Robust conditional inference for location parameters

L. Ventura

2001.14

Dipartimento di Scienze Statistiche
Università degli Studi
Via C. Battisti 241-243
35121 Padova

Ottobre 2001

ROBUST CONDITIONAL INFERENCE FOR LOCATION PARAMETERS

Laura Ventura
Dipartimento di Scienze Statistiche
Università degli Studi di Padova
Via C. Battisti, 241 – 35121 Padova
e-mail: ventura@stat.unipd.it

Lavoro presentato al Meeting del Progetto Europeo "Geometry and Computer Algebra in Statistics and Probability", Barcellona (Spagna), 16-20 settembre 1996.

Robust conditional inference for location parameters

Laura VENTURA

Department of Statistical Sciences, University of Padova, Italy

Key words and phrases: Ancillary statistic, conditional inference, interval estimation, location model, robust inference.

ABSTRACT

The robustness properties of conditional normal-theory procedures of inference for a location parameter are considered; in particular, a robust conditional density is proposed to be used instead of the classical methods based on the assumption of normality. The new density is conditional on a robust ancillary and its properties are studied in comparison to the exact conditional density but also under slight violations of the normal model assumption.

RÉSUMÉ

Nous considérons les propriétés de robustesse des procédures de l'inférence basées sur l'hypothèse de normalitè pour un paramètre de location; en particulier, nous proposons une densité conditionnelle robuste à utiliser au lieu des méthodes classiques basées sur le modèle gaussien. La nouvelle densité est conditionnelle a une ancillaire robuste et ses propriétés sont étudies en comparaison de la densité conditionnelle exact, mais aussi pour des observations non gaussiens.

1. INTRODUCTION

Let $(y_1,...,y_n)$ denote a random sample of size n from a random variable Y with density belonging to the parametric family $\{p_\varepsilon(y-\mu),\ \varepsilon\in E,\ \Re\supseteq E\}$, where $\mu\in\Re$ is an unknown location parameter and $p_\varepsilon(\cdot)$ a density function on the real line indexed by the known scalar parameter ε. Suppose, in addition, that ε=0 corresponds to the standard normal density.

Since we are considering a location model, the conditionality principle can be applied for inference about μ (see, e.g., Cox and Hinkley 1974, § 2.3). The statistic $a=(a_1,...,a_n)=(y_1\text{-}\bar{y},...,y_n\text{-}\bar{y})$, where $\bar{y}=\Sigma_{i=1}^n y_i/n$ denotes the sample mean, is ancillary for μ; note that a is an ancillary statistic for any density $p_\varepsilon(\cdot)$ and that $\bar{y}$ can be substituted by any equivariant estimator of μ. Conditional inference about μ is based on the exact conditional distribution of the pivotal quantity $t=\bar{y}-\mu$ given a

$$f_\varepsilon(t|a) = c_\varepsilon(a) \prod_{i=1}^{n} p_\varepsilon(t + a_i) \quad , \tag{1.1}$$

where $c_\varepsilon(a)^{-1} = \int_{-\infty}^{+\infty} \prod p_\varepsilon(r + a_i)dr$.

To carry out inference about the location parameter μ on the basis of the observed sample, one convenient way to proceed is to take $\varepsilon=0$ so that t and a are independent and hence inference about μ uses the fact that (1.1) is the normal density with mean 0 and variance n^{-1}. Observe that, in this case, it would sufficient to write $f_0(t)$ but, in the following, we will continue to use the notation $f_0(t|a)$ to emphasize that inference based on the marginal distribution of t is, by default, conditional on a.

Our interest in focused on the behavior of the normal-theory conditional procedures of inference when the assumption that $p_\varepsilon(\cdot)$ is the standard normal density is not satisfied, i.e. when $\varepsilon\neq 0$. Although there exist now a great variety of approaches towards the robustness problem (see Hampel *et al.* 1986, § 1.1), the robustness approach to conditional inference has received little attention. An exception is Severini (1992) who proposes specific measures of conditional robustness of normal-theory inference for a location parameter; in particular, the author derives a measure of the robustness of the coverage probability of the conditional confidence intervals for μ based on the assumption that the observations are normally distributed, that is $\varepsilon=0$, when the data are distributed according to $p_\varepsilon(\cdot)$, $\varepsilon\neq 0$. This measure depends on the observed value of the ancillary statistic and hence the robustness or non-robustness of a conditional analysis is strictly related on the value of a.

The aim of this paper is to suggest a new density conditional on a robust transformation of the ancillary statistic to make inference about location parameters under the assumption of normality. The starting point is to consider a bounded transformation of a, which has generic component y_i-$\bar{y}$ (i=1,...,n), so that the new value of the ancillary is less sensitive to the presence of non-normal observations in the sample. Hence we define a robust conditional density based on the transformed ancillary and we will show that, under the hypothesis $\varepsilon=0$, inference about μ based on the new density is more conditionally robust that inference based on $f_0(t|a)$, when the true underlying model is assumed to lie in some neighborhood of the normal model. To determine the properties of the procedures of the robust conditional density we have also considered the case in which the assumed parametric model in correctly specified: in fact, since the new density is not the exact one, we will show that the two densities lead to similar inferential conclusions about μ when $\varepsilon=0$.

