The Effects of Rounding on Likelihood Procedures

L. Pace, A. Salvan, L. Ventura

2002.21

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

Dicembre 2002

The Effects of Rounding on Likelihood Procedures

L. PACE

Department of Statistics, University of Udine, Italy

A. SALVAN and L. VENTURA[1]

Department of Statistics, University of Padova, Italy

SUMMARY. The aim of this paper is to investigate the robustness properties of likelihood inference with respect to rounding effects. Attention is focused on exponential families and on inference about a scalar parameter of interest, also in the presence of nuisance parameters. A summary value of the influence function of a given statistic, the local-shift sensitivity, is considered. It accounts for small fluctuations in the observations. The main result is that the local-shift sensitivity is bounded for the usual likelihood-based statistics, i.e. the directed likelihood, the Wald and score statistics. It is bounded also for the modified directed likelihood, which is a higher-order adjustment of the directed likelihood. The practical implication is that likelihood inference is expected to be robust with respect to rounding effects. Theoretical analysis is supplemented and confirmed by a number of Monte Carlo studies, performed to assess the coverage probabilities of confidence intervals based on likelihood procedures when data are rounded. In addition, simulations

[1] *Address for correspondence*: Laura Ventura, Department of Statistics, University of Padova, Via C. Battisti 241, 35121 Padova, Italy, e-mail: ventura@stat.unipd.it.

indicate that the directed likelihood is less sensitive to rounding effects than the Wald and score statistics. This provides another criterion for choosing among first-order equivalent likelihood procedures. The modified directed likelihood shows the same robustness as the directed likelihood, so that its gain in inferential accuracy does not come at the price of an increase in instability with respect to rounding.

Key words: Directed likelihood; Exponential family; Higher-order asymptotics; Influence function; Maximum likelihood estimator; Modified directed likelihood; Robustness; Score test; Wald test.

1 Introduction

In many practical situations, data are subject to rounding. This operation consists in changing the values of the observations slightly, as happens in rounding to a fixed number of decimal places, or grouping or due to some local inaccuracies, e.g., of the measuring instrument. The question arises of how this reduced precision in the recorded data may affect classical likelihood based procedures for testing and interval estimation.

So far, studies on the effects of the reduced precision of recorded data on statistical inference have concentrated largely on point estimation and inference in Gaussian models. Heitjan (1989) provided a survey of the current state of the art regarding grouped data. Tricker (1984, 1990, 1992, 1995) presents the results of several simulation studies to assess the effects of rounding on some classical statistical techniques. The results indicate that less precision of the recorded data is required than the precision which is usually available. No general results seem however to be available concerning likelihood procedures for testing and interval estimation. Some natural tools for tackling this issue are provided by robust statistics (Hampel et al., 1986).

Although in the last two decades the amount of statistical research devoted to robustness has increased considerably, the issue of rounding the observed data has been somewhat neglected. Most of the reasearch effort has focused on robust estimation and on robust testing procedures; for recent reviews see Markatou, Stahel and Ronchetti (1991), Markatou and Ronchetti (1997), and Ronchetti (1997). A basic tool used to formalize the robustness properties is the *influence function* (IF), which can be used to investigate the local stability of a statistic, such as an estimator or a test statistic. Whereas there exists a large body of literature on the use of the IF to assess the effect of small deviations from the assumed parametric model on the performance of classical estimation and testing procedures (see, e.g., Hampel et al., 1986), the issue of "wiggle around" with the observations has received very little attention. One exception is the paper of Victoria-Feser and Ronchetti (1997), that studies the sensitivity of a class of minimum power divergence estimators for grouped data to grouping effects.

Here we address the effects of rounding on likelihood procedures from a robust statistics point of view. We restrict attention to exponential families when inference on a scalar parameter of interest, also in the presence of nuisance parameters, is desired. We use the *local-shift sensitivity* derived from the IF (see Hampel, 1974; Hampel et al., 1986, chap. 2) as a basic tool to assess the robustness properties of the classical likelihood inference procedures. In particular, we focus on the directed likelihood, the Wald and score statistics. Moreover, we also consider the modified directed likelihood (see e.g. Barndorff-Nielsen and Cox, 1994, chap. 6, and Severini, 2000, chap. 7), which is a higher-order adjustment of the directed likelihood, having standard normal, $N(0,1)$, null distribution with third-order accuracy.

The IFs of the above mentioned statistics are unbounded, and therefore small deviations from the assumed model can have drastic effects on likelihood based procedures. The main theoretical result of this paper is that, on the contrary, their local-shift sensitivities are bounded. Likelihood based procedures are therefore expected to be robust with respect to rounding. Theoretical analysis is supplemented and confirmed by a number of Monte Carlo studies. They indicate that the directed likelihood is less sensitive to rounding effects than the Wald and score statistics. This provides another criterion for choosing among first-order equivalent likelihood procedures. Moreover, the modified directed likelihood shows the same robustness as the directed likelihood, so that the gain in inferential accuracy does not come at the price of an increase in instability with respect to rounding.

The paper is organized as follows. In Section 2 the IF and the local-shift sensitivity are briefly discussed. In Section 3 we calculate the local-shift sensitivity of likelihood procedures in the context of one-parameter exponential families. Multiparameter exponential families are considered in Section 4. Section 5 is devoted to the higher-order modification of the directed likelihood, i.e. the modified directed likelihood. Section 6 illustrates the results of a number of Monte Carlo studies performed to assess the stability of the coverage levels of confidence intervals based on likelihood procedures when data are rounded to a fixed number of decimal places. Some final remarks are given in Section 7.

2 Robustness concepts

There are at least two basic operations which occur in discussing the robustness properties of any statistic. The first one consists in "adding" (or "deducting") a small mass of arbitrary contamination at a point x to assess the effect of outliers, gross errors, bad values or whatever one wants to call them. The second basic operation consists in "wiggling" around with the observations, i.e. to change their values slightly, as happens in rounding or grouping. As noted by Hampel (1974), this second operation can be reduced to the first one, since to shift an observation slightly from the point x to some neighboring point y is the same as to add an observation at y and to remove one at x. Therefore, the effect of wiggling is about the difference of the effects of adding and deducting at two neighboring points. Since the normalized effect of adding a small contamination at some point x is described by the value of the IF, the normalized effect of wiggling somewhere is described by a normalized difference or simply the slope of the IF at that point. This explains the central role of the IF in the study of local robustness problems.

Consider the following basic model. Let $y = (y_1, \ldots, y_n)$ be a random sample of size n such that, for each $i = 1, \ldots, n$, y_i has a continuous distribution $F_\theta = F(\cdot;\theta)$, with $\theta \in \Theta \subseteq \mathrm{I\!R}^p$, $p \geq 1$. Let us denote by $\hat{F}_n = \hat{F}_n(x)$ the empirical distribution function. In the following, we focus on statistics $T_n = T_n(y)$ which can be represented (at least asymptotically) as functionals of the empirical distribution function, i.e. $T_n = T(\hat{F}_n)$.

One way of assessing the robustness properties of a statistic is by means of the IF (see Hampel, 1974; Hampel et al., 1986). Let δ_x be a point mass at x and let F_ϵ denote mixtures of F_θ and δ_x, i.e. $F_\epsilon = (1-\epsilon)F_\theta + \epsilon\delta_x$, with $0 < \epsilon < 1$. The IF of T at F_θ is defined by

$$IF_T(x) = \frac{\partial}{\partial\epsilon} T\left(F_\epsilon\right)\Big|_{\epsilon=0} . \tag{1}$$

Equation (1) describes the effect of a small contamination at the point x on T, standardized by the mass of the contamination. A desiderable robustness property for a statistic is a bounded IF. Accordingly, the most important summary value of the IF is the supremum of its absolute value, i.e. the *gross-error sensitivity* $\gamma^* = \sup_x |IF_T(x)|$. It measures the worst approximate influence which an infinitesimal amount of contamination can have on the value of T.

