

## ON A DATA BASED POWER TRANSFORMATION FOR REDUCING SKEWNESS

JOAN G. STANISWALIS\*

*Department of Mathematical Sciences, University of Texas, El Paso 79968-0514*

THOMAS A. SEVERINI

*Department of Statistics, Northwestern University, 2006 Sheridan Road,  
Evanston, Illinois 60208-4070*

PANAGIS G. MOSCHOPOULOS

*Department of Mathematical Sciences, University of Texas, El Paso 79968-0514*

*(Received March 18, 1992; in final form September 14, 1992)*

In this paper, we study the application of a power transformation for the purpose of accelerating the rate of convergence of a statistic's sampling distribution to that of its limiting normal. The power transformation is chosen so that the sampling distribution of the transformed test statistic is less skewed. Scale invariant sufficient conditions on the cumulants of the statistic are given which guarantee the reduction of skewness. Unfortunately, this power transformation depends on the first three moments of the test statistic for which exact expressions are not always available. We propose the estimation of these moments via a parametric bootstrap. The effectiveness of this data based power transformation in an application to goodness-of-fit testing is established through computer simulations. We demonstrate that the resulting normal approximation to the sampling distribution of the transformed goodness-of-fit test statistic is better than the approximation provided by the bootstrapped sampling distribution based on 100 bootstrap samples. This computational savings is important in applications for which each bootstrap realization is computationally intensive.

KEY WORDS: Local likelihood; Kernel estimator; Weighted likelihood; Lack-of-fit; Deviance

### 1. INTRODUCTION

In many situations the limiting distribution of an estimator or test statistic is normal under certain regularity conditions. It is not uncommon for the limiting normal approximation to provide a poor approximation to the sampling distribution. Rather than working on better approximations to the sampling distribution of a statistic, or *carefully* bootstrapping its sampling distribution, we consider applying a power transformation which accelerates the rate of convergence to normality.

---

\* This research was supported in part by NSF Grant DMS 9048731 Amendment Number 9048731.  
AMS Subject Classification Codes: 62E20, 62E25, 62G99.

An application is considered here for which the power transformation is estimated from a (relatively speaking) small bootstrap sample.

The power transformation was first introduced by Wilson and Hilferty (1931) to symmetrize the distribution of a chi-square. In this case it was then established that the cube-root transformation was more effective for reducing skewness than the square-root transformation usually ascribed to Fisher. The transformation was reconsidered by Jensen and Solomon (1972) where it was shown that, if the power of the transformation is allowed to depend on the first three moments of the distribution in question, then a remarkably accurate normal approximation results. This was demonstrated in the case of positive definite quadratic forms. Since then, the power transformation has been used in a number of other important applications; for example, Mudholkar and Trivedi (1980, 1981), Moschopoulos and Mudholkar (1983). An important refinement has been considered in Moschopoulos (1983).

The power transformation is introduced in Section 2 of this paper. Section 3 describes the test statistic of Staniswalis and Severini (1991) to which the power transformation is applied in Section 4. Section 4 contains the results of a small computer simulation study which demonstrates that the proposed transformation does in fact accelerate the rate of convergence to normality of the Staniswalis and Severini (1991) goodness-of-fit test statistic. It is expected that the results in this paper will extend to other settings where the sampling distribution of a test statistic is not approximated well by the limiting normal distribution; for example, Eubank and Spiegelman (1990).

## 2. THE DATA BASED POWER TRANSFORMATION

Let  $T_n$  denote a generic nonnegative "test statistic" with mean  $\theta_n$  and variance  $\sigma_n^2$  such that  $(T_n - \theta_n)/\sigma_n \rightarrow_D N(0,1)$ . Without loss of generality, it is assumed that  $\lim_{n \rightarrow \infty} \theta_n > 0$  since an appropriate constant can always be added to  $T_n$ . The power transformation that we study is

$$P(T_n) = (T_n/\theta_n)^h, \quad \text{where } h > 0.$$

Then by the Mann-Wald theorem  $P(T_n)$  is also approximately normal for large  $n$ . To accelerate convergence to normality,  $h$  is chosen so that the leading term in a series representation of the third central moment of  $P(T_n)$  (and hence skewness) vanishes. This choice of  $h$  depends on the first three cumulants of  $T_n$ .

For convenience, the dependence on  $n$  is suppressed in the remainder of the paper. Let  $\kappa_2, \kappa_3, \dots$  denote the cumulants of a nonnegative random variable  $T$ . Then using a simple expansion it can be shown that the mean  $\mu_1(h)$  and the variance  $\mu_2(h)$  of  $P(T)$  are given as follows.

