AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
<http://ageconsearch.umn.edu>
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

*Department of Agricultural &
Resource Economics, UCB*
CUDARE Working Papers
(University of California, Berkeley)

Year 2006

Paper 1012

Large Deviations Theory and Empirical
Estimator Choice

Marian Grendar *

George G. Judge †

*Institute of Measurement Sciences SAS, Bratislava, Slovakia

†University of California, Berkeley and Giannini Foundation

This paper is posted at the eScholarship Repository, University of California.

http://repositories.cdlib.org/are_uchb/1012

Copyright ©2006 by the authors.

Large Deviations Theory and Empirical Estimator Choice

Abstract

Criterion choice is such a hard problem in information recovery and in estimation and inference. In the case of inverse problems with noise, can probabilistic laws provide a basis for empirical estimator choice? That is the problem we investigate in this paper. Large Deviations Theory is used to evaluate the choice of estimator in the case of two fundamental situations-problems in modelling data. The probabilistic laws developed demonstrate that each problem has a unique solution-empirical estimator. Whether other members of the empirical estimator family can be associated a particular problem and conditional limit theorem, is an open question.

Large Deviations Theory and Empirical Estimator Choice*

Marian Grendar[†]

George Judge[‡]

Abstract

Criterion choice is such a hard problem in information recovery and in estimation and inference. In the case of inverse problems with noise, can probabilistic laws provide a basis for empirical estimator choice? That is the problem we investigate in this paper. Large Deviations Theory is used to evaluate the choice of estimator in the case of two fundamental situations-problems in modelling data. The probabilistic laws developed demonstrate that each problem has a unique solution-empirical estimator. Whether other members of the empirical estimator family can be associated a particular problem and conditional limit theorem, is an open question.

*Date: Jan 5, 2006. JEL Classification: C111; C10; C13; C24. Keywords: Entropy; Information Theory; Large Deviations; Empirical Likelihood; Boltzmann Jaynes Inverse Problem; Probabilistic Laws

[†]Dept. of Mathematics, FPV UMB, Banska Bystrica; Inst. of Mathematics and CS of Slovak Academy of Sciences (SAS), Banska Bystrica; Institute of Measurement Sciences SAS, Bratislava, Slovakia. Email: marian.grendar@savba.sk. Supported by VEGA 1/0264/03 grant.

[‡]207 Giannini Hall, University of California, Berkeley, CA, 94720. Email: judge@are.berkeley.edu.

1 Introduction

Within the context of Large Deviations Theory, which involves rare events [11], we consider two fundamental situations that capture two basic approaches dealing with data and a (parametric) statistical model. We designate these two situations as α and β problems. The α -problem concerns selecting an empirical distribution from a parametric, in general, set when the source of the data is known. A problem of this form was first considered by Boltzmann in the 1870's. A heuristic approach discussed in [14] led Boltzmann to select an empirical distribution with the highest value of a quantity which some 70 years later was emphasized by Shannon [33]. At about the same time the entropy maximization method (MaxEnt) was promoted outside the area of the original problem by Jaynes [19]. Though many other researchers contributed substantially to this area it seems fair to call the α -problem a Boltzmann Jaynes Inverse Problem. In this work we are interested in an empirical variant of the α -problem, where instead of the true source only a data-based estimate is available. The other β problem, we investigate, concerns a single empirical distribution and a parametric, in general, set of its possible sources. The focus of the empirical α -problem is data-centric in attitude, as it uses data in the position of a source (generator) and a model in the position of data. In contrast the β -problem is model-oriented, as it uses the model in the form of a set of distributions in the source position.

In a recent work Kitamura and Stutzer [23], [24] recognize that Large Deviations Theory for Empirical Measures (Types) through its fundamental result which is commonly known as the Gibbs' Conditioning Principle [9], [11], can be used to provide a probabilistic justification for an estimator we call Empirical MaxMaxEnt (EMME) estimator. No other empirical estimator satisfies Gibbs' Conditioning Principle. Stated from a different point of view: any empirical estimator, other than EMME, violates a fundamental probabilistic law. However, the Gibbs' Conditioning Principle is only pertinent to the α -problem. It is thus correct to conclude that in the area of the empirical α -problem an application of any empirical estimator other than EMME violates the Gibbs' Conditioning Principle.

In the case of the β -problem recently developed Large Deviations Theorems for Sources [16] provide a probabilistic justification of the Empirical Likelihood (EL) method. This implies that in this context any empirical estimator other than EL violates these probabilistic laws.

