13 Tracking

As mentioned in the previous section, adaptive filters and beamformers can be seen as devices for estimating unknown parameters. In this case, however, the parameters are constants. If the unknown parameters are time varying, the problem is one of *tracking*.

Since the estimation of N parameters requires at least N pieces of data, it is not possible to estimate more than one arbitrary time-varying parameter from a single time series. It is therefore conventional to assume that the parameters evolve in a known manner; for example, $\theta(n) = F(\theta(n-1) \mid \Phi)$, where Φ are (known) parameters of the function F. Given this model for the time evolution of the parameter, it is then possible to formulate a parameter-estimation algorithm. As with adaptive filtering and beamforming, one can take a deterministic (i.e., least-squares) approach or a Bayesian approach. In the former case one ends up with the well-known *Kalman filter*, which is optimum for linear systems and Gaussian noise. In the latter case one ends up with a more powerful algorithm but with the computational issues mentioned above.

Further Reading

Bernardo, J. M., and A. F. Smith. 2000. *Bayesian Theory*. New York: John Wiley.

Bunch, J. R., R. C. Le Borne, and I. K. Proudler. 2001. A conceptual framework for consistency, conditioning and stability issues in signal processing. *IEEE Transactions on Signal Processing* 49(9):1971–81.

Haykin, S. 2001. *Adaptive Filter Theory*, 4th edn. Englewood Cliffs, NJ: Prentice-Hall.

——. 2006. *Adaptive Radar Signal Processing*. New York: John Wiley.

McWhirter, J. G., P. D. Baxter, T. Cooper, S. Redif, and J. Foster. 2007. An EVD algorithm for para-Hermitian polynomial matrices. *IEEE Transactions on Signal Processing* 55(6):2158–69.

Proakis, J. G., C. Rader, F. Ling, and C. Nikias. 1992. *Advanced Digital Signal Processing*. London: Macmillan.

Proakis, J. G., and M. Salehi. 2007. *Digital Communications*, 5th edn. Columbus, OH: McGraw-Hill.

Rabiner, L. R., and B. Gold. 1975. *Theory and Application of Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall.

Shannon, C. E., and W. Weaver. 1949. *The Mathematical Theory of Communication*. Champaign, IL: University of Illinois Press.

Skolnik, M. I. 2002. *Introduction to Radar Systems*. Columbus, OH: McGraw-Hill.

IV.36 Information Theory

Sergio Verdú

1 "A Mathematical Theory of Communication"

Rarely does a scientific discipline owe its existence to a single paper. Authored in 1948 by Claude Shannon (1916–2001), "A mathematical theory of communication" is the Magna Carta of the information age and information theory's big bang. Using the tools of probability theory, it formulates the central optimization problems in data compression and transmission, and finds the best achievable performance in terms of the statistical description of the information sources and communication channels by way of information measures such as entropy and mutual information. After a glimpse at the state of the art as it was in 1948, we elaborate on the scope of Shannon's masterpiece in the rest of this section.

1.1 Communication Theory before the Big Bang

Motivated by the improvement in telegraphy transmission rate that could be achieved by replacing the Morse code by an optimum code, both Nyquist (1924) and Hartley (1928) recognized the need for a measure of information devoid of "psychological factors" and put forward the logarithm of the number of choices as a plausible alternative. Küpfmüller (1924), Nyquist (1928), and Kotel'nikov (1933) studied the maximum telegraph signaling speed sustainable by band-limited linear systems at a time when Fourier analysis of signals was already a standard tool in communication engineering. Inspired by the telegraph studies, Hartley put forward the notion that the "capacity of a system to carry information" is proportional to the time–bandwidth product, a notion further elaborated by Gabor (1946). However, those authors failed to grapple with the random nature of both noise and the information-carrying signals. At the same time, the idea of using mathematics to design linear filters for combatting additive noise optimally had been put to use by Kolmogorov (1941) and Wiener (1942) for minimum mean-square error estimation and by North (1943) for the detection of radar pulses.

Communication systems such as FM and PCM in the 1930s and spread spectrum in the 1940s had opened up the practical possibility of using transmission bandwidth as a design parameter that could be traded off for reproduction fidelity and robustness against noise.

