Biometrika Advance Access published January 14, 2010

Biometrika (2009), *pp*. 1–8
doi: 10.1093/biomet/asp074

Pseudo-score confidence intervals for parameters in discrete statistical models

By ALAN AGRESTI

Department of Statistics, University of Florida, Gainesville, Florida 32611, U.S.A.
aa@stat.ufl.edu

AND EUIJUNG RYU

Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, Minnesota 55905, U.S.A.
ryu.euijung@mayo.edu

Summary

We propose pseudo-score confidence intervals for parameters in models for discrete data. The confidence interval is obtained by inverting a test that uses a Pearson chi-squared statistic to compare fitted values for the working model with fitted values of the model when a parameter of interest takes various fixed values. For multinomial models, the pseudo-score method simplifies to the score method when the model is saturated and otherwise it is asymptotically equivalent to score and likelihood ratio test-based inferences. For cases in which ordinary score methods are impractical, such as when the likelihood function is not an explicit function of model parameters, the pseudo-score method is feasible. We illustrate the method for four such examples. Generalizations of the method are also presented for future research, including inference for complex sampling designs using a quasilikelihood Pearson statistic that compares fitted values for two models relative to the variance of the observations under the simpler model.

Some key words: Categorical data; Complex sampling; Contingency table; Multinomial model; Pearson chi-squared statistic; Quasilikelihood.

1. Introduction

For contingency tables, large-sample score-test-based inference performs well, often even for relatively small samples. For goodness-of-fit tests of generalized linear models, the score test statistic is the Pearson statistic comparing observed and fitted values (Smyth, 2003). For interval estimation of a parameter in a multinomial model, we propose a pseudo-score method based on inverting a Pearson statistic that compares the model-fitted values to the values fitted assuming a particular value of that parameter. The inferences are asymptotically equivalent alternatives to actual score and likelihood ratio test-based inferences. When score confidence intervals are difficult to construct, such as when the likelihood function is not an explicit function of the model parameters, it is still possible to conduct pseudo-score inference.

After introducing the pseudo-score method and discussing the connection between ordinary score and pseudo-score inferences, we illustrate pseudo-score confidence intervals for four models for which the ordinary score method is difficult to implement. The emphasis of this article is on modelling categorical response variables having a multinomial distribution. However, we also present a generalization of the pseudo-score method for complex sampling designs and suggest future research on a quasilikelihood generalization that uses an extended Pearson statistic comparing fitted values for two nested models relative to the estimated covariance matrix of the observations under the simpler model. Such methods are useful in applications for which likelihood-based methods such as profile likelihood confidence intervals are not available.

2. Pseudo-score inference using the Pearson chi-squared statistic

For a contingency table, let $\{n_i\}$ denote cell counts and let $\{\hat{\mu}_i\}$ denote maximum likelihood estimates of expected cell counts for a particular model. We assume a multinomial distribution for $\{n_i\}$. If one variable is a response and the others are explanatory, we assume independent multinomial distributions at all combinations of settings of the explanatory variables.

Let $\{\hat{\mu}_{i0}\}$ denote maximum likelihood fitted values assuming a simpler, null model. The likelihood ratio statistic for testing the null hypothesis that the simpler model holds against the alternative hypothesis that the full model holds equals $G^2 = 2\sum_i \hat{\mu}_i \log(\hat{\mu}_i/\hat{\mu}_{i0})$. A corresponding Pearson form of statistic is

$$X^2 = \sum_i \frac{(\hat{\mu}_i - \hat{\mu}_{i0})^2}{\hat{\mu}_{i0}}.$$

When the full model is saturated, such as unspecified multinomial probabilities in a 2×2 table, X^2 is the Pearson statistic for testing the goodness-of-fit of the simpler model. By contrast, the use of X^2 for comparing two unsaturated models has received little attention. Rao (1961) proposed this version of X^2 and derived its asymptotic distribution. Haberman (1977) further developed theory and showed the asymptotic null equivalence of G^2 and X^2 for large, sparse tables.

