Asymptotic Estimates in Information Theory with Non-Vanishing Error Probabilities

Vincent Y. F. Tan
Department of Electrical and Computer Engineering
Department of Mathematics
National University of Singapore
Singapore 119077
Email: vtan@nus.edu.sg

the essence of knowledge
Boston — Delft

Foundations and Trends® in Communications and Information Theory

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

V. Y. F. Tan. *Asymptotic Estimates in Information Theory with Non-Vanishing Error Probabilities*. Foundations and Trends® in Communications and Information Theory, vol. 11, no. 1-2, pp. 1–184, 2014.

This Foundations and Trends® issue was typeset in LaTeX using a class file designed by Neal Parikh. Printed on acid-free paper.

ISBN: 978-1-60198-853-9

Foundations and Trends® in Communications and Information Theory

Volume 11, Issue 1-2, 2014

Editorial Scope

Topics

Foundations and Trends® in Communications and Information Theory publishes survey and tutorial articles in the following topics:

- Coded modulation
- Coding theory and practice
- Communication complexity
- Communication system design
- Cryptology and data security
- Data compression
- Data networks
- Demodulation and Equalization
- Denoising
- Detection and estimation
- Information theory and statistics
- Information theory and computer science
- Joint source/channel coding
- Modulation and signal design
- Multiuser detection
- Multiuser information theory
- Optical communication channels
- Pattern recognition and learning
- Quantization
- Quantum information processing
- Rate-distortion theory
- Shannon theory
- Signal processing for communications
- Source coding
- Storage and recording codes
- Speech and Image Compression
- Wireless Communications

Information for Librarians

Foundations and Trends® in Communications and Information Theory, 2014, Volume 11, 4 issues. ISSN paper version 1567-2190. ISSN online version 1567-2328. Also available as a combined paper and online subscription.

Foundations and Trends® in Communications and
Information Theory
Vol. 11, No. 1-2 (2014) 1–184

DOI: 10.1561/0100000086

Asymptotic Estimates in Information Theory with Non-Vanishing Error Probabilities

Vincent Y. F. Tan
Department of Electrical and Computer Engineering
Department of Mathematics
National University of Singapore
Singapore 119077
Email: vtan@nus.edu.sg

Contents

Abstract

This monograph presents a unified treatment of single- and multi-user problems in Shannon's information theory where we depart from the requirement that the error probability decays asymptotically in the blocklength. Instead, the error probabilities for various problems are bounded above by a non-vanishing constant and the spotlight is shone on achievable coding rates as functions of the growing blocklengths. This represents the study of *asymptotic estimates with non-vanishing error probabilities*.

In Part I, after reviewing the fundamentals of information theory, we discuss Strassen's seminal result for binary hypothesis testing where the type-I error probability is non-vanishing and the rate of decay of the type-II error probability with growing number of independent observations is characterized. In Part II, we use this basic hypothesis testing result to develop second- and sometimes, even third-order asymptotic expansions for point-to-point communication. Finally in Part III, we consider network information theory problems for which the second-order asymptotics are known. These problems include some classes of channels with random state, the multiple-encoder distributed lossless source coding (Slepian-Wolf) problem and special cases of the Gaussian interference and multiple-access channels. Finally, we discuss avenues for further research.

V. Y. F. Tan. *Asymptotic Estimates in Information Theory with Non-Vanishing Error Probabilities*. Foundations and Trends® in Communications and Information Theory, vol. 11, no. 1-2, pp. 1–184, 2014.

DOI: 10.1561/0100000086.

Part I

Fundamentals

1

Introduction

Claude E. Shannon's epochal *"A Mathematical Theory of Communication"* [141] marks the dawn of the digital age. In his seminal paper, Shannon laid the theoretical and mathematical foundations for the basis of all communication systems today. It is not an exaggeration to say that his work has had a tremendous impact in communications engineering and beyond, in fields as diverse as statistics, economics, biology and cryptography, just to name a few.

It has been more than 65 years since Shannon's landmark work was published. Along with impressive research advances in the field of *information theory*, numerous excellent books on various aspects of the subject have been written. The author's favorites include Cover and Thomas [33], Gallager [56], Csiszár and Körner [39], Han [67], Yeung [189] and El Gamal and Kim [49]. Is there sufficient motivation to consolidate and present another aspect of information theory systematically? It is the author's hope that the answer is in the affirmative.

To motivate why this is so, let us recapitulate two of Shannon's major contributions in his 1948 paper. First, Shannon showed that to *reliably* compress a discrete memoryless source (DMS) $X^n = (X_1, \ldots, X_n)$ where each X_i has the same distribution as a common random variable X, it is sufficient to use $H(X)$ bits per source symbol in the limit of large blocklengths n, where $H(X)$ is the Shannon entropy of the source. By *reliable*, it is meant that the probability of incorrect decoding of the source sequence tends to zero as the blocklength n grows. Second, Shannon showed that it is possible to *reliably* transmit a message $M \in \{1, \ldots, 2^{nR}\}$ over a discrete memoryless channel (DMC) W as long as the message rate R is smaller than the capacity of the channel $C(W)$. Similarly to the source compression scenario, by *reliable*, one means that the probability of incorrectly decoding M tends to zero as n grows.

There is, however, substantial motivation to revisit the criterion of having error probabilities vanish asymptotically. To state Shannon's source compression result more formally, let us define $M^*(P^n, \varepsilon)$ to be the minimum code size for which the length-n DMS P^n is compressible to within an error probability $\varepsilon \in (0, 1)$. Then, Theorem 3 of Shannon's paper [141], together with the strong converse for lossless source coding [49, Ex. 3.15], states that

$$\lim_{n\to\infty} \frac{1}{n} \log M^*(P^n, \varepsilon) = H(X), \quad \text{bits per source symbol.} \tag{1.1}$$

Similarly, denoting $M^*_{\mathrm{ave}}(W^n, \varepsilon)$ as the maximum code size for which it is possible to communicate over a DMC W^n such that the average error probability is no larger than ε, Theorem 11 of Shannon's paper [141], together with the strong converse for channel coding [180, Thm. 2], states that

$$\lim_{n\to\infty} \frac{1}{n} \log M^*_{\mathrm{ave}}(W^n, \varepsilon) = C(W), \quad \text{bits per channel use.} \tag{1.2}$$

In many practical communication settings, one does not have the luxury of being able to design an arbitrarily long code, so one must settle for a non-vanishing, and hence finite, error probability ε. In this *finite blocklength* and *non-vanishing error probability* setting, how close can one hope to get to the asymptotic limits $H(X)$ and $C(W)$? This is, in general a difficult question because exact evaluations of $\log M^*(P^n, \varepsilon)$ and $\log M^*_{\mathrm{ave}}(W^n, \varepsilon)$ are intractable, apart from a few special sources and channels.

In the early years of information theory, Dobrushin [45], Kemperman [91] and, most prominently, Strassen [152] studied approximations to $\log M^*(P^n, \varepsilon)$ and $\log M^*_{\text{ave}}(W^n, \varepsilon)$. These beautiful works were largely forgotten until recently, when interest in so-called *Gaussian approximations* were revived by Hayashi [75, 76] and Polyanskiy-Poor-Verdú [122, 123].[1] Strassen showed that the limiting statement in (1.1) may be refined to yield the *asymptotic expansion*

$$\log M^*(P^n, \varepsilon) = nH(X) - \sqrt{nV(X)}\Phi^{-1}(\varepsilon) - \frac{1}{2}\log n + O(1), \quad (1.3)$$

where $V(X)$ is known as the *source dispersion* or the *varentropy*, terms introduced by Kostina-Verdú [97] and Kontoyiannis-Verdú [95]. In (1.3), Φ^{-1} is the inverse of the Gaussian cumulative distribution function. Observe that the first-order term in the asymptotic expansion above, namely $H(X)$, coincides with the (first-order) fundamental limit shown by Shannon. From this expansion, one sees that if the error probability is fixed to $\varepsilon < \frac{1}{2}$, the extra rate above the entropy we have to pay for operating at finite blocklength n with admissible error probability ε is approximately $\sqrt{V(X)/n}\,\Phi^{-1}(1-\varepsilon)$. Thus, the quantity $V(X)$, which is a function of P just like the entropy $H(X)$, quantifies how fast the rates of optimal source codes converge to $H(X)$. Similarly, for well-behaved DMCs, under mild conditions, Strassen showed that the limiting statement in (1.2) may be refined to

$$\log M^*_{\text{ave}}(W^n, \varepsilon) = nC(W) + \sqrt{nV_\varepsilon(W)}\Phi^{-1}(\varepsilon) + O(\log n) \quad (1.4)$$

and $V_\varepsilon(W)$ is a channel parameter known as the *ε-channel dispersion*, a term introduced by Polyanskiy-Poor-Verdú [123]. Thus the backoff from capacity at finite blocklengths n and average error probability ε is approximately $\sqrt{V_\varepsilon(W)/n}\,\Phi^{-1}(1-\varepsilon)$.

1.1 Motivation for this Monograph

It turns out that Gaussian approximations (first two terms of (1.3) and (1.4)) are *good proxies* to the true non-asymptotic fundamental limits ($\log M^*(P^n, \varepsilon)$ and $\log M^*_{\text{ave}}(W^n, \varepsilon)$) at moderate blocklengths and

[1] Some of the results in [122, 123] were already announced by S. Verdú in his Shannon lecture at the 2007 International Symposium on Information Theory (ISIT) in Nice, France.

moderate error probabilities for some channels and sources as shown by Polyanskiy-Poor-Verdú [123] and Kostina-Verdú [97]. For error probabilities that are not too small (e.g., $\varepsilon \in [10^{-6}, 10^{-3}]$), the Gaussian approximation is often better than that provided by traditional *error exponent* or *reliability function* analysis [39, 56], where the code rate is fixed (below the first-order fundamental limit) and the exponential decay of the error probability is analyzed. Recent refinements to error exponent analysis using exact asymptotics [10, 11, 135] or saddlepoint approximations [137] are alternative proxies to the non-asymptotic fundamental limits. The accuracy of the Gaussian approximation in *practical* regimes of errors and finite blocklengths gives us motivation to study refinements to the first-order fundamental limits of other single- and multi-user problems in Shannon theory.

