Variations on a Theme by Massey

Olivier Rioul, *Member, IEEE*

Abstract—In 1994, Jim Massey proposed the guessing entropy as a measure of the difficulty that an attacker has to guess a secret used in a cryptographic system, and established a well-known inequality between entropy and guessing entropy. Over 15 years before, in an unpublished work, he also established a well-known inequality for the entropy of an integer-valued random variable of given variance. In this paper, we establish a link between the two works by Massey in the more general framework of the relationship between discrete (absolute) entropy and continuous (differential) entropy. Two approaches are given in which the discrete entropy (or Rényi entropy) of an integer-valued variable can be upper bounded using the differential (Rényi) entropy of some suitably chosen continuous random variable. As an application, lower bounds on guessing entropy and guessing moments are derived in terms of entropy or Rényi entropy (without side information) and conditional entropy or Arimoto conditional entropy (when side information is available).

Index Terms—Arikan's inequality, discrete vs. differential entropies, generalized Gaussian densities, generalized exponential densities, guessing entropy, guessing moments, guessing with side information, Kullback's inequality, Massey's inequality, Poisson summation formula, Rényi entropies, Rényi-Arimoto conditional entropies.

I. Introduction

I N AN unpublished work in the mid-1970s, later published in the late 1980s [1], James L. Massey proved the following bound on the entropy of an integer-valued random variable X with variance σ^2:

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

This inequality establishes an interesting connection between the entropy of X and that of a Gaussian random variable. After more than a decade, Massey also established an important inequality for the guessing entropy [2]:

$$G(X) \geqslant 2^{H(X)-2} + 1 \text{ when } H(X) \geqslant 2 \text{ bits,} \tag{2}$$

where again an integer-valued random variable (number of guesses) is involved, the guessing entropy $G(X)$ being defined as the minimum average number of guesses. Perhaps surprisingly, the two Massey inequalities can be seen as part of a common framework which relates discrete (absolute) and continuous (differential) entropies.

The question of making the link between the entropy $H(X)$ of a discrete random variable X and the entropy $h(\mathcal{X})$ of a continuous random variable $\mathcal{X}$ is not new. The usual setting is to consider a discrete random variable X whose values are regularly spaced Δ apart, with some probability distribution $p(x) = \mathbb{P}(X = x)$ having finite entropy. As $\Delta \to 0$, X may approach in distribution a continuous random variable $\mathcal{X}$ with density f. How then the discrete (absolute) entropy

$$H(X) \triangleq \sum_x p(x) \log \frac{1}{p(x)} \tag{3}$$

is related to the continuous (differential) entropy

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

and how can $H(X)$ be evaluated from $h(\mathcal{X})$? Similarly (or more generally), for any fixed $\alpha > 0$, how is the discrete Rényi α-entropy

$$H_\alpha(X) \triangleq \frac{1}{1-\alpha} \log \sum_x p(x)^\alpha \tag{5}$$

related to the continuous Rényi α-entropy

$$h_\alpha(\mathcal{X}) \triangleq \frac{1}{1-\alpha} \log \int f(x)^\alpha \, \mathrm{d}x \tag{6}$$

and how can $H_\alpha(X)$ be evaluated from $h_\alpha(\mathcal{X})$? The limiting case $\alpha \to 1$ gives $H_1(X) = H(X)$ and $h_1(\mathcal{X}) = h(\mathcal{X})$.

For Shannon's entropy, the classical answer to this question dates back to the 1961 textbook by Reza [3, §8.3], and has also been presented in the classical textbooks [4, §1.3] and [5, §8.3]. The approach is to first consider the continuous variable $\mathcal{X}$ having density f, and then *quantize* it to obtain the discrete X with step size Δ. It follows that the integral in (4) or in (6) can approximated by a Riemann sum. Appendix A generalizes the argument to Rényi entropies. One obtains the well-known approximation $H(X) \approx h(\mathcal{X}) - \log \Delta$ for small Δ, and more generally,

$$H_\alpha(X) \approx h_\alpha(\mathcal{X}) - \log \Delta \tag{7}$$

for any $\alpha > 0$. Reza's approximation (7), however appealing as it may be, is not so convenient for evaluating the discrete entropy of X from the continuous one: It requires an arbitrary small Δ and the resulting values of X are in fact not necessarily regularly spaced since they correspond to mean values (Eq. (145) in Appendix A).

Massey's approach, in an unpublished work in the mid-1970s [1], is to write density f as a staircase function whose values are the discrete probabilities. Compared to Reza's, Massey's approach somehow 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 [5, Exercice 8.7], [6] which also credits an unpublished work by Frans Willems. By doing so, the discrete entropy is expressed in terms of the continuous one. Remarkably, as stated in Theorem 1 below, (7) becomes an *exact* equality

$$H_\alpha(X) = h_\alpha(\mathcal{X}) - \log \Delta \qquad (8)$$

where Δ needs not be arbitrarily small.

This paper presents various Massey-type bounds on the Shannon entropy as well as on the Rényi entropy of an arbitrary positive order $\alpha > 0$, of a discrete random variable using a version of *Kullback's inequality* for exponential families applied to $\mathcal{X}$. An alternative bounding technique is to apply Kullback's inequality not to the continuous variable but directly to an integer-valued variable X using the same exponential family density, combined with the *Poisson summation formula* from Fourier analysis.

As an application, Massey's original inequality (1) can be recovered and 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 (Equation (86) below):

$$H(X) < \frac{1}{2}\log(2\pi e \sigma^2) + \frac{2\log e}{e^{2\pi^2 \sigma^2} - 1}. \qquad (9)$$

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

The natural generalization of (1) to Rényi entropies is also easily obtained, e.g.,

$$H_{\frac{1}{2}}(X) < \frac{1}{2}\log\left(4\pi^2\left(\sigma^2 + \frac{1}{12}\right)\right) \qquad (10)$$

(see (72) below for the general case). This particular inequality can be improved as (Equation (96) below)

$$H_{\frac{1}{2}}(X) < \log(2\pi\sigma) + \frac{2\log e}{e^{2\pi\sigma} - 1}. \qquad (11)$$

The method is not only applicable when X has fixed variance but also when $X > 0$ has fixed mean μ (and more generally with some fixed ρ-th order moment). It follows that Massey's lower bound (2) for the guessing entropy can be easily improved as (Equation (103) below):

$$G(X) > \frac{2^{H(X)}}{e} + \frac{1}{2}. \qquad (12)$$

valid for any value of $H(X)$. This inequality also holds in the presence of an observed output Y of a side channel using conditional quantities (Equation (104) below):

$$G(X|Y) > \frac{2^{H(X|Y)}}{e} + \frac{1}{2}. \qquad (13)$$

The improvement over Massey's original inequality (2) is particularly important for large values of entropy, by the factor $4/e$. It is quite startling to notice that the approach followed by Massey back in the 1970s [1] can improve the result of his 1994 paper [2] so much.

The natural generalization to Rényi entropy $H_\alpha(X)$ (without side information) and to Arimoto's conditional entropy $H_\alpha(X|Y)$ (in the presence of some side information Y) reads, e.g.,

$$G(X|Y) > \frac{4}{9}2^{H_2(X|Y)} + \frac{1}{2} \qquad (14)$$

$$G(X|Y) > \frac{1}{4}2^{H_{\frac{2}{3}}(X|Y)} + \frac{1}{2} \qquad (15)$$

(see (108) below for the general case). As shown in this paper, such lower bounds depending of $H_\alpha(X|Y)$ cannot hold in general when $\alpha \leqslant 1/2$, because the support of X may be infinite. For X with *finite* support of size M, Arikan's inequality [7]:

$$G(X|Y) \geqslant \frac{2^{H_{\frac{1}{2}}(X|Y)}}{1 + \ln M} \qquad (16)$$

can be recovered and generalized to values $\alpha < 1/2$ by the method of this paper, e.g.,

$$G(X|Y) > \frac{2^{2H_{\frac{1}{3}}(X|Y)}}{2(2M + 1)} \qquad (17)$$

(see (119) below for a general case). Inequalities relating guessing entropy to (Rényi) entropies have become increasingly popular for practical applications because of scalability properties of entropy (see, e.g., [8], [9]).

The techniques of this paper can also be applied to the guessing ρ-th moment $G_\rho(X|Y)$. While Arikan's inequality

$$G_\rho(X|Y) \geqslant \frac{2^{\rho H_{\frac{1}{1+\rho}}(X|Y)}}{1 + \ln M}, \qquad (18)$$

holds for X with finite support size M, lower bounds independent of M and valid for infinite supports can be obtained for any $\alpha > \frac{1}{1+\rho}$, e.g.,

$$G_2(X|Y) > 2 \cdot \frac{2^{2H(X|Y)}}{\pi e} \qquad (19)$$

$$G_3(X|Y) > \frac{9}{2} \cdot \frac{2^{3H_{1/2}(X|Y)}}{\sqrt{3}\,\pi^3} \qquad (20)$$

$$G_4(X|Y) > \frac{10000}{59049} \cdot 2^{4H_2(X|Y)} \qquad (21)$$

among many other inequalities of this kind (see (127) and (131) below for the general case).

The remainder of this paper is organized as follows. Based on Massey's approach, a general method for establishing Massey-type inequalities for entropies and α-entropies is presented in Section II. An alternative "mixed" bounding technique using the Poisson summation formula is presented in Section III. Section IV applies the method to integer-valued random variables with fixed moment, support length, variance, or mean. Improved inequalities for fixed variance are derived in Section V. Application to guessing is presented in Section VI, where lower bounds are derived for guessing

entropy and ρ-guessing entropy (guessing moment of order ρ). Section VII concludes and suggests perspectives.

II. GENERAL APPROACH TO MASSEY'S INEQUALITIES

A. Massey's Equivalence

A general approach to Massey-type bounds first consists in identifying discrete entropies to continuous ones as follows.

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

$$\mathcal{X} = X + \mathcal{U} \qquad (22)$$

where $\mathcal{U}$ is a continuous random variable independent of X, with support of finite length $\leqslant \Delta$. Then

$$H_\alpha(X) = h_\alpha(\mathcal{X}) - h_\alpha(\mathcal{U}). \qquad (23)$$

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

$$H_\alpha(X) = h_\alpha(\mathcal{X}) - \log \Delta \qquad (24)$$

holds for any $\alpha > 0$.

Proof: See Appendix B. ∎

Remark 1: Theorem 1 shows a peculiar additivity property of entropy:

$$h_\alpha(X + \mathcal{U}) = H_\alpha(X) + h_\alpha(\mathcal{U}), \qquad (25)$$

which does not hold in general when $\mathcal{U}$ has support length $> \Delta$.

Remark 2: The identity (24) is invariant by *scaling*: if $s > 0$, $H_\alpha(sX) = h_\alpha(s\mathcal{X}) - \log(s\Delta)$ is the same as (24) because of the scaling property $h_\alpha(s\mathcal{X}) = h_\alpha(\mathcal{X}) + \log s$. As a result, one can always set $\Delta = 1$ and consider an *integer-valued* random variable X. Hereafter whenever $\mathcal{U}$ is taken uniform we shall always make this assumption. As a result, (24) simply writes

$$H_\alpha(X) = h_\alpha(\mathcal{X}) \qquad (26)$$

when $\mathcal{U}$ is uniformly distributed in an interval of length 1. This is the original remark by Massey [1] that discrete and continuous entropies coincide in this case.

