MiniMax Entropy and Maximum Likelihood

Complementarity of Tasks, Identity of Solutions

Marian Grendár
Institute of Measurement Science, Slovak Academy of Sciences, Dúbravská cesta 9, 842 19 Bratislava, Slovakia (umergren@savba.sk)

Marián Grendár
Railways of Slovak Republic, DDC, Klemensova 8, 813 61 Bratislava, Slovakia (grendar.marian@zsr.sk)

Abstract. Concept of exponential family is generalized by simple and general exponential form. Simple and general potential are introduced. Maximum Entropy and Maximum Likelihood tasks are defined. ML task on the simple exponential form and ME task on the simple potentials are proved to be complementary in set-up and identical in solutions. ML task on the general exponential form and ME task on the general potentials are weakly complementary, leading to the same necessary conditions. A hypothesis about complementarity of ML and MiniMax Entropy tasks and identity of their solutions, brought up by a special case analytical as well as several numerical investigations, is suggested in this case.

MiniMax Ent can be viewed as a generalization of MaxEnt for parametric linear inverse problems, and its complementarity with ML as yet another argument in favor of Shannon's entropy criterion.

Keywords: simple and general exponential form, simple potential, general potential, Maximum Likelihood task, Maximum Entropy task, MiniMax Entropy task, complementarity

Abbreviations: ML – Maximum Likelihood; ME – Maximum Entropy; MiniMax Ent – MiniMax Entropy; MMM – Modified Method of Moments; FOC – First Order Condition

AMS codes: Primary 62-02; Secondary 62A10, 62A99, 62F10

1. Introduction

A relationship between Maximum Likelihood (ML) and Maximum Entropy (ME, MaxEnt) methods has been noted and investigated many times. Yet it seems to be intricate and puzzling. Jaynes,(Jaynes, 1982), is worth long quoting on the subject

> ..., any MaxEnt solution also defines a particular model for which the predictive distribution using the ML estimates of the parameters, is identical with the MaxEnt distribution. This is essentially the Pitman-Koopman theorem used backwards; given any data the MaxEnt distribution having exponential form, in effect creates a model for which those data would have been sufficient statistics.

> This can give one deeper understanding of the terms 'information' and 'sufficiency' in statistics, but only after some deep thought. As a result, almost every conceivable opinion about the relationship between MaxEnt and ML can be found expressed in the current literature.

Some of the opinions (with different level of generality) can be found at (Kullback, 1968), (Barndorff-Nielsen, 1978), (Dutta, 1966), (Golan, 1998), (Campbell, 1970), (Nishii, 1989), (Mohammad-Djafari and Idier, 1991), (Mohammad-Djafari, 1998). Adding to it other views on MaxEnt itself (like interpreting Shannon's entropy function as minus expected log-likelihood, or restrictive interpretation of the MaxEnt recovered distribution as Maxwell-Boltzmann special member of exponential family, or insisting on non-solvability of Jaynes' die problem by ML method) makes investigation of relationship between MaxEnt and ML adventurous.

In the present article we make a clear distinction between operational mode of MaxEnt and ML methods, by defining MaxEnt task (as a simple instance of MaxEnt method) and also ML task. An analogy between Boltzmann's deduction of equilibrium distribution of an ideal gas in an external potential field and probability distribution leads us to extending exponential family into *general exponential form*, and introducing a notion of *simple potential* and *general potential*. Concept of complementarity is introduced, and complementarity of ME task on simple potential and ML task on simple exponential form is proved. Finally, a hypothesis about complementarity of MiniMaxEnt task on general potential and ML task on general exponential form, suggested by a simple case analytical as well as several numerical calculations, is put forward. The results instantaneously extends to Relative Entropy Maximization (REM)/I-divergence minimization.

2. DEFINITIONS AND NOTATION

The notion of exponential family is extended into simple and general exponential forms.

