Guessing, Entropy and Large Deviations

Vinita Jakhar

May, 2025

Abstract

The problem of guessing the value of a discrete random variable X by sequentially querying possible values until the correct one is identified has been a topic of significant interest in information theory and cryptography. This problem, originally introduced by Massey in 1994 his seminal work "Guessing and Entropy, explores the relationship between the average number of guesses required to identify X and its entropy $H(X)$. He also showed that no meaningful upper bound exists for $E[G]$ in terms of $H(X)$. Massey studied a setting where X can assume a countably infinite set of possible values.

Building on Massey's foundation, subsequent research has expanded the scope of the guessing problem in several directions.In 1996 Arikan investigated the role of Rényi entropy in bounding the moments of guesswork, revealing deep connections between guessing strategies and information-theoretic measures.However, Arikan restricted to a setting where X can assume only a finitely many values. In 2017 Huleihel, Salamatian, and Médard explored the scenario of a memoryless guesser, where each guess is independent of previous attempts. They derived optimal guessing strategies for this setting and established connections to Rényi entropy, demonstrating that a memoryless guesser can asymptotically perform as well as one with perfect memory.

Christiansen and Duffy (2013) further advanced this line of inquiry by establishing a large deviation principle (LDP) for the logarithm of guesswork, providing precise estimates of the guesswork distribution for long sequences. Their work showed that the rate function governing the LDP exhibits a specific structure, encapsulating the underlying probabilistic nature of the problem.

Further extensions include the study of conditional guesswork, where the guesser has access to correlated side information Y. Li (2017) investigated large deviation principles for conditional guesswork, generalizing earlier results to this setting. The analysis revealed that

the scaled cumulant generating function of conditional guesswork is linked to conditional Rényi entropy, providing a framework for understanding how side information influences guessing performance.