B. Inequalities of the Kullback Type

The next step in the general approach to Massey's inequalities is to bound continuous entropies $h_\alpha(\mathcal{X})$ using appropriate bounding techniques. The case $\alpha = 1$ is familiar:

Theorem 2 (Kullback's Inequality): Let $\mathcal{X}$ be a continuous random variable with differential entropy $h(\mathcal{X})$ and $T(x)$ be a nonnegative function such that the "moment" $\mathbb{E}[T(\mathcal{X})] = m$ is a fixed quantity. Then

$$h(\mathcal{X}) \leqslant m \log e + \log Z \qquad (27)$$

where $Z = \int e^{-T(x)} \,\mathrm{d}x$. Equality holds if and only if $\mathcal{X}$ has density

$$\varphi(x) \triangleq \frac{e^{-T(x)}}{Z}. \qquad (28)$$

Proof: Let $D(f\|\varphi) = \int f \log \frac{f}{\varphi}$ be the relative entropy (or Kullback-Leibler divergence) between the density f of $\mathcal{X}$ and

density φ. The *information inequality* [5, Thm. 2.6.3] states that $D(f\|\varphi) \geqslant 0$ with equality iff (if and only if) $f = \varphi$ a.e. This gives the well known Gibbs inequality

$$h(\mathcal{X}) \leqslant -\mathbb{E} \log \varphi(\mathcal{X}) \qquad (29)$$

with equality iff $f = \varphi$ a.e. Applying Gibbs' inequality to (28) proves the theorem. ∎

Remark 3: Inequality (27) is well known (see, e.g., [10, §21]) and can be seen as a version of *Kullback's inequality* [11, §4] (or the *Kullback-Sanov inequality* [12, pp. 23–24], [13, Chap. 3, Thm. 2.1]) for exponential families parameterized by some $\theta \in \mathbb{R}$. It is more general in the sense that one does not use the condition on "partition function" $Z = Z(\theta)$ which would be required for equality to hold. Such a condition would read $\frac{\mathrm{d}}{\mathrm{d}\theta} \log Z(\theta) = -m$ in the case of a natural exponential family $\varphi(x) = e^{-\theta T'(x)}/Z(\theta)$ where T' does not depend on θ.

The natural generalization to Rényi entropies is as follows.

Theorem 3 (α-Kullback's Inequality): Let $\mathcal{X}$ be a continuous random variable with differential α-entropy $h_\alpha(\mathcal{X})$ and $T(x)$ be a nonnegative function such that the "moment" $\mathbb{E}[T(\mathcal{X})] = m$ is a fixed quantity. Then

$$h_\alpha(\mathcal{X}) \leqslant \frac{\alpha}{1-\alpha} \log m + \log Z_\alpha \qquad (30)$$

where $Z_\alpha = \int T(x)^{\frac{\alpha}{\alpha-1}} \,\mathrm{d}x$. Equality holds iff $\mathcal{X}$ has density

$$\varphi(x) \triangleq \frac{T(x)^{\frac{1}{\alpha-1}}}{Z} \qquad (31)$$

where $Z = \int T(x)^{\frac{1}{\alpha-1}} \,\mathrm{d}x$.

Proof: Let $D_\alpha(f\|\varphi) = \frac{1}{\alpha-1} \log \int f^\alpha \varphi^{1-\alpha}$ be the Rényi α-divergence [14] between the density f of $\mathcal{X}$ and density φ. We have $D_\alpha(f\|\varphi) \geqslant 0$ with equality iff $f = \varphi$ a.e. Denoting the "escort" densities of exponent α by $f_\alpha = \frac{f^\alpha}{\int f^\alpha}$ and $\varphi_\alpha = \frac{\varphi^\alpha}{\int \varphi^\alpha}$, the *relative α-entropy* [15][1] between f and φ is defined as

$$\Delta_\alpha(f\|\varphi) \triangleq D_{1/\alpha}(f_\alpha\|\varphi_\alpha) \qquad (32)$$

which is nonnegative and vanishes iff $f = \varphi$ a.e. Expanding $D_{1/\alpha}(f_\alpha\|\varphi_\alpha)$ gives the α-Gibbs' inequality [18, Prop. 8] which generalizes Gibbs' inequality (29):

$$h_\alpha(\mathcal{X}) \leqslant \frac{\alpha}{1-\alpha} \log \mathbb{E}\, \varphi_\alpha^{1-\frac{1}{\alpha}}(\mathcal{X}) \qquad (33)$$

with equality iff $f = \varphi$ a.e. Applying α-Gibbs' inequality to (31) proves the theorem. ∎

Remark 4: Notice that both $T(x)^{\frac{1}{\alpha-1}}$ and $T(x)^{\frac{\alpha}{\alpha-1}}$ need to be Lebesgue-integrable over the given support interval for Z and Z_α to be well defined and finite.

If the relation $\mathbb{E}[T(\mathcal{X})] = m$ is also satisfied when $\mathcal{X} \sim \varphi$, then

$$\frac{Z_\alpha}{Z} = \frac{1}{Z} \int T(x) T(x)^{\frac{1}{\alpha-1}} \,\mathrm{d}x = \mathbb{E}[T(\mathcal{X})] = m \qquad (34)$$

---

[1]Also named Sundaresan's divergence [16]. For $\alpha = 2$, $D_2(f\|\varphi) = \log \frac{\int f^2 \int g^2}{(\int fg)^2}$ was previously known as the Cauchy-Schwarz divergence [17, Eq. (31) p. 38].

so that in this case (30) simplifies to

$$h_\alpha(\mathfrak{X}) \leqslant \frac{\log m}{1-\alpha} + \log Z. \qquad (35)$$

C. Examples of Inequalities of the Kullback Type

A general maximization statement of α-entropies subject to constraints is given in [19]. A fairly general example is obtained when $\mathfrak{X}$ is parametrized by ρth-order moment $\theta = \mathbb{E}(|\mathfrak{X}|^\rho)$ where $\rho > 0$ is arbitrary.

Theorem 4: For $\theta = \mathbb{E}(|\mathfrak{X}|^\rho)$ with $0 < \rho < +\infty$, and $\alpha > \frac{1}{1+\rho}$, both (30) and (35) reduce to

$$h_\alpha(\mathfrak{X}) \leqslant \begin{cases} \frac{1}{\rho}\log\big(\frac{(1+\rho)\alpha-1}{1-\alpha}\theta\big) + \frac{1}{1-\alpha}\log\frac{\rho\alpha}{(1+\rho)\alpha-1} \\ \quad + \log\frac{2\cdot\Gamma(\frac{1}{\rho}+1)\Gamma(\frac{1}{1-\alpha}-\frac{1}{\rho})}{\Gamma(\frac{1}{1-\alpha})} \quad \text{for } \frac{1}{1+\rho} < \alpha < 1; \\[2ex] \frac{1}{\rho}\log\big(\frac{(1+\rho)\alpha-1}{\alpha-1}\theta\big) + \frac{1}{\alpha-1}\log\frac{(1+\rho)\alpha-1}{\rho\alpha} \\ \quad + \log\frac{2\cdot\Gamma(\frac{1}{\rho}+1)\Gamma(\frac{\alpha}{\alpha-1})}{\Gamma(\frac{\alpha}{\alpha-1}+\frac{1}{\rho})} \quad \text{for } \alpha > 1, \end{cases}$$
$$(36)$$

with equality iff $\mathfrak{X}$ is a generalized α-Gaussian random variable. Inequality (27) reduces to

$$h(\mathfrak{X}) \leqslant \frac{1}{\rho}\log(\rho e\theta) + \log\big(2\Gamma(1+\tfrac{1}{\rho})\big) \qquad (37)$$

with equality iff $\mathfrak{X}$ is a generalized Gaussian random variable.

In case of the one-side constraint $\mathfrak{X} \geqslant 0$ with $\theta = \mathbb{E}(\mathfrak{X}^\rho)$, the same inequalities hold when the factor **2** inside the logarithm is removed.

Proof: See Appendix C, where the generalized α-Gaussian is given in (153). The limiting case $\alpha \to 1$ gives (37). The case $\alpha = 1$ is also proved directly by setting $T(x) = \frac{1}{\rho}\frac{|x|^\rho}{\theta}$ so that $m = \frac{1}{\rho}$ and $Z = 2\Gamma(1+\frac{1}{\rho})(\rho\theta)^{1/\rho}$ in (27). ∎

Let $\mu_{\mathfrak{X}}$ and $\sigma_{\mathfrak{X}}^2$ denote the mean and variance of $\mathfrak{X}$, respectively. We illustrate Theorem 4 in three classical situations:

1) Support Length Parameter: This can be seen as a particular case of Theorem 4 by setting $\rho = +\infty$ in the case of a finite support $(-1, 1)$. More generally, suppose $\mathfrak{X}$ has finite support: $\mathfrak{X} \in (a, b)$ a.s.; letting $\ell(\cdot)$ denote the support length, the corresponding parameter is $\theta = \ell(\mathfrak{X}) = b - a$. For $\alpha = 1$, we set $T(x) = 0$ if $x \in (a, b)$ and $= +\infty$ otherwise. Then φ is the uniform distribution on (a, b), moment $m = 0$, partition $Z = b - a$ and (27) reduces to the known bound [5, Ex. 12.2.4]

$$h(\mathfrak{X}) \leqslant \log(b - a) \qquad (38)$$

with equality iff $\mathfrak{X}$ is uniformly distributed in (a, b).

For $\alpha \neq 1$, we set $T(x) = 1$ if $x \in (a, b)$ and $= 0$ otherwise, so that $\varphi = \varphi_\alpha$ is the uniform distribution on (a, b), moment $m = 1$, $Z = Z_\alpha = b - a$ and (30) or (35) reduces to

$$h_\alpha(\mathfrak{X}) \leqslant \log(b - a) \qquad (39)$$

with equality iff $\mathfrak{X}$ is uniformly distributed in (a, b).
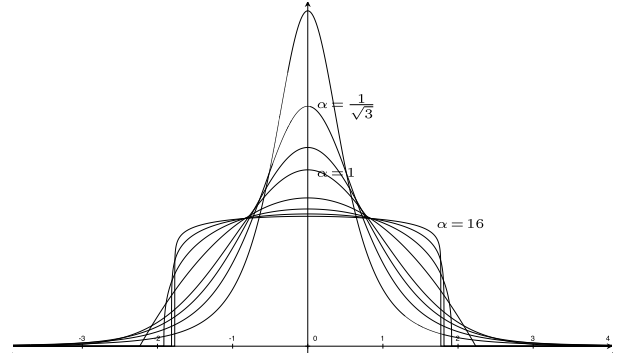

Fig. 1. α-Gaussian distributions (154) for $\alpha = 3^{-3/4}, 3^{-1/2}, 3^{-1/4}, 1, 2, 4, 8, 16$.

