JOURNAL DE LA SOCIÉTÉ FRANÇAISE DE STATISTIQUE

STEPHAN MORGENTHALER

Comments on "Sur une limitation très générale de la dispersion de la médiane" by M. Fréchet

Journal de la société française de statistique, tome 147, nº 2 (2006), p. 65-66.

<http://www.numdam.org/item?id=JSFS_2006__147_2_65_0>

COMMENTS ON «SUR UNE LIMITATION TRÈS GÉNÉRALE DE LA DISPERSION DE LA MÉDIANE» BY M. FRÉCHET

Stephan MORGENTHALER *

Maurice Fréchet's paper is remarkable and I congratulate the French Statistical Society for locating and reprinting it. The methods used in the paper and the ideas expressed in it anticipate developments of statistical theory that would later – sometimes very much later – be pursued in great detail.

The paper is divided into two parts. The first part contains some remarks about the state of mathematical statistics at the beginning of the second world war. Fréchet deplores the lack of logical and mathematical rigour in some of the writings of statisticians. In an oblique reference to Fisher's fiducial method he remarks that it appears to equate the probabilities of an event in two different probability spaces, which to him is inacceptable. In another critique of mathematical statisticians, he expresses his reservations about the tendency to establish the optimality of some selected method under some defined conditions and then to interpret such results as attributing exclusive validity to these methods. As an example, he cites the famous couple formed by the arithmetic mean and the standard deviation, which at the time of the reading of his paper had completely taken over.

In the second part, he goes on to discredit this type of argument by introducing the following result. If M_n denotes the median of n independent replicates of a random variable X and $\bar{X}_n$ denotes their mean, then

$$\sup_{X\sim F\in\mathcal{F}} \frac{\mathrm{IQR}_{\mathrm{asy}}(M_n)}{\mathrm{IQR}(X)} = 1.35 < \sup_{X\sim F\in\mathcal{F}} \frac{\mathrm{IQR}_{\mathrm{asy}}(\bar{X}_n)}{\mathrm{IQR}(X)} = \infty\,. \tag{1}$$

Here $\mathrm{IQR}_{\mathrm{asy}}$ denotes the asymptotic interquartile range and the set $\mathcal{F}$ of distributions of X contains all probability laws with bounded density, finite variance, and the property that the upper bound of the density equals the value of the density at the median of the law.

He contrasts this with the well-known result about the asymptotic variance of mean and median when the data are normally distributed

$$\frac{\mathrm{VAR}_{\mathrm{asy}}(\bar{X}_n)}{\mathrm{VAR}(X)} = 1 < \frac{\mathrm{VAR}_{\mathrm{asy}}(M_n)}{\mathrm{VAR}(X)} = \frac{\pi}{2}\,.$$

* Ecole Polytechnique Fédérale de Lausanne. E-mail : stephan.morgenthaler@epfl.ch

He also cites the following generalization, which is natural in the context of his previous result (1)

$$\sup_{X\sim F\in\mathcal{F}} \frac{\mathrm{VAR}_{\mathrm{asy}}(\bar{X}_n)}{\mathrm{VAR}(X)} < \sup_{X\sim F\in\mathcal{F}} \frac{\mathrm{VAR}_{\mathrm{asy}}(M_n)}{\mathrm{VAR}(X)} = 3\,. \tag{2}$$

Whereas this second inequality (2) gives a moderate to large advantage to the mean, (1) is overwhelmingly in favor of the median. Merely changing the way in which the performance of an estimator is judged, the relative merits of estimators can be turned upside down. Following the reasoning of Maurice Fréchet, it is indeed hard to understand the mathematical statistician who, arguing along the lines of (2) – or merely citing the result in the Gaussian case – goes on to deduce a kind of universal optimality of the mean.

It is a pity that some of Maurice Fréchet's statistical works have not been more widely read and absorbed at the time of their writing. In Fréchet (1958) he gently points this out himself and the present paper is another example of an oversight. I could not find a single reference to it in the mainstream literature, even though it contains important ideas. A list of these ideas together with some key works that followed in their wake is as follows :

- the use of nonparametric sets of distributions (Lehmann, 1953);
- the derivation of upper bounds when comparing the efficiency of a nonparametric method with a parametric one (Hodges & Lehmann, 1956);
- the idea of contamination and mixtures (Tukey, 1960);
- the recognition of the possibility of a discontinuity in the performance of a statistical method, or, in other words, an extreme sensitivity of the performance to small perturbations of the underlying distribution (Tukey, 1960 and Huber, 1964).

References

FRÉCHET M. (1958). Leter to the editor. *Econometrica*, 26 :590-591.

HODGES P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35 :73-101.

LEHMANN E. L. (1953). The power of rank tests. *Annals of Mathematical Statistics*, 24 :23-43.

TUKEY J. W. (1960). A survey of sampling from contaminated distributions. In Olkin I., Ghurye S. G., Hoeffing W., Madow W. G., and Mann H. B., editors, *Contributions to Probability and Statistics, Essays in Honor of Harold Hotelling.* Stanford University Press.