# Exploring Marginal Guesswork with the Theorem of Berry-Esséen

Timo Glaser
Ruhr-University Bochum
timo.glaser@rub.de

**Abstract**

In 2000, Pliam showed that there does not exist an upper or lower bound in terms of Shannon entropy alone for the number of guesses required in order to guess some randomly sampled element $s$ with certainty $0 < q < 1$, denoted as marginal guesswork. In comparison, it was shown by Arikan in 1996 that, if the distribution is defined over finite support, the expected guesswork, that is the expected amount of trials it takes to guess some random element $s$ with certainty, is bound in terms of Rényi entropy $\mathrm{H}_{1/2}$. In this paper, we utilize the Theorem of Berry-Esséen to amend Pliams result. We show that there exists some asymptotic expression for the marginal guesswork in terms of Shannon entropy, variance in information and the probit function, provided $q$ is fixed.

***Index terms—*** Marginal Guesswork, Expected Guesswork, Theorem of Berry-Esséen, Shannon Entropy, Rényi Entropy, Probit Function, Variance in Information

## I. Introduction

Let us assume that some target element $s$ was sampled from some known probability distribution $\mathcal{P}$ and that we would like to find out which element was sampled, assuming we had access to a truthful oracle that, on input $X$, returns "yes" if $X = s$ and "no" otherwise. An algorithm like this can sometimes be found as a subroutine in cryptographic attacks on e.g. LWE schemes [1], [2].

In 1994, Massey [3] proved that the expected guesswork $\mathbb{E}[G(X)]$, i.e. the *expected amount of trials* required to find $X$ such that $X = s$, is bound from below by

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

if $\mathrm{H}(X) \geq 2$. This lower bound was later improved by Rioul [4] to

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

for *all* distributions $\mathcal{P}$.

Massey [3] also showed that no matching upper bound in terms of Shannon entropy alone could exist.

In 1996, Arikan [5] gave bounds for $\mathbb{E}[G(X)]$ that depend solely on the size $M = |\mathcal{X}|$ of the support $\mathcal{X}$ of $\mathcal{P}$ and the *Renyi* Entropy $\mathrm{H}_{1/2}(X)$ of order $\frac{1}{2}$. He showed that, for *any* $X \sim \mathcal{P}$, if we guess in order such that the most likely element was guessed first and then all other elements were guessed in order of sampling probability, then

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

In the same paper, Arikan generalized this expression for any moment of $G$, i.e. he found expressions of $\mathbb{E}[G(X)^\rho]$ in terms of $\mathrm{H}_{\frac{1}{1+\rho}}(X)$. For distributions that are defined over *infinite* support, Sason and Verdu [6] and Rioul [4] found similar expressions for lower bounds.

Another measure of uncertainty that is of interest for this setting is that of *Marginal Guesswork* $\mathrm{MG}(q)$, i.e. the question as to *how many* elements one has to trial in order to find the desired element with some predetermined success probability $q \in [0, 1]$. This number is equal to the size of the smallest set $\mathcal{S}_q \subseteq \mathcal{X}$ with probability mass of

$$p(\mathcal{S}_q) = \sum_{x \in \mathcal{S}_q} p(\mathbf{x}) \geq q.$$

Pliam [7] showed that there exists no bounds for $\mathrm{MG}(q)$ in terms of Shannon entropy alone. However, if $\mathbf{X}$ is a random vector instead with i.i.d. sampled coordinates $X_1, X_2, \ldots, X_n$, then we know from Shannon [8] that

$$\lim_{n\to\infty} \frac{\log |\mathcal{S}_q^n|}{n} = \mathrm{H}(X).$$

In 2023, Glaser, May and Nowakowski [9] proved the existence of some set $\mathcal{C}^n \subseteq \mathcal{X}^n$, denoted the *core set*, of size

$$|\mathcal{C}^n| \leq 2^{\mathrm{H}(X)n}$$

with a probability mass of

$$p(\mathcal{C}^n) = \frac{1}{2} \pm \mathcal{O}\Big(\frac{1}{\sqrt{n}}\Big),$$

provided the distribution is finite and non-uniform. Furthermore, for any $\frac{1}{2} < q < 1$, they showed the existence of a set $\mathcal{C}_q^n$ of size

$$2^{\mathrm{H}(X)n+\Theta(\sqrt{n})}$$

and probability mass

$$p(\mathcal{C}_q^n) = q \pm \mathcal{O}\Big(\frac{1}{\sqrt{n}}\Big).$$

*Our Contribution*

In this paper, we extend the result of Glaser, May and Nowakowski for most discrete probability distributions. We find an expression for $\mathrm{MG}_n(q)$ that is asymptotic in $n$, provided $\mathcal{P}$ is either non-uniform and has a finite support or presents a distribution function where the probabilities tend to 0 *fast enough*. In particular, we show that, for some fixed $q \in (0,1)$, there exists some constants $r_q, R_q, c_q$ such that

$$\frac{r_q}{\sqrt{n}} 2^{\mathrm{H}(X)n+c_q\sqrt{n}} \leq \mathrm{MG}_n(q) \leq \frac{R_q \log n}{\sqrt{n}} 2^{\mathrm{H}(X)n+c_q\sqrt{n}},$$

provided $n$ is sufficiently large. In comparison, Shannon's [8] result only states that

$$\mathrm{MG}_n(q) = 2^{\mathrm{H}(X)n \pm o(n)}.$$

We also show that, for some fixed $0 < q^- < q^+ < 1$, there exists $r, R$ such that, when replacing $r_q, R_q$ with $r, R$, the expression still holds true for all $q \in [q^-, q^+]$.

At last, we observe what happens when $q$ gets arbitrarily close to 0 or 1. In particular, we show that, if $r, R$ exist such that the above statement holds true for all $q \in [0,1]$, then these constants can be arbitrarily close to the support size (or its reciprocal, for the lower bound).

*Organization of our Paper*

In Section II, we present basic notation and recall some relevant results. In Section III, we provide an asymptotic expression for $\mathrm{MG}_n(q)$ that is tight up to some logarithmic factor. In Section IV, we then show that this bound only holds in the asymptotic setting and that the expression $2^{\mathrm{H}(X)n+c\sqrt{n}}$ can over- or underestimate $\mathrm{MG}(q)$ by a factor of almost $2^{\log M}$ or $2^{-\log M}$, respectively. In Section V, we conclude by presenting three open questions that could not be answered over the course of this paper.

## II. Preamble

### A. Notation

We denote random variables with $X$ and probability distributions with $\mathcal{P}$. Unless specified otherwise, all probability distributions are assumed to be discrete.

The support of a random variable $X$ is denoted with

$$\mathcal{X} = \{x \mid \Pr[X = x] > 0\}.$$

W.l.o.g., with the exception of examples, we assume that the support is always either

$$\mathcal{X} = \{1, \ldots, M\} \qquad \text{or} \qquad \mathcal{X} = \mathbb{N}.$$

We denote random vectors with $\mathbf{X}$. Unless specified otherwise, all random vectors have their coordinates sampled i.i.d. from some probability distribution $\mathcal{P}$. The support of a random vector of length $n$ is denoted with $\mathcal{X}^n$.

We denote the probability distribution function (p.d.f.) of some $x \in \mathcal{X}$ with

$$p(x) = \Pr_{X \leftarrow \mathcal{P}}[X = x]$$

and that of some $\mathbf{x} \in \mathcal{X}^n$ with

$$p(\mathbf{x}) = \Pr_{X_i \leftarrow \mathcal{P}}[(X_1, \ldots, X_n) = \mathbf{x}].$$

W.l.o.g., we assume that the support is sorted in order of descending probability, i.e. that $p(x) \geq p(x+1)$ for all $x \in \mathcal{X}$.

For some set $\mathcal{S} \subseteq \mathcal{X}$, we denote the *probability mass* of $\mathcal{S}$ (w.r.t. p.d.f. $p(x)$) as

$$p(\mathcal{S}) = \sum_{x \in \mathcal{S}} p(x) = \Pr[X \in \mathcal{S}].$$

We express the natural Logarithm with $\ln$ and the logarithm w.r.t. base 2 with $\log$. The *Information* of $x \in \mathcal{X}$ is denoted with

$$\mathrm{I}(x) := -\log(p(x)).$$

The Shannon entropy and the Variance in Information of $X$ are denoted with

$$\mathrm{H}(X) = \sum_{x \in \mathcal{X}} p(x)\,\mathrm{I}(X) = \mathbb{E}[\mathrm{I}(X)],$$
$$\sigma_{\mathrm{I}}^2 = \sum_{x \in \mathcal{X}} p(x)(\mathrm{I}(X) - \mathrm{H}(X))^2 = \mathrm{Var}[\mathrm{I}(X)].$$

For some $\alpha > 0, \alpha \neq 1$, we denote the Rényi entropy of order $\alpha$ of $X$ with

$$\mathrm{H}_\alpha(X) = \frac{1}{1-\alpha} \log \sum_{x \in \mathcal{X}} p(x)^\alpha.$$

We extend this notion for $\alpha = 0, 1, \infty$

$$\mathrm{H}_\alpha(X) = \lim_{\alpha' \to \alpha} \mathrm{H}_{\alpha'}(X)$$

and find

$$\mathrm{H}_0(X) = \log |\mathcal{X}|, \qquad \mathrm{H}_1(X) = \mathrm{H}(X), \qquad \mathrm{H}_\infty(X) = \inf_{x \in \mathcal{X}} \mathrm{I}(x).$$

The cumulative distribution function (c.d.f.) of the standard normal distribution is denoted with

$$\Phi(x) = \Pr_{X \leftarrow \mathcal{N}(0,1)}[X \leq x] = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\,\mathrm{d}t.$$

The inverse of $\Phi$ is denoted with $\Phi^{-1}$ and is referred to as the *Probit Function*.

### *B. Guessing Functions, Marginal Guesswork and Expected Guesswork*

Consider Algorithm 1. This high-level algorithm enumerates all elements from $\mathcal{X}$ in some order until one particular element $s \in \mathcal{X}$ – that was sampled randomly from some known distribution $\mathcal{P}$ – was found or until some predefined amount of guesses would be exceeded. We assume that we have access to some oracle $\tau$ that tells us as to whether $X = s$ as well as some method to enumerate all elements of $\mathcal{X}$ in order.

**Definition 1** (Guessing Function)**.** Let $x_i$ be the $i$-th element that is tested against $\tau$ in Algorithm 1. We call the function

$$G : \mathcal{X} \to \{1, \ldots, |\mathcal{X}|\}$$

*Guessing Function* if

$$G(x_i) = i \qquad \text{for all } 1 \leq i \leq |\mathcal{X}|.$$

If $G(x) < G(x')$ for all $x, x' \in \mathcal{X}$ such that $p(x) > p(x')$ then we call $G$ *optimal*.

*Remark* 1. $G(x)$ tells us How many iterations it takes for Algorithm 1 to test $x$ against $\tau$, so the amount of iterations Algorithm 1 makes before it terminates can always be expressed as $\min\{G(s), b\}$. It is not hard to see that the expected value $\mathbb{E}[G(X)]$ is minimized if $G$ is an optimal guessing function, hence the notion of optimality.

From here on, we always assume that Algorithm 1 enumerates the elements from $\mathcal{X}$ in order of some optimal guessing function $G$, i.e. that the algorithm traverses $\mathcal{X}$ in order of non-increasing probability.

**Algorithm 1:** Enumeration

**Input:** Access to $\tau$, $b \in \mathbb{N} \cup \{\infty\}$
**Output:** $x$ or $\perp$

```
Set counter = 0;
for x ∈ X in some predetermined order do
    if τ(x) == "Yes" then
        return x
    else
        Increment counter by 1;
        if counter == b then
            return ⊥
```

**Definition 2** (Marginal and Expected Guesswork)**.** Let $G$ be some optimal guessing function w.r.t. some probability distribution $\mathcal{P}$ and p.d.f. $p(x)$. For some $0 < q < 1$, we denote the *Marginal Guesswork* (sometimes also called *work factor*) with

$$\mathrm{MG}(q) := \min\{k \in \mathbb{N} \mid \Pr[G(X) < k] \geq q\}.$$

The marginal guesswork of some $n$-dimensional random vector with $i.i.d.$ coordinates is analogously denoted with $\mathrm{MG}_n(q)$.

Furthermore, the *(Expected) Guesswork* denotes the expected amount of trials required to find $\mathbf{s}$ if we traverse $\mathcal{X}$ in order of $G$, i.e.

$$\mathbb{E}[G(X)] = \sum_{x \in \mathcal{X}} p(x)G(x).$$

In terms of Algorithm 1, $\mathrm{MG}(q)$ denotes the minimal size of input $b$ to ensure that we find the target vector with probability at least $q$. Equivalently, $\mathrm{MG}(q)$ denotes the size of the smallest set $\mathcal{S}_q \subseteq \mathcal{X}$ with a probability mass of at least $q$.

Regarding the size of this set – and thus the size of $\mathrm{MG}_n(q)$ – in the $n$-dimensional case, Shannon [8, Theorem 4] proved the following statement.

**Lemma 1** (Shannon)**.** *For any probability distribution $\mathcal{P}$ and any $X \leftarrow \mathcal{P}, 0 < q < 1$, it holds true that*

$$\lim_{n\to\infty} \frac{\log|\mathcal{S}_q^n|}{n} = \mathrm{H}(X) \qquad \textit{for all } 0 < q < 1.$$

In the case of arbitrary one-dimensional distributions, Pliam [7, Theorem 1] proved that marginal guesswork could not be bound in terms of Shannon entropy alone.

