

One Sample Tests for the Location of Modes of Nonnormal Data

ANTHONY M. CAROLAN

anthonyc@uow.edu.au

School of Mathematics and Applied Statistics, University of Wollongong, Australia

J. C. W. RAYNER

john_rayner@uow.edu.au

School of Mathematics and Applied Statistics, University of Wollongong, Australia

Abstract. There are few techniques available for testing if modes take specified values. We show that standard tests of location such as the t-test and the Wilcoxon test, which test for the mean and median respectively, can perform poorly as tests for modes when the data is other than unimodal and symmetric. Carolan and Rayner [1] proposed a score test of location for symmetric nonnormal data. We consider a family of distributions similar to those considered by Carolan and Rayner [1] and propose a test for the mode or modes of data from multimodal or skewed distributions and demonstrate by way of simulations that it is reasonably effective.

Keywords: Score Test, Wald Test, Skewed Distributions, Multimodal Distributions, t-test, Wilcoxon Test

1. Introduction

For symmetric data the mean and median are obvious measures of location. When the appropriate assumptions are satisfied, standard parametric and nonparametric tests can be used to test for the mean and median. The t-test assumes normality (which of course implies symmetry) while the Wilcoxon signed rank test assumes just symmetry. There are also non-parametric procedures that can be used to test for the median of asymmetric data - the sign test for example. For more details of these tests see Lehmann [5] or Daniel [3]. In the case of asymmetric data the data analyst must carefully consider what exactly it is that they wish to test. More often than not the mean is of little value as a measure of location for skewed data and the median is more useful. On other occasions the mode or modes of skewed data maybe of more interest. For example consider a population which is thought to be roughly normal but with ‘contamination’ affecting some observations, leading to a skewed or multimodal sample. If the underlying population without the contamination is of interest then the mode of the sample may give the best idea of the true location. In this

paper we suggest one possible model for skewed or multimodal data and use it to develop tests for the location of modes.

Carolan and Rayner [1] proposed a test of location for nonnormal symmetric data based on the model,

$$g(x; \mu, \sigma, \boldsymbol{\theta}) = C(\boldsymbol{\theta}) \exp\left(\sum_{i=3}^k \theta_i h_i(x; \mu, \sigma)\right) f(x; \mu, \sigma). \quad (1)$$

In (1) $f(x; \mu, \sigma)$ is the normal probability density function with mean μ and variance σ^2 , $h_i(x; \mu, \sigma) = H_i\left(\frac{x-\mu}{\sigma}\right)$ is the normalised Hermite-Chebyshev polynomial of degree i , orthonormal on $f(x; \mu, \sigma)$; $\boldsymbol{\theta} = (\theta_3, \theta_4, \dots, \theta_k)^T$ is a vector of parameters and $C(\boldsymbol{\theta})$ is a normalising constant. For $C(\boldsymbol{\theta})$ to exist and for (1) to define a proper probability density function it is required that k , the order of the highest order polynomial included in the model, be even and that $\theta_k < 0$ so that $g_k(x) \rightarrow 0$ as $x \rightarrow \pm\infty$. The idea is that instead of the data following a $N(\mu, \sigma^2)$ distribution, they follow a ‘smooth alternative’ to normality. The greater k , the more complex and potentially nonnormal the model may be. Carolan and Rayner [1] derived a score test of $H_0 : \mu = \mu_0$ against $K : \mu \neq \mu_0$ for data from this model. In particular they considered the symmetric case where only θ_4 is included in the model.

Here we will consider a variation of the model (1) and how the score test behaves when skewed distributions are permitted: that is, the model includes a term involving θ_i for at least one odd i . In the symmetric case μ and the mean both take the same value but when θ_3 (or any odd degree term) is included in (1), the model becomes skewed and there is no clear interpretation of the parameter μ as a measure of location. Replacing the Hermite polynomials in (1) by simple powers of $(x - \mu)$ leads to a model where it can easily be shown that μ represents the mode: see Section 2. Preliminary results suggested some numerical instability in estimating the parameter σ so for ease of computation a further alteration was made to the model, the parameter σ being replaced by θ_2 ; $(\theta_2 - \frac{1}{2})$ replaces $\frac{1}{2\sigma^2}$ as the coefficient of $(x - \mu)^2$ in the exponent. Under this reparameterisation $\theta_2 = 0$ corresponds to $\sigma = 1$ and $\theta_2 \rightarrow \frac{1}{2}$ corresponds to $\sigma^2 \rightarrow \infty$. So the model we use in this paper is

$$g_k(x; \mu, \boldsymbol{\theta}) = C(\boldsymbol{\theta}) \exp\left(\sum_{i=2}^k \theta_i (x - \mu)^i\right) f(x; \mu, 1), \quad (2)$$

again with k even and $\theta_k < 0$ (k being the order of highest order polynomial included in the model). Curves of some probability density functions of

this form are given in several subsequent figures throughout this paper. In Section 2 we show that when we restrict θ_2 to take values less than $\frac{1}{2}$, the parameter μ represents the mode or modes of the distribution. The other parameters in the model $\theta_3, \dots, \theta_{k-1}$ can take any real values. In Section 3 we discuss how, for data from such a model, we can use score and Wald tests to test for the location of modes. Section 4 gives some examples of how these tests perform for particular data sets. Simulation results for size and power properties are presented in Section 5.

2. Modes of the Distribution and the Region of Unimodality

In this section we demonstrate that when $\theta_2 < \frac{1}{2}$, the parameter μ is a mode of the distribution specified by (2). We also find the region of the parameter space where the distribution is unimodal.

