Contents lists available at ScienceDirect

Information Processing Letters

journal homepage: www.elsevier.com/locate/ipl

An information theoretic proof of the Chernoff-Hoeffding inequality

Olivier Rioul [a,*], Patrick Solé [a,b]

[a] LTCI, Télécom Paris, Institut Polytechnique de Paris, France
[b] Institut de Mathématiques de Marseille (CNRS, Aix-Marseille University), Marseilles, France

ARTICLE INFO

MSC:
primary 68, P30
secondary 94, B25

Keywords:
Chernoff bound
Hoeffding's inequality
Shannon entropy
Hilbert entropy function
Kullback-Leibler divergence

ABSTRACT

The Chernoff bound is a well-known upper bound on the tail of binomial distributions of parameter $1/2$ involving the binary entropy function. Hoeffding's inequality (or the Chernoff-Hoeffding inequality) is a generalization for binomial distributions of parameter $1-1/q$, involving the q-ary entropy function (with $q \geq 2$), which can be written in terms of the Kullback-Leibler divergence and is related to the bound in Fano's inequality. We give an information theoretic proof of that bound, and sketch some applications to channel and source coding. We also derive a refined bound which is always sharper.

1. Introduction

The Chernoff bound is a concentration inequality for sums of symmetric Bernoulli random variables (of parameter $\frac{1}{2}$), of popular use in information theory [3,11,10], extremal combinatorics [1], and theoretical computer science in general [7]. It can be written as

$$\sum_{i=0}^{d} \binom{n}{i} \leq \exp\bigl(nh_2(\tfrac{d}{n})\bigr) \tag{1}$$

where $h_2(x)$ is the binary entropy function defined for $0 < x < 1$ as

$$h_2(x) = -x\log x - (1-x)\log(1-x). \tag{2}$$

Hoeffding's inequality [4] for sums of Bernoulli random variables of parameter $1 - \frac{1}{q}$ is a similar inequality:

$$\sum_{i=0}^{d} \binom{n}{i} (q-1)^i \leq \exp\bigl(nh_q(\tfrac{d}{n})\bigr) \tag{3}$$

where $h_q(x)$ is the q-ary entropy function defined for $0 < x < \frac{q-1}{q}$ as

$$h_q(x) = -x\log x - (1-x)\log(1-x) + x\log(q-1), \tag{4}$$

also called *Hilbert entropy function* in [6, § 13.5] and in [12, p. 14]—although we could not find an explanation as to why such a function is named after Hilbert.

Remark 1 *(Base of Logarithms and Exponentials).* Throughout this note, logarithms and exponentials can be taken to *any* base. The reciprocal function of the logarithm $\log(\cdot)$ is denoted $\exp(\cdot)$ in the same base. Thus for example,

- to base 2, $\log 2 = 1$ and $\exp(x) = 2^x$;
- to base q, $\log q = 1$ and $\exp(x) = q^x$;
- to natural base e, $\log x = \ln x$ and $\exp(x) = e^x$.

It is customary, in coding theory [6,9,12], to use logarithms to base 2 in the expression (2) of the binary entropy, and logarithms to base q in the expression (4) of the q-ary entropy. With this convention, the upper bound in Chernoff's inequality (1) writes $2^{nh_2(\frac{d}{n})}$ while the upper bound in Hoeffding's inequality (3) writes $q^{nh_q(\frac{d}{n})}$.

In this work, we recall the classical analytic proof of Hoeffding's inequality, which is in fact a naive saddle point method [5]. Then we give an information theoretic proof, which yields a sharper inequality for a

E-mail address: olivier.rioul@telecom-paris.fr (O. Rioul).

https://doi.org/10.1016/j.ipl.2025.106582
Received 23 February 2025; Received in revised form 25 April 2025; Accepted 28 April 2025

modified parameter d, hereby denoted d'. We give several applications of this bound, notably to combinatorial coding and source coding.

