Reliable Inference from Empirical Likelihoods

S.A. Corcoran and A.C. Davison

Department of Statistics, 1 South Parks Road, Oxford OX1 3TG, UK

and

R.H. Spady

Nuffield College, Oxford OX1 1Nffi, UK

Draft of March 24, 1995

Summary

Theoretical work has shown that similar asymptotic theory applies to empirical likelihood and parametric likelihood, and in particular that the higher-order properties of the empirical likelihood ratio statistic can be improved by Bartlett adjustment. We use simulation to show that for small and moderate samples the adjusted statistic performs worse than a related statistic that does not have the same 'good' higher-order behaviour. The general conclusion is that standard asymptotic calculations are not useful in understanding the small-sample behaviour of nonparametric likelihood statistics.

Some key words: Bartlett adjustment; Empirical likelihood; Empirical exponential family; Likelihood ratio statistic; Nonparametric likelihood.

1 Introduction

Theoretical work has established that in many situations the usual distributional approximations for likelihood quantities apply to empirical likelihood (Owen, 1988, 1990). For example, if $Y_1, \ldots, Y_n$ is a random sample of size n from an unknown distribution F with mean θ_0 and a finite third moment, the empirical likelihood ratio statistic $W_{EL}(\theta_0)$ satisfies

$$\mathrm{pr}\{W_{EL}(\theta_0) \leq c_p\} = p + O(n^{-1}), \tag{1}$$

where c_p is the p quantile of the chi-squared distribution with one degree of freedom, χ_1^2. Under stronger moment conditions, $W_{EL}(\theta_0)$ is Bartlett-correctable, that is, there exists a quantity b such that

$$\mathrm{pr}\{W_{EL}(\theta_0)(1 + \hat{b}/n)^{-1} \leq c_p\} = p + O(n^{-2}), \tag{2}$$

where $\hat{b}$ is an estimate of b. Equations (1) and (2) hold under the usual regularity conditions guaranteeing validity of Edgeworth expansions, i.e. existence of certain moments of F and that F satisfies Cramér's condition (Hall, 1992, p. 45). Thus in this situation the same likelihood approximations apply to $W_{EL}(\theta_0)$, a nonparametric likelihood ratio statistic, as would to the likelihood ratio statistic for a correct parametric model.

Figure 1 poses questions for this theory. To construct it we generated 5000 samples of sizes $n = 20$ and $n = 40$ from chi-squared distribution with one degree of freedom, and compared the quantiles of several statistics based on those samples with their asymptotic distribution. In each panel the solid vertical lines are at $c_{0.95}$ and $c_{0.99}$, and the solid 45° line represents the asymptotic quantiles of the statistic as $n \to \infty$. In the top left panel the solid lines above the diagonal are quantiles of $W_{EL}(\theta_0)$, while quantiles of the Bartlett-adjusted statistics are shown as dotted lines. The chi-squared approximation is poor at the higher quantiles, whose curvature shows that simple scale adjustment as in (2) cannot make a dramatic difference. There is a marked contrast with the top right panel, which shows that for both sample sizes the parametric likelihood ratio statistic is so close to χ_1^2 that Bartlett adjustment will make little difference. Thus although (1) and (2) apply to both parametric and empirical likelihoods, they are not equally relevant to practice even in this very simple case. We discuss the lower panels below.

Approximations like (1) and (2) are usually used to set confidence intervals for θ_0. For this purpose the coverage of confidence sets based on inequalities such as $W(\theta)_{EL} \leq c_{0.90}$ is more important than the precise distribution of the empirical likelihood ratio statistic. Table 1 shows the coverage of such sets for various situations, based on $W_{EL}(\theta_0)$ without Bartlett adjustment, the modified form $W_{EL}(\theta_0)(1 + b/n)^{-1}$ with the theoretical Bartlett adjustment, and the modified form $W_{EL}(\theta_0)(1 + \hat{b}/n)^{-1}$ in which b is replaced by an estimate. Both modifications give better coverage, although a smaller improvement is given by the version that would be used in practice, involving $\hat{b}$.

