Gaussian Bounds for Discrete Entropies

Olivier Rioul
LTCI, Télécom Paris, Institut Polytechnique de Paris, F-91120, Palaiseau, France.
Email: olivier.rioul@telecom-paris.fr

***Abstract*—It is well known that the Gaussian distribution has the largest differential entropy amongst all distributions of equal variance. In this paper, we derive similar (generalized) Gaussian upper bounds for discrete (Rényi) entropies of integer-valued variables. Using a mixed discrete-continuous bounding technique and the Poisson summation formula from Fourier analysis, it is proved that in many cases, such Gaussian bounds hold with an additive term that vanishes exponentially as the variance increases.**

I. Introduction

As is well known, the Gaussian distribution has the largest differential entropy amongst all distributions of given variance: For any real-valued random variable $\mathscr{X}$ with variance σ^2 and density f,

$$h(\mathscr{X}) \triangleq \int f(x)\log\frac{1}{f(x)}\,\mathrm{d}x \leqslant \tfrac{1}{2}\log(2\pi e\sigma^2). \tag{1}$$

One may wonder whether the same inequality could hold for discrete distributions: For a given discrete random variable X of variance σ^2 with probabilities $p(x) = \mathbb{P}(X = x)$, does one have a similar bound of the (discrete) entropy:

$$H(X) \triangleq \sum_x p(x)\log\frac{1}{p(x)} \leqslant \tfrac{1}{2}\log(2\pi e\sigma^2)? \tag{2}$$

The answer is obviously *no* in general, since for a deterministic random variable ($\sigma^2 = 0$) one has $H(X) = 0$ and $\frac{1}{2}\log(2\pi e\sigma^2) = -\infty$.

More generally, this raises the question of the precise relation between discrete (absolute) and continuous (differential) entropies. The classical answer to this question dates back to the 1961 textbook by Reza [19, § 8.3], and has also been presented in the classical textbooks [13, § 1.3] and [5, § 8.3]. If X is a quantized version of $\mathscr{X}$ with step size Δ, one obtains the well-known approximation $h(\mathscr{X}) \approx H(X) + \log\Delta$ for small Δ (see § II-A below). This explains why $h(\mathscr{X})$ may become unbounded negative as $\Delta \to 0$, and why (2) is false in general.

However, a well-known Gaussian bound similar to (2) has been derived by Jim Massey in an unpublished work in the mid-1970s—later published in the late 1980s [12]:

$$H(X) < \tfrac{1}{2}\log\bigl(2\pi e(\sigma^2 + \tfrac{1}{12})\bigr) \tag{3}$$

for *any integer-valued* random variable of variance σ^2. The classical textbook [5, Exercice 8.7] also credits an unpublished work by Frans Willems[1].

As it turns out, (3) was in fact first published by A. G. Djačkov [7] from the Russian school of information theory, in his 1975 work on coin weighing. Djačkov even stated that

> *For some discrete probability distributions (binomial, hypergeometric, and son on) one can prove that* (3) *is approximately exact if* $\sigma \to \infty$.

In fact, it was long observed from numerical experiments that many discrete distributions over the integers (such as Poisson, binomial, etc.) do actually satisfy the Gaussian bound (2) provided that σ^2 is "not too small".

It was pointed out recently to the author that inequality (2) (without the $1/12$ constant) was in fact rigorously established in the particular case where X follows a binomial $\mathscr{B}(n, 1/2)$ of parameter $p = 1/2$ and even length n by Chang and Weldon in their 1979 work on the multiple-user adder channel [3]. More than fifteen years later it was also established for odd n using the same method based on the theory of Jacobi theta functions [16], and then generalized under some constraints on mean and variance by Mow in [17].

Remark 1 (Two seemingly unrelated applications)**.** Gaussian bounds of the discrete entropy were derived independently at about the same time (mid 1970s) by Djačkov in Russia for solving a weighing problem and by Chang and Weldon in the U.S.A. for solving an adder channel coding problem. Specifically, these problems were as follows:

- (Djačkov) Consider M coins, among which $T \geqslant 2$ are counterfeit coins. There are two distinct (known) weights w_0 (true) and w_1 (counterfeit). (One can always set $w_0 = 0$ and $w_1 = 1$.) We can weigh subsets of coins in a spring scale. Determine all counterfeit coins in a minimal number n of weighings (more precisely, find an asymptotic lower bound on n when M is large);
- (Chang & Weldon) Consider a T-user multiple access noiseless adder channel using a binary uniquely decodable (n, M) code C. The output of the channel is the component-wise integer addition of the T binary codewords of C of length n. Determine the maximum sum-rate (more precisely, find the asymptotic capacity).

Interestingly, it turns out that these two problems are identical! To see this, consider a $n \times M$ matrix $\mathbf{A}$ of zeros and ones, x a column vector of length M of zeros and ones with Hamming weight T, and let $y = \mathbf{A}x$. In the weighing problem, the components of x are the unknown weights of each coin, each row of A gives the corresponding selection of coins for each weighing, and the weighing results are the components of y. In the coding problem, x gives the corresponding selection of codewords for each user, the columns of A are the codewords

[1]Possibly related to his work on multiple access channels, see remark 1.

of C, and y is the channel output. In both cases, we assume that x can be recovered from $y = \mathbf{A}x$ (all counterfeit coins are identified by the weighings, the coding scheme is uniquely decodable), and the problem is to find an asymptotic lower bound on n for large M, or an asymptotic upper bound of the rate $(\log_2 M)/n$.

Such asymptotic bounds were determined in both cases from Gaussian bounds on the discrete entropy. Djačkov used a hypergeometric distribution while Chang and Weldon used the binomial distribution. We refer to [7] and [3] for more details.

In this paper, we establish generalizations of the Djačkov-Massey-Willems (DMW) inequality (3) and simple conditions under which the genuine Gaussian bound (2) actually holds for integer-valued random variables. With the help of the *Poisson summation formula* from Fourier analysis, the original DMW inequality (3) is improved by removing the constant $\frac{1}{12}$ inside the logarithm at the expense of an additional constant which is exponentially small as σ^2 increases. The resulting inequality takes the form

$$H(X) < \tfrac{1}{2}\log(2\pi e\sigma^2) + O(e^{-2\pi^2\sigma^2})$$

(Equation (25) below). The additional constant can become negative under some mild conditions and the Gaussian bound $H(X) < \frac{1}{2}\log(2\pi e\sigma^2)$, which was classically obtained for continuous random variables, still holds for many examples of integer-valued random variables including ones whose distribution satisfies an entropic central limit theorem.