Definition 1. Let X be a random variable with pmf/pdf $f_X(x)$. If $f_X(x)$ can be written in the form of

$$f_X(x|\boldsymbol{\lambda}) = k(\boldsymbol{\lambda})e^{-U(x,\boldsymbol{\lambda})}$$

where $U(x, \boldsymbol{\lambda})$ is

$$U(x, \boldsymbol{\lambda}) = \boldsymbol{\lambda}'\mathbf{u}(x)$$

a linear combination of functions $\mathbf{u}(x)$ not depending on other parameters, and $k(\boldsymbol{\lambda})$ is normalizing factor, then it has *simple exponential form* . $u(x)$ is called *simple potential.*

If the pmf/pdf can be written in the form of

$$f_X(x|\boldsymbol{\lambda},\boldsymbol{\alpha}) = k(\boldsymbol{\lambda},\boldsymbol{\alpha})e^{-U(x,\boldsymbol{\lambda},\boldsymbol{\alpha})}$$

where $U(x,\boldsymbol{\lambda},\boldsymbol{\alpha})$ is

$$U(x,\boldsymbol{\lambda},\boldsymbol{\alpha}) = \boldsymbol{\lambda}'\mathbf{u}(x,\boldsymbol{\alpha})$$

a linear combination of functions $\mathbf{u}(x,\boldsymbol{\alpha})$ depending on other parameters $\boldsymbol{\alpha}$, and $k(\boldsymbol{\lambda},\boldsymbol{\alpha})$ is normalizing factor, then it has *general exponential form* . $u(x,\boldsymbol{\alpha})$ is called *general potential* .

The $U(\cdot)$ function is called *total potential.*

Note. Any class of pmf/pdf which can be written in the exponential form is equivalently characterized by its exponential form pmf/pdf or by its potentials.

Example 1. $\Gamma(\alpha,\beta)$ distribution has simple exponential form, with total potential $U(x,\boldsymbol{\lambda}) = \lambda_1 x + \lambda_2 \ln x$; $\lambda_1 = \frac{1}{\beta}$ and $\lambda_2 = 1-\alpha$; $u_1(x) = x$ and $u_2(x) = \ln x$ are the potentials. The normalizing factor $k(\lambda_1,\lambda_2) = \frac{1}{\Gamma(1-\lambda_2)\lambda_1^{\lambda_2-1}}$.

Logistic (μ,β) distribution has general exponential form with total potential $U(x,\boldsymbol{\lambda},\boldsymbol{\alpha}) = \lambda_1 u_1(x,\boldsymbol{\alpha}) + \lambda_2 u_2(x,\boldsymbol{\alpha})$, with $\boldsymbol{\lambda} = [\frac{1}{\alpha_2}, 2]$, and the potentials $u_1(\cdot) = \frac{x-\alpha_1}{\alpha_2}$, $u_2(\cdot) = \ln(1+e^{-\frac{x-\alpha_1}{\alpha_2}})$, and $\boldsymbol{\alpha} = [\mu,\beta]$. $k(\alpha_2) = 1/\alpha_2$.

Discrete normal distribution $dn(\lambda,\alpha)$, defined over a support by

$$f_X(x_i|\lambda) = \frac{e^{-\lambda(x_i-\alpha)^2}}{\sum_i e^{-\lambda(x_i-\alpha)^2}}$$

has total potential $U(x,\lambda,\alpha) = \lambda(x-\alpha)^2$. It can be equivalently expressed in simple form with $U(x,\lambda_1,\lambda_2) = \lambda_1 x + \lambda_2 x^2$, where $\lambda_1 = -2\alpha\lambda$ and $\lambda_2 = \lambda$. $\diamondsuit$

Standard definitions of moment and sample mean are extended.

Definition 2. V-moment of random variable X, $\mu(V)$, is for any function $V(X,\boldsymbol{\alpha})$ defined as

$$\mu(V) = \mathrm{E}\, V(X,\boldsymbol{\alpha})$$

Definition 3. Sample V-moment of random variable X, $m(V)$, is for any function $V(X, \boldsymbol{\alpha})$ defined as

$$m(V) = \sum_{i=1}^{m} r_i V(X_i, \boldsymbol{\alpha})$$

where r_i is frequency of i-th element of support in sample.

Definition 4. Let $\mu(V)$, $m(V)$ are V-moment and sample V-moment, respectively. Then requirement of their equality

$$\mu(V) = m(V)$$

will be called *V-moment consistency condition.*

Notation. $\boldsymbol{\lambda}$, $\mathbf{u}(\cdot)$, $\boldsymbol{\mu}(\cdot)$ and $\mathbf{m}(\cdot)$ are $[J, 1]$ vectors, indexed by j. $\mathbf{x}$, $\mathbf{p}$ and $\mathbf{r}$ are $[m, 1]$ vectors, indexed by i, with m finite or infinite. $\boldsymbol{\alpha}$ is $[T, 1]$ vector indexed by t.