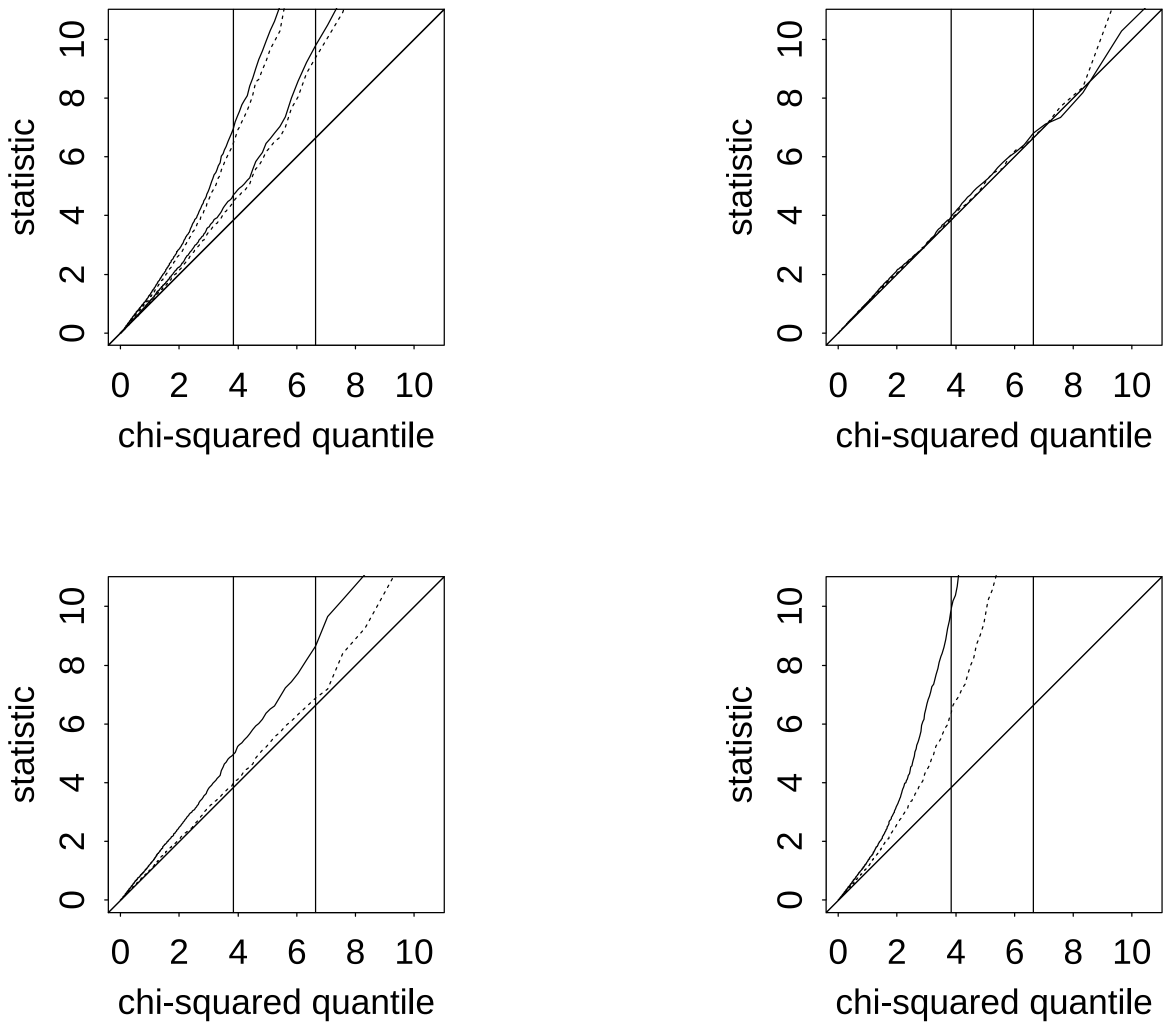

Figure 1: Empirical quantiles of likelihood ratio and robust Wald statistics plotted against quantiles of their asymptotic distribution, for 5000 samples of sizes 20 and 40 from the χ^2_1 distribution. The top left panel shows quantiles for the empirical likelihood ratio statistic without (solid) and with Bartlett adjustment (dots); the lines closer the diagonal are for $n = 40$. The top right panel shows quantiles for the parametric likelihood ratio statistics. The bottom panels show quantiles for empirical and parametric robust Wald statistics defined in §3. In each panel the solid diagonal line gives the quantiles for $n = \infty$, and the solid vertical lines give the 0.95 and 0.99 quantiles of the limiting χ^2_1 distribution. In the top right and bottom panels the solid line is for $n = 20$.

Table 1: Observed coverages (%) for two-sided confidence intervals for θ_0 based on the empirical likelihood ratio statistic, W_{EL}, with theoretical and estimated Bartlett adjustment, and on a robust Wald statistic T^2_{EEF} for an empirical exponential family. The results are obtained from 5000 simulated samples of size n from the standard normal, χ^2_1 and t_5 distributions. The coverages in parentheses are from DiCiccio, Hall and Romano (1991). The Monte Carlo standard errors for columns 3–6 are 0.57, 0.42, 0.31 and 0.14.

