Twenty Questions with Random Error

Other titles in Foundations and Trends® in Communications and Information Theory

Power Control for Battery-Limited Energy Harvesting Communications
Shengtian Yang and Jun Chen
ISBN: 978-1-63828-520-5

A Toolbox for Refined Information-Theoretic Analyses
Neri Merhav and Nir Weinberger
ISBN: 978-1-63828-500-7

Channel Simulation: Theory and Applications to Lossy Compression and Differential Privacy
Cheuk Ting Li
ISBN: 978-1-63828-486-4

Maximizing Entropy with an Expectation Constraint and One-Parameter Exponential Families of Distributions: A Reexamination
David L. Neuhoff
ISBN: 978-1-63828-480-2

Codes for Adversaries: Between Worst-Case and Average-Case Jamming
Bikash Kumar Dey, Sidharth Jaggi, Michael Langberg, Anand D. Sarwate and Yihan Zhang
ISBN: 978-1-63828-460-4

Universal Features for High-Dimensional Learning and Inference
Shao-Lun Huang, Anuran Makur, Gregory W. Wornell and Lizhong Zheng
ISBN: 978-1-63828-176-4

Twenty Questions with Random Error

Lin Zhou
Beihang University
lzhou@buaa.edu.cn

Alfred O. Hero
University of Michigan
hero@eecs.umich.edu

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

L. Zhou and A. O. Hero. *Twenty Questions with Random Error*. Foundations and Trends® in Communications and Information Theory, vol. 22, no. 4, pp. 394–604, 2025.

ISBN: 978-1-63828-547-2

Foundations and Trends® in Communications and Information Theory
Volume 22, Issue 4, 2025

Editorial Scope

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, 2025, Volume 22, 4 issues. ISSN paper version 1567-2190. ISSN online version 1567-2328 . Also available as a combined paper and online subscription.

Contents

Twenty Questions with Random Error

Lin Zhou[1] and Alfred O. Hero[2]

[1] *School of Cyber Science and Technology, Beihang University; lzhou@buaa.edu.cn*

[2] *Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor; hero@eecs.umich.edu*

ABSTRACT

Twenty Questions originated as a parlor game between two players. The game starts from a player named an oracle, who privately thinks of a secret. The other player, called the questioner, tries to guess the secret by querying the oracle with at most twenty questions having Yes/No answers. Early versions of the game can be traced to ancient Greece and ancient Rome. Motivated by the Hungarian version of this game, in the middle of the twentieth century, Rényi formulated the game as a mathematical problem of guessing an integer from a finite set, where the oracle could lie either randomly to each question or lie to a finite number of questions. The mathematical study of Twenty Questions is motivated by current applications in many domains: communications; ma-chine learning; and computer vision.

The game with an oracle who is allowed a fixed number of lies was also studied by Ulam and Berlekamp and is known as the Rényi-Ulam-Berlekamp game. In contrast, the setting where the oracle lies randomly is less understood. In this monograph, we summarize recent advances in the

Lin Zhou and Alfred O. Hero (2025), "Twenty Questions with Random Error", Foundations and Trends® in Communications and Information Theory: Vol. 22, No. 4, pp 394–604. DOI: 10.1561/0100000144.

information theoretical analysis of Twenty Questions with random error. In particular, focusing on the practical application of sensor network target localization, we study a query-dependent channel to model oracle's noisy response behavior, such as providing a wrong answer or declining to answer a question. We concentrate on non-adaptive query procedures where all questions are designed prior to posing questions. We cover settings relevant to estimating a single target, a single moving target, and multiple targets over the unit cube of a finite dimension. We also consider adaptive querying for a single target to illustrate the benefit of adaptivity. In adaptive querying, each question is designed sequentially using responses to all previous questions. All of our theoretical results are illustrated using numerical examples.

Finally, we discuss future research directions. These include geometry constraints for query sets, low-complexity query procedures, connections to group testing, and practical applications in machine learning and communications.

1

Introduction

1.1 What is Twenty Questions?

1.1.1 Origin

The modern version of the Twenty Questions game was proposed in the mid twentieth century as a parlor game between two players. The game starts with one player, the oracle, who privately thinks of a secret, which could be, for example, an object, a location, or the answer to a question that is unknown to the other player, the questioner. The task of the questioner, is to figure out the secret by posing a sequence of at most twenty questions to the oracle. In the parlor game, the players are assumed to be honest and lying is prohibited.

A query procedure consists of a policy to formulate questions and a decision rule to estimate the secret using the sequence of answers of the oracle. A query procedure can be either adaptive or non-adaptive. In an adaptive query procedure, the questioner designs the current question based on oracle's responses to all previous questions. In a non-adaptive query procedure, the questioner designs each question independently of other questions and responses. Consider the example of guessing an integer that takes values in $\{1,\ldots,8\}$. Assume that the secret chosen by the oracle is 3. A classical adaptive query procedure is bisection. In bisection, the first question could be "Is the number no greater than 4?" Given the correct answer "yes" from the oracle, the questioner knows that the secret must be in $\{1, \ldots, 4\}$ and poses the next question as "Is the number no greater than 2?" Again, given the correct answer "no", the questioner knows that the secret is either 3 or 4 and poses the final question "Is the number 3?". Given the answer "yes" from the oracle, the questioner has successfully estimated the secret. The above bisection procedure successively shrinks the search region using the entire sequence of answers from the oracle.

A non-adaptive query procedure, called the dyadic procedure, also finds the correct answer to this simple game in three steps. To formulate the non-adaptive query procedure, the questioner converts the set of decimal integers $\{1, \ldots, 8\}$ into binary form as $\{000, \ldots, 111\}$, where 000 corresponds to 1 and 111 corresponds to 8. Subsequently, the questioner poses three questions at once: the first question is whether the first bit is one, the second question is whether the second bit is one and the third question is whether the third bit is one. In other words, the first question asks whether the secret is in $\{5, 6, 7, 8\}$, the second question asks whether the secret is in $\{3, 4, 7, 8\}$ and the third question asks whether the secret is in $\{2, 4, 6, 8\}$. When the secret is 3, the answers of the oracle would be "no, yes, no", which corresponds to the binary number 010 and thus the decimal number 3. The dyadic procedure achieves the same performance as the bisection adaptive procedure. The above example generalizes to guessing any integer in a finite range. In particular, if the secret is an integer with M possible values, the Twenty Questions game requires only $\lceil \log_2 M \rceil$ questions to determine the secret when either dyadic or bisection query procedure is used.

1.1.2 History

Games similar to Twenty Questions existed in ancient times, which may have influenced the rediscovery of the parlor game in the twentieth century. In ancient Greece, the philosopher Plato (4th century BCE) described a similar game called "Erotema". In his treatise The Phaedra, in the form of a dialog, Plato wrote that: "The art of question-asking, which is what I call Erotema, is a powerful thing, Phaedrus. It is the midwife of the soul. Just as a midwife brings a child to birth in the body, so the questioner brings to birth ideas in the soul." The game of Erotema was presented as a powerful tool for understanding the nature of rhetoric and for improving one's own rhetorical skills. Plato explained that rhetoric is a form of persuasion, and that the best way to persuade someone is to ask them a series of questions that will lead them to the desired conclusion. In this sense, Erotema corresponds to adaptive query procedures and goes beyond the parlor game of Twenty Questions by providing a model for teaching and learning.

In ancient Rome, riddles were popular and were described in the works of several Roman poets and playrights. A compound riddle is a form where several independent questions are given in the riddle that jointly narrow down the answer. Publius Ovidius Naso (43 BCE to 17 AD), or simply Ovid, was a prolific Roman poet who wrote compound riddles. Ovid's riddles are thought-provoking, and often combine two or more different concepts into one entity. For example, in one riddle, Ovid asks: "What is that which is born of water, yet lives on land?" "What is that which has no mouth, yet speaks?" "What is that which has no feet, yet walks?" The answer to the riddle is a boat. As observed, each question specifies a certain property and successive questions are not dependent. Thus, compound riddles are analogous to non-adaptive query procedures of the Twenty Questions game described above.

In [79, Page 13], Rényi reported the following legendary story of Bar Kochba (135 BC), who played a trick similar to Twenty Questions: *"In 135 BC, the Jews started a war of independence against Romans under the leadership of Bar Kochba. The Romans, in superior numbers, laid siege to a fortress which was defended historically by Bar Kochba at the head of a small garrison. It is also said that Bar Kochba sent out a scout to the Roman camp who was captured and tortured, having his tongue cut out. He escaped from captivity and reported back to Bar Kochba, but being unable to talk, he could not tell in words what he had seen. Bar Kochba accordingly asked him questions which he could answer by nodding or shaking his head. Thus he acquired from his mute scout the information he needed to defend the fortress."*

1.1.3 Rényi's Two Formulations

The story of Bar Kochba led to the Hungarian version game of Twenty Questions named the Bar-Kochba game. Inspired by the game, Rényi [77] pioneered the study of Twenty Questions with errors, which assumes that the secret is a random variable taking values in a finite set of integers, the questioner poses yes or no questions, and the oracle can lie and give incorrect answers *randomly* to each question with a certain probability. In particular, Rényi wrote that *"Two players are playing the game, let us call them A and B. A thinks of something and B must guess it. B can ask questions which can be answered by yes or no and he must find out what A had thought from the answers. It is better to suppose that a given percentage of the answers are wrong (because A misunderstands the questions or does not know certain facts."*

Subsequently, in [79, Page 53], Rényi reformulated the problem so that the oracle lies up to a fixed number of questions and pointed out the connection between the Twenty Questions problem and channel coding by writing the following paragraph: *"I tried, when thinking about what I heard today, to make a connection between information transmission through a channel and our game. I made up the following version, which I called 'Bar-Kochba with lies'. Assume that the number of questions which can be asked to figure out the "something" being thought of is fixed and the one who answers is allowed to lie a certain number of times. The questioner, of course, doesn't know which answer is true and which is not. Moreover the one answering is not required to lie as many times as is allowed. "*

1.2 Scope, Previous Work and Significance of This Monograph

1.2.1 Scope of This Monograph

The Twenty Questions problem with a fixed number of lies was also studied by Berlekamp [11] as a quiet-question-noisy-answer game and subsequently popularized by Ulam in his autobiography [98]. The Twenty Questions problem with a fixed number of lies is therefore called the Rényi-Berlekamp-Ulam game. Over the past six decades, the Rényi-Berlekamp-Ulam game has been well studied under various assumptions on the type of the questions, the error patterns and the search space [1], [2], [30], [33], [37], [38], [39], [57], [60], [61], [68]. Many of these advancements have borrowed ideas from coding theory and have been comprehensively summarized in surveys and monographs, e.g., [23], [32], [70].

In contrast, the original version of Twenty Questions with random error, proposed by Rényi in [77], has been less studied, e.g., in [69]. In 2012, roughly 50 years after the introduction of the problem, the Twenty Questions problem with random error was revived by Jedynak *et al.* [51] and subsequently generalized to many different settings [17], [18], [21], [54], [76], [94], [95], [99] in the past decade. These analyses have borrowed tools from Shannon theory [86] as contrasted to the coding theory based analyses used in the case of fixed number of errors.

These previous studies on Twenty Questions with random error have limitations, which include the use of an indirect estimation accuracy measure, the posterior entropy for estimation accuracy [76], [94], [95], and the assumption that number of questions is unbounded [17], [18], [21], [54]. These assumptions facilitated the application of Shannon theory. This monograph overcomes these limitations by borrowing tools from non-asymptotic and second-order asymptotic analyses for channel coding [73], [88], [92]. This allows us to obtain theoretical benchmarks for query procedures with *finitely* many questions under the direct accuracy measure of absolute error. We develop this theory for four different settings, unifying and integrating the work reported in [89], [107], [108], [109], [112], [113], [114].

1.2.2 Previous Work and Significance of This Monograph

Below we summarize previous work that sets the context for this monograph. Jedynak *et al.* [51] studied the problem of Twenty Questions with random error introduced by Rényi [77]. In [51], it was assumed that the secret is a continuous random variable taking values in the unit interval. The posterior entropy of the questioner's estimate of the random variable was proposed as the performance criterion, and optimal adaptive and non-adaptive query procedures were shown to achieve identical performance. The results of Jedynak *et al.* have been generalized in several directions, including a collaborative setting [94] with multiple oracles, a distributed setting [95] with multiple questioners and a multiple target setting with more than one secret random variable [76].

However, the posterior entropy is an indirect measure of estimation accuracy, which leads to the dilemma that smaller entropy does not guarantee more accurate estimation [20]. To solve this problem, the square error or the absolute error criterion have been adopted as alternatives to posterior entropy. In particular, Variani [99] proposed a non-adaptive query procedure that ensured sub-exponential decay of the mean square error (MSE) with respect to the number of questions. The result of [99] was refined by Chung *et al.*, who used superposition coding to construct a non-adaptive query procedure that ensures exponential decay of MSE. All above studies assumed that the random noise is independent of each question. In this case, the Twenty Questions game with random error is essentially a channel coding problem and the theoretical results of both problems are closely related.

