A robust conditional approximation of marginal tail probabilities

A. R. Brazzale, L. Ventura

2001.12

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

Luglio 2001

A robust conditional approximation of marginal tail probabilities

A.R. Brazzale
Institute of Systems Science and Biomedical Engineering, National Research Council, Padova, Italy
L. Ventura
Department of Statistics, University of Padova, Padova, Italy

Abstract: The aim of this contribution is to derive a robust approximate conditional procedure used to eliminate nuisance parameters in regression and scale models. Unlike the approximations to exact condtional solutions based on the likelihood function and on the maximum likelihood estimator, the robust conditional approximation of marginal tail probabilities does not suffer from lack of robustness to model misspecification. To assess the performance of the proposed robust conditional procedure the results of sensitivity analyses are discussed.

Keywords: Ancillary; Asymptotic Theory; Conditioning; *M*-estimator; Regression and Scale Model; Robustness; Sensitivity Analysis

1 Introduction

Conditioning is a very natural concept, and most statisticians, theoretical and applied, use it in their everyday work. With a few exceptions, all analyses condition on the observed sample size. Regression analyses in observational studies are almost always performed conditionally on the observed values of the covariates. Conditioning arguments are present in all paradigms of statistical inference. Bayesians condition on the observed data. The Neyman-Pearson approach uses conditioning to retrieve optimality of some test criteria. It is almost impossible to cover all the roles of conditioning in inference. An excellent review is given in Reid (1995).

As first outlined by Fisher (1934), the conditioning argument provides an elegant way to make inference in regression and scale models, that is, in models characterized by a linear predictor and not necessarily normal errors. Conditioning on the sample configuration reduces the dimension of the original data to the dimension of the parameter. Nuisance parameters are then eliminated by integrating, that is, by finding the marginal density related to the component of interest. Applications of regression and scale models are found in many areas of statistics as, for example, survival analysis or industrial applications. Frequent choices for the error distribution include the Gumbel or extreme value distribution and Student's t distribution.

Although well defined, exact conditional inference is seldom feasible in practice because of the computational burden involved in deriving the marginal distribution of interest. Recently developed higher-order asymptotics (see Cox, 1988, and Reid, 1995, for a thorough account) provide excellent approximations to exact conditional solutions. These methods have been found to suffer from lack of robustness to model misspecification (Brazzale and Ventura, 2001). The aim of our research is to derive a robust approximation of marginal tail probabilities applicable when, by conditioning arguments, inference about unknown parameters is achieved through exact pivots whose joint density is known except, perhaps, for a normalizing constant. The robust conditional approximation considered here is based on the results of DiCiccio et al. (1990).

The paper organizes as follows. Sections 2 and 3 give a short review of exact and approximate conditional inference in regression and scale models. Our research is discussed in Section 4. Sections 5 gives the results of sensitivity analyses carried out to assess the performance of the proposed robust conditional approximation.

2 Regression and scale models

Regression and scale models belong to the wider class of transformation models, which represent an important class of statistical models to which the principles of model and data reduction apply. A regression and scale model has the form

$$y = X\beta + \sigma\varepsilon \ , \tag{1}$$

where X is a fixed $n \times p$ matrix, $\beta \in \mathbb{R}^p$ an unknown regression coefficient, $\sigma > 0$ a scale parameter, and ε represents an n-dimensional vector of errors that are independent and identically distributed according to a known density $p_0(\cdot)$ on $\mathbb{R}$. For the ith response we write $y_i = x_i^{\mathsf{T}}\beta + \sigma\varepsilon_i$, where x_i^{T} is the ith row of X. Given a sample $y = (y_1, \dots, y_n)$, the sample configuration $a = (a_1, \dots, a_n)$, with $a_i = (y_i - x_i^{\mathsf{T}}\hat{\beta})/\hat{\sigma}$, where $(\hat{\beta}, \hat{\sigma})$ is the maximum likelihood (or any equivariant) estimate of the parameters β and σ, is an ancillary for β and σ (Fraser, 1979). Given a, the pivot

