Journal of Informetrics 10 (2016) 1241–1242

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Correspondence

Some further aspects of sampling: Comment on Williams and Bornmann

1. Introduction

I greatly appreciate the article by Williams and Bornmann (2016), which highlights the importance of sampling in bibliometric analysis in its different facets. This contribution further develops ideas from a paper by Bornmann and Mutz (2013) to the same topic, particularly by introducing the statistical concept of power analysis as a method among others for calculating sample sizes. Although I am in full agreement with this account, I want to emphasize the importance of four specific aspects relevant to this approach: Sampling and statistical inference, sampling and bibliometric data, observed power analysis, and statistical modeling.

2. Sampling and statistical inference

Sampling is an important tool in statistics (Levy & Lemeshow, 1999), especially in classical frequentist statistics with its clear notion of parameter and of probability (Gelman et al., 2014). While a parameter is regarded as a fixed quantity or point estimate, probability is defined as a long-run relative frequency. The relative frequency of heads, for instance, obtained by tossing a fair coin an infinite number of times approaches the probability of 0.5. If the assumption of repeated samples or repeated events is justified in principle, and the sampling process is random, the estimated parameter value approximates the true parameter in the population with increasing sample size. Therefore, it is compelling to discuss sampling within the realm of classical frequentist statistics, as the authors implicitly do by introducing the more sophisticated power analysis tool discussed in the realm of null hypothesis statistical testing. In my view these ideas of repeated sampling and asymptotic behavior make sense in bibliometric analysis, as well. Another and quite different kind of statistical inference, equipped with a clear notion of probability and parameter as well, is offered by Bayesian statistics. In this regard probability is the degree of subjective belief about a parameter value. A parameter is not a fixed value, but a random variable following a pre-defined statistical distribution, which expresses the uncertainty of the parameter. Bayesian parameter estimation is more or less a learning process in the sense that a prior belief of a parameter is updated by the data (likelihood) to obtain a posterior belief of a parameter. Bayesian statistics is, therefore, not restricted, as frequentist statistics, to random sampling, but can be applied to populations or to arbitrarily drawn samples as well (Berk, Western, & Weiss, 1995, p. 433f., also mentioned by Williams & Bornmann, 2016). Then again, sampling is not restricted to a certain kind of statistical inference.

3. Sampling and bibliometric data

Sampling in bibliometric analysis should take into account the special nature of bibliometric data. Publications are not per se independent units, but interwoven, for instance, with bibliographic coupling. If two publications cite a common third publication, bibliographic coupling occurs. The more references two publications share, the more similar the two publications are (e.g. field, content), probably with respect to their bibliometric data, as well. Doubling of publications in data analysis due to co-authorship might provoke further dependency in measurements. Indeed, simulation studies are required to assess the statistical behavior of tests and parameter estimations (e.g. bias in mean, bias in standard error) under certain sampling conditions (e.g. varying percentages of doublets of publications in the data). Williams and Bornmann recommend bootstrapping as a tool to estimate the standard error. In addition, bibliometric data are characterized by different sections or strata, for example, scientific fields. Within a scientific field bibliometric data (e.g. citations) are assumed to be distributed more homogeneously than between fields. This heterogeneity in the population requires samples not to be drawn by a simple random sampling but rather by a stratified sampling procedure, which takes into account the different strata within the population.

http://dx.doi.org/10.1016/j.joi.2016.09.007

4. Observed power analysis

The well-known statistical procedure of power analysis was provided by Williams and Bornmann for prospectively planning the sample size of a bibliometric study. This procedure reflects a more sophisticated treatment of null hypothesis significance testing (NHST), which is usually approached in empirical studies by simply flagging an empirical result with a star. Often observed power analysis is recommended by journals and reviewers in order to estimate the power of empirical tests in already published studies. From a statistical point of view the post hoc calculated power does not provide any useful information and should be avoided (Sun, Pan, & Wang, 2011; Yuan & Maxwell, 2005). "…[T]he present study demonstrates that observed power is usually not as informative or helpful as we think because (a) observed power for a nonsignificant test is generally low and, therefore, does not provide additional information to the test; and (b) a low observed power does not always indicate that the test is underpowered." (Yuan & Maxwell, 2011, p. 81). Even if an observed power analysis is not fruitful, a statistically non-significant test result, under the condition of high sample size and high power, most likely indicates no effect at all.

5. Statistical modeling

Last but not least, a statistical modeling approach appears to be more appealing than classical standard tests in bibliometric analysis, which might capture the various statistical problems (e.g. measurement dependencies, Mutz & Daniel, 2015). Simple statistical tests could be replaced by model comparison tests with information criteria (e.g. Akaike's Information Criterion) or Log-likelihood Ratio Tests (e.g. Bornmann, Mutz, & Daniel, 2013). In this regard sampling could be important as well. The use of bibliometric data for comparison purposes and (world-wide) rankings (e.g. Leiden Ranking), respectively, or for the analysis of research networks requires huge and complex structured data. In light of these enormous data statistical computation could easily run out of memory. By reducing the sample size, sampling could facilitate more complex statistical modeling (be it Bayesian or frequentist) without significant loss of estimation accuracy.

6. Conclusion

Beyond question sampling is an essential tool in statistics. It is the merit of Williams and Bornmann to have posed sampling issues in bibliometrics. However, further empirical and simulation studies are required, which should address the special properties of bibliometric data (e.g. bibliographic coupling, doubling of publications) and their impact on the sampling and statistical estimation process.

References

Berk, R., Western, B., & Weiss, R. E. (1995). Statistical inference for apparent populations. *Sociological Methodology*, *25*, 421–458.

Bornmann, L., & Mutz, R. (2013). The advantage of the use of samples in evaluative bibliometric studies. *Journal of Informetrics*, *7*(1), 89–90.

Bornmann, L., Mutz, R., & Daniel, H.-D. (2013). A multilevel-statistical reformulation of citation-based university rankings: the Leiden Ranking 2011/2012. *Journal of the American Society for Information Science and Technology*, *64*, 1649–1658.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis*. London: Chapman & Hall.

Levy, P. S., & Lemeshow, S. (1999). *Sampling of populations. Methods and applications*. New York: Wiley.

Mutz, R., & Daniel, H.-D. (2015). What is behind the curtain of the Leiden Ranking? *Journal of the American Society for Information Science and Technology*, *66*, 1950–1953.

Sun, S., Pan, W., & Wang, L. L. (2011). Rethinking observed power. Concept, practice, and implications. *Methodology*, *7*(3), 81–87.

Williams, R., & Bornmann, L. (2016). Sampling issues in bibliometric analysis. *Journal of Informetrics*.

Yuan, K.-H., & Maxwell, S. (2005). On the post hoc power in testing mean differences. *Journal of Educational and Behavioral Statistics*, *30*(2), 141–167.

Rüdiger Mutz
ETH Zurich, Professorship for Social Psychology and Research on Higher Education, Muehlegasse 21, CH-8001 Zurich, Switzerland
E-mail address: mutz@gess.ethz.ch

Available online 15 October 2016