5-31-2024

Information theoretic bounds for capacity and bayesian risk

Ian Zieder
New Jersey Institute of Technology, ihz3@njit.edu

ABSTRACT

INFORMATION THEORETIC BOUNDS FOR CAPACITY AND BAYESIAN RISK

by
Ian Zieder

In this dissertation, the problem of finding lower error bounds on the minimum mean-squared error (MMSE) and the maximum capacity achieving distribution for a specific channel is addressed. Presented are two parts, a new lower bound on the MMSE and upper and lower bounds on the capacity achieving distribution for a Binomial noise channel. The new lower bound on the MMSE is achieved via use of the Poincaré inequality. It is compared to the performance of the well known Ziv-Zakai error bound. The second part considers a binomial noise channel and is concerned with the properties of the capacity-achieving distribution. In particular, for the binomial channel, it is not known if the capacity-achieving distribution is unique since the output space is finite (i.e., supported on integers 0, ... , n) and the input space is infinite (i.e., supported on the interval [0, 1]), and there are multiple distributions that induce the same output distribution. This paper shows that the capacity-achieving distribution is unique by appealing to the total positivity property of the binomial kernel. In addition, we provide upper and lower bounds on the cardinality of the support of the capacity-achieving distribution. Specifically, an upper bound of order $\frac{n}{2}$ is shown, which improves on the previous upper bound of order n due to Witsenhausen. Moreover, a lower bound of order $\sqrt{n}$ is shown. Finally, additional information about the locations and probability values of the support points is established.

INFORMATION THEORETIC BOUNDS FOR CAPACITY AND BAYESIAN RISK

by
Ian Zieder

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
in Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Electrical Engineering

Helen and John C. Hartmann
Department of Electrical and Computer Engineering

May 2024

APPROVAL PAGE

INFORMATION THEORETIC BOUNDS FOR CAPACITY AND BAYESIAN RISK

Ian Zieder

Dr. Alexander M. Haimovich, Dissertation Advisor — Date
Distinguished Professor of Electrical and Computer Engineering, NJIT

Dr. Ali Abdi, Committee Member — Date
Professor of Electrical and Computer Engineering, NJIT

Dr. Joerg Kliewer, Committee Member — Date
Professor of Electrical and Computer Engineering, NJIT

Dr. Martina Cardone, Committee Member — Date
Assistant Professor of Electrical and Computer Engineering,
University of Minnesota - Twin Cities, Minneapolis, MN

Dr. Alex Dytso, Committee Member — Date
Engineer, Qualcomm Flarion Technologies, Inc., Bridgewater, NJ

BIOGRAPHICAL SKETCH

Author: Ian Zieder

Degree: Doctor of Philosophy

Date: May 2024

Undergraduate and Graduate Education:

- Doctor of Philosophy in Electrical Engineering,
  New Jersey Institute of Technology, Newark, NJ, 2024
- Master of Science in Electrical Engineering,
  New Jersey Institute of Technology, Newark, NJ, 2018
- Bachelor of Science in Electrical & Computer Engineering,
  Rutgers University, New Brunswick, NJ, 2015

Major: Electrical Engineering

Presentations and Publications:

Ian Zieder, Antonino Favano, Luca Barletta, Alex Dytso, "Binomial Channel: On the Capacity-Achieving Distribution and Bounds on the Capacity," *https://doi.org/10.48550/arXiv.2401.12818* , January 2024

Alex Dytso, Martina Cardone, Ian Zieder, "Meta Derivative Identity for the Conditional Expectation," *IEEE Transactions on Information Theory*, vol. 196, issue: 7, pp 4284 - 4302 , July 2023

Alex Dytso, Martina Cardone, Ian Zieder, "High-Noise Asymptotics of the Ziv-Zakai Bound," *IEEE Signal Processing Letters*, vol. 29, pp 1933 - 1937 , September 2022

Ian Zieder, Alex Dytso, Martina Cardone, "An MMSE Lower Bound via Poincaré Inequality," *IEEE International Symposium on Information Theory (ISIT)*, vol. 29, pp 1933 - 1937 , June 2022

Dedicated to my entire family. With a special dedication to my Mom and Dad, who gave me a copy of this poem many years ago:

<u>Don't Quit</u>

When things go wrong, as they sometimes will,
when the road you're trudging seems all uphill,
when the funds are low and the debts are high,
and you want to smile but you have to sigh,
when care is pressing you down a bit,
rest if you must, but don't you quit.

Life is queer with its twists and turns.
As everyone of us sometimes learns,
And many a fellow turns about,
when he might have won had he stuck it out.
Don't give up though the pace seems slow,
you may succeed with another blow.

Often the goal is nearer than it seems to a faint and faltering man;

Often the struggler has given up,
when he might have captured the victor's cup;
and he learned too late when the night came down,
how close he was to the golden crown.
Success is failure turned inside out,
the silver tint of the clouds of doubt,
and when you never can tell how close you are,
it may be near when it seems afar;
so stick to the fight when you're hardest hit,
it's when things seem worst, you must not quit.

Edgar Albert Guest

ACKNOWLEDGMENTS

First, I would like to express my deep sense of gratitude to my advisor, Dr. Alexander Haimovich, for providing me with the opportunity to join his lab, funding my research, and guiding me throughout my Ph.D. studies. I have been fortunate to be supervised by him, two separate times. He has been supportive since the days when I was a Master's student sitting in the front of his class.

I would also like to sincerely thank Dr. Alex Dytso who has remained a huge source of support throughout his time as a professor and my time as a Ph.D. candidate. His energy and passion towards not just the subjects, but also towards accomplishing goals is something that I can not completely describe on paper. During some very difficult times in my life, both Dr. Haimovich and Dr. Dytso have given me the support and encouragement I needed to continue. They both embody the poem, *Don't Quit* by Edgar Albert Guest, written on the previous page.

I would like to thank the rest of my dissertation committtee memebers: Dr. Ali Abdi, Dr. Joerg Kliewer, and Dr. Martina Cardone. Each one has contributed immensely, in their own way, towards helping me build my confidence from the ground up and towards pushing me to succeed.

I would like to thank Dr. Durga Misra, Dr. Cong Wang, Dr. Sotirios Ziavras, Ms. Teri Bass, and Ms. Clarisa Gonzalez-Lenahan; for all of their funding, help, and support throughout my Master's and Ph.D. programs.

A special shout out to my friends and colleagues in the CWiP lab. A special thanks to Anne Ghosh, Kyle Wensell, Zareen Khan, Sara Ghasvarianjahromi, Yusuf Ozkan, Ms. Kathy Bosco, and Mr. Armando Blandon. You guys inspired me so much to keep pushing.

A final shoutout to my family-friend, Rob Geiger. He's always rooting for me, and I just want to tell him that I'm also always rooting for him too.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Minimum-Mean Squared Error Estimator and the Exponential Family

The minimum mean squared error (MMSE) is an essential and ubiquitous fidelity criterion in statistical signal processing. However, the MMSE is often difficult to compute in closed-form, and we often need to rely on bounds. In terms of bounds, the attention typically falls on lower bounds as deriving a tight lower bound can often be a difficult task.

In this work, a novel lower bound is derived on the MMSE of estimating $\mathbf{X} \in \mathbb{R}^d$ from the noisy observation $\mathbf{Y} \in \mathbb{R}^k$. Towards this end, an alternative representation of the MMSE $\mathrm{mmse}(\mathbf{X}|\mathbf{Y})$ is presented and studied. This new representation provides a new line of attack for direct computation of the MMSE and, together with the Poincaré inequality, a new lower bound on the MMSE is derived.

The focus is on the exponential family. The class of probability models $\mathcal{P} = \left\{P_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}, \mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^d\right\}$ supported on $\mathbf{y} \in \mathcal{Y} \subseteq \mathbb{R}^k$ is a continuous exponential family if the probability density function (pdf) of it can be written as

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = h(\mathbf{y})\mathrm{e}^{\langle \mathbf{x}, \mathbf{T}(\mathbf{y})\rangle - \phi(\mathbf{x})}, \tag{1.1}$$

where $\mathbf{T} : \mathcal{Y} \to \mathbb{R}^d$ is the sufficient statistic function; $\phi : \mathcal{X} \to \mathbb{R}$ is the log-partition function; $h : \mathcal{Y} \to [0, \infty)$ is the base measure; and $\langle \cdot, \cdot \rangle$ is the inner product.

Loosely speaking, there are three approaches for finding lower bounds on the MMSE, which result in three different families of lower bounds.

The *first* family is known as Weiss–Weinstein family [1], and it includes important bounds such as the Bayesian Cramér-Rao bound [2] (also known as the Van Trees bound), the Bobrovsky–Zakai bound [3], the Barankin bound [4], and the Bobrovsky-Mayer-Wolf-Zakai bound [5]. The Weiss–Weinstein family relies on the Cauchy-Schwarz inequality, which establishes the following variational representation of the MMSE,

$$\mathrm{mmse}(\mathbf{X}|\mathbf{Y}) = \sup_{\psi \in \mathcal{C}} \frac{\|\mathbb{E}[\psi(\mathbf{X},\mathbf{Y})\mathbf{X}^{\mathsf{T}}]\|}{\mathbb{E}[\|\psi(\mathbf{X},\mathbf{Y})\|^2]},$$

where

$$\begin{aligned} \mathcal{C} = \{\psi : & \mathcal{X} \times \mathcal{Y} \to \mathcal{X} : \mathbb{E}[\psi(\mathbf{X},\mathbf{Y})|\mathbf{Y}=\mathbf{y}] = 0, \mathbf{y} \in \mathcal{Y}, \\ & \mathbb{E}[\|\psi(\mathbf{X},\mathbf{Y})\|^2] < \infty\}. \end{aligned}$$

The aforementioned lower bounds are then attained by a clever choice of the function ψ that results in a computationally feasible bound. One of the drawbacks of this family of bounds is that choosing the right ψ can be challenging. In particular, to the best of our knowledge, all of the existing bounds require that the random vector $\mathbf{X}$ has a pdf; as such, these bounds do not, for example, hold for discrete or mixed random vectors.

The *second* family of lower bounds is known as Ziv-Zakai and it was originally proposed in [6] and later improved in [7–9]. Ziv-Zakai bounds rely on connecting estimation to binary hypothesis testing. While this family of lower bounds is typically very tight, it suffers from several drawbacks. First, it can be difficult to compute in closed-form. Second, while there are vector generalizations of this bound [10], typically, these generalizations contain another layer of optimization, which can make the computation difficult. Third, these bounds assume that the input has a density and cannot be used to study the MMSE of discrete or mixed random variables.

The *third* family of lower bounds uses a variational approach and it works by minimizing the MSE subject to a constraint on a suitably chosen divergence measure, for example, the Kullback–Leibler (KL) divergence [11]. Similar to the previous bounds, also this family only holds if the input has a density and hence, it is not suitable for studying the MMSE of discrete or mixed random variables.

The key ingredient in the proof of our new lower bound on the MMSE is the Poincaré inequality [12]. The Poincaré inequality has found a number of applications in information theory and signal processing and the interested reader is referred to [13–20] and references therein. Recently, the authors of [21] developed bounds on the MMSE using the log-Sobolev inequality, which has deep connections with the Poincaré inequality.

1.2 Lower Error Bounds

Loosely speaking, there are three approaches for finding lower bounds on the MMSE, which result in three different families of lower bounds.

The *first* family is known as Weiss–Weinstein family [1], and it includes important bounds such as the Bayesian Cramér-Rao bound [2] (also known as the Van Trees bound), the Bobrovsky–Zakai bound [3], the Barankin bound [4], and the Bobrovsky-Mayer-Wolf-Zakai bound [5]. The Weiss–Weinstein family relies on the Cauchy-Schwarz inequality, which establishes the following variational representation of the MMSE,

$$\mathrm{mmse}(\mathbf{X}|\mathbf{Y}) = \sup_{\psi \in \mathcal{C}} \frac{\|\mathbb{E}[\psi(\mathbf{X},\mathbf{Y})\mathbf{X}^{\mathsf{T}}]\|}{\mathbb{E}[\|\psi(\mathbf{X},\mathbf{Y})\|^2]},$$

where

$$\begin{aligned}\mathcal{C} = \{\psi :& \mathcal{X} \times \mathcal{Y} \to \mathcal{X} : \mathbb{E}[\psi(\mathbf{X},\mathbf{Y})|\mathbf{Y}=\mathbf{y}] = 0, \mathbf{y} \in \mathcal{Y},\\ &\mathbb{E}[\|\psi(\mathbf{X},\mathbf{Y})\|^2] < \infty\}.\end{aligned}$$

The aforementioned lower bounds are then attained by a clever choice of the function ψ that results in a computationally feasible bound. One of the drawbacks of this family of bounds is that choosing the right ψ can be challenging. In particular, to the best of our knowledge, all of the existing bounds require that the random vector $\mathbf{X}$ has a pdf; as such, these bounds do not, for example, hold for discrete or mixed random vectors.

The *second* family of lower bounds is known as Ziv-Zakai and it was originally proposed in [6] and later improved in [7–9]. Ziv-Zakai bounds rely on connecting estimation to binary hypothesis testing. While this family of lower bounds is typically very tight, it suffers from several drawbacks. First, it can be difficult to compute in closed-form. Second, while there are vector generalizations of this bound [10], typically, these generalizations contain another layer of optimization, which can make the computation difficult. Third, these bounds assume that the input has a density and cannot be used to study the MMSE of discrete or mixed random variables.

The *third* family of lower bounds uses a variational approach and it works by minimizing the MSE subject to a constraint on a suitably chosen divergence measure, for example, the Kullback–Leibler (KL) divergence [11]. Similar to the previous bounds, also this family only holds if the input has a density and hence, it is not suitable for studying the MMSE of discrete or mixed random variables.

The key ingredient in the proof of our new lower bound on the MMSE is the Poincaré inequality [12]. The Poincaré inequality has found a number of applications in information theory and signal processing and the interested reader is referred to [13–20] and references therein. Recently, the authors of [21] developed bounds on the MMSE using the log-Sobolev inequality, which has deep connections with the Poincaré inequality.

1.3 Binomial Channel

A binomial channel is a type of communication channel characterized by discrete inputs and outputs, where the noise or interference is modeled as a binomial distribution. The capacity of a communication channel refers to the maximum rate at which information can be reliably transmitted over the channel.

The capacity-achieving distribution for a binomial channel refers to the probability distribution of input symbols that achieves the maximum possible rate of information transmission (i.e., the capacity) for that channel. This distribution depends on the specific characteristics of the channel, such as the probability of error or noise.

Bounds on the capacity of a binomial channel provide limits on the maximum achievable transmission rate given certain constraints or assumptions. These bounds can be derived using various techniques, such as information theory and channel coding theory.

For example, the Shannon capacity formula provides an upper bound on the capacity of a channel based on its bandwidth and signal-to-noise ratio.

In practice, finding the exact capacity-achieving distribution for a binomial channel and determining tight bounds on its capacity can be challenging and may require advanced mathematical techniques. These properties have been studied to understand the fundamental limits of communication systems and to design efficient coding and modulation schemes that approach these limits.

1.4 Organization and Contributions

This dissertation is organized according to and following the table of contents. Presented are the two parts of my research, lower error bounds and binomial channel capacity. Problem Formulation (PART I) contains information on the new lower bound on the MMSE via Poincaré Inequality. Discussed is the background and motivation behind finding a new lower error bound on the MMSE, the new representation of the MMSE, the tightness in the high-noise regime, and how it compares to other lower error bounds such as the Ziv-Zakai and Cramer-Rao bounds.

Problem Formulation (PART II) discusses the background of a binomial channel and the properties of the optimal input and output distributions.

The results (PARTS I and II) contain the contributions of the research into lower error bounds and capacity of the binomial channel.

CHAPTER 2

PROBLEM FORMULATION (PART I)

2.1 A New Lower Bound on the MMSE

2.1.1 Poincaré Inequality

Consider a class of functions $\mathcal{A}$. We say that a probability distribution $P_{\mathbf{U}}$ satisfies a Poincaré inequality with respect to $\mathcal{A}$ with a constant $\kappa \geq 0$ if [12]

$$\mathrm{Var}(f(\mathbf{U})) \leq \frac{1}{\kappa} \mathbb{E}\left[||\nabla f(\mathbf{U})||^2\right], \forall f \in \mathcal{A}. \tag{2.1}$$

If $\kappa = 0$, we treat the right-hand side of eq. (2.1) as infinity.

We are here interested in the conditional version of the Poincaré inequality, i.e., for a class of functions $\mathcal{A}$ we say that a conditional probability $P_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}$ (for a fixed $\mathbf{x} \in \mathcal{X}$) satisfies a Poincaré inequality in eq. (2.1) with respect to $\mathcal{A}$ with a constant $\kappa(\mathbf{x}) \geq 0$ if

$$\mathrm{Var}(f(\mathbf{Y})|\mathbf{X}=\mathbf{x}) \leq \frac{1}{\kappa(\mathbf{x})}\mathbb{E}\left[||\nabla f(\mathbf{Y})||^2|\mathbf{X}=\mathbf{x}\right], \forall f \in \mathcal{A}.$$

Since $\mathbf{x} \in \mathcal{X}$ can be treated as a parameter of the distribution, the conditional and unconditional versions hold under the same conditions. There exist several sufficient conditions on $\mathcal{A}$ and $P_{\mathbf{Y}|\mathbf{X}}$, which guarantee that a Poincaré inequality holds, and which identify the constant $\kappa(\mathbf{x})$. We next list a few of these.

• Convex Poincaré [22]. Let $P_{\mathbf{Y}|\mathbf{X}}$ be a product distribution and $\mathcal{A}$ be a set of functions such that $f \in \mathcal{A}$ is $f : [0,1]^k \to \mathbb{R}$, separately convex and the partial derivatives of which exist. Then, $\kappa_{\mathrm{C}}(\mathbf{x}) = 1$.

• Bakry-Émery condition [23]: Let $\mathcal{A}$ be a class of continuously differentiable functions. Then,

$$\kappa_{\mathrm{BE}}(\mathbf{x}) = \max\left\{\kappa : \nabla^2_{\mathbf{y}} \log\left(\frac{1}{f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})}\right) \succeq \kappa \mathsf{I}_k, \forall \mathbf{y} \in \mathcal{Y}\right\}. \tag{2.2}$$

If the set in eq. (2.2) is empty, we set $\kappa_{\rm BE}(\mathbf{x}) = 0$. We note that the condition in eq. (2.2) simply requires that the distribution is strongly log-concave. As an example, the Bakry-Émery constant for the exponential family is given by the next proposition.

