Review of the book
"Information Theory:
Coding Theorems for Discrete Memoryless Systems"
by Imre Csiszár and János Körner
Second Edition
Cambridge University Press, 2011
ISBN:978-0-521-19681-9

Review by
Serdar Boztaş

August 21, 2013

1 Background on Information Theory

The modern study of transmission and storage of information through noisy communication channels goes back to Claude E. Shannon [4]. It is very rare that an intellectual discipline is initiated from scratch, and its framework fully given, as well as its many basic theorems proven by the initiator–but that's exactly what Shannon did.

This has made Shannon a very tough act to follow. By its nature, Information Theory can be very dry and terse. However, Shannon strongly believed in making the intuition behind results clear, and his writing was precise and lucid at the same time. He never sacrificed clarity at the altar of rigour, except when absolutely necessary. Over the past 60 years, the field of information theory has developed along many directions, and once engineering practice caught up with theory, the modern information society, where personal communicators such as mobile phones, tablets are commonplace would not have emerged without the research output of information theorists.

The first edition [2] of this text was very rigorous, and it was also very direct and effective in obtaining the strongest results possible, as a reward

for the rigour used. The authors introduced and used the method of types to great effect in obtaining their results in the first edition, and also stripped down the problems of information theory to their bare-bones essentials, where the combinatorial nature of many of the problems becomes very plain. A review [1] of the first edition states "Theorem 2.15 [...] is a universal version of the block source coding theorem complete with the exponentially tight bound on the optimum attainable decay with block length of the probability of ambiguous encoding. That all this happens within a scant ten pages reveals [...] the power of the authors' approach." The second edition is also written in the same rigorous fashion.

Finally, the authors are both Shannon award winners of the IEEE Information Theory Society, the highest award the society bestows on any researcher, based on the totality of their research career. They have been active contributors of major research results in information theory, over the past four decades and are uniquely qualified to write a research monograph on information theory.

2 Summary of the book

We now give a brief overview of the topics covered in the book. Firstly we note that this book is the expanded second edition of the classic published by Academic Press in 1981 [2]. The new book still has the same basic organisation into three parts, but there are two new chapters, Chapter 11 and Chapter 17, as well as some additions to the old chapters, mainly in the form of new problems. There are also historical notes on the results at the end of each chapter.

Part I: Information Measures in Simple Coding Problems This part forms the probabilistic foundations of information theory.

Chapter 1: Source coding and Hypothesis Testing introduces the basic definitions required to specify a discrete memoryless source (DMS) and proves the relationship that must be satisfied by the code rate of any source code for correct reproduction of the source with high probability. A brief discussion of the relationship between the source coding and the hypothesis testing problems is also provided.

Chapter 2: Types and typical sequences introduces the concept of types, which is central to the book, from a combinatorial point of view. It is

then proved that any subset of the set of sequences with probability at least $1-\varepsilon$ must have a large overlap with the "typical set". the set of sequences which almost surely form the output of a DMS.

Chapter 3: Formal properties of Shannon's information measures considers these measures–such as entropy, conditional entropy, mutual information, and establishes their basic properties including the chain rule, the data processing inequality, etc. The classical "Fano's inequality" is also proved here.

Chapter 4: Non-block source coding introduces non-block source coding in the context of a separable code, i.e., a prefix code.Most of the classical results on these codes, such as the Kraft's inequality are relegated to the problems following the Chapter, while much of the chapter itself focuses on proving more general results on source coding with variable symbol costs. In the simplest incarnation, if the source X has entropy $H(X)$ the rate required to encode it reliably–i.e, compress it if $H(X) < \log M$–is also $H(X)$ bits per symbol, where M is the size of the input alphabet.

Chapter 5: Blowing up lemma: a combinatorial digression discusses the topic in the title, which is a combinatorial concentration of measure result which applies to finite alphabet product spaces. Informally it says that any set A within the space, with large enough measure–i.e., constant measure instead of exponentially falling measure in terms of the dimension n–can be extended into its neighborhood by including those points which are at a small, say constant c, Hamming distance away from *some* point in A and the resulting set A_c will have measure approaching 1.

Part II: Two terminal systems Here one considers a single transmitter and a single receiver.

Chapter 6: The noisy channel coding problem proves the basic noisy channel coding theorem by proving that the maximum value of the common information between the input and output distributions is the $\varepsilon-$capacity of the discrete memoryless channel (DMC), for any $\varepsilon > 0$. This $\varepsilon-$capacity is simply the maximum data rate required if one tolerates a probability of block error equal to $\varepsilon > 0$. In this context, it is interesting to note that that in the last 2 decades or so, a number of actual constructive codes have finally been devised (Turbo codes and more recently Polar codes) and some rediscovered (LDPC codes, which date back to Gallager from the 1960's) which

approach, and in a practical sense achieve, channel capacity at practical codeword lengths.

Chapter 7: Rate-distortion tradeoff in source coding and the source-channel transmission problem focuses on the joint problem of source encoding-to remove redundancy and achieve efficiency-followed by channel coding to combat noise on the transmission channel. Rate-distortion functions encapsulate how much distortion on average results at a given transmission rate. The main result of the chapter is the source-channel coding theorem which states that the two problems can be considered separately at no loss of source fidelity or error performance.

Chapter 8: Computation of channel capacity and $\Delta-$distortion rates describes the capacity computing algorithm of Blahut and Arimoto and related algorithms, such as the $\Delta-$distortion rate computation algorithm. The main interest in these algorithms is that apart from a few special cases, it is impossible to derive channel capacities analytically, and numerical optimization must be used. These algorithms are iterative algorithms that converge to the required capacity region and are of great practical importance.

