,

Guessing, Entropy and Large Deviations

VINITA JAKHAR
under the supervision of
M Ashok Kumar

May, 2025

INTRODUCTION

- The problem of guessing originally introduced by Massey in 1994 his seminal work "Guessing and Entropy" , explores the relationship between the expected number of guesses required to guess X and its entropy $H(X)$.
- Arikan (1996) studied the problem for random variable with finite support and established the role of Rényi entropy in bounding the moments of guesswork.
- Huleihel, Salamatian, and Médard (2017) studied memoryless guessing where each guess is independent of previous attempts. They derived optimal guessing strategies for this setting and established connections to Rényi entropy.

- Christiansen and Duffy (2013) established a large deviation principle (LDP) for the logarithm of guesswork, providing precise estimates of the guesswork distribution for long sequences.
- Y. Li (2017) investigated large deviation principles for conditional guesswork where the guesser has access to correlated side information.

GUESSING AND ENTROPY

- Let X be a random variable that assumes values from a countably infinite set $\mathcal{X}$.
- Consider the problem of guessing X in one trial of a random experiment by asking questions of the form "Did X take on its i-th possible value?" until the answer is "yes".
- Let G be the number of guesses required to guess X correctly.
- Our objective is to minimize $E[G]$ over all 'guessing strategies' .
- The optimal guessing strategy is to guess X according to the decreasing order of probabilities

A Lower Bound On E[G]

Massey(1994) show that

$$E[G] \geq \frac{1}{4} 2^{H(X)} + 1$$

- The proof was based on maximum entropy problem.

Remarks :

1. The idea of Massey cannot be extended to finite support
2. $E[G]$ may be arbitrarily large when H(X) is an arbitrarily small positive number so that there is no interesting upper bound on $E[G]$ in terms of $H(X)$.

Example:

Let X follow:

$$P(X = 0) = 1 - \epsilon$$

$$P(X = k) = \epsilon/(M - 1) \qquad k = 1, \ldots, M - 1$$

Then:

$$H(X) \approx \epsilon \log(M/\epsilon) \to 0 \quad \text{as} \quad \epsilon \to 0$$

$$E[G(X)] \approx 1 + \epsilon M/2 \to \infty \quad \text{as} \quad M \to \infty$$

Theorem[Arikan(1996)]

Theorem

Consider the random variable $\mathcal{X} = \{x_1, x_2, \ldots, x_M\}$
Then for $\rho \geq 1$,

$$\mathbb{E}[G(X)^\rho] \geq (1 + \log M)^{-\rho} \left[\sum_x P_X(x)^{\frac{1}{1+\rho}} \right]^{1+\rho},$$

where P_X *is the probability distribution of* X

- **Rényi entropy** of order $\alpha > 0$ is defined as

$$H_\alpha(X) := \frac{1}{1-\alpha} \log \left[\sum_{x \in \mathcal{X}} P_X(x)^\alpha \right]$$

Hence:

$$\mathbb{E}[G(X)] \geq \frac{2^{H_{\frac{1}{2}}(X)}}{1 + \log M}.$$

Lower Bound on Moments in Terms of Shannon Entropy

Here I am using the same method as Arikan to find the bound in terms of the Shannon entropy.
Consider the random variable with finite range $\mathcal{X} = \{x_1, x_2, \ldots, x_M\}$ and distribution $P_X(x_i) = p_i$ for $i = 1, \ldots, M$.
Assume $G(x_j) = i$ (i.e., x_j is the i-th guess).
Then,

$$\mathbb{E}[G(X)^\rho] = \sum_x P(x) \exp\left(-\log\frac{1}{G(x)^\rho}\right)$$

Using Jensen's inequality, we get

$$\mathbb{E}[G(X)^\rho] \geq \exp\left(H(\mathbf{p}) - \rho\log\sum_x \frac{1}{G(x)}\right)$$

$$\geq \frac{2^{H(\mathbf{p})}}{(1+\log M)^\rho}$$

Letting $\rho = 1$, we get

$$\mathbb{E}[G(X)] \geq \frac{2^{H(\mathbf{p})}}{1 + \log M}$$

Remarks :

- The lower bound that comes in terms of Shannon entropy is not a tight bound .
  By the monotone decreasing property of Rényi entropy (i.e. as value of α increases Rényi entropy value decreases.)

$$\mathbb{E}[G(X)] \geq \frac{2^{H_{\frac{1}{2}}(X)}}{1 + \log M} \geq \frac{2^{H_1(X)}}{1 + \log M} = \frac{2^{H(X)}}{1 + \log M}.$$

Theorem[Arikan(1996)]

Theorem

Let $X_1, \ldots, X_n$ be a sequence of i.i.d. random variables over a finite set. Let $G^*(X_1, \ldots, X_n)$ be an optimal guessing function. Then, for any $\rho > 0$,

$$\lim_{n\to\infty} \frac{1}{n\rho} \ln \mathbb{E}\left[(G^*(X_1, \ldots, X_n))^\rho\right] = H_{1/(1+\rho)}(X),$$

MEMORYLESS GUESSING

- Suppose Bob thinks of a random variable X between 1 and M with probability distribution P_X.
- Alice tries to guess it with probability distribution $\hat{P}$ by asking questions only of the form "is $X = x$?" with every guess independent of the previous guesses.
- The setting we consider is one in which Alice knows the distribution P_X and presents a sequence of i.i.d guesses $\hat{X}_1, \hat{X}_2$ $\ldots$ drawn from some distribution $\hat{P}(.)$.

- We define *Hitting time*

$$\mathcal{G}(X, \hat{X}_1^\infty) := \inf\{k \geq 1 : \hat{X}_k = X\} \ ,$$

That is, the number of guesses until a success.