		Coverage for $N(0,1)$ data			
n	Statistic	80%	90%	95%	99%
10	W_{EL}	74.34 (74.70)	84.46 (84.46)	90.14 (89.64)	95.70 (95.42)
	$W_{EL}(1+b/n)^{-1}$	77.88 (77.96)	87.40 (87.06)	92.08 (91.82)	96.82 (96.50)
	$W_{EL}(1+\hat{b}/n)^{-1}$	76.94 (79.38)	86.72 (88.02)	91.60 (92.46)	96.50 (96.96)
	T^2_{EEF}	78.58	89.14	94.74	99.32
20	W_{EL}	77.52 (78.34)	86.88 (88.30)	93.04 (93.42)	98.00 (98.04)
	$W_{EL}(1+b/n)^{-1}$	79.00 (80.06)	88.20 (89.62)	94.30 (94.18)	98.30 (98.44)
	$W_{EL}(1+\hat{b}/n)^{-1}$	78.86 (80.34)	87.84 (89.80)	94.02 (94.24)	98.24 (98.48)
	T^2_{EEF}	79.56	89.50	95.82	99.24
		Coverage for χ^2_1 data			
n	Statistic	80%	90%	95%	99%
20	W_{EL}	73.90 (73.14)	83.54 (82.88)	88.66 (88.78)	94.92 (95.36)
	$W_{EL}(1+b/n)^{-1}$	79.38 (78.72)	87.90 (87.72)	92.46 (92.62)	96.96 (97.06)
	$W_{EL}(1+\hat{b}/n)^{-1}$	75.66 (76.34)	84.82 (85.46)	90.18 (90.34)	95.60 (96.16)
	T^2_{EEF}	75.70	86.46	92.16	97.86
40	W_{EL}	76.48 (76.44)	86.64 (86.80)	91.98 (92.36)	97.40 (97.40)
	$W_{EL}(1+b/n)^{-1}$	79.30 (79.10)	88.56 (88.96)	93.44 (94.18)	97.90 (98.00)
	$W_{EL}(1+\hat{b}/n)^{-1}$	77.94 (78.04)	87.48 (87.88)	92.64 (93.34)	97.74 (97.74)
	T^2_{EEF}	78.08	88.30	94.26	98.86
		Coverage for t_5 data			
n	Statistic	80%	90%	95%	99%
15	W_{EL}	74.96 (75.16)	85.40 (85.02)	91.26 (90.94)	97.32 (96.92)
	$W_{EL}(1+b/n)^{-1}$	82.54 (82.66)	91.40 (91.06)	95.70 (95.44)	98.74 (98.62)
	$W_{EL}(1+\hat{b}/n)^{-1}$	77.06 (78.98)	87.36 (88.84)	92.84 (93.48)	97.72 (97.94)
	T^2_{EEF}	77.34	89.36	95.32	99.42
30	W_{EL}	77.70 (77.68)	87.70 (87.84)	92.84 (93.08)	97.84 (97.80)
	$W_{EL}(1+b/n)^{-1}$	80.90 (81.14)	90.08 (90.42)	94.66 (94.96)	98.50 (98.66)
	$W_{EL}(1+\hat{b}/n)^{-1}$	79.02 (79.54)	88.56 (89.24)	93.66 (94.22)	98.16 (98.32)
	T^2_{EEF}	79.26	89.50	94.92	98.98

DiCiccio, Hall and Romano (1991) were the first to show that the coverage error for confidence regions based on $W_{EL}(\theta)$ could be reduced from $O(n^{-1})$ to $O(n^{-2})$ by Bartlett adjustment. Simulation results from their paper are given in parentheses in our Table 1. Their results agree closely with ours except for Bartlett adjusted results for normal data. But if the labels on these rows are switched in their Table 1, their results agree with ours to Monte Carlo accuracy, and we find that results for theoretical Bartlett adjustment are generally superior to those for estimated adjustment, as seems plausible.

Nonparametric likelihood inferences can be based on statistics other than $W_{EL}(\theta_0)$. The lower left panel of Figure 1 shows the quantiles of $T^2_{EEF}(\theta_0)$, a robust Wald statistic for an empirical exponential family (DiCiccio and Efron, 1992). The basis for this is outlined in Section 3. The lower right panel shows the corresponding parametric Wald statistic, which is clearly further from its asymptotic distribution. The figure suggests that $T^2_{EEF}(\theta_0)$ matches its asymptotic distribution more closely than $W_{EL}(\theta_0)$, and this is confirmed by Table 1: coverages for $T^2_{EEF}(\theta_0)$ are everywhere closer to the nominal levels than those for $W_{EL}(\theta_0)$ and $W_{EL}(\theta_0)(1+\hat{b}/n)^{-1}$. Moreover, the straightness of the empirical quantiles in the lower left panel shows that scale adjustment can improve the coverage of $T^2_{EEF}(\theta_0)$. However, calculations outlined in Section 3 show that

$$\text{pr}\left\{T^2_{EEF}(\theta_0) \le c_p\right\} = p + O(n^{-1}), \tag{3}$$

so that the asymptotic error is larger than in (2). Thus results such as (2) and (3) are of doubtful practical relevance.

