



Good things peak in pairs: a note on the bimodality coefficient

Roland Pfister^{1*}, Katharina A. Schwarz², Markus Janczyk¹, Rick Dale³ and Jonathan B. Freeman⁴

¹ Department of Psychology III, Institute of Psychology, Julius Maximilians University of Würzburg, Würzburg, Germany

² University Medical Center Hamburg-Eppendorf, Hamburg, Germany

³ Cognitive and Information Sciences, University of California, Merced, Merced, CA, USA

⁴ Department of Psychological & Brain Sciences, Dartmouth College, Hanover, NH, USA

*Correspondence: roland.pfister@psychologie.uni-wuerzburg.de

Edited by:

Holmes Finch, Ball State University, USA

Keywords: distribution analysis, bimodality

A commentary on

Assessing bimodality to detect the presence of a dual cognitive process by Freeman, J. B., and Dale, R. (2013). *Behav. Res. Methods* 45, 83–97. doi: 10.3758/s13428-012-0225-x

DISTRIBUTION ANALYSES AND BIMODALITY

Distribution analyses are becoming increasingly popular in the psychological literature as they promise invaluable information about hidden cognitive processes (e.g., Ratcliff and Rouder, 1998; Ratcliff et al., 1999; Wagenmakers et al., 2005; Miller, 2006; Freeman and Dale, 2013). One particular approach probes distributions for uni- vs. bi-modality, because bimodality often results from the contribution of dual processes underlying the observed data (Larkin, 1979; Freeman and Dale, 2013; see Knapp, 2007, for a historical overview). Although several statistical tools for this purpose exist, it remains unclear which one can be considered as a gold standard for assessing bimodality in practice.

Freeman and Dale (2013) have recently shed some light on the utility of three different measures of bimodality known as the *bimodality coefficient* (BC; SAS Institute Inc, 1990), *Hartigan's dip statistic* (HDS; Hartigan and Hartigan, 1985), and *Akaike's information criterion* (AIC; Akaike, 1974) as applied to one-component and two-component Gaussian mixture distribution models (McLachlan and Peel, 2000). Overall, their analyses favored the HDS but also credited the BC with considerable utility. Notably, however, rather different formulas for the BC can be found in the literature (SAS

Institute Inc, 1990, 2012; Knapp, 2007; Bimodal distribution, 2013; Freeman and Dale, 2013)—certainly a potential source of confusion among researchers using the BC.¹ Additionally, the Appendix of Freeman and Dale (2013) gives a slightly ambiguous formula for the BC because their approach used non-standard MATLAB functions that are not widely accessible. The present article aims at clarifying and correcting these issues in an attempt to prevent misunderstanding and confusion. Further, methodological issues in using this measure are sketched to provide an intuition about its behavior. Note that the current paper does not intend to argue in favor of the BC as compared to other measures (see Freeman and Dale, 2013, for a thorough comparison). Rather, we want to point out pitfalls and limitations of this measure that can easily be overlooked.

THE BC AND ITS CAVEATS

The computation of the BC is easy and straightforward as it only requires three numbers: the sample size n , the skewness of the distribution of interest, and its excess kurtosis² (see DeCarlo, 1997, and Joanes and Gill, 1998, for a detailed description of the latter two statistics). First appearing as part of the SAS procedure CLUSTER under the headline

¹The corresponding Wikipedia article (Bimodal distribution, 2013) used a wrong formula throughout, but has been corrected as part of preparing this article.

²Excess kurtosis and Pearson's original kurtosis differ only as to whether the distribution's fourth scaled moment is normalised to a value of 0 for normal distributions or not (with "excess" indicating that a value of three has been subtracted for normalisation). The present article assumes all statistics to represent excess kurtosis if not explicitly indicated otherwise.

"Miscellaneous Formulas" of the SAS User's Guide (SAS Institute Inc, 1990, p. 561), the original formulation of the BC is

$$BC = \frac{m_3^2 + 1}{m_4 + 3 \cdot \frac{(n-1)^2}{(n-2)(n-3)}}$$

with m_3 referring to the skewness of the distribution and m_4 referring to its excess kurtosis (see Knapp, 2007, for critical remarks about this notation), with both moments being corrected for sample bias (cf. Joanes and Gill, 1998). The BC of a given empirical distribution is then compared to a benchmark value of $BC_{crit} = 5/9 \approx 0.555$ that would be expected for a uniform distribution; higher numbers point toward bimodality whereas lower numbers point toward unimodality.