2) Variance Parameter: This can be seen as a particular case of Theorem 4 by setting $\rho = 2$ for the centered variable $\mathfrak{X} - \mu_{\mathfrak{X}}$. A direct derivation is as follows. We assume that $\mathfrak{X} \in \mathbb{R}$ with parameter $\theta = \sigma_{\mathfrak{X}}$. For $\alpha = 1$ we set $T(x) = \frac{1}{2}(\frac{x-\mu_{\mathfrak{X}}}{\sigma_{\mathfrak{X}}})^2$, so that $\varphi = \mathcal{N}(\mu_{\mathfrak{X}}, \sigma_{\mathfrak{X}}^2)$ is the Gaussian density, moment $m = \frac{1}{2}$, partition $Z = \sqrt{2\pi\sigma_{\mathfrak{X}}^2}$, and (27) reduces to the well-known Shannon bound [20, §20.5]

$$h(\mathfrak{X}) \leqslant \frac{1}{2}\log(2\pi e\sigma_{\mathfrak{X}}^2) \qquad (40)$$

with equality iff $\mathfrak{X}$ is Gaussian.

For $\alpha \neq 1$ we set $T(x)$ in the form $T(x) = 1 + \beta \cdot (\frac{x-\mu_{\mathfrak{X}}}{\sigma_{\mathfrak{X}}})^2$ so that $m = 1 + \beta$ and β is such that (31) has finite variance $\sigma_{\mathfrak{X}}^2$. The corresponding density φ is known as the α-*Gaussian* density [21]. Under these assumptions, one has $\alpha > \frac{1}{3}$, $\beta = \frac{1-\alpha}{3\alpha-1}$, and both (30) and (35) reduce to the following

Corollary 1: For any continuous random variable $\mathfrak{X}$ with differential α-entropy $h_\alpha(\mathfrak{X})$,

$$h_\alpha(\mathfrak{X}) \leqslant \begin{cases} \frac{1}{2}\log\big(\frac{3\alpha-1}{1-\alpha}\pi\sigma_{\mathfrak{X}}^2\big) + \frac{1}{1-\alpha}\log\frac{2\alpha}{3\alpha-1} \\ \quad + \log\frac{\Gamma(\frac{1}{1-\alpha}-\frac{1}{2})}{\Gamma(\frac{1}{1-\alpha})} \quad \text{for } \frac{1}{3} < \alpha < 1; \\[2ex] \frac{1}{2}\log\big(\frac{3\alpha-1}{\alpha-1}\pi\sigma_{\mathfrak{X}}^2\big) + \frac{1}{\alpha-1}\log\frac{3\alpha-1}{2\alpha} \\ \quad + \log\frac{\Gamma(\frac{\alpha}{\alpha-1})}{\Gamma(\frac{\alpha}{\alpha-1}+\frac{1}{2})} \quad \text{for } \alpha > 1, \end{cases}$$
$$(41)$$

with equality iff $\mathfrak{X}$ is α-Gaussian.

Proof: See Appendix C, where the expression of the α-Gaussian is given in (154). ∎

Fig. 1 plots α-Gaussian densities for different values of α.

Example 1: When $\alpha \to 1$ we recover (40) attained for the Gaussian density. As other examples we have

$$h_{\frac{1}{2}}(\mathfrak{X}) \leqslant \log(2\pi\sigma_{\mathfrak{X}}) \qquad (42)$$

$$h_{\frac{2}{3}}(\mathfrak{X}) \leqslant \log\big(\frac{8\pi\,\sigma_{\mathfrak{X}}}{3\sqrt{3}}\big) \qquad (43)$$

$$h_2(\mathfrak{X}) \leqslant \log\big(\frac{5\sqrt{5}\,\sigma_{\mathfrak{X}}}{3}\big) \qquad (44)$$

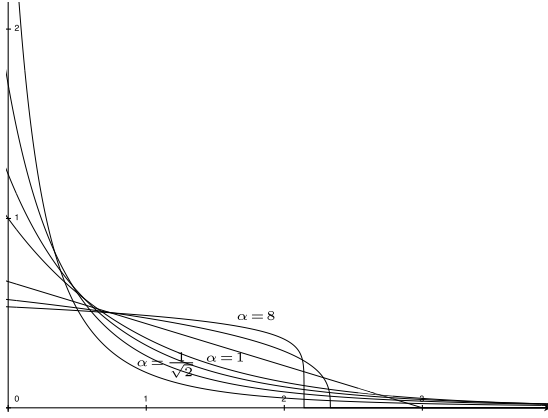$$h_3(\mathfrak{X}) \leqslant \log\big(\frac{2\pi\,\sigma_{\mathfrak{X}}}{\sqrt{3}}\big) \qquad (45)$$

Fig. 2. α-exponential distributions (156) for $\alpha = 2^{-3/4}, 2^{-1/2}, 2^{-1/4}, 1,$ 2, 4, 8.

with equality iff $\mathcal{X}$ is $\frac{1}{2}$-Gaussian, $\frac{2}{3}$-Gaussian, 2-Gaussian and 3-Gaussian, respectively.

3) Mean Parameter: This can be seen as a particular case of Theorem 4 by setting $\rho = 1$ under the one-sided constraint $\mathcal{X} \geqslant 0$. A direct derivation is as follows. We assume that $\mathcal{X} > 0$ a.s. with parameter $\theta = \mu_{\mathcal{X}}$. For $\alpha = 1$ we set $T(x) = \frac{x}{\mu_{\mathcal{X}}}$ so that φ is the exponential density, moment $m = 1$, partition $Z = \mu_{\mathcal{X}}$ and (27) reduces to another Shannon bound [20, § 20.7]

$$h(\mathcal{X}) \leqslant \log(e\mu_{\mathcal{X}}) \tag{46}$$

with equality iff $\mathcal{X}$ is exponential.

For $\alpha \neq 1$, we set $T(x)$ in the form $T(x) = 1 + \beta \cdot \frac{x}{\mu_{\mathcal{X}}}$ so that $m = 1 + \beta$ and β is such that (31) has finite mean $\mu_{\mathcal{X}}$. The corresponding density φ can be named "α-exponential". Under these assumptions, one has $\alpha > \frac{1}{2}$, $\beta = \frac{1-\alpha}{2\alpha-1}$, and both (30) and (35) reduce to the following

Corollary 2: For any continuous random variable $\mathcal{X}$ with differential α-entropy $h_\alpha(\mathcal{X})$,

$$h_\alpha(\mathcal{X}) \leqslant \log \mu_{\mathcal{X}} + \frac{\alpha}{1-\alpha} \log \frac{\alpha}{2\alpha-1} = \log \mu_{\mathcal{X}} + \frac{\alpha}{\alpha-1} \log \frac{2\alpha-1}{\alpha} \tag{47}$$

with equality iff $\mathcal{X}$ is α-exponential.

Proof: See Appendix C, where the expression of the α-exponential is given in (156). ∎

Fig. 2 plots α-exponential densities for different values of α. For $\alpha < 1$, φ is a Pareto Type II distribution with shape parameter $\frac{\alpha}{1-\alpha}$, also known as the Lomax density.

Example 2: When $\alpha \to 1$ we recover (46) attained for the exponential density. As other examples we have

$$h_{\frac{2}{3}}(\mathcal{X}) \leqslant \log(4\mu_{\mathcal{X}}) \tag{48}$$

$$h_{\frac{3}{4}}(\mathcal{X}) \leqslant \log \frac{27\mu_{\mathcal{X}}}{8} \tag{49}$$

$$h_2(\mathcal{X}) \leqslant \log \frac{9\mu_{\mathcal{X}}}{4} \tag{50}$$

with equality iff $\mathcal{X}$ is $\frac{2}{3}$-exponential, $\frac{3}{4}$-exponential, and 2-exponential, respectively.

III. ALTERNATIVE BOUNDING TECHNIQUES

A. Mixed Discrete-Continuous Inequalities of the Kullback Type

Instead of applying Kullback inequalities (27) or (30) 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.

Theorem 5 (Case $\alpha = 1$): Let X be a discrete random variable and let $\mathcal{X}$ be the random variable having density (28):

$$f(x) \triangleq \frac{e^{-T(x)}}{Z} \tag{51}$$

such that the "moment" $\mathbb{E}[T(X)] = \mathbb{E}[T(\mathcal{X})] = m$ is a fixed quantity. Then

$$H(X) \leqslant h(\mathcal{X}) + \log Z' \tag{52}$$

where

$$Z' = \sum_x f(x), \tag{53}$$

the sum being over all discrete values x of X.

Proof: Apply the information inequality $D(p\|q) \geqslant 0$ to $p(x) = \mathbb{P}(X = x)$, 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 $-\mathbb{E}\log f(X) = \mathbb{E}[T(X)]\log e + \log Z = \mathbb{E}[T(\mathcal{X})]\log e + \log Z = h(\mathcal{X})$ by the equality case in (27). ∎

Theorem 6 (Case $\alpha \neq 1$): Let X be a discrete random variable and let $\mathcal{X}$ be the random variable having density (31):

$$f(x) \triangleq \frac{T(x)^{\frac{1}{\alpha-1}}}{Z} \tag{54}$$

such that the "moment" $\mathbb{E}[T(X)] = \mathbb{E}[T(\mathcal{X})] = m$ is a fixed quantity. Then

$$H_\alpha(X) \leqslant h_\alpha(\mathcal{X}) + \log Z'_\alpha \tag{55}$$

where

$$Z'_\alpha = \sum_x f_\alpha(x), \tag{56}$$

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: Let $D_\alpha(p\|q) = \frac{1}{\alpha-1}\log\sum p^\alpha(x)q^{1-\alpha}(x)$ be the Rényi α-divergence [14] between the distribution p of a discrete random variable X and some probability distribution q defined over the same alphabet. We have $D_\alpha(p\|q) \geqslant 0$ with equality iff $p = q$ a.e. 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* [15] between p and q is defined as

$$\Delta_\alpha(p\|q) \triangleq D_{1/\alpha}(p_\alpha\|q_\alpha) \geqslant 0 \tag{57}$$

with equality $= 0$ iff $p = q$ a.e. Expanding $D_{1/\alpha}(p_\alpha\|q_\alpha)$ similarly as in [18, Prop. 8] gives the following α-Gibbs' inequality which generalizes the discrete Gibbs inequality:

$$H_\alpha(X) \leqslant \frac{\alpha}{1-\alpha}\log\mathbb{E}\,q_\alpha^{1-\frac{1}{\alpha}}(X) \tag{58}$$

with equality iff $p = q$ a.e. Now apply (58) to $p(x) = \mathbb{P}(X = x)$, 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 $\frac{\alpha}{1-\alpha} \log \mathbb{E} f_\alpha^{1-\frac{1}{\alpha}}(X) = \frac{\alpha}{1-\alpha} \log \mathbb{E}[T(X)] + \log Z_\alpha = \frac{\alpha}{1-\alpha} \log \mathbb{E}[T(\mathcal{X})] + \log Z_\alpha = h_\alpha(\mathcal{X})$ by the equality case in (30). ∎

Remark 5: Similary as for (24), notice that (52) and (55) are invariant by *scaling*: if $\Delta > 0$, $H_\alpha(\Delta X) = H_\alpha(X)$ while $h_\alpha(\Delta \mathcal{X}) = h_\alpha(\mathcal{X}) + \log \Delta$, hence under scaling by Δ, Z'_α is divided by Δ, and the r.h.s. of (55) becomes $h_\alpha(\mathcal{X}) + \log \Delta + \log(Z'_\alpha/\Delta) = h_\alpha(\mathcal{X}) + \log Z'_\alpha$.