- For a given integer $\rho \geq 1$, we define the quantity , called ρ^{th} **factorial moment function**.

$$V_\rho(X, \hat{X}_1^\infty) := \frac{1}{\rho!} \prod_{l=0}^{\rho-1} (\mathcal{G}(X, \hat{X}_1^\infty) + l)$$

Theorem

For any integer $\rho \geq 1$,

$$\log \mathbb{E}\left\{V_\rho^*(X, \hat{X}_1^\infty)\right\} = \rho \cdot H_{\frac{1}{1+\rho}}(X),$$

and for any $x \in X$,

$$\hat{P}^*(x) = \frac{P_X(x)^{\frac{1}{1+\rho}}}{\sum_{x' \in \mathcal{X}} P_X(x')^{\frac{1}{1+\rho}}}.$$

where

$$E\{V_\rho^*(X, \hat{X}_1^\infty)\} := \inf_{\hat{P} \in \mathcal{P}} \mathbb{E}\{V_\rho(X, \hat{X}_1^\infty)\}$$

and $\mathcal{P}$ is the set of all probability distributions on $\mathcal{X}$

Guessing a Sequence of Random Variables:

Now, we can consider the case of guessing a sequence $X^n = (X_1, \ldots, X_n)$ of i.i.d random variables distributed according to P_{X^n}.

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left\{ V_\rho^*(X^n, \hat{X}_1^\infty) \right\} = \rho \cdot H_{\frac{1}{1+\rho}}(X).$$

- **Remark :** 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.

GUESSWORK AND LARGE DEVIATIONS

- If a password W_k is chosen at random from a finite set $\mathbb{A}^k = \{1, 2, \ldots, m^k\}$, how hard is it to guess W_k?
- If $P(W_k = w)$ is known, then an optimal strategy is to guess passwords in decreasing order of probability.
- Let $G(w)$ denote the number of attempts required before correctly guessing $w \in \mathbb{A}^k$.

Scaled Cumulant generating function

- **Scaled Cumulant generating function:** Consider the sequence of random variables $\{k^{-1} \log G(W_k)\}$ and the scaled cumulant generating function (sCGF) of this sequence:

$$\Lambda(\alpha) := \lim_{k \to \infty} \frac{1}{k} \log \mathbb{E}\left[e^{\alpha \log G(W_k)}\right]$$

$$= \lim_{k \to \infty} \frac{1}{k} \log \mathbb{E}\left[G(W_k)^{\alpha}\right]$$

- **Rate Function:** We define the candidate rate function as the Legendre-Fenchel transform of the sCGF

$$\Lambda^*(x) := \sup_{\alpha \in \mathbb{R}} \{\alpha x - \Lambda(\alpha)\}$$

Theorem

The sequence $\{k^{-1}\log G(W_k)\}$ *satisfies a LDP with rate function* Λ^*. *i.e*

$$\lim_{\epsilon\downarrow 0}\limsup_{k\to\infty}\frac{1}{k}\log P\left(\frac{1}{k}\log G(W_k)\in B_\epsilon(x)\right)$$
$$=\lim_{\epsilon\downarrow 0}\liminf_{k\to\infty}\frac{1}{k}\log P\left(\frac{1}{k}\log G(W_k)\in B_\epsilon(x)\right)$$
$$=-\Lambda^*(x)$$

- **Direct Estimates on Guesswork :**
  LDP for the sequence $\frac{1}{k}\log G(W_k)$ used to develop the more valuable direct estimate of the distribution of each $G(W_k)$. From the LDP, we have the approximation that for large k:

$$P(G(W_k) = n) \approx \frac{1}{n} \exp\left(-k\Lambda^* \left(\frac{1}{k} \log n\right)\right)$$

Remarks:

1. As this calculation only involves the determination of Λ^*, to approximately calculate the probability of the n-th most likely word in words of length k, one does not have to identify the word itself.
2. LDP gives direct estimates on the guesswork distribution $P(G(W_k) = n)$ for large k.

- X is the random variable to be guessed by a series of truthfully answered questions of the form "Is $X = x$?", while Y is a correlated random variable that is directly observed.
- We call $G(X|Y)$ a guessing function of X given Y.
  For example, in sequential decoding, we can think of X as channel input and Y as channel output.

- **Setup:** Let $X \in \{000, 001, \ldots, 111\}$ (3-bit strings). Channel flips each bit independently with probability $p = 0.1$. Suppose $X = 010$ is sent and $Y = 000$ is received.
  Without side information the guesser guesses uniformly over all 8 codewords.
  With side information $Y = 000$, $X \in \{001, 010, 100\}$ so guesser need at most 3 attempts to guess the codeword correctly.

Future work:

1. Finding an upper bound on $E(G)$ for a finite number of objects in terms of Renyi entropy.
2. Finding both upper and lower bounds on $E(G)$ for an infinite number of objects in terms of Renyi entropy.
3. Memoryless guessing problem for an infinite number of objects.

REFERENCES:

1. J.L. Massey. Guessing and entropy. In Proceedings of 1994 IEEE International Symposium on Information Theory, pages 204–, 1994.
2. E. Arikan. An inequality on guessing and its application to sequential decoding. IEEE Transactions on Information Theory, 42(1):99–105, 1996.
3. Wasim Huleihel, Salman Salamatian, and Muriel M ´edard. Guessing with limited memory. In 2017 IEEE International Symposium on Information Theory (ISIT), pages 2253–2257, 2017.
4. Mark M. Christiansen and Ken R. Duffy. Guesswork, large deviations, and shannon entropy. IEEE Transactions on Information Theory, 59(2):796–802, 2013.
5. Jiange Li. Large deviations for conditional guesswork. Statistics Probability Letters, 153:7–14, 2019

Thank You!