Mecánica Computacional, Volumen XIX, pp. 359-366
Compilado por: F. Quintana y S. Felicelli
San Carlos de Bariloche, Noviembre de 2000

DENSITY ESTIMATION USING NON-CONVENTIONAL STATISTICS

Danilo R. Velis
Departamento de Geofísica Aplicada, Observatorio Astronómico
Universidad Nacional de La Plata, Paseo del Bosque s/n, La Plata 1900, Argentina

ABSTRACT

In this work the use of non-conventional statistics for estimating the probability density function of a given process is investigated. It is well known that the calculation of higher-order statistics, like skewness and kurtosis (which I call C-moments), is very sensitive to the presence of outliers and very dependent on sample size. Recently, the development of L-moments estimators (which are linear combinations of ordered statistics) had a significant impact on the use of sample statistics to infer probabilities. In contrast to C-moments, L-moments provide more robust and consistent sample estimators. I take advantage of this fact to obtain superior non-parametric pdf estimates via the principle of maximum entropy. The potential use of alternative skewness and kurtosis measures is also explored. The results obtained from simulation studies are discussed.

INTRODUCTION

Probability density functions (pdf) are of central importance in science and engineering. Assessing the pdf shape of a sample of data is useful, for example, for screening data for outliers, describing information about asymmetry and tail weight, determining how well a distribution fits the data, etc. One way of estimating the pdf is by means of the calculation of moments (including mean and variance) from the available data, which are in turn used as constraints the pdf must satisfy. The calculation of higher-order statistics, such as skewness and kurtosis (which I call C-moments), is very sensitive to the presence of outliers and very dependent on sample size. Recently, the development of L-moments estimators (which are based on ordered statistics) had a significant impact on the use of sample statistics to infer probabilities. In contrast to C-moments, L-moments are more robust to the presence of outliers in the data, more consistent for small samples, and less biased [4]. I couple these facts together with the principle of maximum entropy to obtain conservative pdf's which are superior, in most cases, to C-moments derived ones.

The method described in this work consists of selecting, among all distributions that agree with the available information (sample moments), the one which has the least information content. This is the so-called principle of maximum entropy [5, 2]. The usual procedure is to maximize the entropy of the unknown distribution subject to constraints on its moments. The key is the use of L-moments instead of C-moments to constrain the unknown pdf [12]. This paper also explores the potential use of alternative skewness and kurtosis measures for estimating densities using the maximum entropy method. These new measures, which I call S-measures for convenience, were defined in [10]. The results obtained from several simulation studies with symmetric and asymmetric distributions are discussed.

BACKGROUND

Quantifying the shape of a distribution is important in data analysis. Classical skewness and kurtosis are two of the main indices that characterize shape and are included in most statistical packages. The first one is associated with the symmetry (or lack of symmetry), and the second one is usually associated to tail heaviness, pdf peakedness, bimodality, or any combination of the three concepts[1]. Classical skewness and kurtosis are higher-order moments which are difficult to estimate when sample size is small, specially because they are too sensitive to moderate fluctuations in the tail of the distribution (where outliers may be present). This is the main motivation for developing new skewness and kurtosis measures which are more robust, less biased and more consistent for small sample sizes.

C-moments

Let X be a random variable with probability density function $p(x)$ and mean μ. The r-th moment about the mean of $p(x)$ is defined as the expectation of $(X-\mu)^r$, that is $E[(X-\mu)^r] = \int (x-\mu)^r p(x)dx$. Denoting the variance by $\sigma^2 = E[(X-\mu)^2]$, the usual indices of skewness, $\sqrt{\beta_1}$, and kurtosis, β_2, are $E[Z^3]$ and $E[Z^4]$, respectively, where $Z = \frac{X-\mu}{\sigma}$ is the standardized variable, a random process with zero-mean and unit variance.