Since entropy maximization can be reasonably constrained by constraints other than the moment consistency constraints (see for instance (Golan, Judge and Miller, 1996), (Mittelhammer et al., 2000), (Golan, Judge and Perloff, 1996) or proceedings of MaxEnt conferences), in order to be specific, we will speak about an *ME task*. Also, *ML task* is defined. The complementarity results obtained for the ME task easily extends to the more general constraints used with the Shannon's entropy maximization criterion.

Definition 5. ML task on $f_X(x|\boldsymbol{\theta})$. Let $X_1, X_2, \ldots, X_n$ be a random sample from population $f_X(x|\boldsymbol{\theta})$. The maximum likelihood task on $f_X(x|\boldsymbol{\theta})$ is to find maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$, given the sample.

Definition 6. ME task on $\mathbf{u}(\cdot)$*.* Given a sample and a vector of known potential functions $\mathbf{u}(\cdot)$, the maximum entropy task is to find the most entropic distribution $\mathbf{p}$ consistent with the set of $\mathbf{u}$-moment consistency conditions.

3. ML TASK AND ME TASK

3.1. Simple exponential form, simple potential case

THEOREM 1. Complementarity of ML and ME tasks, identity of solutions

Let $X_1, X_2, \ldots, X_n$ be a random sample. Then,

i) complementarity of tasks

a) ML estimator $\hat{\boldsymbol{\lambda}}$ of $\boldsymbol{\lambda}$ on simple exponential form $f_X(x|\boldsymbol{\lambda}) = k(\boldsymbol{\lambda})e^{-\boldsymbol{\lambda}'\mathbf{u}}$ is obtained as a solution of system of J u_j-moment consistency conditions,

b) the most entropic distribution $\mathbf{p}$ satisfying the system of J u_j-moment consistency conditions is the simple exponential form pmf/pdf $f_X(x|\hat{\boldsymbol{\lambda}})$.

ii) identity of solutions

necessary and sufficient conditions for ML task on simple exponential form pmf/pdf $f_X(x|\boldsymbol{\lambda}) = k(\boldsymbol{\lambda})e^{-\boldsymbol{\lambda}'\mathbf{u}(x)}$ and ME task on the simple potentials $\mathbf{u}(x)$ are identical, and they are

$$\mu(u_j) = m(u_j) \qquad j = 1, 2, \ldots, J$$

Proof.

Discrete r.v. case.

1. ML task.

$$\max_{\boldsymbol{\lambda}} \; l(\boldsymbol{\lambda}) = \ln(k(\boldsymbol{\lambda})) - \sum_{j=1}^{J}\sum_{i=1}^{m} \lambda_j r_i u_j(x_i)$$

leads to system of J first order conditions (FOC)

$$\mu(u_j) = m(u_j) \qquad j = 1, 2, \ldots, J$$

The corresponding hessian matrix of second derivatives of loglikelihood function with respect to (wrt) $\boldsymbol{\lambda}$ is

$$H_{ML} = -\begin{pmatrix} \mathrm{Var}\,(u_1) & \mathrm{Cov}\,(u_1, u_2) & \ldots & \mathrm{Cov}\,(u_1, u_J) \\ \mathrm{Cov}\,(u_2, u_1) & \mathrm{Var}\,(u_2) & \ldots & \mathrm{Cov}\,(u_2, u_J) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}\,(u_J, u_1) & \mathrm{Cov}\,(u_J, u_2) & \ldots & \mathrm{Var}\,(u_J) \end{pmatrix}$$

negative definite, assuring that unique global maximum was attained.

Thus, ML task on simple exponential form of pmf is identical with solving a system of J non-linear equations, the u_j-moment consistency conditions.

2. ME task.

$$\max_{\mathbf{p}} \; H(\mathbf{p}) = -\sum_{i=1}^{m} p_i \ln p_i$$
$$\text{subject to} \tag{1}$$
$$\mu(u_j) = m(u_j) \qquad j = 1, 2, \ldots, J$$

which can be accomplished by means of Lagrangean

$$L(\mathbf{p}) = -\sum_{i=1}^{m} p_i \ln p_i + \sum_{j=1}^{J} \lambda_j (m(u_j) - \mu(u_j))$$

leading to system of m FOC

$$p_i = e^{-\boldsymbol{\lambda}'\mathbf{u}(x_i)} \qquad i = 1, 2, \ldots, m$$

which, after a normalization gives the simple exponential form as the solution.

The corresponding hessian matrix of second derivatives of the Lagrangean wrt $\mathbf{p}$ is

$$H_{ME} = \begin{pmatrix} -1/p_1 & 0 & \ldots & 0 \\ 0 & -1/p_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & -1/p_m \end{pmatrix}$$

for $p_i > 0$ negative definite, satisfying the sufficient conditions for a unique global maximum.

