cramér von Mises norm. In fact, the absolutely continuous distributions with compact supports form a dense set in the subspace. Any such distribution can be approximated in turn by densities induced by tangents. We further suppose $T(F) = 0$

Theorem 3.1. (Representation Theorem) *Suppose the functional L is continuous on the set $\mathcal{G} - F$ with respect to the CvM norm. Then there is a continuous function $\psi_F : R^k \to R$ with finite variation such that for every distribution $G \in \mathcal{G}$ we have*

$$L(G-F) = \int \psi_F \, d(G-F).$$

Proof. Define $\psi_F(s) = L(x_s)$, where $x_s(t) = I_{\mathcal{A}_s}(t) - F(t)$ with $\mathcal{A}_s = \{t : t \geq s\}$. Notice that the function $\psi_F(s)$ has to be bounded because of continuity of the functional. Moreover, since for $s_n \to s$, x_{s_n} converges to x_s in the CvM norm, we conclude continuity of ψ_F.

The function ψ_F is also of finite variation (we prove this in the one dimensional case first). Let $\pi_n = \{s_i\}_{i=0}^n$ be a strictly increasing sequence of real numbers. Define a sequence of functions

$$z_n(t) = \sum_{k=1}^{n} (x_{s_k}(t) - x_{s_{k-1}}(t))\epsilon_k,$$

where $\epsilon_k = sgn[\psi_F(s_k) - \psi_F(s_{k-1})], k = 1, ..., n$. Then

$$\sum_{k=1}^{n} |\psi_F(s_k) - \psi_F(s_{k-1})| = \sum_{k=1}^{n} \epsilon_k[\psi_F(s_k) - \psi_F(s_{k-1})]$$

$$= \sum_{k=1}^{n} \epsilon_k[L(x_{s_k}) - L(x_{s_{k-1}})] = Lz_n \leq \|L\| \cdot \|z_n\| \leq \|L\|.$$

Take now an element, say $G - F$ of $\mathcal{G} - F$, and define

$$G_n(t) = \sum_{k=0}^{n-1} G(s_i)[x_{s_i}(t) - x_{s_{i+1}}(t)].$$

The integral $\int \psi_F(t)d(G - F)(t)$ is well defined and it has to be a limit of $\int \psi_F(t)d(G_n - F)(t)$ since G_n converges to G at continuity points of G. Now $\int (G_n - G)^2 dF$ converges to zero as the partitions π_n become finer in the sense that the largest difference between consequitive s_i converges to zero and the smallest and the largest points tend to $-\infty$ and $+\infty$ respectively. Therefore by continuity of the functional L we obtain

$$L(G - F) = \int \psi_F(t)\, d(G - F)(t).$$

The bounded variation argument goes similarly in the multivariate case. For instance in R^2 case we take any finite set of disjoint bounded rectangles instead of the increasing sequences $\{s_i\}$. Instead of the differences $x_{s_i}(t) - x_{s_{i+1}}(t)$ we take the corresponding double differences.

The following theorem establishes the Fréchet differentiability of the functional T with respect to the CvM norm.

Theorem 3.2. *Let $\mathcal{G}_n$ denote a shrinking family of neighbourhoods around the distribution F satisfying conditions given in the begining of this section;*

$$\mathcal{G}_n = \{G : \|G - F\|_{CvM} \leq c/\sqrt{n}\}.$$

Suppose that for a probability measure M and for all sequences $G_n \in \mathcal{G}_n$ such that $dG_n/dP_0 = 1 + g_n$ for a bounded g_n, we have

$$\sqrt{n}(T(F_n) - T(G_n))_{|G_n^{\otimes n}} \to_w M$$

and

$$\sqrt{n}T(F_n)_{|G_n^{\otimes n}} - \sqrt{n}L(G_n - F) \to_w M,$$

where the linear functional L *extended to* $\mathcal{H}_0$ *is continuous for the CvM norm. Then*

$$T(G) - T(F) = \int \psi_F(t)\, d(G-F)(t) + o(\|G-F\|_{CvM}),$$

where $\psi_F(t)$ *is given in the conclusion of the representation theorem.*

Proof. Take any sequence $\{G_n'\}$ such that $\sqrt{n}\|G_n' - F\|_{CvM} \le c$. Take a bounded interval $[a,b]$ so that $F(b) - F(a) > 1 - \epsilon/2$ for a given ϵ. We can find a differentiable distribution function on $[a,b]$ with bounded derivative, with $G_n(a) = 0$, $G_n(b) = 1$ and such that $\int_a^b (G_n' - G_n)^2 dF < \epsilon/2$. Therefore $\|G_n' - G_n\|_{CvM} < \epsilon$.

