A robust predictive density based on the saddlepoint approximation for M-estimators

BY ELVEZIO RONCHETTI

Department of Econometrics, University of Geneva, Blv. Carl-Vogt 102, CH-1211 Geneva, Switzerland

e-mail: Elvezio.Ronchetti@metri.unige.ch

AND PAOLO VIDONI

Department of Statistics, University of Udine, via Treppo 18, I-33100 Udine, Italy

e-mail: vidoni@dss.uniud.it

SUMMARY

The robust predictive density proposed in this paper is the analogous of the parametric bootstrap predictive function introduced by Harris (1989), when the maximum likelihood estimator is substituted by a suitable M-estimator. An useful closed form approximation to this new proposal is derived, which consists of approximating the sampling distribution of the M-estimator by the associated saddlepoint approximation and then using the Laplace method to evaluate the integral. By means of a simple comparative study, we show that this fully robust predictive procedure improves the usual robust predictive distributions, obtained by simply substituting the unknown parameter by a suitable robust estimator.

Some key words: Approximate p^* predictive density; Data contamination; Huber estimator; Laplace method; Parametric bootstrap predictive density.

1. INTRODUCTION

The purpose of this paper is to define a robust predictive density for a future or as yet unobserved random variable in such a way that the associated prediction statements do not suffer seriously from data contamination. In particular, we consider

the simple situation where the data $x_1,...,x_n$ are independent realizations of a random variable *X*, with probability density function $f(x;\omega), \omega \in \Omega \subseteq \mathbf{R}^d, d \geq 1$. The future random variable *Z* is independent of *X* and has the same distribution as *X*, which depends on the unknown parameter ω. Any estimator of the true probability density function $f(z;\omega)$ of *Z*, based on the observable sample, is called a predictive density function.

The main perspective here is to study robust predictive densities, that is predictive densities which present a good behavior in the presence of small deviations from the original model. More precisely, we assume that the data, on which the prediction are based, may be contaminated, while the future observation is supposed to come from the true model.

The robust predictive density proposed in this paper is the analogous of the parametric bootstrap predictive density introduced by Harris (1989), when the maximum likelihood estimator is substituted by a suitable M-estimator. Unlike the predictive density given by Basu & Harris (1994), this new proposal is fully robust since it is based on the sampling distribution of the robust estimator. Indeed, an useful closed form approximation is derived, which consists of approximating the sampling distribution of the M-estimator by the associated saddlepoint approximation and then using the Laplace method to evaluate the integral. Preliminary results, related to a simple example involving the normal model, show that this fully robust predictive distribution constitutes an improvement over predictive distributions obtained by simply substituting the unknown parameter ω with a suitable robust estimator.

2. A REWIEV ON PREDICTIVE DENSITIES

The simplest and most intuitive approach to prediction consists of using the estimative probability density function $f(z;\hat{\omega})$ obtained by substituting $\hat{\omega}$ for ω; $\hat{\omega}$ is an appropriate estimator of ω based on the observable sample, usually the maximum likelihood estimator. In spite of its intuitiveness, $f(z;\hat{\omega})$ may not be entirely adequate for prediction since, especially when the dimension of ω is large in comparison with *n*, it may provide inaccurate prediction results. A number of recent papers aim to improve

the estimative density; see, for example, Harris (1989), Vidoni (1995), Komaki (1996) and Corcuera & Giummolè (1998).

In particular, Harris (1989) proposed the parametric bootstrap predictive density, given by

$$f_B(z;\omega) = \int f(z;t)p(t;\omega)dt, \tag{2.1}$$

computed at $\omega = \hat{\omega}$, where $p(\cdot;\omega)$ is the probability density function of the maximum likelihood estimator. Density (2.1) has some desirable properties; in particular, within natural exponential models, it is asymptotically superior to the estimative density $f(z;\hat{\omega})$ in terms of average Kullback-Leibler divergence. Unfortunately, it is usually not in a reasonable closed form and it needs to be computed numerically even for simple models.

