



Generalised Smooth Tests of Goodness of Fit

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

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

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

Received: February 25, 2008 Revised: December 15, 2008

Abstract

Smooth tests of goodness of fit may be constructed by defining an order k alternative to the hypothesised probability density function and deriving the score test to assess whether or not the data are consistent with the hypothesised probability density function. For many important distributions the form of the score test statistic is the sum of squares of components that are asymptotically independent and asymptotically standard normal. Moreover each component has a moment interpretation that assists with interpreting rejection of the null hypothesis. Here a sufficient condition is given for the score test statistic to have this form and for the components to have this simple and convenient moment interpretation. Alternative approaches, using generalised score tests, are given for when the sufficient condition is not satisfied. This enables the construction of convenient tests of fit for distributions not from exponential families of distributions, such as the logistic and extreme value distributions.

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

Key-words: Order k alternative; Extreme value distribution; Logistic distribution; M-estimation; Score and generalised score tests.

1. Introduction

A feature of the smooth tests of goodness of fit first described by Neyman (1937), with more recent developments in Rayner and Best (1989) and Rayner, Thas and Best (2009), is

that testing is via a test statistic that is the sum of squares of asymptotically independent directional components. Frequently these components have a ready interpretation that enables a complete and informative assessment of how the data set differs from the hypothesised model.

This convenient decomposition does not always occur. As Boulerice and Ducharme (1995) noted, examples include testing for the logistic and Laplace distributions. Additional examples include testing for the extreme value (see Best, Rayner and Thas, 2007a) and zero inflated Poisson (see Thas and Rayner, 2005) distributions. In the next section a sufficient condition for the fortuitous decomposition is given in Theorem 2.1.

The approach of Rayner and Best (1989) is to nest a hypothesised probability density function in a more complex probability density function that is indexed by a parameter θ . When $\theta = 0$ the hypothesised probability density function is obtained. Testing for the null probability density function may be based on the score test of $H_0 : \theta = 0$ against $K : \theta \neq 0$. When the fortuitous decomposition is not obtained from the score test, we will show how to use the generalised score test to obtain such a decomposition.

Some users will be happy to test for a hypothesised distribution in a data analytic fashion using a fixed number of components: Neyman (1937) recommended four. Others will prefer to choose the number of components using model selection techniques such as the data driven techniques described in Ledwina (1994) and subsequent papers. In this article we attempt to satisfy the first group of users; the others should see the companion article by Thas, Rayner, Best and DeBoeck (2009). Our preference is to use these techniques within the class of test statistics that are sums of squares of conveniently distributed readily interpreted components rather than quadratic forms. See Rayner, Thas and Best (2009).

A sufficient condition for the convenient decomposition is given in section 2. Generalised score tests are described in section 3. Section 4 gives generalised smooth tests of fit that may be applied when the usual score test does not give a convenient decomposition. Finally, in section 5, generalised smooth tests for the logistic and extreme-value distributions are given and applied to data sets. These test statistics do have convenient decompositions.

2. A sufficient condition for a convenient decomposition

The smooth tests of goodness of fit described in Rayner and Best (1989) test for a probability density function $f(x; \beta)$, $\beta = (\beta_1, \dots, \beta_q)^T$ by first nesting $f(x; \beta)$ within what is called an order k alternative probability density function

$$g_k(x; \theta, \beta) = C(\theta, \beta) \exp \left\{ \sum_{i=1}^k \theta_i h_i(x; \beta) \right\} f(x; \beta) \quad (2.1)$$

Here $\{h_i(x; \beta)\}$ is a set of functions orthonormal on $f(x; \beta)$ with $h_0(x; \beta) = 1$ for all x , and the existence of a normalising constant $C(\theta, \beta)$ is assumed. Given a random sample X_1, \dots, X_n the score test statistic for testing $H_0 : \theta = (\theta_1, \dots, \theta_k)^T = \theta$ against $K : \theta \neq 0$ is given by Rayner and Best (1989, Theorem 6.1.1). This test is based on the $k \times 1$ score vector $\hat{U}_\theta = U_\theta(\hat{\beta}_0)$ where $U_\theta(\beta)$ has r th element $\{h_r(X_1; \beta) + \dots + h_r(X_n; \beta)\} / \sqrt{n}$, and where $\hat{\beta}_0$ is the maximum likelihood (ML) estimator of β estimated under the null hypothesis. Since β is estimated by ML the score vector coincides with the efficient score vector. Here

and subsequently the zero subscript indicates evaluation under the null hypothesis. The asymptotic covariance matrix of U_θ under the null hypothesis is given by

$$\Sigma = I_k - \text{cov}_0 \left(h, \frac{\partial \log f}{\partial \beta} \right) \text{var}_0^{-1} \left(\frac{\partial \log f}{\partial \beta} \right) \text{cov}_0 \left(\frac{\partial \log f}{\partial \beta}, h \right),$$

in which I_k is the $k \times k$ identity matrix and $h^T = (h_1(X; \beta), \dots, h_k(X; \beta))$. The usable form of the score test is $\hat{S}_k = S(\hat{\beta}_0) = \hat{U}_\theta^T \hat{\Sigma}_0^{-1} \hat{U}_\theta$ where $\hat{\Sigma}_0$ is the matrix Σ with the unknown β replaced by $\hat{\beta}_0$. The matrix $\hat{\Sigma}_0$ is required to be non-singular. If the null hypothesis is rejected then (2.1) gives an alternative model on which inference may be based. See, for example, Carolan and Rayner (2000a, 2000b, 2001) and the companion paper Rayner and Carolan (2009).

For many important distributions the form of the score test statistic is the sum of squares of components that are asymptotically independent and asymptotically standard normal. Distributions with $S(\hat{\beta}_0)$ of this form include the normal, exponential, Poisson and geometric, all discussed in Rayner and Best (1989, sections 6.2 to 6.5), and the binomial, discussed in Best and Rayner (1997). This follows from Theorem 2.1 below. See Best and Rayner (2003, 2005, 2006 and 2007) for some recent developments in smooth testing for these distributions. In all of the cases mentioned, the sufficient condition of Theorem 2.1 is satisfied.