The study of *asymptotic estimates with non-vanishing error probabilities*—or more succinctly, *fixed error asymptotics*—also uncovers several interesting phenomena that are not observable from studies of first-order fundamental limits in single- and multi-user information theory [33, 49]. This analysis may give engineers deeper insight into the design of practical communication systems. A non-exhaustive list includes:

1. Shannon showed that *separating* the tasks of source and channel coding is optimal rate-wise [141]. As we see in Section 4.5.2 (and similarly to the case of error exponents [35]), this is not the case when the probability of excess distortion of the source is allowed to be non-vanishing.

2. Shannon showed that feedback does not increase the capacity of a DMC [142]. It is known, however, that variable-length feedback [125] and full output feedback [8] improve on the fixed error asymptotics of DMCs.

3. It is known that the entropy can be achieved *universally* for fixed-to-variable length almost lossless source coding of a DMS [192], i.e., the source statistics do not have to be known. The redundancy has also been studied for prefix-free codes [27]. In the fixed error setting (a setting complementary to [27]), it was shown by

Kosut and Sankar [100, 101] that universality imposes a penalty in the *third-order* term of the asymptotic expansion in (1.3).

4. Han showed that the output from any source encoder at the optimal coding rate with asymptotically vanishing error appears almost completely random [68]. This is the so-called *folklore theorem*. Hayashi [75] showed that the analogue of the folklore theorem does not hold when we consider the second-order terms in asymptotic expansions (i.e., the second-order asymptotics).

5. Slepian and Wolf showed that separate encoding of two correlated sources incurs no loss rate-wise compared to the situation where side information is also available at all encoders [151]. As we shall see in Chapter 6, the fixed error asymptotics in the vicinity of a corner point of the polygonal Slepian-Wolf region suggests that side-information at the encoders may be beneficial.

None of the aforementioned books [33, 39, 49, 56, 67, 189] focus exclusively on the situation where the error probabilities of various Shannon-theoretic problems are upper bounded by $\varepsilon \in (0,1)$ and asymptotic expansions or second-order terms are sought. This is what this monograph attempts to do.

1.2 Preview of this Monograph

This monograph is organized as follows: In the remaining parts of this chapter, we recap some quantities in information theory and results in the *method of types* [37, 39, 74], a particularly useful tool for the study of discrete memoryless systems. We also mention some probability bounds that will be used throughout the monograph. Most of these bounds are based on refinements of the central limit theorem, and are collectively known as *Berry-Esseen theorems* [17, 52]. In Chapter 2, our study of asymptotic expansions of the form (1.3) and (1.4) begins in earnest by revisiting Strassen's work [152] on binary hypothesis testing where the probability of false alarm is constrained to not exceed a positive constant. We find it useful to revisit the fundamentals of hypothesis testing as many information-theoretic problems such as source and channel coding are intimately related to hypothesis testing.

Part II of this monograph begins our study of information-theoretic problems starting with lossless and lossy compression in Chapter 3. We emphasize, in the first part of this chapter, that (fixed-to-fixed length) lossless source coding and binary hypothesis testing are, in fact, the same problem, and so the asymptotic expansions developed in Chapter 2 may be directly employed for the purpose of lossless source coding. Lossy source coding, however, is more involved. We review the recent works in [86] and [97], where the authors independently derived asymptotic expansions for the logarithm of the minimum size of a source code that reproduces symbols up to a certain distortion, with some admissible probability of excess distortion. Channel coding is discussed in Chapter 4. In particular, we study the approximation in (1.4) for both discrete memoryless and Gaussian channels. We make it a point here to be precise about the third-order $O(\log n)$ term. We state conditions on the channel under which the coefficient of the $O(\log n)$ term can be determined exactly. This leads to some new insights concerning optimum codes for the channel coding problem. Finally, we marry source and channel coding in the study of source-channel transmission where the probability of excess distortion in reproducing the source is non-vanishing.

Part III of this monograph contains a sparse sampling of fixed error asymptotic results in network information theory. The problems we discuss here have conclusive second-order asymptotic characterizations (analogous to the second terms in the asymptotic expansions in (1.3) and (1.4)). They include some channels with random state (Chapter 5), such as Costa's writing on dirty paper [30], mixed DMCs [67, Sec. 3.3], and quasi-static single-input-multiple-output (SIMO) fading channels [18]. Under the fixed error setup, we also consider the second-order asymptotics of the Slepian-Wolf [151] distributed lossless source coding problem (Chapter 6), the Gaussian interference channel (IC) in the strictly very strong interference regime [22] (Chapter 7), and the Gaussian multiple access channel (MAC) with degraded message sets (Chapter 8). The MAC with degraded message sets is also known as the *cognitive* [44] or *asymmetric* [72, 167, 128] MAC (A-MAC). Chapter 9

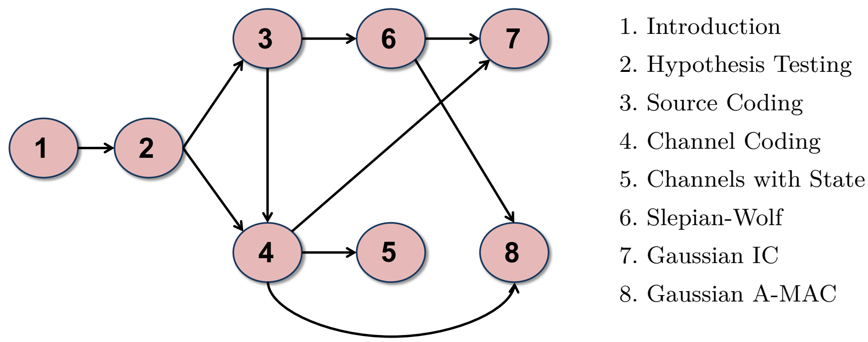

Figure 1.1: Dependence graph of the chapters in this monograph. An arrow from node s to t means that results and techniques in Chapter s are required to understand the material in Chapter t.

concludes with a brief summary of other results, together with open problems in this area of research. A dependence graph of the chapters in the monograph is shown in Fig. 1.1.

This area of information theory—*fixed error asymptotics*—is vast and, at the same time, rapidly expanding. The results described herein are not meant to be exhaustive and were somewhat dependent on the author's understanding of the subject and his preferences at the time of writing. However, the author has made it a point to ensure that results herein are *conclusive* in nature. This means that the problem is *solved* in the information-theoretic sense in that an operational quantity is *equated* to an information quantity. In terms of asymptotic expansions such as (1.3) and (1.4), by *solved*, we mean that either the second-order term is known or, better still, both the second- and third-order terms are known. Having articulated this, the author confesses that there are many relevant information-theoretic problems that can be considered solved in the fixed error setting, but have not found their way into this monograph either due to space constraints or because it was difficult to meld them seamlessly with the rest of the story.

1.3 Fundamentals of Information Theory

In this section, we review some basic information-theoretic quantities. As with every article published in the *Foundations and Trends in Communications and Information Theory*, the reader is expected to have some background in information theory. Nevertheless, the only prerequisite required to appreciate this monograph is information theory at the level of Cover and Thomas [33]. We will also make extensive use of the method of types, for which excellent expositions can be found in [37, 39, 74]. The measure-theoretic foundations of probability will not be needed to keep the exposition accessible to as wide an audience as possible.

1.3.1 Notation

The notation we use is reasonably standard and generally follows the books by Csiszár-Körner [39] and Han [67]. Random variables (e.g., X) and their realizations (e.g., x) are in upper and lower case respectively. Random variables that take on finitely many values have alphabets (support) that are denoted by calligraphic font (e.g., $\mathcal{X}$). The cardinality of the finite set $\mathcal{X}$ is denoted as $|\mathcal{X}|$. Let the random vector X^n be the vector of random variables $(X_1, \ldots, X_n)$. We use bold face $\mathbf{x} = (x_1, \ldots, x_n)$ to denote a realization of X^n. The set of all distributions (probability mass functions) supported on alphabet $\mathcal{X}$ is denoted as $\mathscr{P}(\mathcal{X})$. The set of all conditional distributions (i.e., channels) with the input alphabet $\mathcal{X}$ and the output alphabet $\mathcal{Y}$ is denoted by $\mathscr{P}(\mathcal{Y}|\mathcal{X})$. The joint distribution induced by a marginal distribution $P \in \mathscr{P}(\mathcal{X})$ and a channel $V \in \mathscr{P}(\mathcal{Y}|\mathcal{X})$ is denoted as $P \times V$, i.e.,

$$(P \times V)(x, y) := P(x)V(y|x). \tag{1.5}$$

The marginal output distribution induced by P and V is denoted as PV, i.e.,

$$PV(y) := \sum_{x \in \mathcal{X}} P(x)V(y|x). \tag{1.6}$$

If X has distribution P, we sometimes write this as $X \sim P$.

Vectors are indicated in lower case bold face (e.g., $\mathbf{a}$) and matrices in upper case bold face (e.g., $\mathbf{A}$). If we write $\mathbf{a} \geq \mathbf{b}$ for two vectors $\mathbf{a}$

and $\mathbf{b}$ of the same length, we mean that $a_j \geq b_j$ for every coordinate j. The transpose of $\mathbf{A}$ is denoted as $\mathbf{A}'$. The vector of all zeros and the identity matrix are denoted as $\mathbf{0}$ and $\mathbf{I}$ respectively. We sometimes make the lengths and sizes explicit. The ℓ_q-norm (for $q \geq 1$) of a vector $\mathbf{v} = (v_1, \ldots, v_k)$ is denoted as $\|\mathbf{v}\|_q := (\sum_{i=1}^k |v_i|^q)^{1/q}$.

We use standard asymptotic notation [29] in this monograph: $a_n \in O(b_n)$ if and only if (iff) $\limsup_{n\to\infty} |a_n/b_n| < \infty$; $a_n \in \Omega(b_n)$ iff $b_n \in O(a_n)$; $a_n \in \Theta(b_n)$ iff $a_n \in O(b_n) \cap \Omega(b_n)$; $a_n \in o(b_n)$ iff $\limsup_{n\to\infty} |a_n/b_n| = 0$; and $a_n \in \omega(b_n)$ iff $\liminf_{n\to\infty} |a_n/b_n| = \infty$. Finally, $a_n \sim b_n$ iff $\lim_{n\to\infty} a_n/b_n = 1$.

