Dewar dice

(A probabilistic look at Maximum Entropy Production)

M. Grendár*

Dec 5, 2007

Abstract

Recently Bruers proposed a simple setup, illustrating Dewar's Maximum Entropy Production (MaxEP). The setup is used as a framework for discussing Dewar dice problem – an analogue of the well-known Jaynes dice, – from a probabilistic point view, which rests on Conditional Law of Large Numbers and Maximum Probability/Maximum Entropy asymptotic correspondence. A couple of examples is worked out. It is noted that in Bruers' setup, MaxEP distribution can be obtained without solving constrained optimization problem, utilizing its independence property.

1 Bruers' example of Dewar's MaxEP

Let us begin with a quick, simplified presentation of Bruers' [2] interesting example of Dewar's [8], [9] Maximum Entropy Production (MaxEP) principle/model.

Let there be a system of ι sites. Let time t be discrete. At every t there is a random flux $f_{ij} = \pm 1$ between site i and j. The sign of flux is stochastic. Also, $f_{ij} = -f_{ji}$. For a site i and j, a microscopic path Γ of fluxes is a sequence of values $+1$, -1. The path space is space of all possible paths. For a path Γ, the time average is $\overline{f_{ij,\Gamma}} = \frac{1}{\tau}\sum_t f_{ij}(t)$, where τ is length of the path. For each microscopic path Γ, we assign a probability p_Γ. The path ensemble average $\langle\overline{f_{ij}}\rangle \triangleq \sum_\Gamma p_\Gamma \overline{f_{ij,\Gamma}}$.

Dewar suggests to select such a probability measure p_Γ over the path space which maximizes Shannon's entropy $-\sum_\Gamma p_\Gamma \log p_\Gamma$, subject to the following constraints:

$$\sum_\Gamma p_\Gamma = 1 \tag{1}$$

$$\sum_\Gamma p_\Gamma \overline{f_{ij,\Gamma}} = F_{ij}. \tag{2}$$

The second constraint means (cf. [2]) 'that the time and path ensemble average of the flux from i to j is measured to be the numerical value F_{ij}'.

Once the optimization problem is set up one can go on, find its solution and develop further results/theory; cf. [8], [9], [2].

*Department of Mathematics, FPV UMB, Tajovského 40, 974 01 Banská Bystrica, Slovakia; Institute of Mathematics and CS of the Slovak Academy of Sciences (SAS), Institute of Measurement Science of SAS. Email: marian.grendar@savba.sk.

2 Jaynes dice, Dewar dice

Dewar's presentation of his MaxEP [8], [9] stresses that MaxEP was proposed in analogy with Jaynes' Maximum Entropy (MaxEnt). There is a toy example of what is MaxEnt about, due to Jaynes – the well-known Jaynes dice [17]. Jaynes dice problem is a restatement of the problem which Boltzmann faced in 1870's. In the Jaynes dice setup, there are two 'players', with different information[1]. The less informed player has to select a frequency distribution from a set of such a distributions, when some additional information is available.

First we describe the Jaynes dice problem and then use Bruers' setup to discuss "Dewar dice".

2.1 Jaynes dice

Let X be a random variable taking on values from a set $\mathcal{X} = \{x_1, \ldots, x_m\}$, with probabilities $q = (q_1, \ldots, q_m)$. Let X_1^n be a random sample drawn identically and independently from q. Given $X_1^n = x_1^n$, player A calculates sample average $\frac{1}{n}\sum_{i=1}^n x_i$ and finds it to be a. The sample is not given to the other player B. All the information available to player B can be summarized by a quadruple $\{\mathcal{X}, q, n, \frac{1}{n}\sum_{i=1}^n x_i = a\}$. Given this information (and nothing else), the player B is asked to make his best guess of frequency distribution of the m possible outcomes, which the sample x_1^n induced.