The outline of the paper is as follows. Section 2 proposes the robust conditional density to be used instead of classical normal-theory methods to achieve robustness; moreover the properties of this density are studied when ε=0. Following the approach based on influence functions, section 3 presents an expansion of the robust conditional density to derive measures of robustness and these are evaluated for two specific choices for the family of densities $\{p_\varepsilon(\cdot)\}$; moreover, the coverage probabilities of the normal-theory confidence intervals based on the new density when $\varepsilon \neq 0$ are estimated via Monte Carlo simulations. Finally, the last section contains a Monte Carlo study to evaluate the efficiency loss of the robust conditional confidence intervals with respect to the classical exact methods.

2. ROBUST CONDITIONAL DENSITY

Of course the robustness of a conditional analysis depends on the observed value of a. Hence to define a robust conditional method, the idea is first to consider a transformation of a that, in some sense, produces a robust ancillary statistic and then to define a conditional density based on this new ancillary. Instead of the usual ancillary a, we suggest to use in (1.1) a transformation $g=(g_1,\dots,g_n)=(g^*(a_1),\dots,g^*(a_n))$, with $g^*(\cdot)$ monotone. In this study, we will focus on the transformation

$$g^*(x) = \arctan\left[2k\frac{x}{\pi}\right] , \tag{2.1}$$

where k is a suitable constant. The transformation (2.1) is suggested by the robust theory of Huber's estimators, i.e. by the transformation $\psi(\cdot)$ that defines the M-estimator $\hat{\mu}_m$ for a location parameter, given by $\Sigma_{i=1}^n \psi(a_i)=0$, where $a_i=y_i-\hat{\mu}_m$ $(i=1,\dots,n)$, $\psi(x)=x\min\{1,k/|x|\}$ and k a suitable constant (see Hampel *et al.* 1986, § 2.3). The main difference between $g^*(\cdot)$ and $\psi(\cdot)$ is that (2.1) is one-to-one so that $g=(g_1,\dots g_n)$ is still ancillary. Of course other choices for $g^*(\cdot)$ in addition to (2.1) can be made; in particular, in the following, some desirable properties for this function will be given.

The aim of this section is to propose a new conditional density for t based on the robust ancillary to be used instead of (1.1) for inference about μ. We define the *robust conditional density of t given g* as

$$\tilde{f}_\varepsilon(t|g) = \tilde{c}_\varepsilon(g) \prod_{i=1}^{n} p_\varepsilon(t + g_i) , \tag{2.2}$$

where $\tilde{c}_\varepsilon(g)^{-1}=\int\prod p_\varepsilon(r+g_i)dr$. The (2.2) can be defined for any transformation $g^*(\cdot)$; in particular, taking the identity function, i.e. $g^*(x)=x$, (2.2) gives (1.1).

Observe that (2.2) is not the true conditional density of t given g, which is $f_\varepsilon(t|g)=c_\varepsilon(g)\prod p_\varepsilon(t+g^{*-1}(g_i))$, where $g^{*-1}(\cdot)$ denotes the inverse function of (2.1) and $c_\varepsilon(g)$ the normalizing constant which incorporates also the jacobian $\partial a/\partial g$.

For $\varepsilon=0$, i.e. under the assumption of normality, (2.2) defines a normal density with mean $-\bar{g}$ and variance n^{-1}, where $\bar{g}=\Sigma_{i=1}^n g^*(a_i)/n$, i.e.,

$$\tilde{f}_0(t|g)=\sqrt{\frac{n}{2\pi}}\exp\left\{-\frac{n}{2}(t+\bar{g})^2\right\}. \tag{2.3}$$

Our aim is to determine the properties of the procedures based on (2.3). Since (2.3) is not an exact conditional density, the first natural request is that, under the assumption $\varepsilon=0$, (2.3) is close to the usual conditional density $f_0(t|a)$ so that the functions $\tilde{f}_0(t|g)$ and $f_0(t|a)$ lead to similar inferential conclusions about μ when the assumed parametric model is correctly specified. Let us denote by $\tilde{F}_0(t|g)$ and $F_0(t|a)$ the distribution functions of t given g and of t given a, respectively. We want to check whether the difference

$$\left|\tilde{F}_0(k_0(g)|g)-F_0(k_0(a)|a)\right| \tag{2.4}$$

is small, where for a fixed α $(0<\alpha<1)$ $k_0(a)$ indicates the α-quantile of the normal density $f_0(t|a)$ and $k_0(g)$ the α-quantile of (2.3). Since $f_0(t|a)$ is the exact distribution of t, we have $F_0(k_0(a)|a)=\alpha$. The quantity (2.4) measures the difference between the coverage probability of the one-sided confidence intervals for μ based on (2.3) and the coverage probability of the exact ones based on the true conditional density $f_0(t|a)$.

Let us denote by k_0 the α-quantile of the standard normal density $p_0(\cdot)$ and by $P_0(\cdot)$ the standard normal distribution function. Then, (2.4) may be written as

$$\left|P_0(\sqrt{n}(k_0+\bar{g}))-P_0(\sqrt{n}k_0)\right|\sim\sqrt{n}\,|\bar{g}|\,p_0(\sqrt{n}k_0)\quad, \tag{2.5}$$

where the right-hand side in (2.5) results from a Taylor expansion of $P_0(\cdot)$ as a function of $\bar{g}$ around $\bar{g}=0$. The approximation given by (2.5) holds provided $\bar{g}=(\bar{g}-\bar{a})=(1/n)\Sigma_{i=1}^n(g_i-a_i)$ is small; this quantity has to be studied also to establish if the right-hand side in (2.5), and thus the difference (2.4), is small. To this end, we focus only on values of a_i $(i=1,\dots,n)$ in the closed interval $[-2.3,2.3]$; indeed it is shown in the Appendix that the probability that a_i $(i=1,\dots,n)$ belongs to $[-2.3,2.3]$, when $\varepsilon=0$ and

$n \leq 50$, is approximately 0.99.

