Research Article

Generalized Active Information: Extensions To Unbounded Domains

Daniel Andrés Díaz-Pachón[1] and Robert J. Marks II[2]*
[1]Division of Biostatistics, University of Miami, Miami, FL, USA
[2]Electrical & Computer Engineering, Baylor University, Waco, TX, USA

Abstract

In the last three decades, several measures of complexity have been proposed. Up to this point, most of such measures have only been developed for finite spaces. In these scenarios the baseline distribution is uniform. This makes sense because, among other things, the uniform distribution is the measure of maximum entropy over the relevant space. Active information traditionally assumes a finite interval universe of discourse but can be extended to other cases where maximum entropy is defined. Illustrating this is the purpose of this paper. Disequilibrium from maximum entropy, measured as active information, can be evaluated from baselines with unbounded support.

INTRODUCTION

Conventional thermodynamic maximum entropy brings to mind the diffusion of an aerosol to be uniformly distributed throughout a closed room. Similarly, Shannon's maximum entropy occurs when the probabilities of an outcome of a fixed number of events are evenly distributed and equal. Maximum entropy and probability as intertwined concepts goes as far back as Jacob Bernoulli, who stated in 1713 [2]:

> "... in the absence of any prior knowledge, we must assume that the events ... have equal probability."

Bernoulli's *Principle of Insufficient Reason* (PrOIR) [40] states that, in the absence of further knowledge, the wisest assumption we can do on the distribution of a finite set of points is that its elements are all equiprobable. Bernoulli's idea is so intuitive that it is difficult to dispute. Attempts to do so have been repudiated [29]. Interestingly, several developments in science and mathematics during the twentieth century have confirmed its validity. For example:

1. *Information theory* tells us that equiprobability is the distribution with maximum entropy (maxent) [22, 23].

2. *Learning and optimization* tell us that no search can do better uniformly on average than a blind search (i.e., an equiprobable search), which is the average of all existing distributions of a finite sample space [46, 47].

3. *Bayesian inference* uses equiprobability because it constitutes a noninformative prior over a finite sample space [3] (although, of course, other noninformative priors are possible).

The PrOIR is based on absence of further knowledge. We can generalize it to the principle of maximum entropy (maxent), which can be more powerfully asserted "for the positive reason that [the maxent probability distribution] is uniquely determined as the one which is maximally noncommittal with respect to missing information" [22]. A distribution with maximum entropy is such that it is the least biased estimate possible on some given knowledge, albeit partial (see, e.g., [5, 23, 41]). Such partial knowledge usually takes the form of moments restrictions on a distribution living in a given space. Table 1 presents the maxent distributions under some moments restrictions over the most relevant spaces [6].

In terms of active information, i.e. the information gap between two different distributions, where usually

Space	Restrictions	Distributions
Finite	None	Equiprobability
Finite interval $[a, b]$	None	$\mathcal{U}(a, b)$
$\mathbb{Z}^+$	$\mathbf{E}X = \mu$	$\text{Geom}(1/\mu)$
$\mathbb{R}^+$	$\mathbf{E}X = \mu$	$\text{Exp}(1/\mu)$
$\mathbb{R}$	$\mathbf{E}X = \mu;\ \mathbf{E}(X - \mu)^2 = \sigma^2$	$\mathcal{N}(\mu, \sigma^2)$

Table 1: Maximum entropy distributions over some relevant spaces.

$\mathbb{Z}^+ = \{1, 2, \cdots\}$ and $\mathcal{N}(\mu, \sigma^2)$ is normal with mean μ and variance σ^2.

one of the two distributions is taken as a baseline, one may choose the appropriate maxent distribution of Table 1 as this baseline. This change of perspective from the PrOIR to the more general maxent viewpoint serves multiple purposes.