Conditional Law of Large Numbers (CLLN) ([23], [22], [4], [5]) implies that at least for n large, player B should select such a distribution from the feasible set (specified here by the mean value constraint) which maximizes (with respect to p) the relative entropy $H(p\,||\,q) \triangleq -\sum_{i=1}^m p_i \log \frac{p_i}{q_i}$. This is the probabilistic (large deviations, cf. [11], [12], [7]) *justification* of Relative Entropy Maximization (REM/MaxEnt) method[2]. REM/MaxEnt can be viewed as an asymptotic instance of Maximum Probability (MaxProb) method ([1], [24], [14], [15]; for a broad view of entropy and MaxProb see [19], [20]). This is the probabilistic *interpretation* of REM/MaxEnt.

2.2 Dewar dice

2.2.1 Univariate case

First, a probabilistic formulation of Dewar dice problem for the simplest case of Bruers' model: $\iota = 2$.

Let $\gamma = \{-1, 1\}$, and $\mathit{\Gamma}$ be the cartesian product of τ of γ's, i.e., set of all binary sequences of length τ. For each element Γ of $\mathit{\Gamma}$, called *path*, let $\overline{\Gamma} \triangleq \frac{1}{\tau}\sum_{i=1}^{\tau} \Gamma_i$ denote average of the τ values which constitute Γ; i.e., the path average. Let $\mathit{\Gamma}$ be support of a random variable S with probability mass function (pmf) $q = (q_1, \ldots, q_{2^\tau})$. Let S_1^n be a random sample drawn identically and independently (*iid*) from pmf q. S_1^n represents n tosses of 2^τ-sided Dewar dice. With sample S_1^n, a sequence of averages of paths $\overline{\Gamma}_1, \ldots, \overline{\Gamma}_n$, is associated. Player A calculates average $\frac{1}{n}\sum_{i=1}^n \overline{\Gamma}_i$ of the n observed path averages, to be F.

[1] For a formal game theoretic view of MaxEnt see [16].

[2] CLLN justification of REM/MaxEnt, well-known in Shannon theory community as well as among some statistical physicists, appears to be unknown in MaxEP community. For this reason a short appendix is attached to this note, discussing CLLN.

Player B does not know the observed sequence $S_1^n = s_1^n$. Rather, player A gave him only the value F of the average of observed path averages. Player B has in possession only the following information: $\{\mathit{\Gamma}, q, n, \frac{1}{n}\sum_{i=1}^{n}\overline{\Gamma}_i = F\}$. Given this information, player B should make his best guess of frequency distribution $\nu^n = (n_1, \dots, n_\tau)/n$ of the 2^τ possible paths, which the sample s_1^n induced.

Following Boltzmann Vincze method of Maximum Probability (MaxProb), player B might want to select the type(s) which can be generated by q with the highest probability, i.e., $\hat{\nu}^n = \arg\sup_{\nu^n:\sum_{i=1}^{\tau}\nu_i^n\overline{\Gamma}_i=a} \pi(\nu^n; q)$, where $\pi(\nu^n; q) = n!\prod_{i=1}^{\tau}\frac{q_i^{n_i}}{n_i!}$. This selection scheme is justified by Conditional Law of Large Numbers; cf. Appendix. Furthermore, for $n \to \infty$, MaxProb type(s) converges to pmf(s) which maximize relative entropy $\hat{p} = \arg\sup_{p:\sum_{i=1}^{\tau}p_i\overline{\Gamma}_i=a} H(p\,||\,q)$. Thus, for large n, MaxProb and Relative Entropy Maximization method 'coincide'.