Having defined a measure for the worst possible effect of "adding" contamination, we also need a measure for the worst effect of "wiggling" the observations. An important summary value of the IF which accounts for small fluctuations in the observations is provided by the *local-shift sensitivity* defined as

$$\lambda^* = \sup_{x \neq y} \frac{|IF_T(x) - IF_T(y)|}{|x-y|} . \tag{2}$$

This measure is particularly relevant when one considers the local effects of rounding or grouping.

When attention is focused on testing hypotheses in a general parametric model, robust statistics introduces a *level influence function* and a *power influence function* that describe the influence af a small amount of contamination at some point x on the asymptotic level and power of the test statistic. However, it turns out that the level influence function and the power influence function are proportional to the IF of the test statistic (see e.g. Hampel et al., 1986, chap. 3, and Ronchetti, 1997). Therefore the robustness properties of a testing procedure depend on the properties of the IF of the test statistic.

3 Results for one-parameter exponential families

Suppose that, possibly after a sufficiency reduction, the reference model is a one-parameter exponential family with density

$$p(t;\theta) = p_0(t)\exp\{\theta t - nK(\theta)\}\ , \tag{3}$$

where $\theta \in \Theta$, with $\Theta \subseteq \mathbb{R}$ the natural parameter space, $p_0(t)$ a function depending on the canonical statistic $t = t(y)$ only, and $K(\theta)$ the cumulant function.

The log-likelihood function based on model (3) is $\ell(\theta) = \ell(\theta; t) = \theta t - nK(\theta)$. We will write $\hat{\theta}$ for the maximum likelihood estimator (MLE) of θ and we assume that it exists with probability one. The MLE is then the unique solution of the likelihood equation $nK'(\hat{\theta}) = t$. In the following, a hat over a likelihood quantity will denote evaluation at $\hat{\theta}$. In view of the likelihood equation, the log-likelihood function for θ can be written as $\ell(\theta) = \ell(\theta;\hat{\theta}) = \theta nK'(\hat{\theta}) - nK(\theta)$. For setting confidence regions or for testing hypotheses about θ, the *directed likelihood*

$$\begin{aligned} r &= r(\theta_0;\hat{\theta}) = \mathrm{sgn}(\hat{\theta}-\theta_0)\sqrt{2\{\ell(\hat{\theta})-\ell(\theta_0)\}} \\ &= \mathrm{sgn}(\hat{\theta}-\theta_0)\sqrt{2n}\left\{K'(\hat{\theta})(\hat{\theta}-\theta_0)+K(\theta_0)-K(\hat{\theta})\right\}^{1/2} \end{aligned} \tag{4}$$

may be used. For instance, confidence regions with nominal coverage $1-\alpha$ for θ can be constructed as $\{\theta : -z_{1-\alpha/2} \le r(\theta) \le z_{1-\alpha/2}\}$, where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the $N(0,1)$ distribution. Two statistics closely related to the directed likelihood are the *Wald statistic*

$$r_e = r_e(\theta_0;\hat{\theta}) = (\hat{\theta}-\theta_0)i(\hat{\theta})^{1/2} = (\hat{\theta}-\theta_0)\left\{nK''(\hat{\theta})\right\}^{1/2} , \tag{5}$$

where $i(\theta) = nK''(\theta)$ is Fisher information, and the *score statistic*

$$r_u = r_u(\theta_0;\hat{\theta}) = \ell'(\theta_0)i(\theta_0)^{-1/2} = \sqrt{n}\left\{K'(\hat{\theta}) - K'(\theta_0)\right\}\{K''(\theta_0)\}^{-1/2} ,$$

where $\ell'(\theta) = (\partial/\partial\theta)\ell(\theta) = n\{K'(\hat{\theta}) - K'(\theta)\}$ is the score function. Both r_e and r_u have the same asymptotic null distribution as r, i.e. they are asymptotically distributed according to a $N(0,1)$ distribution.

Observe that the MLE can be written as a functional of the form $\hat{\theta} = T(\hat{F}_n)$. Moreover, r, r_e and r_u are functionals too, since they depend on t only through the MLE $\hat{\theta}$. Therefore, their IFs can easily be computed using definition (1) and some basic differentiation rules. Let us denote by θ the true parameter value, that is $\theta = T(F_\theta)$. The following expressions for the IF of the statistics r, r_e and r_u can be derived:

$$\begin{aligned} IF_r(x) &= \left.\frac{\partial}{\partial\epsilon} r(\theta_0;T(F_\epsilon))\right|_{\epsilon=0} = C\{x-K'(\theta)\} , \\ IF_{r_e}(x) &= \left.\frac{\partial}{\partial\epsilon} r_e(\theta_0;T(F_\epsilon))\right|_{\epsilon=0} = C_e\{x-K'(\theta)\} , \\ IF_{r_u}(x) &= \left.\frac{\partial}{\partial\epsilon} r_u(\theta_0;T(F_\epsilon))\right|_{\epsilon=0} = C_u\{x-K'(\theta)\} , \end{aligned}$$

where C, C_e and C_u are three terms that do not depend on x, given by

$$\begin{aligned} C &= \sqrt{\frac{n}{2}}|\theta - \theta_0| \left\{K'(\theta)(\theta - \theta_0) + K(\theta_0) - K(\theta)\right\}^{-1/2} , \\ C_e &= \sqrt{n}K''(\theta)^{-1} \left\{K''(\theta)^{1/2} + (\theta - \theta_0)K''(\theta)^{-1/2}K'''(\theta)\right\} , \\ C_u &= \sqrt{n}\{K''(\theta_0)\}^{-1/2} . \end{aligned}$$

When θ approaches θ_0, we see that $C = C_e = C_u = \sqrt{n}\{K''(\theta_0)\}^{-1/2}$. The IFs of the three statistics are all linear functions of x. Therefore, they are unbounded. This means that a small deviation from the assumed model F can have a drastic effect on r, r_e and r_u.

The local-shift sensitivity λ^* for r, r_e and r_u is simply given by

$$C \ , \ |C_e| \text{ and } C_u \ , \tag{6}$$

respectively. Therefore, for the three statistics under consideration λ^* is bounded. Its value, however, can be large: for a given θ, it depends on θ_0 and on n. To illustrate this, Figure 1 gives the plots of the local-shift sensitivities (6) computed for the gamma distribution (with fixed shape) and the inverse Gaussian distribution (with fixed shape). From Figure 1, it may be noted that the local-shift sensitivity associated with the statistic r, C, is less dependent on θ_0 than $|C_e|$ and C_u. Moreover, it can be noted that the Wald statistic r_e appears to be more sensitive with respect to rounding effects than the directed likelihood and the score test. The same behaviour for the the local-shift sensitivity associated to the statistics r, r_e and r_u has been observed also for other one-parameter exponential families not reported here. This provides another criterion for choosing among first-order equivalent likelihood procedures.

(Figure 1 about here)

4 Results for multiparameter exponential families

Consider an exponential family of order p with density

$$p(t, u; \theta) = p_0(t, u) \exp\{\tau t + \zeta \cdot u - nK(\tau, \zeta)\} , \qquad (7)$$

where $\theta = (\tau, \zeta) \in \Theta \subseteq \mathbb{R}^p$, with Θ the natural parameter space. The component $t = t(y)$ of the natural observation $(t, u) = (t(y), u(y))$ as well as the component τ of the natural parameter are assumed to be scalars.

Consider τ as the parameter of interest and ζ as a nuisance parameter. In the following, for partial derivatives of the cumulant function $K(\tau, \zeta)$ we will use the succinct notations $K_\tau = \partial K(\tau, \zeta)/\partial\tau$, $K_\zeta = \partial K(\tau, \zeta)/\partial\zeta$, $K_{\tau\tau} = \partial^2 K(\tau, \zeta)/\partial\tau^2$, $K_{\tau\zeta} = \partial^2 K(\tau, \zeta)/\partial\tau\partial\zeta^{\mathsf{T}}$, $K_{\zeta\zeta} = \partial^2 K(\tau, \zeta)/\partial\zeta\partial\zeta^{\mathsf{T}}$, and so on, where the superscript $^{\mathsf{T}}$ denotes transposition.