Vidoni (1995) pointed out that, although Harris's proposal is often unsuitable for exact calculations, it allows fairly simple approximations through straightforward asymptotic arguments. In particular, when $\hat{\omega}$ is a sufficient statistic, such as within natural exponential models, it is possible to derive an high-order, closed form approximation to (2.1), which consists of approximating $p(\cdot;\omega)$ by Barndorff-Nielsen's (1983) p^*-formula and then using a Laplace approximation with $O(n^{-1})$ correction terms for integrating out the parameter. This procedure may be extended to the more general case where the maximum likelihood estimator is not itself a sufficient statistic.

These predictive densities are based on the maximum likelihood estimator, which is usually a non-robust estimator. Thus, they may give poor prediction results, whenever the observed data present small deviations from the assumed statistical model. As a first solution to this problem, Basu & Harris (1994) proposed a robust version for the parametric bootstrap predictive density, namely $f_B(z;\hat{\omega}_H)$. This is obtained by computing (2.1) at the observed value of $\hat{\omega}_H$, the minimum Hellinger distance estimator of ω. The sampling distribution of $\hat{\omega}_H$ cannot be computed explicitly and, for this reason, the weighting function which is considered is still the

probability density function of the maximum likelihood estimator. Since $\hat{\omega}_H$ and $\hat{\omega}$ are first order equivalent, this robust version of the parametic bootstrap predictive density maintains the superiority over the estimative one, in terms of the average Kullback-Liebler divergence within natural exponential models. Moreover, it may give better robustness results than $f(z;\hat{\omega})$ and $f_B(z;\hat{\omega})$, which involve the maximum likelihood estimator. However, since $f_B(z;\hat{\omega}_H)$ is (2.1) computed at the observed value of $\hat{\omega}_H$, it maintains the same computational drawbacks of the parametric bootstrap predictive density. Indeed, $f_B(z;\hat{\omega}_H)$ does not take into account the sampling distribution of $\hat{\omega}_H$.

3. A NEW ROBUST PREDICTIVE DENSITY BASED ON M-ESTIMATORS

In this section, we introduce a fully robust predictive distribution. More precisely, we define a robust parametric bootstrap predictive density based on a suitable M-estimator and its probability distribution. Unlike the predictive density given by Basu & Harris (1994), which involves as weighing function the density of the maximum likelihood estimator, this new proposal is based on the sampling distribution of the robust estimator. Although this distribution is usually not explicitly known, the associated saddlepoint approximation enables the determination of an approximate robust predictive density, which is accurate to third-order under ordinary repeated sampling.

An M-estimator for ω is defined as the solution $\hat{\omega}_M$ of the system of equations

$$\sum_{i=1}^{n} \psi_r(x_i;\omega) = 0, \quad r = 1,\ldots,d,$$

where $\psi_r(\cdot)$ is the rth component of a suitable $\mathbf{R}^d$-valued function $\psi(\cdot)$. By setting $\psi_r(x_i;\omega) = \partial \log p(x_i;\omega)/\partial\omega^r$, $\hat{\omega}_M$ becomes the maximum likelihood estimator.

The robust parametric bootstrap predictive density based on a given M-estimator $\hat{\omega}_M$ is

$$f_{BM}(z;\omega) = \int f(z;t)p_M(t;\omega)dt\,, \tag{3.1}$$

computed at $\omega = \hat{\omega}_M$, where $p_M(\cdot;\omega)$ is the probability density function of $\hat{\omega}_M$. Density (3.1) is usually unsuitable for exact calculations, however it allows a relatively simple approximation through the asymptotic arguments outlined in the following.