**Lemma 2** (Pliam)**.** *For any $0 < q < 1, N \in \mathbb{N}$, there exist a distribution $\mathcal{P}$ such that, for $X \leftarrow \mathcal{P}$,*

$$\mathrm{H}(X) - \log \mathrm{MG}(q) > N.$$

*Similarly, there exists a distribution $\mathcal{P}'$ such that, for $Y \leftarrow \mathcal{P}'$,*

$$\log \mathrm{MG}(q) - \mathrm{H}(Y) > N.$$

As stated above, $\mathbb{E}[G(X)]$ denotes the expected amount of iterations Algorithm 1 makes before finding the target element $s$, assuming we do not abort (i.e. we set $b = \infty$). Arikan [5] found the following bounds for this value.

**Lemma 3** (Arikan)**.** *Let $\mathcal{P}$ be sone distribution over finite support of size $M$. For any $\rho \geq 0$, the $\rho$-th moment of $G$ is bound by*

$$2^{\rho\,\mathrm{H}_{\frac{1}{1+\rho}}(\alpha)X} > \mathbb{E}[G(X)^\rho] \geq \frac{1}{(1+\log M)^\rho} 2^{\rho\,\mathrm{H}_{\frac{1}{1+\rho}}(\alpha)X}.$$

These bounds have since been improved [10, Theorem 3] [4, Corollary 6], though we content ourselves with Arikan's original result.

For distributions over infinite support, Arikan's lower bound becomes trivial, though there exist variants of this inequality that include such distributions by Sason and Verdú [6, Theorem 1] and Rioul [4, Theorem 15].

### *C. The Core Set*

In 2023, Glaser, May and Nowakowski [9] introduced the *Core Set*.

**Definition 3** (Core Set)**.** Let $n \in \mathbb{N}, \delta \in \mathbb{R}$ and let $X$ be such that $\mathrm{H}(X) < \infty$. We define the *Core Set* of dimension $n$ as the set of random vectors with sampling probability at least $2^{-\mathrm{H}(X)n-\delta n}$, i.e.

$$\mathcal{C}^{n,\delta} = \{\mathbf{x} \in \mathcal{X}^n \mid p(\mathbf{x}) \geq 2^{-\mathrm{H}(X)n-\delta n}\}.$$

For our purposes, we slightly alter the notation of the core set for distributions with finite variance in information.

**Definition 4** (Core Set (alternate definition))**.** Let $n \in \mathbb{N}$ and let $X$ be such that $\mathrm{H}(X), \sigma_{\mathrm{I}}^2 < \infty$. Let $\Phi^{-1}$ denote the probit function. For any $0 < q < 1$, we define the *Core Set* of dimension $n$ as the set of random vectors with sampling probability at least $2^{-\mathrm{H}(X)n-\sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}}$, i.e.

$$\mathcal{C}_q^n = \{\mathbf{x} \in \mathcal{X}^n \mid p(\mathbf{x}) \geq 2^{-\mathrm{H}(X)n-\sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}}\}.$$

If $q = \frac{1}{2}$, we drop the subscript and write

$$\mathcal{C}^n = \mathcal{C}_{\frac{1}{2}}^n.$$

Furthermore, we extend the definition for $q = 0, 1$ to

$$\mathcal{C}_0^n = \varnothing, \qquad \mathcal{C}_1^n = \mathcal{X}^n.$$

*Remark* 2. Both definitions can be used interchangeably. Indeed, for $\delta = \frac{\sigma_{\mathrm{I}}\Phi^{-1}(q)}{\sqrt{n}}$, we find

$$\mathcal{C}^{n,\delta} = \mathcal{C}^n_q.$$

*Remark* 3. The core set shares some similarity to the *Typical Set*, see e.g. [11]. Whereas the latter only cares about elements with a sampling probability of around $2^{\mathrm{H}(X)n}$, the former disregards all elements with a sampling probability of *less than* $2^{\mathrm{H}(X)n+\Theta(\sqrt{n})}$. In particular, both sets are interesting as their asymptotic probability mass can be easily determined – 1 in the case of the typical set and $q$ in the case of the core set, see Lemma 5 and Theorem 2.

**Example 1.** Let $\mathcal{P}$ be such that $\sigma_{\mathrm{I}}^2 = 0$, this necessitates that

$$p(x)(-\log p(x) - \mathrm{H}(X))^2 = 0 \quad \text{for all } x \in \mathcal{X}.$$

Thus, $\sigma_{\mathrm{I}}^2 = 0$ if and only if, for all $x \in \mathcal{X}$, $p(x) = \frac{1}{2^{\mathrm{H}(X)}}$, i.e. if $\mathcal{P}$ samples uniformly from $\mathcal{X}$. In that case, we find

$$p(\mathbf{x}) = \frac{1}{M^n} = 2^{-\log M \cdot n} = 2^{-\mathrm{H}(X)n} \qquad \text{for all } \mathbf{x} \in \mathcal{X}^n,$$

so $\mathcal{C}^n = \mathcal{X}^n$. Since $\sigma_{\mathrm{I}} = 0$, we also have $\mathcal{C}^n_q = \mathcal{X}^n$ for all $0 < q \leq 1$.

**Example 2.** Let $\mathbf{X}$ have its coordinates sampled from the binomial distribution $\mathcal{B}(\frac{1}{2}, 2)$, this distribution has Shannon entropy 1.5. We can easily show that the probability mass of $\mathcal{C}^n$ is at least

$$p(\mathcal{C}^n) \geq \frac{1}{2}$$

with equality if and only if $n$ is odd.

To this end, note that any $\mathbf{x}$ with $k$ many coordinates equal to 1 has sampling probability

$$p(\mathbf{x}) = \frac{1}{2^k} \cdot \frac{1}{4^{n-k}} = \frac{1}{2^{2n-k}},$$

so

$$\mathbf{x} \in \mathcal{C}^n \quad \Leftrightarrow \quad -2n + k \geq -1.5n \quad \Leftrightarrow \quad k \geq 0.5n.$$

Hence, any vector $\mathbf{x}$ is contained in $\mathcal{C}^n$ if and only if at least half of its coordinates are equal to 1. Also, note that there are exactly $2^{n-k}\binom{n}{k}$ many such vectors for all $k = 1, \ldots, n$. Thus, if $n$ is odd, the probability mass of $\mathcal{C}^n$ amounts to

$$p(\mathcal{C}^n) = \sum_{\mathbf{x} \in \mathcal{C}^n} p(\mathbf{x}) = \sum_{k=\frac{n+1}{2}}^{n} \frac{1}{2^{2n-k}} 2^{n-k} \binom{n}{k} = \frac{\sum_{k=\frac{n+1}{2}}^{n} \binom{n}{k}}{2^n}.$$

The sum contains one half of the $n$-th level of the pascal triangle, so this yields $p(\mathcal{C}^n) = \frac{1}{2}$. With a similar argument, we can show that, if $n$ is even, then $p(\mathcal{C}^n) > \frac{1}{2}$.

### *D. Berry-Esséen Distributions*

The special case found in Example 2 highlights a strong property of the core set, namely that, for *any* $0 < q < 1$, the probability mass of $\mathcal{C}^n_q$ is around $q$. We will prove this statement with the help of the Theorem of Berry-Esséen [12], [13].

**Lemma 4** (Berry-Esséen)**.** *Let $X_1, \ldots, X_n$ be some i.i.d. real-valued random variables with mean $\mu$ and variance $\sigma^2$ such that*

- $\mu < \infty$,
- $0 < \sigma^2 < \infty$ *and*
- $\mathbb{E}[|X_i - \mathbb{E}[X_i]|^3] < \infty$,

*then there exists a constant $c_{\mathrm{BE}}$ such that, for any $x \in \mathbb{R}$,*

$$\left| \Pr\left[ \frac{\sum X_i - n\mu}{\sigma\sqrt{n}} \leq x \right] - \Phi(x) \right| \leq c_{\mathrm{BE}} \frac{\mathbb{E}[|X_i - \mathbb{E}[X_i]|^3]}{\sigma^3\sqrt{n}}.$$

*The constant $c_{\mathrm{BE}}$ does not depend on our choice of $\mathcal{P}$.*

*Remark* 4. The best known lower and upper bound for $c_{\mathrm{BE}}$ are

$$0.4097 < c_{\mathrm{BE}} < 0.4748$$

and were discovered by Esséen [13] and Shevtsova [14], respectively.

With the help of Lemma 4, Glaser, May and Nowakowski [9] have already proven that, for any $\delta \geq 0$, the probability mass of $\mathcal{C}^{n,\delta}$ is $\gtrsim \frac{1}{2}$ and can be arbitrarily close to 1 for appropriate choices of $\delta$.

**Lemma 5** (Glaser-May-Nowakowski)**.** *Let $\mathcal{P}$ be a probability distribution over finite support. For every $\delta \geq 0$, it holds true that*

$$P(\mathcal{C}^{n,\delta}) \geq \frac{1}{2} \pm \mathcal{O}\Big(\frac{1}{\sqrt{n}}\Big)$$

*Additionally, for every $\varepsilon \in (0,1)$, there exists some $\delta = \Theta\big(\sqrt{\frac{\ln(\varepsilon^{-1})}{n}}\big)$ such that*

$$P(\mathcal{C}^{n,\delta}) \geq 1 - \varepsilon.$$

For our purposes, the result of Lemma 5 is insufficient in that we require appropriate bounds for $\mathcal{C}_q^n$ when $q$ is fixed. For a large subset of probability distributions, we can prove a tighter bound on $p(\mathcal{C}_q^n)$.

**Definition 5** (BE-Distributions)**.** Let $\mathcal{P}$ be some probability distribution. We say that $\mathcal{P}$ is a *Berry-Esséen-Distribution (BE-Distribution)* if, for $X \sim \mathcal{P}$,

- $\mathrm{H}(X) < \infty$,
- $0 < \sigma_{\mathrm{I}} < \infty$ and
- $\mathbb{E}[|\,\mathrm{I}(X) - \mathbb{E}[\mathrm{I}(X)]|^3] < \infty$.

**Theorem 1** (Extension of Glaser-May-Nowakowski)**.** *Let $X_1, \dots, X_n$ be some i.i.d. random variables that are sampled from some BE-Distribution. Then, for*

$$c_{\mathcal{P}} := c_{\mathrm{BE}} \frac{\mathbb{E}[|\,\mathrm{I}(X_i) - \mathrm{H}(X_i)|^3]}{\sigma_{\mathrm{I}}^3}, \tag{1}$$

*it holds true that*

$$q - \frac{c_{\mathcal{P}}}{\sqrt{n}} \leq p(\mathcal{C}_q^n) \leq q + \frac{c_{\mathcal{P}}}{\sqrt{n}}.$$

*Proof.* We prove the statement by showing that

$$|p(\mathcal{C}_q^n) - q| \leq c_{\mathrm{BE}} \frac{\mathbb{E}[|\,\mathrm{I}(X_i) - \mathrm{H}(X_i)|^3]}{\sigma_{\mathrm{I}}^3 \sqrt{n}}.$$

Let $Y_n = \sum_{i=1}^n \mathrm{I}(X_i)$ and let $\mathbf{X}_n = (X_1, \dots, X_n)$. Then, we find

$$\begin{aligned}
&p(\mathcal{C}_q^n) \\
&= \Pr[\mathbf{X}_n \in \mathcal{C}_q^n] \\
&= \Pr\Big[\prod_{i=1}^n p(X_i) \geq 2^{-\mathrm{H}(X)n - \sigma_{\mathrm{I}} \Phi^{-1}(q)\sqrt{n}}\Big] \\
&= \Pr\Big[\sum_{i=1}^n \mathrm{I}(X_i) \leq \mathrm{H}(X)n + \sigma_{\mathrm{I}} \Phi^{-1}(q)\sqrt{n}\Big] \\
&= \Pr\Big[\frac{Y_n - \mathrm{H}(X)n}{\sigma_{\mathrm{I}}\sqrt{n}} \leq \Phi^{-1}(q)\Big].
\end{aligned}$$

With Lemma 4, we now find

$$\begin{aligned}
|p(\mathcal{C}_q^n) - q| &= \Big|\Pr\Big[\frac{Y_n - \mathrm{H}(X)n}{\sigma_{\mathrm{I}}\sqrt{n}} \leq \Phi^{-1}(q)\Big] - q\Big| \\
&\leq c_{\mathrm{BE}} \frac{\mathbb{E}[|\,\mathrm{I}(X_i) - \mathrm{H}(X_i)|^3]}{\sigma_{\mathrm{I}}^3 \sqrt{n}},
\end{aligned}$$

proving the claim. □

*Remark* 5. For any distribution that satisfies the properties of Lemma 5, there is an intricate relationship between $\mathrm{MG}_n(q)$ and $\mathcal{C}_q^n$. Clearly, the former fixes the success probability in finding the target vector but leaves the amount of vectors to be tested unclear, whereas the latter fixes the amount of vectors to be tested to $\approx 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}} \Phi^{-1}(q)\sqrt{n}}$ but leaves some uncertainty in the success probability. As it turns out, this relationship is what enables us to find an asymptotic expression for $\mathrm{MG}_n(q)$.