Inspired by practical noise accumulation phenomena in sensor networks, Chiu and Javidi [17] proposed the query-dependent noise model, where the random noise depends on each question though a function of its size, and analyzed the achievable performance of an adaptive query procedure. Subsequently, under the same noise model, for the first time ever in the studies of Twenty Questions with random error, Kaspi *et al.* [54] demonstrated the benefit of adaptivity by showing that adaptive query procedures yields superior performance over optimal non-adaptive query procedures. Furthermore, Kaspi *et al.* also studied a multiple target setting [53] and a moving target setting [54, Theorem 3], where the secret is the position of a target that changes linearly with a certain speed. For non-adaptive query procedures, they studied the asymptotic limit of $-\log \delta^*(n,\varepsilon)/n$, where n is the number of questions, $\varepsilon \in (0,1)$ is the probability of estimating the target variable incorrectly and $\delta^*(n,\varepsilon)$ denotes the estimation accuracy using n questions subject to the error probability constraint ε. Such results are known as first-order asymptotics, since they only account for the first dominant term in the asymptotic expansion of the estimation accuracy, i.e., $-\log \delta^*(n,\varepsilon) = nC + o(n)$ for some constant C, when $n \to \infty$ and

$\varepsilon \to 0$. Analogously, for adaptive query procedures, they studied the first dominant term for the asymptotic limit of $-\log \delta_{\mathrm{a}}^*(l, \varepsilon)/l$, where l is the average number of questions of an adaptive query procedure, $\varepsilon \in (0, 1)$ is the probability of estimating the target variable incorrectly, and $\delta^*(l, \varepsilon)$ denotes the estimation accuracy of an optimal adaptive query procedure with average number of questions no greater than l subject to the error probability constraint ε. Thus, all above-mentioned existing results are only *asymptotically* optimal when the number of questions is unbounded.

The critical practical question that forms the basis for this monograph is: what is the theoretical performance of optimal query procedures under query-dependent random noise and a direct accuracy measure when the number of questions is *finite*? This problem is of interest since any practical search problem must be completed in finite time, requiring a finite number n of questions. To achieve this goal, we need to bound the estimation accuracy $\delta^*(n, \varepsilon)$ for non-adaptive query procedures with finite $n \in \mathbb{N}$ and bound $\delta_{\mathrm{a}}^*(l, \varepsilon)$ for adaptive query procedures with finite $l \in \mathbb{R}_+$. In recent studies [109], [110], [111], we obtained such non-asymptotic bounds and second-order asymptotic approximations when one searches for a target in a finite dimensional unit cube of dimension $d \in \mathbb{N}$ and the corresponding theoretical accuracy benchmarks were denoted as $\delta^*(n, d, \varepsilon)$ for non-adaptive query procedures and $\delta_{\mathrm{a}}^*(l, d, \varepsilon)$ for adaptive query procedures. In particular, the second-order asymptotic result demonstrates that $-d \log \delta^*(n, d, \varepsilon) = nC + \sqrt{n}L + O(\log n)$ for non-adaptive query procedures and $-d \log \delta_{\mathrm{a}}^*(l, d, \varepsilon) \geq \frac{lC}{1-\varepsilon} + O(\log n)$ for adaptive query procedures, where C is a function of the channel that models the noisy response behavior of the oracle and L is a function of both the channel and ε.

These second-order asymptotic results refine the first-order asymptotic studies in several aspects. Firstly, the second-order asymptotic result provides a tight approximation to the non-asymptotic limits $\delta^*(n, d, \varepsilon)$ for finite n and $\delta_{\mathrm{a}}^*(l, d, \varepsilon)$ for finite l. Secondly, for non-adaptive query procedures, the second-order asymptotic result brings the new insights that there exists phase transition: asymptotically the estimation error probability tends to either zero or one depending on whether or not the target accuracy exponent $\frac{-\log \delta^*(n,d,\varepsilon)}{d}$ is greater than the theoretical benchmark $\frac{C}{d}$. Thirdly, the second-order asymptotic result reveals the benefit of adaptivity even for query-independent channels. To illustrate, in Figure 1.1, we plot the first and second-order asymptotic approximations to estimation accuracy for an optimal non-adaptive query procedure over a Bernoulli noise channel and compare the results with the non-asymptotic simulated performance of an optimal query procedure in Algorithm 1 of Section 2. As observed, second-order asymptotics provides a good approximation to the non-asymptotic performance of optimal query procedures.

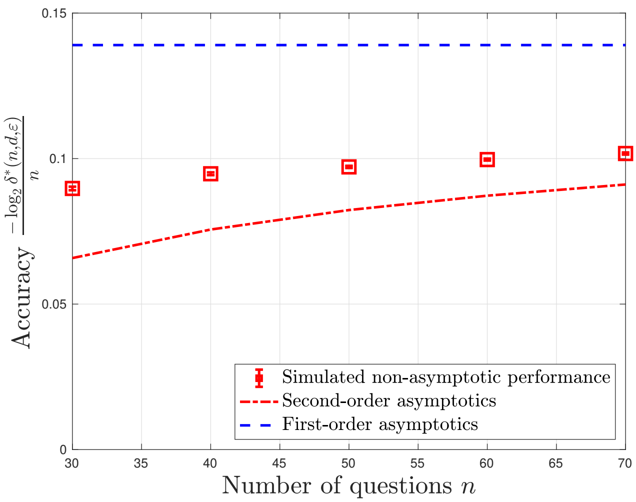


Figure 1.1: The second-order asymptotic analysis (red dot dashed) of a non-adaptive query procedure (Algorithm 1 in Section 2) is significantly better than the first-order analysis (blue dashed) in capturing estimation accuracy (red squares).

Subsequently, in a series of recent studies [89], [108], [112], [113], [114], these theoretical benchmarks for non-adaptive querying of a single target were generalized to a moving target and to multiple targets, following the same framework of non-asymptotic and second-order asymptotic analyses. When specialized to moving target search, the impact of maximal speed on the estimation accuracy is demonstrated. When specialized to multiple target search, the impact of the number of targets on estimation accuracy is demonstrated. Table 1.1 summarizes the

Settings	First-order asymptotics	Second-order asymptotics
Non-Adaptive Single Target	$\lim_{\varepsilon\to 0} -\log \delta^*(n,1,\varepsilon) = nC + o(n)$ [54, Theorem 1] for Bernoulli Noise	$-d\log \delta^*(n,d,\varepsilon) = nC + \sqrt{n\mathrm{V}_\varepsilon}\Phi^{-1}(\varepsilon) + O(\log n)$ Section 2 for arbitrary discrete and Gaussian noise
Adaptive Single Target	$\lim_{\varepsilon\to 0} -\log \delta_{\mathrm{a}}^*(l,1,\varepsilon) \geq lC_0 + o(n)$ [17] for Bernoulli noise, [59] for Gaussian noise	$-d\log \delta_{\mathrm{a}}^*(l,d,\varepsilon) \geq \frac{lC}{1-\varepsilon} + O(\log n)$ Section 3 for arbitrary discrete and Gaussian noise
Non-Adaptive Moving Target	$\lim_{\varepsilon\to 0} -k\log \delta^*(n,1,\varepsilon) = nC + o(n)$ [54, Theorem 3] for Bernoulli noise	$-2d\log \delta^*(n,d,\varepsilon) = nC + \sqrt{n\mathrm{V}_\varepsilon}\Phi^{-1}(\varepsilon) + O(\log n)$ Section 4 for arbitrary discrete and Gaussian noise
Non-Adaptive Multiple Targets	$\lim_{\varepsilon\to 0} -\log \delta^*(n,1,\varepsilon) = nC(k) + o(n)$ [53] for Bernoulli noise	$-kd\log \delta^*(n,d,\varepsilon) = nC(k) + \sqrt{n\mathrm{V}(k,\varepsilon)}\Phi^{-1}(\varepsilon) + O(\log n)$ Section 5 for arbitrary discrete and Gaussian noise

Table 1.1: Summary of theoretical benchmarks in this monograph

theoretical benchmarks for the four settings studied in this monograph, and compares to prior results for the case of one-dimensional single target search. In the table, C will be defined in (2.15), V_ε will be defined in (2.16), $C(k)$ will be defined in (5.10) and $\mathrm{V}(k,\varepsilon)$ will be defined in (5.11). Furthermore, Bernoulli noise refers to the query-dependent BSC channel in Definition 1.1, Gaussian noise refers to the query-dependent AWGN channel in Definition 1.4 while arbitrary discrete noise refers to any query-dependent channel defined in Section 1.4.3. The second-order asymptotic results developed here refined the existing first-order asymptotic results by providing approximations to performance of optimal tests with finitely many questions and an arbitrary error probability and generalizing the results from one-dimensional search with Bernoulli noise to finite-dimensional search with arbitrary discrete noise and with Gaussian noise. This monograph therefore demonstrates the impact of multi-dimension, moving targets and multiple targets on the theoretical benchmarks of optimal query procedures for Twenty Questions with random error.

In summary, this monograph provides a self-contained in-depth treatment of non-asymptotic and second-order asymptotic analyses of Twenty Questions with random error. The concluding section discusses future research directions, including opportunities for researches in information theory, communications, computer vision and machine learning. To our best knowledge, this monograph is the first to summarize advances for Twenty Questions with random error and finitely many questions, which complements the existing surveys of Twenty Questions with a fixed number of errors treated in [23], [32], [70] and provides several new or extended applications.

1.3 Practical Motivations of Twenty Questions

The study of Twenty Questions with random error has been motivated by several timely applications, including: fault-tolerant communication (channel coding) [17], [21], [54], target localization using sensor networks [94], [95], object tracking using cameras or satellite remote sensing networks [45], [90], face localization in images [76], and beam alignment in mmWave multiple antenna communication [19], [80]. We discuss these motivating applications below and defective elements detection in an intelligent reflecting surface [106].

1.3.1 Channel Coding

As shown in Figure 1.2, the task of point-to-point channel coding is to transmit a message reliably over a channel. For channel coding, random error means that the transmission of each bit of a message can be erroneous with a certain probability, governed by a transition matrix named the channel matrix induced by the channel. As commented by Rényi [78] and used in a sequence of other follow-up studies [21], [54], [99], non-adaptive querying for Twenty Questions with random error corresponds to channel coding [86] while adaptive querying corresponds to variable length channel coding with feedback [50]. In both cases, the questioner is equivalent to the encoder and the decoder while noisy response behavior of the oracle is modeled by the channel. Furthermore, the secret to be guessed is the transmitted message while the formulation of questions and the estimation procedure based on noisy answers correspond to the encoder and the decoder, respectively.

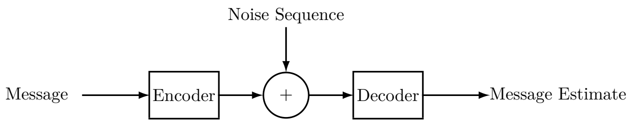


Figure 1.2: Channel coding interpretation of Twenty Questions. The goal is to transmit the message reliably over a channel to protect message symbols from noisy corruption via the encoder and decoder. The message corresponds to the target random variable, the noise corrupts the oracle's true answers and the encoder and decoder correspond to the query procedure.

For Twenty questions with random error, we prove its relationship with channel coding in Section 1.5 by showing that the theoretical limit of Twenty Questions with random error is bounded by the corresponding theoretical limit of channel coding. Thus, when one designs a query procedure for Twenty Questions with random error, the procedure can be modified to yield a code for channel coding and vice versa. The theoretical formulation summarized in this monograph builds on the connection between Twenty Questions and channel coding.

1.3.2 Target Localization

As shown in Figure 1.3, when one uses a sensor network to locate a stationary target, spatial coverage of the sensor network corresponds to the valid search space and the location of the target corresponds to the secret. Estimating the target location can be thus transformed as a Twenty Questions problem with random error, where error is associated with sensor measurement noise, a question corresponds to a subset of sensors to be queried and the noisy response of the oracle corresponds to the sensor measurements. The decision rule in Twenty Questions corresponds to the estimation function for target localization. Thus, the problem of stationary target localization with a sensor network can be mapped to the problem of Twenty Questions with random noise. This framework easily generalizes to multiple targets.

1.3.3 Object Tracking

As shown in Figure 1.4, when one tracks a moving target with unknown location and speed using either cameras or satellites, the task can be mapped to design of a query procedure for Twenty Questions with noise. Unlike target localization in Section 1.3.2, the target location changes over time and the goal is to estimate the target trajectory. Thus, in this case, the secret corresponds to the unknown target trajectory and all other parts remain unchanged relative to object localization in Section 1.3.2. As we shall show in Section 4, tracking a moving target is theoretically equivalent to locating a stationary target in a two-dimensional space.

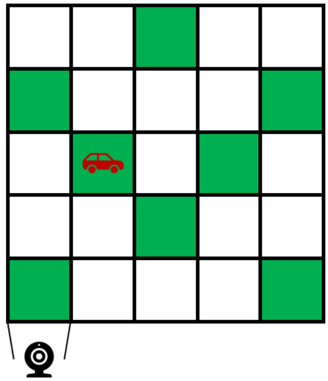


Figure 1.3: Locating a target (automobile, car) in space from sensor measurements. The search space is divided into equal-sized sub-regions and a target detection sensor is placed in each sub-region. At each time, the questioner chooses a subset of sensors and asks whether a target exists in the covering sub-regions (in green) of selected sensors.

