Minds and Machines (2018) 28:787–793
https://doi.org/10.1007/s11023-018-9477-2

COMMENTARY

Info-Metrics for Modeling and Inference

Amos Golan[1,2,3]

Published online: 11 September 2018

Abstract

Info-metrics is a framework for rational inference based on insufficient information. The complete info-metric framework, accompanied with many interdisciplinary examples and case studies, as well as graphical representations of the theory appear in the new book "Foundations of Info-Metrics: Modeling, Inference and Imperfect Information," Oxford University Press, 2018. In this commentary, I describe that framework in general terms, demonstrate some of the ideas via simple examples, and provide arguments for using it to transform information into useful knowledge.

Keywords Decision making · Inference · Information · Info-metrics · Modeling · Optimization

A young couple returns to their house by the beach. The kitchen floor is very wet, a light is flashing on the dishwasher panel, a half-gallon fishbowl lies broken on the counter and the goldfish is gone. What happened while they were at the beach? Many stories could explain the scene. How we deduce which explanation is correct is a fundamental problem in solving mystery stories and scientific puzzles.

Unburying a hidden story is complicated because insufficient information—including limited, incomplete, complex, noisy, and uncertain information—is the norm for most investigations across all disciplines. In this commentary, I describe a framework for converting information to useful knowledge. It is a framework for rational inference based on partial, and often uncertain and noisy, information, called *info-metrics*. I begin by taking a closer look at underlying inference problem in a more precise way via simple examples.

✉ Amos Golan
agolan@american.edu

1 Info-Metrics Institute and Economics, American University, Washington, DC, USA

2 Santa Fe Institute, Santa Fe, USA

3 Pembroke College, Oxford, UK

1 The Underlying Problem: Insufficient Observed Information

Consider the problem of figuring out the ages of Adam and Eve from the information that their joint age is forty-two. This problem is mathematically underdetermined: there is more than a single solution that can be logically consistent with this information. In fact, this problem generates a continuum of solutions.

Even more challenging, consider the case where all we know is that the joint age of Adam and Eve is more-or-less forty-two. The ‘more-or-less’ captures the uncertainty about our information, the inherent noise in the observed information, or both. Going back to the missing fish example, if rather than the couple returns to their home, imagine that a guest entered the home before the young couple. She would not know whether a single goldfish or many were silently calling for help somewhere on the kitchen floor. Or think of modeling a social system of interacting agents. We never know the exact relationships among the agents, other processes within the system, and the properties of the system. That uncertainty, and noise in the observable evidence, complicate the way we go about solving the problem. We already know that the simpler problem is underdetermined. But unlike the noiseless case where more information leads to a unique solution (say, Adam is 22), in the uncertain and noisy case the problem remains underdetermined regardless of the amount of noisy evidence we have (say, Adam is more-or-less 22).

The info-metrics framework provides a rational inference framework for dealing with mathematically underdetermined problems caused by insufficient information. It provides a logical framework for converting input information—including noisy and uncertain information—into useful knowledge.

There are numerous other approaches to the inference problem. These methods vary across disciplines and problems. Some approaches are well known: maximum likelihood, least squares, Bayesian, and various computational methods. Each one of these approaches imposes different information on the inferential method; each one uses different assumptions and soft information; each one uses a different criterion function. Unsurprisingly, each yields a different solution. By soft information I mean information acquired before we saw the evidence, the assumptions and structures we impose on the solution (Adam is at least 5 years older than Eve), our own interpretation of the evidence (most of the water on the floor is from the broken dishwasher), our values and beliefs (the cat would never attack the fish!), and any other form of unobserved, information, such as private information not available to the investigator or the scientist (the husband hates the goldfish). Much of that soft information is subjective. Together, the observed (or circumstantial) and soft information, comprise the total information we use. I call it ‘input information.’