First we show μ is a mode. Differentiating (2) with respect to x we get

$$\frac{dg_k}{dx} = \left(\sum_{i=2}^k i\theta_i (x - \mu)^{i-1} - (x - \mu) \right) g_k(x; \mu, \boldsymbol{\theta}).$$

Clearly when $x = \mu$, $\frac{dg_k}{dx} = 0$ so there is a stationary point at μ . Also

$$\begin{aligned} \frac{d^2g_k}{dx^2} &= \left(\sum_{i=2}^k i\theta_i (x - \mu)^{i-1} - (x - \mu) \right)^2 g_k(x; \mu, \boldsymbol{\theta}) \\ &\quad + \left(\sum_{i=2}^k i(i-1)\theta_i (x - \mu)^{i-2} - 1 \right) g_k(x; \mu, \boldsymbol{\theta}). \end{aligned}$$

So when $x = \mu$,

$$\frac{d^2g_k}{dx^2} = (2\theta_2 - 1) g_k(\mu; \mu, \boldsymbol{\theta}).$$

This is negative whenever $\theta_2 < \frac{1}{2}$ since $g_k(x; \mu, \boldsymbol{\theta})$, being a probability density function, is positive for all x . So if we restrict $\theta_2 < \frac{1}{2}$ then $g_k(x; \mu, \boldsymbol{\theta})$ has a local maximum at $x = \mu$ and μ is a mode. Without this restriction μ can take the value of both local maxima and local minima. Multimodality is possible because there are stationary points other than at $x = \mu$. The region of the parameter space where $g_k(x)$ is unimodal is that where

$$\left(\sum_{i=2}^k i\theta_i (x - \mu)^{i-1} - (x - \mu) \right) g_k(x; \mu, \boldsymbol{\theta}) = 0$$

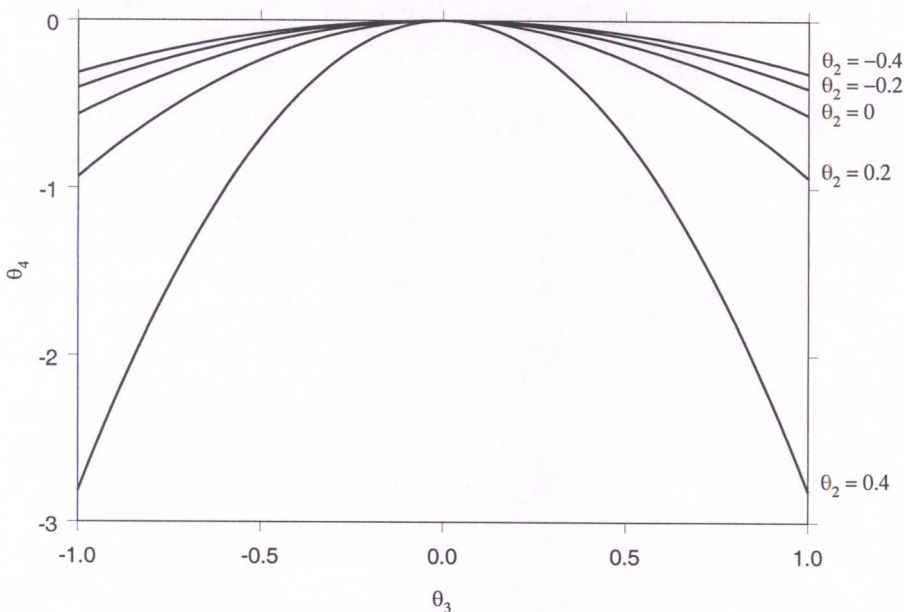


Figure 1. $g_k(x; \mu, \theta)$ will be unimodal for θ_3, θ_4 pairs that fall inside the parabola for the appropriate value of θ_2 .

has only one solution, or equivalently the region in which

$$\left(\sum_{i=2}^k i\theta_i (x - \mu)^{i-2} - 1 \right) g_k(x; \mu, \theta) = 0$$

has no solutions. Dividing by $g_k(x; \mu, \theta)$, which is positive for all x , we require

$$\sum_{i=2}^{k+1} i\theta_i (x - \mu)^{i-2} - 1 = 0$$

to have no real solutions. Considering the case where $k = 4$, this polynomial is the quadratic

$$4\theta_4 (x - \mu)^2 + 3\theta_3 (x - \mu) + 2\theta_2 - 1 = 0$$

which has no real solution when the discriminant is negative. So we require

$$9\theta_3^2 - 32\theta_4 \left(\theta_2 - \frac{1}{2} \right) < 0.$$

Some examples of this region for different values of θ_2 are sketched in Figure 1. For $k > 4$ the region where $g_k(x)$ is unimodal involves higher degree polynomials and cannot be so easily defined. Considerable deviation from normality can be modeled by the $k = 4$ model which we will use here.

3. Applying the Score and Wald Tests to Skewed Data

A detailed exposition of the score and Wald tests is given in Cox and Hinkley [2]. Deriving a score test of $H_0 : \mu = \mu_0$ against $K : \mu \neq \mu_0$ based on (2) follows the same steps as in Carolan and Rayner [1]. Maximum likelihood estimates under the null hypothesis are calculated numerically and substituted into the score vector and information matrix from which the score statistic \hat{S} is calculated. The score statistic is given by

$$\hat{S} = \{U_\mu(\hat{\gamma}_0)\}^T \{\Sigma(\hat{\gamma}_0)\}^{-1} \{U_\mu(\hat{\gamma}_0)\}$$

where γ is the vector of all parameters and is partitioned into the parameter of interest here, μ , and nuisance parameters, θ ; that is $\gamma = (\mu, \theta^T)^T$. Also $U(\hat{\gamma}_0)$ is the score vector with γ replaced by $\hat{\gamma}_0$, its estimate under the null hypothesis; $U_\mu(\hat{\gamma}_0)$ is the component of the score vector which relates to the parameter of interest and $\Sigma(\gamma) = I_{\mu\mu}(\gamma) I_{\theta\theta}^{-1}(\gamma) I_{\theta\mu}(\gamma)$ where $I(\gamma)$ is the information matrix. Again subscripts μ and θ refer to the parameters of interest and nuisance parameters respectively. Under H_0 , \hat{S} will be asymptotically chi-squared distributed with one degree of freedom.