In this paper, we also provide a variant of Theorem 1 that requires only the first two properties. The proof is similar to that of Theorem 1 but replaces the need of the Theorem of Berry-Esséen with the Central Limit Theorem. We omit the proof to the appendix.

**Theorem 2.** *Let $X_1, X_2, \dots$ be i.i.d. such that*

- $\mathrm{H}(X_i) < \infty$ *and*
- $0 < \sigma_{\mathrm{I}} < \infty$,

*then*

$$\lim_{n\to\infty} p(\mathcal{C}_q^n) = q.$$

*Remark* 6. Theorem 2 describes a variant of the Asymptotic Equipartition Property (AEP) [11, Theorem 3.1.1] for core sets rather than typical sets.

**Example 3.** Since $\sigma_{\mathrm{I}} = 0$ if and only if $X$ was sampled uniformly, every non-uniform probability distribution over finite support is a BE-distribution.

**Example 4.** Let $X$ be sampled from the geometric distribution $\mathrm{Geo}(p)$, i.e. $x \in \mathbb{N}$ has sampling probability

$$p(x) = (1-p)^{x-1}p,$$

then

$$\mathrm{H}(X) = \frac{-p\log p - (1-p)\log(1-p)}{p} = \frac{\mathrm{H}(p)}{p} < \infty.$$

Furthermore, the series

$$\begin{aligned}
&\mathbb{E}[|\,\mathrm{I}(X) - \mathbb{E}[\mathrm{I}(X)]|^k] \\
&= \sum_{x=1}^{\infty} p(x)|\,\mathrm{I}(x) - \mathrm{H}(X)|^k \\
&= p\log^k \frac{1}{1-p} \sum_{x=1}^{\infty} (1-p)^x \Big| x - \frac{1-p}{p}\Big|^k
\end{aligned}$$

converges for any $k \in \mathbb{N}$, so $\sigma_{\mathrm{I}}^2, \mathbb{E}[|\,\mathrm{I}(X) - \mathbb{E}[\mathrm{I}(X)]|^3] < \infty$, and $\mathrm{Geo}(p)$ is a BE-distribution.

**Example 5.** To see that not every distribution is a BE-distribution, let $\ell > 1$ and let $X$ be sampled from the distribution $\mathcal{P}_\ell$ with p.d.f.

$$p(x) = \frac{\frac{1}{x\log^\ell x}}{\sum_{y\geq 2}\frac{1}{y\log^\ell y}} =: c\frac{1}{x\log^\ell x}, \qquad x \geq 2.$$

The series $\sum \frac{1}{y\log^\ell y}$ converges for all $\ell > 1$, so this indeed induces a probability distribution.

Now, if $\ell \leq 2$, then

$$\begin{aligned}
\mathrm{H}(X) &= -\log c + \sum_{x\geq 2} \frac{\log x + \ell\log\log x}{x\log^\ell x} \\
&\geq -\log c + \sum_{x\geq 2} \frac{1}{x\log^{\ell-1} x}.
\end{aligned}$$

Clearly, this series is divergent. With a similar agument, for $\ell \leq 3$, the distribution has diverging variance in information and, for $\ell \leq 4$, the third absolute central moment of $\mathrm{I}(X)$ diverges. Hence, for any $1 < \ell \leq 4$, $\mathcal{P}_\ell$ is *not* a BE-distribution.

From Example 5, it is readily apparent that any distribution where

$$p(x) = \mathcal{O}\Big(\frac{1}{x\log^\ell x}\Big) \qquad \text{for some } \ell > 4$$

yields a BE-distribution. Thus, we do not consider the exclusion of any non-BE-distribution to be a significant limitation.

## III. Bounding the Marginal Guesswork

As we have stated previously, $\mathcal{C}_q^n$ and $\mathrm{MG}_n(q)$ share a relationship where one fixes what the other leaves ambiguous and vice versa. Unsurprisingly, in this Section, we will find that, for most distributions,

$$\mathrm{MG}_n(q) \approx |\mathcal{C}_q^n|.$$

The advantage we have now is that we are actually able to calculate some upper and lower bound for $|\mathcal{C}_q^n|$, bar some constant factors that depend on our choice of $q$. Once we know the size of $\mathcal{C}_q^n$, we can then proceed to find $\mathrm{MG}_n(q)$.

**Theorem 3** (The size of $\mathcal{C}_q^n$)**.** *Let $0 < q < 1$ and let $\mathcal{C}_q^n$ be the core set for some random vector $\mathbf{X}$ with coordinates sampled from $\mathcal{P}$.*

*1)* $|\mathcal{C}_q^n| \leq 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}}$.

*2) If $\mathcal{P}$ is a BE-distribution, then there exists some constants $r_q > 0$, $n_q \in \mathbb{N}$ such that, for all $n \geq n_q$,*

$$|\mathcal{C}_q^n| \geq \frac{r_q}{\sqrt{n}} 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}}.$$

*3) If $\mathcal{P}$ is a BE-distribution, then there exists some constants $R_q > 0, N_q \in \mathbb{N}$ such that, for all $n \geq N_q$,*

$$|\mathcal{C}_q^n| \leq \frac{R_q \log n}{\sqrt{n}} 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}}.$$

*All constants $r_q, R_q, n_q, N_q$ depend on the distribution and on the choice of $q$ alone.*

*Proof.* To prove the first statement, consider

$$1 = \sum_{\mathbf{x} \in \mathcal{X}^n} p(\mathbf{x}) \geq \sum_{\mathbf{x} \in \mathcal{C}_q^n} p(\mathbf{x}) \geq |\mathcal{C}_q^n| 2^{-\mathrm{H}(X)n - \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}}.$$

Multiplying both sides with $2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}}$ proves the claim.

For the second and third statement, let $c_{\mathcal{P}}$ be as in Equation (1). According to Theorem 1, we have the chain of inequalities

$$q + \frac{c_{\mathcal{P}}}{\sqrt{n}} \geq p(\mathcal{C}_q^n) \geq q - \frac{c_{\mathcal{P}}}{\sqrt{n}}.$$

In particular, for *any two* $0 < q' < q < 1$, this yields

$$q - q' + 2\frac{c_{\mathcal{P}}}{\sqrt{n}} \geq p(\mathcal{C}_q^n \backslash \mathcal{C}_{q'}^n) \geq q - q' - 2\frac{c_{\mathcal{P}}}{\sqrt{n}}. \tag{2}$$

We now prove the second statement. To this end, let

$$q_n = q - \frac{2c_{\mathcal{P}} + 1}{\sqrt{n}}.$$

For

$$n_q = \min\left\{ n \in \mathbb{N} \middle| q > \frac{2c_{\mathcal{P}} + 1}{\sqrt{n}} \right\},$$

we find $0 < q_n < q$. Now, consider

$$\begin{aligned}
\frac{1}{\sqrt{n}} &\leq p(\mathcal{C}_q^n \backslash \mathcal{C}_{q_n}^n) \\
&= \sum_{\mathbf{x} \in \mathcal{C}_q^n \backslash \mathcal{C}_{q_n}^n} p(\mathbf{x}) \\
&\leq \sum_{\mathbf{x} \in \mathcal{C}_q^n \backslash \mathcal{C}_{q_n}^n} 2^{-\mathrm{H}(X)n - \sigma_{\mathrm{I}}\Phi^{-1}(q_n)\sqrt{n}} \\
&= |\mathcal{C}_q^n \backslash \mathcal{C}_{q_n}^n| 2^{-\mathrm{H}(X)n - \sigma_{\mathrm{I}}\Phi^{-1}(q_n)\sqrt{n}} \\
&\leq |\mathcal{C}_q^n| 2^{-\mathrm{H}(X)n - \sigma_{\mathrm{I}}\Phi^{-1}(q_n)\sqrt{n}}.
\end{aligned} \tag{3}$$

Multiplying both sides with $2^{-\mathrm{H}(X)n - \sigma_{\mathrm{I}}\Phi^{-1}(q_n)\sqrt{n}}$ yields

$$\begin{aligned}
|\mathcal{C}_q^n| &\geq \frac{1}{\sqrt{n}} 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q_n)\sqrt{n}} \\
&\geq \frac{2^{-\sigma_{\mathrm{I}}(\Phi^{-1}(q) - \Phi^{-1}(q_n))\sqrt{n}}}{\sqrt{n}} 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}}.
\end{aligned}$$

It remains to show that

$$\inf_{n \geq n_q} 2^{-\sigma_{\mathrm{I}}(\Phi^{-1}(q) - \Phi^{-1}(q_n))\sqrt{n}} > 0,$$

this immediately follows from the identity

$$\lim_{n \to \infty} 2^{-\sigma_{\mathrm{I}}(\Phi^{-1}(q) - \Phi^{-1}(q_n))\sqrt{n}} = 2^{-\sigma_{\mathrm{I}}(2c_{\mathcal{P}} + 1)\sqrt{2\pi} e^{\frac{\Phi^{-1}(q)^2}{2}}}.$$

We prove this equation with the help of Lemma 10 in the Appendix (where $f(n) = 2c_{\mathcal{P}} + 1$).

In order to prove the third statement, let us consider Equation (2) one more time. Let

$$b_q = \frac{1}{\sqrt{2\pi}\sigma_{\mathrm{I}}} e^{\frac{\Phi^{-1}(q)^2}{2}} \qquad \text{and} \qquad q_n' = q - \frac{b_q \log n}{\sqrt{n}}.$$

If $n \in \mathbb{N}$ is such that $q_n > 0$ then, with a chain of inequalities similar to that of Equation (3), we find that

$$|\mathcal{C}_q^n \backslash \mathcal{C}_{q'_n}^n| 2^{-\,\mathrm{H}(X)n - \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}} \leq \frac{(2c_{\mathcal{P}} + b_q)\log n}{\sqrt{n}}.$$

This inequality can be rearranged to

$$|\mathcal{C}_q^n| \leq \frac{(2c_{\mathcal{P}} + b_q)\log n}{\sqrt{n}} 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}} + |\mathcal{C}_{q'_n}^n|.$$

We want to show that, for $n$ sufficiently large,

$$\frac{(2c_{\mathcal{P}} + b_q)\log n}{\sqrt{n}} 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}} \geq |\mathcal{C}_{q'_n}^n|. \tag{4}$$

According to Lemma 10 (with $f(n) = b_q \log n$) and the first statement of Theorem 3, there exists some value $\xi_n \in [q'_n, q]$,

$$\begin{aligned}
|\mathcal{C}_{q'_n}^n| &\leq 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q'_n)\sqrt{n}} \\
&= 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}} 2^{-\sigma_{\mathrm{I}}(\Phi^{-1}(q) - \Phi^{-1}(q'_n))\sqrt{n}} \\
&= 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}} 2^{-\sigma_{\mathrm{I}} b_q \sqrt{2\pi} e^{-\frac{\Phi^{-1}(\xi_n)^2}{2}} \log n} \\
&= 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}} 2^{-e^{\frac{\Phi^{-1}(q)^2 - \Phi^{-1}(\xi_n)^2}{2}} \log n} \\
&= \frac{1}{n^{e^{\frac{\Phi^{-1}(q)^2 - \Phi^{-1}(\xi_n)^2}{2}}}} 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}}.
\end{aligned}$$

Hence, Equation (4) is guaranteed to hold true if

$$\log n \geq \frac{1}{2c_{\mathcal{P}} + b_q} \qquad \text{and} \qquad \frac{\Phi^{-1}(q)^2 - \Phi^{-1}(\xi_n)^2}{2} \geq -\ln 2.$$

Since $\xi_n \xrightarrow{n\to\infty} q$, the second inequality is guaranteed to hold true eventually, so if we set

$$N_q := \max\left\{ n \in \mathbb{N} \;\middle|\; \begin{array}{rrl} \frac{\Phi^{-1}(q)^2 - \Phi^{-1}(\xi_n)^2}{2} & & < -\ln 2 \\ \text{or} & \log n & \leq \frac{1}{2c_{\mathcal{P}} + b_q} \\ \text{or} & q'_n & \leq 0 \end{array} \right\} + 1,$$

then, for $n \geq N_q$ and

$$R_q := 2(2c_{\mathcal{P}} + b_q) = 2\Big(2c_{\mathcal{P}} + \frac{1}{\sqrt{2\pi}\sigma_{\mathrm{I}}} e^{\frac{\Phi^{-1}(q)^2}{2}}\Big),$$

$\mathcal{C}_q^n$ is bound by

$$\begin{aligned}
|\mathcal{C}_q^n| &\leq \frac{(2c_{\mathcal{P}} + b_q)\log n}{\sqrt{n}} 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}} + |\mathcal{C}_{q'_n}^n| \\
&\leq \frac{R_q \log n}{\sqrt{n}} 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}}.
\end{aligned}$$

□

*Remark* 7. The first statement was already proven in [9]. Furthermore, for typical sets, there exist a similar statement that uses the exact same proof idea, see e.g. [11, Theorem 3.1.2.].

*Remark* 8. Note that our lower bound for $r$ can become very small as $q$ tends to 0 as the constant is essentially doubly exponential in $-\Phi^{-1}(q)^2$. Similarly, the upper bound $R$ is exponential in $\Phi^{-1}(q)^2$ and can become very large when $q$ is close to 0 or 1. In Figure 1, we can see that these bounds are not necessarily tight and that, in some cases, $|\mathcal{C}_q^n| \approx \frac{1}{\sqrt{n}} 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}}$ already yields a sufficient approximation, although we believe that there do not exist *general* bounds for $r_q, R_q$ that are tighter than the ones we have derived in Theorem 3.