It is well known that the probability density function of an M-estimator $\hat{\omega}_M$ may be approximated by means of the general saddlepoint approximation technique derived by Field (1982), Field & Ronchetti (1990) and Almudevar, Field & Robinson (1997). Under suitable regularity assumptions, it corresponds to

$$\tilde{p}_M(t;\omega) = d(\omega)\exp\{-n\log c(t;\omega)\}\left|\det[w_{rs}(t;\omega)]\right|\left|\det[v_{rs}(t;\omega)]\right|^{-1/2}, \tag{3.2}$$

where

$$c(t;\omega) = \left\{\int \exp\{\alpha_r(t;\omega)\psi_r(x;t)\}f(x;\omega)dx\right\}^{-1},$$

$\alpha(t;\omega) = (\alpha_1(t;\omega),\ldots,\alpha_d(t;\omega))^T$ is the solution of

$$\int \psi_s(x;t)\exp\{\alpha_r(t;\omega)\psi_r(x;t)\}f(x;\omega)dx = 0\,, \qquad s = 1,\ldots,d\,,$$

and the summation notation is used. Moreover,

$$w_{rs}(t;\omega) = E_h\left\{\partial\psi_r(x;t)/\partial t^s\right\}, \qquad v_{rs}(t;\omega) = E_h\left\{\psi_r(x;t)\psi_s(x;t)\right\},$$

where the expectations are with respect to the conjugate density

$$h(x;t,\omega) = c(t;\omega)\exp\{\alpha_r(t;\omega)\psi_r(x;t)\}f(x;\omega).$$

The term $d(\omega)$ is the normalizing constant of expression (3.2); with this choice, the saddlepoint approximation usually presents a relative error of order $O(n^{-3/2})$.

Thus, using (3.2) as a weighting function in (3.1), the robust saddlepoint predictive density can be defined as

$$f_S(z;\omega) = \int f(z;t)\tilde{p}_M(t;\omega)dt\,, \tag{3.3}$$

computed at $\omega = \hat{\omega}_M$. Function (3.3) is an approximation to the robust parametric bootstrap predictive density (3.1), with a relative error of order $O(n^{-3/2})$. This predictive density can be further approximated, retaining the same order of error, by Laplace method, giving a relatively simple closed form expression. The technique is similar to that used by Vidoni (1995) in order to define the approximate *p** predictive density. The approximation is (see the Appendix for the derivation)

$$\tilde{f}_S(z;\omega) = f(z;\tilde{t})\{1 + \tfrac{1}{2}M(\tilde{t};z,\omega)\}, \tag{3.4}$$

computed at $\omega = \hat{\omega}_M$, with $\tilde{t} = \tilde{t}(\omega)$ and

$$M(\tilde{t};z,\omega) = \left[\ell_{ij}(\tilde{t};z) + \ell_i(\tilde{t};z)\ell_j(\tilde{t};z) + 2\ell_i(\tilde{t};z)D_j(\tilde{t};\omega)\{D(\tilde{t};\omega)\}^{-1}\right] r^{ij}(\tilde{t};\omega)n^{-1}$$
$$- \ell_i(\tilde{t};z)r_{jhk}(\tilde{t};\omega)r^{jh}(\tilde{t};\omega)r^{ki}(\tilde{t};\omega)n^{-1}.$$

Here, $\ell(t;z) = \log f(z;t)$ and $\ell_{I_m}(\tilde{t};z)$, $I_m = (i_1,\ldots,i_m)$, $m \in \mathbf{N}^+$, is the *m*th partial derivative with respect to the corresponding components of *t*, evaluated at $t = \tilde{t}$. The minimum point $\tilde{t} = \tilde{t}(\omega)$ and the coefficients $D(\tilde{t};\omega)$, $D_j(\tilde{t};\omega)$, $r^{ij}(\tilde{t};\omega)$ and $r_{jhk}(\tilde{t};\omega)$ are defined in the Appendix. The term $M(\tilde{t};z,\hat{\omega}_M)$ is of order $O(n^{-1})$.

The robust predictive density (3.4) is a modification of the estimative density based on $\tilde{t}(\hat{\omega}_M)$ and presents the same structure as the approximate *p** predictive density (2.3) in Vidoni (1995). More precisely, with regard to the modifying $O(n^{-1})$ term, we can see that the coefficients differs from those of the approximate *p** density,

while the variable terms, based on the first two partial derivatives of $\log f(z;\omega)$ with respect to ω, are the same. Moreover, within natural exponential models, if the M-estimator equals the maximum likelihood estimator, then $\tilde{f}_S(z;\hat{\omega}_M)$ coincides, as one would expect, with the approximate *p** predictive density.