This paper also considers the differential Rényi α-entropy

$$h_\alpha(\mathcal{X}) \triangleq \frac{1}{1-\alpha}\log\int f(x)^\alpha \,\mathrm{d}x \tag{4}$$

which (see Theorem 2 below) is maximized for a given variance $\sigma^2_{\mathcal{X}}$ when $f = \varphi$ is a *generalized Gaussian* distribution :

$$\varphi(x) = \begin{cases} \sqrt{\dfrac{\beta}{\pi\sigma^2_{\mathcal{X}}}}\,\dfrac{\Gamma(\frac{1}{1-\alpha})}{\Gamma(\frac{1}{1-\alpha}-\frac{1}{2})}\,\dfrac{1}{\big(1+\beta(\frac{x-\mu_{\mathcal{X}}}{\sigma_{\mathcal{X}}})^2\big)^{\frac{1}{1-\alpha}}} & \text{for } \frac{1}{3}<\alpha<1; \\ \sqrt{\dfrac{|\beta|}{\pi\sigma^2_{\mathcal{X}}}}\,\dfrac{\Gamma(\frac{\alpha}{\alpha-1}+\frac{1}{2})}{\Gamma(\frac{\alpha}{\alpha-1})}\,\big(1-|\beta|(\frac{x-\mu_{\mathcal{X}}}{\sigma_{\mathcal{X}}})^2\big)_+^{\frac{1}{\alpha-1}} & \text{for } \alpha>1, \end{cases} \tag{5}$$

where $\beta = \frac{1-\alpha}{3\alpha-1}$. Fig. 1 plots some α-Gaussian densities.

We study similar generalized Gaussian bounds applied to the discrete Rényi α-entropy[2]

$$H_\alpha(X) \triangleq \frac{1}{1-\alpha}\log\sum_x p(x)^\alpha \tag{6}$$

To illustrate, the natural generalization of (3) to the Rényi entropy of order $\frac{1}{2}$ reads

$$H_{\frac{1}{2}}(X) < \tfrac{1}{2}\log\big(4\pi^2(\sigma^2+\tfrac{1}{12})\big)$$

[2]The limiting case $\alpha\to 1$ gives $H_1(X) = H(X)$ and $h_1(\mathcal{X}) = h(\mathcal{X})$. For simplicity this is hereafter referred to as the case of $\alpha = 1$ (by continuity).

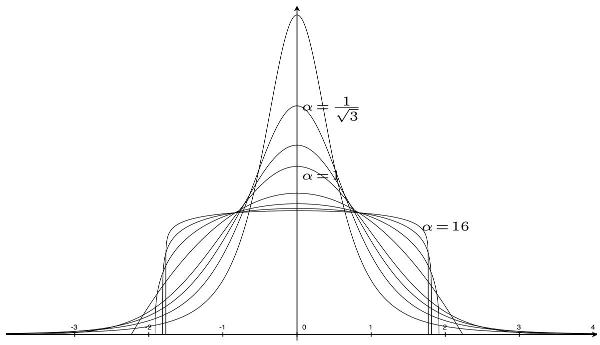

Fig. 1. α-Gaussians (5) for $\alpha = 3^{-3/4}$, $3^{-1/2}$, $3^{-1/4}$, 1, 2, 4, 8, 16.

(see (12) below for the general case), which can be improved as

$$H_{\frac{1}{2}}(X) < \log(2\pi\sigma) + O(e^{-2\pi\sigma})$$

(see (37) below). Such improvements depends on the availability of simple expressions of Fourier transform pairs with sufficient decay at infinity.

II. Massey-Type Inequalitites

A. Links Between Discrete vs. Continuous Entropies

As recalled in the introduction, the classical approach for establishing a link between the (discrete) entropy $H(X)$ and the (continuous) differential entropy $h(\mathcal{X})$ is to "quantize" $\mathcal{X}$ with step size Δ to obtain the discrete X with distribution $p(x_k) = \mathbb{P}(X = x_k) = \int_{k\Delta}^{(k+1)\Delta} f(x)\,\mathrm{d}x$ where the discrete values x_k correspond to mean values $f(x_k) = \frac{1}{\Delta}\int_{k\Delta}^{(k+1)\Delta} f(x)\,\mathrm{d}x = \frac{p(x_k)}{\Delta}$.

It is easily seen that this approach can equally be applied to Rényi entropies of any order $\alpha > 0$. If the density f of $\mathcal{X}$ is continuous within each bin of length Δ and the integral defining the differential entropy exists, then by the continuity assumption, the values x_k are well defined and given by the mean value theorem; since the integral defining $h(\mathcal{X})$ (resp. $h_\alpha(\mathcal{X})$) converges and f is piecewise continuous, $f\log f$ (resp. f^α) is Riemann-integrable and the integral can be respectively approximated by the Riemann sum $\sum_k \Delta\cdot f(x_k)\log\frac{1}{f(x_k)} = H(X) + \log\Delta$ (resp. $\frac{1}{1-\alpha}\log\sum_k \Delta\cdot f^\alpha(x_k) = H_\alpha(X) + \log\Delta$) which tends to $h(\mathcal{X})$ (resp. $h_\alpha(\mathcal{X})$) as $\Delta\to 0$. Therefore,

$$\lim_{\Delta\to 0}\{H_\alpha(X) + \log\Delta\} = h_\alpha(\mathcal{X})$$

for any $\alpha > 0$, which gives the approximation $h_\alpha(\mathcal{X}) \approx H_\alpha(X) + \log\Delta$ for small Δ.

Compared to the classical approach, Massey's proof [12] of his bound (3) (similar to Djačkov's proof in [7]) goes in the opposite direction: Instead of deriving the discrete X from the continuous $\mathcal{X}$ and expressing the continuous entropy in terms of the discrete one, it starts from the discrete random variable X with regularly spaced values, and adds an independent uniformly distributed random perturbation $\mathcal{U}$ to obtain a "dithered" continuous random variable $\mathcal{X} = X + \mathcal{U}$. This is explained in detail in [5, Exercice 8.7], [6]. By doing so, the discrete entropy is expressed in terms of the continuous one. Remarkably, the approximation $h_\alpha(\mathcal{X}) \approx H_\alpha(X) + \log\Delta$ becomes an *exact* equality $h_\alpha(\mathcal{X}) = H_\alpha(X) + \log\Delta$ where Δ needs not be arbitrarily small:

Theorem 1. *Let X be a discrete random variable whose values are regularly spaced Δ apart, and define $\mathcal{X}$ by*

$$\mathcal{X} = X + \mathcal{U} \tag{7}$$

where $\mathcal{U}$ is a continuous random variable independent of X, with support of finite length $\leqslant \Delta$. Then one has the following "additivity" property of entropy:

$$h_\alpha(\mathcal{X}) = H_\alpha(X) + h_\alpha(\mathcal{U}). \tag{8}$$

In particular, if $\mathcal{U}$ is uniformly distributed in an interval of length Δ, then $h_\alpha(\mathcal{U}) = \log\Delta$ and the exact equality

$$h_\alpha(\mathcal{X}) = H_\alpha(X) + \log\Delta \tag{9}$$

holds for any $\alpha > 0$.