Before we move on to replace the size of $\mathcal{C}^n$ with the marginal guesswork, we need to generalize the statement of Theorem 3 to a broader statement. In particular, we need to show that the statement of Theorem 3 holds true for *all* values of $q$ in a fixed range, as long as this range may not get arbitrarily close to 0 or 1.

**Lemma 6.** *Let $\mathcal{P}$ be a BE-distribution. For some $0 < q^- < q^+ < 1$, there exists $r, R > 0, n_0, N_0 \in \mathbb{N}$ such that, for all $q \in [q^-, q^+]$,*

$$|\mathcal{C}_q^n| \geq \frac{r}{\sqrt{n}} 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}} \qquad \textit{for all } n \geq n_0$$

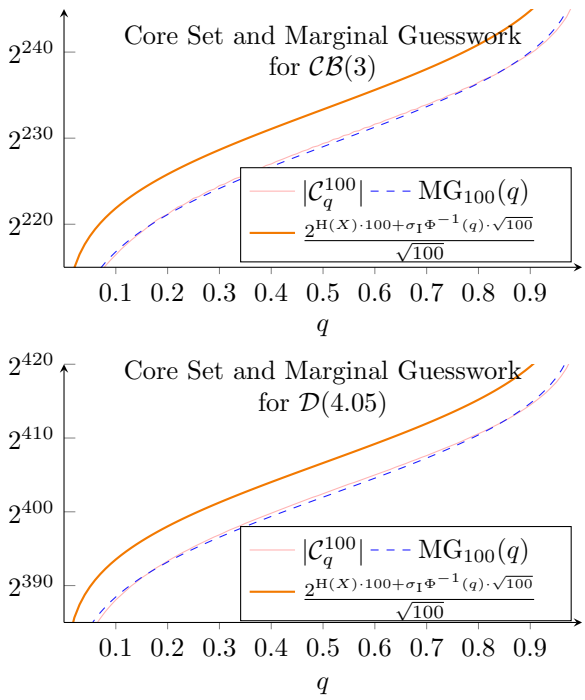


Fig. 1. Size of $\mathcal{C}^n$ and $\mathrm{MG}_n(q)$ for $n = 100$ and distributions $\mathcal{CB}(3)$ and $\mathcal{D}(4.05)$, the distribution used in KYBER512 [15] and FALCON512 [16], respectively. The graphs verify the results from Theorem 3, namely that $|\mathcal{C}_q^n| \approx \mathrm{MG}_n(q) \approx \frac{1}{\sqrt{n}} 2^{\mathrm{H}(X)n+\sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}}$.

*and*

$$|\mathcal{C}_q^n| \leq \frac{R \log n}{\sqrt{n}} 2^{\mathrm{H}(X)n+\sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}} \qquad \textit{for all } n \geq N_0.$$

*Proof.* We only prove the upper bound, the lower bound can be proven analogously.

For $q \in [q^-, q^+]$, let $R_q$ be as in Theorem 3 and let

$$R := \sup_{q\in[q^-,q^+]} R_q = \sup_{q\in[q^-,q^+]} 2\Big(2c_{\mathcal{P}} + \frac{1}{\sqrt{2\pi}\sigma_{\mathrm{I}}} 2^{\frac{\Phi^{-1}(q)^2}{2}}\Big).$$

Since $\Phi^{-1}(q)^2$ is convex, the supremum is attained at either $q^-$ or $q^+$, so $R$ is indeed finite as $q^-, q^+ \neq 0, 1$.

Now, we fix $q \in [q^-, q^+]$ and let

$$b_q = \frac{1}{\sqrt{2\pi}\sigma_{\mathrm{I}}} e^{\frac{\Phi^{-1}(q)^2}{2}}.$$

Remember that, in Theorem 3, we have shown that, if $n$ is such that

- $$\frac{\Phi^{-1}(q)^2 - \Phi^{-1}(\xi_n)^2}{2} \geq -\ln 2$$

  for $\xi_n$ such that

$$\Phi'(\Phi^{-1}(\xi_n)) = \left(\Phi^{-1}(q) - \Phi^{-1}\Big(q - \frac{b_q \log n}{\sqrt{n}}\Big)\right)\sqrt{n},$$

- $n > 2^{\frac{1}{2c_{\mathcal{P}}+b_q}}$ and
- $q - \frac{b_q \log n}{\sqrt{n}} > 0$,

then $R_q$ satisfies

$$|\mathcal{C}_q^n| \leq \frac{R_q \log n}{\sqrt{n}} 2^{\mathrm{H}(X)n+\sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}}.$$

We also have defined $N_q$ such that no $n \geq N_q$ violates any of these conditions.

Now, let

$$N_0 := \sup_{q\in[q^-,q^+]} N_q,$$

i.e. the above conditions are satisfied for all $q \in [q^-, q^+]$, $n \geq N_0$. Thus, if $q \in [q^-, q^+], n \geq N_0 \geq N_q$, then

$$\begin{aligned}|\mathcal{C}_q^n| &\leq \frac{R_q \log n}{\sqrt{n}} 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}} \Phi^{-1}(q)\sqrt{n}} \\ &\leq \frac{R \log n}{\sqrt{n}} 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}} \Phi^{-1}(q)\sqrt{n}},\end{aligned}$$

so if we can prove that $N_0 < \infty$, we are done.

Regarding the third condition, we clearly have

$$\begin{aligned} & q - \frac{\frac{1}{\sqrt{2\pi}\sigma_{\mathrm{I}}} e^{\frac{\Phi^{-1}(q)^2}{2}} \log n}{\sqrt{n}} && > 0 \\ \Leftrightarrow \quad & q\sqrt{2\pi}\sigma_{\mathrm{I}} e^{-\frac{\Phi^{-1}(q)^2}{2}} && > \tfrac{\log n}{\sqrt{n}}. \end{aligned}$$

The infimum over $[q^-, q^+]$ of the L.H.S. is greater than 0 as $q \geq q^- > 0$. Hence, as the R.H.S. converges to 0, there exists some $N_3$ such that, for $n \geq N_3$, the inequality holds true for all $q \in [q^-, q^+]$.

In a similar vein, it is easy to see that $2^{\frac{1}{2c_{\mathcal{P}} + b_q}}$ is bound from above by

$$2^{\frac{1}{2c_{\mathcal{P}} + b_q}} \leq \left\lceil 2^{\frac{1}{2c_{\mathcal{P}} + \frac{1}{\sqrt{2\pi}\sigma_{\mathrm{I}}}}} \right\rceil =: N_2,$$

so for any $n \geq N_2$, the second condition is satisfied.

It remains to show that, for some $N_1 \in \mathbb{N}$, the first condition is satisfied for all $n \geq N_1, q \in [q^-, q^+]$. We omit the proof for this statement to the Appendix, see Lemma 11.

In conclusion, this proves that

$$N_0 = \sup_{q \in [q^-, q^+]} N_q \leq \max\{N_1, N_2, N_3\} < \infty. \qquad \square$$

We now finish this section by showing that, as long as $q$ can not grow arbitrarily close to 0 or 1, $\mathrm{MG}_n(q)$ grows essentially as fast as $|\mathcal{C}_q^n|$, except for some factor that is at most logarithmic in $n$. In the next section, we show that fixing $q$ is required as otherwise, $\mathrm{MG}_n(q)$ can grow either larger or smaller than $2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}} \Phi^{-1}(q)\sqrt{n}}$ and $|\mathcal{C}_q^n|$.

**Theorem 4.** *Let $\mathcal{P}$ be a BE-distribution.*

*1) For some $0 < q < 1$, there exist some constants $r'_q, n'_q, R'_q, N'_q$ such that*

$$\mathrm{MG}_n(q) \geq \frac{r'_q}{\sqrt{n}} 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}} \Phi^{-1}(q)\sqrt{n}} \quad \textit{for all } n \geq n'_q$$

*and*

$$\mathrm{MG}_n(q) \leq \frac{R'_q \log n}{\sqrt{n}} 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}} \Phi^{-1}(q)\sqrt{n}} \quad \textit{for all } n \geq N'_q.$$

*2) For some constant $0 < q^- \leq q^+ < 1$, there exist some constants $r', n'_0, R', N'_0$ such that, for all $q \in [q^-, q^+]$,*

$$\mathrm{MG}_n(q) \geq \frac{r'}{\sqrt{n}} 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}} \Phi^{-1}(q)\sqrt{n}} \quad \textit{for all } n \geq n'_0$$

*and*

$$\mathrm{MG}_n(q) \leq \frac{R' \log n}{\sqrt{n}} 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}} \Phi^{-1}(q)\sqrt{n}} \quad \textit{for all } n \geq N'_0.$$

*Proof.* We again only prove the upper bound, the lower bound has an analogous proof. Furthermore, the first statement is a special case of the second statement where $q^- = q^+$, so we will only prove the second statement.

Let $\delta > 0$ be arbitrary and let $N''_0 \in \mathbb{N}$ be such that, for all $n \geq N''_0$,

$$0 < q^- - \frac{c_{\mathcal{P}} + \delta}{\sqrt{n}} \quad \text{and} \quad q^+ + \frac{c_{\mathcal{P}} + \delta}{\sqrt{n}} < 1.$$

We also define

$$q^{++} := q^+ + \frac{c_{\mathcal{P}} + \delta}{\sqrt{N''_0}}.$$

According to Lemma 6, there exists some $R, N_0$ such that, for all $q' \in [q^-, q^{++}]$ and $n \geq N_0$,

$$|\mathcal{C}_{q'}^n| \leq \frac{R \log n}{\sqrt{n}} 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}} \Phi^{-1}(q')\sqrt{n}}.$$

For these $R, N_0$, we define

$$R' := R \cdot \sup_{q' \in [q^-, q^{++}]} 2^{\sigma_{\mathrm{I}}(c_{\mathcal{P}}+\delta)\sqrt{2\pi} e^{\frac{\Phi^{-1}(q')^2}{2}}}$$
$$\text{and} \quad N_0' := \max\{N_0, N_0''\}.$$

Clearly, $R'$ and $N_0'$ are finite.

Now, let

$$q_n := q + \frac{c_{\mathcal{P}} + \delta}{\sqrt{n}}.$$

If $q_n < 1$, then, by Theorem 1,

$$p(\mathcal{C}_{q_n}^n) \geq q + \frac{\delta}{\sqrt{n}} > q.$$

By definition of $\mathrm{MG}_n(\cdot)$, this implies

$$\mathrm{MG}_n(q) \leq |\mathcal{C}_{q_n}^n|.$$

If $n \geq N_0'$, then a slight variation of Lemma 10 tells us that there exists some $\xi_n \in [q, q_n] \subseteq [q^-, q^{++}]$ that satisfies

$$\begin{aligned}
&\mathrm{MG}_n(q) \\
&\leq \frac{R \log n}{\sqrt{n}} 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}} 2^{-\sigma_{\mathrm{I}}(\Phi^{-1}(q) - \Phi^{-1}(q_n))\sqrt{n}} \\
&= \frac{R \log n}{\sqrt{n}} 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}} 2^{\sigma_{\mathrm{I}}(c_{\mathcal{P}}+\delta)\sqrt{2\pi} e^{\frac{\Phi^{-1}(\xi_n)^2}{2}}} \\
&\leq \frac{R' \log n}{\sqrt{n}} 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}}.
\end{aligned}$$

□

In the next sections, we will see what happens when we allow $q$ to tend arbitrarily close to 0 or 1. In particular, we will see that the results from this section no longer hold when $q$ is not fixed.

## IV. Lack of Non-Trivial Bounds for Marginal Guesswork

### A. On Upper Bounds...

In Section III, we have proven the existence of a constant $R$ that, for arbitrary $0 < q^- < q^+ < 1$, yields the inequality

$$\mathrm{MG}_n(q) \leq \frac{R \log n}{\sqrt{n}} 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}} \tag{5}$$

for all $q \in [q^-, q^+]$, assuming $n$ was sufficiently large.

In this section, we will prove that, if such a constant were to exist for $q^+ = 1$, then it would have to be of size at least $2^{n \log |\mathcal{X}| - o(n)}$. In particular, the existence of this constant can be disproven if we do not fix $n \in \mathbb{N}$.

We prove this statement in two steps. The first step is that, for *any* distribution $\mathcal{P}$, there exists some value $\zeta$ such that, over some interval $I \subseteq (0, 1)$, $\mathrm{MG}_n(q)$ is bound by

$$\mathrm{MG}_n(q) \gtrapprox 2^{\zeta n} 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}}$$

Then, we show that $\zeta$ can come arbitrarily close to $\log M$ for some distributions, thus showing that Equation (5) is false in general.

For this section, we require the identities

$$\mathbb{E}[G(X)^\rho] = \int_0^1 \mathrm{MG}_n(q)^\rho \, \mathrm{d}q, \tag{6}$$

$$\mathbb{E}[G(X)^\rho \mid G(X) \leq y] = \frac{\int_0^{\Pr[G(\mathbf{X}) \leq y]} \mathrm{MG}_n(q)^\rho \, \mathrm{d}q}{\Pr[G(\mathbf{X}) \leq y]}. \tag{7}$$

A proof of these is provided in Lemma 12 in the Appendix.