Let us denote by $\ell(\tau, \zeta) = \ell(\tau, \zeta; t, u) = \tau t + \zeta \cdot u - nK(\tau, \zeta)$ the log-likelihood based on model (7). We will write $(\hat\tau, \hat\zeta)$ for the unconstrained MLE of (τ, ζ), given as the solution of the likelihood equations

$$t = nK_\tau(\hat\tau, \hat\zeta) , \quad u = nK_\zeta(\hat\tau, \hat\zeta) . \qquad (8)$$

We will write $\hat\zeta_\tau$ for the partial MLE of ζ for a given value of τ, defined as the solution of the likelihood equation $u = nK_\zeta(\hat\tau, \hat\zeta)$ which, in view of (8), can be written as

$$K_\zeta(\tau, \hat\zeta_\tau) = K_\zeta(\hat\tau, \hat\zeta) . \qquad (9)$$

The profile log-likelihood for τ is $\ell_P(\tau) = \ell(\tau, \hat\zeta_\tau) = n\tau\tilde K_\tau + n\hat\zeta_\tau \cdot \hat K_\zeta - n\tilde K$, where a tilde over a likelihood quantity is used when the quantity is evaluated at $(\tau, \hat{\zeta}_\tau)$, while a hat denotes evaluation at $(\hat{\tau}, \hat{\zeta})$.

Under model (7), the ***profile directed likelihood***

$$
\begin{aligned}
r_P &= r_P(\tau_0; \hat{\tau}, \hat{\zeta}) = \text{sgn}(\hat{\tau} - \tau_0)\sqrt{2\{\ell_P(\hat{\tau}) - \ell_P(\tau_0)\}} \\
&= \text{sgn}(\hat{\tau} - \tau_0)\sqrt{2n}\left\{(\hat{\tau} - \tau_0)\hat{K}_\tau + (\hat{\zeta} - \hat{\zeta}_{\tau_0})\hat{K}_\zeta - \hat{K} + \tilde{K}\right\}^{1/2}
\end{aligned}
\tag{10}
$$

may be used as in the scalar case for setting confidence regions or for testing hypotheses on the scalar parameter of interest τ. A statistic closely related to the profile directed likelihood is the ***Wald statistic***

$$
\begin{aligned}
r_{Pe} &= r_{Pe}(\tau_0; \hat{\tau}, \hat{\zeta}) = (\hat{\tau} - \tau_0)\sqrt{j_P(\hat{\tau})} \\
&= (\hat{\tau} - \tau_0)\left\{n\hat{K}_{\tau\tau} - n\hat{K}_{\tau\zeta}\hat{K}_{\zeta\zeta}^{-1}\hat{K}_{\zeta\tau}\right\}^{1/2} ,
\end{aligned}
\tag{11}
$$

where $j_P(\tau) = -\partial^2\ell_P(\tau)/\partial\tau^2 = n\tilde{K}_{\tau\tau} - n\tilde{K}_{\tau\zeta}\tilde{K}_{\zeta\zeta}^{-1}\tilde{K}_{\zeta\tau}$ is the profile information. Another statistic of interest is the ***score statistic***

$$
\begin{aligned}
r_{Pu} &= r_{Pu}(\tau_0; \hat{\tau}, \hat{\zeta}) = \ell_P'(\tau_0)/\sqrt{j_P(\tau_0)} \\
&= n(\hat{K}_\tau - \tilde{K}_\tau)\left\{n\tilde{K}_{\tau\tau} - n\tilde{K}_{\tau\zeta}\tilde{K}_{\zeta\zeta}^{-1}\tilde{K}_{\zeta\tau}\right\}^{-1/2} ,
\end{aligned}
\tag{12}
$$

where $\ell_P'(\tau) = (\partial/\partial\tau)\ell_P(\tau) = n(\hat{K}_\tau - \tilde{K}_\tau)$ is the profile score function. Under regularity conditions, the statistics (11) and (12) both have the same asymptotic $N(0,1)$ null distribution as the profile directed likelihood (10).

As in the scalar case, $\hat{\tau}$ and $\hat{\zeta}$ can be written as a functionals of the empirical distribution function, of the form $\hat{\tau} = T_1(\hat{F}_n)$ and $\hat{\zeta} = T_2(\hat{F}_n)$. Moreover, from equation (9), it may be noted that also $\hat{\zeta}_\tau$ can be written as a functional, since for a fixed τ it is a function of $\hat{\tau}$ and $\hat{\zeta}$. Finally, also r_P, r_{Pe} and r_{Pu} can be written as functionals, since they depend on the observations only through the MLE $(\hat{\tau}, \hat{\zeta})$. Therefore, their IFs can be computed using definition (1) and some basic differentiation rules. Let us denote by (τ, ζ) the true parameter value, i.e. $\tau = T_1(F_\theta)$ and $\zeta = T_2(F_\theta)$, and let us denote by ζ_{τ_0} the limit in probability of $\hat{\zeta}_{\tau_0}$, when (τ, ζ) is the true parameter value. Moreover, let us denote by $IF_{T_1}(x)$ and $IF_{T_2}(x)$ the IFs of the MLEs of τ and ζ, respectively. It can be shown that $IF_{T_1}(x) = I_1 x + I_2$ and that $IF_{T_2}(x) = I_3 x + I_4$, where the scalar terms I_1 and I_2 and the vectors I_3 and I_4 do not depend on x (their expressions are given in the Appendix). Moreover, it can be shown that the IFs of the statistics r_P, r_{Pe} and r_{Pu} are of the form:

$$\begin{aligned}
IF_{r_P}(x) &= \frac{\partial}{\partial \epsilon} r_P(\tau_0; T_1(F_\epsilon), T_2(F_\epsilon))\Big|_{\epsilon=0} = A_1 IF_{T_1}(x) + A_2 \cdot IF_{T_2}(x) \\
IF_{r_{Pe}}(x) &= \frac{\partial}{\partial \epsilon} r_{Pe}(\tau_0; T_1(F_\epsilon), T_2(F_\epsilon))\Big|_{\epsilon=0} = B_1 IF_{T_1}(x) + B_2 \cdot IF_{T_2}(x) \\
IF_{r_{Pu}}(x) &= \frac{\partial}{\partial \epsilon} r_{Pu}(\tau_0; T_1(F_\epsilon), T_2(F_\epsilon))\Big|_{\epsilon=0} = D_1 IF_{T_1}(x) + D_2 \cdot IF_{T_2}(x) ,
\end{aligned}$$

where the scalar terms A_1, B_1, D_1 and the vectors A_2, B_2, and D_2 do not depend on x (their expressions are given in the Appendix). Note that the IFs of the three statistics depend linearly on x, since both $IF_{T_1}(x)$ and $IF_{T_2}(x)$ are linear functions of x. Therefore, the IFs are unbounded for all the three statistics under consideration.

It is easy to see that the local-shift sensitivity λ^* for r_P, r_{Pe} and r_{Pu} is given by

$$|A_1 I_1 + A_2 \cdot I_3| \ , \ |B_1 I_1 + B_2 \cdot I_3| \text{ and } |D_1 I_1 + D_2 \cdot I_3| \ ,$$

respectively. Therefore, as in the scalar case, for the three statistics r_P, r_{Pe} and r_{Pu} the local-shift sensitivity λ^* is bounded. Figure 2 gives the plot of the local-shift sensitivity for r_P, r_{Pe} and r_{Pu} computed for the gamma distribution with shape parameter of interest and with scale nuisance parameter. From Figure 2, it may be noted that the λ^* has a similar behaviour as in the scalar case. Moreover, it can also be noted that the profile directed likelihood statistic r_P appears less sensitive to rounding effects than r_{Pe} and r_{Pu}. The same behaviour for the the local-shift sensitivity associated to the statistics r_P, r_{Pe} and r_{Pu} has been observed also for other multidimensional exponential families not reported here.

(Figure 2 about here)

5 Results for higher-order approximate pivots

In many cases, the accuracy of the usual first-order normal approximation for the distributions of r or r_P is questionable. It is, however, possible to consider modified versions of r or r_P, that have a null $N(0,1)$ distribution to a high order of approximation, while still retaining the essential character of the directed likelihood.

