

Empty set problem of maximum empirical likelihood methods

Marian Grendár*†

*Department of Mathematics, Faculty of Natural Sciences, Bel University, Tajovského 40,
SK-974 01 Banská Bystrica, Slovakia,*

Institute of Mathematics and CS, Slovak Academy of Sciences (SAS),

Institute of Measurement Sciences of SAS, Bratislava, Slovakia

e-mail: marian.grendar@savba.sk

and

George Judge

*Professor in the Graduate School, 207 Giannini Hall, University of California,
Berkeley, CA, 94720*

e-mail: gjudge@berkeley.edu

Abstract: In an influential work, Qin and Lawless (1994) proposed a general estimating equations (GEE) formulation for maximum empirical likelihood (MEL) estimation and inference. The formulation replaces a model specified by GEE with a set of data-supported probability mass functions that satisfy empirical estimating equations (E^3). In this paper we use several examples from the literature to demonstrate that the set may be empty for some E^3 models and finite data samples. As a result, MEL does not exist for such models and data sets. If MEL and other E^3 -based methods are to be used, then models will have to be checked on case-by-case basis for the absence or presence of the empty set problem.

AMS 2000 subject classifications: Primary 62F10.

Keywords and phrases: empirical estimating equations, generalized minimum contrast, empirical likelihood, euclidean empirical likelihood, generalized empirical likelihood, affine empty set problem, empirical likelihood bootstrap, model selection.

Received November 2009; revised December 2009.

Contents

1	Introduction	1543
1.1	Empirical estimating equations approach	1543
1.2	Paper format	1545
2	Empty set problem	1545
3	Some models from the literature	1546
3.1	QL, Example 1	1546

*Corresponding author.

†Supported by VEGA 1/0077/09 grant.

3.2	QL, Example 2	1548
3.3	QL, Example 3	1548
	3.3.1 EL confidence intervals and tests	1549
3.4	Mean and median	1549
4	Some other models with ESP	1550
	4.1 Estimation of a location parameter from noisy data	1550
	4.2 Score E^3	1550
	4.3 Discrete random variable	1551
	4.4 Restricted parameter space	1551
5	Implications for theory and practice	1552
	Acknowledgments	1553
	References	1553

1. Introduction

Maximum Empirical Likelihood (MEL) and MEL-like methods have attracted considerable amount of interest and applications over the last decade; for recent literature see e.g., [Cheng, Small, Tan et al. \(2009\)](#), [Crepet, Harari-Kermadec and Tressou \(2009\)](#), [Devereux and Tripathi \(2009\)](#), [Hjort, McKeague and van Keilegom \(2009\)](#), [Wang and Chen \(2009\)](#), [Wong, Liu, Chen et al. \(2009\)](#), [Zhou, Zhou, Chen et al. \(2009\)](#), [Chen and van Keilegom \(2009\)](#). The methods operate within the empirical estimating equations (E^3) approach to estimation and inference. The approach allows for a flexible semi-parametric modelling, and is attractive wherever a model can be formulated in terms of the general estimating equations. The E^3 approach replaces a model specified by general estimating equations with a set of data-supported probability mass functions that satisfy their empirical counterparts. In this paper, we demonstrate that the E^3 approach may not work for every statistical model and data set and thereby may face what we call an empty set problem (ESP). Due to ESP, studies of finite-sample properties of the E^3 -based methods, may be distorted for some models. In applications of MEL and other E^3 -based methods, this means the model should be checked for emptiness.

1.1. Empirical estimating equations approach

In statistics and other fields such as econometrics, it is rather common to formulate a probabilistic model for the random variable $X \in \mathbb{R}^d$ with a probability distribution $r_X(x; \theta)$ that is parametrized by $\theta \in \Theta \subseteq \mathbb{R}^K$, as a set $\Phi(\Theta)$ of probability distributions with certain moment properties. This is accomplished through general estimating functions $u(X; \theta) \in \mathbb{R}^J$ of parameter θ . The estimating functions are employed to form the set $\Phi(\Theta) = \bigcup_{\theta \in \Theta} \Phi(\theta)$ of parametrized probability distributions $F(x; \theta)$, where $\Phi(\theta)$ is defined through the estimating equations (cf. [Godambe and Kale \(1991\)](#)) as

$$\Phi(\theta) = \left\{ F(x; \theta) : \int u(x; \theta) dF(x; \theta) = 0 \right\}.$$