Proposition 1. *Assume that $P_{\mathbf{Y}|\mathbf{X}}$ has a pdf of the form in eq.* (1.1). *Then, for $\mathbf{x} \in \mathcal{X}$, we have*

$$\kappa_{\rm BE}(\mathbf{x}) = \max\left\{0, \tilde{\kappa}_{\rm BE}(\mathbf{x})\right\},$$
$$\tilde{\kappa}_{\rm BE}(\mathbf{x}) = \min_{\mathbf{y}\in\mathcal{Y}} \lambda_{\min}\left(\nabla^2_{\mathbf{y}} \log\left(\frac{1}{h(\mathbf{y})}\right) - \nabla^2_{\mathbf{y}}\langle \mathbf{x}, \mathbf{T}(\mathbf{y})\rangle\right),$$

where $\nabla^2_{\mathbf{y}}$ denotes the Hessian.

We have

$$\begin{aligned}
&\nabla^2_{\mathbf{y}} \log\left(\frac{1}{f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})}\right)\\
&= -\nabla^2_{\mathbf{y}}\left(\log\left(h(\mathbf{y})\right) - \nabla^2_{\mathbf{y}}(\langle \mathbf{x},\, \mathbf{T}(\mathbf{y})\rangle) - \phi(\mathbf{x})\right)\\
&\succeq \lambda_{\min}\left(\nabla^2_{\mathbf{y}} \log\left(\frac{1}{h(\mathbf{y})}\right) - \nabla^2_{\mathbf{y}}\langle \mathbf{x}, \mathbf{T}(\mathbf{y})\rangle\right) \mathsf{I}_k,
\end{aligned}$$

which concludes the proof of Proposition 1.

• Laplace distribution [24]: Let $P_{Y|X=x}$ have a Laplace pdf (i.e., $f_{Y|X}(y|x) = \frac{1}{2}\mathrm{e}^{-|y-x|}$) and $\mathcal{A}$ be a set of all functions $f : \mathbb{R} \to \mathbb{R}$ that are continuously differentiable and $\lim_{x\pm\infty} \mathrm{e}^{-|x|} f(x) = 0$. Then, $\kappa_{\rm Lap}(x) = \frac{1}{4}$.

2.1.2 New lower bound via Poincaré Inequality

We here leverage the result in Theorem 1 to derive a new lower bound on the MMSE for the exponential family (i.e., we assume that $P_{\mathbf{Y}|\mathbf{X}}$ has a pdf of the form in eq. (1.1)). Our new lower bound on the MMSE is given by the next theorem. The key advantage of this new lower bound is that it holds for all distributions on $\mathbf{X}$ (not necessarily continuous).

Theorem 1. *Assume that the following three conditions hold:*

1. *For all $\mathbf{x} \in \mathcal{X}$ the distribution $P_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}$ has a pdf of the form in eq.* (1.1) *and it satisfies a Poincaré inequality with respect to $(\mathcal{A}, \kappa(\mathbf{x}))$;*

2. $\mathbf{y} \mapsto \iota_{P_{\mathbf{XY}}}(\mathbf{x}; \mathbf{y}) \in \mathcal{A}$ *for every $\mathbf{x}$ such that $\kappa(\mathbf{x}) > 0$;*

3. *There exists a $\rho \geq 0$ such that for all $\mathbf{y} \in \mathcal{Y}$, $\sigma_{\min}\left((\mathbf{J}_{\mathbf{y}}\mathbf{T}(\mathbf{y}))^{+}\right) \geq \rho$.*

Then,

$$\text{mmse}(\mathbf{X}|\mathbf{Y}) \geq \rho^2 \mathbb{E}\left[\kappa(\mathbf{X})\text{Var}(\iota_{P_{\mathbf{XY}}}(\mathbf{X};\mathbf{Y})|\mathbf{X})\right].$$

We have

$$\begin{aligned}
\text{mmse}(\mathbf{X}|\mathbf{Y}) &= \mathbb{E}\left[||\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]||^2\right] \\
&\overset{(a)}{=} \mathbb{E}\left[||(\mathbf{J}_{\mathbf{Y}}\mathbf{T}(\mathbf{Y}))^{+}\nabla_{\mathbf{Y}}\iota_{P_{\mathbf{XY}}}(\mathbf{X};\mathbf{Y})]||^2\right] \\
&\overset{(b)}{\geq} \rho^2 \mathbb{E}\left[\|\nabla_{\mathbf{Y}}\iota_{P_{\mathbf{XY}}}(\mathbf{X};\mathbf{Y})]\|^2\right] \\
&= \rho^2 \mathbb{E}\left[\mathbb{E}\left[||\nabla_{\mathbf{Y}}\iota_{P_{\mathbf{XY}}}(\mathbf{X};\mathbf{Y})||^2|\mathbf{X}\right]\right] \\
&\overset{(c)}{\geq} \rho^2 \mathbb{E}\left[\kappa(\mathbf{X})\text{Var}(\iota_{P_{\mathbf{XY}}}(\mathbf{X};\mathbf{Y})|\mathbf{X})\right],
\end{aligned}$$

where the labeled (in)equalities follow from: (a) applying Theorem 1; (b) using condition 3) in Theorem 1 and the inequality $\|\mathbf{A}\mathbf{x}\| \geq \sigma_{\min}(\mathbf{A})\|\mathbf{x}\|$; and (c) using a Poincaré inequality and conditions 1) and 2) in Theorem 1. This concludes the proof of Theorem 1.

Remark 1. *Theorem 1 holds provided that three conditions are satisfied. Conditions 1) and 2) are required for the application of a Poincaré inequality in the proof of the bound. We have listed a number of sufficient conditions for 1) to hold. Condition 2) requires that the information density $\mathbf{y} \mapsto \iota_{P_{\mathbf{XY}}}(\mathbf{x}; \mathbf{y})$ belongs to some regular enough family of functions $\mathcal{A}$. Interestingly, such conditions are not difficult to find. Moreover, often these conditions only depend on $P_{\mathbf{Y}|\mathbf{X}}$ and are independent of $P_{\mathbf{X}}$. For example, the information density for the exponential family in eq.* (1.1) *is known to be infinitely differentiable for all distributions on $P_{\mathbf{X}}$ [25]. Finally, condition 3) imposes a requirement on the sufficient statistics $\mathbf{T}(\mathbf{y})$. This condition, for example, holds when $\mathbf{T}(\mathbf{y})$ is a linear function (e.g., Gaussian, Wishart).*

2.1.3 Tightness in the high-noise regime

We here show an example of $P_{\mathbf{Y}|\mathbf{X}}$ for which our lower bound in Theorem 1 is tight in the high-noise regime. Towards this end, we consider a scenario where

$$\mathbf{Y} = \mathbf{X} + \mathbf{N}, \tag{2.3}$$

where $\mathbf{X}$ and $\mathbf{N}$ are independent and $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_k, \sigma_N^2 \mathsf{I}_k)$. In this case, the lower bound in Theorem 1 reduces to

$$\mathrm{mmse}(\mathbf{X}|\mathbf{Y}) \geq \sigma_N^2 \mathbb{E}\left[\mathrm{Var}(\iota(\mathbf{X};\mathbf{Y})|\mathbf{X})\right]. \tag{2.4}$$

It is noted that (2.4) is a new representation of the MMSE. Conditions 1)-3) are verified as follows. First, we note that for the model in eq. (2.3), we have that $\mathbf{T}(\mathbf{y}) = \mathbf{y}/\sigma_N^2$, which implies $\rho = \sigma_N^2$ in condition 3). To verify conditions 1) and 2), we use the Bakry-Émery condition presented in (2.2), which first requires that $\mathbf{y} \mapsto \iota_{P_{\mathbf{XY}}}(\mathbf{x};\mathbf{y})$ is continuously differentiable for every $\mathbf{x}$, which is a well-known fact, see for example [26]. Second, since $h(\mathbf{y}) = \frac{1}{(2\pi\sigma_N^2)^{k/2}} \mathrm{e}^{-\frac{\|\mathbf{y}\|^2}{2\sigma_N^2}}$, we can find the Bakry-Émery constant by applying Proposition 1, namely $\kappa(\mathbf{x}) = \kappa_{\mathrm{BE}}(\mathbf{x}) = 1/\sigma_N^2$. The quantity $\mathbb{E}\left[\mathrm{Var}(\iota(\mathbf{X};\mathbf{Y})|\mathbf{X})\right]$ has appeared in the past in [27], where it was termed as conditional information variance.

We now argue that the bound in eq. (2.4) is tight in the high-noise regime. To do this, we recall the following high-noise behavior of the MMSE in the Gaussian noise setting [28]

$$\lim_{\sigma_N \to \infty} \mathrm{mmse}(\mathbf{X}|\mathbf{Y}) = \mathrm{Var}\left(\mathbf{X}\right).$$

At this point, it is interesting to point out that for the Gaussian noise setting, the Cramér-Rao bound is given by [2]

$$\mathrm{mmse}(\mathbf{X}|\mathbf{Y}) \geq \frac{k^2 \sigma_N^2}{k + \sigma_N^2 J(\mathbf{X})},$$

where $J(\mathbf{X})$ is the Fisher information of $\mathbf{X}$. Hence, we obtain

$$\lim_{\sigma_N \to \infty} \frac{k^2 \sigma_N^2}{k + \sigma_N^2 J(\mathbf{X})} = \frac{k^2}{J(\mathbf{X})},$$

which is equal to the variance if and only if $\mathbf{X}$ is isotropic Gaussian [26]. Thus, the Cramér-Rao bound is only tight for the class of isotropic Gaussian inputs in the high-noise regime, and otherwise is sub-optimal.

The next result shows that the lower bound in Theorem 1 is tight for a large family of prior distributions on $\mathbf{X}$.

Theorem 2. *Assume that* $\mathbf{X}$ *is sub-Gaussian*[1]*. Then,*

$$\lim_{\sigma_N \to \infty} \sigma_N^2 \mathbb{E}\left[\mathrm{Var}(\iota(\mathbf{X};\mathbf{Y})|\mathbf{X})\right] = \mathrm{Var}\left(\mathbf{X}\right).$$

Proof. To simplify the proof, without loss of generality, we assume that $\mathbb{E}[\mathbf{X}] = \mathbf{0}_k$. In addition, since we are looking at $\sigma_N \to \infty$, we assume that $\sigma_N > 1$ when we derive our inequalities. Now, let $g(\mathbf{y}) = (2\pi\sigma_N^2)^{\frac{k}{2}} f_{\mathbf{Y}}(\mathbf{y})$ and note that

$$\begin{aligned}
&\mathrm{Var}(\iota(\mathbf{X};\mathbf{Y})|\mathbf{X}=\mathbf{x}) \\
&= \mathrm{Var}\left(-\frac{\|\mathbf{Y}-\mathbf{X}\|^2}{2\sigma_N^2} - \log f_{\mathbf{Y}}(\mathbf{Y}) \middle| \mathbf{X}=\mathbf{x}\right) \\
&= \mathrm{Var}\left(\frac{\|\mathbf{Z}\|^2}{2} + \log g(\mathbf{x}+\sigma_N\mathbf{Z})\right).
\end{aligned} \tag{2.5}$$

Next, observe that

$$\begin{aligned}
g(\mathbf{x}+\sigma_N\mathbf{z}) &= (2\pi\sigma_N^2)^{\frac{k}{2}} f_{\mathbf{Y}}(\mathbf{x}+\sigma_N\mathbf{z}) \\
&= (2\pi\sigma_N^2)^{\frac{k}{2}} \mathbb{E}\left[f_{\mathbf{Y}|\mathbf{X}}\left(\mathbf{x}+\sigma_N\mathbf{z}|\mathbf{X}\right)\right] \\
&= \underbrace{\mathbb{E}\left[\mathrm{e}^{-\frac{\|\mathbf{x}-\mathbf{X}\|^2+2(\mathbf{x}-\mathbf{X})^\mathsf{T}\sigma_N\mathbf{z}}{2\sigma_N^2}}\right]}_{\tilde{g}(\mathbf{x}+\sigma_N\mathbf{z})} \mathrm{e}^{-\frac{\|\mathbf{z}\|^2}{2}}.
\end{aligned} \tag{2.6}$$

Therefore, combining eq. (2.5) and eq. (2.6) we arrive at

$$\lim_{\sigma_N \to \infty} \sigma_N^2 \mathbb{E}\left[\mathrm{Var}(\iota(\mathbf{X};\mathbf{Y})|\mathbf{X})\right]$$

[1]A random variable $X \in \mathbb{R}$ is said to be sub-Gaussian with parameter σ_0 if for every $\lambda \in \mathbb{R}$ we have that $\mathbb{E}\left[\mathrm{e}^{\lambda[X-\mathbb{E}[X]]}\right] \leq \mathrm{e}^{\lambda^2\sigma_0/2}$. A random vector $\mathbf{X} \in \mathbb{R}^k$ is said to be sub-Gaussian with parameter σ_0 if $\mathbf{u}^\mathsf{T}\left(\mathbf{X} - \mathbb{E}[\mathbf{X}]\right)$ is sub-Gaussian with parameter σ_0 for any unit vector $\mathbf{u}$.

$$
\begin{aligned}
&= \lim_{\sigma_N \to \infty} \sigma_N^2 \mathbb{E}_{\mathbf{X}} \left[\mathrm{Var}_{\mathbf{Z}} \left(\frac{\|\mathbf{Z}\|^2}{2} + \log g(\mathbf{X} + \sigma_N \mathbf{Z}) \right) \right] \\
&= \lim_{\sigma_N \to \infty} \mathbb{E}_{\mathbf{X}} \left[\mathrm{Var}_{\mathbf{Z}} \left(\frac{\sigma_N \|\mathbf{Z}\|^2}{2} + \sigma_N \log g(\mathbf{X} + \sigma_N \mathbf{Z}) \right) \right] \\
&= \lim_{\sigma_N \to \infty} \mathbb{E}_{\mathbf{X}} \left[\mathrm{Var}_{\mathbf{Z}} \left(\sigma_N \log \left(\tilde{g}(\mathbf{X} + \sigma_N \mathbf{Z}) \right) \right) \right]. \qquad (2.7)
\end{aligned}
$$

We now leverage the following lemma, the proof of which is provided in [29, Appendix A].

Lemma 1. *Assume that* $\mathbf{X}$ *is sub-Gaussian. Then,*

$$
\begin{aligned}
&\lim_{\sigma_N \to \infty} \mathbb{E}_{\mathbf{X}} \left[\mathrm{Var}_{\mathbf{Z}} \left(\sigma_N \log \left(\tilde{g}(\mathbf{X} + \sigma_N \mathbf{Z}) \right) \right) \right] \\
&= \mathbb{E}_{\mathbf{X}} \left[\mathrm{Var}_{\mathbf{Z}} \left(\lim_{\sigma_N \to \infty} \sigma_N \log \left(\tilde{g}(\mathbf{X} + \sigma_N \mathbf{Z}) \right) \right) \right].
\end{aligned}
$$

Moreover, we also have that (see [29] for the details)

$$
\lim_{\sigma_N \to \infty} \sigma_N \log \left(\tilde{g}(\mathbf{x} + \sigma_N \mathbf{z}) \right) = -\mathbf{x}^{\mathsf{T}} \mathbf{z}. \qquad (2.8)
$$

Combining eq. (2.7), Lemma 1 and eq. (2.8), we arrive at

$$
\begin{aligned}
&\lim_{\sigma_N \to \infty} \sigma_N^2 \mathbb{E} \left[\mathrm{Var}(\iota(\mathbf{X}; \mathbf{Y}) | \mathbf{X}) \right] \\
&= \mathbb{E}_{\mathbf{X}} \left[\mathrm{Var}_{\mathbf{Z}} \left(-\mathbf{X}^{\mathsf{T}} \mathbf{Z} \right) \right] = \mathbb{E} \left[\|\mathbf{X}\|^2 \right],
\end{aligned}
$$

which concludes the proof of Theorem 2. □

2.1.4 Numerical evaluation

Theorem 2 shows that our bound in Theorem 1 is tight in the high-noise regime for a large family of prior distributions on $\mathbf{X}$. However, we suspect that such a tightness result holds more generally. To support this, we here provide numerical evaluations for three ‘toy’, yet practically relevant, scenarios.

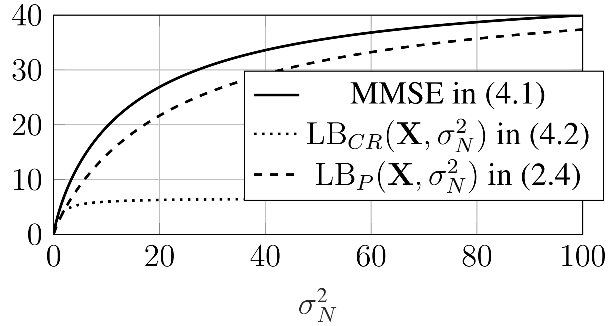

Figure 2.1 $\mathbf{X} \sim \mathcal{N}(\mathbf{0}_6, \boldsymbol{\Sigma}_{\mathbf{X}})$ with a randomly generated $\boldsymbol{\Sigma}_{\mathbf{X}}$.

1. *Gaussian Input.* We assume that $\mathbf{X} \sim \mathcal{N}(\mathbf{0}_k, \boldsymbol{\Sigma}_{\mathbf{X}})$, i.e., $\mathbf{X}$ is a Gaussian random vector. For this scenario, the MMSE is obtained as [30],

$$\text{mmse}(\mathbf{X}|\mathbf{Y}) = \mathsf{Tr}\left[\boldsymbol{\Sigma}_{\mathbf{X}}\left(\mathsf{I}_k + \frac{1}{\sigma_N^2}\boldsymbol{\Sigma}_{\mathbf{X}}\right)^{-1}\right], \tag{2.9}$$

and the Cramér-Rao lower bound evaluates to [2]

$$\text{mmse}(\mathbf{X}|\mathbf{Y}) \geq \frac{k^2}{\frac{k}{\sigma_N^2} + \mathsf{Tr}[\boldsymbol{\Sigma}_{\mathbf{X}}^{-1}]}. \tag{2.10}$$

In Figure 2.1, we plot the MMSE in eq. (4.1) (solid line), the Cramér-Rao bound in eq. (4.2) (dotted line), and our bound on the MMSE in eq. (2.4) (dashed line) versus different values of σ_N^2 for $k = 6$ and a randomly generated $\boldsymbol{\Sigma}_{\mathbf{X}}$.

2. *BPSK Input.* We let $k = 1$ and assume that $X \in \{-1, 1\}$ with equal probability, i.e., X is a Binary Phase Shift Keying (BPSK) signal with $P_X(1) = P_X(-1) = 1/2$. For this scenario, the MMSE is obtained as [31],

$$\text{mmse}(X|Y) = 1 - \int_{-\infty}^{\infty} \frac{\mathrm{e}^{-\frac{y^2}{2}}}{\sqrt{2\pi}} \tanh\left(\frac{1}{\sigma_N^2} - \frac{y}{\sigma_N}\right) \, \mathrm{d}y. \tag{2.11}$$

In Figure 2.2, we plot the MMSE in eq. (4.3) (solid line) and our lower bound in eq. (2.4) (dashed line) versus σ_N^2.

