Published in final edited form as:

Biometrics. 2020 June ; 76(2): 599–601. doi:10.1111/biom.13254.

Discussion on “Predictively consistent prior effective sample sizes,” by Beat Neuenschwander, Sebastian Weber, Heinz Schmidli, and Anthony O’Hagan

Gary L. Rosner,

Division of Biostatistics & Bioinformatics, Department of Oncology, Johns Hopkins University School of Medicine, Baltimore, Maryland, U.S.A.

Peter Müller

Department of Statistics & Data Science, University of Texas, Austin, Texas, U.S.A.

Summary:

Neuenschwander et al. (2020) address a seemingly easy but often complicated problem in applied Bayesian methodology. We discuss some issues that relate to the question of why one might care about the effective sample size (*ESS*) in a Bayesian model and the motivation for reporting the *ESS*.

We thank the authors for their interesting and useful contribution to applied Bayesian methodology. We would like to focus our discussion on the question, “Why do we care about estimating effective sample size (*ESS*)?”

There are several reasons one might be interested in determining *ESS* in a study.

- Sample size is an easy-to-understand quantity outside of statistics.
- Precision of an estimator is often proportional to the actual number of observations in a sample.
- When designing a study, the sample size is a primary concern.
- In many simple conjugate exponential models that are in common use, there is a nice and intuitive relationship between the prior distribution’s model parameters (or a function of them) and the sample size.

The authors propose an abstraction of the relationships between a prior model’s scalar parameter and the sample size in several well-known cases. They focus on the ratio of the expected variance of an estimator in a sample of size N to the variance of the parameter. Specifically, they form an *ESS* estimator based on the expected local-information ratio. This estimator (ESS_{ELIR}) is the expected ratio of the prior information to the expected Fisher information for a single unit of information.

The ESS_{ELIR} has some appealing properties. One property is that it reduces to the well-known ESS estimators in the common conjugate cases. A second property is that it satisfies a criterion the authors call *predictive consistency*. An effective sample size estimator ESS^* will satisfy this criterion if the difference between the posterior ESS^* and the sample size N is the prior ESS^* .

The paper includes many non-conjugate examples with single-parameter models to illustrate ESS_{ELIR} . The authors refer readers to other papers for ESS proposals that scale beyond one-dimensional parameters, stating that “Finding predictively consistent ESS for such [multi-parameter] cases requires further research.” Thus, we might infer that the requirement of predictive consistency causes some of the difficulty scaling to models with parameters in two or more dimension.

We would like to focus our attention on measures of information instead of sample size. When carrying out experiments, one hopes to generate information that one can use for inference. From the Bayesian standpoint, the information in an experiment should increase our knowledge about some quantity or phenomenon when we combine that information with what we already know outside of the experiment (i.e., prior information). Appeals to ESS are appeals to replace a fairly coherent but perhaps esoteric (or technical) concept in mathematical statistics with a concept that may have broader appeal, being analogous to the vernacular use of the term “sample size.”

One may argue, however, that we statisticians—especially Bayesian statisticians—should be trying to educate our scientist collaborators to other concepts beyond sample size. That is, we should be speaking with our colleagues in terms of the amount of information in a planned or already conducted experiment. Although planning a study will have to include determining the required number of subjects and/or observations to achieve the study’s goals, collaborating statisticians too often find their roles reduced to providing a simple number of subjects to enroll. This attitude toward collaborating statisticians has probably not had a good effect on the scientific enterprise. Especially among studies with frequentist designs, the sample size reflects the desire to achieve statistical significance, regardless of the scientific value or clinical significance of the outcome of the study. As we know, any difference between two groups may be found to be statistically significant at some level if the sample size is large enough. This is one of the reasons for a methodological crisis known as the replication crisis. See, for example, Held (2020) for a discussion and a proposed solution.