This note is arranged as follows. The next section collects the notions and notations needed in the rest of the paper. Section 3 recalls the analytic proof and Section 4 derives the information theoretic proof. Section 5 discusses an improvement of the information theoretic bound using the modified parameter d'. Finally, Section 6 gives applications to channel and source coding.

2. Information theoretic background

Definition 2 *(q-ary Entropies).* Let $q > 1$ be an integer. The *q-ary entropy* of a probability distribution $(p_0, p_1, \dots, p_{q-1})$ is

$$H(p_0, p_1, \dots, p_{q-1}) = \sum_{i=0}^{q-1} p_i \log \frac{1}{p_i}. \tag{5}$$

The *q-ary relative entropy* (a.k.a. Kullback-Leibler divergence) between two probability distributions $(p_0, p_1, \dots, p_{q-1})$ and $(r_0, r_1, \dots, r_{q-1})$ is

$$D(p_0, p_1, \dots, p_{q-1} \| r_0, r_1, \dots, r_{q-1}) = \sum_{i=0}^{q-1} p_i \log \frac{p_i}{r_i}. \tag{6}$$

In particular, the binary entropy and binary relative entropy are

$$\begin{aligned} h_2(p) &\triangleq H(p, 1-p) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p} \\ d_2(p\|r) &\triangleq D(p, 1-p \| r, 1-r) = p \log \frac{p}{r} + (1-p) \log \frac{1-p}{1-r}. \end{aligned} \tag{7}$$

Note that $h_2(p)$ is defined for $0 \le p \le 1$ and increasing for $0 \le p \le 1/2$.

Maximum entropy is a classical concept (see e.g., [3, Chap. 12], [11, Chap. 4]):

Proposition 3 *(Max q-ary Entropy).*

1. *The maximum q-ary entropy is attained for a uniform distribution:*

$$H(p_0, p_1, \dots, p_{q-1}) \le \log q \tag{8}$$

with equality if and only if $p_i = 1/q$ for all i.

2. *The maximum q-ary entropy under the constraint $p_0 = 1 - \alpha$ is attained for the distribution $(1-\alpha, \frac{\alpha}{q-1}, \dots, \frac{\alpha}{q-1})$:*

$$H(p_0 = \alpha, p_1, \dots, p_{q-1}) \le h_q(\alpha) \triangleq h(\alpha) + \alpha \log(q-1) \tag{9}$$

with equality if and only if $p_i = \frac{\alpha}{q-1}$ for $i \ne 0$.

Proof. We use the basic information inequality [3,11,10]

$$D(p_0, p_1, \dots, p_{q-1} \| r_0, r_1, \dots, r_{q-1}) \ge 0 \tag{10}$$

with equality $= 0$ if and only if $p_i = r_i$ for all i:

1. Take $r_i = 1/q$ for all i:
$D(p_0, \dots, p_{q-1} \| \frac{1}{q}, \dots, \frac{1}{q}) = \log q - H(p_0, \dots, p_{q-1}) \ge 0.$
2. Take $r_1 = p_1 = \alpha$ and $r_i = \frac{\alpha}{q-1}$ for $i \ne 0$:
$D(p_0 \dots p_{q-1} \| \alpha, \frac{\alpha}{q-1} \dots \frac{\alpha}{q-1}) = h(\alpha) + \alpha \log(q-1) - H(p_0 \dots p_{q-1}) \ge 0.$ □

Remark 4 *(Fano Bound).* The second inequality in Proposition 3 is a particular instance of *Fano's inequality* for unconditional entropies, where α represents a probability of error (see e.g., [3, § 2.10], [11, § 4.3], and [10, § 3.6]). For instance, if $X \in \{0, 1, \dots, q-1\}$ is an error random variable having distribution $(p_0, p_1, \dots, p_{q-1})$, then $X = 0$ represents zero error, of probability $p_0 = 1 - \alpha$ while $X \ne 0$ is the error event of probability $p_1 + \dots + p_{q-1} = \alpha$. The Fano inequality writes

$$H(X) \le h_q(\alpha) = h_2(\alpha) + \alpha \log(q-1), \tag{11}$$

where the term $h_2(\alpha)$ is the (binary) entropic uncertainty that an error occurs (or not), and the second term, weighted by the error probability α, the remaining maximum entropy for the remaining possible $q-1$ nonzero values of X, which is $\log(q-1)$.