3. *Sparse Input.* We let $k = 1$ and assume that $X \sim P_X = (1 - \alpha)\delta_0 + \alpha\mathcal{N}(0, 1)$, where $\alpha \in [0, 1]$ and δ_0 is the point measure at 0. Such input distributions are used

to model sparsity and have been studied in [32–35]. To the best of our knowledge, a closed-form expression for the MMSE is not known. In Figure 4.3, we plot the MMSE (solid line) and our bound in eq. (2.4) (dashed line) versus σ_N^2 for $\alpha = 0.4$. The two curves were obtained via a Monte Carlo simulation with $5 \cdot 10^5$ iterations.

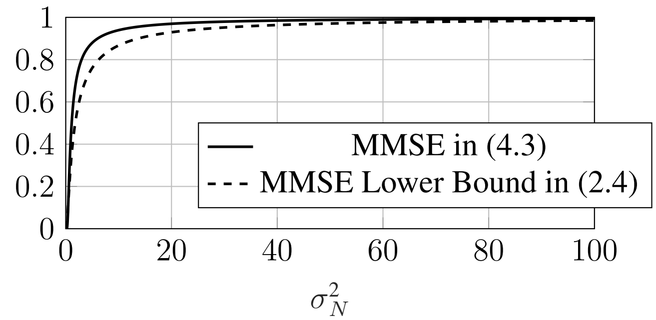

Figure 2.2 $X \sim P_X(x) = 1/2$ for $x \in \{-1, 1\}$.

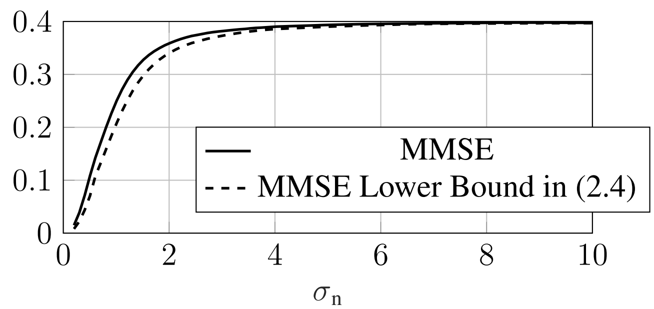

Figure 2.3 $X \sim P_X = (1-\alpha)\delta_0 + \alpha\mathcal{N}(0,1),\ \alpha = 0.4$.

From Figure 2.1, Figure 2.2, and Figure 2.3, we observe that our lower bound in eq. (2.4) well approximates the MMSE even in the finite noise regime. Moreover, for the scenario of a Gaussian input in Figure 2.1, our lower bound in eq. (2.4) remarkably outperforms the well-known Cramér-Rao bound. Finally, we point out that for the scenarios of a BPSK input (Figure 2.2) and a sparse input (Figure 2.3), i.e., where $\mathbf{X}$ has a discrete or a mixed distribution, commonly used lower bounds (e.g., Cramér-Rao) do not hold, whereas ours does. These examples suggest that our lower bound in eq. (2.4) might indeed be tight (or offer a performance guarantee) even in the finite noise regime. Hence, it would also be interesting to characterize the behavior of our lower bound in the low-noise regime.

However, in the low-noise regime, the MMSE has an intricate behavior, for example, it depends on whether the distribution of $\mathbf{X}$ is discrete or continuous [36].

In this section, we prove Theorem 1, which provides a new representation of the MMSE. Towards this end, we leverage the following proposition, which provides an expression for the gradient of the information density.

Proposition 2. *For* $\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}$*, we have that*

$$\nabla_{\mathbf{y}} \iota_{P_{\mathbf{XY}}}(\mathbf{x}; \mathbf{y}) = \mathbf{J}_{\mathbf{y}} \mathbf{T}(\mathbf{y})(\mathbf{x} - \mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{y}]).$$

Fix some $\mathbf{x} \in \mathcal{X}$. Then,

$$\begin{aligned}
&\nabla_{\mathbf{y}} \iota_{P_{\mathbf{XY}}}(\mathbf{x}; \mathbf{y}) \\
&\overset{(a)}{=} \nabla_{\mathbf{y}} \log h(\mathbf{y}) + \nabla_{\mathbf{y}} \langle \mathbf{x}, \mathbf{T}(\mathbf{y}) \rangle - \nabla_{\mathbf{y}} \log f_{\mathbf{Y}}(\mathbf{y}) \\
&\overset{(b)}{=} \mathbf{J}_{\mathbf{y}} \mathbf{T}(\mathbf{y})\mathbf{x} - \nabla_{\mathbf{y}} \log \frac{f_{\mathbf{Y}}(\mathbf{y})}{h(\mathbf{y})} \\
&\overset{(c)}{=} \mathbf{J}_{\mathbf{y}} \mathbf{T}(\mathbf{y})\mathbf{x} - \mathbf{J}_{\mathbf{y}} \mathbf{T}(\mathbf{y}) \mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{y}] \\
&= \mathbf{J}_{\mathbf{y}} \mathbf{T}(\mathbf{y})(\mathbf{x} - \mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{y}]),
\end{aligned}$$

where the labeled equalities follow from: (a) the definition of the exponential family in eq. (1.1); (b) using the fact that $\nabla_{\mathbf{y}} \langle \mathbf{x}, \mathbf{T}(\mathbf{y}) \rangle = \mathbf{J}_{\mathbf{y}} \mathbf{T}(\mathbf{y})\mathbf{x}$; and (c) using the TRE identity. This concludes the proof of Proposition 6.

Proof of Theorem 1. Since $\mathbf{J}_{\mathbf{Y}} \mathbf{T}(\mathbf{Y})$ has full rank a.s. $\mathbf{Y}$, then the pseudo inverse $(\mathbf{J}_{\mathbf{Y}} \mathbf{T}(\mathbf{Y}))^{+}$ exists a.s. $\mathbf{Y}$. Using Proposition 6, we have that a.s. $\mathbf{Y}$,

$$(\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]) = (\mathbf{J}_{\mathbf{Y}} \mathbf{T}(\mathbf{Y}))^{+} \nabla_{\mathbf{Y}} \iota_{P_{\mathbf{XY}}}(\mathbf{X}; \mathbf{Y}). \tag{2.12}$$

Now, taking the norm squared and the expectation of both sides of (4.4), and recalling that $\mathrm{mmse}(\mathbf{X}|\mathbf{Y}) = \mathbb{E}\left[||\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]||^2\right]$ we arrive at the desired result. This concludes the proof of Theorem 1.

To show an application of the new representation of the MMSE in Theorem 1, we here consider the following model,

$$Y = \frac{Z}{\sqrt{2X}}, \tag{2.13}$$

where Z is the standard normal random variable, i.e., $Z \sim \mathcal{N}(0,1)$ and X is the unknown variance drawn from the gamma distribution with $\alpha > 0$ shape, and $\beta > 0$ rate, i.e.,

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \mathrm{e}^{-\beta x}, \tag{2.14}$$

where Γ is the gamma function.

Using the channel model in eq. (4.5), we therefore obtain,

$$f_{Y|X}(y|x) = \sqrt{\frac{x}{\pi}} \mathrm{e}^{-xy^2}. \tag{2.15}$$

The conditional pdf in eq. (4.7) can be mapped to the exponential family in eq. (1.1) through the following mapping,

$$h(y) = \sqrt{\frac{1}{\pi}},\ \phi(x) = -\log\left(\sqrt{x}\right),\ T(y) = -y^2. \tag{2.16}$$

We now evaluate the MMSE expression in Theorem 1 for our model in eq. (4.5) with the mapping in eq. (4.8). We note that

$$\left(\frac{\mathrm{d}}{\mathrm{d}y} T(y)\right)^{-1} = -\frac{1}{2y}. \tag{2.17}$$

We now focus on deriving $\frac{\mathrm{d}}{\mathrm{d}y} \iota_{P_{XY}}(x;y)$, where

$$\iota_{P_{XY}}(x;y) = \log\left(f_{Y|X}(y|x)\right) - \log\left(f_Y(y)\right),$$

where $f_{Y|X}$ is defined in eq. (4.7). Hence, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}y} \log\left(f_{Y|X}(y|x)\right) = -\frac{\mathrm{d}}{\mathrm{d}y}\left(xy^2\right) = -2xy. \tag{2.18}$$

To complete the evaluation of the MMSE in Theorem 1, we need $\frac{\mathrm{d}}{\mathrm{d}y} \log\left(f_Y(y)\right)$, where (see [29] for the details)

$$f_Y(y) = \sqrt{\frac{1}{\pi}} \beta^{\alpha} \frac{\Gamma\left(\alpha + \frac{1}{2}\right)}{\Gamma(\alpha)} \frac{1}{(y^2 + \beta)^{\alpha + \frac{1}{2}}}.$$

Thus, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}y} \log\left(f_Y(y)\right) = \frac{1}{f_Y(y)} \frac{\mathrm{d}}{\mathrm{d}y} f_Y(y) = -\frac{y(2\alpha + 1)}{y^2 + \beta}. \tag{2.19}$$

Finally, by substituting eq. (4.9), eq. (4.10) and eq. (4.11) inside the MMSE expression in Theorem 1, we obtain (see [29] for the details)

$$\mathrm{mmse}(X|Y) = \frac{\alpha(\alpha + 1)}{\beta^2 \left(\alpha + \frac{3}{2}\right)}.$$

We note that, in order to compute the MMSE above, we did not need to compute $\mathbb{E}[X|Y]$ (which is needed by the classical representation of the MMSE), but only the marginal pdf f_Y, which can be done by simple computations. Moreover, we also highlight that the MMSE above is in closed-form, and this expression highlights that the MMSE only depends on the parameters of the gamma distribution.

2.2 Analysis of the Ziv-Zakai Lower Bound

2.2.1 Noise models

An interesting bound that beautifully connects binary hypothesis testing and estimation is known as the Ziv-Zakai bound [6–10, 37]. This bound is believed to be one of the tightest bounds available in the literature. Before presenting the bound we need the following two notions. The *valley-filling* function acting on a function $f : \mathbb{R} \to \mathbb{R}$ is defined as

$$\mathcal{V}\{f(x)\} = \sup_{\varepsilon \geq 0} f(x + \varepsilon),\ x \in \mathbb{R}. \tag{2.20}$$

For a given $x_0, x_1 \in \mathbb{R}$, $P_e\left[x_0, x_1, p_0, p_1\right]$ denotes *the minimum probability of error* (obtained by using the optimal likelihood ratio test) for the following binary hypothesis testing problem,

$$H_0 : \mathbf{Y} \sim P_{\mathbf{Y}|X}(\cdot|x_0),$$
$$H_1 : \mathbf{Y} \sim P_{\mathbf{Y}|X}(\cdot|x_1),$$

where

$$\mathsf{Pr}(H_0) = p_0,\ \mathsf{Pr}(H_1) = 1 - \mathsf{Pr}(H_0) = p_1.$$

The general Ziv-Zakai lower bound is stated next [8, 10]. *(Ziv-Zakai Lower Bound.)* Consider a pair of random variables $(X, \mathbf{Y})$ where X has probability density function (pdf) $f_X(x)$ and where the noisy observation model $\mathbf{Y}|X = x$ is governed by the distribution $P_{\mathbf{Y}|X}(\cdot|x)$. Then, we have

$$\mathrm{mmse}(X|\mathbf{Y}) \geq \overline{\mathsf{LB}_{\mathsf{ZZ}}}(X|\mathbf{Y}), \tag{2.21}$$

where

$$\overline{\mathsf{LB}_{\mathsf{ZZ}}}(X|\mathbf{Y}) = \frac{1}{2}\int_0^\infty \mathcal{V}\left\{\int_{-\infty}^{\infty} P_e\left[x, x+h, p_0(x,h), p_1(x,h)\right] \cdot \left(f_X(x) + f_X(x+h)\right)\, \mathrm{d}x\right\} h\, \mathrm{d}h, \tag{2.22}$$

with

$$p_0(x,h) = \frac{f_X(x)}{f_X(x) + f_X(x+h)},\ p_1(x,h) = 1 - p_0(x,h).$$

The valley-filling function in Theorem 2.2.1 introduces an extra layer of optimization which can make the bound difficult to evaluate. Thus, one often considers a loosened version of the Ziv-Zakai bound that drops the valley-filling function, that is,

$$\mathsf{LB}_{\mathsf{ZZ}}(X|\mathbf{Y}) = \frac{1}{2}\int_0^\infty \int_{-\infty}^{\infty} P_e\left[x, x+h, p_0(x,h), p_1(x,h)\right] \cdot \left(f_X(x) + f_X(x+h)\right) h\, \mathrm{d}x\, \mathrm{d}h. \tag{2.23}$$

In this work, our main goal is to understand the behavior of the Ziv-Zakai bounds in eq. (2.22) and in eq. (2.23) in the practically relevant high-noise regime. Different from the low-noise regime where several lower bounds are known to perform well [11, 36], in the high-noise the same is not true in general. Thus, it is of interest to understand if the Ziv-Zakai bound is tight in high-noise. The analysis of this regime is an important benchmark for the performance, especially in wireless scenarios, where high-noise represents a weak signal scenario, and has received some attention in various contexts [28, 38–40].

Notation. Random variables are denoted by upper case letters and their instances by lower case letters. The expected value of a random variable X and its variance are denoted by $\mathbb{E}[X]$ and $\mathsf{Var}(X)$, respectively.

We here describe a family of noise distributions for which our results hold. First, in order to take limits and quantify the strength of the noise, we need to be able to parameterize our model and hence, we make the following assumption,

A1: $P_{\mathbf{Y}|X}$ can be parameterized in terms of the parameter $\eta \geq 0$, i.e., $P_{\mathbf{Y}|X}(\mathbf{y}|x) = P_{\mathbf{Y}|X}(\mathbf{y}|x;\eta)$ for all $(x, \mathbf{y})$. We refer to the parameter η as the *noise level*.

Second, we require that the performance of our system degrades as the noise level increases. Towards this end, we make the following assumption,

A2: For the sequence of noisy observation models $\{P_{\mathbf{Y}|X}(\cdot|\cdot;\eta)\}_{\eta\geq 0}$, we parameterize $P_e\left[x_0, x_1, p_0, p_1\right] = P_e\left[\eta; x_0, x_1, p_0, p_1\right]$, and we assume that the following holds. For every (x_0, x_1, p_0, p_1) where $x_0 \neq x_1$:

A2a: $\eta \mapsto P_e[\eta; x_0, x_1, p_0, p_1]$ is non-decreasing; and

$$\textbf{A2b:} \lim_{\eta\to\infty} P_e[\eta; x_0, x_1, p_0, p_1] = \min\{p_0, p_1\}. \tag{2.24}$$

All of the assumptions above are rather natural. In particular, assumption **A2a** simply states that the probability of error for binary detection can not decrease as the noise level increases. Assumption **A2b** states that when the observation is completely dominated by the noise, the best strategy is to guess the $x_i, i \in \{0,1\}$ with the largest probability.

Most of the observation models encountered in practice satisfy the above assumptions. We now give a few examples.

- *Additive White Gaussian Model:* Let

$$Y = X + \sqrt{\eta} Z, \tag{2.25}$$

  with X and Z being independent, and Z being a standard Gaussian random variable. In this case, the noise level parameter $\eta > 0$ is known as the *noise power* [30].

- *Poisson Noise Model:* For $x \geq 0,\ y \in \mathbb{N}_0$, let

$$P_{Y|X}(y|x) = \frac{(x+\eta)^y \mathrm{e}^{-(x+\eta)}}{y!}; \tag{2.26}$$

  in this case, the noise level parameter $\eta > 0$ is known as the *dark current* parameter [41, 42].

- *Binary Symmetric Model:* For $x \in \{0,1\}$ and $y \in \{0,1\}$, let

$$P_{Y|X}(y|x) = \begin{cases} 1-\eta & x = y, \\ \eta & x \neq y, \end{cases} \tag{2.27}$$

  where the noise level parameter $\eta \in \left(0, \frac{1}{2}\right)$ is known as the *cross over probability* [43].

See also [44] for an example on how to define the noise level parameter η in the context of the exponential family.

For the models that satisfy all of the above assumptions, we parameterize $\mathrm{mmse}(X|\mathbf{Y})$ as a function of X and η. Thus, in what follows we denote it as $\mathrm{mmse}(X,\eta)$. Similarly, we use $\overline{\mathsf{LB}_{\mathsf{ZZ}}}(X,\eta)$ and $\mathsf{LB}_{\mathsf{ZZ}}(X,\eta)$ for $\overline{\mathsf{LB}_{\mathsf{ZZ}}}(X|\mathbf{Y})$ in eq. (2.22) and $\mathsf{LB}_{\mathsf{ZZ}}(X|\mathbf{Y})$ in eq. (2.23), respectively.

2.2.2 High noise asymptotics

The next theorem provides the behavior of the two Ziv-Zakai lower bounds. Under the assumptions **A1** and **A2**, we have the following,

$$\mathbb{V}(X) = \frac{1}{4}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \min\{f_X(y), f_X(x)\}\,|y-x|\,\mathrm{d}x\,\mathrm{d}y,$$
$$\overline{\mathbb{V}}(X) = \frac{1}{4}\int_{-\infty}^{\infty}\mathcal{V}\Big\{\int_{-\infty}^{\infty}\min\{f_X(x), f_X(x+h)\}\,\mathrm{d}x\Big\}\,|h|\,\mathrm{d}h.$$

The examples below show that, while there are cases for which even the loosened version of the Ziv-Zakai bound is tight (i.e., $\mathbb{V}(X) = \mathsf{Var}(X)$), in general, neither of the bounds is tight.