The paper is organized as follows: first, the α -problem is stated, the relevant Large Deviations Theorems for Types are mentioned and their consequences for how the α -problem should be solved are demonstrated. Next, the empirical form of the α -problem is introduced, and it is noted that application of any other method than the EMME would violate the Large Deviations Theorems. The continuous case of the empirical α -problem presents a challenge and two ways how to achieve a solution are discussed. Then, the β -problem is formulated and it is discussed along the lines used in the α -problem. Relevant Large Deviations Theorems for Sources are briefly surveyed and their implications for

solving the β -problem are drawn, next. The paper concludes with a summary that discusses the implications of our application of the probabilistic laws. An Appendix contains formal statements regarding relevant Large Deviations Theorems.

2 The α -problem and Large Deviations Theorems for Types

Assume a random sample $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ is drawn from probability mass function (pmf) q , that is defined on a finite support $\mathcal{X} \triangleq \{x_1, x_2, \dots, x_m\}$. In Information Theory parlance, q is called a source, $\mathcal{X}$ is the alphabet, and the random sample is a sequence of letters. Information theorists have an apt name for a frequency distribution, they call it a type or n -type and stress that it is induced by a sequence of length n . Thus, a type v^n is an m -element vector, whose i -th element is the relative frequency with which the i -th letter occurred in a sequence of length n .

Given the basic terminology, the α -problem can be informally stated as follows: there is the source q and a set Π_n of n -types and the objective is to select an n -type (one or more) from the set Π_n . To solve the α -problem it is necessary to provide an algorithm-technique for selecting a type(s) from Π_n when only an information-quadruple $\{n, \mathcal{X}, q, \Pi_n\}$ is supplied.

Remarks: 1) If Π_n contains more than one type, then the problem becomes under-determined, and in this sense ill-posed.

2) It is more advantageous to state the problem in terms of a set of pmf's Π , to which the n -types have to belong. Obviously, $\Pi_n \subset \Pi$.

3) The α -problem becomes non-trivial if Π does not contain the source q . Such a feasible set is said to be rare.

Example 1. [15] Let $n = 6$, $\mathcal{X} = [1 \ 2 \ 3 \ 4]$, $q = [0.1 \ 0.6 \ 0.2 \ 0.1]$ and let $\Pi \triangleq \{p : 0 \leq p_1 \leq 1/3, 0 \leq p_2 \leq 2/3, 0 \leq p_3 \leq 2/3, 1/3 \leq p_4 \leq 2/3\}$. *The α -problem is to select a type(s) from Π , when only the information-quadruple $\{n, \mathcal{X}, q, \Pi\}$ is given.*

The part of Probability Theory commonly known as the Large Deviations Theory is concerned with rare events. Sanov's Theorem, the basic Theorem of Large Deviations Theory for Types, states that the probability that the source q generates an n -type from an open set Π decays exponentially fast, as $n \rightarrow \infty$. The decay rate is determined by the infimal value of the information divergence (Kullback-Leibler distance, minus relative entropy etc.) $I(p||q) \triangleq \sum_{i=1}^m p_i \log \frac{p_i}{q_i}$ over $p \in \Pi$.

Provided that the source q generated a type from a rare set Π , it is of interest to know how the conditional probability is spread among the types; in particular as $n \rightarrow \infty$. For rare sets of certain form, this issue is covered by the Conditional Limit Theorem. The theorem states that if q generated a type from a rare, convex and closed set Π , then for $n \rightarrow \infty$ it is with probability one a type which

is arbitrarily close to the I -projection $\hat{p}$ of the source q on Π . The information projection, or I -projection for short, is defined as $\hat{p} \triangleq \arg \inf_{p \in \Pi} I(p||q)$.

The Conditional Limit Theorem (CoLT) suggests that, conditional on the rare Π , it is the I -projection $\hat{p}$ rather than q , which should be considered as the true *iid* source of data. Gibbs' Conditioning Principle (GCP), an important strengthening of CoLT, captures this 'intuition'.

Let us finally note that the Large Deviations Theorems for Types (i.e., CoLT and GCP) also hold for the case of a continuous random variable X . In this case, the concept of type is defined via discretization and a suitable topology has to be introduced on the set of all probability measures with support $\mathcal{X}$.

The most commonly considered feasible set Π is defined via moments; $\Pi \triangleq \{p : \sum_{i=1}^m p_i(u_j(x_i) - a_j) = 0, 1 \leq j \leq J\}$, where $u_j(\cdot)$ is a u -moment and a is a J -vector of given numbers. In this case the I -projection of q on Π is known to belong to the exponential family $\mathcal{E}(X, \lambda, u) \triangleq k(\lambda)q(X) \exp(-\sum_{j=1}^J \lambda_j u_j(X))$, where $k(\lambda) \triangleq \sum_{i=1}^m q(x_i) \exp(-\sum_{j=1}^J \lambda_j u_j(x_i))$ is the normalizing constant.