A Taylor expansion of the function (2.1) around $x=0$ gives

$$g^*(x) - x = \frac{x^3}{3!}\left(- \frac{2(1+\tilde{x}^2) - 8\tilde{x}^2}{(1+\tilde{x}^2)^3} \right) = \frac{x^3}{3} \frac{(3\tilde{x}^2 - 1)}{(1+\tilde{x}^2)^3} = \frac{x^3}{3} R(\tilde{x}) \ , \quad \text{with } \tilde{x} \in [-2.3, 2.3] \qquad (2.6)$$

where $R(x)=(3x^2-1)/(1+x^2)^3$ is such that $|R(x)| \leq 1$. Given any value of x in $[-2.3,2.3]$ we have $|g^*(x)-x| \leq |(x^3/3)R(\tilde{x})| \leq |x^3/3| \leq 4.05$ and, consequently, $|\bar{g}|=|\bar{g}-\bar{a}| \leq 4.05$. Using this inequality, it can be shown that the right-hand side in (2.5) is a function of k_0 and n which tends to zero for $k_0>0.8$ and for small values of n (such as 5 or 10). Hence, we may argue that if we use the function (2.1) the difference (2.4) is small and, consequently, the functions $\tilde{f}_0(t|g)$ and $f_0(t|a)$ lead to similar inferential conclusions about μ.

This analysis is confirmed also by the Monte Carlo results presented in Table 1, in which the coverage probability of the conditional confidence intervals for μ when $\varepsilon=0$ have been estimated by integrating the densities (1.1) and (2.3), respectively. In general, for some fixed α on (0,1), the equation $\tilde{F}_0(z|g)-\tilde{F}_0(-z|g)=1-\alpha$ produces a robust conditional confidence interval for μ of nominal size $1-\alpha$ of the form

$$C(y) = (\hat{\mu} - z, \hat{\mu} + z) \ , \qquad (2.7)$$

where z indicates the $(1-\alpha/2)$-quantile of the normal density (2.3). The results presented in Table 1, and in the following simulations, are all based on 5000 Monte Carlo trials, with $\mu=0$.

As a consequence of the expansion (2.6), desirable properties for a general function $g^*(\cdot)$ such that the difference (2.4) is small are that $g^*(0)=0$, $g^{*\prime}(0)>0$, $g^{*\prime\prime}(0)=0$ and that $|g^*(x)| \leq k$, where k is a suitable constant.

	α=0.05	α=0.1
	***n*=5**	
a	0.955	0.903
g	*0.954*	*0.901*
	***n*=10**	
a	0.950	0.906
g	*0.946*	*0.900*
	***n*=20**	
a	0.950	0.902
g	*0.944*	*0.899*

Table 1. Estimates of the coverage probability of (2.7) for ε=0.

3. EXPANSION OF THE ROBUST CONDITIONAL DENSITY

The aim of this section is to show that inference about μ based on the density (2.3) is more robust than inference based on $f_0(t|a)$. In particular, using von Mises expansions for functionals (see Hampel *et al.* 1986), we will consider a Taylor expansion of the distribution (2.2) around ε=0 and we will evaluate the quantity

$$\left| \lim_{\varepsilon\to 0} \frac{\tilde{F}_\varepsilon(t|g) - \tilde{F}_0(t|g)}{\varepsilon} \right| . \tag{3.1}$$

Quantity (3.1) represents a reasonable measure of the robustness of the conditional distribution since small values of (3.1) indicate that the conditional distribution is relatively insensitive to the value of ε. More precisely, (3.1) measures the asymptotic bias of the functional caused by model misspecification (i.e. $\varepsilon\neq 0$). In this context, we prefer to use (3.1) in place of the influence function since it is of interest to consider various departures from normality. We recall that the influence function of a functional describes only the effect of an infinitesimal contamination at a point x_o, that is it assumes the contaminated density $p_\varepsilon(x)=(1-\varepsilon)p_0(x)+\varepsilon\delta_{x_o}(x)$, where $\delta_{x_o}(x)$ denotes the density function which puts mass 1 at the point x_o.

Expansion of (2.2) around ε=0 gives

$$\begin{aligned}
\tilde{f}_\varepsilon(t|g) &= \tilde{f}_0(t|g) + \varepsilon \frac{\partial}{\partial\varepsilon}\tilde{f}_\varepsilon(t|g)\Big|_{\varepsilon=0} + R \\
&= \tilde{f}_0(t|g) + \varepsilon \Big[\Big(\frac{\partial}{\partial\varepsilon}\tilde{c}_\varepsilon(g)\Big)\Big(\prod_{i=1}^{n} p_\varepsilon(t+g_i)\Big) + \tilde{c}_\varepsilon(g)\Big(\frac{\partial}{\partial\varepsilon}\prod_{i=1}^{n} p_\varepsilon(t+g_i)\Big)\Big]\Big|_{\varepsilon=0} + R \\
&= \tilde{f}_0(t|g) + \varepsilon \Big[-\tilde{c}_0(g)^2 \Big(\int_{-\infty}^{+\infty} \prod_{j=1}^{n} p_0(r+g_j) \sum_{i=1}^{n} \frac{p_0'}{p_0}(r+g_i)\, dr\Big)\Big(\prod_{i=1}^{n} p_0(t+g_i)\Big) \\
&\qquad + \tilde{c}_0(g)\prod_{j=1}^{n} p_0(t+g_j) \sum_{i=1}^{n} \frac{p_0'}{p_0}(t+g_i)\Big] + R \\
&= \tilde{f}_0(t|g) + \varepsilon \Big[\tilde{f}_0(t|g) \sum_{i=1}^{n} \frac{p_0'}{p_0}(t+g_i) - \tilde{f}_0(t|g) \int_{-\infty}^{+\infty} \tilde{f}_0(r|g) \sum_{i=1}^{n} \frac{p_0'}{p_0}(r+g_i)\, dr\Big] + R
\end{aligned} \tag{3.2}$$