Chapter 9: A covering lemma and the error exponent in source coding generalizes and sharpens the source coding theorem by considering more general source models and an arbitrary distortion measure, and evaluates more precisely the asymptotically best code performance. The chapter makes extensive use of the technique of random coding to obtain its results. The sharpening refers to rates of convergence, as opposed to saying that an error probability goes to zero the rate is given.

Chapter 10: A packing lemma and the error exponent in channel coding revisits the coding theorem for a DMC and obtains rates of convergence of the probability of error to: (a) zero, at rates below capacity; and (b) one at rates above capacity. The point is made that the channel coding problem here is more complex than the corresponding source coding problem in Chapter 9, and has not yet been fully settled.

Chapter 11: The compound channel revisited: zero-error information theory and extremal combinatorics It may be a bit surprising that for certain channels it is possible to communicate at zero error. This topic lies at the intersection of information theory and extremal combinatorics. The

zero error capacity of a channel is simply its $\varepsilon-$capacity for $\varepsilon = 0$. Determining the asymptotics of the largest cardinality (hence allowable rate, i.e., zero error capacity, of a zero error code $C \subset X^n$ is a genuinely combinatorial and hard problem. Graph theory, graph colourings, fractional colourings and coverings, all play a part.

Chapter12: Arbitratily Varying Channels Instead of a DMC, or a compound DMC (i.e., a collection of DMCs) arbitrarily varying channels are such that the channel parameters can change from symbol interval to symbol interval in an arbitrary and unknown manner. Even then an $\varepsilon-$capacity for such a channel can be defined and analyzed with respect to either average or maximum probability of error. The two approaches will in general yield different results.

Part III: Multi-terminal systems In today's networked world, these systems are the most relevant, but also the hardest to analyze.

Chapter 13: Separate coding of correlated sources In the typical setup here, there are two sources, X and Y which are correlated. If they are jointly encoded we can use $H(X, Y)$ bits per source symbol which is strictly less then encoding them separately which requires $H(X) + H(Y)$ bits per source symbol. In the case that the side information Y is available at the decoder but not at the encoder, it is in fact possible to encode at the rate $H(X|Y)$ bits per source symbol. In the more general network setting, such as acyclic networks where sources enter at terminal nodes, the idea of "rate slicing" can be useful sometimes.

Chapter 14: Multiple-access channels are channels where different transmitters jointly access a commmon receiver, and the question is how should they encode the information to be as efficient as possible provided each transmitters information can be decoded with vanishing error rates. Capacity regions of of such channels are subsets of the positive orthant which are the convex closure of achievable rate points and have been determined for most channels of interest.

Chapter15: Entropy and image size characterization is a technical chapter which treats the problem of entropy characterization for source networks, and solves it completely for the 3-source case. The main tool is the so-called single letterization lemma, but a number of general problems are still unsolved, for more than 3 sources.

Chapter 16: Source and channel networks A number of source and channel networks which are "normal" in some sense are solved here, by utilizing the techniques from Chapter 15. Some network problems are solved for the case of more general fidelity criteria than probability of error. Note that the capacity region of the two-output broadcast channel, with say 3 input sources and a central node that transmits to two receivers, each interested in only two of the 3 input sources, is still unknown, even though it has been considered since the early 1960s.

Chapter 17: Information theoretic security describes the area of provable security (in the information theoretic sense, sometimes called perfect secrecy in the case of encryption in the traditional cryptographic literature). The two main problems considered are secure transmission over insecure channels and secret key generation taking advantage of public communications. The adversary (say eavesdropper) may obtain a noisier version of the transmitted signal, or may be restricted in some other way, and the main quantities studied are asymptotic maximum rates of secure communication and asymptotic maximum rates of key generation possible under these scenarios.

3 What is the book like (style)?

The book is written very rigorously in theorem/proof fashion with motivation mostly relegated to the historical discussions at the end of each chapter. Its coverage is as broad as any other current information theory book currently in print, but it is decidedly a book about theory not applications.

4 Would you recommend this book?

In my opinion, this is a book that can be recommended to serious researchers in information theory. I think Chapter 17 on information theoretic security should be of interest to researchers in cryptology, and provides a wealth of theorems in this area. In this area of cryptology there are gaps to be bridged between the computer science research community and the information theory research community. Some of the work covered in Chapter 17 was published in the traditional cryptography literature (for example by Maurer and Wolf [3]) but mostly it was published in the information theoretic literature.

Finally, the fact that the problems in each chapter are high level research problems helps to motivate further research in information theory.

This would be a difficult book for non-specialist in the area, or the researchers working mostly in applications of information theory. It will be a useful reference for the researcher, as well as a very good textbook for a rigorous graduate level course in the area. I recommend the book heartily to specialists and beginning researchers in the area who want to make their mark by learning the strongest techniques available.

The reviewer is an academic at RMIT University whose interests include cryptography, sequence design, and coding and information theory.

References

[1] Book Review: Information Theory:Coding Theorems for Discrete Memoryless Systems, Imre Csiszár and János Körner. *IEEE Transactions on Information Theory* 30(4):693–694, 1984.

[2] *Information Theory:Coding Theorems for Discrete Memoryless Systems*, Imre Csiszár and János Körner, Academic Press, 1981.

[3] Information-theoretic key agreement: from weak to strong secrecy for free. *EUROCRYPT 2000*, *Lecture Notes in Computer Science* 1807:351-368.

[4] C. E. Shannon, A Mathematical Theory of Communication, Bell Systems Technical Journal, 27:379–423 (Part I) and 27:623–656 (Part II), 1948.