Example 2. Let $\mathcal{X}$, q and n be the same as in the Example 1. Let $j = 1$, $u(x) = x$, $a = 2.6$. Thus, $\Pi \triangleq \{p : \sum_{i=1}^m p_i(x_i - 2.6) = 0\}$. The set Π_6 contains eight 6-types, among which a type has to be selected.

The moment function $u(\cdot)$ can be viewed as a special case of a general, parametric $u(X, \theta)$ -moment function, where $\theta \in \Theta$ is a (vector) parameter. The moment function $u(X, \theta)$ is commonly known as an unbiased estimating function. Given this base Π turns into a parametric feasible set $\Pi(\Theta) \triangleq \{p(\theta) : \sum_{i=1}^m p_i(\theta)u_j(X, \theta) = 0, \theta \in \Theta, 1 \leq j \leq J\}$.

Example 3. Let $\mathcal{X}$, q and n be the same as in the Example 1. Let $j = 1$, $u(x, \theta) = x - \theta$, $\theta \in \Theta$, where $\Theta \triangleq]1.7, 2.6]$. Thus, $\Pi(\theta) \triangleq \{p(\theta) : \sum_{i=1}^m p_i(\theta)(x_i - \theta) = 0\}$.

The Conditional Limit Theorem and Gibbs' Conditioning Principle have a significant bearing for the α -problem, as they imply that at least for sufficiently large n the problem has to be solved by means of the I -projection of q on Π . If instead of the Relative Entropy Maximization method the Maximum Non-parametric Likelihood method which selects $\tilde{p} \triangleq \arg \inf_{p \in \Pi} I(q||p)$ is used, then this method would for n sufficiently large select a type that has conditionally almost a zero probability of occurring – an unfortunate modelling strategy.

2.1 Empirical α -problem

The source q is rarely known. If from q a random sample of size N can be drawn then it is possible to estimate the source by the sample-based N -type, and use it in place of q in the α -problem. This way the problem turns into the *empirical* α -problem. Replacing q by its sample estimate v^N has an important implication: even if Π was such that q belongs to it, with respect to the sample estimate of q the feasible set Π will be rare.

The empirical α -problem consists of two 'ingredients': the estimate v^N of the source based on a sample of size N and a set Π of n -types. To make the

problem a bit harder, a parametric $\Pi(\Theta)$ will be assumed; and to make it a bit easier we assume that n is sufficiently large so that instead of the set of n -types that are of necessity rational the set $\Pi(\Theta)$ of pmf's can be used. This instance of the empirical α -problem can be viewed as a problem of selecting a parametric pmf from $\Pi(\Theta)$ when an estimate v^N of the true source q is known. It seems reasonable to call $\Pi(\Theta)$ a parametric model of the source q , and view the problem of selecting $p(\theta)$ as a problem of estimating the 'true' value of θ . Note that the model is used to form the parametric feasible set of types/pmf's.

Example 4. Let $\mathcal{X}$, q , $\Pi(\Theta)$ be the same as in the Example 3. Let $N = 13$ and an N -sample induce $v^{13} = [2 \ 9 \ 1 \ 1]$. This is the estimate of the true source q . Let n be sufficiently large so that $\Pi(\Theta)$ is effectively a set of pmf's. The empirical α -problem is to select a pmf/type from $\Pi(\Theta)$ given the information-quadruple $\{n, \mathcal{X}, v^N, \Pi(\Theta)\}$.

Provided that $\Pi(\Theta)$ is a convex, closed set, if CoLT and GCP are applied to the empirical α -problem this implies that

$$\hat{p}(\theta) = \arg \inf_{p \in \Pi(\Theta)} I(p(\theta) || v^N) \quad (1a)$$

with $\theta = \hat{\theta}$, where

$$\hat{\theta} = \arg \inf_{\theta \in \Theta} I(\hat{p}(\theta) || v^N) \quad (1b)$$

must be selected. The estimator $\hat{\theta}$ given by (1b) is known as the Maximum Entropy Empirical Likelihood [28] or Exponentially Tilted estimator [1], [2], [18], [23]; we call it Empirical Maximum Maximum Entropy (EMME) estimator. Information divergence is also known as minus relative entropy and minimization of the information divergence is equivalent to maximizing relative entropy. The suggested name arises from the fact that the equations (1a) and (1b), which define the EMME estimator, can be equivalently written in the form of a double maximization of the relative entropy, where the inner maximization is taken over $p \in \Pi(\Theta)$ and the outer maximization over $\theta \in \Theta$. The empirical attribute of the estimator stems for the fact that the source q is replaced by its empirical estimate v^N .