In a practical context, given the sample $\{x_1, \ldots, x_n\}$, skewness and kurtosis are estimated using average values where μ is replaced by $\bar{x} = \sum x_i/n$, and σ^2 by $\bar{s}^2 = \sum(x_i - \bar{x})^2/n$. So

$$\sqrt{\beta_1} = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i-\bar{x})^3}{\bar{s}^3} \quad \text{and} \quad \beta_2 = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i-\bar{x})^4}{\bar{s}^4}. \tag{1}$$

L-moments

Let $X_{1:1} \leq \cdots \leq X_{1:n}$ be the population order statistics corresponding to random samples of size n drawn from $p(x)$ (note that these quantities are clearly distinct from values $x_{1:1}, \ldots, x_{1:n}$ obtained in actual sampling). The L-moments of X are defined by $l_r = \frac{1}{r}\sum_{j=0}^{r-1}(-1)^j\binom{r-1}{j}E[X_{r-j:r}]$, $(r = 1, 2, \ldots)$, where $X_{j:n}$ is the j-th smallest observation in a sample of size n. Note that we are estimating a linear combination of order statistics. In fact the "L" in L-moments represents exactly this linearity. Akin to the definition of conventional normalized moments, $\tau_3 = \frac{l_3}{l_2}$ and $\tau_4 = \frac{l_4}{l_2}$ are statistics related to the skewness and kurtosis of the pdf, usually called L-skewness and L-kurtosis.

Hosking [4] has shown that the first four L-moments are $l_1 = \gamma_0$, $l_2 = 2\gamma_1 - \gamma_0$, $l_3 = 6\gamma_2 - 5\gamma_1 + \gamma_0$, and $l_4 = 20\gamma_3 - 30\gamma_2 + 12\gamma_1 - \gamma_0$, respectively, where $\gamma_r = \int_0^1 x(F)F^r dF$ are the probability weighting moments (PWM) defined by [3], and F is the cumulative distribution. Finally, given ranked samples of X, $x_1 \leq \cdots \leq x_n$, Landwehr et al. [7] have shown that the unbiased estimator of γ_r is given by

$$\hat{\gamma}_r = \frac{1}{n}\sum_{i-1}^{n}\frac{(i-1)(i-2)\cdots(i-r)}{(n-1)(n-2)\cdots(n-r)}x_i. \tag{2}$$

Sample L-moments may now be computed by means of equation (2).

S-measures

Seier [10] proposed a family of skewness and kurtosis measures of the form $E[g(f(Z))]$, where g is a linear function, f is an odd or even continuous function, and Z is the standardized variable. Some common skewness and kurtosis measures are identified as members of this family, such as $\sqrt{\beta_1} = E[Z^3]$ and $\beta_2 = E[Z^4]$. Two other interesting members, which I call S-skewness and S-kurtosis for convenience, are $\theta_1 = E[a_1 Z|Z|^{b_1}]$ and $\theta_2 = E[a_2 b_2^{-|Z|}]$, respectively. Here a_1,

[1] How kurtosis is related to shape is far from clear (see for example [9, 1]).

b_1, a_2 and b_2 are constants that I choose to be: $a_1 = 2$, $b_1 = 0.5$, $a_2 = 5.7344$, and $b_2 = e$ ($a_2 = 5.7344$ is to honor $E[a_2 e^{-|Z|}] = 3$ for a normal distribution [10]).