Example 1. Suppose that $f_X(x) = g(|x|)$ where $g : [0,\infty) \to [0,\infty)$ is non-increasing. Then,

$$\mathbb{V}(X) = \overline{\mathbb{V}}(X) = \mathsf{Var}(X).$$

Thus, both versions of the Ziv-Zakai bound agree and are tight, i.e., the valley-filling function is not needed in this case. This example encompasses a broad range of widely used symmetric distributions (e.g., Gaussian, Laplace, generalized normal). □

Example 2. Suppose that $f_X(x)$ is non-increasing on (a,b) where $-\infty < a < b \le \infty$, and zero elsewhere. Then,

$$\mathbb{V}(X) = \overline{\mathbb{V}}(X) = \frac{\mathsf{Var}(X) + (a - \mathbb{E}[X])^2}{4}.$$

Thus, the two versions of the Ziv-Zakai bound agree, but are *not* tight. For instance, assume that X has an exponential pdf with parameter λ; we have that $\mathsf{Var}(X) = 1/\lambda^2$ and $\mathbb{V}(X) = \overline{\mathbb{V}}(X) = 1/(2\lambda^2)$, i.e., the Ziv-Zakai bound is off by a factor of two and hence, it can be substantially suboptimal. □

Example 3. Suppose that $0 < a < b$ and

$$f_X(x) = \frac{1}{2(b-a)}\left(\mathrm{rect}(x; -b, -a) + \mathrm{rect}(x; a, b)\right),$$

where $x \mapsto \mathrm{rect}(x; a, b)$ is the unit-height rectangle with support over the interval (a, b). In other words, the distribution of X is a mixture of two uniform distributions. Then, for $0 < a < b < 3a$, we have that

$$\mathsf{Var}(X) = \frac{a^2 + ab + b^2}{3},$$

and

$$\begin{aligned}\mathbb{V}(X) &= \mathsf{Var}(X) - \frac{a(a+b)}{2},\\ \overline{\mathbb{V}}(X) &= \mathbb{V}(X) + \frac{7a^2 + 10ab - b^2}{32}\\ &= \mathsf{Var}(X) - \frac{(3a+b)^2}{32}.\end{aligned}$$

Thus, the Ziv-Zakai bound with the valley-filling function can be strictly better than the one without it, yet not optimal. $\square$

We were not able to identify an example for which $\mathbb{V}(X) < \overline{\mathbb{V}}(X) = \mathsf{Var}(X)$, i.e., a case for which the Ziv-Zakai bound with a valley-filling function is optimal, but the bound without the valley-filling function is strictly sub-optimal.

2.2.3 Comparison with the Cramer-Rao lower bound

An interesting question that arises is: Does the Ziv-Zakai bound outperform the CR bound[2]? To show this analytically, one would need to demonstrate the following inequality,

$$\frac{1}{J(X)} \leq \overline{\mathbb{V}}(X).$$

In Figure 2.4, this inequality is numerically verified for a mixed Gaussian distribution.

Before proceeding with the proof, we will need the following facts. First, by using eq. (2.20), note that if $f(x) \leq g(x)$ for all x, then for all x it holds that

$$\mathcal{V}\{f(x)\} \leq \mathcal{V}\{g(x)\}. \tag{2.28}$$

[2]We note that the Ziv-Zakai bound holds for a larger family of distributions than the CR bound as the pdf does not need to be differentiable or even continuous.

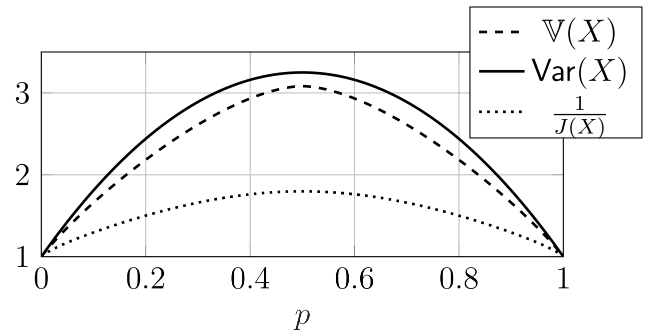

Figure 2.4 Mixed Gaussian $f_X = p\mathcal{N}(-1,1) + (1-p)\mathcal{N}(2,1)$ where $p \in [0,1]$.

Second, the valley-filling function is lower semicontinuous, i.e., for any sequence of functions $\{f_n\}_{n=1}^{\infty}$, we have that

$$\liminf_{n\to\infty} \mathcal{V}\{f_n(x)\} \geq \mathcal{V}\{\liminf_{n\to\infty} f_n(x)\}. \tag{2.29}$$

To see this recall that $\liminf_{n\to\infty} f_n(x) = \sup_{n\geq 0} \inf_{m\geq n} f_m(x)$ and note that

$$\begin{aligned}\liminf_{n\to\infty} \mathcal{V}\{f_n(x)\} &= \sup_{n\geq 0} \inf_{m\geq n} \sup_{\varepsilon\geq 0} f_m(x+\varepsilon) \\ &\geq \sup_{\varepsilon\geq 0} \sup_{n\geq 0} \inf_{m\geq n} f_m(x+\varepsilon) \\ &= \mathcal{V}\{\liminf_{n\to\infty} f_n(x)\},\end{aligned} \tag{2.30}$$

where the inequality follows from the max-min inequality.

We now consider the behavior of $\overline{\mathsf{LB}_{\mathsf{ZZ}}}(X,\eta)$. We have

$$\begin{aligned}&2\overline{\mathsf{LB}_{\mathsf{ZZ}}}(X,\eta) \\ &= \int_0^\infty \mathcal{V}\left\{\int_{-\infty}^\infty P_e\left[\eta; x, x+h, p_0(x,h), p_1(x,h)\right] \right. \\ &\qquad \left. \cdot \left(f_X(x) + f_X(x+h)\right) \mathrm{d}x\right\} h \; \mathrm{d}h \\ &\overset{\text{(a)}}{\leq} \int_0^\infty \mathcal{V}\left\{\int_{-\infty}^\infty P_e\left[\infty; x, x+h, p_0(x,h), p_1(x,h)\right] \right. \\ &\qquad \left. \cdot \left(f_X(x) + f_X(x+h)\right) \mathrm{d}x\right\} h \; \mathrm{d}h \\ &\overset{\text{(b)}}{=} \int_0^\infty \mathcal{V}\left\{\int_{-\infty}^\infty \min\left\{f_X(x), f_X(x+h)\right\} \mathrm{d}x\right\} h \; \mathrm{d}h,\end{aligned} \tag{2.31}$$

where (a) follows by using eq. (2.28) and by noting that

$$P_e\left[\eta; x, x+h, p_0, p_1\right] \leq P_e\left[\infty; x, x+h, p_0, p_1\right],$$

which is a consequence of the assumption **A2a**; and (b) follows from the assumption **A2b**.

Next, we note that

$$\begin{aligned}
&\liminf_{\eta\to\infty}\ 2\overline{\mathsf{LB}_{\mathsf{ZZ}}}(X,\eta) \\
&\overset{(\mathrm{c})}{\geq} \int_0^{\infty} \liminf_{\eta\to\infty} \mathcal{V}\left\{\int_{-\infty}^{\infty} P_e\left[\eta; x, x+h, p_0(x,h), p_1(x,h)\right]\right. \\
&\qquad \left. \cdot \left(f_X(x) + f_X(x+h)\right) \mathrm{d}x\right\} h \,\mathrm{d}h \\
&\overset{(\mathrm{d})}{\geq} \int_0^{\infty} \mathcal{V}\left\{\liminf_{\eta\to\infty} \int_{-\infty}^{\infty} P_e\left[\eta; x, x+h, p_0(x,h), p_1(x,h)\right]\right. \\
&\qquad \left. \cdot \left(f_X(x) + f_X(x+h)\right) \mathrm{d}x\right\} h \,\mathrm{d}h \\
&\overset{(\mathrm{e})}{\geq} \int_0^{\infty} \mathcal{V}\Big\{\int_{-\infty}^{\infty} \min\left\{f_X(x), f_X(x+h)\right\} \mathrm{d}x\Big\}\, h \,\mathrm{d}h,
\end{aligned} \tag{2.32}$$

where: (c) follows by using Fatou's lemma; (d) follows from eq. (2.29); and (e) follows by using Fatou's lemma, (2.28) and assumption **A2b**.

Combining the upper bound on the limit in eq. (2.31) and the lower bound on the limit in eq. (2.32) we arrive at

$$\begin{aligned}
&\lim_{\eta\to\infty} 2\overline{\mathsf{LB}_{\mathsf{ZZ}}}(X,\eta) \\
&= \int_0^{\infty} \mathcal{V}\Big\{\int_{-\infty}^{\infty} \min\left\{f_X(x), f_X(x+h)\right\} \mathrm{d}x\Big\}\, h \,\mathrm{d}h \\
&= \frac{1}{2}\int_{-\infty}^{\infty} \mathcal{V}\Big\{\int_{-\infty}^{\infty} \min\left\{f_X(x), f_X(x+h)\right\} \mathrm{d}x\Big\}\, |h| \,\mathrm{d}h,
\end{aligned}$$

where in the last step we have used that

$$\begin{aligned}
&\int_{-\infty}^{\infty} \min\left\{f_X(x), f_X(x+h)\right\}\ \mathrm{d}x \\
&= \int_{-\infty}^{\infty} \min\left\{f_X(x), f_X(x-h)\right\}\ \mathrm{d}x.
\end{aligned}$$

This concludes the proof of the limit for $\overline{\mathsf{LB}_{\mathsf{ZZ}}}(X,\eta)$. To obtain the limit for $\mathsf{LB}_{\mathsf{ZZ}}(X,\eta)$, we simply drop the valley-filling function. With this we arrive at

$$\begin{aligned}
&\lim_{\eta\to\infty} 2\mathsf{LB}_{\mathsf{ZZ}}(X,\eta) \\
&= \frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \min\left\{f_X(x), f_X(x+h)\right\} |h| \,\mathrm{d}x \,\mathrm{d}h \\
&= \frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \min\left\{f_X(x), f_X(y)\right\} |y-x| \,\mathrm{d}x \,\mathrm{d}y.
\end{aligned}$$

This concludes the proof of Theorem 2.2.2.

2.3 Examples for the High-Noise Regime

2.3.1 Example 1

We will show that $\mathbb{V}(X) = \mathsf{Var}(X)$, which will also characterize $\overline{\mathbb{V}}(X)$. By substituting $f_X(x) = g(|x|)$ inside the expression of $\mathbb{V}(X)$ in Theorem 2.2.2, we arrive at

$$\begin{aligned}
4\,\mathbb{V}(X) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \min\{g(|x|), g(|y|)\}\ |y-x| \,\mathrm{d}x \,\mathrm{d}y \\
&= 4\int_{0}^{\infty}\int_{-y}^{y} g(|y|)\ |y-x| \,\mathrm{d}x \,\mathrm{d}y \\
&= 8\int_{0}^{\infty} g(|y|)\ y^2 \,\mathrm{d}y \overset{\text{(a)}}{=} 4\,\mathsf{Var}(X),
\end{aligned}$$

where (a) follows since $\mathbb{E}[X] = 0$ from the structure of $f_X(x)$.

2.3.2 Example 2

By using the expression of $\mathbb{V}(X)$ in Theorem 2.2.2, we arrive at

$$\begin{aligned}
4\,\mathbb{V}(X) &= \int_{a}^{b}\int_{a}^{b} \min\{f_X(x), f_X(y)\}\ |y-x| \,\mathrm{d}x \,\mathrm{d}y \\
&= 2\int_{a}^{b}\int_{a}^{y} f_X(y)\ |y-x| \,\mathrm{d}x \,\mathrm{d}y \\
&= \int_{a}^{b} f_X(y)\ |y-a|^2 \,\mathrm{d}y \\
&= \mathbb{E}[(X-a)^2] = \mathsf{Var}(X) + (a-\mathbb{E}[X])^2.
\end{aligned}$$

By using the expression of $\overline{\mathbb{V}}(X)$ in Theorem 2.2.2, we have

$$
\begin{aligned}
2\overline{\mathbb{V}}(X) &\overset{\text{(a)}}{=} \int_0^\infty \mathcal{V}\Big\{ \int_a^b \min\left\{f_X(x), f_X(x+h)\right\} \mathrm{d}x \Big\}\; h \;\mathrm{d}h \\
&\overset{\text{(b)}}{=} \int_0^\infty \mathcal{V}\Big\{ \int_a^{\max\{a,b-h\}} f_X(x+h)\mathrm{d}x \Big\}\; h \;\mathrm{d}h \\
&= \int_0^\infty \mathcal{V}\Big\{ \int_{a+h}^{\max\{a+h,b\}} f_X(x)\mathrm{d}x \Big\}\; h \;\mathrm{d}h \\
&\overset{\text{(c)}}{=} \int_0^\infty \mathcal{V}\Big\{ \mathbb{P}[X \geq a+h] \Big\}\; h \;\mathrm{d}h \\
&\overset{\text{(d)}}{=} \int_0^\infty \mathbb{P}[X \geq a+h]\; h \;\mathrm{d}h \\
&\overset{\text{(e)}}{=} \frac{1}{2}\mathbb{E}[(X-a)^2] = \frac{1}{2}\left(\mathsf{Var}(X) + (a - \mathbb{E}[X])^2\right),
\end{aligned}
$$

where the labeled equalities follow from: (a) using the fact that X is supported on (a, b); (b) the assumption that $f_X(x)$ is a non-increasing function; (c) the fact that X is supported on (a, b) and hence, we can drop the upper limit; (d) the fact that $h \mapsto \mathbb{P}[X \geq a+h]$ is a non-increasing function and hence, the valley-filling function can be dropped; and (e) the following alternative representation of the second moment of non-negative random variables: for a random variable $U \geq 0$, we have that $\mathbb{E}[U^2] = 2\int_0^\infty \mathbb{P}[U \geq h]\; h \;\mathrm{d}h$.

CHAPTER 3

PROBLEM FORMULATION (PART II)

3.1 Preliminaries

We consider a channel for which the relationship between the input $X \in [0,1]$ and the output $Y \in \{0,\ldots,n\}$ is described by the binomial distribution:

$$P_{Y|X}(y|x) = \binom{n}{y} x^y (1-x)^{n-y}. \tag{3.1}$$

In this work, we are interested in studying the capacity of this channel as a function of the number of trials n, that is

$$C(n) = \max_{P_X \colon X \in [0,1]} I(X;Y). \tag{3.2}$$

In addition to studying capacity, we are also interested in studying properties of an optimal capacity-achieving distribution distribution denoted by $P_{X^\star}$.

3.1.1 Literature Review

The binomial channel naturally arises in molecular communications and the interested reader is referred to [45–48] and references therein. The channel is also useful in the study of the deletion channel [49, 50].

The capacity of the binomial channel was first considered in [51] where the authors used minimax redundancy theorem in [52] to argue that asymptotically the capacity scales as $\frac{1}{2}\log n$. The exact capacity for the $n = 1$ case was computed in [45] where binary distribution with support on $\{0,1\}$ was shown to be capacity-achieving. To the best of our knowledge, there are no firm bounds on the capacity of the binomial channel.

Properties of the capacity-achieving distribution have also been looked at. For example, the authors of [45] have designed an algorithm for computing capacity and a capacity-achieving distribution by using a dual representation of the maximization problem. It is also known that, by using the Witsenhausen technique [53], there exists a capacity-achieving distribution with at most $n + 1$ mass points. We note, however, that the Witsenhausen technique does not guarantee that the optimal input distribution is unique. In fact, for the binomial channel, uniqueness has not been shown; note that uniqueness is important not just for theoretical purposes, but also for algorithmic purposes. A conventional way to show that the capacity-achieving distribution is unique is by establishing that the mutual information is a *strictly* concave function of the input distribution. However, as will be shown by an example, for the binomial channel, the mutual information is not strictly concave. Other properties, such as location of the support points, are also not well understood. The main goal of this work is to close some of these gaps.

In this work, we also rely on estimation theoretic quantities such as the conditional expectation. For the estimation theoretic treatments of the binomial channel, the interested reader is referred to [54,55]. Recently, deterministic identification capacity for the binomial channel has been studied in [56].

3.1.2 KKT conditions

The key that allows one to study properties of the support of an optimal input distribution is the following lemma which contains the KKT conditions for our optimization problem [57].

$P_{X^\star}$ is a capacity-achieving input distribution if and only if the following conditions hold:

$$i(x; P_{Y^\star}) \le C(n), \qquad x \in [0,1], \tag{3.3}$$

$$i(x; P_{Y^\star}) = C(n), \qquad x \in \mathsf{supp}(P_{X^\star}) \tag{3.4}$$

where $P_{X^\star} \to P_{Y|X} \to P_{Y^\star}$ (i.e., the optimal output distribution) and

$$i(x; P_{Y^\star}) = P_{Y|X}(\cdot|x) P_{Y^\star}. \tag{3.5}$$

We also define the following set, which would be useful in our study of the uniqueness of $P_{X^\star}$:

$$\mathcal{A}_n = \{x \in [0,1] : \ i(x; P_{Y^\star}) - C(n) = 0\}. \tag{3.6}$$

The importance of $\mathcal{A}_n$ is demonstrated in the following lemma. For a given n

- $\mathcal{A}_n$ is unique; and
- $\mathsf{supp}(P_{X^\star}) \subseteq \mathcal{A}_n$ for every $P_{X^\star}$.

Proof. Note that, for a given n, both $P_{Y^\star}$ and $C(n)$ are unique (even if $P_{X^\star}$ is not unique) [58] and, since $\mathcal{A}_n$ only depends on these quantities, the uniqueness follows.

The second part follows from the KKT conditions in Lemma 3.1.2, because $x \in \mathsf{supp}(P_{X^\star})$ implies $x \in \mathcal{A}_n$. □

3.1.3 Estimation theoretic preliminaries

Estimation theoretic quantities will play an important role in our analysis. In what follows, the quantity $\mathbb{E}^{n-1}\left[f(Y) \mid X = x\right]$ denotes expectation with respect to a binomial distribution with $n-1$ trials and success probability x per trial, and

$$\ell_b(x, \hat{x}) = x \log\left(\frac{x(1-\hat{x})}{(1-x)\hat{x}}\right) - \frac{x-\hat{x}}{1-\hat{x}}, \quad (x, \hat{x}) \in (0,1)^2 \tag{3.7}$$

represents the Bregman divergence for the binomial channel.

We now summarize some of these preliminary results.

Proposition 3. *For $n \geq 2$ and $x \in (0,1)$, we have*

$$\begin{aligned} i'(x; P_Y) = &\frac{n}{x}\mathbb{E}^{n-1}\left[\ell_b(x, \mathbb{E}^{n-1}\left[X \mid Y\right]) \,\middle|\, X = x\right] \\ &+ \frac{n}{x}\mathbb{E}^{n-1}\left[\frac{x - \mathbb{E}^{n-1}\left[X \mid Y\right]}{1 - \mathbb{E}^{n-1}\left[X \mid Y\right]} \,\middle|\, X = x\right] \end{aligned} \tag{3.8}$$

and

$$i''(x; P_Y) = \frac{n}{x(1-x)} + \frac{1}{(1-x)^2}G(x) \tag{3.9}$$

$$G(x) = \tag{3.10}$$
$$\mathbb{E}\left[(n-Y)(n-Y-1)\log \frac{\mathbb{E}\left[X \mid Y=Y\right]}{\mathbb{E}\left[1-X \mid Y=Y+1\right]} \frac{\mathbb{E}\left[1-X \mid Y=Y+2\right]}{\mathbb{E}\left[X \mid Y=Y+1\right]} \,\middle|\, X=x\right]$$

The Bregman divergence in eq.(3.8) appeared previously in a different but related result, specifically in [54, Prop. 2] it was shown that for $a \in (0,1)$

$$\frac{\partial}{\partial a} I(X; \mathcal{B}_n(aX)) = \frac{n}{a}\mathbb{E}\left[\ell_b\left(aX, \mathbb{E}[aX|\mathcal{B}_{n-1}(aX')]\right)\right] \tag{3.11}$$

where $Y = \mathcal{B}_n(aX)$ denotes the transformation of input aX through a binomial channel with n trials.

