Additional details on functional boxplots

Ali Gharouni[1] and Benjamin M. Bolker[1,2,3]

[1]Department of Mathematics & Statistics, McMaster University, Hamilton, Canada
[2]Department of Biology, McMaster University, Hamilton, Canada
[3]Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, Canada

October 21, 2021

Juul et al. (2021) pointed out shortcomings to the standard ways that researchers draw confidence intervals for ensembles of curves, with specific examples drawn from the output of stochastic epidemic models. In particular, they showed that fixed-time approaches (e.g., computing pointwise quantiles) can fail to capture the uncertainty in key features of an epidemic such as the timing and magnitude of epidemic peaks. As an alternative to fixed-time approaches, the authors illustrated methods to compute the *central set* of an ensemble of curves, a high-dimensional analogue of interquartile range or confidence interval. There is a large body of literature on this topic under the rubrics of *functional depth* and *functional boxplots* for high dimensional data (Fraiman and Muniz, 2001; López-Pintado and Romo, 2007, 2009; Sun and Genton, 2011; Sun et al., 2012). While Juul *et al.* do cite this literature (Sun and Genton, 2011), exploring it in more depth led us to several useful practical and theoretical points that could be useful for researchers interested in using these approaches.

Juul *et al.*'s primary method for determining a curve's centrality (and hence whether it should be included in a central set of curves for display) was to measure the fraction of times that a given curve is completely included within the envelope of a set of other curves randomly sampled from the ensemble. Juul *et al.* chose the number of curves [or *tuning parameter*] as $J = 50$ (they use the notation N_{curves}), and chose $N_{\text{samples}} = 100$ such samples to compute the fraction (*band depth*) for each curve. They provided open-source Python code that implements this method, as well as some of the weighted variants they discuss. For the simple (unweighted) case, however, there are already mature open source implementations available in R (Ramsay et al., 2020; Ieva et al., 2019), Matlab (`https://www.psych.mcgill.ca/misc/fda/downloads/FDAfuns/`), and Python (Seabold and Perktold, 2010). In general these packages use the same functional band depth measure as Juul *et al.*, but substituting $J = 2$, which is robust (López-Pintado and Romo, 2009) and allows the use of a computationally efficient algorithm for large data sets (Sun et al., 2012). It is unclear why Juul *et al.* chose larger values of J (10 and 50), although the dimensions of their examples are small enough that the computational burden is not important.

Juul *et al.* also suggest ranking according to a single, one-dimensional feature of interest such as the maximum values of newly hospitalized cases in a single day (their Fig. 2e). We suggest that this approach could be extended to incorporate multiple features of interest. Functional band depth could again be used on this reduced set of features; here we use the *Mahalanobis distance* (Mahalanobis, 1936), which measures distance from a centroid accounting both for variation in the scales or typical magnitudes of different features and for correlation among features. Our example uses a feature set including the peak value of incidence (new infections), the time at which the peak occurs, and the initial growth rate, duration, and final size of the epidemic. While these are typical epidemiological features of interest, researchers can and should choose the features that are most closely connected to their particular research questions (Probert et al., 2016).

References

Fraiman, R. and Muniz, G. (2001). Trimmed means for functional data. *Test*, 10(2):419–440.

Ieva, F., Paganoni, A. M., Romo, J., and Tarabelloni, N. (2019). roahd Package: Robust Analysis of High Dimensional Data. *The R Journal*, 11(2):291–307.

Juul, J. L., Græsbøll, K., Christiansen, L. E., and Lehmann, S. (2021). Fixed-time descriptive statistics underestimate extremes of epidemic curve ensembles. *Nature Physics*, 17(1):5–8.

López-Pintado, S. and Romo, J. (2007). Depth-based inference for functional data. *Computational Statistics & Data Analysis*, 51(10):4957–4968.

López-Pintado, S. and Romo, J. (2009). On the concept of depth for functional data. *Journal of the American statistical Association*, 104(486):718–734.

Mahalanobis, P. C. (1936). On the generalized distance in statistics. National Institute of Science of India.

Probert, W. J., Shea, K., Fonnesbeck, C. J., Runge, M. C., Carpenter, T. E., Dürr, S., Garner, M. G., Harvey, N., Stevenson, M. A., Webb, C. T., et al. (2016). Decision-making for foot-and-mouth disease control: objectives matter. *Epidemics*, 15:10–19.

Ramsay, J. O., Graves, S., and Hooker, G. (2020). *fda: Functional Data Analysis*. R package version 5.1.9.

Seabold, S. and Perktold, J. (2010). statsmodels: Econometric and statistical modeling with Python. In *9th Python in Science Conference*.

