The empty set and zero likelihood problems in maximum empirical likelihood estimation

Wicher Bergsma

Statistics Department, London School of Economics and Political Science
Houghton Street, WC2A 2AE London, UK
e-mail: w.p.bergsma@lse.ac.uk
url: stats.lse.ac.uk/bergsma

and

Marcel Croon and L. Andries van der Ark

Department of Methodology and Statistics
Tilburg School of Social and Behavioral Sciences
Tilburg University,al PO Box 90153, Tilburg, the Netherlands
e-mail: m.a.croon@uvt.nl *;* a.vdark@uvt.nl
url: http://www.tilburguniversity.edu/webwijs/show/?uid=m.a.croon
url: http://www.tilburguniversity.edu/webwijs/show/?uid=a.vdark

Abstract: We describe a previously unnoted problem which, if it occurs, causes the empirical likelihood method to break down. It is related to the empty set problem, recently described in detail by Grendár and Judge (2009), which is the problem that the empirical likelihood model is empty, so that maximum empirical likelihood estimates do not exist. An example is the model that the mean is zero, while all observations are positive. A related problem, which appears to have gone unnoted so far, is what we call the zero likelihood problem. This occurs when the empirical likelihood model is nonempty but all its elements have zero empirical likelihood. Hence, also in this case inference regarding the model under investigation breaks down. An example is the model that the covariance is zero, and the sample consists of monotonically associated observations. In this paper, we define the problem generally and give examples. Although the problem can occur in many situations, we found it to be especially prevalent in marginal modeling of categorical data, when the problem often occurs with probability close to one for large, sparse contingency tables.

AMS 2000 subject classifications: Primary 62G05, 62G10; secondary 62H12, 62H15, 62H17.
Keywords and phrases: Empirical likelihood, empty set problem, zero likelihood problem, marginal model, optimization with constraints.

Received June 2012.

Contents

1. Description of the empirical likelihood method

With $\mathcal{P}$ a set of probability distributions, a subset of $\mathcal{P}$ is called a model in $\mathcal{P}$. A useful and general way to define a model is as

$$\Phi(\theta, \mathcal{P}) = \{P \in \mathcal{P} \,|\, \theta(P) = 0\}$$

where θ is an appropriate map from $\mathcal{P}$ into a Hilbert space. For a topological space Ω let $\mathcal{P}(\Omega)$ be the set of Borel probability measures on Ω, called the saturated model. With

$$\Phi_\Omega = \Phi(\theta, \mathcal{P}(\Omega))$$

we now consider maximum empirical likelihood (MEL) estimation of a distribution $P \in \Phi_\Omega$

Consider a sample of iid observations $X_1, \ldots, X_n$ ($X_i \in \Omega$) with unknown common distribution $P \in \mathcal{P}(\Omega)$. The empirical likelihood sample space is the finite set $\mathcal{X} = \{X_1, \ldots, X_n\} \subseteq \Omega$, and the empirical likelihood of $Q \in \mathcal{P}(\mathcal{X})$ is

$$L(Q) = \prod_{i=1}^{n} Q(X_i).$$

Note that, for all $Q \in \mathcal{P}(\mathcal{X})$,

$$\sum_{x \in \mathcal{X}} Q(x) = 1$$

The unrestricted MEL estimator of $P \in \mathcal{P}(\Omega)$ is defined as

$$\hat{P}_1 = \{Q \in \mathcal{P}(\mathcal{X}) | L(Q) \geq L(Q') \forall Q' \in \mathcal{P}(\mathcal{X})\}$$

Via the inequality of arithmetic and geometric means, it is straightforward to show that $\hat{P}_1$ is the empirical distribution, i.e., $\hat{P}_1(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i = x)$ where I is the indicator function (in the continuous case, $\hat{P}_1(X_i) = 1/n$ for all i with probability 1).

The empirical likelihood model $\Phi_\mathcal{X} \subseteq \Phi_\Omega$ is defined as

$$\Phi_\mathcal{X} = \Phi(\theta, \mathcal{P}(\mathcal{X}))$$

Thus, $\Phi_\mathcal{X}$ consists of those probability distributions in Φ_Ω whose support is $\mathcal{X}$. The MEL estimator of $P \in \mathcal{P}(\Omega)$ in Φ_Ω is

$$\hat{P}_0 = \{Q \in \Phi_\mathcal{X} | L(Q) \geq L(Q') \forall Q' \in \Phi_\mathcal{X}\}$$

A particularly important application of MEL estimation is the testing of the null hypothesis that $P \in \Phi_\Omega$ against the alternative hypothesis that $P \in \mathcal{P}(\Omega) \setminus \Phi_\Omega$. For this purpose, it is common to use the log likelihood ratio test statistic

$$T = -2 \log \frac{L(\hat{P}_0)}{L(\hat{P}_1)},$$

In this context, it is assumed that $\theta : \mathcal{P}(\Omega) \to \mathbb{R}^p$ for some $p \geq 1$; then if $P \in \Phi_\Omega$, and θ satisfies certain smoothness conditions, T has an asymptotic chi-square distribution with p degrees of freedom (Qin and Lawless, 1994; Owen, 2001). (NB: these authors gave conditions on estimating equations which imply $\theta(P) = 0$; so those are actually *implicit* conditions on θ.) So-called inversion of the likelihood ratio test can be used to construct confidence intervals for $\theta(P)$.