Finally, we also need to show the monotonicity of the conditional mean.

The function $y \mapsto \mathbb{E}\left[X \mid Y=y\right]$ is non-decreasing.

Proof. First of all, note that

$$\mathbb{E}\left[X \mid Y=y\right] = \frac{\mathbb{E}\left[X^{y+1}(1-X)^{n-y}\right]}{\mathbb{E}\left[X^{y}(1-X)^{n-y}\right]}. \tag{3.12}$$

Let us now introduce the functions f_1, f_2, g_1, g_2 as follows:

$$f_1(x) = x^y, \quad f_2(x) = x^{y+1}, \tag{3.13}$$

$$g_1(x) = (1-x)^{n-y}, \quad g_2(x) = x(1-x)^{n-y-1}, \tag{3.14}$$

and note that the functions

$$\frac{f_2(x)}{f_1(x)} = x, \qquad \frac{g_2(x)}{g_1(x)} = \frac{x}{1-x} \tag{3.15}$$

are both increasing and non-negative for $x \in [0,1]$. As a consequence, the entries and the determinant of the matrices

$$\begin{bmatrix} f_1(x_1) & f_1(x_2) \\ f_2(x_1) & f_2(x_2) \end{bmatrix}, \qquad \begin{bmatrix} g_1(x_1) & g_1(x_2) \\ g_2(x_1) & g_2(x_2) \end{bmatrix}, \tag{3.16}$$

are non-negative for any choice of $0 \leq x_1 < x_2 \leq 1$. By using the *basic composition formula* of [59, Ch. 3.1], we can also say that the entries and the determinant of the matrix

$$\begin{bmatrix} \mathbb{E}\left[f_1(X)g_1(X)\right] & \mathbb{E}\left[f_1(X)g_2(X)\right] \\ \mathbb{E}\left[f_2(X)g_1(X)\right] & \mathbb{E}\left[f_2(X)g_2(X)\right] \end{bmatrix} \tag{3.17}$$

are non-negative. Therefore, we have

$$\frac{\mathbb{E}\left[f_2(X)g_2(X)\right]}{\mathbb{E}\left[f_1(X)g_2(X)\right]} \geq \frac{\mathbb{E}\left[f_2(X)g_1(X)\right]}{\mathbb{E}\left[f_1(X)g_1(X)\right]} \tag{3.18}$$

or

$$\frac{\mathbb{E}\left[X^{y+2}(1-X)^{n-y-1}\right]}{\mathbb{E}\left[X^{y+1}(1-X)^{n-y-1}\right]} \geq \frac{\mathbb{E}\left[X^{y+1}(1-X)^{n-y}\right]}{\mathbb{E}\left[X^{y}(1-X)^{n-y}\right]}, \tag{3.19}$$

which, by using eq.(3.12), is the same as

$$\mathbb{E}\left[X \mid Y = y+1\right] \geq \mathbb{E}\left[X \mid Y = y\right]. \tag{3.20}$$

This concludes the proof. □

3.2 Properties of the Capacity-Achieving Distributions

In this section we study properties of capacity-achieving distributions.

3.2.1 Discreteness

As already mentioned in section 3.1.1, from the Witsenhausen approach we only know that there exists a discrete distribution with at most $n+1$ mass points. This, however, does not rule out the existence of other capacity-achieving distributions (e.g., continuous capacity-achieving distributions).

We now show that all capacity-achieving distributions are discrete and provide a preliminary bound on the support.

Proposition 4. $|\mathcal{A}_n| \leq n + 1$.

3.2.2 Uniqueness of the Optimal Input Distribution

In this section, we show and discuss uniqueness of the capacity-achieving input distribution. To aid our discussion, it is useful to parameterize the mutual information in terms of distributions instead of random variables, that is

$$I(P_X; P_{Y|X}) = I(X; Y). \tag{3.21}$$

We also let $\mathcal{P}_{\mathcal{X}}$ be the set of all distributions over the set $\mathcal{X}$. In particular, the optimization in eq.(3.2) can be written as

$$\max_{P_X \in \mathcal{P}_{[0,1]}} I(P_X; P_{Y|X}). \tag{3.22}$$

A typical way to show that there is a unique maximizer is to show that the mapping $P_X \mapsto I(P_X; P_{Y|X})$ over the set $\mathcal{P}_{[0,1]}$ is *strictly* concave [60]. However, due to the fact that the output space of the binomial channel is finite and the input space is uncountable, the mutual information is not strictly concave over $\mathcal{P}_{[0,1]}$. For example, when $n = 1$ any distribution symmetric around $x = \frac{1}{2}$ will induce

$$P_Y(0) = P_Y(1) = \frac{1}{2} \tag{3.23}$$

which is the capacity-achieving output distribution for $n = 1$. Therefore, to show uniqueness of the capacity-achieving input distribution a new or slightly different argument is needed.

We begin by showing the following result.

Proposition 5. *Consider an arbitrary sequence* $0 \leq x_1 < \ldots < x_{n+1} \leq 1$ *and define the matrix* $\mathsf{A} \in \mathbb{R}^{n+1 \times n+1}$ *as*

$$[\mathsf{A}]_{i,k} = P_{Y|X}(i-1|x_k), \quad i \in [n+1],\, k \in [n+1]. \tag{3.24}$$

Then, A *is full rank.*

Proof. First of all, we argue that considering $x_1 = 0$ and $x_{n+1} = 1$ comes without loosing generality. In fact, in this case the first and last columns of A are $\mathbf{e}_1$ and $\mathbf{e}_{n+1}$, respectively, where $\mathbf{e}_i$ is a zero vector with a 1 in the i-th position. As a consequence, we have $\det(\mathsf{A}) = \det(\tilde{\mathsf{A}})$, where

$$[\tilde{\mathsf{A}}]_{i,k} = [\mathsf{A}]_{i+1,k+1}, \quad i \in [n-1],\, k \in [n-1]. \tag{3.25}$$

Next, note that we can rewrite the binomial law as

$$P_{Y|X}(y|x) = \binom{n}{y}(1+^{t})^{-nty} \tag{3.26}$$

where $x = \frac{t}{1+^{t}}$. The matrix B with $[\mathsf{B}]_{y,k} =^{t_k y}$ and $y \in [n-1]$ is a Vandermonde matrix, which is full rank since the t_k's are all distinct [61]. Thanks to the multilinear property of the determinant, we can write that

$$\det(\tilde{\mathsf{A}}) = \det(\mathsf{B}) \prod_{y=1}^{n-1} \binom{n}{y} \prod_{k=2}^{n} (1+^{t_k})^{-n} > 0 \tag{3.27}$$

where the last step is due to $\det(\mathsf{B}) > 0$ and to the positivity of the products. As a consequence, A is a full rank matrix. □

With the aid of Proposition 5, we show the following result.

Theorem 3. *Let $\mathcal{X} \subset [0,1]$ be a discrete set of cardinality $n+1$. Then, $P_X \mapsto I(P_X; P_{Y|X})$ is* strictly *concave over $\mathcal{P}_{\mathcal{X}}$.*

Proof. Let $P_X, Q_X \in \mathcal{P}_{\mathcal{X}}$, and let $P_X^{\epsilon} = (1-\epsilon)P_X + \epsilon Q_X$ for $\epsilon \in (0,1)$, which is also in $\mathcal{P}_{\mathcal{X}}$. Moreover, let $P_X \to P_{Y|X} \to P_Y$, $Q_X \to P_{Y|X} \to Q_Y$ and $P_X^{\epsilon} \to P_{Y|X} \to P_Y^{\epsilon}$. Then, first note that

$$\begin{aligned} & I(P_X^{\epsilon}; P_{Y|X}) \\ & - (1-\epsilon) I(P_X; P_{Y|X}) - \epsilon I(Q_X; P_{Y|X}) && (3.28) \\ & = D(P_{Y|X} \| P_Y^{\epsilon} | P_X^{\epsilon}) \\ & - (1-\epsilon) D(P_{Y|X} \| P_Y | P_X) - \epsilon D(P_{Y|X} \| Q_Y | Q_X) && (3.29) \\ & = (1-\epsilon) D(P_Y \| P_Y^{\epsilon}) + \epsilon D(Q_Y \| P_Y^{\epsilon}). && (3.30) \end{aligned}$$

We now show that every $P_X \in \mathcal{P}_{\mathcal{X}}$ induces a distinct output distribution (i.e., $P_X \to P_{Y|X} \to P_Y$ is an injective mapping), which implies that eq.(3.30) is strictly positive and, therefore, the mutual information is strictly concave. Define the following:

$$\mathbf{p}_X = [P_X(x_1), \ldots, P_X(x_{n+1})], \quad x_k \in \mathcal{X}, \tag{3.31}$$

$$\mathbf{p}_Y = [P_Y(0), \ldots, P_Y(n)]. \tag{3.32}$$

Then, the mapping $P_X \to P_{Y|X} \to P_Y$ can be written as the following system of linear equations:

$$\mathsf{A}\mathbf{p}_X = \mathbf{p}_Y \tag{3.33}$$

where the matrix $\mathsf{A} \in \mathbb{R}^{n+1 \times n+1}$ is such that

$$[\mathsf{A}]_{i,k} = P_{Y|X}(i-1|x_k),\ i \in [n+1],\ x_k \in \mathcal{X}. \tag{3.34}$$

From Proposition 5, we know that A is full rank for any $\mathcal{X}$ of cardinality $n+1$. Therefore, from standard linear algebra result, it follows that the mapping in eq.(3.33) is injective (i.e., every $\mathbf{p}_X$ induces a distinct $\mathbf{p}_Y$). Therefore, we conclude that eq.(3.30) is positive and mutual information is strictly concave. □

Note that since by Proposition 4, $\mathcal{A}_n$ has cardinality of at most $n+1$, from Theorem 3 we have the following corollary.

Corollary 1. $P_X \mapsto I(P_X; P_{Y|X})$ *is strictly concave over* $\mathcal{P}_{\mathcal{A}_n}$. *Consequently,* P_{X^*} *is unique.*

Table 3.1 Capacity and Capacity-Achieving Distributions

n	$C(n)$	$\mathcal{X} \equiv \mathsf{supp}(P_{X^\star})$	$\{P_{X^\star}(x),\, x \in \mathcal{X}\}$	$\{P_{Y^\star}(y),\, y \in \{0\} \cup [n]\}$
1	$\log(2)$	$\{0,1\}$	$\left\{\frac{1}{2}, \frac{1}{2}\right\}$	$\left\{\frac{1}{2}, \frac{1}{2}\right\}$
2	$\log\left(\frac{17}{8}\right)$	$\left\{0, \frac{1}{2}, 1\right\}$	$\left\{\frac{15}{34}, \frac{2}{17}, \frac{15}{34}\right\}$	$\left\{\frac{8}{17}, \frac{1}{17}, \frac{8}{17}\right\}$
3	$\log\left(\frac{19}{8}\right)$	$\left\{0, \frac{1}{2}, 1\right\}$	$\left\{\frac{15}{38}, \frac{4}{19}, \frac{15}{38}\right\}$	$\left\{\frac{8}{19}, \frac{3}{38}, \frac{3}{38}, \frac{8}{19}\right\}$

CHAPTER 4

RESULTS (PART I)

4.1 Lower Bound via Poincaré Inequality

Theorem 2 shows that our bound in Theorem 1 is tight in the high-noise regime for a large family of prior distributions on $\mathbf{X}$. However, we suspect that such a tightness result holds more generally. To support this, we here provide numerical evaluations for three 'toy', yet practically relevant, scenarios.

1. *Gaussian Input.* We assume that $\mathbf{X} \sim \mathcal{N}(\mathbf{0}_k, \mathbf{\Sigma}_\mathbf{X})$, i.e., $\mathbf{X}$ is a Gaussian random vector. For this scenario, the MMSE is obtained as [30],

$$\mathrm{mmse}(\mathbf{X}|\mathbf{Y}) = \mathsf{Tr}\left[\mathbf{\Sigma}_\mathbf{X}\left(\mathsf{I}_k + \frac{1}{\sigma_N^2}\mathbf{\Sigma}_\mathbf{X}\right)^{-1}\right], \tag{4.1}$$

and the Cramér-Rao lower bound evaluates to [2]

$$\mathrm{mmse}(\mathbf{X}|\mathbf{Y}) \geq \frac{k^2}{\frac{k}{\sigma_N^2} + \mathsf{Tr}[\mathbf{\Sigma}_\mathbf{X}^{-1}]}. \tag{4.2}$$

In Figure 4.1, we plot the MMSE in eq. (4.1) (solid line), the Cramér-Rao bound in eq. (4.2) (dotted line), and our bound on the MMSE in eq. (2.4) (dashed line) versus different values of σ_N^2 for $k = 6$ and a randomly generated $\mathbf{\Sigma}_\mathbf{X}$.

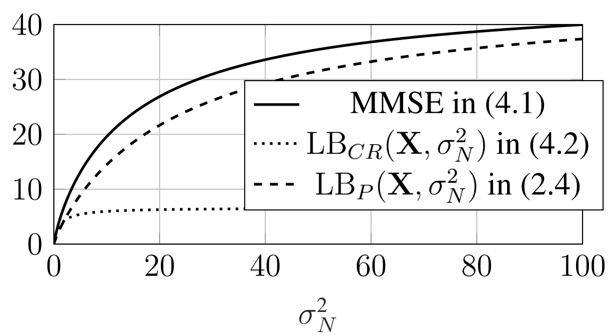

Figure 4.1 $\mathbf{X} \sim \mathcal{N}(\mathbf{0}_6, \mathbf{\Sigma}_\mathbf{X})$ with a randomly generated $\mathbf{\Sigma}_\mathbf{X}$.

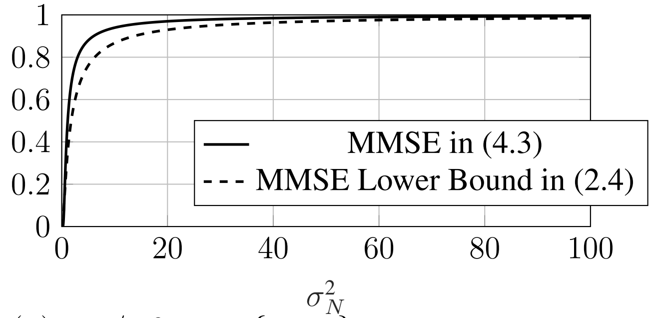

Figure 4.2 $X \sim P_X(x) = 1/2$ for $x \in \{-1, 1\}$.

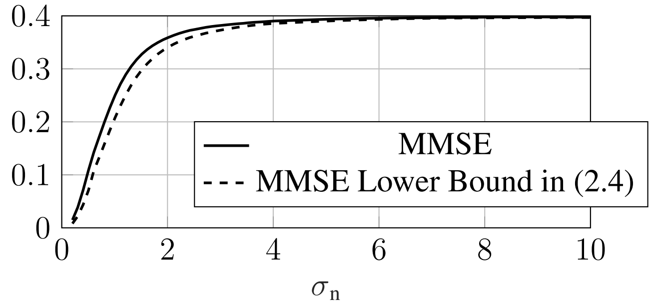

Figure 4.3 $X \sim P_X = (1-\alpha)\delta_0 + \alpha\mathcal{N}(0,1),\ \alpha = 0.4$.

2. *BPSK Input.* We let $k = 1$ and assume that $X \in \{-1, 1\}$ with equal probability, i.e., X is a Binary Phase Shift Keying (BPSK) signal with $P_X(1) = P_X(-1) = 1/2$. For this scenario, the MMSE is obtained as [31],

$$\mathrm{mmse}(X|Y) = 1 - \int_{-\infty}^{\infty} \frac{\mathrm{e}^{-\frac{y^2}{2}}}{\sqrt{2\pi}} \tanh\left(\frac{1}{\sigma_N^2} - \frac{y}{\sigma_N}\right) \mathrm{d}y. \tag{4.3}$$

   In Figure 4.2, we plot the MMSE in eq. (4.3) (solid line) and our lower bound in eq. (2.4) (dashed line) versus σ_N^2.

3. *Sparse Input.* We let $k = 1$ and assume that $X \sim P_X = (1-\alpha)\delta_0 + \alpha\mathcal{N}(0,1)$, where $\alpha \in [0,1]$ and δ_0 is the point measure at 0. Such input distributions are used to model sparsity and have been studied in [32–35]. To the best of our knowledge, a closed-form expression for the MMSE is not known. In Figure 4.3, we plot the MMSE (solid line) and our bound in eq. (2.4) (dashed line) versus σ_N^2 for $\alpha = 0.4$. The two curves were obtained via a Monte Carlo simulation with $5 \cdot 10^5$ iterations.

From Figure 4.1, Figure 4.2, and Figure 4.3, we observe that our lower bound in eq. (2.4) well approximates the MMSE even in the finite noise regime. Moreover, for the scenario of a Gaussian input in Figure 4.1, our lower bound in eq. (2.4) remarkably outperforms the well-known Cramér-Rao bound. Finally, we point out that for the scenarios of a BPSK input (Figure 4.2) and a sparse input (Figure 4.3), i.e., where $\mathbf{X}$ has a discrete or a mixed distribution, commonly used lower bounds (e.g., Cramér-Rao) do not hold, whereas ours does. These examples suggest that our lower bound in eq. (2.4) might indeed be tight (or offer a performance guarantee) even in the finite noise regime. Hence, it would also be interesting to characterize the behavior of our lower bound in the low-noise regime. However, in the low-noise regime, the MMSE has an intricate behavior, for example, it depends on whether the distribution of $\mathbf{X}$ is discrete or continuous [36].

4.2 New Representation of the MMSE and its Applicability

In this section, we prove Theorem 1, which provides a new representation of the MMSE. Towards this end, we leverage the following proposition, which provides an expression for the gradient of the information density.