1.2 The Medium

In the title of Shannon's paper, "communication" refers to

- communication across space, namely, *information-transmission* systems like radio and television broadcasting, telephone wires, coaxial cables, optical fibers, microwave links, and wireless telephony; and
- communication across time, namely, *information-storage* systems, which typically employ magnetic (tape and disks), optical (CD, DVD, and BD), and semiconductor (volatile and flash) media.

Although, at some level, all transmission and storage media involve physical continuously variable analog quantities, it is useful to model certain media such as optical disks, computer memory, or the Internet as digital media that transmit or record digital signals (zeros/ones or data packets) with a certain reliability level.

1.3 The Message

The message to be stored or transmitted may be

- analog (such as sensor readings, audio, images, video, or, in general, any message intended for the human ear/eye) or
- digital (such as text, software, or data files).

An important difference between analog and digital messages is that, since noise is unavoidable in both sensing and transmission, it is impossible to reconstruct exactly the original analog message from the recorded or transmitted information. Lossy reproduction of analog messages is therefore inevitable. Even when, as is increasingly the case, sensors of analog signals output quantized information, it is often conceptually advantageous to treat those signals as analog messages.

1.4 The Coat of Arms

Shannon's theory is a paragon of *e pluribus unum*. Indeed, despite the myriad and diversity of communication systems encompassed by information theory, its key ideas and principles are all embracing and are applicable to any of them.

Reproduced from Shannon's paper, figure 1 encompasses most cases (see section 9) of communication across time or space between one sender and one destination.

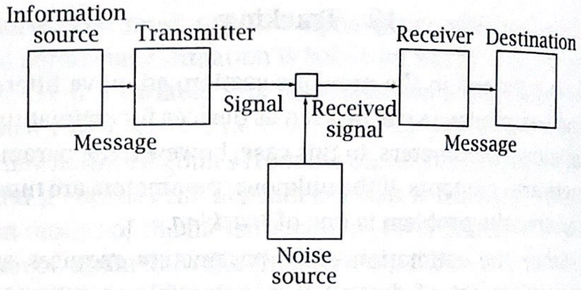

Figure 1 A schematic of a general communication system (this is figure 1 in "A mathematical theory of communication").

The purpose of the *encoder* (or *transmitter*, in figure 1) is to translate the message into a signal suitable for the transmission or storage medium. Conversely, the *decoder* (or *receiver*, in figure 1) converts the received signal into an exact or approximate replica of the original message.

The communication medium that connects the transmitter to the receiver is referred to as the *channel*. Several notable examples, classified according to the various combinations of the nature of message and medium, are listed below.

Analog message, analog medium. Radio broadcasting and long-distance telephony were the primary applications of the first analog modulation systems, such as AM, SSB, and FM, developed in the early twentieth century. With messages intended for the ear/eye and the radio frequency spectrum as the medium, all current systems for radio and television (wireless) broadcasting are also examples of this case. However, in most modern systems (such as DAB and HDTV) the transmitter and receiver perform an internal intermediate conversion to digital, for reasons that are discussed in section 4.

Analog message, digital medium. This classification includes the audio compact disc, MP3, DVD, and Voice over Internet Protocol (VoIP). So "digital audio" or "digital video" refers to the medium rather than the message.

Digital message, analog medium. The earliest examples of optical and electrical systems for the transmission of digital information were the wired telegraph systems invented in the first half of the nineteenth century, while the second half of the century saw the advent of Marconi's wireless telegraph.

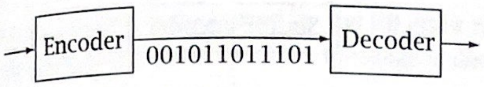

Figure 2 A data-compression system.

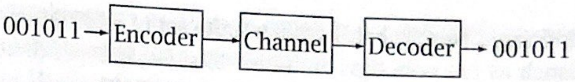

Figure 3 A data-transmission system.

Other examples developed prior to 1948 include teletype, fax, and spread spectrum. The last four decades of the twentieth century saw the development of increasingly fast general-purpose modems to transmit bit streams through analog media such as the voice-band telephone channel and radio frequency bands. Currently, modems that use optical, DSL, and CATV media to access the Internet are ubiquitous.

Digital message, digital medium. This classification includes data storage in an optical disk or flash memory.