where $p_o'(y)=\partial p_\varepsilon(y)/\partial\varepsilon|_{\varepsilon=0}$ and the remainder R is a term of order $O(\varepsilon^2)$. By integrating (3.2) we obtain the following expansion for the distribution function

$$\tilde{F}_\varepsilon(t|g) = \int_{-\infty}^{t} \tilde{f}_\varepsilon(r|g)\, dr$$

$$= \tilde{F}_0(t|g) + \varepsilon \Big[\int_{-\infty}^{t} \tilde{f}_0(r|g) \sum_{i=1}^{n} \frac{p_0'}{p_0}(r+g_i)\, dr - \tilde{F}_0(t|g) \int_{-\infty}^{+\infty} \tilde{f}_0(r|g) \sum_{i=1}^{n} \frac{p_0'}{p_0}(r+g_i)\, dr \Big] + R$$

$$= \tilde{F}_0(t|g) + \varepsilon \dot{F}_0(t|g) + R, \tag{3.3}$$

where

$$\dot{F}_0(t|g) = \int_{-\infty}^{t} \tilde{f}_0(r|g) \sum_{i=1}^{n} \frac{p_0'}{p_0}(r+g_i)\, dr - \tilde{F}_0(t|g) \int_{-\infty}^{+\infty} \tilde{f}_0(r|g) \sum_{i=1}^{n} \frac{p_0'}{p_0}(r+g_i)\, dr$$

$$= \sum_{i=1}^{n} \Big[\int_{-\infty}^{t} \tilde{f}_0(r|g) \frac{p_0'}{p_0}(r+g_i)\, dr - \tilde{F}_0(t|g) \int_{-\infty}^{+\infty} \tilde{f}_0(r|g) \frac{p_0'}{p_0}(r+g_i)\, dr \Big]$$

$$= \sqrt{\frac{n}{2\pi}} \sum_{i=1}^{n} \Big[\int_{-\infty}^{t} \exp\{-\tfrac{n}{2}(r+\bar{g})^2\} \frac{p_0'}{p_0}(r+g_i) dr$$

$$- \tilde{F}_0(t|g) \int_{-\infty}^{+\infty} \exp\{-\tfrac{n}{2}(r+\bar{g})^2\} \frac{p_0'}{p_0}(r+g_i)\, dr \Big] . \tag{3.4}$$

Quantity (3.4), with $g^*(x)=x$, has been used by Severini (1992) to define a measure $\dot{\alpha}_0(a)$ of conditional robustness of the coverage probability of normal-theory confidence intervals. For the robust conditional distribution, this measure becomes

$$\dot{\alpha}_0(g) = \dot{F}_0(z|g) - \dot{F}_0(-z|g)\,, \tag{3.5}$$

where z denotes the $(1-\alpha/2)$-quantile of the normal distribution (2.3). Following Severini, a small value of $|\dot{\alpha}_0(g)|$ indicates that the coverage probability is relatively insensitive to the choice $\varepsilon=0$, while a large value of $|\dot{\alpha}_0(g)|$ indicates that inferential conclusions may be very sensitive to the assumption of normality.

Our goal is to study the quantity (3.1) or equivalently, in view of (3.3), the quantity (3.4); in particular, we want to show that (3.4) is smaller if we use the robust conditional distribution (2.3) for inference about μ instead of the usual $f_0(t|a)$, for every value of t. Moreover we want to show that a similar result holds if we consider the measure of robustness (3.5).

If the family of densities $\{p_\varepsilon(\cdot)\}$ is not specified, it is not possible to evaluate the measures of robustness (3.4) and (3.5) since the function $p_0'/p_0(\cdot)$ is not known. Therefore, in the following, to study the properties of (3.4) and (3.5) we will focus on two specific choices for the family of densities $\{p_\varepsilon(\cdot)\}$, which indicates the departures from normality that are of interest.

i) Exponential Power Densities::

$$p_\varepsilon(y) = c(\varepsilon)\exp\left\{-\frac{1}{2-\varepsilon}|y|^{2-\varepsilon}\right\},\quad y\in\Re\,,\ \varepsilon<1\,, \tag{3.6}$$

with $c(\varepsilon)^{-1}=2(2-\varepsilon)^{1/(2-\varepsilon)}\Gamma(1+1/(2-\varepsilon))$ (Box 1953, Turner 1960); note that $p_\varepsilon(\cdot)$ is the standard normal density when $\varepsilon=0$, the double exponential density when $\varepsilon=1$, a moderately heavy or lighter tailed symmetric density when $\varepsilon<0$ or $\varepsilon>0$, respectively.

ii) Contaminated normal densities:

$$p_\varepsilon(y) = (1-\varepsilon)\,p_0(y) + \varepsilon\, p_1(y)\,,\qquad y\in\Re\,,\ 0\leq\varepsilon\leq 1\,, \tag{3.7}$$

where $p_1(\cdot)$ is a normal density with mean 0 and known variance σ^2. Hence, $p_\varepsilon(\cdot)$ is the density function of a random variable that is distributed according to a standard normal distribution with probability $(1-\varepsilon)$ and according to a normal distribution with mean 0 and variance σ^2 with probability ε. This model may be appropriate when it is believed that the data are normally distributed except for a few outliers.