First, it answers an old criticism by Häggström on Dembski's preliminary work. Häggström claimed that "there is absolutely no a priori reason to expect that the 'blind forces of nature' should produce a fitness landscape distributed [uniformly]" [20], and repeated his criticism somewhat less rancorously in [21]. The truth is that we expect a maxent distribution in every aspect of nature, and an explanation is required when some natural process does not follow it. Moreover, from a subjective point of view, maxent is the safest and most conservative assumption to make, since any other option will introduce selection bias that has not been accounted for. Either way, either in the ontological reality of nature or in our epistemological apprehension of such reality, departures from maxent cannot be explained away. It is not that out-of-equilibrium explanations are not allowed, it is that they must be accounted for. Thus, criticisms to active information, if they aspire to be successful, must be grounded on something else than the baseline distribution being of maxent.

Second, it liberates active information from having to deal with criticisms to the PrOIR [24, 45]. Many criticisms have been repudiated [18, 28, 29]. For example, Keynes devoted a whole chapter in his *Treatise on Probability* to debunk it [25]. Keynes' notions have been addressed by Marks *et al.* [29].

Third, maxent invites a generalization of active information to other type of spaces beyond the finite one considered until now. With this change, all of the previous results for active information continue being true under the umbrella of a more universal maxent principle. We can evaluate active information under noncompact spaces and justify the use of some endogenous distributions once we have acquired relevant knowledge. Accordingly, we propose to expand the theory of active information using maxent distributions as baselines.

The manifestation of maxent can be viewed thermodynamically.

Example 1. The second law of thermodynamics famously states that the entropy of a gas in a closed room reaches maximum entropy. Maximum entropy is also manifest in other domains as documented in Table 1.

Example 2. Barometric pressure is measured from the surface of the earth to space so is an example from Table 1 of maximum entropy distribution of $\text{Exp}(1/\mu)$. When the temperature lapse rate is zero, the equation for percent pressure decrease from sealevel is

$$\frac{P(h)}{P_0} = \exp\left(\frac{-gMh}{RT}\right)$$

where

- $P(h) = $ pressure at elevation h,

- $P_0 = $ static pressure at sealevel,

- $T = $ temperature in degrees Kelvin,

- $h = $ height above sea level,

- $R = $ universal gas constant: 8.3144598 J/(mol · K),

- $g = $ gravitational acceleration: 9.80665 m/s^2, and

- $M = $ molar mass of Earth's air: 0.0289644 kg/mol.

Example 3. The Maxwell-Boltzmann distribution describes the velocity in three dimensions by the vector $\vec{v} = [v_x\ v_y\ v_z]'$ of ideal gas particles as

$$f(\vec{v}) = \left(\frac{m}{2\pi kT}\right)^{3/2} e^{-\frac{m\vec{v}'\vec{v}}{2kT}},$$

where $\vec{v}'$ denotes transposition of the vector and

- $m = $ particle mass,

- $T = $ temperature,

- $k = $ Boltzman's constant: $1.38064852 \times 10^{-23}$ m^2 kg s^{-2} K^{-1}

The velocity in each dimension is in $\mathbb{R}$ and, in accordance with Table 1, the corresponding maxent displayed in the Maxwell-Boltzmann distribution is normal.

Example 4. Bernoulli's PrOIR relates to the success of a single trial, p. Repeating Bernoulli trials using the same null information assumption of a Bernoulli trial until the first success is, as shown in Table 1, $\text{Geom}(p)$ on $\mathbb{Z}^+$ and is, as an extension of a single Bernoulli trial, therefore maxent.

GENERALIZATION

Equilibria in search and learning [15, 29, 33, 42] as dictated by conservation of information and popularized by the No Free Lunch Theorems [9, 10, 46, 47] dictates that, predicated on maximum Shannon entropy, no search algorithm will outperform any other on average. Active information was introduced to measure the deviation from equilibria, thereby providing a quantification of the information infused in the algorithm to make it work better than average.