THEOREM 2.1 *Mudholkar and Trivedi (1980)*

If  $\phi_r = \kappa_r/\theta$  ( $r = 2, 3, \dots$ ) are bounded as  $\theta \rightarrow \infty$ , then

$$\begin{aligned}\mu_1(h) &= 1 + h(h-1)\phi_2/(2\theta) \\ &\quad + h(h-1)(h-2)[4\phi_3 + 3(h-3)\phi_2^2]/(24\theta^2) + R \\ \mu_2(h) &= h^2\phi_2/\theta + h^2(h-1)/(2\theta^2)[2\phi_3 + (3h-5)\phi_2^2] + R,\end{aligned}\quad (1)$$

where  $R = O(\theta^{-3})$ . Further, the choice  $h = 1 - \theta\kappa_3/(3\kappa_2^2)$  is such that the leading term for the third centered moment  $\mu_3(h)$  vanishes resulting in a skewness of the order  $O(\theta^{-3})$ .

The approximation for the mean and variance of  $P(T)$  as provided by Theorem 2.1 uses the sufficient condition that  $\theta \rightarrow \infty$ . However, we also need such approximations to the mean and variance of  $P(T)$  in situations (an example of which is given in Section 3) where  $\theta$  is bounded in  $n$ . Therefore, we wish to replace the sufficient conditions provided by Theorem 2.1 with scale invariant sufficient conditions on the cumulants of  $T$ . Following is our generalization; the proof is in the Appendix.

THEOREM 2.2 *If*

$$\begin{aligned}k_i/\theta^i &= o(n^{-\alpha_i}) \quad (i = 2, 3, \dots), \quad \text{where } 0 < \alpha_2 < \alpha_3 < \dots, \\ \sum_{k=0}^{\infty} |\theta^{-k}| E[|T - \theta^k|] \binom{r}{k} &= c_{r,n} < \infty \quad (r = 1, 2, 3),\end{aligned}\quad (2)$$

there exists  $c_r < \infty$  such that  $c_{r,n} \rightarrow c_r$  as  $n \rightarrow \infty$ , and

$$\sup_n E[|P(T_n)|^r] < \infty \quad \text{for } r = 1, 2, 3,$$

then the mean  $\mu_1(h)$  and the variance  $\mu_2(h)$  of  $P(T)$  are given by equation (1) with  $R = O(\kappa_4/\theta^4)$ . Furthermore, if  $h = 1 - \theta\kappa_3/(3\kappa_2^2)$ , the skewness of  $P(T)$  is  $O(\kappa_4/\theta^4)$ .

This theorem establishes scale invariant sufficient conditions such that for large  $n$ , the sampling distribution of  $P(T)$  is approximately normal with mean and variance given by (1). The practical choice of  $h = 1 - \theta\kappa_3/(3\kappa_2^2)$  (Mudholkar and Trivedi 1980) is guided by the fact that the best normal approximation is attained when the skewness of  $P(T)$  is small.

The application considered in the next section is a prime example of a situation where this type of transformation may be helpful. The statistic considered there is a sum of chi-squares that has significant skewness that makes the limiting normal distribution a poor approximation to the sampling distribution. The choice of  $h$  given in the theorems accelerates the convergence of  $P(T)$  to the limiting normal distribution by reducing the skewness of its sampling distribution.

### 3. APPLICATION

Let  $(X_i, Y_i); i = 1, \dots, n$ , denote independent random variables where  $X \in \mathbf{U} = [0, 1]^d$  is a vector of explanatory variables and  $Y$  is a real valued response variable. Here  $X_1, \dots, X_n$  are either iid random variables or design variables on a lattice. The distribution of  $Y|X = x$  is assumed to be a member of a family of distributions depending on the parameter  $s(x)$ ,  $x \in \mathbf{U}$ . Let  $f(Y|s(x))$  denote the density of  $Y|X = x$ , where  $f$  is known.

In parametric regression, a functional form for  $s(x)$  is postulated which depends upon a vector  $\nu$  of  $q < n$  unknown parameters. The data  $(x_i, y_i), i = 1, \dots, n$ , are reduced to a vector  $\hat{\nu}$  of  $q$  parameter estimates which frequently have physical interpretations. However, if the postulated parametric model does not provide a good approximation to the regression function within the experimental region, then the subsequent inferences are unreliable. We concern ourselves with developing diagnostic procedures for checking the goodness-of-fit of a parametric function for  $s(x)$  against a nonparametric alternative in likelihood based models.

