



Introduction to Modern Goodness of Fit Methods

J. C. W. Rayner, *School of Mathematical and Physical Sciences,
University of Newcastle, NSW 2308, Australia.*
Email: John.Rayner@newcastle.edu.au

O. Thas, *Department of Applied Mathematics, Biometrics and
Process Control, Ghent University, B-9000 Gent, Belgium.*
Email: olivier.thas@UGent.be

D. J. Best, *School of Mathematical and Physical Sciences,
University of Newcastle, NSW 2308, Australia.*
Email: John.Best@newcastle.edu.au

Received: November 6, 2008 Revised: April 14, 2009

Abstract

We set the context for this special issue on modern goodness of fit methods, or modern methods of assessing statistical models.

AMS Subject Classification: 62F03; 62G07; 62G10.

Key-words: Empirical distribution function tests; Exponential distribution; Generalized two-sided power distribution; Hypothesis testing; Logistic distribution; Model selection; Negative binomial distribution; Nonparametric regression; Score and generalised score tests; Smooth tests for location-scale families; Tests for ARCH assumptions in financial time series; Circular von Mises distribution.

In itself goodness of fit has an old-fashioned connotation, in part no doubt because the discipline of statistics could be argued to date from the classic Pearson (1900) paper. To avoid that label, perhaps those of us with this interest should talk about ‘assessing statistical models’. In this context ‘modern’ could be interpreted to mean ‘using current tools’ or ‘designed for current classes of models’.

What do we get from having applied a goodness of fit test and having accepted the specified model?

- a *compact description* of the data;
- *powerful parametric procedures* are valid if the data are consistent with the assumed parametric model;
- *light may be shed on the mechanisms generating the data.*

See Rayner and Best (1989, p.4) and Rayner, Thas and Best (2009, Section 1.1). What if the specified model is rejected? An informative goodness of fit procedure should provide suggestions as to how the specified model could be altered so that an improved model, i.e. a parsimonious model that described the data better, can be put forward. Do we find these characteristics in the old and the current goodness of fit methods?

Some aspects of the articles that are the foundation of goodness of fit are obviously and necessarily dated, but some are decidedly modern. The notion of a p -value is inherent in Pearson (1900). The test statistic of Neyman (1937) has components that are asymptotically independent and standard normal, assessing moment differences between the data and the hypothesised model. These components can therefore be used for exploratory data analysis or as tests of fit in their own right. Moreover, the components still turn up in decompositions of test statistics, not only in 'old' statistics like Anderson-Darling, but also in some of the more recent statistics. This occurs, for example, with statistics based on the empirical characteristic function (ECF) and the Wasserstein distance. See Gürtler and Henze (2000) and Meintanis (2004), among others.

In the early days goodness of fit only referred to the one-sample problem, which is concerned with testing that a sample comes from a hypothesised distribution. Nowadays the assessment of the quality of the fit of any parametric or semiparametric model falls under its umbrella. Sometimes the term 'lack-of-fit' is used instead.

There are several approaches that have emerged over time, but have reached what might be called maturity during the past two decades. D'Agostino and Stephens (1986) discussed the following types of tests for the one-sample problem:

- X^2 type tests
- tests based on the empirical distribution function
- regression and correlation methods
- transformation methods

Rayner and Best (1989) and Rayner, Thas and Best (2009) focus on smooth tests for the one-sample problem. The order of the smooth test to apply is an issue, and this has received considerable attention in the development of the data dependent approach. See, for example, Ledwina (1994) and Claeskens and Hjort (2004). The use of components and the data-driven selection of the order has also been studied in the context of the assessment of the quality of the fit of regression models; see Hart (1997) for an overview. Another issue is whether or not the components are diagnostic. See Henze and Klar (1996). Other recent work, for example by Henze and Meintanis (2002), proposes a general approach for constructing tests of fit based on the empirical Laplace transform.

The Rayner, Thas and Best preference is to first apply what we call an omnibus test (for example, a smooth test of order two, or four, or perhaps the Anderson-Darling test) and then scrutinise the data by using components in an EDA manner. Our recent research focuses on better tools: using generalised score tests, to construct generalised smooth tests to assess consistency of the data with virtually any model. Enough moments need to be finite so as to be able to construct the orthonormal polynomials needed to in turn construct interpretable components that are asymptotically independent and asymptotically standard normal and whose sums of squares are excellent test statistics in their own right. This approach also fits into a semiparametric view on goodness of fit testing.

We, the editors of this special edition, have recently investigated testing for distributions with current applications: the ZIP, logistic, Laplace and extreme-value. Current and future interest is on constructing improved models: for example, when components are significant to adjust the hypothesized model and ultimately select a model using modern model selection techniques crafted to our approach. This approach balances on the edge between hypothesis testing and density estimation.

Thus in our goodness of fit research our aims are thus at least three-fold.

- To build parsimonious valid models (cf *compact description*).
- To facilitate inference that is based on valid improved models that is more powerful than the standard parametric and nonparametric options (cf *powerful parametric procedures*).
- To achieve improved understanding of the model (cf *mechanisms generating the data*).