Model (2) permits multimodal distributions which make tests of location based on the mode difficult to interpret. Such a distribution with, say, m modes, can be rewritten in m different ways with the parameter μ taking the value of a different mode in each case (see Section 4 for an example). So testing $\mu = \mu_0$ against $\mu \neq \mu_0$ really tests whether or not there is ‘a mode’ rather than ‘the mode’ at μ_0 . A distribution with m modes at $\mu_1, \mu_2, \dots, \mu_m$ say, can often be reasonably well approximated by a distribution with $m + 1$ modes at $\mu_1, \mu_2, \dots, \mu_m, \mu_{m+1}$ where the spurious mode at μ_{m+1} is the result of sampling variation or involves only a small deviation from the true shape of the distribution. For example μ_{m+1} may be a mode of negligible height outside the range of the data. This flexibility of our model can cause problems for the score test, especially for small sample sizes and if the choice of k is too large. In particular, if the data has fewer modes than the maximum permitted by the choice of k , then

because of spurious modes the score test may not be effective in detecting large deviations from the null hypothesis. For this reason we would expect our test to perform best for a large sample from a m -modal distribution with clearly defined modes and with the choice of k permitting a maximum of m modes. In general g_k can have up to $\frac{k}{2}$ modes, noting that k is always chosen to be even.

In order to try and avoid these problems with multimodal distributions we have investigated restricting the maximum likelihood estimation in the calculation of the score statistic to the regions of the parameter space where $g_k(x)$ is unimodal. We will denote this statistic by \hat{S}_R . In Section 5 we present simulation results, for data from a truly unimodal distribution, which show the test size to be about right, but the power very poor. As μ becomes more distant from μ_0 a large proportion of the parameter estimates fall on the boundary of the unimodal region.

A second possible solution is to consider the Wald test. The Wald test, with test statistic,

$$\hat{W} = (\hat{\mu} - \mu_0)^T \Sigma(\hat{\gamma}) (\hat{\mu} - \mu_0),$$

is asymptotically equivalent to the score test and, as we will see, is somewhat more convenient in our case. Because the Wald test requires only the calculation of the unconditional maximum likelihood estimates and not maximum likelihood estimates under the null hypothesis, it avoids some of the problems mentioned above. Given the freedom to estimate all parameters, maximum likelihood estimation is more likely to fit a distribution with the correct number of modes than if μ is fixed at an unreasonable value. Also in the case of the score test, when μ takes an unreasonable value there can be numerical problems in finding the maximum likelihood estimates under the null hypothesis, $\hat{\gamma}_0$, as well as in the calculation of $\{\Sigma(\hat{\gamma}_0)\}^{-1}$. Estimation under the full model, while slightly more computationally time consuming (one extra parameter needs to be estimated), will encounter such problems less frequently.

As mentioned above if the fitted distribution is multimodal there may be more than one set of parameter estimates which identify the same distribution - which set do we choose in calculating \hat{W} ? There are a few obvious ways of proceeding. First, if it is known the population is unimodal, then it is sensible to choose the $\hat{\mu}$ which corresponds to the largest mode of the fitted distribution. The Wald statistic calculated in this fashion will be denoted \hat{W}_M . This test will ignore smaller secondary modes of the fitted distribution which, given the assumption of unimodality, must be spurious. This version of the Wald test will be appropriate if there is prior knowledge to suggest the assumption of unimodality is reasonable or if a test for the

number of modes has been performed and unimodality accepted (for an example of such a test see Silverman [8]). If on the other hand there is no information available about the hypothesised mode or the number of modes, it would appear reasonable to proceed with the Wald test using the parameter estimates with $\hat{\mu}$ closest to the hypothesised value μ_0 - this will give the smallest possible value of the statistic. This statistic calculated using the nearest $\hat{\mu}$ will be denoted \hat{W}_N . In this case the misidentification of a unimodal distribution as multimodal is simply an example of a type II error and will become less likely as the sample size is increased. Such misidentification should occur a lot less frequently when unconditional maximum likelihood estimates are used rather than the estimates under the null hypothesis as in the score test.

Other choices of $\hat{\mu}$ in evaluating the Wald statistic may apply in particular instances - for example if it were hypothesised that μ_0 is the left most mode of the underlying distribution then in calculating \hat{W} we would choose the left most value of $\hat{\mu}$. This flexibility of the Wald test makes it much easier to apply and interpret than the score test. The score test will be most similar to the application of the Wald test using the \hat{W}_N statistic.

In the next section we give some examples of how the score and Wald tests perform in particular cases of data from unimodal and multimodal distributions.

4. Examples

In this section we consider three different situations and how the score test and Wald test for modes perform in each. First, in Section 4.1, applying these tests to a clearly bimodal distribution we demonstrate an instance where both techniques appear to behave reasonably well. For unimodal data we give an example in Section 4.2 that shows the score test to be somewhat less effective. We also show how the score test with restricted parameter space performs for the same example. In Section 4.3 we consider a real data set to show how the Wald test maybe applied in practice.