This article applies X^2 to cases in which $\{\hat{\mu}_i\}$ refers to maximum likelihood fitted values for the multinomial model and $\{\hat{\mu}_{i0}\}$ refers to maximum likelihood fitted values for a special case of this model in which a particular scalar parameter β from the model takes a fixed value β_0 in the interior of the parameter space. That is, $\{\hat{\mu}_{i0}\}$ are found by fitting the model under the constraint that $\beta = \beta_0$. Let $\chi^2_{\nu,a}$ denote the $(1-a)$ quantile of the chi-squared distribution with ν degrees of freedom. The set of β_0 values for which

$$X^2(\beta_0) = \sum_i \frac{(\hat{\mu}_i - \hat{\mu}_{i0})^2}{\hat{\mu}_{i0}} \leqslant \chi^2_{1,a}$$

is an asymptotic $100(1-a)\%$ confidence interval for β. When the full model is saturated, $X^2(\beta_0)$ is the score statistic for testing H_0: $\beta = \beta_0$ (Cox & Hinkley, 1974, p. 326; Smyth, 2003). Inverting this test, for the possible β_0 values, yields the score confidence interval. For example, see Cornfield (1956) for the odds ratio, Mee (1984) for the difference of proportions, Koopman (1984) for the relative risk and Lang (2008) for generic measures of association.

When the full model is unsaturated, which is the situation of interest in this article, $X^2(\beta_0)$ is not the score statistic. We refer to the test using $X^2(\beta_0)$ to compare models in that case as a pseudo-score test and the confidence interval obtained by inverting this test as a pseudo-score confidence interval. In the case of a generalized linear model with canonical link function, Lovison (2005) gave a formula for the score statistic that resembles the Pearson statistic, being a quadratic form comparing fitted values for the two models. Let X be the model matrix for the full model and let $\hat{V}_0$ be the diagonal matrix of estimated variances of the observations under the null model, with fitted values $\hat{\mu}$ for the full model and $\hat{\mu}_0$ for the reduced model. Then, the score statistic can be expressed as $(\hat{\mu} - \hat{\mu}_0)^T X (X^T \hat{V}_0 X)^{-1} X^T (\hat{\mu} - \hat{\mu}_0)$. See also Lang et al. (1999) for the loglinear case. For this canonical-link case, Lovison also showed that the score statistic bounds below the $X^2(\beta_0)$ statistic comparing the models.

Although $X^2(\beta_0)$ as defined above refers to models applied to a contingency table, like the likelihood ratio statistic G^2, it often applies even with continuous explanatory variables or highly sparse contingency tables. For the contingency table representation of the data, the difference between the numbers of parameters in the two models must be fixed as n increases, and those parameters cannot fall on the boundary of the parameter space. The asymptotic equivalence with G^2 under these conditions follows from standard results (Haberman, 1977) by which the Pearson statistic is a quadratic approximation for the likelihood ratio statistic such that the difference between the statistics converges in probability to zero under the simpler model and certain regularity conditions. In addition, when β is d-dimensional rather than scalar, the method extends to a confidence region by referring $X^2(\beta_0)$ to $\chi^2_{d,a}$.

We believe that pseudo-score methods are useful for three reasons: First, for models such as considered in this article, which are not generalized linear models with canonical link, ordinary score methods are

Table 1. *Contingency table for illustrating pseudo-score inferences for models for ordinal data*

	Health care			
Environment	1	2	3	Total
1	21	13	12	46
2	10	10	11	31
3	9	7	29	45
Total	40	30	52	122

not practical but the pseudo-score methods can be implemented with the same level of difficulty as profile likelihood confidence intervals. The next section shows examples of four such models. Second, as §5 discusses, extensions apply to settings in which profile likelihood methods are not available. Third, research has shown that ordinary score inferences, when available, perform well for a variety of measures for discrete data, in terms of actual error probabilities being close to nominal levels. In fact, they often perform better than likelihood ratio test-based inference and much better than Wald test-based inference when sample sizes are not large. This may reflect the fact that for canonical models, the score statistic is a standardization of a sufficient statistic that is a linear combination of the observations, and uses a null rather than non-null standard error. For example, see Koehler & Larntz (1980) for testing independence in two-way contingency tables, Newcombe (1998a) for confidence intervals for binomial proportions, Newcombe (1998b) and Agresti & Min (2005) for confidence intervals for the difference of proportions and relative risk and Miettinen & Nurminen (1985) and Agresti & Min (2005) for confidence intervals for the odds ratio. Our simulations have suggested that similar good performance occurs for pseudo-score confidence intervals.