Like in the computation of $\sqrt{\beta_1}$ and β_2, θ_1 and θ_2 are estimated using average values, i.e.

$$\theta_1 = \frac{2}{n}\sum_{i=1}^{n}\frac{(x_i-\bar{x})|x_i-\bar{x}|^{1/2}}{\bar{s}^{3/2}} \quad \text{and} \quad \theta_2 = \frac{5.7344}{n}\sum_{i=1}^{n} e^{-\frac{|x_i-\bar{x}|}{s}}. \tag{3}$$

At this point it is worth mentioning that often τ_3 and τ_4 are less sensitive to small changes in the tails of the distributions than $\sqrt{\beta_1}$ and β_2. Thus, L-measures are more robust to the presence of outliers than the classical measures. Also τ_3 and τ_4 have less variability, are less biased and are less sensitive to sample size than $\sqrt{\beta_1}$ and β_2. On the other hand $\sqrt{\beta_1}$ and β_2 tend to be highly correlated when the distribution is skewed. This is because C-moments focus on the tails, whereas τ_3 and τ_4 focus on the tails and on the peak of the distribution, respectively. These and other issues motivated some authors to suggest the replacement of $\sqrt{\beta_1}$ and β_2 by τ_3 and τ_4 in all statistical packages [8]. S-skewness and S-kurtosis exhibit a strong correlation with τ_3 and τ_4. As a consequence, θ_1 and θ_2 are also quite robust and consistent shape indices, being suitable alternative measures for quantifying shape [10]. For a detailed review the reader is referred to [4, 8, 10].

MAXIMUM ENTROPY PRINCIPLE

A very useful method for conservatively assigning probabilities consists of maximizing the entropy of the unknown distribution subject to constraints on its moments [5]. The available information is given by the natural constraint $\int p(x)dx = 1$, and the moment constraints $\int h(x;r)p(x)dx = \mu_r$, $(r = 1, \ldots, K)$, where $h(x;r) = x^r$ and K is the number of moments that are taken into account, and μ_r are estimated from the data using sample statistics. Lagrange multipliers are then used to solve this constrained optimization problem. The replacement of $h(x;r)$ by $f(g(x))$ for $r = 3, 4$ only, allows one to use the same algorithm for setting S-skewness and S-kurtosis constraints instead of the classical indices.

Because L-moments are defined in terms of $F(x)$ rather than in terms of $p(x)$, L-moments constraints cannot be easily incorporated into the optimization problem through Lagrange multipliers [12]. Rather, the problem is transformed into an unconstrained optimization problem by minimizing the following cost function with respect to the unknown distribution:

$$\Phi[p(x); F(x)] = -H + \alpha\left[\left(1 - \int_a^b p(x)dx\right)^2 + \sum_r \left(l_r - \hat{l}_r\right)^2\right], \tag{4}$$

where H is the entropy and α is a constant.

SIMULATIONS

I now illustrate the behavior of the non-parametric density estimation using the maximum entropy criterion with C-, L- and S-measures constraints. I choose four distributions that reflect only kurtosis (Laplace and Logistic) and both skewness and kurtosis (Exponential and Lognormal) (see for example [6]). Clearly, this is a rather limited investigation, but, I am primarily interested in some broad rather than detailed conclusions. In all cases I set $K = 4$, and $n = \{20, 50, 100, 250, 500, 1000\}$. Also, for each sample I tested the methods under the presence of outliers. For comparison, all examples show the pdf derived using a kernel approach, specifically the Epanechnikov kernel [11].