4.1. Example of Score Test and Wald Test with Bimodal Data

Consider a sample of 100 observations from $g_4(x; \mu, \theta)$ where $\mu = 0$ and $\theta = (0, 0.25, -0.03)^T$. This distribution and histogram of the sample are shown in Figure 2. This distribution has two modes, at 0 and 4.322, and could have been written as $g_4(x; \mu^*, \theta^*)$ with $\mu^* = 4.322$ and $\theta^* = (-0.121, -0.269, -0.03)^T$. In Figure 3 we have plotted values of the score

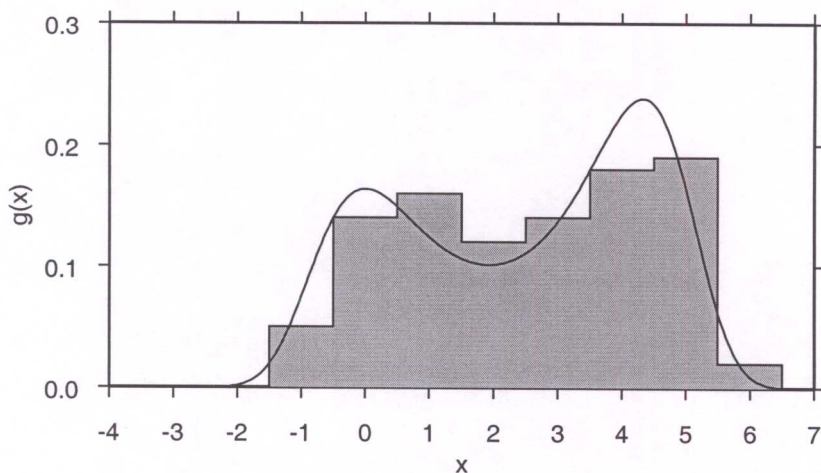


Figure 2. Probability density function and a histogram of 100 observations from $g_A(x; \mu, \theta)$ where $\mu = 0$ and $\theta = (0, 0.25, -0.03)^T$.

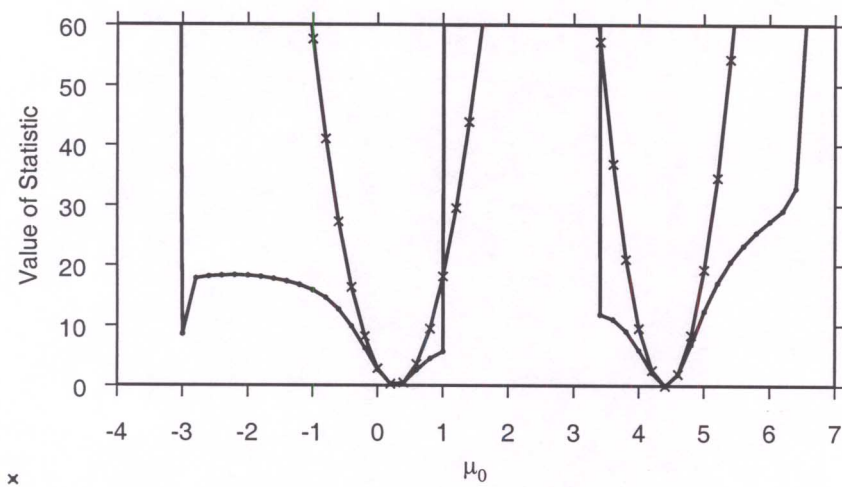


Figure 3. Behaviour of the score statistic (dots) and Wald statistic (crosses) for a sample of size 100 from $g_A(x; \mu, \theta)$ where $\mu = 0$ and $\theta = (0, 0.25, -0.03)^T$ as μ_0 is varied over the range of the data.

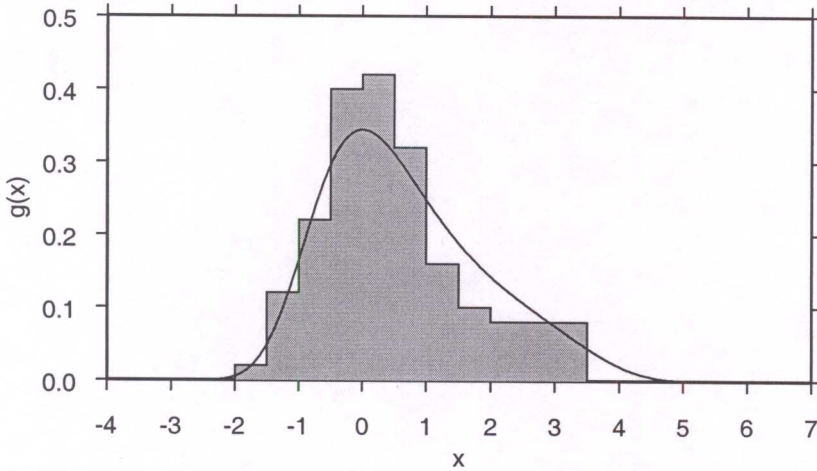


Figure 4. Probability density function and a histogram of 100 observations from $g_A(x; \mu, \theta)$ where $\mu = 0$ and $\theta = (0, 0.2, -0.03)^T$.

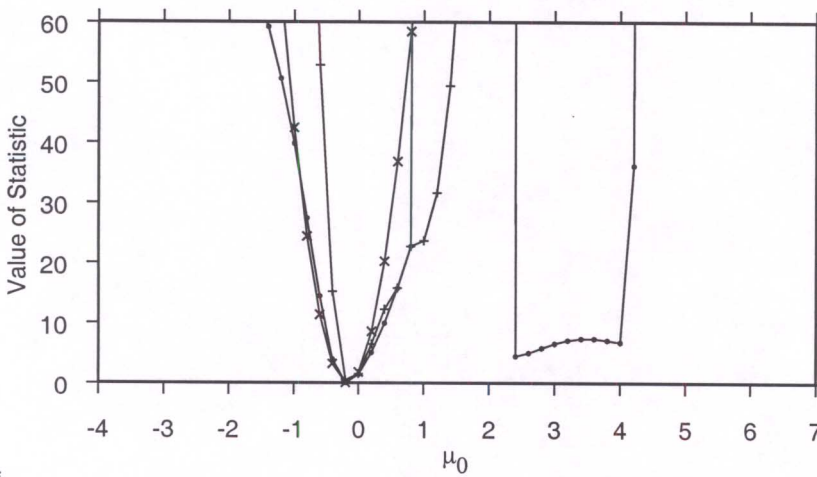


Figure 5. Behaviour of the unrestricted score statistic (dots), restricted score test (pluses) and Wald statistic (crosses) for a sample of size 100 from $g_A(x; \mu, \theta)$ where $\mu = 0$ and $\theta = (0, 0.2, -0.03)^T$ as μ_0 is varied over the range of the data.