Theorem 2.1. *Suppose F is a $q \times q$ matrix of constants of full rank and $H_1 = (h_1, \dots, h_q)^T$ in which h_r is written for $h_r(x; \beta), r = 1, \dots, k$. If*

$$\frac{\partial \log f}{\partial \beta} = FH_1, \tag{2.2}$$

the score test statistic for testing $H_0 : \theta = 0$ against $K : \theta \neq 0$ in the model (2.1) is not well defined. However, if the model is modified to

$$g_k(x; \theta, \beta) = C(\theta, \beta) \exp \left\{ \sum_{i=q+1}^{q+k} \theta_i h_i(x; \beta) \right\} f(x; \beta) \tag{2.3}$$

and if $\hat{V}_r = \sum_{j=1}^n h_r(X_j, \hat{\beta}_0) / \sqrt{n}$ in which $\hat{\beta}_0$ is the ML estimator of β under the null hypothesis, the score test statistic for this model is

$$\hat{S}_k = \hat{V}_{q+1}^2 + \dots + \hat{V}_{q+k}^2.$$

Proof. Routine calculations show that when $\partial \log f / \partial \beta = FH_1$,

$$\text{cov}_0(h, \partial \log f / \partial \beta) = (F | 0_{q \times (k-q)})^T,$$

$$\text{var}_0(\partial \log f / \partial \beta) = FF^T, \quad \text{and}$$

$$\Sigma_0(\beta) = n(0_{q \times q} \oplus I_{k-q})$$

where $0_{r \times s}$ is an $r \times s$ matrix of zeros. Hence the score test statistic, that requires an inverse of $\Sigma_0(\beta)$, is not well defined.

The model (2.3) removes $\theta_1, \dots, \theta_q$ from $g_k(x; \theta, \beta)$ in model (2.1), giving a redefined $g_k(x; \theta, \beta)$ (k in \hat{S}_k and g_k because there are $k\theta_i$). The derivation of the score test statistic then proceeds precisely as in Rayner and Best (1989, Theorem 6.1.1). \square

The reason for adopting the model (2.3) is that we are seeking to define a k dimensional space of alternatives. However the q dimensional spaces spanned by β and the first q θ_i are the same.

Theorem 2.1 shows that when (2.2) is true the score test statistic is the sum of squares of components that are asymptotically independent and asymptotically standard normal.

Moreover if ML and method of moments (MOM) estimation coincide, then each component has a simple moment interpretation that assists in interpreting rejection of the null hypothesis. Rayner, Best and Mathews (1995) argue that, when the $h_r(x, \beta)$ are the orthonormal polynomials on $f(x; \beta)$, a significant \hat{V}_r^2 suggests the data differ from the hypothesised distribution in the r th moment, although it may in fact differ in moments up to the $2r$ th. For a deeper discussion of the interpretation of components, see Henze and Klar (1996), Henze (1997) and Klar (2000), and the companion paper Thas, Rayner, Best and De Boeck (2009).

We now demonstrate that for an important class of distributions desirable decompositions do occur.

Example 2.1 (Testing for an exponential family of distributions). Suppose that, based on a random sample X_1, \dots, X_n , we wish to test for distributions in the natural parameterisation of a q parameter exponential family with probability density function

$$f(x; \beta) = K(\beta) \exp \left\{ \sum_{i=1}^q \beta_i t_i(x) \right\} u(x)$$

in which $K(\beta)$ is a normalising constant, and $u(x), t_1(x), \dots, t_q(x)$ are functions that do not involve $\beta = (\beta_1, \dots, \beta_q)^T$. As is well-known, the $\sum_{j=1}^n t_i(X_j)$ are sufficient statistics for the $\beta_i, i = 1, \dots, q$. We further assume that the $t_i(x)$ are linearly independent.

We calculate

$$\frac{\partial \log f}{\partial \beta_r} = \frac{\partial \log K}{\partial \beta_r} + t_r(x) = t_r(x) - E_0[t_r(X)] \quad \text{for } r = 1, \dots, q$$

using Lehmann and Romano (2005, Problem 2.16, p.55). If h_1, \dots, h_q span the same space as $t_1(x), \dots, t_q(x)$ then writing $H = (h_1, \dots, h_q)$ and $T = (t_1(X), \dots, t_q(X))^T$ it follows that $T = FH$, where F is a $q \times q$ matrix of constants of rank q since the $t_i(X)$ are linearly independent. Now

$$\frac{\partial \log f}{\partial \beta} = T - E_0[T] = FH - FE_0[H] = FH.$$

The last equality follows because the orthonormality, $E_0[h_r h_s] = 1$ for $r = s$ and zero otherwise, implies, with $r = i$ and $s = 0$, $E_0[h_i] = 0$ for all i , and hence $E_0[H] = 0$.

Thus for densities in an exponential family, if the $t_i(X)$ are linearly independent then (2.2) is satisfied and smooth tests of goodness of fit can be constructed in such a way that the test statistic is the sum of squares of components that are asymptotically independent and asymptotically standard normal.

We note that if the elements of T include h_1, \dots, h_q , or, equivalently, t, t^2, \dots, t^q , then the ML and MOM estimators will coincide and the components will have a convenient moment interpretation.

Example 2.2 (Testing for the binomial distribution). With $0 < p < 1$ the binomial (m, p) distributions form an exponential family with natural parameter $\beta = \log\{p/(1-p)\}$, $t(x) = x$, $K(\beta) = (1 + e^\beta)^{-m}$ and $u(x) = {}^m C_x$. We have

$$\frac{\partial \log f}{\partial \beta} = x - mp = h_1(x; \beta) \sqrt{\{mp(1-p)\}}$$

since the first orthonormal polynomial is $h_1(x; \beta) = (x - \mu)/\sigma$ where $\mu = E[X] = mp$ and $\sigma^2 = \text{var}(X) = mp(1-p)$. Theorem 2.1 advises us to remove $h_1(x; \beta)$ from (2.1), and then the test statistic is the sum of squares of components based on orthonormal polynomials of order two and higher. The statistic for testing goodness of fit is $\hat{S}_k = \hat{V}_2^2 + \dots + \hat{V}_{k+1}^2$, where the \hat{V}_r use the Krawtchouk polynomials, orthonormal on the binomial (m, p) .

Example 2.3 (Testing for the gamma distribution). Suppose we wish to test for the gamma distributions

$$f(x; \beta) = \{\Gamma(\beta_1)\}^{-1} \beta_2^{-\beta_1} x^{\beta_1-1} \exp\{-x/\beta_2\}$$

for $x > 0$ and zero otherwise, in which both $\beta_1 > 0$ and $\beta_2 > 0$. These densities form an exponential family with $t_1(x) = \log x$ and $t_2(x) = x$. Clearly $t_1(x)$ and $t_2(x)$ are not linearly independent, and so we cannot proceed as in the first example above.