Whether one is dealing with messages, channel inputs, or channel outputs, Shannon recognized that it is mathematically advantageous to view continuous-time analog signals as living in a finite-dimensional vector space. The simplest example is a real-valued signal of bandwidth B and (approximate) duration T, which can be viewed as a point in the Euclidean space of dimension $2BT$. To that end, Shannon gave a particularly crisp version of the sampling theorem, precursors of which had been described by E. Whittaker (1915), J. Whittaker (1929), and Kotelnikov (1933), who discovered how to interpolate losslessly the sampled values of band-limited functions.

Three special cases of figure 1, dealt with in each of the next three sections, merit particular attention.

2 Lossless Compression

Although communication across time or space is always subject to errors or failures, it is useful to consider the idealized special case of figure 1 shown in figure 2, in which there is no channel and the input to the decoder is a digital sequence equal to the encoder output. This setup, also known as *source coding*, models the paradigm of compression in which the encoder acts as the compressor and the decoder acts as the decompressor. The task of the encoder is to remove *redundancy* from the message, which can be recovered exactly or approximately at the decoder from the compressed data itself.

Lossless, or reversible, conversion is possible only if the message is digital. Morse, Huffman, TIFF, and PDF are examples of lossless compression systems, where message redundancy (unequal likelihoods of the various choices) is exploited to compact the data by assigning shorter binary strings to more likely messages. As we discuss more precisely in section 6, the goal is to obtain a compression/decompression algorithm that generates, on average, the shortest encoded version of the message.

If the source is stationary, *universal* data compressors exploit its redundancy without prior knowledge of its probabilistic law. Found in every computer operating system (e.g., ZIP), the most widely used universal data compressors were developed by Lempel and Ziv between 1976 and 1978.

3 Lossy Compression

Depending on the nature of the message, we can distinguish two types of lossy compression.

Analog-to-digital. Early examples of analog-to-digital coding (such as the vocoder and pulse-code modulation (PCM)) were developed in the 1930s. The vocoder was the precursor to the speech encoders used in cellular telephony and in VoIP, while PCM remains in widespread use in telephony and in the audio compact disc. The conceptually simplest analog-to-digital compressor, used in PCM, is the scalar quantizer, which partitions the real line in 2^k segments, each of which is assigned a unique k-bit label. JPEG [VII.7 §5] and MPEG are contemporary examples of lossy compressors for images and audio/video, respectively. Even if the inputs to those algorithms are finite-precision numbers, their signal processing treats them as real numbers.

Digital-to-digital. Even in the case of digital messages, one may be willing to tolerate a certain loss of information for the sake of economy of transmission time or storage space (e.g., when emailing a digital image or when transmitting the analog-to-digitally compressed version of a sensor reading).

4 Data Transmission

Figure 3 depicts the paradigm, also known as *channel coding*, in which the message input to the encoder is incompressible or nonredundant, in the sense that it is chosen equiprobably from a finite set of alternatives (such as fair coin flips or "pure" bits, i.e., independent binary digits equally likely to be 0 or 1). The task of the encoder is to add redundancy to the message in order to protect it from channel noise and facilitate its recovery by the decoder from the noisy channel output. In general, this is done by assigning *codewords* to each possible message, which are different enough to be distinguishable at the decoder as long as the noise is not too severe. For example, in the case of a digital medium the encoder may use an error-correcting code that appends redundant bits to the binary message string. In the case of an analog medium such as a telephone channel, the codewords are continuous-time waveforms. Based on the statistical knowledge of the channel and the codebook (assignment of messages to codewords) used by the encoder, the decoder makes an intelligent guess about the transmitted message.

Remarkably, Shannon predicted the performance of the best possible codes at a time when very few error-correcting codes were known. Hamming, a coworker at Bell Laboratories, had just invented his namesake code (see APPLIED COMBINATORICS AND GRAPH THEORY [IV.37 §4]) that appends three parity-check bits to every block of four information bits in a way that makes all sixteen codewords differ from each other in at least three positions. Therefore, the decoder can correct any single error affecting every encoded block of seven bits.