statistic and Wald statistic as the hypothesised mode, μ_0 , is varied. We expect both to take small values in the region of the modes and large values elsewhere. Note that the asymptotic critical value for a 5% level test is $\chi_{1,0.05}^2 = 3.84$. This plot exhibits some interesting behaviour. The score statistic behaves well in the region of the true modes, 0 and 4.322, returning small non-critical values. As μ_0 moves away from both modes it appears that the test is performing reasonably well with the score statistic increasing in magnitude. When μ_0 falls in the intervals $[-4, -3]$, $[1.2, 3.2]$ and $[6.4, 7]$ the maximum likelihood estimate of θ_2 is greater than $\frac{1}{2}$. So the restriction that $\theta_2 < \frac{1}{2}$ (forcing $\hat{\theta}_2 = 0.4999$ in our implementation) leads to the score statistic becoming unstable and in the majority of these cases it falls outside the range of the plot. This behaviour is not too unreasonable; $\theta_2 \rightarrow \frac{1}{2}$ suggests that for the particular choice of μ_0 being considered the model proposed in (2) is not applicable and therefore we would hope for large values of \hat{S} . These problems are avoided altogether by the Wald test which requires estimation under the full model. As mentioned in the previous section computational and numerical difficulties are far less likely to arise than in the score test case where maximum likelihood estimation is required with μ fixed at what maybe a unreasonable value. Because the Wald statistic is just a quadratic function of the distance between μ_0 and the nearest $\hat{\mu}$ it behaves in a far more predictable fashion than the score statistic.

4.2. Example of Score and Wald Tests with Unimodal Data

Now consider data from a skewed but unimodal distribution. Choosing $\mu = 0$ and $\boldsymbol{\theta} = (0, 0.2, -0.03)^T$ in $g_4(x; \mu, \boldsymbol{\theta})$ gives us such a distribution. For unimodality we require

$$9\theta_3^2 - 32\theta_4 \left(\theta_2 - \frac{1}{2} \right) < 0.$$

In this case

$$9\theta_3^2 - 32\theta_4 \left(\theta_2 - \frac{1}{2} \right) = -0.12 < 0,$$

so the distribution is indeed unimodal. This can be seen in Figure 4 where the probability density function and a histogram of 100 observations from this distribution are plotted.

Figure 5 shows the values of \hat{S} , \hat{S}_R and \hat{W}_M for various values of μ_0 for a sample of size 100. In the case of the unrestricted score statistic, there

are again intervals where the choice of μ_0 forces $\hat{\theta}_2 \rightarrow \frac{1}{2}$ and very large values of \hat{S} result. It also appears that as μ_0 takes some larger positive values, outside the range of the majority of the data (for example, the interval $[2.4, 4]$), there is still sufficient flexibility in the model to allow \hat{S} to take reasonably small values. In this particular case these values of \hat{S} are still large enough to be significant at the 5% level but do suggest a possible weakness of the score test. This behaviour is typical, can often be more extreme and does lead to poor power properties. As mentioned in the previous section, the model with $k = 4$ permits up to two modes. So for a small sample of unimodal data a hypothesised mode distant from the true mode can often be accommodated by the model without a major distortion in the shape of the distribution. Again the Wald test appears to be better behaved than the score test and just as effective in detecting deviations from the true mode.

Using the score test with the parameter estimates restricted to the region where the distribution is unimodal improves its performance in this case. In the next section we conduct a small simulation study to show that the power of the score test with the restricted parameter space is actually quite poor. This is due mainly to erratic behaviour of the score statistic when one or more of the parameters lie on the boundary of the permissible region.

4.3. Peruvian Indian Pulse Rates

Here we consider applying the Wald test for the location of modes to a real set of data. We use data from Hand et al. [4], (originally from Ryan et al. [7]) that gives the pulse rate for 39 Peruvian Indians. No further information is given regarding the sample. A histogram of the data (Figure 6) shows it to be quite skewed. It may be that the skewed nature of the data is due to elderly or ill people in the sample. If we had previous information regarding heart rates for young healthy Indians we maybe more interested in comparing the old benchmark to the mode of this new data set rather than to its mean or median. For the sake of this example let us assume that it is known that a healthy heart rate for a healthy young Peruvian Indian is around 65. A naively conducted t-test of $H_0 : \mu = 65$ against $K : \mu \neq 65$ yields a p-value of 0.0012. If we use the Wald test for modes with the nearest mode method we obtain a p-value (based on 500 bootstrap simulations) of 0.58. These p-values lead to clearly different conclusions. The reason for this is clear in Figure 6. The skewed nature of the data has dragged the mean away from the peak in the distribution.

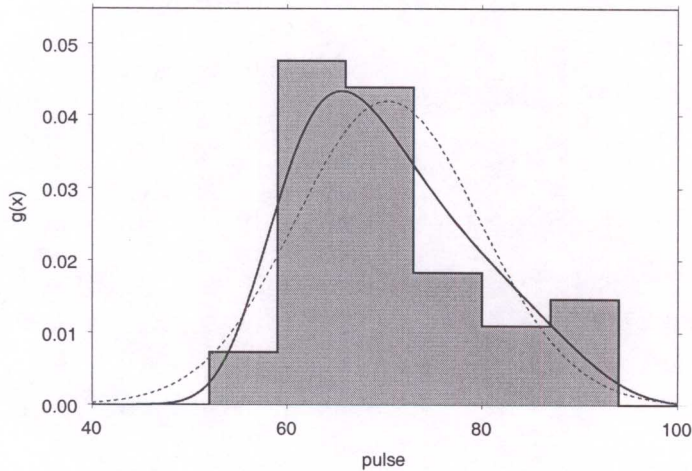


Figure 6. A histogram and fitted probability density functions (normal - dashed; g_4 - solid) for the Peruvian Indian pulse data.