Example 1. Let $\tau = 3$ to make things simple yet sufficiently rich. Then, $\mathit{\Gamma}$ consists of eight paths: $\{(1,1,1), (1,1,-1), (1,-1,1), (-1,1,1), (-1,-1,1), (-1,1,-1), (1,-1,-1), (-1,-1,-1)\}$. With each path, its path average is associated; thus there is the twin set $\left\{1, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}, -1\right\}$. Let q be uniform, $q_1 = \dots = q_8 = 1/8$. Let $n = 10^{23}$, and the value of average of observed sample of path averages let be $F = 1/3$. There is a large number of frequency distributions of the eight paths, compatible with the observed average value F of the sample of path averages. Player B should utilize also his knowledge of q, to pick up one (or more) of the frequency distributions as his best guess of the observed frequency distribution, induced by the observed sample s_1^n. Since $n = 10^{23}$, player B can without much of harm ignore discrete nature of frequency distribution, and instead try to select a real-valued probability mass function $p(\Gamma)$ from the set $\Pi = \{p(\Gamma) : \sum_\Gamma p(\Gamma)\overline{\Gamma} = F, \sum_\Gamma p(\Gamma) = 1\}$.

As the mean $\phi = E_q\overline{f_\Gamma}$ of path averages is 0, Law of Large Numbers implies that a rare sample s_1^n occurred. Another Law of Large Numbers, the Conditional one, says that if such a rare event happened, then for such a large n, it could have happened essentially only in such a way, that the observed frequency distribution is virtually indistinguishable from the probability mass function $\hat{p}(\Gamma)$ which among pmf's from the feasible set Π maximizes the relative entropy $H(p\,||\,q) \triangleq -\sum_\Gamma p(\Gamma)\log\frac{p(\Gamma)}{q(\Gamma)}$. Asymptotic correspondence between Maximum Probability and Relative Entropy Maximization (REM/MaxEnt) solutions permits to claim that $\hat{p}(\Gamma)$ selected by REM/MaxEnt is essentially the frequency distribution of the eight paths, which – among all frequency distributions which are based on samples of size $n = 10^{23}$ and has average value of path averages equal 1/3, – can be drawn from q (uniform, in this case) with highest probability.

REM/MaxEnt distribution is $\hat{p}(\Gamma) = (0.296, 0.148, 0.148, 0.148, 0.074, 0.074, 0.074, 0.037)$; this is the Maximum (Relative) Entropy Production (Max(R)EP) solution. Mode of the distribution $\hat{p}(\Gamma)$ is at the path $\Gamma_1 = (1,1,1)$. Using Bruers' terminology, (1,1,1) is thus the most probable path of fluxes (from site 1 to site 2, say). If order of values within a path does not matter, then the second, third and fourth paths are indistinguishable and form a single super-path (frequency distribution), which has probability $3 \cdot 0.148 = 0.444$, higher than is the probability of the first path. If order does not matter, then it would be possible to begin considerations with elementary outcomes (i.e. $\mathit{\Gamma}$) formed by the frequency distributions of fluxes.

Finally, let us note that for a uniform q, REM/MaxEnt/MaxEP distribution for Dewar dice can be obtained directly, without any need for solving the constrained optimization problem. This comes from an independence property of Shannon entropy (cf. [3], Thm. 2.6.6. (Independence bound on entropy)) and the fact that γ is two-element set. Let $P = (P_-, P_+)$ be probability distribution on $\gamma = \{-1, 1\}$. Assume that $p(\Gamma)$ can be factorized; in other words that elements which constitute a path Γ were drawn identically and independently from P. Thus, for instance $p(\Gamma = (1, -1, 1)) = (P_+)^2 P_-$. Then, the constraint $\sum_\Gamma p(\Gamma)\overline{\Gamma} = F$ uniquely determines P_+, which in turn leads to a probability distribution $p(\Gamma)$. The distribution $p(\Gamma)$ obtained in such a way is the same as that selected by REM/MaxEnt. For the above Example, the constraint gives that $P_+ = \frac{F+1}{2}$. As $F = 1/3$, $P_+ = 2/3$. Thus, for instance $p(\Gamma = (1, -1, 1)) = 0.148$, which is just the value $\hat{p}((1, -1, 1))$ obtained by REM/MaxEnt.

2.2.2 Bivariate case

Now, assume Bruers' model with $\iota = 3$; i.e., there is a flux from say site 1 to site 2, and from 2 to 3. The following is a probabilistic model of this setting.