$$(Q_1, Q_2) = \left(\frac{\hat{\beta} - \beta}{\hat{\sigma}}, \frac{\hat{\sigma}}{\sigma} \right) \ ,$$

has exact conditional density

$$p(q_1, q_2 | a) = c(a) q_2^{n-1} \prod_{i=1}^{n} p_0 \left\{ (x_i^{\mathsf{T}} q_1 + a_i) q_2 \right\} \ , \tag{2}$$

where $c(a)$ is a normalizing constant. Several authors, among whom are Fisher (1934) and Fraser (1979), suggest that inference on the parameters (β, σ) should be made conditionally on the observed value a. Exact conditional confidence intervals for any single parameter, say β_r, $r = 1, \dots, p$, or σ, are based on the marginal density of the related pivot obtained by integrating out the remaining components in (2). Examples thereof are given in Lawless (1972, 1973, 1978) for applications to Cauchy, logistic, Weibull and extreme value distributions and Kappenman (1975) for the Laplace distribution, in the case of a scale and location model, and Fraser (1979) for linear models.

Exact calculation of the marginal distribution for the pivots of interest in general involves multidimensional numerical integration, and can be hard to do in practice. For instance, the normalizing constant $c(a)$ is given by

$$c(a)^{-1} = \int_0^{\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} q_2^{n-1} \prod_{i=1}^{n} p_0 \left\{ (x_i^{\top} q_1 + a_i) q_2 \right\} dq_{11} \ldots dq_{1p} dq_2 \ ,$$

where $q_{1r} = (\hat{\beta}_r - \beta_r)/\hat{\sigma}$, $r = 1, \ldots, p$. Otherwise, if only a subset of the parameters is of interest, it is straighforward to eliminate the remaining ones by integrating (2) with respect to the components of Q_1 and Q_2 that depend on the nuisance parameters, leading to a marginal conditional distribution that only depends on the parameters of interest. Yet, though well defined, exact marginal inference for linear regression models is seldom feasible in practice. The computational efforts required grow rapidly beyond what is feasible in a reasonable amount of time and with acceptable accuracy, especially if the number of parameters is large and the dimension of the parameter of interest low. One ways to overcome this problem is by resorting to small-sample approximations.

3 Higher-order approximation of marginal distributions

Recently, several higher-order asymptotics have been developed that provide extremely accurate approximations to exact conditional inference, hence allowing us to by-pass numerical integration of (2). In the following, we focus on the approach of DiCiccio et al. (1990). Their results are derived in a general setting, i.e. for a real-valued vector variable $X = (X_1, \ldots, X_d)$ such that $X_i = O_p(n^{-1/2})$, for $i = 1, \ldots, d$, as some parameter n increases indefinitely. In applications, n usually represents a number of observations. Suppose that the density of X is given by

$$p(x) \propto \exp\{l(x)\} \ , \ x = (x_1, \ldots, x_d) \ , \tag{3}$$

where function $l(\cdot)$ is known. It is assumed that $l(x)$ is of order $O(n)$ for each fixed x and that $l(x)$ attains its maximum at $x = 0$. We are interested in approximating the marginal distribution function of a single component of X, say X_1, that is, in approximating integrals of the form

$$Pr(X_1 \leq x_1) = \int_{-\infty}^{x_1} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} p(z_1, \ldots, z_d) dz_1 dz_2 \ldots dz_d \ . \tag{4}$$

Integrals of this form are for instance common in the analysis of the linear regression models and show up when we derive the marginal distribution of one of the pivots (Q_1, Q_2), whose joint distribution is (2).