B. Examples of Mixed Inequalities of the Kullback Type

As in the preceding section, we illustrate the bounding method for an integer-valued X in three situations:

1) Support Length Parameter: X has finite support $\{k, \dots, k+\ell\}$ of length $\ell \geqslant 0$, $\mathcal{X}$ is uniformly distributed on an interval (a, b) that includes $\{k, k+\ell\}$. Then $h(\mathcal{X}) = h_\alpha(\mathcal{X}) = \log(b-a)$, $f = f_\alpha$, $Z'_\alpha = \sum_x \frac{1}{b-a} = \frac{\ell+1}{b-a}$ so that (52) and (55) reduce to the known bound $H_\alpha(X) \leqslant \log(b-a) + \log \frac{\ell+1}{b-a} = \log(\ell+1)$ achieved when X is equiprobable.

2) Variance Parameter:

Corollary 3: Let X be integer-valued with finite mean μ and variance σ^2. Then

$$H(X) \leqslant \tfrac{1}{2} \log(2\pi e \sigma^2) + \log \sum_x \frac{e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}}{\sqrt{2\pi\sigma^2}}, \qquad (59)$$

which can be simplified as

$$H(X) \leqslant \frac{\log e}{2} + \log \sum_x e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}, \qquad (60)$$

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,

$$H_\alpha(X) \leqslant \frac{\alpha}{1-\alpha} \log \frac{2\alpha}{3\alpha-1} + \log \sum_x \left(1 + \frac{1-\alpha}{3\alpha-1}\left(\frac{x-\mu}{\sigma}\right)^2\right)_+^{\frac{\alpha}{\alpha-1}} \qquad (61)$$

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

Proof: For $\alpha = 1$ we take $\mathcal{X} \sim \mathcal{N}(\mu, \sigma^2)$ of differential entropy $h(\mathcal{X}) = \frac{1}{2} \log(2\pi e \sigma^2)$. Theorem 5 then gives (59).

For $\alpha \neq 1$ we take $\mathcal{X}$ to be α-Gaussian of parameters $(\mu_{\mathcal{X}} = \mu, \sigma_{\mathcal{X}}^2 = \sigma^2)$ and differential entropy $h_\alpha(\mathcal{X}) = \frac{\alpha}{1-\alpha} \log(1 + \beta) + \log Z_\alpha$, given by the r.h.s. of (41). From the expression of an α-Gaussian (154), we have $f_\alpha(x) = \frac{1}{Z_\alpha}\left(1 + \beta(\frac{x-\mu}{\sigma})^2\right)_+^{\frac{\alpha}{\alpha-1}}$ where $\beta = \frac{1-\alpha}{3\alpha-1}$ and Z_α is given by (155). Therefore, Theorem 6 gives (61). ∎

Remark 6: It may appear peculiar that the upper bound in (59), (60) or (61) depends on the mean $\mu = \mathbb{E}(X)$ while the entropy $H_\alpha(X)$ should 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
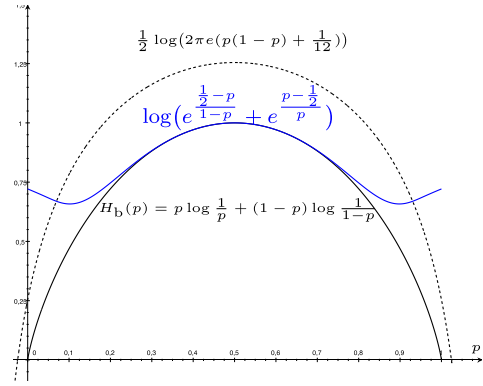

Fig. 3. Moustache bound (62) (blue) vs. Massey's bound (71) (dashed) on the binary entropy function (in bits).

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 (59), (60) or (61) depends only on μ's fractional part $\{\mu\} = \mu \bmod 1$.

Remark 7: The sum in (59), (60) or (61) 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 (61) is restricted to values x in the interval $|x - \mu| < \sqrt{\frac{3\alpha-1}{\alpha-1}}$.

Remark 8: For large variance, the unsimplified expression (59) 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 (60) can be valuable. For example, when $X \sim \mathcal{B}(p)$ is a Bernoulli random variable of entropy $H_b(p) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}$, the sum in (60) has only two terms:

$$H_b(p) \leqslant \log\left(e^{\frac{\frac{1}{2}-p}{1-p}} + e^{\frac{p-\frac{1}{2}}{p}}\right). \qquad (62)$$

This is illustrated in Fig. 3. 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 (71) is much looser.

3) Mean Parameter:

Corollary 4: Let $X \geqslant 0$ be integer-valued with finite mean μ. Then

$$H(X) \leqslant \log(e\mu) + \log \sum_x \frac{e^{-x/\mu}}{\mu} = \log e + \log \sum_x e^{-x/\mu} \qquad (63)$$

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

For $\alpha > \frac{1}{2}$ and any integer-valued $X \geqslant 0$ with finite mean μ,

$$H_\alpha(X) \leqslant \frac{\alpha}{1-\alpha} \log \frac{\alpha}{2\alpha-1} + \log \sum_x \left(1 + \frac{1-\alpha}{2\alpha-1} \cdot \frac{x}{\mu}\right)_+^{\frac{\alpha}{\alpha-1}} \qquad (64)$$

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

Proof: For $\alpha = 1$ we take $\mathfrak{X}$ with exponential density $\frac{e^{-x/\mu}}{\mu}$ of differential entropy $h(\mathfrak{X}) = \log(e\mu)$. Theorem 5 gives (63).

For $\alpha \neq 1$, $\mathfrak{X}$ is α-exponential of mean $\mu_{\mathfrak{X}} = \mu$ and differential entropy given by the r.h.s. of (47). From the expression of an α-exponential (156), we have $f_\alpha(x) = \frac{1}{Z_\alpha}\left(1 + \beta\frac{x}{\mu}\right)_+^{\frac{\alpha}{\alpha-1}}$ where $\beta = \frac{1-\alpha}{2\alpha-1}$ and Z_α is given by (157), that is, $Z_\alpha = \mu$. Theorem 6 gives (64). ∎

Remark 9: Again the sum in (63) or (64) does not need to be taken over *all* $x \in \mathbb{N}$ if the support of X is limited. In particular, when $\alpha > 1$, the sum in (64) is restricted to values x in the interval $0 \leqslant x < \frac{2\alpha-1}{\alpha-1}\mu$.

If, however, the sum is to be taken over $\mathbb{N}$, then evaluating the geometric sum $\sum_{x\in\mathbb{N}} e^{-x/\mu} = \frac{1}{1-e^{-1/\mu}}$ in (63) gives the inequality

$$H(X) \leqslant \log e - \log(1 - e^{-1/\mu}), \qquad (65)$$

As seen in Subsection V-B below, however, this bound turns out to be always weaker then the corresponding Massey-type inequality (78).

C. Use of the Poisson Summation Formula

When σ^2 or μ is large, then the additional logarithmic term $\log Z'$ in (52) 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

$$\hat{f}(t) \triangleq \int_{-\infty}^{+\infty} f(x)\, e^{-2i\pi tx}\,\mathrm{d}x \qquad (66)$$

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:

$$\sum_{x\in\mathbb{Z}} f(x) = \sum_{x\in\mathbb{Z}} \hat{f}(x) \qquad (67)$$

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 given in Table I.

Example 3: As an example, using the first Fourier transform pair of Table I in Poisson's formula (67) 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 [23, 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 (59) is in fact exponentially small.

TABLE I
SOME FOURIER TRANSFORM PAIRS

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

IV. INEQUALITIES OF THE MASSEY TYPE

In this section, we apply the techniques described in Section II to obtains inequalities of the Massey type. In keeping with Remark 2, we assume that X is integer-valued, with mean μ and variance σ^2, and we apply Theorem 1 in the form $H_\alpha(X) = h_\alpha(\mathfrak{X}) - h_\alpha(\mathfrak{U})$ where $\mathfrak{U}$ has support of finite length $\ell(\mathfrak{U}) = \Delta \leqslant 1$. Then Kullback's inequality (27) or (30) applied to $\mathfrak{X} = X + U$ provides various upper bounds on the discrete entropy $H(X)$ from upper bounds on $h(\mathfrak{X})$.

We illustrate this approach here in the three classical situations a), b), c) of Subsection II-C, where we respectively have

a) Support length $\ell(\mathfrak{X}) = \ell(X) + \ell(\mathfrak{U}) = \ell(X) + \Delta$;
a) Variance $\sigma_{\mathfrak{X}}^2 = \sigma^2 + \sigma_{\mathfrak{U}}^2$;
a) Mean $\mu_{\mathfrak{X}} = \mu + \mu_{\mathfrak{U}}$.

A. Inequalities for Fixed Support Length

Suppose that X has finite support $\{k, \ldots, k+\ell\}$ of length $\ell \geqslant 0$. Since $\ell(\mathfrak{X}) = \ell(X) + \ell(\mathfrak{U}) = \ell + \Delta$, by Theorem 1 and inequality (38) or (39), we have

$$H_\alpha(X) \leqslant \log(\ell + \Delta) - h(\mathfrak{U}) \qquad (68)$$

for any $\alpha > 0$. Since $\mathfrak{U}$ has support length $\Delta \leqslant 1$, from (38) or (39) we always have $h(\mathfrak{U}) \leqslant \log\Delta \leqslant \log 1 = 0$ with equality iff $\mathfrak{U}$ is uniformly distributed in an interval of length $\Delta = 1$. Thus, given Δ, the best upper bound in (68) is $\log(\ell + \Delta) - \log\Delta$, which is minimized when Δ is maximum $= 1$. One obtains the well-known bound

$$H_\alpha(X) \leqslant \log(\ell + 1) \qquad (69)$$

achieved when X is equiprobable (hence $\mathfrak{X} = X + \mathfrak{U}$ is uniformly distributed).

Remark 10: Interestingly, achievability of $h(X + \mathfrak{U}) = \log(\ell + 1)$ for $\alpha = 1$ is at the basis of the analysis done in [24, Thm. 1] on Shannon's vs. Hartley's formula.