5 Compression/Transmission

Figure 4 illustrates another special case of figure 1 in which the transmitter consists of the source encoder, or compressor, followed by the channel encoder, and the receiver consists of the channel decoder followed by the source decoder, or decompressor. This architecture capitalizes on the solutions found in the special cases in sections 2, 3, and 4. To that end, in the scheme shown in figure 4 the interfaces between source and channel encoders, and between channel and source decoders, are digital regardless of the message or medium. Inspired by the teachings of information theory, in which the bit emerges as the universal currency, the modular design in figure 4 is prevalent in most modern systems for the transmission of analog messages through either digital or analog media. It allows the source encoding/decoding system to be tailored particularly to the message, disregarding the nature of the channel. Analogously, it allows the channel encoding/decoding system to be focused on the reliable transmission of nonredundant bits by combatting the channel noise disregarding the nature of the original message. In this setup, the source encoder removes redundancy from the message in a way that is tuned to the information source, while the channel encoder adds redundancy in a way that is tuned to the channel. Under widely applicable sufficient conditions, such modular design is asymptotically optimal (in the sense of section 6) in the limit in which the length of the message goes to infinity and when both source and channel operate in the ergodic regime.

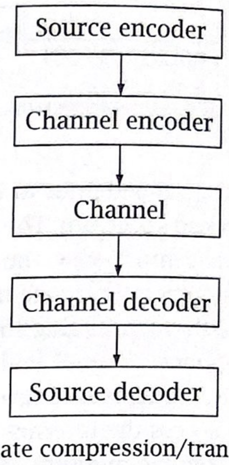

Figure 4 A separate compression/transmission system.

6 Performance Measures

The basic performance measures depend on the type of system under consideration.

Lossless compression. The compression *rate* (in bits per symbol) is the ratio of encoded bits to the number of symbols in the digital message.

Lossy compression. The quality of reproduction is measured by a distortion function of the original and reproduced signals, e.g., in the case of analog signals, the mean-square error (energy of the difference signal), and in the case of binary messages, the bit error rate. The rate (in bits per second, or per symbol) of a lossy compression system is the ratio of encoded bits to the duration of the message.

Data transmission. For a given channel and assuming that the message is incompressible, the performance of a data-transmission system is determined by the rate and the error probability. The rate (in bits per second, or per symbol) is the ratio of message duration to the time it takes to send it through the channel. Depending on the application, the reliability of the transmission is measured by the bit error rate or by the probability that the entire message is decoded correctly.

Joint compression/transmission. In the general case, the rate is measured as in the data-transmission case, with reliability measured either by a distortion measure or by the probability that the entire message is decoded correctly, depending on the nature of the message and the application.

7 Fundamental Limits

Instead of delving into the analysis and design of specific transmission systems or codes, the essence of Shannon's mathematical theory is to explore the best performance that an optimum encoder/decoder system (simply referred to as the *code*) can achieve. Information theory obtains *fundamental limits* without actually deriving the optimal codes, which are often unknown. For the three problems formulated by Shannon, the fundamental limits are as follows.

Lossless compression: the minimum achievable compression rate.

Lossy compression: the *rate-distortion function*, which is the minimum compression rate achievable as a function of the allowed average level of distortion.

Data transmission: the *channel capacity*, defined as the maximum transmission rate compatible with vanishing error probability. Capacity is often given in terms of channel parameters such as transmitted power. Before Shannon's paper, the common wisdom was that vanishing error probability would necessarily entail vanishing rate of information transmission.

The fundamental limits are very useful to the engineer because they offer a comparison of the performance of any given system with that ultimately achievable. Although, in Shannon's formulation, the growth of computational complexity as a function of the message size is not constrained in any way, decades of research on the constructive side of compression and transmission have yielded algorithms that can approach the Shannon limits with linear complexity. Often, information theory leads to valuable engineering conclusions that reveal that simple (or modular) solutions may perform at or near optimum levels. For example, as we mentioned, there is no loss in achievable performance if one follows the principle of separate compression/transmission depicted in figure 4. Fundamental limits can be, and often are, used to sidestep the need for cumbersome analysis in order to debunk performance claims made for a given system.

