Descriptive statistics for nonparametric models. The impact of some Erich Lehmann's papers

Hannu Oja

1 Introduction

In this review we discuss the six papers: Lehmann (1963) and Bickel and Lehmann (1974, 1975, 1976a, 1976b, 1976c).

The first paper deals with confidence intervals based on nonparametric tests, and the other papers discuss descriptive statistics for nonparametric models.

In an estimation problem, sample statistics $\hat{\theta}$ may often be seen as values of a functional $\theta(F)$ at the sample cumulative distribution function F_n. For example, for the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and sample variance $S^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$ the functionals are the mean functional $\mu(F) = \int x dF(x)$ and variance functional $\sigma^2(F) = \int (x - \mu(F))^2 dF(x)$, respectively. Under general assumptions on F, two functionals $\theta_1(F)$ and $\theta_2(F)$ may give the same value (e.g., mean functional and median functional under the symmetry assumption) but the statistical properties (e.g. efficiency and robustness) of the sample statistics $\hat{\theta}_1 = \theta_1(F_n)$ and $\hat{\theta}_2 = \theta_2(F_n)$ may be totally different. How should one then choose between the estimates/functionals $\theta_1(F)$ and $\theta_2(F)$? When can one be convinced that, in wide nonparametric models, the functional provides a measure of a natural aspect of the population values? These questions are posed and discussed in the series of the last five papers.

2 Nonparametric confidence intervals for a shift parameter

In this paper, Lehmann considers the two-sample location problem and gives exact expressions as well as large-sample approximations for the nonparametric confidence

H. Oja
Tampere School of Health Sciences, 33014 University of Tampere, Finland
e-mail: Hannu.Oja@uta.fi

J. Rojo (ed.), *Selected Works of E. L. Lehmann*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1412-4_40, © Springer Science+Business Media, LLC 2012

intervals for a shift parameter Δ. The confidence interval is obtained by "inverting" the two-sample Wilcoxon-Mann-Whitney test which is strictly distribution-free test for testing the null hypothesis that all the observations are i.i.d.

The paper proposes that different confidence intervals (with the same covering probabilities) should be compared using the lengths of the intervals. It is proven that, in regular cases, the length of the confidence interval standardized in an appropriate way converges in probability to a constant, and these constants can be used for asymptotic efficiency comparisons. The asymptotical efficiencies of the intervals based on Wilcoxon test and on the t test are then compared and shown to be the same as in the testing case. As a consequence of the results in the paper, a consistent estimate for the important constant $[\int f^2(x)dx]^{-1}$ is found. This estimation technique is still used in practice; see e.g. the recent book by Hettmansperger (1988).

3 Descriptive statistics for nonparametric models

3.1 The literature before Bickel and Lehmann

We first briefly discuss the literature just before the publication of these papers. von Mises (1947) may be the first one to develop a general theory of statistical functionals. Tukey (1962) discussed the properties of several location functionals in case of "spotty data". Huber (1964) gave a new approach for robust estimation, introduced the M-estimates for location and found the asymptotic theory for the estimates.

He also found the "most robust" estimate in a contaminated normal distribution. (The estimate minimizes the maximum asymptotic variance in a contaminated normal neighborhood model.) Instead of considering asymptotic variances in neighborhood models, robustness is often seen as a continuity requirement of the functional θ. Small changes in F should result in small changes in $\theta(F)$. For first alternative formulations of this requirement and their analysis, see Hampel (1971). In his review, Huber (1972) discussed an early history of robust estimation and considered three classes of location estimates in semiparamateric models, the maximum likelihood type estimates (M-estimates), linear combinations of order statistics (L-estimates), and the estimates derived from rank tests (R-estimates). In almost all studies discussed so far, the assumed model was semiparametric, with natural location parameter θ, and the location functionals $T(F)$ were used in the estimation of unknown θ.

Bickel and Lehmann then introduced the idea that one should, in wide nonparametric models, define measures (functionals) of different characteristics of the population and then use the corresponding sample statistics as estimators. The series of papers on descriptive statistics in nonparametric models together with van Zwet's (1964) work on skewness was a main source of inspiration for future research in the area.