Notice that $h_q(\alpha)$ is well-defined and concave for $0 \le \alpha \le 1$ but increasing only for $0 \le \alpha \le 1 - \frac{1}{q}$. In the binary case one recovers of course the binary entropy.

By the above derivation of the maximum entropy attained by the uniform distribution when $(p_0, p_1, \dots, p_{q-1}) = (1-\alpha, \frac{\alpha}{q-1}, \dots, \frac{\alpha}{q-1})$, one has

$$\begin{aligned} D(1-\alpha, \tfrac{\alpha}{q-1}, \dots, \tfrac{\alpha}{q-1} \| \tfrac{1}{q}, \dots, \tfrac{1}{q}) &= \log q - H(1-\alpha, \tfrac{\alpha}{q-1}, \dots, \tfrac{\alpha}{q-1}) \\ &= \log q - h_q(\alpha) \end{aligned} \tag{12}$$

This divergence is, therefore, equal to the difference $\log q - h_q(\alpha)$, which can be interpreted as the loss of the maximal q-ary entropy due to the error probability constraint $p_1 = 1 - \alpha$ (instead of $1/q$). We make the following trivial, but apparently not so well-known remark:

Lemma 5 *(Max Entropy Loss). The loss of maximal q-ary entropies due to the error probability constraint $p_1 = 1 - \alpha$ (instead of $1/q$) equals the binary divergence between the probability parameters $1 - \alpha$ and $1/q$:*

$$\log q - h_q(\alpha) = d_2(1-\alpha \| \tfrac{1}{q}) = d_2(\alpha \| 1 - \tfrac{1}{q}) \tag{13}$$

Direct Proof by Calculation. $\log q - h_2(\alpha) - \alpha \log(q-1) = \alpha \log(\alpha \frac{q}{1-q}) + (1-\alpha) \log(q(1-\alpha)) = d_2(\alpha || 1 - \frac{1}{q}) = d_2(1-\alpha \| \frac{1}{q}).$ □

The above proof does not actually explain the fundamental reason why these two divergences $D(1-\alpha, \frac{\alpha}{q-1}, \dots, \frac{\alpha}{q-1} \| \frac{1}{q}, \dots, \frac{1}{q})$ and $d_2(\alpha || 1 - \frac{1}{q})$ coincide. The following information theoretic proof is more satisfactory in this respect, and applies to any divergence $D(P\|Q)$ between probability distributions P and Q satisfying the *data processing inequality* (DPI):

$$D(W(P) \| W(Q)) \le D(P\|Q) \tag{14}$$

where W is any Markov channel [3,10,11]. Not only does the Kullback-Leibler divergence satisfies the DPI, but more generally, any Csiszár f-divergence or Rényi α-divergence also satisfies the DPI [10].

Information Theoretic Proof of Lemma 5. On one hand, by the DPI applied to the channel $X \in \{0, 1, \dots, q-1\} \to 1_{X \ne 0} = Y \in \{0, 1\}$ we have

$$d_2(1-\alpha \| \tfrac{1}{q}) = D(1-\alpha, \alpha \| \tfrac{1}{q}, \tfrac{q-1}{q}) \le D(1-\alpha, \tfrac{\alpha}{q-1}, \dots, \tfrac{\alpha}{q-1} \| \tfrac{1}{q}, \dots, \tfrac{1}{q}). \tag{15}$$

On the other hand, by the DPI applied to the channel $Y \in \{0, 1\} \to X = YU \in \{0, 1, \dots, q-1\}$ where U is uniformly distributed in $\{1, \dots, q-1\}$, we obtain the opposite inequality:

$$\begin{aligned} D(1-\alpha, \tfrac{\alpha}{q-1}, \dots, \tfrac{\alpha}{q-1} \| \tfrac{1}{q}, \dots, \tfrac{1}{q}) &\le D(1-\alpha, \alpha \| \tfrac{1}{q}, \tfrac{q-1}{q}) \\ &= d_2(1-\alpha \| \tfrac{1}{q}). \end{aligned} \tag{16}$$

