



Entropy-based information recovery in complex behavioural systems

George Judge & Wendy K. Tam

To cite this article: George Judge & Wendy K. Tam (06 Jan 2026): Entropy-based information recovery in complex behavioural systems, Applied Economics Letters, DOI: [10.1080/13504851.2026.2613069](https://doi.org/10.1080/13504851.2026.2613069)

To link to this article: <https://doi.org/10.1080/13504851.2026.2613069>



© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 06 Jan 2026.



Submit your article to this journal [↗](#)



Article views: 546



View related articles [↗](#)



View Crossmark data [↗](#)

Entropy-based information recovery in complex behavioural systems

George Judge^a and Wendy K. Tam^{b,c}

^aGraduate School and Giannini Foundation, University of California Berkeley, Berkeley, CA; ^bDepartments of Computer Science and Political Science and Law School, Vanderbilt University, Nashville, TN, USA; ^cDepartment of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA

ABSTRACT

Reigning reductionist econometric paradigms greatly simplify complex behavioural processes. Using concepts from entropy methods and nonequilibrium systems dynamics, we develop a theoretical framework for self-organizing and optimizing macro-level behaviour that is based on nonlinear micro-level behaviour.

KEYWORDS

Causal entropy maximization; information theoretic dynamics; minimum power divergence; statistical equilibrium

JEL CLASSIFICATION

C10; C20

We present entropy methods and nonequilibrium dynamics for recovering information in complex behavioural systems. Leveraging Boltzmann's equation and information-theoretic entropy (Golan, Judge, and Miller 1996) with nonlinear probabilistic reasoning, we model emergent collective behaviour. Unlike moment-based estimation, KL-divergence bounds large-deviation rate functions and controls moments of all orders. Divergence functions over discrete distributions on a common space then yield a rigorous macro-economic index of behaviour.

Most behavioural systems follow reductionist models that express the behaviour of the whole from that of its parts. Their dynamic stochastic equilibrium counterparts appear in functional form with empirical sample moments-constraints such as

$$h(Y, X, Z; \beta) = n^{-1} [Z'(Y - X\beta)] \xrightarrow{p} 0, \quad (1)$$

where Y , X , and Z are respectively matrices of dependent variables, explanatory variables, and instruments, and β is an unknown and unobservable vector. The moment conditions converge in probability to 0 as the sample size grows. A solution may be based on parametric and semi-parametric likelihood functions and a range of

observed data sampling processes that involve a finite set of moments. The moments are usually based on indirect noisy observations in a sample, while our interest is in the underlying population. Higher-order moments are assumed to be well-behaved. It is difficult to obtain a meaningful probabilistic basis for a realistic dynamic behavioural system from this reductionist mix of model, data, and methods. We propose an alternative information theoretic framework based on nonlinear stochastic macro-level dynamics and stochastic individual micro-level behaviour.

Jaynes (1978) developed the Maximum Entropy Principle, which, as we will show, leads to a powerful statistical approach to modelling complex systems that addresses many of the deficiencies of the method of moments summarized in equation (1). By using entropy as a means of statistical inference on multinomial processes, he transformed the objective to identifying the most likely observable microstates of a system, where a microstate is a particular configuration of the system's components. Georgescu-Roegen (1976) noted that a theoretical model of economic behaviour is akin to random processes that obey the laws of stochastic processes and entropy. Wissner-Gross and Freer (2013) connected adaptive

intelligent behaviour and causal entropy maximization for self-organization/emergence and optimization in multi-agent systems. This idea that a system adapts behaviour according to an optimizing principle leads to self-organized equilibrium seeking behaviour and establishes an entropy driven inferential framework (Judge 2016). In this context, behavioural systems are equilibrium stationary state seeking, and the entropic force quantifies the intuitive notion that systems are more likely to move from low to high probability states.

System predictability in terms of entropy function(s) is then equivalent to the study of its nonequilibrium nature. For example, if we characterize a macro state by its probability distribution, entropy reflects the combinatorial arrangements of possible microstates leading to a macro state. This view permits us to recognize that data arise from systems with dynamic non-deterministic adaptive behaviour, involve uncertain quantifiable information, and are driven towards a stationary state associated with a functional structure. The Wissner-Gross and Freer (2013) economic behaviour entropy connection allows us to use entropy as the optimizing criterion. A uniform distribution of the microstates corresponds to a macro state with maximum entropy and minimum system information. Accordingly, causal entropy maximization becomes an optimizing behavioural system with a large number of agents interacting in finite time.