For the Exponential Power densities (3.6) one obtains $p_0'/p_0(y)=c'-y^2/4+(y^2\log y^2)/4$, where $c'\cong 0.0676$ and, inserting this expression into (3.4), we have

$$\dot{F}_0(t|g) = \frac{1}{4}\sum_{i=1}^{n}\Big[\int_{-\infty}^{t}\tilde{f}_0(t|g)\,(r+g_i)^2\log(r+g_i)^2\,dr - \tilde{F}_0(t|g)\int_{-\infty}^{+\infty}\tilde{f}_0(t|g)\,(r+g_i)^2\log(r+g_i)^2\,dr\Big]$$
$$+\frac{1}{4}\tilde{f}_0(t|g)\,(t+\bar{g}) + n\,\tilde{F}_0(t|g)\,\bar{g}^2\,. \tag{3.8}$$

Severini's measure of robustness (3.5) is in this case given by

$$\dot{\alpha}(g) = \frac{1}{4}\sum_{i=1}^{n}\Big[\int_{-z}^{z}\tilde{f}_0(t|g)\,(r+g_i)^2\log(r+g_i)^2\,dr - (1-\alpha)\int_{-\infty}^{+\infty}\tilde{f}_0(t|g)\,(r+g_i)^2\log(r+g_i)^2\,dr\Big]$$
$$+\frac{1}{2}z\,\tilde{f}_0(z|g) + (1-\alpha)n\,\bar{g}^2\,, \tag{3.9}$$

where z is the $(1-\alpha/2)$-quantile of the distribution (2.3).

To study if (3.8) is smaller if evaluated with (2.1) than with $g^*(x)=x$, for every value of t, a Monte Carlo simulation has been performed. In particular, in Figure 2 the Monte Carlo estimates of the absolute value of (3.8) are plotted, computed with $g^*(x)=\arctan(2kx/\pi)$ and with $g^*(x)=x$, for $\varepsilon=0.5$ and $n=5$. The plot shows that, for every value of t, quantity (3.1) is smaller if we use the robust conditional distribution (2.3) instead of the usual conditional distribution $f_0(t|a)$. Here, only the plot with $\varepsilon=0.5$ and

n=5 has been considered but similar results hold for other choices of ε and n.

To study the behavior of the measure of robustness (3.9) another Monte Carlo simulation has been performed. In Table 2 the estimates of (3.9) are given, based on samples drawn from the density (3.6) for different values of ε and they show that the conditional confidence intervals based on the robust density (2.3) with (2.1) are more robust than the usual ones with $g^*(x)=x$.

Consider now the contaminated normal family of densities (3.7). We have

$$\frac{p_0'}{p_0}(y) = \frac{p_1}{p_0}(y) - 1 = \frac{1}{\sigma}\exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2}-1\right)y^2\right\} - 1 = \frac{1}{\sigma}\exp\{-\tfrac{1}{2}cy^2\} - 1, \qquad c=\frac{1}{\sigma^2}-1 ,$$

so that

$$\dot{F}_0(t|g) = \frac{1}{\sigma}\sqrt{\frac{n}{2\pi}}\sum_{i=1}^{n}\Big[\int_{-\infty}^{t}\exp\left\{-\frac{n}{2}(r+\bar{g})^2\right\}\exp\left\{-\frac{c}{2}(r+g_i)^2\right\}dr$$

$$-\tilde{F}_0(t|g)\int_{-\infty}^{+\infty}\exp\left\{-\frac{n}{2}(r+\bar{g})^2\right\}\exp\left\{-\frac{c}{2}(r+g_i)^2\right\}dr\ \Big] .$$

Since

$$\exp\left\{-\frac{n}{2}(r+\bar{g})^2-\frac{c}{2}(r+g_i)^2\right\} = \exp\left\{-\frac{nc}{2(n+c)}(\bar{g}-g_i)^2\right\}\exp\left\{-\frac{n+c}{2}\left(r+\frac{n\bar{g}+cg_i}{n+c}\right)^2\right\},$$

we have

$$\dot{F}_0(t|g)=\frac{1}{\sigma}\sqrt{\frac{n}{n+c}}\sum_{i=1}^{n}\exp\left\{-\frac{nc}{2(n+c)}(\bar{g}-g_i)^2\right\}\left[P_0\left(\sqrt{n+c}\left(t+\frac{n\bar{g}+cg_i}{n+c}\right)\right)-P_0\left(\sqrt{n}(t+\bar{g})\right)\right] . \tag{3.10}$$

Consequently, Severini's measure of robustness for this family is given by

$$\dot{\alpha}(g) = \frac{1}{\sigma}\sqrt{\frac{n}{n+c}}\sum_{i=1}^{n}\exp\left\{-\frac{nc}{2(n+c)}(\bar{g}-g_i)^2\right\}\Big[P_0\left(\sqrt{n+c}\left(z+\frac{n\bar{g}+cg_i}{n+c}\right)\right)$$

$$-P_0\left(\sqrt{n+c}\left(-z+\frac{n\bar{g}+cg_i}{n+c}\right)\right) - (1-\alpha)\Big]. \tag{3.11}$$

Also in this case a simulation study has been performed to compare (3.10) and (3.11) with (2.1) and with $g^*(x)=x$. In particular, in Table 3 estimates of (3.11), computed on the logarithmic scale, are given; these results are based on samples drawn from the density (3.7), for different values of ε, and they show that the measure (3.11) is always larger if we use the usual ancillary statistics, based on $g^*(x)=x$.

