Peerage of Science
Peer review, for the Peers, by the Peers

Review by Peer 195 on manuscript:

[redacted]

Revision recommendation: Withdraw

ADDED INFO ABOUT PUBLICLY FEATURED REVIEW

Author of this peer review is

a post-doctoral researcher at a research institute in France.

PEQ = 4.1 / 5

Peer reviewed by 5 Peers.

Introduction

This paper [1] endeavours to detect [redacted]. While I found the title and topic of the paper appealing, I must confess disappointment both because I feel the method is unclear, and the only graphic of the paper is poor. Statistical graphics, what Edward Tufte calls 'Beautiful Evidence' [2] can be very powerful and illuminating. The current manuscript need some major revisions.

Merits score: 3.0 / 5

The main merit of [1] is to try to propose a [redacted].

Critique score: 4.8 / 5

1.1 Assumptions of the Bootstrap

[1] states (page 9): "there is no need for computationally intensive processes, such as bootstrapping, to make assumptions about the data". I disagree. Since the bootstrap is sampling with replacement [3], it implicitly assumes that all the potential values of the data have been observed: "this method of generating the sampling distribution of a statistic essentially assumes that the sample cdf is the population cdf" [4]. In other words, every [redacted] where a [redacted] is in the sample! This is a rather strong assumption: it basically assumes that detection probability of [redacted] is perfect, and the sample we have is not a convenience sample. While the detection probability of [redacted] may be close to one in [redacted], this may not be the case elsewhere. This strong assumption should be discussed more. Currently, detection probability is mentioned briefly in the introduction and quickly dismissed on page 8, with the premise that the authors' approach is freed from any assumptions, when in fact it makes a strong one. There are always data generating

mechanisms, some more complicated than others. What the bootstrap circumvents is to make assumptions about the sampling distribution of a statistic, not to make assumptions about the data.

1.2 Bootstrapping and Posterior Inference

The bootstrap as presented in [1] is confusing. I think that the authors can't save a bit of notation to clarify their method. Please see attached pdf for optimal display of the equations.

1.3 False Absences and Detection Probability

The authors are discussing false absences, and linking it correctly to detection probability. Then they conclude that the latter is no issue with their method because they randomly subsampled the data. There is some confusion about what the bootstrap does (see attached pdf), but it seems strange to conclude of robustness when the assumption of random sampling has never been tested properly and was assumed throughout. On page 6, I don't think that the authors have "simulated [...] false absences". They have merely sub-sampled the observed data. Moreover, sampling is typically not random with presence-only data! Any sub-sampling procedures should try to mimic this in a realistic way. This issue is linked to that of detection probability, which is not discussed enough. See for example [7,8,9], for a discussion of the analysis of presence-only data. I don't think the proposed methodology is addressing this issue properly.

1.4 [redacted] methods

The only figure is rather disappointing. I think the proposed [redacted] is too conservative. In the univariate case, a rule of thumb to assess statistical significance between two independent means is that their error bars should not overlap more than half the length of one arm [10]. Interesting work has been done in the univariate case, and additional refinement for multivariate data would be much welcome! This is a line of investigation that would be very useful, and in my opinion would correspond to the title of the paper. As it stands, the proposed [redacted] is in all likelihood overly conservative. This is no small defect in the context of [redacted] !

2.1 Running title:

I'm afraid I find the running title meaningless. Statistical significance is the probability of finding a statistics as least as extreme as the observed one under the null. What is the null hypothesis for [redacted]?

2.2 page 3: what is a "poor dataset"?