The purpose of this paper is to investigate this further. We use $W_{EL}(\theta)$ and $T^2_{EEF}(\theta_0)$ as examples to show that in the nonparametric context conventional asymptotics do not help to predict when approximations such as (1), (2) and (3) will be reliable. In Section 2 we outline the basis of empirical likelihood, and give some theoretical results that underlie (2). Section 3 describes some related statistics, with the corresponding theoretical results. Section 4 contains some concluding discussion.

2 Empirical Likelihood

2.1 Definition

Let $y_1, \ldots, y_n$ be the observed values of a random sample from a continuous univariate distribution F with mean θ_0 and finite third moment. The log empirical likelihood $\ell_{EL}(\theta)$ is obtained by maximizing $\sum \log p_j$ subject to the constraint that $\sum p_j y_j = \theta$, where $p_1, \ldots, p_n$, is the multinomial distribution that puts probability mass p_j on y_j. Straightforward application of Lagrange multipliers shows that

$$p_j = \frac{1}{n\{1+\lambda_\theta(y_j-\theta)\}}, \quad j = 1, \ldots, n, \tag{4}$$

where λ_θ satisfies the equation

$$\sum \frac{y_j - \theta}{1 + \lambda_\theta (y_j - \theta)} = 0. \tag{5}$$

Expression (5) has a solution only when $\min y_j < \theta < \max y_j$, and $\ell_{EL}(\theta)$ is defined only in this range. The log empirical likelihood is maximized at $\theta = \bar{y}$, the sample average, at which point $\lambda_\theta = 0$, and $p_j \equiv n^{-1}$. The empirical likelihood ratio statistic for θ is

$$W_{EL}(\theta) = 2 \sum \log\{1 + \lambda_\theta (y_j - \theta)\}, \quad \min y_j < \theta < \max y_j. \tag{6}$$

For any n there is positive probability that the interval $(\min y_j, \max y_j)$ will not contain θ_0, in which case $W_{EL}(\theta_0)$ is infinite. Thus $\mathrm{E}\{W_{EL}(\theta_0)\}$ is infinite. Nonetheless, as DiCiccio *et al.*(1991) point out, $W_{EL}(\theta_0)$ can be approximated in a distributional sense by a Taylor series expansion, and adjustment of $W_{EL}(\theta_0)$ by the expected value of the leading terms of this expansion improves the accuracy of the distributional approximation.

2.2 Asymptotic properties

Asymptotic analysis of $W_{EL}(\theta)$ starts from inversion of (5) to obtain an expansion for λ_θ in terms of $\bar{y} - \theta$, $n^{-1}\sum(y_j - \bar{y})^2$, and so forth. Substitution of this series into (6) and replacement of θ by $\theta_0 + n^{-1/2}\delta$ leads to a series expansion for $W_{EL}(\theta_0 + n^{1/2}\delta)$ in descending powers of $n^{1/2}$ that depends on δ and the empirical moments of $y_1, \ldots, y_n$. Conditions on the moments of F beyond those in Section 2 are needed for what follows: we assume that any necessary moments are finite.

The higher-order properties of $W_{EL}(\theta)$ are best studied via those of the expansion for the corresponding directed likelihood statistic, $R_{EL}(\theta) = \mathrm{sign}(\bar{y} - \theta) W_{EL}(\theta)^{1/2}$. Expansions along the lines above lead to expressions such as

$$R_{EL}(\theta_0) = Z + n^{-1/2} \frac{k_3}{3k_2^{3/2}} Z^2 + n^{-1} \frac{1}{36 k_2^3} \left(16k_3^2 - 9k_2k_4 - 9k_2^3\right) Z^3 + O_p(n^{-3/2}), \tag{7}$$

where $Z = n^{-1/2} k_2^{-1/2} (\bar{y} - \theta_0)$, and $k_2 = n^{-1}\sum(y_j - \bar{y})^2$, $k_3 = n^{-1}\sum(y_j - \bar{y})^3$, and $k_4 = n^{-1}\sum(y_j - \bar{y})^4 - 3k_2^2$ are empirical cumulants of $y_1, \ldots, y_n$. A tedious calculation straightforwardly implemented in `Mathematica 2.0` (Wolfram, 1991) gives the leading terms of the cumulants of $R_{EL}(\theta_0)$. These are shown in Table 2 to an accuracy of $O(n^{-5/2})$, for reasons we shall describe later.