In general, we assume that $r_X(x; \theta)$ need not belong to $\Phi(\Theta)$. Given a random sample $X_1^n = X_1, X_2, \dots, X_n$ from $r_X(x; \theta)$ the objective is to estimate θ and draw inferences. To this end, the *Empirical Estimating Equations* (E^3) approach replaces the model $\Phi(\Theta)$ with its empirical, data-based counterpart $\Phi_q(\Theta) = \bigcup_{\theta \in \Theta} \Phi_q(\theta)$ where

$$\Phi_q(\theta) = \{q(x; \theta) : \text{ch}(u(X_1^n; \theta)) \cap \{0\}\},$$

where $\text{ch}(A)$ denotes the convex hull of the set $A \subseteq \mathbb{R}^J$, and 0 is the J -dimensional zero vector. Note that this approach replaces the set $\Phi(\Theta)$ of probability distributions with the set $\Phi_q(\Theta)$ of probability mass functions (pmf's) $q(x; \theta)$ that are supported on the data X_1^n . An estimate $\hat{\theta}$ of θ is then obtained by means of a method that selects a pmf $\hat{q}(\cdot; \hat{\theta})$ from $\Phi_q(\Theta)$. One prominent method is the Maximum Empirical Likelihood (MEL) method (cf. [Qin and Lawless \(1994\)](#), [Owen \(2001\)](#)) which selects

$$\hat{q}(\cdot; \hat{\theta})_{\text{MEL}} = \arg \sup_{q(\cdot; \theta) \in \Phi_q(\Theta)} \frac{1}{n} \sum_{i=1}^n \log q(x_i; \theta). \tag{1}$$

By means of the convex duality (cf. [Rockafellar \(1970\)](#)), $\hat{\theta}_{\text{MEL}}$ can be obtained as

$$\hat{\theta}_{\text{MEL}} = \arg \inf_{\theta \in \Theta} \sup_{\lambda \in \mathbb{R}^J} -\frac{1}{n} \sum_{i=1}^n \log q(x_i; \theta, \lambda), \tag{2}$$

where

$$q(\cdot; \theta, \lambda) = \left[n \left(1 - \sum_{j=1}^J \lambda_j u_j(\cdot; \theta) \right) \right]^{-1}.$$

The Maximum Empirical Likelihood (MEL) estimator is a non-linear function of the data and can be obtained by a numerical solution. Asymptotic distributional properties of the MEL estimator are known (cf. [Qin and Lawless \(1994\)](#)), and provide a basis for inference.

Methods other than MEL can be used for estimation within the E^3 approach. In particular, the objective function in (1) can be replaced by another measure of closeness of $q(\cdot)$ to the uniform distribution, supported on the sample; cf. [Back and Brown \(1990\)](#), [Brown and Chen \(1998\)](#), [Corcoran \(2000\)](#), [Imbens, Spady and Johnson \(1998\)](#), [Kitamura and Stutzer \(1997\)](#), [Mittelhammer, Judge and Miller \(2000\)](#), [Owen \(2001\)](#), [Schennach \(2005\)](#). Most common are the measures of closeness (also known as divergences) encapsulated by the Cressie Read (CR) family [Cressie and Read \(1984\)](#) or the so-called convex statistical distances, which lead to the Generalized Minimum Contrast (GMC) estimators [Bickel, Klaassen, Ritov et al. \(1993\)](#), [Corcoran \(1998\)](#), [Kitamura \(2007\)](#). The GMC class contains MEL and Maximum Entropy Empirical Likelihood (also known as the Exponential Tilt, cf. [Imbens, Spady and Johnson \(1998\)](#), [Kitamura and Stutzer \(1997\)](#), [Corcoran \(2000\)](#), [Mittelhammer, Judge and Miller \(2000\)](#)) as special cases. The two estimators are contained also in the

Generalized Empirical Likelihood [Smith \(1997\)](#) class that is based on the CR family.

Finally, note a modification of E^3 , without the usual nonnegativity constraints where $\Phi_q(\theta)$ is replaced by

$$\Phi_q^m(\theta) = \{q(x; \theta) : \text{ah}(u(X_1^n; \theta)) \cap \{0\}\}.$$

In this case, $\text{ah}(A)$ is the affine hull (cf. [Rockafellar \(1970\)](#)) of the set $A \subseteq \mathbb{R}^J$. A method such as the Euclidean Empirical Likelihood (cf. [Brown and Chen \(1998\)](#)), or some other member of the CR class of estimators operates within the modified E^3 approach (mE^3) to provide a basis for estimation and inference.

1.2. Paper format

The paper is organized as follows: in [Sect. 2](#), we define the empty set problem and the affine empty set problem. [Section 3](#) is devoted to a detailed analysis of three examples from [Qin and Lawless \(1994\)](#) that are subject to the empty set problem. An E^3 model by [Brown and Chen \(1998\)](#) is shown to be free of ESP. [Section 4](#) contains some other models that are contaminated with ESP. In [Section 5](#) we draw implications of ESP for practical applications of E^3 and mE^3 -based methods.

2. Empty set problem

The replacement of $\Phi(\Theta)$ by $\Phi_q(\Theta)$ may seem natural¹. However, because $\Phi_q(\Theta)$ is data-dependent, it can be empty for some model and data and therefore may be subject to an *empty set problem* (ESP); cf. [Grendar and Judge \(2009b\)](#). In other words, ESP introduces the possibility that for a sample size n , there may exist $X_1^n \sim r_X(x; \theta)$ such that $\Phi_q(\Theta) = \emptyset$. Consequently for such a model the very existence of an estimator and corresponding inferences that are obtained by methods that operate within E^3 , are data-dependent.