Let $\tilde{x}(x_1)$ be the point at which $l(x)$ is maximized subject to the constraint that the first component of x equals the specified value x_1. Let us define the statistic

$$r = r(x_1) = \text{sgn}(x_1)\left\{2[l(0) - l(\tilde{x}(x_1))]\right\}^{1/2} ,$$

which, as shown in DiCiccio et al. (1990), has a standard normal distribution to the first order, that is

$$Pr(X_1 \leq x_1) = \Phi(r)\{1 + O(n^{-1/2})\} ,$$

where Φ denotes the cumulative distribution function of the standard normal. In the same paper, an improved version is given that has error of order $O(n^{-3/2})$ and takes a form similar to the tail area approximation discussed, for instance, in Barndorff-Nielsen and Cox (1994). The approximation involves only first and second derivatives of $l(x)$, and is given by

$$Pr(X_1 \leq x_1) = \Phi(r) + \phi(r)(r^{-1} - u^{-1}) + O(n^{-3/2}) , \tag{5}$$

where ϕ is the density of the standard normal and

$$u = -\frac{l_1(\tilde{x}(x_1))|[-l_{st}(\tilde{x}(x_1))]|^{1/2}}{|[-l_{ab}(0)]|^{1/2}} .$$

In the former expression $l_1(x) = dl(x)/dx_1$ and $l_{ab}(x) = \partial^2 l(x)\partial x_a \partial x_b$, with $a, b = 1, \ldots, d$ and $s, t = 2, \ldots, d$. An asymptotically equivalent version is

$$Pr(X_1 \leq x_1) = \Phi(r^*)\{1 + O(n^{-3/2})\} , \tag{6}$$

where

$$r^* = r + r^{-1}\log(u/r) .$$

Expressions (5) and (6) remain valid even if the density $c\exp\{l(x)\}$ only approximates the true density $p(x)$ with relative error $O(n^{-3/2})$, and they can be easily generalized in cases where the joint density of the variables is maximized at a point other than 0; see DiCiccio et al. (1990, formula (16)).

DiCiccio et al. (1990) have also computed (5) in the specific context of linear regression models (1), using the parameterisation (β, θ), where $\theta = \log\sigma$ (see also DiCiccio and Field, 1991). Let $Q_1 = (\hat{\beta} - \beta)/e^{\hat{\theta}}$ and $Q_2 = \hat{\theta} - \theta$ be the associated pivots based on the maximum likelihood estimate for (β, θ). They are of the order $O_p(n^{-1/2})$. In this case, the joint conditional distribution of the pivots satisfies (3), with

$$l(q_1, q_2) = nq_2 - \sum_{i=1}^{n} g\left\{(x_i^{\mathsf{T}} q_1 + a_i)e^{q_2}\right\} , \tag{7}$$

where $g(x) = -\log p_0(x)$. Straightforward calculation shows that (7) is maximized at 0, that is if $\hat{\beta} = \beta$ and $\hat{\theta} = \theta$. Note that this would not hold if we decided not to work on a logarithmic scale for the parameter σ.

Let us now fix $Q_j = Q_{1j} = (\hat{\beta}_j - \beta_j)/e^{\hat{\theta}}$, i.e. one of the components of Q_1, or $Q_j = Q_2 = \hat{\theta} - \theta$. The constrained estimates of the remaining components in (7) are equivalent to replacing β and θ with the constrained maximum likelihood estimates obtained by maximizing the loglikelihood function subject to the constraint that $Q_j = q_j$. Approximations to the marginal conditional distribution of the pivot of interest can be obtained from the tail area approximation (5) or, equivalently, by the corresponding r^*-type version (6). In this case the errors in (5) and (6) are of order $O(n^{-3/2})$ conditionally on a and unconditionally. For computational aspects of (5) and (6) in regression and scale models see DiCiccio et al. (1990, Sec. 4), DiCiccio and Field (1991) and also Brazzale (2000, Sec. 3.2.2).