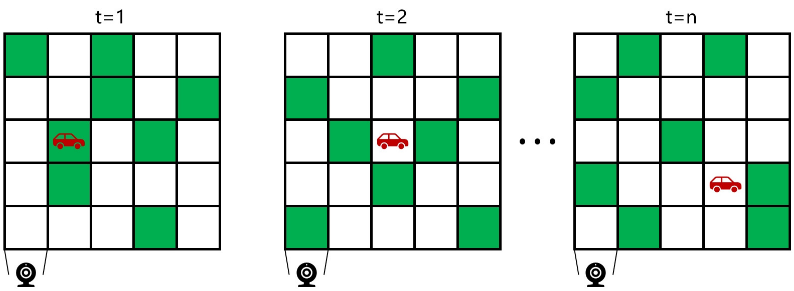


Figure 1.4: Localization of a moving target consists of a sequence of questions over time. The search space is divided into equal-sized sub-regions and a sensor is placed in each sub-region. At each time, the questioner chooses a set of sensors and asks whether a target currently lies in the covering sub-regions of selected sensors, which are denoted in green at each time point.

1.3.4 Face Localization

As shown in Figure 1.5, when one localizes a face in an image, one can first separate the whole image into equal-sized sub-regions, chooses a certain set of sub-regions and checks whether a face exists in the chosen sub-regions using a face detector. The task of localizing a face

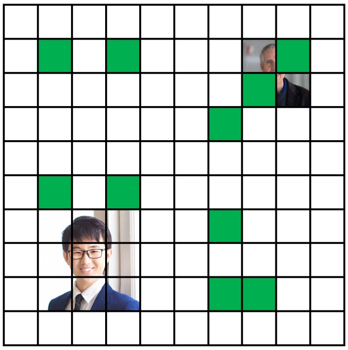


Figure 1.5: Face localization in an image. The image is separated into equal-sized sub-regions. At each time point, the questioner chooses a set of sub-regions and asks whether portraits of face exist in the chosen regions (denoted in green).

in an image corresponds to Twenty Questions with noise. Specifically, the search region is the whole image, the secret is the sub-regions that contain the face, the question corresponds to the chosen sub-regions and the oracle's response corresponds to the output of a certain face detector that could result in incorrect detection. Thus, we can modify query procedures for Twenty Questions with noise to design algorithms to localize faces in an image, with the help of face detectors.

1.3.5 Beam alignment

Beamforming is a critical technology that mitigates interference and improves signal coverage for multi-antenna communication. Beam alignment is the first step for beamforming, which finds critical applications in massive antenna communication in millimeter wave communications. As shown in the left part of Figure 1.6, in the uplink communication system, the task of beam alignment is to locate the angles of arrival of radio waves at the antenna. As illustrated in the right part of Figure 1.6, beam alignment corresponds to Twenty Questions with random noise, where the search region is the three-dimensional space centered at the base station and is divided into equal-sized sub-regions of angles. Here the search directions are unit vectors on the surface of a sphere.

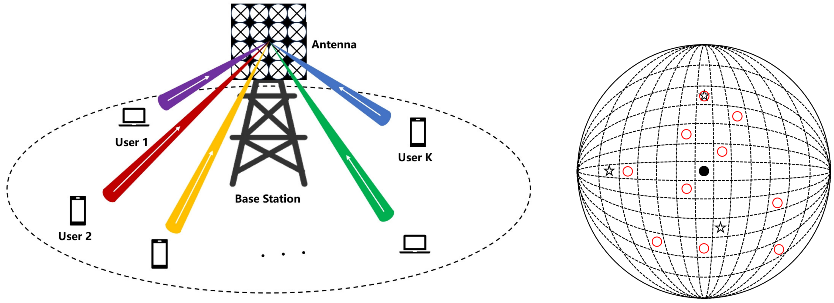


Figure 1.6: Beam direction detection in the uplink of a wireless communication system. To provide high-quality communication, the base station needs to estimate the angle of arrival of each device, which is known as beam alignment (left figure). The right figure illustrates of analogy between twenty questions with error and beam alignment, where the black dot denotes the base station and the stars denote the angles of arrived waves sent by user devices. Here each question is a possible set of sub-regions (red circles) in the beam direction space that contains radio waves.

The secret is the sub-regions that contain the true directions of arrived waves, the question is whether any wave exists in the chosen sub-regions.

1.3.6 Defective Element Detection in IRS

Intelligent reflecting surface (IRS) is a key technology for 6G communications, which has significantly improved the performance of wireless communications [62], [101]. As shown in the left part of Figure 1.7, in an IRS-assisted communication system, the transmitter improves the communication performance via the help of IRS by constructing an additional transmission link. However, the reflecting elements in IRS are susceptible to defection and could not generate the desired phase shifts [106]. To fix the problem, one needs to localize all defective elements. As illustrated in the right part of Figure 1.7, defective element detection in IRS is related to Twenty Questions with random error. The search region is all possible combinations of elements in an IRS, the secret is the defective elements denoted in gray, each question asks whether any defective element exists in a particular chosen sub-region and oracle's noisy response corresponds to potentially incorrect detection due to noise in the transmission channel.

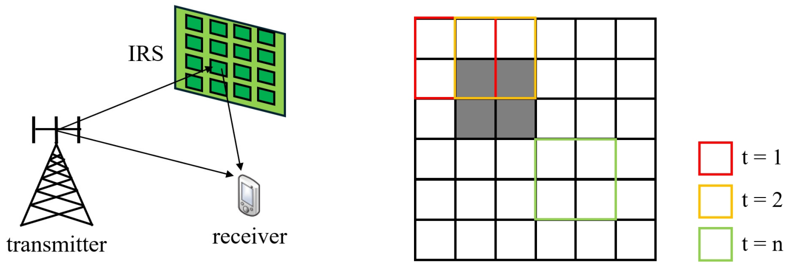


Figure 1.7: Defective element detection in an IRS-assisted communication system. To ensure performance improvement, the reflecting elements in an IRS should work properly and generates desired phase shifts as required (left figure). However, these elements are susceptible to failure and lead to poor performance. To solve this problem, query procedures for twenty questions can be used and the right figure illustrates of analogy between twenty questions with error and defective element detection in IRS, where the gray region denotes the defective elements. Here each question corresponds to a possible combination of the reflecting elements that can contain defective ones. See the squares with contours different colors for the query sets of t-th question.

1.4 Twenty Questions with Random Error

In this section, we present Rényi's formulation of Twenty Questions with random error for a continuous secret random variable, present the definition of the noise model that characterizes the noisy response behavior of the oracle, clarify the differences between adaptive and non-adaptive query procedures and define the theoretical limits.

1.4.1 Notation

Random variables and their realizations are denoted by upper case variables (e.g., X) and lower case variables (e.g., x), respectively. Sets are denoted in calligraphic font (e.g., $\mathcal{X}$). We use $X^n := (X_1, \ldots, X_n)$ to denote a random vector with n components. We use $\Phi(\cdot)$ and $\Phi^{-1}(\cdot)$ to denote the cumulative distribution function (cdf) of the standard Gaussian and its inverse function, respectively. We use $\mathbb{R}$, $\mathbb{R}_+$ and $\mathbb{N}$ to denote the sets of real numbers, positive real numbers and integers respectively. Given any two integers $(m, n) \in \mathbb{N}^2$, we use $[m : n]$ to denote the set of integers $\{m, m+1, \ldots, n\}$ and use $[m]$ to denote $[1:m]$. Given any $(m,n) \in \mathbb{N}^2$, $m \le n$, for any m by n matrix $\mathbf{a} = \{a_{i,j}\}_{i\in[m],j\in[n]}$, the infinity norm is defined as $\|\mathbf{a}\|_\infty := \max_{i\in[m],j\in[n]} |a_{i,j}|$. The set of all probability mass functions on a finite set $\mathcal{X}$ is denoted as $\mathcal{P}(\mathcal{X})$ and the set of all conditional probability distributions from $\mathcal{X}$ to $\mathcal{Y}$ is denoted as $\mathcal{P}(\mathcal{Y}|\mathcal{X})$. Furthermore, we use $\mathcal{F}(\mathcal{S})$ to denote the set of all probability density functions on a continuous set $\mathcal{S}$. All logarithms are base e. Finally, we use $\mathbb{1}(x \in \mathcal{S})$ to denote the indicator function that a variable x is contained in a set $\mathcal{S}$.

1.4.2 Problem Formulation

Let $S \in [0,1]$ be a continuous scalar random variable with arbitrary probability density function (pdf) P_S. As shown in Figure 1.8, in Twenty Questions with random error, a questioner aims to accurately estimate the secret random variable S by posing a sequence of questions to an oracle knowing S, where for each $t \in [n]$, the questioner asks whether the secret S lies in a certain query set $\mathcal{A}_t \subseteq [0,1]$. After receiving the query set $\mathcal{A}_t$, the oracle gives the binary answer $X_t = \mathbb{1}(S \in \mathcal{A}_t)$ but it is corrupted by noise through a channel with a transition matrix $P_{Y|X}^{\mathcal{A}_t}$, yielding noisy responses Y_t. If a total of n questions are posed by the questioner, based on the sequence of noisy responses Y^n from the oracle, the questioner uses a decision rule $g : \mathcal{Y}^n \to [0,1]$ to obtain an estimate $\hat{S}$ of the target variable S. When the pdf P_S is chosen so that the search space $[0,1]$ is quantized into a number of equal-sized regions, the above problem reduces to Twenty Questions estimation for a discrete secret random variable with bounded range, originally studied in the Rényi-Berlekamp-Ulam game.

A query procedure for Twenty Questions with random error consists of the query sets $\mathcal{A}^n = (\mathcal{A}_1, \ldots, \mathcal{A}_n) \subseteq [0,1]^n$ and a decision rule $g : \mathcal{Y}^n \to [0,1]$. The procedures can be classified into two categories: non-adaptive and adaptive. In non-adaptive querying, the questioner needs to determine the number of questions n and design all the query sets $\mathcal{A}^n$ simultaneously. In contrast, in adaptive querying, the questions are designed sequentially and the number of questions is a variable. Specifically, when designing the t-th query set $\mathcal{A}_t$, the questioner

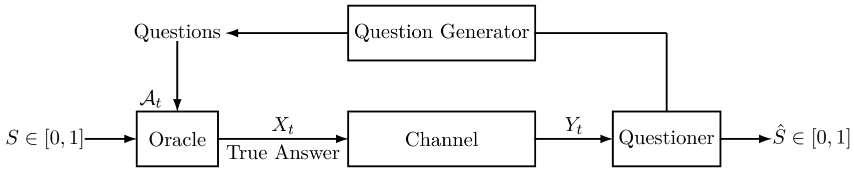
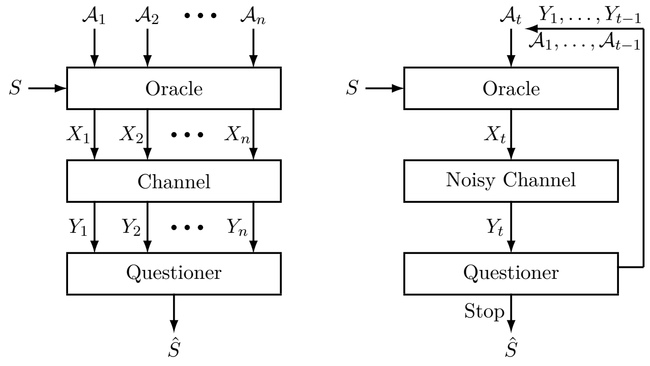


Figure 1.8: Block diagram for the problem of Twenty Questions with random error for estimation of a scalar variable S taking values in the unit interval $[0, 1]$. The questioner generates an estimate $\hat{S}$ by posing a sequence of questions to the oracle, who responses with true answer but the true answer is received with error by the questioner due to noise in the channel.



(a) Non-adaptive (b) Adaptive

Figure 1.9: Illustration of query procedures for Twenty Questions with random error. In the non-adaptive case (a), a secret random variable S is known to the oracle who responds to a sequence of questions with query sets $\mathcal{A}_1, \ldots, \mathcal{A}_n$ and provides binary responses $X_1, \ldots, X_n$, respectively. These responses are corrupted by a channel that outputs $Y_1, \ldots, Y_n$, which are subsequently used by the questioner to produce an estimate $\hat{S}$. In the adaptive case (b), the questions are posed sequentially and the questioner needs to determine a stopping time, specifying when to stop the query procedure.

can use previous questions and the noisy responses from the oracle, i.e., $\{(\mathcal{A}_1, Y_1), \ldots, (\mathcal{A}_{t-1}, Y_{t-1})\}$. Furthermore, the questioner needs to choose a potentially stochastic stopping criterion to determine when to stop asking questions. We illustrate the difference between non-adaptive and adaptive query procedures in Figure 1.9.

1.4.3 Query-Dependent Channel

The noisy response behavior of the oracle is modeled by a query-dependent channel having the transition probability $P^{\mathcal{A}}_{Y|X}$, where $\mathcal{A}$ is a query set, X is the oracle's response and Y is the noisy response received by the questioner. In this subsection, we present the definition and examples of $P^{\mathcal{A}}_{Y|X}$ in [111]. Such a channel can arise in settings where certain responses of the oracle are received with higher corruption than others.