Consider first an exponential family model with scalar parameter θ. A modified version of r for model (3) is given by (see e.g. Pierce and Peters, 1992, or Severini, 2000, Sec. 7.3)

$$r^* = r^*(\theta_0; \hat{\theta}) = r + \frac{1}{r} \log \left(\frac{r_e}{r} \right) , \tag{13}$$

where r and r_e are defined by (4) and (5), respectively. Consider now a model parameterised by a scalar parameter of interest τ together with a nuisance parameter ζ (as in Section 4). The modified directed likelihood based on model (7) is then (see e.g. Pierce and Peters, 1992, or Severini, 2000, Sec. 7.4)

$$r_P^* = r_P^*(\tau_0; \hat{\tau}, \hat{\zeta}) = r_P + \frac{1}{r_P} \log \left(\frac{u_P}{r_P} \right) , \tag{14}$$

where $u_P(\tau_0;\hat{\tau},\hat{\zeta}) = v(\tau_0;\hat{\tau},\hat{\zeta}) r_{Pe}(\tau_0;\hat{\tau},\hat{\zeta})$, with $v(\tau_0;\hat{\tau},\hat{\zeta}) = |\hat{\imath}_{\zeta\zeta}|^{1/2}/|\tilde{\imath}_{\zeta\zeta}|^{1/2}$. The modified directed likelihoods (13) and (14) are higher-order pivotal quantities, having a null $N(0,1)$ distribution with error of order $O(n^{-3/2})$.

The statistics r^* and r_P^* can be written as functionals, since they depend on the observations only through the MLEs. They assume simple expressions in the context of exponential families, so that their IF can easily be computed. For a one-parameter model, we obtain that the IF of r^* is

$$IF_{r^*}(x) = \frac{\partial}{\partial \epsilon} r^*(\theta_0; T(F_\epsilon)) \Big|_{\epsilon=0} = C^* \{x - K'(\theta)\} \, , \tag{15}$$

where C^* does not depend on x, and is given by

$$C^* = C \left\{ 1 - \frac{1}{r(\theta_0;\theta)^2} \log \frac{r_e(\theta_0;\theta)}{r(\theta_0;\theta)} - \frac{1}{r(\theta_0;\theta)^2} \right\} + \frac{C_u}{r_e(\theta_0;\theta) r(\theta_0;\theta)} \, .$$

Note that the IF of r^* has a form similar to the IFs of the statistics r, r_e and r_u, since it depends linearly on x. It is important to note that the IF (15) is unbounded. This finding is in agreement with Ronchetti and Ventura (2001), who show that small deviations from the assumed model can wipe out the improvements of the accuracy obtained by second-order approximations to the distribution of classical statistics.

The local-shift sensitivity for r^* in this case is simply given by $|C^*|$. Therefore λ^* is bounded. Figure 3 gives the plots of the local-shift sensitivity of r and r^* for the same scalar exponential families as those considered in Figure 1. From Figure 3, it may be noted that the local-shift sensitivities associated with r^* and with r are almost equal. This means that r^* improves on r maintaining its robustness with respect to rounding. The same behaviour for the the local-shift sensitivity associated to the statistics r and r^*

has been observed also for other one-parameter exponential families not reported here.

(Figure 3 about here)

Consider now the case in which there is a scalar parameter of interest τ and a nuisance parameter ζ. Then, the IF of r_P^* is of the form

$$\begin{aligned} IF_{r_P^*}(x) &= \left.\frac{\partial}{\partial \epsilon} r_P^*(\tau_0; T_1(F_\epsilon), T_2(F_\epsilon))\right|_{\epsilon=0} \\ &= \{A_1 IF_{T_1}(x) + A_2 \cdot IF_{T_2}(x)\} \left\{ 1 - \frac{1}{r_P(\tau_0;\tau,\zeta)^2} \log \frac{u_P(\tau_0;\tau,\zeta)}{r_P(\tau_0;\tau,\zeta)} \right. \\ &\quad \left. - \frac{1}{r_P(\tau_0;\tau,\zeta)^2} \right\} + \frac{E_1 IF_{T_1}(x) + E_2 \cdot IF_{T_2}(x)}{v(\tau_0;\tau,\zeta) r_P(\tau_0;\tau,\zeta)} \\ &\quad + \frac{B_1 IF_{T_1}(x) + B_2 \cdot IF_{T_2}(x)}{r_P(\tau_0;\tau,\zeta) r_{Pe}(\tau_0;\tau,\zeta)}, \end{aligned} \tag{16}$$

where the scalar term E_1 and the vector E_2 do not depend on x (see the Appendix for their expression). Note that the IF of r_P^* has a form similar to the IFs of r_P, r_{Pe} and r_{Pu}. In fact, (16) depends linearly on x, since both $IF_{T_1}(x)$ and $IF_{T_2}(x)$ are linear functions of x. It is important to note that the IF (16) is unbounded, while, as in the scalar case, it is easy to show that the local-shift sensitivity λ^* for r_P^* is bounded. For several examples not reported here, our experience suggests that the local-shift sensitivities associated with r_P^* and with r_P are almost indistinguishable, paralleling what has been seen in Section 4 (see Figure 3). Therefore, r_P^* improves on the first order pivot r_P maintaining its robustness with respect to rounding.

6 Monte Carlo evidence

Some Monte Carlo studies have been performed to assess the stability of coverage levels of confidence intervals based on likelihood procedures with respect to rounding effects. Each simulation study is based on 100000 Monte Carlo trials. We assume that the recorded data z is

$$z = \varepsilon < y/\varepsilon > ,$$

where $< x >$ is the nearest integer to x and ε is the length of the rounding interval. Then, if $\varepsilon = 1$ the observation is discretised, while if $\varepsilon = 10^{-d}$, then z is y rounded to d decimal places. In the simulation studies, we consider the following situations: T when $z = y$, i.e. we use the true data y; R when $d = 1$, i.e. when y is rounded to the first decimal place; I when $d = 0$, i.e. y is rounded to the nearest integer. In our study, the true data T have six decimal places, i.e. $d = 6$. Values $1 < d < 6$ are not reported, since preliminary simulations have shown the same results as under T.

Other simulation experiments have been carried out using $\varepsilon = 2^{-d}$ (with $d = 1, 2, 3$), in order to represent data which are recorded with a precision which is intermediate between R and I. The results were found to be very similar to the ones discussed in this section and are not reported here. In particular, 2^{-3} produces the same results as T, 2^{-2} produces the same results as R, and 2^{-1} produces results quite similar to I.

Subsection 6.1 illustrates examples both with one-parameter models and with models having a scalar parameter of interest τ and a nuisance parameter ζ, for small or moderate sample sizes. The main conclusion is that, for such values of n, the reduced precision of the recorded data scarcely affects likelihood based procedures, particularly r (and r_P) and r^* (and r_P^*). This finding indicates that we can use less precision in the recorded data than has been realised and still apply standard likelihood procedures, as pointed out also by Tricker (1990, 1992, 1995) for particular inferential procedures. However, the effects of rounding on likelihood procedures depend on the number of classes after grouping. In particular, the number of classes decreases as the ratio between the lenght of the rounding interval and the standard deviation of the population, i.e.

$$\rho = \frac{\varepsilon}{\sigma} ,$$

increases. The quantity ρ is a natural measure of the severity of rounding (see e.g. Tricker, 1984, 1990, 1992). For small or moderate sample sizes, the simulation results described in Subsection 6.1 below indicate that when the number of classes after grouping is very small, such as 2 or 3, and accordingly ρ is greater than 1.5, results are unsatisfactory. On the other hand, when the number of classes is greater than 7, and accordingly ρ is smaller than 0.5, the results based on I are quite similar to those based on R.

We have to recall, however, that when continuous data are grouped inference should be based on the multinomial distribution with cell probabilities expressed in terms of F_θ. If not, MLEs are no more consistent and likelihood based procedures lose their usual first-order asymptotic behaviour. Hence, the effects of rounding on likelihood procedures depend on the sample size. This point is illustrated in Subsection 6.2.