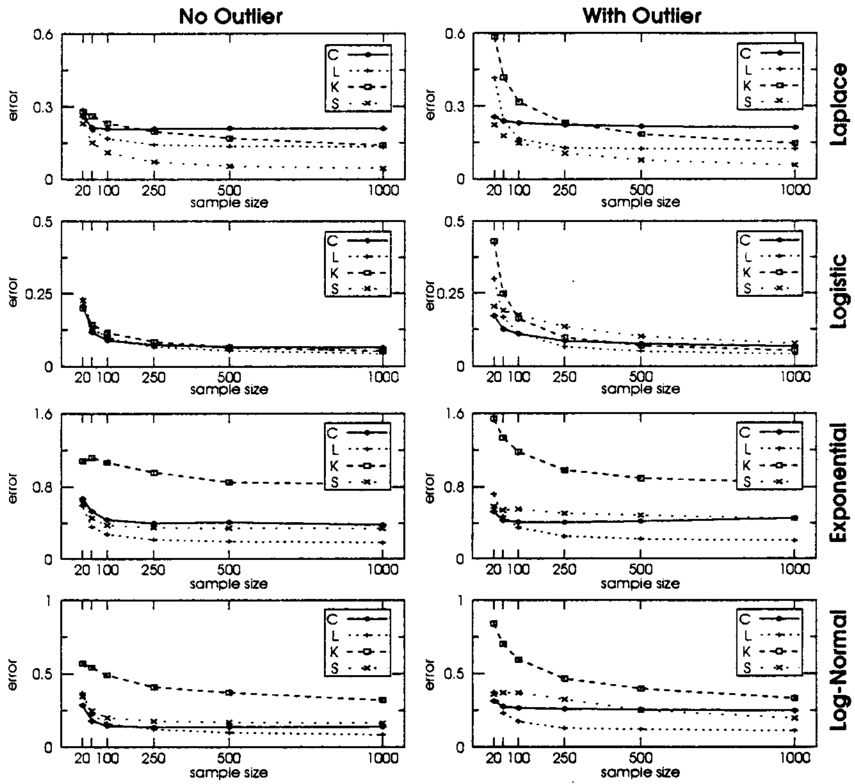

Figure 1: Pdf results from numerical simulations. The curves show $\bar{e}_{rms}$, equation (5), vs sample size n for various distributions, with and without outlier. "C" (C-moments), "L" (L-moments), "S" (S-measures), and "K" (Kernel) denote the method utilized.

Symmetric distributions

Here, samples of different sizes are drawn from Laplace and Logistic distributions, without outliers and with an outlier at $X = 1.75$. The results of the computations are shown in Figure 1 (top two rows), where I have plot the root-mean-square error, $\bar{e}_{rms}$, between true and estimated pdf's, after averaging 500 independent realizations, that is

$$\bar{e}_{rms} = \frac{1}{500}\sum_{j=1}^{500}\sqrt{\frac{1}{m}\sum_{i=1}^{m}(p_i - \hat{p}_i^j)^2}, \tag{5}$$

where $\hat{p}_i^j$ is the estimated pdf i-th coordinate point of the j-th realization. Figure 2 shows the measures used as constraints in the estimation of the Laplace distribution for $n = 50$ and $n = 250$ (500 realizations). Note that classical skewness and kurtosis are highly correlated. Also, note that L-moments and S-measures are less sensitive to the presence of the outlier that