In the section ahead, we demonstrate that models considered in a seminal MEL paper [Qin and Lawless \(1994\)](#) (abbreviated QL), are subject to the empty set problem. A few other models are used to illustrate the problem and provide a basis for further discussion.

There are models, such as QL, [Example 1](#), where even the modified E^3 approach is subject to an empty set problem. In this case the empty set problem concerns the possibility that for samples of size n , there may exist $X_1^n \sim r_X(x; \theta)$ such that $\Phi_q^m(\Theta) = \emptyset$. This form of ESP will be referred to as the *affine Empty Set Problem* (aESP). Whenever there is the affine ESP for a data set, then there is also ESP; the opposite need not be true.

Three items are worth emphasizing:

¹In particular, when it is phrased in context of the Generalized Minimum Contrast estimation, cf. [Kitamura \(2007\)](#), [Corcoroan \(1998\)](#), [Bickel, Klaassen, Ritov et al. \(1993\)](#).

Remark 1) The empty set problem is a substantive extension of the convex hull problem, that is known in the Empirical Likelihood literature; cf. Owen (2001), see also Sect. 3.3.1. Indeed, the convex hull problem can be seen as the empty set problem, in the case when $\Theta = \{\theta_0\}$, i.e., when the parametric space comprises a single point. It is well-known that in this case the set $\Phi_q(\theta_0)$ may be empty, for some data set, and some θ_0 . Our objective is to demonstrate that for some models there are data, for which $\Phi_q(\Theta)$ may be empty, for the entire Θ .

Remark 2) Another substantive point concerns the affine form $\Phi_q^m(\Theta)$ of the model. We show that even in this case there are models that are subject to the affine empty set problem. This means, that the modifications of MEL, that are designed to mitigate the convex hull problem by expanding it into the affine hull, cannot, in general, serve as a rescue.

Remark 3) It is well known that methods of estimation other than those that operate within E^3 may break down for some models and data sets. For instance, in some models, the Maximum Likelihood estimator (MLE) may not exist for some data sets. However, non-existence of MLE relates to the absence of a finite maximum of the likelihood function, and the impossibility to attain supremum of the likelihood function for any finite value of the parameter vector, for the data set. There are also the E^3 models where the Maximum Empirical Likelihood (MEL) estimator may not exist for some data sets due to absence of a finite (point of) maximum. However, unlike MLE, MEL, or any other E^3 -based estimator may not exist for some data because of emptiness of the model set $\Phi_q(\Theta)$.

3. Some models from the literature

3.1. QL, Example 1

QL, Example 1, p. 301 and pp. 309-311, consider the following setup: there is a random variable X , with pdf $r(X; \theta)$, where $\theta \in \Theta = \mathbb{R}$, and a random sample $X_1^n = X_1, \dots, X_n$, drawn from $r_X(x; \theta)$. A researcher specifies a pair of estimating functions

$$\begin{aligned} u_1(X; \theta) &= X - \theta, \\ u_2(X; \theta) &= X^2 - (2\theta^2 + 1). \end{aligned}$$

The estimating functions, through unbiased estimating equations, define a set of pdf's (i.e., the model) $\Phi(\Theta) = \bigcup_{\theta \in \Theta} \Phi(\theta)$ where $\Theta = \mathbb{R}$ and

$$\Phi(\theta) = \{f_X(x; \theta) : Eu_1(X; \theta) = 0; Eu_2(X; \theta) = 0\},$$

into which, in the researcher's view, the 'true' sampling distribution should belong. In this case $J > K$ and the model is referred to in the econometric literature as over-identified.

As noted in the Introduction, the Empirical Estimation Approach (E³) proceeds by replacing the set $\Phi(\theta)$ by its empirical analogue

$$\Phi_q(\theta) = \left\{ q(x; \theta) : \sum_{i=1}^n q(x_i; \theta)u_1(x_i; \theta) = 0; \sum_{i=1}^n q(x_i; \theta)u_2(x_i; \theta) = 0; \sum_{i=1}^n q(x_i; \theta) = 1; q(x_i; \theta) \geq 0, 1 \leq i \leq n \right\}.$$

Our objective is to derive a condition on data X_1^n under which the set $\Phi_q(\Theta)$ is empty. To this end, note that the two empirical estimating equations $\sum_{i=1}^n q(x_i; \theta)(x_i - \theta) = 0$, $\sum_{i=1}^n q(x_i; \theta)(x_i^2 - (2\theta^2 + 1)) = 0$ can be combined into a single equation

$$\sum_{i=1}^n q(x_i; \cdot)x_i^2 - 2 \left(\sum_{i=1}^n q(x_i; \cdot)x_i \right)^2 = 1. \tag{3}$$