*Remark* 9. Let $\mathcal{P}$ be a non-uniform distribution over finite support $\mathcal{X}$. Since $2^{\sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}}$ tends to 0 for $q$ small and $\infty$ for $q$ large, the existence of bounds such that

$$r 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}} \leq \mathrm{MG}_n(q) \leq R 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}}$$

is easily disproven for $q$ sufficiently small or large as naturally, the left inequality can not hold when $q$ is large, and similarly, the right inequality can not hold when $q$ is small.

To this end, we only regard upper bounds w.r.t. values of $q$ taken from intervals $(1-\varepsilon, 1) \subseteq [p(1)^n, 1]$, as $q < p(1)^n$ satisfies $\mathrm{MG}_n(q) = 1$ (or $\mathrm{MG}_n(q) = 0$ if $q = 0$) anyway and does not require any bounds.

In a similar vein, we only regard lower bounds w.r.t. intervals $(0, \varepsilon) \subseteq [0, 1-p(|\mathcal{X}|)^n]$. Again, we note that, if $q > 1-p(|\mathcal{X}|)^n$, then $\mathrm{MG}_n(q) = |\mathcal{X}|^n$, so we do not require any bounds in this domain.

**Lemma 7.** *Let $\mathcal{P}$ be some finite, non-uniform probability distribution and, for some $\rho > 1$, let*

$$\zeta := \mathrm{H}_{\frac{1}{1+\rho}}(X) - \mathrm{H}(X) - \frac{\rho}{2 \log e} \sigma_{\mathrm{I}}^2.$$

*Then, for any $n \in \mathbb{N}, \varepsilon > 0$, there exists some interval $I \subseteq (1-\varepsilon, 1)$ such that, for all $q \in I$,*

$$\mathrm{MG}_n(q) \geq \frac{\varepsilon^{\frac{1}{\rho}} 2^{\zeta n}}{1 + \log M^n} 2^{\mathrm{H}(X) n + \sigma_{\mathrm{I}} \Phi^{-1}(q) \sqrt{n}}.$$

*Proof.* We start by noting that for any real valued, non-decreasing function $f$ that is defined over $[0, 1]$ and for any $\delta \in [0, 1)$, it is intuitive that

$$\int_0^1 f(x) \, \mathrm{d}x \leq \int_\delta^1 \frac{1}{1-\delta} f(x) \, \mathrm{d}x.$$

We provide a short proof in Lemma 13.

Since $\mathrm{MG}_n(q)^\rho$ is non-decreasing for any $\rho > 0$, we may use Equation (6) to find

$$\mathbb{E}[G(X)^\rho] = \int_0^1 \mathrm{MG}_n(q)^\rho \, \mathrm{d}q \leq \int_{1-\varepsilon}^1 \frac{1}{\varepsilon} \mathrm{MG}_n(q)^\rho \, \mathrm{d}q.$$

Let $c := \frac{\rho}{\log e} \sigma_{\mathrm{I}} \sqrt{n}$. By substituting $q = \Phi(t)$, we can express the integral over $2^{\rho(\mathrm{H}(X)n + \sigma_{\mathrm{I}} \Phi^{-1}(q) \sqrt{n})}$ for all $q \in [0, 1]$ as

$$\begin{aligned}
&\int_0^1 2^{\rho(\mathrm{H}(X)n + \sigma_{\mathrm{I}} \Phi^{-1}(q) \sqrt{n})} \, \mathrm{d}q \\
&= 2^{\rho \, \mathrm{H}(X) n} \int_0^1 e^{c \Phi^{-1}(q)} \, \mathrm{d}q \\
&= 2^{\rho \, \mathrm{H}(X) n} \int_{-\infty}^{\infty} e^{ct} \Phi'(t) \, \mathrm{d}t \\
&= 2^{\rho \, \mathrm{H}(X) n} e^{\frac{c^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(t-c)^2}{2}} \, \mathrm{d}t \\
&= 2^{\rho \, \mathrm{H}(X) n} e^{\frac{c^2}{2}} \\
&= 2^{\rho \, \mathrm{H}(X) n + \frac{\rho^2}{2 \log e} \sigma_{\mathrm{I}}^2 n} \\
&= 2^{\rho \, \mathrm{H}_{\frac{1}{1+\rho}}(X) n - \rho \zeta n}.
\end{aligned}$$

With Arikan's inequality, we get

$$\begin{aligned}
&\int_{1-\varepsilon}^1 2^{\rho(\mathrm{H}(X)n + \sigma_{\mathrm{I}} \Phi^{-1}(q) \sqrt{n})} \, \mathrm{d}q \\
&\leq \int_0^1 2^{\rho(\mathrm{H}(X)n + \sigma_{\mathrm{I}} \Phi^{-1}(q) \sqrt{n})} \, \mathrm{d}q \\
&= 2^{\rho \, \mathrm{H}_{\frac{1}{1+\rho}}(X) n - \rho \zeta n} \\
&\leq (1 + \log M^n)^\rho \mathbb{E}[G(X)^\rho] 2^{-\rho \zeta n} \\
&\leq \int_{1-\varepsilon}^1 \frac{(1 + \log M^n)^\rho}{\varepsilon} \mathrm{MG}_n(q)^\rho 2^{-\rho \zeta n} \, \mathrm{d}q.
\end{aligned}$$

Both integrals are taken over the same domain, so there needs to exist some interval $I \subseteq (1-\varepsilon, 1)$ such that, for $q \in I$,

$$2^{\rho(\mathrm{H}(X)n + \sigma_{\mathrm{I}} \Phi^{-1}(q) \sqrt{n})} \leq \frac{(1 + \log M^n)^\rho}{\varepsilon} \mathrm{MG}_n(q)^\rho 2^{-\rho \zeta n}.$$

Multiplying both sides with $\frac{\varepsilon \cdot 2^{\rho \zeta n}}{(1 + \log M^n)^\rho}$ and exponentiating with $\frac{1}{\rho}$ finishes the proof. □

*Remark* 10. The denominator $1 + \log M^n$ has been derived from Arikan [5] and thus is not optimal, as has been brought up by Rioul [4, Corollary 6]. Since either version yields a denominator of the form $\Theta(n \log M)$, we content ourselves with Arikan's original bound.

**Theorem 5.** *For any $n \in \mathbb{N}$ and $\varepsilon, \delta > 0$, there exists some distribution $\mathcal{P}$ such that*

- *$\mathcal{P}$ is defined over a finite support $\mathcal{X}$ of size $M > 2^{\delta}$ and*
- *there exists an interval $I \subseteq (1-\varepsilon, 1)$ such that, for all $q \in I$,*

$$\mathrm{MG}_n(q) \geq \frac{2^{(\log M - \delta)n}}{1 + \log M^n} 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}} \Phi^{-1}(q)\sqrt{n}}.$$

*Proof.* Let $A > 1$ and let $M > 2A - 1$ be an integer such that

$$\log(M-1) \geq \log M - \frac{\delta}{3}, \quad \mathrm{H}\Big(\frac{2(A-1)}{M}\Big) \leq \frac{\delta}{3}$$

For $p = \frac{M-2(A-1)}{M}$, let $\mathcal{P}(A, M)$ be the distribution with support $\mathcal{X} = \{1, \dots, M\}$ and p.d.f.

$$p(1) = p, \qquad p(x) = \frac{1-p}{M-1} \quad \text{for } 2 \leq x \leq M.$$

We want to show that, for $\zeta$ as in Lemma 7, there exists values for $\rho, A, M$ such that

$$\varepsilon^{\frac{1}{\rho}} \cdot 2^{\zeta n} \geq 2^{(\log M - \delta)n}.$$

To that end, we need to give appropriate bounds for $\mathrm{H}_{\frac{1}{1+\rho}}(X), \mathrm{H}(X), \sigma_{\mathrm{I}}^2$.

The Rényi-Entropy of order $\alpha < \frac{1}{2}$ of $\mathcal{P}$ equals

$$\begin{aligned} &\mathrm{H}_{\alpha}(X) \\ &= \frac{\log\left((M-1)\frac{(2(A-1))^{\alpha}}{(M-1)^{\alpha}} + (M - 2(A-1))^{\alpha}\right) - \log(M^{\alpha})}{1-\alpha}, \end{aligned}$$

this term is bound from below by

$$\begin{aligned} \mathrm{H}_{\alpha}(X) &\geq \log(M-1) - \frac{\alpha}{1-\alpha}\log M. \\ &\geq \frac{1-2\alpha}{1-\alpha}\log M - \frac{\delta}{3}. \end{aligned}$$

For $\alpha = \frac{1}{1+\rho}$, this translates to

$$\mathrm{H}_{\frac{1}{1+\rho}}(X) \geq \frac{\rho - 1}{\rho}\log M - \frac{\delta}{3}.$$

On the other hand, we can bound $\mathrm{H}(X)$ and $\sigma_{\mathrm{I}}^2$ by

$$\begin{aligned} \mathrm{H}(X) &= \mathrm{H}(p) + (1-p)\log(M-1), \\ &= \mathrm{H}\Big(\frac{2(A-1)}{M}\Big) + \frac{2(A-1)}{M}\log(M-1) \\ &\leq \frac{\delta}{3} + \frac{2A\log M}{M} \end{aligned}$$

and

$$\begin{aligned} \sigma_{\mathrm{I}}^2 &= p(1-p)\Big(\log\frac{p}{1-p} + \log(M-1)\Big)^2 \\ &= \frac{(M-2(A-1))2(A-1)\log^2\frac{(M-2(A-1))(M-1)}{2(A-1)}}{M^2} \\ &\leq \frac{8A\log^2 M}{M}. \end{aligned}$$

Now, let

$$\begin{aligned} \zeta_{\rho} &:= \mathrm{H}_{\frac{1}{1+\rho}}(X) - \mathrm{H}(X) - \frac{\rho}{2\log e}\sigma_{\mathrm{I}}^2 \\ &\geq \frac{\rho-1}{\rho}\log M - \frac{2A\log M(1 + \frac{2\rho}{\log e}\log M)}{M} - \frac{2\delta}{3}. \end{aligned}$$

For $\rho = \log^2 M$, this expression becomes

$$\zeta_{\rho} \geq \log M - \frac{1}{\log M} - \frac{2A\log M(1 + \frac{2}{\log e}\log^3 M)}{M} - \frac{2\delta}{3},$$

the second and third term vanish when $M$ grows large. Similarly, the term $\varepsilon^{\frac{1}{\rho}}$ comes arbitrarily close to 1 as $\rho$ grows large. Thus, when $M$ is chosen sufficiently large, the expression $\varepsilon^{\frac{1}{\rho}} 2^{\zeta_\rho n}$ from Lemma 7 is bound by

$$\begin{aligned}\varepsilon^{\frac{1}{\rho}} 2^{\zeta_\rho n} &\geq 2^{\frac{\log \varepsilon}{\log^2 M} + \left( \log M - \frac{1}{\log M} - \frac{2A \log M (1 + \frac{2}{\log e} \log^3 M)}{M} - \frac{2\delta}{3} \right) n} \\ &\geq 2^{(\log M - \delta) n}.\end{aligned}$$

Plugging this into Lemma 7 proves the theorem. □

*Remark* 11. For either proof of Lemma 7 and Theorem 5, we have treated $n$ as a constant. In particular, either statement holds true *for any* $n \in \mathbb{N}$. If $n = 1$, we get a statement that extends Pliams's statement [7], namely that $\mathrm{MG}(q)$ can not be expressed in terms of $\mathrm{H}(X)$ and $\sigma_{\mathrm{I}} \cdot \Phi^{-1}(q)$ (whereas Pliam only proved this for $\mathrm{H}(X)$).

On the other hand, if we let $n$ get arbitrarily large while fixing $M$, the statement from Theorem 5 can be simplified a bit to

$$\mathrm{MG}_n(q) \geq 2^{(\log M - \delta) n} 2^{\mathrm{H}(X) n + \sigma_{\mathrm{I}} \Phi^{-1}(q) \sqrt{n}}.$$

In particular, in this case, we find that Theorem 4 can not be generalized for intervals $[q^-, q^+]$ that contain either 0 or 1.

**Corollary 1.** *Let $\mathcal{P}$ be a probability distribution. There does not exist a constant $R$ that depends on $\mathcal{P}$ alone such that, for all $n \in \mathbb{N}, q \in [0, 1]$,*

$$\mathrm{MG}_n(q) \leq R \cdot 2^{\mathrm{H}(X) n + \sigma_{\mathrm{I}} \Phi^{-1}(q) \sqrt{n}}$$

*Remark* 12. Theorem 4 and Theorem 5 work in a complementary way. Whereas the former shows that, for fixed $q$, the approximation $\mathrm{MG}_n(q) \approx \frac{1}{\sqrt{n}} 2^{\mathrm{H}(X) n + \sigma_{\mathrm{I}} \Phi^{-1}(q) \sqrt{n}}$ is tight up to some logarithmic factor, the latter theorem shows that, when fixing $n$ and letting $q$ grow arbitrarily close to 1, the approximation can become arbitrarily inaccurate, to a point where a factor of at least $\approx \frac{M^n}{\log M^n}$ is required to fix the upper bound.

### *B. ... and Lower Bounds*

So far, we have seen that the upper bound from Theorem 4 for fixed $q$ and variable $n$ can not be applied to $\mathrm{MG}_n(q)$ when the roles are reversed, i.e. when $n$ is fixed and when $q$ is variable.