2. The empty set and zero likelihood problems

For finite samples, there may be problems which cause the MEL method to break down. Grendár and Judge (2009) considered in detail the problem that $\Phi_\mathcal{X} = \emptyset$, which they called the empty set problem. Another problem which can occur and which we have not seen explicitly described before, is that $L(P) = 0$ for all $P \in \Phi_\mathcal{X}$. We call this the *zero likelihood problem*. Note that every $P \in \Phi_\mathcal{X}$ then trivially maximizes the empirical likelihood. Evidently, the zero likelihood problem occurs if and only if for all $P \in \mathcal{P}(\mathcal{X})$, there exists an $1 \leq i_P \leq n$ such that $P(X_{i_P}) = 0$. The next example illustrates the two problems and how they compare.

Example 1. Suppose $\Omega = \mathbb{R}$ and θ is the mean, i.e., Φ_Ω is the model that the mean is zero. The empty set problem occurs if and only if $X_i > 0$ for all i or $X_i < 0$ for all i, i.e., 0 lies outside the convex hull of the data points (Qin and Lawless, 1994). Alternatively, the zero likelihood problem occurs if and only if $X_i \geq 0$ for all i, with $X_i = 0$ for at least one i and $X_i > 0$ for at least one i, or the same with reverse inequalities: then $\Phi_\mathcal{X}$ consists of the degenerate probability distribution with all mass at 0. If the population mean is zero, the zero likelihood problem can occur if and only if $P(X = 0) > 0$, $P(X > 0) > 0$ and $P(X < 0) > 0$.

To take a specific example, suppose we have three observations $X_1 = 0$, $X_2 = 1$, and $X_3 = 2$. The empirical likelihood is

$$p_1 p_2 p_3,$$

where $p_i = P(X = X_i | X_i)$. We seek $\{p_1, p_2, p_3\}$ subject to the zero mean constraint

$$0 * p_1 + 1 * p_2 + 2 * p_3 = 0$$

and

$$p_1, p_2, p_3 \geq 0, \quad p_1 + p_2 + p_3 = 1.$$

The only solution is $p_1 = 1$ and $p_2 = p_3 = 0$. Hence, the empirical likelihood equals zero. A numerical example is

$$P(X = 0) = 0.5, \ P(X = 1) = 0.2, \ P(X = 2) = 0.2, \ P(X = -6) = 0.1.$$

Then for sample size three, it can be verified that the probability of the zero likelihood problem occurring would be 0.63 (i.e., the probability that exactly one or exactly two observations are zero).

We next give three examples with $\Omega \subseteq \mathbb{R}^2$, denoting the sample points as $(X_1, Y_1), \ldots, (X_n, Y_n)$, where each (X_i, Y_i) is an independent replication of (X, Y).

Example 2. Let $\Omega = \mathbb{R}^2$ and let θ be the covariance, i.e., $\Phi_{\mathcal{X}}$ is the model that the covariance is zero. Since, in general, the covariance of a degenerate distribution on a point mass is zero, the empty set problem cannot occur. On the other hand, the zero likelihood problem occurs if $n \geq 2$ and the observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ are monotonically associated, in the sense that there is a permutation $(i_1, \ldots, i_n)$ of $(1, \ldots, n)$ such that

$$X_{i_1} < X_{i_2} < \cdots < X_{i_n} \text{ and } Y_{i_1} < Y_{i_2} < \cdots < Y_{i_n}$$

or that

$$X_{i_1} < X_{i_2} < \cdots < X_{i_n} \text{ and } Y_{i_1} > Y_{i_2} > \cdots > Y_{i_n}$$

In particular, $\Phi_{\mathcal{X}}$ consists of the degenerate probability distributions which have all probability mass at a single point (X_i, Y_i), so $L(P) = 0$ for all $P \in \Phi_{\mathcal{X}}$.

Example 3. Let $\Omega = \{0, 1\} \times \mathbb{R}$ and let θ be the differences in conditional variances, conditioning on X, i.e.,

$$\theta = \text{var}(Y|X = 1) - \text{var}(Y|X = 0).$$

Hence, Φ_{Ω} is the model of equal conditional variances. The empty set problem occurs if there is no i such that $X_i = 0$ or if there is no i such that $X_i = 1$. The zero likelihood problem occurs if there is exactly one (X_i, Y_i) with $X_i = 0$, and there are $(X_i, Y_i) \neq (X_j, Y_j)$ with $X_i = X_j = 1$, or vice versa with the roles of 0 and 1 interchanged. In the first case, in the empirical likelihood model $\text{var}(Y|X = 0) = 0$, and hence also $\text{var}(Y|X = 1) = 0$. The second case is analogous. Thus, in either case, $\Phi_{\mathcal{X}}$ consists of those probability distributions whose probability mass is nondegenerately distributed on two points, one with $X = 0$ and one with $X = 1$.