3.2 *Statistical models*

When Bickel and Lehmann wrote the series of papers on descriptive statistics in nonparametric models, the statisticians and the practical users of statistical testing and estimation procedures did not trust any more in the dogma of normally distributed errors and often did not even believe in any parametrical statistical model.

Bickel and Lehmann discussed the model selection problem and distinguished a number of possibilities: Parametric models (e.g., normal model), neighborhood models (e.g., contaminated normal model), semiparametric models with some natural parameters (e.g. symmetry assumption), or totally nonparametric models. In the totally nonparametric model one can describe different interesting properties of the distribution F just by defining and using a corresponding measure or functional $\theta(F)$. Classical moment based measures of location, scale, skewness and kurtosis, μ, σ, β_1 and β_2 serve as examples here. Bickel and Lehmann find it convenient, however, first to try to define when a distribution G possesses the interesting attribute more strongly than a distribution F. Also, the measures should satisfy certain equivariance conditions.

3.3 *Partial orderings of distributions*

The idea to order the distributions with respect to the considered property is not very old. Mann and Whitney (1947) introduced the notion of stochastic ordering. Birnbaum (1948) proposed a dispersion ordering. The skewness and kurtosis orderings by van Zwet (1964) have proved useful and now accepted widely. The first step in the general strategy by Bickel and Lehmann is to try to define when a distribution G possesses the interesting attribute more strongly than a distribution F. This is then denoted by $F \prec G$ and it is required that a measure of this property must preserve this ordering, that is,

$$\theta(F) \le \theta(G) \text{ whenever } F \prec G.$$

The authors then introduce and discuss natural orderings for location, dispersion (symmetric models) and spread (asymmetric models). Skewness is not discussed, and measures of kurtosis are seen as ratios of two dispersion measures.

Inspired by this work Oja (1981) considers the concepts of location, scale, skewness and kurtosis of univariate distributions. For two random variables with strictly increasing cumulative distribution functions F and G, these comparisons were made using functions

$$R(x) = G^{-1}(F(x)) \text{ and } \Delta(x) = R(x) - x.$$

Doksum (1975) called $\Delta(x)$ the shift function as $X \sim F$ when shifted by $\Delta(X)$ has the distribution G, that is,

$$X \sim F \;\Rightarrow\; R(X) = X + \Delta(X) \sim G.$$

See also Doksum and Sievers (1976). It is then interesting to note that the two distributions can be naturally ordered using the function $\Delta(x)$: The orderings for location ("stochastic ordering"), spread (Bickel and Lehmann), and skewness (van Zwet) are given by

$$\Delta(x) \geq 0, \text{ for all } x, \quad \Delta'(x) \geq 0, \text{ for all } x, \quad \text{and } \Delta''(x) \geq 0, \text{ for all } x.$$

The ordering given by $\Delta^{(3)}(x) \geq 0$ may be seen as an ordering for kurtosis, and so on. Orderings of skewness are used widely also in reliability theory, see Barlow and Proschan (1975). The approach based on partial orderings have inspired statisticians for example to define and consider, also in the multivariate case, dispersion orderings (see e.g. Oja (1983), Zuo and Serfling (2000), and Romanazzi (2009), as well as kurtosis orderings (see Wang and Serfling (2005) and Wang (2009) and references therein). See also Serfling (2004).

3.4 Equivariance of the measures under linear transformations

In the following we write F_X for the cdf of X. Bickel and Lehmann required that a *measure of location* $\mu(F)$ should satisfy

1. $\mu(F) \leq \mu(G)$ whenever $F \prec_{loc} G$ and
2. $\mu(F_{aX+b}) = a\mu(F_X) + b$ for all a and b.

Similarly it was required that a *measure of spread* $\sigma(F)$ satisfies

1. $\sigma(F) \leq \sigma(G)$ whenever $F \prec_{spread} G$ and
2. $\mu(F_{aX+b}) = |a|\mu(F_X)$ for all a and b.

