

ON THE ESTIMATION OF THE STRUCTURE PARAMETER OF A NORMAL DISTRIBUTION OF ORDER p

Angelo M. Mineo

1. INTRODUCTION

In statistical inference the usual hypotheses that we make on sample observations are that they are drawn from a population distributed as a normal, are homoskedastic and independent. However, in many real situations the hypothesis of normality is not met by data that we are provided, so we have the problem to find alternative methods. In literature an used approach is to apply robust methods. Another approach is to hypothesize a different distribution for the observations and to look for deriving suitable methods by beginning from this hypothesis. This approach can be considered as alternative to the robust methods, since rather than referring, implicitly or explicitly, to the theory of the so called outliers, seeks distributional models more general than the normal one, how is pointed out by many researchers. Among them, the most authoritative is certainly Sir Ronald A. Fisher (1922) that, against the practice of excluding from analysis the so called outliers, says this: “as a statistical measure, however, the rejection of observations is too crude to be defended and unless there are other reasons for rejection than mere divergence from the majority, it would be more philosophical to accept these extreme values, not as gross errors, but as indications that the distribution of errors is not normal”. In this sense, the family of normal distributions of order p (Vianelli, 1963; Lunetta, 1963), known as exponential power distribution in the anglo-saxon literature (Box and Tiao, 1992; Gonin and Money, 1989), constitutes a valid alternative to the gaussian normal distribution. A recent review on this distribution family can be found on Chiodi (2000).

In this paper it is faced the problem of the parameter estimation of a normal distribution of order p and particularly that, harder, of the estimation of the structure parameter p , by comparing some of the most interesting proposals existing in literature. In particular, after doing some considerations on the normal distribution of order p and on the maximum likelihood estimators of its parameters, we compare three methods based respectively on the likelihood function, on the profile likelihood function and on the conditional profile likelihood function and a fourth suggestion based on the use of a particular index of kurtosis.

2. NORMAL DISTRIBUTION OF ORDER p

In 1923 Subbotin proposed a distribution family in which every component represents a random error distribution that generalizes the normal one. Subbotin, beginning from these two axioms:

1. the probability of a random error depends only on the dimension of the same error and can be expressed by a function $\varphi(\varepsilon)$ having the first derivative continuous in general;
2. the most probable value of a quantity, of which are known direct measures, must not depend on the used measure unit (in literature this axiom is known as the Schiaparelli second axiom);

that are equal to those used by Gauss with the exception of the second part of the first axiom that in our case is more general (for this part Gauss settled down the condition that the best way to combine observations is to use the arithmetic mean), has derived the distribution that has density function:

$$f(\varepsilon) = \frac{mb}{2\Gamma(1/m)} \exp(-b^m |\varepsilon|^m) \quad (1)$$

In 1963 Lunetta, following the procedure introduced by Pearson (1895) to derive new probability distributions, solves the differential equation:

$$\frac{d \log f}{dx} = p \frac{\log f - \log a}{x - c} \quad (2)$$

that brings to the density function:

$$f(x) = \frac{b^{1/p}}{2\Gamma(1+1/p)} \exp(-b |x - c|^p) \quad (3)$$

By using a different parameterization, the (3) assumes the form:

$$f(x) = \frac{1}{2p^{1/p} \sigma_p \Gamma(1+1/p)} \exp\left(-\frac{|x - \mu|^p}{p\sigma_p^p}\right) \quad (4)$$

for $-\infty < x < \infty$, $-\infty < \mu < \infty$, $\sigma_p > 0$ and $p > 0$, with

$$\mu = E[X] = \int_{-\infty}^{\infty} xf(x)dx \quad (5)$$

the location parameter,

$$\sigma_p = \{E[|X - \mu|^p]\}^{1/p} = \left\{ \int_{-\infty}^{\infty} |x - \mu|^p f(x)dx \right\}^{1/p} \quad (6)$$

the scale parameter and p the structure parameter. It is easy to see how (1) coincides with (4) if we consider the following substitutions: $\varepsilon = x - \mu$, $m = p$ e $h = (p^{1/p} \sigma_p)^{-1}$.

The distributions described by (4) have been called by Vianelli (1963) normal distributions of order p . This distribution family, as is known, describes both leptokurtic ($0 < p < 2$) and platykurtic ($p > 2$) distributions, providing as special cases the Laplace distribution for $p=1$, the gaussian normal distribution for $p=2$ and the uniform distribution for $p \rightarrow \infty$. It is worth noting how this distribution family is often used in literature only for $1 \leq p < 2$, since these distributions present heavier tails than those ones of the normal distribution (see among others Hogg, 1974; D'Agostino and Lee, 1977). However, the case of platykurtic normal distributions of order p is also interesting by a practical point of view, since in the reality we can have samples with observations that we can think drawn either from leptokurtic or platykurtic distributions, how verified by Cox (1967) that by checking different data sets says that "the surprising conclusion was that while there are frequent departures from normality these were about equally often toward long-tailed and toward short-tailed distributions", or by Box (1967), according to whom "platykurtic distributions do occur in practice because of deliberate or unconscious truncation and these ought not to be ruled out a priori".