If we try to verify (2.2) directly we need the derivatives of $\log f$. These are

$$\frac{\partial \log f}{\partial \beta_1} = \log \beta_2 - \frac{\Gamma'(\beta_1)}{\Gamma(\beta_1)} + \log x \quad \text{and} \quad \frac{\partial \log f}{\partial \beta_2} = -\frac{\beta_1}{\beta_2} + \frac{x}{\beta_2^2}.$$

These derivatives are not orthogonal, and so cannot be normalised to give orthonormal functions $h_1(x)$ and $h_2(x)$. The Gram-Schmidt method could be applied to construct orthonormal functions from the derivatives, and then (2.2) will be satisfied. The smooth test statistic will then be a sum of squares of components, but as the orthonormal functions are not polynomials, the components will not in general have a convenient moment interpretation.

The result above shows that if testing for a distribution from a q parameter exponential family and if the elements of T include h_1, \dots, h_q , or, equivalently, t, t^2, \dots, t^q , then the convenient decomposition with interpretable components will be obtained. If this is not the case it should *not* be expected that the convenient decomposition will be obtained. We demonstrate a particular case of the converse. If, for example, there is only one nuisance parameter ($q = 1$), and if the $h_r(x; \beta)$ are the orthonormal polynomials, suppose $\partial \log f / \partial \beta = A(\beta)h_1(x; \beta) = b(\beta)x + c(\beta)$. Integrating gives, in an obvious notation, $\log f = B(\beta)x + C(\beta) + D(x)$. Exponentiating shows that the densities of $f(x; \beta)$ form a one parameter exponential family. In the Appendix the converse is proven for arbitrary q under the assumption that $E[(\log f)^2] < \infty$.

Now it may be that $\partial \log f / \partial \beta \neq FH_1$. The score test statistic is still a quadratic form that may be expressed as a sum of squares by diagonalising the covariance matrix, but

the components are now the squares of linear combinations of the \hat{V}_r , and the informative moment interpretation is no longer available. Use of a different orthonormal system may assist in finding useful and interpretable components, but which system will depend on the particular $f(x; \beta)$.

In section 4 we will give an alternative formulation in which the generalised score test statistic is a sum of squares of components that do have a convenient moment interpretation. It may be necessary to estimate θ less than efficiently. The derivations in section 4 require the notions of M-estimators and generalised score tests, and these are briefly outlined in the next section.

3. M-Estimators and generalised score tests

The class of M-estimators includes both ML and MOM estimators. An M-estimator $\tilde{\gamma}$ satisfies

$$\sum_{j=1}^n \Psi(X_j, \tilde{\gamma}) = 0,$$

where X_1, \dots, X_n are independent but not necessarily identically distributed, Ψ is a known $p \times 1$ function not depending on j or n , and γ is a p -dimensional parameter. The estimating function Ψ must be sufficiently 'smooth'. In particular, its derivatives up to second order, and their expectations, must exist. Hence the matrices A and B defined subsequently are assumed to exist. Also, the expectation of the second order derivatives must be bounded in probability. More technical details on M-estimators may be found in van der Vaart (1998, Chapter 5).

In our setting we assume that $\gamma = (\theta^T, \beta^T)^T$ and that we wish to test $H_0 : \theta = 0$ against $K : \theta \neq 0$ with θ being the $k \times 1$ vector of primary interest, with β a $q \times 1$ vector of nuisance parameters, and with $p = k + q$. Write E_0 for expectation under the null hypothesis. The generalised score test is based on the partial M-estimator that satisfies

$$\sum_{j=1}^n \Psi_{\beta}(X_j, \tilde{\gamma}_0) = 0$$

where Ψ and γ are similarly partitioned, so that $\Psi^T = (\Psi_{\theta}^T, \Psi_{\beta}^T)$, and where $\tilde{\gamma}_0 = (0^T, \tilde{\beta}_0^T)^T$.

Define

$$U_{\theta}(\gamma) = \sum_{j=1}^n \Psi_{\theta}(X_j, \gamma),$$

$$A(\gamma) = -E_0 \left[\frac{\partial \Psi(X, \gamma)}{\partial \gamma} \right] = \begin{pmatrix} A_{\theta\theta} & A_{\theta\beta} \\ A_{\beta\theta} & A_{\beta\beta} \end{pmatrix},$$

$$B(\gamma) = E_0[\Psi\Psi^T] = \begin{pmatrix} B_{\theta\theta} & B_{\theta\beta} \\ B_{\beta\theta} & B_{\beta\beta} \end{pmatrix}, \text{ and}$$

$$\Sigma_{GS}(\gamma) = n\{B_{\theta\theta} - A_{\theta\beta}A_{\beta\beta}^{-1}B_{\beta\theta} - B_{\theta\beta}(A_{\beta\beta}^{-1})^T A_{\theta\theta}^T + A_{\theta\beta}A_{\beta\beta}^{-1}B_{\beta\beta}(A_{\beta\beta}^{-1})^T A_{\theta\beta}^T\}.$$

The generalised score test statistic is

$$\tilde{S}_k = U_{\theta}^T(\tilde{\gamma}_0)\Sigma_{GS}^{-1}(\tilde{\gamma}_0)U_{\theta}(\tilde{\gamma}_0).$$

The asymptotic distribution of \tilde{S}_k under H_0 is χ_k^2 .

If $A = B$ then $\Sigma_{GS}(\gamma)$ reduces to $n\{B_{\theta\theta} - B_{\theta\beta}B_{\beta\beta}^{-1}B_{\beta\theta}\}$. Importantly this occurs when $\Psi(X, \gamma)$ is the logarithm of the probability density function, $g_k(x; \theta, \beta)$ subsequently, in which case B is the usual information matrix. If ML estimation is used, \hat{S}_k is the score statistic of the previous section. See Boos (1992) for a comfortable exposition of generalised score tests.

4. Generalised smooth tests of goodness of fit

As mentioned in section 2, the sufficient condition of Theorem 2.1 is satisfied by at least the normal, exponential, Poisson, geometric and the binomial distributions. In such cases the ML estimator of the nuisance parameters under the null hypothesis is the solution to $\hat{V}_1 = 0$ when there is just one nuisance parameter, and to $\hat{V}_1 = \hat{V}_2 = 0$ when there are two nuisance parameters. Clearly this need not always be the case. When this convenient situation does not occur, the generalised score test can be used to obtain convenient and parsimonious tests of fit.