3. Examples of pseudo-score confidence intervals

As mentioned above, for inference about a parameter in an unsaturated model, the score statistic is not the Pearson statistic and it is often impractical to find it. Even for simple models, ordinary score confidence intervals are not available with common statistical software. Using such software, it is possible to implement the pseudo-score method by fitting the model and then fitting the reduced model for various β_0 values and comparing their fitted values using $X^2(\beta_0)$.

This section illustrates pseudo-score confidence intervals by presenting four relatively simple models for which ordinary score-test-based inference is difficult. We illustrate the models by fitting them to Table 1 from the 2006 General Social Survey in which U.S. respondents were asked, ‘How successful do you think the government in America is nowadays in (a) Providing health care for the sick? (b) Protecting the environment?’ The outcome categories are $1 =$ successful, $2 =$ neither successful nor unsuccessful, $3 =$ unsuccessful. Table 1 shows results for subjects of ages 18–25. We treat the nine cell counts as a multinomial sample.

Let (y_1, y_2) denote the bivariate ordinal response in Table 1. One type of model analyzes the association between the variables, such as by assuming a common global log odds ratio,

$$\log\left\{\frac{\mathrm{pr}(y_1 \leqslant i, y_2 \leqslant j)\mathrm{pr}(y_1 > i, y_2 > j)}{\mathrm{pr}(y_1 \leqslant i, y_2 > j)\mathrm{pr}(y_1 > i, y_2 \leqslant j)}\right\} = \beta$$

for all i, j. The likelihood function is a complex function of the model parameter, making ordinary score methods difficult. Pseudo-score inferences are relatively simple, by fitting the model for various fixed β_0 using methods for maximizing a multinomial likelihood function subject to constraints. To fit the model, we used the R function mph.fit written by Prof. Joseph Lang at the University of Iowa, which uses methods developed in Lang (2004, 2005). Here, the constraints equate all four of the global log odds ratios to a common unknown value for the general model and equate them to a fixed value β_0 for the special cases. The model with unspecified β fits Table 1 well, with Pearson goodness-of-fit statistic 1·75 based on 3 degrees of freedom and maximum likelihood estimate $\hat{\beta} = 1{\cdot}181$ with standard error 0·318. The 95% pseudo-score confidence interval for β is (0·556, 1·796). For example, the model under the constraint $\beta_0 = 0{\cdot}556$ or under the constraint $\beta_0 = 1{\cdot}796$ has maximum likelihood fitted values for which $X^2(\beta_0) = 3{\cdot}84$. The profile likelihood interval is (0·562, 1·809), quite similar.

An alternative type of model compares the marginal distributions of this square table. The cumulative logit marginal model

$$\text{logit}\{\text{pr}(y_1 \leqslant j)\} = \alpha_j, \quad \text{logit}\{\text{pr}(y_2 \leqslant j)\} = \alpha_j + \beta \quad (j = 1, 2),$$

is designed to detect a location shift in the two marginal distributions. The multinomial likelihood function for the nine cell probabilities cannot be explicitly expressed in terms of the model parameters, which limits the applicability of ordinary score-test-based inference. We used Lang's R function mph.fit to fit the model and special cases under the constraint that cumulative logits in the two margins differ by a fixed value β_0. This marginal model with unconstrained β fits these data adequately, with goodness-of-fit statistic $X^2 = 0{\cdot}01$ based on 1 degree of freedom and with $\hat{\beta} = -0{\cdot}230$ having standard error 0·194. The 95% pseudo-score confidence interval for β is (−0·617, 0·157). The profile likelihood confidence interval is (−0·616, 0·153).

An alternative way to compare the response distributions is with a random effects model. For responses (y_{1s}, y_{2s}) by subject s, a random intercept model is

$$\text{logit}\{\text{pr}(y_{1s} \leqslant j)\} = u_s + \alpha_j, \quad \text{logit}\{\text{pr}(y_{2s} \leqslant j)\} = u_s + \alpha_j + \beta \quad (j = 1, 2),$$

where $\{u_s\}$ are independent from a $N(0, \sigma)$ distribution with unknown standard deviation σ. The marginal likelihood obtained by integrating out the random effects does not have closed form, and score confidence intervals are difficult for this model. It is straightforward to obtain a pseudo-score confidence interval for β or σ when $\sigma > 0$ by fitting the model for various fixed values of the parameter of interest. The unrestricted model fits Table 1 adequately, with goodness-of-fit $X^2 = 2{\cdot}0$ based on 4 degrees freedom and with $\hat{\beta} = -0{\cdot}319$ having standard error 0·267 and $\hat{\sigma} = 1{\cdot}43$ having standard error 0·31. The 95% pseudo-score confidence interval for β is (−0·833, 0·200), and the profile likelihood confidence interval is (−0·848, 0·201). The corresponding confidence intervals for σ are (0·83, 2·08) and (0·84, 2·10).