4. A NUMERICAL RESULT

The robust parametric bootstrap predictive density (3.1) and the associated higher-order approximation $\tilde{f}_S(z;\hat{\omega}_M)$ are based on the M-estimator $\hat{\omega}_M$ and on its probability distribution. Thus, the robust correction of the estimative density introduced in this context concerns both the estimate of the unknown parameter ω and the functional form of the predictive density. Therefore, this new robust predictive distribution is supposed to give good results under data contamination. In particular, we expect better results than those given by the parametric bootstrap predictive density (2.1) and the associated approximate *p** predictive distribution, computed at $\omega = \hat{\omega}_M$. We shall verify these statements with regard to a simple example.

Let $X_1,\ldots,X_n,Z$ be mutually independent according to a location model with common probability density function $f(x;\mu) = f_0(x-\mu)$, $x \in \mathbf{R}$, $\mu \in \mathbf{R}$, and $\omega = \mu$; in this context, ψ-functions of the type $\psi(x;\omega) = \psi_0(x-\omega)$ are usually considered. In particular, $\psi(x;\omega) = (x-\omega)\min(1, b/|x-\omega|)$, with $b > 0$ a given constant, defines the Huber (1964) estimator of location $\hat{\omega}_M$, obtained as the solution of the equation

$$\sum_{i=1}^{n} (x_i - \hat{\omega}_M)\min(1, b/|x_i - \hat{\omega}_M|) = 0 \ .$$

This estimator has robustness properties (see Hampel, Ronchetti, Rousseeuw & Stahel, 1986, § 2.4), and, for $b = +\infty$, coincides with the mean. In order to define the robust approximate predictive density (3.4), the first step is to abtain the solution $\alpha(t;\omega)$ of

$$\int (x-t)\min\left(1,b/|x-t|\right)\exp\left\{\alpha(t;\omega)(x-t)\min\left(1,b/|x-t|\right)\right\} f_0(x-\omega)dx = 0$$

and then to calculate the expectations $w(t;\omega) = E_h\{\partial\psi(x;t)/\partial t\}$ and $v(t;\omega) = E_h\left\{\psi(x;t)^2\right\}$, with respect to the conjugate density $h(x;t,\omega)$. Once these functions have been computed, usually by means of numerical methods, it is not difficult to obtain the corresponding derivatives up to order two and to calculate the minimum point $\tilde{t} = \tilde{t}(\omega)$. By considering the first two derivatives of $\ell(t;z) = \log f_0(z-t)$ with respect to *t*, it is possible to obtain the modifying term $\frac{1}{2} M\left(\tilde{t};z,\hat{\omega}_M\right)$ and to define the robust approximate predictive density $\tilde{f}_S\left(z;\hat{\omega}_M\right)$.

As a simple application, let $X_1,\ldots,X_n,Z$ be mutually independent and normally distributed with unknown mean μ and known variance $\sigma^2 = 1$ and let us consider the Huber estimator $\hat{\omega}_M$. In this case, by using the procedure previously outlined, it is possible to determine the robust predictive density (3.4) associated to a given observed sample. More precisely, it is

$$\tilde{f}_S(z;\omega) = f\left(z;\tilde{t}\right)\left[1+\tfrac{1}{2}\left\{(z-\tilde{t})^2 + (z-\tilde{t})(2\tilde{d}-\tilde{r}_3/\tilde{r}_2) - 1\right\}(\tilde{r}_2 n)^{-1}\right], \qquad (4.1)$$

computed at $\omega = \hat{\omega}_M$, with $\tilde{t} = \omega$; $\tilde{r}_2$, $\tilde{r}_3$ and $\tilde{d}$ are, respectively,

$$r_2(t;\omega) = -\partial\{\alpha(t;\omega)w(t;\omega)\}/\partial t\,,$$
$$r_3(t;\omega) = -\partial^2\{\alpha(t;\omega)w(t;\omega)\}/\partial t^2\,,$$
$$d(t;\omega) = \{\partial w(t;\omega)/\partial t\}\{w(t;\omega)\}^{-1} - \tfrac{1}{2}\{\partial v(t;\omega)/\partial t\}\{v(t;\omega)\}^{-1},$$