Bickel and Lehmann consider only location and spread (scale) measures. This is sufficient, however, as secondary properties, like skewness and kurtosis, can often be seen trough the comparisons of the behavior of different location and scale measures. Let μ_1 and μ_2 be two different location measures, and σ, σ_1 and σ_2 different scale measures. Then

$$\beta_1(F) = \frac{\mu_1(F) - \mu_2(F)}{\sigma(F)} \quad \text{and} \quad \beta_2(F) = \frac{\sigma_1(F)}{\sigma_2(F)}$$

can be used to measure skewness and kurtosis, respectively. The measures of skewness and kurtosis then naturally satisfy invariance properties

$$\beta_1(F_{aX+b}) = sgn(a)\beta_1(F_X) \quad \text{and} \quad \beta_2(F_{aX+b}) = \beta_2(F_X),$$

respectively.

One can easily find in the literature a lot of studies on location, scale, skewness and kurtosis including robust measures of skewness and kurtosis Brys *et al.* (2004) and Brys *et al.* (2006). Also equivariance properties for multivariate location vectors, scatter matrices and regression coefficient matrices are commonly required in the literature. See e.g. Maronna *et al.* (2006). Natural multivariate skewness and kurtosis measures were proposed by Mardia (1970). Nordhausen *et al.* (2011) considered multivariate skewness and kurtosis measures based on two different location vectors and two different scatter matrices.

3.5 Robustness and efficiency comparisons

How can one then compare different measures of location, dispersion and spread? Is the comparison always possible? The comparison of two measures is definitely reasonable if their values coincide under our model assumptions. Under the symmetry assumption for example, the sample mean and the sample median estimate the same population quantity and their comparison as estimates is reasonable. But what about measures of spread or measures of location for asymmetric distributions?

Bickel and Lehmann wish to find measures which are robust and at the same time could be estimated efficiently. In the location case efficiency means a guaranteed high efficiency relative to the mean (using the asymptotic variance). For dispersion and spread measures, the comparison was to the standard deviation (and the criterion was the standardized asymptotic variance). According to their general strategy, the authors wished to have a partial ordering for robustness as well: A measure μ_2 is called more robust than μ_1 if and only if μ_2 is continuous with respect to the topology induced by μ_1. Three classes of location estimates, L-estimates, R-estimates and M-estimates were compared. Trimmed standard deviations and pth power deviations were compared to the standard deviation in the dispersion case.

In the community of robust statisticians, the measures of location and spread (vectors and matrices in the multivariate case) are usually chosen to satisfy some equivariance properties, but relevant partial orderings are not discussed. The measures of spread (or scatter matrices in the multivariate case) are often rescaled so that they estimate the standard deviation (or covariance matrix) in the (multivariate) normal case. Then of course the efficiency comparisons of different estimates become reasonable in a neighborhood of a normal model. The most popular tools for robustness used today are the breakdown point, the influence function, gross-error sensitivity, asymptotic bias, etc.

4 Final comments

When Bickel and Lehmann wrote the series of papers on descriptive statistics in non-parametric models, the statisticians and the users of statistical testing and estimation procedures in practice did not trust any more in the dogma of normally distributed observations. As shown by my short review, the papers considered here had a very strong impact on future ideas and work on developing new measures and estimates for the description of population and data under general nonparametric models. Personally speaking, these papers together with van Zwet's (1964) work on skewness were the main source of inspiration in my own thesis work, Oja (1981), on the concepts of location, scale, skewness and kurtosis of the univariate distributions and also in my later work on analysis methods for multivariate data.