Since the identity (8) or (9) is obviously invariant by scaling, one can always set $\Delta = 1$ and restrict oneself to an *integer-valued* random variable X, for which one can identify $H_\alpha(X) = h_\alpha(\mathcal{X})$.

Proof of Theorem 1: By direct calculation (see [21, App. B]). A simpler proof is as follows in the case $\alpha = 1$: By the support assumption, X can be recovered by rounding $X + \mathcal{U}$, hence is a deterministic function of $\mathcal{X}$. Therefore, $H(X|\mathcal{X}) = 0$ and $H(X) = H(X) - H(X|\mathcal{X}) = I(X;\mathcal{X}) = h(\mathcal{X}) - h(\mathcal{X}|X) = h(\mathcal{X}) - h(\mathcal{U})$. ∎

The content of Theorem 1 was recently used in [9] for $\alpha = 1$ in connection with the central limit theorem and also independently in [2] for Rényi entropies of order $\alpha > 1$.

B. A Closer Look at Massey's Inequality (3)

Now suppose that X has finite mean μ and variance σ^2 and define $\mathcal{X}$ by (7). The mean and variance of $\mathcal{X}$ are, respectively, $\mu_{\mathcal{X}} = \mu + \mu_{\mathcal{U}}$ and $\sigma^2_{\mathcal{X}} = \sigma^2 + \sigma^2_{\mathcal{U}}$. Theorem 1 and (1) give

$$H(X) \leqslant \tfrac{1}{2}\log\bigl(2\pi e(\sigma^2 + \sigma^2_{\mathcal{U}})\bigr) - h(\mathcal{U}) \tag{10}$$

where $\mathcal{U}$ has support length $\leqslant 1$. Here the best compromise between maximum possible $h(\mathcal{U})$ and minimum possible $\sigma^2_{\mathcal{U}}$ depends on the value of σ^2. But it can be observed that the obtained bound can never be tight for small values of σ^2. Indeed when $\sigma^2 = 0$, X is deterministic, $H(X) = 0$ and the upper bound in (10) becomes $\frac{1}{2}\log(2\pi e\sigma^2_{\mathcal{U}}) - h(\mathcal{U})$ which is necessarily strictly positive since $\mathcal{U}$ cannot be Gaussian when it has finite support.

Therefore, for large σ^2, the best asymptotic upper bound in (10) is obtained when $h(\mathcal{U})$ is maximum $= \log 1 = 0$. As is well known, $\mathcal{U}$ is then necessarily uniformly distributed in an interval of length 1. In this case $\sigma^2_{\mathcal{U}} = \frac{1}{12}$ and one recovers Massey's inequality (3)—which is strict because $\mathcal{X} = X + \mathcal{U}$ cannot be Gaussian. Notice that (3) gives an $O(\frac{1}{\sigma^2})$ additive term as σ^2 increases since $H(X) < \frac{1}{2}\log\bigl(2\pi e(\sigma^2 + \frac{1}{12})\bigr) < \frac{1}{2}\log(2\pi e\sigma^2) + \frac{\log e}{24\sigma^2}$.

Not only is (3) asymptotically the best possible result using this method, but it is also asymptotically tight for large σ^2: As an example, for Poisson distributed X we have [8] $H(X) = \frac{1}{2}\log(2\pi e\sigma^2) + O(\frac{1}{\sigma^2})$. However, it can still be improved as shown below in Section IV—the $\frac{1}{12}$ constant in (3) will be replaced by an arbitrary small constant as σ gets large.

C. Generalization to Rényi Entropies

Theorem 2 (Generalized Gaussian Bound)**.** *For any continuous random variable $\mathcal{X}$ with differential α-entropy $h_\alpha(\mathcal{X})$,*

$$h_\alpha(\mathcal{X}) \leqslant \begin{cases} \frac{1}{2}\log\bigl(\frac{3\alpha-1}{1-\alpha}\pi\sigma^2_{\mathcal{X}}\bigr) + \frac{1}{1-\alpha}\log\frac{2\alpha}{3\alpha-1} \\ \quad + \log\frac{\Gamma(\frac{1}{1-\alpha}-\frac{1}{2})}{\Gamma(\frac{1}{1-\alpha})} & \text{for } \frac{1}{3} < \alpha < 1; \\ \frac{1}{2}\log\bigl(\frac{3\alpha-1}{\alpha-1}\pi\sigma^2_{\mathcal{X}}\bigr) + \frac{1}{\alpha-1}\log\frac{3\alpha-1}{2\alpha} \\ \quad + \log\frac{\Gamma(\frac{\alpha}{\alpha-1})}{\Gamma(\frac{\alpha}{\alpha-1}+\frac{1}{2})} & \text{for } \alpha > 1, \end{cases} \tag{11}$$

with equality iff $\mathcal{X}$ is α-Gaussian (with density (5)).

Proof: See e.g., [21]. This was first established in [15] for integer $\alpha > 1$ and in [4] in the multivariate case. ∎

Example 1. When $\alpha \to 1$ we recover the usual Gaussian bound (1). As other examples we have

$$h_{\frac{1}{2}}(\mathcal{X}) \leqslant \log(2\pi\sigma_{\mathcal{X}}) \qquad h_{\frac{2}{3}}(\mathcal{X}) \leqslant \log\Bigl(\frac{8\pi\,\sigma_{\mathcal{X}}}{3\sqrt{3}}\Bigr)$$

$$h_2(\mathcal{X}) \leqslant \log\Bigl(\frac{5\sqrt{5}\,\sigma_{\mathcal{X}}}{3}\Bigr) \qquad h_3(\mathcal{X}) \leqslant \log\Bigl(\frac{2\pi\,\sigma_{\mathcal{X}}}{\sqrt{3}}\Bigr)$$

with equality iff $\mathcal{X}$ is $\frac{1}{2}$-Gaussian, $\frac{2}{3}$-Gaussian, 2-Gaussian and 3-Gaussian, respectively. Letting $\alpha \to +\infty$ one obtains $h_\infty(\mathcal{X}) < \log(2\sqrt{3}\,\sigma_{\mathcal{X}})$.