3. ESTIMATION OF THE PARAMETERS OF A NORMAL DISTRIBUTION OF ORDER p

As it is known, in Statistics the most used estimation method is the maximum likelihood, because it provides estimators with suitable properties, at least asymptotically. The derivation of the maximum likelihood estimators does not give big problems in the case of the normal distribution of order p parameters, even though usually we obtain estimators not expressible in a closed form. Indeed, let's suppose to have a sample of n i.i.d. observations drawn from (4): then the likelihood function is given by:

$$L(x; \mu, \sigma_p, p) = [2 p^{1/p} \sigma_p \Gamma(1 + 1/p)]^{-n} \exp\left(-\frac{\sum_{i=1}^n |x_i - \mu|^p}{p \sigma_p^p}\right) \quad (7)$$

and the log-likelihood function is given by:

$$l(x; \mu, \sigma_p, p) = \log L(x; \mu, \sigma_p, p) = -n \log[2 p^{1/p} \sigma_p \Gamma(1 + 1/p)] - \frac{\sum_{i=1}^n |x_i - \mu|^p}{p \sigma_p^p} \quad (8)$$

If we want to determine the maximum likelihood estimators, we can derive the log-likelihood function respect to the three parameters (μ, σ_p, p) and equal to zero the obtained expressions:

$$\frac{\partial l(x)}{\partial \mu} = -\frac{1}{\sigma_p^p} \sum_{i=1}^n |x_i - \mu|^{p-1} \text{sign}(x_i - \mu) = 0 \quad (9)$$

$$\frac{\partial l(x)}{\partial \sigma_p} = -\frac{n}{\sigma_p} + \frac{1}{\sigma_p^{p+1}} \sum_{i=1}^n |x_i - \mu|^p = 0 \quad (10)$$

$$\begin{aligned} \frac{\partial l(x)}{\partial p} &= \frac{n}{p^2} [\log p + \Psi(1 + 1/p) - 1] + \\ &+ \frac{1}{p\sigma_p^p} \left[\frac{1}{p} \sum_{i=1}^n |x_i - \mu|^p + \log \sigma_p \sum_{i=1}^n |x_i - \mu|^p - \sum_{i=1}^n |x_i - \mu|^p \log |x_i - \mu| \right] = 0 \end{aligned} \quad (11)$$

with $\psi(\cdot)$ the digamma function, i.e. the first derivative of the logarithm of the gamma function (Abramowitz and Stegun, 1972):

$$\Psi(x) = \frac{\partial \ln \Gamma(x)}{\partial x} = \frac{\Gamma'(x)}{\Gamma(x)} \quad (12)$$

Equations (9) and (11) do not give estimators in a closed form, while (10) gives the maximum likelihood estimator for σ_p :

$$\hat{\sigma}_p = \left(\frac{\sum_{i=1}^n |x_i - \mu|^p}{n} \right)^{1/p} \quad (13)$$

The quantity $\hat{\sigma}_p$ is also called power deviation of order p and it can be seen as a general variability index (Vianelli, 1963).

It is also possible to compute the inverse of the Fisher information matrix (Agrò, 1995) that defines the asymptotic variance matrix of the maximum likelihood estimators $(\hat{\mu}, \hat{\sigma}_p, \hat{p})$:

$$\mathbf{I}^{-1} = \begin{bmatrix} \frac{\sigma_p^2 \Gamma(1/p) p^{(2-p)/p}}{(p-1)\Gamma(1-1/p)} & 0 & 0 \\ 0 & \frac{\sigma_p^2}{p} \left\{ 1 + \frac{[\log p + \Psi(1+1/p)]^2}{(1+1/p)\Psi'(1+1/p)-1} \right\} & p\sigma_p \frac{[\log p + \Psi(1+1/p)]^2}{(1+1/p)\Psi'(1+1/p)-1} \\ 0 & p\sigma_p \frac{[\log p + \Psi(1+1/p)]^2}{(1+1/p)\Psi'(1+1/p)-1} & \frac{p^3}{(1+1/p)\Psi'(1+1/p)-1} \end{bmatrix} \quad (14)$$

with $\psi'(\cdot)$ the trigamma function, i.e. the second derivative of the logarithm of the gamma function.

It has been noted (Capobianco, 2000) how in general the asymptotic variance

of the maximum likelihood estimator of the normal distribution of order p scale parameter is larger than the corresponding estimator of the Laplace distribution and normal distribution scale parameter; in fact, the supposed loss of efficiency of the maximum likelihood estimator for σ_p is only due to the need of estimating the structure parameter p . Indeed, by supposing to know the "true" value of the parameter p , the information matrix is given by:

$$\mathbf{I} = \begin{bmatrix} I_{\mu\mu} & I_{\mu\sigma_p} \\ I_{\sigma_p\mu} & I_{\sigma_p\sigma_p} \end{bmatrix} \quad (15)$$

with

$$I_{\mu\mu} = -E \left[\frac{\partial^2 \ln f(x)}{\partial \mu^2} \right] = \frac{(p-1)p^{(p-2)/p} \Gamma(1-1/p)}{\sigma_p^2 \Gamma(1/p)} \quad (16)$$

$$I_{\mu\sigma_p} = I_{\sigma_p\mu} = -E \left[\frac{\partial^2 \ln f(x)}{\partial \mu \partial p} \right] = 0 \quad (17)$$

$$I_{\sigma_p\sigma_p} = -E \left[\frac{\partial^2 \ln f(x)}{\partial \sigma_p^2} \right] = \frac{p}{\sigma_p^2} \quad (18)$$

and, by inverting the information matrix, we obtain:

$$\mathbf{I}^{-1} = \begin{bmatrix} \frac{\sigma_p^2 \Gamma(1/p) p^{(2-p)/p}}{(p-1) \Gamma(1-1/p)} & 0 \\ 0 & \frac{\sigma_p^2}{p} \end{bmatrix} \quad (19)$$

that for $p=1$ becomes:

$$\mathbf{I}^{-1} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_1^2 \end{bmatrix}$$