4 A robust conditional tail area approximation

Recent research pointed out that some higher-order approximations suffer lack of robustness to model misspecification. For instance, Ronchetti and Ventura (2001) discuss the effects of model misspecification on higher-order asymptotic approximations of the distribution of classical statistics. They show that small deviations from the assumed model can wipe out the improvements of the accuracy obtained by second-order approximations. Moreover, Brazzale and Ventura (2001) investigate the stability with respect to single point contamination of the p-values based on (5) and (6) in the context of linear regression models by means of a sensitivity analysis. They show that even a small change of one single influential observation can have a strong impact on the test results based on (5) and (6). These findings concern the marginal performance of the approximate procedures and not the conditional behaviour given the configuration statistic a. On the contrary, the conditional simulation study performed in Brazzale and Ventura (2001) by means of MCMC techniques suggests that higher-order solutions are less infuenced by contamination of the error distribution if only the conditional sample space is taken into account and that robustness properties of (5) and (6) depend on the nature of the contaminating distribution.

The aim of this contribution is to discuss a robust version of the approximate marginal distribution (5) or, equivalently, of (6), that presents robustness properties both from a marginal point of view and also conditionally on an ancillary a. The approximation is based on the results given in DiCiccio et al. (1990). Robustness of the proposed approximation is investigated in Sections 4 and 5 by means of sensitivity analyses and also by simulation studies.

Expression (2) for the exact conditional density holds for any equivariant estimate of the parameters β and σ used to define the configuration ancillary a. In the following, to achieve robustness, we focus on robust equivariant estimators of β and σ. In particular, we consider the general class of robust M-stimators for linear regression models (see Hampel et al., 1986, chap. 4). An M-estimator is defined as the solution of the unbiased estimating function

$$\sum_{i=1}^{n} \psi(y_i;\beta,\sigma) = 0 \ , \tag{8}$$

where $\psi(\cdot)$ is a given function. Under broad conditions, M-estimators are consistent and asymptotically normal. Moreover, M-estimators are B-robust if and only if the corresponding $\psi(\cdot)$ function is bounded.

Let us consider the parameterisation (β, θ) and let $Q_1 = (\hat{\beta} - \beta)/e^{\hat{\theta}}$ and $Q_2 = \hat{\theta} - \theta$ be the associated pivots, based on the M-estimates for β and θ. They are still of order $O_p(n^{-1/2})$ and their joint exact conditional distribution still satisfies (3), with $l(q_1, q_2)$ given in (7). In this case, function (7), unlike the standard case based on maximum likelihood estimates discussed in the former section, does not attain its maximum value at 0. Indeed, differentiation of (7) with respect to (q_1, q_2) gives

$$l_{q_r} = -e^{q_2} \sum_{i=1}^{n} g^{(1)} \left[(x_i^{\mathsf{T}} q_1 + a_i) e^{q_2} \right] x_{ir} \quad , r = 1, \ldots , p \ , \tag{9}$$

$$l_{q_2} = n - e^{q_2} \sum_{i=1}^{n} g^{(1)} \left[(x_i^{\mathsf{T}} q_1 + a_i) e^{q_2} \right] (x_i^{\mathsf{T}} q_1 + a_i) \ , \tag{10}$$

where $g^{(1)}(x) = dg(x)/dx$. If (Q_1, Q_2) is based on the maximum likelihood estimate of (β, θ), then (9) and (10) are the usual likelihood score equations and (7) attains its maximum at 0. But if (Q_1, Q_2) is based on an M-estimate for (β, θ), then the solution of $l_{q_r} = 0$, $r = 1, \ldots , p$, and $l_{q_2} = 0$ must be computed numerically. In the following, we denote by δ the maximum of (7) and, since the maximum likelihood estimate is a particular case of (8), we continue to use the parameterisation (β, θ).