Let $\gamma_0 = \{-1, 1\}$, and let $\gamma = \gamma_0 \otimes \gamma_0$, be the cartesian product of γ_0 and γ_0; this is the basic object, now.

$$\gamma = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right\}.$$

Linking it to Bruers' model, an element of γ specifies what is the flux from 1 to 2 (the upper number in the pair), and what is the flux from 2 to 3 (the lower number in the pair). $\mathit{\Gamma}$ is the set of all possible paths of length τ, i.e., the cartesian product of τ of γ's. Note that each path Γ comprises two sequences of length τ. Let $\overline{\Gamma} = \frac{1}{\tau}\sum_{i=1}^{\tau}\Gamma$, be the path average; it is a two-element vector. $\mathit{\Gamma}$ is support of a random variable S, with pmf q. And there is a random sample S_1^n of size n drawn *iid* from q. Based on an observed sample $S_1^n = s_1^n$, player A calculates $\frac{1}{n}\sum_{i=1}^{n}\overline{\Gamma}_i = F$; F is vector of two values, the first is sample average of path averages of fluxes from site 1 to site 2, the second value is sample average of path averages of fluxes from site 2 to site 3. Player's B information is $\{\mathit{\Gamma}, q, n, \frac{1}{n}\sum_{i=1}^{\tau}\overline{\Gamma}_i = F\}$. The objective is to make the best guess of frequency distribution of 2^τ paths, which the sample s_1^n induced.

Example 2. Let $\tau = 2$. Then $\mathit{\Gamma}$ consists of 16 paths:

$$\left\{ \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad \ldots \quad , \begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix} \right\}.$$

In a path Γ, the upper pair of values represents sequence (of length $\tau = 2$) of fluxes from site 1 to site 2, the lower pair of values are fluxes from site 2 to site 3. With each path a pair of path averages is associated; thus there is the twin set of path averages:

$$\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \ldots \quad , \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right\}.$$

Let the distribution q over $\mathit{\Gamma}$ be uniform, i.e., $q_1 = \cdots = q_{16} = 1/16$. Let there be a random sample of size $n = 10^{23}$ drawn from q. The observed average of path averages let be $F = (0.2, 0.7)'$. Player B can freely ignore discrete nature of frequency distribution, and consider a problem of selecting pmf $p(\Gamma)$ from $\Pi \triangleq \{p(\Gamma) : \sum_\Gamma p(\Gamma)\overline{\Gamma} = F, \sum_\Gamma p(\Gamma) = 1\}$. Conditional Law of Large Numbers dictates to select such a $\hat{p}(\Gamma)$ that in Π maximizes relative entropy. The entropy maximizing distribution is in Table 1. There, a path Γ is coded by a string of four numbers ± 1; the first couple is the sequence of fluxes from 1 to 2, the second pair is sequence of fluxes from 2 to 3.

Table 1: Example 2. MaxEnt/MaxEP distribution $\hat{p}(\Gamma)$.

$\hat{p}(\Gamma)$	Γ
0.2601	(1 1 1 1)
0.0459	(1 1 1 -1)
0.0459	(1 1 -1 1)
0.0081	(1 1 -1 -1)
0.1734	(1 -1 1 1)
0.0306	(1 -1 1 -1)
0.0306	(1 -1 -1 1)
0.0054	(1 -1 -1 -1)
0.1734	(-1 1 1 1)
0.0306	(-1 1 1 -1)
0.0306	(-1 1 -1 1)
0.0054	(-1 1 -1 -1)
0.1156	(-1 -1 1 1)
0.0204	(-1 -1 1 -1)
0.0204	(-1 -1 -1 1)
0.0036	(-1 -1 -1 -1)