The fundamental limits turn out to depend crucially on the duration of the message. Since Shannon's 1948 paper, information theory has focused primarily, but not exclusively, on the fundamental limits in the regime of asymptotically long messages. By their very nature, the fundamental limits for a given source or channel are not technology dependent, and they do not become obsolete with improvements in hardware/software. On the contrary, technological advances pave the way for the design of coding systems that approach the ideal fundamental limits increasingly closely. Although the optimum compression and transmission systems are usually unknown, the methods of proof of the fundamental limits often suggest features that near-optimum practical communication systems ought to have, thereby offering design guidelines to approach the fundamental limits. Shannon's original proof of his channel coding theorem was one of the first nontrivial instances of the *probabilistic method*, now widely used in discrete mathematics; to show the existence of an object that satisfies a certain property it is enough to find a probability distribution on the set of all objects such that those satisfying the property have nonzero probability. In his proof, Shannon computed an upper bound to the error probability averaged with respect to an adequately chosen distribution on the set of all codes; at least one code must have error probability not exceeding the bound.

8 Information Measures

The fundamental performance limits turn out to be given in terms of so-called information measures, which have units such as bits. In this section we list the three most important information measures.

Entropy: a measure of the randomness of a discrete distribution P_X defined on a finite or countably infinite alphabet $\mathcal{A}$, defined as

$$H(X) = \sum_{a \in \mathcal{A}} P_X(a) \log \left(\frac{1}{P_X(a)} \right).$$

In the limit as $n \to \infty$, a stationary ergodic random source $(X_1, \dots, X_n)$ can be losslessly encoded at its entropy rate

$$\lim_{n \to \infty} \frac{1}{n} H(X_1, \dots, X_n),$$

a limit that is easy to compute in the case of Markov chains. In the simplest case, asymptotically, n flips of a coin with bias p can be compressed losslessly at any rate exceeding $h(p)$ bits per coin flip with

$$h(p) = p \log \frac{1}{p} + (1-p) \log \left(\frac{1}{1-p}\right),$$

which is the entropy of the biased coin source. The ubiquitous linear-time Lempel–Ziv universal data-compression algorithms are able to achieve, asymptotically, the entropy rate of ergodic stationary sources. Therefore, at least in the long run, universality incurs no penalty.

Relative entropy: a measure of the dissimilarity between two distributions P and Q defined on the same measurable space $(\mathcal{A}, \mathcal{F})$, defined as

$$D(P\|Q) = \int \log \left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right) \mathrm{d}P.$$

Relative entropy plays a central role not only in information theory but also in the analysis of the ability to discriminate between data models, and in particular in large-deviation results, which explore the exponential decrease (in the number of observations) of the probability of very unlikely events. Specifically, if n independent data samples are generated with probability distribution Q, the probability that they will appear to be generated from a distribution in some class $\mathcal{P}$ behaves as

$$\exp\left(-n \inf_{P \in \mathcal{P}} D(P\|Q)\right).$$

Relative entropy was introduced by Kullback and Leibler in 1951 with the primary goal of extending Shannon's measure of information to nondiscrete cases.

Mutual information: a measure of the dependence between two (not necessarily discrete) random variables X and Y given by the relative entropy between the joint measure and the product of the marginal measures:

$$I(X;Y) = D(P_{XY} \| P_X \times P_Y).$$

Note that $I(X;X) = H(X)$ if X is discrete.

For stationary channels that behave ergodically, the channel capacity is given by

$$C = \lim_{n\to\infty} \frac{1}{n} \max I(X_1, \dots, X_n; Y_1, \dots, Y_n),$$

where the maximum is over all joint distributions of $(X_1, \dots, X_n)$, and $(Y_1, \dots, Y_n)$ are the channel responses to $(X_1, \dots, X_n)$. If the channel is stationary memoryless, then the formula boils down to

$$C = \max I(X;Y).$$

The capacity of a channel that erases a fraction δ of the codeword symbols (drawn from an alphabet $\mathcal{A}$) is

$$C = (1-\delta) \log |\mathcal{A}|,$$

as long as the location of the erased symbols is known to the decoder and the nonerased symbols are received error free. In the case of a binary channel that introduces errors independently with probability δ, the capacity is given by

$$C = 1 - h(\delta),$$

while in the case of a continuous-time additive Gaussian noise channel with bandwidth B, transmission power P, and noise strength N, the capacity is