Let Q_j be the component of (Q_1, Q_2) that is of interest and of which we want to compute the approximate marginal distribution. Let $\tilde{\delta}(q_j)$ be the point at which (7) attains its maximum value subject to the constraint that the component Q_j equals the specific value q_j. Then, provided $q_j - \delta_j$ is $O(n^{-1/2})$, the robust approximation of the marginal conditional tail probability of Q_j given the ancillary a, is given by

$$Pr(Q_j \le q_j | a) = \Phi(r_b) + \phi(r_b)(r_b^{-1} - u_b^{-1}) + O(n^{-3/2}) \ , \tag{11}$$

where

$$r_b = r_b(q_j) = \mathrm{sgn}(q_j - \delta_j) \left[2(l(\delta) - l(\tilde{\delta}(q_j))) \right]^{1/2}$$

and

$$u_b = u_b(q_j) = -\frac{l_j(\tilde{\delta}(q_j)) |[-l_{st}(\tilde{\delta}(q_j))]|^{1/2}}{|[-l_{ab}(\delta)]|^{1/2}} \ ,$$

and a, b vary over $1, \ldots, p, p+1$ and s, t vary over $1, \ldots, p, p+1 \neq j$. The second derivatives of (7) are given by

$$
\begin{aligned}
l_{q_r q_s} &= -e^{2q_2} \sum_{i=1}^{n} g^{(2)} \left[(x_i^{\mathsf{T}} q_1 + a_i) e^{q_2} \right] x_{ir} x_{is} \; , r, s = 1, \ldots, p \; , \\
l_{q_r q_2} &= -e^{q_2} \sum_{i=1}^{n} g^{(1)} \left[(x_i^{\mathsf{T}} q_1 + a_i) e^{q_2} \right] x_{ir} - \\
&- e^{2q_2} \sum_{i=1}^{n} g^{(2)} \left[(x_i^{\mathsf{T}} q_1 + a_i) e^{q_2} \right] (x_i^{\mathsf{T}} q_1 + a_i) x_{ir} \; , r = 1, \ldots, p \; , \\
l_{q_2 q_2} &= -e^{2q_2} \sum_{i=1}^{n} g^{(2)} \left[(x_i^{\mathsf{T}} q_1 + a_i) e^{q_2} \right] (x_i^{\mathsf{T}} q_1 + a_i)^2 - \\
&- e^{q_2} \sum_{i=1}^{n} g^{(1)} \left[(x_i^{T} q_1 + a_i) e^{q_2} \right] (x_i^{\mathsf{T}} q_1 + a_i) \; ,
\end{aligned}
$$

where $g^{(2)} = d^2 g(x)/dx^2$.

An asymptotically equivalent version of (11) is given by

$$
Pr(Q_j \leq q_j | a) = \Phi(r_b^*) \{ 1 + O(n^{-3/2}) \} \; , \tag{12}
$$

where

$$
r_b^* = r_b + r_b^{-1} \log(u_b / r_b) \; .
$$

The errors in (11) and (12) are of order $O(n^{-3/2})$ conditionally on a and unconditionally.

For the univariate case, i.e. in the simple location model or in the simple scale model, (11) reduces to

$$
Pr(Q \leq q | a) = \Phi(r_b) + \phi(r_b) \left[\frac{1}{r_b} + \frac{[-l^{(2)}(\delta)]^{1/2}}{l^{(1)}(q)} \right] + O(n^{-3/2}) \; , \tag{13}
$$

where r_b is now defined as $r_b = r_b(q) = \mathrm{sgn}(q - \delta)[2(l(\delta) - l(q))]^{1/2}$ and $l^{(1)} = -\sum g^{(1)}(q + a_i)$, $l^{(2)} = -\sum g^{(2)}(q + a_i)$ in the location model, or $l^{(1)} = n - e^q \sum g^{(1)}(a_i e^q)$, $l^{(2)} = -e^q \sum g^{(1)}(a_i e^q) - e^{2q} \sum g^{(2)}(a_i e^q)$ in the scale model.