2.3 page 9: "Furthermore, knowing that a change in [redacted] is statistically significant is a vital first step before trying, erroneously, to imply causation to a pattern that is actually non-significant". This ignores the literature about the shortcomings of statistical significance [11,12,13,14,15]. The authors seems to succumb to a fallacy; to paraphrase [16] "a lack of statistical significance - statistical insignificance - is easily though often mistaken to show a lack of cause and effect when in fact there is one". I can't disagree more with [1] on this sentence. I fail to see any vital point in establishing statistical significance. Practical significance, what are the ecological consequences of [redacted], is the crucial point; and not, to paraphrase Sir Harold Jeffreys, "that a hypothesis that may be true may be rejected because it has not predicted observable results that have not occurred".

Discussion score: 4.5 / 5

I think the authors should change the focus of their analysis to directly simulate the sampling distribution of the difference between the CoG observed two decades apart. That is, the parameter of interest should be directly: Delta(CoG) = CoG_obs,2 - CoG_obs,1. The bootstrap could be used to simulate the sampling distribution of this statistics, and then to compare how extreme is the observed value. Weights informed by how the data were collected could be incorporated in the analysis. In my opinion, more, not less, assumptions would strengthen the analysis.

As it stands, the paper is confusing. I think that the authors should provide for example R code to make their research reproducible. Finally, one could also use a bivariate normal distribution to estimate CoG and use the existing R package ellipse to directly compute confidence ellipse. This approach may also enable to study more ecological relevant aspect of [redacted] such as [redacted]. An abstract construct like the CoG is likely to lie in the middle of [redacted] for [redacted], which is unfortunate. A contour of land masses would have been nice on the figure. I very much encourage the authors to further develop [redacted] to assess statistical significance.

References

[1] Anonymous authors (2012) A [redacted] (unpublished manuscript) - Peerage of Science

[2] Tufte, E. (2006) Beautiful Evidence. 1st edition. Graphic Press.

[3] Efron, B. (1979) Bootstrap Methods: Another Look at the Jackknife. Annals of Statistics 7:1-26.

[4] Rubin, D. (1981) The Bayesian Bootstrap. The Annals of Statistics 9:130-134.

[5] Efron, B. (2011a) Bayesian Inference and the Parametric Bootstrap. Technical report, Stanford University.

[6] Efron, B. (2011b) The Bootstrap and Markov chain Monte Carlo. Journal of Biopharmaceutical Statistics 21:1052-1062.

[7] Royle, J., and R. Dorazio. (2006) Hierarchical Models of Animal Abundance and Occurrence. Journal of Agricultural, Biological, and Environmental Statistics 11:249-263.

[8] Royle, J., and R. Dorazio. (2008) Hierarchical Modeling and Inference in Ecology. 1st edition. Academic Press.

[9] Royle, J., R. Chandler, C. Yackulic, and J. Nichols. (2012) Likelihood Analysis of Species Occurrence Probability from Presence-Only Data for Modelling Species Distributions. Methods in Ecology and Evolution (early view)

[10] Cumming, G. (2009) Inference by Eye: Reading the Overlap of Independent Confidence Intervals. Statistics in Medicine 28:205-220.

[11] Yoccoz, N. (1992) Use, Overuse, and Misuse of Significance Tests in Evolutionary Biology and Ecology. Bulletin of the Ecological Society of America 72:106-111.

[12] Burnham, K., and D. Anderson. (2002) Model Selection and Multimodel Inference. A Practical Information-Theoretic Approach. 2nd edition. Springer.

[13] Stephens, P., S. Buskirk, and C. Martinez del Rio. (2007) Inference in Ecology and Evolution. Trends in Ecology and Evolution 22:192-197.

[14] Whittingham, M., P. Stephens, R. Bradbury, and R. Freckleton. (2006) Why Do We Still Use Stepwise Modelling in Ecology and Behaviour? Journal of Animal Ecology 75:1182-1189.

[15] Martinez Abrain, A. (2010) Statistical Significance and Biological Relevance: a Call for a more Cautious Interpretation of Results in Ecology. Acta Oecologica 34:9-11.

[16] Ziliak, S. (2011) Matrixx vs. Siracusano and Student vs. Fisher. Significance 8:133-134.