Proposition 6. *For $\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}$, we have that*

$$\nabla_{\mathbf{y}} \iota_{P_{\mathbf{XY}}}(\mathbf{x}; \mathbf{y}) = \mathbf{J}_{\mathbf{y}} \mathbf{T}(\mathbf{y})(\mathbf{x} - \mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{y}]).$$

Fix some $\mathbf{x} \in \mathcal{X}$. Then,

$$\begin{aligned}
&\nabla_{\mathbf{y}} \iota_{P_{\mathbf{XY}}}(\mathbf{x}; \mathbf{y}) \\
&\overset{(a)}{=} \nabla_{\mathbf{y}} \log h(\mathbf{y}) + \nabla_{\mathbf{y}} \langle \mathbf{x}, \mathbf{T}(\mathbf{y}) \rangle - \nabla_{\mathbf{y}} \log f_{\mathbf{Y}}(\mathbf{y}) \\
&\overset{(b)}{=} \mathbf{J}_{\mathbf{y}} \mathbf{T}(\mathbf{y}) \mathbf{x} - \nabla_{\mathbf{y}} \log \frac{f_{\mathbf{Y}}(\mathbf{y})}{h(\mathbf{y})} \\
&\overset{(c)}{=} \mathbf{J}_{\mathbf{y}} \mathbf{T}(\mathbf{y}) \mathbf{x} - \mathbf{J}_{\mathbf{y}} \mathbf{T}(\mathbf{y}) \mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{y}] \\
&= \mathbf{J}_{\mathbf{y}} \mathbf{T}(\mathbf{y})(\mathbf{x} - \mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{y}]),
\end{aligned}$$

where the labeled equalities follow from: (a) the definition of the exponential family in eq. (1.1); (b) using the fact that $\nabla_{\mathbf{y}}\langle \mathbf{x}, \mathbf{T}(\mathbf{y})\rangle = \mathbf{J}_{\mathbf{y}}\mathbf{T}(\mathbf{y})\mathbf{x}$; and (c) using the TRE identity. This concludes the proof of Proposition 6.

Proof of Theorem 1. Since $\mathbf{J}_{\mathbf{Y}}\mathbf{T}(\mathbf{Y})$ has full rank a.s. $\mathbf{Y}$, then the pseudo inverse $(\mathbf{J}_{\mathbf{Y}}\mathbf{T}(\mathbf{Y}))^{+}$ exists a.s. $\mathbf{Y}$. Using Proposition 6, we have that a.s. $\mathbf{Y}$,

$$(\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]) = (\mathbf{J}_{\mathbf{Y}}\mathbf{T}(\mathbf{Y}))^{+}\nabla_{\mathbf{Y}}\iota_{P_{\mathbf{XY}}}(\mathbf{X};\mathbf{Y}). \tag{4.4}$$

Now, taking the norm squared and the expectation of both sides of eq. (4.4), and recalling that $\mathrm{mmse}(\mathbf{X}|\mathbf{Y}) = \mathbb{E}\left[||\mathbf{X}-\mathbb{E}[\mathbf{X}|\mathbf{Y}]||^2\right]$ we arrive at the desired result. This concludes the proof of Theorem 1.

4.2.1 Example: Univariate Normal with Unknown Variance with Gamma Prior

To show an application of the new representation of the MMSE in Theorem 1, we here consider the following model,

$$Y = \frac{Z}{\sqrt{2X}}, \tag{4.5}$$

where Z is the standard normal random variable, i.e., $Z \sim \mathcal{N}(0,1)$ and X is the unknown variance drawn from the gamma distribution with $\alpha > 0$ shape, and $\beta > 0$ rate, i.e.,

$$f_X(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1}\mathrm{e}^{-\beta x}, \tag{4.6}$$

where Γ is the gamma function.

Using the channel model in eq. (4.5), we therefore obtain,

$$f_{Y|X}(y|x) = \sqrt{\frac{x}{\pi}}\mathrm{e}^{-xy^2}. \tag{4.7}$$

The conditional pdf in eq. (4.7) can be mapped to the exponential family in eq. (1.1) through the following mapping,

$$h(y) = \sqrt{\frac{1}{\pi}},\ \phi(x) = -\log\left(\sqrt{x}\right),\ T(y) = -y^2. \tag{4.8}$$

We now evaluate the MMSE expression in Theorem 1 for our model in eq. (4.5) with the mapping in eq. (4.8). We note that

$$\left(\frac{\mathrm{d}}{\mathrm{d}y}T(y)\right)^{-1} = -\frac{1}{2y}. \tag{4.9}$$

We now focus on deriving $\frac{\mathrm{d}}{\mathrm{d}y}\iota_{P_{XY}}(x;y)$, where

$$\iota_{P_{XY}}(x;y) = \log\left(f_{Y|X}(y|x)\right) - \log\left(f_Y(y)\right),$$

where $f_{Y|X}$ is defined in eq. (4.7). Hence, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}y}\log\left(f_{Y|X}(y|x)\right) = -\frac{\mathrm{d}}{\mathrm{d}y}\left(xy^2\right) = -2xy. \tag{4.10}$$

To complete the evaluation of the MMSE in Theorem 1, we need $\frac{\mathrm{d}}{\mathrm{d}y}\log\left(f_Y(y)\right)$, where (see [29] for the details)

$$f_Y(y) = \sqrt{\frac{1}{\pi}}\beta^{\alpha}\frac{\Gamma\left(\alpha+\frac{1}{2}\right)}{\Gamma(\alpha)}\frac{1}{(y^2+\beta)^{\alpha+\frac{1}{2}}}.$$

Thus, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}y}\log\left(f_Y(y)\right) = \frac{1}{f_Y(y)}\frac{\mathrm{d}}{\mathrm{d}y}f_Y(y) = -\frac{y(2\alpha+1)}{y^2+\beta}. \tag{4.11}$$

Finally, by substituting eq. (4.9), eq. (4.10) and eq. (4.11) inside the MMSE expression in Theorem 1, we obtain (see [29] for the details)

$$\mathrm{mmse}(X|Y) = \frac{\alpha(\alpha+1)}{\beta^2\left(\alpha+\frac{3}{2}\right)}.$$

We note that, in order to compute the MMSE above, we did not need to compute $\mathbb{E}[X|Y]$ (which is needed by the classical representation of the MMSE), but only the marginal pdf f_Y, which can be done by simple computations. Moreover, we also highlight that the MMSE above is in closed-form, and this expression highlights that the MMSE only depends on the parameters of the gamma distribution.

CHAPTER 5

RESULTS (PART II)

5.1 Capacity and Bounds on the Capacity

In this section, we provide exact values of the capacity for $n \leq 3$. For the remaining regime we provide upper and lower bounds on capacity.

5.1.1 Exact Capacity for $n \leq 3$

The exact capacity can be computed by first making a guess of the capacity-achieving distribution according to the properties outlined in Section 3.2. Then, this guess can be checked against the sufficient and necessary KKT conditions in Lemma 3.1.2. These, somewhat tedious, computations are performed in Appendix B.1 and Table 3.1 displays the results.

5.1.2 Bounds on the Capacity

We now provide bounds on the capacity. Our upper bound relies on the dual representation of the capacity as:

$$C(n) = \inf_{q} \max_{x \in [0,1]} P_{Y|X}(\cdot|x)q, \tag{5.1}$$

which, by properly choosing an auxiliary output distribution q, often leads to order-tight bounds. The reader is referred to [62–64] for applications to other channels. It will also be convenient to work with continuous output, and we will use the following channel output: $\tilde{Y} = Y + U$, where $U \sim \mathcal{U}(0,1)$. Note that because the distance between original Y points is one, such additive noise can be completely filtered out, and we have $I(X;Y) = I(X;Y+U)$ for all X. This trick has been used before in the context of the Poisson channel in [62].

The lower bound on the capacity will follow from choosing a convenient input distribution. The exact computation, however, will not be possible, and some further bounds on the entropy of the binomial distribution will be needed. Therefore, in Appendix C.1, we also provide a new upper bound on the entropy of a binomial distribution. Bounds on the entropy of a binomial distribution have been considered before in [65, 66].

Theorem 4. *For $n \geq 1$, the channel capacity is bounded below by*

$$C(n) \geq \Bigg\{ \log(2), \log(\pi n) - \frac{1}{2}\log\left(2\pi\left(\frac{n}{8}+\frac{1}{12}\right)\right) + \frac{1}{\sqrt{\pi\left(n+\frac{1}{4}\right)}}\log\left(\frac{1}{16n^2}\right) - \log(4) - 1 \Bigg\} \tag{5.2}$$

and bounded above by

$$C(n) \leq \min\Bigg\{ \log\left(3 + \left\lfloor\frac{(n-1)}{2}\right\rfloor\right), \log(\pi(n+1)) - \frac{1}{2}\log(n) + \frac{3}{2} + \frac{1}{2^{n+1}}\log(n) + \frac{1}{2}\log\left(\frac{3}{2}\left(1+\frac{1}{n}\right)\right)\Bigg\}. \tag{5.3}$$

Proof. A first lower bound follows from observing that $P_{Y|X}(0|0) = 1$ and $P_{Y|X}(n|1) = 1$. Hence the input distribution $P_X(0) = P_X(1) = \frac{1}{2}$ gives a mutual information of $I(X;Y) = \log(2)$ nats for all $n \geq 1$.

For an alternative capacity lower bound, pick an input pdf as

$$f_X(x) = \frac{1}{\pi\sqrt{x(1-x)}}, \qquad x \in (0,1), \tag{5.4}$$

which is a Beta distribution with shape parameters $\alpha = \beta = \frac{1}{2}$. Then, the capacity can be lower-bounded as follows:

$$C(n) = \max_{P_X} I(X;Y) \tag{5.5}$$

$$\geq I(X;Y) \tag{5.6}$$

$$
\begin{aligned}
&= H(Y) - H(Y|X) && (5.7)\\
&\geq H(Y) - \frac{1}{2}\mathbb{E}\left[\log\left(2\pi\left(nX(1-X)+\frac{1}{12}\right)\right)\right] && (5.8)\\
&\geq H(Y) - \frac{1}{2}\log\left(2\pi\left(n\mathbb{E}\left[X(1-X)\right]+\frac{1}{12}\right)\right) && (5.9)\\
&= H(Y) - \frac{1}{2}\log\left(2\pi\left(\frac{n}{8}+\frac{1}{12}\right)\right) && (5.10)
\end{aligned}
$$

where in eq. (5.8) we have used the upper bound on the entropy of a binomial distribution, which is given in Appendix C.1; in eq. (5.9) we have applied Jensen's inequality; and in the last step we have used that $\mathbb{E}\left[X(1-X)\right] = \frac{1}{8}$ from the Beta distribution in eq. (5.4).

As for the output entropy, write

$$
\begin{aligned}
H(Y) &= -\mathbb{E}\left[\log P_Y(Y)\right] && (5.11)\\
&= -\mathbb{E}\left[\log\binom{n}{Y}\right] - \mathbb{E}\left[\log \mathbb{E}_X[X^Y(1-X)^{n-Y}]\right]. && (5.12)
\end{aligned}
$$

Now note that

$$
\begin{aligned}
\mathbb{E}\left[X^y(1-X)^{n-y}\right] &= \int_0^1 \frac{1}{\pi\sqrt{x(1-x)}} x^y(1-x)^{n-y} x && (5.13)\\
&= \frac{1}{\pi n}\frac{\Gamma\left(y+\frac{1}{2}\right)\Gamma\left(n-y+\frac{1}{2}\right)}{\Gamma\left(n\right)} && (5.14)
\end{aligned}
$$

which, by expanding the binomial coefficient in terms of gamma functions, gives

$$
\begin{aligned}
H(Y) &= \log(\pi n) - \mathbb{E}\left[\log\left(n\frac{\Gamma\left(Y+\frac{1}{2}\right)}{\Gamma\left(Y+1\right)}\frac{\Gamma\left(n-Y+\frac{1}{2}\right)}{\Gamma\left(n-Y+1\right)}\right)\right] && (5.15)\\
&\geq \log(\pi n) - \mathbb{E}\left[\log\left(\frac{n}{\left(Y+\frac{1}{4}\right)^{\frac{1}{2}}\left(n-Y+\frac{1}{4}\right)^{\frac{1}{2}}}\right)\right] && (5.16)\\
&= \log(\pi n) + \frac{1}{2}\mathbb{E}\left[\log\left(\left(\frac{Y}{n}+\frac{1}{4n}\right)\left(1-\frac{Y}{n}+\frac{1}{4n}\right)\right)\right] && (5.17)\\
&= \log(\pi n) + \mathbb{E}\left[\log\left(\frac{Y}{n}+\frac{1}{4n}\right)\right] && (5.18)\\
&\geq \log(\pi n) + \mathbb{E}\left[\mathbb{1}(Y=0)\log\left(\frac{1}{4n}\right)\right]
\end{aligned}
$$

$$+ \mathbb{E}\left[\mathbb{1}(0 < Y \leq n) \log\left(\frac{Y}{n}\right)\right] \tag{5.19}$$

$$\geq \log(\pi n) + \mathbb{E}\left[(1-X)^n\right] \log\left(\frac{1}{4n}\right)$$

$$+ \mathbb{E}\left[(1-(1-X)^n) \log(X)\right] - 1 \tag{5.20}$$

$$\geq \log(\pi n) + \frac{\Gamma\left(n+\frac{1}{2}\right)}{\sqrt{\pi}\Gamma\left(n+1\right)} \log\left(\frac{1}{16n^2}\right) + \mathbb{E}\left[\log(X)\right] - 1 \tag{5.21}$$

$$\geq \log(\pi n) + \frac{1}{\sqrt{\pi\left(n+\frac{1}{4}\right)}} \log\left(\frac{1}{16n^2}\right) - \log(4) - 1 \tag{5.22}$$

where in eq. (5.16) and in eq. (5.22) we used Kershav's inequality [67]

$$\frac{\Gamma(x+s)}{\Gamma(x+1)} \leq \frac{1}{\left(x+\frac{s}{2}\right)^{1-s}} \tag{5.23}$$

for $x > 0$ and $s \in (0,1)$; in eq. (5.18) we have used the symmetry of the output pmf $Y \overset{d}{=} (n-Y)$; in eq. (5.20) we have used Lemma C.1; in eq. (5.21) we have have used $\mathbb{E}\left[(1-X)^n\right] = \frac{\Gamma\left(n+\frac{1}{2}\right)}{\sqrt{\pi}\Gamma(n+1)}$ and the fact that $\log(X) \leq 0$; finally, in the last step we have used $\mathbb{E}\left[\log(X)\right] = -\log(4)$.

To sum up, the capacity lower bound is given by eq. (5.2).

A first upper bound on $C(n)$ follows by noting that

$$C(n) \leq H(X^\star) \leq \log\left(3 + \left\lfloor\frac{(n-1)}{2}\right\rfloor\right), \tag{5.24}$$

where the last upper bound is due to Theorem 2.

For an alternative capacity upper bound, choose the auxiliary output pdf

$$q(t) = \frac{1}{\pi(n+1)}\left(\frac{t}{n+1}\left(1-\frac{t}{n+1}\right)\right)^{-\frac{1}{2}}, \ t \in (0, n+1), \tag{5.25}$$

and, by introducing $U \sim \mathcal{U}[0,1]$ independent of X and Y, write

$$C(n) = \max_{P_X} I(X;Y) \tag{5.26}$$

$$= \max_{P_X} I(X;Y+U) \tag{5.27}$$

$$
\begin{aligned}
&\leq \max_{x\in[0,1]} P_{Y+U|X}(\cdot|x)q && (5.28)\\
&= \max_{x\in[0,1]} -H(Y|X=x) - \mathbb{E}\left[\log q(Y+U) \mid X=x\right] && (5.29)\\
&\leq \max_{x\in[0,1]} \Bigg\{ \log(\pi(n+1)) - (1-(1-x)^n - x^n)\frac{1}{2}\log(2\pi n)\\
&\quad -\frac{1}{2}(1-(1-x)^n)\log(x) - \frac{1}{2}(1-x^n)\log(1-x) + 1\\
&\quad +\frac{1}{2}\mathbb{E}\left[\log\left(\frac{Y+U}{n+1}\left(1-\frac{Y+U}{n+1}\right)\right) \,\middle|\, X=x\right]\Bigg\} && (5.30)\\
&\leq \max_{x\in[0,1]} \Bigg\{ \log(\pi(n+1)) - (1-(1-x)^n - x^n)\frac{1}{2}\log(2\pi n)\\
&\quad -\frac{1}{2}(1-(1-x)^n)\log(x) - \frac{1}{2}(1-x^n)\log(1-x) + 1\\
&\quad +\frac{1}{2}\log\left(\frac{nx+\frac{1}{2}}{n+1}\left(1-\frac{nx+\frac{1}{2}}{n+1}\right)\right)\Bigg\} && (5.31)\\
&= \log(\pi(n+1)) - \frac{1}{2}\log(2\pi n) + \max_{x\in[0,1]}\{g_n(x)\} && (5.32)
\end{aligned}
$$

where in eq. (5.28) we have used the dual formulation of capacity; in eq. (5.29) we used $h(Y+U|X=x) = H(Y|X=x)$; in eq. (5.30) we have used the lower bound on the entropy of a binomial distribution.C.1; in eq. (5.31) we have used Jensen's inequality; and in eq. (5.32) we have introduced the function

$$
\begin{aligned}
g_n(x) = &\frac{((1-x)^n + x^n)}{2}\log(2\pi n) - \frac{(1-(1-x)^n)}{2}\log(x)\\
&-\frac{(1-x^n)}{2}\log(1-x) + \frac{1}{2}\log\left(\frac{nx+\frac{1}{2}}{n+1}\left(1-\frac{nx+\frac{1}{2}}{n+1}\right)\right) \qquad (5.33)
\end{aligned}
$$

for $x \in [0,1]$. □

APPENDIX A

PROOF OF LEMMA 1

Our objective is to show that the following two exchanges of the limit and expectation, and the limit and variance are permissible,

$$\begin{aligned}&\lim_{\sigma_N\to\infty}\mathbb{E}_{\mathbf{X}}\left[\mathrm{Var}_{\mathbf{Z}}\left(\sigma_N\log\left(\tilde{g}(\mathbf{X}+\sigma_N\mathbf{Z})\right)\right)\right]\\&=\mathbb{E}_{\mathbf{X}}\left[\mathrm{Var}_{\mathbf{Z}}\left(\lim_{\sigma_N\to\infty}\sigma_N\log\left(\tilde{g}(\mathbf{X}+\sigma_N\mathbf{Z})\right)\right)\right].\end{aligned}$$

Equivalently, by using the definition of the variance, we have to show that

$$\lim_{\sigma_N\to\infty}\mathbb{E}\left[f_{\sigma_N}(\mathbf{X},\mathbf{Z})\right]=\mathbb{E}\left[\lim_{\sigma_N\to\infty}f_{\sigma_N}(\mathbf{X},\mathbf{Z})\right], \tag{A.1a}$$

where

$$f_{\sigma_N}(\mathbf{x},\mathbf{z})=\left(\sigma_N\log\tilde{g}(\mathbf{x}+\sigma_N\mathbf{z})-\mathbb{E}[\sigma_N\log\tilde{g}(\mathbf{X}+\sigma_N\mathbf{Z})]\right)^2. \tag{A.1b}$$