II. Information theoretic minimum divergence family

Behavioural systems have a natural connection with nonlinear stochastic processes. To link the model to a family of likelihood functions, we use the Cressie-Read (CR) family of entropic power divergence measures given by

$$I(\mathbf{p}, \mathbf{q}, \gamma) = \frac{1}{\gamma(\gamma + 1)} \sum_{i=1}^n p_i \left[\left(\frac{p_i}{q_i} \right)^\gamma - 1 \right], \quad (2)$$

where γ indexes members of the CR family, p_i represents the subject data probability distribution, and q_i are reference non-sample prior probabilities. We assume $p_i, q_i \in [0, 1]$ i , $\sum_{i=1}^n p_i = 1$, and

$\sum_{i=1}^n q_i = 1$. As γ varies, the resulting CR family of estimators that minimize power divergence exhibit qualitatively different sampling behaviour (Judge and Mittelhammer 2012). The CR family of power divergences is equivalent to the Pearson's chi-squared statistic, X^2 , when $\gamma = 1$ (Pearson 1904), the log-likelihood ratio, G^2 , when $\gamma = 0$ (Wilks 1938), the Freeman-Tukey statistic, T^2 , when $\gamma = -1/2$ (Freeman and Tukey 1950), the modified chi-squared statistic, N^2 , when $\gamma = -2$ (Neyman 1949), and the modified log-likelihood ratio statistic, M^2 , when $\gamma = -1$ (Kullback 1959).

The CR measure also exhibits formal convexity in $\mathbf{p}$, for all values of γ and $\mathbf{q}$, and embodies the required probability system characteristics, such as invariance with respect to monotonic transformation of the divergence measures. As γ varies, the optimized value of $I(\mathbf{p}, \mathbf{q}, \gamma)$ represents a range of data sampling processes and likelihood functions. The estimators that minimize power divergence exhibit qualitatively different sampling behaviour. As $\gamma \rightarrow 0$, the solution of the first-order condition leads to maximum entropy and the logistic expression for the conditional probabilities. The Maximum Exponential Empirical Likelihood (MEEL) estimator is the most likely combinatorial distribution.

III. Entropy based probability density functions

We describe how to recover the micro income probability density function and the macro income inequality measure from a sample of discretized income data.¹ The microstates are a multiplicity of the macro state. We begin by representing the income probability distribution by separating the data into K bins that span the sample space, forming a histogram. Samples from these partitions yield the discrete random variable, d_j , for $j = 1, 2, \dots, n$. Under repeated observation, one of n histograms of the micro configurations associated with the macro state income is observed with probability p_j . After a large number of trials, we recover the first-moment mean value, $\sum_{j=1}^K d_j p_j = \bar{d}$. Here, as $\gamma \rightarrow 0$, $I(\mathbf{p}, \mathbf{q}, \gamma)$ converges

¹Elsewhere, we provide an empirical example with country-based micro income data from Eurostat where we recover the income distribution of European countries using entropy methods on household income data (Villas-Boas et al. 2018).

to an estimation criterion equivalent to the MEEL metric, $H(\mathbf{p}) = -\sum_{j=1}^K p_j \ln(p_j)$. The entropy function may then be formulated as

$$\max_{\mathbf{p}} \left[-\sum_{j=1}^K p_j \ln p_j \mid \sum_{j=1}^K p_j d_j = \bar{d}, \sum_{j=1}^K p_j = 1, p_j > 0, \forall j \right], \quad (3)$$

with corresponding Lagrange function

$$\begin{aligned} L(\mathbf{p}, \eta, \lambda) = & -\sum_{j=1}^K p_j \ln(p_j) + \lambda \left(\bar{d} - \sum_{j=1}^K p_j d_j \right) \\ & + \eta \left(1 - \sum_{j=1}^K p_j \right). \end{aligned} \quad (4)$$

Solving the first-order conditions yields the exponential result

$$\hat{p}_j = \frac{\exp(-d_j \hat{\lambda})}{\sum_{j=1}^K \exp(-d_j \hat{\lambda})} \quad (5)$$

for the j th outcome and the mean-related income distribution. Using the CR($\gamma \rightarrow 0$) entropy functional, the joint histograms and the mean income value, the resulting probability density income distribution, and the entropy measure, $\sum_{j=1}^K p_j \ln(p_j)$, provides a measure of income inequality. One could also calculate the Theil index, which assesses inequality as the difference between our maximum entropy measure and the observed entropy.