First we note that if we use the model given by (2.1) and derive the generalised score statistic, the score vector $U_{\theta}(\tilde{\gamma}_0)$ contains elements that are identically zero. We choose to delete terms from the model to avoid this. In this our treatment and results parallel those in Rayner and Best (1989) when deriving the smooth tests. To this end we define $(\Psi)_r = (\Psi_{\theta})_r = \partial \log g_k / \partial \theta_r$ for $r = 1, \dots, k$. Since there are q nuisance parameters we further define $(\Psi)_{k+s} = (\Psi_{\beta})_s = h_s(X, \beta)$ for $s = 1, \dots, q$. As previously, $V_r(\beta) = \sum_{j=1}^n h_r(X_j, \beta) / \sqrt{n}$ for $r = 1, \dots, p$, so that the nuisance parameters β are estimated by solving $\tilde{V}_r = V_r(\hat{\beta}_0) = 0$ for $r = 1, \dots, q$, provided these equations are of full rank. Subsequently we assume this is the case. If they are not, alternative approaches need to be developed on a case by case basis. One option is to equate further \tilde{V}_r to zero until all parameters are estimable. Note that if the h_r are orthonormal polynomials as described subsequently, $\tilde{\beta}_0$ is the MOM estimator.

It follows that

$$U_{\theta}(\gamma) = \left(\sum_{j=1}^n \frac{\partial \log g_k(X_j; \theta, \beta)}{\partial \theta_1}, \dots, \sum_{j=1}^n \frac{\partial \log g_k(X_j; \theta, \beta)}{\partial \theta_k} \right)^T.$$

Calculations similar to those in Rayner and Best (1989, section 6.1) yield

$$\frac{\partial \log g_k(x; \theta, \beta)}{\partial \theta_r} = h_r(X, \beta) - E[h_r(X, \beta)]$$

and, noting that $E_0[h_r(X, \beta)] = 0$, this choice of $g_k(x; \theta, \beta)$ results in the first q elements of $U_{\theta}(\tilde{\gamma}_0)$ being identically zero. This is resolved by removing the first q elements of θ from the model $g_k(x; \theta, \beta)$ defined by (2.1), and adopting the model given by (2.3). The resulting generalised score test statistic is given in Theorem 4.1.

Theorem 4.1. Write \tilde{V}_r for $V_r(\tilde{\beta}_0)$ and \tilde{V} for (\tilde{V}_r) . The generalised score test statistic for testing $H_0 : \theta = 0$ against $K : \theta \neq 0$ for model (2.3) is

$$\tilde{S}_k = V^T(\tilde{\beta}_0)\Sigma^{-1}(\tilde{\beta}_0)V(\tilde{\beta}_0)$$

in which $\Sigma(\beta_0) = I_k + \text{cov}_0\left(H_2, \frac{\partial \log f}{\partial \beta}\right) \text{cov}_0^{-1}\left(\frac{\partial \log f}{\partial \beta}, H_1\right) \text{cov}_0^{-1}\left(H_1, \frac{\partial \log f}{\partial \beta}\right) \text{cov}_0\left(\frac{\partial \log f}{\partial \beta}, H_2\right)$.

Proof. Recall that $(\Psi)_r = (\Psi_\theta)_r = \partial \log g_k / \partial \theta_{q+r} = h_{q+r}(X, \beta)$ for $r = 1, \dots, k$, and $(\psi)_{k+s} = (\Psi_\beta)_s = h_s(X, \beta)$ for $s = 1, \dots, q$. It follows that $U_\theta(\tilde{\gamma}_0) = (\tilde{V}_{q+1}\sqrt{n}, \dots, \tilde{V}_{q+k}\sqrt{n})^T = \tilde{V}^T \sqrt{n}$. Again writing h_r for $h_r(X, \beta)$, define H by $H^T = (H_1^T, H_2^T)$, in which $H_1 = (h_1, \dots, h_q)^T$ and $H_2 = (h_{q+1}, \dots, h_{q+k})^T$. Calculations similar to those in Rayner and Best (1989, section 6.1) give

$$\begin{aligned} A &= - \begin{pmatrix} E_0 \left[\frac{\partial \Psi_\theta}{\partial \theta} \right] & E_0 \left[\frac{\partial \Psi_\theta}{\partial \beta} \right] \\ E_0 \left[\frac{\partial \Psi_\beta}{\partial \theta} \right] & E_0 \left[\frac{\partial \Psi_\beta}{\partial \beta} \right] \end{pmatrix} \\ &= -n \begin{pmatrix} \left(E_0 \left[\frac{\partial^2 \log g_k}{\partial \theta_r \partial \theta_s} \right] \right) & \left(E_0 \left[\frac{\partial^2 \log g_k}{\partial \theta_r \partial \beta_s} \right] \right) \\ \left(E_0 \left[\frac{\partial h_r}{\partial \theta_s} \right] \right) & \left(E_0 \left[\frac{\partial h_r}{\partial \beta_s} \right] \right) \end{pmatrix} \\ &= -n \begin{pmatrix} \text{var}_0(H_2) & \text{cov}_0\left(H_2, \frac{\partial \log f}{\partial \beta}\right) \\ \text{cov}_0(H_1, H_2) & \text{cov}_0\left(H_1, \frac{\partial \log f}{\partial \beta}\right) \end{pmatrix} \end{aligned}$$

and

$$\begin{aligned} B &= n \begin{pmatrix} \left(E_0 \left[\left(\frac{\partial \log g_k}{\partial \theta_r} \right) \left(\frac{\partial \log g_k}{\partial \theta_s} \right) \right] \right) & \left(E_0 \left[\left(\frac{\partial \log g_k}{\partial \theta_r} \right) h_s \right] \right) \\ \left(E_0 \left[h_r \left(\frac{\partial \log g_k}{\partial \theta_s} \right) \right] \right) & \left(E_0 [h_r h_s] \right) \end{pmatrix} \\ &= \begin{pmatrix} \text{var}_0(H_2) & \text{cov}_0(H_2, H_1) \\ \text{cov}_0(H_1, H_2) & \text{var}_0(H_1) \end{pmatrix} \\ &= n \begin{pmatrix} I_k & 0 \\ 0 & I_q \end{pmatrix} = nI_p. \end{aligned}$$

It follows that $\Sigma_{GS}(\gamma) = n\Sigma(\beta_0)$ and \tilde{S}_k is as given. □

The asymptotic covariance matrix of $\tilde{V}, \Sigma(\tilde{\beta}_0)$, is not diagonal unless, as is familiar, (2.2) holds when $\beta_0 = \tilde{\beta}_0$. So for generalised smooth tests when the orthonormal functions are not polynomials, constructing the orthonormal system so that $\partial \log f / \partial \beta = FH_1$ is one way of achieving a test statistic that is the sum of squares of the components \tilde{V}_r . However we would usually prefer to work with orthonormal polynomials, because of their convenience, familiarity and ease of interpretation of the resulting components. In this case, when testing for a distribution not from an exponential family and when using orthonormal polynomials, the Cholesky decomposition provides a strategy for constructing convenient components.