Given any question with a query set $\mathcal{A}$, the correct answer is $X = \mathbb{1}(S \in \mathcal{A})$ and the potentially erroneous response Y from the oracle is the output of passing X over the channel $P^{\mathcal{A}}_{Y|X}$, which is assumed a function of the query set. Specifically, given any $\mathcal{A} \subseteq [0,1]$, let $|\mathcal{A}| = \int_{t\in\mathcal{A}} \mathrm{d}t$ be its size. Assume that the query-dependent channel $P^{\mathcal{A}}_{Y|X}$ depends on the query set $\mathcal{A}$ only through its size $|\mathcal{A}|$. In this case, $P^{\mathcal{A}}_{Y|X}$ is equivalent to a channel with state $P^{q}_{Y|X}$ [40, Chapter 7], where the state q is a function of the query set size, i.e., $q = f(|\mathcal{A}|)$, and we assume that $f : [0,1] \to \mathbb{R}_+$ is a bounded Lipschitz continuous function with parameter K.

Four types of query-dependent channels are analyzed in this monograph. The first type is discrete with symmetric Bernoulli noise.

Definition 1.1. Given any $\mathcal{A} \subseteq [0,1]$, the channel $P^{\mathcal{A}}_{Y|X}$ is said to be a query-dependent Binary Symmetric Channel (BSC) with parameter $\zeta \in (0,1]$ if $\mathcal{X} = \mathcal{Y} = \{0,1\}$ and for any $(x,y) \in \mathcal{X} \times \mathcal{Y}$,

$$P^{\mathcal{A}}_{Y|X}(y|x) = (\zeta f(|\mathcal{A}|))^{\mathbb{1}(y\neq x)}(1-\zeta f(|\mathcal{A}|))^{\mathbb{1}(y=x)}. \tag{1.1}$$

Note that the output of a query-dependent BSC with parameter ζ is the input flipped with probability $\zeta f(|\mathcal{A}|)$. The channel models an error pattern corresponding to a lying oracle, where the oracle lies randomly to each question. The channel was introduced in [111] and generalizes the query-dependent Bernoulli noise model in [54], where $\zeta = 1$ and $f(|\mathcal{A}|) = |\mathcal{A}|$.

The second type is discrete with asymmetric Bernoulli noise.

Definition 1.2. Given any $\mathcal{A} \subseteq [0, 1]$, the channel $P^{\mathcal{A}}_{Y|X}$ is said to be a query-dependent Z-channel with parameter $\zeta \in [0, 1]$ if $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ and for any $(x, y) \in \mathcal{X} \times \mathcal{Y}$,

$$P^{\mathcal{A}}_{Y|X}(y|x) = (1 - \zeta f(|\mathcal{A}|))^{\mathbb{1}(y=x=1)}(\zeta f(|\mathcal{A}|))^{\mathbb{1}(y=0,x=1)} \times (0)^{\mathbb{1}(y=1,x=0)}. \tag{1.2}$$

The binary output bit of a query-dependent Z-channel is flipped with probability $\zeta|\mathcal{A}|$ only if the input is $x = 1$. The channel models the half-lie error pattern of the oracle [37], where the oracle randomly lies to each question only if the correct answer to the question is yes.

The third type is discrete with erasure noise.

Definition 1.3. Given any $\mathcal{A} \subseteq [0, 1]$, the channel $P^{\mathcal{A}}_{Y|X}$ is said to be a query-dependent Binary Erasure Channel (BEC) with parameter $\tau \in [0, 1]$ if $\mathcal{X} = \{0, 1\}$, $\mathcal{Y} = \{0, 1, \mathrm{e}\}$ and for any $(x, y) \in \mathcal{X} \times \mathcal{Y}$,

$$P^{\mathcal{A}}_{Y|X}(y|x) = (1 - \tau f(|\mathcal{A}|))^{\mathbb{1}(y=x)}(f(|\mathcal{A}|))^{\mathbb{1}(y=\mathrm{e})} \tag{1.3}$$

The binary output bit of a query-dependent BEC with parameter τ is erased with probability $\tau f(|\mathcal{A}|)$. The channel models the error pattern of the oracle, where the oracle randomly refuses to answer the question, corresponding to an erasure "e".

The fourth type is continuous with Gaussian noise.

Definition 1.4. Given any $\mathcal{A} \subseteq [0, 1]$, the channel $P^{\mathcal{A}}_{Y|X}$ is said to be a query-dependent AWGN channel with parameter $\sigma \in \mathbb{R}_+$ if $\mathcal{X} = \{-1, 1\}$, $\mathcal{Y} = \mathbb{R}$ and for any $(x, y) \in \mathcal{X} \times \mathcal{Y}$,

$$P^{\mathcal{A}}_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi(f(|\mathcal{A}|)\sigma)^2}} \exp\left(-\frac{(y-x)^2}{2(f(|\mathcal{A}|)\sigma)^2}\right). \tag{1.4}$$

The output of a query-dependent AWGN channel with parameter σ corrupts the correct answer with a random Gaussian noise with mean zero and variance $\sigma f(|\mathcal{A}|)$. The channel models the error pattern of the oracle in practical applications such as localizing targets using a sensor network, where the correct answers provided by each sensor is corrupted by additive Gaussian noise during the transmission process. This definition first appeared in [59].

For the special case where $f(\mathcal{A})$ equals a constant, the above definitions of a query-dependent channel reduces to a query-independent channel that was used many early studies of Twenty Questions with random error, e.g., [21], [51], [69], [77], [99]. In this case, regardless of the question, the oracle has the same error pattern. In Section 1.5, we show that for the query-independent channel, the theoretical benchmarks of optimal non-adaptive and adaptive query procedures for Twenty Questions with random error are bounded by theoretical benchmarks of channel coding and variable-length channel coding with feedback, respectively.

1.4.4 Theoretical Benchmarks for Non-Adaptive Query Procedures

A non-adaptive query procedure is formally defined as follows.

Definition 1.5. Given any $(n, \delta, \varepsilon) \in \mathbb{N} \times \mathbb{R}_+ \times [0, 1]$, an (n, δ, ε)-non-adaptive query procedure consists of

- n questions with query sets $(\mathcal{A}_1, \ldots, \mathcal{A}_n) \in [0, 1]^n$
- an estimator $g : \mathcal{Y}^n \to [0, 1]$

such that the probability that the estimation error exceeds δ after n questions satisfies

$$\mathrm{P_e}(n, \delta) := \sup_{P_S \in \mathcal{F}([0,1])} \Pr\left\{|\hat{S} - S| > \delta\right\} \leq \varepsilon, \tag{1.5}$$

where $\hat{S} = g(Y^n)$ is the estimate and δ is called a resolution parameter.

Note that the probability in (1.5) is calculated with respect to distributions of the secret random variable S and the channel output Y^n since $\hat{S} = g(Y^n)$. Inspired by the definition of the excess-distortion probability for the rate-distortion problem [10], [56], [115], we call the probability on the left hand side of (1.5) the excess-resolution probability.

Definition 1.5 differs from the original setting of Rényi [77]. Specifically, Rényi considered a discrete secret random variable S that takes values in a finite set of integers. Definition 1.5 specializes to the setting of Rényi by restricting the set $\mathcal{F}([0,1])$ to probability mass functions with uniform quantization of $[0,1]$.

Motivated by practical scenarios where the number n of questions are limited, one is interested in the following non-asymptotic theoretical benchmark on the estimation error:

$$\delta^*(n,\varepsilon) := \inf\{\delta : \exists \text{ an } (n,\delta,\varepsilon)\text{−non−adaptive−procedure}\}. \quad (1.6)$$

Note that $\delta^*(n,\varepsilon)$ denotes the best resolution one can achieve with probability at least $1-\varepsilon$ using any non-adaptive query procedure with n questions. In other words, it is the resolution of optimal non-adaptive query procedures tolerating an excess-resolution probability of $\varepsilon \in [0,1]$. Dual to (1.6), one can also study the sample complexity by deriving the minimal number of questions required to achieve resolution δ with probability at least $1-\varepsilon$, i.e.,

$$n^*(\delta,\varepsilon) := \inf\{n : \exists \text{ an } (n,\delta,\varepsilon)\text{−non−adaptive−procedure}\}. \quad (1.7)$$

It follows that for any $(\delta,\varepsilon) \in \mathbb{R}_+ \times [0,1]$,

$$n^*(\delta,\varepsilon) = \inf\{n : \delta^*(n,\varepsilon) \leq \delta\}. \quad (1.8)$$

This motivates us to study $\delta^*(n,\varepsilon)$ as a theoretical benchmark.

1.4.5 Theoretical Benchmarks for Adaptive Query Procedures

An adaptive query procedure is formally defined as follows.

Definition 1.6. Given any $(l,\delta,\varepsilon) \in \mathbb{R}_+^2 \times [0,1]$, an (l,δ,ε)-adaptive query procedure consists of

- a sequence of questions where for each $t \in \mathbb{N}$, the design of the query set $\mathcal{A}_t \subseteq [0,1]$ depends on previous query sets $\{\mathcal{A}_j\}_{j\in[t-1]}$ and noisy responses Y^{t-1} from the oracle
- a sequence of decoding functions $g_t : \mathcal{Y}^t \to [0,1]$ for each $t \in \mathbb{N}$
- a random stopping time τ whose distribution depends on noisy responses $\{Y_t\}_{t\in\mathbb{N}}$

such that the average number of questions satisfies

$$\sup_{P_S \in \mathcal{F}([0,1])} \mathbb{E}[\tau] \leq l, \tag{1.9}$$

and analogously to the non-adaptive query procedure, the excess-resolution probability satisfies

$$\mathrm{P}_{\mathrm{e,a}}(l,\delta) := \sup_{P_S \in \mathcal{F}([0,1])} \Pr\{|\hat{S} - S| > \delta\} \leq \varepsilon. \tag{1.10}$$

Similar to the theoretical benchmark in (1.6) for non-adaptive querying, given any $(l,\varepsilon) \in \mathbb{R}_+ \times [0,1)$, the theoretical benchmark for adaptive querying is defined as follows:

$$\delta_{\mathrm{a}}^*(l,\varepsilon) := \inf\{\delta \in \mathbb{R}_+ : \exists \text{ an } (l,\delta,\varepsilon)\text{−adaptive query procedure}\}. \tag{1.11}$$

Thus, $\delta_{\mathrm{a}}^*(l,\varepsilon)$ is the minimal achievable resolution of an optimal adaptive query procedure, tolerating an excess-resolution probability of at most ε. Correspondingly, the sample complexity of adaptive querying is defined as

$$l^*(\delta,\varepsilon) := \inf\{l \in \mathbb{R}_+ : \exists \text{ an } (l,\delta,\varepsilon)\text{−adaptive query procedure}\}. \tag{1.12}$$

1.5 Relationship with Channel Coding

As pointed out by Rényi [78] and rediscovered in [21], [54], non-adaptive querying for Twenty Questions with random error is closely related to channel coding [86] and adaptive querying is closely related to variable-length channel coding with feedback (VLF) [50]. In this section, for the query-independent channel, we formally present this relationship by showing that the non-asymptotic theoretical benchmarks defined in (1.6) and (1.11) for non-adaptive and adaptive query procedures are bounded by non-asymptotic theoretical benchmarks for channel coding and VLF, respectively.

1.5.1 Channel Coding and Non-Adaptive Querying

Channel coding is a classical problem in information theory, proposed and studied by Shannon [86]. The goal of channel coding is to transmit a message W reliably over a memoryless channel defined by a transition matrix $P_{Y|X}$ where $X \in \mathcal{X}$ is known as the channel input and $Y \in \mathcal{Y}$ is known as the channel output.

Formally, a code that specifies the encoding and decoding operations of a source with M values for channel coding is defined as follows.

Definition 1.7. Given any $(n, M, \varepsilon) \in \mathbb{N}^2 \times [0, 1]$, an (n, M, ε)-code for the channel coding problem consists of

- an encoder $e : [M] \to \mathcal{X}^n$
- a decoder $\phi : \mathcal{Y}^n \to [M]$ such that

$$\mathrm{P}_{\mathrm{e}}^{(n)} := \sup_{P_W \in \mathcal{P}([M])} \Pr\{W \neq \phi(Y^n)\} \leq \varepsilon. \tag{1.13}$$

Note that in (1.13), the error probability is calculated with respect to the distributions of the message W and the channel output Y^n, which induced by the channel $P_{Y|X}$ and a channel input distribution P_X. In traditional channel coding, one often considers a uniformly distributed message W. However, in (1.13), we consider an *arbitrary* message distribution. Allowing nonuniform message distribution enables us to relate the theoretical benchmarks for non-adaptive query procedures for Twenty Questions with random error and channel coding through (1.5).

Given $(n, \varepsilon) \in \mathbb{N} \times [0, 1]$, the non-asymptotic theoretical benchmark for channel coding is defined as

$$M^*(n, \varepsilon) := \sup\{M : \ \exists \text{ an } (n, M, \varepsilon)\text{−code}\}. \tag{1.14}$$

Note that $M^*(n, \varepsilon)$ is the maximal number of messages that can be transmitted over n channel uses such that regardless of the message distribution, the average error probability is no greater than ε. When the message W is distributed uniformly over the set $[M]$, $M^*_{\mathrm{unif}}(n, \varepsilon)$ is defined similarly to maximal number of messages of any code with average error probability bounded by ε.

