Leinster, Tom
Entropy and diversity. The axiomatic approach. (English) Zbl 07298513
Cambridge: Cambridge University Press (ISBN 978-1-108-83270-0/hbk; 978-1-108-96557-6/pbk; 978-1-108-96355-8/ebook). xiv, 442 p. (2021)

This book was born of the author's research in category theory, brought to life by the author's vigorous debate on how to quantify biological diversity, given strength by information theory, and fed by the scientific field of functional equations, applying the power of the axiomatic method to a biological problem of pressing concern while presenting new advances in pure mathematics. The book shows that the theory of diversity measurement is fertile soil for new mathematics, just as much as the neighboring but far more thoroughly worked field of information theory. The problem of quantifying diversity is to take a biological community and extract from it a numerical measure of its diversity, which is certainly beset with troublesome practical problems and statistical difficulties. The book focuses on a fundamental conceptual problem. The spectrum of possible interpretations of the concept of diversity ranges from species to communities, which is encoded by a continuous one-parameter family $(D_q)_{q\in[0,\infty]}$ of diversity measures [Ecology 93, 477–489 (2012)], different values of the viewpoint parameter $q \in [0, \infty]$ producing different judgements on which of the two distributions is more diverse. One might wish to evaluate an ecological community in a way that takes into account some notion of the values of the species such as phylogenetic distinctiveness, for which there is a sensible family (σ_q) of measures that does the job.

Information theory surely helps to analyze the diversity of ecological communities made up of a number of smaller communities such as geographical regions, the established notions of relative entropy, conditional entropy and mutual information providing meaningful measures of the structure of a metacommunity. The author does more than simply translate information theory into ecological language. By way of example, the new characterization of the Rényi entropies is a byproduct of the characterization theorem for measures of ecological value.

This book grew out of a general category-theoretic study of size. Sets have cardinality, vector spaces have dimension, subset of Euclidean space have volume, topological spaces have Euler characteristic, and so on. Some unification is achieved by defining a notion of size of categories themselves (called magnitude or Euler characteristic) [*T. Leinster*, Doc. Math. 13, 21–49 (2008; Zbl 1139.18009)]. The theory of magnitude of categories is closely related with the theory of Möbius-Rota inversion for partially ordered sets [*G.-C. Rota*, Z. Wahrscheinlichkeitstheor. Verw. Geb. 2, 340–368 (1964; Zbl 0121.02406); *T. Leinster*, Bull. Belg. Math. Soc. - Simon Stevin 19, No. 5, 909–933 (2012; Zbl 1269.18002)]. Furthermore, the decisive, unifying step is the generalization of the concept of magnitude to enriched categories [*T. Leinster*, Doc. Math. 18, 857–905 (2013; Zbl 1284.51011)], including not only categories but also metric spaces, graphs and additive categories. Magnitude for enriched categories is also to be realized as the Euler characteristic of a certain Hochschild-like homology theory of enriched categories, in the same sense that the Jones polynomial for knots is the Euler characteristic of Khovanov homology [*M. Khovanov*, Duke Math. J. 101, No. 3, 359–426 (2000; Zbl 0960.57005)], which was established by *M. Shulman* and the author ["Magnitude homology of enriched categories and metric spaces", Preprint, `arXiv:1711.00802`], building on the case of magnitude homology for graphs [*R. Hepworth* and *S. Willerton*, Homology Homotopy Appl. 19, No. 2, 31–60 (2017; Zbl 1377.05088)].

The connections between diversity measurement on the one hand, and information theory and category theory on the other, are fruitful for both mathematics and biology. Much of the book is about characterization theorems for entropies, diversities and means, and the conditions characterizing these quantities are functional equations.

A synopsis of the book, consisting of 12 chapters together with an appendix on proofs of background facts, go as follows:

- Chapter 1 solves three classical, fundamental, functional equations. §1.1 is concerned with Cauchy's equation on a function $f : \mathbb{R} \to \mathbb{R}$

$$f(x + y) = f(x) + f(y)$$

§1.2 deals with the functional equation

$$f(mn) = f(m) + f(n)$$

on a sequence $(f(n))_{n \geq 1}$. §1.3 is engaged in the functional equation

$$f(xy) = f(x) + g(x) f(y)$$

in two unknown functions $f, g : (0, \infty) \to \mathbb{R}$.