(in this case to obtain the first element of the matrix, $I_{1,1}^{-1}$, we have to compute a simple limit), while for $p=2$ becomes:

$$\mathbf{I}^{-1} = \begin{bmatrix} \sigma_2^2 & 0 \\ 0 & \frac{\sigma_2^2}{2} \end{bmatrix}$$

At this point, it is easy to see how the asymptotic variance of the maximum likelihood estimator of the normal distribution of order p scale parameter, both for $p=1$ and for $p=2$, results equal to that of the corresponding maximum likelihood estimator of the scale parameter of the Laplace and normal distribution, respectively. Therefore, it is evident that if we have preliminary information such that we can believe with a reasonable certainty that the sample at disposal has been drawn from a normal or a Laplace distribution, then it is needless to use normal distribution of order p . Maybe this preliminary information could also concern the location and scale parameter and so we do not need to make inference at all. However, usually we do not dispose of this information and then it is necessary to consider also p unknown.

For the estimation of p , besides the use of the maximum likelihood estimator (Agrò, 1995), have been proposed other two procedures (Agrò, 1999) based the former on the profile log-likelihood (Barndorff-Nielsen, 1988):

$$l_p(x; \mu, \sigma_p, p) = -n \left\{ \log 2 + \log \Gamma(1 + 1/p) + \frac{1}{p} \left[\log p + 1 + \log \left(\frac{\sum_{i=1}^n |x_i - \mu|^p}{n} \right) \right] \right\} \quad (20)$$

the latter on the conditional profile log-likelihood (Cox and Reid, 1987):

$$l_{CP}(x; \mu, \sigma_p, p) = -n \log 2 - \frac{1}{2} \log \left[n(p-1) \frac{\sum_{i=1}^n |x_i - \mu|^{p-2}}{\sum_{i=1}^n |x_i - \mu|^p} \right] - \log[\sqrt{np}] + \\ - (n-1) \left\{ \log \Gamma(1 + 1/p) + \frac{1}{p} \left[\log p + 1 + \log \left(\frac{\sum_{i=1}^n |x_i - \mu|^p}{n} \right) \right] \right\} \quad (21)$$

In the conditional profile log-likelihood the parameters of interest have to be orthogonal to the nuisance parameters and then, in our case, σ_p has to be orthogonal to p , in such a way making possible, according to Agrò, finite estimates of p even for samples with smaller size ($n=30$) than those considered when we use the (20) or when we apply directly the (8). However, we think that the requirement of finite estimates of p is not convincing, since it is possible to have values of $p \rightarrow \infty$, by defining in this way a normal distribution of order p corresponding to an uniform distribution, as we have already seen.

Anyway, it is necessary to note that in order to estimate p exist other proposals in literature, based on the computation of particular indices of kurtosis (see for example Mineo A.M., 1994, 1995 and 1996). These estimation procedures are to be seen as procedures based on the method of moments that first look for determining the most proper value of p by means of the sample observations, afterwards go through the use of the maximum likelihood estimators to estimate the location parameter μ and the scale parameter σ_p . The indices of kurtosis more used for this aim are:

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\Gamma(1/p)\Gamma(5/p)}{[\Gamma(3/p)]^2} \quad (22)$$

$$\sqrt{VI} = \frac{\sqrt{\mu_2}}{\mu_1} = \frac{\sqrt{\Gamma(1/p)\Gamma(3/p)}}{\Gamma(2/p)} \quad (23)$$

$$\beta_p = \frac{\mu_{2p}}{\mu_p^2} = p + 1 \quad (24)$$

whose moment estimators are given by:

$$\hat{\beta}_2 = \frac{n \sum_{i=1}^n (x_i - M)^4}{[\sum_{i=1}^n (x_i - M)^2]^2} \quad (25)$$

$$\sqrt{\hat{VI}} = \frac{\sqrt{n \sum_{i=1}^n (x_i - M)^2}}{\sum_{i=1}^n |x_i - M|} \quad (26)$$

$$\hat{\beta}_p = \frac{n \sum_{i=1}^n |x_i - M|^{2\hat{p}}}{[\sum_{i=1}^n |x_i - M|^{\hat{p}}]^2} = \hat{p} + 1 \quad (27)$$

From these relationships it is evident the nature of the structure parameter p , that essentially is itself an index of kurtosis, and the fact that p is positively correlated with σ_p , as it can be seen from (14), too. With these considerations it is evident as an estimation method for p based on the use of an index of kurtosis results very promising. In particular, in the simulation study described in the next section we have used the estimation procedure of p based on the index of kurtosis \sqrt{VI} , with value of M in the (26) given by the estimate of the location parameter μ of the corresponding normal distribution of order p (for more details on the method and particularly on the reason why we have chosen the index \sqrt{VI} among the others see Mineo A.M., 1996).