Active information can be viewed as a generalized instantiation of *anomaly detection* [1, 4, 13, 32] otherwise known as *novelty filtering* [19, 30, 43, 44]. The status quo of probabilistic uniformity is set and any significant deviation is flagged as novel. The degree of deviation from normalcy is measured by the active information. Anomaly detection typically requires training data to establish the quiescence of normalcy. Such is not the case with active information. Normalcy, rather, is defined by conservation of information which, in turn, defines the operational equilibrium of the non-informed search. We will show that this same approach can be applied elsewhere where maximum entropy defined over different domains and under different conditions uniquely establishes a performance equilibria. Montañez has nicely extended the analysis to the information capacity of search spaces and machine learning in general [34–39].

To derive active information, let Ω be a *search space* (for instance, one of those considered in Table 1) and $T \subset \Omega$ be a target, where T is a measurable set. Find the maxent distribution ϕ for Ω, subject to whatever is known. The *endogenous* information is then defined as

$$I_\Omega = -\log \phi(T). \qquad (1)$$

The units are dependent on the base of the log. When base 2, the information is in bits. Base e gives units of nats and base 10 gives Hartleys. The endogenous information represents the inherent difficulty of the problem to reach the target T. Any additional knowledge might modify the search generating a new distribution φ that will assign probability $\varphi(T)$ of reaching the target T. The information required to reach the target when applying this knowledge under φ is then *exogenous* and defined as

$$I_1 = -\log \varphi(T). \qquad (2)$$

The difference between (1) and (2) naturally defines what is called the *active information* at the target T:

$$I_+(\varphi|\phi)(T) = I_\Omega - I_1 = \log \frac{\varphi(T)}{\phi(T)}. \qquad (3)$$

Unless otherwise stated, all logarithms are taken to be base 2, so that information is measured in bits. Out of Equation (3) the following properties emerge.

1. A search is improved iff the assisted search has a higher probability that the unassisted search. In other words the active information is positive, i.e. $I_+(\varphi(T)|\phi(T)) \geq 0$ iff $\varphi(T) \geq \phi(T)$.

2. A negative active information means that the search induced by the exogenous probability is deleterious e.g., the information about the location of T was not accurate. This occurs when $\varphi(T) < \phi(T)$. The active information can approach $-\infty$ as $\varphi(T) \to 0$.

3. The maximum active information occurs when the probability of success in the alternative search equals one, i.e. $\varphi(T) = 1$. Then $I_+ = I_\Omega$ and all available information has been extracted from the search space.

4. Finally, a null active information means that the exogenous information did not contribute anything new in order to reach T. This occurs when there is no improvement and $\varphi(T) = \phi(T)$ and the exogenous and endogenous information are identical.

Knowledge cannot always be translated into the search. An interesting example in which additional knowledge about the structure of the search space does not alter the chances of reaching the target can be seen in the next example:

Example 5. Let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ be the initial search space and let $T \subset \Omega$ be the target. Thus the endogenous distribution ϕ is given by the equiprobability of all singletons, so that $\phi(T) = |T|/|\Omega|$. Although equiprobability is sometimes referred to as a discrete uniform distribution, in our context equiprobability is not uniformity, since a uniform r.v. must live in a metric space with some distance metric d, and the elements of Ω must satisfy that $d(\omega_i, \omega_{i-1}) = d(\omega_j, \omega_{j-1})$, for $2 \leq i, j \leq n$ (see, e.g., [7]). Suppose further that we learn that $\omega_i - \omega_{i-1} = 1$, for $2 \leq i \leq n$. Thus ϕ can be replaced by a uniform distribution $\mathbf{U}$. The uniform distribution incorporates the knowledge about the equidistance between points in the space but such knowledge is not detected by active information since $I_+ = \log \mathbf{U}(T)/\phi(T) = 0$.