At first glance, the problem as to whether a similar statement can be made *in regards to a lower bound* might appear trivial now, as it appears intuitive to take the statement from Lemma 7 and reverse-engineer a lower bound solution in the same manner. However, note that we are especially interested in lower bounds *in the domain* $(0, \varepsilon)$, as compared to the case of upper bounds where the domain of interest was $(1 - \varepsilon, 1)$. The vast difference here is that, for any $\varepsilon > 0$, we can make a rough estimate

$$\mathbb{E}[G(X)^\rho] \approx \int_{1-\varepsilon}^{1} \mathrm{MG}_n(q)^\rho \, \mathrm{d}q,$$

since the most significant part of this integral – by several orders of magnitudes – lies in the very small interval just to the left of 1. Thus, cutting off almost all of this integral from the left does not impact the size of the integral except for some small factor, which we more than account for by multiplying with $\frac{1}{\varepsilon}$.

In comparison, if we were to follow the same strategy with the *leftmost* part of the integral, we would bound

$$\int_0^{\varepsilon} \mathrm{MG}_n(q)^\rho \, \mathrm{d}q \leq \varepsilon \mathbb{E}[G(X)^\rho] \leq \varepsilon 2^{\rho \, \mathrm{H}_{\frac{1}{1+\rho}}(X) n}.$$

With Theorem 4 in mind, we can immediately make a guess that this inequality suppresses a factor on a scale of roughly $2^{\mathrm{H}_{\frac{1}{1+\rho}}(X) n - \mathrm{H}(X) n - \sigma_{\mathrm{I}} \Phi^{-1}(q) \sqrt{n}} = 2^{\Theta(n)}$, so we have to find a better inequality. We discuss our results in Lemma 9.

*Remark* 13. It is natural to assume that no lower bound for $\mathrm{MG}_n(q)$ can exist that is greater than $\frac{1}{M^n} 2^{\mathrm{H}(X) n + \sigma_{\mathrm{I}} \Phi^{-1}(q) \sqrt{n}}$ and indeed, this is a tight lower bound when $\mathcal{P}$ samples uniformly from $M$ elements since any $r > \frac{1}{M}$ creates the inequality

$$\mathrm{MG}\left(\frac{1}{M}\right) = 1 < rM.$$

However, since $\sigma_{\mathrm{I}} = 0$ in that case, we do not learn anything about whether there exist better lower bound that takes $\sigma_{\mathrm{I}}$ into account and whether or not the uniform distribution should be considered to be the *worst* of the *best* case scenario (or neither). To account for that, we always consider our results w.r.t. non-uniform distributions.

For the upcoming Lemma, we require an approximation on $\frac{1}{\Phi(-x)}$ when $x > 0$. To that end, we use a special case of *Mills Ratio* (see e.g. [17]).

**Lemma 8** (Mills' Ratio for the standard normal distribution)**.** *Let*

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

*and let $x > 0$. Then,*

$$\frac{x}{x^2+1} \leq \frac{\Phi(-x)}{\phi(x)} \leq \frac{1}{x}.$$

**Lemma 9.** *Let $n \in \mathbb{N}$ and let $\mathcal{P}$ be some non-uniform distribution with finite Shannon entropy and variance in information. Then, for*

- *$\rho > 0, c = \frac{\rho}{\log e}\sigma_{\mathrm{I}}\sqrt{n}$ and $0 < \varepsilon < 1$,*
- *$x_0 := c - \Phi^{-1}(\varepsilon)$ and*

$$\overline{m}(\varepsilon,\rho,n) := \begin{cases} \sqrt{2\pi}e^{\frac{x_0^2}{2}}\frac{x_0^2+1}{x_0} & x_0 > 0 \\ 2 & \textit{else} \end{cases}$$

- *some optimal guessing function $G$,*
- *the distribution $\mathcal{P}_\varepsilon^n$ with support*

$$\mathcal{X}_\varepsilon^n = \{\mathbf{x} \in \mathcal{X}^n \mid G(\mathbf{x}) \leq \mathrm{MG}_n(\varepsilon)\},$$

  *p.d.f. $p_\varepsilon(\mathbf{x}) = \frac{p(\mathbf{x})}{p(\mathcal{X}_\varepsilon^n)}$ and Rényi entropy $\mathrm{H}_{\frac{1}{1+\rho}}(\mathbf{X}_\varepsilon)$ and*
- 

$$\nu := \frac{\mathrm{H}_{\frac{1}{1+\rho}}(\mathbf{X}_\varepsilon)}{n} - \mathrm{H}(X) - \frac{\rho}{2\log e}\sigma_{\mathrm{I}}^2,$$

*there exists an interval $I \subseteq (0,\varepsilon)$ such that, for all $q \in I$,*

$$\mathrm{MG}_n(q) \leq p(\mathcal{X}_\varepsilon^n)^{\frac{1}{\rho}}\overline{m}(\varepsilon,\rho,n)^{\frac{1}{\rho}}2^{\nu n}2^{\mathrm{H}(X)n+\sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}}.$$

*Proof.* Similar to the proof of Lemma 7, we want to express $\mathbb{E}[G(\mathbf{X}) \mid G(\mathbf{x}) \leq \mathrm{MG}_n(\varepsilon)]$ in terms of an integral defined over the domain $(0,\varepsilon)$. To this end, let us again substitute $q = \Phi(t)$ to find

$$\begin{aligned}\int_0^\varepsilon 2^{\rho\sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}}\,\mathrm{d}q &= \int_{-\infty}^{\Phi^{-1}(\varepsilon)} e^{ct}\Phi'(t)\,\mathrm{d}t \\ &= e^{\frac{c^2}{2}}\int_{-\infty}^{\Phi^{-1}(\varepsilon)}\frac{1}{\sqrt{2\pi}}e^{\frac{(t-c)^2}{2}}\,\mathrm{d}t \\ &= 2^{\frac{\rho^2}{2\log e}\sigma_{\mathrm{I}}^2 n}\Phi(\Phi^{-1}(\varepsilon)-c) \\ &= 2^{\frac{\rho^2}{2\log e}\sigma_{\mathrm{I}}^2 n}\Phi(-x_0).\end{aligned}$$

For any $\varepsilon > 0$, we find

$$2^{\frac{\rho^2}{2\log e}\sigma_{\mathrm{I}}^2 n} = \frac{1}{\Phi(-x_0)}\int_0^\varepsilon 2^{\rho\sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}}\,\mathrm{d}q.$$

Now, if $x_0 \leq 0$, then $\frac{1}{\Phi(-x_0)} \leq 2$. Otherwise, let

$$m(\varepsilon,\rho,n) := \frac{\Phi(-x_0)}{\phi(x_0)}$$

be the Mills Ratio of $x_0$. Then, we have

$$\frac{1}{\Phi(-x_0)} = \frac{1}{\phi(x_0)m(\varepsilon,\rho,n)},$$

this function is bound from above by $\frac{x_0^2+1}{\phi(x_0)x_0}$, i.e. we get the inequality

$$\frac{1}{\Phi(-x_0)} \leq \sqrt{2\pi}e^{\frac{x_0^2}{2}}\frac{x_0^2+1}{x_0}.$$

In either case, this yields

$$\frac{1}{\Phi(-x_0)} \leq \overline{m}(\varepsilon,\rho,n).$$

Note that, by definition of $\mathcal{X}_\varepsilon^n$, $\varepsilon$ is not greater than $p(\mathcal{X}_\varepsilon^n)$. With Arikan's inequality and Equation (7), we thus find

$$\begin{aligned}
&\int_0^\varepsilon \mathrm{MG}_n(q)^\rho \,\mathrm{d}q \\
&\leq \int_0^{p(\mathcal{X}_\varepsilon^n)} \mathrm{MG}_n(q)^\rho \,\mathrm{d}q \\
&= \Pr[G(\mathbf{x}) \leq \mathrm{MG}_n(\varepsilon))]\mathbb{E}[G(\mathbf{X})^\rho \mid G(\mathbf{x}) \leq \mathrm{MG}_n(\varepsilon)] \\
&= p(\mathcal{X}_\varepsilon^n)\mathbb{E}[G(\mathbf{X}_\varepsilon)^\rho] \\
&\leq p(\mathcal{X}_\varepsilon^n) 2^{\rho\,\mathrm{H}_{\frac{1}{1+\rho}}(\mathbf{X}_\varepsilon)} \\
&= p(\mathcal{X}_\varepsilon^n) 2^{\rho\cdot\nu n} 2^{\rho\,\mathrm{H}(X)n + \frac{\rho^2}{2\log e}\sigma_{\mathrm{I}}^2 n} \\
&= \int_0^\varepsilon p(\mathcal{X}_\varepsilon^n) \frac{1}{\Phi(-x_0)} 2^{\rho\cdot\nu n} 2^{\rho\,\mathrm{H}(X)n + \rho\sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}} \,\mathrm{d}q \\
&\leq \int_0^\varepsilon p(\mathcal{X}_\varepsilon^n) \overline{m}(\varepsilon,\rho,n) 2^{\rho\cdot\nu n} 2^{\rho\,\mathrm{H}(X)n + \rho\sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}} \,\mathrm{d}q.
\end{aligned}$$

The remainder of the proof is analogous to that of Lemma 7. □

*Remark* 14. The inequality $x_0 > 0$ is equivalent to

$$\rho > \frac{\log e}{\sigma_{\mathrm{I}}\sqrt{n}}\Phi^{-1}(\varepsilon),$$

this is guaranteed to happen when we fix $\varepsilon$ and $n$ while $\rho$ grows large. Thus, for the sake of simplicity, we ignore the case where $x_0 \leq 0$ and just assume that

$$\overline{m}(\varepsilon,\rho,n) = \sqrt{2\pi} e^{\frac{x_0^2}{2}} \frac{x_0^2+1}{x_0}.$$

*Remark* 15. If we fix $\varepsilon$ and $n$, then $p(\mathcal{X}_\varepsilon^n)^{\frac{1}{\rho}}$ and most of the terms of $\overline{m}(\varepsilon,\rho,n)^{\frac{1}{\rho}}$ tend to 1 when $\rho$ grows large. In particular, both terms $\sqrt{2\pi}^{\frac{1}{\rho}}$ and $\left(\frac{(c-\Phi^{-1}(\varepsilon))^2+1}{c-\Phi^{-1}(\varepsilon)}\right)^{\frac{1}{\rho}} \sim \rho^{\frac{1}{\rho}}$ are asymptotically equivalent to 1.

For the exponent of $e^{\frac{(c-\Phi^{-1}(\varepsilon))^2}{2\rho}} 2^{\nu n}$ on the other hand, we find

$$\begin{aligned}
&\log e \frac{(c-\Phi^{-1}(\varepsilon))^2}{2\rho} + \nu n \\
&= -\sigma_{\mathrm{I}}\Phi^{-1}(\varepsilon)\sqrt{n} + \log e \frac{\Phi^{-1}(\varepsilon)^2}{2\rho} + \mathrm{H}_{\frac{1}{1+\rho}}(\mathbf{X}_\varepsilon) - \mathrm{H}(X)n.
\end{aligned}$$

For $\rho$ large, $\log e \frac{\Phi^{-1}(\varepsilon)^2}{2\rho}$ vanishes too, so asymptotically in $\rho$, the inequality becomes

$$\mathrm{MG}_n(q) \leq (1+o(1)) 2^{\sigma_{\mathrm{I}}(\Phi^{-1}(q)-\Phi^{-1}(\varepsilon))\sqrt{n} + \mathrm{H}_{\frac{1}{1+\rho}}(\mathbf{X}_\varepsilon)}. \tag{8}$$

*Remark* 16. It is important to note that the bounds of $I$ depend on our choice of $\varepsilon$ and $\rho$ and can vanish for $\rho = \infty$. For example, if $\varepsilon$ is such that $\mathrm{MG}_n(\varepsilon) = 1$, then Equation (8) becomes

$$1 = \mathrm{MG}_n(q) \leq \mathrm{MG}_n(\varepsilon) \cdot 2^{\sigma_{\mathrm{I}}(\Phi^{-1}(q)-\Phi^{-1}(\varepsilon))\sqrt{n}},$$

this inequality is false for all $0 < q < \varepsilon$.