evaluated at $t = \tilde{t}$. Whenever $b = +\infty$, $\hat{\omega}_M$ coincides with the maximum likelihood estimator $\hat{\omega} = n^{-1}\sum X_i$. Thus, in this particular instance, the equation

$$\int (x-t)\exp\left\{\alpha(t;\omega)(x-t)\right\}(2\pi)^{-1/2}\exp\left\{-\tfrac{1}{2}(x-\omega)^2\right\}dx = 0$$

has solution $\alpha(t;\omega) = (t - \omega)$ and the conjugate density $h(x;t,\omega)$ corresponds to a normal distribution with mean *t*, unit variance and normalizing constant $c(t;\omega) = \exp\left\{\frac{1}{2}(\omega - t)^2\right\}$. Moreover, since in this case $\tilde{t} = \omega$ and the determinants of $v(t;\omega)$ and $w(t;\omega)$ correspond to unity, the robust predictive density (4.1) becomes

$$f(z;\omega)\left[1 + \tfrac{1}{2}\left\{(z - \omega)^2 - 1\right\}n^{-1}\right], \tag{4.2}$$

evaluated at $\omega = \hat{\omega}$, which equals, as one would expect, the approximate *p** predictive density. Since, within natural exponential models, the approximate *p** predictive density is an approximation to the parametric bootstrap predictive density, function (4.2), computed at $\omega = \hat{\omega}_H$, is an approximation to the robust predictive density proposed by Basu & Harris (1994), whenever an underlying normal distribution is considered.

Let us now discuss the results of a small simulation study for this case. We assume that the true model is the standard normal distribution, abbreviated as $N(0,1)$, and that the observable data, on which the prediction are based, may be contaminated; the future observation is supposed to come from the true model. We consider pseudo random samples of size *n* from the $0.9\,N(0,1) + 0.1\,N(\mu,\sigma^2)$ distribution, with $\mu = 0, 3$ and $\sigma^2 = 1, 10$; $(\mu,\sigma^2) = (0,1)$ determines the uncontaminated case. All the results are based on 10,000 replications and the sample sizes are $n = 10, 20$. The variance of the future observation is supposed to be known. For each combination, we compare the estimative and the approximate *p** predictive densities, computed both at the observed value of the maximum likelihood estimator $\hat{\omega}$ and at the observed value of the Huber estimator $\hat{\omega}_M$, with the new robust predictive density (4.1) evaluated at $\omega = \hat{\omega}_M$. As a measure of fit, we adopt the average Kullback-Liebler divergence (see, for example, Harris, 1989 and Basu & Harris, 1994) and we determine the ratio of the average Kullback-Liebler divergence of the other four methods to that of the estimative distribution, based on the maximum likelihood estimator. The average Kullback-Liebler divergences involving the estimative and the approximate *p** predictive densities, evaluated at $\omega = \hat{\omega}$, can be computed exactly. For the other predictive distributions, the

divergence is estimated by simulation. The standard error for the corresponding ratios is given in parentheses.

The results are presented in Table 1. A part from the uncontaminated case, we observe that the robust predictive distributions perform substantially better than those based on the maximum likelihood estimator. With regard to the robust methods, it appears that the approximate *p** predictive density (4.2) and the fully robust predictive density (4.1), computed at $\omega = \hat{\omega}_M$, improve the estimative one based on the same estimator. Note that, in accordance with Basu & Harris (1994), the improvement is greater when the contamination is location-based rather than scale-based. A direct comparison between the fully robust and the approximate *p** predictive densities, computed at $\omega = \hat{\omega}_M$, shows that the first one performs slightly better in almost all the contamination cases. The superiority is more pronounced when the contamination concerns the mean of the true model, while it almost disappears when the contamination concerns the variance.