The situation in Example 5 occurs because information is defined in terms of the *probability* of events, but

not on the events as such. Also, notice that active information *might be* introduced when we gain knowledge on the structure of the search (sample) space or on the location of the target. Example 5 deals with the first case: active information that *could be* added because we learned something about the space. Possible practical implications of this situation might be seen, for example, when comparing two different categorical r.v.'s over the same space (e.g., from nominal to ordinal, as when we jump from a space in which the only thing we knew about it was its cardinality but then learned that its elements are ordered) —or even a categorical r.v. against a uniform one over the same space provided both r.v.'s assign equal probability to each singleton (e.g., from nominal to uniform, when we learn that the elements of the space are equidistant). Thus, although a nonzero active information implies a change of distribution, not all acquired knowledge about the search space implies a nonzero active information.

Definition 1 (Stochastic domination)**.** Let ϕ and φ be two (cumulative probability) distributions over Ω. ϕ is stochastically dominated by φ if $\phi(x) \geq \varphi(x)$ for all x.

Example 6. Let $\Omega = \{1, 2, 3\}$. And consider two r.v.'s X and X' on Ω, with mass functions f_X and $f_{X'}$ given in Table 2. Calling F_X and $F_{X'}$ the respective distributions of X and X', then $F_X(x)$ is stochastically dominated by $F_{X'}(x)$.

	1	2	3
f_X	2/3	1/6	1/6
$f_{X'}$	1/3	1/3	1/3

Table 2: $F_X(x) \geq F_{X'}(x)$ **for all** $x \in \Omega$

Example 7. Let X, Y, and Z be exponential r.v.'s with intensity parameters $\lambda = 1/\mu = 0.5, 1, 1.5$ respectively, and distributions F_X, F_Y, F_Z, respectively. Then F_Z is stochastically dominated by F_Y, and F_Y in turn is stochastically dominated by F_X. See Figure 1.

Lemma 1. *Let ϕ and φ as in Definition 1. φ is stochastically dominated by ϕ if and only if $I_+(\varphi|\phi)((-\infty, x]) \geq 0$ for all x.*

Thus, whenever φ is stochastically dominated by ϕ, any target of the form $T = (-\infty, x]$ has non-negative active information, and negative active information for any target of the form $T' = (x, \infty)$.

ACTIVE INFORMATION AS A COMPLEXITY MEASURE

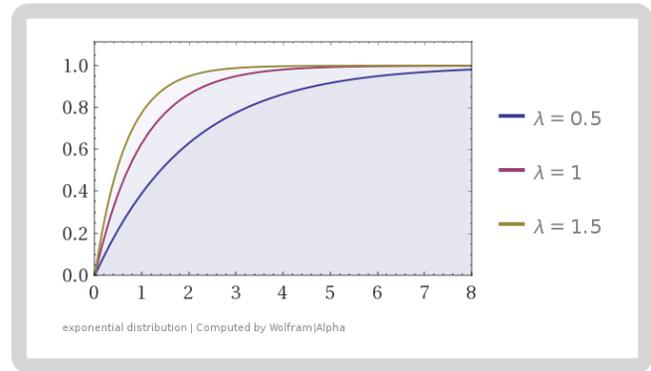Active information can also be seen as a statistical complexity measure. Feldman and Crutchfield argue that

Figure 1: Three CDF's of the exponential r.v.'s with intensity parameters 0.5, 1, and 1.5. Clearly the one with parameter 1.5 is dominated by the one with parameter 1, which in turn is dominated by the one with parameter 0.5. **doi:**10.5048/BIO-C.2020.3.f1

at least three things are needed to build a statistical complexity measure: (i) it needs to vanish on the extremes, (ii) it needs to have a clear interpretation, and (iii) it needs to have a specified use [17]. Active information, $I_+ = \log \varphi(T)/\phi(T)$, as a statistical measure of complexity *per event* satisfies these requirements:

(i) The active information is well defined. It is 0 when the endogenous probability equals the exogenous one $(\phi(T) = \varphi(T)$.) On the other hand, when $\varphi(T) = 1$ we say that the search is perfect and $I_+ = I_\Omega$, which means that all information available in the problem was extracted in the modified search for the target.