6.1 Small or moderate sample sizes

Tables 1–3 report the simulation results for various one-parameter exponential families. In particular, random samples $y = (y_1, \ldots, y_n)$ have been generated with: y_i having an exponential distribution with mean $1/\lambda$, for $\lambda = 1$ and $\lambda = 4$; $\exp(y_i)$ having a gamma distribution with scale parameter $\lambda = 1$ and shape parameter $k = 0.2$ and $k = 1.5$ (log-gamma distribution for y_i); y_i having a Gaussian distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$ and $\sigma^2 = 16$. For each distribution, coverage probabilities of confidence intervals based on r, r_e, r_u, and r^* have been estimated.

From Tables 1–3 the following general comments emerge. In all the situations considered r is preferable to r_e and r_u, and r^* improves on r. The scenario R gives always results very similar to those given by T. This is not surprising since λ^* is bounded for all the statistics considered. Moreover, the difference between results with data T and I is smaller for r and r^* than for r_e and r_u. Finally, as expected, when the number of classes after grouping is very restricted (e.g. exponential distribution with $\lambda = 4$), and accordingly ρ is about 2 or larger, operation I gives rise to unsatisfactory results for all the test statistics considered.

(Table 1 about here)

(Table 2 about here)

(Table 3 about here)

Tables 4–5 give the simulation results when the reference model is a multiparameter exponential family and inference procedures about a scalar component of the canonical parameter are considered. In particular, two multiparameter exponential families have been considered: independent observations $y_i, i = 1, \ldots, n$, having a log-gamma distribution, and independent observations $y_i, i = 1, \ldots, n$, with $\exp(y_i)$ having an inverse Gaussian distribution (log-inverse Gaussian distribution for y_i).

The results for data T and R are very similar both for r_P and for r_P^*. This is not true for the two other test statistics. Moreover, the difference between results for data T and I is always smaller for r_P and r_P^* than for r_{Pe} and r_{Pu}. When the number of classes after grouping is very restricted (e.g. inverse Gaussian with $\lambda = 1$), and accordingly ρ is about 2 or larger, operation I gives rise to unsatisfactory results for all the test statistics considered. In this case it would be necessary to use the multinomial distribution.

(Table 4 about here)

(Table 5 about here)

It is important noticing that the results discussed here, and other simulations experiments not reported , indicate that the effects of rounding on likelihood procedures depend both on the severity of rounding ρ and on the skewness of the distribution. Generally as ρ decreases, so does the effect of rounding. On the other hand, for a given value of ρ, the effect of rounding increases as the skewness of the distribution increases. As an example, for the log-gamma distribution, asymmetry increases as k approaches zero, and thus the rounding process causes more effects on likelihood procedures as the distribution becomes more asymmetrical. The same result holds also for the exponential and the log-inverse gaussian distribution. This was not unexpected since Tricker (1984, 1992) showed that the effects on rounding on some classical inferential procedures decrease as the distribution becomes less skewed.

6.2 Large sample sizes

In order to assess how the effects of discretising depend on the sample size, Table 6 gives the results of simulation studies using the same distributions as in Section 6.1 with larger values of n when data are discretized. The results allow us to see how fast coverage levels of confidence intervals based on likelihood procedures decrease as n increases.

It is important to note that this behaviour was also anticipated by the local-shift sensitivity derived from the IF. In fact, the expressions of λ^* obtained in Sections 3–5 depend on the sample size. In particular, in all the situations considered, the local-shift sensitivity increases as n increases and therefore likelihood based procedures become more sensitive to discretising effects when the sample size is large.

(Table 6 about here)

7 Final Remarks

Likelihood procedures are robust to rounding, as has been shown both by analytic tools and by Monte Carlo studies. The main result is that, for the usual likelihood-based statistics, the local-shift sensitivity is bounded, while the IF is unbounded. This means that these statistics are not robust with respect to small model deviations, outliers and influent observations, but are robust with respect to rounding to a fixed number of decimal places. In view of this, especially when n is small or moderate, we can use far less precision in the recorded data than has been realised and still apply the likelihood based analyses holding for continuous data. Moreover, inference based on the directed likelihood and on its higher-order modification appears more robust than inference based on the Wald or on the score statistic.

Although this paper focuses on exponential families, which allow a simple computation of the local-shift sensitivity, the results obtained are expected to maintain outside this class of models. Some simulation results in the context of scale and regression models are in accordance with our conjecture.

Appendix

It can be shown that (see Hampel et al., 1986, chap. 4) the IF of $(\hat{\tau}, \hat{\zeta})$ is $(IF_{T_1}(x), IF_{T_2}(x)) = i(\tau,\zeta)^{-1}(x - K_\tau, x - K_\zeta)$, where $i(\tau,\zeta)$ is the Fisher information with blocks $i_{\tau\tau} = K_{\tau\tau}$, $i_{\tau\zeta} = K_{\tau\zeta}$, $i_{\zeta\zeta} = K_{\zeta\zeta}$ and $i_{\zeta\tau} = i_{\tau\zeta}^{\mathsf{T}}$. In the following, to give the expressions of the terms involved in the IFs discussed in Section 4 and 5, it is convenient to use index notation and the Einstein summation convention (Barndorff-Nielsen and Cox, 1989, Sec. 5.3). The components of the nuisance parameter ζ are denoted by ζ^a, the corresponding components of K_ζ are K_a and the derivatives of K_τ and K_a with respect to τ or to the components of ζ are denoted by

$$K_{ab} = \frac{\partial K_a}{\partial \zeta^b}, \quad K_{\tau a} = \frac{\partial K_\tau}{\partial \zeta^a}, \quad K_{\tau\tau} = \frac{\partial K_\tau}{\partial \tau},$$
$$K_{\tau\tau\tau} = \frac{\partial^2 K_\tau}{\partial \tau^2}, \quad K_{\tau\tau a} = \frac{\partial^2 K_\tau}{\partial \tau \partial \zeta^a}, \quad K_{\tau ab} = \frac{\partial^2 K_\tau}{\partial \zeta^a \partial \zeta^b},$$

and so on, where the indices $a, b, \ldots$ range over $1, \ldots, p-1$. Then we write $i_{\tau\tau} = K_{\tau\tau}$, $i_{\tau a} = K_{\tau a} = i_{a\tau}$ and $i_{ab} = K_{ab}$. In addition, $i^{\tau\tau}$, $i^{\tau a}$, i^{ab} and $i^{a\tau} = i^{\tau a}$ denote the components of the blocks of $i(\tau,\zeta)^{-1}$. The symbol $\vee$ over a likelihood quantity is used when the quantity is evaluated at $(\tau, \zeta_{\tau 0})$. Finally, $IF^a_{T_2}(x)$ denote the components of $IF_{T_2}(x)$.

Using index notation, the terms I_i, $i = 1,2,3,4$, involved in the IF of the MLEs of τ

and ζ can be written as

$$I_1 = i^{\tau\tau} + \sum_{a=1}^{p-1} i^{\tau a}, \quad I_2 = -i^{\tau\tau} K_\tau - i^{\tau a} K_a ,$$
$$I_3^a = i^{\tau a} + \sum_{b=1}^{p-1} i^{ba}, \quad I_4^a = -i^{\tau a} K_\tau - i^{ab} K_b .$$