Thus, ME task with the system of J u_j-moment consistency conditions leads to the simple exponential form, where $\boldsymbol{\lambda}$, the Lagrange multipliers, have to be found out of the system of nonlinear equations (1).

Continuous r.v. case.

1. ML task – in analogy with the discrete case proof.

2. ME task.

$$\max_{f_X(x)} \; H(f_X(x)) = -\int f_X(x) \ln(f_X(x))\, dx$$
$$\text{subject to}$$
$$\mu(u_j) = m(u_j) \qquad j = 1, 2, \ldots, J$$

which can be accomplished by means of Lagrangean functional

$$L(f_X(x)) = -f_X(x) \ln(f_X(x)) + \sum_{j=1}^{J} \lambda_j u_j(x) f_X(x)$$

leading to Euler's equation (FOC)

$$\frac{\partial L(\cdot)}{\partial f_X(x)} = 0$$

which, after a normalization gives the simple exponential form

$$f_X(x|\boldsymbol{\lambda}) = \frac{e^{-\boldsymbol{\lambda}'\mathbf{u}(x)}}{\int e^{-\boldsymbol{\lambda}'\mathbf{u}(x)}\,dx}$$

Note. ML task on simple exponential form and ME task on simple potentials are complementary in the sense, that where one starts the other one ends, and vice versa. ML starts with known simple exponential form of pmf/pdf and ends up with ML estimators of the parameters, found out of the potential moment consistency equations. ME, working on the sample, starts with assumed form of potential functions, forming potential moment consistency constraints. The most entropic distribution resolved is just the exponential form pmf/pdf ML has assumed. And the ME estimators of its parameters are the same as the ML estimators. We say that ML task on simple exponential form pmf/pdf and ME task on simple potentials are *complementary.*

ML and ME tasks are complementary in set-up but identical in solution. Both the tasks end up with the same mathematical problem of solving estimators of $\boldsymbol{\lambda}$ out of the system of potential moment consistency equations (1).

Example 2. Let $X_1, X_2, \dots, X_n$ be a random sample of size n from discrete normal distribution $dn(\lambda_1, \lambda_2)$, taken in the simple exponential form.

ML task of estimation leads to solving λ_1, λ_2 out of system of equations

$$\begin{aligned}
\frac{\sum_{i=1}^m x_i e^{-(\lambda_1 x_i + \lambda_2 x_i^2)}}{\sum_{i=1}^m e^{-(\lambda_1 x_i + \lambda_2 x_i^2)}} &= \sum_{i=1}^m r_i x_i \\
\frac{\sum_{i=1}^m x_i^2 e^{-(\lambda_1 x_i + \lambda_2 x_i^2)}}{\sum_{i=1}^m e^{-(\lambda_1 x_i + \lambda_2 x_i^2)}} &= \sum_{i=1}^m r_i x_i^2
\end{aligned}$$

which is just the system of x-moment and x^2-moment consistency conditions.

ME task constrained by system of x-moment, and x^2-moment consistency conditions

$$\sum_{i=1}^{m} p_i x_i = \sum_{i=1}^{m} r_i x_i \qquad \sum_{i=1}^{m} p_i x_i^2 = \sum_{i=1}^{m} r_i x_i^2 \tag{2}$$

finds the most entropic distribution consistent with the constraints to have form (after normalization)

$$p_i = \frac{e^{-(\lambda_1 x_i + \lambda_2 x_i^2)}}{\sum_{i=1}^{m} e^{-(\lambda_1 x_i + \lambda_2 x_i^2)}} \tag{3}$$

where, λ_1, λ_2 should be found out of the system (2), after plugging (3) in. $\Diamond$

In passing we mention an identity of ML and modified method of moments (MMM) in the case of exponential family, discovered by (Huzurbazar, 1949) and explored further by (Davidson and Solomon, 1974). The identity holds also for the simple exponential form, making ME complementary to both ML and MMM. Note that MMM starts with a moment consistency conditions, where understanding of moments is enhanced as done here by Definitions 2, 3, 4.

3.2. General exponential form, general potential case

Complementarity of the general exponential form ML task and general potential ME task can not be assessed analytically in full extent, for sufficient conditions for maximum of likelihood or entropy function do not allow, in general, for it. We show, analytically, that ML task on the general exponential form and ME task on the general potentials lead to the same FOC's. This could be called 'weak complementarity'.