Let  $C^k(\mathbf{U})$  denote that set of real valued functions on  $\mathbf{U}$  with continuous partial derivatives of order  $k \geq 2$  and  $\mathfrak{S} = \{s(\cdot; \nu); \nu \in \Omega \subset \mathbf{R}^q\}$ , where  $s(\cdot; \nu)$  is a postulated functional form on  $\mathbf{U}$  that depends on a vector  $\nu$  of  $q$  unknown parameters. It is assumed throughout that  $\mathfrak{S}$  is a subset of  $C^k(\mathbf{U})$ . Of interest is a test of

$$H_0 : s \in \mathfrak{S} \quad (3)$$

$$H_1 : s \in C^k(\mathbf{U}) \text{ not in } \mathfrak{S}.$$

Our test of (3) compares an estimate of the regression function under the null parametric model against an estimate under an unspecified smooth alternative.

We apply the power transformation to the test statistic  $\Lambda_w$  introduced in the paper of Staniswalis and Severini (1991) for detection of global lack-of-fit of a parametric regression equation against a nonparametric alternative. This goodness-of-fit test was developed from a likelihood perspective using weighted (local) likelihoods; Staniswalis (1989), Tibshirani and Hastie (1987). The statistic  $\Lambda_w$  measures the level of agreement between the parametric model fit to the data and the nonparametric kernel estimator (Gasser and Muller 1979) at  $m$  values of the independent variable.  $\Lambda_w$  is a measure of the difference between a nonparametric deviance and its limit in probability under the null model. Its limiting normal distribution with  $m = n$  was reported to be a poor approximation to the sampling distribution.

In order to apply the power transformation to  $\Lambda_w$ , the first three moments are needed. It is well known that this power transformation is very sensitive to misspecification of the moments. In Staniswalis and Severini (1991), the asymptotic first moment did not provide a good approximation for the first moment of  $\Lambda_w$  obtained through simulation. Useful expressions for the first moment of  $\Lambda_w$  are not available much less expressions for the second and third moments.

The parametric bootstrap will be used to estimate the first three moments of the goodness-of-fit statistic conditional on  $X_1 = x_1, \dots, X_n = x_n$ . Let  $\hat{\nu}$  denote a  $n^{1/2}$  consistent estimator of  $\nu$ ; for example, the maximum likelihood estimator

(MLE).  $B$  bootstrap samples of the form  $Y_i^*$ ;  $i = 1, \dots, n$  are generated from the known conditional distribution of  $Y|X = x_i$ , namely  $f(y|s(x_i; \hat{\nu}))$  ( $i = 1, \dots, n$ ). A goodness-of-fit statistic  $\Lambda_w^*$  is computed from each of the  $B$  bootstrap samples. The bootstrap realizations of  $\Lambda_w$  are used to construct estimates of its moments in the obvious way. The bootstrapped moment estimates are then used to estimate  $h = 1 - \theta\kappa_3/(3\kappa_2^2)$  for the power transformation of Mudholkar and Trivedi (1980). Let  $P^*(\cdot)$  denote the data based power transformation. Finally,  $P^*(\Lambda_w)$  is standardized by its mean and variance (1) and compared to critical values from a standard normal table in order to decide whether the null parametric model is rejected for lack of fit.

It is assumed throughout that the bandwidth parameter for the kernel estimator goes to zero as  $n \rightarrow \infty$  as in Staniswalis and Severini (1991). As a result of a personal communication with Jeff Hart, we do not recommend the use of  $\Lambda_w$  when this is not the case. This would occur for example under an additive error model when  $E(Y|X = x)$  is a polynomial of order  $q$  in  $x$  under the null model, and a kernel of order  $k > q$  is used. In which case, the bandwidth minimizing the mean squared error of the kernel estimator does not go to zero as  $n \rightarrow \infty$  and hence the limiting distribution of the goodness-of-fit test statistic using an optimal bandwidth is not known. The reader is referred to Raz (1990) and Edwards (1989) for randomization tests on the residuals that can be expected to work well in the latter situation.