The question is whether for an observed sample $X_1^n = x_1^n$ the value on the Right-Hand Side of (3) can be attained, for some $q(\cdot, \cdot)$. If not, then for the data x_1^n the set $\Phi_q(\Theta)$ is empty. The expression on the Left-Hand Side of (3) can attain its maximal value for such a $q(x; \cdot)$ that the only non-zero elements of $q(x; \cdot)$ are $q_{(1)} \triangleq q(x_{(1)}; \cdot)$ and $q_{(n)} \triangleq q(x_{(n)}; \cdot)$; there $x_{(1)}$ denotes the lowest, $x_{(n)}$ the largest value in x_1^n . The value $\hat{q}_{(1)}$ of $q_{(1)}$ for which the LHS of (3) attains its maximum is

$$\hat{q}_{(1)} = \arg \max_{q_{(1)} \in [0,1]} L(q_{(1)}), \tag{4}$$

where

$$L(q_{(1)}) \triangleq q_{(1)}x_{(1)}^2 + (1 - q_{(1)})x_{(n)}^2 - 2(q_{(1)}x_{(1)} + (1 - q_{(1)})x_{(n)})^2. \tag{5}$$

Without a $[0, 1]$ constraint on the range of the values that $q_{(1)}$ can take on, the optimal value $\hat{q}_{(1)}^m$ is

$$\hat{q}_{(1)}^m = \frac{\frac{x_{(1)}^2 - x_{(n)}^2}{4(x_{(1)} - x_{(n)})} - x_{(n)}}{x_{(1)} - x_{(n)}},$$

if $x_{(1)} \neq x_{(n)}$. Then $\hat{q}_{(n)}^m = 1 - \hat{q}_{(1)}^m$. Without the $[0, 1]$ -range constraint the maximal value of the LHS of (3) is given by inserting $\hat{q}_{(1)}^m$ into (5); we denote the result by v . Under the range constraint the maximal value of the LHS of (3) can be only smaller or equal to v . If the data $X_1^n \sim r_X(x; \theta)$ are such that v is smaller than 1 (i.e., the RHS of (3)) then $\Phi_q^m(\Theta)$ as well as $\Phi_q(\Theta)$ is empty for such data. Consequently, for such data there is no mE³-based or E³-based estimator.

For a given $r(X; \theta)$ and n , we are interested in the probability $\Pr(v < 1)$ that the LHS maximal value is smaller than 1, the RHS value of (3); i.e., the

probability that $\Phi_q^m(\Theta)$ (as well as $\Phi_q(\Theta)$) is empty. The probability can be estimated by means of a Monte Carlo simulation from $r(X; \theta)$. For $r(X; \theta)$ being $n(0, 1)$ and $n = 15$, as in QL, the Monte Carlo estimate of the probability is 0.0173, based on $M = 10000$ samples. Thus, 17 of 1000 samples of size $n = 15$ drawn from $n(0, 1)^2$ are such that $\Phi_q^m(\Theta)$ (as well as $\Phi_q(\Theta)$) is empty, and hence it is meaningless in these samples to look for EL, or any other E^3 -based (or mE^3 -based) estimate.

QL performed an MC study of small-sample properties of MEL, with the objective of comparing it with other estimators (sample mean and ML). The above results raise questions concerning the meaning of the numbers in Table 1. Since MEL does not exist if $\Phi_q(\Theta)$ is empty, this puts into question the comparison of MEL with other competitive estimators. Indeed, the entire Empirical Estimating Equations approach is questionable, except asymptotically³, where the feasible set $\Phi_q(\Theta)$ becomes 'always' non-empty. The same holds for mE^3 .

The setup considered by QL in their Example 1 is a simple one. Its simplicity permits the empty set problem of E^3 to be illustrated analytically. The E^3 approach, and its most common instance - the Maximum Empirical Likelihood Estimation, - are commonly studied with more complicated models, where the question whether $\Phi_q(\Theta)$ is non-empty for a particular data, may be much harder to answer.

3.2. QL, Example 2

In the second QL example, there is a random sample $(X, Y)_1^n$ of bivariate observations, such that $E(X) = E(Y) = \theta$, $\theta \in \Theta = \mathbb{R}$. The authors suggest using the bivariate estimating function $u(x, y; \theta) = (X - \theta, Y - \theta)$ and estimating θ by MEL. However, due to the problem of the empty set $\Phi_q(\Theta)$ this is not always possible. For instance, let $X \sim n(0, \sigma_1^2)$, $Y \sim n(0, \sigma_2^2)$. Then, as noted in Qin and Lawless (1994), $\Phi_q(\Theta)$ will be empty for every sample, such that $X_i - Y_i > 0$, or $X_i - Y_i < 0$, for all $i = 1, \dots, n$. For $n = 10$ the probability is $2(0.5)^{10} = 0.002$. The probability increases, as the model becomes misspecified. For instance, assume that $X \sim n(-0.3, 0.1)$ and $Y \sim n(0, 0.1)$, then $X - Y \sim n(-0.3, 0.2)$ and the probability $\Pr(X - Y < 0) = 0.749$, so that $\Pr(X - Y > 0) = 0.251$. Consequently, for $n = 10$, the probability that $\Phi_q(\Theta)$ is empty is $0.749^{10} + 0.251^{10} = 0.056$.