THEOREM 2. *Let $X_1, X_2, \ldots, X_n$ be a random sample. Then, necessary conditions for*

a) ML task on general exponential form pmf/pdf $f_X(x|\boldsymbol{\lambda}, \boldsymbol{\alpha}) = k(\boldsymbol{\lambda}, \boldsymbol{\alpha}) e^{-\boldsymbol{\lambda}' \mathbf{u}(x, \boldsymbol{\alpha})}$

b) ME task on the general potentials $\mathbf{u}(x, \boldsymbol{\alpha})$

are identical, and they are

$$\mu(u_j) = m(u_j) \qquad j = 1, 2, \ldots, J$$

$$\boldsymbol{\lambda}' \boldsymbol{\mu}\left(\frac{\partial \mathbf{u}}{\partial \alpha_t}\right) = \boldsymbol{\lambda}' \mathbf{m}\left(\frac{\partial \mathbf{u}}{\partial \alpha_t}\right) \qquad t = 1, 2, \ldots T$$

Proof.
Discrete r.v. case.
1. ML task.

$$\max_{\boldsymbol{\lambda},\boldsymbol{\alpha}} \ l(\boldsymbol{\lambda},\boldsymbol{\alpha}) = \ln(k(\boldsymbol{\lambda},\boldsymbol{\alpha})) - \sum_{j=1}^{J}\sum_{i=1}^{m} \lambda_j r_i u_j(x_i,\boldsymbol{\alpha})$$

leads to system of $J+T$ first order conditions

$$\begin{aligned} \mu(u_j) &= m(u_j) \qquad j=1,2,\ldots,J \\ \boldsymbol{\lambda}'\boldsymbol{\mu}\left(\frac{\partial \mathbf{u}}{\partial \alpha_t}\right) &= \boldsymbol{\lambda}'\mathbf{m}\left(\frac{\partial \mathbf{u}}{\partial \alpha_t}\right) \qquad t=1,2,\ldots T \end{aligned} \tag{4}$$

2. ME task.

$$\max_{\mathbf{p}(\boldsymbol{\alpha})} \ H(\mathbf{p}(\boldsymbol{\alpha})) = -\sum_{i=1}^{m} p_i \ln p_i$$

subject to

$$\mu(u_j) = m(u_j) \qquad j=1,2,\ldots,J$$

which can be accomplished by means of Lagrangean

$$L(\mathbf{p}(\boldsymbol{\alpha})) = -\sum_{i=1}^{m} p_i \ln p_i + \sum_{j=1}^{J} \lambda_j (m(u_j) - \mu(u_j))$$

leading to system of $m+T$ FOC's

$$\begin{aligned} p_i &= e^{-\boldsymbol{\lambda}'\mathbf{u}(x_i,\boldsymbol{\alpha})} \qquad i=1,2,\ldots,m \\ & -\sum_{i=1}^{m}\left(\frac{\partial p_i}{\partial \alpha_t}\ln p_i + \frac{\partial p_i}{\partial \alpha_t}\right) + \end{aligned} \tag{5}$$

$$+\sum_{j=1}^{J}\lambda_j\left(\frac{\partial m(u_j)}{\partial \alpha_t} - \sum_{i=1}^{m}\left\{p_i\frac{\partial u_j(x_i,\boldsymbol{\alpha})}{\partial \alpha_t} + \frac{\partial p_i}{\partial \alpha_t}u_j(x_i,\boldsymbol{\alpha})\right\}\right) = 0 \qquad \forall t$$

The most entropic distribution after normalization takes general exponential form

$$p_i = \frac{e^{-\boldsymbol{\lambda}'\mathbf{u}(x_i,\boldsymbol{\alpha})}}{\sum_{i=1}^{m} e^{-\boldsymbol{\lambda}'\mathbf{u}(x_i,\boldsymbol{\alpha})}} \qquad i=1,2,\ldots,m$$

where 'ME estimators' of λ have to be found out of the system of (5).

The T of equations of the system (5) simplifies heavily into

$$\boldsymbol{\lambda}'\boldsymbol{\mu}\left(\frac{\partial \mathbf{u}}{\partial \alpha_t}\right) = \boldsymbol{\lambda}'\mathbf{m}\left(\frac{\partial \mathbf{u}}{\partial \alpha_t}\right) \qquad t = 1, 2, \ldots, T$$

which are the same as the T equations of FOC's for ML task (4).

Thus, the ME and ML tasks indeed lead to the same necessary conditions (4).

Continuous r.v. case.