This shows the desired equality of divergences for any type of divergence satisfying the data processing inequality. □

3. Hoeffding's inequality: classical proof

Theorem 6 *(Hoeffding's Inequality [4]). Let d be a natural integer such that $\frac{d}{n} \le 1 - \frac{1}{q}$. Then*

$$\sum_{i=0}^{d}\binom{n}{i}(q-1)^i \le \exp\bigl(n\,h_q(\tfrac{d}{n})\bigr). \tag{17}$$

Remark 7. The classical form of Hoeffding's inequality is rather an upper bound on the cumulative binomial distribution of parameter p:

$$\sum_{i=0}^{d}\binom{n}{i}p^i(1-p)^{n-i} \le \exp\bigl(-n\,d_2(\tfrac{d}{n}\|p)\bigr) \tag{18}$$

but this amounts the same for $p = 1-\frac{1}{q}$ since then $p^i(1-p)^{n-i} = (q-1)^i q^{-n}$ and by Lemma 5,

$$\begin{aligned} q^n \exp\bigl(-n\,d_2(\tfrac{d}{n}\|1-\tfrac{1}{q})\bigr) &= \exp\bigl(n(\log q - d_2(\tfrac{d}{n}\|1-\tfrac{1}{q}))\bigr)\\ &= \exp\bigl(n\,h_q(\tfrac{d}{n})\bigr). \end{aligned} \tag{19}$$

The classical proof of Hoeffding's inequality uses a (generalized) Chernoff inequality for sum of i.i.d. Bernoulli variables, but it can be rewritten as a simple proof as follows.

"Classical Proof". Let $\alpha = \frac{d}{n}$ and $x \in (0,1]$. Since for $i \le d$, $x^{i-d} \ge 1$, we have the following Markov inequality:

$$\begin{aligned} \sum_{i=0}^{d}\binom{n}{i}(q-1)^i &\le \sum_{i=0}^{n}\binom{n}{i}(q-1)^i x^{i-d}\\ &= (1+(q-1)x)^n x^{-d} = \exp(n\,\varphi(x)) \end{aligned} \tag{20}$$

where

$$\varphi(x) = -\alpha\log x + \log(1+(q-1)x) \tag{21}$$

has derivative

$$\varphi'(x) = -\frac{\alpha}{x} + \frac{q-1}{1+(q-1)x} \gtrless 0 \iff x \gtrless \frac{1}{q-1}\cdot\frac{\alpha}{1-\alpha} \tag{22}$$

Thus $x^* = \frac{1}{q-1}\cdot\frac{\alpha}{1-\alpha}$, which lies in $(0,1]$ since $\alpha \le 1-\frac{1}{q}$, achieves the minimum

$$\phi(x^*) = -\alpha\log\Bigl(\frac{1}{q-1}\cdot\frac{\alpha}{1-\alpha}\Bigr) + \log\frac{1}{1-\alpha} = h_q(\alpha). \quad \square \tag{23}$$

Remark 8 *(Saddle Point Method).* In hindsight, x^* is a saddle point in the sense of [5] for the generating functions $f(z) = (1+(q-1)z)$ and $g(z) = 1$. Using the techniques in [5] it is possible to derive an asymptotic equivalent for $n \to \infty$ and fixed q of the left-hand side of Hoeffding's inequality. This equivalent coincides with the right-hand side up to subexponential terms as is already clear from [6, Lemma 2.10.3 p. 91].

4. Hoeffding's inequality: information theoretic proof

The above classical proof works for any real $q > 1$ but does not explain why this particular q-ary entropy $h_q(\cdot)$ comes into play. The following proof of Hoeffding's inequality is in this respect much more satisfactory.