Among other models for which the pseudo-score method is simpler to implement than the ordinary score method, because the likelihood function is not an explicit function of the model parameters, are models that assign scores $\{v_j\}$ to ordered categories and describe mean responses. For Table 1, a regression form of the model describes how the mean response for y_2 changes linearly in the score for y_1, and a marginal model compares the marginal mean responses. The first of these models is $\sum_j v_j \text{pr}(y_2 = v_j \mid y_1 = v_i) = \alpha + \beta v_i$. For this model we treat each row of the table as an independent multinomial sample. Lipsitz (1992) and Lang (2005) presented algorithms for maximum likelihood fitting of such models. For Table 1, we treat the health care opinion as the response variable and use the row numbers and column numbers as the scores. The mean response model fits well, with Pearson goodness-of-fit statistic $X^2 = 0{\cdot}29$ based on 1 degree of freedom. The 95% pseudo-score confidence interval for β is (0·143, 0·475). The profile likelihood confidence interval is (0·145, 0·479).

4. Simulation example

We promoted pseudo-score inference by suggesting that it applies in cases in which ordinary score inference is impractical to implement, and for many cases ordinary score inference has been seen to perform well, even with relatively small samples. The previous section showed four examples of its use when score inference is impractical, with the pseudo-score results being similar to profile likelihood confidence intervals. To check whether the good performance may also apply to pseudo-score inference, we conducted some simulations using models mentioned in the previous section. We found that the pseudo-score method performed similarly to the profile likelihood interval and sometimes a bit better when the sample size is quite small.

We illustrate typical results with the cumulative logit marginal model, $\text{logit}[\text{pr}(y_1 \leqslant j)] = \alpha_j$ and $\text{logit}[\text{pr}(y_2 \leqslant j)] = \alpha_j + \beta$. We simulated pseudo-score, profile likelihood and Wald confidence intervals

Table 2. *Simulated coverage percentages, with percentages of underestimation and overestimation in parentheses, for interval estimation of β in marginal cumulative logit model with sample size n and underlying bivariate normal distribution with correlation* 0 *or* 0·5 *in* 2×2 *and* 3×3 *tables*

2×2		Correlation, with $\beta = 0{\cdot}0$		Correlation, with $\beta = 0{\cdot}5$	
n	Method	0·0	0·5	0·0	0·5
20	Pseudo-score	95·6 (2·2, 2·2)	95·1 (2·4, 2·5)	95·1 (2·4, 2·5)	95·0 (2·9, 2·1)
	Profile likelihood	92·7 (3·7, 3·7)	91·6 (4·2, 4·2)	93·6 (3·6, 2·9)	93·2 (3·8, 3·2)
	Wald	93·3 (3·3, 3·3)	92·9 (3·5, 3·5)	94·6 (2·4, 3·0	93·7 (2·9, 3·4)
50	Pseudo-score	94·7 (2·7, 2·7)	95·0 (2·5, 2·5)	95·1 (2·4, 2·5)	95·1 (2·6, 2·3)
	Profile likelihood	94·7 (2·7, 2·7)	94·3 (2·8, 2·8)	94·6 (2·8, 2·6)	94·7 (2·9, 2·5)
	Wald	94·7 (2·7, 2·7)	94·5 (2·8, 2·7)	94·8 (2·6, 2·6)	94·6 (2·6, 2·8)
3×3		Correlation, with $\beta = 0{\cdot}0$		Correlation, with $\beta = 0{\cdot}5$	
n	Method	0·0	0·5	0·0	0·5
20	Pseudo-score	92·6 (3·7, 3·7)	91·5 (4·2, 4·3)	93·7 (3·7, 2·5)	94·2 (3·2, 2·6)
	Profile likelihood	91·4 (4·4, 4·3)	87·2 (6·3, 6·5)	92·1 (5·1, 2·8)	90·5 (6·3, 3·2)
	Wald	93·0 (3·5, 3·5)	90·7 (4·5, 4·8)	93·9 (3·4, 2·6)	93·4 (3·4, 3·2)
50	Pseudo-score	94·5 (2·8, 2·7)	93·9 (3·0, 3·0)	94·5 (2·8, 2·7)	94·1 (3·3, 2·6)
	Profile likelihood	94·1 (2·9, 2·9)	93·4 (3·3, 3·3)	94·2 (3·1, 2·7)	93·5 (3·7, 2·8)
	Wald	94·3 (2·9, 2·8)	93·8 (3·1, 3·1)	94·3 (2·9, 2·8)	94·1 (2·9, 3·0)