3.3. QL, Example 3

The third QL example concerns the selection of a 'representative member' from

$$\Phi(\Theta = \{a\}) = \{f_X(x; \theta) : E_f u(X; \theta) = 0\},$$

²Note, that the data sampling distribution $n(0, 1)$ belongs to $\Phi(\Theta)$. When it is replaced by $n(0.1, \sigma = 0.9)$, the MC estimate of the probability is 0.0453.

³The point that MEL is an asymptotic method has already been made in Grendár and Judge (2009a), albeit from a different point of view.

where $a \in \mathbb{R}$ is known, and there is a random sample X_1^n from unknown $r_X(x)$. In the E^3 approach, the model $\Phi(a)$ is replaced by its empirical analogue

$$\Phi_q(a) = \left\{ q(x; a) : \sum_{i=1}^n q(x_i; a)u(x_i; a) = 0, \sum_{i=1}^n q(x_i; a) = 1, q(x_i; a) \geq 0, \forall i \right\},$$

and a pmf is selected from $\Phi_q(a)$ by some method, such as MEL, or Minimum Discriminant Information (cf. [Haberman \(1984\)](#)). Even in this simple setting the empty set problem appears. For the sake of illustration, let us assume that $r_X(x) = [0.025, 0.025, 0.15, 0.8]$ is a pmf on $\mathcal{X} = \{1, 2, 3, 4\}$. Let $u(X) = X - a$, $a = 2.0$ and let the sample size $n = 40$. The probability that $\Phi_q(2)$ is empty is the probability that the sample contains only values greater than 2, or only values smaller than 2, which is 0.129.

3.3.1. EL confidence intervals and tests

The empty set problem in the setting of this Example also has bearing for constructing confidence intervals and tests by those in applied work who use the Empirical Likelihood (EL) method. Although we are concerned here with MEL estimation and inference, it is worth noting that EL inference is also undermined by the empty set problem. For instance, assume - in the context of the above Example - that the null hypothesis $H_0 : EX = \mu = 2.0$ is to be tested. The EL test and confidence interval are based on the Non-parametric Likelihood Ratio statistic (cf. [Owen \(2001\)](#)), which involves computation of the value of the Non-parametric Likelihood at $\mu = 2.0$. If the data are such that $\Phi_q(2.0)$ is empty then no confidence interval or test can be performed. In the above setting this happens with probability 0.129, for the data of size $n = 40$.

Existence of the empty set problem in the construction of confidence intervals and statistical tests by EL is not new, and in [Owen \(2001\)](#) (cf. Sect. 3.14, Sect. 10.4, Chap. 12, among others), it is referred to as the convex hull condition. Recently, [Chen, Variyath and Brown \(2008\)](#) suggested an adjustment of EL, with the aim of mitigating this problem; see also [Zhou, Zhou, Chen et al. \(2009\)](#), and [Liu and Chen \(2009\)](#), [Emerson and Owen \(2009\)](#) for different ways of mitigating the convex hull restriction. The authors suggest adding to the data an additional observation which is the negative multiple of the sample average. The multiplication constant serves to rescue EL from the empty set problem. However, the value of the constant which achieves the goal is also data dependent.

3.4. Mean and median

[Brown and Chen \(1998\)](#) investigated the problem of estimating a location parameter of a sampling distribution by a data-based combination of the mean

and the median. Their empirical model is

$$\Phi_q(\theta) = \left\{ q(x; \theta) : \sum_{i=1}^n q(x_i; \theta)(x_i - \theta) = 0; \sum_{i=1}^n q(x_i; \theta)\text{sgn}(x_i - \theta) = 0; \sum_{i=1}^n q(x_i; \cdot) = 1; q(x_i; \cdot) \geq 0, 1 \leq i \leq n \right\},$$

and $\theta \in \Theta = \mathbb{R}$. In their model $J > K$ and thus is over-determined. Unless Θ is restricted to a subspace of \mathbb{R} (cf. Sect. 4.4), there is no problem of an empty set, and hence in this case MEL always exists. However, an EL confidence interval/test need not exist for every θ .

4. Some other models with ESP

4.1. Estimation of a location parameter from noisy data

Motivated by Gzyl and Ter Horst (2009), consider the following data-generating process: $Y = X + \epsilon$, where $X \sim \text{Exp}(\theta = 1)$, $\epsilon \sim n(0, \sigma^2)$ and $X \perp \epsilon$. A researcher observes a random sample Y_1^n and would like to estimate the location parameter θ of the distribution $f_X(x; \theta)$ of X ; σ is assumed known. Since neither $f_X(x; \theta)$ nor the distribution of ϵ are known, the researcher chooses an over-determined model based on the first two moments of Y :

$$\Phi(\theta) = \{f_Y(y; \theta) : E(Y - \theta) = 0; E(Y^2 - (2\theta^2 + \sigma^2)) = 0\},$$

and $\theta \in \Theta = (0, \infty)$. To make it operative, $\Phi(\theta)$ is replaced, in the usual way, by its empirical analogue $\Phi_q(\theta)$. Observe that the model is just a modification of Example 1 of QL, so the argument of Sect. 3.1 can be employed to find probability that in this setting $\Phi_q(\Theta)$ is empty. The sets $\Phi_q^m(\Theta)$ and $\Phi_q(\Theta)$ are empty for those data sets for which $L(\hat{q}_{(1)}^m)$ from (5) is smaller than σ^2 . For $\sigma = 3$ and $n = 100$ the probability is 0.215 (estimated by 10000 MC runs).