When q is uniform, it is again possible to find MaxEnt/MaxEP distribution without actually solving the optimization problem. Let $P = (P_{(+1,+1)'}, P_{(+1,-1)'}, P_{(-1,+1)'}, P_{(-1,-1)'})$ be distribution over γ. Assume that the bivariate distribution can be factored into product of independent marginals $P^1 = (P^1_+, P^1_-)$ and $P^2 = (P^2_+, P^2_-)$; then for instance $P_{(-1,+1)'} = P^1_- P^2_+$. Take the following two constraints:

$$\sum_{\Gamma_1} P^1(\Gamma_1)\overline{\Gamma_1} = F_1, \tag{3}$$

$$\sum_{\Gamma_2} P^2(\Gamma_2)\overline{\Gamma_2} = F_2, \tag{4}$$

where the index indicates the respective marginal. Thus, Γ_1 is, in the physics terminology, a path of fluxes from site 1 to site 2; Γ_2 is a path of fluxes from site 2 to site 3. Assume that elements of a marginal path Γ_j are drawn *iid* from P^j, $j = 1, 2$. Then the system of constraints (3), (4) determines P^1, P^2 uniquely. For Example 2, it gives $P^1_+ = (F_1 + 1)/2 = 0.6$, and $P^2_+ = (F_2 + 1)/2 = 0.85$. Thus, for instance $P_{(+1,+1)'} = P^1_+ P^2_+ = 0.6 \cdot 0.85 = 0.51$. Hence, $P_{(+1,+1)'} \cdot P_{(+1,+1)'} = 0.51^2 = 0.2601$, which is just the value of MaxEnt/MaxEP $\hat{p}(\cdot)$ for the path $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. The other values of $\hat{p}(\cdot)$ can be obtained in the same way.

The above observation can alternatively be explicated as follows: the bivariate MaxEnt/MaxEP solution $\hat{p}(\Gamma)$ is such as if the sequence Γ_1 of length

τ of fluxes from site 1 to site 2 was drawn from uniform q_1, independently of sequence of fluxes Γ_2 of length τ, drawn from uniform q_2. Hence, the bivariate MaxEP solution $\hat{p}(\Gamma)$ is cartesian product of univariate solutions of two isolated MaxEP problems, with feasible sets specified by (3), (4). And we have already seen (cf. Sect 2.2.1) how a univariate MaxEP solution can be obtained directly, algebraically. For instance, $\hat{p}(\Gamma_1 = (1,1)) = 0.36$, by the univariate MaxEnt/MaxEP (for $F_1 = 0.2$); $\hat{p}(\Gamma_2 = (1,1)) = 0.7225$, by the other univariate MaxEnt/MaxEP (for $F_2 = 0.7$). Thus, the bivariate MaxEnt/MaxEP solution $\hat{p}(\Gamma = ((1,1)', (1,1)')) = 0.36 \cdot 0.7225 = 0.2601$, which is indeed the value obtained by the bivariate MaxEP.

3 Value added

Bruers' [2] example of Dewar's [8], [9] Maximum Entropy Production (MaxEP) was viewed from a probabilistic point of view, which rests on Conditional Law of Large Numbers and Maximum Probability/Maximum Entropy asymptotic correspondence. The example served as a source of narrative about a fictive game (Dewar dice problem) between two players, analogous to that which is played in Jaynes dice problem. A univariate and bivariate cases of Dewar dice problem were studied in detail, and a couple of simple illustrative examples was worked out. It was observed that MaxEP distribution has an independence property, which in Bruers' setting permits to find the distribution algebraically, without solving optimization problem.

4 Appendix: Conditional Law of Large Numbers, Relative Entropy Maximization method

It is well-known in Shannon theory community ([22], [5], [3]) as well as among some statistical physicists ([11], [12]) that Relative Entropy Maximization (REM) method possesses a probabilistic justification, via Conditional Law of Large Numbers (CLLN) ([23], [22], [5], [7], [11], [12]). Shannon Entropy Maximization is a special, uniform q, case of REM. The following is a brief discussion of CLLN, based on [6], [3].