B. Inequalities for Fixed Variance

Suppose that X has finite variance σ^2. Since $\sigma_{\mathfrak{X}}^2 = \sigma^2 + \sigma_{\mathfrak{U}}^2$, by Theorem 1 and inequality (40), we have

$$H(X) \leqslant \tfrac{1}{2}\log\big(2\pi e(\sigma^2 + \sigma_{\mathfrak{U}}^2)\big) - h(\mathfrak{U}) \qquad (70)$$

where $\mathcal{U}$ has support length $\leqslant 1$. Here the best choice of $\mathcal{U}$—the best compromise between maximum possible $h(\mathcal{U})$ and minimum possible $\sigma_{\mathcal{U}}^2$—depends on the value of σ^2. But it can be observed that the obtained bound cannot be tight for small values of σ^2. Indeed when $\sigma^2 = 0$, X is deterministic, $H(X) = 0$ and the upper bound in (70) becomes $\frac{1}{2}\log(2\pi e\sigma_{\mathcal{U}}^2) - h(\mathcal{U})$ which from (40) is strictly positive since $\mathcal{U}$ cannot be Gaussian when it has finite support.

Therefore, for large σ^2, the best asymptotic upper bound in (70) is obtained when $h(\mathcal{U})$ is maximum $= \log 1 = 0$. From the equality case in (38) $\mathcal{U}$ is then uniformly distributed in an interval of length 1. In this case $\sigma_{\mathcal{U}}^2 = \frac{1}{12}$ and one recovers *Massey's inequality* [1]

$$H(X) < \tfrac{1}{2}\log\big(2\pi e(\sigma^2 + \tfrac{1}{12})\big) \tag{71}$$

for any fixed σ^2, where the strictness of the inequality follows from the fact that $\mathcal{X} = X + \mathcal{U}$ is not Gaussian.

Remark 11: The bound (71) is asymptotically tight for large σ^2: As an example, for Poisson distributed X we have [25] $H(X) = \frac{1}{2}\log(2\pi e\sigma^2) + O(\frac{1}{\sigma^2})$. However, it can still be improved: Section V shows that the $\frac{1}{12}$ constant in (71) can be replaced by an arbitrary small constant as σ gets large.

The natural generalization of Massey's inequality (71) to α-entropies is given by the folllowing

Theorem 7: For any integer-valued X with finite variance σ^2,

$$H_\alpha(X) < \begin{cases} \frac{1}{2}\log\big(\frac{3\alpha-1}{1-\alpha}\pi(\sigma^2+\frac{1}{12})\big) + \frac{1}{1-\alpha}\log\frac{2\alpha}{3\alpha-1} \\[4pt] \quad + \log\frac{\Gamma(\frac{1}{1-\alpha}-\frac{1}{2})}{\Gamma(\frac{1}{1-\alpha})} \qquad \text{for } \frac{1}{3} < \alpha < 1 \\[10pt] \frac{1}{2}\log\big(\frac{3\alpha-1}{\alpha-1}\pi(\sigma^2+\frac{1}{12})\big) + \frac{1}{\alpha-1}\log\frac{3\alpha-1}{2\alpha} \\[4pt] \quad + \log\frac{\Gamma(\frac{\alpha}{\alpha-1})}{\Gamma(\frac{\alpha}{\alpha-1}+\frac{1}{2})} \qquad \text{for } \alpha > 1. \end{cases} \tag{72}$$

Proof: With 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. Hence (26) holds, and since $\sigma_{\mathcal{X}}^2 = \sigma^2 + \sigma_{\mathcal{U}}^2 = \sigma^2 + \frac{1}{12}$, (41) gives (72). The strictness of the inequality follows from the fact that $\mathcal{X} = X + \mathcal{U}$ (which has a staircase density) cannot be α-Gaussian. ∎

Example 4: Thus, referring to Example 1,

$$H_{\frac{1}{2}}(X) < \frac{1}{2}\log\Big(4\pi^2\big(\sigma^2 + \tfrac{1}{12}\big)\Big) \tag{73}$$

$$H_{\frac{2}{3}}(X) < \frac{1}{2}\log\Big(\frac{64}{27}\pi^2\big(\sigma^2 + \tfrac{1}{12}\big)\Big) \tag{74}$$

$$H_2(X) < \frac{1}{2}\log\Big(\frac{125}{9}\big(\sigma^2 + \tfrac{1}{12}\big)\Big) \tag{75}$$

$$H_3(X) < \frac{1}{2}\log\Big(\frac{4}{3}\pi^2\big(\sigma^2 + \tfrac{1}{12}\big)\Big). \tag{76}$$

Remark 12: 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 $\alpha \leqslant \frac{1}{3}$.

C. Inequalities for Fixed Mean

Suppose that $X \geqslant 0$ has finite mean μ. Since $\mu_{\mathcal{X}} = \mu + \mu_{\mathcal{U}}$, by Theorem 1 and inequality (46), we have

$$H(X) \leqslant \log\big(e(\mu + \mu_{\mathcal{U}})\big) - h(\mathcal{U}) \tag{77}$$

provided that $\mathcal{U} \geqslant 0$ a.s. with support length $\leqslant 1$.

Again the best choice of $\mathcal{U}$ (the best compromise between maximum possible $h(\mathcal{U})$ and minimum possible $\mu_{\mathcal{U}}$) depends on the value of the parameter $\mu \geqslant 0$. Also the obtained bound cannot be tight for small values of μ: When $\mu = 0$, $X = 0$ a.s., $H(X) = 0$ and the upper bound in (77) becomes $\log\big(e\mu_{\mathcal{U}}\big) - h(\mathcal{U})$ which from (46) is strictly positive because $\mathcal{U}$ cannot be exponential when it has finite support.

For large μ, the best asymptotic upper bound in (77) is again obtained when $h(\mathcal{U})$ is maximum $= \log 1 = 0$. From the equality case in (38) $\mathcal{U} \geqslant 0$ is then uniformly distributed in an interval of length 1. In this case the minimum value of $\mu_{\mathcal{U}}$ is achieved when $\mathcal{U} \geqslant 0$ is uniformly distributed in $(0,1)$, which gives $\mu_{\mathcal{U}} = \frac{1}{2}$. We obtain the following variation of Massey inequality.

Theorem 8: For any integer-valued $X \geqslant 0$ with finite mean μ,

$$H(X) < \log\big(e(\mu + \tfrac{1}{2})\big) \tag{78}$$

Here the strictness of the inequality follows from the fact that $\mathcal{X} = X + \mathcal{U}$ is not exponential, hence (46) cannot be achieved with equality.

Remark 13: The bound (78) is asymptotically tight for large μ: As an example, for geometric X we have $H(X) = \mu H_{\mathrm{b}}(1/\mu) = \log(e\mu) + O(\frac{1}{\mu})$ where $H_{\mathrm{b}}(p) = p\log\frac{1}{p} + (1-p)\log\frac{1}{1-p}$ is the binary entropy function.

The natural generalization of (78) to α-entropies is given by the following

Theorem 9: For any integer-valued $X \geqslant 0$ with mean μ and any $\alpha > \frac{1}{2}$,

$$\begin{aligned} H_\alpha(X) &< \log(\mu + \tfrac{1}{2}) + \frac{\alpha}{1-\alpha}\log\frac{\alpha}{2\alpha-1} \\ &= \log(\mu + \tfrac{1}{2}) + \frac{\alpha}{\alpha-1}\log\frac{2\alpha-1}{\alpha}. \end{aligned} \tag{79}$$

Proof: For large μ, as in the case $\alpha = 1$ above, the best upper bound in Theorem 1 is obtained when $\mathcal{U}$ is uniformly distributed in $(0,1)$. Hence (26) holds, and since $\mu_{\mathcal{X}} = \mu + \mu_{\mathcal{U}} = \mu + \frac{1}{2}$, (47) gives (79) for any $\alpha > \frac{1}{2}$, where the strictness of the inequality follows from the fact that $\mathcal{X} = X + \mathcal{U}$ (which has a staircase density) cannot be α-exponential. ∎

Example 5: Thus, referring to Example 2,

$$H_{\frac{2}{3}}(X) < \log(4\mu + 2) \tag{80}$$

$$H_{\frac{3}{4}}(X) < \log\frac{27(\mu + \frac{1}{2})}{8} \tag{81}$$

$$H_2(X) < \log\frac{9(\mu + \frac{1}{2})}{4}. \tag{82}$$

Remark 14: Such inequalities cannot exist in general when $\alpha \leqslant \frac{1}{2}$. To see this, consider the discrete random variable

$X \geqslant 1$ with distribution $\mathbb{P}(X = k) = \frac{c}{(k \log k)^2}$ where $c = \sum_{k>0} \frac{1}{(k \log k)^2}$ is a normalization constant. Then X has finite mean $\mu = \sum_{k>0} \frac{c}{k \log^2 k} < +\infty$ but $\sum_{k>0} \sqrt{\mathbb{P}(X = k)} = \sum_{k>0} \frac{1}{k \log k} = +\infty$, hence $H_\alpha(X) \geqslant H_{\frac{1}{2}}(X) = +\infty$ for all $\alpha \leqslant \frac{1}{2}$.

V. IMPROVED INEQUALITIES

In this section, we apply the alternative bounding techniques described in Section III with the aim to improve the previous inequalities of the Massey type. Applying Theorem 5 or 6 will have the effect of removing the constant $\frac{1}{12}$ in (71) and $\frac{1}{2}$ in (78) at the expense of an additional additive constant $\log Z'$ or $\log Z'_\alpha$ in the upper bound.

We again consider an integer-valued variable under the three classical situations *a)*, *b)*, *c)* of Subsection III-B.

A. Inequalities for Fixed Support Length

In case *a)* we have already seen in Subsection III-B that one obtains the known inequality $H_\alpha(X) \leqslant \log(\ell + 1)$ achieved when X of support length ℓ is equiprobable. Thus in this case, no improvement is possible: We obtain the same result as in Subsection IV-A.

B. Inequalities for Fixed Mean

Here we assume $X \geqslant 0$ with fixed mean μ. For $\alpha = 1$, inequality (63) applies with $Z' = \frac{1}{\mu} \sum_{x \in \mathbb{N}} e^{-x/\mu}$. Using the second Fourier transform pair of Table I in Poisson's formula (67) we obtain $\sum_{x \in \mathbb{Z}} \frac{e^{-|x|/\mu}}{\mu} = \sum_{x \in \mathbb{Z}} \frac{2}{1 + (2\pi\mu x)^2}$, which gives

$$Z' = \frac{1}{\mu} \sum_{x \in \mathbb{N}} e^{-x/\mu} = 1 + \frac{1}{2\mu} + 2 \sum_{x=1}^{+\infty} \frac{1}{1 + (2\pi\mu x)^2}. \quad (83)$$

Here we have applied Poisson's formula to the symmetrized density $\frac{1}{2}(f(x) + f(-x))$ to ensure that the decay condition at infinity holds for the Fourier transform. It follows from (83) that

$$\sum_{x \in \mathbb{N}} \frac{e^{-x/\mu}}{\mu} > 1 + \frac{1}{2\mu}, \quad (84)$$

which implies that (65) is strictly weaker than the Massey-type inequality (78): In fact, (78) already reads $H(X) < \log(e\mu) + \log(1 + \frac{1}{2\mu})$.

