

Nonparametric Goodness-of-Fit Tests for Discrete Null Distributions

Taylor B. Arnold^{1,*}

1. Yale University, Department of Statistics

*Contact author: taylor.arnold@yale.edu

Keywords: Kolmogorov-Smirnov, Numerical Instability, Power Analysis

Goodness-of-fit tests are used to assess whether data are consistent with a hypothesized null distribution. The χ^2 test is the best-known parametric goodness-of-fit test, while the most popular nonparametric tests are the classic test proposed by Kolmogorov and Smirnov followed closely by several variants on Cramér-von Mises tests.

In their most basic forms, these nonparametric goodness-of-fit tests are intended for continuous hypothesized distributions, but they have also been adapted for discrete distributions by Conover (1972), Choulakian et al. (1994), and Gleser (1985). Unfortunately, most modern statistical software packages and programming environments have failed to incorporate these discrete versions. As a result, researchers would typically rely upon the χ^2 test or a nonparametric test designed for a continuous null distribution. For smaller sample sizes, in particular, both of these choices can produce misleading inferences.

We will present a revision of R's `ks.test()` function and a new `cvm.test()` function to fill this void for researchers and practitioners in the R environment. This work was motivated by the need for such goodness-of-fit testing in a study of Olympic figure skating scoring (Emerson and Arnold, 2010). We will first present overviews of the theory and general implementation of the discrete Kolmogorov-Smirnov and Cramér-von Mises tests. We then will discuss the particular implementation of the tests in R and provide examples paying close attention to the many numerical issues that arise in the implementation.

References

- Arnold, T. B. and J. W. Emerson (2010). *ks.test: A Kolmogorov Goodness-of-Fit Test for Discontinuous Distributions*. R package version 1.0.
- Choulakian, V., R. A. Lockhart, and M. A. Stephens (1994). Cramér-von Mises statistics for discrete distributions. *The Canadian Journal of Statistics* 22(1), 125–137.
- Conover, W. J. (1972). A Kolmogorov goodness-of-fit test for discontinuous distributions. *Journal of the American Statistical Association* 67(339), 591–596.
- Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton, NJ: Princeton University Press.
- Emerson, J. W. and T. B. Arnold (2010, August). The power of human nature in a study of olympic figure skating. Unpublished manuscript.
- Gleser, L. (1985). Exact power of goodness-of-fit tests of Kolmogorov type for discontinuous distributions. *Journal of the American Statistical Association* 80(392), 954–958.
- Imhof, J. P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika* 48, 419–426.
- Kolmogorov, A. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell' Istituto Italiano degli Attuari* 4, 83–91.