Towards this end, we start by noting that

$$\begin{aligned}&|\sigma_N\log\tilde{g}(\mathbf{x}+\sigma_N\mathbf{z})|\\&\overset{(a)}{=}\sigma_N\max\left\{\log\mathbb{E}\left[\mathrm{e}^{-\frac{\|\mathbf{x}-\mathbf{X}\|^2+2(\mathbf{x}-\mathbf{X})^{\mathsf{T}}\sigma_N\mathbf{z}}{2\sigma_N^2}}\right],\right.\\&\qquad\left.-\log\mathbb{E}\left[\mathrm{e}^{-\frac{\|\mathbf{x}-\mathbf{X}\|^2+2(\mathbf{x}-\mathbf{X})^{\mathsf{T}}\sigma_N\mathbf{z}}{2\sigma_N^2}}\right]\right\}\\&\overset{(b)}{\le}\sigma_N\max\left\{\log\mathbb{E}\left[\mathrm{e}^{-\frac{2(\mathbf{x}-\mathbf{X})^{\mathsf{T}}\sigma_N\mathbf{z}}{2\sigma_N^2}}\right],\right.\\&\qquad\left.-\log\mathrm{e}^{-\frac{\mathbb{E}\left[\|\mathbf{x}-\mathbf{X}\|^2+2(\mathbf{x}-\mathbf{X})^{\mathsf{T}}\sigma_N\mathbf{z}\right]}{2\sigma_N^2}}\right\}\\&\overset{(c)}{=}\sigma_N\max\left\{\log\mathbb{E}\left[\mathrm{e}^{-\frac{(\mathbf{x}-\mathbf{X})^{\mathsf{T}}\mathbf{z}}{\sigma_N}}\right],-\log\mathrm{e}^{-\frac{\mathbb{E}\left[\|\mathbf{x}-\mathbf{X}\|^2+2\mathbf{x}^{\mathsf{T}}\sigma_N\mathbf{z}\right]}{2\sigma_N^2}}\right\}\end{aligned}$$

$$= \sigma_N \max\left\{-\frac{\mathbf{x}^\mathsf{T}\mathbf{z}}{\sigma_N}+\log\mathbb{E}\left[\mathrm{e}^{\frac{\mathbf{X}^\mathsf{T}\mathbf{z}}{\sigma_N}}\right], \frac{\mathbf{x}^\mathsf{T}\mathbf{z}}{\sigma_N}+\frac{\mathbb{E}\left[\|\mathbf{x}-\mathbf{X}\|^2\right]}{2\sigma_N^2}\right\}$$

$$\overset{(d)}{\le} \|\mathbf{x}\|\|\mathbf{z}\| + \sigma_N \max\left\{\log\mathbb{E}\left[\mathrm{e}^{\frac{\mathbf{X}^\mathsf{T}\mathbf{z}}{\sigma_N}}\right], \frac{\mathbb{E}\left[\|\mathbf{x}-\mathbf{X}\|^2\right]}{2\sigma_N^2}\right\}$$

$$\overset{(e)}{\le} \|\mathbf{x}\|\|\mathbf{z}\| + \sigma_N \max\left\{\frac{B\|\mathbf{z}\|^2}{\sigma_N^2}, \frac{2\|\mathbf{x}\|^2+2\mathbb{E}\left[\|\mathbf{X}\|^2\right]}{2\sigma_N^2}\right\}$$

$$\overset{(f)}{\le} \|\mathbf{x}\|\|\mathbf{z}\| + \max\left\{B\|\mathbf{z}\|^2, \|\mathbf{x}\|^2 + \mathbb{E}\left[\|\mathbf{X}\|^2\right]\right\}, \tag{A.2}$$

where the labeled inequalities/equalities follow from: (a) using the property that $|\log(x)| = \max\{\log(x), -\log(x)\}$; (b) using the bound $\mathrm{e}^{-\frac{\|\mathbf{x}-\mathbf{X}\|^2}{2\sigma_N^2}} \le 1$ on the first logarithm, and using Jensen's inequality on the second logarithm; (c) the assumption that $\mathbb{E}[\mathbf{X}] = \mathbf{0}_k$; (d) using Cauchy-Schwarz inequality; (e) the assumption that $\mathbf{X}$ is sub-Gaussian for some constant $B > 0$ and using the bound $\|\mathbf{x}-\mathbf{X}\|^2 \le 2\|\mathbf{x}\|^2 + 2\|\mathbf{X}\|^2$; and (f) the assumption that $\sigma_N > 1$.

Now, we can use (2.4) to bound $f_{\sigma_N}(\mathbf{x}, \mathbf{z})$ in (A.1). We obtain

$$f_{\sigma_N}(\mathbf{x}, \mathbf{z})$$

$$= (\sigma_N \log\tilde{g}(\mathbf{x}+\sigma_N\mathbf{z}) - \mathbb{E}[\sigma_N \log\tilde{g}(\mathbf{X}+\sigma_N\mathbf{Z})])^2$$

$$\overset{(a)}{\le} 2\left(\sigma_N \log\tilde{g}(\mathbf{x}+\sigma_N\mathbf{z})\right)^2 + 2\left(\mathbb{E}[\sigma_N \log\tilde{g}(\mathbf{X}+\sigma_N\mathbf{Z})]\right)^2$$

$$\overset{(b)}{\le} 2\left(\sigma_N \log\tilde{g}(\mathbf{x}+\sigma_N\mathbf{z})\right)^2 + 2\mathbb{E}[|\sigma_N \log\tilde{g}(\mathbf{X}+\sigma_N\mathbf{Z})|^2]$$

$$\overset{(c)}{\le} 2\left(\|\mathbf{x}\|\|\mathbf{z}\| + \max\left\{B\|\mathbf{z}\|^2, \|\mathbf{x}\|^2 + \mathbb{E}\left[\|\mathbf{X}\|^2\right]\right\}\right)^2$$

$$+2\mathbb{E}\left[\left(\|\mathbf{X}\|\|\mathbf{Z}\|+\max\left\{B\|\mathbf{Z}\|^2, \|\mathbf{X}\|^2+\mathbb{E}\left[\|\mathbf{X}\|^2\right]\right\}\right)^2\right], \tag{A.3}$$

where the labeled inequalities follow from: (a) the fact that $(a-b)^2 \le 2a^2+2b^2$; (b) using Jensen's inequality; and (c) using the bound in (2.4).

Now, note that under the assumption that $\mathbf{X}$ is sub-Gaussian all moments are finite and hence, the quantity in (A.3) is integrable. Consequently, under the assumption that $\mathbf{X}$ is sub-Gaussian, the random variable $f_{\sigma_N}(\mathbf{X}, \mathbf{Z})$ is bounded by an integrable random variable, and we can apply the dominate convergence theorem to exchange the limit and the expectation and arrive at

$$
\begin{aligned}
&\lim_{\sigma_N \to \infty} \mathbb{E}_{\mathbf{X}} \left[\mathrm{Var}_{\mathbf{Z}} \left(\sigma_N \log \left(\tilde{g}(\mathbf{X} + \sigma_N \mathbf{Z}) \right) \right) \right] \\
&= \mathbb{E}_{\mathbf{X}} \left[\mathrm{Var}_{\mathbf{Z}} \left(\lim_{\sigma_N \to \infty} \sigma_N \log \left(\tilde{g}(\mathbf{X} + \sigma_N \mathbf{Z}) \right) \right) \right] \qquad \square
\end{aligned}
$$

which concludes the proof of Lemma 1.

APPENDIX B

CAPACITY COMPUTATION FOR $n \leq 3$

B.1 Capacity Computation for $n \leq 3$

B.1.1 The Case of $n = 3$

For $n = 3$, we infer that

$$\mathsf{supp}(P_{X^\star}) \subseteq \left\{0, \frac{1}{2}, 1\right\}.$$

Now let $p = P_{X^\star}(\frac{1}{2})$. Corollary 1 and direction computations imply that

$$P_{Y^\star}(0) = P_{Y^\star}(3) =^{-C(3)}, \tag{B.1}$$

$$P_{Y^\star}(1) = P_{Y^\star}(2) = \frac{3}{8}p. \tag{B.2}$$

Now using above and the fact that $\sum_{y=0}^{3} P_{Y^*}(y) = 1$, we have that

$$p = \frac{4}{3}\left(1 - 2^{-C(3)}\right). \tag{B.3}$$

Next, it can be shown that

$$i\left(\frac{1}{2}; P_{Y^\star}\right) = \frac{1}{4}\log\left(\frac{^{C(3)}}{8p^3}\right). \tag{B.4}$$

From the KKT equality condition in (3.4), we have that

$$C(3) = i\left(\frac{1}{2}; P_{Y^\star}\right) = \frac{1}{4}\log\left(\frac{^{C(3)}}{8p^3}\right) \tag{B.5}$$

using the expression for p in (B.3) and simplifying, we arrive at

$$C(3) = \log\left(\frac{1}{\frac{8}{3}\left(1 - 2^{-C(3)}\right)}\right). \tag{B.6}$$

Solving for $C(3)$ we arrive at

$$C(3) = \log\left(\frac{19}{8}\right). \tag{B.7}$$

We also have that

$$P_{Y^\star}(0) = P_{Y^\star}(3) = \frac{8}{19}, \tag{B.8}$$

$$P_{Y^\star}(1) = P_{Y^\star}(2) = \frac{3}{38}, \tag{B.9}$$

$$P_{X^*}(0) = P_{X^*}(1) = \frac{15}{38}, \tag{B.10}$$

$$P_{X^*}\left(\frac{1}{2}\right) = \frac{4}{19}. \tag{B.11}$$

APPENDIX C

A UNIFORM BOUND ON $g_n(x)$

For $n \geq 1$, we have

$$\max_{x\in[0,1]} g_n(x) \leq \frac{1}{2}\log(2\pi) + \frac{1}{2} + \frac{1}{2^{n+1}}\log(n) + \frac{1}{2}\log\left(\frac{3}{2}\left(1+\frac{1}{n}\right)\right). \tag{C.1}$$

Proof. Since $g_n(x) = g_n(1-x)$, we can limit our analysis in the interval $x \in \left[0, \frac{1}{2}\right]$. By applying the substitution $x = \frac{\alpha}{n}$ for $\alpha \in \left[0, \frac{n}{2}\right]$, we get

$$2g_n\left(\frac{\alpha}{n}\right) \tag{C.2}$$

$$\begin{aligned}
&= \left(\left(1-\frac{\alpha}{n}\right)^n + \left(\frac{\alpha}{n}\right)^n\right)\log(2\pi n) \\
&- \left(1-\left(1-\frac{\alpha}{n}\right)^n\right)\log\left(\frac{\alpha}{n}\right) - \left(1-\left(\frac{\alpha}{n}\right)^n\right)\log\left(1-\frac{\alpha}{n}\right) \\
&\quad + \log\left(\frac{\alpha+\frac{1}{2}}{n+1}\left(1-\frac{\alpha+\frac{1}{2}}{n+1}\right)\right)
\end{aligned} \tag{C.3}$$

$$\begin{aligned}
&\leq \log\left(\frac{2\pi n}{n+1}\right) - \left(1-\left(1-\frac{\alpha}{n}\right)^n\right)\log(\alpha) \\
&\quad - \left(1-\left(\frac{\alpha}{n}\right)^n\right)\log(n-\alpha) \\
&\quad + \log\left(\alpha+\frac{1}{2}\right) + \log\left(n-\alpha+\frac{1}{2}\right)
\end{aligned} \tag{C.4}$$

$$\begin{aligned}
&\leq \log(2\pi) - \left(1-\left(1-\frac{\alpha}{n}\right)^n\right)\log(\alpha) \\
&\quad + \left(\frac{\alpha}{n}\right)^n\log(n-\alpha) + \log\left(\alpha+\frac{1}{2}\right) \\
&\quad + \log\left(1+\frac{1}{2(n-\alpha)}\right)
\end{aligned} \tag{C.5}$$

$$\begin{aligned}
&\leq \log(2\pi) - \left(1-\left(1-\frac{\alpha}{n}\right)^n\right)\log(\alpha) + \frac{1}{2^n}\log(n) \\
&\quad + \log\left(\alpha+\frac{1}{2}\right) + \log\left(1+\frac{1}{n}\right).
\end{aligned} \tag{C.6}$$

When $\alpha \in [0,1]$, the term $\log(\alpha)$ is negative, and by using $\left(1-\frac{\alpha}{n}\right)^n \geq 1-\alpha$ we can further upper-bound as follows:

$$2g_n\left(\frac{\alpha}{n}\right) \leq \log(2\pi) - \alpha\log(\alpha) + \frac{1}{2^n}\log(n) + \log\left(\alpha + \frac{1}{2}\right) + \log\left(1+\frac{1}{n}\right) \tag{C.7}$$

$$\leq \log(2\pi) + {}^{-1} + \frac{1}{2^n}\log(n) + \log\left(1+\frac{1}{2}\right) + \log\left(1+\frac{1}{n}\right), \tag{C.8}$$

which is bounded in n. When $\alpha \in \left[1, \frac{n}{2}\right]$, then $\log(\alpha)$ is positive and, by using $e^{-x} \geq \left(1-\frac{x}{n}\right)^n$ for all $n \geq 1$ and all $x \geq 0$, we have that

$$2g_n\left(\frac{\alpha}{n}\right) \leq \log(2\pi) + \left(\left(1-\frac{\alpha}{n}\right)^n - 1\right)\log(\alpha) + \frac{1}{2^n}\log(n) + \log\left(\alpha+\frac{1}{2}\right) + \log\left(1+\frac{1}{n}\right) \tag{C.9}$$

$$\leq \log(2\pi) + \left({}^{-\alpha} - 1\right)\log(\alpha) + \frac{1}{2^n}\log(n) + \log\left(\alpha+\frac{1}{2}\right) + \log\left(1+\frac{1}{n}\right) \tag{C.10}$$

$$= \log(2\pi) + {}^{-\alpha}\log(\alpha) + \frac{1}{2^n}\log(n) + \log\left(1+\frac{1}{2\alpha}\right) + \log\left(1+\frac{1}{n}\right) \tag{C.11}$$

$$\leq \log(2\pi) + {}^{-\alpha}\log(\alpha) + \frac{1}{2^n}\log(n) + \log\left(1+\frac{1}{2}\right) + \log\left(1+\frac{1}{n}\right) \tag{C.12}$$

$$\leq \log(2\pi) + 1 + \frac{1}{2^n}\log(n) + \log\left(1+\frac{1}{2}\right) + \log\left(1+\frac{1}{n}\right), \tag{C.13}$$

which is bounded in n. Since (C.13) is strictly larger than (C.8), we can conclude the result in (C.1). □

C.1 Bounds on the Entropy of a Binomial Random Variable

First of all we need the following result. Let $P_{Y|X}(\cdot|x)$ be a Binomial pmf with n trials and success probability x per trial. Then,

$$\mathbb{E}\left[\mathbb{1}(0 < Y \leq n) \log\left(\frac{Y}{n}\right) \,\middle|\, X = x\right]$$
$$\geq (1-(1-x)^n)\log(x) - 1. \tag{C.14}$$

Proof. Inspired by the approach of [62, Appendix B], we bound the expectation as follows:

$$\mathbb{E}\left[\mathbb{1}(0 < Y \leq n) \log\left(\frac{Y}{n}\right) \,\middle|\, X = x\right]$$
$$= \mathbb{E}\left[\mathbb{1}(0 < Y \leq n) \log(x) \mid X = x\right]$$
$$+ \mathbb{E}\left[\mathbb{1}(0 < Y \leq n) \log\left(\frac{Y}{nx}\right) \,\middle|\, X = x\right] \tag{C.15}$$
$$= (1-(1-x)^n)\log(x)$$
$$+ \mathbb{E}\left[\mathbb{1}(0 < Y \leq n) \log\left(\frac{Y}{nx}\right) \,\middle|\, X = x\right] \tag{C.16}$$
$$= (1-(1-x)^n)\log(x) + \sum_{y=1}^{n-1} P_{Y|X}(y|x) \log\left(\frac{y}{nx}\right) \tag{C.17}$$
$$\geq (1-(1-x)^n)\log(x) + \int_0^n P_{Y|X}(\lfloor y \rfloor|x) \log\left(\frac{y}{nx}\right) y \tag{C.18}$$
$$= (1-(1-x)^n)\log(x) + n\int_0^1 P_{Y|X}(\lfloor nt \rfloor|x) \log\left(\frac{t}{x}\right) t \tag{C.19}$$

where the inequality holds because $x \mapsto \log(x)$ is an increasing function and negative for $x \in (0,1)$.

Now introduce the continuous rv Z with pdf $f_Z(z) = nP_{Y|X}(\lfloor nz \rfloor|x)$ for $z \in [0,1]$. Then, the integral of (C.19) becomes:

$$n\int_0^1 P_{Y|X}(\lfloor nt \rfloor|x) \log\left(\frac{t}{x}\right) t$$
$$= \int_0^1 f_Z(t) \log\left(\frac{t}{x}\right) t \tag{C.20}$$
$$= \int_0^x f_Z(t) \log\left(\frac{t}{x}\right) t + \int_x^1 f_Z(t) \log\left(\frac{t}{x}\right) t. \tag{C.21}$$

Let us now bound the two integrals separately. For the first integral, by integrating by parts we have

$$\begin{aligned}
&\int_0^x f_Z(t)\log\left(\frac{t}{x}\right)t \\
&= \left[\Pr(Z\le t)\log\left(\frac{t}{x}\right)\right]_0^x - \int_0^x \Pr(Z\le t)\frac{1}{t}t && \text{(C.22)}\\
&\ge -\int_0^x\int_0^t nP_{Y|X}(\lfloor nz\rfloor|x)z\frac{1}{t}t && \text{(C.23)}\\
&\ge -\int_0^x nP_{Y|X}(\lfloor nt\rfloor|x)t && \text{(C.24)}\\
&= -\int_0^x f_Z(t)t && \text{(C.25)}\\
&\ge -1 && \text{(C.26)}
\end{aligned}$$

where in (C.24) we used that $\int_0^t nP_{Y|X}(\lfloor nz\rfloor|x)z \le tnP_{Y|X}(\lfloor nt\rfloor|x)$ thanks to the following lemma and to $t\le x$: Let $P_{Y|X}$ be a Binomial pmf. Then, $y\mapsto P_{Y|X}(y|x)$ is increasing for $y\le\lfloor(n+1)x\rfloor$, and decreasing for $y\ge\lceil(n+1)x\rceil$.