#### 4. SIMULATIONS

A small computer simulation study was performed in order to explore the power and the size of the global goodness-of-fit test based on the normal approximation with mean and variance given by equation (1) to the sampling distribution of  $P^*(\Lambda_w)$ . The simulations were conducted in Fortran on a Sun SparcStation 1+. This example uses  $n = 100$ ,  $d = 2$  with  $X_1$  and  $X_2$  independent uniform random variables on  $[0,1]$ . The  $Y|X = x$  were taken to be Bernoulli random variables with parameter  $s(x)$ ,

$$\log \left( \frac{s(x)}{1 - s(x)} \right) = 4(x_1 + x_2) - 12x_1x_2 \quad (4)$$

for  $x = (x_1, x_2)$ .

The logistic regression models

$$\log \left( \frac{s(x)}{1 - s(x)} \right) = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 \quad (5)$$

and

$$\log \left( \frac{s(x)}{1 - s(x)} \right) = b_0 + b_1x_1 + b_2x_2 \quad (6)$$

**Table 1** Size of the Goodness-of-Fit Test

<i>Level</i>	$P^*(\Lambda_w)$	$h = 1/3$	<i>Bootstrapped p-value</i>
.5	.54	.61	.58
.4	.42	.51	.47
.3	.31	.43	.37
.2	.19	.33	.27
.1	.12	.26	.13
.05	.06	.21	.09
.04	.05	.20	.06
.02	.02	.18	.04
.01	.01	.16	.03

were fit to the data generated with the uniform and binomial random number generators in SAS. Four hundred and fifty realizations of the parametric MLE (using PROC CATMOD in SAS) and nonparametric kernel estimator of  $s(x)$  were generated to verify that the size of the goodness-of-fit test was close to the nominal level. With 450 Monte Carlo realizations there is 0.80 power at the 0.05 level to detect a 0.0115 difference between the size and a 0.05 nominal level. Two hundred realizations were used to investigate the power. Refer to Staniswalis and Severini (1991) for details about the statistic and the bandwidth computation. The pseudo-spline kernel of Messer and Goldstein (1989) was used for the kernel estimator. It has the advantage of providing a simple boundary modification to the kernel estimator.

One hundred ( $B = 100$ ) bootstrap realizations were then used to compute the data based power transformation  $P^*$  for each of the Monte Carlo realizations of  $\Lambda_w$ . The data based global bandwidth used to compute  $\Lambda_w$  was also used to compute the 100 bootstrap realizations of  $\Lambda_w^*$ . However, the bandwidth varied over the Monte Carlo realizations of  $\Lambda_w$ .

Table 1 lists the number of realizations out of 448 (two of the 450 computations yielded a value of  $h$  which resulted in a negative variance when substituted into equation (1) and therefore were not useable Monte Carlo realizations) which incorrectly rejected the null model given by equation (5) based on the value of  $P^*(\Lambda_w)$ . The p-value of the Z statistic used to test that the size of the test is equal to the nominal level of the test for  $\alpha = 0.05$  is 0.16. The size of the test is not significantly different from the 0.05 nominal level.

The bootstrapped value of  $h$  has a median of .29, mean of .26, and standard error of the mean of .01. This indicates that the value of  $h = 1/3$ , used for symmetrizing a chi-square distribution is not appropriate here. This is also demonstrated by Table 1 where the size of the test using  $h = 1/3$  as estimated from the simulation is listed.

In addition, for each of the 450 Monte Carlo realizations of the test statistic, a p-value was computed based on the 100 bootstrapped replications. Table 1 lists the size of the test which rejects when the p-value based on the 100 bootstrap replications is less than the nominal level. The p-value of the Z statistic used to test that the size of this test is equal to the nominal level of the test for  $\alpha = 0.05$  is very close to zero ( $Z = 3.89$ ). The size of the test is significantly different from

**Table 2** Power of the Goodness-of-Fit Test

<i>Standard Normal Critical Value</i>	<i>Level</i>	<i>P*(<math>\Lambda_w</math>)</i>
1.282	.2	.92
1.645	.1	.89
1.96	.05	.84
2.054	.04	.81
2.326	.02	.75
2.576	.01	.70

the 0.05 nominal level. This indicates that our bootstrap sample of size 100 is not effective for estimating the sampling distribution of the test statistic for the purpose of hypothesis testing.

In these simulations, the global bandwidth was computed for each Monte Carlo realization of the test statistic. However, the global bandwidth was held fixed during the bootstrap resampling. This means that the bootstrap sampling distribution is that of a fixed bandwidth test statistic. It is possible that the fixed bandwidth test statistic has a different sampling distribution than that of a data based bandwidth test statistic. This may explain why the size of the test procedure based on a p-value computed from the 100 bootstrap samples did not equal the nominal level. If this is the case, the p-value based on a bootstrap sample where the bandwidth is allowed to vary might give better results. However, there is no indication that this avenue, which has increased level of computational complexity, should be explored further given the favorable results of our bootstrapped power transformation to normality.