This example demonstrates how tests for modes may sometimes be more appropriate than standard tests of location.

5. Simulations

In this section we present some simulated size and power properties of the score and Wald tests for the situations discussed in Section 4. The examples given in Section 4 suggest that the score test may perform reasonably well for large sample sizes when the model is not over parameterised. On the other hand, including too many parameters θ_i in the model makes spurious modes more likely and may be costly in terms of power, especially for smaller sample sizes. Simulation results confirm this. We also find that the power of the score test is very poor when parameter estimation is restricted to the unimodal region. The Wald test is shown to have similar size and power properties to the score test but is computationally more convenient.

The Wald test uses only the parameter estimates under the full model. For each simulated sample maximum likelihood estimation need only be performed once and the Wald test can be calculated very efficiently for a number of alternative values of μ . For the score test, on the other hand, each choice of μ requires a new set of maximum likelihood estimates. So

simulations can be performed far more efficiently for the Wald test than for the score test. Because of this, far more points have been used for the Wald test power curves (200 points for unimodal data and 400 for bimodal data) than for the score test curves (11 points for unimodal data and 16 points for bimodal data). Of course this advantage is only relevant for simulation studies such as ours. The Wald test does require the estimation of one more parameter than the score test and so should be slightly slower in most practical applications. However as we have already noted the score test, because it requires estimation under what may be an unreasonable null hypothesis, is more prone to estimation problems and numerical instability than the Wald test.

First we apply the score and Wald tests to repeated simulations from the two distribution used in Section 4. Sample sizes of 20, 50, and 100 were used. Figures 7 and 9 show the power of the 5% level score test for data from these two distributions. Figures 8 and 10 give the same curves for the Wald test. We use the W_N version of the Wald test for the bimodal data where we are interested in detecting both modes and the W_M version for the unimodal data where smaller secondary modes in the fitted distribution can be assumed to be spurious. Test sizes for these tests are presented in Table 1.

5.1. Simulation Results for the Score and Wald tests for Data from a Multimodal Distribution

The power curves for both the score and Wald tests in the bimodal data (Figures 7 and 8) show two minima corresponding to the two modes at 0 and 4.32. This is an unusual shape for a power curve but is not unreasonable given that we are interested in detecting all modes and so ideally there will be a low probability of rejecting H_0 if either of the modes is close to the hypothesised mode. For this reason, in Table 1, two ‘sizes’ are reported for each test in the case of the bimodal distribution. As expected both the tests become more powerful and the test sizes approach their asymptotic levels as sample size increases. The power functions for the score and Wald test have a very similar shape in this case and there appears to be little difference in their performance.

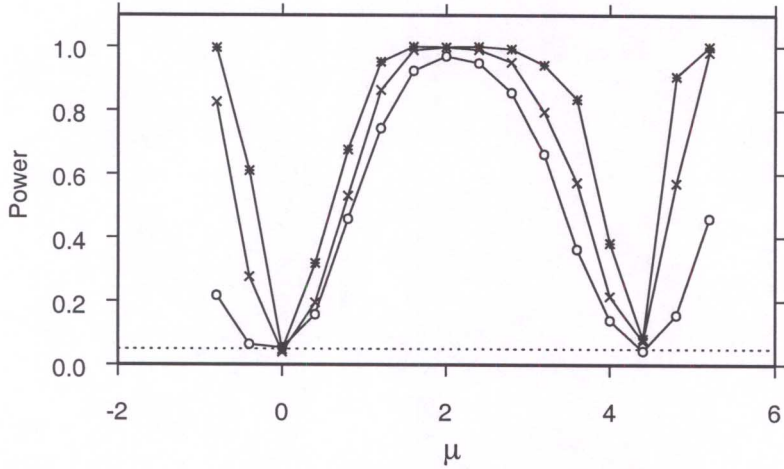


Figure 7. Comparison of power curves of the score test for samples of size 20 (circles), 50 (crosses) and 100 (stars) from the bimodal distribution $g_4(x; \mu, \theta)$ with $\theta = (0, 0.25, -0.03)^T$; $\alpha = 0.05$; 1000 simulations.

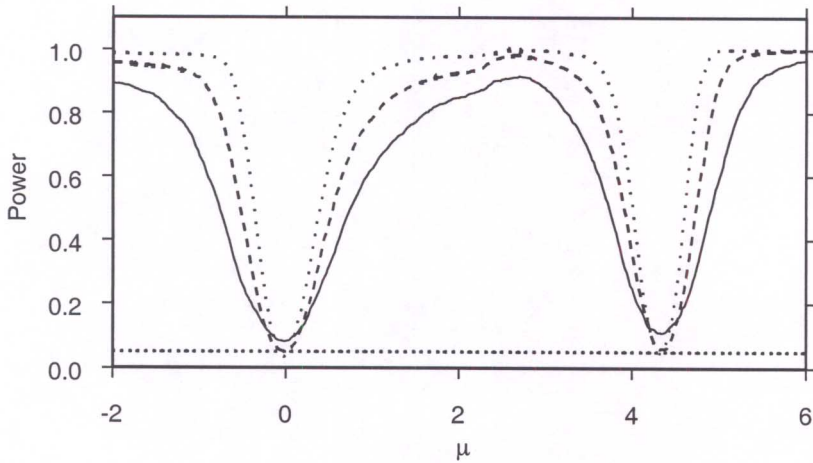


Figure 8. Comparison of power curves of the Wald test using the nearest mode technique (W_N) for samples of size 20 (solid), 50 (dashes) and 100 (dots) from the bimodal distribution $g_4(x; \mu, \theta)$ with $\theta = (0, 0.25, -0.03)^T$; $\alpha = 0.05$; 1000 simulations.