In analogy to the proof of Theorem 1.

COROLLARY 1. *Due to the linearity of* $U(x, \boldsymbol{\lambda}, \boldsymbol{\alpha})$ *in* $\boldsymbol{\lambda}$*, the necessary conditions (4) can be rewritten in a compact form*

$$\mu\left(\frac{\partial U}{\partial \lambda_j}\right) = m\left(\frac{\partial U}{\partial \lambda_j}\right) \qquad j = 1, 2, \ldots, J$$
$$\mu\left(\frac{\partial U}{\partial \alpha_t}\right) = m\left(\frac{\partial U}{\partial \alpha_t}\right) \qquad t = 1, 2, \ldots T$$

Example 3. Let $X_1, X_2, \ldots, X_n$ be a random sample from discrete normal distribution $dn(\lambda, \alpha)$, taken in the general exponential form, so $u(x, \alpha) = (x - \alpha)^2$.

ML task of estimation leads to solving λ, α out of the system of equations

$$\begin{aligned} \mu(u) &= m(u) \\ \mu\left(\frac{\partial u}{\partial \alpha_t}\right) &= m\left(\frac{\partial u}{\partial \alpha_t}\right) \end{aligned} \tag{6}$$

ME task constrained by moment consistency condition

$$\sum_{i=1}^{m} p_i(x_i - \alpha)^2 = \sum_{i=1}^{m} r_i(x_i - \alpha)^2$$

leads to the FOC's

$$\begin{aligned} p_i &= e^{-\lambda(x_i - \alpha)^2} \\ \mu\left(\frac{\partial u}{\partial \alpha_t}\right) &= m\left(\frac{\partial u}{\partial \alpha_t}\right) \end{aligned}$$

where λ, α has to be found out of (6), after normalizing p's.

So, ML and ME tasks lead to the same necessary conditions. Also, note that the ML and ME estimators are the same as in the Example 2, where $dn(\cdot)$ was taken in the simple exponential form. $\Diamond$

Regarding the sufficient conditions, following Theorem states the second derivatives for the both tasks. Whether they are identical can not be in general analytically assessed.

THEOREM 3. *Second derivatives for the ML task are*

$$\frac{\partial^2 l(\boldsymbol{\lambda}, \boldsymbol{\alpha})}{\partial \lambda_j^2} = -\,Var(U'_{\lambda_j})$$

$$\frac{\partial^2 l(\boldsymbol{\lambda}, \boldsymbol{\alpha})}{\partial \lambda_j \partial \lambda_\iota} = -\,Cov(U'_{\lambda_j} U'_{\lambda_\iota})$$

$$\frac{\partial^2 l(\boldsymbol{\lambda}, \boldsymbol{\alpha})}{\partial \alpha^2} = -\left(Var(U'_\alpha) + m(U''_\alpha) - \mu(U''_\alpha)\right)$$

$$\frac{\partial^2 l(\boldsymbol{\lambda}, \boldsymbol{\alpha})}{\partial \alpha_t \partial \alpha_\tau} = -\mu(U'_{\alpha_t})\mu(U'_{\alpha_\tau}) - \sum_{j=1}^{J} \lambda_j \mu(U''_{\lambda_j \alpha_t} U''_{\alpha_t \alpha_\tau}) - m(U''_{\alpha_t \alpha_\tau}) + \mu(U''_{\alpha_t \alpha_\tau})$$

$$\frac{\partial^2 l}{\partial \lambda_j \partial \alpha_t} = -\lambda_j\, Cov(U'_{\lambda_j}, U''_{\lambda_j \alpha_t}) - \sum_{k \neq j}^{J} \lambda_k \mu(U'_{\lambda_j} U''_{\lambda_k \alpha_t}) - m(U''_{\lambda_j \alpha_t}) + \mu(U''_{\lambda_j \alpha_t})$$

and for the ME task they are

$$\frac{\partial^2 L(\mathbf{p}(\boldsymbol{\alpha}))}{\partial p_i^2} = -\frac{1}{p_i}$$

$$\frac{\partial^2 L(\mathbf{p}(\boldsymbol{\alpha}))}{\partial \alpha_t^2} = Var(U'_{\alpha_t}) + m(U''_{\alpha_t}) - \mu(U''_{\alpha_t})$$

$$\frac{\partial^2 L(\mathbf{p}(\boldsymbol{\alpha}))}{\partial \alpha_t \partial \alpha_\tau} = m(U''_{\alpha_t \alpha_\tau}) - \mu(U''_{\alpha_t \alpha_\tau}) + Cov(U'_{\alpha_t}, U'_{\alpha_\tau})$$

Proof. Differentiating twice the loglikelihood function, and the Lagrange function lead to the stated results.