(ii) The interpretation of I_+ is evident from its definition and the four properties listed in the Generalization section. It measures how many bits of difference there are between the endogenous and the exogenous information at the event T. The active information is a measure of the degree to which a search has been assisted to succeed.

(iii) When Wolpert and Macready popularized the No Free Lunch Theorems (NFLT), they established the impossibility of one search algorithm to outperform any other without additional knowledge [46, 47]. That is, in absence of additional information, no algorithm should do better, on average, in finding a target than blind search. In their words, "the performance of any two algorithms... is, on average, identical" [46]. In reaching a target, however, search algorithms usually do better than blind chance. Why is this so? Again, using the words of Wolpert and Macready, it is because of the direct input of information, by "incorporating problem-specific knowledge into the behavior of

the [optimization or search] algorithm" [47]. Active information determines the information gap between the search of a target by pure chance and the input of an expert/dumb programmer.

The usual way to proceed with measures of complexity on a bounded domain is to account for some kind of difference between equilibrium and non-equilibrium distributions, i.e. equiprobability versus non-equiprobability, for finite settings. For instance, López-Ruiz et al. define *disequilibrium* (i.e., the sum of the quadratic Euclidean distance of the probability of atomic events), D_E, as part of their complexity measure [26]; Martin et al. replace the Euclidean distance by Wootters' [48] to obtain their disequilibirum D_W, while keeping the difference between equilibrium and non-equilibrium distributions [31]. And finally *active information* is the difference of the information for an event under equilibrium and non-equilibrium [8, 9, 14, 16, 27, 34]. In fact, just as the informational entropy is the average of the self-information, the Kullback–Leibler distance ([6], chap. 12), referenced to the maximum entropy distribution, is the average of the active information over all the elements of a finite alphabet.

CONCLUSION

We have proposed extension of active information for domains other than those confined to a finite interval. Maximum entropy (maxent) is defined on domains other than a finite interval. These maxent manifestations are evidenced in both applications and nature. Using these generalized maxent measures, active information formerly associated with finite domain distributions can be extended. The extension conforms to the criteria for measuring statistical complexity.

Active information has recently been expanded as a multidimensional mode hunting tool [11]. Briefly speaking, in a finite space every deviation from equiprobability constitutes a local mode. Thus in every event of positive active information there is a local mode. In a subsequent paper, the generalized measure is applied specifically to the Wright-Fisher model of population genetics [12].

ACKNOWLEDGMENTS

1. Allen LV, Tilbury DM (2012) Anomaly detection using model generation for event-based systems without a pre-existing formal model. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans 42(3):654–668. **doi:**10.1109/TSMCA.2011.2170418

2. Bernoulli J (1713) Ars Conjectandi.

3. Boos DD, Stefanski LA (2013) Essential Statistical Inference. Springer. **doi:**978-1-4614-4818-1

4. Chandola V, Banerjee A, Kumar V (2009) Anomaly detection: A survey. ACM Computing Surveys (CSUR) 41(3):15. **doi:**10.1145/1541880.1541882

5. Conrad K (2004) Probability distributions and maximal entropy. http://www.math.uconn.edu/~kconrad/blurbs/analysis/entropypost.pdf.

6. Cover TM, Thomas JA (2012) Elements of Information Theory. John Wiley & Sons. **doi:**10.1002/047174882X

7. Dembski WA (1990) Uniform probability. Journal of Theoretical Probability 3(4):611–626. **doi:**10.1007/BF01046100

8. Dembski WA, Marks II RJ (2009) Conservation of Information in search: Measuring the cost of success. IEEE Transactions on Systems, Man and Cybernetics A, Systems & Humans 5(5):1051–1061. **doi:**10.1109/TSMCA.2009.2025027

9. Dembski WA, Marks II RJ (2009b) Bernoulli's Principle of Insufficient Reason and Conservation of Information in computer search. Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics. San Antonio, TX, 2647–2652. **doi:**10.1109/ICSMC.2009.5346119