Note that the only difficulty in implementing (11) rather than the usual expression based on the maximum likelihood estimators (5) is the calculation of the maximizing point δ and of $\tilde{\delta}(q_j)$. In fact, the determination of the first two derivatives of (7) is the same in the two cases. On the other hand, it must be noted that the higher-order conditional approximation (11) is much easier to compute than the approximation for the marginal distribution function for M-estimators discussed in Field and Ronchetti (1990, chap. 6), based on saddleponit techniques.

5 Sensitivity analyses

Procedures that are not overly influenced by clearly outlying observations are of practical interest. The theory of robust statistics is concerned with statistical procedures that are stable even under small deviations from the assumed statistical model.

To assess the robustness properties of a method, a useful tool is the *empirical influence function* (Hampel et al., 1982). A natural way to study the stability of a method is to move a single observation x of the observed data and then explore the effect of changing x. Formally, choose a fixed sample, either a real data set, a simulated sample, or made-up values. Choose the observation x of the observed sample (in a linear regreggion model x is typically an influential observation), move x over a grid of values, find the values of a quantity of interest (for instance a p-value) and plot them as a function of the position of x. This will typically show whether the procedure does qualitatively what it is expected to do. If it fails, it may become clear what part of the procedure should be improved. A related tool is Tukey's (1977) ***sensitivity curve***. It is simply the normalized effect of the observation x on the quantity of interest, i.e. the normalized version of the empirical influence function. It can be shown that in many situations for $n \to \infty$ the limit of the sensitivity curve is the influence function. The influence curve is a fundamental tool to assess the robustness properties of a method (Hampel et al., 1986). However, although the definition of the influence function is natural in the case of marginal inference, its definition is not clear in the case of the conditioning on the configuration ancillary and further research is required on the appropriate definition in this case. This difficulty also arises in the context of time series (see Hampel et al., 1986, Section 8.3) or in the context of mixed linear models (see Welsh and Richardson, 1997) where there is no entirely satisfactory definition of the influence function. For this reason, in the following we focus on the empirical influence function to investigate the stability of (11) with respect to single point contaminations.

Example 1: Sea Levels in Venice

To compare (5) and (11), we perform an empirical sensitivity analysis of the associated p-values for testing problems. This study is illustrated by an application to the data discussed in Smith (1986), which concern the maximum sea levels measured in Venice; see also Pirazzoli (1982). It should be made clear that the main purpose of what follows is to illustrate the application of our statistical technique, not to provide a substantial reappraisal of the flood danger in Venice. Thus we shall take no account of physical effects such as geological subsidence, the pumping of water from underground sources or the construction of new channels. On the other hand, the problems which we do discuss (stability with respect to single point contamination) arise in many statistical analyses of this kind, so our discussion of these problems is relevant in a broader context.

Venice data consist in the ten largest values (with a few exceptions) of sea levels for the years 1887-1981. Following Smith (1986), the data should be divided into two halves: roughly constant sea level with irregular fluctuations up to 1930, steadily increasing thereafter. More detailed analysis of the maxima confirms this impression, and it was therefore decided to analyse only the data from 1931-1981. Smith (1986) shows that a good reference model is a linear regression model with extreme value distributed error, i.e. with density function $p_0(x) = \exp(-x - \exp(-x))$, with $x \in \mathbb{R}$. The linear predictor includes a quadratic trend in the years (Smith, 1986, model 3.1) plus a 19-year seasonal component due to periodic tidal fluctuations. Maximum likelihood estimation is discussed in Smith (1986). Interest focuses on the quadratic term.

We investigate the robustness of tests based on (5) and on (11), with respect to single point contaminations. This corresponds to a contamination $\varepsilon = 1/(n-1) = 0.02$, that is of around 2% of the data. In particular, the analysis is run with respect to the most influential observation, corresponding to year 1932. The corresponding observed sea level in Venice is 78cm. This observation is perturbed and allowed to vary in the range 40cm to 120cm. At each time (5) and (11) are recomputed. The resulting p-values for assessing the significance of the scalar parameter of interest are shown in Figure 1. The M-estimators considered are the well-known robust Huber estimators for linear regression models (see Hampel et al., 1986, Section 6.2).