Table 1. Actual sizes of the score and Wald tests with a nominal test size of $\alpha = 0.05$ for 1000 simulated samples of size 20, 50 and 100 from $g_4(x; \mu, \theta)$ with $\theta = (0, 0.2, -0.03)^T$ and $\theta = (0, 0.25, -0.03)^T$.

θ	Mode	n	\hat{S}	\hat{W}_N	\hat{W}_M
$(0, 0.2, -0.03)^T$ (unimodal)	0	20	0.10	.	0.18
		50	0.06	.	0.12
		100	0.04	.	0.08
$(0, 0.25, -0.03)^T$ (bimodal)	0	20	0.05	0.08	.
		50	0.04	0.05	.
		100	0.05	0.04	.
	4.32	20	0.05	0.11	.
		50	0.07	0.06	.
		100	0.05	0.05	.

5.2. Simulation Results for the Score and Wald tests for Data from a Unimodal Distribution

For unimodal data there is only one minimum in each power function, as expected. Again as sample size is increased both tests become more powerful and sizes approach their asymptotic levels. Examining the test sizes in Table 1 it appears that the score test is performing better than the Wald test. This is due mainly to a test bias in the case of the Wald test. In both cases the test sizes based on the asymptotic χ^2 approximation do not appear satisfactory for sample sizes of less than 50. In these cases, for both the unimodal and multimodal data, we would recommend a resampling approach to obtain p-values. Bootstrap p-values can be obtained by resampling from the fitted model and looking at the quantiles of the empirical distribution of the test statistic. This of course can be quite computationally time consuming. In terms of power, especially for small sample sizes, the score test (Figure 9) performs poorly when μ is distant from 0. As μ increases past a value around 1.2 (the true value of μ being 0) the power even starts to decrease. Although the unimodal data does come from a g_4 distribution, at small sample sizes there is insufficient information in the data for the score test to discern between unimodal distributions with μ taking the value of the true mode and bimodal distributions with μ at a spurious mode and a second mode at the true value. As the sample size increases, distributions with spurious modes no longer have likelihood comparable to the true density, so the test becomes more powerful. The

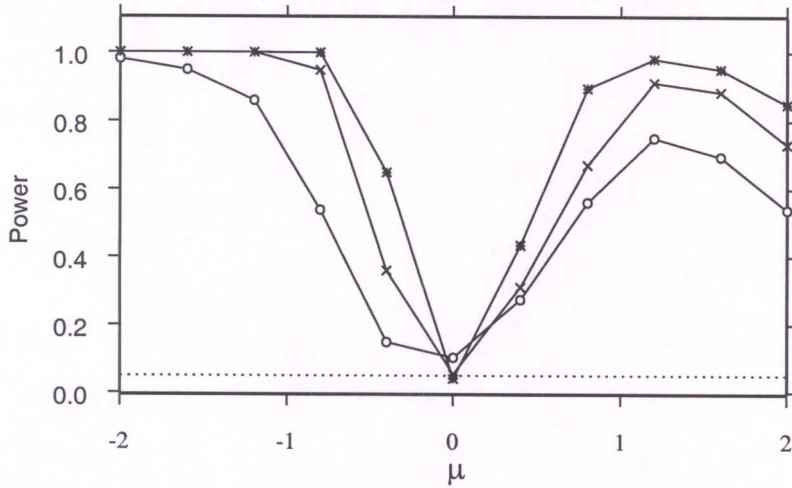


Figure 9. Comparison of power curves of the score test for samples of size 20 (circles), 50 (crosses) and 100 (stars) from the unimodal distribution $g_A(x; \mu, \theta)$ with $\theta = (0, 0.2, -0.03)^T$; $\alpha = 0.05$; 1000 simulations.

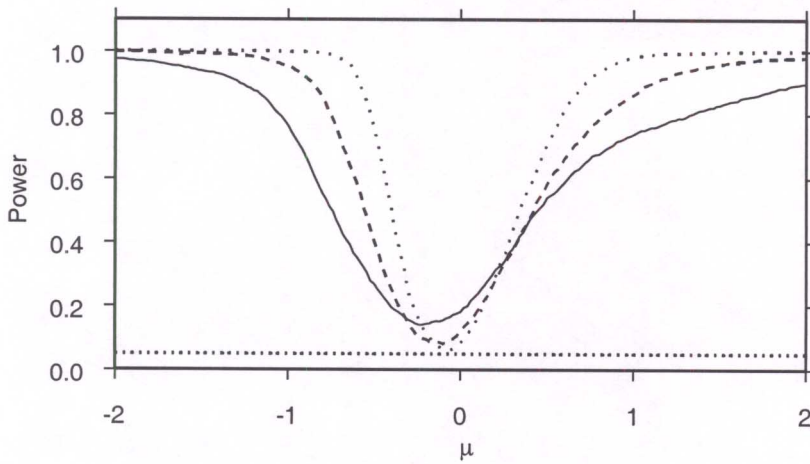


Figure 10. Comparison of power curves of the Wald test using maximum mode technique (W_M) for samples of size 20 (solid), 50 (dashes) and 100 (dots) from the unimodal distribution $g_A(x; \mu, \theta)$ with $\theta = (0, 0.2, -0.03)^T$; $\alpha = 0.05$; 1000 simulations.