A similar phenomenon occurs when $\alpha \neq 1$. In fact, comparing (64) to (79) one has

$$\sum_{x \in \mathbb{N}} \left(1 + \frac{1-\alpha}{2\alpha-1} \cdot \frac{x}{\mu}\right)_+^{\frac{\alpha}{\alpha-1}} > \mu + \frac{1}{2}. \quad (85)$$

for any $\alpha > 1/2$ (See Appendix D for a simple proof). Therefore, unfortunately, the approach of this section cannot improve the result in Subsection IV-C.

C. Improved Inequalities for Fixed Variance

For large variance σ^2, Massey's original inequality (71) reads $H(X) \leqslant \frac{1}{2} \log(2\pi e(\sigma^2 + \frac{1}{12})) < \frac{1}{2} \log(2\pi e\sigma^2) + \frac{\log e}{24\sigma^2}$. Now (59) together with Poisson's formula (67) 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 10: For any integer-valued X of variance $\sigma^2 > 0$,

$$H(X) < \frac{1}{2} \log(2\pi e\sigma^2) + \frac{2 \log e}{e^{2\pi^2\sigma^2} - 1}. \quad (86)$$

Proof: Using the first Fourier transform pair of Table I in Poisson's formula (67) one obtains

$$\frac{1}{\sqrt{2\pi\sigma^2}} \sum_{x \in \mathbb{Z}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} = 1 + 2 \sum_{x=1}^{+\infty} e^{-2(\pi\sigma x)^2} \cos 2\pi\mu x \quad (87)$$

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 (59) and using the inequality $\log(1 + z) < (\log e)z$ (when $z > 0$) gives the result. ∎

Example 6: As a illustration, consider a binomial $X \sim \mathcal{B}(n, p)$ of variance $\sigma^2 = npq$ (where $p + q = 1$). The best known upper bound on $H(X)$ is [26, Eq. (7)]

$$H(X) < \frac{1}{2} \log(2\pi enpq) + \frac{\log e}{12n} + \frac{\log(pq)}{2n} + \frac{\log e}{6npq} \quad (88)$$

which (86) considerably improves for large n since all $O(\frac{1}{n})$ terms are replaced by $O(e^{-2\pi^2 npq})$:

$$H(X) < \frac{1}{2} \log(2\pi enpq) + \frac{2 \log e}{e^{2\pi^2 npq} - 1}. \quad (89)$$

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

Corollary 5: If the integer-valued variable $X \in \mathbb{N}$ is non-negative and $\frac{\mu}{\sigma^2}$ is bounded by a constant $< 2\pi$, then for large enough σ^2,

$$H(X) < \frac{1}{2} \log(2\pi e\sigma^2). \quad (90)$$

Proof: Apply (59) where the sum can be taken only over $x \in \mathbb{N}$. Then by (87),

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

To obtain (90) 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 7: As an example, if $X \sim \mathcal{P}(\lambda)$ is Poisson-distributed then $\frac{\mu}{\sigma^2} = \frac{\lambda}{\lambda} = 1 < 2\pi$ so that for large enough λ,

$$H(X) < \frac{1}{2} \log(2\pi e\lambda). \quad (91)$$

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

Example 8: Similarly, if $X \sim \mathcal{B}(n, p)$ 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 5, for large enough n,

$$H(X) < \frac{1}{2} \log(2\pi enpq). \tag{92}$$

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

Remark 15: For the last two examples, Takano's strong central limit theorem [27, Thm. 2] implies that

$$H(X) = \frac{1}{2}\log(2\pi e\sigma^2) + o\left(\frac{1}{\sigma^{1+\varepsilon}}\right) \tag{93}$$

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

We now illustrate the use of the Poisson summation formula (67) in (61) for α-entropies, in the two cases $\alpha = \frac{1}{2}$ and $\alpha = \frac{2}{3}$.

Lemma 2: One has the following Poisson summation formulas:

$$Z'_{\frac{1}{2}} = \frac{1}{\pi\sigma}\sum_{x\in\mathbb{Z}} \frac{1}{1+(\frac{x-\mu}{\sigma})^2} = 1 + 2\sum_{x=1}^{+\infty} e^{-2\pi\sigma x}\cos 2\pi\mu x. \tag{94}$$

$$Z'_{\frac{2}{3}} = \frac{2}{\pi\sigma}\sum_{x\in\mathbb{Z}} \frac{1}{(1+(\frac{x-\mu}{\sigma})^2)^2} = 1 + 2\sum_{x=1}^{+\infty}(1+2\pi\sigma x)e^{-2\pi\sigma x} \\ \times \cos 2\pi\mu x. \tag{95}$$

Proof: By (154) 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 (67) 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 (94).

By (154) 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 (67) 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 (95). ∎

In the two cases $\alpha = \frac{1}{2}$ and $\frac{2}{3}$, the Massey-type inequalities (73) and (74) write $H_{\frac{1}{2}}(X) \leqslant \frac{1}{2}\log\left(4\pi^2(\sigma^2 + \frac{1}{12})\right) < \log(2\pi\sigma) + \frac{\log e}{24\sigma^2}$ and $H_{\frac{2}{3}}(X) \leqslant \frac{1}{2}\log\left(\frac{64}{27}\pi^2(\sigma^2 + \frac{1}{12})\right) < \log(\frac{8}{3\sqrt{3}}\pi\sigma) + \frac{\log e}{24\sigma^2}$, 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 11: For any integer-valued X of variance $\sigma^2 > 0$,

$$H_{\frac{1}{2}}(X) < \log(2\pi\sigma) + \frac{2\log e}{e^{2\pi\sigma}-1} \tag{96}$$

$$H_{\frac{2}{3}}(X) < \log\left(\frac{8\pi\sigma}{3\sqrt{3}}\right) + \frac{4(1+\pi\sigma)\log e}{e^{2\pi\sigma}-1}. \tag{97}$$

Proof: The sum in the r.h.s. of (94) is bounded by $\sum_{x\geqslant 1}e^{-2\pi\sigma x} = \frac{1}{e^{2\pi\sigma}-1}$. Substituting in (61) and using the inequality $\log(1+z) < (\log e)z$ (when $z > 0$) gives (96).

Likewise, the sum in the r.h.s. of (95) 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 (61) and using the inequality $\log(1+z) < (\log e)z$ (when $z > 0$) gives (97). ∎

Remark 16: Using the Poisson summation formula on other Fourier transform pairs, it is possible to generalize Theorem 11 to any value of the form $\alpha = \frac{k+1}{k+2}$ $(k = 0, 1, \ldots)$ and prove that

$$H_{\frac{k+1}{k+2}}(X) < \log(c_k\pi\sigma) + O(\sigma^k e^{-2\pi\sigma}) \tag{98}$$

where the constant c_k is given by

$$c_k = 4\sqrt{2k+1}\binom{2k}{k}\left(\frac{k+1}{2(2k+1)}\right)^{k+1}. \tag{99}$$

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.

VI. APPLICATION TO GUESSING

A. Improved Massey's Inequality for Guessing

Inequality (78) can be thought of as an improvement of Massey's inequality for the guessing entropy [2]. To see this, let $\mathcal{G}(X)$ be the number of successive guesses of some (discrete valued) secret X before the actual value of X is found, and define the *guessing entropy* as the minimum average number of guesses for a given probability distribution of X:

$$G(X) \triangleq \min \mathbb{E}\big(\mathcal{G}(X)\big). \tag{100}$$

Massey's original inequality reads [2]

$$G(X) \geqslant 2^{H(X)-2} + 1 \quad \text{when } H(X) \geqslant 2 \text{ bits.} \tag{101}$$

A more general situation described by Arikan in [7] is when one guesses X given the observed output Y of some side channel. The corresponding (conditional) guessing entropy is [7]

$$G(X|Y) \triangleq \mathbb{E}\big(G(X|Y=y)\big) \tag{102}$$

where the expectation is over Y's distribution.

Theorem 12 (Improvement of Massey's Inequality): When $H(X)$ or $H(X|Y)$ is expressed in bits,

$$G(X) > \frac{2^{H(X)}}{e} + \frac{1}{2}. \tag{103}$$

$$G(X|Y) > \frac{2^{H(X|Y)}}{e} + \frac{1}{2}. \tag{104}$$

Proof: As explained in [2] the optimal strategy leading to the minimum (100) requires k guesses with probability

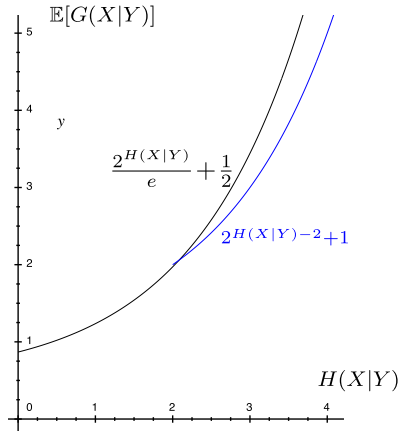$$\mathbb{P}(\mathcal{G}(X) = k) = p_{(k)} \quad (k = 1, 2, \ldots) \tag{105}$$

Fig. 4. Massey's original (blue) and improved (black) lower bounds.

where $p_{(k)}$ is the kth largest probability in X's distribution. Applying (78) to $\mathcal{G}(X) - 1 \geqslant 0$, and noting that $\mu = G(X) - 1$ and $H(\mathcal{G}(X)) = H(X)$ yields

$$H(X) < \log\big(e(G(X) - \tfrac{1}{2})\big) \tag{106}$$

which is (103). Applying (103) to $X|Y = y$ for every y, taking the expectation over Y's distribution and applying Jensen's inequality to the exponential function gives (104). ∎

Remark 17: Inequality (103) improves Massey's original inequality (101) as soon as $H(X) \geqslant \log \frac{2e}{4-e} \approx 2.0846\ldots$ bits and is also valid for $H(X) < 2$ bits. Fig. 4 shows that the improvement over Massey's original inequality is particularly important for large values of entropy, by the factor $4/e$. It is quite startling to notice that the approach followed by Massey back in the 1970s [1] can improve the result of his 1994 paper [2] so much.

Massey's inequality was already improved by the author, with a very different proof, in the (weaker) form $G(X|Y) > \frac{2^{H(X|Y)}}{e}$, see [28] and [29]. See also [29], [30] for a different kind of improvement.

Inequality (103) or (104) can be shown to be the best among all possible bounds of the form $G > a \cdot b^H + c$ [9]. In particular, for large values of entropy, the gain factor $\frac{4}{e}$ of (103) over (101) is optimal, as well as the additive constant $\frac{1}{2}$.

B. Generalization to Rényi Entropies

In this Subsection, we consider Rényi's entropy $H_\alpha(X)$ as well as Arimoto's conditional entropy $H_\alpha(X|Y)$ [31], [32] of order $\alpha > 0$ which finds natural application to guessing with side information [7], [33], [34].