The discipline of goodness of fit is undeniably about building and understanding better models, or, as Michael Stephens puts it in the interview taken from him and published in this issue: “A practical statistician knows that the model is only an approximation, and wants to feel that he or she may safely use the model without dangerous errors; if the sample is very large and the test rejects, the model might still be good enough for practical use.” Stephens has suggested the term ‘good-enoughness-of-fit’. The application of goodness of fit techniques to current questions of significance is what makes them modern and essential.

We believe the contributors to this special edition have reflected this vision and we now give a brief summary of special issue papers.

- Lockhart, O’Reilly and Stephens revisit tests for the circular von Mises distribution. However, the focus of the paper is on the approximation of exact conditional p -values. They find that these p -values come very close to those obtained with the parametric bootstrap.
- Best, Rayner and Thas compare, predominantly from the perspective of smooth tests, new and existing tests of fit for the negative binomial distribution.
- A modification of the Greenwood statistic based on higher order spacings is used by Gulati and Shapiro to test for the logistic distribution.
- Meintanis and Iliopoulos construct a test for the generalized two-sided power distribution using the empirical moment process.
- In a commonly used nonparametric regression model Dette and Wiczorek give a new test for the hypothesis of a constant coefficient of variation. This test can also be used to construct tests for ARCH assumptions in financial time series.
- Isotones are used by Ketterer, Klar and Henze to construct tests of univariate and bivariate normality against skew-normal distributions.
- Michael Stephens shows how to approximate the asymptotic power of certain tests based on the empirical distribution function. This is demonstrated by calculating the approximate asymptotic power of the Anderson-Darling, Cramér-von Mises and Watson tests when testing for exponentiality against Weibull and gamma alternatives.
- Janic, Ledwina and Kopernika consider data-driven smooth tests for location-scale families. Their presentation permits quite general estimation schemes, including method of moments.



The editors; from left to right: D.J. Best, O. Thas and J.C.W. Rayner. This picture was taken in the office of John Rayner during Olivier's visit in April 2009.

- Rayner, Best and Thas show that the smooth tests of fit are based on sums of squares of components that are asymptotically independent and standard normal and moreover that are readily interpretable if and only if method of moments and maximum likelihood estimation coincide - essentially if testing is for a distribution from an exponential family. If that is not the case then generalized score tests can be used to construct generalized smooth tests that have convenient components.
- Hart tests the null hypothesis that all of a function's Fourier coefficients are zero using a Laplace approximation to the posterior probability of the null hypothesis. One application is to lack of fit tests in a regression context. The test statistics are weighted sums of exponentiated squared Fourier coefficients. This test can be very powerful against high frequency alternatives.
- Thas, Rayner, Best and De Boeck discuss how to construct a data-driven smooth model and incidentally reexamines the diagnostic properties of smooth tests of goodness of fit. While emphasizing the relation between hypothesis testing and density estimation, they propose a weighted mean integrated squared error (MISE) as an order selection rule for the data-driven smooth test.
- Carolan and Rayner consider inference on means after a goodness of fit test has been applied, and the hypothesised model rejected. The 'usual' parametric analysis is no longer valid, but using the model inherent in the smooth model, even more powerful analysis is possible.
- Finally Michael Stephens is interviewed by Richard Lockhart and John Spinelli.

The editors would like to thank the contributors for submitting their quality articles to this special edition, and for their patience throughout the reviewing process. We would also like to thank Editor-in-Chief of the *Journal of Statistical Theory and Practice*, Dr. Sat Gupta of the University of North Carolina at Greensboro for commissioning this issue and seeing it through to completion.

References

- Claeskens, G., Hjort, N., 2004. Goodness of fit via non-parametric likelihood ratios. *Scandinavian Journal of Statistics*, 31, 487–513.
- D'Agostino, R.B., Stephens, M.A., 1986. *Goodness-of-Fit Techniques*. Marcel Dekker, New York.
- Gürtler, N., Henze, N., 2000. Goodness-of-fit tests for the Cauchy distribution based on the empirical characteristic function. *Annals of the Institute of Statistical Mathematics*, 52, 267–286.
- Hart, J., 1997. *Nonparametric Smoothing and Lack-of-Fit Tests*. Springer, Berlin.
- Henze, N., Klar, B., 1996. Properly rescaled components of smooth tests of fit are diagnostic. *Austral. J. Statist.*, 38, 61–74.
- Henze, N., Meintanis, S., 2002. Tests of fit for exponentiality based on the empirical Laplace transformation. *Statistics*, 36, 147–161.
- Ledwina, T., 1994. Data-driven version of Neyman's smooth test of fit. *J. Amer. Statist. Assoc.* 89, 1000–1005.
- Meintanis, S., 2004. Goodness-of-fit tests for the logistic distribution based on empirical transforms. *Sankhyā, Series B*, 66, 306–326
- Neyman, J., 1937. 'Smooth' test for goodness of fit. *Skand. Aktuarietidskr.*, 20, 150–199.
- Pearson, K., 1900. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can reasonably be supposed to have arisen from random sampling. *Philos. Mag.*, 5th ser., 50, 157–175.
- Rayner, J.C.W., Best, D.J., 1989. *Smooth Tests of Goodness of Fit*. Oxford University Press, New York.
- Rayner, J.C.W., Thas, O., Best, D.J., 2009. *Smooth Tests of Goodness of Fit: Using R* (2nd edition). Wiley, Singapore.