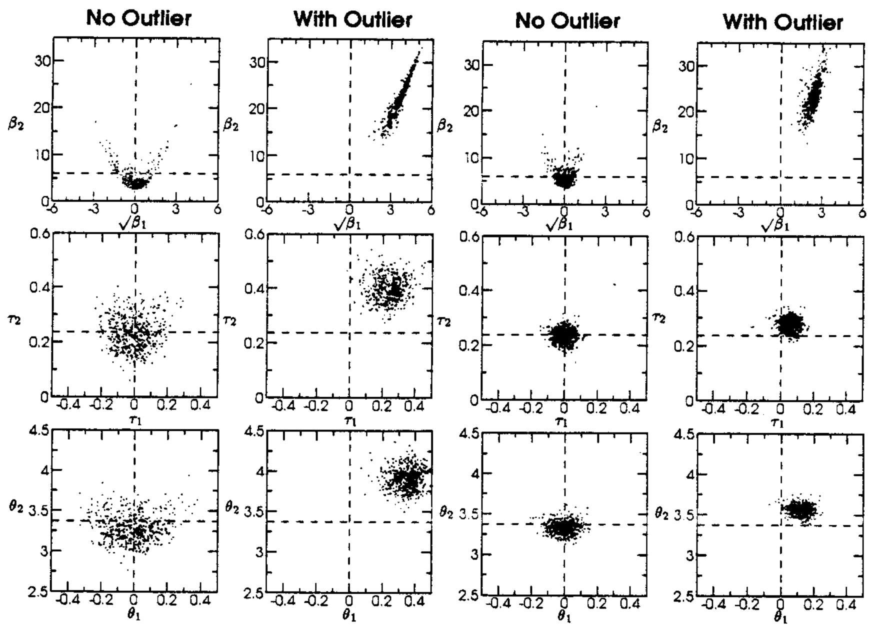

Figure 2: Skewness vs kurtosis (row 1), L-skewness vs L-kurtosis (row 2), and S-skewness vs S-kurtosis (row 3) for the Laplace distribution (500 realizations) using $n = 50$ (columns 1 and 2) and $n = 250$ (columns 3 and 4). Dashed lines represent theoretical values.

C-moments. This explains, in part, why L-moments and S-measures derived pdf's are better estimates than C-moments derived ones, as illustrated in Figure 1, specially for the Laplace distribution. Estimated Laplace distributions ($n = 250$) are displayed in Figure 3. Note how well it has been estimated when using L-moments and S-measures. Conventional statistics fail to characterize the peak of the distribution.

Skewed distributions

In these experiments, samples are drawn from Exponential and Log-normal distributions, without outliers and with an outlier at $X' = 1.75$ and $X = 2.5$, respectively. Figure 1 shows $\bar{e}_{rms}$ as a function of n for the various methods. Note that L-moments derived pdf's are superior estimates of the true pdf for almost all sample sizes. Besides, the sensitivity to the presence of the outlier is smaller. Here, S-measures derived pdf's are not better than C-moments derived ones. However, $\bar{e}_{rms}$ decreases faster with increasing n when using S-measures than when using C-moments under the presence of the outlier. Figure 4 displays the estimated Log-normal pdf's ($n = 250$). It is clear that L-moments derived estimators are much more accurate than C-moments derived ones, specially in the outlier case.

CONCLUSIONS

I have investigated the use of non-conventional statistics to obtain density estimators in the maximum entropy method. The approach I have used is to compare the rms error between

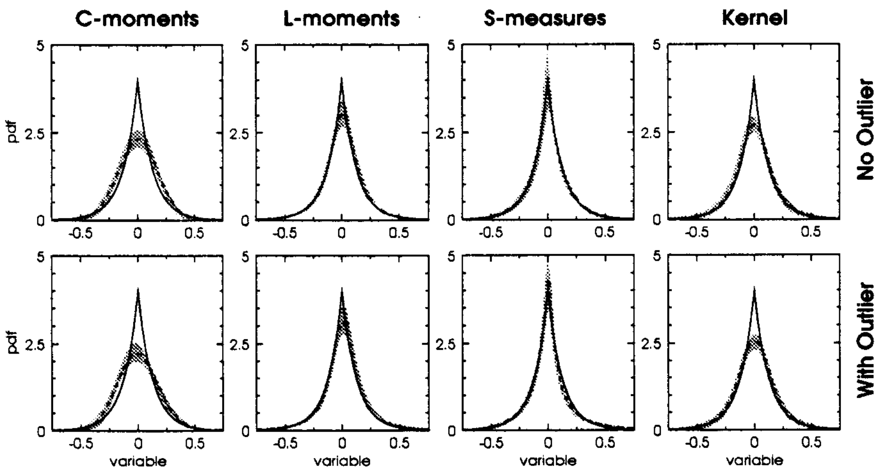

Figure 3: Laplace pdf estimates ($n = 250$). The shaded area is limited by 10 and 90 percentiles of the results (500 realizations), while the dashed line shows the median solution and the solid line the theoretical distribution.

the theoretical pdf and the corresponding estimates, when selecting alternative sets of moment constraints. In general, L-moment constraints provide superior pdf estimates. C-moment constraints, on the contrary, provide limited accuracy even for large sample sizes, specially under the presence of outliers. The performance of S-measures is intermediate. When using S-skewness and S-kurtosis as constraints for the Laplace case, the results are excellent, but they tend to be poor in the Log-normal case. However, I believe they represent an interesting alternative to L-moments, because the latter lead to a more complicated optimization problem requiring penalty terms to constrain the maximum entropy solution.
The results show not only how powerful it is the maximum entropy method for estimating the pdf given a (possible small) data sample, but also how non-conventional statistics that measure shape are utilized to obtain superior density estimates than using classical skewness and kurtosis.