Theorem 13: When $H_\alpha(X)$ and $H_\alpha(X|Y)$ are expressed in bits, for any $\alpha > \frac{1}{2}$,

$$G(X) > \frac{2^{H_\alpha(X)}}{(1+\frac{\alpha-1}{\alpha})^{\frac{\alpha}{\alpha-1}}} + \frac{1}{2} = (1-\tfrac{1-\alpha}{\alpha})^{\frac{\alpha}{1-\alpha}} \cdot 2^{H_\alpha(X)} + \frac{1}{2}. \tag{107}$$

$$G(X|Y) > \frac{2^{H_\alpha(X|Y)}}{(1+\frac{\alpha-1}{\alpha})^{\frac{\alpha}{\alpha-1}}} + \frac{1}{2} = (1-\tfrac{1-\alpha}{\alpha})^{\frac{\alpha}{1-\alpha}} \cdot 2^{H_\alpha(X|Y)} + \frac{1}{2}. \tag{108}$$

Proof: Similarly as in the preceding Subsection VI-A, the $\mu + \frac{1}{2}$ term in (79) is replaced by $G(X) - \frac{1}{2}$, and one immediately obtains (107).

Arimoto's conditional α-entropy [31] satisfies $H_\alpha(X|Y) = \frac{\alpha}{1-\alpha} \log \mathbb{E} \exp \frac{1-\alpha}{\alpha} H_\alpha(X|Y = y)$. Thus if $H_\alpha(X|Y)$ is expressed in bits, one has

$$2^{H_\alpha(X|Y)} = \left(\mathbb{E}\, 2^{\frac{1-\alpha}{\alpha} H_\alpha(X|Y=y)}\right)^{\frac{\alpha}{1-\alpha}} \tag{109}$$

where the expectation is over Y's distribution. Applying (107) to $X|Y = y$ for every y, taking the expectation over Y's distribution and applying Jensen's inequality to the function $x \mapsto x^{\frac{\alpha}{1-\alpha}}$, which is strictly convex when $\alpha > \frac{1}{2}$, gives (108). ∎

Remark 18: Since the factor $(1 + \frac{\alpha-1}{\alpha})^{\frac{\alpha}{\alpha-1}}$ converges to e as $\alpha \to 1$, Theorem 12 is recovered by letting $\alpha \to 1$. This factor is nonincreasing in α, and since $1 + x < e^x$ for $x \neq 0$, the term $(1 + \frac{\alpha-1}{\alpha})^{\frac{\alpha}{\alpha-1}} = \frac{1}{(1-\frac{1-\alpha}{\alpha})^{\frac{\alpha}{1-\alpha}}}$ is greater than e for $\alpha < 1$ and less than e for $\alpha > 1$. Since $H_\alpha(X)$ is also nonincreasing in α, none of the inequalities (107) (or (108)) is a trivial consequence of another for a different value of α.

Example 9: Thus, referring to Example 5,

$$G(X) > \frac{1}{4} 2^{H_{\frac{2}{3}}(X)} + \frac{1}{2} \tag{110}$$

$$G(X) > \frac{8}{27} 2^{H_{\frac{3}{4}}(X)} + \frac{1}{2} \tag{111}$$

$$G(X) > \frac{4}{9} 2^{H_2(X)} + \frac{1}{2} \tag{112}$$

and similarly for $X|Y$, where $\frac{9}{4} < e < \frac{27}{8} < 4$.

Remark 19: By Remark 14, no inequality of the type (107) or (108) can generally hold for $\alpha \leqslant \frac{1}{2}$. This does not contradict Arikan's inequality [7] for the limiting case $\alpha = \frac{1}{2}$, which reads

$$G(X|Y) \geqslant \frac{2^{H_{\frac{1}{2}}(X|Y)}}{1 + \ln M}, \tag{113}$$

because it was established when X takes a *finite* number M of possible values. As $M \to +\infty$ the r.h.s. vanishes. In other words, it is impossible to improve Arikan's inequality (113) with some positive constant independent of M.

C. Arikan-Type Inequalities for Rényi Entropies of Small Orders

By Remark 14 and 19, the results of the previous subsection cannot generalize to $\alpha \leqslant \frac{1}{2}$. However, when X takes values in a *finite* alphabet of size M, Arikan's inequality (113) for $\alpha = \frac{1}{2}$ and extensions of it for $\alpha < \frac{1}{2}$ can still be obtained using Theorem 1 (equation (26)) applied to $\mathcal{G}(X)$, on top of the α-Kullback inequality (Theorem 3). In this case the density (31) has to be constrained in an interval of finite length which depends on M.

A derivation is as follows. Recall that $\mathcal{G}(X) \geqslant 1$ has mean $G(X)$ and α-entropy $H_\alpha(X)$. For simplicity consider $\mathcal{U}$ to be zero-mean, uniformly distributed in $(-\frac{1}{2}, \frac{1}{2})$, so that $\mathcal{X}$ has the

same mean $G(X)$ and is supported in the interval $(\frac{1}{2}, M+\frac{1}{2})$. Now consider

$$\varphi(x) = \frac{x^{\frac{1}{\alpha-1}}}{Z} \qquad (114)$$

restricted in the same interval $(\frac{1}{2}, M+\frac{1}{2})$. Then (30) gives $H_\alpha(X) = h_\alpha(\mathcal{X}) < \frac{\alpha}{1-\alpha} \log G(X) + \log Z_\alpha$, where the strictness of the inequality follows from the fact that $\mathcal{X} = X + \mathcal{U}$ (which has a staircase density) cannot have density φ. Since $\alpha < 1$ the latter inequality reads

$$G(X) > \frac{2^{\frac{1-\alpha}{\alpha} H_\alpha(X)}}{Z_\alpha^{\frac{1-\alpha}{\alpha}}} \qquad (115)$$

where $Z_\alpha = \int_{1/2}^{M+1/2} x^{-\frac{\alpha}{1-\alpha}} \, dx$. In particular we have the following

Corollary 6 (Arikan's Inequality [7], Slightly Improved): For $\alpha = \frac{1}{2}$,

$$G(X) > \frac{2^{H_{\frac{1}{2}}(X)}}{\ln(2M+1)} \qquad (116)$$

$$G(X|Y) > \frac{2^{H_{\frac{1}{2}}(X|Y)}}{\ln(2M+1)}. \qquad (117)$$

Proof: Plugging $Z_{\frac{1}{2}} = \ln\frac{M+1/2}{1/2}$ in (115) gives (116). Since by (109), $2^{H_{\frac{1}{2}}(X|Y)} = \mathbb{E}\, 2^{H_{\frac{1}{2}}(X|Y=y)}$, this immediately gives (117). ∎

Remark 20: Inequality (117) slightly improves Arikan's original inequality (113) for $M > 1$ because $\ln(2M+1) < \ln(eM) = \ln M + 1$. It can be found from Arikan's derivation [7] that the optimal constant in the denominator is in fact $1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{M} = \ln M + 0.5772\ldots + O(\frac{1}{M})$ (see [34, Eqn. (47)]). Here $\ln(2M+1) = \ln M + 0.6931\ldots + O(\frac{1}{M})$ is not optimal but fairly close.

For even smaller Rényi orders we have the following

Corollary 7: For any $0 < \alpha < \frac{1}{2}$,

$$G(X) > \left(1 - \frac{\alpha}{1-\alpha}\right)^{\frac{1-\alpha}{\alpha}} \cdot \frac{2^{\frac{1-\alpha}{\alpha} H_\alpha(X) + \frac{1-2\alpha}{\alpha}}}{(2M+1)^{\frac{1-2\alpha}{\alpha}}} \qquad (118)$$

$$G(X|Y) > \left(1 - \frac{\alpha}{1-\alpha}\right)^{\frac{1-\alpha}{\alpha}} \cdot \frac{2^{\frac{1-\alpha}{\alpha} H_\alpha(X|Y) + \frac{1-2\alpha}{\alpha}}}{(2M+1)^{\frac{1-2\alpha}{\alpha}}}. \qquad (119)$$

Proof: One has $Z_\alpha = \int_{1/2}^{M+1/2} x^{-\frac{\alpha}{1-\alpha}} \, dx < (M+\frac{1}{2})^{\frac{1-2\alpha}{1-\alpha}}$. Plugging this in (115) gives (118). The second inequality then follows from (109), which reads $2^{\frac{1-\alpha}{\alpha} H_\alpha(X|Y)} = \mathbb{E}\, 2^{\frac{1-\alpha}{\alpha} H_\alpha(X|Y=y)}$. ∎

Example 10: For any M-ary random variable X,

$$G(X) > \frac{2^{2H_{\frac{1}{3}}(X)}}{2(2M+1)} \qquad (120)$$

$$G(X) > \frac{32}{27} \cdot \frac{2^{3H_{\frac{1}{4}}(X)}}{(2M+1)^2} \qquad (121)$$

and similarly for $X|Y$.

Remark 21: The method of this Subsection also works for $\frac{1}{2} < \alpha < 1$. In this case Z_α is bounded by $\frac{1-\alpha}{2\alpha-1} 2^{\frac{2\alpha-1}{1-\alpha}}$ (independently of M) and applying (115) gives

$$G(X) > \left(\frac{\alpha}{1-\alpha} - 1\right)^{\frac{1-\alpha}{\alpha}} \cdot 2^{\frac{1-\alpha}{\alpha} H_\alpha(X) - \frac{2\alpha-1}{\alpha}} \qquad (122)$$

$$G(X|Y) > \left(\frac{\alpha}{1-\alpha} - 1\right)^{\frac{1-\alpha}{\alpha}} \cdot 2^{\frac{1-\alpha}{\alpha} H_\alpha(X|Y) - \frac{2\alpha-1}{\alpha}}. \qquad (123)$$

However, it can be verified that these inequalities are always weaker than (107) and (108), respectively. This is not surprising since the derivation of the latter in the preceding subsection used, instead of (114), the optimal α-exponential density achieving equality in (30).

D. Generalization to Guessing Moments

While entropy $H(X)$ is generalized by the α-entropy $H_\alpha(X)$ for any $\alpha > 0$, the guessing entropy $G(X)$ can be generalized by the ρ-guessing entropy for any $\rho > 0$, defined as the ρth order moment [7]

$$G_\rho(X) \triangleq \min \mathbb{E}\big(\mathcal{G}^\rho(X)\big). \qquad (124)$$

Again the minimum occurs when the guessing function is a ranking function: $\mathcal{G}(x) = k$ iff $p(x) = \mathbb{P}(X = x)$ is the kth largest probability in X's distribution. The conditional version given side information Y is given by [7]

$$G_\rho(X|Y) \triangleq \mathbb{E}\big(G_\rho(X|Y = y)\big). \qquad (125)$$

Theorem 14: When $H(X)$ is expressed in bits,

$$G_\rho(X) > \frac{2^{\rho H(X)}}{\rho\big(\Gamma(1 + \frac{1}{\rho})\big)^\rho e} \qquad (126)$$

$$G_\rho(X|Y) > \frac{2^{\rho H(X|Y)}}{\rho\big(\Gamma(1 + \frac{1}{\rho})\big)^\rho e}. \qquad (127)$$

Proof: Applying Theorem 1 to $\mathcal{G}(X)$ for uniformly distributed $\mathcal{U}$ over the interval $(-1, 0)$, one has $0 \leqslant \mathcal{X} = \mathcal{G}(X) + \mathcal{U} \leqslant \mathcal{G}(X)$ with $h(\mathcal{X}) = H(\mathcal{G}(X)) = H(X)$. Since $\theta = \mathbb{E}(\mathcal{X}^\rho) \leqslant \mathbb{E}(\mathcal{G}^\rho(X)) = G_\rho(X)$, (37) of Theorem 4 (one-sided case) gives (126). The inequality is strict because the staircase density of $\mathcal{G}(X) + \mathcal{U}$ cannot coincide with the (one-sided) α-Gaussian achieving equality in (37). Applying (126) to $X|Y = y$ for every y, taking the expectation over Y's distribution and applying Jensen's inequality to the exponential function gives (127). ∎

Remark 22: During the revision process of this paper, the author became aware that (126) (with an additional $o(1)$ term as $G_\rho(X) \to +\infty$) was obtained by Weinberger and Shayevitz in [35, Lemma 2] using a similar method.