Before ending this section, it is worth noting that about the estimation problem of the normal distribution of order p parameters has been proposed recently an approach based on the use of a genetic algorithm (Vitranò and Baragona, 2001) that, however, does not present substantial improvements in comparison to any other numerical method used to solve the estimation problem by means of the likelihood function.

4. SIMULATION PLAN AND RESULTS

In order to compare the four different approaches to estimate the parameter p , described in the previous section, we have conducted a simulation study by drawing 1000 samples of size 10(10)50, 100, 200 from a normal distribution of order p , with values of $p=(1.5(0.5)3.5)$, location parameter $\mu=50$ and scale parameter $\sigma_p=2$. The samples have been drawn by using the method of the inverse of the distribution function, by exploiting the relationship linking normal distribution of order p and gamma distribution (Mineo A. 1978). Besides the parameter p , the location and the scale parameters have been estimated by using the maximum likelihood estimators. To solve these optimization problems we have chosen the simplex method (Nelder and Mead, 1965) implemented in the function `optim()` of the R software (Ihaka and Gentleman, 1996). Since the simplex method is an unconstrained optimization method, a suitable reparameterization on the functions to be optimized that present constraints on σ_p e p has been necessary (see for a similar example Everitt, 1987, pp. 32-35). The starting points for the simplex method have been the least squares estimates, i.e. respectively the arithmetic mean, the standard deviation and $p=2$.

The obtained results seem very interesting. While it seems that there are not substantial differences among the approaches in the estimation of the location parameter (results not shown), there are some differences in the estimation of the remaining two parameters.

In particular, for the scale parameter we can see (table 1) how for large sample sizes ($n=100, 200$) the whole log-likelihood and the profile log-likelihood seem to give unbiased estimates, while the method based on the index $\hat{V}T$ results competitive. For these sample sizes the conditional profile log-likelihood behave in a good way, but for $p=1.5$. Concerning the estimate variances, all the methods seem to provide very similar values of variance.

For smaller sample sizes ($n=10(10)50$) all the estimates seem biased, with values related to the whole log-likelihood and to the profile log-likelihood biased to the top, while the values related to the conditional profile log-likelihood and to the fourth method based on the index $\hat{V}T$ seem biased to the bottom. However, it is worth noting that for sake of comparison we have used the relationship (13) for all the four methods, by replacing μ with the corresponding maximum likelihood estimate; but for small sample sizes some corrections are necessary, since the maximum likelihood estimator of σ_p^p is biased, as it is known; in particular have been proposed corrections that recall that one used in the case of the maximum likelihood estimator of the variance (see Mineo A.M., 1996), or asymptotic correction (see Chiodi, 1988), that seem to adjust, at least partially, the bias. However, these corrections would not adjust the bias of the estimates derived by using the whole log-likelihood or the profile log-likelihood, since they would increase the mean values reported in table 1. Concerning the variances of these estimates, it is worth noting as the lowest are that derived from the fourth method (estima-

tion based on the index $\hat{V}I$), but for $n=10$ where the lowest variances are that derived from the conditional profile log-likelihood, that anyway have the great drawback described next.

TABLE 1

Arithmetic means (M) and variances (V) of the power deviation of order p estimates for the four used methods (I=whole log-likelihood, II=profile log-likelihood, III=conditional profile log-likelihood, IV=method based on the index $\hat{V}I$)