If, however, one appeals to information (or the related concept of entropy), then one can state the goal of the study in terms of increasing knowledge, rather than achieving statistical significance. The difference is particularly meaningful in early studies. In clinical and translational research, scientists often carry out pilot studies or proof-of-concept studies either to gather preliminary data or to generate further hypotheses. As pointed out by Piantadosi (2005), evaluating such early studies based on changes in information (or entropy) may well be more appropriate than appealing to concepts of sample size and the latter’s common reference to statistical significance and power. Entropy may also provide a useful criterion for decision-theoretic study designs, as illustrated by Ventz et al. (2019)

in their proposal for Bayesian uncertainty-directed trial designs. Furthermore, information measures extend to multi-parameter settings.

The increase in information about a quantity of subject-matter importance is the goal of experimentation, whether the quantity is a model parameter or a function of model parameters. While translating change in information to a quantity that one can think of as a sample size may be appealing to some, the truly relevant quantity is the information in the experiment. The authors have provided a new conception of what an effective sample size means, what properties it should possess, and how we statisticians might measure it. We do, however, caution against making the *ESS* the ultimate aim, rather than a convenient transformation of the actual measure of importance, namely, the information in the experiment.

Closely related to the why and the motivation for reporting an *ESS* is another choice one may make. The authors develop the proposed expected local ESS_{ELIR} for a univariate parameter. In both examples in the paper, this focus is entirely appropriate, in part because of the motivation for reporting an *ESS*, as implied in the discussion of the examples. If, in contrast, the aim were to justify an informative prior in a clinical trial design, one might be more concerned with the question of the *ESS* for the entire multivariate prior model—and this focus could significantly complicate the discussion. Consider, for example, a normal linear regression model (e.g., example 9 in Morita et al., 2008). While an *ESS* for each factor in a typical prior definition is easy to consider, *ESS* for the joint prior on regression parameters and residual variance is actually not obvious. Of course, the definition of ESS_{ELIR} could easily be extended to multivariate priors using a suitable scalar summary of the information matrices in the paper's equation (7) that defines ESS_{ELIR} .

Another important issue related to the why of *ESS* evaluation is well illustrated in the example in Section 3.2 of the paper. In the hierarchical beta-binomial model, like in any other hierarchical model, one could focus on several different equivalent statements of the model. In particular, one could ask for a prior *ESS* for the prior distribution of the hyperparameters, defining the rest of the model as the sampling model after marginalizing with respect to the next-level parameters. Alternatively one could focus on the implied joint prior on second-level parameters, after marginalizing with respect to common hyperparameters. In the case of the sarcoma trial in Section 3.2, a discussion of *ESS* for the same trial (using a slightly different model) reports an *ESS* of 2.6 and 3.7, respectively, depending on the why (Morita et al., 2012). In Section 3.2. of the present paper, the authors discuss yet another interesting *ESS* computation for the same model. The difference, again, rests on the why.

In summary, it is remarkable how many different perspectives one could have in answering the seemingly easy and basic question of reporting an equivalent sample size for informative priors. We thank the authors for a stimulating and very concise discussion of the related issues, including the excellent and very enlightening summary of previous literature.

Acknowledgements

GLR acknowledges partial support from grant P30 CA006973 from the U.S. National Cancer Institute. PM acknowledges partial support from grant R01 CA132897 from the U.S. National Cancer Institute.

References

- Held L (2020). A new standard for the analysis and design of replication studies. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* in press,.
- Morita S, Thall PF, and Mueller P (2008). Determining the effective sample size of a parametric prior. *Biometrics* 64, 595–602. [PubMed: 17764481]
- Morita S, Thall PF, and Mueller P (2012). Prior effective sample size in conditionally independent hierarchical models. *Bayesian Analysis* 7, 591–614.
- Piantadosi S (2005). Translational clinical trials: An entropy-based approach to sample size. *Clinical Trials* 2, 182–192. [PubMed: 16279140]
- Ventz S, Cellamare M, Bacallado S, and Trippa L (2019). Bayesian uncertainty directed trial designs. *Journal of the American Statistical Association* 114, 962–974.