In the next theorem, the theoretical benchmark $\delta^*(n, \varepsilon)$ in (1.6) for optimal non-adaptive query procedures is related to the theoretical benchmarks $M^*(n, \varepsilon)$ and $M^*_{\mathrm{unif}}(n, \varepsilon)$ of channel coding. In this theorem, the channel input takes values in the binary input alphabet, i.e., $\mathcal{X} = \{0, 1\}$.

Theorem 1.1. The following claims hold for non-adaptive query procedures of Twenty Questions with random error for a query-independent channel.

1. For any $(n, M, \varepsilon) \in \mathbb{N}^2 \times [0, 1]$, any (n, M, ε)-code for channel coding specifies an $(n, \frac{1}{M}, \varepsilon)$-non-adaptive query procedure for Twenty Questions with random error using n questions, achieving excess-resolution probability no greater than ε with respect to the resolution $\delta = \frac{1}{M}$.

2. For any $(n, \delta, \varepsilon) \in \mathbb{N} \times \mathbb{R}_+ \times [0, 1]$, any (n, δ, ε)-non-adaptive query procedure for Twenty Questions with random error specifies an $(n, \lceil \frac{\beta}{\delta} \rceil, \varepsilon + 2\beta)$-code for channel coding for any $\beta \in (0, \frac{1-\varepsilon}{2})$, which ensures that the average error probability of the channel code is no greater than $\varepsilon + 2\beta$ when blocklength is n and the number of messages to be transmitted is $M = \lceil \frac{\beta}{\delta} \rceil$.

3. Given any $(n, \varepsilon) \in \mathbb{N} \times [0, 1]$, for any $\beta \in (0, \frac{1-\varepsilon}{2})$,

$$\log M^*(n, \varepsilon) \leq -\log \delta^*(n, \varepsilon) \leq \log M^*_{\mathrm{unif}}(n, \varepsilon + 2\beta) - \log \beta. \tag{1.15}$$

The proof of Theorem 1.1 is given in Section 1.6.

Claim 1 is proved by showing that a non-adaptive query procedure per Definition 1.5 for Twenty Questions with random error can be constructed using a channel code per Definition 1.7 with proper quantization of the secret random variable S. The intuition is that we can partition the unit interval $[0, 1]$ into equal size sub-intervals and treat the index of the sub-interval where the secret random variable S lies in as the message to be transmitted over the channel. Claim 2 is proved by showing that a channel code can be constructed from a non-adaptive query procedure for Twenty Questions, when the random secret random variable is specialized to be uniformly distributed. The parameter β appears since the quantization interval for Claim 2 differs from that of Claim 1. Claim 3 follows directly Claims 1 and 2, and gives a bound on the minimal achievable resolution of an optimal non-adaptive query procedure in terms of the maximal message size of an optimal channel code.

The result in (1.15) allows one to approximate the performance of an optimal non-adaptive query procedure with a finite number n of questions using theoretical benchmarks for finite blocklength channel coding [49], [72], assuming that the noisy response of the oracle does not the depend on questions.

1.5.2 Variable-Length Channel Coding with Feedback and Adaptive Querying

Compared with channel coding per Definition 1.7, in VLF, the encoder can transmit codewords whose lengths can vary for each message. For VLF, the decoder needs to determine the stopping time at which point it will make an estimate of the transmitted message. A formal definition of a code for VLF [74] is as follows.

Definition 1.8. Given any $(l, M, \varepsilon) \in \mathbb{R}_+ \times \mathbb{N} \times [0, 1]$, an (l, M, ε)-code for VLF consists of

- a sequence of encoders $e_n : [M] \times \mathcal{Y}^{n-1} \to \mathcal{X}$ for each $n \in \mathbb{N}$
- a sequence of decoders $\phi_n : \mathcal{Y}^n \to [M]$ for each $n \in \mathbb{N}$
- a random stopping time τ as a function of the channel outputs $\{Y_1, Y_2, \ldots\}$ such that

$$\sup_{P_W \in \mathcal{P}([M])} \mathsf{E}[\tau] \leq l \tag{1.16}$$

and

$$\mathrm{P}_{\mathrm{e}}^{(\tau)} := \sup_{P_W \in \mathcal{P}([M])} \Pr\{W \neq \phi(Y^\tau)\} \leq \varepsilon. \tag{1.17}$$

Analogous to $M^*(n, \varepsilon)$ in (1.14), for any $(l, \varepsilon) \in \mathbb{R}_+ \times [0, 1]$, the non-asymptotic theoretical benchmark on the maximal message size for VLF is defined as

$$M_{\mathrm{f}}^*(l, \varepsilon) := \sup\{M : \ \exists \text{ an } (l, M, \varepsilon)\text{−code}\}. \tag{1.18}$$

Furthermore, let $M^*_{\mathrm{f,unif}}(n,\varepsilon)$ denote the theoretical benchmark on the maximal message size when the message W is uniformly distributed.

Analogous to Theorem 1.1, we have the following theorem, which relates optimal codes for VLF to the theoretical benchmarks for optimal adaptive query procedures for Twenty Questions with random error.

Theorem 1.2. The following claims hold for adaptive query procedures of Twenty Questions with random error for a query-independent channel.

1. For any $(l,M,\varepsilon)\in\mathbb{R}_+\times\mathbb{N}\times[0,1]$, any (l,M,ε)-code for VLF specifies an $(l,\frac{1}{M},\varepsilon)$-adaptive query procedure for Twenty Questions with random error, where the average number of questions is no more than l, the resolution is $\delta=\frac{1}{M}$ and the excess-resolution probability is no greater than ε.

2. For any $(l,\delta,\varepsilon)\in\mathbb{R}_+^2\times[0,1]$, any (l,δ,ε)-adaptive query procedure for Twenty Questions with random error specifies an $(l,\lceil\frac{\beta}{\delta}\rceil,\varepsilon+2\beta)$-code for VLF for any $\beta\in(0,\frac{1-\varepsilon}{2})$, where the average blocklength is upper bounded by l, the message size is $M=\lceil\frac{\beta}{\delta}\rceil$ and the average error probability is no greater than $\varepsilon+2\beta$.

3. Given any $(l,\varepsilon)\in\mathbb{R}_+\times[0,1]$, for any $\beta\in(0,\frac{1-\varepsilon}{2})$,
$$\log M^*_{\mathrm{f}}(l,\varepsilon)\le-\log\delta^*_{\mathrm{a}}(n,\varepsilon)\le\log M^*_{\mathrm{f,unif}}(l,\varepsilon+2\beta)-\log\beta. \tag{1.19}$$

The proof of Theorem 1.2 parallels that of Theorem 1.1 and is thus omitted. The remarks of Theorem 1.1 also apply here.

Theorems 1.1 and 1.2 together provide the key link between channel coding and Twenty Questions with random error. However, both theorems only hold for the restrictive case where the noisy response behavior of the oracle does not depend on the questions.

In many practical applications such as target localization using a sensor network, the noisy response behavior of the oracle may depend on the questions. This could be due to a "lying oracle" or through variations in noise power when questions involve different sensors [17]. For this case discussed in Section 1.4.3, the theoretical benchmarks for Twenty Questions with random error are more complicated and no corresponding finite blocklength results for channel coding or VLF currently exist. In the following four sections of this monograph, we present recent advances on this query-dependent channel case for four different settings of Twenty Questions with random error by generalizing the intuition conveyed in Theorems 1.1 and 1.2 towards practical applications mentioned in Section 1.3. In particular, we consider a multi-dimensional secret random variable, which corresponds to a real-valued target in three-dimensional space; we consider a time varying secret random variable, which models the trajectory of a moving target, e.g., changing face locations in a video; we consider multiple secret random variables, which models multiple objects to be located or multiple beams to be aligned.

1.6 Proof of Relationship with Channel Coding

In this technical section, we formally prove Theorem 1.1, where the noisy response behavior of the oracle does not depend on the questions.

1.6.1 Construct a Non-adaptive Query Procedure Using a Channel Code

Claim 1 and the first inequality of Claim 3 of Theorem 1.1 are implied by the following lemma.

Lemma 1.3. For any $(n, M, \varepsilon) \in \mathbb{N}^2 \times [0, 1]$, given any (n, M, ε)-code for channel coding per Definition 1.7, we can construct an $(n, \frac{1}{M}, \varepsilon)$-non-adaptive query procedure per Definition 1.5. It follows that, for any $(n, \varepsilon) \in \mathbb{N} \times [0, 1]$,

$$-\log \delta^*(n, \varepsilon) \geq \log M^*(n, \varepsilon). \tag{1.20}$$

Proof. Fix any (n, M, ε)-code with encoding function e and decoding function ϕ. For each $m \in [M]$, let $X^n(m) := e(m)$ be the channel input (codeword) corresponding to message m and let $\mathcal{C} := \{X^n(m)\}_{m\in[M]}$ be the collection of all codewords. In the following, we construct a non-adaptive query procedure using the (n, M, ε)-code for channel coding.

We first partition the unit interval $\mathcal{S} := [0, 1]$ into M equal-sized sub-intervals $\{\mathcal{S}_m\}_{j\in[M]}$, each with length $\frac{1}{M}$. Given any $s \in [0, 1]$, define the following quantization function

$$\mathrm{q}(s) := \lceil sM \rceil, \tag{1.21}$$

which is exactly the index of the sub-interval in which s lies.

For each $t \in [n]$, the t-th query set is designed as follows:

$$\mathcal{A}_t := \bigcup_{m \in [M]: X_t(m)=1} \mathcal{S}_m, \tag{1.22}$$

where $X_t(m)$ denotes the t-th element of the m-th codeword $X^n(m)$. Thus, the t-th question is whether the secret random variable S lies in the union of sub-intervals with indices of the codewords whose t-th element are one. Hence, for each $t \in [n]$, the t-th element of each codeword is an indicator function for whether a particular sub-interval is queried in the t-th question, with one being affirmative and zero being negative.

Fix any $s \in \mathcal{S}$ and $w \in [M]$ such that $\mathrm{q}(s) = w$. For each $t \in [n]$, using the query procedure in (1.22), the correct response Z_t from the oracle is

$$Z_t = 1\{s \in \mathcal{A}_t\} \tag{1.23}$$

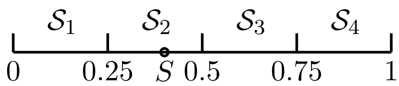
$$= 1\Big\{s \in \bigcup_{j \in [M]: X_t(j)=1} \mathcal{S}_j\Big\} \tag{1.24}$$

$$= \begin{cases} 1 & \text{if } X_t(w) = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1.25}$$

Thus, the correct response is $Z^n = X^n(\mathrm{q}(s))$, i.e., the $\mathrm{q}(s)$-th codeword is the true answer the n questions. We provide an illustration through an example in Figure 1.10.

The oracle's correct responses Z^n are corrupted by random noise in the channel, yielding noisy response Y^n that are provided to the questioner. Given Y^n, the questioner uses the channel decoder ϕ to produce an estimate $\hat{S}$ of s by first decoding $\mathrm{q}(s)$ as $\hat{W} = \phi(Y^n)$ and then declaring the mean value in $\hat{W}$-th sub-interval as the final estimate, i.e.,

$$\hat{S} = g(Y^n) = \frac{\phi(Y^n)}{2M} = \frac{2\hat{W} - 1}{2M}. \tag{1.26}$$



$X^n(1)$	1	0	0	1	1	0
$X^n(2)$	0	1	1	1	0	1
$X^n(3)$	1	0	1	0	1	1
$X^n(4)$	1	1	1	0	1	0

Figure 1.10: Illustration of question design and correct responses when $s = 0.4$, $M = 4$ and $n = 6$. The top figure shows that the unit interval $[0, 1]$ is partitioned into $M = 4$ equal size sub-intervals: $\mathcal{S}_1 = [0, 0.25)$, $\mathcal{S}_2 = [0.25, 0.5)$, $\mathcal{S}_3 = [0.5, 0.75)$ and $\mathcal{S}_4 = [0.75, 1]$. Since $s = 0.4$, the quantized value is thus $\mathrm{q}(s) = 2$ and s lies in the second sub-interval $\mathcal{S}_2$. The bottom figure shows encoder matrix for $M = 4$ codewords $(X^n(1), \ldots, X^n(4))$, each of which have $n = 6$ components. According to question design in (1.22), the first question is whether s lies in the union of sub-intervals with indices $m \in [M]$ such that $X_1(m) = 1$, i.e., the union of the first, third and the fourth sub-intervals. The correct response to the first question is negative, i.e., $Z_1 = 0 = X_1(2)$. Continuing this process of querying the oracle, the correct responses Z^n is $X^n(2)$, corresponding to the codeword with index $\mathrm{q}(s) = 2$.