The advantage of the info-metrics framework over other approaches to the inference problem is that the info-metrics framework allows us to process the available information with minimal reliance on assumptions that cannot be validated. It converts the underdetermined problem to a well-behaved one, guaranteeing a unique solution. In doing so, basic tenets of science, requiring a basic set of desired axioms to be satisfied, are fulfilled. We want this set of axioms to be minimal in the sense that once imposed, the framework satisfies all the desired properties; no additional axioms are needed. We also want our solution to be conservative in the sense that out of all possible solutions, that are consistent with the entire observed evidence, we choose the one that is based on minimal assumptions to fully characterize the system under study and enable inference. In more commonly used language, we want the solution, or the model yielding the solution, that is the least complex. For example, if a model has three parameters and the other has two, but other than that they tell the exact same story, the two-parameter model is preferred. This is the model that the info-metrics framework will solve for. All other approaches introduce more soft information into their processing method and therefore risk introducing incorrect information, assumptions or structures. Not only does the info-metrics approach minimize assumptions and the use of soft information, it is also computationally and statistically efficient. It holds good statistical properties, provides a way to validate the assumptions and the model, and is grounded in rational optimization framework.

2 The Info-Metrics Framework

The info-metrics framework is grounded in ordinary notions of optimization of a decision criterion. In info-metrics this decision criterion is an information-theoretic one called entropy (Shannon 1948). The process of choosing the solution that optimizes our decision criterion is called an optimization process. We optimize that decision criterion—entropy—while considering the *entire* observed information together with the other information required to fully characterize the problem under study. No hidden assumptions, such as assumptions about structure, are imposed in the optimization process. This optimization process is the info-metrics characterization of rational inference: we want to minimize imposed assumed structure. All the information we have, including assumptions, enters as constraints within our optimization. These constraints can be thought of as conservation rules; they conserve the available information. Optimization of the entropy decision function selects the best solution. (In more statistical terms, it is the constraints, together with the decision criterion, that determine the likelihood.)

A simple, yet representative, example helps to illustrate the approach. Consider the well-known six-sided die example, first introduced by Boltzmann (in the 1870's) and then popularized by Jaynes (in the 1960's). Suppose a six-sided die is tossed 50 times. We only know the mean value of the number of spots showing on the upper face, say 3.7. Given that information, we want to infer the probabilities that a specific value of the face will show up in the next toss of the die. There are infinitely many combinations of probabilities (i.e., solutions) that satisfy a certain mean. Take for example, the unlikely case of a loaded die, with mean observed value of 5.3. Both probability distributions (0.008, 0.019, 0.044, 0.104, 0.245, 0.580) and (0, 0, 0.010, 0.170, 0.330, 0.490) corresponding to the six possible values (1, 2, 3, 4, 5, 6) satisfy this mean value, and the requirement that the probabilities add up to exactly one. Which one should we choose? What number would you bet on? The info-metrics approach tells us to choose the solution that is a result of minimal assumptions. This means, the chosen solution is the most uniform one—the most uncertain. In our example, it is the solution that is closest as possible to (1/6, 1/6, 1/6, 1/6, 1/6, 1/6). That solution is the first one above, which is the info-metrics one. Any other approach will yield a less uniform solution, a solution that is built on more assumptions. In fact, the second solution above is the commonly used least squares solution. As can be easily observed, that solution is much more 'informed'—it is further away from the uniform, uninformed, solution. That additional information did not come from evidence but rather via the least squares assumptions. Since we do not have that information, it is better not to include it in the model. (See http://info-metrics.org/ for an interactive version of this example, computer codes, and additional examples.)