Information Theoretic Proof. The l.h.s. of Hoeffding's inequality is the number of words x in $(\mathbb{F}_q)^n$ of Hamming weight $w_H(x) \le d$ and length n. Let $X = (X_1, X_2, \dots, X_n)$ be chosen uniformly at random from this set, where each $X_k \in \mathbb{F}_q$. Since X is uniformly distributed, its entropy equals

$$H(X) = \log\sum_{i=0}^{d}\binom{n}{i}(q-1)^i. \tag{24}$$

By symmetry, the X_k's are identically distributed, and we have the following well-known entropic inequality [3,11,10]

$$\begin{aligned} H(X) &= H(X_1, X_2, \dots, X_n)\\ &\le H(X_1) + H(X_2) + \dots + H(X_n) = nH(X_1). \end{aligned} \tag{25}$$

Again by symmetry, conditioned on fixed Hamming weight $w_H(X) = i$, the probability that $X_1 \ne 0$ is $\frac{i}{n}$. Thus

$$\mathbb{P}(X_1 = 0 \mid w_H(X) = i) = 1 - \frac{i}{n}. \tag{26}$$

It follows from Proposition 3 (part 2) that, conditioned on $w_H(X) = i$,

$$H(X_1 \mid w_H(X) = i) \le h_q\bigl(\frac{i}{n}\bigr). \tag{27}$$

(In fact equality holds, but this is not needed in the proof). Since $h_q(\alpha)$ is increasing for $\alpha \le 1 - \frac{1}{q}$,

$$H(X_1 \mid w_H(X) = i) \le h_q\bigl(\frac{d}{n}\bigr) \tag{28}$$

where the r.h.s. does not depend on $i = w_H(X)$. Therefore, taking the expectation over $w_H(X) \in \{0, 1, \dots, d\}$ we obtain

$$H(X_1) \le h_q\bigl(\frac{d}{n}\bigr) \tag{29}$$

which ends the proof. $\square$

Remark 9 *(Reduction from the q-ary to the Binary Case).* Hoeffding's inequality in the q-ary case may also be easily deduced from the corresponding inequality in the *binary* case, but only provided that $\frac{d}{n} \le \frac{1}{2}$:

$$\sum_{i=0}^{d}\binom{n}{i} \le \exp\bigl(nh_2(\tfrac{d}{n})\bigr) \tag{30}$$

Indeed, if the latter inequality holds, then

$$\begin{aligned} \sum_{i=0}^{d}\binom{n}{i}(q-1)^i &\le (q-1)^d\sum_{i=0}^{d}\binom{n}{i} \le (q-1)^d\exp\bigl(nh_2(\tfrac{d}{n})\bigr)\\ &= \exp\bigl(nh_q(\tfrac{d}{n})\bigr) \end{aligned} \tag{31}$$

Thus, if $\frac{d}{n} \le \frac{1}{2}$, we may use the same information theoretic proof as above, simplified to the binary case $q = 2$, where part 2 of Proposition 3 is not even needed. Note, however, that this simplified proof does *not* extend to values of d such that $\frac{1}{2} < \frac{d}{n} \le 1 - \frac{1}{q}$.

5. Improved bound

As a consequence of the above information theoretic proof in the preceding section, we have the following refined bound.

Corollary 10. *For any $d \le n$ define*

$$d' = \frac{\sum_{i=0}^{d} i\binom{n}{i}(q-1)^i}{\sum_{i=0}^{d}\binom{n}{i}(q-1)^i} \le d. \tag{32}$$

Then the following bound improves (17) with a smaller exponent $d' \le d$:

$$\sum_{i=0}^{d}\binom{n}{i}(q-1)^i \le \exp\bigl(n\,h_q(\tfrac{d'}{n})\bigr). \tag{33}$$

Proof. A closer look at the above information theoretic proof in the preceding section shows in fact the inequality

$$\sum_{i=0}^{d}\binom{n}{i}(q-1)^i \le \exp(n\sum_{i=0}^{d} p_i h_q(\tfrac{i}{n})), \tag{34}$$

where

Table 1
Bounds for $n = 20$, $q = 3$.