Table 2 lists the number of realizations out of 195 (five of the two hundred computations yielded a value of  $h$  which resulted in a negative variance when substituted into equation (1) and therefore were not useable Monte Carlo realizations) which correctly rejected the null model given by equation (6) based on the value of  $P^*(\Lambda_w)$ . The power of the goodness-of-fit test ranged between 0.70 and 0.92 for levels of significance between 0.2 and 0.01. The bootstrapped value of  $h$  had a median of 0.27, a mean of .23, and the mean had a standard error of .015.

## 5. DISCUSSION

Through simulation, we have been able to show that the sampling distribution of the test statistic obtained via a data based power transformation of  $\Lambda_w$  (Staniswalis and Severini 1991), namely,  $P^*(\Lambda_w)$ , is approximated well by the normal distribution with mean and variance given by (1). The advantage of using the data based power transformation is that fewer bootstrap samples were needed to estimate the moments of the test statistic than for estimation of tail probabilities from the bootstrapped sampling distribution. In many applications; for example, those involving nonparametric regression, this computational savings is important. Although the

simulations were very narrow in scope, we expect similar results will be obtained for other estimators which satisfy the conditions specified in Section 2.

In our simulations, a few of the bootstrapped realizations yielded a value of  $h$  which resulted in a negative variance when substituted into equation (1). In all of these cases, the computed value of  $h$  was also an outlier of the Monte Carlo realizations. If this were to happen with a data set being analyzed in practice, we suggest either reporting the bootstrapped p-value or changing the value of the seed of the pseudo-random number generator used to construct the bootstrapped realizations.

## 6. APPENDIX

### *Proof of Theorem 2.2*

First, we show that

$$E[(P(T))^r] = \sum_{k=0}^{\infty} \binom{rh}{k} \mu_k / \theta^k + O(\kappa_4 / \theta^4), \quad (7)$$

where  $\mu_k = E[(T - \theta)^k]$ ,

$$\binom{rh}{0} = 1 \quad \text{and} \quad \binom{rh}{k} = \frac{rh(rh-1)\dots(rh-k+1)}{k!}, \quad k \geq 1.$$

This nonstandard use of the binomial coefficient notation for nonintegral values of  $rh$  is used by Trench (1978). Observe that  $P(T)^r = \sum_{k=0}^{\infty} \binom{rh}{k} [(T - \theta)/\theta]^k$  for  $|(T - \theta)/\theta| < 1$  and any  $rh \in \mathbb{R}$  (Trench 1978; p 107, p 279). Let  $\nu_T$  denote the distribution of  $T$ . Then

$$\begin{aligned} E[P(T)^r] &= \sum_{k=0}^{\infty} \int_{\{t: |(t-\theta)/\theta| < 1\}} \binom{rh}{k} [(T - \theta)/\theta]^k d\nu_T(t) \\ &\quad + \int_{\{t: |(t-\theta)/\theta| > 1\}} P(t)^r d\nu_T(t). \end{aligned}$$

Using Markov's inequality,

$$\int_{\{t: |(t-\theta)/\theta| > 1\}} P(t)^r d\nu_T(t) \leq \sup_n E[P(T_n)^r] \mu_{2K} / \theta^{2K},$$

where  $K > 2$  is an integer satisfying  $\alpha_4 \leq K\alpha_2$ , and  $\alpha_4 \leq \alpha_3\pi_3 + \alpha_2\pi_2$  for  $3\pi_3 + 2\pi_2 = 2K$ . Since

$$\mu_{2K} = \sum_{m=1}^{2K} \sum (\kappa_{p_1}/p_1!)^{\pi_1} (\kappa_{p_2}/p_2!)^{\pi_2} \cdots (\kappa_{p_m}/p_m!)^{\pi_m} (2K)! / (\pi_1! \cdots \pi_m!),$$

where the second summation extends over all non-negative values of the  $\pi$ 's such that  $p_1\pi_1 + \cdots + p_m\pi_m = 2K$  (Kendall, Stuart, and Ord 1987, p 85-87), it follows that  $\mu_{2K}/\theta^{2K} = O(\kappa_4/\theta^4)$ .