Any other way of solving the empirical α -problem would violate the conditional limit theorems.

2.1.1 Empirical α -problem: the continuous case

To construct the analogous empirical α -problem for continuous $\mathcal{X}$ is challenging. Though the true source $q(X)$ can be estimated by an N -type obtained by a discretization of the random sample, the feasible set contains continuous probability density functions (pdf's).

There are two possibilities to solve this conflict: 1) construct a continuous estimate of the source (e.g., by a kernel estimator), or 2) use the Empirical Estimation trick (cf. [31], [28], Ch. 12) of creating another layer. This is done

by forcing the observed random sample to become a support $\mathcal{S}$ of a random variable S with a uniform distribution u . (Alternatively, the trick could be explicated as forming a Dirac-type estimator of the continuous source.) The feasible set $\Pi(\Theta)$ of pdf's thus turns into a set $\Pi_S(\Theta) \triangleq \{p_S(\theta) : \sum_{l=1}^N p_S(s_l; \theta) u_j(x_l; \theta) = 0, 1 \leq j \leq J\}$ of pmf's on the support $\mathcal{S}$.

To the continuous- X case of the empirical α -problem, transformed via the Empirical Estimation trick, the Large Deviations Theorems for Types can be applied. These Theorems dictate that we select

$$\hat{p}_S(\theta) = \arg \inf_{p_S \in \Pi_S(\Theta)} I(p(\theta) || u), \quad (2a)$$

with $\theta = \hat{\theta}$, where

$$\hat{\theta} = \arg \inf_{\theta \in \Theta} I(\hat{p}_S(\theta) || u) \quad (2b)$$

is the continuous-case EMME estimator.

Note that the Empirical Estimation trick could as well be applied in the case of discrete $\mathcal{X}$ where it, obviously, can be collapsed into the formulation given in Section 2.1.

Kitamura and Stutzer [23] avoid the need for the two-level trick, by solving the discrete-continuous conflict as follows: assume that the continuous source $q(X)$ is known. Then we face the standard α -problem, albeit with continuous random variable X . The Large Deviations Theorems single out the MaxMaxEnt estimator (cf. Eq. 3):

$$\hat{p}(X; \theta) = \arg \inf_{p(X) \in \Pi(\Theta)} I(p(X; \theta) || q(X)),$$

with $\theta = \hat{\theta}$, where

$$\hat{\theta} = \arg \inf_{\theta \in \Theta} I(\hat{p}(X; \theta) || q(X)). \quad (3)$$

There, the continuous I -divergence $I(p(X) || q(X)) \triangleq \int_{\mathcal{X}} p(X) \log \frac{p(X)}{q(X)} \lambda(dx)$ is used. The convex dual problem to (3) is:

$$\hat{\theta} = \arg \inf_{\theta \in \Theta} \sup_{\lambda \in \mathbb{R}} E_q \log \hat{p}(X; \lambda, \theta), \quad (4)$$

where $\hat{p}(X; \lambda, \theta)$ belongs to the exponential family $\mathcal{E}(X, \lambda, u(\theta))$. If instead of the continuous source $q(X)$ a random sample of size N exists, only, (thus we face the continuous empirical α -problem) – it seems reasonable to replace the expectation of (4) by its estimator $\frac{1}{N} \sum_{l=1}^N \log \hat{p}(X_l; \lambda, \theta)$. It is straightforward to see that the estimator obtained this way is identical to EMME estimator given by Eq. (2b).

In general, any empirical estimator based on a convex statistical discrepancy measure for which the implied probability distribution is not self-referential, can be obtained in the Kitamura Stutzer way.

3 The β -problem and Large Deviations Theorems for Sources

In the α -problem, there is a source q and a set Π of n -types. The task is to select an n -type. The β -problem is, in a sense, opposite to the α -problem.

In the β -problem there is an N -type v^N and a set $\mathcal{Q}$ of its possible n -sources, where an n -source is a rational pmf, such that all of its m denominators are n . The task is to select an n -source from the set. Note that this problem is already in empirical form, as the type v^N is an N -sample-based estimate of the true source r . To solve the β -problem a method for selecting an n -source from $\mathcal{Q}$ is required, when the information quadruple $\{v^N, \mathcal{X}, n, \mathcal{Q}\}$ and nothing else is supplied. The most common case is when n is sufficiently large, so that the set of rational n -sources can effectively be identified with the set of pmf's.

