The statistical power of replications in difference tests

Per Bruun Brockhoff

Department of Mathematics and Physics, The Royal Veterinary and Agricultural University, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark

It has been argued that the binomial test with N=nk observations is a valid test when n assessors each perform k replicated difference tests, for instance triangular tests, see Kunert and Meyners (1999). However, the statistical power of this test is unknown when k>1. The objective of this work is to fill in this gap.

Various models and approaches to account for the replications have been suggested in the sensory literature. Ennis and Bi (1998) (and other papers by these authors) recommends the beta-binomial model, Hunter *et al.* (2000) recommends generalized linear models and Brockhoff and Schlich (1998) recommends an adjusted over-dispersion approach. A problem of the latter is the lack of a formal model. It is argued in this paper that a problem of a straightforward use of the two former is the lack of a formal handling of the fact that individual correct answer probabilities cannot go below a certain limit c (c=1/3 for the triangle test). In this work, corrected versions of the beta-binomial model and the generalized linear mixed model are suggested that do take this properly into account. In Kunert and Meyners (1999) a binomial mixture model was suggested that also formally handled the problem, but the approach was not investigated in much detail. In the present paper maximum likelihood methods are applied.

All these different approaches are compared theoretically and by applications on real data. A striking result is the similarity between beta-binomial models and generalized linear models: The beta-binomial model assumes that the true individual correct answer probabilities follow a beta-distribution. For the generalized linear model the corresponding density is deduced, and despite the apparent difference between the mathematical formulae, plots of the densities show that there is hardly any difference at all between the models induced by the two approaches, see Figure 1.

It is shown how the statistical power of the binomial test can easily be computed for the various approaches using Monte Carlo methods and standard software. These power calculations show little difference between the three main approaches: beta-binomial models, generalized linear models and binomial mixture models. They also together with the theoretical comparison show how the simple extreme version of the binomial mixture model can be seen as the common extreme case for all three approaches. This common extreme case scenario corresponds to the situation where each individual is assumed to be either a discriminator (having probability one of correct answer) or a non-discriminator (having probability c of correct answer). Although this is not the proper description of the data generating process it does provide a lower limit of power for a given combination of n and k. Tables of these limit of power is provided for combinations of n=5,..,50 and k=1,..,5.

It is shown how this lower limit is high enough to be of practical importance. For instance with n=12 assessors and k=4 replications for each assessor the power of the 0.05-level binomial test with N=48 for an effect size of 25% above chance is 77%. For the extreme case the (lower limit) power is 69%, hence only a moderate loss of power is seen. The power of

69% for a 0.05-level test is obtained with around 40-42 observations, so 12 assessors and 4 replicates pr. assessor gives at least the same power as around 40-42 assessors. This gives a strong and direct tool for the sensory practitioner to control the power of the standard binomial test in case of replications and more specifically to increase the power by introducing replications in experiments.

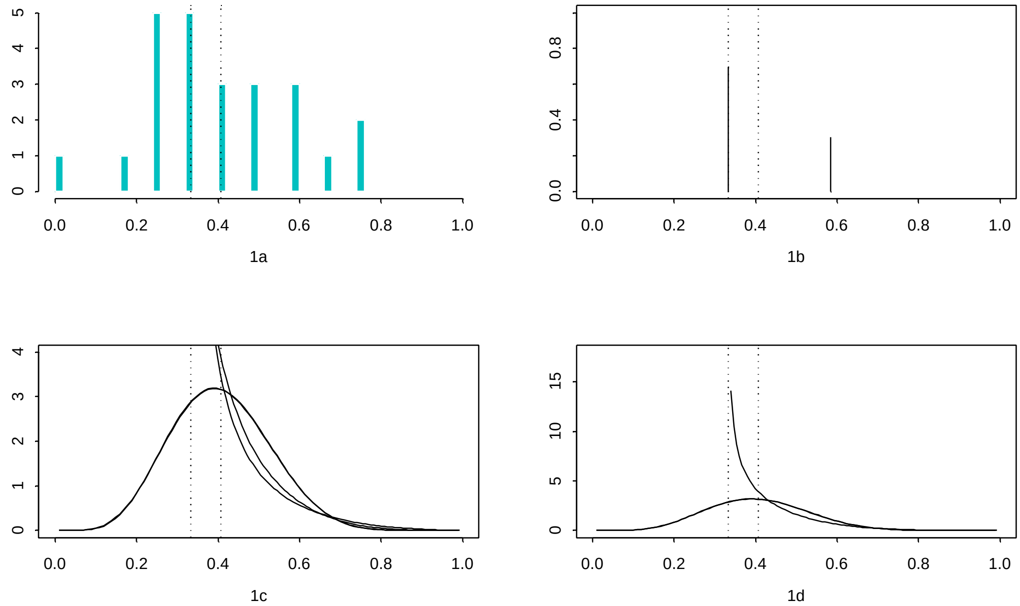

Figure 1. a: Observed individual relative frequencies of correct answers (1a) for the triangle test data from Experiment 2 in Hunter *et al.* (2000), 24 assessors, 12 triangle tests for each assessor. **b:** The estimated proportions of the binomial mixture model.
c and d: Fitted densities for the straightforward use of the beta-binomial and the generalized linear mixed model (the two bell-shaped densities) and the two re-scaled approaches(the two skew densities). 1c and 1d show the same four densities on two different scales. The leftmost vertical dotted line shows the chance level c=1/3, the rightmost dotted vertical line shows the overall relative frequency of correct answers 0.4062.

References

- Brockhoff, P.B. and Schlich, P. (1998). Handling replications in discrimination tests. *Food Quality and Preference*, **9**(5), 303-312.
- Ennis D.M. and Bi J. (1998). The Beta-Binomial model: Accounting for inter-trial variation in replicated difference and preference tests: *Journal of Sensory Studies*, **13**: 3819-412.
- Hunter, E.A., Piggot, J.R. and Lee, K.Y.M. (2000). Analysis of discrimination tests. *In Proceedings: 6th Conference on Food Industry and Statistics*, p. 9-9_8, January 19-21, Pau, France.
- Kunert J. and Meyners, M. (1999). On the triangle test with replications. *Food Quality and Preference*, **10** (6): 477-482.