IV. Reference distribution

The income pdf was obtained using a uniform reference distribution, $\mathbf{q}$. Additional information from a non-uniform reference distribution may be incorporated with the minimum cross-entropy Kullback-Liebler (KL) directed divergence (Good 1963; Kullback 1959; Kullback and Leibler 1951). Both Caticha and Golan (2014) and Scharfenaker and Foley (2017) provide an example of how to use Lagrange Multipliers for solving the constrained optimization problem of maximizing entropy. Caticha and Golan (2014) further discuss using cross entropy in their model of economies. Here, one would choose the estimate of $\mathbf{p}$ that can be discriminated from $\mathbf{q}$ with minimum difference.

The KL estimation objective function is a particular case of the more general CR family of entropy measures. Then, instead of a uniform distribution, $\mathbf{q}$, we use the distribution, q_{ML} , as the reference distribution in (2). Replacing the uniform distribution with q_{ML} , and $\gamma \rightarrow -1$ gives us the KL empirical likelihood criterion

$$\begin{aligned} \lim_{\gamma \rightarrow -1} I(\mathbf{p}, \mathbf{q}_{ML}, \gamma) &= \sum_{j=1}^n q_{jML} \ln(p_j/q_{jML}) \\ &= \sum_{j=1}^n q_{jML} \ln(p_j) \\ &\quad - \sum_{j=1}^n q_{jML} \ln(q_{jML}), \end{aligned} \quad (6)$$

where $\sum_{j=1}^n q_{jML} \ln(q_{jML})$, is a constant. Using this criterion yields

$$\begin{aligned} \hat{p}_{jML}(\bar{d}, \hat{\lambda}) &= q_{jML} \left(1 + \hat{\lambda}(d_j - \bar{d}) \right)^{-1} \\ &\text{for } j = 1, \dots, n \end{aligned} \quad (7)$$

where $\hat{\lambda}$ is such that $\hat{p}_{jML}(\bar{d}, \hat{\lambda})$ satisfies the mean constraint. The solution, $\hat{p}_{jML}$, may then be used to estimate the income distribution and an entropy measure of income inequality.

V. Entropic prior

If we are observing a system on a long timescale, the current system state depends largely on the closest previous state. After discretizing income levels, a Markov process with finite discrete state and time is obtained. We denote this Markov process with the Markov matrix $\mathbf{P}$. We can then demonstrate a monotonically decreasing CR entropy family, $\gamma \geq 0$, with Markov dynamics, where $p(t)$ and $q(t)$ are two distributions following the Markov chain, with different initial distributions, $p(0)$ and $q(0)$. The function $\varphi(x) = x^{\gamma+1}$ is convex and additive. Therefore,

$$\begin{aligned} I(\{p_i(t+1)\}, \{q_i(t+1)\}, \gamma) &= \frac{1}{\gamma} \left\{ \sum_{i=1}^n q_i(t+1) \left(\frac{p_i(t+1)}{q_i(t+1)} \right)^{\gamma+1} \right\} \\ &\leq \frac{1}{\gamma} \left\{ \sum_{i=1}^n \sum_{j=1}^n q_j(t) P_{ji} \left(\frac{p_j(t)}{q_j(t)} \right)^{\gamma+1} - 1 \right\} \\ &= \frac{1}{\gamma} \left\{ \sum_{j=1}^n q_j(t) \left(\frac{p_j(t)}{q_j(t)} \right)^{\gamma+1} - 1 \right\} = I(\{p_i(t)\}, \{q_i(t)\}, \gamma) \end{aligned} \quad (8)$$

where the inequality in (8) follows from the convexity of $(\cdot)^{\gamma+1}$ (Cressie and Pardo 2000).

In many cases, we are only interested in a periodic, irreducible Markov processes because the probability of a state being strictly periodic is practically zero. Further, if the entire system can be reduced into subsystems, we can study the subsystems individually. In those cases, the CR entropy approaches zero as p and q approach the unique stationary distribution. Finally,

$$\lim_{\gamma \rightarrow 0} I(\mathbf{p}, \mathbf{q}, \gamma) = \sum_{i=1}^n p_i \ln \left(\frac{p_i}{q_i} \right), \quad (9)$$