As we did earlier, if we are to be realistic we must consider noise and uncertainty as well. I will modify the die example to demonstrate that issue. Instead of a single measurement, we have two measurements of the same die, each based on the total of the numbers on the upper face of the die after several throws. The first measurement is exactly 30 after 10 throws (mean is 3). The second is 80 following another 20 throws (mean is 4). These measurements are independent of each other. We can view it as two different individuals who counted independently with possible recording or summing-up errors, or as the same individual who counted both trials independently but may have made errors in recordings or calculations. In a more statistical language, these are two independent samples taken from the same underlying distribution, where the first sample is half the size of the second one. These two measurements, or pieces of observable information, are not consistent with each other. Though they came out of the same die, they reflect two different underlying distributions; distributions that will lead to opposite conclusions for a decision maker trying to infer the next outcome. To solve the problem we need to somehow combine these two seemingly contradictory, noisy pieces of information. To do so with other methods demands the imposition of information we don't have, such as assuming that the uncertainty about the two measurements is a direct function of the sample size. The info-metrics framework provides us with a way to solve the problem without imposing unrealistic assumptions.

The info-metrics modeling approach I discuss here has its roots in information theory in conjunction with the Bernoulli–Laplace–Keynes principle of indifference, also known as the 'principle of insufficient reason.' The principle of indifference instructs us to assign equal probabilities to all events unless we know a priori that some events are more probable than others. Additional knowledge, that comes from our understanding of the rules governing the target system, is expressed in the constraints. This is also the logic underlying the maximum entropy principle, proposed in 1957 by Ed Jaynes. The major difference is that the info-metrics framework is a generalization of that principle. It extends it to solve a much wider class of problems under all types of uncertainties. Info-metrics is an inherently interdisciplinary framework that emerged from the intersection of information theory, statistical inference, and decision-making under uncertainty.

3 Why Info-Metrics?

Info-metrics is the only inference and modeling framework satisfying the requirements specified above. It unburies the most conservative hidden story in the evidence we have. Most conservative meaning the info-metrics approach imposes the minimal assumptions needed to identify the solution out of the continuum of solutions satisfying the evidence. (In information theory, it is referred to as the most 'uninformed' solution that satisfies the evidence.) In addition to providing a unique solution, that framework allows us to test our model and prediction. It provides us with the tools to validate our hypotheses and to improve our model and interpretations if our initial set of hypotheses is rejected. In line with Popper's (1959) arguments that a scientific theory can never be proved correct, the info-metrics framework allows us to eliminate and falsify theories based on empirical evidence. But somewhat different than Popper's approach, my point of view here is that there are different hypotheses concerning some system, and rival hypotheses are continuously eliminated by new evidence. Thus, success is the successful falsification of an increasing number of rival theories, increasingly corroborating the surviving one. I now provide additional arguments in favor of the info-metrics approach.

Info-metrics satisfies the notion of simplicity that goes back to the famous "Occam's razor," which says that there is no need for a more complex explanation (model) if a simple one already provides the same answer. That idea is attributed to William of Occam, a Franciscan monk who lived in the thirteenth century. Though the idea is clear and simple, at times it is not easy to choose among models that have slightly different predictive power and are of different levels of complexity.

With the above in mind, we can go back to the simplicity requirement of the constrained optimization approach advocated here. The above notions of simplicity can be defined as the minimal number of imposed constraints (or pieces of information) needed to adequately answer our question or explain the system we model, or its theory, to its fullest level of complexity. It can be thought of as the model with the minimal number of parameters needed to adequately describe the system under study. The info-metrics approach ensures that simplicity. Simply stated, there is a parameter for each piece of information, or constraint, used. Given that these constraints (by definition) are not redundant, this means that there could be no simpler way to characterize the solution to the problem or the model.

The above argument proves that if we use entropy as our decision criterion, and if the minimal set of correct constraints (governing rules) is used, then our solution (model or theory) is adequately explained. Though the notion of 'adequately explained' is not fully defined here, I mean that our solution captures the complete information we need for understanding the system we study or to solve our problem. It is the simplest possible solution or model. It is the one that is based on minimal imposed assumptions. In a more statistical language, it is flattest possible likelihood that satisfies the observed information. Referring to the six-sided die example, the flattest likelihood means the most uniform distribution, of the six-sided die, satisfying the two pieces of information simultaneously.