Aside: The Cholesky decomposition is usually met in the form that when A is a symmetric positive definite matrix there exists a lower triangular matrix L such that $A = LL^T$. A dual result is that for such matrices A there exists an upper triangular matrix M such that $A = MM^T$.

Corollary to Theorem 4.1. *Given \tilde{V} and $\Sigma(\beta_0)$ as defined in Theorem 4.1, apply the dual of the Cholesky decomposition to $\Sigma^{-1}(\tilde{\beta}_0)$ so that $\Sigma^{-1}(\tilde{\beta}_0) = MM^T$. Then if $\tilde{V}^* = M^T \tilde{V}$*

$$\tilde{S}_k = \tilde{V}^{*T} \tilde{V}^* = (\tilde{V}_{q+1}^*)^2 + \dots + (\tilde{V}_{q+k}^*)^2.$$

The components \tilde{V}_{q+r}^* are asymptotically independent and asymptotically standard normal, as it is routine to show that $cov(\tilde{V}^*) = I_k$. Moreover \tilde{V}_{q+r}^* is a linear combination of $\tilde{V}_{q+1}, \dots, \tilde{V}_{q+r}$ and hence involves polynomials up to the $(q+r)$ th. Thus \tilde{V}_{q+r}^* may be interpreted as suggesting moment differences between the data and the hypothesised distribution up to the $(q+r)$ th, given that, since $\tilde{V}_1 = \dots = \tilde{V}_q = 0$, they agree in moments up to the r th. However note the caveats discussed after Theorem 2.1.

One difference between the \hat{S}_k defined in section 2 and \tilde{S}_k here is that in the former the parameters are estimated by ML estimation, while here β is estimated by solving $\tilde{V}_r = V_r(\tilde{\beta}_0) = 0$ for $r = 1, \dots, q$. The latter are MOMs because \tilde{V}_r involves sample moments up to the r th, and $E_0[\tilde{V}_r] = 0$, so that the r th equation equates a polynomial of degree r with its expected value. It is possible that the MOM and the ML estimators coincide, and it is possible that MOM estimators are fully efficient, but neither would usually be the case.

The possibly less than efficient estimation used in calculating \tilde{S}_k is offset by the fact that \tilde{S}_k is the sum of squares of k components, each of which is readily interpreted, asymptotically mutually independent and asymptotically standard normal. Of course, Theorem 4.1 is still true if the orthonormal functions are not the orthonormal polynomials; \tilde{S}_k still has the same convenient form but the interpretation of significant components will depend on the orthonormal functions used. Another advantage of the use of MOMs is that the generalised smooth tests are also defined in a semiparametric setting in which the null hypothesis is formulated in terms of the first k moments of the distribution for which we test. See Klar (2000), and Rayner, Thas and Best (2009, Chapter 9).

Thas and Rayner (2005) use generalised smooth tests to test for the zero inflated Poisson distribution. In the next section we show how to use Theorem 4.1 and its corollary to construct generalised smooth tests for location-scale families in general, and for a logistic distribution in particular.

5. Examples

5.1. Generalised smooth tests for the logistic distribution

Suppose that we wish to test whether or not the random sample X_1, \dots, X_n comes from the two-parameter logistic distribution with probability density function given by

$$f(x; \beta) = \frac{\exp\{-(x - \mu)/\sigma\}}{\sigma\{1 + \exp[-(x - \mu)/\sigma]\}^2} \quad \text{for } -\infty < x < \infty,$$

where μ and σ are the location and scale parameters respectively, and we put $\beta^T = (\mu, \sigma)$.

The first five orthonormal polynomials for the standard logistic distribution with $\mu = 0$ and $\sigma = 1$ are

$$\begin{aligned} h_0(z) &= 1, h_1(z) = z(\sqrt{3})/\pi, h_2(z) = (z^2 - \pi^2/3)3/(\pi^2\sqrt{3.2}), \\ h_3(z) &= (z^3 - 7\pi^2z/5)3/(\pi^2\sqrt{3.2}), \text{ and} \\ h_4(z) &= (z^4 - 26\pi^2z^2/7 + 27\pi^4/35)35/(64\pi^4). \end{aligned}$$

The location parameter μ has MOM estimator $\tilde{\mu}$ which is the solution to

$$V_1(\tilde{\beta}) = \frac{1}{\sqrt{n}} \sum_{j=1}^n h_1\left(\frac{(X_j - \tilde{\mu})}{\tilde{\sigma}}\right) = \frac{\sqrt{3}}{\pi\sqrt{n}} \sum_{j=1}^n (X_j - \tilde{\mu}) = 0.$$

Hence $\tilde{\mu} = \bar{X}$, the sample mean. Here $\tilde{\mu} \neq \hat{\mu}$: the ML and MOM estimators do not agree. The use of MOM estimators rather than the ML estimators may result in some loss in estimation efficiency, but, in this particular case, Lehmann (1999, Example 7.3.3) has shown that the sample mean is an efficient estimator of μ .

The MOM estimator $\tilde{\sigma}$ of σ is the solution to $V_2(\tilde{\beta}) = 0$, and is given by $\tilde{\sigma} = S(\sqrt{3})/\pi$, where S satisfies $S^2 = \sum_{j=1}^n (X_j - \bar{X})^2/n$.

Writing $m_r = \sum_{j=1}^n (X_j - \tilde{\mu})^r/n$, the third and fourth order components satisfy

$$\tilde{V}_3 = 5m_3\sqrt{7n}/(12\pi^3) \quad \text{and} \quad \tilde{V}_4 = 35(m_4 - 7\pi^4/15)\sqrt{n}/(64\pi^4).$$

These components are contrasts between the central sample and population moments of the logistic distribution.