A general inspection of Figure 1 shows steep p-values curves for (5), and almost constant p-values for (11). Therefore, for the standard higher-order approximation (5) the most striking feature is that even a small change in the observed sea level has a strong impact on the observed significance level. This finding is in agreement with Ronchetti and Ventura (2001). On the contrary, for the robust conditional approximation (11) a small change of the sea level does not imply a substantial change of the p-values.

This analysis confirms that when testing contaminated models by (5) even small changes of a single observations can have a strong impact on the test results. This is a valid reason to justify the use of the robust approximation (11)

Example 2: Further evidence with House price data

Sen and Srivastava (1990, page 32) consider a data set (26 observations) on house prices. Among the variables examined are the selling price in thousands of dollars (y), the number of bedrooms (x_1), the floor space in square feet (x_2), the total number of rooms (x_3) and the front footage of lot in feet (x_4). The model can be written as

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \sigma \varepsilon_i \ , i = 1, \dots , 26 \ , \qquad (14)$$

where ε_i is taken to be standard Student's t with 5 degrees of freedom, to allow for longer tails and for extreme values.

Again, the main purpose of what follows is to illustrate the application of our statistical technique, not to provide a substantial reappraisal of the house prices. We investigate the robustness of tests based on (5) and on (11), with

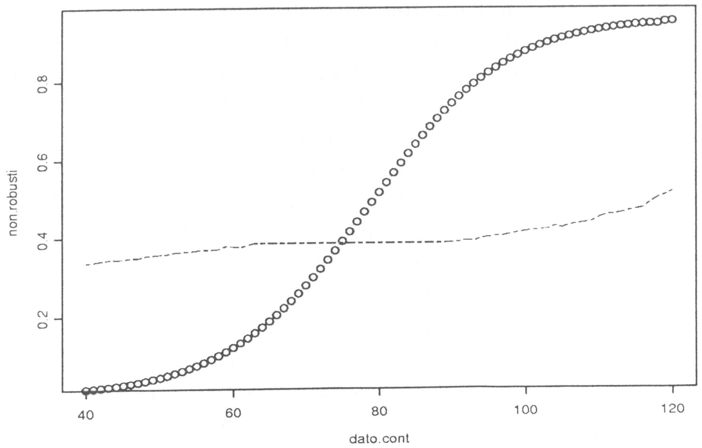

Figure 1: Sensitivity analysis for the Sea Level Data in Venice. The year 1932 observation was contaminated.

respect to single point contaminations (observation 8). For the observation 8 the observed price is 70 thousands of dollars. We vary this observation between 30 to 90, and recompute each time the p-values based on (5) and on (11) for assessing the significance of the scalar parameter β_4 of model (14) associated with the variable of interest x_4. Also for this example, a general inspection of Figure 2 shows varying p-values for (5) and more stable p-values for (11). Therefore, for the standard higher-order approximation (5) a small change of the price on the x-axis implies a significant change of the p-values on the y-axis. On the contrary, for the robust conditional approximation (11) a small change of the price does not imply a substantial change of the p-values.

References

Barndorff-Nielsen, O.E., Cox, D.R. (1994). *Inference and Asymptotics.* Chapman and Hall, London.

Brazzale, A.R. (2000). *Practical Small-Sample Parametric Inference.* Ph.D. Thesis N. 2230, Department of Mathematics, Swiss Federal Institute of Technology Lausanne.

Brazzale, A.R., Ventura, L. (2001). How robust is conditional inference?. *Proceedings of the 16th Meeting of Statistical Modelling*, Odense (Danemark), July 2-6 2001, to appear.