for β with 2×2 and 3×3 tables having $\beta = 0$ and 0·5, $n = 20$ and 50, based on a joint distribution that satisfied the model with uniform values for the row marginal probabilities and for joint cell probabilities based on underlying bivariate normal distributions with correlations 0 and 0·5. Table 2 shows the estimated coverage percentages based on 50 000 simulations; an estimator of a true coverage percentage of 95·0 has standard error 0·1. For each case, the table also reports the directional error rates, estimating the percentages of cases in which the interval falls below, or above, the parameter value. In the 3×3 case with $\beta = 0{\cdot}5$ and $n = 20$, in about 2% of the simulations the table was such that the model-fitting algorithm did not converge or at least one of the methods broke down, typically because of an infinite $\hat{\beta}$, so in that case results apply conditional on this not happening. We do not report results with unbalanced marginal probabilities, because then the frequency of such problematic tables was excessive. Table 2 does not show substantially different results for the three methods, but in many small-sample cases the pseudo-score method performed somewhat better than the profile likelihood method. In this simulation and in others we conducted, when the directional error rates differed substantially, the imbalance was a bit worse for the profile likelihood method than for the pseudo-score method, as illustrated for the cases for 3×3 tables with $n = 20$ and $\beta = 0{\cdot}5$.

While there is no guarantee that similar behaviour occurs for other models or cases, our simulations were promising. The pseudo-score confidence interval is simple to construct and it shows promise of being a good, general-purpose method for inference with categorical response data when the ordinary score method is not feasible. The simulation results do not suggest that methodologists should abandon profile likelihood methods in favour of the pseudo-score method. However, as discussed next, there is scope for extending the pseudo-score method to cases in which likelihood-based methods are not available.

5. Generalizations of pseudo-score inference

5·1. *Pseudo-score inference for other discrete distributions*

Suppose $y_1, \ldots, y_n$ are independent observations assumed to have some discrete distribution other than the multinomial. A Pearson-type pseudo-score statistic for comparing models has the form

$$X^2 = \sum_i \frac{(\hat{\mu}_i - \hat{\mu}_{i0})^2}{v(\hat{\mu}_{i0})} = (\hat{\mu} - \hat{\mu}_0)^{\mathrm{T}} \hat{V}_0^{-1} (\hat{\mu} - \hat{\mu}_0),$$

where $v(\hat{\mu}_{i0})$ denotes the estimated variance of y_i under the null distribution for y_i and $\hat{V}_0$ is the diagonal matrix of such values (Lovison, 2005). For some cases, the pseudo-score confidence intervals of this paper for multinomial models extend to other models for discrete data using this generalized statistic. An example is estimating parameters of Poisson regression models.

One application for which this extended statistic is especially useful is analyzing discrete data obtained with a complex sampling scheme, by replacing $\hat{V}_0$ by an appropriately inflated or nondiagonal estimated covariance matrix. To illustrate, §3 treated the General Social Survey as a simple random sample, but in reality it is a multi-stage sample that employs stratification and clustering. The codebook for the survey suggests that because of its design, sampling variances of estimates are approximately 50% larger than obtained with simple random sampling. For the examples in §3, we can obtain more relevant 95% confidence intervals from the set of β_0 with $\sum_i(\hat{\mu}_i - \hat{\mu}_{i0})^2/(1{\cdot}5\hat{\mu}_{i0}) \leqslant 3{\cdot}84$. For the marginal cumulative logit model, for instance, we get the interval $(-0{\cdot}708, 0{\cdot}248)$ for β instead of $(-0{\cdot}617, 0{\cdot}157)$. For such complex sampling designs, profile likelihood confidence intervals are not usually available.