1.3.2 Information-Theoretic Quantities

Information-theoretic quantities are denoted in the usual way [39, 49]. All logarithms and exponential functions are to the base 2. The *entropy* of a discrete random variable X with probability distribution $P \in \mathscr{P}(\mathcal{X})$ is denoted as

$$H(X) = H(P) := -\sum_{x\in\mathcal{X}} P(x)\log P(x). \tag{1.7}$$

For the sake of clarity, we will sometimes make the dependence on the distribution P explicit. Similarly given a pair of random variables (X, Y) with joint distribution $P \times V \in \mathscr{P}(\mathcal{X} \times \mathcal{Y})$, the *conditional entropy* of Y given X is written as

$$H(Y|X) = H(V|P) := -\sum_{x\in\mathcal{X}} P(x)\sum_{y\in\mathcal{Y}} V(y|x)\log V(y|x). \tag{1.8}$$

The *joint entropy* is denoted as

$$H(X, Y) := H(X) + H(Y|X), \quad \text{or} \tag{1.9}$$

$$H(P \times V) := H(P) + H(V|P). \tag{1.10}$$

The *mutual information* is a measure of the correlation or dependence between random variables X and Y. It is interchangeably denoted as

$$I(X;Y) := H(Y) - H(Y|X), \quad \text{or} \tag{1.11}$$

$$I(P, V) := H(PV) - H(V|P). \tag{1.12}$$

Given three random variables (X, Y, Z) with joint distribution $P \times V \times W$ where $V \in \mathscr{P}(\mathcal{Y}|\mathcal{X})$ and $W \in \mathscr{P}(\mathcal{Z}|\mathcal{X} \times \mathcal{Y})$, the *conditional mutual information* is

$$I(Y; Z|X) := H(Z|X) - H(Z|XY), \quad \text{or} \tag{1.13}$$

$$I(V, W|P) := \sum_{x \in \mathcal{X}} P(x) I\big(V(\cdot|x), W(\cdot|x, \cdot)\big). \tag{1.14}$$

A particularly important quantity is the *relative entropy* (or *Kullback-Leibler divergence* [102]) between P and Q which are distributions on the same finite support set $\mathcal{X}$. It is defined as the expectation with respect to P of the log-likelihood ratio $\log \frac{P(x)}{Q(x)}$, i.e.,

$$D(P\|Q) := \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}. \tag{1.15}$$

Note that if there exists an $x \in \mathcal{X}$ for which $Q(x) = 0$ while $P(x) > 0$, then the relative entropy $D(P\|Q) = \infty$. If for every $x \in \mathcal{X}$, if $Q(x) = 0$ then $P(x) = 0$, we say that P is absolutely continuous with respect to Q and denote this relation by $P \ll Q$. In this case, the relative entropy is finite. It is well known that $D(P\|Q) \geq 0$ and equality holds if and only if $P = Q$. Additionally, the *conditional relative entropy* between $V, W \in \mathscr{P}(\mathcal{Y}|\mathcal{X})$ given $P \in \mathscr{P}(\mathcal{X})$ is defined as

$$D(V\|W|P) := \sum_{x \in \mathcal{X}} P(x) D\big(V(\cdot|x)\|W(\cdot|x)\big). \tag{1.16}$$

The mutual information is a special case of the relative entropy. In particular, we have

$$I(P, V) = D(P \times V\|P \times PV) = D(V\|PV|P). \tag{1.17}$$

Furthermore, if $U_{\mathcal{X}}$ is the uniform distribution on $\mathcal{X}$, i.e., $U_{\mathcal{X}}(x) = 1/|\mathcal{X}|$ for all $x \in \mathcal{X}$, we have

$$D(P\|U_{\mathcal{X}}) = -H(P) + \log|\mathcal{X}|. \tag{1.18}$$

The definition of relative entropy $D(P\|Q)$ can be extended to the case where Q is not necessarily a probability measure. In this case non-negativity does not hold in general. An important property we exploit is the following: If μ denotes the counting measure (i.e., $\mu(\mathcal{A}) = |\mathcal{A}|$ for $\mathcal{A} \subset \mathcal{X}$), then similarly to (1.18)

$$D(P\|\mu) = -H(P). \tag{1.19}$$

1.4 The Method of Types

For finite alphabets, a particularly convenient tool in information theory is the *method of types* [37, 39, 74]. For a sequence $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$ in which $|\mathcal{X}|$ is finite, its *type* or *empirical distribution* is the probability mass function

$$P_{\mathbf{x}}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{x_i = x\}, \qquad \forall\, x \in \mathcal{X}. \tag{1.20}$$

Throughout, we use the notation $\mathbb{1}\{\text{clause}\}$ to mean the *indicator function*, i.e., this function equals 1 if "clause" is true and 0 otherwise. The set of types formed from n-length sequences in $\mathcal{X}$ is denoted as $\mathscr{P}_n(\mathcal{X})$. This is clearly a subset of $\mathscr{P}(\mathcal{X})$. The *type class* of P, denoted as $\mathcal{T}_P$, is the set of all sequences of length n for which their type is P, i.e.,

$$\mathcal{T}_P := \{\mathbf{x} \in \mathcal{X}^n : P_{\mathbf{x}} = P\}. \tag{1.21}$$

It is customary to indicate the dependence of $\mathcal{T}_P$ on the *blocklength* n but we suppress this dependence for the sake of conciseness throughout. For a sequence $\mathbf{x} \in \mathcal{T}_P$, the set of all sequences $\mathbf{y} \in \mathcal{Y}^n$ such that $(\mathbf{x}, \mathbf{y})$ has joint type $P \times V$ is the V*-shell*, denoted as $\mathcal{T}_V(\mathbf{x})$. In other words,

$$\mathcal{T}_V(\mathbf{x}) := \{\mathbf{y} \in \mathcal{Y}^n : P_{\mathbf{x},\mathbf{y}} = P \times V\}. \tag{1.22}$$

The conditional distribution V is also known as the *conditional type* of $\mathbf{y}$ given $\mathbf{x}$. Let $\mathscr{V}_n(\mathcal{Y}; P)$ be the set of all $V \in \mathscr{P}(\mathcal{Y}|\mathcal{X})$ for which the V-shell of a sequence of type P is non-empty.

We will often times find it useful to consider information-theoretic quantities of empirical distributions. All such quantities are denoted using hats. So for example, the *empirical entropy* of a sequence $\mathbf{x} \in \mathcal{X}^n$ is denoted as

$$\hat{H}(\mathbf{x}) := H(P_{\mathbf{x}}). \tag{1.23}$$

The *empirical conditional entropy* of $\mathbf{y} \in \mathcal{Y}^n$ given $\mathbf{x} \in \mathcal{X}^n$ where $\mathbf{y} \in \mathcal{T}_V(\mathbf{x})$ is denoted as

$$\hat{H}(\mathbf{y}|\mathbf{x}) := H(V|P_{\mathbf{x}}). \tag{1.24}$$

The *empirical mutual information* of a pair of sequences $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ with joint type $P_{\mathbf{x},\mathbf{y}} = P_{\mathbf{x}} \times V$ is denoted as

$$\hat{I}(\mathbf{x} \wedge \mathbf{y}) := I(P_{\mathbf{x}}, V). \tag{1.25}$$

The following lemmas form the basis of the method of types. The proofs can be found in [37, 39].

Lemma 1.1 (Type Counting). The sets $\mathscr{P}_n(\mathcal{X})$ and $\mathscr{V}_n(\mathcal{Y};P)$ for $P \in \mathscr{P}_n(\mathcal{X})$ satisfy

$$|\mathscr{P}_n(\mathcal{X})| \le (n+1)^{|\mathcal{X}|}, \quad \text{and} \quad |\mathscr{V}_n(\mathcal{Y};P)| \le (n+1)^{|\mathcal{X}||\mathcal{Y}|}. \tag{1.26}$$

In fact, it is easy to check that $|\mathscr{P}_n(\mathcal{X})| = \binom{n+|\mathcal{X}|-1}{|\mathcal{X}|-1}$ but (1.26) or its slightly stronger version

$$|\mathscr{P}_n(\mathcal{X})| \le (n+1)^{|\mathcal{X}|-1} \tag{1.27}$$

usually suffices for our purposes in this monograph. This key property says that the number of types is polynomial in the blocklength n.

Lemma 1.2 (Size of Type Class). For a type $P \in \mathscr{P}_n(\mathcal{X})$, the type class $\mathcal{T}_P \subset \mathcal{X}^n$ satisfies

$$|\mathscr{P}_n(\mathcal{X})|^{-1} \exp\big(nH(P)\big) \le |\mathcal{T}_P| \le \exp\big(nH(P)\big). \tag{1.28}$$

For a conditional type $V \in \mathscr{V}_n(\mathcal{Y};P)$ and a sequence $\mathbf{x} \in \mathcal{T}_P$, the V-shell $\mathcal{T}_V(\mathbf{x}) \subset \mathcal{Y}^n$ satisfies

$$|\mathscr{V}_n(\mathcal{Y};P)|^{-1} \exp\big(nH(V|P)\big) \le |\mathcal{T}_V(\mathbf{x})| \le \exp\big(nH(V|P)\big). \tag{1.29}$$

This lemma says that, on the exponential scale,

$$|\mathcal{T}_P| \cong \exp\big(nH(P)\big), \quad \text{and} \quad |\mathcal{T}_V(\mathbf{x})| \cong \exp\big(nH(V|P)\big), \tag{1.30}$$

where we used the notation $a_n \cong b_n$ to mean equality up to a polynomial, i.e., there exists polynomials p_n and q_n such that $a_n/p_n \le b_n \le q_n a_n$. We now consider probabilities of sequences. Throughout, for a distribution $Q \in \mathscr{P}(\mathcal{X})$, we let $Q^n(\mathbf{x})$ be the product distribution, i.e.,

$$Q^n(\mathbf{x}) = \prod_{i=1}^n Q(x_i), \qquad \forall\, \mathbf{x} \in \mathcal{X}^n. \tag{1.31}$$

Lemma 1.3 (Probability of Sequences). If $\mathbf{x} \in \mathcal{T}_P$ and $\mathbf{y} \in \mathcal{T}_V(\mathbf{x})$,

$$Q^n(\mathbf{x}) = \exp\big(-nD(P\|Q) - nH(P)\big) \quad \text{and} \tag{1.32}$$

$$W^n(\mathbf{y}|\mathbf{x}) = \exp\big(-nD(V\|W|P) - nH(V|P)\big). \tag{1.33}$$

This, together with Lemma 1.2, leads immediately to the final lemma in this section.