A complementary argument, in favor of the info-metrics framework, is that the rules that govern the problem (or system) under study and imposed as constraints (the information used), are sufficient statistics. Generally speaking, a statistic is a certain function of the values of an observed information, such as the observed sample's expectation or median. It is the mean value of the six-sided die rolls, or the joint, or mean, age of Adam and Eve. A sufficient statistic is a statistic that summarizes all the information in the observed sample, such that given that statistic, the observed sample is maximally random and can convey no further information about the underlying population or system studied. It provides a way to significantly reduce the dimensionality of the observed information while preserving all the information contained in the observed sample. In the info-metrics approach, the constraints are sufficient statistics, assuming they are correctly specified.

The next argument is to do with unobserved, but essential information. What determines a rule (constraint) within info-metrics is not that its value happens to be observed or known, but rather that it represents a relevant piece of information that constrains the system. It is one of the governing laws of the system we study or the theory we are formulating. Therefore, we must impose it in our model in order to have an adequately described model at the desired level of complexity. In practice, there are many problems where the expected value is not observed. For example, consider a problem in statistical mechanics where thermodynamic equilibrium requires a constraint on the expected energy (The conservation of energy law). However, the value of the expected energy is not known. Regardless, we still use such information (constraint) because it is necessary; the energy level affects the distribution of interest, say a certain type of gas. Including that constraint allows us to determine a family of solutions that depend on temperature as a parameter. Since temperature is measurable, we can then solve the problem empirically. Another example from the social and behavioral sciences is expected utility, which is unobserved, yet necessary for characterizing and modeling such systems.

So far, I have emphasized the inference and modeling issues. But the info-metrics framework also facilitates constructing theories. The info-metrics framework can be viewed as a "meta-theory"—a theory of how to construct theories and models given the imperfect information we have. That frame-work provides a rational perspective that helps us to identify the elements needed for building a reasonably sound theory or a model. Though in slightly different context, I discussed this above in the setting of testing and falsification.

Finally, unlike most other approaches, the info-metrics framework allows us to accommodate all possible uncertainties, including possibly the most important one: model uncertainty. This may be done by allowing all the information imposed as constraints to be specified a stochastic (the joint age of Adam and Eve is more-or-less 42). When more than a single constraint is used, such a formulation can be essential, and it allows us to find the optimal solution (under the criteria discussed earlier) with minimal assumptions. It also allows us to infer the solution without imposing specific statistical assumptions. That approach is not just mechanical, but rather it is based on the following argument. It is rare that we know the underlying model or theory. But any approach we take will find a certain solution even if our model or theory (often called the 'functional form' of the equations used) is incorrect. That leads to a misspecified model and a wrong conclusion. The info-metrics approach allows us to find the best approximate model to the underlying theory. It is the best model out of possible models that satisfy the information at hand. It is the one that enables us to learn the most about the still unknown theory. From a practical point of view, such an approach is essential, as our observed information is constantly evolving, complex and with much noise. The info-metrics framework provides us with all the necessary tools to learn from complex and evolving evidence.

4 Summary

Info-metrics is a powerful and complete framework for modeling and inference, rather than a problem-specific model. It allows us to construct theories and models and to perform consistent inferences and predictions with all types of information and uncertainty. Naturally, each problem is different and demands its own information and structure, but the info-metrics framework provides us with the general logical foundations and tools for approaching all inference problems. It also guides us toward a correct specification of the constraints, which is a nontrivial problem.

The discussion in this commentary is based on my new book "*Foundations of Info-Metrics: Modeling, Inference and Imperfect Information*," Oxford University Press, 2018. The complete framework, its rational, development and many interdisciplinary examples are discussed in that book. The presentation of the theory is supported by a collection of intuitively accessible figures. The book is therefore suitable for the curious as well as for serious practitioners.

(So, what happened to the fish? The husband did it.)

References

Golan, A. (2018). *Foundations of info-metrics: Modeling, inference, and imperfect information*. Oxford: Oxford University Press.

Popper, K. R. (1959). *The logic of scientific discovery*. London: Hutchinson.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27,* 379–423.