Rejoinder

Erich L. Lehmann*

Department of Statistics, University of California, Berkeley, CA 94720, USA

(Received 21 February 2009; final version received 21 February 2009)

I am grateful to the four discussants for their interesting and helpful comments. Their contributions have a common structure. They take cognisance of the narrow focus of my survey, and discuss various aspects and issues that I did not cover. Let me be more specific.

1. Testing for goodness (or lack) of fit

Professor Eubank points to the distinction between nonparametric and distribution-free, and to the fact that the first of these covers, for example, nonparametric function estimation. So as not to go too far afield, he focuses on a situation at the intersection of the two concepts, testing for goodness-of-fit. He considers various ways to compare different such tests. These comparisons constitute a difficult problem on which Professor Eubank states that ‘the final word has yet to be spoken’. (A discussion of some of the relevant issues can be found in Chapter 14 of Lehmann and Romano [1].)

The same testing problem is also considered in Professor Hart’s contribution where, in particular, he reviews the order selection test proposed by Eubank and Hart.

2. Full adaptation

This is another topic considered by Professor Hart. It was discussed in my survey for the case of a two-sample shift model, where the fully adaptive test does not seem very practicable. Hart treats, instead, a different setting where an adaptive rank test seems useful. It is a multiple testing situation that is of considerable practical importance.

3. Multiple comparisons

This is an extension considered by three of the discussants. I have already mentioned Professor Hart’s multiple comparisons example in the context of full adaptation.

*Email: shaffer@stat.berkeley.edu

Multiple comparisons is the central topic of the contribution by Professors Lin and Dicker. The authors discuss protein data for 1450 proteins from four obese and four lean mice, and rise the question whether Wilcoxon rank-sum statistics could be used instead of t -statistics. They enumerate some of the difficulties with such an approach. To begin with, the Bonferroni correction is useless since no p -value could ever be significant. Even the FDR is too conservative. Other difficulties are different variances and lack of independence. Although they do not mention it, a way of dealing with these issues through resampling has been proved by Westfall and Young [2]. However, they do not consider rank-based methods.

Finally, Professor Romano considers the efficiency of Wilcoxon to t in a simple multiple comparisons setting. In this situation, he finds the efficiency to be the same as it is for individual tests.

4. Infinite variance

An additional topic treated by Professor Romano is the efficiency of Wilcoxon to t when the underlying distributions have infinite variances. He shows that under fairly general conditions this efficiency is infinite.

Through their extensions, the discussants have greatly broadened the coverage of my survey and increased its utility. They have also raised many questions, which I hope others will take up.

I am most grateful to them and to the Editor, Professor Rojo, for their help.

References

- [1] E.L. Lehmann and J.P. Romano, *Testing Statistical Hypotheses*, 3rd ed., Springer, New York, 2005.
- [2] P.H. Westfall and S.S. Young, *Resampling-Based Multiple Testing*, Wiley, New York, 1993.