Lemma 1.4 (Probability of Type Classes)**.** For a type $P \in \mathscr{P}_n(\mathcal{X})$,

$$|\mathscr{P}_n(\mathcal{X})|^{-1} \exp\big(-nD(P\|Q)\big) \le Q^n(\mathcal{T}_P) \le \exp\big(-nD(P\|Q)\big). \quad (1.34)$$

For a conditional type $V \in \mathscr{V}_n(\mathcal{Y};P)$ and a sequence $\mathbf{x} \in \mathcal{T}_P$, we have

$$\begin{aligned} |\mathscr{V}_n(\mathcal{Y};P)|^{-1} \exp\big(-nD(V\|W|P)\big) &\le W^n(\mathcal{T}_V(\mathbf{x})|\mathbf{x}) \\ &\le \exp\big(-nD(V\|W|P)\big). \end{aligned} \quad (1.35)$$

The interpretation of this lemma is that the probability that a random i.i.d. (independently and identically distributed) sequence X^n generated from Q^n belongs to the type class $\mathcal{T}_P$ is exponentially small with exponent $D(P\|Q)$, i.e.,

$$Q^n(\mathcal{T}_P) \cong \exp\big(-nD(P\|Q)\big). \quad (1.36)$$

The bounds in (1.35) can be interpreted similarly.

1.5 Probability Bounds

In this section, we summarize some bounds on probabilities that we use extensively in the sequel. For a random variable X, we let $\mathsf{E}[X]$ and $\mathsf{Var}(X)$ be its expectation and variance respectively. To emphasize that the expectation is taken with respect to a random variable X with distribution P, we sometimes make this explicit by using a subscript, i.e., E_X or E_P.

1.5.1 Basic Bounds

We start with the familiar Markov and Chebyshev inequalities.

Proposition 1.1 (Markov's inequality)**.** Let X be a real-valued non-negative random variable. Then for any $a > 0$, we have

$$\Pr(X \ge a) \le \frac{\mathsf{E}[X]}{a}. \quad (1.37)$$

If we let X above be the non-negative random variable $(X - \mathsf{E}[X])^2$, we obtain Chebyshev's inequality.

Proposition 1.2 (Chebyshev's inequality)**.** Let X be a real-valued random variable with mean μ and variance σ^2. Then for any $b > 0$, we have

$$\Pr\left(|X - \mu| \geq b\sigma\right) \leq \frac{1}{b^2}. \tag{1.38}$$

We now consider a collection of real-valued random variables that are i.i.d. In particular, let $X^n = (X_1, \ldots, X_n)$ be a collection of independent random variables where each X_i has distribution P with zero mean and finite variance σ^2.

Proposition 1.3 (Weak Law of Large Numbers)**.** For every $\epsilon > 0$, we have

$$\lim_{n\to\infty} \Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| > \epsilon\right) = 0. \tag{1.39}$$

Consequently, the average $\frac{1}{n}\sum_{i=1}^{n} X_i$ converges to 0 in probability.

This follows by applying Chebyshev's inequality to the random variable $\frac{1}{n}\sum_{i=1}^{n} X_i$. In fact, under mild conditions, the convergence to zero in (1.39) occurs exponentially fast. See, for example, Cramer's theorem in [43, Thm. 2.2.3].

1.5.2 Central Limit-Type Bounds

In preparation for the next result, we denote the *probability density function* (pdf) of a univariate Gaussian as

$$\mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}. \tag{1.40}$$

We will also denote this as $\mathcal{N}(\mu, \sigma^2)$ if the argument x is unnecessary. A *standard Gaussian distribution* is one in which the mean $\mu = 0$ and the standard deviation $\sigma = 1$. In the multivariate case, the pdf is

$$\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^k |\boldsymbol{\Sigma}|}} \mathrm{e}^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})} \tag{1.41}$$

where $\mathbf{x} \in \mathbb{R}^k$. A *standard multivariate Gaussian distribution* is one in which the mean is $\mathbf{0}_k$ and the covariance is the identity matrix $\mathbf{I}_{k\times k}$.

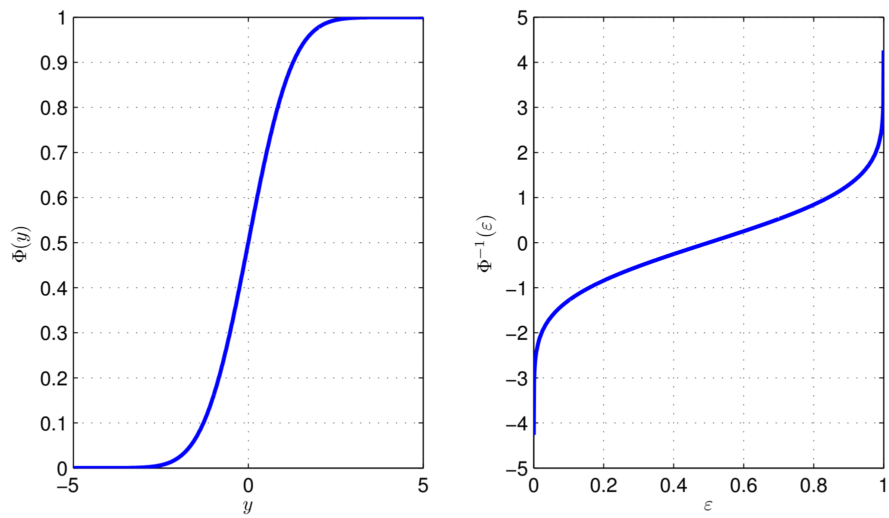

Figure 1.2: Plots of $\Phi(y)$ and $\Phi^{-1}(\varepsilon)$

For the univariate case, the *cumulative distribution function* (cdf) of the standard Gaussian is denoted as

$$\Phi(y) := \int_{-\infty}^{y} \mathcal{N}(x; 0, 1)\, \mathrm{d}x. \tag{1.42}$$

We also find it convenient to introduce the inverse of Φ as

$$\Phi^{-1}(\varepsilon) := \sup\{y \in \mathbb{R} : \Phi(y) \le \varepsilon\} \tag{1.43}$$

which evaluates to the usual inverse for $\varepsilon \in (0, 1)$ and extends continuously to take values $\pm\infty$ for ε outside $(0, 1)$. These monotonically increasing functions are shown in Fig. 1.2.

If the scaling in front of the sum in the statement of the law of large numbers in (1.39) is $\frac{1}{\sqrt{n}}$ instead of $\frac{1}{n}$, the resultant random variable $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i$ converges in distribution to a Gaussian random variable. As in Proposition 1.3, let X^n be a collection of i.i.d. random variables where each X_i has zero mean and finite variance σ^2.

Proposition 1.4 (Central Limit Theorem)**.** For any $a \in \mathbb{R}$, we have

$$\lim_{n\to\infty} \Pr\left(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n} X_i < a\right) = \Phi(a). \tag{1.44}$$

In other words,

$$\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n} X_i \xrightarrow{\mathrm{d}} Z \tag{1.45}$$

where $\xrightarrow{\mathrm{d}}$ means convergence in distribution and Z is the standard Gaussian random variable.

Throughout the monograph, in the evaluation of the non-asymptotic bounds, we will use a more quantitative version of the central limit theorem known as the Berry-Esseen theorem [17, 52]. See Feller [54, Sec. XVI.5] for a proof.

Theorem 1.5 (Berry-Esseen Theorem (i.i.d. Version))**.** Assume that the third absolute moment is finite, i.e., $T := \mathsf{E}\big[|X_1|^3\big] < \infty$. For every $n \in \mathbb{N}$, we have

$$\sup_{a\in\mathbb{R}}\left|\Pr\left(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n} X_i < a\right) - \Phi(a)\right| \le \frac{T}{\sigma^3\sqrt{n}}. \tag{1.46}$$

Remarkably, the Berry-Esseen theorem says that the convergence in the central limit theorem in (1.44) is uniform in $a \in \mathbb{R}$. Furthermore, the convergence of the distribution function of $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i$ to the Gaussian cdf occurs at a rate of $O(\frac{1}{\sqrt{n}})$. The constant of proportionality in the $O(\cdot)$-notation depends *only* on the variance and the third absolute moment and not on any other statistics of the random variables.

There are many generalizations of the Berry-Esseen theorem. One which we will need is the relaxation of the assumption that the random variables are identically distributed. Let $X^n = (X_1, \ldots, X_n)$ be a collection of independent random variables where each random variable has zero mean, variance $\sigma_i^2 := \mathsf{E}[X_i^2] > 0$ and third absolute moment $T_i := \mathsf{E}\big[|X_i|^3\big] < \infty$. We respectively define the average variance and average third absolute moment as

$$\sigma^2 := \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2, \quad \text{and} \quad T := \frac{1}{n}\sum_{i=1}^{n} T_i. \tag{1.47}$$

Theorem 1.6 (Berry-Esseen Theorem (General Version))**.** For every $n \in \mathbb{N}$, we have

$$\sup_{a\in\mathbb{R}}\left|\Pr\left(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n} X_i < a\right) - \Phi(a)\right| \le \frac{6\,T}{\sigma^3\sqrt{n}}. \tag{1.48}$$

Observe that as with the i.i.d. version of the Berry-Esseen theorem, the remainder term scales as $O(\frac{1}{\sqrt{n}})$.

The proof of the following theorem uses the Berry-Esseen theorem (among other techniques). This theorem is proved in Polyanskiy-Poor-Verdú [123, Lem. 47]. Together with its variants, this theorem is useful for obtaining third-order asymptotics for binary hypothesis testing and other coding problems with non-vanishing error probabilities.