Example 5. Let $\mathcal{X}$, $j = 1$, $u(x) = x$, $a = 2.\bar{6}$ be the same as in the Example 2. And let $N = 13$, v^{13} be the same as in the Example 4. Thus, the set of sources is $\mathcal{Q}(\theta) \triangleq \{q : \sum_{i=1}^m q_i x_i = a\}$. Let n be sufficiently large, so that $\mathcal{Q}$ can be identified with a set of pmf's. The β -problem is to select an n -source q from $\mathcal{Q}$ when the information-quadruple $\{n, \mathcal{X}, v^N, \mathcal{Q}\}$ is supplied.

If $\mathcal{Q}$ does not contain v^N then it is a rare set. In contrast to the Large Deviations Theorems for Types there are Large Deviations Theorems for Sources [16]. The Static Sanov's Theorem for Sources (LST) describes the asymptotic exponential decay of a probability that an n -source q^n of the N -type v^N belongs to an open set $\mathcal{Q}$. The decay rate is provided by the supremal value of the L -divergence $L(q||v^N) \triangleq \sum_{i=1}^m v_i^N \log q_i$ of q with respect to v^N , over $q \in \mathcal{Q}$. The Static Conditional Limit Theorem for Sources (LCoLT) states that given that the N -type and an n -source from a rare convex closed set $\mathcal{Q}$ has occurred, then asymptotically it is almost impossible to find a source other than those arbitrarily close to the L -projection $\hat{q}$ of v^N on $\mathcal{Q}$. The likelihood projection, or L -projection for short, is defined as $\hat{q} \triangleq \arg \sup_{q \in \mathcal{Q}} L(q||v^N)$.

Thus, the β -problem should for sufficiently large n be solved by selecting the L -projection of the observed N -type on the set of sources $\mathcal{Q}$. Any other selection scheme would select a source which is asymptotically conditionally impossible. It is assumed in the Large Deviations Theorems for Sources that there is a prior distribution of sources.

Extensions of the Limit Theorems for Sources to the case of the continuous random variable X are also possible; cf. [16].

Next consider the parametric case of the β -problem, where the feasible set of sources is defined by means of the $u(X, \theta)$ -moments; $\mathcal{Q}(\Theta) \triangleq \{q(\theta) : q_i(\theta) u_j(X, \theta) = 0, 1 \leq j \leq J\}$. The set $\mathcal{Q}(\Theta)$ can be interpreted as a parametric model of the true source r . The β -problem thus becomes a problem of selecting a source from $\mathcal{Q}$ of the observed N -type v^N . It is useful to compare this problem with the parametric empirical α -problem (Sect 2.1).

Provided that $\mathcal{Q}(\Theta)$ is a convex, closed, rare set, the Static CoLT for Sources

applies and suggests we select

$$\hat{q}(\theta) = \arg \sup_{q \in \mathcal{Q}(\Theta)} L(q(\theta) || v^N)$$

with $\theta = \hat{\theta}$, where

$$\hat{\theta} = \arg \sup_{\theta \in \Theta} L(\hat{q}(\theta) || v^N). \quad (5)$$

The estimator (5) is known as the Empirical Likelihood (EL) estimator.

In the above discussion $\mathcal{X}$ was assumed to be a finite set. The case of continuous $\mathcal{X}$ poses a challenge, similar to that of Sect 2.1.1. The random sample has to be discretized in order to form an n -type; on the other hand $\mathcal{Q}(\Theta)$ is a set of pdf's. The discrete-continuous conflict can be resolved either via the Empirical Estimation trick, or through the Kitamura Stutzer [23] approach. They lead to the same continuous-case EL estimator, which is commonly considered in the literature, cf. [31], [28].

4 Summary

In an influential paper on the Empirical Likelihood (EL) method, Owen [30] noted that for the construction of an empirical estimator a distance-divergence measures other than the non-parametric likelihood could be used. Since this seminal article, several such discrepancy measures in the context of empirical estimation have been investigated. These include Kullback-Leibler divergence, Euclidean distance, and the Cressie-Read (CR) class of divergence measures [7], [8] that encapsulates the two previously mentioned and many other commonly used distance-divergence measures.