10. Dembski WA, Ewert W, Marks II RJ (2013) A general theory of information cost incurred by successful search. In: Marks RJ II, Behe MJ, Dembski WA, Gordon BL, Sanford JC, eds. Biological Information: New Perspectives. World Scientific (Singapore) pp 26–63. **doi:**10.1142/9789814508728_0002

11. Díaz-Pachón DA, Sáenz JP, Rao JS, Dazard J-E (2019) Mode hunting through active information. Appl Stochastic Models Bus Ind 35:376–393. **doi:**10.1002/asmb.2430

12. Díaz-Pachón DA, Marks RJ. Active information requirements for success of the Wright-Fisher model of population genetics. BIO-Complexity (in review).

13. Esponda F, Forrest S, Helman P (2004) A formal framework for positive and negative detection schemes. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 34(1):357–373. **doi:**10.1109/tsmcb.2003.817026

14. Ewert WA, Dembski WA, Marks II RJ (2012) Climbing the Steiner Tree—Sources of active information in a genetic algorithm for solving the Euclidean Steiner Tree Problem. BIO-Complexity 2012(1):1–14. **doi:**10.5048/BIO-C.2012.1.e2

15. Ewert W, Dembski WA, Marks II RJ (2013) Conservation of Information in relative search performance. Proceedings of the 2013 IEEE 45th Southeastern Symposium on Systems Theory (SSST). Waco, TX, 41–50. **doi:**10.1109/SSST.2013.6524963

16. Ewert WA, Dembski WA, Marks II RJ (2013) Active information in metabiology. BIO-Complexity 2013(4):1–10. **doi:**10.5048/BIO-C.2013.4

17. Feldman DP, Crutchfield JP (1998) Measures of statistical complexity: Why? Phys Letters A 238(4):244–252. **doi:**10.1016/S0375-9601(97)00855-4

18. Fisher A (1922) The Mathematical Theory of Probabilities and Its Application to Frequency Curves and Statistical Methods, Volume 1. The Macmillan Company. **doi:**10.1177/000271622310700150

19. Guttormsson S, Marks II RJ, El-Sharkawi MA, Kerszenbaum I (1999) Elliptical novelty grouping for on-

line short-turn detection of excited running rotors. IEEE Transactions on Energy Conversion 14(1):16–22. **doi:**10.1109/60.749142

20. Häggström O (2005) Intelligent Design and NFL theorems: Debunking Dembski. http://www.math.chalmers.se/~olleh/Dembski.pdf.

21. Häggström O (2007) Biol Philos 22:217.

22. Jaynes ET (1957) Information theory and statistical mechanics. Physical Review Series II 106(4):620–630. **doi:**10.1103/PhysRev.106.620

23. Jaynes ET (1957) Information theory and statistical mechanics II. Physical Review Series II 108(2):171–190. **doi:**10.1103/PhysRev.108.171

24. Kasner E, James RN (2001) Mathematics and the Imagination. Dover Publications.

25. Keynes JM (1921) A Treatise On Probability. Macmillan.

26. López-Ruiz R, Macini HL, Calbet X (1995) A statistical measure of complexity. Phys Letters A 209(5–6):321–326. **doi:**10.1016/0375-9601(95)00867-5

27. Marks II RJ (2010) Evolutionary computation: A perpetual motion machine for design information? In: Dembski WA, Licona MR, eds. Evidence for God: 50 Arguments for Faith from the Bible, History, Philosophy, and Science. Baker Books, 91–96.