The terms involved in the IF of the statistic r_P are

$$A_1 = A_3 \left\{ (\tau - \tau_0) K_{\tau\tau} + (\zeta^a - \zeta^a_{\tau_0}) K_{a\tau} - \check{K}^{ab} K_{b\tau} K_a \right\} ,$$
$$A_2^a = A_3 \left\{ (\tau - \tau_0) K_{\tau a} + (\zeta^b - \zeta^b_{\tau_0}) K_{ab} - \check{K}^{bc} K_{ca} K_b \right\} ,$$

where $A_3 = \sqrt{n/2}\mathrm{sgn}(\tau - \tau_0) \left\{ (\tau - \tau_0) K_\tau + (\zeta^a - \zeta^a_{\tau_0}) K_a + \check{K} - K \right\}^{-1/2}$. The terms involved in the IF of the statistic r_{Pe} are

$$B_1 = (i^{\tau\tau})^{-1/2} + B_3 \left\{ K_{\tau\tau\tau} - 2K_{\tau\tau a} K^{ab} K_{b\tau} + K_{\tau a} K^{ab} K^{dc} K_{c\tau} K_{bd\tau} \right\} ,$$
$$B_2^a = B_3 \left\{ K_{\tau\tau a} - 2K_{\tau ba} K^{bc} K_{c\tau} + K_{\tau e} K^{ec} K^{db} K_{b\tau} K_{cda} \right\} ,$$

where $B_3 = (n/2)(\tau - \tau_0)(i^{\tau\tau})^{1/2}$. The terms involved in the IF of the statistic r_{Pu} are

$$\begin{aligned} D_1 &= D_3 \left\{ K_{\tau\tau} - \check{K}_{\tau a} \check{K}^{ab} K_{b\tau} \right\} - D_4 \left\{ \check{K}_{\tau\tau a} \check{K}^{ab} K_{b\tau} \right. \\ &- \left. 2\check{K}_{\tau ac} \check{K}^{ab} \check{K}_{b\tau} \check{K}^{cd} K_{d\tau} + \check{K}_{\tau a} \check{K}^{ac} \check{K}^{db} \check{K}_{cde} \check{K}_{b\tau} \check{K}^{ef} K_{f\tau} \right\} , \\ D_2^a &= D_3 \left\{ K_{\tau a} - \check{K}_{\tau c} \check{K}^{cb} K_{ba} \right\} - D_4 \left\{ \check{K}_{\tau\tau c} \check{K}^{cb} K_{ba} \right. \\ &- \left. 2\check{K}_{\tau ec} \check{K}^{eb} \check{K}_{b\tau} \check{K}^{cd} K_{da} + \check{K}_{\tau g} \check{K}^{gc} \check{K}^{db} \check{K}_{cde} \check{K}_{b\tau} \check{K}^{ef} K_{fa} \right\} . \end{aligned}$$

where

$$D_3 = \sqrt{n} \left\{ \check{K}_{\tau\tau} - \check{K}_{\tau a} \check{K}^{ab} \check{K}_{b\tau} \right\}^{-1/2} ,$$
$$D_4 = \frac{\sqrt{n}}{2} (K_\tau - \check{K}_\tau) \left\{ \check{K}_{\tau\tau} - \check{K}_{\tau a} \check{K}^{ab} \check{K}_{b\tau} \right\}^{-3/2} .$$

Finally, the two terms involved in the IF of the statistic r_P^* are

$$E_1 = \frac{1}{2}\left\{K^{ab}K_{ab\tau} - \breve{K}^{ab}\breve{K}_{abc}\breve{K}^{cd}K_{d\tau}\right\} ,$$
$$E_2^a = \frac{1}{2}\left\{K^{cb}K_{cba} - \breve{K}^{eb}\breve{K}_{ebc}\breve{K}^{cd}K_{da}\right\} .$$

References

BARNDORFF-NIELSEN, O.E. & COX, D.R. (1989) *Asymptotic Techniques for Use in Statistics*, Chapman and Hall, London.

BARNDORFF-NIELSEN, O.E. & COX, D.R. (1994) *Inference and Asymptotics*, Chapman and Hall, London.

HAMPEL, F.R. (1974) The influence curve and its role in robust estimation, **Journal of the American Statistical Association**, 69, pp. 383–393.

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

HEITJAN, A.F. (1989) Inference from grouped continuous data: a review, **Statistical Science**, 4, pp. 164–183

MARKATOU, M., STAHEL, W. & RONCHETTI, E. (1991) Robust M-type testing procedures for linear models, in W. Stahel and S. Weisberg (Ed.) **Directions in Robust Statistics and Diagnostics (Part I)**, IMA Volumes in Mathematics and its Applications No. 33 (Springer, New York), pp. 201–220.

MARKATOU, M. & RONCHETTI, E. (1997) Robust inference: The approach based on influence functions, in G.S. Maddala and C.R. Rao (Ed.), **Handbook of Statistics**, Vol. 15 (North Holland), pp. 49–75.

PIERCE, D.A. & PETERS, D. (1992) Practical use of higher order asymptotics for multiparameter exponential families (with discussion), **Journal of the Royal Statistical Society Series B**, 54, pp. 701–737.

RONCHETTI, E. (1997) Robust inference by influence functions, **Journal of Statistical Planning and Inference**, 57, pp. 59–72.

RONCHETTI, E. & VENTURA, L. (2001) Between stability and higher-order asymptotics, **Statistics and Computing**, 11, pp. 427–433.

SEVERINI, T.A. (2000) *Likelihood Methods in Statistics*, Oxford University Press.

TRICKER, A.R. (1984) Effects of rounding on the moments of a probability distribution, **The Statistician**, 33, pp. 381–390.

TRICKER, A.R. (1990) The effect of rounding on the significance level of certain normal test statistics, **Journal of Applied Statistics**, 17, pp. 31–38.

TRICKER, A.R. (1992) Estimation of parameters for rounded data from non-normal distributions, **Journal of Applied Statistics**, 19, pp. 465–471.

TRICKER, A.R. (1995) Effect on Bayesian inference of the degree of precision of recorded data, **Journal of Applied Statistics**, 22, pp. 235–244.

VICTORIA-FESER, M.P. & RONCHETTI, E. (1997) Robust estimation for grouped data, **Journal of the American Statistical Association**, 92, pp. 333–340.

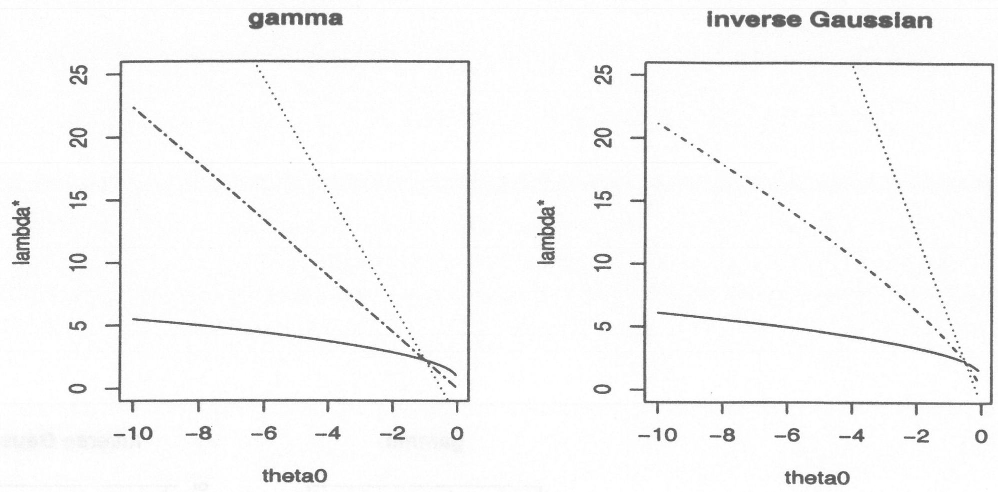

Figure 1: Local-shift sensitivity λ^* for r (solid line), r_e (dotted line) and r_u (dashed line) for two well-known one-parameter exponential families for $n = 10$. The three curves intersect at $\theta_0 = \theta$.

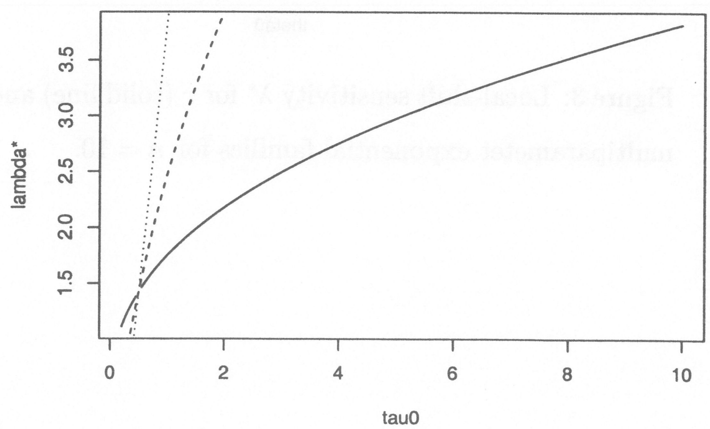

Figure 2: Local-shift sensitivity for r_P (solid line), r_{Pe} (dotted line) and r_{Pu} (dashed line) for the gamma distribution, with shape parameter of interest, scale parameter unknown, and $n = 10$. The three curves intersect at $\theta_0 = \theta$.