The natural generalization of Massey's inequality (3) to α-entropies is given by a similar reasoning as above in the case $\alpha = 1$. For large σ^2, the best upper bound in Theorem 1 is obtained when $\mathcal{U}$ is uniformly distributed in an interval of length 1 so that $\sigma^2_{\mathcal{X}} = \sigma^2 + \frac{1}{12}$. Thus (11) gives the following

Theorem 3. *For integer-valued X with finite variance σ^2,*

$$H_\alpha(X) < \begin{cases} \frac{1}{2}\log\bigl(\frac{3\alpha-1}{1-\alpha}\pi(\sigma^2 + \frac{1}{12})\bigr) + \frac{1}{1-\alpha}\log\frac{2\alpha}{3\alpha-1} \\ \quad + \log\frac{\Gamma(\frac{1}{1-\alpha}-\frac{1}{2})}{\Gamma(\frac{1}{1-\alpha})} & \text{for } \frac{1}{3} < \alpha < 1 \\ \frac{1}{2}\log\bigl(\frac{3\alpha-1}{\alpha-1}\pi(\sigma^2 + \frac{1}{12})\bigr) + \frac{1}{\alpha-1}\log\frac{3\alpha-1}{2\alpha} \\ \quad + \log\frac{\Gamma(\frac{\alpha}{\alpha-1})}{\Gamma(\frac{\alpha}{\alpha-1}+\frac{1}{2})} & \text{for } \alpha > 1. \end{cases} \tag{12}$$

The strictness of the inequality follows from the fact that $\mathcal{X} = X + \mathcal{U}$ has a staircase density and cannot be α-Gaussian.

Example 2. Thus, referring to Example 1,

$$H_{\frac{1}{2}}(X) < \tfrac{1}{2}\log\Bigl(4\pi^2\bigl(\sigma^2 + \tfrac{1}{12}\bigr)\Bigr),\ H_{\frac{2}{3}}(X) < \tfrac{1}{2}\log\Bigl(\tfrac{64}{27}\pi^2\bigl(\sigma^2 + \tfrac{1}{12}\bigr)\Bigr)$$

$$H_2(X) < \tfrac{1}{2}\log\Bigl(\tfrac{125}{9}\bigl(\sigma^2 + \tfrac{1}{12}\bigr)\Bigr),\ H_3(X) < \tfrac{1}{2}\log\Bigl(\tfrac{4}{3}\pi^2\bigl(\sigma^2 + \tfrac{1}{12}\bigr)\Bigr).$$

and letting $\alpha \to +\infty$, $H_\infty(X) < \frac{1}{2}\log\big(12\sigma^2+1\big)$.

Remark 2. Since the upper bounds of Theorem 2 are tight for α-entropies of continuous variables, a simple quantization argument shows that the upper bounds of (12) (e.g., the constants in Example 2) are asymptotically *tight* for integer-valued variables X as $\sigma^2 \to +\infty$.

This means that for any $\alpha > \frac{1}{3}$, the coefficient multiplying σ^2 inside the logarithm:

$$\frac{3\alpha-1}{|\alpha-1|}\pi$$

is optimal. In [2], the more general identity (8) was used for $\alpha > 1$ and α-Gaussian $\mathcal{U}$ to show that $H_\alpha(X) \leqslant \frac{1}{2}\log(4\frac{3\alpha-1}{\alpha-1}\sigma^2+1)$ where the constant $4\frac{3\alpha-1}{\alpha-1}$ is suboptimal for finite $\alpha > 1$.

Remark 3. Such inequalities cannot exist in general when $\alpha \leqslant \frac{1}{3}$. To see this, consider the discrete random variable $X \geqslant 1$ having distribution $\mathbb{P}(X=k) = \frac{c}{(k\log k)^3}$ with normalization constant $c = \sum_{k>0}\frac{1}{(k\log k)^3}$. Then X has finite second moment $\sum_{k>0}\frac{c}{k\log^3 k} < +\infty$ hence finite variance, but $\sum_{k>0}\sqrt[3]{\mathbb{P}(X=k)} = \sum_{k>0}\frac{1}{k\log k} = +\infty$, hence $H_\alpha(X) \geqslant H_{\frac{1}{3}}(X) = +\infty$ for all $\alpha \leqslant \frac{1}{3}$.

III. An Alternative Bounding Technique

A. Derivation

Instead of applying the usual entropy bounds on $\mathcal{X} = X + \mathcal{U}$, it is possible, as an alternative, to apply a similar inequality directly on the discrete entropy of X but using the same probability density functions. This novel general bounding technique has its own interest.

Theorem 4 (Case $\alpha = 1$)**.** *Let X be a discrete random variable and let $\mathcal{X}$ be the random variable having density*

$$f(x) \triangleq \frac{e^{-T(x)}}{Z} \tag{13}$$

where $Z = \int e^{-T(x)}\mathrm{d}x$, such that $\mathbb{E}[T(X)] = \mathbb{E}[T(\mathcal{X})] = m$ is a fixed quantity. Then

$$H(X) \leqslant h(\mathcal{X}) + \log Z' \tag{14}$$

where $Z' = \sum_x f(x)$, the sum being over all discrete values x of X.

Proof: Apply the information inequality $D(p\|q) \geqslant 0$ to $p(x)$, the probability distribution of X, and to $q(x) = \frac{f(x)}{Z'}$, which is also a discrete probability distribution on the same alphabet because of the normalization constant Z'. We obtain *Gibbs' inequality* in the form $H(X) \leqslant -\mathbb{E}\log q(X) = -\mathbb{E}\log f(X) + \log Z'$ where $-\mathbb{E}\log f(X) = \mathbb{E}[T(X)]\log e + \log Z = \mathbb{E}[T(\mathcal{X})]\log e + \log Z = h(\mathcal{X})$. ∎

Theorem 5 (Case $\alpha \neq 1$)**.** *Let X be a discrete random variable and let $\mathcal{X}$ be the random variable having density*

$$f(x) \triangleq \frac{T(x)^{\frac{1}{\alpha-1}}}{Z} \tag{15}$$

where $Z = \int T(x)^{\frac{1}{\alpha-1}}\mathrm{d}x$, such that $\mathbb{E}[T(X)] = \mathbb{E}[T(\mathcal{X})] = m$ is a fixed quantity. Then

$$H_\alpha(X) \leqslant h_\alpha(\mathcal{X}) + \log Z'_\alpha \tag{16}$$

where $Z'_\alpha = \sum_x f_\alpha(x)$, and $f_\alpha = \frac{f^\alpha}{\int f^\alpha}$ is the α-escort density of f, the sum being over all discrete values x of X.