Let X be a discrete random variable, which can take on values from a finite set (a.k.a. support or alphabet) $\mathcal{X} \triangleq \{x_1, \ldots, x_m\}$, with probabilities $q \triangleq (q_1, \ldots, q_m)$. Let $X_1^n \triangleq (X_1, \ldots, X_n)$ be a random sample of size n and drawn from q. Let $\mathcal{P}(\mathcal{X})$ be set of all probability mass functions (pmf's) with support $\mathcal{X}$. Let $\mathcal{P}_n(\mathcal{X})$ be set of all frequency distributions (with support $\mathcal{X}$) which could be induced by sample of size n. Let $\Pi \subseteq \mathcal{P}(\mathcal{X})$. In Jaynes dice (cf. Sect. 2.1), Π is set of all such pmf's (on $\mathcal{X}$) which have mean value a. Let $\Pi_n \triangleq \Pi \cap \mathcal{P}_n(\mathcal{X})$. Thus, in Jaynes dice, Π_n is the set of frequency distributions (with support $\mathcal{X}$), which are induced by samples of size n, and their sample average is a. Recall the problem player B faces (cf. Sect 2.1): given $\{\mathcal{X}, q, n, \Pi\}$, the objective is to select a frequency distribution ν^n (a.k.a. type) from Π. Conditional Law of Large Numbers (CLLN) implies that for $n \to \infty$, it is the pmf $\hat{p} \triangleq \arg\sup_{p \in \Pi} H(p \,||\, q)$, which should be selected.

CLLN: Let Π be convex, closed set; q is not in Π. Let $\epsilon > 0$. Then,

$$\lim_{n\to\infty} \pi(|\nu_i^n - \hat{p}_i| < \epsilon, i = 1, 2, \ldots, m \,|\, \nu^n \in \Pi) = 1.$$

CLLN shows, informally stated, that for large n, if a frequency distribution ν^n from Π occurred, then with probability approaching one, it should be such a frequency distribution which is close to the relative entropy maximizing pmf $\hat{p}$.

Dewar [10], following Jaynes [18], uses Asymptotic Equipartition Property (AEP) to provide justification of the Shannon entropy maximization method, as well as to its MaxEP extension. AEP is another probabilistic result, where entropy pops up. However, there it is Shannon entropy $H(p) = -\sum_{i=1}^{m} p_i \log p_i$, rather than the relative entropy. AEP ([21], [3]) is Shannon's restatement of Law of Large Numbers.

AEP: If X_1^n are identically and independently drawn from q, then

$$-\frac{1}{n}\log q(X_1^n) \to H(q), \quad \text{in probability.}$$

AEP shows [3], that asymptotically, in a sequence X_1^n there should be approximately nq_i outcomes x_i, $i = 1, \ldots, m$, and that all such sequences have roughly the same probability $e^{-nH(q)}$. AEP holds also for ergodic stochastic processes, where it is known as Shannon-McMillan-Breiman Theorem. Recently (cf. [13]), AEP was established also for semi-Markov processes.

Unlike to CLLN, AEP is irrelevant for the problem of selection a frequency distribution, when $\{\mathcal{X}, q, n, \Pi\}$ is available. It should be noted that Jaynes, through his Entropy Concentration Theorem (ECT), made a step in the direction of CLLN justification of MaxEnt. ECT is, however, a weaker result than CLLN.

5 Acknowledgement

I learned about MaxEP from Robert K. Niven. Stimulating discussions on MaxEP with Filip J. R. Meysman during *Facets of Entropy* workshop are gratefully acknowledged. Financial support of participation at the workshop from Niels Bohr Institute (Copenhagen), is also gratefully acknowledged. Formally supported by VEGA 1/3016/06 grant.

References

[1] Boltzmann, L. (1877) "Über die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheilichkeitsrechnung respektive den Sätzen über das Wärmegleichgewicht," *Wiener Berichte*, 2, pp. 373-435.

[2] Bruers, S. (2007) "A discussion on maximum entropy production and information theory", arxiv:0705.3226, cond-mat, stat-mech.

[3] Cover, T. and Thomas, J. (1991) *Elements of Information Theory*, Wiley.