A considerable effort has been devoted to finding which of the discrepancy measures should be recommended. In order to solve this Criterion Choice Problem, investigations have focused on the finite sampling properties of the resulting empirical estimators [28]. Among other things it was found out that *i*) all empirical estimators based on CR discrepancies share the same first-order asymptotic properties [2], *ii*) EL is the only Bartlett correctable in the CR class [13], [20], [2], [4] *iii*) EL should typically have lower bias than any other CR-based empirical estimator [29], *iv*) Euclidean distance and other members of CR with negative value of CR parameter could lead to negative probabilities [31], *v*) Euclidean Likelihood is easy to compute [25], [31], *vi*) EL is less robust than EMME (the KL-based empirical estimator), in terms of Huber's influence function [18], [34], *vii*) across the CR class resulting empirical estimators exhibit for small samples different sampling behavior [2], [3], [26]; see also [21], [27], *viii*) the empirical likelihood tests are more powerful than the alternative tests, in Hoeffding's sense [22].

In this work we followed the observation of Kitamura and Stutzer [24], that Large Deviations Theorems for Types (i.e., Conditional Limit Theorem for Types and Gibbs' Conditioning Principle) can be used to provide a probabilistic justification of the Empirical Maximum Entropy method. If this observation

is taken literally, it would imply that any other empirical estimator violates the fundamental probabilistic laws. However, Large Deviations Theorems for Types are pertinent to a certain, though very general, problem, only. We call it the α -problem (cf. Sect 2).

There is also an equally general β -problem, which is in a sense opposite to the empirical α -problem, to which recently developed Large Deviations Theorems for Sources apply. The Theorems imply that the β -problem has to be solved by maximization of the Non-parametric Likelihood criterion, which results into the Empirical Likelihood estimator.

To sum up: the two empirical estimators EMME and EL provide a solution to the respective two empirical problems. Any other empirical estimator when applied to the particular empirical problem would violate the respective Conditional Limit Theorem. For instance application of EL or any other estimator to the empirical α -problem would lead to selection of such a parametric type/pmf which has asymptotically conditionally zero probability of happening – an unfortunate modelling strategy.

The probabilistic laws demonstrate that each of the two problems we have considered has a unique solution-estimator. The laws say nothing about the statistical properties of the resulting estimators. Neither the probabilistic laws say how one should choose among the two problems, which brings us face to face with the real world. Our interpretation of the empirical α and β -problems suggests that if one trusts the data more than the model, then the Data-centric empirical α -problem and hence EMME estimator have to be considered. If one has more confidence in the model than in the data, then the Model-oriented β -problem and hence EL estimator should be selected. Statistical investigations of the robustness of EL and EMME with respect to incorrect problem choice (cf. [34]) could help to provide the needed guidance.

Empirical estimators other than EL and EMME must occupy problem spaces outside of the two problems we have considered. It is not known whether some of the remaining empirical estimators such as the Rényi or integer or non integer Cressie-Read Empirical Estimators can be associated with particular problems and conditional limit theorems. It is possible under certain conditions that the Exponential Empirical Likelihood (EEL) estimator [5] can be justified by sequential application of the Large Deviations Theorems for Types and Sources. Questions such as this relate to work in progress.

References

- [1] Back, K and Brown, D., Estimating distributions from moment restrictions, working paper, Graduate School of Business, Indiana University, 1990.
- [2] Baggerly K. A., Empirical Likelihood as a goodness-of-fit measure, *Biometrika*, 85/3, pp. 535-547, 1998.
- [3] Baggerly K. A., Studentized Empirical Likelihood and Maximum Entropy, technical report, Rice University, Dept. of Statistics, 1999.

- [4] Bravo F, Bartlett-type adjustments for Empirical Discrepancy test statistics, Disc. paper in Economics 2004/14, The University of York, 2004.
- [5] Corcoran S. A., Empirical exponential family likelihood using several moment conditions, *Stat. Sinica*, 10, pp. 545-557, 2000.
- [6] Cover T. and Thomas J., *Elements of Information Theory*, Wiley, NY, 1991.
- [7] Cressie N. and Read T., Multinomial goodness of fit tests, *JRSS Ser. B*, 46, pp. 440-464, 1984.
- [8] Cressie N. and Read T., *Goodness-of-Fit Statistics for Discrete Multivariate Data*, Springer-Verlag, 1988.
- [9] Csiszár I., Sanov property, generalized I -projection and a conditional limit theorem, *Ann. Probab.*, 12, 768-793, 1984.
- [10] Csiszár I. and Shields P., Information Theory and Statistics: A tutorial, *Foundations and Trends in Communications and Information Theory*, 1/4, pp. 1-111, 2004.
- [11] Dembo A. and Zeitouni O., *Large Deviations Techniques and Applications*, Springer, NY, 1998.
- [12] DiCiccio T. J. and Romano J., Nonparametric confidence limits by resampling methods and least favorable families. *I.S.I. Review*, 58, pp. 59-76, 1990.
- [13] DiCiccio T. J., Hall P. J. and Romano J., Empirical Likelihood is Bartlett-correctable, *Ann. Statist.*, 19, pp. 1053-1061, 1991.
- [14] Ellis, R. S., The theory of large deviations: from Boltzmann's 1877 calculation to equilibrium macrostates in 2D turbulence, *Physica D*, 106-136, 1999.
- [15] Grendár M. Jr and Grendár, M., Maximum probability/entropy translating of contiguous categorical observations into frequencies, *Appl. Math. Comput.*, 161, pp. 347-351, 2005.
- [16] Grendár M., Conditioning by rare sources, submitted, 2005.
- [17] Imbens G. W., A new approach to Generalized Method of Moments estimation, Harvard Institute of Economic Research working paper 1633, 1993.
- [18] Imbens, G. W., Spady, R. H. and Johnson, P., Information theoretic approaches to inference in moment condition models, *Econometrica*, 66/2, pp. 333-357, 1998.
- [19] Jaynes, E. T., 1957, Information Theory and Statistical Mechanics, *Phys. Rev.*, Vol. 106, 620 and Vol. 108, 171.