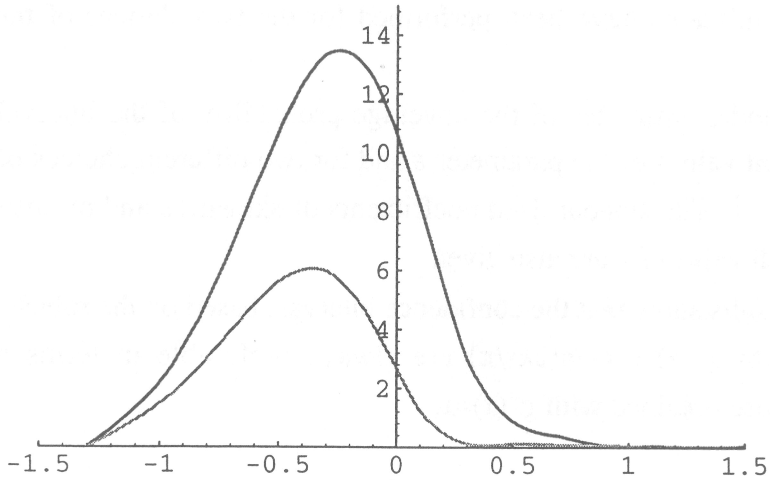

Fig. 1. Plot of (3.8) for the family (3.6); the black line refers to the ancillary a, while the gray one to g.

	$\varepsilon = -0.5$	$\varepsilon = 0.5$	$\varepsilon = 1.5$
		$n = 5$	
a	4.10	4.35	4.47
g	0.08	0.09	0.11
		$n = 10$	
a	7.84	8.74	9.04
g	0.11	0.13	0.15

Table 2. Severini's measure for the family (3.6); k=1.2; α=0.05.

	$\varepsilon = 0.01$	$\varepsilon = 0.05$	$\varepsilon = 0.1$
		$n = 5$	
a	43.41	78.35	99.12
g	2.12	2.29	2.30
		$n = 10$	
a	63.92	112.45	116.05
g	2.73	2.77	2.81

Table 3. Severini's measure for the family (3.7); k=1.2; α=0.05 (logarithmic scale).

3.1 Monte Carlo simulations

To study the properties of the normal-theory confidence intervals (2.7) when $\varepsilon \neq 0$, we evaluate their coverage probability when the data are distributed according to $p_\varepsilon(\cdot)$, $\varepsilon \neq 0$; this approach is called *criterion robustness* by Box and Tiao (1964). To this end Monte Carlo simulations have been performed for the two choices of the family of densities $\{p_\varepsilon(\cdot)\}$.

In Tables 4 and 5 estimates of the coverage probability of the intervals (2.7) are given for different values of the parameter ε and for two different choices of the family of densities $\{p_\varepsilon(\cdot)\}$. The standardized coefficients of skewness and curtosis ρ_3 and ρ_4 associated to each value of ε are also given.

Simulation results show that the confidence intervals based on the robust conditional density (2.3) with $g^*(x)=\arctan(2kx/\pi)$ are always preferable in terms of *criterion robustness* to those obtained with $g^*(x)=x$.

α		ε=-1.5 ρ_3=0, ρ_4=-0.71	ε=-0.5 ρ_3=0, ρ_4=-0.36	ε=0.5 ρ_3=0, ρ_4=0.76
	***n*=5**			
0.05	*a*	0.921	0.939	0.966
	g	*0.938*	*0.949*	*0.967*
0.1	*a*	0.885	0.894	0.915
	g	*0.899*	*0.901*	*0.925*
	***n*=10**			
0.05	*a*	0.918	0.934	0.940
	g	*0.923*	*0.941*	*0.949*
0.1	*a*	0.883	0.891	0.908
	g	*0.900*	*0.907*	*0.917*
	***n*=20**			
0.05	*a*	0.913	0.931	0.943
	g	*0.927*	*0.942*	*0.952*
0.1	*a*	0.882	0.898	0.907
	g	*0.896*	*0.906*	*0.910*

Table 4. Estimates of the coverage probability for the Exponential Power densities (3.6).

α		ε=0.01 ρ_3=0, ρ_4=-0.11	ε=0.05 ρ_3=0, ρ_4=-0.5	ε=0.1 ρ_3=0, ρ_4=-0.86
	***n*=5**			
0.05	*a*	0.930	0.860	0.804
	g	*0.936*	*0.886*	*0.845*
0.1	*a*	0.890	0.800	0.753
	g	*0.892*	*0.816*	*0.795*
	***n*=10**			
0.05	*a*	0.927	0.850	0.764
	g	*0.934*	*0.882*	*0.830*
0.1	*a*	0.878	0.785	0.694
	g	*0.892*	*0.823*	*0.754*
	***n*=20**			
0.05	*a*	0.922	0.830	0.733
	g	*0.940*	*0.891*	*0.813*
0.1	*a*	0.870	0.759	0.654
	g	*0.883*	*0.818*	*0.733*

Table 5. Estimates of the coverage probability for the contaminated normal densities (3.7).