[4] Csiszár, I. (1984) Sanov property, generalized I-projection and a conditional limit theorem, *Ann. Probab.*, 12, 768-793.

[5] Csiszár, I. (1996) "MaxEnt, mathematics and information theory," *Maximum Entropy and Bayesian Methods*, K. M. Hanson and R. N. Silver (eds.), pp. 35-50, Kluwer.

[6] Csiszár, I. (1998) The method of types, *IEEE Trans. Inform. Theory*, vol. IT-44, pp. 2505-2523.

[7] Dembo, A. and Zeitouni, O. (1998) *Large Deviations Techniques and Applications*, Springer, NY.

[8] Dewar, R. (2003) Information theory explanation of the fluctuation theorem, maximum entropy production and self-organized criticality in non-equilibrium stationary states *J. Phys. A*, 36, 631-641.

[9] Dewar, R. (2005) Maximum entropy production and the fluctuation theorem *J. Phys. A*, 38, L371-L381.

[10] Dewar, R. (2005) Maximum Entropy Production and Non-equilibrium Statistical Mechanics in *Non-equilibrium Thermodynamics and the Production of Entropy*, A. Kleidon, R. D. Lorenz (eds.), pp. 41-55, Springer.

[11] Ellis, R. S. (2005) *Entropy, Large Deviations and Statistical Mechanics*, Springer, NY. 2-nd ed.

[12] Ellis, R. S. (1999) The theory of large devitions: from Boltzmann's 1877 calcualtion to equilibrium macrostates in 2D turbulence, *Physica D*, 106-136.

[13] Girardin, V. and Limnios, N. (2006) Entropy for semi-Markov processes with Borel state spaces: asymptotic equirepartition properties and invariance principles, *Bernoulli*, 12(3), 515-533.

[14] Grendár, M., Jr. and Grendár, M. (2001) "What is the question that MaxEnt answers? A probabilistic interpretation", in *Bayesian inference and Maximum Entropy methods in science and engineering*, A. Mohammad-Djafari (ed.), pp. 83-94, AIP, Melville. Also at arXiv:math-ph/0009020.

[15] Grendár, M. (2004) *Acta U. Belii Ser. Math.*, 11, 3-7.

[16] Harremoës, P. and Topsøe, F. (2001) Maximum Entropy fundamentals, *Entropy*, 3, 191-226.

[17] Jaynes, E. T. (1979) *Papers on Probability, Statistics and Statistical Physics*, R. D. Rosenkratz (ed.), D. Reidel, Dordrecht.

[18] Jaynes, E. T. (1979) "Where do we stand on maximum entropy?", in *The maximum entropy principle*, R. D. Levine and M. Tribus (eds.), pp. 15-118, MIT, Cambridge MA. Also, visit bayes.wustl.edu, a web-page maintained by L. Bretthorst, for a collection of Jaynes' papers.

[19] Niven, R. K. (2005) "Combinatorial Information Theory: I. philosophical basis of cross-entropy and entropy," arXiv:cond-mat/0512017.

[20] Niven, R. K. (2007) "Origins of the combinatorial basis of entropy," in *Bayesian inference and Maximum Entropy methods in science and engineering*, K. H. Knuth et al. (eds.), pp. 133-142, AIP, Melville.

[21] Shannon, C. E. (1948) Bell Sys. Tech. Jorunal, 27, 479-423, 623-658.

[22] van Campenhout, J. M. and Cover, T. M. (1981) Maximum entropy and conditional probability, IEEE IT, 27, 483-489.

[23] Vasicek, O. A. (1980) A conditional law of large numbers, *Ann. Probab.*, 8, 142-147.

[24] Vincze, I. (1972) "On the maximum probability principle in statistical physics," *Coll. Math. Soc. J. Bolyai*, **9**, European Meeting of Statisticians, Budapest, 869-893.

[25] Vincze, I. (1997) Indistinguishability of particles or independence of the random variables?, *J. Math. Sci.*, 84, 1190-1196.