- [20] Jing B.-Y. and Wood T. A., Exponential Empirical Likelihood is not Bartlett correctable, *Ann. Stat.*, 24, pp. 365-369, 1996.
- [21] Judge G. G. and Mittelhammer R. C., A semiparametric basis for combining estimation problems under quadratic loss, *JASA*, 99, pp. 479-487, 2004.
- [22] Kitamura Y., Asymptotic optimality of empirical likelihood for testing moment restrictions, *Econometrica*, 69, pp. 1661-1672, 2001.
- [23] Kitamura Y. and Stutzer M., An information-theoretic alternative to Generalized Method of Moments estimation, *Econometrica*, 65, pp. 861-874, 1997.
- [24] Kitamura Y. and Stutzer M., Connections between entropic and linear projections in asset pricing estimation, *J. Econometrics*, 107, pp. 159-174, 2002.
- [25] Lindsay B. G. and Qu A., Inference functions and quadratic score tests, *Statistical Science*, 18/3, pp. 394-410, 2003.
- [26] Mittelhammer R. C., Judge G. G. and Schoenberg R., Empirical evidence concerning the finite sample performance of EL-type structural equations estimation and inference methods, In *Festschrift in Honor of Thomas Rothenberg*, CUP, 2005.
- [27] Mittelhammer R. C. and Judge G. G., Robust Empirical Likelihood estimation of models with non-orthogonal noise components, Volume in Honor of Henri Theil, *J. Agricult. Appl. Econ.*, 2001.
- [28] Mittelhammer R. C., Judge G. G. and Miller D. J., *Econometric Foundations*, CUP, 2000.
- [29] Newey W. and Smith R. J., Higher-order properties of GMM and Generalized Empirical Likelihood Estimators, *Econometrica*, 72, pp. 219-255, 2004
- [30] Owen A. B., Empirical Likelihood for linear models, *Ann. Statist.*, 19, pp. 1725-1747, 1991.
- [31] Owen A. B., *Empirical Likelihood*, Chapman-Hall/CRC, 2001.
- [32] Qin J. and Lawless J., Empirical Likelihood and General Estimating Equations, *Ann. Statist.*, 22, pp. 300-325, 1994.
- [33] Shannon C. E., A mathematical theory of communication, *Bell System Technical Journal*, vol. 27, pp. 379-423 and 623-656, July and October, 1948.
- [34] Schennach S. M., Exponentially Tilted Empirical Likelihood, working paper, Dept. of Economics, University of Chicago, July 2004.

5 Appendix: Large Deviations Theorems

The basic reference on Large Deviations is [11]. A grasp of what is going on there can be gained from [6], [14], [10]. The results on the Large Deviations for Sources can be found at [16].

Let $\mathcal{P}(\mathcal{X})$ be a set of all probability mass functions on the finite alphabet $\mathcal{X} \triangleq \{x_1, x_2, \dots, x_m\}$. For a $p \in \mathcal{P}(\mathcal{X})$ the support is $S(p) \triangleq \{p_i : p_i > 0, 1 \leq i \leq m\}$. Let $\pi(v^n; q)$ denote the probability that the source q generates an n -type v^n , i.e., $\pi(v^n; q) \triangleq \Gamma(v^n) \prod_{i=1}^m (q_i)^{n v_i^n}$ where $\Gamma(v^n) \triangleq \frac{n!}{\prod_{i=1}^m n_i!}$. Consequently, for $A \subseteq B \subseteq \mathcal{P}(\mathcal{X})$, $\pi(v^n \in A | v^n \in B; q) = \frac{\pi(v^n \in A; q)}{\pi(v^n \in B; q)}$; provided that $\pi(v^n \in B; q) \neq 0$. $\Pi \subseteq \mathcal{P}(\mathcal{X})$ is rare if it does not contain q , the true source. On $\mathcal{P}(\mathcal{X})$ the relative topology is assumed.