In the following simple instance of the general potential the sufficient conditions are analytically tractable, showing that at the points chosen by the necessary conditions (4) entropy function attains its *maximum* in $\mathbf{p}(\boldsymbol{\alpha})$, and *minimum* in $\boldsymbol{\alpha}$, hence the chosen distribution has minimal entropy in the class of the most entropic distributions, consistent with the moment consistency constraints. Likelihood function at the points attains its maximum.

Example 4. Find the sufficient conditions for the Example 3 set-up.
The general total potential is $U(x, \lambda, \alpha) = \lambda(x - \alpha)^2$, so the potential is $u(x, \alpha) = (x - \alpha)^2$. The second derivatives stated in the above

Theorem then simplifies into

$$\frac{\partial^2 l(\lambda,\alpha)}{\partial\lambda^2} = -\mathrm{Var}\,(u)$$
$$\frac{\partial^2 l(\lambda,\alpha)}{\partial\alpha^2} = -(\lambda^2\mathrm{Var}\,(u'_\alpha) + \lambda(m(u''_\alpha) - \mu(u''_\alpha)))$$
$$\frac{\partial^2 l(\lambda,\alpha)}{\partial\lambda\partial\alpha} = -\left(\lambda\mathrm{Cov}\,(u,u'_\alpha) + m(u'_\alpha) - \mu(u'_\alpha)\right)$$

for the ML task, and into

$$\frac{\partial^2 L(\mathbf{p}(\alpha))}{\partial p_i^2} = -\frac{1}{p_i}$$
$$\frac{\partial^2 L(\mathbf{p}(\alpha))}{\partial\alpha^2} = \lambda^2\mathrm{Var}\,(u'_\alpha) + \lambda(m(u''_\alpha) - \mu(u''_\alpha))$$

for the ME task. Furthermore, in this case

$$m(u''_\alpha) - \mu(u''_\alpha) = 0$$

and also, due to the FOC's (4)

$$m(u'_\alpha) - \mu(u'_\alpha) = 0$$

Thus, the second derivatives for the ML task form a hessian matrix

$$H_{ML} = -\begin{pmatrix} \mathrm{Var}\,(u) & \lambda\mathrm{Cov}\,(u,u') \\ \lambda\mathrm{Cov}\,(u,u') & \lambda^2\mathrm{Var}\,(u') \end{pmatrix}$$

which is negative definite, assuring in this case, that the global maximum was attained.

ME task second derivatives are

$$\frac{\partial^2 L(\mathbf{p}(\alpha))}{\partial p_i^2} = -\frac{1}{p_i}$$
$$\frac{\partial^2 L(\mathbf{p}(\alpha))}{\partial\alpha^2} = 4\lambda^2\mathrm{Var}\,(x)$$

showing that entropy attains its maximum in distribution $\mathbf{p}$, and minimum in α, at the same point where likelihood attains its maximum.

This result was also supported by numerical investigations, elucidating the behavior. In the α suggested by FOC's entropy function attains its *minimum*, whilst the maximum is attained for an $\tilde{\alpha}$ degenerating $\mathbf{p}$ into an uniform distribution. No surprise, since the value of parameter α of $u(x,\alpha)$ is free to choose, and attaining the goal of maximal entropy the value is set up such that the uniform distribution is reached. $\Diamond$

The above analytically tractable case of the sufficient conditions and several numerical investigations of more complex general potentials lead us to propose a *hypothesis* about complementarity of ML and *MiniMax Entropy* tasks and identity of their solutions, under the general exponential form, general potentials.

For the sake of completeness, the MiniMax Ent task is defined.

Definition 7. MiniMax Entropy task. Given a sample and a vector of known general potentials $\mathbf{u}(x, \boldsymbol{\alpha})$, the MiniMax Entropy task is to find in the class of all most entropic distributions $\mathbf{p}(\boldsymbol{\alpha})$ consistent with the set of $\mathbf{u}$-moment consistency conditions, a pmf/pdf with minimal entropy.

Note. If the potentials are simple, MiniMax Ent task reduces into the ME task on simple potentials.

4. CONCLUSIONS

As a way of concluding we sum up the main points of the presented work:

1) In light of the physical analogy mentioned in the Introduction traditional statistical notion of exponential family (see for instance (Brown, 1986), (Barndorff-Nielsen, 1978)) appeared to be too restrictive. An extension to *general exponential form*, driven by the analogy was proposed. Also, simple and general potential were introduced in the vocabulary of statistics.