Freeman and Dale (2013) gave information about computation of the BC with Matlab, but unfortunately two problems likely arise from using their code (for more information and examples of calculation with different software packages, see the online material): First, the call

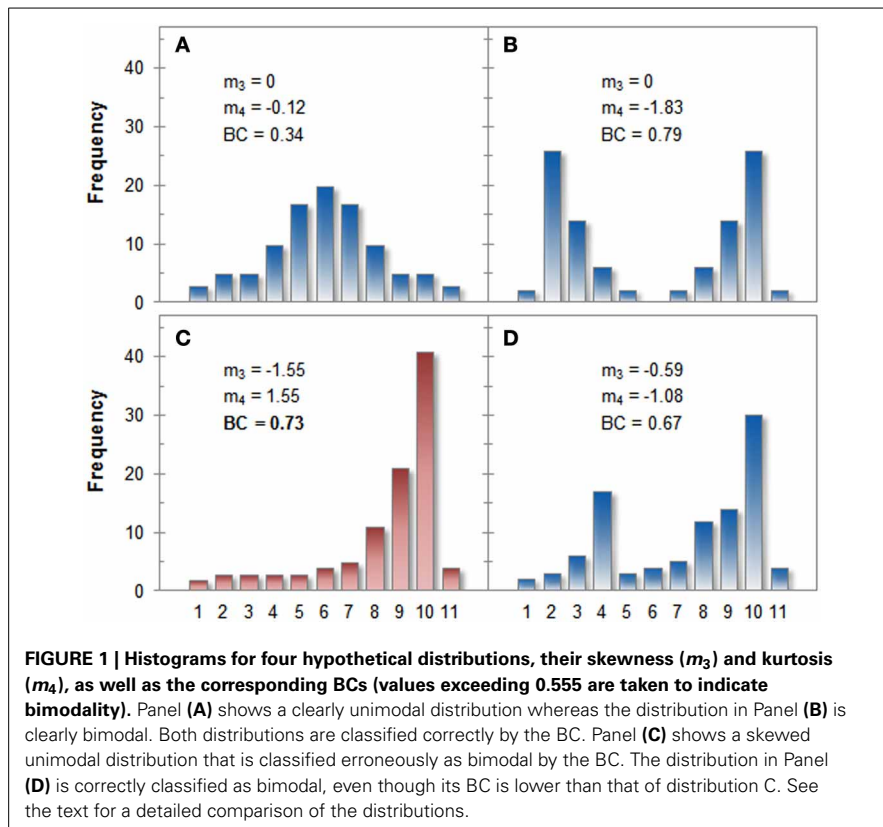
$$m_3 = \text{skew}(x);$$

likely results in an error, as `skew()` is not a native Matlab function. The correct call should be

$$m_3 = \text{skewness}(x, 0);$$

where the second input parameter 0 prompts the necessary correction for sample bias. Secondly, `kurtosis()` computes Pearson's original kurtosis (The MathWorks Inc., 2012). To obtain the correct and sample-bias corrected value, the call should be

$$m_4 = \text{kurtosis}(x, 0) - 3;$$



Irrespective of these computational issues, the above-mentioned formula reveals that the BC is directly influenced by both, skewness and kurtosis: Higher BCs result from high absolute values of skewness and low or negative values of kurtosis. Especially the influence of skewness can result in undesired behavior of the BC. As an illustration, four hypothetical distributions of 100 values each (range 1–11) are plotted in **Figure 1**, including their skewness, kurtosis, and the resulting BC (see Appendix for the raw data).

Comparing distribution A and B reveals the expected behavior of the BC: The two obvious modes in distribution B decrease kurtosis and increase the BC. Distribution C, however, is clearly unimodal when inspected by eye but its heavy skew also leads to a large BC. In terms of the BC, distribution C is even more bimodal than distribution D even though distribution D clearly has two modes, but otherwise both are very similar. The additional mode, however, decreases skewness thereby lowering the BC as long as it is not compensated by (negative) kurtosis.

CONCLUSIONS

As described above, empirical values of $BC > 0.555$ are taken to indicate bimodality. A probability density function for the BC, however, cannot be derived (Knapp, 2007). This is a major drawback because it precludes a thorough null-hypothesis significance test.

A suitable alternative test for bimodality is the *dip test* (Hartigan and Hartigan, 1985) that probes for deviations from unimodality (see also Freeman and Dale, 2013, for a more detailed description). An algorithm for this test was proposed after its publication (Hartigan, 1985) and this algorithm has meanwhile been adopted for MATLAB (Mechler, 2002). Additionally, an up-to-date, bug-corrected version was recently published as an R package (*diptest*, Maechler, 2012).

A direct comparison of the BC and the *dip test* (Freeman and Dale, 2013) revealed that both measures have merit for assessing bimodality but neither statistic is perfectly sensitive and specific at the same time. Accordingly, one may assess empirical distributions with both measures and diagnose bimodality especially

in case of convergent results. Should the results not converge, it seems the best strategy to investigate distributions for other measures, such as skewness and kurtosis individually (as well as their appearance when inspected by eye), to determine whether the result of the BC might be biased in one or the other direction.