Sanov's Theorem (ST) [10] reads:

(ST). *Let Π be a set such that its closure is equal to the closure of its interior. Let q be such that $S(q) = \mathcal{X}$. Then,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \pi(v^n \in \Pi; q) = -I(\Pi | q).$$

Conditional Limit Theorem for Types (CoLT) [6], [11], is a direct consequence of ST.

(CoLT). *Let Π be a convex, closed rare set. Let $B(\hat{p}, \varepsilon)$ be a closed ε -ball defined by the total variation metric, centered at I -projection $\hat{p}$ of q on Π . Then for any $\varepsilon > 0$,*

$$\lim_{n \rightarrow \infty} \pi(v^n \in B(\hat{p}, \varepsilon) | v^n \in \Pi; q) = 1.$$

Gibbs' Conditioning Principle (GCP) [9], [11]:

(GCP). *Let $\mathcal{X}$ be a finite set. Let Π be closed, convex rare set. Let $n \rightarrow \infty$. Then for a fixed t*

$$\lim_{n \rightarrow \infty} \pi(X_1 = x_1, \dots, X_t = x_t | v^n \in \Pi; q) = \prod_{l=1}^t \hat{p}_{x_l}.$$

In order to state the Large Deviations Theorems for Sources, it is necessary to introduce further terminology.

A probability mass function (pmf) from $\mathcal{P}(\mathcal{X})$ is rational if it belongs to the set $\mathcal{R} \triangleq \mathcal{P}(\mathcal{X}) \cap \mathbb{Q}^m$. A rational pmf is n -rational, if denominators of all its m elements are n . Set of all n -rational pmf's q_n – called n -sources – will be denoted by $\mathcal{R}_n$. Let $\mathcal{Q} \subseteq \mathcal{P}(\mathcal{X})$ be the set of sources.

n -sources q^n are assumed to have a uniform prior distribution $\pi(q^n)$. If from $\mathcal{R}_n$ n -source q^n occurs, then the source generates n -type v^n with probability $\pi(v^n | q^n) \triangleq \Gamma(v^n) \prod_{i=1}^m (q_i^n)^{n v_i^n}$.

The Large Deviations for Sources are concerned about the asymptotic behavior of probability $\pi(q^n \in B | (q^n \in \mathcal{Q}) \wedge v^n)$ that if the n -type v^n and an n -source q^n from a rare set $\mathcal{Q}$ occurred, then the n -source belongs to a subset B of $\mathcal{Q}$. Note

that $\pi(q^n \in B | (q^n \in \mathcal{Q}) \wedge v^n) = \frac{\pi(q^n \in B | v^n)}{\pi(q^n \in \mathcal{Q} | v^n)}$; provided that $\pi(q^n \in \mathcal{Q} | v^n) > 0$. The posterior probability $\pi(q^n | v^n)$ is related to already defined probabilities $\pi(v^n | q^n)$ and $\pi(q^n)$ via Bayes's Theorem.

Static Sanov's Theorem for Sources (Static LST) [16]:

(Static LST). *Let v^{n_0} be a type. Let $\mathcal{Q}$ be an open set of sources. Then, for $n \rightarrow \infty$ over a subsequence $n = kn_0$, $k \in \mathbb{N}$,*

$$\frac{1}{n} \log \pi(q^n \in \mathcal{Q} | v^n) = L(\mathcal{Q} || v^{n_0}) - L(\mathcal{P} || v^{n_0}).$$

$\mathcal{Q}$ is rare if it does not contain v^{n_0} . Static Conditional Limit Theorem for Sources (LCoLT) [16]:

(Static LCoLT). *Let v^{n_0} be a type. Let $\mathcal{Q}$ be a convex, closed rare set of sources. Let $\hat{q}$ be the L -projection of v^{n_0} on $\mathcal{Q}$ and let $B(\hat{q}, \varepsilon)$ be a closed ε -ball defined by the total variation metric, centered at $\hat{q}$. Then, for $\varepsilon > 0$ and $n \rightarrow \infty$ over a subsequence $n = kn_0$, $k \in \mathbb{N}$,*

$$\pi(q^n \in B(\hat{q}, \varepsilon) | (q^n \in \mathcal{Q}) \wedge v^n) = 1.$$