*Remark* 17. In Theorem 5 we have seen that some distributions do not allow for upper bounds of the form

$$\mathrm{MG}_n(q) \lessapprox 2^{\log M^n(1-\delta)} 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}}.$$

For *lower* bounds, we can only prove that this is the case for $n = 1$ as we do not have appropriate bounds for $\mathrm{H}_{\frac{1}{1+\rho}}(\mathbf{X}_\varepsilon)$, other than

$$\mathrm{H}_{\frac{1}{1+\rho}}(\mathbf{X}_\varepsilon) \leq \log \mathrm{MG}_n(\varepsilon).$$

In particular, in the next theorem, we will deliberately choose $\mathcal{P}_\varepsilon$ such that the element with the highest sampling probability is sampled with probability $\varepsilon$, i.e.

$$\mathcal{X}_\varepsilon = \{1\}, \qquad \mathrm{H}_{\frac{1}{1+\rho}}(X_\varepsilon) = \log \mathrm{MG}(\varepsilon) = 0.$$

This equality gets lost when we have to deal with random vectors instead. On a side note, this also means that we can not disprove the existence of $r > 0$ such that

$$\mathrm{MG}_n(q) \geq \frac{r}{\sqrt{n}} 2^{\mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}}, \quad q \geq p(1)^n.$$

**Theorem 6.** *Let* $1 > \varepsilon > 0, \delta > 0$ *and let*

$$f(\varepsilon) = 1 - \varepsilon + \sqrt{\varepsilon(1-\varepsilon)}\Phi^{-1}(\varepsilon).$$

*There exists a distribution* $\mathcal{P}$ *such that*

- $\mathcal{P}$ *is defined over a finite support of size* $M$ *and*
- *there exists an interval* $I \subseteq (0, \varepsilon)$ *such that, for* $q \in I$,

$$\mathrm{MG}(q) \leq 2^{(\delta - f(\varepsilon))\log M} 2^{\mathrm{H}(X) + \sigma_{\mathrm{I}}\Phi^{-1}(q)}.$$

*Proof.* Let $M > \frac{1}{\varepsilon} > 1$ be an integer and $0 < \delta' \leq \delta$ be such that

$$(\delta' - f(\varepsilon))\log(M-1) \leq (\delta - f(\varepsilon))\log M \tag{9}$$

$$\text{and} \qquad \varepsilon \geq \frac{1-\varepsilon}{M-1}$$

Let $\mathcal{P}(\varepsilon, M)$ be the distribution with support $\mathcal{X} = \{1, \ldots, M\}$ and p.d.f.

$$p(1) = \varepsilon, \quad p(x) = \frac{1-\varepsilon}{M-1} \qquad \text{for } 2 \leq x \leq M.$$

This distribution has Shannon entropy and variance in information

$$\mathrm{H}(X) = \mathrm{H}(\varepsilon) + (1-\varepsilon)\log(M-1)$$
$$\sigma_{\mathrm{I}}^2 = \varepsilon(1-\varepsilon)\log^2\frac{\varepsilon(M-1)}{1-\varepsilon}.$$

Since $p(1) = \varepsilon$, for all $0 < q < \varepsilon$, it holds true that

$$\mathrm{MG}_n(q) = 1.$$

Let $\mathcal{P}_\varepsilon, \mathcal{X}_\varepsilon$ be as in Lemma 9. We find that $\mathcal{X}_\varepsilon = \{1\}$, so the distribution $\mathcal{P}_\varepsilon$ has Rényi entropy

$$\mathrm{H}_\alpha(X_\varepsilon) = 0 \qquad \text{for all } \alpha \geq 0.$$

Let $c, x_0, \nu, \overline{m}$ be as in Lemma 9. Let $I \subseteq (0, \varepsilon)$ be such that, for all $q \in I$,

$$\mathrm{MG}(q) \leq p(1)^{\frac{1}{\rho}}\overline{m}(\varepsilon, \rho, 1)^{\frac{1}{\rho}} 2^{\nu} 2^{\mathrm{H}(X) + \sigma_{\mathrm{I}}\Phi^{-1}(q)}.$$

According to Remark 15, we may choose $\rho > \frac{\log e}{\sigma_{\mathrm{I}}}\Phi^{-1}(\varepsilon)$ such that

$$\left(p(1)\sqrt{2\pi}e^{\frac{\Phi^{-1}(\varepsilon)^2}{2}}\frac{x_0^2+1}{x_0}\right)^{\frac{1}{\rho}} \leq 2^{\frac{\delta'}{2}} \leq 2^{\frac{\delta'}{2}\log(M-1)}.$$

This yields the upper bound

$$p(1)^{\frac{1}{\rho}}\overline{m}(\varepsilon, \rho, 1)^{\frac{1}{\rho}} 2^{\nu} \leq 2^{-\mathrm{H}(X) - \sigma_{\mathrm{I}}\Phi^{-1}(\varepsilon) + \frac{\delta'}{2}\log(M-1)}.$$

It remains to show that, if $M$ is sufficiently large, then

$$-\mathrm{H}(X) - \sigma_{\mathrm{I}}\Phi^{-1}(\varepsilon) \leq \left(\frac{\delta'}{2} - f(\varepsilon)\right)(\log M - 1).$$

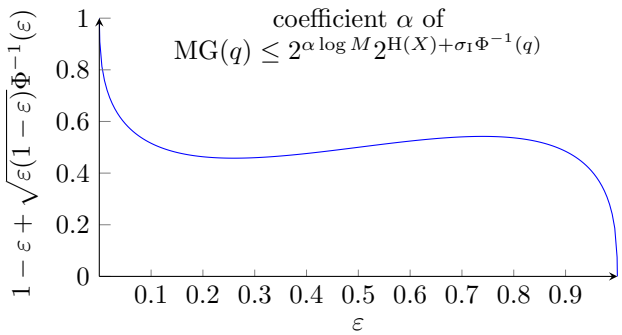


Fig. 2. The auxiliary function $f(\varepsilon)$ as discussed in Theorem 6. The graph tends to 1 as $\varepsilon$ tends to 0, so for any $\delta > 0$, there exists a distribution $\mathcal{P}$ and some $\varepsilon > 0$ such that, for some $q < \varepsilon$, we have $\mathrm{MG}(q) \leq 2^{(\delta-1)\log M} 2^{\mathrm{H}(X) + \sigma_{\mathrm{I}}\Phi^{-1}(q)}$.

To that end, consider

$$\begin{aligned}
&-\mathrm{H}(X) - \sigma_{\mathrm{I}}\Phi^{-1}(\varepsilon) \\
=&-\mathrm{H}(\varepsilon) - (1-\varepsilon)\log(M-1) \\
&-\sqrt{\varepsilon(1-\varepsilon)}\Big(\log(M-1) + \Big|\log\frac{\varepsilon}{1-\varepsilon}\Big|\Big)\Phi^{-1}(\varepsilon) \\
=&-\mathrm{H}(\varepsilon) - \sqrt{\varepsilon(1-\varepsilon)}\Phi^{-1}(\varepsilon)\Big|\log\frac{\varepsilon}{1-\varepsilon}\Big| \\
&-f(\varepsilon)\log(M-1) \\
=&-\frac{\mathrm{H}(\varepsilon) + \sqrt{\varepsilon(1-\varepsilon)}\Phi^{-1}(\varepsilon)|\log\frac{\varepsilon}{1-\varepsilon}|}{\log(M-1)}\log(M-1) \\
&-f(\varepsilon)\log(M-1).
\end{aligned}$$

The first expression tends to

$$-\frac{\mathrm{H}(\varepsilon) + \sqrt{\varepsilon(1-\varepsilon)}\Phi^{-1}(\varepsilon)|\log\frac{\varepsilon}{1-\varepsilon}|}{\log(M-1)} \xrightarrow{M\to\infty} 0,$$

so if $M$ is sufficiently large, then the term bound by

$$-\frac{\mathrm{H}(\varepsilon) + \sqrt{\varepsilon(1-\varepsilon)}\Phi^{-1}(\varepsilon)|\log\frac{\varepsilon}{1-\varepsilon}|}{\log(M-1)} \leq \frac{\delta'}{2}.$$

Overall, with Lemma 9 and Equation (9), this yields the desired bound

$$\begin{aligned}
\mathrm{MG}(q) &\leq p(1)^{\frac{1}{\rho}}\overline{m}(\varepsilon,\rho,1)^{\frac{1}{\rho}}2^{\nu}2^{\mathrm{H}(X)+\sigma_{\mathrm{I}}\Phi^{-1}(q)} \\
&\leq 2^{(\delta'-f(\varepsilon))\log(M-1)}2^{\mathrm{H}(X)+\sigma_{\mathrm{I}}\Phi^{-1}(q)} \\
&\leq 2^{(\delta-f(\varepsilon))\log M}2^{\mathrm{H}(X)+\sigma_{\mathrm{I}}\Phi^{-1}(q)}.
\end{aligned}$$

□

*Remark* 18. The function $f(\varepsilon)$ tends to 1 as $\varepsilon$ tends to 0 (see Figure 2), so the theorem statement can be rephrased such that, for any $\delta > 0$, there exists some distribution $\mathcal{P}, \varepsilon > 0$ and some interval $I \subseteq (0,\varepsilon)$ such that, for all $q \in I$,

$$\mathrm{MG}(q) \leq 2^{(\delta-1)\log M}2^{\mathrm{H}(X)+\sigma_{\mathrm{I}}\Phi^{-1}(q)}.$$

## V. Discussion

Before we conclude this paper, we would like to present some questions that arose over the past few sections and have yet to be answered.

*Problem* 1. Let $\mathcal{P}$ be any BE-distribution and let $0 < q < 1$. Theorem 3 shows the existence of constants $r_q, R_q$ such that $|\mathcal{C}_q^n|$ is bound by

$$\frac{r}{\sqrt{n}}2^{\mathrm{H}(X)n+\sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}} \leq |\mathcal{C}_q^n| \leq \frac{R\log n}{\sqrt{n}}2^{\mathrm{H}(X)n+\sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}}.$$

These constants only depend on $\sigma_{\mathrm{I}}, \mathbb{E}[|\,\mathrm{I}(X) - \mathrm{H}(X)|^3]$ and $\Phi^{-1}(q)$ and are of the form

$$r_q \approx 2^{-\sigma_{\mathrm{I}}(2c_{\mathcal{P}}+1)\sqrt{2\pi}e^{\frac{\Phi^{-1}(q)^2}{2}}}, \quad R_q = 2\Big(2c_{\mathcal{P}} + \frac{1}{\sqrt{2\pi}}e^{\frac{\Phi^{-1}(q)^2}{2}}\Big).$$

However, we have seen, e.g. in Figure 1, that these bounds are not always tight, so we wonder as to whether there exists a general formula that expresses $r_q, R_q$ more moderately while still maintaining the inequalities from Theorem 3.

With Theorem 4, a similar question can be asked w.r.t. $\mathrm{MG}_n(q)$.

*Problem* 2. Let $\mathcal{P}$ be a distribution over finite support of size $M$. Theorem 5 and Theorem 6 show that, if constants $r, R$ exist such that

$$r2^{\mathrm{H}(X)+\sigma_{\mathrm{I}}\Phi^{-1}(q)} \leq \mathrm{MG}(q) \leq R2^{\mathrm{H}(X)+\sigma_{\mathrm{I}}\Phi^{-1}(q)}, \tag{10}$$

then these constants can be arbitrarily close to

$$r \approx \frac{1}{M}, \qquad R \approx M.$$

However, we were unable to find examples where Equation (10) was incorrect for $r = \frac{1}{M}, R = M$, and we assume that, for these $r, R$, the inequality chain is always satisfied. More formally, we assume that the following Conjecture holds true.

**Conjecture 1.** *Let $\mathcal{P}$ be a distribution over support $\mathcal{X}$ of size $M$ with p.d.f. $p$ and let*

$$\min_{\mathbf{x}\in\mathcal{X}^n} p(\mathbf{x}) \leq q \leq \max_{\mathbf{x}\in\mathcal{X}^n} p(\mathbf{x}),$$

*then, for any $n \in \mathbb{N}$,*

$$\frac{1}{M^n} 2^{\mathrm{H}(X)n+\sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}} \leq \mathrm{MG}_n(q) \leq M^n 2^{\mathrm{H}(X)n+\sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}}.$$

*Remark* 19. Conjecture 1 removes the critical ranges $(0,\varepsilon)$ and $(1-\varepsilon,1)$ for lower and upper bounds, respectively. In particular, if $q, q'$ are such that

$$\begin{aligned} 0 &\leq \mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n} \\ \log M^n &\geq \mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q')\sqrt{n}, \end{aligned}$$

then it immediately holds true that

$$\begin{aligned} \mathrm{MG}_n(q) \leq M^n \leq M^n 2^{\mathrm{H}(X)n+\Phi^{-1}(q)\sqrt{n}} \\ \text{and} \quad \mathrm{MG}_n(q') \geq 1 \geq \frac{1}{M^n} 2^{\mathrm{H}(X)n+\sigma_{\mathrm{I}}\Phi^{-1}(q')\sqrt{n}}. \end{aligned}$$

Thus, in order to prove Conjecture 1, it is sufficient to prove it for *the remaining, intermediate* values of $q$.

*Problem* 3. The last problem we were unable to answer is as to whether Corollary 1 can be formulated w.r.t. lower bounds. More precisely, we assume that the following conjecture also holds true.