For the logistic the covariance matrix of Theorem 4.1 above is given in Rayner, Thas and Best (2009, Appendix C), and is fortuitously diagonal, although not the identity. The order three and four components are $\tilde{V}_3^* = \tilde{V}_3/\sqrt{1.064185}$ and $\tilde{V}_4^* = \tilde{V}_4/\sqrt{1.088889}$, and these are asymptotically independent and asymptotically standard normal under the null hypothesis.

On the other hand, the score statistic is not nearly as appealing. The asymptotic covariance matrix $\Sigma(\hat{\beta}_0)$ is not diagonal, as we no longer have $V_1(\hat{\mu}) = 0$, although in practice $V_1(\bar{X})$ will often be small.

A simulation study in the companion paper Thas, Rayner, Best and De Boeck (2009) demonstrates that the convergence to the asymptotic null distribution of the generalised smooth test statistics is sometimes so slow that in general bootstrap methods are recommended for the practical application of these tests. The same study recommended the Anderson-Darling test as a good omnibus test for both the one and two parameter logistic distributions.

Life-testing of incandescent lamps data. Bain, Easterman, and Engelhardt (1973) present data on the life-testing of incandescent lamps. They consider the logistic distribution as a possible life-testing model. The observed failure times are 785, 855, 905, 918, 919, 920, 929, 936, 948 and 950. The same data set has also been used by Engelhardt (1975), who assumed that the data are well described by a logistic distribution. We now formally test the null hypothesis that the data are sampled from the two-parameter logistic distribution. For all smooth tests, bootstrap p -values are computed based on 1,000 bootstrap runs.

The bootstrap p -values of the Anderson-Darling, Watson and Cramer-von Mises tests are 0.024, 0.046 and 0.046, respectively. Thus, at the 5% level of significance, these tests all suggest that the logistic distribution does not describe the data well. Smooth tests were applied and result in p -values of 0.017 for the test based on \hat{S}_4 , that uses ML estimation, and 0.027 for the test based on \hat{S}_2 , that uses MOM estimation. The test based on \hat{S}_4 uses alternatives $g_k(x; \theta, \beta)$ that involve $h_i(x; \beta), i = 1, \dots, 4$ while the test based on \hat{S}_2 uses alternatives $g_k(x; \theta, \beta)$ that involve $h_3(x; \beta)$ and $h_4(x; \beta)$ but not $h_1(x; \beta)$ and $h_2(x; \beta)$.

The bootstrap p -values of \tilde{V}_3^2 and \tilde{V}_3^{*2} (and hence of $(\tilde{V}_3^*)^2$ and $(\tilde{V}_4^*)^2$ also) are 0.025 and 0.949 respectively. Since only \tilde{V}_3^2 results in significance, we suggest that the sample data do not agree with the logistic distribution in the third moment (skewness). Note that this conclusion is based on a test statistic which has estimated nuisance parameters by MOM estimation so that the first and the second sample moments fit exactly. Thus, this conclusion is conditional on the correctness of the first two moments.

The ‘ordinary’ smooth tests using ML estimation gave p -values for the $\hat{V}_1^2, \hat{V}_2^2, \hat{V}_3^2$ and \hat{V}_4^2 component tests, of 0.012, 0.031, 0.0328 and 0.037 respectively. The significance of \hat{V}_1 and \hat{V}_2 may be hard to understand. ML uses information about the entire distribution in estimating parameters. In particular, the moments estimated by ML may be well removed from the sample moments.

Further analysis, using data driven tests and model selection techniques are described in Rayner, Thas and Best (2009, section 11.2).

5.2. Generalised smooth tests for the extreme value distribution

The extreme value distribution is a commonly applied skewed distribution. See, for example, Coles (2004, section 4 of Chapter 3 and the list on page 1). Tests for the extreme value distribution have been suggested by Stephens (1977), Shapiro and Brain (1987) and Liao and Shimokawa (1999).

The usual extreme value distribution has two parameters: a location parameter μ and a dispersion or scale parameter σ . Its density function is given by

$$f(x; \beta) = \exp \left\{ -\frac{(x - \mu)}{\sigma} - \exp \left[-\frac{(x - \mu)}{\sigma} \right] \right\} \quad \text{for } -\infty < x < \infty,$$

where, as with the logistic, μ and σ are the location and scale parameters respectively, and we put $\beta^T = (\mu, \sigma)$. If we let X have the extreme value distribution then the standardized extreme value variable $T = (X - \mu)/\sigma$ has probability density function

$$f_T(t) = \exp\{-t - \exp(-t)\}, \quad -\infty < t < \infty.$$

For the extreme value distribution the MOM estimators $\tilde{\mu}$ of μ and $\tilde{\sigma}$ of σ are given by $\tilde{\sigma} = S(\sqrt{6})/\pi$ and $\tilde{\mu} = \bar{X} - \gamma\tilde{\sigma}$, where \bar{X}, S^2 are as in section 5.1 above and where γ is Euler’s constant, approximately 0.57722.

If MOM estimation is used then \tilde{V}_1 and \tilde{V}_2 are identically zero, and it follows from a more general argument in Rayner, Thas and Best (2009, section 11.1) that the first two non-trivial smooth test statistics \tilde{V}_3 and \tilde{V}_4 are contrasts between central sample and population moments up to the third and fourth respectively. If the ML estimators of μ and σ are used it

is not true that $\hat{V}_1 \equiv 0$ and $\hat{V}_2 \equiv 0$. Moreover \hat{V}_3 and \hat{V}_4 don't have simple links to contrasts between sample and population moments, and can no longer be simply interpreted.

A simulation study Best, Rayner and Thas (2007) shows that the convergence to the asymptotic variances of \tilde{V}_3 and \tilde{V}_4 is so slow that the parametric bootstrap is recommended for finding p -values. However the asymptotic covariance matrix is not diagonal, so \tilde{V}_3 and \tilde{V}_4 are *not* asymptotically independent. Use of the Cholesky decomposition in the Corollary to Theorem 4.1 shows that

$$\tilde{V}_3^* = 0.794076776\tilde{V}_3 \quad \text{and} \quad \tilde{V}_4^* = 0.238736899\tilde{V}_3 - 0.918707359\tilde{V}_4$$

are asymptotically independent and asymptotically standard normal.