Cox, D.R. (1988). Some aspects of conditional and asymptotic inference. *Sankhya*, **50**, 314–337.

DiCiccio, T.J., Field, C.A., Fraser, D.A.S. (1990). Approximations of marginal tail probabilities and inference for scalar parameters. *Biometrika*, **77**, 77–

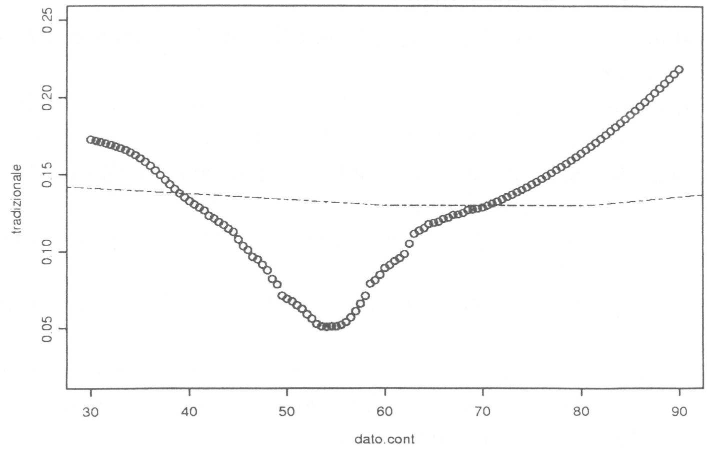

Figure 2: Sensitivity analysis for the House Price Data. The observation 8 was contaminated.

95.

DiCiccio, T.J., Field, C.A. (1991). An accurate method for approximate conditional nad Bayesian inference about linear regression models from censored data. *Biometrika*, **78**, 903–910.

Fisher, R.A. (1934). Two new properties of mathematical likelihood. *Proceedings of the Royal Society, Series A*, **144**, 285–307.

Field, C., Ronchetti, E. (1990). *Small Sample Asymptotics*. IMS lecture notes – monograph series 13, IMS, Hayward.

Fraser, D.A.S. (1979). *Inference and Linear Models*. McGraw Hill, New York.

Hampel, F.R., Marazzi, A., Ronchetti, E., Rousseeuw, P.J., Stahel, W.A., Welsh, R.E. (1982). Handouts for the instructional meeting on robust statistical methods. *15th European Meeting of Statisticians*, Palermo, Italy.

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

Kappenman, R.F. (1975). Conditional confidence intervals for the double exponential distribution parameters. *Technometrics*, **17**, 233–235.

Lawless, J.F. (1972). Conditional confidence interval procedures for the location and scale parameters of the Cauchy and logistic distributions. *Biometrika*, **59**, 377–386.

Lawless, J.F. (1973). Conditional versus unconditional confidence intervals for the parameters of the Weibull distribution. *Journ. Amer. Statist. Assoc.*, **68**, 655–669.

Lawless, J.F. (1978). Confidence interval estimation for the Weibull and extreme value distributions. *Technometrics*, **20**, 355–368.

Pirazzoli, P. (1982). Maree estreme a Venezia (periodo 1872-1981). *Acqua Aria*, **10**, 1023-1039.

Reid, N. (1995). The roles of conditioning in inference. *Statistical Science*, **10**, 138–199.

Ronchetti, E. and Ventura, L. (2001). Between stability and higher-order asymptotics. *Statistics and Computing*, **11**, 427–433.

Sen, A., Srivastava, M. (1990). *Regression Analysis: Theory, Methods and Applications*. Springer-Verlag, New York.

Smith, R.L. (1986). Extreme vaue theory based on the largest r annual events. *J. Hydrol.*, **86**, 27-43.

Tukey, J.W. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, Mass.

Welsh, A.H., Richardson, A.M. (1997). Approaches to the Robust Estimation of Mixed Models. In *Handbook of Statistics*, Vol. 15, Maddala and Rao Eds., Elsevier, North-Holland.