APPENDIX

Derivation of formula (3.4)

In order to obtain the approximate robust predictive density (3.4), we consider the following procedure. First, it is convenient to rewrite expression (3.3) as a ratio of integrals in the standard form (see Tierney, Kass & Kadane, 1989) such as

$$f_S(z;\omega) = \frac{\int N(t;z,\omega)\exp\{-r_n(t;\omega)\}dt}{\int D(t;\omega)\exp\{-r_n(t;\omega)\}dt},$$

where $r_n(t;\omega) = n\,r(t;\omega) = n\log c(t;\omega)$, $N(t;z,\omega) = f(z;t)D(t;\omega)$ and $D(t;\omega) = \left|\det\left[w_{ij}(t;\omega)\right]\right|\left|\det\left[v_{ij}(t;\omega)\right]\right|^{-1/2}$. If $r(t;\omega)$ is smooth and presents a unique minimum in the interior of the domain at $t = \tilde{t}$ then, by applying formula (A.1) in Vidoni (1995) to the numerator and to the denominator, we obtain

$$f_S(z;\omega) = f(z;\tilde{t})\frac{1 + A_n + \{N_i(\tilde{t};z,\omega)B_n^i + \frac{1}{2}N_{ij}(\tilde{t};z,\omega)C_n^{ij}\}\{N(\tilde{t};z,\omega)\}^{-1}\{1+O(n^{-2})\}}{1 + A_n + \{D_i(\tilde{t};\omega)B_n^i + \frac{1}{2}D_{ij}(\tilde{t};\omega)C_n^{ij}\}\{D(\tilde{t};\omega)\}^{-1}\{1+O(n^{-2})\}}.$$

$N_{I_m}(\tilde{t};z,\omega)$ and $D_{I_m}(\tilde{t};\omega)$, $I_m = (i_1,\ldots,i_m)$, $m \in \mathbf{N}^+$, are the mth partial derivatives of $N(t;z,\omega)$ and $D(t;\omega)$ with respect to the components of t with indices in I_m, computed at $t = \tilde{t}$. The $O(n^{-1})$ correction terms A_n, B_n^i and C_n^{ij} are given by

$$A_n = \tfrac{1}{24}\left\{3\tilde{r}_{ijh}\tilde{r}_{klm} + 2\tilde{r}_{ihl}\tilde{r}_{jkm}\right\}\tilde{r}^{ij}\tilde{r}^{hk}\tilde{r}^{lm}n^{-1} - \tfrac{1}{8}\tilde{r}_{ijhk}\tilde{r}^{ij}\tilde{r}^{hk}n^{-1},$$

$$B_n^i = -\tfrac{1}{2}\tilde{r}_{jhk}\tilde{r}^{jh}\tilde{r}^{ki}n^{-1}, \qquad C_n^{ij} = \tilde{r}^{ij}n^{-1},$$

where $\tilde{r}_{I_m} = r_{I_m}(\tilde{t};\omega)$, with $r_{I_m}(t;\omega)$, $I_m = (i_1,\ldots,i_m)$, $m \in \mathbf{N}^+$, the mth partial derivative of $r(t;\omega)$, with respect to the corresponding components of t and $\tilde{r}^{ij}$ the (i,j) element of the inverse of the matrix $\left[r_{ij}(\tilde{t};\omega)\right]$.

Secondly, a further expansion for the denominator allows us to cancel A_n and to simplify some terms involving the partial derivatives of $N(t;z,\omega)$ and $D(t;\omega)$. Thus, with the quantities B_n^i and C_n^{ij} given explicitly, the approximation (3.4) is obtained.