References

[1] Barlow, R. E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing*. Holt, Rinehart and Winston, New York.
[2] Bickel, P. J. and Lehmann, E. L. (1974). Measures of location and scale. *Proc. Prague Symposium on Asymptotic Statist,* J. Hajek, Ed., Charles University, Prague, **1**, 23–36.
[3] Bickel, P. J. and Lehmann, E. L. (1975). Descriptive statistics for nonparametric models. I. Introduction. *Ann. Statist.*, **3**, 1038–1044.
[4] Bickel, P. J. and Lehmann, E. L. (1976a). Descriptive statistics for nonparametric models II. Location. *Ann. Statist.*, **4**, 1045–1069.
[5] Bickel, P. J. and Lehmann, E. L. (1976b). Descriptive statistics for nonparametric models. III. Dispersion. *Ann. Statist.*, **4**, 1139–1158.
[6] Bickel, P. J. and Lehmann, E. L. (1976c). Descriptive statistics for nonparametric models, IV. Spread. In *Contributions to Statistics* , Jaroslav Hajek Memorial Volume, Prague.
[7] Birnbaum, Z. W. (1943). On random variables with comparable peakedness. *Ann. Math. Stat.* **19**, 76–81.
[8] Brys, G., Hubert, M. and Stryuf, A. (2004). A robust measure of skeness. *Journal of Computational and Graphical Statistics* **1**3, 996–1017.
[9] Brys, G., Hubert, M. and Stryuf, A. (2006). Robust measures of tailweight. *Computational Statistics & Data Analysis* **5**0, 733–759.
[10] Doksum, K. A. (1975). Measures of location and asymmetry. *Scand. J. Statist.* **2**, 11-22.
[11] Doksum, K. A. and Sievers, G. L. (1975). Plotting with confidence: Graphical comparisons of two populations. *Biometrika*, **6**3, 421–434.
[12] Hampel, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Stat.*, **42**, 1887–1896.
[13] Hettmansperger, T. P. and McKean, J. W. (1998). *Robust Nonparametric Statistical Methods*. Arnold, London.
[14] Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Stat.*, **35**, 73–101.
[15] Huber, P. J. (1972). Robust statistics: A review. *Ann. Math. Stat.*, **43**, 1041–1067.
[16] Huber, P. J. (1981). *Robust Statistics.* Wiley, New York.
[17] Lehmann, E. L. (1963). Nonparametric confidence intervals for a shift parameter. *Ann. Math. Stat.*, **34**, 1507–1512.
[18] Mann, H. B. and Whitney, D. R. (1947). On tests of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.*, **18**, 50–60.
[19] Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, **57**, 519–530.
[20] Maronna, R. A., Martin, R. D. and Yohai, V. Y. (2006). *Robust Statistics. Theory and Methods*. Wiley, New York.
[21] Nordhausen, K., Oja, H. and Ollila, E. (2011). Multivariate Models and the First Four Moments. In Hunter, D. R., Richards, D. S. R. and Rosenberger, J. L. (editors) "Nonparametric Statistics and Mixture Models: A Festschrift in Honor of Thomas P. Hettmansperger", 267–287, World Scientific, Singapore.
[22] Oja, H. (1981). On location, scale, skewness and kurtosis of univariate distributions. *Scandinavian Journal of Statistics*, **8**, 154–168.

[23] Oja, H. (1983). Descriptive statistics for multivariate distributions. *Stat. Prob. Letters* **1**, 327–332.
[24] Romanazzi, M. (2009). Data depth, random simplices and multivariate dispersion. *Stat. Prob. Letters* **7**9, 1473–1479.
[25] Serfling, R. (2004). Nonparametric multivariate descriptive measures based on spatial quantiles. *Journal of Statistical Planning and Inference*, bf 123, 259-278.
[26] Tukey, J. W. (1962). The future of data analysis. *Ann. Math. Stat.*, **33**, 1–67.
[27] van Zwet, W. R. (1964). *Convex transformations of random variables.* Math Centrum, Amsterdam.
[28] von Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Stat.*, **18**, 309–348.
[29] Wang, J. (2009). A family of kurtosis orderings for multivariate distributions. *Journal of Multivariate Analysis*, **100**, 509–517.
[30] Wang, J. and Serfling, R. (2005). Nonparametric multivariate kurtosis and tailweight measures. *Journal of Nonparametric Statistics*, **17**, 441–456.
[31] Zuo, Y. and Serfling, R. (2000). Nonparametric notions of multivariate scatter measure and more scattered based on statistical depth functions. *Journal of Multivariate Analysis*, **75**, 62–78.