Stephens (1977) has suggested that tests of fit for the extreme value distribution should be based on statistics themselves based on the empirical distribution function. As with other distributions, the test based on the Anderson-Darling statistic, A^2 say, seems to perform well for most alternatives. Best, Rayner and Thas (2007) compared five tests of fit for the extreme value distribution, and marginally preferred the Anderson-Darling test to those based on \tilde{V}_3 and \tilde{V}_4 . In the following example the data are analysed by the tests based on A^2 , \tilde{V}_3 , \tilde{V}_4 , \tilde{V}_3^* and \tilde{V}_4^* .

Port Pirie annual maximum sea-levels data. The annual maximum sea-levels recorded at Port Pirie (just north of Adelaide), South Australia for 1923-1987 are given in Table 5.1. These are typical of data that are modelled by the extreme value distribution.

We find $(\hat{\mu}, \hat{\sigma}) = (3.87, 0.186)$ as does Coles (2004, p.63). Further, $(\tilde{V}_3^*)^2 = 0.38$, $(\tilde{V}_4^*)^2 = 0.14$, and $A^2 = 0.17$ with parametric bootstrap p -values, based on 1,000 bootstrap runs, of 0.41, 0.63 and 0.95 respectively. It appears the Port Pirie data are well fitted by the extreme value distribution and this agrees with the graphical assessment of Coles (2004, figure 3.6).

6. Conclusion

For many distributions, the smooth tests of goodness of fit have test statistics that are sums of squares of components that are asymptotically independent and asymptotically standard normal. However several distributions have been identified in the literature for which this desirable decomposition does not occur. Here we have given a condition for this desirable decomposition and have shown that, if the smooth alternative is defined using orthogonal polynomials, the condition is satisfied if and only if the null distribution being tested for is from an exponential family of distributions.

When the condition is not true we have shown how to construct generalised smooth tests of goodness of fit based on MOM estimation. Using the Cholesky decomposition, components of the test statistic that are asymptotically independent and asymptotically standard normal may be constructed. In addition, using MOM estimation results in components that are more interpretable than those resulting from using ML.

The examples in section 5 involve location-scale families of distributions, and this means that in applying Theorem 4.1 only the covariance matrix $\Sigma(\beta_0)$ for the standardised distribution is required. In a companion paper in this issue, Best, Rayner and Thas (2009), examples are given for the negative binomial distribution. The covariance matrix there is more complicated in that it depends on the nuisance parameter, the probability of success in the negative binomial distribution.

Table 5.1. Annual maximum sea-levels, recorded in metres at Port Pirie, South Australia 1923-1987.

4.03	3.96	4.06	4.11	3.91
3.83	3.85	3.71	4.24	3.72
3.65	3.93	3.96	3.96	4.00
3.88	3.75	4.06	4.21	3.66
4.01	3.63	4.55	3.74	3.62
4.08	3.57	3.79	3.85	4.33
4.18	4.25	3.89	3.88	4.55
3.80	3.97	4.11	3.66	3.75
4.36	4.05	3.85	4.11	4.08
3.96	4.24	3.86	3.71	3.90
3.98	4.22	3.86	4.18	3.88
4.69	3.73	4.21	3.90	3.94
3.85	4.37	4.01	3.78	4.33

Appendix: If the sufficient condition for a desirable decomposition is true, sampling must be from an exponential family

In the text we demonstrated that sampling from an exponential family is sufficient for a desirable decomposition. Here we demonstrate that, provided $E[(\log f)^2] < \infty$, it is also necessary; that is, if (2.2) holds, then sampling must be from a q parameter exponential family.

From (2.2), in obvious notation,

$$\frac{\partial \log f}{\partial \beta_i} = (FH)_i = \sum_{j=1}^q f_{ij} h_j(x; \beta) = \sum_{j=0}^q f_{ij}^*(\beta) x^j \quad \text{for } i = 1, 2, \dots, q.$$

Integrating the r th of these q equations with respect to β_r gives

$$\log f = \sum_{j=0}^q F_{rj}^*(\beta) x^j + C_r(x) = \sum_{j=0}^q F_{rj}(\beta) h_j + C_r(x; \beta_{-r}),$$

where $C_r(x; \beta_{-r})$ may involve x and β_j other than β_r . Since $E[(\log f)^2] < \infty$, the constant of integration $C_r(x; \beta_{-r})$ may be expressed as a linear combination of the basis functions, and in turn, as a linear combination of powers of x :

$$\begin{aligned} \log f &= \sum_{j=0}^q F_{rj}(\beta) h_j + \sum_{j=0}^{\infty} c_{rj}(x; \beta_{-r}) h_j \\ &= \sum_{j=0}^q F_{rj}^*(\beta) x^j + \sum_{j=0}^{\infty} c_{rj}^*(x; \beta_{-r}) x^j \end{aligned}$$

Differentiating with respect to β_s gives

$$\frac{\partial \log f}{\partial \beta_s} = \sum_{j=0}^q \frac{\partial F_{rj}^*(\beta)}{\partial \beta_s} x^j + \sum_{j=0}^{\infty} \frac{\partial c_{rj}^*(x; \beta_{-r})}{\partial \beta_s} x^j$$

$$= \sum_{j=0}^q \frac{\partial F_{rj}(\beta)}{\partial \beta_s} h_j + \sum_{j=0}^{\infty} \frac{\partial c_{rj}(x; \beta_{-r})}{\partial \beta_s} h_j.$$

Taking expectation with respect to h_t for $t > q$ gives

$$\frac{\partial c_{rt}(x; \beta_{-r})}{\partial \beta_s} = 0:$$

the first order derivatives of the constants of integration $C_r(x; \beta_{-r})$ cannot involve orthonormal polynomials and hence powers higher than the q th. It follows that both the infinite sums of $\partial \log f / \partial \beta_s$ above are truncated at the q th term. Thus

$$\begin{aligned} \frac{\partial \log f}{\partial \beta_s} &= \sum_{j=0}^q \frac{\partial F_{rj}(\beta)}{\partial \beta_s} h_j + \sum_{j=0}^q \frac{\partial c_{rj}(x; \beta_{-r})}{\partial \beta_s} h_j \\ &= \sum_{j=0}^q \frac{\partial F_{rj}^*(\beta)}{\partial \beta_s} x^j + \sum_{j=0}^q \frac{\partial c_{rj}^*(x; \beta_{-r})}{\partial \beta_s} x^j \quad \text{and} \\ \log f &= \sum_{j=0}^q F_{rj}^{**}(\beta) x^j + \sum_{j=0}^q c_{rj}^{**}(x; \beta_{-r}) x^j \\ &= \sum_{j=0}^q F_{rj}(\beta) h_j + \sum_{j=0}^q c_{rj}(x; \beta_{-r}) h_j \\ &= \sum_{j=0}^q \{F_{rj}(\beta) + c_{rj}(x; \beta_{-r})\} h_j \end{aligned}$$