Since, from the definition of $\alpha(t;\omega)$ it follows that $\alpha_{r,i}(t;\omega)\int \psi_r(x;t)h(x;t,\omega)dx = 0$, $i = 1,\ldots,d$, the partial derivatives of $r(t;\omega)$ may be expressed explicitly giving,

$$r_i(t;\omega) = -\alpha_r(t;\omega)w_{ir}(t;\omega),$$

$$r_{ij}(t;\omega) = -\left\{\alpha_{r,j}(t;\omega)w_{ir}(t;\omega) + \alpha_r(t;\omega)w_{ir,j}(t;\omega)\right\},$$

$$r_{ijh}(t;\omega) = -\left\{\alpha_{r,jh}(t;\omega)w_{ir}(t;\omega) + \alpha_{r,j}(t;\omega)w_{ir,h}(t;\omega)[2] + \alpha_r(t;\omega)w_{ir,jh}(t;\omega)\right\},$$

where $\alpha_{r,I_m}(t;\omega)$ and $w_{ir,I_m}(t;\omega)$, $I_m = (i_1,\ldots,i_m)$, $m \in \mathbf{N}^+$, are the mth partial derivatives of $\alpha_r(t;\omega)$ and $w_{ir}(t;\omega)$ with respect to the corresponding components of t

and $[2]$ indicates a sum of two terms obtained by permutation of the indices *j* and *h*. The minimum $\tilde{t} = \tilde{t}(\omega)$ is such that $r_i(\tilde{t};\omega) = 0$, $i = 1,\ldots,d$, and that $\left[r_{ij}(\tilde{t};\omega)\right]$ is a positive definite matrix. Furthermore, using the differentiation rule for determinants, it follows that

$$D_j(t;\omega)\left\{D(t;\omega)\right\}^{-1} = w^{rs}(t;\omega)w_{rs,j}(t;\omega) - \tfrac{1}{2}v^{rs}(t;\omega)v_{rs,j}(t;\omega),$$

where $v_{rs,j}(t;\omega)$, is the partial derivative of $v_{rs}(t;\omega)$ with respect to the *j*th component of t and $w^{rs}(t;\omega)$ and $v^{rs}(t;\omega)$ indicate the (r,s) elements of the inverse of the matrices $\left[w_{rs}(t;\omega)\right]$ and $\left[v_{rs}(t;\omega)\right]$, respectively.

The Laplace expansion used in this context, assures that (3.4) is an approximation, with relative error of order $O(n^{-2})$, to $f_S(z;\hat{\omega}_M)$ and, with relative error of order $O(n^{-3/2})$, to $f_{BM}(z;\hat{\omega}_M)$. As a consequence of the balance relations linking the joint moments of the derivatives $\ell_{I_m}(t;z)$, this approximation is a probability density function, since it is exactly normalized.

REFERENCES

ALMUDEVAR, A., FIELD, C.A. & ROBINSON, J. (1997). Saddlepoint density approximation for multivariate M-estimates. Manuscript.

BARNDORFF-NIELSEN, O.E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, **70**, 343-65.

BASU, A. & HARRIS, I.R. (1994). Robust predictive distributions for exponential families. *Biometrika*, **81**, 790-4.

CORCUERA, J. M. & GIUMMOLÈ, F. (1998). On the relationship between α-connections and the asymptotic properties of predictive distributions. *Bernoulli* (to appear).

FIELD, C.A. (1982). Small sample asymptotic expansions for multivariate M-estimates. *Ann. Statist.*, **10**, 672-82.

FIELD, C.A. & RONCHETTI, E.M. (1990). *Small Sample Asymptotics.* Lecture Notes-Monograph Series, vol.13. Institute of Mathematical Statistics, Hayward, California.

HAMPEL, F.R., RONCHETTI, E.M., ROUSSEEUW, P.J. & STAHEL, W.A. (1986). *Robust Statistics: the Approach Based on Influence Functions.* New-York: Wiley.

HARRIS, I.R. (1989). Predictive fit for natural exponential families. *Biometrika*, **76**, 675-84.

HUBER, P.J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.*, **35**, 73-101.

KOMAKI, F. (1996). On asymptotic properties of predictive distributions. *Biometrika* **83**, 299-313.

TIERNEY, L., KASS, R.E. & KADANE, J.B. (1989). Fully exponential Laplace approximations to expectations and variances of non-positive functions. *J. Am. Statist. Assoc.*, **84**, 710-6.

VIDONI, P. (1995). A simple predictive density based on the p^* formula. *Biometrika*, **82**, 855-63.