By continuity of T with respect to CvM norm for each G_n' we can now find G_n so that

$$\|G_n' - G_n\|_{CvM} < c/\sqrt{n},$$

$$\sqrt{n}[T(G_n') - T(G_n)] \to 0$$

and

$$L[\sqrt{n}(G_n' - F)] - L[\sqrt{n}(G_n - F)] \to 0$$

For the sequence G_n we have $dG_n/dF = 1 + g_n$, where g_n is bounded for each n. By the assumptions

$$\{\sqrt{n}[T(F_n) - T(F)] - \sqrt{n}[T(G_n) - T(F)]\}_{|G_n^{\otimes n}} \to_w M$$

$$\{\sqrt{n}[T(F_n) - T(F)] - \sqrt{n}L(G_n - F)\}_{|G_n^{\otimes n}} \to_w M$$

Since the difference of the above random variables is a number depending on n we obtain

$$\{\sqrt{n}[T(G_n) - T(F)] - \sqrt{n}L(G_n - F)\} \to 0$$

and finally

$$\sqrt{n}[T(G_n') - T(F)] - \sqrt{n}\int \psi_F(t) d(G_n' - F)(t) \to 0$$

as $n \to \infty$. This terminates the proof.

Theorem 3.2 should be interpreted as follows. If a functional (estimator) satisfies **uniformly** the usual regularity conditions required for estimation in the parametric case, then necessarily the functional is Fréchet differentiable. The theorem is not meant to be a tool in the search for differentiable functionals; in general there is no difficulty in construction of Fréchet differentiable functionals. It partly expains to what extent this particular differentiability notion is natural in statistical estimation.

REFERENCES

[1] T. Bednarski, *Fréchet differentiability and robust estimation*, in: *Asymptotic Statistics*, Proceedings of the Fifth Prague Symposium, eds. P. Mandl and M. Husková, Springer Lecture Notes (1993), 49–58.

[2] T. Bednarski, *Robust estimation in Cox's regression model* Scand. J. Statist. **20** (1993), 213–225.

[3] T. Bednarski and S. Zontek, *Robust estimation of parameters in a mixed unbalanced model*, to appear in the Annals of Statistics (1995).

[4] T. Bednarski, B.R. Clarke, W. and Kolkiewicz, *Statistical expansions and locally uniform Fréchet differentiability*, J. Australian Math. Soc., Ser. A **50** (1991), 88–97.

[5] D.D Boos and R.J. Serfling, *A note on differentials and the CLT and LIL for statistical functions with application to M-estimates*, Ann. Statist. **8** (1980), 618–624.

[6] B.R Clarke, *Uniqueness and Fréchet differentiability of functional solutions to maximum likelhood type equations*, Ann. Statist. **11** (1983), 1196–1205.

[7] B.R. Clarke and C.R. Heathcote, *Robust estimation of k-component univariate normal mixtures*, Ann. Inst. Statist. Math. **46** (1994), 83–93.

[8] R.M. Dudley, *Fréchet differentiability, p-variation and uniform Donsker classes*, Ann. Probab. **20** (1992), 1968–1982.

[9] R.M. Dudley, *The order of the remainder in derivatives of composition and inverse operators for p-variation norms*, Ann. Statist. **22** (1994), 1–20.

[10] A. Dvoretzky, J. Kiefer and J. Wolfowitz, *Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator*, Ann. Math. Statist. **27** (1956), 642–669.

[11] L.T. Fernholz, *Von Mises Calculus for Statistical Functionals*, Lecture Notes in Statistics, Springer-Verlag, New York, Vol. 19 (1983).

[12] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw and W.A. Stahel, *Robust Statistics - The Approach Based on Influence Functions*, Wiley, New York 1986.

[13] P.J. Huber, *Robust Statistics*, Wiley, New York 1981.

[14] G. Kallianpur and C.R. Rao, *On Fisher's lowerbound to asymptotic variance of a consistent estimate*, Sankhya **15** (1955), 331–342.

[15] L. Le Cam, *Théorie Asymptotique de la Décision Statistique*, Les Presses de l'Université de Montréal 1969.

[16] L. Le Cam, *Asymptotic Methods in Statistical Decision Theory*, Springer-Verlag, New York 1986.

[17] P.W. Millar, *Robust estimation via minimum distance methods*, Z. Wahrsch. verw. Gebiete **55** (1981), 73–89.

[18] D. Pollard, *The minimum distance method of testing*, Metrika **27** (1980), 43–70.

[19] H. Rieder, *Robust Asymptotic Statistics*, Springer-Verlag, New York 1994.

[20] R. Von Mises, *On the asymptotic distribution of differentiable statistical functions*, Ann. Math. Statist. **18** (1947), 309–348.

Received 27 February 1998