**Conjecture 2.** *Let $\mathcal{P}$ be a probability distribution. There does not exist a constant $r > 0$ that depends on $\mathcal{P}$ alone such that, for all $n \in \mathbb{N}, q \in [\min p(\mathbf{x}), \max p(\mathbf{x})]$,*

$$\mathrm{MG}_n(q) \geq \frac{r}{\sqrt{n}} \cdot 2^{\mathrm{H}(X)n+\sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}}.$$

## APPENDIX

*Proof of Theorem 2.* Let $Y_n = \sum_{i=1}^n \mathrm{I}(X_i)$ and let $\mathbf{X}_n = (X_1, \ldots, X_n)$. Via Central Limit Theorem, we find

$$\begin{aligned} & p(\mathcal{C}_q^n) \\ &= \Pr[\mathbf{X}_n \in \mathcal{C}_q^n] \\ &= \Pr\Big[\prod_{i=1}^n p(X_i) \geq 2^{-\mathrm{H}(X)n-\sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}}\Big] \\ &= \Pr\Big[\sum_{i=1}^n \mathrm{I}(X_i) \leq \mathrm{H}(X)n + \sigma_{\mathrm{I}}\Phi^{-1}(q)\sqrt{n}\Big] \\ &= \Pr\Big[\frac{Y_n - \mathrm{H}(X)n}{\sigma_{\mathrm{I}}\sqrt{n}} \leq \Phi^{-1}(q)\Big] \\ &\xrightarrow{n\to\infty} \Phi(\Phi^{-1}(q)) = q. \end{aligned}$$

□

**Lemma 10.** *Let $0 < q < 1$ and let $q_n = q - \frac{f(n)}{\sqrt{n}}$ where $f(n)$ is non-negative and such that*

$$q_n > 0, \qquad \lim_{n\to\infty} \frac{f(n)}{\sqrt{n}} = 0.$$

*Then, there exists some $q_n \leq \xi_n \leq q$ such that*

$$(\Phi^{-1}(q) - \Phi^{-1}(q_n))\sqrt{n} = f(n)\sqrt{2\pi} e^{\frac{\Phi^{-1}(\xi_n)^2}{2}}.$$

*Furthermore, if $\lim_{n\to\infty} f(n) = C$, then*

$$\lim_{n\to\infty} (\Phi^{-1}(q) - \Phi^{-1}(q_n))\sqrt{n} = C\sqrt{2\pi} e^{\frac{\Phi^{-1}(q)^2}{2}}.$$

*Proof.* We may rephrase the expression via

$$\begin{aligned} (\Phi^{-1}(q) - \Phi^{-1}(q_n))\sqrt{n} &= f(n)\frac{\Phi^{-1}(q) - \Phi^{-1}(q_n)}{q - (q - \frac{f(n)}{\sqrt{n}})} \\ &= f(n)\frac{\Phi^{-1}(q) - \Phi^{-1}(q_n)}{q - q_n}. \end{aligned}$$

According to the mean value theorem, there exists some value $\xi_n \in [q_n, q]$ such that

$$\begin{aligned}(\Phi^{-1}(q) - \Phi^{-1}(q_n))\sqrt{n} &= f(n)(\Phi^{-1})'(\xi_n) \\ &= f(n)\sqrt{2\pi}e^{\frac{\Phi^{-1}(\xi_n)^2}{2}}.\end{aligned}$$

For the second statement, it is not hard to see that $\lim_{n\to\infty} \xi_n = q$ since

$$q = \lim_{n\to\infty} q_n \le \lim_{n\to\infty} \xi_n \le q.$$

Thus, if $f$ converges to $C$, then

$$\begin{aligned}\lim_{n\to\infty}(\Phi^{-1}(q) - \Phi^{-1}(q_n))\sqrt{n} &= \lim_{n\to\infty} f(n)\sqrt{2\pi}e^{\frac{\Phi^{-1}(\xi_n)^2}{2}} \\ &= C\sqrt{2\pi}e^{\frac{\Phi^{-1}(q)^2}{2}}.\end{aligned}$$

□

**Lemma 11.** *Let $0 < \sigma_{\mathrm{I}} < \infty$ and let $0 < q^- < q^+ < 1$. For arbitrary $q \in [q^-, q^+]$, we denote $b_q = \frac{1}{\sqrt{2\pi}\sigma_{\mathrm{I}}}e^{\frac{\Phi^{-1}(q)^2}{2}}$. Then, for any $q - \frac{b_q \log n}{\sqrt{n}} \le \xi_{n,q} \le q$, there exists some $N_1 \in \mathbb{N}$ such that, for $n \ge N_1$,*

$$\frac{\Phi^{-1}(q)^2 - \Phi^{-1}(\xi_{n,q})^2}{2} \ge -\ln 2$$

*where $N_1$ does not depend on the choice of $q$.*

*Proof.* Let $\lambda_q \in [0, 1]$ be such that

$$\xi_{n,q} = q - \lambda_q \frac{b_q \log n}{\sqrt{n}}.$$

If $\lambda_q = 0$, the inequality is trivially correct. Otherwise, note that $\Phi^{-1}(\cdot)^2$ is convex, so we may utilize the inequality

$$\frac{f(x) - f(y)}{x - y} \le \frac{f(x') - f(y')}{x' - y'} \qquad \text{for all } x < x', y < y'$$

of convex functions. Let

$$b := \sup_{q\in[q^-,q^+]} b_q = \max\{b_{q^-}, b_{q^+}\}.$$

Consider

$$\begin{aligned}&\frac{\Phi^{-1}(q)^2 - \Phi^{-1}\big(q - \lambda_q \frac{b_q \log n}{\sqrt{n}}\big)^2}{2} \\ &= \frac{1}{2}\cdot\lambda_q\frac{b_q\log n}{\sqrt{n}}\frac{\Phi^{-1}(q)^2 - \Phi^{-1}\big(q - \lambda_q \frac{b_q \log n}{\sqrt{n}}\big)^2}{q - \big(q - \lambda_q \frac{b_q \log n}{\sqrt{n}}\big)} \\ &\ge \frac{1}{2}\cdot\lambda_q\frac{b_q\log n}{\sqrt{n}}\frac{\Phi^{-1}(q^-)^2 - \Phi^{-1}\big(q^- - \frac{b \log n}{\sqrt{n}}\big)^2}{q^- - \big(q^- - \frac{b \log n}{\sqrt{n}}\big)} \\ &= \lambda_q\frac{b_q}{b}\frac{\Phi^{-1}(q^-)^2 - \Phi^{-1}\big(q^- - \frac{b \log n}{\sqrt{n}}\big)^2}{2}\end{aligned}$$

If $q^- > \frac{1}{2}$, then

$$\Phi^{-1}(q^-)^2 \ge \Phi^{-1}\Big(q^- - \frac{b\log n}{\sqrt{n}}\Big)^2$$

for all but finitely many $n$, so $N_1$ can be chosen to exclude all $n$ not satisfying this inequality. Otherwise, if $q^- \le \frac{1}{2}$, then $\Phi^{-1}(q^-)^2 - \Phi^{-1}(q^- - \frac{b\log n}{\sqrt{n}})^2$ is *always* negative, so

$$\begin{aligned}&\lambda_q\frac{b_q}{b}\frac{\Phi^{-1}(q^-) - \Phi^{-1}\big(q^- - \frac{b \log n}{\sqrt{n}}\big)^2}{2} \\ &\ge \frac{\Phi^{-1}(q^-) - \Phi^{-1}\big(q^- - \frac{b \log n}{\sqrt{n}}\big)^2}{2}.\end{aligned}$$

This term converges to 0 as $n$ grows large, so there exists some $N_1$ such that, for all $q \in [q^-, q^+], n \geq N_1$,

$$\begin{aligned}
&\frac{\Phi^{-1}(q)^2 - \Phi^{-1}\big(q - \lambda_q \frac{b_q \log n}{\sqrt{n}}\big)^2}{2} \\
&\geq \frac{\Phi^{-1}(q^-) - \Phi^{-1}\big(q^- - \frac{b \log n}{\sqrt{n}}\big)^2}{2} \\
&\geq -\ln 2.
\end{aligned}$$

□

**Lemma 12.** *Let $\mathcal{P}$ be some arbitrary probability distribution and let $n \in \mathbb{N}$. Let $G$ be some optimal guessing function, then*

$$\mathbb{E}[G(\mathbf{X})^\rho] = \int_0^1 \mathrm{MG}_n(q)^\rho \,\mathrm{d}q$$

*and, for any $y \in \mathbb{N}$,*

$$\mathbb{E}[G(\mathbf{X})^\rho \mid G(\mathbf{X}) \leq y] = \frac{\int_0^{\Pr[G(\mathbf{X}) \leq y]} \mathrm{MG}_n(q)^\rho \,\mathrm{d}q}{\Pr[G(\mathbf{X}) \leq y]}.$$

*Proof.* We only prove the second identity as the first describes a special case where $y = \infty$.

For some arbitrary $0 < q < 1$, let $\mathbf{x} \in \mathcal{X}^n$ be such that

$$\Pr[G(\mathbf{X}) < G(\mathbf{x})] < q \leq \Pr[G(\mathbf{X}) \leq G(\mathbf{x})],$$

by definition of $\mathrm{MG}_n(q)$, this implies $\mathrm{MG}_n(q) = G(\mathbf{x})$. Furthermore, since $G$ is bijective, the sampling probability of $\mathbf{x}$ can be expressed as

$$\begin{aligned}
p(\mathbf{x}) &= \Pr[G(\mathbf{X}) = G(\mathbf{x})] \\
&= \Pr[G(\mathbf{X}) \leq G(\mathbf{x})] - \Pr[G(\mathbf{X}) < G(\mathbf{x})],
\end{aligned}$$

so the term $p(\mathbf{x})G(\mathbf{x})^\rho$ can be rephrased as

$$\begin{aligned}
p(\mathbf{x})G(\mathbf{x})^\rho &= \int_{\Pr[G(\mathbf{X}) < G(\mathbf{x})]}^{\Pr[G(\mathbf{X}) \leq G(\mathbf{x})]} G(\mathbf{x})^\rho \,\mathrm{d}q \\
&= \int_{\Pr[G(\mathbf{X}) < G(\mathbf{x})]}^{\Pr[G(\mathbf{X}) \leq G(\mathbf{x})]} \mathrm{MG}_n(q)^\rho \,\mathrm{d}q.
\end{aligned}$$

Thus, we find

$$\begin{aligned}
\mathbb{E}[G(X)^\rho \mid G(X) \leq y] &= \sum_{\mathbf{x} \in \mathcal{X}^n, G(\mathbf{x}) \leq y} \frac{p(\mathbf{x})}{\Pr[G(\mathbf{X}) \leq y]} G(\mathbf{x})^\rho \\
&= \frac{\int_0^{\Pr[G(\mathbf{X}) \leq y]} \mathrm{MG}_n(q)^\rho \,\mathrm{d}q}{\Pr[G(\mathbf{X}) \leq y]}.
\end{aligned}$$

□

**Lemma 13.** *Let $f(x)$ be non-decreasing on the interval $(0, 1)$, then, for any $0 < \delta < 1$,*

$$\int_0^1 f(x) \,\mathrm{d}x \leq \int_\delta^1 \frac{1}{1-\delta} f(x) \,\mathrm{d}x.$$

*Proof.* This statement can easily be proven via integration by substitution. Indeed, let $g(x) = \delta + (1-\delta)x$, then $g(x) \geq x$ for all $x \in [0, 1]$ and thus

$$\int_\delta^1 f(x) \,\mathrm{d}x = \int_0^1 g'(x) f(g(x)) \,\mathrm{d}x \geq \int_0^1 (1-\delta) f(x) \,\mathrm{d}x.$$

Dividing both sides by $1 - \delta$ proves the claim. □

## REFERENCES


[1] M. IDF, "Report on the security of lwe:improved dual lattice attack," 2022. https://zenodo.org/record/6412487#.ZCrT7-xBxqs.

[2] D. J. Bernstein, "Asymptotics of hybrid primal lattice attacks." Cryptology ePrint Archive, Paper 2023/1892, 2023.

[3] J. L. Massey, "Guessing and entropy," in *Proceedings of 1994 IEEE International Symposium on Information Theory*, p. 204, IEEE, 1994.

[4] O. Rioul, "Variations on a theme by massey," *IEEE Transactions on Information Theory*, vol. 68, no. 5, pp. 2813–2828, 2022.

[5] E. Arikan, "An inequality on guessing and its application to sequential decoding," *IEEE Transactions on Information Theory*, vol. 42, no. 1, pp. 99–105, 1996.

[6] I. Sason and S. Verdú, "Improved bounds on lossless source coding and guessing moments via rényi measures," *IEEE Transactions on Information Theory*, vol. 64, no. 6, pp. 4323–4346, 2018.

[7] J. O. Pliam, "On the incomparability of entropy and marginal guesswork in brute-force attacks," in *Progress in Cryptology —INDOCRYPT 2000*, (Berlin, Heidelberg), pp. 67–79, Springer Berlin Heidelberg, 2000.

[8] C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.
[9] T. Glaser, A. May, and J. Nowakowski, “Entropy suffices for guessing most keys.” Cryptology ePrint Archive, Paper 2023/797, 2023.
[10] S. Boztas, “Comments on ”an inequality on guessing and its application to sequential decoding”,” *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 2062–2063, 1997.
[11] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley, 2012.
[12] A. C. Berry, “The accuracy of the gaussian approximation to the sum of independent variates,” *Transactions of the american mathematical society*, vol. 49, no. 1, pp. 122–136, 1941.
[13] C.-G. Esseen, “Fourier analysis of distribution functions. a mathematical study of the laplace-gaussian law,” 1945.
[14] I. Shevtsova, “On the absolute constants in the berry–esseen inequality and its structural and nonuniform improvements,” *Informatika i Ee Primeneniya*, vol. 7, pp. 124–125, 2013.
[15] J. Bos, L. Ducas, E. Kiltz, T. Lepoint, V. Lyubashevsky, J. M. Schanck, P. Schwabe, G. Seiler, and D. Stehlé, “Crystals-kyber: a cca-secure module-lattice-based kem,” in *2018 IEEE European Symposium on Security and Privacy (EuroS&P)*, pp. 353–367, IEEE, 2018.
[16] P.-A. Fouque, J. Hoffstein, P. Kirchner, V. Lyubashevsky, T. Pornin, T. Prest, T. Ricosset, G. Seiler, W. Whyte, and Z. Zhang, “Falcon: Fast-Fourier Lattice-based Compact Signatures over NTRU,” in *Submission to NIST’s post-quantum cryptography standardization process*, 2018.
[17] J. D. Cook, “Upper and lower bounds for the normal distribution function.” https://www.johndcook.com/blog/norm-dist-bounds/. Accessed: 2025-05-20.