- The single most significant property of Shannon entropy is the chain rule which is a formula for the entropy of a composite distribution. The goal of Chapter 2 is to show that Shannon entropy is essentially the unique quantity abiding by the chain rule. §2.1 reviews probability distributions and composition of distributions. §2.2 derives the chain rule itself, along with other basic properties of Shannon entropy. The fundamental concepts and theorems of coding theory were set out in [*C. E. Shannon*, Bell Syst. Tech. J. 27, 379–423, 623–656 (1948; Zbl 1154.94303)], with rigor and detail added soon afterwards in [*A. I. Khinchin*, Mathematical foundations of information theory. New York: Dover Publications, Inc (1957; Zbl 0088.10404); *A. Feinstein*, Foundations of information theory. McGraw-Hill Electrical and Electronic Engineering Series. New York-Toronto-London: McGraw-Hill Book Company, Inc (1958; Zbl 0082.34602)] and so on. §2.3 presents part of that early work, giving Shannon's source coding theorem. Entropies of various kinds have been used to measure biological diversity for almost as long as diversity measures have been considered. §2.4 explains entropy in terms of diversity. The unique characterization of Shannon entropy by the chain rule is established in §2.5.
- The notion of relative entropy allows of comparing two probability distributions on the same space. Relative entropy goes by a remarkable number of names such as "Kullback-Leibler information" [*M. J. Schervish*, Theory of statistics. New York, NY: Springer-Verlag (1995; Zbl 0834.62002)], "Kullback-Leibler distance" [*T. M. Cover* and *J. A. Thomas*, Elements of information theory. New York: John Wiley &| Sons, Inc. (1991; Zbl 0762.94001)], "Kullback-Leibler divergence" [*M. Lovric* (ed.), International encyclopedia of statistical science. Berlin: Springer (2011; Zbl 1241.62001), pages 720–722], "directed divergence" [*S. Kullback*, Information theory and statistics. London: Chapman & Sons, Inc. (1959; Zbl 0088.10406)], "information divergence" [*M. Grendár* and *R. K. Niven*, Inf. Sci. 180, No. 21, 4189–4194 (2010; Zbl 1204.94051)], "amount of information" [*A. Rényi*, in: Proc. 4th Berkeley Symp. Math. Stat. Probab. 1, 547–561 (1961; Zbl 0106.33001)], "relative information" [*I. J. Taneja* and *P. Kumar*, Inf. Sci. 166, No. 1–4, 105–125 (2004; Zbl 1101.68947)], "information gain" [*A. Rényi*, Probability theory. English translation by László Vekerdi. Amsterdam-London: North-Holland Publishing Company (1970; Zbl 0206.18002)], "discrimination distance" [*M. Kanter*, Probab. Theory Relat. Fields 86, No. 3, 403–422 (1990; Zbl 0685.60040)], "error" [*D. F. Kerridge*, J. R. Stat. Soc., Ser. B 23, 184–194 (1961; Zbl 0112.10302)], among others. Chapter 3 provides multiple explanations and applications of relative entropy, as well as a theorem pinpointing what makes relative entropy uniquely useful. §3.1 gives the definition and basic properties of relative entropy without motivation. §3.2 gives explanation of relative entropy in terms of coding. §3.3 interprets the exponential of relative entropy as a measure of how diverse or atypical one community is when seen from a viewpoint of another. elaborating the ideas of [*R. Reeve* et al., "How to partition diversity", Preprint, `arXiv:1404.6520`]. §3.4 gives brief accounts of roles played by relative entropy in three other subjects, namely, in measure theory, in geometry and in statistics. §3.5 gives a characterization theorem for relative entropy [Zbl 1409.94838].
- Chapter 4 presents two one-parameter families of entropies, both including Shannon entropy as the case $q = 1$. The author begins with the q-logarithmic entropies $(S_q)_{q \in \mathbb{R}}$ [*J. Havrda* and *F. Charvát*, Kybernetika 3, 30–35 (1967; Zbl 0178.22401)] in §4.1. §4.2 collects some basic facts about power means, which streamline later material on Rényi entropies and diversity measures. §4.3 gives the other main family of deformation of Shannon entropy, namely, the Rényi entropies $(H_q)_{q \in [-\infty, \infty]}$, introducing the Hill number D_q of order q. §4.4 establishes the main properties of tng ihe Hill numbers by exploiting properties of the power means. This book proves two characterization theorems for the Hill numbers, the first one being established in §4.5 while the second one being postponed to §7.4.
- The ideal of the axiomatic approach to diversity measurement is to be able to say "any measure of diversity abiding by properties X, Y and Z must be one of the following." Theorems of this type

in this book stand on the shoulders of characterization theorems for means. Therefore chapter 5 is concerned with means, beginning with the classical quasiarithmetic means in §5.1. §5.2–§5.4 concern general unweighted means, culminating in four characterization theorems. §5.5 develops a method for converting characterization theorems for unweighted means into those for weighted means, one of the resulting four characterization theorems of weighted means going back to [*G. H. Hardy* et al., Inequalities. Cambridge: Cambridge University Press (1934; Zbl 0010.10703)].