The first term yields (7) upon observing that

$$\left| \sum_{k=0}^{\infty} \int_{\{t:|(t-\theta)/\theta|>1\}} \binom{rh}{k} [(T-\theta)/\theta]^k d\nu_T(t) \right| \leq \nu_T(\{t:|(t-\theta)/\theta| \geq 1\}) \sum_{k=0}^{\infty} \theta^{-k} E[|T-\theta|^k] \binom{rh}{k} = O(\kappa_4/\theta^4).$$

Next, rewrite the terms of the absolutely convergent series in equation (7) in terms of the cumulants and truncate the series as follows:

$$E[P(T)^r] = 1 + a_{1,r}\kappa_2/\theta^2 + a_{2,r}\kappa_2^2/\theta^4 + a_{3,r}\kappa_3/\theta^3 + R \tag{8}$$

where the  $a_{i,r}$  depend on  $h$ , but not  $n$ . Since  $(T_n - \theta_n)/\sigma_n \rightarrow_D N(0,1)$  and the fourth order cumulant of a standard normal random variable is zero, it follows that  $2\alpha_2 < \alpha_4$ . Together with assumption (2) this implies that  $R = O(\kappa_4/\theta^4)$ .

Finally, from (8) a truncated series approximation for the third centered moment  $\mu_3(h)$  of  $P(T)$  is derived in terms of the first three cumulants of  $T$  and as a function of  $h$ . The power  $h$  is then chosen to make the leading term in  $\mu_3(h)$  zero; see for example, Moschopoulos (1983). Similar arguments show that (1) provide approximate expressions for the mean and variance of  $P(T)$  with  $R = O(\kappa_4/\theta^4)$ .

**References**

Edwards, Eric (1989). A Monte Carlo approach to hypothesis testing in nonparametric regression. Ph.D. thesis: Virginia Commonwealth University, Department of Biostatistics, Richmond, Virginia, 23298-0032.

Eubank, R. L., and Spigleman, C. H. (1990). Testing the goodness-of-fit of a linear model via nonparametric regression techniques. *Journal of the American Statistical Association*, **85**(410), 387-392.

Gasser, T. and Müller, H. G. (1979). Kernel estimation of regression functions. *Lecture Notes in Mathematics: Smoothing Techniques for Curve Estimation* (T. Gasser and M. Rosenblatt, eds.). New York: Springer-Verlag, **757**, 23-68.

Jensen, D. R. and Solomon, Herbert (1972). A gaussian approximation to the distribution of a definite quadratic form. *Journal of the American Statistical Association*, **67**(340), 898-902.

Kendall, Maurice, Stuart, Alan, and Ord, J. Keith (1987). *Kendall's Advanced Theory of Statistics*, Fifth Edition of Volume 1. New York: Oxford University Press.

- Messer, K. and Goldstein, L. (1989). A new class of kernels for nonparametric curve estimation. To appear in *Annals of Statistics*.
- Moschopoulos, Panagis G. (1983). On a new transformation to normality. *Communications in Statistics—Theory and Methods*, **12**(16), 1873–1878.
- Moschopoulos, Panagis G., Mudholkar, G. S. (1983). A likelihood ratio-based normal approximation for the non-null distribution of the multiple correlation coefficient. *Communications in Statistics—Simulation and Computation*, **12**(3), 355–371.
- Mudholkar, G. S., and Trivedi, M. C. (1981). A gaussian approximation to the distribution of the sample variance from non-normal populations. *Journal of the American Statistical Association*, **76**(374), 479–485.
- Mudholdar, G. S., Trivedi, M. C. (1980). A normal approximation for the distribution of the likelihood ratio statistic in multivariate analysis of variance. *Biometrika*, **67**(2), 485–488.
- Raz, Jonathan (1990). Testing for no effect when estimating a smooth function by nonparametric regression: A randomization approach. *Journal of the American Statistical Association*, **85**(409), 132–138.
- Staniswalis, J. G. (1989). On the kernel estimate of a regression function in likelihood based models. *Journal of the American Statistical Association*, **84**(405), 276–283.
- Staniswalis, J. G. and Severini, T. A. (1991). Diagnostics for assessing regression models. *Journal of the American Statistical Association*, **84**(205), 276–283.
- Tibshirani, R. and Hastie, T. (1987). Local likelihood estimation. *Journal of the American Statistical Association*, **82**, 559–568.
- Trench, William F. (1978). *Advanced Calculus*. Harper & Row, New York.
- Wilson E. B. and Hilferty, M. M. (1931). The distribution of chi-square. *Proceedings of the National Academy of Science*, **17**, 684–688.