d	$\sum_{i=0}^{d} \binom{n}{i}(q-1)^i$	$\exp\big(n\,h_q(\frac{d'}{n})\big)$ (33)	$\exp\big(n\,h_q(\frac{d}{n})\big)$ (Hoeffding)
1	41	96	106
2	801	2,288	2,664
3	99,201	30,512	37,557
5	583,569	1,806,877	2,452,059
6	3,064,209	9,133,781	12,944,339
7	12,986,769	36,516,145	53,809,632
8	45,235,089	117,823,292	179,450,647
9	131,230,609	311,350,993	485,699,607
10	320,420,753	681,702,205	1,073,741,823

$$p_i = \mathbb{P}(w_H(X) = i | w_H(X) \le d) = \frac{\binom{n}{i}(q-1)^i}{\sum_{i=0}^{d} \binom{n}{i}(q-1)^i}. \tag{35}$$

Since $h_q(\alpha)$ is concave,

$$\sum_{i=0}^{d} p_i h_q(\tfrac{i}{n}) \le h_q(\sum_{i=0}^{d} p_i \tfrac{i}{n}), \tag{36}$$

and, combining with (34) we obtain

$$\sum_{i=0}^{d} \binom{n}{i}(q-1)^i \le \exp(nh_q(\tfrac{d'}{n})) \tag{37}$$

where

$$d' = \mathbb{E}(w_H(X)|w_H(X) \le d) = \sum_{i=0}^{d} ip_i \tag{38}$$

is given by (32). □

Remark 11. From (38) we may interpret d' as the average Hamming weight in the Hamming ball of radius d, whose volume V_d is the denominator in the expression (32).

The numerics like those of Table 1 show that (33) is significantly tighter than Hoeffding's inequality.

Some elementary properties of the $d \mapsto d'$ transformation are as follows.

Proposition 12. *Consider the mapping $\Phi : d \mapsto d'$ for any $d \in \{0, 1, \dots, n\}$. Then*

1. *The map $d \mapsto d'$ is strictly increasing;*
2. *$n' = n\frac{q-1}{q}$, where $n' = \Phi(n)$;*
3. *$\frac{d'}{n} \le \frac{q-1}{q}$ (even when $\frac{d}{n}$ does not satisfy this inequality);*
4. *If $\frac{d}{n} \to \alpha \in (0, 1 - \frac{1}{q}]$ when $n \to \infty$ then $h_q(\frac{d'}{n}) \to h_q(\alpha)$;*
5. *$d' \ge \sqrt{\frac{d}{8(1-d/n)}}$;*
6. *for any $0 \le k \le d$, $d' \le d - k\frac{V_{d-k}}{V_d}$ where V_k denotes the volume of the Hamming ball of radius k. In particular, letting $\epsilon_k = \frac{\binom{n}{k}(q-1)^k}{V_d} \ll 1$ be the proportion of words of weight k in the Hamming ball of radius d, we have*

$$d' \le \min\Big(d - 1 + \epsilon_d,\ d - 2 + 2(\epsilon_{d-1} + \epsilon_d)\Big). \tag{39}$$

Proof.

1. Let $V = \sum_{i=0}^{d} \binom{n}{i}(q-1)^i$ be the volume of the Hamming ball of radius d, $N = \sum_{i=0}^{d} i\binom{n}{i}(q-1)^i$, and $h = \binom{n}{d+1}(q-1)^{d+1}$. We need to check that $d' < (d+1)'$, that is,

$$\frac{N}{V} < \frac{N + (d+1)h}{V + h}, \tag{40}$$

that is, $Nh < (d+1)Vh$ by clearing denominators. Now we have, by definition, $N \le dV < (d+1)V$. The result follows.
2. Here n' is the average of a binomial law of parameter $p = 1 - 1/q$, which equals $np = n\frac{q-1}{q}$.
3. Immediate by 1 and 2.
4. Since $d' \le d$, by the monotonicity of the q-ary entropy we have $h_q(\frac{d'}{n}) \le h_q(\frac{d}{n})$, and therefore