Sun, Y. and Genton, M. G. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics*, 20(2):316–334.

Sun, Y., Genton, M. G., and Nychka, D. W. (2012). Exact fast computation of band depth for large functional datasets: How quickly can one million curves be ranked? *Stat*, 1(1):68–74.
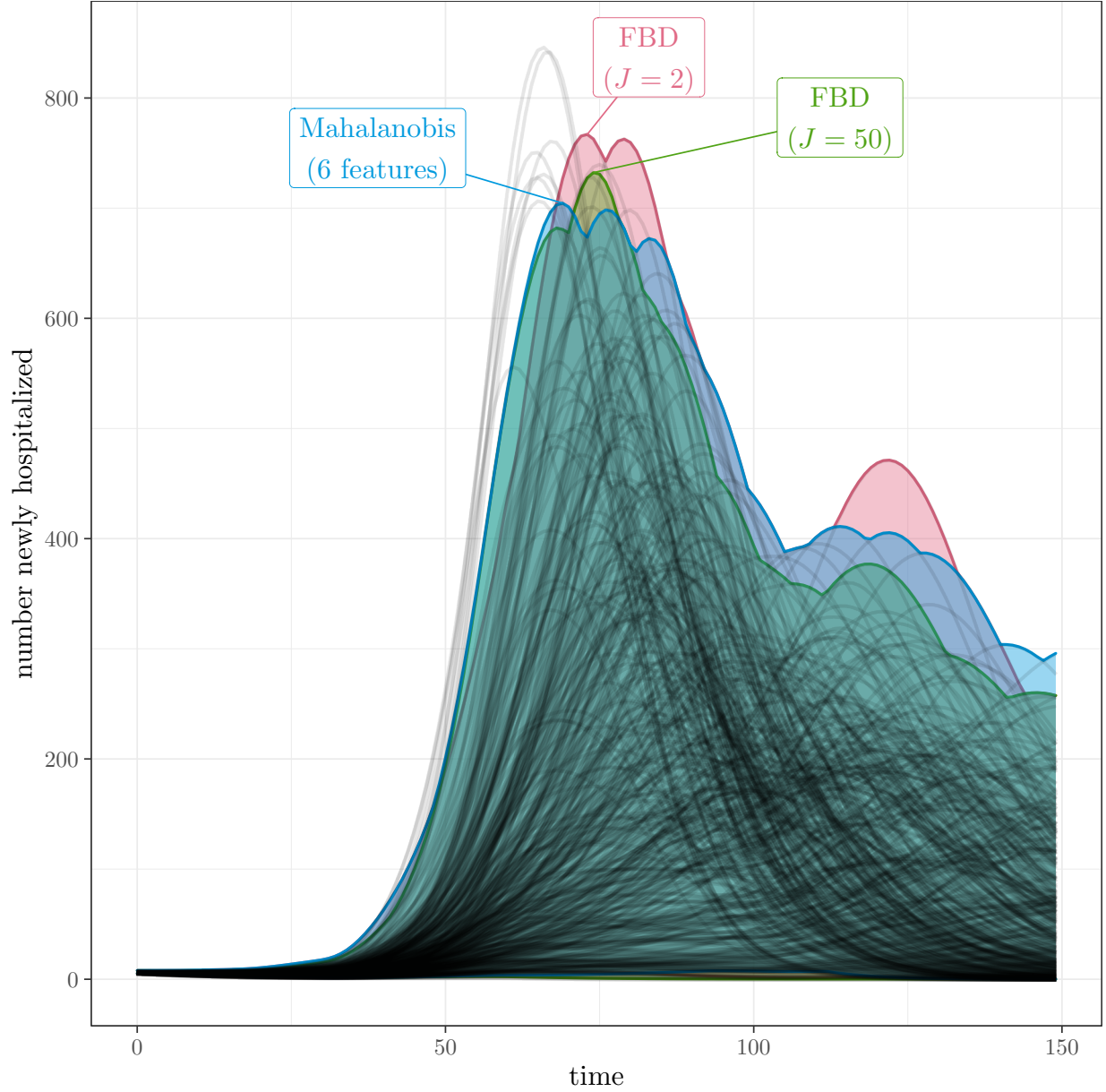
Figure 1: Comparison of 90% central regions computed with different functional boxplot methods, using epidemic curve ensembles from Juul *et al.*. FBD = functional band distance; J = number of curves used for centrality calculation. Curve with $J = 2$ computed via the `roahd` package (Ieva et al., 2019): curve with $J = 50$ used our own implementation of the functional band distance algorithm described by Juul *et al.*.