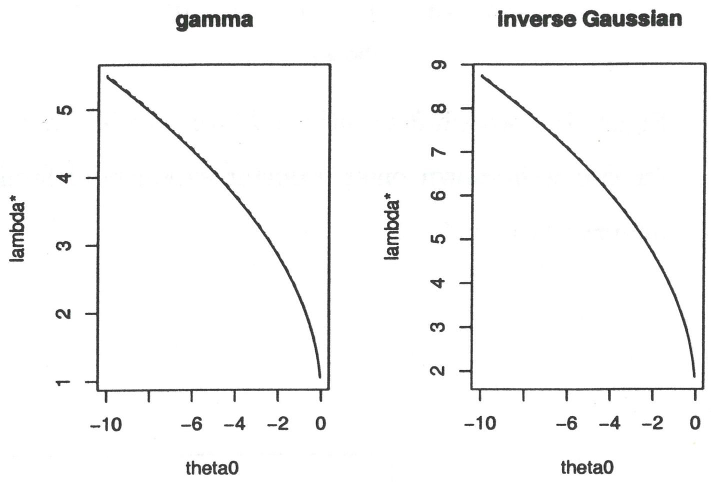

Figure 3: Local-shift sensitivity λ^* for r (solid line) and r^* (dotted line) for two well-known multiparameter exponential families for $n = 10$.

			$k = 0.2$			$k = 1.5$		
n	data	test	0.900	0.950	0.990	0.900	0.950	0.990
5	T	r	0.891	0.945	0.988	0.898	0.948	0.990
	R	r	0.891	0.945	0.988	0.898	0.948	0.991
	I	r	0.890	0.943	0.989	0.888	0.925	0.980
	T	r_e	0.905	0.947	0.982	0.894	0.944	0.984
	R	r_e	0.905	0.948	0.982	0.894	0.943	0.985
	I	r_e	0.907	0.945	0.980	0.880	0.911	0.978
	T	r_u	0.911	0.952	0.982	0.898	0.950	0.986
	R	r_u	0.911	0.952	0.982	0.899	0.948	0.986
	I	r_u	0.914	0.950	0.984	0.881	0.916	0.979
	T	r^*	0.897	0.949	0.989	0.898	0.949	0.989
	R	r^*	0.897	0.949	0.989	0.898	0.949	0.989
	I	r^*	0.896	0.947	0.990	0.888	0.928	0.982
10	T	r	0.894	0.948	0.990	0.901	0.951	0.990
	R	r	0.894	0.948	0.990	0.903	0.951	0.991
	I	r	0.892	0.949	0.990	0.891	0.938	0.987
	T	r_e	0.903	0.949	0.985	0.898	0.949	0.987
	R	r_e	0.904	0.950	0.985	0.898	0.947	0.986
	I	r_e	0.900	0.940	0.982	0.883	0.932	0.976
	T	r_u	0.906	0.951	0.985	0.901	0.951	0.988
	R	r_u	0.905	0.951	0.986	0.899	0.951	0.989
	I	r_u	0.901	0.945	0.982	0.883	0.934	0.982
	T	r^*	0.899	0.949	0.990	0.900	0.950	0.990
	R	r^*	0.899	0.949	0.990	0.901	0.950	0.990
	I	r^*	0.895	0.950	0.990	0.893	0.940	0.987
20	T	r	0.901	0.948	0.990	0.900	0.950	0.990
	R	r	0.901	0.948	0.990	0.900	0.950	0.990
	I	r	0.899	0.950	0.990	0.883	0.939	0.985
	T	r_e	0.905	0.950	0.989	0.899	0.949	0.988
	R	r_e	0.905	0.950	0.989	0.897	0.949	0.988
	I	r_e	0.900	0.945	0.987	0.878	0.937	0.983
	T	r_u	0.906	0.951	0.989	0.901	0.949	0.989
	R	r_u	0.905	0.951	0.989	0.900	0.949	0.989
	I	r_u	0.901	0.945	0.987	0.881	0.937	0.983
	T	r^*	0.900	0.950	0.990	0.900	0.950	0.990
	R	r^*	0.900	0.950	0.990	0.900	0.950	0.990
	I	r^*	0.899	0.951	0.990	0.885	0.940	0.985

Table 2: Coverage levels of confidence intervals for the shape parameter k of the (log-) gamma distribution (scale parameter known and fixed to 1). In situation I, the average number of classes after grouping is about 10 when $k = 0.2$ ($\rho = 0.2$) and about 4 when $k = 1.5$ ($\rho = 1.1$).

			$\sigma = 1$			$\sigma = 4$		
n	data	test	0.900	0.950	0.990	0.900	0.950	0.990
5	T	r	0.888	0.944	0.987	0.887	0.943	0.987
	R	r	0.889	0.942	0.987	0.887	0.942	0.988
	I	r	0.905	0.926	0.994	0.887	0.944	0.986
	T	r_e	0.826	0.862	0.911	0.825	0.864	0.913
	R	r_e	0.828	0.862	0.911	0.825	0.863	0.913
	I	r_e	0.811	0.946	0.946	0.829	0.860	0.910
	T	r_u	0.932	0.954	0.978	0.931	0.952	0.976
	R	r_u	0.931	0.954	0.978	0.931	0.952	0.976
	I	r_u	0.919	0.916	0.969	0.929	0.950	0.971
	T	r^*	0.889	0.947	0.989	0.886	0.942	0.988
	R	r^*	0.889	0.947	0.989	0.886	0.942	0.988
	I	r^*	0.903	0.925	0.994	0.885	0.941	0.987
10	T	r	0.894	0.945	0.988	0.892	0.945	0.986
	R	r	0.895	0.944	0.988	0.891	0.945	0.988
	I	r	0.884	0.945	0.987	0.890	0.945	0.988
	T	r_e	0.853	0.900	0.945	0.862	0.903	0.947
	R	r_e	0.855	0.900	0.945	0.861	0.904	0.947
	I	r_e	0.865	0.936	0.937	0.864	0.901	0.945
	T	r_u	0.920	0.957	0.981	0.919	0.953	0.981
	R	r_u	0.921	0.957	0.981	0.920	0.953	0.981
	I	r_u	0.899	0.927	0.971	0.918	0.950	0.980
	T	r^*	0.895	0.948	0.989	0.894	0.946	0.989
	R	r^*	0.895	0.948	0.989	0.894	0.946	0.989
	I	r^*	0.885	0.947	0.989	0.892	0.945	0.989
20	T	r	0.896	0.947	0.988	0.893	0.946	0.988
	R	r	0.896	0.947	0.988	0.894	0.946	0.988
	I	r	0.887	0.936	0.989	0.894	0.946	0.988
	T	r_e	0.880	0.926	0.965	0.881	0.922	0.964
	R	r_e	0.882	0.925	0.965	0.881	0.922	0.964
	I	r_e	0.904	0.949	0.972	0.882	0.924	0.964
	T	r_u	0.909	0.954	0.985	0.907	0.954	0.987
	R	r_u	0.909	0.954	0.986	0.908	0.954	0.987
	I	r_u	0.877	0.927	0.971	0.907	0.953	0.986
	T	r^*	0.899	0.950	0.990	0.900	0.950	0.990
	R	r^*	0.899	0.950	0.989	0.900	0.950	0.990
	I	r^*	0.889	0.940	0.989	0.901	0.951	0.989

Table 3: Coverage levels of confidence intervals for the scale parameter of the Gaussian distribution (μ known and fixed to 0). In situation I, the average number of classes after grouping is about 4 when $\sigma = 1$ ($\rho = 1$) and about 10 when $\sigma = 4$ ($\rho = 0.25$).