Theorem 1.7. Assume the same setup as in Theorem 1.6. For any $\gamma \geq 0$, we have

$$\mathsf{E}\left[\exp\left(-\sum_{i=1}^{n} X_i\right)\mathbb{1}\left\{\sum_{i=1}^{n} X_i > \gamma\right\}\right] \leq 2\left(\frac{\log 2}{\sqrt{2\pi}} + \frac{12T}{\sigma^2}\right)\frac{\exp(-\gamma)}{\sigma\sqrt{n}}. \tag{1.49}$$

It is trivial to see that the expectation in (1.49) is upper bounded by $\exp(-\gamma)$. The additional factor of $(\sigma\sqrt{n})^{-1}$ is crucial in proving coding theorems with better third-order terms. Readers familiar with strong large deviation theorems or exact asymptotics (see, e.g., [23, Thms. 3.3 and 3.5] or [43, Thm. 3.7.4]) will notice that (1.49) is in the same spirit as the theorem by Bahadur and Ranga-Rao [13]. There are two advantages of (1.49) compared to strong large deviation theorems. First, the bound is purely in terms of σ^2 and T, and second, one does not have to differentiate between lattice and non-lattice random variables. The disadvantage of (1.49) is that the constant is worse but this will not concern us as we focus on asymptotic results in this monograph, hence constants do not affect the main results.

For multi-terminal problems that we encounter in the latter parts of this monograph, we will require vector (or multidimensional) versions of the Berry-Esseen theorem. The following is due to Götze [63].

Theorem 1.8 (Vector Berry-Esseen Theorem I)**.** Let $X_1^k, \ldots, X_n^k$ be independent $\mathbb{R}^k$-valued random vectors with zero mean. Let

$$S_n^k = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i^k. \tag{1.50}$$

Assume that S_n^k has the following statistics

$$\mathsf{Cov}(S_n^k) = \mathsf{E}\big[S_n^k (S_n^k)'\big] = \mathbf{I}_{k\times k}, \quad \text{and} \quad \xi := \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}\big[\|X_i^k\|_2^3\big]. \quad (1.51)$$

Let Z^k be a standard Gaussian random vector, i.e., its distribution is $\mathcal{N}(0^k, \mathbf{I}_{k\times k})$. Then, for all $n \in \mathbb{N}$, we have

$$\sup_{\mathscr{C}\in\mathfrak{C}_k} \Big|\Pr\left(S_n^k \in \mathscr{C}\right) - \Pr\left(Z^k \in \mathscr{C}\right)\Big| \le \frac{c_k\,\xi}{\sqrt{n}}, \quad (1.52)$$

where $\mathfrak{C}_k$ is the family of all convex subsets of $\mathbb{R}^k$, and where c_k is a constant that depends only on the dimension k.

Theorem 1.8 can be applied for random vectors that are independent but not necessarily identically distributed. The constant c_k can be upper bounded by $400\,k^{1/4}$ if the random vectors are i.i.d., a result by Bentkus [15]. However, its precise value will not be of concern to us in this monograph. Observe that the scalar versions of the Berry-Esseen theorems (in Theorems 1.5 and 1.6) are special cases (apart from the constant) of the vector version in which the family of convex subsets is restricted to the family of semi-infinite intervals $(-\infty, a)$.

We will frequently encounter random vectors with non-identity covariance matrices. The following modification of Theorem 1.8 is due to Watanabe-Kuzuoka-Tan [177, Cor. 29].

Corollary 1.9 (Vector Berry-Esseen Theorem II)**.** Assume the same setup as in Theorem 1.8, except that $\mathsf{Cov}(S_n^k) = \mathbf{V}$, a positive definite matrix. Then, for all $n \in \mathbb{N}$, we have

$$\sup_{\mathscr{C}\in\mathfrak{C}_k} \Big|\Pr\left(S_n^k \in \mathscr{C}\right) - \Pr\left(Z^k \in \mathscr{C}\right)\Big| \le \frac{c_k\,\xi}{\lambda_{\min}(\mathbf{V})^{3/2}\sqrt{n}}, \quad (1.53)$$

where $\lambda_{\min}(\mathbf{V}) > 0$ is the smallest eigenvalue of $\mathbf{V}$.

The final probability bound is a quantitative version of the so-called *multivariate delta method* [174, Thm. 5.15]. Numerous similar statements of varying generalities have appeared in the statistics literature (e.g., [24, 175]). The simple version we present was shown by Molavian-Jazi and Laneman [112] who extended ideas in Hoeffding and Robbins' paper [81, Thm. 4] to provide rates of convergence to Gaussianity under appropriate technical conditions. This result essentially says that a differentiable function of a normalized sum of independent random vectors also satisfies a Berry-Esseen-type result.

Theorem 1.10 (Berry-Esseen Theorem for Functions of i.i.d. Random Vectors)**.** Assume that $X_1^k, \ldots, X_n^k$ are $\mathbb{R}^k$-valued, zero-mean, i.i.d. random vectors with positive definite covariance $\mathsf{Cov}(X_1^k)$ and finite third absolute moment $\xi := \mathsf{E}[\|X_1^k\|_2^3]$. Let $\mathbf{f}(\mathbf{x})$ be a vector-valued function from $\mathbb{R}^k$ to $\mathbb{R}^l$ that is also twice continuously differentiable in a neighborhood of $\mathbf{x} = \mathbf{0}$. Let $\mathbf{J} \in \mathbb{R}^{l \times k}$ be the Jacobian matrix of $\mathbf{f}(\mathbf{x})$ evaluated at $\mathbf{x} = \mathbf{0}$, i.e., its elements are

$$J_{ij} = \frac{\partial f_i(\mathbf{x})}{\partial x_j}\bigg|_{\mathbf{x}=\mathbf{0}}, \tag{1.54}$$

where $i = 1, \ldots, l$ and $j = 1, \ldots, k$. Then, for every $n \in \mathbb{N}$, we have

$$\sup_{\mathscr{C} \in \mathfrak{C}_l} \left| \Pr\left(\mathbf{f}\left(\frac{1}{n} \sum_{i=1}^n X_i^k \right) \in \mathscr{C} \right) - \Pr\left(Z^l \in \mathscr{C} \right) \right| \le \frac{c}{\sqrt{n}} \tag{1.55}$$

where $c > 0$ is a finite constant, and Z^l is a Gaussian random vector in $\mathbb{R}^l$ with mean vector and covariance matrix respectively given as

$$\mathsf{E}[Z^l] = \mathbf{f}(\mathbf{0}), \quad \text{and} \quad \mathsf{Cov}(Z^l) = \frac{\mathbf{J}\,\mathsf{Cov}(X_1^k)\mathbf{J}'}{n}. \tag{1.56}$$

In particular, the inequality in (1.55) implies that

$$\sqrt{n}\left(\mathbf{f}\left(\frac{1}{n} \sum_{i=1}^n X_i^k \right) - \mathbf{f}(\mathbf{0}) \right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(\mathbf{0}, \mathbf{J}\,\mathsf{Cov}(X_1^k)\mathbf{J}' \right), \tag{1.57}$$

which is a canonical statement in the study of the multivariate delta method [174, Thm. 5.15].

Finally, we remark that Ingber-Wang-Kochman [87] used a result similar to that of Theorem 1.10 to derive second-order asymptotic results for various Shannon-theoretic problems. However, they analyzed the behavior of functions of *distributions* instead of functions of *random vectors* as in Theorem 1.10.

References

[1] R. Ahlswede. Multiway communication channels. In *Proceedings of the International Symposium on Information Theory*, pages 23–51, Tsahkadsor, Armenia, Sep 1971.

[2] R. Ahlswede. The capacity of a channel with two senders and two receivers. *Annals of Probability*, 2(5):805–814, 1974.

[3] R. Ahlswede. An elementary proof of the strong converse theorem for the multiple access channel. *Journal of Combinatorics, Information & System Sciences*, 7(3):216–230, 1982.

[4] R. Ahlswede and I. Csiszár. Common randomness in information theory and cryptography–I: Secret sharing. *IEEE Transactions on Information Theory*, 39(4):1221–1132, 1993.

[5] R. Ahlswede and G. Dueck. Identification via channels. *IEEE Transactions on Information Theory*, 35(1):15–29, 1989.

[6] R. Ahlswede, P. Gács, and J. Körner. Bounds on conditional probabilities with applications in multi-user communication. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 34(3):157–177, 1976.

[7] R. Ahlswede and J. Körner. Source coding with side information and a converse for the degraded broadcast channel. *IEEE Transactions on Information Theory*, 21(6):629–637, 1975.

[8] Y. Altuğ and A. B. Wagner. Feedback can improve the second-order coding performance in discrete memoryless channels. In *Proceedings of the International Symposium on Information Theory*, pages 2361–2365, Honolulu, HI, Jul 2014.

[9] Y. Altuğ and A. B. Wagner. Moderate deviations in channel coding. *IEEE Transactions on Information Theory*, 60(8):4417–4426, 2014.

[10] Y. Altuğ and A. B. Wagner. Refinement of the random coding bound. *IEEE Transactions on Information Theory*, 60, 2014.

[11] Y. Altuğ and A. B. Wagner. Refinement of the sphere-packing bound: Asymmetric channels. *IEEE Transactions on Information Theory*, 60(3):1592–1614, 2014.

[12] Y. Altuğ and A. B. Wagner. The third-order term in the normal approximation for singular channels. In *Proceedings of the International Symposium on Information Theory*, pages 1897–1901, Honolulu, HI, Jul 2014. `arXiv:1309.5126 [cs.IT]`.

[13] R. R. Bahadur and R. Ranga Rao. On deviations of the sample mean. *Annals of Mathematical Statistics*, 31(4):1015–1027, 1980.

[14] D. Baron, M. A. Khojastepour, and R. G. Baraniuk. How quickly can we approach channel capacity? In *Proceedings of Asilomar Conference on Signals, Systems and Computers*, pages 1096–1100, Monterey, CA, Nov 2004.

[15] V. Bentkus. On the dependence of the Berry-Esseen bound on dimension. *Journal of Statistical Planning and Inference*, 113:385–402, 2003.

[16] T. Berger. *Rate-Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, N.J.: Prentice-Hall, 1971.

[17] A. C. Berry. The accuracy of the Gaussian approximation to the sum of independent variates. *Transactions of the American Mathematical Society*, 49(1):122–136, 1941.

[18] E. Biglieri, J. Proakis, and S. Shamai (Shitz). Fading channels: information-theoretic and communications aspects. *IEEE Transactions on Information Theory*, 44(6):2619–2692, 1998.

[19] M. Bloch and J. Barros. *Physical-Layer Security: From Information Theory to Security Engineering*. Cambridge University Press, 2011.

[20] S. Boucheron and M. R. Salamatian. About priority encoding transmission. *IEEE Transactions on Information Theory*, 46(2):699–705, 2000.

[21] M. V. Burnashev. Information transmission over a discrete channel with feedback. *Problems of Information Transmission*, 12(4):10–30, 1976.