$$C = B \log \left(1 + \frac{P}{BN}\right) \text{ bits per second},$$

a formula that dispels the pre-1948 notion that the information-carrying capacity of a communication channel is proportional to its bandwidth and that is reminiscent of the fact that in a cellular phone the stronger the received signal the faster the download. In lossy data compression of a stationary ergodic source $(X_1, X_2, \dots)$, the rate compatible with a given per-sample distortion level d under a distortion measure $d \colon \mathcal{A}^2 \to [0, \infty]$ is given by

$$R(d) = \lim_{n\to\infty} \frac{1}{n} \min I(X_1, \dots, X_n; Y_1, \dots, Y_n),$$

where the minimum is taken over the joint distribution of source X^n and reproduction Y^n, with given P_{X^n}, and such that

$$\frac{1}{n} \sum_{i=1}^{n} d(X_i, Y_i) \leqslant d.$$

For stationary memoryless sources, just as for capacity we obtain a "single-letter" expression $R(d) = \min I(X;Y)$.

It should be emphasized that the central concern of information theory is not the definition of information measures but the theorems that use them to describe the fundamental limits of compression and transmission. However, it is rewarding that entropy, mutual information, and relative information, as well as other related measures, have found applications in many fields beyond communication theory, including probability theory, statistical inference, ergodic theory, computer science, physics, economics, life sciences, and linguistics.

9 Beyond Figure 1

Work on the basic paradigm in figure 1 continues to this day, not only to tackle source and channel models inspired by new applications and technologies but in furthering the basic understanding of the capabilities of coding systems, particularly in the nonasymptotic regime. However, in order to analyze models of interest in practice, many different setups have been studied since 1948 that go beyond the original. We list a few of the ones that have received the most attention.

Feedback. A common feature of many communication links is the availability of another communication channel from receiver to transmitter. In what way can knowledge of the channel output aid the transmitter in a more efficient selection of codewords? In 1956 Shannon showed that, in the absence of channel memory, capacity does not increase even if the encoder knows the channel output instantaneously and noiselessly. Nevertheless, feedback can be quite useful to improve transmission rate in the nonasymptotic regime and in the presence of channel memory.

Separate compression of dependent sources of information. Suppose that there is one decompressor that receives the encoded versions of several sources produced by individual compressors. If, instead, a single compressor had access to all the sources, it could exploit the statistical dependence among them to encode at a rate equal to the overall entropy. Surprisingly, in 1973 Slepian and Wolf showed that even in the completely decentralized setup the sum of the encoded rates can be as low as in the centralized setting and still the decompressor is able to correctly decode with probability approaching 1. In the lossy setting the corresponding problem is not yet completely solved.

Multiple-access channel. If, as in the case of a cellular wireless telephony system, a single receiver obtains a signal with mutually interfering encoded streams produced by several transmitters, there is a trade-off among the achievable rates. The channel capacity is no longer a scalar but a *capacity region*.

Interference channel. As in the case of a wired telephone system subject to crosstalk, in this model there is a receiver for each transmitter, and the signal it receives not only contains the information transmitted by the desired user but is contaminated by the signals of all other users. It does not reduce to a special case of the multiple-access setup because each receiver is required to decode reliably only the message of its desired user.

Broadcast channel. A single transmitter sends a codeword, which is received by several geographically separated receivers. Each receiver is therefore connected to the transmitter by a different communication channel, but all those channels share the same input. If the broadcaster intends to send different messages to the various destinations, there is again a trade-off among the achievable rates.

Relay channel. The receiver obtains both a signal from the transmitter and a signal from a relay, which itself is allowed to process the signal it receives from the transmitter in any way it wants. In particular, the relay need not be able to fully understand the message sent by the transmitter.

Inspired by various information technologies, a number of information-theoretic problems have arisen that go beyond issues of eliminating redundancy (for compression) or adding redundancy (for transmission in the presence of noise). Some examples follow.

Secrecy. Simultaneously with communication theory, Shannon established the basic mathematical theory of cryptography and showed that iron-clad privacy requires that the length of the encryption key be as long as that of the message. Most modern cryptographic algorithms do not provide that level of security; they rely on the fact that certain computational problems, such as integer factorization, are believed to be inherently hard. A provable level of security is available using an information-theoretic approach pioneered by Wyner (1975), which guarantees that the eavesdropper obtains a negligible amount of information about the message.