Proof: Denoting the "escort" distributions of exponent α by $p_\alpha(x) = \frac{p^\alpha(x)}{\sum p^\alpha(x)}$ and $q_\alpha(x) = \frac{q^\alpha(x)}{\sum q^\alpha(x)}$, the *relative α-entropy* [11] between p and q is defined as

$$\Delta_\alpha(p\|q) \triangleq D_{1/\alpha}(p_\alpha\|q_\alpha) \geqslant 0 \tag{17}$$

with equality $= 0$ iff $p = q$ a.e. Here D_α denotes the Rényi α-divergence [24]. Expanding $D_{1/\alpha}(p_\alpha\|q_\alpha)$ similarly as in [20, Prop. 8] gives the following α-Gibbs' inequality which generalizes the Gibbs inequality:

$$H_\alpha(X) \leqslant \frac{\alpha}{1-\alpha}\log\mathbb{E}\, q_\alpha^{1-\frac{1}{\alpha}}(X) \tag{18}$$

with equality iff $p = q$ a.e. Now apply (18) to $p(x)$, the probability distribution of X, and to $q(x) = \frac{f(x)}{Z'}$ with the normalization constant $Z' = \sum_x f(x)$, which is also a discrete probability distribution on the same alphabet. Since $q_\alpha(x) = \frac{f_\alpha(x)}{Z'_\alpha}$, we obtain $H_\alpha(X) \leqslant \frac{\alpha}{1-\alpha}\log\mathbb{E}\, q_\alpha^{1-\frac{1}{\alpha}}(X) = \frac{\alpha}{1-\alpha}\log\mathbb{E}\, f_\alpha^{1-\frac{1}{\alpha}}(X) + \log Z'_\alpha$ where $\frac{\alpha}{1-\alpha}\log\mathbb{E}\, f_\alpha^{1-\frac{1}{\alpha}}(X) = \frac{\alpha}{1-\alpha}\log\mathbb{E}[T(X)] + \log Z_\alpha = \frac{\alpha}{1-\alpha}\log\mathbb{E}[T(\mathcal{X})] + \log Z_\alpha = h_\alpha(\mathcal{X})$. ∎

B. Examples

Corollary 1. *Let X be integer-valued with finite mean μ and variance σ^2. Then*

$$H(X) \leqslant \tfrac{1}{2}\log(2\pi e\sigma^2) + \log\sum_x \frac{e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}}{\sqrt{2\pi\sigma^2}}, \tag{19}$$

which can be simplified as

$$H(X) \leqslant \frac{\log e}{2} + \log\sum_x e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}, \tag{20}$$

the sums being taken over all nonnegative integer values x of X.

For $\alpha > \frac{1}{3}$ and any integer-valued X with mean μ and variance σ^2,

$$H_\alpha(X) \leqslant \frac{\alpha}{1-\alpha}\log\frac{2\alpha}{3\alpha-1} + \log\sum_x\Big(1+\frac{1-\alpha}{3\alpha-1}\big(\frac{x-\mu}{\sigma}\big)^2\Big)_+^{\frac{\alpha}{\alpha-1}} \tag{21}$$

where the sum is taken over all integer values x of X.

Proof: For $\alpha = 1$ we take $\mathcal{X}$ Gaussian of parameters (μ, σ^2) and differential entropy $h(\mathcal{X}) = \frac{1}{2}\log(2\pi e\sigma^2)$. Theorem 4 then gives (19).

For $\alpha \neq 1$ we take $\mathcal{X}$ to be α-Gaussian of parameters $(\mu_{\mathcal{X}} = \mu, \sigma^2_{\mathcal{X}} = \sigma^2)$ and differential entropy $h_\alpha(\mathcal{X}) = \frac{\alpha}{1-\alpha}\log(1+\beta) + \log Z_\alpha$, given by the r.h.s. of (11). From the expression of an α-Gaussian (5), we have $f_\alpha(x) = \frac{1}{Z_\alpha}\big(1+\beta(\frac{x-\mu}{\sigma})^2\big)_+^{\frac{\alpha}{\alpha-1}}$ where $\beta = \frac{1-\alpha}{3\alpha-1}$. Therefore, Theorem 5 gives (21). ∎

Remark 4. It may appear peculiar that the upper bound in (19), (20) or (21) depends on the mean $\mu = \mathbb{E}(X)$ while the entropy $H_\alpha(X)$ should normally not. But this upper bound is, in fact, invariant by translation $X + c$ (where $c \in \mathbb{Z}$ because of the constraint of integer-valued variables), as is readily seen by making a change of variables in the sum, e.g., $\sum_x e^{-\frac{1}{2}(\frac{x-(\mu+c)}{\sigma})^2} = \sum_x e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$. In other words, the upper bound in (19), (20) or (21) depends only on μ's fractional part $\{\mu\} = \mu \bmod 1$. The constraint of integer-valued variables makes it impossible to tighten the bound by minimizing over μ since the only possible changes by translation are integer shifts.

Remark 5. The sum in (19), (20) or (21) does not need to be taken over *all* integers if the support of X is limited. A tighter bound always results if one takes the sum only on those integers actually taken by the variable. In particular, when $\alpha > 1$, the sum in (21) is restricted to values x in the interval $|x-\mu| < \sqrt{\frac{3\alpha-1}{\alpha-1}}$.

Remark 6. For large variance, the unsimplified expression (19) is perhaps preferable because its second term can be made small (see Example 3 below). It should be noted, however, that for moderate values of the variance, the obtained bound in the simplified expression (20) can be valuable. For example, when $X \sim \mathcal{B}(p)$ is a Bernoulli random variable of entropy $H_{\rm b}(p) = p \log \frac{1}{p} + (1-p)\log\frac{1}{1-p}$, the sum in (20) has only two terms:

$$H_{\rm b}(p) \leqslant \log\bigl(e^{\frac{\frac{1}{2}-p}{1-p}} + e^{\frac{p-\frac{1}{2}}{p}}\bigr). \tag{22}$$

This is illustrated in Fig. 2. On the scale of the figure, when the variance is not too small ($|p-\frac{1}{2}| < 0.2$), the two curves are indistinguishable, while in comparison Massey's original bound (3) is much looser.

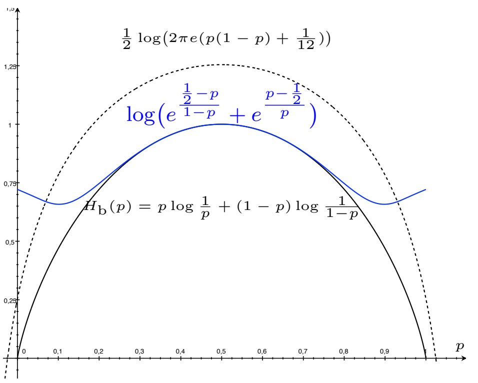

Fig. 2. Moustache bound (22) (blue) vs. Massey's bound (3) (dashed) on the binary entropy function (in bits).

C. Use of the Poisson Summation Formula

When σ^2 or μ is large, then the additional logarithmic term $\log Z'$ in (14) is likely to be small because of the approximation $Z' = \sum_x f(x) \approx \int f(x)\,\mathrm{d}x = 1$. In order to evaluate this precisely, the *Poisson summation formula* can be used.