[22] A. B. Carleial. A case where interference does not reduce capacity. *IEEE Transactions on Information Theory*, 21:569–570, 1975.

[23] N. R. Chaganty and J. Sethuraman. Strong large deviation and local limit theorems. *Annals of Probability*, 21(3):1671–1690, 1993.

[24] L. H. Y. Chen and Q.-M. Shao. Normal approximation for nonlinear statistics using a concentration inequality approach. *Bernoulli*, 13(2):581–599, 2007.

[25] H. Chernoff. Measure of asymptotic effiency tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, 23:493–507, 1952.

[26] H.-F. Chong, M. Motani, H. K. Garg, and H. El Gamal. On the Han-Kobayashi region for the interference channel. *IEEE Transactions on Information Theory*, 54(7):3188–3195, 2008.

[27] B. S. Clarke and A. R. Barron. Information-theoretic asymptotics of bayes methods. *IEEE Transactions on Information Theory*, 36(3):453–471, 1990.

[28] J. H. Conway and N. J. A. Sloane. *Sphere packings, lattices and groups.* Springer Verlag, 2003.

[29] T. Cormen, C. Leiserson, R. Rivest, and C. Stein. *Introduction to Algorithms.* McGraw-Hill Science/Engineering/Math, 2nd edition, 2003.

[30] M. Costa. Writing on dirty paper. *IEEE Transactions on Information Theory*, 29(3):439–441, 1983.

[31] T. Cover. Broadcast channels. *IEEE Transactions on Information Theory*, 18(1):2–14, 1972.

[32] T. M. Cover. A proof of the data compression theorem of Slepian and Wolf for ergodic sources. *IEEE Transactions on Information Theory*, 21(3):226–228, 1975.

[33] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley-Interscience, 2nd edition, 2006.

[34] I. Csiszár. On an extremum problem of information theory. *Studia Sci. Math. Hungarica*, 9(1):57–71, 1974.

[35] I. Csiszár. Joint source-channel error exponent. *Problems of Control and Information Theory*, 9:315–328, 1980.

[36] I. Csiszár. Linear codes for sources and source networks: Error exponents, universal coding. *IEEE Transactions on Information Theory*, 28(4), 1982.

[37] I. Csiszár. The method of types. *IEEE Transactions on Information Theory*, 44(6):2505–23, 1998.

[38] I. Csiszár and J. Körner. Graph decomposition: A new key to coding theorems. *IEEE Transactions on Information Theory*, 27:5–11, 1981.

[39] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.

[40] I. Csiszár and Z. Talata. Context tree estimation for not necessarily finite memory processes, via BIC and MDL. *IEEE Transactions on Information Theory*, 52(3):1007–1016, 2006.

[41] P. Cuff. Distributed channel synthesis. *IEEE Transactions on Information Theory*, 59(11):7071–7096, 2012.

[42] M. Dalai. Lower bounds on the probability of error for classical and classical-quantum channels. *IEEE Transactions on Information Theory*, 59(12):8027–8056, 2013.

[43] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer, 2nd edition, 1998.

[44] N. Devroye, P. Mitran, and V. Takokh. Achievable rates in cognitive radio channels. *IEEE Transactions on Information Theory*, 52(5):1813–1827, 2006.

[45] R. L. Dobrushin. Mathematical problems in the Shannon theory of optimal coding of information. In *Proc. 4th Berkeley Symp. Math., Statist., Probabil.,*, pages 211–252, 1961.

[46] G. Dueck. Maximal error capacity regions are smaller than average error capacity regions for multi-user channels. *Problems of Control and Information Theory*, 7(1):11–19, 1978.

[47] G. Dueck. The strong converse coding theorem for the multiple-access channel. *Journal of Combinatorics, Information & System Sciences*, 6(3):187–196, 1981.

[48] F. Dupuis, L. Kraemer, P. Faist, J. M. Renes, and R. Renner. Generalized entropies. In *Proceedings of the XVIIth International Congress on Mathematical Physics*, 2012.

[49] A. El Gamal and Y.-H. Kim. *Network Information Theory*. Cambridge University Press, Cambridge, U.K., 2012.

[50] E. O. Elliott. Estimates of error rates for codes on burst-noise channels. *The Bell Systems Technical Journal*, 42:1977–97, Sep 1963.

[51] U. Erez, S. Litsyn, and R. Zamir. Lattices which are good for (almost) everything. *IEEE Transactions on Information Theory*, 51(10):3401–3416, 2005.

[52] C.-G. Esseen. On the Liapunoff limit of error in the theory of probability. *Arkiv för matematik, astronomi och fysik*, A28(1):1–19, 1942.

[53] A Feinstein. A new basic theorem of information theory. *IEEE Transactions on Information Theory*, 4(4):2–22, 1954.

[54] W. Feller. *An Introduction to Probability Theory and Its Applications*. John Wiley and Sons, 2nd edition, 1971.

[55] G. D. Forney. Exponential error bounds for erasure, list, and decision feedback schemes. *IEEE Transactions on Information Theory*, 14:206–220, 1968.

[56] R. G. Gallager. *Information Theory and Reliable Communication*. Wiley, New York, 1968.

[57] R. G. Gallager. Capacity and coding for degraded broadcast channels. *Problems of Information Transmission*, 10(3):3–14, 1974.

[58] R. G. Gallager. Source coding with side information and universal coding. Technical report, MIT LIDS, 1976.

[59] S. Gelfand and M. Pinsker. Coding for channel with random parameters. *Problems of Control and Information Theory*, 9(1):19–31, 1980.

[60] E. N. Gilbert. Capacity of burst-noise channels. *The Bell Systems Technical Journal*, 39:1253–1265, Sep 1960.

[61] A. J. Goldsmith and P. P. Varaiya. Capacity of fading channels with channel side information. *IEEE Transactions on Information Theory*, 43(6):1986–1992, 1997.

[62] V. D. Goppa. Nonprobabilistic mutual information without memory. *Problems of Control and Information Theory*, 4:97–102, 1975.

[63] F. Götze. On the rate of convergence in the multivariate CLT. *Annals of Probability*, 19(2):721–739, 1991.

[64] W. Gu and M. Effros. A strong converse for a collection of network source coding problems. In *Proceedings of the International Symposium on Information Theory*, pages 2316–2320, Seoul, S. Korea, Jun–Jul 2009.

[65] E. Haim, Y. Kochman, and U. Erez. A note on the dispersion of network problems. In *Convention of Electrical and Electronics Engineers in Israel*, pages 1–9, Eilat, Nov 2012.

[66] T. S. Han. An information-spectrum approach to capacity theorems for the general multiple-access channel. *IEEE Transactions on Information Theory*, 44(7):2773–2795, 1998.

[67] T. S. Han. *Information-Spectrum Methods in Information Theory*. Springer Berlin Heidelberg, Feb 2003.

[68] T. S. Han. Folklore in source coding: Information-spectrum approach. *IEEE Transactions on Information Theory*, 51(2):747–753, 2005.

[69] T. S. Han and S.-I. Amari. Statistical inference under multiterminal data compression. *IEEE Transactions on Information Theory*, 44(6):2300–2324, 1998.

[70] T. S. Han and K. Kobayashi. A new achievable rate region for the interference channel. *IEEE Transactions on Information Theory*, 27(1):49–60, 1981.

[71] T. S. Han and S. Verdú. Approximation theory of output statistics. *IEEE Transactions on Information Theory*, 39(3):752–772, 1993.

[72] E. A. Haroutunian. Error probability lower bound for the multiple-access communication channels. *Problems of Information Transmission*, 11(2):22–36, 1975.

[73] E. A. Haroutunian. A lower bound on the probability of error for channels with feedback. *Problems of Information Transmission*, 3(2):37–48, 1977.

[74] E. A. Haroutunian, M. E. Haroutunian, and A. N. Harutyunyan. Reliability criteria in information theory and statistical hypothesis testing. *Foundations and Trends® in Communications and Information Theory*, 4(2–3):97–263, 2007.

[75] M. Hayashi. Second-order asymptotics in fixed-length source coding and intrinsic randomness. *IEEE Transactions on Information Theory*, 54(10):4619–4637, 2008.

[76] M. Hayashi. Information spectrum approach to second-order coding rate in channel coding. *IEEE Transactions on Information Theory*, 55(11):4947–4966, 2009.

[77] M. Hayashi and H. Nagaoka. General formulas for capacity of classical-quantum channels. *IEEE Transactions on Information Theory*, 49(7):1753–68, 2003.

[78] M. Hayashi, H. Tyagi, and S. Watanabe. Secret key agreement: General capacity and second-order asymptotics. In *Proceedings of the International Symposium on Information Theory*, pages 1136–1140, Honolulu, HI, Jul 2014.

[79] M. Hayashi, H. Tyagi, and S. Watanabe. Strong converse for degraded wiretap channel via active hypothesis testing. In *Proceedings of Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct 2014.

[80] D.-K. He, L. A. Lastras-Montaño, E.-H. Yang, A. Jagmohan, and J. Chen. On the redundancy of Slepian-Wolf coding. *IEEE Transactions on Information Theory*, 55(12):5607–5627, 2009.

[81] W. Hoeffding and H. Robbins. The central limit theorem for dependent random variables. *Duke Mathematical Journal*, 15(3):773–780, 1948.

[82] J. Hoydis, R. Couillet, and P. Piantanida. Bounds on the second-order coding rate of the MIMO Rayleigh block-fading channel. In *Proceedings of the International Symposium on Information Theory*, pages 1526–1530, Istanbul, Turkey, Jul 2013. `arXiv:1303.3400 [cs.IT]`.

[83] J. Hoydis, R. Couillet, P. Piantanida, and M. Debbah. A random matrix approach to the finite blocklength regime of MIMO fading channels. In *Proceedings of the International Symposium on Information Theory*, pages 2181–2185, Cambridge, MA, Jul 2012.

[84] Y.-W. Huang and P. Moulin. Finite blocklength coding for multiple access channels. In *Proceedings of the International Symposium on Information Theory*, pages 831–835, Cambridge, MA, Jul 2012.

[85] A. Ingber and M. Feder. Finite blocklength coding for channels with side information at the receiver. In *Proceedings of the Convention of Electrical and Electronics Engineers in Israel*, pages 798–802, Eilat, Nov 2010.