Let ω_r denote the rth cumulant of $R_{EL}(\theta_0)$. Then standard identities (McCullagh, 1987, p. 254) can be used to show that the mean and variance of $W_{EL}(\theta_0)$ are $\omega_2 + \omega_1^2$ and $\omega_4 + 4\omega_3\omega_1 + 2\omega_2^2 + 4\omega_2\omega_1^2$. Now we see from Table 2 that the leading terms of the ω_r can be written as $\omega_1 = a_1 n^{-1/2}$, $\omega_2 = 1 + a_2 n^{-1}$, $\omega_3 = a_3 n^{-3/2}$ and $\omega_4 = a_4 n^{-2}$, where $a_r \neq 0$. Therefore $\mathrm{E}\{W_{EL}(\theta_0)\} = 1 + (a_1^2 + a_2)n^{-1} + O(n^{-2})$ and

$$\mathrm{var}\{W_{EL}(\theta_0)\} = 2 \left[\mathrm{E}\left\{W_{EL}(\theta_0)\right\}\right]^2 + O(n^{-2}). \tag{8}$$

Table 2: Cumulants of directed empirical likelihood statistics.

$R_{EL}(\theta_0)$	
First	$-n^{-1/2}\kappa_3/(6\kappa_2^{3/2}) + n^{-3/2}(-420\kappa_2^3\kappa_3 - 20\kappa_3^3 + 75\kappa_2\kappa_3\kappa_4 - 66\kappa_2^2\kappa_5)/(240\kappa_2^{9/2})$
Second	$1 + n^{-1}(54\kappa_2^3 - 13\kappa_3^2 + 18\kappa_2\kappa_4)/(36\kappa_2^3) + n^{-2}(13680\kappa_2^6 - 25920\kappa_1\kappa_2^4\kappa_3 - 420\kappa_2^3\kappa_3^2$ $-20\kappa_3^4 + 4320\kappa_1^2\kappa_2^3\kappa_4 + 15480\kappa_2^4\kappa_4 + 2880\kappa_1\kappa_2^2\kappa_3\kappa_4 + 795\kappa_2\kappa_3^2\kappa_4 - 360\kappa_2^2\kappa_4^2 - 4320\kappa_1\kappa_2^3\kappa_5 - 1506\kappa_2^2\kappa_3\kappa_5 + 1200\kappa_2^3\kappa_6)/(720\kappa_2^6)$
Third	$n^{-3/2}(-10530\kappa_2^3\kappa_3 - 1255\kappa_3^3 + 3240\kappa_1\kappa_2^2\kappa_4 + 2970\kappa_2\kappa_3\kappa_4 - 1944\kappa_2^2\kappa_5)/(1080\kappa_2^{9/2})$
Fourth	$n^{-2}(72900\kappa_2^6 - 492480\kappa_1\kappa_2^4\kappa_3 - 2700\kappa_2^3\kappa_3^2 - 5395\kappa_3^4 + 82080\kappa_1^2\kappa_2^3\kappa_4 + 91800\kappa_2^4\kappa_4 + 39600\kappa_1\kappa_2^2\kappa_3\kappa_4 + 17820\kappa_2\kappa_3^2\kappa_4$ $-6300\kappa_2^2\kappa_4^2 - 77760\kappa_1\kappa_2^3\kappa_5 - 14256\kappa_2^2\kappa_3\kappa_5 + 8640\kappa_2^3\kappa_6)/(1080\kappa_2^6)$
$T_{EEF}(\theta_0)$	
First	$n^{-3/2}(-48\kappa_2^3\kappa_3 - 87\kappa_3^3 + 80\kappa_2\kappa_3\kappa_4 - 21\kappa_2^2\kappa_5)/(24\kappa_2^{9/2})$
Second	$1 + n^{-1}(-12\kappa_2^3 + 3\kappa_3^2 - 4\kappa_2\kappa_4)/(4\kappa_2^3) + n^{-2}(756\kappa_2^6 - 504\kappa_1\kappa_2^4\kappa_3 - 267\kappa_2^3\kappa_3^2 + 1290\kappa_3^4 + 84\kappa_1^2\kappa_2^3\kappa_4 + 1332\kappa_2^4\kappa_4$ $+72\kappa_1\kappa_2^2\kappa_3\kappa_4 - 836\kappa_2\kappa_3^2\kappa_4 - 55\kappa_2^2\kappa_4^2 - 84\kappa_1\kappa_2^3\kappa_5 - 196\kappa_2^2\kappa_3\kappa_5 + 156\kappa_2^3\kappa_6)/(12\kappa_2^6)$
Third	$n^{-1/2}\kappa_3/\kappa_2^{3/2} + n^{-3/2}3(-50\kappa_2^3\kappa_3 - 104\kappa_3^3 + 8\kappa_1\kappa_2^2\kappa_4 + 85\kappa_2\kappa_3\kappa_4 - 20\kappa_2^2\kappa_5)/(8\kappa_2^{9/2})$
Fourth	$n^{-1}6(-3\kappa_2^3 + \kappa_3^2 - \kappa_2\kappa_4))/\kappa_2^3 + n^{-2}(12528\kappa_2^6 - 12960\kappa_1\kappa_2^4\kappa_3 - 6696\kappa_2^3\kappa_3^2 + 18405\kappa_3^4 + 2160\kappa_1^2\kappa_2^3\kappa_4 + 18288\kappa_2^4\kappa_4$ $+1824\kappa_1\kappa_2^2\kappa_3\kappa_4 - 13728\kappa_2\kappa_3^2\kappa_4 + 304\kappa_2^2\kappa_4^2 - 2064\kappa_1\kappa_2^3\kappa_5 - 2268\kappa_2^2\kappa_3\kappa_5 + 1824\kappa_2^3\kappa_6)/(24\kappa_2^6)$