Lemma 1 (Poisson Summation Formula [22, p. 252])**.** *Let f be Lebesgue-integrable and let*

$$\hat{f}(t) \triangleq \int_{-\infty}^{+\infty} f(x)\, e^{-2i\pi t x}\,\mathrm{d}x \tag{23}$$

be the Fourier transform of $f(x)$. If both f and $\hat{f}$ have $O(\frac{1}{|x|^{1+\varepsilon}})$ decay at infinity then Poisson's summation formula holds:

$$\sum_{x\in\mathbb{Z}} f(x) = \sum_{x\in\mathbb{Z}} \hat{f}(x) \tag{24}$$

where the $x=0$ term in the r.h.s. is $\hat{f}(0) = \int f(x)\,\mathrm{d}x = 1$.

The Fourier transform pairs used in this paper are in Table I.

TABLE I
SOME FOURIER TRANSFORM PAIRS.

$f(x)$	$\hat{f}(x)$
$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$	$e^{-2i\pi\mu x} e^{-2(\pi\sigma x)^2}$
$\frac{1}{\pi\sigma}\frac{1}{1+(\frac{x-\mu}{\sigma})^2}$	$e^{-2i\pi\mu x} e^{-2\pi\sigma\|x\|}$
$\frac{2}{\pi\sigma}\frac{1}{(1+(\frac{x-\mu}{\sigma})^2)^2}$	$e^{-2i\pi\mu x}(1+2\pi\sigma\|x\|) e^{-2\pi\sigma\|x\|}$

Example 3. As an example, using the first Fourier transform pair of Table I in Poisson's formula (24) one obtains $\sum_{x\in\mathbb{Z}} \frac{e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}}{\sqrt{2\pi\sigma^2}} = \sum_{x\in\mathbb{Z}} e^{-2i\pi\mu x} e^{-2(\pi\sigma x)^2} = 1 + 2\sum_{x=1}^{+\infty} e^{-2(\pi\sigma x)^2} \cos 2\pi\mu x$. This identity is historically the very first occurence of the formula in 1823 by Poisson [18, Eq. (15)] which was later generalized by other mathematicians to other Fourier transform pairs. It shows that for large variance, the second term inthe r.h.s. of (19) is in fact exponentially small.

IV. Improved (Generalized) Gaussian Bounds

In this section, we apply the alternative bounding technique described in Section III with the aim to improve the previous (generalized) Gaussian bounds established in Section II. Applying Theorem 4 or 5 will have the effect of removing the constant $\frac{1}{12}$ in (3) at the expense of a small additional additive constant $\log Z'$ or $\log Z'_\alpha$ in the upper bound.

A. Gaussian Bounds for the Discrete Entropy

For large variance σ^2, Massey's original inequality (3) reads $H(X) \leqslant \frac{1}{2}\log\bigl(2\pi e(\sigma^2+\frac{1}{12})\bigr) < \frac{1}{2}\log(2\pi e\sigma^2) + \frac{\log e}{24\sigma^2}$. Now (19) together with Poisson's formula (24) greatly improves Massey's inequality, since the $O(\frac{1}{\sigma^2})$ term can be replaced by the exponentially small $O(e^{-2\pi^2\sigma^2})$:

Theorem 6. *For any integer-valued X of variance $\sigma^2 > 0$,*

$$H(X) < \frac{1}{2}\log(2\pi e\sigma^2) + \frac{2\log e}{e^{2\pi^2\sigma^2}-1}. \tag{25}$$

Proof: Using the first Fourier transform pair of Table I in Poisson's formula (24) one obtains

$$\frac{1}{\sqrt{2\pi\sigma^2}}\sum_{x\in\mathbb{Z}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} = 1 + 2\sum_{x=1}^{+\infty} e^{-2(\pi\sigma x)^2}\cos 2\pi\mu x \tag{26}$$

The sum in the r.h.s. is bounded by $\sum_{x\geqslant 1} e^{-2(\pi\sigma x)^2} \leqslant \sum_{x\geqslant 1} e^{-2(\pi\sigma)^2 x} = \frac{1}{e^{2\pi^2\sigma^2}-1}$. Substituting in (19) and using the inequality $\log(1+z) < (\log e)z$ (when $z>0$) gives the result. ∎

Example 4. As a illustration, consider a binomial $X \sim \mathscr{B}(n,p)$ of variance $\sigma^2 = npq$ (where $p+q=1$). The best known upper bound on $H(X)$ is [1, Eq. (7)]

$$H(X) < \frac{1}{2}\log(2\pi enpq) + \frac{\log e}{12n} + \frac{\log(pq)}{2n} + \frac{\log e}{6npq} \quad (27)$$

which (25) considerably improves for large n since all $O(\frac{1}{n})$ terms are replaced by $O(e^{-2\pi^2 npq})$:

$$H(X) < \frac{1}{2}\log(2\pi enpq) + \frac{2\log e}{e^{2\pi^2 npq}-1}. \quad (28)$$

The exponentially small term can even be made disappear under mild conditions. For example:

Corollary 2. *If the integer-valued variable $X \in \mathbb{N}$ is nonnegative and $\frac{\mu}{\sigma^2}$ is bounded by a constant $< 2\pi$, then for large enough σ^2,*

$$H(X) < \frac{1}{2}\log(2\pi e\sigma^2). \quad (29)$$

More precisely, this holds true as soon as

$$\Big(\frac{\mu+1}{\sigma^2}\Big)^2 < (2\pi)^2 - \frac{\log(8\pi\sigma^2)}{\sigma^2} \quad (30)$$

Proof: Apply (19) where the sum can be taken only over $x \in \mathbb{N}$. Then by (26),

$$\sum_{x\in\mathbb{N}} \frac{e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}}{\sqrt{2\pi\sigma^2}} \leqslant 1 + 2\sum_{x=1}^{+\infty} e^{-2(\pi\sigma x)^2} - \sum_{x=1}^{+\infty} \frac{e^{-\frac{1}{2}(\frac{x+\mu}{\sigma})^2}}{\sqrt{2\pi\sigma^2}}.$$

To obtain (29) it is sufficient to prove that $2e^{-2(\pi\sigma x)^2} < \frac{e^{-\frac{1}{2}(\frac{x+\mu}{\sigma})^2}}{\sqrt{2\pi\sigma^2}}$, i.e., $2(\pi\sigma x)^2 - \frac{1}{2}(\frac{x+\mu}{\sigma})^2 > \log\sqrt{8\pi\sigma^2}$ for all $x \geqslant 1$. When $2\pi\sigma^2 > 1$ we have $2(\pi\sigma)^2 > 1/2\sigma^2$ and it is enough to prove the required inequality for $x=1$, i.e., $(2\pi\sigma)^2 > (\frac{\mu+1}{\sigma})^2 + \log(8\pi\sigma^2)$. This will hold for large enough σ^2 provided that $2\pi\sigma^2 > (1+\varepsilon)\mu$ for some $\varepsilon > 0$. ∎

Example 5. As an example, if $X \sim \mathscr{P}(\lambda)$ is Poisson-distributed then $\frac{\mu}{\sigma^2} = \frac{\lambda}{\lambda} = 1 < 2\pi$ so that for large enough λ,

$$H(X) < \frac{1}{2}\log(2\pi e\lambda). \quad (31)$$

It is found numerically that this inequality holds as soon as $\lambda > 0.1312642451\ldots$.