[86] A. Ingber and Y. Kochman. The dispersion of lossy source coding. In *Proceedings of the Data Compression Conference (DCC)*, pages 53–62, Snowbird, UT, Mar 2011. `arXiv:1102.2598 [cs.IT]`.

[87] A. Ingber, D. Wang, and Y. Kochman. Dispersion theorems via second order analysis of functions of distributions. In *Proceedings of Conference on Information Sciences and Systems*, pages 1–6, Princeton, NJ, Mar 2012.

[88] A. Ingber, R. Zamir, and M. Feder. Finite dimensional infinite constellations. *IEEE Transactions on Information Theory*, 59(3):1630–1656, 2013.

[89] J. Jiang and T. Liu. On dispersion of modulo lattice additive noise channels. In *Proceedings of the International Symposium on Wireless Communication Systems*, pages 241–245, Aachen, Germany, Nov 2011.

[90] B. Kelly and A. Wagner. Reliability in source coding with side information. *IEEE Transactions on Information Theory*, 58(8):5086–5111, 2012.

[91] J. H. B. Kemperman. Studies in Coding Theory I. Technical report, University of Rochester, NY, 1962.

[92] G. Keshet, Y. Steinberg, and N. Merhav. Channel coding in the presence of side information. *Foundations and Trends® in Communications and Information Theory*, 4(6):445–486, 2007.

[93] I. Kontoyiannis. Second-order noiseless source coding theorems. *IEEE Transactions on Information Theory*, 43(4):1339–1341, 1997.

[94] I. Kontoyiannis. Pointwise redundancy in lossy data compression and universal lossy data compression. *IEEE Transactions on Information Theory*, 46:136–152, 2000.

[95] I. Kontoyiannis and S. Verdú. Optimal lossless data compression: Non-asymptotics and asymptotics. *IEEE Transactions on Information Theory*, 60(2):777–795, 2014.

[96] V. Kostina. *Lossy Data Compression: Non-asymptotic fundamental limits*. PhD thesis, Princeton University, 2013.

[97] V. Kostina and S. Verdú. Fixed-length lossy compression in the finite blocklength regime. *IEEE Transactions on Information Theory*, 58(6):3309–3338, 2012.

[98] V. Kostina and S. Verdú. Channels with cost constraints: strong converse and dispersion. In *Proceedings of the International Symposium on Information Theory*, pages 1734–1738, Istanbul, Turkey, Jul 2013. `arXiv:1401.5124 [cs.IT]`.

[99] V. Kostina and S. Verdú. Lossy joint source-channel coding in the finite blocklength regime. *IEEE Transactions on Information Theory*, 59(5):2545–2575, 2013.

[100] O. Kosut and L. Sankar. Universal fixed-to-variable source coding in the finite blocklength regime. In *Proceedings of the International Symposium on Information Theory*, pages 649–653, Istanbul, Turkey, Jul 2013.

[101] O. Kosut and L. Sankar. New results on third-order coding rate for universal fixed-to-variable source coding. In *Proceedings of the International Symposium on Information Theory*, pages 2689–2693, Honolulu, HI, Jul 2014.

[102] S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86, 1951.

[103] S.-Q. Le, V. Y. F. Tan, and M. Motani. Second-order asymptotics for the Gaussian interference channel with very strong interference. In *Proceedings of the International Symposium on Information Theory*, pages 2514–2518, Honolulu, HI, Jul 2014.

[104] Y. Liang, H. V. Poor, and S. Shamai (Shitz). Information-theoretic security. *Foundations and Trends® in Communications and Information Theory*, 5(4–5):355–580, 2008.

[105] H. H. J. Liao. *Multiple access channels*. PhD thesis, University of Hawaii, Honolulu, 1972.

[106] K. Marton. Error exponent for source coding with a fidelity criterion. *IEEE Transactions on Information Theory*, 20(2):197–199, 1974.

[107] K. Marton. A simple proof of the blowing-up lemma. *IEEE Transactions on Information Theory*, 32(3):445–446, 1986.

[108] W. Matthews. A linear program for the finite block length converse of Polyanskiy-Poor-Verdú via nonsignaling codes. *IEEE Transactions on Information Theory*, 58(2):7036–7044, 2012.

[109] N. Merhav. Universal decoding for memoryless Gaussian channels with a deterministic interference. *IEEE Transactions on Information Theory*, 39(4):1261–1269, 1993.

[110] S. Miyake and F. Kanaya. Coding theorems on correlated general sources. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E78-A(9):1063–1070, 1995.

[111] E. MolavianJazi and J. N. Laneman. Simpler achievable rate regions for multiaccess with finite blocklength. In *Proceedings of the International Symposium on Information Theory*, pages 36–40, Cambridge, MA, Jul 2012.

[112] E. MolavianJazi and J. N. Laneman. A finite-blocklength perspective on Gaussian multi-access channels. *Submitted to the IEEE Transactions on Information Theory*, 2014. `arXiv:1309.2343 [cs.IT]`.

[113] P. Moulin. The log-volume of optimal codes for memoryless channels, within a few nats. *Submitted to the IEEE Transactions on Information Theory*, Nov 2013. `arXiv:1311.0181 [cs.IT]`.

[114] P. Moulin and J. A. O'Sullivan. Information-theoretic analysis of information hiding. *IEEE Transactions on Information Theory*, 49(3):563–593, 2003.

[115] P. Moulin and Y. Wang. Capacity and random-coding exponents for channel coding with side information. *IEEE Transactions on Information Theory*, 53(4):1326–1347, 2007.

[116] M. Mushkin and I. Bar-David. Capacity and coding for the Gilbert-Elliott channels. *IEEE Transactions on Information Theory*, 35(6):1277–1290, 1989.

[117] R. Nomura and T. S. Han. Second-order resolvability, intrinsic randomness, and fixed-length source coding for mixed sources: Information spectrum approach. *IEEE Transactions on Information Theory*, 59(1):1–16, 2013.

[118] R. Nomura and T. S. Han. Second-order Slepian-Wolf coding theorems for non-mixed and mixed sources. *IEEE Transactions on Information Theory*, 60(9):5553–5572, 2014.

[119] Y. Polyanskiy. *Channel coding: Non-asymptotic fundamental limits*. PhD thesis, Princeton University, 2010.

[120] Y. Polyanskiy. On dispersion of compound dmcs. In *Proceedings of Allerton Conference on Communication, Control, and Computing*, pages 26–32, Monticello, IL, Oct 2013.

[121] Y. Polyanskiy. Saddle point in the minimax converse for channel coding. *IEEE Transactions on Information Theory*, 59(5):2576–2595, 2013.

[122] Y. Polyanskiy, H. V. Poor, and S. Verdú. New channel coding achievability bounds. In *Proceedings of the International Symposium on Information Theory*, pages 1763–1767, Toronto, ON, Jul 2008.

[123] Y. Polyanskiy, H. V. Poor, and S. Verdú. Channel coding rate in the finite blocklength regime. *IEEE Transactions on Information Theory*, 56(5):2307–2359, 2010.

[124] Y. Polyanskiy, H. V. Poor, and S. Verdú. Dispersion of the Gilbert-Elliott channel. *IEEE Transactions on Information Theory*, 57(4):1829–48, 2011.

[125] Y. Polyanskiy, H. V. Poor, and S. Verdú. Feedback in the non-asymptotic regime. *IEEE Transactions on Information Theory*, 57(8):4903–4925, 2011.

[126] Y. Polyanskiy, H. V. Poor, and S. Verdú. Minimum energy to send k bits through the Gaussian channel with and without feedback. *IEEE Transactions on Information Theory*, 57(8):4880–4902, 2011.

[127] Y. Polyanskiy and S. Verdú. Channel dispersion and moderate deviations limits for memoryless channels. In *Proceedings of Allerton Conference on Communication, Control, and Computing*, pages 1334–1339, Monticello, IL, Oct 2010.

[128] V. V. Prelov. Transmission over a multiple-access channel with a special source hierarchy. *Problems of Information Transmission*, 20(4):3–10, 1984.

[129] M. Raginsky and I. Sason. Concentration of measure inequalities in information theory, communications and coding. *Foundations and Trends® in Communications and Information Theory*, 10(1–2):1–247, 2013.

[130] T. J. Riedl, T. P. Coleman, and A. C. Singer. Finite block-length achievable rates for queuing timing channels. In *Proceedings of the IEEE Information Theory Workshop*, pages 200–204, Paraty, Brazil, Oct 2011.

[131] S. Sarvotham, D. Baron, and R. G. Baraniuk. Non-asymptotic performance of symmetric Slepian-Wolf coding. In *Proceedings of Conference on Information Sciences and Systems*, Baltimore, MD, Mar 2005.

[132] S. Sarvotham, D. Baron, and R. G. Baraniuk. Variable-rate universal Slepian-Wolf coding with feedback. In *Proceedings of Asilomar Conference on Signals, Systems and Computers*, pages 8–12, Pacific Grove, CA, Nov 2005.

[133] I. Sason. Moderate deviations analysis of binary hypothesis testing. In *Proceedings of the International Symposium on Information Theory*, pages 821–825, Cambridge, MA, Jul 2012.

[134] J. Scarlett. On the dispersion of dirty paper coding. In *Proceedings of the International Symposium on Information Theory*, pages 2282–2286, Honolulu, HI, Jul 2014. `arXiv:1309.6200 [cs.IT]`.

[135] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas. A derivation of the asymptotic random-coding prefactor. In *Proceedings of Allerton Conference on Communication, Control, and Computing*, pages 956–961, Monticello, IL, Oct 2013. `arXiv:1306.6203 [cs.IT]`.

[136] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas. Second-order rate region of constant-composition codes for the multiple-access channel. In *Proceedings of Allerton Conference on Communication, Control, and Computing*, pages 588–593, Monticello, IL, Oct 2013. `arXiv:1303.6167 [cs.IT]`.

[137] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas. The saddlepoint approximation: Unified random coding asymptotics for fixed and varying rates. In *Proceedings of the International Symposium on Information Theory*, pages 1892–1896, Honolulu, HI, Jul 2014. `arXiv:1402.3941 [cs.IT]`.

[138] J. Scarlett and V. Y. F. Tan. Second-order asymptotics for the Gaussian MAC with degraded message sets. *IEEE Transactions on Information Theory*, 2014.