$$\limsup_{n\to\infty} h_q(\frac{d'}{n}) \le h_q(\alpha).$$

The inequality

$$\liminf_{n\to\infty} h_q(\frac{d'}{n}) \ge h_q(\alpha)$$

is obtained by taking logs and dividing by n the inequality (33), upon letting $n \to \infty$ and invoking [6,5]

$$\lim_{n\to\infty} \frac{1}{n} \log(\sum_{i=0}^{d} \binom{n}{i}(q-1)^i) = h_q(\alpha).$$

The result follows by combining the inequalities of superior and inferior limits.
5. Writing the definition $d' = \frac{N}{V}$. We know by Hoeffding's inequality that $V \le \exp\big(n\,h_q(\alpha)\big)$ where $\alpha = \frac{d}{n}$. Bounding below the sum N by its last term yields $N \ge d\binom{n}{d}(q-1)^d$. By [9, Chapter 10, Lemma 7 p. 309] we get $\binom{n}{d} \ge \frac{2^{nh_2(\alpha)}}{\sqrt{8n\alpha(1-\alpha)}}$. The result follows then upon noticing that like in Remark 9 (Equation (31)) we have $(q-1)^d \exp\big(nh_2(\frac{d}{n})\big) = \exp\big(nh_q(\frac{d}{n})\big)$.
6. By splitting the sum $N_d = \sum_{i=0}^{d} i\binom{n}{i}(q-1)^i$ for $i \le d-k$ and $i > d-k$,

$$N_d \le (d-k)V_{d-k} + d(V_d - V_{d-k}) = dV_d - kV_{d-k}. \tag{41}$$

Dividing by V_d gives the announced inequality $d' \le d - k\frac{V_{d-k}}{V_d}$. Numerical calculations show that the optimal value of k is typically $k = 1$ or 2 for reasonable values of q, n and d. Since $\frac{V_{d-1}}{V_d} = 1 - \epsilon_d$ and $\frac{V_{d-2}}{V_d} = 1 - \epsilon_d - \epsilon_{d-1}$ this yields (39). □

6. Applications

6.1. Improvement of Liu et al.'s cover metric bound

In [8, Lemma 2], Liu et al. proved a cover metric bound of the form

$$|\mathcal{B}_C(A,d)| \le (d+1)\exp\big((m+n)h_2(\frac{d}{m+n})\big)q^{md} \tag{42}$$

where $\mathcal{B}_C(A,d)$ is the cover metric ball of center A (a $m \times n$ matrix with entries in $\mathbb{F}_q$) with radius d of a cover metric code C. Their proof is based on the simple inequality

$$\binom{m+n}{d} \le 2^{(m+n)h_2(\frac{d}{m+n})} \tag{43}$$

but as seen in the following, Hoeffding's inequality is stronger. The following lemma was also proved in [13].

Lemma 13 *(Improved Lemma of Liu et al.).*

$$|\mathcal{B}_C(A,d)| \le \exp\big((m+n)h_2(\tfrac{d}{m+n})\big)q^{md} \tag{44}$$

Proof. Proceed as in the proof of [8, Lemma 2] to show that

$$|\mathcal{B}_C(A,d)| \le \sum_{r=0}^{d} \binom{m+n}{r} q^{mr} \tag{45}$$

then apply Hoeffding's inequality with $m+n$ in place of n and q^m in place of $(q-1)$ gives

$$|\mathcal{B}_C(A,d)| \le \exp\big((m+n)h_{q^m+1}(\tfrac{d}{m+n})\big) \tag{46}$$

where

$$h_{q^m+1}(\tfrac{d}{m+n}) = h_2(\tfrac{d}{m+n}) + \tfrac{d}{m+n}\log(q^m) \tag{47}$$

which gives the announced inequality. □

6.2. Rate-distortion theory

When a block code is used as a codebook for data compression an important parameter is the *covering radius*, which measures the largest possible distortion [2]. A lower bound on the covering radius ρ of a q-ary $[n,k]$ code is the sphere covering bound