			$k=0.5$			$k=5$		
n	data	test	0.900	0.950	0.990	0.900	0.950	0.990
	T	r_P	0.825	0.896	0.968	0.833	0.904	0.969
	R	r_P	0.824	0.896	0.968	0.832	0.903	0.969
	I	r_P	0.818	0.888	0.972	0.837	0.908	0.971
	T	r_{Pe}	0.958	0.978	0.987	0.965	0.976	0.994
	R	r_{Pe}	0.957	0.972	0.986	0.966	0.979	0.994
5	I	r_{Pe}	0.945	0.962	0.980	0.953	0.970	0.986
	T	r_{Pu}	0.958	0.971	0.986	0.965	0.976	0.994
	R	r_{Pu}	0.957	0.972	0.986	0.966	0.979	0.994
	I	r_{Pu}	0.945	0.967	0.980	0.953	0.970	0.986
	T	r_P^*	0.896	0.947	0.989	0.893	0.940	0.990
	R	r_P^*	0.896	0.947	0.989	0.892	0.939	0.989
	I	r_P^*	0.888	0.938	0.996	0.887	0.935	0.987
	T	r_P	0.867	0.927	0.984	0.868	0.925	0.976
	R	r_P	0.867	0.928	0.983	0.870	0.925	0.976
	I	r_P	0.874	0.938	0.988	0.856	0.920	0.983
	T	r_{Pe}	0.932	0.970	0.989	0.932	0.978	0.995
	R	r_{Pe}	0.932	0.969	0.989	0.932	0.973	0.995
10	I	r_{Pe}	0.923	0.960	0.983	0.920	0.959	0.985
	T	r_{Pu}	0.931	0.969	0.988	0.932	0.978	0.995
	R	r_{Pu}	0.931	0.969	0.988	0.932	0.973	0.995
	I	r_{Pu}	0.920	0.959	0.983	0.920	0.959	0.985
	T	r_P^*	0.897	0.948	0.990	0.900	0.941	0.990
	R	r_P^*	0.898	0.948	0.989	0.900	0.941	0.990
	I	r_P^*	0.895	0.940	0.988	0.891	0.937	0.987
	T	r_P	0.882	0.935	0.986	0.885	0.949	0.991
	R	r_P	0.884	0.934	0.986	0.882	0.946	0.991
	I	r_P	0.882	0.940	0.986	0.868	0.952	0.991
	T	r_{Pe}	0.908	0.959	0.988	0.916	0.967	0.991
	R	r_{Pe}	0.908	0.959	0.988	0.919	0.967	0.991
20	I	r_{Pe}	0.896	0.941	0.979	0.934	0.950	0.989
	T	r_{Pu}	0.907	0.958	0.987	0.916	0.967	0.991
	R	r_{Pu}	0.907	0.958	0.988	0.919	0.967	0.991
	I	r_{Pu}	0.924	0.943	0.979	0.934	0.950	0.989
	T	r_P^*	0.900	0.950	0.990	0.900	0.950	0.990
	R	r_P^*	0.900	0.950	0.990	0.900	0.950	0.990
	I	r_P^*	0.897	0.945	0.988	0.907	0.953	0.988

Table 4: Coverage levels of confidence intervals for the shape parameter k of the log-gamma distribution. The scale parameter is unknown (fixed to 1 in the simulations). In situation I, the average number of classes after grouping is about 8 when $k = 0.5$ ($\rho = 0.5$) and about 3 when $k = 5$ ($\rho = 2$).

n	data	test	$\lambda = 1$ 0.900	0.950	0.990	$\lambda = 10$ 0.900	0.950	0.990
	T	r_P	0.829	0.893	0.971	0.825	0.894	0.965
	R	r_P	0.832	0.890	0.971	0.828	0.892	0.967
	I	r_P	0.877	0.902	0.984	0.875	0.925	0.990
	T	r_{Pe}	0.961	0.979	0.989	0.958	0.974	0.988
	R	r_{Pe}	0.955	0.968	0.985	0.955	0.974	0.988
5	I	r_{Pe}	0.903	0.944	0.974	0.879	0.879	0.951
	T	r_{Pu}	0.961	0.979	0.989	0.958	0.974	0.988
	R	r_{Pu}	0.955	0.968	0.986	0.955	0.974	0.988
	I	r_{Pu}	0.904	0.944	0.974	0.879	0.879	0.951
	T	r_P^*	0.898	0.949	0.988	0.898	0.949	0.989
	R	r_P^*	0.896	0.946	0.988	0.896	0.948	0.988
	I	r_P^*	0.877	0.956	0.994	0.923	0.962	0.995
	T	r_P	0.860	0.920	0.980	0.869	0.929	0.985
	R	r_P	0.860	0.921	0.980	0.870	0.929	0.985
	I	r_P	0.852	0.918	0.984	0.850	0.906	0.980
	T	r_{Pe}	0.931	0.973	0.989	0.928	0.970	0.988
	R	r_{Pe}	0.930	0.972	0.989	0.929	0.967	0.988
10	I	r_{Pe}	0.906	0.943	0.970	0.853	0.863	0.938
	T	r_{Pu}	0.931	0.973	0.989	0.928	0.970	0.988
	R	r_{Pu}	0.930	0.972	0.989	0.929	0.967	0.988
	I	r_{Pu}	0.906	0.943	0.970	0.853	0.864	0.939
	T	r_P^*	0.901	0.950	0.991	0.898	0.951	0.990
	R	r_P^*	0.899	0.949	0.991	0.899	0.950	0.990
	I	r_P^*	0.880	0.948	0.993	0.878	0.932	0.984
	T	r_P	0.873	0.931	0.985	0.882	0.939	0.984
	R	r_P	0.870	0.931	0.985	0.883	0.939	0.985
	I	r_P	0.859	0.915	0.979	0.781	0.844	0.957
	T	r_{Pe}	0.906	0.956	0.988	0.914	0.958	0.989
	R	r_{Pe}	0.906	0.955	0.988	0.913	0.959	0.989
20	I	r_{Pe}	0.860	0.912	0.959	0.707	0.821	0.902
	T	r_{Pu}	0.906	0.956	0.988	0.914	0.958	0.989
	R	r_{Pu}	0.906	0.955	0.988	0.913	0.959	0.989
	I	r_{Pu}	0.860	0.912	0.959	0.707	0.821	0.902
	T	r_P^*	0.900	0.950	0.990	0.900	0.950	0.990
	R	r_P^*	0.899	0.950	0.990	0.899	0.950	0.990
	I	r_P^*	0.879	0.946	0.882	0.798	0.862	0.971

Table 5: Coverage levels of confidence intervals for the parameter λ of the log- inverse Gaussian distribution (ϕ unknown and fixed to 1 in the simulations). In situation I, the average number of classes after grouping is about 5 when $\lambda = 1$ ($\rho = 1.1$) and about 3 when $\lambda = 10$ ($\rho = 1.9$).

		$n = 100$	$n = 500$	$n = 1000$
Exponential ($\lambda = 1$)	r, r^*	0.844	0.748	0.624
	r_e, r_u	0.829	0.736	0.613
Gamma ($k = 1.5$)	r, r^*	0.891	0.886	0.789
	r_e, r_u	0.890	0.886	0.785
Gaussian ($\sigma = 1$)	r, r^*	0.867	0.672	0.460
	r_e, r_u	0.830	0.617	0.429
Gaussian ($\mu = 0$)	r, r^*, r_e, r_u	0.882	0.873	0.865
Gamma ($k = 5$)	r_P, r_P^*	0.684	0.584	0.217
	r_{Pe}, r_{Pu}	0.576	0.474	0.172
Inv.Gaussian ($\lambda = 10$)	r_P, r_P^*	0.821	0.492	0.257
	r_{Pe}, r_{Pu}	0.799	0.468	0.249

Table 6: Coverage levels of confidence intervals when data are discretised for large values of n and $1 - \alpha = 0.90$. Only the results for r (and r_P) and r_u (and r_{Pu}) are given since r^* (and r_P^*) and r_e (and r_{Pe}) give results quite similar to r (and r_P) and r_u (and r_{Pu}), respectively.