For any $s \in \mathcal{S}$, it is easy to verify that the estimate $\hat{S}$ is within $\frac{1}{M}$ of s if $\hat{W} = \mathrm{q}(s)$. Thus, it follows from the definition of an (n, M, ε)-code in Definition 1.7 that

$$\sup_{P_S} \Pr\left\{|g(Y^n) - S| > \frac{1}{M}\right\} \leq \sup_{P_S} \Pr\{\mathrm{q}(S) \neq \phi(Y^n)\} \tag{1.27}$$

$$= \sup_{P_W} \Pr\{W \neq \phi(Y^n)\} \tag{1.28}$$

$$\leq \varepsilon, \tag{1.29}$$

Therefore, we have constructed an $(n, \frac{1}{M}, \varepsilon)$-non-adaptive query procedure per Definition 1.5 using an (n, M, ε)-code for channel coding per Definition 1.7. Thus, using (1.6) and (1.14), it follows that for any $(n, \varepsilon) \in \mathbb{N} \times [0, 1]$,

$$-\log \delta^*(n, \varepsilon) \geq \log M^*(n, \varepsilon). \tag{1.30}$$

The proofs for Claim 1 and the first inequality in Claim 3 of Theorem 1.1 is now completed. □

1.6.2 Construct a Channel Code Using a Non-adaptive Query Procedure

Claim 2 and the second inequality in Claim 3 of Theorem 1.1 are implied by the following lemma.

Lemma 1.4. For any $(n,\delta,\varepsilon)\in\mathbb{N}\times\mathbb{R}_+\times[0,1]$, given an (n,δ,ε)-non-adaptive query procedure for Twenty Questions with random error, for any $\beta\in(0,\frac{1-\varepsilon}{2})$, we can construct an $(n,\lceil\frac{\beta}{\delta}\rceil,\varepsilon+2\beta)$-code for channel coding for a uniformly distributed message W. Therefore, given any $(n,\varepsilon)\in\mathbb{N}\times[0,1]$, for any $\beta\in(0,\frac{1-\varepsilon}{2})$,

$$-\log\delta^*(n,\varepsilon)\le\log M^*_{\mathrm{unif}}(n,\varepsilon+2\beta)-\log\beta. \tag{1.31}$$

Proof. Given any $\varepsilon\in[0,1]$, let $\beta\in(0,\frac{1-\varepsilon}{2})\subseteq(0,0.5)$ be arbitrary. Partition the unit interval $\mathcal{S}=[0,1]$ interval into $\tilde{M}:=\lceil\frac{\beta}{\delta}\rceil$ equal-sized sub-intervals, each with length $\frac{1}{\tilde{M}}$. For any $s\in\mathcal{S}$, define the following quantization function

$$\mathrm{q}_\beta(s):=\lceil s\tilde{M}\rceil. \tag{1.32}$$

Let P_S^{u} be the uniform distribution over $\mathcal{S}$. For any (n,δ,ε)-non-adaptive query procedure with query sets $\{\mathcal{A}_t\}_{t\in[n]}$ and a decoder $g:\mathcal{Y}^n\to\mathcal{S}$, it follows that

$$\varepsilon\ge\sup_{P_S\in\mathcal{F}(\mathcal{S})}\Pr\{|S-\hat{S}|>\delta\}\ge\Pr_{P_S^{\mathrm{U}}}\{|S-\hat{S}|>\delta\}. \tag{1.33}$$

In the following, we show how to construct the encoding function e and the decoding function ϕ for channel coding using the above (n,δ,ε)-non-adaptive query procedure. The encoding function $e:[\tilde{M}]\to\{0,1\}^n$ is constructed such that for each $w\in[\tilde{M}]$, $e(w)$ is the correct response to question $\{\mathcal{A}_t\}_{t\in[n]}$ when the secret variable is $s=\frac{2w-1}{2\tilde{M}}$, i.e.,

$$e(w)=(Z_1,\ldots,Z_n) \tag{1.34}$$
$$=\left(1\left\{\frac{2w-1}{2\tilde{M}}\in\mathcal{A}_1\right\},\ldots,1\left\{\frac{2w-1}{2\tilde{M}}\in\mathcal{A}_n\right\}\right). \tag{1.35}$$

Given noisy responses Y^n, the channel decoding function $\phi:\mathcal{Y}^n\to[\tilde{M}]$ is constructed as follows:

$$\phi(Y^n):=\mathrm{q}_\beta(g(Y^n)). \tag{1.36}$$

We next bound the error probability of the constructed channel code. For simplicity, let $\hat{S} = g(Y^n)$, $\tilde{W} = \mathrm{q}_\beta(S)$ and $\hat{\tilde{W}} = \phi(Y^n)$. When the secret random variable S is uniformly distributed over $\mathcal{S}$, the induced random variable $\tilde{W}$ is uniformly distributed over $[\tilde{M}]$. It follows that

$$
\begin{aligned}
&\Pr\{W \neq g(Y^n)\} \\
&= \Pr\{\hat{W} \neq W,\ |\hat{S} - S| > \delta\} + \Pr\{W \neq \hat{W},\ |\hat{S} - S| \leq \delta\} && (1.37) \\
&\leq \Pr\{|\hat{S} - S| > \delta\} + \Pr\{\hat{W} \neq W,\ |\hat{S} - S| \leq \delta\} && (1.38) \\
&\leq \varepsilon + \Pr\{W \neq \hat{W},\ |\hat{S} - S| \leq \delta\} && (1.39) \\
&\leq \varepsilon + \frac{2\delta}{\tilde{M}} && (1.40) \\
&\leq \varepsilon + 2\beta, && (1.41)
\end{aligned}
$$

where (1.39) follows from (1.33), (1.40) follows since i) the events $\hat{W} \neq W$ and $|S - \hat{S}| \leq \delta$ occur simultaneously only when S is within δ to the boundaries (left and right) of the sub-interval with indices $W = \mathrm{q}_\beta(S)$, ii) S is uniformly distributed over $\mathcal{S} = [0, 1]$ and thus iii) the probability of the event $\{\hat{W} \neq W,\ |S - \hat{S}| \leq \delta\}$ is upper bounded by $\frac{2\delta}{\tilde{M}}$, and (1.41) follows from the definition of $\tilde{M}$. A figure illustrating (1.40) is provided in Figure 1.11.

Figure 1.11: Figure illustration of (1.40). Let $\delta = \frac{1}{600}$ and $\beta = 6$. Thus, we partition the unit interval $\mathcal{S}$ into $\tilde{M} = 100$ sub-intervals each with length $\beta\delta = \frac{1}{100}$. In the figure, we plot three consecutive sub-intervals with indices $(k-1, k, k+1)$ for some $k \in [2 : \tilde{M} - 1]$. Note that the k-th interval starts from $\beta(k-1)\delta$ and end at $\beta k\delta$ and contains β smaller intervals, each of length δ. Suppose S lies in k-th sub-interval. One can find $\hat{S}$ in adjacent sub-interval such that $|\hat{S} - S| \leq \delta$ and $\mathrm{q}_\beta(\hat{S}) = \hat{W} \neq W = \mathrm{q}_\beta(S)$ only if S are with $\delta = \frac{1}{100}$ of the boundaries in k-th sub-interval, denoted by shaded color.

Using (1.41), we conclude that an $(n, \tilde{M}, \varepsilon + 2\beta)$-code for the channel coding with uniformly distributed message W can be constructed from an (n, δ, ε)-non-adaptive query procedure for Twenty Questions with random error. Thus, given any (n, ε), for any $\beta \in (0, \frac{1-\varepsilon}{2})$,

$$\log M^*_{\mathrm{unif}}(n, \varepsilon + 2\beta) \geq \log \left\lceil \frac{\beta}{\delta^*(n, \varepsilon)} \right\rceil \geq -\log \delta^*(n, \varepsilon) + \log \beta. \quad (1.42)$$

□

1.7 Basic Definitions and Mathematical Tools

The main tools that we use are the method of types [28] to bound the performance loss of transforming a query-dependent channel to a particular query-independent channel and the Berry-Esseen theorem to derive approximations to non-asymptotic theoretical benchmarks. Some of the mathematical background is provided below.

1.7.1 Basic Definitions

Given any distribution $P_X \in \mathcal{P}(\mathcal{X})$ defined on a finite alphabet $\mathcal{X}$, the entropy is defined as

$$H(X) = H(P_X) := -\sum_{x \in \mathrm{supp}(P_X)} P_X(x) \log P_X(x). \quad (1.43)$$

Note that the notation $H(X)$ is used in classical textbook [27] and the notation $H(P_X)$ is often used to clarify the dependence of the entropy on the distribution [29]. We use both notations for the entropy and other information theoretical quantities interchangeably. Analogously, given a joint probability mass function (PMF) $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ defined on a finite alphabet $\mathcal{X} \times \mathcal{Y}$, the joint entropy is defined as

$$H(X, Y) = H(P_{XY}) = -\sum_{(x,y) \in \mathrm{supp}(P_{XY})} P_{XY}(x, y) \log P_{XY}(x, y), \quad (1.44)$$

and the conditional entropy of X given Y is defined as

$$\begin{aligned} H(X|Y) &= H(P_{X|Y}|P_Y) \\ &= -\sum_{(x,y) \in \mathrm{supp}(P_{XY})} P_{XY}(x, y) \log P_{X|Y}(x|y), \end{aligned} \quad (1.45)$$

where $(P_{X|Y}, P_Y)$ are the induced conditional and marginal distributions of P_{XY}. Furthermore, the mutual information that measures dependence of two random variables (X,Y) with joint distribution P_{XY} is defined as

$$I(X;Y) = I(P_X, P_{X|Y}) = H(P_X) - H(P_{X|Y}|P_Y). \tag{1.46}$$

Note that mutual information $I(X;Y)$ is symmetric so that $I(X;Y) = I(Y;X)$. Similar to the definition of entropy, we use $I(X;Y)$ and the distribution dependent version $I(P_X, P_{X|Y})$ interchangeably. Analogously, given the joint distribution P_{XYZ} of three random variables (X,Y,Z) defined on a finite alphabet $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, define the conditional mutual information $I(X;Y|Z)$ as

$$I(X;Y|Z) = I(P_{X|Z}, P_{X|YZ}|P_Z) \tag{1.47}$$

$$= H(P_{X|Z}|P_Z) - H(P_{X|YZ}|P_{YZ}), \tag{1.48}$$

where all distributions are induced by the joint distribution P_{XYZ}.

1.7.2 Method of Types

Since we focus on binary input discrete memoryless channel as described in Section 1.4.3, the method of types plays a critical role in the analyses in subsequent sections of this monograph. We recall some of the machinery here, see also [28], [27, Chapter 11] and [29, Chapter 2]). Fix a finite alphabet $\mathcal{X}$ and an integer $n \in \mathbb{N}$. Given a length-n binary sequence $x^n \in \mathcal{X}^n$, the empirical distribution (type) $\hat{T}_{x^n}$ is defined such that for each $a \in \mathcal{X}$,

$$\hat{T}_{x^n}(a) = \frac{1}{n} \sum_{t\in[n]} 1\{x_t = a\}. \tag{1.49}$$

The set of types formed from length-n sequences that take values in $\mathcal{X}$ is denoted by $\mathcal{P}_n(\mathcal{X})$. Given a type $P \in \mathcal{P}_n(\mathcal{X})$, the set of all sequences x^n such that $\hat{T}_{x^n} = P$ forms the type class, which is denoted by $\mathcal{T}_P^n$. For any $n \in \mathbb{N}$, the number of types satisfies

$$|\mathcal{P}_n(\mathcal{X})| \leq (n+1)^{|\mathcal{X}|}. \tag{1.50}$$

For any type $P \in \mathcal{P}_n(\mathcal{X})$, the size of the type class $\mathcal{T}_P^n$ satisfies

$$(n+1)^{-|\mathcal{X}|} \exp(nH(P_X)) \leq |\mathcal{T}_{P_X}^n| \leq \exp(nH(P_X)). \tag{1.51}$$

For any sequence x^n that is generated i.i.d. from a distribution $P_X \in \mathcal{P}(\mathcal{X})$, its probability satisfies

$$P_X^n(x^n) = \exp\big(-n(D(\hat{T}_{x^n}\|P_X) + H(\hat{T}_{x^n}))\big). \tag{1.52}$$

Thus, for any type $Q \in \mathcal{P}_n(\mathcal{X})$, the probability of the type class $\mathcal{T}_Q^n$ satisfies

$$(n+1)^{-|\mathcal{X}|} \le \frac{P_X^n(\mathcal{T}_Q^n)}{\exp(-nD(Q\|P_X))} \le 1. \tag{1.53}$$

Fix another finite alphabet $\mathcal{Y}$. Given any two sequences $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$, the joint empirical distribution (type) $\hat{T}_{x^n y^n}$ is defined such that for any $(a,b) \in \mathcal{X} \times \mathcal{Y}$,

$$\hat{T}_{x^n y^n}(a,b) = \frac{1}{n}\sum_{t\in[n]} 1\{(x_t, y_t) = (a,b)\}. \tag{1.54}$$

Given any $x^n \in \mathcal{X}^n$ and conditional distribution $V \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$, the set of all sequences $y^n \in \mathcal{Y}^n$ such that $\hat{T}_{x^n y^n} = \hat{T}_{x^n} \times V$ forms the conditional type class, which is denoted by $\mathcal{T}_V(x^n)$. Fix any $P \in \mathcal{P}_n(\mathcal{X})$. The set of all conditional distributions $V \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ such that the set $\mathcal{T}_V(x^n)$ is not empty for some $x^n \in \mathcal{T}_P^n$ forms the set of conditional types given the marginal type P, which is denoted by $\mathcal{V}^n(\mathcal{Y}; P)$.