Random number generation for system simulation. Random processes with prescribed distributions can be generated by a deterministic algorithm driven by a source of random bits. A key quantity that quantifies the "complexity" of the generated random process is the minimal rate of the source of random bits necessary to accomplish the task. The *resolvability* of a system is defined as the minimal randomness required to generate any desired input so that the output distributions are approximated with arbitrary accuracy. In 1993 Han and Verdú showed that the resolvability of a system is equal to its channel capacity.

Minimum description length. In the 1960s Kolmogorov and others took a nonprobabilistic approach to the compression of a message, which, like universal lossless compression, uses no prior knowledge: the algorithmic complexity of the message is the length of the shortest program that will output the message. Although this notion is useful only asymptotically, it has important links with information theory and has had an impact in statistical inference, primarily through the minimum description length statistical modeling principle put forward by Rissanen in 1978: the message is compressed according to a certain distribution, which is chosen from a predetermined model class and is also communicated to the decompressor. The distribution is chosen so that the sum of the lengths of its description and the compressed version of the message are minimized.

Inequalities and convex analysis. A principle satisfied by information measures is that processing cannot increase either the dependence between input and output as measured by mutual information or the relative entropy between any pair of distributions governing the input of the processor. Mathematically, the nonnegativity of relative entropy and those *data processing principles* are translated into convex inequalities, which have been used successfully in the rederivation of various inequalities, such as those of Hadamard and Brunn-Minkowski, and in the discovery of new inequalities.

Portfolio theory. One possible approach to PORTFOLIO SELECTION [V.10] (for a given number of stocks) is to choose the *log-optimal* portfolio, which maximizes the asymptotic appreciation growth rate. When their distribution is known, a simplistic model of independent identically distributed stock prices leads to limiting results with a strong information-theoretic flavor. Just as in data compression, under assumptions of stationarity and ergodicity, it is possible to deal with more realistic scenarios in which the distribution is not known a priori and the stock prices are interdependent.

Identification. Suppose that the transmitter sends the identity of an addressee to a multitude of possible users. Each user is interested only in finding out whether it is indeed the addressee or not. Allowing, as usual, a certain error probability, this setup can be captured as in figure 1, except that the decoder is free to declare a list of several messages (addresses) to be simultaneously "true." Each user simply checks whether its identity is in the list or not. How many messages can be transmitted while guaranteeing vanishing probability of erroneous information? The surprising answer found by Ahlswede and Dueck in 1989 is that the number of addresses grows doubly exponentially with the number of channel uses. Moreover, the second-order exponent is equal to the channel capacity.

Finally, we mention the discipline of *quantum information theory*, which deals with the counterparts of the fundamental limits discussed above for quantum mechanical models of sources and channels. Probability measures, conditional probabilities, and bits translate into density matrices, self-adjoint linear operators, and qubits. The quantum channel coding theorem was proved by Holevo in 1973, while the quantum source coding theorem was proved by Schumacher in 1995.

Further Reading

Cover, T. M., and J. Thomas. 2006. *Elements of Information Theory*, 2nd edn. New York: Wiley Interscience.

Shannon, C. E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27:379–423, 623–56.

Verdú, S. 1998. Fifty years of Shannon theory. *IEEE Transactions on Information Theory* 44(6):2057–78.

IV.37 Applied Combinatorics and Graph Theory

Peter Winkler

1 Introduction

Combinatorics and graph theory are the cornerstones of *discrete mathematics*, which has seen an explosion of activity since the middle of the twentieth century. The main reason for this explosion is the plethora of applications in a world where digital (as opposed to analog) computing has become the norm. Once considered more "recreational" than serious, combinatorics and graph theory now boast many fundamental and useful results, adding up to a cogent theory. Our objective here is to present the most elementary of these results in a format useful to those who may run into combinatorial problems in applications but have not studied combinatorics or graph theory.

Accordingly, we will begin each section with a (not necessarily serious, but representative) problem, introducing the basic techniques, algorithms, and theorems of combinatorics and graph theory in response.

We will assume basic familiarity with mathematics but none with computer science. Proofs, sometimes informal, are included when they are useful and short.