Example 6. Similarly, if $X \sim \mathscr{B}(n,p)$ is binomial, we may always assume that $p \leqslant \frac{1}{2}$ since considering $n-X$ in place of X permutes the roles of p and $q = 1-p$ without changing $H(X)$. Then $\frac{\mu}{\sigma^2} = \frac{np}{npq} = \frac{1}{q} \leqslant 2 < 2\pi$, and by Corollary 2, for large enough n,

$$H(X) < \frac{1}{2}\log(2\pi enpq). \quad (32)$$

It is found numerically that this inequality holds for all $n > 0$ as soon as $|p - \frac{1}{2}| < 0.304449\ldots$.

Remark 7. For the last two examples, Takano's strong central limit theorem [23, Thm. 2] implies that

$$H(X) = \frac{1}{2}\log(2\pi e\sigma^2) + o\Big(\frac{1}{\sigma^{1+\varepsilon}}\Big) \quad (33)$$

for every $\varepsilon > 0$. The above inequalities show that the $o\big(\frac{1}{\sigma^{1+\varepsilon}}\big)$ term is actually negative for large enough σ.

Remark 8. When X has finite support length N, i.e., $0 \leqslant X \leqslant N-1$, a similar calculation as in the proof of Corollary 2 gives Mow's result [16, Cor. 1]

$$\Big(\frac{\max(\mu+1, N-\mu)}{\sigma^2}\Big)^2 < (2\pi)^2 - \frac{\log(2\pi\sigma^2)}{\sigma^2} \quad (34)$$

By contrast, the sufficient condition (30) does not depend on the support length and is thus valid for any N.

B. Generalized Gaussian Bounds for Discrete Rényi Entropies

We now illustrate the use of the Poisson summation formula (24) in (21) for α-entropies, in the two cases $\alpha = \frac{1}{2}$ and $\alpha = \frac{2}{3}$.

Lemma 2. *One has the following Poisson summations:*

$$Z'_{\frac{1}{2}} = \frac{1}{\pi\sigma}\sum_{x\in\mathbb{Z}} \frac{1}{1+(\frac{x-\mu}{\sigma})^2} = 1 + 2\sum_{x=1}^{+\infty} e^{-2\pi\sigma x}\cos 2\pi\mu x. \quad (35)$$

$$Z'_{\frac{2}{3}} = \frac{2}{\pi\sigma}\sum_{x\in\mathbb{Z}} \frac{1}{(1+(\frac{x-\mu}{\sigma})^2)^2} = 1 + 2\sum_{x=1}^{+\infty}(1+2\pi\sigma x)e^{-2\pi\sigma x}\cos 2\pi\mu x. \quad (36)$$

Proof: By (5) the $\frac{1}{2}$-Gaussian density is of the form $f(x) = \frac{1}{Z}(1+(\frac{x-\mu}{\sigma})^2)^{-2}$. It follows that $f_{\frac{1}{2}}(x) = \frac{1}{Z_\alpha}(1+(\frac{x-\mu}{\sigma})^2)^{-1} = \frac{1}{\pi\sigma}\frac{1}{1+(\frac{x-\mu}{\sigma})^2}$. Using the third Fourier transform pair of Table I in Poisson's formula (24) one obtains $\sum_{x\in\mathbb{Z}} \frac{1}{\pi\sigma}\frac{1}{1+(\frac{x-\mu}{\sigma})^2} = \sum_{x\in\mathbb{Z}} e^{-2i\pi\mu x}e^{-2\pi\sigma|x|}$, which is (35).

By (5) the $\frac{2}{3}$-Gaussian density is of the form $f(x) = \frac{1}{Z}(1+\beta(\frac{x-\mu}{\sigma})^2)^{-3}$ where $\beta = \frac{1}{3}$. It follows that $f_{\frac{2}{3}}(x) = \frac{1}{Z_\alpha}(1+\beta(\frac{x-\mu}{\sigma})^2)^{-2} = \frac{2}{\pi\sigma}\frac{1}{(1+(\frac{x-\mu}{\sigma})^2)^2}$. Using the fourth Fourier transform pair of Table I in Poisson's formula (24) one obtains $\sum_{x\in\mathbb{Z}} \frac{2}{\pi\sigma}\frac{1}{(1+(\frac{x-\mu}{\sigma})^2)^2} = \sum_{x\in\mathbb{Z}} e^{-2i\pi\mu x}(1+2\pi\sigma|x|)e^{-2\pi\sigma|x|}$, which is (36). ∎

In the two cases $\alpha = \frac{1}{2}$ and $\frac{2}{3}$, the Massey-type inequalities in Example 2 write $H_{\frac{1}{2}}(X) \leqslant \frac{1}{2}\log\big(4\pi^2(\sigma^2+\frac{1}{12})\big) < \log(2\pi\sigma) + \frac{\log e}{24\sigma^2}$ and $H_{\frac{2}{3}}(X) \leqslant \frac{1}{2}\log\Big(\frac{64}{27}\pi^2\big(\sigma^2+\frac{1}{12}\big)\Big) < \log(\frac{8}{3\sqrt{3}}\pi\sigma) + \frac{\log e}{24\sigma^2}$, respectively. In these inequalities, the $O(\frac{1}{\sigma^2})$ term can be replaced by the exponentially small $O(e^{-2\pi\sigma})$ and $O(\sigma e^{-2\pi\sigma})$, respectively:

Theorem 7. *For any integer-valued X of variance $\sigma^2 > 0$,*

$$H_{\frac{1}{2}}(X) < \log(2\pi\sigma) + \frac{2\log e}{e^{2\pi\sigma}-1} \quad (37)$$

$$H_{\frac{2}{3}}(X) < \log\Big(\frac{8\pi\sigma}{3\sqrt{3}}\Big) + \frac{4(1+\pi\sigma)\log e}{e^{2\pi\sigma}-1}. \quad (38)$$

Proof: The sum in the r.h.s. of (35) is bounded by $\sum_{x\geqslant 1} e^{-2\pi\sigma x} = \frac{1}{e^{2\pi\sigma}-1}$. Substituting in (21) and using the inequality $\log(1+z) < (\log e)z$ (when $z>0$) gives (37).