1.7.3 Mathematical Tools

In the proof of theoretical results of the monograph, we use a sequence of concentration inequalities [31]: the weak law of large numbers, the Markov inequality and the Berry-Esseen theorem. In this subsection, we recall these mathematical tools.

Let $X^n = (X_1, \ldots, X_n)$ be a collection of n i.i.d. random variables with zero mean and variance σ^2 and let the normalized sum of these n random variables be

$$S_n := \frac{1}{n}\sum_{t\in[n]} X_t. \tag{1.55}$$

We first recall the weak law of large numbers, which states that the normalized sum S_n converges in probability to its mean.

Theorem 1.5 (The Weak Law of Large Numbers)**.** For any positive real number $\delta \in \mathbb{R}_+$,

$$\lim_{n\to\infty} \Pr\{S_n > \delta\} = 0. \tag{1.56}$$

In the proofs of many theorems, the Markov inequality is used.

Theorem 1.6 (The Markov Inequality)**.** For any non-negative real number $\theta \in \mathbb{R}_+$ and any positive real number t,

$$\Pr\{S_n > t\} \le \frac{\mathsf{E}[\exp(\theta S_n)]}{\exp(t\theta)}. \tag{1.57}$$

The Berry-Esseen Theorem for i.i.d. random variables [12], [41] will be used to derive theoretical benchmarks.

Theorem 1.7 (The Berry-Esseen Theorem)**.** Assume that the third absolute moment of X_1 is finite, i.e., $T := \mathbb{E}|X_1|^3 < \infty$. For each $n \in \mathbb{N}$,

$$\sup_{t\in\mathbb{R}} \left| \Pr\left\{ S_n < t\sqrt{\frac{\sigma^2}{n}} \right\} - \Phi(t) \right| \le \frac{T}{\sigma^3\sqrt{n}}, \tag{1.58}$$

where $\Phi(t)$ is the standard Gaussian cumulative distribution function evaluated at a real number $t \in \mathbb{R}$.

To tackle certain problems, we need to consider independent but not identically distributed (i.n.i.d.) random variables. Let $X^n = (X_1, \ldots, X_n)$ be a sequence of random variables, where for each $m \in [n]$, the random variable X_t has zero mean, variance $\sigma_m^2 := \mathsf{E}[X_m^2] > 0$ and finite third-absolute moment $T_m := \mathsf{E}[|X_m|^3]$. Define the average variance and third-absolute moment as follows:

$$\sigma^2 := \frac{1}{n}\sum_{m\in[n]} \sigma_m^2, \tag{1.59}$$

$$T := \frac{1}{n}\sum_{m\in[n]} T_m. \tag{1.60}$$

The Berry-Esseen theorem for i.n.i.d. random variables is as follows.

Theorem 1.8. For each $n \in \mathbb{N}$,

$$\sup_{t\in\mathbb{R}} \left| \Pr\left\{ S_n < t\sqrt{\frac{\sigma^2}{n}} \right\} - \Phi(t) \right| \le \frac{6T}{\sigma^3\sqrt{n}}. \tag{1.61}$$

1.8 Definitions of Symbols

In this section, we summarize the definitions of various symbols used throughout the monograph. We use $[0,1]^d$ to denote the unit cube of dimension d and use $\mathbf{S} \in [0,1]^d$ to denote a d-dimensional target. When there are $k \in \mathbb{N}$ targets, we use $\mathbf{S}^k = (\mathbf{S}_1, \ldots, \mathbf{S}_k) \in ([0,1]^d)^k$ to denote the collection of k targets. For moving targets, we use $\mathbf{v}$ to denote the collection of velocities over different dimensions. For non-adaptive query procedures, we use n to denote the number of questions and for adaptive query procedures, we use l to denote the upper bound on the average number of questions. For both query procedures, we use $\delta \in \mathbb{R}_+$ to denote the desired resolution (absolute error) in each dimension and $\varepsilon \in (0,1)$ to denote the excess-resolution probability where the estimate of any dimension has resolution greater than δ. As a result, for non-adaptive query procedures, the theoretical benchmark is usually denoted as $\delta^*(n,d,\varepsilon)$, which corresponds to the minimal achievable absolute estimation error in any dimension when n questions are used and when an excess-resolution probability ε is tolerated. Analogously, for adaptive query procedures, we use $\delta_{\mathrm{a}}^*(l,d,\varepsilon)$. In all our proposed query procedures, we use M to denote the number of quantization levels in each dimension, use $\mathcal{A} \in (0,1)$ to denote the query set, use X^n to denote the binary vector to construct query sets, use Y^n to denote the noisy response of the oracle and use $\hat{S}_i$ to denote the estimate of the target $\mathbf{S}$ in the i-th dimension.

1.9 Organization for the Rest of the Monograph

Section 2 focuses on non-adaptive querying for a single target. We consider a multiple-dimensional target under the query-dependent channel, present characterizations of the performance of optimal non-adaptive querying, discuss the significance of these results beyond previous results and illustrate these theoretical results via numerical examples. Finally, we provide an accessible proof sketch. A major practical implication is that, when searching for a target in a multidimensional region, it is optimal to search in all dimensions simultaneously while it is only asymptotically optimal to search over each dimension separately. Section 2 draws on and synthesizes [54], [109], [111].

Section 3 focuses on adaptive querying for a single target. We formulate the problem, introduce two adaptive query procedures, present the achievability performance of both adaptive query procedures, demonstrate the benefit of using adaptive query procedures over their non-adaptive counterparts, and compare the performance of both for different settings, numerically and analytically. We show that in general, neither of the two adaptive query procedures is superior to the other one for all cases. Furthermore, we demonstrate that there is a cost associated with the superior performance of adaptive querying: high complexity in making decisions and the need for the perfect knowledge of the query-dependent channel transition probability matrix. Thus, although adaptive querying yields better performance, non-adaptive querying is often more practical due to the above two properties since it can be applied in a parallel and channel-agnostic manner. Section 3 draws on and synthesizes [17], [18], [110].

Section 4 focuses on non-adaptive querying for a moving target. Motivated by practical applications where the target moves, e.g., searching for a missing airplane or searching for a face in a video, we propose a piecewise linear constant velocity, characterize the performance of an optimal query procedure, and illustrate the results via numerical examples. A major practical implication is that although it is intuitively true that the performance of searching for a moving target is equivalent to searching for a stationary target in a two dimensional region, the searching complexity in the former scenario is much higher. Section 4 draws on and synthesizes [54], [112], [114].

Section 5 focuses on non-adaptive querying for multiple targets. This setting is motivated by practical applications, e.g, searching for multiple faces in an image or searching for multiple missing targets with a sensor network. We formulate the multi-target search problem, present the results with discussions and illustrate the results via numerical examples. A practical implication is that searching for multiple targets is equivalent to transmitting an unknown number of messages over a random access channel [104]. Section 5 draws on and synthesizes [108], [113]

Finally, Section 6 discusses future research directions that might be worthwhile to pursue.

References

[1] R. Ahlswede, F. Cicalese, and C. Deppe, "Searching with lies under error cost constraints," *Discret. Appl. Math.*, vol. 156, no. 9, 2008, pp. 1444–1460.

[2] R. Ahlswede, F. Cicalese, C. Deppe, and U. Vaccaro, "Two batch search with lie cost," *IEEE Trans. Inf. Theory*, vol. 55, no. 4, 2009, pp. 1433–1439.

[3] M. Aigner, "Searching with lies," *J. Combin. Theory Ser.A*, vol. 74, 1996, pp. 43–56.

[4] M. Aldridge, "Individual testing is optimal for nonadaptive group testing in the linear regime," *IEEE Trans. Inf. Theory*, vol. 65, no. 4, 2018, pp. 2058–2061.

[5] M. Aldridge, O. Johnson, and J. Scarlett, "Group testing: An information theory perspective," *Foundations and Trends® in Communications and Information Theory*, vol. 15, no. 3–4, 2019, pp. 196–392.

[6] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, 2009, pp. 3051–3073.

[7] G. K. Atia and V. Saligrama, "Boolean compressed sensing and noisy group testing," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, 2012, pp. 1880–1901.

[8] L. Bai, J. Diao, and L. Zhou, "Achievable error exponents for almost fixed-length binary classification," in *IEEE ISIT*, pp. 1336–1341, 2022.
[9] J. L. Bentley and A. C.-C. Yao, "An almost optimal algorithm for unbounded searching," *Inf. Process Lett.*, vol. 5, no. SLAC-PUB-1679, 1976.
[10] T. Berger, *Rate-Distortion Theory*. Wiley Online Library, 1971.
[11] E. R. Berlekamp, "Block coding for the binary symmetric channel with noiseless, delayless feedback," *Error-correcting codes*, E. H. B. Mann, Ed., 1968, pp. 61–68.
[12] A. C. Berry, "The accuracy of the Gaussian approximation to the sum of independent variates," *Trans. Am. Math. Soc.*, vol. 49, no. 1, 1941, pp. 122–136.
[13] M. V. Burnashev, "Data transmission over a discrete channel with feedback: Random transmission time," *Problemy Peredachi Informatsii*, vol. 12, no. 4, 1976, pp. 10–30.
[14] M. V. Burnashev and H. Yamamoto, "On using feedback in a gaussian channel," *Problemy Peredachi Informatsii*, vol. 50, no. 3, 2014, pp. 19–34.
[15] M. V. Burnashev and K. Zigangirov, "An interval estimation problem for controlled observation," *Problemy Peredachi Informatsii*, vol. 10, no. 3, 1974, pp. 51–64.
[16] S.-E. Chiu, "Noisy binary search: Practical algorithms and applications," Ph.D. dissertation, UC San Diego, 2019.
[17] S.-E. Chiu and T. Javidi, "Sequential measurement-dependent noisy search," in *IEEE ITW*, pp. 221–225, 2016.
[18] S.-E. Chiu and T. Javidi, "Low complexity sequential search with size-dependent measurement noise," *IEEE Trans. Inf. Theory*, vol. 67, no. 9, 2021, pp. 5731–5748.
[19] S.-E. Chiu, N. Ronquillo, and T. Javidi, "Active learning and CSI acquisition for mmwave initial alignment," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 11, 2019, pp. 2474–2489.
[20] H. W. Chung, B. M. Sadler, and A. O. Hero, "Bounds on variance for unimodal distributions," *IEEE Trans. Inf. Theory*, vol. 63, no. 11, 2017, pp. 6936–6949.

[21] H. W. Chung, B. M. Sadler, L. Zheng, and A. O. Hero, "Unequal error protection querying policies for the noisy 20 questions problem," *IEEE Trans. Inf. Theory*, vol. 64, no. 2, 2018, pp. 1105–1131.

[22] F. Cicalese, "Q-ary searching with lies," in *Proc. 6th Italian Conf. on Theoretical Computer Science*, pp. 228–240, 1998.

[23] F. Cicalese, *Fault-Tolerant Search Algorithms: Reliable Computation with Unreliable Information*. Springer Berlin, Heidelberg, 2013.

[24] F. Cicalese and D. Mundici, "Perfect two-fault tolerant search with minimum adaptiveness," *Adv. Appl. Math*, vol. 25, 2000, pp. 65–101.

[25] F. Cicalese, D. Mundici, and U. Vaccaro, "Least adaptive optimal search with unreliable tests," *Theor. Comput. Sci.*, vol. 270, 2002, pp. 877–893.

[26] F. Cicalese and U. Vaccaro, "Optimal strategies against a liar," *Theor. Comput. Sci.*, vol. 230, 1999, pp. 167–193.

[27] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[28] I. Csiszar, "The method of types [information theory]," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, 1998, pp. 2505–2523.

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

[30] Y. Dagan, Y. Filmus, D. Kane, and S. Moran, "The entropy of lies: Playing twenty questions with a liar," *arXiv: 1811.02177*, 2018.

[31] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, vol. 38. Springer, 2009.

[32] C. Deppe, "Coding with feedback and searching with lies," in *Entropy, Search, Complexity*, vol. 17, János Bolyai Mathematical Society and Springer, 2007, pp. 27–70.

[33] C. Deppe, V. Lebedev, G. Maringer, and N. Polyanskii, "Coding with noiseless feedback over the Z-channel," *IEEE Trans. Inf. Theory*, vol. 68, no. 6, 2022, pp. 3731–3739.

[34] J. Diao, L. Zhou, and L. Bai, "Achievable error exponents for almost fixed-length M-ary classification," in *IEEE ISIT*, 2023.

[35] J. Diao, L. Zhou, and L. Bai, "Achievable error exponents for almost fixed-length M-ary hypothesis testing," in *IEEE ICASSP*, 2023.

[36] R. Dorfman, "The detection of defective members of large populations," *Ann. Stat.*, vol. 14, no. 4, 1943, pp. 436–440.