28. Marks II RJ, Behe M, Dembski WA, Gordon BL, Sanford JC, eds (2016) Biological Information—New Perspectives. Cornell University. World Scientific. **doi:**10.1142/8818

29. Marks II RJ, Dembski WA, Ewert W (2016) Introduction to Evolutionary Informatics. World Scientific (Singapore). **doi:**10.1142/9974

30. Marsland S (2003) Novelty detection in learning systems. Neural Computing Surveys 3(2):157–195.

31. Martin MT, Plastino A, Rosso OA (2003) Statistical complexity and desequilibrium. Phys Letters A 311(2–3):126–132. **doi:**10.1016/S0375-9601(03)00491-2

32. Meneganti M, Saviello FS, Tagliaferri R (1998) Fuzzy neural networks for classification and detection of anomalies. IEEE Transactions on Neural Networks 9(5):848–861. **doi:**10.1109/72.712157

33. Mitchell TM (1990) The need for biases in learning generalizations. Technical Report CBM-TR-117, Department of Computer Science, Rutgers University (1980). p.59. Reprinted In: Shavlik JW, Dietterich TG. Readings in Machine Learning (Morgan Kauffmann Series in Machine Learning) pp 184–190.

34. Montañez GD, Ewert W, Dembski WA, Marks II RJ (2010) A vivisection of the ev computer organism: Identifying sources of active information. BIO-Complexity 2010(3):1–6. **doi:**10.5048/BIO-C.2010.3.e1

35. Montañez GD (2011) Information storage capacity of genetic algorithm fitness maps. Masters diss. http://hdl.handle.net/2104/8233.

36. Montañez GD (2013) Information transmission through genetic algorithm fitness maps. 2013 IEEE Congress on Evolutionary Computation (CEC), 756–763. **doi:**10.1109/CEC.2013.6557644

37. Montañez GD (2013) Bounding the number of favorable functions in stochastic search. 2013 IEEE Congress on Evolutionary Computation (CEC), 3019–3026. **doi:**10.1109/CEC.2013.6557937

38. Montañez GD (2017) Why Machine Learning Works. PhD diss., Carnegie Mellon University.

39. Montañez GD (2017) The famine of forte: Few search problems greatly favor your algorithm. 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 477–482. **doi:**10.1109/SMC.2017.8122651

40. Papoulis A (1991) Probability, Random Variables, and Stochastic Processes, 3rd ed. McGraw-Hill (New York).

41. Park SY, Bera AK (2009) Maximum entropy autoregressive conditional heteroskedasticity model. J of Econometrics 150:219–230. **doi:**10.1016/j.jeconom.2008.12.014

42. Schaffer C (1994) A conservation law for generalization performance. In: Hirsh H, Cohen WW, eds. Proc 11th Int Conf Mach Learn, 259–265.

43. Streifel RJ, Marks II RJ, El-Sharkawi MA and Kerszenbaum I (1996) Detection of shorted-turns in the field winding of turbine-generator rotors using novelty detectors: Development and field test. IEEE Transactions on Energy Conversion 11(2):312–317. **doi:**10.1109/60.507183

44. Thompson BB, Marks II RJ, Choi JJ, El-Sharkawi MA (2002) Implicit learning in autoencoder novelty assessment. Proceedings of the 2002 International Joint Conference on Neural Networks, 2002 IEEE World Congress on Computational Intelligence, Honolulu, 2878–2883. **doi:**10.1109/IJCNN.2002.1007605

45. Ulrych TJ, Clayton RW (1976) Time series modeling and maximum entropy. Physics of the Earth and Planetary Interiors 12(2):188–200. **doi:**10.1016/0031-9201(76)90047-9

46. Wolpert DH, Macready WG (1995) No Free Lunch Theorems for search. Technical Report SFI-TR-95-02-010. Sante Fe Institute (Santa Fe, NM). http://delta.cs.cinvestav.mx/~ccoello/compevol/wolpert96.pdf.

47. Wolpert DH, Macready WG (1997) No Free Lunch Theorems for optimization. IEEE Transactions on Evolutionary Computation 1(1):67–82. **doi:**10.1109/4235.585893

48. Wootters WK (1981) Statistical distance and Hilbert space. Phys Rev D 23(2):357–362. **doi:**10.1103/PhysRevD.23.357