References

[1] Kevin P. Balanda and H.L. MacGillivray. Kurtosis: a critical review. *The Am. Stat.*, 42(2):111–119, 1988.

[2] W. Gouveia, F.S. de Moraes, and John Scales. Entropy, information and inference. 1998. (Preprint).

[3] J.A. Greenwood, J.M. Landwehr, N.C. Matalas, and J.R. Wallis. Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form. *Water Resour. Res.*, 15:1049–1054, 1979.

[4] J.R.M. Hosking. L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Int. J. R. Stat. Soc., Ser. B*, 52:105–124, 1990.

[5] E.T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106:620–630, 1957.

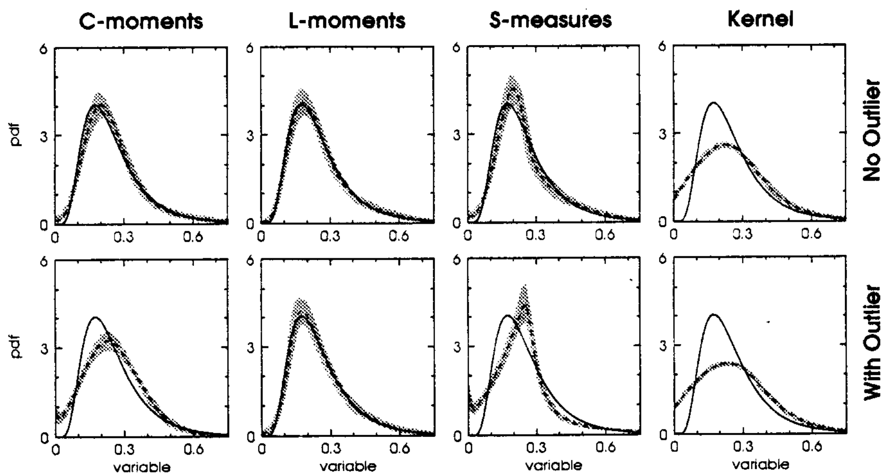

Figure 4: Log-normal pdf estimates ($n = 250$). The shaded area is limited by 10 and 90 percentiles of the results (500 realizations), while the dashed line shows the median solution and the solid line the theoretical distribution.

[6] N.L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous univariate distributions*, volume 1 & 2 of *Wiley Series in Probability and Mathematical Statistics*. John Wiley & Sons, Inc., New York, 2nd edition, 1994.

[7] J.M. Landwehr, N.C. Matalas, and J.R. Wallis. Probability-weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. *Water Resour. Res.*, 15(5):1055–1064, 1979.

[8] P. Royston. Which measures of skewness and kurtosis are best? *Statistics in Medicine*, 11:333–343, 1992.

[9] David Ruppert. What is kurtosis? An influence function approach. *The Am. Stat.*, 41(1):1–5, 1987.

[10] Edith Seier. *A family of skewness and kurtosis measures.* PhD thesis, University of Wyoming, May 1998.

[11] B.W. Silverman. *Density estimation for statistics and data analysis.* Chapman & Hall, London, 1986. 46–47.

[12] T.J. Ulrych, D.R. Velis, A.D. Woodbury, and M.D. Sacchi. L-moments and C-moments. *Stoch. Envir. Res. and Risk Assessment*, 14(1):50–68, March 2000.