n		$p=1.5$		$p=2.0$		$p=2.5$		$p=3.0$		$p=3.5$	
		$M(\hat{\sigma}_p)$	$V(\hat{\sigma}_p)$	$M(\hat{\sigma}_p)$	$V(\hat{\sigma}_p)$	$M(\hat{\sigma}_p)$	$V(\hat{\sigma}_p)$	$M(\hat{\sigma}_p)$	$V(\hat{\sigma}_p)$	$M(\hat{\sigma}_p)$	$V(\hat{\sigma}_p)$
10	I	2.5079	1.1643	2.4249	0.8737	2.3859	0.6910	2.3006	0.5674	2.1943	0.4566
	II	2.9228	1.3359	2.7239	0.8335	2.5960	0.6655	2.4727	0.5027	2.3485	0.3916
	III	1.7055	0.2714	1.5870	0.2198	1.5394	0.1947	1.4791	0.1834	1.4139	0.1629
	IV	1.8086	0.4195	1.7161	0.3960	1.7095	0.3824	1.6299	0.3351	1.6016	0.3185
20	I	2.3364	0.7564	2.3396	0.6897	2.3914	0.6087	2.3973	0.5067	2.4228	0.4236
	II	2.5116	1.0883	2.4786	0.8365	2.5645	0.7114	2.5380	0.5488	2.5663	0.4486
	III	1.8550	0.2868	1.7798	0.3086	1.8026	0.3255	1.8128	0.3347	1.8277	0.2954
	IV	1.8811	0.2583	1.8382	0.2989	1.8530	0.3126	1.8378	0.3141	1.8471	0.3019
30	I	2.1342	0.4337	2.1983	0.5083	2.2954	0.4698	2.3111	0.4346	2.3473	0.4072
	II	2.1941	0.5722	2.2700	0.6198	2.4169	0.5832	2.4200	0.5161	2.4593	0.4699
	III	1.8429	0.2167	1.8073	0.2505	1.8886	0.2962	1.8884	0.3004	1.8999	0.2661
	IV	1.8805	0.1806	1.8407	0.2017	1.8939	0.2461	1.8875	0.2672	1.9041	0.2818
40	I	2.1385	0.3194	2.1295	0.3099	2.2152	0.3591	2.2050	0.3390	2.2427	0.3166
	II	2.1617	0.3796	2.1820	0.4055	2.2629	0.4249	2.2821	0.4263	2.3097	0.3713
	III	1.8748	0.1733	1.8507	0.2017	1.9070	0.2066	1.9213	0.2225	1.9480	0.2084
	IV	1.9348	0.1401	1.9037	0.1721	1.9200	0.1922	1.9319	0.2259	1.9517	0.2688
50	I	2.0754	0.1974	2.1165	0.2640	2.1164	0.2169	2.1814	0.2724	2.2054	0.2730
	II	2.0862	0.2216	2.1311	0.2989	2.1592	0.2873	2.2195	0.3207	2.2638	0.3320
	III	1.8651	0.1361	1.9074	0.1982	1.9203	0.1598	1.9657	0.1731	1.9890	0.1692
	IV	1.9409	0.1122	1.9378	0.1627	1.9214	0.1412	1.9616	0.2091	1.9693	0.2268
100	I	2.0092	0.0538	2.0351	0.0540	2.0268	0.0620	2.0488	0.0780	2.0522	0.0716
	II	2.0092	0.0538	2.0365	0.0569	2.0321	0.0723	2.0520	0.0829	2.0565	0.0792
	III	1.9021	0.0702	1.9802	0.0678	1.9872	0.0575	2.0004	0.0558	2.0008	0.0517
	IV	1.9536	0.0464	1.9676	0.0571	1.9536	0.0757	1.9758	0.1019	2.0080	0.1476
200	I	2.0073	0.0276	2.0148	0.0250	2.0105	0.0233	2.0153	0.0199	2.0281	0.0236
	II	2.0072	0.0276	2.0148	0.0250	2.0105	0.0233	2.0153	0.0199	2.0281	0.0236
	III	1.9474	0.0445	2.0032	0.0252	1.9963	0.0213	1.9993	0.0189	2.0094	0.0224
	IV	1.9788	0.0258	1.9864	0.0256	1.9748	0.0252	1.9875	0.0340	2.0095	0.0562

For the structure parameter p , we can note as all the methods show some samples that join the theoretical bounds that we imposed can assume the parameter p , i.e. 1 and $+\infty$ (we have imposed to p as lower bound 1 because even if p could assume values up to 0, defining in this way normal distributions of order p of some statistical interest since we have cuspidate distributions with tails heavier than the Laplace distribution, the estimation for $0 < p < 1$ involves remarkable computational problems): we are not particularly worried about this occurrence, since for $p=1$ and for $p \rightarrow \infty$, as we have seen, we have special probability distributions very used in Statistics. It is evident that a good p estimator has to behave in such a way that these values do not happen very often when we have drawn samples with a finite value of $p \neq 1$.

The simulation results concerning the estimation of p are reported in tables 2 and 3.

TABLE 2

Arithmetic means (M) and variances (V) of the structure parameter p estimates for the four used methods (I=whole log-likelihood, II=profile log-likelihood, III=conditional profile log-likelihood, IV=method based on the index $\hat{V}I$).