- Chapter 6 is concerned with species similarlty and magnitude. The author encodes the similarities between species as a real matrix Z, continuing to represent the relative abundances of the species as a probability distribution $\boldsymbol{p}$. With this model of a community, the author defines for each $q \in [0, \infty]$ a measure $D_q^Z(\boldsymbol{p})$ of the diversity of the community. Under the extreme hypothesis that different species never have anything in common, Z is the identity matrix I and the diversity $D_q^I(\boldsymbol{p})$ reduces to the Hill number $D_q(\boldsymbol{p})$, in which sense these similarity-sensitive diversity measures generalize the Hill numbers. These are discussed in §6.1 and §6.2. §6.3 discusses that every similarity matrix is of an unambiguous diversity, independent of q, and a distribution maximazing the diversity of all orders q simultaneously. §6.4 and §6.5 are a broad-brush survey of magnitude, demonstrating that maximum diversity, far from being tethered to ecology, is of profound connections with fundamental invariants of geometry.
- Chapter 7 considers what the value of the whole is in terms of its parts. §7.1 defines the value measure σ_q, analyzing some special cases with important examples from both ecology and the analysis of social welfare. §7.2 describes the relationship between value, relative entropy and some other quantities. §7.3 proves that the only value measure with reasonable properties are those belonging to the family (σ_q), from which §7.4 deduces that for communities modelled as their relative abundance distributions, the only reasonable measures of diversity are the Hill numbers.
- Chapter 8 is concerned with mutual information and metacommunities. The starting point is the classical information-theoretic concept of mutual information (a measure of the dependence between two random variables) and the closely related concepts of conditional and joint entropy, as is discussed in §8.1. Next, the author takes exponentials of all these quantities, producing a suite of meaningful measures of ecological metacommunity (large community) divided into smaller subcommunities, where the two random variables in play correspond to a choice of species and a choice of subcommunity. It is argued in §8.2 that some of the measures reflect features of individual subcommunities, while it is claimed in §8.3 that others encapsulate information concerning the entire metacommunity. §8.5 reduces all of the entropies and diversities in this chapter to relative entropy, where reducing the various metacommunity and subcommunity measures to one single concept provides new insights into their their ecological meaning while in the diversity case, they are also usefully expressed in terms of value, as was observed in the previous chapter. The diversity measures treated in this chapter are a very special case of those introduced by *R. Reeve* et al. ["How to partition diversity", Preprint, `arXiv:1404.6520`], allowing a general q (variable emphasis on rare or common species) and a general Z (to model the varying similarities between species), which is sketched in §8.6.
- Chapter 9 addresses how to solve certain functional equations using results from probability theory after [*G. Aubrun* and *I. Nechita*, Confluentes Math. 3, No. 4, 637–647 (2011; Zbl 1245.46011)]. The key theorem from probability theory is a variational formula for the moment generating function, as is discussed in §9.1. This formula is to be understood conceptually as the convex conjugate of Cramer's large derivation theorem, as is seen in §9.2. The probabilistic method is applied to characterize the l^p norms in §9.3 and the power norms in §9.4.
- A measure-preserving map between finite probability spaces is to be regarded as a deterministic process, losing information as such. Chapter 10 attempts to quantify how much information is lost. The main theorem claims that, as soon as a few reasonable requirements are imposed on this quantity, it is highly constrained, meaning that, up to a constant factor, it must be the difference between the entropies of the domain and the codomain, which is another characterization of Shannon entropy [*J. C. Baez* et al., Entropy 13, No. 11, 1945–1957 (2011; Zbl 1301.94043)]. §10.1 is a review of measure-preserving maps, defining information loss. After recording a few simple properties of information loss, §10.2 proves that they characterize it uniquely.
- Chapter 11 defines the entropy of any probability distribution whose probabilities are not real numbers but integers modulo a prime p. §11.1 and §11.2 characterize entropy and information loss modulo p. §11.3 defines a sense in which certain real numbers can be said to have residues mod$p$.

§11.4 finishes this chapter by developing an alternative but equivalent approach to entropy modulo a prime.

- Chapter 12 describes a general category-theoretic construction which, when given as input the real line and the notion of finite probability distribution, automatically produces as output the notion of Shannon entropy. The categorical construction involves operads and their algebras. §12.1 sets out definitions of operads and algebras. For an operad P, there are notions of categorical P-algebra $\boldsymbol{A}$ (a category acted on by P) and of internal algebra in $\boldsymbol{A}$, as is discussed in§12.2. Theorem 12.3.1 claims that for the operad Δ of simplicies and the categorical Δ-algebra $\mathbb{R}$, the internal algebras in $\mathbb{R}$ are precisely the scalar multiples of Shannon entropy. §12.4 describes the free categorical Δ-algebra containing an internal algebra.

Reviewer: Hirokazu Nishimura (Tsukuba)

MSC:

18-02 Research exposition (monographs, survey articles) pertaining to category theory
92-02 Research exposition (monographs, survey articles) pertaining to biology
92D40 Ecology
94-02 Research exposition (monographs, survey articles) pertaining to information and communication theory
94A17 Measures of information, entropy

Full Text: DOI