5·2. *Quasilikelihood inference for marginal modelling*

An interesting problem for future research is to extend pseudo-score inference to other quasilikelihood analyses. A possible application is marginal modelling of clustered categorical responses. A popular approach for marginal modelling uses generalized estimating equations. Because of the lack of a likelihood function, Wald methods are commonly employed, together with a sandwich estimator of the covariance matrix of model parameter estimates. For binary data, let y_{it} denote observation t in cluster i, for $t = 1, \ldots, T_i$ and $i = 1, \ldots, n$. Let $y_i = (y_{i1}, \ldots, y_{iT_i})^T$ and let $\mu_i = E(y_i) = (\mu_{i1}, \ldots, \mu_{iT_i})^T$. Let V_i denote the $T_i \times T_i$ covariance matrix of y_i. For a particular marginal model, let $\hat{\mu}_i$ denote an estimate of μ_i, such as the maximum likelihood estimate under the naive assumption that the $\sum_i T_i$ observations are independent. Let $\hat{\mu}_{i0}$ denote the corresponding estimate under the constraint that a particular parameter β takes value β_0. Let $\hat{V}_{i0}$ denote an estimate of the covariance matrix of y_i under this null model. The main diagonal elements of $\hat{V}_{i0}$ are $\hat{\mu}_{it0}(1 - \hat{\mu}_{it0})$ $(t = 1, \ldots, T_i)$. Separate estimation is needed for the null covariances, which are not part of the marginal model. With categorical explanatory variables, an estimate of $\text{cov}(y_{it}, y_{iu})$ is the sample mean value of $(y_{at} - \hat{\mu}_{at0})(y_{au} - \hat{\mu}_{au0})$ for the set of all clusters a that have the same values of between-cluster explanatory variables as cluster i. This is also the sample estimate of the covariance for the multinomial distribution for the 2×2 joint distribution of (y_{at}, y_{au}) for all such clusters.

Now, consider $X^2(\beta_0) = \sum_i(\hat{\mu}_i - \hat{\mu}_{i0})^{\mathrm{T}}\hat{V}_{i0}^{-1}(\hat{\mu}_i - \hat{\mu}_{i0})$. With categorical explanatory variables, $X^2(\beta_0)$ applies to two sets of fitted marginal proportions for the contingency table obtained by cross-classifying the multivariate binary response with the various combinations of explanatory variable values. The set of β_0 values for which $X^2(\beta_0) \leqslant \chi^2_{1,a}$ is a confidence interval for β. Unlike the generalized estimating equations approach, this pseudo-score method does not require using the sandwich estimator, which can be unreliable unless the number of clusters is large (Firth, 1993). Even with consistent estimation of V_{i0}, however, the limiting null distribution of $X^2(\beta_0)$ need not be exactly chi-squared because the fitted values result from inefficient estimates, but preliminary simulations suggest that the chi-squared often provides a good approximation.

5·3. *Confidence intervals based on power divergence statistics*

For testing goodness-of-fit of a multinomial model for a contingency table with counts $\{n_i\}$, Cressie & Read (1984) presented a family of power divergence statistics that have the same asymptotic null chi-squared distribution as X^2 and G^2. The power divergence statistic for testing that a particular parameter takes value β_0 is

$$P_\lambda(\beta_0) = \frac{2}{\lambda(\lambda+1)} \sum \hat{\mu}_i\{(\hat{\mu}_i/\hat{\mu}_{i0})^\lambda - 1\}, \quad -\infty < \lambda < \infty.$$

Cressie et al. (2003) used such a statistic for tests comparing log-linear models. We define a power divergence confidence interval to be the set of β_0 values having $P_\lambda(\beta_0) \leqslant \chi^2_{1,a}$. This encompasses profile likelihood ($\lambda = 0$) and pseudo-score ($\lambda = 1$) intervals. For various models, it would be useful to see if this method tends to work particularly well for a certain λ value.