n		$p=1.5$		$p=2.0$		$p=2.5$		$p=3.0$		$p=3.5$	
		$M(\hat{p})$	$V(\hat{p})$	$M(\hat{p})$	$V(\hat{p})$	$M(\hat{p})$	$V(\hat{p})$	$M(\hat{p})$	$V(\hat{p})$	$M(\hat{p})$	$V(\hat{p})$
10	I	1.3775	0.3543	1.4467	0.3889	1.5447	0.4582	1.5441	0.4929	1.5610	0.4467
	II	1.1655	0.0971	1.2291	0.1390	1.2483	0.1167	1.2829	0.1546	1.2608	0.1146
	III	1.4070	2.1852	1.5101	2.7133	1.6758	3.5216	1.9513	4.4857	1.9619	4.7065
	IV	1.4902	1.1806	1.6036	1.3460	1.8126	2.0320	1.8678	1.9395	1.9411	2.0547
20	I	1.8891	1.2460	2.2987	1.7187	2.1619	2.7002	2.8312	2.4184	3.1309	2.8716
	II	1.6825	0.6388	2.0439	0.9994	2.2147	1.0311	2.4067	1.3207	2.4691	1.4221
	III	1.5686	2.2364	1.8647	3.0941	2.3234	4.4346	2.8098	5.6338	3.1737	6.2729
	IV	1.5759	1.0477	1.8106	1.2657	2.1967	2.1823	2.4691	2.6015	2.6210	2.7663
30	I	1.8924	1.2147	2.4361	2.0545	3.1064	2.6579	3.4806	3.0479	3.7801	3.4101
	II	1.8071	0.9103	2.2642	1.5510	2.8532	2.1050	3.1619	2.2953	3.3588	2.3221
	III	1.4732	1.3572	1.9298	2.5157	2.6107	3.8825	2.8756	4.0623	3.4471	5.1812
	IV	1.4905	0.5402	1.9122	1.2425	2.3716	1.8949	2.6564	2.6130	2.7783	2.2672
40	I	1.9027	1.0457	2.5003	1.7885	3.0385	2.0904	3.6320	3.2244	3.9720	2.9550
	II	1.8626	0.8943	2.3973	1.4506	2.9742	1.9978	3.4045	2.5245	3.7923	2.5638
	III	1.4763	1.1450	1.9377	1.7780	2.6128	2.9507	3.1624	3.8031	3.5505	3.8479
	IV	1.5308	0.4524	1.9448	0.9258	2.3705	1.3920	2.8432	2.4154	2.9820	2.4394
50	I	1.8009	0.6341	2.4816	1.5094	3.0969	2.1141	3.5966	2.3678	4.1509	3.3259
	II	1.7836	0.5743	2.4314	1.2568	3.0068	1.8247	3.5000	2.1246	3.9910	3.0510
	III	1.4152	0.6951	2.0925	1.7419	2.6259	2.1699	3.2479	2.8675	3.7632	3.2419
	IV	1.5446	0.4192	2.0769	1.1608	2.4343	1.3294	2.9261	2.2080	3.2036	2.3135
100	I	1.6062	0.1524	2.2638	0.5549	2.8208	0.8190	3.4217	1.3420	4.0820	1.9837
	II	1.6061	0.1524	2.2613	0.5488	2.8133	0.8064	3.4125	1.3022	4.0733	1.9664
	III	1.4058	0.2215	2.1072	0.5151	2.6643	0.6764	3.2365	1.1450	3.8002	1.4947
	IV	1.4862	0.1169	2.0362	0.4528	2.4901	0.7157	3.0220	1.3878	3.5818	2.4524
200	I	1.5458	0.0632	2.0916	0.1356	2.6295	0.2508	3.2184	0.4741	3.8375	0.7912
	II	1.5458	0.0633	2.0915	0.1356	2.6294	0.2507	3.2184	0.4738	3.8375	0.7916
	III	1.4402	0.1166	2.0580	0.1298	2.5822	0.2705	3.1329	0.4087	3.7091	0.6539
	IV	1.4878	0.0531	2.0112	0.1387	2.5033	0.3135	3.0706	0.6396	3.7026	1.4575

In table 2 we have reported the means and variances of \hat{p} considering all the values of \hat{p} , but $\hat{p} > 10.0$ that have been considered as $\hat{p} \rightarrow \infty$ (for this problem see Mineo A.M., 1996). From these values we can see how for large sample sizes ($n=100, 200$) the method that seems to give the best values of \hat{p} is that based on the use of the index of kurtosis $\hat{V}I$. For smaller sample sizes ($n=10(10)50$) the fourth method seems again better than the others three, also considering that the mean computed on \hat{p} suffers of the great skewness of its sampling distribution: indeed, in a previous simulation study (Mineo A.M., 1995) we have already noted as from the values of \hat{p} obtained with the fourth method we get frequency distributions with mode centered on the “true” value of p used to generate the samples of pseudo-random observations.

Anyway, we consider really interesting the results shown on table 3, that reports the percentage of values with $\hat{p} > 10.0$ (and then $\hat{p} \rightarrow \infty$) and values with $\hat{p} < 1.01$ (and then $\hat{p} = 1$). In fact, in this table we can note as in the case of conditional profile log-likelihood it is very high the percentage of values with $\hat{p} = 1$: for sample sizes $n=10$ this percentage varies from 80% to 90%, with decreasing

percentages as n increases, but that are very high, anyway. According to our opinion, this behaviour shows a fundamental inadequacy of the conditional profile log-likelihood to estimate p , especially when we have samples with medium-small sizes.

TABLE 3

Percentages of samples that present estimates of the structure parameter p greater than 10.0 or lower than 1.01 for the four used methods (I=whole log-likelihood, II=profile log-likelihood, III=conditional profile log-likelihood, IV=method based on the index \hat{VI}).