If also $\sigma \in \Sigma = (0, \infty)$ was unknown, then the condition for presence of ESP becomes $L(\hat{q}_{(1)}^m) < 0$. For the true $\sigma = 3$ and $n = 10$ the probability of observing a data with ESP is 0.084, estimated by 10000 MC runs. For $n = 100$, the probability becomes smaller than 1/10000.

4.2. Score E^3

Let $r_X(x; \theta)$ and Θ be such that the Maximum Likelihood (ML) estimator $\hat{\theta}_{ML}$ of θ is identical with the solution of the score equations. If $\Phi(\Theta)$ is defined by estimating functions which are based on the score equations, then the corresponding $\Phi_q(\Theta)$ is free of the empty set problem, for any random sample X_1^n drawn from $r_X(x; \theta)$. Any of the GMC estimators is then identical with $\hat{\theta}_{ML}$. A trivial example is given by the gaussian $n(0, 1)$ sampling distribution, where the

score equation for the location parameter θ is $\frac{1}{n} \sum_{i=1}^n X_i - \theta = 0$. Then for the estimating function $u(X; \theta) = X - \theta$ the set $\Phi_q(\Theta)$, where $\Theta = \mathbb{R}$, cannot be empty for any X_1^n from $n(0, 1)$. Changing Θ into the halfline $[0, \infty)$ gives rise to ESP; cf. Sect. 4.4.

4.3. Discrete random variable

It is clear that the empty set problem is not a peculiarity of a continuous random variable. As an illustration, consider an over-identified model given as the set of pmfs which satisfy the estimating equations for the estimating functions $u_1(x; \theta) = X - \theta$, $u_2(x; \theta) = X^2 - \theta^2 - \theta$, $\theta \in \Theta = [0, \infty)$, and defined on the support $\mathcal{X} = \{0, \dots, \infty\}$. Its empirical analogue $\Phi_q(\Theta)$ is empty for any data X_1^n , for which

$$q_{(1)}x_{(1)}^2 + (1 - q_{(1)})x_{(n)}^2 - (q_{(1)}x_{(1)} + (1 - q_{(1)})x_{(n)})^2 - q_{(1)}x_{(1)} - (1 - q_{(1)})x_{(n)} < 0,$$

where $q_{(1)} = 0.5 + 0.5/(x_{(n)} - x_{(1)})$. This follows from the same reasoning as in Sect. 3.1. For the Poisson sampling distribution $\text{Poi}(1)$, the probability that for a sample of size $n = 10$ there will be no empirical estimator is 7/1000 (estimated by 10000 MC runs).

4.4. Restricted parameter space

Let the sampling distribution $r_X(x; \theta)$ be parametrized by a parameter θ which can take on any value in $\Theta = \mathbb{R}$. Assume, for the sake of simplicity, the exactly identified model specified by a single estimating function $u(X; \theta) = X - \theta$, and by $\theta \in \tilde{\Theta} = (a, b)$, a, b are finite. The model is misspecified in the sense that $\tilde{\Theta} \subset \Theta$. In the E^3 approach, the model $\Phi(\tilde{\Theta})$ is replaced by its empirical analogue $\Phi_q(\tilde{\Theta})$. The probability that $\Phi_q(\tilde{\Theta})$ is empty depends on $r_X(x; \theta)$, the values of a, b , and the sample size n . Consider the following illustrative example. Let the sampling distribution be $n(0, 1)$, $n = 100$, $\tilde{\Theta} = (2, 3)$. The probability that the set $\Phi_q(\tilde{\Theta})$ is empty is essentially determined by the probability that all the observed values are below 2, which is 0.100. An extreme case of this setting was considered in Sect. 3.3, where $\tilde{\Theta}$ comprised a single element.

Another possibility for getting an empty $\Phi_q(\Theta)$, is by restricting the parameter space Θ of the sampling distribution. As an illustration, assume the gaussian $n(\theta, 1)$ sampling distribution, with $\theta \in \Theta = [0, \infty)$, and let the model be given by the estimating function $u(X; \theta) = X - \theta$. Whenever the sample X_1^n is such that all its elements are smaller than 0, the parametric space restriction cannot be satisfied; cf. also Owen (2001), Sect. 10.5. Note, that for such a sample the parametric space restriction can be satisfied within the modified E^3 , and hence, there is no affine ESP in this model.