6. Conclusion

We proposed a pseudo-score confidence interval for a multinomial model parameter by inverting a test that uses the Pearson statistic. This confidence interval has the advantage that it is feasible in cases in which ordinary score inference is not. Limited simulation studies showed it can provide similar or better performance than profile likelihood confidence intervals, and extensions for complex sampling apply in cases in which likelihood-based methods are not available. The proposed method is not intended to compete with specifically designed small-sample methods or with higher-order refinements of asymptotic methods (Brazzale et al., 2007). Our goal was to present a method that performs well in a wide variety of settings and is simpler to implement than the ordinary score method. The authors can supply a document containing code showing how to use the free software R (R Development Core Team, 2008) to obtain the pseudo-score inferences reported in this article.

Acknowledgement

This article was indirectly inspired by a conversation about meta analysis with Ingram Olkin, to whom this article is dedicated.

References

Agresti, A. & Min, Y. (2005). Frequentist performance of Bayesian confidence intervals for comparing proportions in 2 x 2 contingency tables. *Biometrics* **61**, 515–23.

Brazzale, A. R., Davison, A. C. & Reid, N. (2007). *Applied Asymptotics: Case Studies in Samll-Sample Statistics.* Cambridge: Cambridge University Press.

Cornfield, J. (1956). A statistical problem arising from retrospective studies. In *Proc. 3rd Berkeley Symp. Math. Statist. Prob.*, vol. 4, Ed. J. Neyman, 135–48. University of California Press.

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

Cressie, N., Pardo, L. & Pardo, M. (2003). Size and power considerations for testing loglinear models using ϕ-divergence test statistics. *Statist. Sinica* **13**, 555–70.

Cressie, N. & Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *J. R. Statist. Soc.* B **46**, 440–64.

Firth, D. (1993). Recent developments in quasi-likelihood methods. In *Proceedings of the ISI 49th Session*, 341–58.

Haberman, S. (1977). Log-linear models and frequency tables with small expected cell counts. *Ann. Statist.* **5**, 1148–69.

Koehler, K. & Larntz, K. (1980). An empirical investigation of goodness-of-fit statistics for sparse multinomials. *J. Am. Statist. Assoc.* **75**, 336–44.

Koopman, P. A. R. (1984). Confidence intervals for the ratio of two binomial proportions. *Biometrics* **40**, 513–17.

Lang, J. B. (2004). Multinomial-Poisson homogeneous models for contingency tables. *Ann. Statist.* **32**, 340–83.

Lang, J. B. (2005). Homogeneous linear predictor models for contingency tables. *J. Am. Statist. Assoc.* **100**, 121–34.

Lang, J. B. (2008). Score and profile likelihood confidence intervals for contingency table parameters. *Statist. Med.* **27**, 5975–90.

Lang, J. B., McDonald, J. W. & Smith, P. W. F. (1999). Association-marginal modeling of multivariate categorical responses: a maximum likelihood approach. *J. Am. Statist. Assoc.* **94**, 1161–71.

Lipsitz, S. (1992). Methods for estimating the parameters of a linear model for ordered categorical data. *Biometrics* **48**, 271–81.

Lovison, G. (2005). On Rao score and Pearson X^2 statistics in generalized linear models. *Statist. Papers* **46**, 555–74.

Mee, R. W. (1984). Confidence bounds for the difference between two probabilities (letter). *Biometrics* **40**, 1175–76.

Miettinen, O. & Nurminen, M. (1985). Comparative analysis of two rates. *Statist. Med.* **4**, 213–26.

Newcombe, R. (1998a). Two-sided confidence intervals for the single proportion. *Statist. Med.* **17**, 857–72.

NEWCOMBE, R. (1998b). Interval estimation for the difference between independent proportions: comparison of eleven methods. *Statist. Med.* **17**, 873–90.

R DEVELOPMENT CORE TEAM (2008). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, URL http://www.R-project.org.

RAO, C. R. (1961). A study of large sample test criteria through properties of efficient estimates. *Sankhya* **A23**, 25–40.

SMYTH, G. K. (2003). Pearson's goodness of fit statistic as a score test statistic. In *Science and Statistics: A Festschrift for Terry Speed,* Ed. D. R. Goldstein, 115–26. IMS Lecture Notes Monograph Series 40. Hayward, CA: Institute of Mathematical Statistics.

[*Received December* 2008. *Revised June* 2009]