n	\hat{p}	$p=1.5$		$p=2.0$		$p=2.5$		$p=3.0$		$p=3.5$	
		>10.0	<1.01	>10.0	<1.01	>10.0	<1.01	>10.0	<1.01	>10.0	<1.01
10	I	53.10	23.40	63.30	16.60	71.10	10.50	76.70	9.50	78.00	7.70
	II	75.40	15.00	82.10	9.70	85.50	7.00	89.40	5.10	90.30	4.10
	III	0.30	91.90	1.20	89.00	1.80	85.50	1.80	80.10	1.60	80.40
	IV	7.60	51.60	10.40	42.60	15.20	33.40	16.30	30.10	20.30	26.80
20	I	22.30	18.10	31.30	9.40	42.00	5.30	53.00	3.40	61.80	3.30
	II	32.50	16.90	41.40	9.00	56.50	5.00	64.90	2.90	74.90	3.20
	III	3.60	80.20	5.30	71.30	8.50	57.90	11.20	47.30	14.40	40.40
	IV	2.50	33.90	6.20	18.00	9.20	11.40	11.70	8.50	16.10	7.20
30	I	7.30	16.60	16.40	5.20	25.60	1.40	32.80	1.30	41.10	0.40
	II	11.00	16.30	22.20	5.10	35.80	1.30	43.00	1.30	52.20	0.30
	III	1.40	78.50	3.10	63.20	6.50	42.70	9.80	34.90	10.00	26.50
	IV	1.00	27.90	2.20	9.70	6.00	3.50	8.50	2.00	12.90	0.30
40	I	4.00	9.70	7.30	2.80	14.80	0.90	18.90	0.50	25.90	0.20
	II	5.50	9.70	11.30	2.80	18.50	0.80	26.20	0.50	32.60	0.20
	III	0.50	74.60	1.70	53.30	2.80	34.00	5.40	23.40	8.10	15.60
	IV	0.40	16.10	2.10	4.80	3.60	1.00	6.70	0.40	11.50	0.30
50	I	1.70	7.30	5.20	1.20	6.00	0.20	13.90	0.00	19.50	0.10
	II	2.40	7.30	6.40	1.20	9.60	0.20	17.50	0.00	25.40	0.10
	III	0.30	72.50	1.30	41.30	1.30	25.50	3.60	13.30	6.60	8.60
	IV	0.20	11.00	1.10	1.60	1.90	0.30	5.40	0.10	8.90	0.00
100	I	0.00	2.40	0.00	0.00	0.40	0.00	1.70	0.00	1.90	0.00
	II	0.00	2.40	0.10	0.00	0.80	0.00	2.00	0.00	2.30	0.00
	III	0.00	51.10	0.00	12.60	0.20	2.40	0.60	0.40	0.80	0.10
	IV	0.00	2.70	0.10	0.00	0.60	0.00	1.60	0.00	3.90	0.00
200	I	0.00	0.30	0.00	0.00	0.10	0.00	0.00	0.00	0.20	0.00
	II	0.00	0.30	0.00	0.00	0.10	0.00	0.00	0.00	0.20	0.00
	III	0.00	30.10	0.00	0.60	0.00	0.10	0.00	0.00	0.20	0.00
	IV	0.00	0.10	0.00	0.00	0.00	0.00	0.20	0.00	0.70	0.00

Concerning the methods based on the whole log-likelihood and on the profile log-likelihood, these seem to have the opposite drawback, that is it seems very high the percentage of samples with $\hat{p} > 10.0$, with results that seem better for the method based on the whole log-likelihood, that have yet the drawback of a greater computational complexity, at least in comparison to the approach based on the profile log-likelihood.

The method to estimate p based on the index \hat{VI} seems the best among the four, having neither a great number of samples with $\hat{p} = 1$, neither with $\hat{p} \rightarrow \infty$. Therefore, by summing up all the considerations done so far, the fourth method is surely to prefer in comparison to the others three.

5. CONCLUSION

In this paper we have seen how the family of normal distributions of order p constitutes a valid generalization of the hypothesis of normality that usually is made and that often it is not sustainable on data we have at disposal. The use of the normal distributions of order p constitutes also a parametric alternative to the robust methods, that referring, implicitly or explicitly, to the theory of the so called outliers do not seem suitable to a profitable use in the scientific research. How is dangerous the practice to eliminate automatically the outliers is testified by the following real event (Faraway, 2000, pag. 70): NASA launched the Nimbus 7 satellite to record information on the terrestrial atmosphere. After several years of operation, in 1985 the British Antarctic Survey observed a large decrease of the level of the atmospheric ozone over the Antarctica. NASA astonished on the fact that its satellite did not record such anomaly ever: by examining more careful the satellite data it was found that the data processing program automatically discarded extremely low observations, assuming that they were wrong recordings. With good reason we can believe that this “drawback” retarded the discovery of the so called ozone hole over the Antarctica, delaying, as a result, the adoption of the correct policies to try to reduce it (for example, by banning CFC).

Concerning the normal distributions of order p , however exist some practical problems still open that have precluded a wide use, so far: one of these problems is the estimation of the structure parameter p . In this paper we have compared some interesting proposals among those existing in literature and in particular the estimation based on the whole log-likelihood, on the profile log-likelihood, on the conditional profile log-likelihood and on the index of kurtosis $\hat{V}T$. The results of a simulation study show as the best method is that one based on the index of kurtosis $\hat{V}T$, while, if for any particular reason, we want to use an approach based on the likelihood function, the whole log-likelihood seems behave better than the profile log-likelihood and the conditional profile log-likelihood.

*Dipartimento di Scienze Statistiche e Matematiche “Silvio Vianelli”
Università di Palermo*

ANGELO M. MINEO

REFERENCES

- M. ABRAMOWITZ, I.A. STEGUN (1972), *Handbook of mathematical functions*, Dover Publications, New York.
- G. AGRÒ (1995), *Maximum likelihood estimation for the exponential power distribution*, “Communications in Statistics (simulation and computation)”, 24, part 2, pp. 523-536.
- G. AGRÒ (1999), *Parameter orthogonality and conditional profile likelihood: the exponential power function case*, “Communications in Statistics (theory and methods)”, 28, part 8, pp. 1759-1768.
- O.E. BARNDORFF-NIELSEN (1988), *Parametric statistical models and likelihood*, *Lecture notes in statistics*, Springer, Heidelberg.
- G.E.P. BOX (1967), *Discussion on “Topics in the investigation of linear relations fitted by the method of least squares”*, “Journal of the Royal Statistical Society”, Serie B, 29, part 1, pp. 42-43.