[139] H. M. H. Shalaby and A. Papamarcou. Multiterminal detection with zero-rate data compression. *IEEE Transactions on Information Theory*, 38(2):254 –267, 1992.

[140] X. Shang and B. Chen. Two-user Gaussian interference channels: An information theoretic point of view. *Foundations and Trends® in Communications and Information Theory*, 10(3):247–378, 2013.

[141] C. E. Shannon. A mathematical theory of communication. *The Bell Systems Technical Journal*, 27:379–423, 1948.

[142] C. E. Shannon. The zero error capacity of a noisy channel. *IRE Transactions on Information Theory*, 2(3):8–19, 1956.

[143] C. E. Shannon. Channels with side information at the transmitter. *IBM J. Res. Develop.*, 2:289–293, 1958.

[144] C. E. Shannon. Coding theorems for a discrete source with a fidelity criterion. *IRE Nat. Conv. Rec.*, pages 142–163, 1959.

[145] C. E. Shannon. Probability of error for optimal codes in a Gaussian channel. *The Bell Systems Technical Journal*, 38:611–656, 1959.

[146] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp. Lower bounds to error probability for coding in discrete memoryless channels I-II. *Information and Control*, 10:65–103,522–552, 1967.

[147] Y. Y. Shkel, V. Y. F. Tan, and S. C. Draper. Converse bounds for assorted codes in the finite blocklength regime. In *Proceedings of the International Symposium on Information Theory*, pages 1720–1724, Istanbul, Turkey, Jul 2013.

[148] Y. Y. Shkel, V. Y. F. Tan, and S. C. Draper. Achievability bounds for unequal message protection at finite block lengths. In *Proceedings of the International Symposium on Information Theory*, pages 2519–2523, Honolulu, HI, Jul 2014. `arXiv:1405.0891 [cs.IT]`.

[149] Y. Y. Shkel, V. Y. F. Tan, and S. C. Draper. On mismatched unequal error protection for finite blocklength joint source-channel coding. In *Proceedings of the International Symposium on Information Theory*, pages 1692–1696, Honolulu, HI, Jul 2014.

[150] Z. Shun and P. McCullagh. Laplace approximation of high dimensional integrals. *Journal of the Royal Statistical Society, Series B (Methodology)*, 57(4):749–760, 1995.

[151] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, 19(4):471–80, 1973.

[152] V. Strassen. Asymptotische Abschätzungen in Shannons Informationstheorie. In *Trans. Third Prague Conf. Inf. Theory*, pages 689–723, Prague, 1962. http://www.math.cornell.edu/∼pmlut/strassen.pdf.

[153] V. Y. F. Tan. Achievable second-order coding rates for the wiretap channel. In *IEEE International Conference on Communication Systems*, pages 65–69, Singapore, Nov 2012.

[154] V. Y. F. Tan. Moderate-deviations of lossy source coding for discrete and Gaussian sources. In *Proceedings of the International Symposium on Information Theory*, pages 920–924, Cambridge, MA, Jul 2012. `arXiv:1111.2217 [cs.IT]`.

[155] V. Y. F. Tan and M. Bloch. Information spectrum approach to strong converse theorems for degraded wiretap channels. In *Proceedings of Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct 2014. `arXiv:1406.6758 [cs.IT]`.

[156] V. Y. F. Tan and O. Kosut. The dispersion of Slepian-Wolf coding. In *Proceedings of the International Symposium on Information Theory*, pages 915–919, Cambridge, MA, Jul 2012.

[157] V. Y. F. Tan and O. Kosut. On the dispersions of three network information theory problems. *IEEE Transactions on Information Theory*, 60(2):881–903, 2014.

[158] V. Y. F. Tan and P. Moulin. Second-order capacities for erasure and list decoding. In *Proceedings of the International Symposium on Information Theory*, pages 1887–1891, Honolulu, HI, Jul 2014. `arXiv:1402.4881 [cs.IT]`.

[159] V. Y. F. Tan and M. Tomamichel. The third-order term in the normal approximation for the AWGN channel. In *Proceedings of the International Symposium on Information Theory*, pages 2077–2081, Honolulu, HI, Jul 2014. `arXiv:1311.2337 [cs.IT]`.

[160] V. Y. F. Tan, S. Watanabe, and M. Hayashi. Moderate deviations for joint source-channel coding of systems with Markovian memory. In *Proceedings of the International Symposium on Information Theory*, pages 1687–1691, Honolulu, HI, Jul 2014.

[161] İ. E. Telatar. *Multi-access communications with decision feedback*. PhD thesis, Massachusetts Institute of Technology, 1992.

[162] L. Tierney and J. B. Kadane. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):82–86, Mar 1986.

[163] M. Tomamichel and M. Hayashi. A hierarchy of information quantities for finite block length analysis of quantum tasks. *IEEE Transactions on Information Theory*, 59(11):7693–7710, 2013.

[164] M. Tomamichel and V. Y. F. Tan. A tight upper bound for the third-order asymptotics of most discrete memoryless channels. *IEEE Transactions on Information Theory*, 59(11):7041–7051, 2013.

[165] M. Tomamichel and V. Y. F. Tan. Second-order coding rates for channels with state. *IEEE Transactions on Information Theory*, 60(8):4427–4448, 2014.

[166] H. Tyagi and P. Narayan. The Gelfand-Pinsker channel: Strong converse and upper bound for the reliability function. In *Proceedings of the International Symposium on Information Theory*, pages 1954–1957, Seoul, S. Korea, Jul 2009. `arXiv:0910.0653 [cs.IT]`.

[167] E. C. van der Meulen. Some recent results on the asymmetric multiple-access channel. In *Proceedings of the 2nd joint Swedish-Soviet International Workshop on Information Theory*, 1985.

[168] S. Verdú. Non-asymptotic achievability bounds in multiuser information theory. In *Proceedings of Allerton Conference on Communication, Control, and Computing*, pages 1–8, Monticello, IL, Oct 2012.

[169] S. Verdú and T. S. Han. A general formula for channel capacity. *IEEE Transactions on Information Theory*, 40(4):1147–1157, 1994.

[170] D. Wang, A. Ingber, and Y. Kochman. The dispersion of joint source-channel coding. In *Proceedings of Allerton Conference on Communication, Control, and Computing*, pages 180–187, Monticello, IL, Oct 2011. `arXiv::1109.6310 [cs.IT]`.

[171] D. Wang, A. Ingber, and Y. Kochman. A strong converse for joint source-channel coding. In *Proceedings of the International Symposium on Information Theory*, pages 2117–2121, Cambridge, MA, Jul 2012.

[172] L. Wang, R. Colbeck, and R. Renner. Simple channel coding bounds. In *Proceedings of the International Symposium on Information Theory*, pages 1804–1808, Seoul, S. Korea, July 2009. `arXiv::0901.0834 [cs.IT]`.

[173] L. Wang and R. Renner. One-shot classical-quantum capacity and hypothesis testing. *Physical Review Letters*, 108:200501, May 2012.

[174] L. Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer, 2004.

[175] L. Wasserman, M. Kolar, and A. Rinaldo. Berry-Esseen bounds for estimating undirected graphs. *Electronic Journal of Statistics*, 8:1188–1224, 2014.

[176] S. Watanabe and M. Hayashi. Strong converse and second-order asymptotics of channel resolvability. In *Proceedings of the International Symposium on Information Theory*, pages 1882–1886, Honolulu, HI, Jul 2014.

[177] S. Watanabe, S. Kuzuoka, and V. Y. F. Tan. Non-asymptotic and second-order achievability bounds for coding with side-information. *IEEE Transactions on Information Theory*, 2014. `arXiv:1301.6467 [cs.IT]`.

[178] G. Wiechman and I. Sason. An improved sphere-packing bound for finite-length codes over symmetric memoryless channels. *IEEE Transactions on Information Theory*, 54(5):1962–1990, 2009.

[179] A. R. Williamson, T.-Y. Chen, and R. D. Wesel. Reliability-based error detection for feedback communication with low latency. In *Proceedings of the International Symposium on Information Theory*, pages 2552–2556, Istanbul, Turkey, Jul 2013. `arXiv::1305.4560 [cs.IT]`.

[180] J. Wolfowitz. The coding of messages subject to chance errors. *Illinois Journal of Mathematics*, 1(4):591–606, 1957.

[181] J. Wolfowitz. *Coding Theorems of Information Theory*. Springer-Verlag, New York, 3rd edition, 1978.

[182] A. D. Wyner. On source coding with side information at the decoder. *IEEE Transactions on Information Theory*, 21(3):294–300, 1975.

[183] A. D. Wyner. The wire-tap channel. *The Bell Systems Technical Journal*, 54:1355–1387, 1975.

[184] A. D. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Transactions on Information Theory*, 22(1):1–10, 1976.

[185] H. Yagi and R. Nomura. Channel dispersion for well-ordered mixed channels decomposed into memoryless channels. In *Proceedings of the International Symposium on Information Theory and Its Applications*, Melbourne, Australia, Oct 2014.

[186] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy. Quasi-static MIMO fading channels at finite blocklength. *IEEE Transactions on Information Theory*, 60(7):4232–4265, 2014.

[187] M. H. Yassaee, M. R. Aref, and A. Gohari. Non-asymptotic output statistics of random binning and its applications. In *Proceedings of the International Symposium on Information Theory*, pages 1849–1853, Istanbul, Turkey, Jul 2013. `arXiv:1303.0695 [cs.IT]`.

[188] M. H. Yassaee, M. R. Aref, and A. Gohari. A technique for deriving one-shot achievability results in network information theory. In *Proceedings of the International Symposium on Information Theory*, pages 1287–1291, Istanbul, Turkey, Jul 2013. `arXiv:1303.0696 [cs.IT]`.

[189] R. Yeung. *A First Course on Information Theory.* Springer, 2002.

[190] A. A. Yushkevich. On limit theorems connected with the concept of entropy of Markov chains. *Uspekhi Matematicheskikh Nauk*, 5(57):177–180, 1953.

[191] Z. Zhang, E.-H. Yang, and V. K. Wei. The redundancy of source coding with a fidelity criterion: Known statistics. *IEEE Transactions on Information Theory*, 43(1):71–91, 1997.

[192] J. Ziv and A. Lempel. Compression of individual sequences via variable-rate coding. *IEEE Transactions on Information Theory*, 24(5):530–536, 1978.