2) Maximum Entropy task, as a typical instance of MaxEnt method and Maximum Likelihood task were defined in order to make clear the difference in operational mode of the two methods.

3) Concept of complementarity was introduced and defined (see Note 2 at the Section 3.1). Maximum Entropy task on simple potential and Maximum Likelihood task on simple exponential form were proved to be complementary.

4) Exploration of the complementarity of MaxEnt on general potential and ML on general exponential form (Sect. 3.2) led to a generalization of MaxEnt into MiniMax Ent. It was proved that MiniMaxEnt on general potential and ML on general exponential form lead to the same necessary conditions. Whether the conditions are also sufficient can not be in general analytically assessed. Simple instance of general potential (Example 4) as well as several numerical investigations suggests that it is the case and full extent complementarity of MiniMaxEnt on general

(parametric) potential and ML on general exponential form can be claimed.

5) Finally, we would like to note that the complemantary relationship of MiniMaxEnt/MaxEnt task to the ML task seems to be specific property of Shannon's entropy criterion. In (Grendar and Grendar, 2000) it was shown, that so-called maximum empirical likelihood (MEL) criterion constrained by moment consistency constraints, proposed by (Mittelhammer et al., 2000) in the context of noiseless linear inverse problem, is not complementary with ML on the MEL recovered class of pmf/pdf.

5. ACKNOWLEDGEMENTS

It is a pleasure to thank George Judge, Ali Mohammad-Djafari, Alberto Solana and Viktor Witkovský for valuable discussions.

References

Barndorff-Nielsen, O.: *Information and Exponential Families*, John Wiley & Sons, Chichester, 1978

Brown L. D.: *Fundamentals of Statistical Exponential Families*, Lecture Notes - monograph series, Vol. 9, Institute of Mathematical Statistics, Hayward, CA, 1986.

Campbell, L. L.: 'Equivalence of Gauss's principle and minimum discrimination information estimation of probabilities', *Ann. Stat.*, **Vol. 41**, pp. 1011-1015, 1970

Davidson, R. R. and Solomon, D. L.: 'Moment-Type Estimation in the Exponential Family', *Communications in Statistics*, **Vol. 3**, pp. 1101-1108, 1974

Dutta, M.: 'On maximum (information-theoretic) entropy estimation', *Sankhya*, Series A, **Vol. 28**, pp. 319-328, 1966

Golan A.: 'Maximum Entropy, Likelihood and Uncertainity', in *Maximum Entropy and Bayesian Methods*, Erickson, Rychert & Smith, eds., 1998.

Golan A., Judge G., Miller D.: *Maximum Entropy Econometrics. Robust Estimation With Limited Data*. John Wiley & Sons, New York, 1996

Golan A., Judge G. and Perloff J.: 'A Maximum Entropy approach to recovering information from multinomial response data', *JASA*, **Vol. 91**, pp. 841-853, 1996

Grendár M. and Grendár M.: 'Criterion Choice Problem. An ML-complementarity approach', TechRep 1/2000 of IMS SAS, 2000

Huzurbazar, V. S.: 'On a Property of Distributions Admitting Sufficient Statistics', *Biometrika* **Vol. 36**, pp. 71-74, 1949

Jaynes E. T.: 'On The Rationale of Maximum Entropy Methods', *Proc. IEEE*, **Vol. 70**, pp. 939-952, 1982

Kullback S.: *Information Theory and Statistics*, Dover, New York, 1968

Mittelhammer R., Judge G. and Miller D.: *Econometric Foundations*, Cambridge University Press, NY, 2000

Mohammad-Djafari A. and Idier J.: 'Maximum Likelihood Estimation of the Lagrange Parameters of the Maximum Entropy Distributions', in *Maximum Entropy and Bayesian Methods*, C.R. Smith, G.J. Erikson and P.O. Neudorfer, eds., Kluwer Academic Publishers, pp. 131-140, 1991

Mohammad-Djafari A.: 'Probabilistic methods for data fusion', in *Maximum Entropy and Bayesian Methods*, J. Rychert and G. Erickson, eds., Kluwer Academic Publishers, pp. 57-69, 1998

Nishii, R.: 'A characterization of probability densities with expected log likelihood', *Communications in Statistcis - Theory & Methods*, **Vol. 18**, pp. 2657-2662, 1989