which is the KL divergence measure in the discrete case. This criterion leads to a natural measure of the deviation of the distribution of probabilities $\mathbf{p}$ and $\mathbf{q}$. Under the principle of minimum discriminability, the difference $I(\mathbf{p}, \mathbf{q})$ is minimized. With an informative prior reference distribution and the sample information, the minimum cross-entropy solution may be obtained from (8), subject to the moment consistency constraints and the normalization constraint. The entropy measure gives a monotonicity relation to the Markov process. In contrast to moment-based estimation problems, where matching the leading moments does not guarantee convergence in probability measures, KL-divergence bounds the large deviation rate function, and thus controls moments of all orders. Consequently, divergence functions over discrete probability distributions defined on the same probability space provide a rigorous macroeconomic index for various behaviours. Importantly, for information recovery, this entropy measure provides a divergence framework and a strong basis for integrating information transfer for discrete probability distributions defined on the same probability space from various sources.

VI. Conclusion

Reigning paradigms are difficult to adapt to non-linear dynamics in complex processes. We suggest a framework with entropy-based nonlinear divergence methods for modelling individual behaviour at the micro-level. A Markov decision process and a transitional reference distribution are used to measure the impact of prior information on related macro-outcome variables. An entropy-based minimum divergence method is used in a stochastic

Markov dynamic context as a basis for predicting macroeconomic behaviour.

Author contributions

CRedit: **George Judge**: Conceptualization, Methodology, Writing – original draft, Writing – review & editing; **Wendy K. Tam**: Conceptualization, Methodology, Writing – original draft, Writing – review & editing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

No grants from public, commercial, or not-for-profit sectors agencies.

References

- Caticha, A., and A. Golan. 2014. “An Entropic Framework for Modeling Economies.” *Physica A: Statistical Mechanics and Its Applications* 408:149–163. <https://doi.org/10.1016/j.physa.2014.04.016>.
- Cressie, N. A., and L. Pardo. 2000. “Minimum Phi Divergence Estimator and Hierarchical Testing in Loglinear Models.” *Statistica Sinica* 10 (3): 867–884.
- Freeman, M. F., and J. W. Tukey. 1950. “Transformations Related to the Angular and the Square Root.” *The Annals of Mathematical Statistics* 21 (4): 607–611. <https://doi.org/10.1214/aoms/1177729756>.
- Georgesu-Roegen, N. 1976. *Energy and Economic Myths*. Pergamon Press.
- Golan, A., G. Judge, and D. Miller. 1996. *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. John Wiley.
- Good, I. 1963. “Maximum Entropy for Hypothesis Formulation, Especially for Multidimensional Contingency Tables.” *The Annals of Mathematical Statistics* 34 (3): 911–934. <https://doi.org/10.1214/aoms/1177704014>.
- Jaynes, E. T. 1978. “Where Do We Stand on Maximum Entropy?” In *Maximum Entropy Formalism*, edited by R. D. Levine and M. Tribus, 15–118. MIT Press.
- Judge, G. 2016. “Econometric Information Recovery in Behavioral Networks.” *Econometrics* 4 (3): 38. <https://doi.org/10.3390/econometrics4030038>.
- Judge, G. G., and R. C. Mittelhammer. 2012. “Implications of the Cressie-Read Family of Additive Divergences for Information Recovery.” *Entropy* 14 (12): 2427–2438. <https://doi.org/10.3390/e14122427>.

- Kullback, J. 1959. *Information Theory and Statistics*. John Wiley.
- Kullback, S., and R. Leibler. 1951. "On Information and Sufficiency." *The Annals of Mathematical Statistics* 22 (1): 79–86. <https://doi.org/10.1214/aoms/1177729694>.
- Neyman, J. 1949. "Contribution to the Theory of Certain Test Criteria." *First Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA, 239–273. University of California Press. <https://doi.org/10.1525/9780520327016-030>.
- Pearson, K. 1904. *On the Theory of Contingency and Its Relation to Association and Normal Correlation*. London: Biometric Series, Drapers' Co. Memoirs.
- Scharfenaker, E., and D. K. Foley. 2017. "Quantal Response Statistical Equilibrium in Economic Interactions: Theory and Estimation." *Entropy* 19 (9): 444–459. <https://doi.org/10.3390/e19090444>.
- Wilks, S. S. 1938. "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses." *The Annals of Mathematical Statistics* 9 (1): 60–62. <https://doi.org/10.1214/aoms/1177732360>.
- Wissner-Gross, A. D., and C. E. Freer. 2013. "Causal Entropic Forces." *Physical Review Letters* 110 (16): 167802–1–5. <https://doi.org/10.1103/PhysRevLett.110.168702>.