ACKNOWLEDGMENTS

We are grateful to Ed Huddleston of the SAS Institute Inc. for providing detailed information about the evolution of the BC.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/Quantitative_Psychology_and_Measurement/10.3389/fpsyg.2013.00700/full

REFERENCES

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Control* 19, 716–723. doi: 10.1109/TAC.1974.1100705
- Bimodal distribution. (2013). In *Wikipedia*. Available online at: http://en.wikipedia.org/wiki/Bimodal_distribution. [A correction to the listed formula for the BC has been submitted on February, 12, 2013 as part of writing this article, Retrieved: January 4, 2013].
- DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychol. Methods* 2, 292–307. doi: 10.1037/1082-989X.2.3.292
- Freeman, J. B., and Dale, R. (2013). Assessing bimodality to detect the presence of a dual cognitive process. *Behav. Res. Methods* 45, 83–97. doi: 10.3758/s13428-012-0225-x
- Hartigan, J. A., and Hartigan, P. M. (1985). The dip test of unimodality. *Ann. Stat.* 13, 70–84. doi: 10.1214/aos/1176346577
- Hartigan, P. M. (1985). Computation of the dip statistic to test for unimodality. *J. R. Stat. Soc. Ser. C (Applied Statistics)* 34, 320–325. doi: 10.2307/2347485
- Joanes, D. N., and Gill, C. A. (1998). Comparing measures of sample skewness and kurtosis. *Statistician* 47, 183–189. doi: 10.1111/1467-9884.00122
- Knapp, T. R. (2007). Bimodality revisited. *J. Mod. Appl. Stat. Methods* 6, 8–20.
- Larkin, R. P. (1979). An algorithm for assessing bimodality vs. unimodality in a univariate distribution. *Behav. Res. Methods Instrum.* 11, 467–468. doi: 10.3758/BF03205709
- Maechler, M. (2012). *diptest: Hartigan's dip test statistic for unimodality – corrected code*. R package version 0.75-74. Available online at: <http://CRAN.R-project.org/package=diptest>. [Retrieved: January 4, 2013].
- McLachlan, G., and Peel, D. (2000). *Finite Mixture Models*. Hoboken, NJ: Wiley.
- Mechler, F. (2002). *Hartigan's dip Statistic*. Available online at: <http://nicprice.net/diptest/>. [Retrieved: January 4, 2013].

- Miller, J. (2006). A likelihood ratio test for mixture effects. *Behav. Res. Methods* 38, 92–106. doi: 10.3758/BF03192754
- Ratcliff, R., and Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychol. Sci.* 9, 347–356. doi: 10.1111/1467-9280.00067
- Ratcliff, R., Van Zandt, T., and McKoon, R. (1999). Connectionist and diffusion models of reaction time. *Psychol. Rev.* 106, 261–300. doi: 10.1037/0033-295X.106.2.261
- SAS Institute Inc. (1990). *SAS/STAT User's Guide*, Version 6, 4th Edn. Cary, NC: Author. [The often-found date of 1989 does not seem to be valid. The BC was not implemented in the preceding release in 1988, Version 6, 3rd Edn.]
- SAS Institute Inc. (2012). *SAS/STAT 12.1 User's Guide*. Cary, NC: Author.
- The MathWorks Inc. (2012). *Kurtosis*. Available online at: <http://www.mathworks.de/de/help/stats/kurtosis.html>. [Retrieved: February 11 2013].
- Wagenmakers, E.-J., Grasman, R. P. P. P., and Molenaar, P. C. M. (2005). On the relation between the mean and the variance of a diffusion model response time distribution. *J. Math. Psychol.* 49, 195–204. doi: 10.1016/j.jmp.2005.02.003

Received: 02 September 2013; accepted: 14 September 2013; published online: 02 October 2013.

Citation: Pfister R, Schwarz KA, Janczyk M, Dale R and Freeman JB (2013) Good things peak in pairs: a note

on the bimodality coefficient. *Front. Psychol.* 4:700. doi: 10.3389/fpsyg.2013.00700

This article was submitted to *Quantitative Psychology and Measurement*, a section of the journal *Frontiers in Psychology*.

Copyright © 2013 Pfister, Schwarz, Janczyk, Dale and Freeman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

Table A1 | Frequency data of four hypothetical distributions of 100 values each, with corresponding estimates of skewness (m_3), kurtosis (m_4), and the BC.

Value	Data Set			
	A	B	C	D
1	3	2	2	2
2	5	26	3	3
3	5	14	3	6
4	10	6	3	17
5	17	2	3	3
6	20	0	4	4
7	17	2	5	5
8	10	6	11	12
9	5	14	21	14
10	5	26	41	30
11	3	2	4	4
m_3	0.00	0.00	-1.55	-0.59
m_4	-0.12	-1.83	1.55	-1.08
BC	0.34	0.79	0.73	0.67

Data set C is adapted from Knapp (2007) (Figure 7).