Two possibilities are pertinent also for the over-identified (i.e., $J > K$) and the under-identified (i.e., $J < K$) models. As an example of the under-determined model consider the one given by the estimating function $u(X; \theta$,

$\sigma^2) = (X - \theta)^2 - \sigma^2$, where θ , σ^2 are parameters of the sampling distribution; $\theta \in \Theta = \mathbb{R}$, $\sigma^2 \in \Sigma = (0, \infty)$. Let the space of values which both θ and σ^2 can take be restricted in the model, in such a way that the resulting set of distributions is non-empty. The corresponding empirical counterpart of the model can be empty for any data which, regardless of q , cannot satisfy restrictions of the parametric space.

5. Implications for theory and practice

The empty set problem (ESP) of the empirical estimating equations (E^3) approach to estimation and inference places a limit on the approach and means that every model should be checked for emptiness of the set $\Phi_q(\Theta)$. To the best of our knowledge this check has not been done in the past. As a result, studies of finite-sample properties of the E^3 -based methods may be distorted for those models that are subject to ESP.

If the proposed model $\Phi_q(\Theta)$ is non-empty for all data sets, then it is free of ESP, and, in principle, any E^3 -based method⁴ can be used to select the representative parametrized, data-supported probability mass function $q(\cdot; \theta)$ from $\Phi_q(\Theta)$. If the model is subject to ESP, then no E^3 -based estimate can be obtained for those data sets, for which the set $\Phi_q(\Theta)$ is empty. Nonexistence of the estimate for the data set also affects all the methods that are based on it; e.g., the Empirical Likelihood confidence intervals, bootstrap confidence intervals from the Maximum Nonparametric Likelihood estimator (cf. Owen (2001), Sect. 9.8), model selection by the Empirical Likelihood-like criteria (cf. Hong, Preston and Shum (2003)).

In some models that are subject to ESP an escape route may be provided by lifting the non-additivity constraints; i.e., by replacing the model $\Phi_q(\Theta)$ with the modified model $\Phi_q^m(\Theta)$. However, this may not always help. There are models (cf. Sect. 3.1 for an example) where the modified E^3 approach is subject to the affine ESP. For such models, the existence of estimators that are obtained by methods that operate within the modified E^3 (mE^3) approach can be data-dependent.

If MEL and other E^3 -based methods are to be used, then models should be checked on case-by-case basis for the absence or presence of the empty set problem. In simple models such as those discussed in Sects. 3 and 4 it is possible to check presence of ESP analytically.

Routine numerical computations of E^3 -based estimates in a MC study may not reveal the presence of ESP. For instance, in context of the Example 1 of QL (cf. Sect. 3.1) data for which the set $\Phi_q(\Theta)$ is empty can lead to a meaningfully looking MEL estimate. Only a review of the 'optimal' result of the inner optimization in (2), will reveal that it failed. On the other hand, difficulty with the inner optimization cannot be taken as a proof that the set $\Phi_q(\Theta)$ is empty, for the data in hand.

⁴See Grendár and Judge (2009a) for an asymptotic argument that singles out MEL among the E^3 -based methods.

As a possible exploratory tool of presence of ESP for an E^3 model we suggest that the E^3 equations be combined into a set of equations that does not depend on θ . Then for a batch of samples of a small sample size, check by numerical optimization over q , whether the equations can be satisfied. Clearly, this tool cannot be used to claim nonexistence of ESP if it does not show up in the batch of samples; but it reveals ESP if for some data set the equations cannot be satisfied.

Acknowledgments

We are indebted to Art Owen for comments, questions and suggestions on an earlier version of this work, which resulted in the extension of ESP into aESP. Suggestions from Amos Golan, Peter Hall, Klára Hornišová, Lukáš Lafférs, Nicole Lazar, Ron Mittelhammer, Andrej Pázman, Jing Qin, Giuseppe Ragusa and Viktor Witkovský are gratefully acknowledged. Thanks are also extended to an anonymous reviewer and an associate editor. This note was initiated by a study of [Gzyl and Ter Horst \(2009\)](#).

References

- BACK, K., AND BROWN, D. (1990). Estimating distributions from moment restrictions. Working paper, Graduate School of Business, Indiana University.
- BICKEL, P.J., KLAASSEN, C.A.J., RITOV, Y., AND WELLNER, J.A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: Johns Hopkins University Press. [MR1245941](#)
- BROWN, B.M., AND CHEN, S.X. (1998). Combined and Least Squares Empirical Likelihood. *Ann. Inst. Statist. Math.* 90:443-450. [MR1671990](#)
- CHEN, J., VARIYATH, A.M., AND ABRAHAM, B. (2008). Adjusted Empirical Likelihood and its properties. *J. Comput. Graph. Statist.* 17/2:426-443. [MR2439967](#)
- CHEN, S.X. AND VAN KEILEGOM, I. (2009). A review on empirical likelihood methods for regression (with discussion and rejoinder by the authors). *Test.* 18/3. November, 2009. DOI: 10.1007/s11749-009-0159-5
- CHENG, J., SMALL, D.S., TAN, Z., TEN HAVE T.R. (2009). Efficient nonparametric estimation of causal effects in randomized trials with noncompliance. *Biometrika.* 96/1:19-36.
- CORCORAN, S.A. (1998). Bartlett adjustment of empirical discrepancy statistics. *Biometrika.* 85:967-972.
- CORCORAN, S.A. (2000). Empirical exponential family likelihood using several moment conditions. *Stat. Sinica.* 10:545-557.
- CRÉPET, A., HARARI-KERMADEC, H., AND TRESSOU, J. (2009). Using Empirical Likelihood to combine data: application to food risk assessment. *Biometrics.* 65:257-266.
- CRESSIE, N., AND READ, T. (1984). Multinomial goodness of fit tests. *J. Roy. Statist. Soc. Ser. B.* 46:440-464. [MR0790631](#)