Proof. From the ratio

$$\frac{P_{Y|X}(y|x)}{P_{Y|X}(y-1|x)} = \frac{n-y+1}{y}\frac{x}{1-x} \tag{C.27}$$

we see that the condition $P_{Y|X}(y|x)\ge P_{Y|X}(y-1|x)$ is satisfied for $y\le\lfloor(n+1)x\rfloor$. □

For the second integral, write

$$\int_x^1 f_Z(t)\log\left(\frac{t}{x}\right)t\ge 0. \tag{C.28}$$

Putting together the two results, we get the result in (C.14). □

We are now ready to give the main result of this appendix. For $x\in[0,1]$, the entropy of a Binomial rv is bounded as follows

$$H(Y|X=x)\le\frac{1}{2}\log\left(2\pi\left(nx(1-x)+\frac{1}{12}\right)\right), \tag{C.29}$$

$$H(Y|X=x) \geq (1-(1-x)^n - x^n)\frac{1}{2}\log(2\pi n)$$
$$+\frac{1}{2}(1-(1-x)^n)\log(x)$$
$$+\frac{1}{2}(1-x^n)\log(1-x) - 1. \tag{C.30}$$

Proof. For the upper bound, write

$$H(Y|X=x) = h(Y+U|X=x) \tag{C.31}$$
$$\leq \frac{1}{2}\log\left(2\pi\left(nx(1-x)+\frac{1}{12}\right)\right), \tag{C.32}$$

where (C.31) follows from [62, Lemma 17] with $U \sim \mathcal{U}[0,1]$ is independent of Y; and the last step follows from the Gaussian maximizes entropy principle.

Next we prove the lower bound. First of all, compute

$$-H(Y|X=x) = \mathbb{E}\left[\log\left(\binom{n}{Y}x^Y(1-x)^{n-Y}\right) \,\middle|\, X=x\right] \tag{C.33}$$
$$= \mathbb{E}\left[\log\binom{n}{Y} \,\middle|\, X=x\right] + nx\log(x)$$
$$+ n(1-x)\log(1-x) \tag{C.34}$$
$$\leq \mathbb{E}\left[\log\binom{n}{Y} \,\middle|\, X=x\right] - nH_2(x) \tag{C.35}$$

By using the bound $\binom{n}{Y} \leq \sqrt{\frac{n}{2\pi Y(n-Y)}}^{nH_2(\frac{Y}{n})}$ for $0 < Y < n$ (see, e.g., [68, Problem 5.8]), we can write:

$$\mathbb{E}\left[\log\binom{n}{Y} \,\middle|\, X=x\right]$$
$$= \mathbb{E}\left[\mathbb{1}(0<Y<n)\log\binom{n}{Y} \,\middle|\, X=x\right] \tag{C.36}$$
$$\leq (1-(1-x)^n - x^n)\frac{1}{2}\log\left(\frac{n}{2\pi}\right)$$
$$-\frac{1}{2}\mathbb{E}\left[\mathbb{1}(0<Y<n)\log(Y(n-Y)) \mid X=x\right]$$
$$+ n\mathbb{E}\left[H_2(\frac{Y}{n}) \,\middle|\, X=x\right] \tag{C.37}$$

$$
\begin{aligned}
&= -(1-(1-x)^n - x^n)\frac{1}{2}\log(2\pi n) \\
&\quad - \frac{1}{2}\mathbb{E}\left[\mathbb{1}(0 < Y < n)\log\left(\frac{Y}{n}\frac{n-Y}{n}\right) \,\middle|\, X = x\right] \\
&\quad + n\mathbb{E}\left[H_2\left(\frac{Y}{n}\right) \,\middle|\, X = x\right] && \text{(C.38)} \\
&= -(1-(1-x)^n - x^n)\frac{1}{2}\log(2\pi n) \\
&\quad - \frac{1}{2}\mathbb{E}\left[\mathbb{1}(0 < Y \le n)\log\left(\frac{Y}{n}\right) \,\middle|\, X = x\right] \\
&\quad - \frac{1}{2}\mathbb{E}\left[\mathbb{1}(0 < Y \le n)\log\left(\frac{Y}{n}\right) \,\middle|\, X = 1-x\right] \\
&\quad + n\mathbb{E}\left[H_2(\frac{Y}{n}) \,\middle|\, X = x\right] && \text{(C.39)} \\
&\le -(1-(1-x)^n - x^n)\frac{1}{2}\log(2\pi n) \\
&\quad - \frac{1}{2}\mathbb{E}\left[\mathbb{1}(0 < Y \le n)\log\left(\frac{Y}{n}\right) \,\middle|\, X = x\right] \\
&\quad - \frac{1}{2}\mathbb{E}\left[\mathbb{1}(0 < Y \le n)\log\left(\frac{Y}{n}\right) \,\middle|\, X = 1-x\right] + nH_2(x), && \text{(C.40)}
\end{aligned}
$$

where in (C.39) we used the channel symmetry $P_{Y|X}(y|x) = P_{Y|X}(n-y|1-x)$; and in the last step we used Jensen's inequality and $\mathbb{E}[Y \mid X = x] = nx$.

By using Lemma C.1, we have

$$
\begin{aligned}
&\mathbb{E}\left[\mathbb{1}(0 < Y \le n)\log\left(\frac{Y}{n}\right) \,\middle|\, X = x\right] \\
&\quad \ge (1-(1-x)^n)\log(x) - 1 && \text{(C.41)}
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}\left[\mathbb{1}(0 < Y \le n)\log\left(\frac{Y}{n}\right) \,\middle|\, X = 1-x\right] \\
&\quad \ge (1-x^n)\log(1-x) - 1. && \text{(C.42)}
\end{aligned}
$$

Therefore, we have

$$
\mathbb{E}\left[\log\binom{n}{Y} \,\middle|\, X = x\right]
$$

$$\leq -(1-(1-x)^n - x^n)\frac{1}{2}\log(2\pi n)$$
$$-\frac{1}{2}(1-(1-x)^n)\log(x) - \frac{1}{2}(1-x^n)\log(1-x)$$
$$+1+nH_2(x) \tag{C.43}$$

and

$$-H(Y|X=x)$$
$$\leq -(1-(1-x)^n - x^n)\frac{1}{2}\log(2\pi n)$$
$$-\frac{1}{2}(1-(1-x)^n)\log(x) - \frac{1}{2}(1-x^n)\log(1-x)+1. \tag{C.44}$$

□

REFERENCES

[1] E. Weinstein and A. J. Weiss, “A general class of lower bounds in parameter estimation,” *IEEE Transactions on Information Theory*, vol. 34, no. 2, pp. 338–342, 1988.

[2] H. V. T. K. BELL and Z. Tian, *Detection, Estimation, and Modulation Theory Part I.* John Wiley & Sons, New York, 1968.

[3] B. Bobrovsky and M. Zakai, “A lower bound on the estimation error for certain diffusion processes,” *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 45–52, 1976.

[4] E. Barankin, “Locally best unbiased estimates,” *The Annals of Mathematical Statistics*, vol. 20, no. 4, pp. 477–501, 1949.

[5] B.-Z. Bobrovsky, E. Mayer-Wolf, and M. Zakai, “Some classes of global Cramér-Rao bounds,” *The Annals of Statistics*, pp. 1421–1438, 1987.

[6] J. Ziv and M. Zakai, “Some lower bounds on signal parameter estimation,” *IEEE Transactions on Information Theory*, vol. 15, no. 3, pp. 386–391, 1969.

[7] L. P. Seidman, “Performance limitations and error calculations for parameter estimation,” *Proceedings of the IEEE*, vol. 58, no. 5, pp. 644–652, 1970.

[8] D. Chazan, M. Zakai, and J. Ziv, “Improved lower bounds on signal parameter estimation,” *IEEE Transactions on Information Theory*, vol. 21, no. 1, pp. 90–93, 1975.

[9] S. Bellini and G. Tartara, “Bounds on error in signal parameter estimation,” *IEEE Transactions on Communications*, vol. 22, no. 3, pp. 340–342, 1974.

[10] K. L. Bell, Y. Steinberg, Y. Ephraim, and H. L. Van Trees, “Extended Ziv-Zakai lower bound for vector parameter estimation,” *IEEE Transactions on Information Theory*, vol. 43, no. 2, pp. 624–637, 1997.

[11] A. Dytso, M. Fauß, A. M. Zoubir, and H. V. Poor, “MMSE bounds for additive noise channels under Kullback–Leibler divergence constraints on the input distribution,” *IEEE Transactions on Signal Processing*, vol. 67, no. 24, pp. 6352–6367, 2019.

[12] M. Raginsky and I. Sason, “Concentration of measure inequalities in information theory, communications and coding,” *arXiv preprint arXiv:1212.4663*, 2012.

[13] S. L. Fong and V. Y. Tan, “A proof of the strong converse theorem for Gaussian broadcast channels via the Gaussian poincaré inequality,” *IEEE Transactions on Information Theory*, vol. 63, no. 12, pp. 7737–7746, 2017.

[14] Y. Polyanskiy and S. Verdú, “Empirical distribution of good channel codes with nonvanishing error probability,” *IEEE Transactions on Information theory*, vol. 60, no. 1, pp. 5–21, 2013.

[15] W. Hachem, O. Khorunzhiy, P. Loubaton, J. Najim, and L. Pastur, “A new approach for mutual information analysis of large dimensional multi-antenna channels,” *IEEE Transactions on Information Theory*, vol. 54, no. 9, pp. 3987–4004, 2008.

[16] D. Bakry, F. Barthe, P. Cattiaux, and A. Guillin, “A simple proof of the poincaré inequality for a large class of probability measures,” *Electronic Communications in Probability*, vol. 13, pp. 60–66, 2008.

[17] A. D. Sarwate and A. G. Dimakis, “The impact of mobility on gossip algorithms,” *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1731–1742, 2012.

[18] A. Schlichting, “Poincaré and log–sobolev inequalities for mixtures,” *Entropy*, vol. 21, no. 1, p. 89, 2019.

[19] A. Vidotto, “An improved second-order poincaré inequality for functionals of gaussian fields,” *Journal of Theoretical Probability*, vol. 33, no. 1, pp. 396–427, 2020.

[20] S. Chatterjee, “Fluctuations of eigenvalues and second order poincaré inequalities,” *Probability Theory and Related Fields*, vol. 143, no. 1, pp. 1–40, 2009.

[21] E. Aras, K.-Y. Lee, A. Pananjady, and T. A. Courtade, “A family of Bayesian Cramér-Rao bounds, and consequences for log-concave priors,” in *2019 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2019, pp. 2699–2703.

[22] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford university press, 2013.

[23] D. Bakry and M. Émery, “Diffusions hypercontractives,” pp. 177–206, 1985.

[24] S. Bobkov and M. Ledoux, “Poincaré’s inequalities and Talagrand’s concentration phenomenon for the exponential distribution,” *Probability Theory and Related Fields*, vol. 107, no. 3, pp. 383–400, 1997.

[25] A. Dytso and M. Cardone, “A general derivative identity for the conditional expectation with focus on the exponential family,” *arXiv preprint arXiv:2105.05106*, 2021.

[26] A. Dytso, H. V. Poor, and S. Shamai, “A general derivative identity for the conditional mean estimator in Gaussian noise and some applications,” *arXiv preprint arXiv:2104.01883*, 2021.

[27] Y. Polyanskiy, H. V. Poor, and S. Verdú, “channel coding rate in the finite blocklength regime,” *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[28] D. Guo, Y. Wu, S. Shamai (Shitz), and S. Verdú, “Estimation in Gaussian noise: properties of the minimum mean-square error,” *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2371–2385, 2011.

[29] I. Zieder, A. Dytso, and M. Cardone, "An MMSE lower bound via poincare inequality," in *IEEE International Symposium on Information Theory. ISIT 2022*, 2022, Extended Version.

[30] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddler River, NJ: Prentice-Hall, Inc., 1993.

[31] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1261–1282, 2005.

[32] Y. Wu and S. Verdú, "Optimal phase transitions in compressed sensing," *IEEE Transactions on Information Theory*, vol. 58, no. 10, pp. 6241–6263, 2012.

[33] D. L. Donoho and J. Tanner, "Counting the faces of randomly-projected hypercubes and orthants, with applications," *Discrete & Computational Geometry*, vol. 43, no. 3, pp. 522–541, 2010.

[34] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, 2009.

[35] D. Guo, D. Baron, and S. Shamai, "A single-letter characterization of optimal noisy compressed sensing," in *2009 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2009, pp. 52–59.

[36] W. Yihong and S. Verdú, "MMSE dimension," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 4857–4879, 2011.

[37] E. Weinstein, "Relations between Belini-Tartara, Chazan-Zakai-Ziv, and Wax-Ziv lower bounds," *IEEE Transactions on Information Theory*, vol. 34, no. 2, pp. 342–343, 1988.

[38] H. Steendam and M. Moeneclaey, "Low-SNR limit of the Cramer-Rao bound for estimating the carrier phase and frequency of a PAM, PSK, or QAM waveform," *IEEE Communications Letters*, vol. 5, no. 5, pp. 218–220, 2001.

[39] F. Rice, B. Cowley, B. Moran, and M. Rice, "Cramer-Rao lower bounds for QAM phase and frequency estimation," *IEEE Transactions on Communications*, vol. 49, no. 9, pp. 1582–1591, 2001.

[40] W. Xu, A. B. Baggeroer, and C. D. Richmond, "Bayesian bounds for matched-field parameter estimation," *IEEE Transactions on Signal Processing*, vol. 52, no. 12, pp. 3293–3305, 2004.

[41] A. Dytso and H. V. Poor, "Estimation in Poisson noise: Properties of the conditional mean estimator," *IEEE Transactions on Information Theory*, vol. 66, no. 7, pp. 4304–4323, 2020.

[42] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and conditional mean estimation in Poisson channels," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 1837–1849, 2008.

[43] T. Cover and J. Thomas, *Elements of Information Theory: Second Edition*. Hoboken, NJ: Wiley, 2006.

[44] J. Jiao, K. Venkat, and T. Weissman, "Relations between information and estimation in discrete-time Lévy channels," *IEEE Transactions on Information Theory*, vol. 63, no. 6, pp. 3579–3594, 2017.

[45] N. Farsad, W. Chuang, A. Goldsmith, C. Komninakis, M. Médard, C. Rose, L. Vandenberghe, E. E. Wesel, and R. D. Wesel, "Capacities and optimal input distributions for particle-intensity channels," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 6, no. 3, pp. 220–232, 2020.

[46] A. Einolghozati, M. Sardari, and F. Fekri, "Design and analysis of wireless communication systems using diffusion-based molecular communication among bacteria," *IEEE Transactions on Wireless Communications*, vol. 12, no. 12, pp. 6096–6105, 2013.

[47] N. Farsad, C. Rose, M. Médard, and A. Goldsmith, "Capacity of molecular channels with imperfect particle-intensity modulation and detection," in *2017 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2017, pp. 2468–2472.

[48] V. Jamali, A. Ahmadzadeh, W. Wicke, A. Noel, and R. Schober, "Channel modeling for diffusive molecular communication—A tutorial review," *Proceedings of the IEEE*, vol. 107, no. 7, pp. 1256–1301, 2019.

[49] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," in *Soviet Physics Doklady*, vol. 10, no. 8. Soviet Union, 1966, pp. 707–710.

[50] M. Cheraghchi, "Capacity upper bounds for deletion-type channels," *Journal of the ACM (JACM)*, vol. 66, no. 2, pp. 1–79, 2019.

[51] C. Komninakis, L. Vandenberghe, and R. D. Wesel, "Capacity of the binomial channel, or minimax redundancy for memoryless sources," in *IEEE International Symposium on Information Theory*, 2001, pp. 127–127.

[52] Q. Xie and A. R. Barron, "Minimax redundancy for the class of memoryless sources," *IEEE Transactions on Information Theory*, vol. 43, no. 2, pp. 646–657, 1997.

[53] H. Witsenhausen, "Some aspects of convexity useful in information theory," *IEEE Transactions on Information Theory*, vol. 26, no. 3, pp. 265–271, 1980.

[54] C. G. Taborda, D. Guo, and F. Perez-Cruz, "Information-estimation relationships over binomial and negative binomial models," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2630–2646, 2014.

[55] A. Dytso, M. Fauß, and H. V. Poor, "Bayesian risk with Bregman loss: A Cramér–Rao type bound and linear estimation," *IEEE Transactions on Information Theory*, vol. 68, no. 3, pp. 1985–2000, 2022.

[56] M. J. Salariseddigh, V. Jamali, H. Boche, C. Deppe, and R. Schober, "Deterministic identification for MC Binomial channel," *arXiv preprint arXiv:2304.12493*, 2023.

[57] A. Dytso, M. Goldenbaum, H. V. Poor, and S. S. Shitz, "When are discrete channel inputs optimal? — optimization techniques and some new results," in *2018 52nd Annual Conference on Information Sciences and Systems (CISS)*, 2018, pp. 1–6.

[58] J. Kemperman, "On the Shannon capacity of an arbitrary channel," in *Indagationes Mathematicae (Proceedings)*, vol. 77, no. 2. North-Holland, 1974, pp. 101–115.

[59] S. Karlin, *Total positivity, Vol. I.* Redwood City, CA: Stanford University Press, 1968.

[60] J. G. Smith, "The information capacity of amplitude-and variance-constrained scalar Gaussian channels," *Information and Control*, vol. 18, no. 3, pp. 203–219, 1971.

[61] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: JHU press, 2013.

[62] A. Lapidoth and S. M. Moser, "On the capacity of the discrete-time Poisson channel," *IEEE Transactions on Information Theory*, vol. 55, no. 1, pp. 303–322, 2008.

[63] A. L. McKellips, "Simple tight bounds on capacity for the peak-limited discrete-time channel," in *Proceedings from IEEE International Symposium on Information Theory (ISIT)*, Chicago, IL, Jul. 2004, pp. 348–348.

[64] A. Thangaraj, G. Kramer, and G. Böcherer, "Capacity bounds for discrete-time, amplitude-constrained, additive white Gaussian noise channels," *IEEE Transactions on Information Theory*, vol. 63, no. 7, pp. 4172–4182, 2017.

[65] C. Knessl, "Integral representations and asymptotic expansions for Shannon and Renyi entropies," *Applied Mathematics Letters*, vol. 11, no. 2, pp. 69–74, 1998.

[66] M. Cheraghchi, "Expressions for the entropy of binomial-type distributions," in *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2018, pp. 2520–2524.

[67] D. Kershaw, "Some extensions of W. Gautschi's inequalities for the gamma function," *Mathematics of Computation*, vol. 41, no. 164, pp. 607–611, 1983.

[68] R. G. Gallager, *Information Theory and Reliable Communication*. Hoboken, NJ: John Wiley & Sons, 1968.