- G.E.P. BOX, G.C. TIAO (1992), *Bayesian inference in statistical analysis*, J. Wiley, New York.
- R. CAPOBIANCO (2000), *Robustness aspects of the generalized normal distribution*, "Quaderni di Statistica, Centro per la formazione in Economia e Politica dello sviluppo rurale, Dipartimento di Scienze Statistiche dell'Università di Napoli "Federico II", Dipartimento di Scienze Economiche dell'Università di Salerno", n. 2, pp. 127-145.
- M. CHIODI (1988), *Sulle distribuzioni di campionamento delle stime di massima verosimiglianza dei parametri delle curve normali di ordine p* , Istituto di Statistica di Palermo, Palermo.
- M. CHIODI (2000), *Le curve normali di ordine p come distribuzioni di errori accidentali: una rassegna dei risultati e problemi aperti per il caso univariato e per quello multivariato*, in *Atti della XL Riunione Scientifica della SIS*, Firenze, pp. 289-300.
- D.R. COX (1967), *Discussion on "Topics in the investigation of linear relations fitted by the method of least squares"*, "Journal of the Royal Statistical Society", Serie B, 29, part 1, p. 39.
- D.R. COX, N. REID (1987), *Parameter orthogonality and approximate conditional inference*, "Journal of the Royal Statistical Society", Serie B, 49, pp. 1-39.
- R.B. D'AGOSTINO, A.F.S. LEE (1977), *Robustness of location estimators under changes of population kurtosis*, "Journal of the American Statistical Association", 72, pp. 393-396.
- B.S. EVERITT (1987), *Introduction to optimization methods and their application in statistics*, Chapman and Hall, London.
- J.J. FARAWAY (2000), *Practical regression and anova using R*, Published on the URL: <http://www.stat.lsa.umich.edu/~faraway/book/>.
- R.A. FISHER (1922), *On the mathematical foundations of theoretical statistics*, "Philosophical Transactions of the Royal Society of London", Serie A, 222, pp. 309-368.
- R. GONIN, A.H. MONEY (1989), *Nonlinear L_p -norm estimation*, M. Dekker, New York.
- R.V. HOGG (1974), *Adaptive robust procedures: a partial review and some suggestions for future applications and theory*, "Journal of the American Statistical Association", 69, pp. 909-927.
- R. IHAKA, R. GENTLEMAN (1996), *R: A language for data analysis and graphics*, "Journal of Computational and Graphical Statistics", 5, pp. 299-314.
- G. LUNETTA (1963), *Di una generalizzazione dello schema della curva normale*, "Annali della Facoltà di Economia e Commercio di Palermo", 17, pp. 237-244.
- A. MINEO (1978), *Prontuari delle probabilità integrali delle curve normali di ordine r comprese fra $\pm k, \sigma$, e criteri per la loro valutazione e il loro impiego*, Edigraphica Sud Europa, Palermo.
- A.M. MINEO (1994), *Un nuovo metodo di stima di p per una corretta valutazione dei parametri di intensità e di scala di una curva normale di ordine p* , in *Atti della XXXVII Riunione Scientifica della SIS*, 2, San Remo, pp. 147-154.
- A.M. MINEO (1995), *Stima dei parametri di intensità e di scala di una curva normale di ordine p (p incognito)*, "Annali della Facoltà di Economia e Commercio di Palermo (Area Statistico-Matematica)", 49, pp. 125-159.
- A.M. MINEO (1996), *La migliore combinazione delle osservazioni: curve normali di ordine p e stimatori di norma L_p* , Tesi di Dottorato di Ricerca in Statistica Computazionale e Applicazioni, Università degli Studi di Napoli "Federico II".
- J.A. NELDER, R. MEAD (1965), *A simplex method for function minimization*, "The Computer Journal", 7, pp. 308-313.
- K. PEARSON (1895), *Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material*, "Philosophical Transactions of the Royal Society of London", Serie A, 186, pp. 343-414.
- M.T. SUBBOTIN (1923), *On the law of frequency of errors*, "Matematicheskii Sbornik", 31, pp. 296-301.
- S. VIANELLI (1963), *La misura della variabilità condizionata in uno schema generale delle curve normali di frequenza*, "Statistica", 23, pp. 447-474.
- S. VITRANO, R. BARAGONA (2001), *The genetic algorithm estimates for the parameters of order p normal distributions*, in *Book of Short Papers CLADAG2001*, Palermo, pp. 45-48.

RIASSUNTO

Sulla stima del parametro di struttura di una distribuzione normale di ordine p

In questo lavoro si confrontano quattro differenti approcci per la stima del parametro di struttura di una distribuzione normale di ordine p (spesso chiamata nella letteratura anglosassone exponential power distribution). In particolare, abbiamo considerato la massimizzazione della log-verosimiglianza, della log-verosimiglianza profilo, della log-verosimiglianza profilo condizionale e un metodo basato su un indice di curtosi. I risultati di uno studio di simulazione sembrano indicare la superiorità dell'ultimo approccio.

SUMMARY

On the estimation of the structure parameter of a normal distribution of order p

In this paper we compare four different approaches to estimate the structure parameter of a normal distribution of order p (often called exponential power distribution). In particular, we have considered the maximization of the log-likelihood, of the profile log-likelihood, of the conditional profile log-likelihood and a method based on an index of kurtosis. The results of a simulation study seem to indicate the latter approach as the best.