PETER BUEHLMANN, PETROS DRINEAS, MICHAEL KANE, MARK VAN DER LAAN. **Handbook of Big Data**. Boca Raton: CRC Press.

The term "Big Data" frequently makes the headlines, even in tabloids. Thus, the layman might expect that professionals in the field agree upon what they are talking about. The "Handbook of Big Data" makes it clear that this is not the case and that many facets of this new paradigm remain to be discussed and sharpened. First, there is the purely technical perspective. Using clever mathematical and computational tricks, one tweaks algorithms for the estimation of classical models, for example logistic regression, in a way that makes them applicable in "big n" (many observations) or "big p" (many variables) situations. The result, however, is still a relatively simple model with little benefit from "Big Data," except for perhaps lower variance in the "big n" world. Some chapters, such as "Big-n versus Big-p in Big Data" and a chapter on penalized estimation, adopt this perspective.

Second, one may ask "What can be done in a big data world that is impossible in a small data world?". We are, of course, interested in estimating (or "learning") more complex models in the presence of more information. The handbook devotes several chapters, for example, one on structured distributions and a series of chapters on targeted learning, to this perspective. From a methodological point of view, it is very interesting to study more complex models, as we might hope to describe the phenomena we are interested in more precisely than it was possible with low-dimensional parametric models.

The third aspect is of more philosophical nature, yet with important implications for science in general. The editors devote the prime spot in their handbook, the introductory chapter, to this topic and Richard Starmans takes the reader on a tour-de-force through epistemology and its connections to statistics. Most statisticians will probably agree that data are a footprint, left to be seen for us by the true model. Like the paleontologist, who reconstructs features of dinosaurs from their footprints and other traces, our job is, essentially, to reconstruct this hidden truth from the data. It must be feared that this makes us dinosaurs as well, because in the big data world the data are the truth. Rendering models, and therefore theories, useless has the potential to radically change the way science will be performed in the future, as Starmans explains. But maybe things are all different. The biases naturally inherent in unsystematically collected big data might spoil the party and random samples of small data are the past, present, and future of statistics, or, as John Tukey told Alfred Kinsey, the lead author of the famous "Kinsey Report": "I would trade all your 18,000 case histories for 400 in a probability sample." Even so, the "Handbook of Big Data" is the first compilation on this emerging subject in our field and is therefore highly recommended to all statisticians and computer scientists.

Torsten Hothorn
Epidemiology, Biostatistics and Prevention Institute
University of Zurich
Zurich, Switzerland
Torsten.Hothorn@uzh.ch