4. EFFICIENCY OF THE ROBUST CONDITIONAL CONFIDENCE INTERVALS

Robust methods are preferable to classic procedures, since they behave better under slight violations of the model. However, they do not enjoy any optimality property and, consequently, we have to evaluate their efficiency loss with respect to the exact classical methods. Following the Neyman approach, a 1-α confidence interval $C(y)$ is optimal if the probability that it contains any parameter value μ' different from the exact one is a minimum, i.e. if

$$P\{ \mu' \in C(y) \, ; \mu \} \tag{4.1}$$

is a minimum, where μ denotes the true value of the parameter, $\mu \neq \mu'$.

To evaluate the efficiency loss of the normal-theory confidence intervals (2.7) computed with $g^*(x)=\arctan(2kx/\pi)$ with respect to those computed with $g^*(x)=x$, some Monte Carlo simulations have been performed to estimate the probability (4.1) for different values of ε and the different choices of the family $\{p_\varepsilon(\cdot)\}$. The results given in the following tables have been computed with $\mu=1$, $\mu'=0$ and for a fixed nominal α.

In Table 6 estimates of the probability (4.1) for $\varepsilon=0$ are given and the results show that the robust procedure has an efficiency loss with respect to the classical procedures; however, the estimates (4.1) are not very different.

In Tables 7 and 8 estimates of the probability (4.1) are given for different values of the parameter ε and for different choices of the family of densities $\{p_\varepsilon(\cdot)\}$. The simulation results show that when $\varepsilon \neq 0$ the robust methods do not have always an efficiency loss with respect to the classical procedures; in particular, the tables show that the estimates of (4.1) are smaller for the robust procedure.

	α=0.05	α=0.1
	***n*=5**	
a	0.403	0.281
g	*0.423*	*0.311*
	***n*=10**	
a	0.126	0.070
g	*0.139*	*0.082*
	***n*=20**	
a	0.003	0.001
g	*0.005*	*0.002*

Table 6. Estimates of (4.1) when ε=0.

α		ε=-0.5	ε=0.5	ε=1.5
	***n*=5**			
0.05	*a*	0.400	0.382	0.378
	g	*0.394*	*0.380*	*0.378*
0.1	*a*	0.282	0.254	0.238
	g	*0.277*	*0.248*	*0.236*
	***n*=10**			
0.05	*a*	0.124	0.075	0.055
	g	*0.110*	*0.073*	*0.058*
0.1	*a*	0.074	0.033	0.025
	g	*0.056*	*0.032*	*0.023*
	***n*=20**			
0.05	*a*	0.005	0.003	0.003
	g	*0.003*	*0.002*	*0.002*
0.1	*a*	0.002	0.001	0.001
	g	*0.002*	*0.001*	*0.001*

Table 7. Estimates of (4.1) for the Exponential Power densities (3.6).

α		ε=0.01	ε=0.05	ε=0.1
	***n*=5**			
0.05	*a*	0.379	0.399	0.397
	g	*0.387*	*0.407*	*0.407*
0.1	*a*	0.271	0.297	0.296
	g	*0.273*	*0.303*	*0.300*
	***n*=10**			
0.05	*a*	0.131	0.177	0.228
	g	*0.132*	*0.165*	*0.204*
0.1	*a*	0.081	0.124	0.170
	g	*0.080*	*0.109*	*0.150*
	***n*=20**			
0.05	*a*	0.014	0.050	0.078
	g	*0.010*	*0.029*	*0.050*
0.1	*a*	0.010	0.036	0.061
	g	*0.005*	*0.018*	*0.035*

Table 8. Estimates of (4.1) for the contaminated normal densities (3.7).

5. FINAL REMARKS

In this paper a robust conditional density is suggested; another interpretation of this density is the following.

The basic idea is to approximate the exact $p_\varepsilon(t+a_i)$ by $cp_0(t+g_i)$, for some constant c and some function $g_i=g^*(a_i)$ $(i=1,...,n)$, for all t. E.g., if $p_\varepsilon(t+a_i)=cp_0(t+g_i)$ for all i and all t, then the robust conditional density will coincide with the exact density. Taking the logarithms and expanding in a Taylor series around $t=0$, this is equivalent to require that $t\cdot dp_\varepsilon(a_i)/p_\varepsilon(a_i)+...=t\cdot dp_0(g_i)/p_0(g_i)+...$, where $dp_\varepsilon(x)=\partial p_\varepsilon(x)/\partial x$ and $dp_0(x)=\partial p_0(x)/\partial x$. Since $p_0(\cdot)$ represents the standard normal density, we have that $dp_0(g_i)/p_0(g_i)=-g_i$ and to achieve equality of the first term we need to take $g_i=-dp_\varepsilon(a_i)/p_\varepsilon(a_i)$. The right-hand side of this equation is simply the score function based on the density $p_\varepsilon(\cdot)$. When $p_\varepsilon(\cdot)$ is unknown it is reasonable to use the function $\psi(\cdot)$ of a robust estimator, which is essentially what we have done in equation (2.1). This observation may be useful in suggesting other choices for the robust ancillary g.