Again since $E[(\log f)^2] < \infty$, $\log f$ has an expansion in terms of the $\{h_r\}$:

$$\log f = \sum_{j=0}^{\infty} a_j(\beta) h_j.$$

Taking expectation with respect to h_t for all t in the last two expressions for $\log f$ shows that this sum is both unique and finite, being truncated at the q th term. Taking exponentials, it is now clear that the $f(x; \beta)$ form a q parameter exponential family. The order zero term in $\log f$ may involve the sum of a function of x alone and a sum of β alone; either or both of these may be zero.

References

- Bain, L., Easterman, J., Engelhardt, M., 1973. A study of life-testing models and statistical analyses for the logistic distribution. *Technical Report ARL-73-0009*, Aerospace Research laboratories, Wright Patterson AFB.
- Best, D.J., Rayner, J.C.W., 1997. Goodness of fit for the binomial distribution. *Australian Journal of Statistics*, 39(3), 355–364.
- Best, D.J., Rayner, J.C.W., 2003. Tests of fit for the geometric distribution. *Communications in Statistics - Simulation and Computation*, 32 (4), 1065–1078.
- Best, D.J., Rayner, J.C.W., 2005. Improved testing for the Poisson distribution using chisquared components with data dependent cells. *Communications in Statistics - Simulation and Computation*, 34 (1), 85–96.
- Best, D.J., Rayner, J.C.W., 2006. Improved testing for the binomial distribution using chisquared components with data dependent cells. *Journal of Statistical Computation and Simulation*, 76 (1), 75–81.

- Best, D.J., Rayner, J.C.W., 2007. Chi-squared components as tests of fit for the grouped exponential distribution. *Computational Statistics & Data Analysis*, 51, 3946–3954.
- Best, D.J., Rayner, J.C.W., Thas, O., 2007. Comparison of five tests of fit for the extreme value distribution. *Journal of Statistical Theory and Practice*, 1(1), 89–99.
- Best, D.J., Rayner, J.C.W., Thas, O., 2009. Anscombe's tests of fit for the negative binomial distribution. *Journal of Statistical Theory and Practice*, 3(3), 555–565. Companion paper.
- Boos, D., 1992. On generalised score tests. *The American Statistician*, 47, 327–333.
- Boulerice, B., Ducharme, G.R., 1995. A note on smooth tests of goodness of fit for location scale families. *Biometrika*, 82, 437–438.
- Carolan, A.M., Rayner, J.C.W., 2000a. Wald tests of location for symmetric nonnormal data. *Biometrical Journal*, 42(6), 777–792.
- Carolan, A.M., Rayner, J.C.W., 2000b. One sample tests of location for nonnormal symmetric data. *Communications in Statistics - Theory and Methods*, 29(7), 1569–1581.
- Carolan, A.M., Rayner, J.C.W., 2001. One sample tests for the location of modes of nonnormal data. *Journal of Applied Mathematics and Decision Sciences*, 5(1), 1–19.
- Coles, S., 2004. *An Introduction to Statistical Modelling of Extreme Values*. Springer, New York.
- Engelhardt, M., 1975. Simple linear estimation of the parameters of the logistic distribution from a complete or censored sample. *Journal of the American Statistical Association*, 70, 1084–1093.
- Henze, N., 1997. Do components of smooth tests of fit have diagnostic properties? *Metrika*, 45, 121–130.
- Henze, N., Klar, B., 1996. Properly rescaled components of smooth tests of fit are diagnostic. *Australian & New Zealand Journal of Statistics*, 38, 61–74.
- Klar, B., 2000. Diagnostic smooth tests of fit. *Metrika*, 52, 237–252.
- Ledwina, T., 1994. Data driven version of Neyman's smooth test of fit. *Journal of the American Statistical Association*, 89, 1000–1005.
- Lehmann, E.L., 1999. *Elements of Large-Sample Theory*. Springer, New York.
- Lehmann, E.L., Romano, J., 2005. *Testing Statistical Hypotheses* (3rd ed.). Springer, New York.
- Liao, M., Shimokawa, T., 1999. A new goodness of fit test for Type-I Extreme-Value and Weibull distributions with estimated parameters. *Journal of Statistical Computation and Simulation*, 64, 23–48.
- Neyman, J., 1937. 'Smooth' test for goodness of fit. *Skandinavisk Aktuarietidskrift*, 20, 150–199.
- Rayner, J.C.W., Best, D.J., 1989. *Smooth Tests of Goodness of Fit*. Oxford University Press, New York.
- Rayner, J.C.W., Carolan, A.M., 2009. Partially parametric testing. *Journal of Statistical Theory and Practice*, 3(3), 735–750. Companion paper.
- Rayner, J.C.W., Thas, O., Best, D.J., 2009. *Smooth Tests of Goodness of Fit: Using R* (2nd ed.). Wiley, Singapore.
- Rayner, J.C.W., Best, D.J., Mathews, K.L., 1995. Interpreting the skewness coefficient. *Communications in Statistics - Theory and Methods*, 24(3), 593–600.
- Shapiro, S.S., Brain, C.W., 1987. W-test for the Weibull distribution. *Communications in Statistics - Simulation and Computation*, 16, 209–219.
- Stephens, M.A., 1977. Goodness of fit for the extreme value distribution. *Biometrika*, 64, 583–588.
- Thas, O., Rayner, J.C.W., 2005. Smooth tests for the zero inflated Poisson distribution. *Biometrics*, 61 (3), 808–815.
- Thas, O., Rayner, J.C.W., Best, D.J., De Boeck, B., 2009. Informative statistical analyses using smooth goodness of fit tests. *Journal of Statistical Theory and Practice*, 3(3), 705–733. Companion paper.
- van der Vaart, A.W., 1998. *Asymptotic Statistics*. Cambridge University Press, Cambridge.