- DEVEREUX, P.J. AND TRIPATHI, G. (2009). Optimally combined censored and uncensored datasets. *J. Econometrics*. 151:17-32.
- EMERSON, S.C. AND OWEN, A.B. (2009). Calibration of the empirical likelihood method for a vector mean. *Electron. J. Statist.* 3:1161-1192.
- GODAMBE, V.P., AND KALE, B.K. (1991). Estimating functions: an overview. In *Estimating Functions*, V.P. Godambe (ed.), Oxford University Press, Oxford, UK, pp. 3-20. [MR1163992](#)
- GRENDÁR, M., AND JUDGE, G. (2009). Asymptotic equivalence of Empirical Likelihood and Bayesian MAP. *Ann. Statist.* 37/5A:2445-2457.
- GRENDÁR, M., AND JUDGE, G. (2009). Maximum Empirical Likelihood: Empty Set Problem. UC Berkeley: CUDARE Working Paper No. 1090. Retrieved from: <http://escholarship.org/uc/item/71v338mh>
- GZYL, H., AND TER HORST, E. (2009). Recovering decay rates from noisy measurements with Maximum Entropy in the Mean. *J. Prob. Stat.* Article ID 563281, 13 pages, doi:10.1155/2009/563281.
- HABERMAN, S. (1984). Adjustment by Minimum Discriminant Information. *Ann. Statist.* 12/3:971-988. [MR0751286](#)
- HJORT, N.L., MCKEAGUE, I.W., AND VAN KEILEGOM, I. (2009). Extending the scope of Empirical Likelihood. *Ann. Statist.* 37/3:1079-1111.
- HONG, H., PRESTON, B., AND SHUM, M. (2003). Generalized Empirical Likelihood-based model selection criteria for moment conditions models. *Econometric Theory*. 19:923-943. [MR2015971](#)
- IMBENS, G.W., SPADY, R.H., AND JOHNSON, P. (1998). Information theoretic approaches to inference in moment condition models. *Econometrica*. 66/2:333-357. [MR1612246](#)
- KITAMURA, Y., AND STUTZER, M. (1997). An information-theoretic alternative to Generalized Method of Moments estimation. *Econometrica*. 65:861-874. [MR1458431](#)
- KITAMURA, Y. (2007). Nonparametric Likelihood: efficiency and robustness. *Jap. Econ. Rev.* 58/1:26-45.
- LIU, Y. AND CHEN, J. (2009). Adjusted empirical likelihood with high-order precision. *Ann. Statist.* (to appear).
- MITTELHAMMER, R., JUDGE, G. AND MILLER, D. (2000). *Econometric Foundations*. Cambridge:Cambridge University Press. [MR1789434](#)
- OWEN, A.B. (2001). *Empirical Likelihood*. New York:Chapman-Hall/CRC.
- QIN, J., AND LAWLESS, J. (1994). Empirical Likelihood and General Estimating Equations. *Ann. Statist.* 22:300-325. [MR1272085](#)
- ROCKAFELLAR, R.T. (1970). *Convex Analysis*. New Jersey: Princeton University Press. [MR0274683](#)
- SCHENNACH, S.M. (2005). Bayesian exponentially tilted empirical likelihood. *Biometrika*. 92/1:31-46. [MR2158608](#)
- SMITH, R.J. (1997). Alternative semi-parametric likelihood approaches to Generalized Method of Moments estimation. *Economic J.* 107:503-519.
- ZHOU, H., ZHOU, H., CHEN, J., LI, Y., LIEBERMAN, J., STYNER, M. (2009). Adjusted Exponentially Tilted Likelihood with applications to brain morphology. *Biometrics*. 65:919-927.

- WANG, D. AND CHEN, S.X. (2009). Combining quantitative trait loci analyses and microarray data: an empirical likelihood approach. *Comp. Stat. Data Anal.* 53:1661-1673.
- WONG, H., LIU, F., CHEN, M., IP, W.CH. (2009). Empirical Likelihood based diagnostics for heteroscedasticity in partial linear models. *Comp. Stat. Data Anal.* 53:3466-3477.