Example 4. Let $\Omega = \{0, 1\} \times \{0, 1\}$, and let $\theta = EX - EY$ be the difference in marginal means. Then Φ_{Ω} is the model of equal marginal means for a 2×2 contingency table, which is also called the marginal homogeneity model. It can be verified that Φ_{Ω} consists of those distributions P with $P(0, 1) = P(1, 0)$. The empty set problem occurs if all observations are either in cell $(0, 1)$ or in cell $(1, 0)$. The zero likelihood problem occurs if there is at least one observation in the main diagonal cells $(0, 0)$ or $(1, 1)$, and at least one observation in cell $(0, 1)$ or in cell $(1, 0)$ but not in both.

3. Remarks

We first encountered the zero likelihood problem when trying to fit categorical marginal models (CMMs; see Bergsma, Croon, and Hagenaars, 2009) for large, sparse contingency tables using MEL. The simplest CMM, marginal homogene-

ity for a 2×2 table, is described in Example 4. The reason we used empirical likelihood rather than ordinary maximum likelihood is that the latter is computationally infeasible for very large tables, say with millions of cells. Such tables occur regularly in practice, for example if there are ten variables with five categories each the number of cells is approximately ten million. Unfortunately, we found that the zero likelihood problem occurred regularly, in many cases with probability close to one, and so we decided it was necessary to determine the general nature of the problem; this resulted in the present paper. In general, we would expect the zero likelihood problem to occur most frequently for for MEL estimation of bivariate or multivariate (continuous or discrete) models.

A possible solution to both the zero likelihood and empty set problems is to augment the empirical sample space with one or more well-chosen points. To solve the empty set problem, Chen et al. (2008) proposed adding a single point, namely minus the average of the sample mean (see also Liu and Chen, 2010). We are currently investigating augmenting the empirical sample space for CMMs, for which it is typically necessary to add *more* than one point in order to solve the two problems.

MEL estimation involves maximization of a multinomial likelihood subject to constraints on the multinomial probabilities. Typically in the empirical likelihood literature, this maximization is done with the help of so-called empirical estimating equations (Qin and Lawless, 1994). Using this technique a large class of models can be written via linear constraints on the multinomial probabilities. The Lagrangian dual is then a convex optimization problem which is, in principle, easy to solve. An alternative Lagrange multiplier method for maximizing a multinomial likelihood subject to constraints, which does not require the potentially cumbersome specification of the estimating equations, is described by Bergsma (1997) (see also Bergsma et al., 2009). We found this method to have very good numerical properties in practice. Another Lagrange multiplier method suitable for MEL estimation was described by Bergsma and Rapcsák (2005); this method turns a smooth constrained optimization problem into an unconstrained one.

Besides the aforementioned hypothesis testing, another important application of empirical likelihood is the construction of confidence intervals for a parameter $\theta(P)$. This is typically done via inversion of the likelihood ratio test, a method which can be formulated in terms of the profile likelihood. Evidently, if the zero likelihood problem occurs for the hypothesis $\theta(P) = \theta_0$, then the value θ_0 will necessarily not be in such a confidence interval, regardless of the confidence level used. So, the zero likelihood problem is also a major potential issue in confidence interval construction.

Of some independent interest in this paper may be the formulation of the empirical likelihood method given in Section 1, which is more general than the estimating equations formulation normally used in the literature. For example, the hypothesis that the medians of two probability distributions are equal seems hard to formulate using estimating equations.

References

BERGSMA, W. P. (1997). *Marginal Models for Categorical Data*. Tilburg University Press, Tilburg.

BERGSMA, W. P., CROON, M. A., AND HAGENAARS, J. A. P. (2009). *Marginal models for dependent, clustered and longitudinal categorical data*. Springer, New York.

BERGSMA, W. P. AND RAPCSÁK, T. (2005). An exact penalty method for smooth equality constrained optimization with application to maximum likelihood estimation. *Eurandom technical report*, 2005-001:1–21. http://www.eurandom.tue.nl/reports/2005/001-report.pdf

CHEN, J., VARIYATH, A. M., AND ABRAHAM, B. (2008). Adjusted empirical likelihood and its properties. *Journal of Computational and Graphical Statistics*, 17(2):426–443. MR2439967

GRENDÁR, M. AND JUDGE, G. (2009). Empty set problem of maximum empirical likelihood methods. *Electronic Journal of Statistics*, 3:1542–1555. MR2578837

LIU, Y. AND CHEN, J. (2010). Adjusted empirical likelihood with high-order precision. *The Annals of Statistics*, 38(3):1341–1362. MR2662345

OWEN, A. B. (2001). *Empirical likelihood*, volume 92. CRC press, New York.

QIN, J. AND LAWLESS, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, pages 300–325. MR1272085