$$q^{n-k} \le \sum_{i=0}^{\rho} \binom{n}{i}(q-1)^i \tag{48}$$

Upon using Hoeffding's inequality on the r.h.s. this bound entails, after taking logarithms on both sides,

$$1-R \le h_q(\frac{\rho}{n}) \tag{49}$$

where $R=\frac{k}{n}$ is the code rate. This finite bound is formally the same as the asymptotic version of the sphere covering bound

$$1-R \le H_q(\varrho), \tag{50}$$

where $R=\limsup_{n\to\infty}\frac{k_n}{n}$, and $\varrho=\liminf_{n\to\infty}\frac{\rho_n}{n}$ for a series of $[n,k_n]$ codes of covering radii ρ_n.

6.3. Gilbert-Varshamov bound

The Gilbert-Varshamov (GV) bound states that there are unrestricted codes of length n over an alphabet of size q with a minimum distance d as good as

$$|C| \ge \frac{q^n}{\sum_{i=0}^{d-1}\binom{n}{i}(q-1)^i}.$$

It can actually be regarded as a consequence of the sphere covering bound of the previous paragraph for optimal codes [2]. Upon using Hoeffding's inequality on the r.h.s. this bound entails, after taking logarithms on both sides,

$$1-R \le h_q(\frac{d-1}{n}), \tag{51}$$

where $R=\frac{k}{n}$ is the code rate. This finite bound is formally the same as the asymptotic version of GV bound

$$1-R \le H_q(\delta), \tag{52}$$

where $R=\limsup_{n\to\infty}\frac{k_n}{n}$, and $\delta=\liminf_{n\to\infty}\frac{d_n}{n}$ for a series of length n codes of minimum distances d_n, and sizes $|C_n|=q^{k_n}$.

CRediT authorship contribution statement

Olivier Rioul: Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Conceptualization. **Patrick Solé:** Writing – review & editing, Investigation, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

[1] N. Alon, J.H. Spencer, The Probabilistic Method, John Wiley & Sons, 2016.

[2] G. Cohen, I. Honkala, S. Litsyn, A. Lobstein, Covering Codes, North Holland, Amsterdam, 1997.

[3] T.M. Cover, J.A. Thomas, Elements of Information Theory, 1st ed., John Wiley & Sons, Hoboken, 1990, 2nd ed., 2006.

[4] W. Hoeffding, Probability inequalities for sums of bounded random variables, J. Am. Stat. Assoc. 58 (301) (1963) 13–30.

[5] D. Gardy, P. Solé, Saddle point techniques in asymptotic coding theory, in: Algebraic Coding 1991, in: Springer Lect. Not., vol. 573, 1991, pp. 75–81.

[6] W.C. Huffman, V. Pless, Fundamentals of Error-Correcting Codes, Cambridge Univ. Press, Cambridge, 2003.

[7] S. Jukna, Extremal Combinatorics with Applications in Computer Science, Springer, Berlin, 2011.

[8] S. Liu, C. Xing, C. Yuan, List decoding of cover metric codes up to the singleton bound, IEEE Trans. Inf. Theory 64 (4) (April 2018) 2410–2416.

[9] F.J. MacWilliams, N.J.A. Sloane, The Theory of Error Correcting Codes, North Holland, Amsterdam, 1977.

[10] Y. Polyanskiy, Y. Wu, Information Theory: From Coding to Learning, Cambridge University Press, 2024.

[11] O. Rioul, Théorie de l'information et du codage, Hermes Science - Lavoisier, Paris, Sept. 2007, 286 pp.

[12] J.L. Walker, Codes and Curves, Student Mathematical Library IAS/Park City Mathematics Subseries, vol. 7, American Mathematical Society, June 2000.

[13] M. Shi, P. Wang, P. Solé, The exact value for the volume of the balls of radius d in the cover metric, Adv. Math. Commun. (2025), submitted for publication.