It can be established from the pattern of higher-order cumulants that the rth cumulant of $W_{EL}(\theta_0)$ equals $\gamma_r \mathrm{E}\{W_{EL}(\theta_0)\}^r + O(n^{-2})$, where γ_r is the rth cumulant of the χ_1^2 distribution. Hence division of $W_{EL}(\theta_0)$ by $1+(a_1^2+a_2)n^{-1}$ yields a statistic whose distribution is χ_1^2 apart from terms of $O(n^{-2})$. It follows that the same approximation applies when a_1 and a_2 are replaced by $\sqrt{n}$-consistent estimates (Barndorff-Nielsen and Hall, 1988). We need conditions guaranteeing the existence of certain Edgeworth expansions to deduce the coverage result (2) from these calculations.

3 Related Statistics

3.1 Likelihood-based statistics

Consider inference for the mean θ of a parametric distribution F based on a random sample $y_1, \ldots, y_n$ in which y_j makes contribution $\ell_j(\theta)$ to the overall log likelihood $\ell(\theta)$. Suppose that the true value of θ is θ_0. The likelihood ratio statistic $W(\theta_0) = 2\{\ell(\hat\theta) - \ell(\theta_0)\}$ is one of a number of asymptotically equivalent statistics in common use in this setting (Cox and Hinkley, 1974, Chapter 9). Others are the score and Wald statistics

$$U(\theta_0) = I(\theta_0)^{-1/2}\frac{\partial \ell(\theta_0)}{\partial\theta}, \quad S(\theta_0) = I(\hat\theta)^{1/2}(\hat\theta - \theta_0). \tag{9}$$

where $I(\theta) = -\partial^2\ell(\theta)/\partial\theta^2$ is the observed information at θ and $\hat\theta$ is the maximum likelihood estimate. Both $U(\theta_0)$ and $S(\theta_0)$ depend on the standard large-sample distributions for $\hat\theta$ and $\partial\ell(\theta_0)/\partial\theta$. A further statistic is based on a robust approximation to the asymptotic variance of $\hat\theta$, the so-called sandwich estimator $I^{-1}(\theta_0)J(\theta_0)I^{-1}(\theta_0)$, where $J(\theta) = \sum\{\partial\ell_j(\theta)/\partial\theta\}^2$ (Royall, 1990). The resulting robust Wald statistic is

$$T(\theta_0) = I^{-1}(\hat\theta)J^{1/2}(\hat\theta)(\hat\theta - \theta_0). \tag{10}$$

Although $W(\theta_0)$, $S(\theta_0)$, $T(\theta_0)$, and $U(\theta_0)$ are most common in practice, variants can be obtained by replacing $I(\hat{\theta})$ and $J(\hat{\theta})$ by $I(\theta_0)$ and $J(\theta_0)$ in $S(\theta_0)$ and $T(\theta_0)$, and by using the expected information $\mathrm{E}\{I(\theta_0)\}$ in (9). The second-order properties of these various statistics are somewhat different, but provided that F satisfies suitable regularity conditions, $S(\theta_0)$, $T(\theta_0)$, $U(\theta_0)$ and $R(\theta_0) = \mathrm{sign}(\hat{\theta} - \theta_0)W(\theta_0)^{1/2}$ all have standard normal distributions in large samples. In a particular context, the choice among them and their variants will depend on factors such as higher-order properties, ease of computation, invariance to reparametrization, and whether or not a robust analysis is required.

In a nonparametric setting the statistics described above may be calculated for empirical likelihood, under regularity conditions on F. It can readily be checked that for empirical likelihood the two forms of observed information are equal, i.e. $J_{EL}(\hat{\theta}) = I_{EL}(\hat{\theta})$, so $T_{EL}(\theta_0) = S_{EL}(\theta_0)$.

3.2 Empirical exponential family likelihood

In an empirical exponential family, (4) is replaced by

$$p_j \propto e^{\lambda_\theta (y_j - \theta)}, \quad j = 1, \ldots, n, \tag{11}$$

where the constant of proportionality is chosen so that $\sum p_j = 1$, and λ_θ satisfies the analogue of (5), namely

$$\sum_{j=1}^{n} (y_j - \theta) e^{\lambda_\theta (y_j - \theta)} = 0. \tag{12}$$

The corresponding likelihood ratio statistic is

$$W_{EEF}(\theta) = 2 \sum_{j=1}^{n} \left\{ 1 - e^{\lambda_\theta (y_j - \theta)} \right\}. \tag{13}$$

Such families have been discussed by Efron (1982, Ch. 10), DiCiccio and Romano (1990), and DiCiccio and Efron (1992), among others.