Likewise, the sum in the r.h.s. of (36) is bounded by $\sum_{x\geqslant 1}(1+2\pi\sigma x)e^{-2\pi\sigma x} = \frac{1+2\pi\sigma}{e^{2\pi\sigma}-1} + \frac{2\pi\sigma}{(e^{2\pi\sigma}-1)^2} < 2\frac{1+\pi\sigma}{e^{2\pi\sigma}-1}$ (where we used that $2\pi\sigma < e^{2\pi\sigma}-1$). Substituting in (21) and using the inequality $\log(1+z) < (\log e)z$ (when $z>0$) gives (38). ∎

Remark 9. Using the Poisson summation formula on other Fourier transform pairs, it is possible to generalize Theorem 7 to any value of the form $\alpha = \frac{k+1}{k+2}$ $(k=0,1,\ldots)$ and prove that

$$H_{\frac{k+1}{k+2}}(X) < \log(c_k \pi\sigma) + O(\sigma^k e^{-2\pi\sigma}) \tag{39}$$

where the constant c_k is given by

$$c_k = 4\sqrt{2k+1}\binom{2k}{k}\left(\frac{k+1}{2(2k+1)}\right)^{k+1}. \tag{40}$$

The method of this and the previous section is not easily applicable to many other cases, however, since it depends on the availability of simple expressions of Fourier transform pairs with sufficient decay at infinity.

V. Conclusion

Simple bounds on the differential Shannon or Rényi entropy for a given fixed variance have long been established in connection with the important maximum entropy problem, which has been heavily studied for continuous distributions. By contrast, the similar problem for discrete distributions does not seem to be as popular: With the exception of discrete uniform or geometric laws, few results are known on the maximizing distributions. However, bounding the discrete entropy or discrete Rényi entropy for fixed variance appears as a basic question in information theory.

This paper has shown that improving Massey's approach, many closed-form bounds on discrete entropies or Rényi entropies can be deduced from bounds on the α-entropies of a continuous distribution. Similar derivations can be done for other types of parameter constraints [21].

A variant of Massey's approach together with some Fourier analysis proves very tight Gaussian or generalized Gaussian bounds for large variance—better than what would have been expected from convergence in entropy towards the Gaussian as established by the strong central limit theorem. Therefore, it is likely that Takano's $\sigma^{-1-\varepsilon}$ term [23] can be very much improved in general, at least for integer-valued random variables with finite higher-order moments.

Since Massey-type bounds easily generalize to Rényi entropies with tight α-Gaussian bounds, it would also be interesting to prove some corresponding convergence results in terms of α-entropies and α-Gaussians.

Generalization to multiple dimensions of the inequalities of this paper is also straightforward with the same principles and methodology when $X \in \mathbb{Z}^n$. A multidimensional version of the DMW inequality (3) already appears in [10, Lemma 5]. Inequality (25) was used recently for multidimensional integer lattices in [14].

References

[1] J. A. Adell, A. Lekuona, and Y. Yu, "Sharp bounds on the entropy of the Poisson law and related quantities," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2299–2306, May 2010.

[2] S. G. Bobkov, A. Marsiglietti, and J. Melbourne, "Concentration functions and entropy bounds for discrete log-concave distributions," *Combinatorics, Probability and Computing*, vol. 31, no. 1, pp. 54–72, Jan. 2022.

[3] S.-C. Chang and E. J. Weldon, "Coding for t-user multiple-access channels," *IEEE Transactions on Information Theory*, vol. 25, no. 5, pp. 684–691, Nov. 1979.

[4] J. Costa, A. Hero, and C. Vignat, "On solutions to multivariate maximum α-entropy problems," in *Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR 2003)*, ser. Lecture Notes in Computer Science, A. Rangarajan, M. Figueiredo, and J. Zerubia, Eds., vol. 2683. Springer, 2003, pp. 211–226.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 1st Ed. 1990, 2nd Ed. 2006.

[6] ——, "Elements of information theory: Solutions to problems," Aug. 2007.

[7] A. G. Djačkov, "On a search model of false coins," in *Proc. Colloquia Mathematica Societatis János Bolyai*, ser. Topics in information theory, vol. 16, Keszthely, Hungary, 1975.

[8] R. J. Evans, J. Boersma, N. M. Blachman, and A. A. Jagers, "The entropy of a Poisson distribution: Problem 87-6," *SIAM Review*, vol. 30, no. 2, pp. 314–317, June 1988.

[9] L. Gavalakis and I. Kontoyiannis, "Entropy and the discrete central limit theorem," Available: https://arxiv.org/abs/2106.00514.

[10] B. L. Hughes and A. B. C. III, "Nearly optimal multiuser codes for the binarv adder channel," *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 387–398, March 1996.

[11] A. Lapidoth and C. Pfister, "Two measures of dependence," *Entropy*, vol. 21, no. 778, pp. 1–40, 2019.

[12] J. L. Massey, "On the entropy of integer-valued random variables," in *Proc. Beijing International Workshop of Information Theory*, July 4–7 1988.

[13] R. J. McEliece, *The Theory of Information and Coding*. Cambridge University Press, 1st Ed. 1985, 2nd Ed. 2002.

[14] F. C. C. Meneghetti, H. K. Miyamoto, and S. I. R. Costa, "Information properties of a random variable decomposition through lattices," in *41st International Conference on Bayesian and Maximum Entropy methods in Science and Engineering (MaxEnt 2022)*, 2022.

[15] S. Moriguti, "A lower bound for a probability moment of any absolutely continuous distribution with finite variance," *Ann. Math. Statist.*, vol. 23, no. 2, pp. 286–289, June 1952.

[16] W. H. Mow, "Capacity bound for a noiseless t-user binary adder channel," *Electronic Letters*, vol. 32, no. 13, pp. 1157–1158, June 1996.

[17] ——, "A tight upper bound on discrete entropy," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 775–778, March 1998.

[18] S.-D. Poisson, "Suite du mémoire sur les intégrales définies et sur la sommation des séries," *Journal de l'École Royale Polytechnique*, vol. 19, no. 12, pp. 404–509, Juillet 1823.

[19] F. M. Reza, *An Introduction to Information Theory*. New York: Dover, 1961.

[20] O. Rioul, "Rényi entropy power and normal transport," in *Proc. International Symposium on Information Theory and Its Applications (ISITA2020)*, Oct. 24-27 2020, pp. 1–5.

[21] ——, "Variations on a theme by Massey," *IEEE Transactions on Information Theory*, vol. 68, no. 5, pp. 2813–2828, May 2022.

[22] E. M. Stein and G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*. Princeton Univertsity Press, 1971.

[23] S. Takano, "Convergence of entropy in the central limit theorem," *Yokohama Journal*, vol. 35, pp. 143–148, 1987.

[24] T. van Erven and P. Harremoës, "Rényi divergence and Kullback-Leibler divergence," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 3797–3820, Jul. 2014.