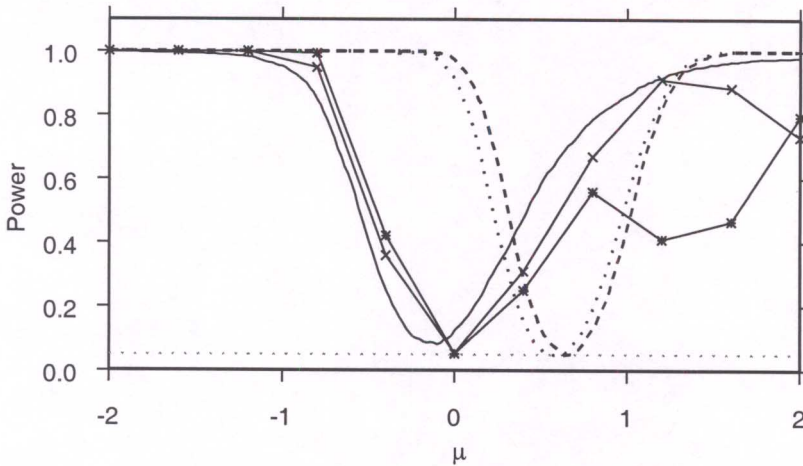


Figure 11. Comparison of power curves of the score test (both with and without restricted parameter space - crosses and stars respectively), the Wald test using the maximum mode technique (W_M - solid line), the t-test (dashed line) and Wilcoxon test (dotted line) for samples of size 50 from the unimodal distribution $g_A(x; \mu, \theta)$ with $\theta = (0, 0.2, -0.03)^T$; $\alpha = 0.05$, 1000 simulations.

Wald test using the maximum mode (W_M) overcomes this problem by ignoring spurious modes and so tends to be more powerful (Figure 10). The power curves for the Wald test indicate a slight test bias which diminishes as n becomes large.

In Figure 11 we present power curves for the score test, the Wald test and the score test with maximum likelihood estimates restricted to the unimodal region of the parameter space (see Section 2) for samples of size 50 from the unimodal distribution discussed above - that is, $g_A(x; \mu, \theta)$ where $\mu = 0$ and $\theta = (0, 0.2, -0.03)^T$. For the sake of comparison, we have also included the power curves of the t-test and Wilcoxon test, two often used tests of location. This comparison serves to demonstrate that for skewed data these tests are testing for different things and if it is the mode which is of interest then neither the t-test nor the Wilcoxon test is applicable. Figure 11 suggests that while the size of the test with restricted parameter space is reasonable, it has poor power properties, worse even than the unrestricted score test for most positive values of μ . This is due to erratic behaviour of the score statistic when one or more of the parameter estimates lie on the boundary of the permissible region.

Table 2. Actual powers of the t, Wilcoxon and Wald tests with a nominal test size of $\alpha = 0.05$ for 2000 simulated samples of size 50 from $N(0,1)$.

μ	\hat{W}_M $\theta_2, \theta_3, \theta_4$	\hat{W}_M θ_2, θ_4	\hat{W}_M θ_2	t-test	Wilcoxon test
0	0.14	0.08	0.06	0.05	0.05
0.2	0.19	0.33	0.30	0.29	0.27
0.4	0.33	0.81	0.81	0.79	0.77
0.6	0.51	0.99	0.99	0.99	0.98
0.8	0.69	1	1	1	1

5.3. Effect of Over Parameterisation

Care should be taken to keep the number of parameters in the model to the smallest reasonable value so as to reduce the risk of misidentifying a unimodal distribution as multimodal - especially if the score test is being used. The importance of a good choice of k is demonstrated clearly in Table 2 where we present powers of the Wald test (again W_M is used since the data is known to be unimodal) along with the corresponding powers of the t and Wilcoxon tests for data from a normal distribution. Using the Wald test with parameters θ_2, θ_3 and θ_4 represents a considerable over parameterisation and poor size and power properties result. When just θ_2 and θ_4 are used the fitted distribution is forced to be symmetric and there is a considerable improvement, although the test size is still slightly higher than the nominal size. When just θ_2 is used the Wald test should be asymptotically equivalent to the optimal test for normal data - the t-test. Applying a smooth goodness of fit test (see Rayner and Best [6]) may give some indication of the degree of departure from normality and an appropriate choice of k . In any case for skewed data it is necessary to include at least θ_2, θ_3 and θ_4 . On the other hand using more than these three parameters may be computationally impractical, especially if resampling is required. So if there is evidence of skewness and a test for modes is required then, unless the data is extremely multimodal, we would recommend using θ_2, θ_3 and θ_4 .

5.4. Conclusion

Standard tests of location like the t-test and Wilcoxon test provide poor tests for the modes of a skewed or multimodal distribution. Here we have shown that, for data that can be well approximated by a distribution of the form given by (2), the score and Wald tests based on this model are reasonably effective in identifying the modes. Using the score test with the parameter space restricted was shown not to have any benefit in terms of power over the unrestricted score test for unimodal data. The score and Wald tests were seen to have similar performances in terms of size and power. The fact that maximum likelihood estimates under the full model will often be useful in their own right and the advantages of the Wald test in terms of flexibility and interpretation suggest the Wald test may be the more appropriate of the two. The tests used in this paper have been implemented in S-PLUS. The authors would be happy to make their code available to any interested reader.

References

1. A. M. Carolan and J. C. W. Rayner. One Sample Tests of Location for Nonnormal Symmetric Data. *Commun. Statist.-Theor. Meth.*, 29(7), 1569-1581, 2000.
2. D. R. Cox and D. V. Hinkley. *Theoretical Statistics*. Chapman and Hall, London, 1974.
3. W. W. Daniel. *Applied Nonparametric Statistics*. PWS-Kent, Boston, 1978.
4. D. J. Hand, F. Daly, A. D. Lunn, K. J. McConway and E. Ostrowski. *A Handbook of Small Data Sets*. Chapman and Hall, London, 1994.
5. E. L. Lehmann. *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day, San Francisco, 1975.
6. J. C. W. Rayner and D. J. Best. *Smooth Tests of Goodness of Fit*. Oxford University Press, New York, 1989.
7. T. A. Ryan Jr., B. L. Joiner and B. F. Ryan. *The Minitab Student Handbook*. Duxbury Press, Boston, 1985.
8. B. W. Silverman. Using Kernel Density Estimates to Investigate Multimodality. *Journal of the Royal Statistical Society B*, 43:97-99, 1981.