Statistics based on (11) are readily calculated *via* a connection to the Poisson distribution. Let Z_1, …, Z_n be independent Poisson variables with means $\exp\{\lambda x_j(\theta)\}$, where for a given θ we set $x_j(\theta) = y_j - \theta$. Then $W_{EEF}(\theta)$ is the deviance for this Poisson model, if the observed values of the Z_j are all zero. Thus (13) and related statistics may obtained by Poisson regression.

The statistics discussed in Section 3.1 can be defined for empirical exponential family likelihoods. We focus on a version of the robust Wald statistic, defined for λ_θ. We set

$$T_{EEF}(\theta) = I^{-1}(\lambda_\theta) J(\lambda_\theta)^{1/2} (\lambda_{\bar{y}} - \lambda_\theta) = -I^{-1}(\lambda_\theta) J(\lambda_\theta)^{1/2} \lambda_\theta, \tag{14}$$

which is the analogue of (10), but defined in terms of λ_θ; note that $\lambda_{\bar{y}} = 0$. In (13) we set

$$J(\lambda) = \sum_{j=1}^{n} \left\{ \frac{\partial \ell_j}{\partial \lambda} \right\}^2, \quad I(\lambda) = -\sum_{j=1}^{n} \frac{\partial^2 \ell_j}{\partial \lambda^2},$$

where for a given value of θ we treat λ as a parameter in the Poisson log likelihood with fixed covariates $x_j(\theta)$.

Standard arguments along the lines outlined in §2.2 give the cumulants of $T_{EEF}(\theta_0)$ and related statistics, which are given in Table 2. The first four cumulants of $T_{EEF}(\theta_0)$ are $O(n^{-3/2})$, $1 + O(n^{-1})$, $O(n^{-1/2})$ and $O(n^{-1})$. It is straightforward to verify that $T^2_{EEF}(\theta_0)$ does not satisfy (8), so division of $T^2_{EEF}(\theta_0)$ by its expectation does not yield (2).

4 Discussion

Since it seems futile to investigate the higher-order properties of more complicated statistics until the simplest case is well-understood, we have considered only the mean of a univariate sample. In this case the empirical likelihood ratio statistic is worse approximated by a χ^2_1 distribution than is the robust Wald statistic $T^2_{EEF}(\theta_0)$. More extensive simulations using various nonparametric analogues of the statistics outlined in Section 3.1 have shown that the simulations presented in Table 1 are representative: $T^2_{EEF}(\theta_0)$ performed best of all the statistics considered. However the usual asymptotic calculations summarized by the contents of Table 2 would not lead us to this conclusion. Since the purpose of studying the higher-order properties of a statistic is to guide its use in practice, we are led to question the usefulness of asymptotic results in this context.

What has gone wrong? One difficulty is that the Taylor series approximation $1 + b/n$ to the mean of $\mathrm{E}\{W_{EL}(\theta_0)\}$ is poor for small samples. Table 3 gives the value of $1+b/n$ and the average of N simulated $W_{EL}(\theta_0)$'s, for random samples from the χ^2_1 distribution. We took $N = 5000$ for $n = 20$, 40, 80, 120, 160, and $N = 15000$ thereafter; $W_{EL}(\theta_0)$ was always finite in our simulations. The theoretical approximation is always closer to one than the simulated value. The table also contains a higher-order theoretical value for the mean, $1 + b/n + b'/n^2$, whose performance is worse. The corresponding figures for $T^2_{EEF}(\theta_0)$ show that by $n = 80$ the simulated average is close to one, and that the theoretical values are poor for $n < 200$. Here the simulated value is always closer to one than is the theoretical $O(n^{-1})$ value.

Are our results wrong? The main theoretical results were obtained by two of us independently using different approaches to the computer algebra, and where there is overlap they agree with the hand calculations of DiCiccio *et al.*(1991). And as mentioned in Section 1, their simulated coverages for $W_{EL}(\theta_0)$ agree with ours, apart from a minor difference. Our simulated results too were obtained by two of us independently. Overall we think it unlikely that our results are seriously incorrect.