[37] I. Dumitriu and J. Spencer, "A halfliar's game," *Theor. Comput. Sci.*, vol. 313, no. 3, 2004, pp. 353–369.

[38] I. Dumitriu and J. Spencer, "The liar game over an arbitrary channel," *Combinatorica*, vol. 25, no. 5, 2005, pp. 537–560.

[39] I. Dumitriu and J. Spencer, "The two-batch liar game over an arbitrary channel," *SIAM J. Discrete Math.*, vol. 19, no. 4, 2005, pp. 1056–1064.

[40] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[41] C.-G. Esseen, *On the Liapounoff limit of error in the theory of probability*. Almqvist & Wiksell, 1942.

[42] R. Gallager, "The random coding bound is tight for the average code (corresp.)," *IEEE Trans. Inf. Theory*, vol. 19, no. 2, 1973, pp. 244–246.

[43] R. G. Gallager, *Low-density parity check codes*. MIT Press, Cambridge, MA, 1963.

[44] O. Gebhard, M. Hahn-Klimroth, O. Parczyk, M. Penschuck, M. Rolvien, J. Scarlett, and N. Tan, "Near-optimal sparsity-constrained group testing: Improved bounds and algorithms," *IEEE Trans. Inf. Theory*, vol. 68, no. 5, 2022, pp. 3253–3280.

[45] D. Geman and B. Jedynak, "Active testing model for tracking roads in satellite images," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 18, no. 1, 1996, pp. 1–14.

[46] W. Gu and M. Effros, "A strong converse for a collection of network source coding problems," in *IEEE ISIT*, pp. 2316–2320, 2009.

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

[48] T. S. Han, "An information-spectrum approach to capacity theorems for the general multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, 2006, pp. 2773–2795.

[49] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, 2009, pp. 4947–4966.

[50] M. Horstein, "Sequential transmission using noiseless feedback," *IEEE Trans. Inf. Theory*, vol. 9, no. 3, 1963, pp. 136–143.

[51] B. Jedynak, P. I. Frazier, and R. Sznitman, "Twenty questions with noise: Bayes optimal policies for entropy loss," *J. Appl. Probab.*, vol. 49, no. 1, 2012, pp. 114–136.

[52] O. Johnson, M. Aldridge, and J. Scarlett, "Performance of group testing algorithms with near-constant tests-per-item," *IEEE Trans. Inf. Theory*, vol. 65, no. 2, 2019, pp. 707–723.

[53] Y. Kaspi, O. Shayevitz, and T. Javidi, "Searching for multiple targets with measurement dependent noise," in *IEEE ISIT*, IEEE, pp. 969–973, 2015.

[54] Y. Kaspi, O. Shayevitz, and T. Javidi, "Searching with measurement dependent noise," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, 2018, pp. 2690–2705.

[55] V. Kostina and S. Verdú, "Lossy joint source-channel coding in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, 2013, pp. 2545–2575.

[56] V. Kostina, "Lossy data compression: Non-asymptotic fundamental limits," Ph.D. dissertation, Department of Electrical Engineering, Princeton University, 2013.

[57] A. Krishnamurthy, T. Lykouris, C. Podimata, and R. Schapire, "Contextual search in the presence of adversarial corruptions," *Oper. Res.*, vol. 71, no. 4, 2023, pp. 1120–1135.

[58] A. Lalitha and T. Javidi, "Reliability of sequential hypothesis testing can be achieved by an almost-fixed-length test," in *IEEE ISIT*, pp. 1710–1714, 2016.

[59] A. Lalitha, N. Ronquillo, and T. Javidi, "Improved target acquisition rates with feedback codes," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 5, 2018, pp. 871–885.

[60] A. Lebedev, V. Lebedev, and N. Polyanskii, "Two-stage coding over the Z-channel," *IEEE Trans. Inf. Theory*, vol. 68, no. 4, 2022, pp. 2290–2299.

[61] V. S. Lebedev, "Coding with noiseless feedback," *Probl. Inf. Transm.*, vol. 52, 2016, pp. 103–113.

[62] C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "A new wireless communication paradigm through software-controlled metasurfaces," *IEEE Commun. Mag.*, vol. 56, no. 9, 2018, pp. 162–169.

[63] A. Malinowski, "K-ary searching with a lie," *Ars Combin.*, vol. 37, 1994, pp. 301–308.

[64] E. MolavianJazi and J. N. Laneman, "A second-order achievable rate region for Gaussian multi-access channels via a central limit theorem for functions," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, 2015, pp. 6719–6733.

[65] S. Muthukrishnan, "On optimal strategies for searching in the presence of errors," in *Proc. 5th Annu. ACM-SIAM Symp. on Discrete Algorithms*, pp. 680–688, 1994.

[66] M. Naghshvar, T. Javidi, and M. Wigger, "Extrinsic Jensen–Shannon divergence: Applications to variable-length coding," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, 2015, pp. 2148–2164.

[67] L. Ozarow, "The capacity of the white gaussian multiple access channel with feedback," *IEEE Trans. Inf. Theory*, vol. 30, no. 4, 1984, pp. 623–629.

[68] A. Paz and L. Peterfreund, "Playing guess who with your kids: Code-word strategy against adversaries," *Theor. Comput. Sci.*, vol. 1016, 2024, p. 114 766.

[69] A. Pelc, "Searching with known error probability," *Theor. Comput. Sci.*, vol. 63, 1989, pp. 185–202.

[70] A. Pelc, "Searching games with errors—fifty years of coping with liars," *Theor. Comput. Sci.*, vol. 270, no. 1-2, 2002, pp. 71–109.

[71] Y. Polyanskiy and S. Verdú, "Channel dispersion and moderate deviations limits for memoryless channels," in *Proc. 48th Annu. Allerton Conf.*, pp. 1334–1339, 2010.

[72] Y. Polyanskiy, "Channel coding: Non-asymptotic fundamental limits," Ph.D. dissertation, Department of Electrical Engineering, Princeton University, 2010.

[73] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, 2010, pp. 2307–2359.

[74] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Feedback in the non-asymptotic regime," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, 2011, pp. 4903–4925.

[75] E. Price and J. Scarlett, "A fast binary splitting approach to non-adaptive group testing," in *RANDOM*, 2020.

[76] P. Rajan, W. Han, R. Sznitman, P. Frazier, and B. Jedynak, "Bayesian multiple target localization," *J. Mach. Learn. Res.*, vol. 37, 2015, pp. 1945–1953.

[77] A. Rényi, "On a problem of information theory," *MTA Mat. Kut. Int. Közl.*, vol. 6B, 1961, pp. 505–516.

[78] A. Rényi, "Lectures on the theory of search," North Carolina State University. Dept. of Statistics, Tech. Rep., 1969.

[79] A. Rényi, *A diary on information theory*. Wiley, New York, 1987, translation of Napló az információelmèletröl (1976).

[80] N. Ronquillo, C.-S. Gau, and T. Javidi, "Integrated beam tracking and communication for (Sub-)mmwave links with stochastic mobility," *IEEE J. Sel. Areas Inf. Theory*, no. 4, 2023, pp. 94–111.

[81] J. Scarlett, V. Y. F. Tan, and G. Durisi, "The dispersion of nearest-neighbor decoding for additive Non-Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 63, no. 1, 2017, pp. 81–92.

[82] J. Scarlett, "Noisy adaptive group testing: Bounds and algorithms," *IEEE Trans. Inf. Theory*, vol. 65, no. 6, 2018, pp. 3646–3661.

[83] J. Scarlett and V. Cevher, "Phase transitions in group testing," in *Proceedings of the Twenty-seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, ser. SODA '16, pp. 40–53, Arlington, Virginia: Society for Industrial and Applied Mathematics, 2016.

[84] J. Scarlett and V. Cevher, "Phase transitions in the pooled data problem," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., Curran Associates, Inc., 2017, pp. 377–385.

[85] J. Schalkwijk and T. Kailath, "A coding scheme for additive notse channels with feedback, part i: No bandwidth constraint," *IEEE Trans. Inf. Theory*, vol. 12, no. 2, 1966, pp. 172–182.

[86] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 1, 1948, pp. 379–423.

[87] O. Shayevitz and M. Feder, "Optimal feedback communication via posterior matching," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, 2011, pp. 1186–1222.

[88] V. Strassen, "Asymptotische abschätzungen in shannons informationstheorie," in *Trans. Third Prague Conf. Information Theory*, pp. 689–723, 1962.

[89] C. Sun and L. Zhou, "Achievable resolution limits of noisy adaptive 20 questions estimation for multiple targets," in *IEEE ISIT*, pp. 1580–1585, 2023.

[90] R. Sznitman and B. Jedynak, "Active testing for face detection and localization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 10, 2010, pp. 1914–1920.

[91] N. Tan, W. Tan, and J. Scarlett, "Performance of group testing algorithms with near-constant tests-per-item," *IEEE Trans. Inf. Theory*, vol. 69, no. 2, 2023, pp. 1224–1243.

[92] 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, 2014, pp. 1–184.

[93] M. Tomamichel and V. Y. F. Tan, "Second-order coding rates for channels with state," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, 2014, pp. 4427–4448.

[94] T. Tsiligkaridis, B. M. Sadler, and A. O. Hero, "Collaborative 20 questions for target localization," *IEEE Trans. Inf. Theory*, vol. 60, no. 4, 2014, pp. 2233–2252.

[95] T. Tsiligkaridis, B. M. Sadler, and A. O. Hero, "On decentralized estimation with active queries," *IEEE Trans. Signal Process.*, vol. 63, no. 10, 2015, pp. 2610–2622.

[96] J. N. Tsitsiklis, K. Xu, and Z. Xu, "Private sequential learning," in *COLT*, 2018.

[97] J. N. Tsitsiklis, K. Xu, and Z. Xu, "Private sequential learning," *Oper. Res.*, vol. 69, no. 5, 2021, pp. 1575–1590.

[98] S. M. Ulam, *Adventures of a Mathematician*. New York: Scribner's, 1976.

[99] E. Variani, K. Lahouel, A. Bar-Hen, and B. Jedynak, "Non-adaptive policies for 20 questions target localization," in *IEEE ISIT*, pp. 775–778, 2015.

[100] D. Williams, *Probability with martingales*. Cambridge university press, 1991.

[101] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, 2021, pp. 3313–3351.

[102] J. Xu, K. Xu, and D. Yang, "Optimal query complexity for private sequential learning against eavesdropping," in *AISTATS*, 2019.

[103] H. Yamamoto and K. Itoh, "Asymptotic performance of a modified schalkwijk-barron scheme for channels with noiseless feedback (corresp.)," *IEEE Trans. Inf. Theory*, vol. 25, no. 6, 1979, pp. 729–733.

[104] R. C. Yavas, V. Kostina, and M. Effros, "Random access channel coding in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 67, no. 4, 2021, pp. 2115–2140.

[105] R. C. Yavas, V. Kostina, and M. Effros, "Variable-length sparse feedback codes for point-to-point, multiple access, and random access channels," *arXiv 2103.09373, to appear in IEEE Trans. Inf. Theory*, 2023.

[106] Z. Y. Zhao, Z. R. Wang, S. W. Zhang, and L. Liang, "Finding defective elements in intelligent reflecting surface via over-the-air measurements," in *IEEE Globecom*, 2024.

[107] L. Zhou and A. O. Hero, "Achievable resolution limits for the noisy adaptive 20 questions problem," in *IEEE ISIT*, pp. 784–789, 2021.

[108] L. Zhou, L. Bai, and A. O. Hero, "Resolution limits of non-adaptive 20 questions search for multiple targets," *IEEE Trans. Inf. Theory*, vol. 68, no. 8, 2022, pp. 4964–4982.

[109] L. Zhou and A. O. Hero, "Resolution limits of non-adaptive querying for noisy 20 questions estimation," in *IEEE ISIT*, pp. 2167–2172, 2020.

[110] L. Zhou and A. O. Hero, "Achievable resolution limits for the noisy adaptive 20 questions problem," in *IEEE ISIT*, pp. 784–789, 2021.

[111] L. Zhou and A. O. Hero, "Resolution limits for the noisy non-adaptive 20 questions problem," *IEEE Trans. Inf. Theory*, vol. 67, no. 4, 2021, pp. 2055–2073.

[112] L. Zhou and A. O. Hero, "Resolution limits of 20 questions search strategies for moving targets," in *IEEE ICASSP*, pp. 5210–5214, 2021.

[113] L. Zhou and A. O. Hero, "Resolution limits of non-adaptive 20 questions estimation for multiple targets," in *IEEE ISIT*, pp. 790–795, 2021.

[114] L. Zhou and A. O. Hero, "Resolution limits of non-adaptive 20 questions search strategies for a moving target," *IEEE Trans. Inf. Theory*, vol. 69, no. 9, 2023, pp. 5463–5484.

[115] L. Zhou and M. Motani, "Finite blocklength lossy source coding for discrete memoryless sources," *Foundations and Trends® in Communications and Information Theory*, vol. 20, no. 3, 2023, pp. 157–389.

[116] L. Zhou, V. Y. F. Tan, and M. Motani, "Exponential strong converse for content identification with lossy recovery," *IEEE Trans. Inf. Theory*, vol. 64, no. 8, 2018, pp. 5879–5897.