The results presented in this study suggest several areas that need further consideration:

1) the choice of the transformation of the ancillary statistic requires more study since the simulation results are promising but not entirely satisfactory;
2) the procedure should be generalized to location and scale models and to the more general linear regression model;
3) other departures from $p_0(\cdot)$ can be considered;
4) it would be interesting to study the properties of the density (2.2) when $p_0(\cdot)$ is a non-normal density.

APPENDIX: MARGINAL DISTRIBUTION OF THE ANCILLARY STATISTIC

We now consider the problem of showing that a generic component of the ancillary statistic a belongs to the closed interval [-2.3,2.3] with approximate probability 0.99, when $\varepsilon=0$ and $n\leq 50$. To this end, we have to determinate the marginal distribution of the ancillary statistic.

Consider the one-to-one transformation $(y_1,...,y_n)\leftrightarrow(\hat{\mu}_0,a_1,...,a_{n-1})$, where $\hat{\mu}_0$ indicates an equivariant estimator of μ under $p_\varepsilon(\cdot)=p_0(\cdot)$. Since $p(y_1,...,y_n;\mu)=\prod p_0(y_i-\mu)$ and $y_i=\hat{\mu}+a_i$ $(i=1,...,n)$, the joint density of $(\hat{\mu}_0,a_1,...,a_{n-1})$ is $p(\hat{\mu}_0,a_1,...,a_{n-1};\mu)=\prod p_0(\hat{\mu}_0-\mu+a_i)|J|$, where $|J|$ is the determinant of the jacobian of the transformation which gives the observations $(y_1,...,y_n)$ in terms of $(\hat{\mu}_0,a_1,...,a_{n-1})$; the component a_n of the ancillary statistic can be studied as a function of $(a_1,...,a_{n-1})$ through the estimating equation for $\hat{\mu}_0$. Since $y_1=\hat{\mu}_0+a_1$, ..., $y_{n-1}=\hat{\mu}_0+a_{n-1}$, $y_n=\hat{\mu}_0+a_n(a_1,...,a_{n-1})$, one obtains

$$|J| = \begin{vmatrix} 1_{n-1} & I_{n-1} \\ 1 & f(a) \end{vmatrix} ,$$

where 1_{n-1} is the $(n-1)$-column vector with elements equal to 1, I_{n-1} is the $(n-1)\times(n-1)$ identity matrix and $f(a)$ is the $(n-1)$-row vector with elements $f_i=\partial a_n(a_1,...,a_{n-1})/\partial a_i$ $(i=1,...,n-1)$. Using a well-known formula for the computation of the determinant of a block-matrix, we have $|J| = |I_{n-1}|\,|1-f(a)I_{n-1}1_{n-1}| = |1-\Sigma f_i|$. Let us denote by $p(a_1,...,a_{n-1})$ the marginal density of $(a_1,...,a_{n-1})$; then

$$p(a_1,...,a_{n-1}) = |J| \int_{-\infty}^{+\infty} \prod_{i=1}^{n} p_0(t+a_i)\,dt = |1-\sum_{i=1}^{n-1} f_i| \int_{-\infty}^{+\infty} \prod_{i=1}^{n} p_0(t+a_i)\,dt \; . \tag{A.1}$$

When $\hat{\mu}_0$ is the sample mean, we have $a_n=a_n(a_1,...,a_n)=-\sum_{i=1}^{n-1}a_i$; hence $f(a)$ reduces to the $(n-1)$-row vector with all the elements equal to -1, and consequently $|J|=|1+n-1|=n$ and with simple calculations (A.1) reduces to

$$p(a_1,...,a_{n-1}) = c\,\exp\left\{-\frac{1}{2}\sum_{i=1}^{n} a_i^2\right\} , \tag{A.2}$$

where $a_n=-\sum_{i=1}^{n-1}a_i$, and c is the normalizing constant which is independent of $a_1,...,a_{n-1}$. The marginal density of a_i $(i=1,...,n-1)$ can be obtained from (A.2) by integration and it is given by

$$p(a_i) = c\,\exp\left\{-\frac{n+1}{2n}\,a_i^2\right\} , \tag{A.3}$$

where c indicates the normalizing constant. The marginal density of the component a_n of a can be obtained by thinking a_n as a function of $(a_1,...,a_{n-1})$, i.e. in this case using the relation $a_n=-\sum_{i=1}^{n-1}a_i$. With simple calculations, for different values of n, it is easy to show that a_i $(i=1,...,n)$ satisfies the required assumption of section 2, i.e. that it takes values in the closed interval [-2.3,2.3] with probability approximately 0.99.

REFERENCES

Box, G.E.P. (1953). A note on regions of kurtosis. *Biometrika*, **40**, 465-468.

Box, G.E.P. and Tiao, G.C. (1964). A note on criterion robustness and inference robustness. *Biometrika*, **51**, 169-173.

Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*. London: Chapman and Hall.

Hampel, F.R., Ronchetti, E.M., Rousseau, J. and Stahel, W.A. (1986). *Robust Statistics: The Approach Based on Influence Function*. New York: Wiley.

Severini, T.A. (1992). Conditional robustness of normal-theory inference for a location parameter. *Technical Report, Dept. of Statistics, Northwestern University, Evanston.*

Turner, M.E. (1960). On heuristic estimation methods. *Biometrics*, **16**, 299-401.