Is empirical likelihood Bartlett-correctable? Try as we might, we have been unable to find a mistake in DiCiccio *et al.*(1991), and we believe that Bartlett adjustment of empirical likelihood does reduce asymptotic coverage error. Now, as Jensen (1993) has pointed out in the parametric case, there are valid Edgeworth-type expansions that justify Bartlett adjust-

Table 3: Theoretical and simulated means for $W_{EL}(\theta_0)$ and $T^2_{EEF}(\theta_0)$, for samples of size n from the χ^2_1 distribution. Simulated results for $n \leq 160$ are based on 5000 simulations, and for $n \geq 200$ on 15000 simulations. The simulation standard error is given as SE.

		n						
		20	40	80	120	160	200	400
$W_{EL}(\theta_0)$	Theoretical $O(n^{-1})$	1.242	1.121	1.060	1.040	1.030	1.024	1.012
	Theoretical $O(n^{-2})$	2.464	1.426	1.136	1.074	1.049	1.036	1.015
	Simulated	1.760	1.288	1.108	1.053	1.070	1.042	1.013
	SE	0.067	0.029	0.023	0.022	0.021	0.012	0.012
$T^2_{EEF}(\theta_0)$	Theoretical $O(n^{-1})$	0.550	0.775	0.888	0.925	0.944	0.955	0.978
	Theoretical $O(n^{-2})$	9.598	3.037	1.453	1.176	1.085	1.045	1.000
	Simulated	1.235	1.115	1.022	0.989	1.021	0.997	0.986
	SE	0.026	0.022	0.020	0.020	0.019	0.011	0.011

ment for likelihood ratio statistics based on data from a continuous distribution F (Chandra and Ghosh, 1979) and also when F has a discrete component, provided that the likelihood ratio does not depend on the discrete part in an essential way. When F is discrete, however, Siotani and Fujikoshi (1984) have shown that the error in the analogue of (1) may be as large as $O(n^{-1/2})$, and Frydenberg and Jensen (1989) have shown numerically that Bartlett adjustment is not necessarily useful. Empirical likelihoods are a hybrid, where a discrete multinomial distribution is placed on a sample assumed to be from a continuous underlying distribution, and perhaps it is not surprising that standard asymptotics do not tell the full story.

One answer to concerns about using confidence regions based on (1) or (2) would be to bootstrap $W_{EL}(\theta_0)$ and replace c_p with the p quantile of the bootstrap replicates. While this will overcome the inaccuracy of the chi-squared approximation, it seems unnecessarily computer-intensive when $T^2_{EEF}(\theta_0)$ provides confidence regions that are accurate enough for many practical purposes.

Acknowledgements

ACD and SAC gratefully acknowledge support from the UK Engineering and Physical Sciences Research Council, through a research award, an Advanced Research Fellowship and a Research Studentship. RHS acknowledges support from the UK Economic and Social Research Council. We thank Dr J.E. Stafford for advice about computer algebra.

References

Barndorff-Nielsen, O. and Hall, P. (1988) On the level error after Bartlett adjustment of the likelihood ratio statistic. *Biometrika*, **75**, 374–378.

Chandra, T.P. and Ghosh, J.K. (1979) Valid asymptotic expansions for the likelihood ratio and other perturbed chi-square variables. *Sankhyā A*, **41**, 22–47.

Cox, D.R. and Hinkley, D.V. (1974) *Theoretical Statistics.* London: Chapman and Hall.

DiCiccio, T.J. and Efron, B. (1992) More accurate confidence intervals in exponential families. *Biometrika*, **79**, 231–245.

DiCiccio, T.J., Hall, P. and Romano, J.P. (1991) Empirical likelihood is Bartlett-correctable. *Ann. Statist.*, **19**, 1053–1061.

DiCiccio, T.J. and Romano, J.P. (1990) Nonparametric confidence limits by resampling methods and least favourable families. *Int. Statist. Rev.*, **58**, 59–76.

Efron, B. (1982) *The Jackknife, the Bootstrap, and Other Resampling Plans*, monograph 38. Philadelphia: SIAM.

Frydenberg, M. and Jensen, J.L. (1989) Is the 'improved likelihood ratio statistic' really improved in the discrete case? *Biometrika*, **76**, 655–661.

Hall, P. (1992) *The Bootstrap and Edgeworth Expansion.* New York: Springer.

Jensen, J.L. (1993) A historical sketch and some new results on the improved log likelihood ratio statistic. *Scand. J. Statist.*, **20**, 1–15.

Owen, A.B. (1988) Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237–249.

Owen, A.B. (1990) Empirical likelihood ratio confidence regions. *Ann. Statist.*, **18**, 90–120.

Royall, (1990) Model robust confidence intervals using maximum likelihood estimators. *Int. Statist. Rev.*, **54**, 221–226.

Siotani, M. and Fujikoshi, Y. (1984) Asymptotic approximations for the distribution of multinomial goodness-of-fit statistics. *Hiroshima Math. J.*, **14**, 115-124.

Wolfram, S. (1991) *Mathematica: A System for Doing Mathematics by Computer, 2nd Ed.* Addison Wesley, New York.