Remark 23: For $\rho = 1$ we recover (103) without the additive constant $1/2$. This suboptimality comes from the fact that $\theta = \mathbb{E}(\mathcal{X}^\rho) = \mathbb{E}\big((\mathcal{G}(X) + \mathcal{U})^\rho\big)$ cannot be determined as a function of $\mathbb{E}(\mathcal{G}^\rho(X)) = G_\rho(X)$ alone when $\rho \neq 1$.

Example 11: For any discrete random variable X,

$$G_2(X) > 2 \cdot \frac{2^{2H(X)}}{\pi e} \qquad (128)$$

$$G_4(X) > 8 \cdot \frac{2^{4H(X)}}{G^2 \pi^3 e} \qquad (129)$$

and similarly for $X|Y$, where $G = 0.834626841674\ldots$ is Gauss's constant.

For α-entropies we have the following

Theorem 15: When $H_\alpha(X)$ and $H_\alpha(X|Y)$ are expressed in bits, and $\alpha > \frac{1}{1+\rho}$,

$$G_\rho(X) > \begin{cases} \dfrac{2^{\rho H_\alpha(X)}}{\left(\frac{(1+\rho)\alpha-1}{1-\alpha}\right)\left(\frac{\rho\alpha}{(1+\rho)\alpha-1}\right)^{\frac{\rho}{1-\alpha}}\left(\frac{\Gamma(\frac{1}{\rho}+1)\Gamma(\frac{1}{1-\alpha}-\frac{1}{\rho})}{\Gamma(\frac{1}{1-\alpha})}\right)^\rho} \\ \qquad\qquad \text{for } \frac{1}{1+\rho} < \alpha < 1; \\[4pt] \dfrac{2^{\rho H_\alpha(X)}}{\left(\frac{(1+\rho)\alpha-1}{\alpha-1}\right)\left(\frac{(1+\rho)\alpha-1}{\rho\alpha}\right)^{\frac{\rho}{\alpha-1}}\left(\frac{\Gamma(\frac{1}{\rho}+1)\Gamma(\frac{\alpha}{\alpha-1})}{\Gamma(\frac{\alpha}{\alpha-1}+\frac{1}{\rho})}\right)^\rho} \\ \qquad\qquad \text{for } \alpha > 1, \end{cases}$$

$$(130)$$

$$G_\rho(X|Y) > \begin{cases} \dfrac{2^{\rho H_\alpha(X|Y)}}{\left(\frac{(1+\rho)\alpha-1}{1-\alpha}\right)\left(\frac{\rho\alpha}{(1+\rho)\alpha-1}\right)^{\frac{\rho}{1-\alpha}}\left(\frac{\Gamma(\frac{1}{\rho}+1)\Gamma(\frac{1}{1-\alpha}-\frac{1}{\rho})}{\Gamma(\frac{1}{1-\alpha})}\right)^\rho} \\ \qquad\qquad \text{for } \frac{1}{1+\rho} < \alpha < 1; \\[4pt] \dfrac{2^{\rho H_\alpha(X|Y)}}{\left(\frac{(1+\rho)\alpha-1}{\alpha-1}\right)\left(\frac{(1+\rho)\alpha-1}{\rho\alpha}\right)^{\frac{\rho}{\alpha-1}}\left(\frac{\Gamma(\frac{1}{\rho}+1)\Gamma(\frac{\alpha}{\alpha-1})}{\Gamma(\frac{\alpha}{\alpha-1}+\frac{1}{\rho})}\right)^\rho} \\ \qquad\qquad \text{for } \alpha > 1. \end{cases}$$

$$(131)$$

Proof: The proof of (130) is similar to the proof of (126) in Theorem 14 using inequality (36) of Theorem 4 (one-sided case). For (131) one proceeds as in the proof of Theorem 13 using (109) and the fact that $x \mapsto x^{\frac{\rho\alpha}{1-\alpha}}$ is strictly convex for all $\alpha > \frac{1}{1+\rho}$. ∎

Example 12: For any discrete random variable X,

$$G_2(X) > \frac{2^{2H_{1/2}(X)}}{\pi^2} \tag{132}$$

$$G_2(X) > \frac{27}{16} \cdot \frac{2^{2H_{2/3}(X)}}{\pi^2} \tag{133}$$

$$G_2(X) > \frac{36}{125} \cdot 2^{2H_2(X)} \tag{134}$$

$$G_2(X) > 3 \cdot \frac{2^{2H_3(X)}}{\pi^2} \tag{135}$$

$$G_3(X) > \frac{9}{2} \cdot \frac{2^{3H_{1/2}(X)}}{\sqrt{3}\,\pi^3} \tag{136}$$

$$G_3(X) > \frac{512}{2401} \cdot 2^{3H_2(X)} \tag{137}$$

$$G_4(X) > \frac{2^{4H_{1/3}(X)}}{G^4\pi^4} \tag{138}$$

$$G_4(X) > \frac{27}{4} \cdot \frac{2^{4H_{1/2}(X)}}{\pi^4} \tag{139}$$

$$G_4(X) > \frac{823543}{82944} \cdot \frac{2^{4H_{2/3}(X)}}{\pi^4} \tag{140}$$

$$G_4(X) > \frac{10000}{59049} \cdot 2^{4H_2(X)} \tag{141}$$

$$G_4(X) > \frac{80}{9} \cdot \frac{2^{4H_5(X)}}{G^4\pi^4} \tag{142}$$

and similarly for $X|Y$, where $G = 0.834626841674\ldots$ is Gauss's constant.

Remark 24: The reason why simple closed-form lower bounds on guessing entropy are obtained is due to the fact that Massey's approach uses bounds on continuous α-entropies. Such simple lower bounds could not obtained by previous methods [34, Rmk. 5].

Remark 25: While Theorem 15 shows that $G_\rho(X)$ can always be lower-bounded by an exponential function of $H_\alpha(X)$ for any $\alpha > \frac{1}{1+\rho}$, such an inequality is impossible for $\alpha \leqslant \frac{1}{1+\rho}$ in general (when the number of possible values of X is infinite). In fact, when X has distribution $\mathbb{P}(X = k) = \frac{c}{(k\log k)^{\rho+1}}$ and $\alpha \leqslant \frac{1}{1+\rho}$, the series $\sum \frac{1}{k(\log k)^{\rho+1}}$ converges—hence $G_\rho(X)$ is finite—while the series $\sum \frac{1}{(k\log k)^{\alpha(\rho+1)}}$ diverges so that $H_\alpha(X) = +\infty$.

As already remarked in [36, p. 476], Arikan's inequality [7] on $G_\rho(X)$:

$$G_\rho(X) \geqslant \frac{2^{\rho H_{\frac{1}{1+\rho}}(X)}}{1+\ln M}, \tag{143}$$

(and similarly for $X|Y$) is for the limiting case $\alpha = \frac{1}{1+\rho}$, but is valid only when X takes a finite number M of possible values. In a manner similar to was done in [34], it is always possible to use the method of Subsection VI-C to obtain inequalities of this kind for any $\alpha \leqslant \frac{1}{1+\rho}$.

VII. CONCLUSION

Simple bounds on the differential entropy or Rényi entropy for a given fixed parameter (such as mean or 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 a given fixed parameter (such as mean or variance) appears as a basic question in information theory. This paper has shown that using Massey's approach, many simple, closed-form bounds on discrete entropies or Rényi entropies can be deduced from bounds on the α-entropies of a continuous distribution. One can envision that many similar derivations can be done for other types of parameter constraints.

Massey's approach gives, in particular, simple lower bounds on the guessing entropy or guessing moments, which are exponential in Rényi (or Rényi-Arimoto) entropies of any order $\alpha > 0$, not just $\alpha = 1$ (Massey's inequality) of $\alpha = \frac{1}{1+\rho}$ (Arikan's inequality). Since similar upper bounds also exist for $\alpha = \frac{1}{1+\rho}$ [7], [34], [37] it would be interesting to similarly upper bound guessing for other values of α in order to obtain

tight evaluations in practical applications where a divide-and-conquer strategy is used [8] to guess a large secret from many small ones.

Finally, a variant of Massey's approach together with some Fourier analysis proves very tight "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 [27] 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.

APPENDIX A
REZA'S EQUIVALENCE EXTENDED TO RÉNYI ENTROPIES

Consider a continuous variable $\mathcal{X}$ having density f, and quantize it to obtain the discrete X with step size Δ, in such a way that

$$p(x_k) = \mathbb{P}(X = x_k) = \int_{k\Delta}^{(k+1)\Delta} f(x)\, \mathrm{d}x \qquad (144)$$

and the discrete values x_k correspond to mean values

$$f(x_k) = \frac{1}{\Delta}\int_{k\Delta}^{(k+1)\Delta} f(x)\, \mathrm{d}x = \frac{p(x_k)}{\Delta}. \qquad (145)$$

Proposition 1: If f is continuous within each bin of length Δ and the integral (in (4) or in (6)) defining $h_\alpha(X)$ exists, then

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

The assumptions are satisfied in particular when f is continuous and compactly supported.

Proof: By the continuity assumption, the values (145) are well defined and given by the mean value theorem. Since the integral in (4) (resp. (6)) converges and f is piecewise continuous, $f\log f$ (resp. f^α) is Riemann-integrable. It follows that the integral in (4) and in (6) can be respectively approximated by the Riemann sum

$$\sum_k \Delta \cdot f(x_k)\log\frac{1}{f(x_k)} = \sum_k p(x_k)\log\frac{\Delta}{p(x_k)}$$
$$= H(X) + \log\Delta \qquad (146)$$

$$\tfrac{1}{1-\alpha}\log\sum_k \Delta \cdot f^\alpha(x_k) = \tfrac{1}{1-\alpha}\log\sum_k \Delta^{1-\alpha}p^\alpha(x_k)$$
$$= H_\alpha(X) + \log\Delta, \qquad (147)$$

which tends to $h(\mathcal{X})$ (resp. $h_\alpha(\mathcal{X})$) as $\Delta \to 0$. ∎

APPENDIX B
MASSEY'S EQUIVALENCE EXTENDED TO RÉNYI ENTROPIES AND ARBITRARY STEP SIZE

Proof of Theorem 1: The density of $\mathcal{X} = X + \mathcal{U}$ is a mixture of the form

$$f(x) = \sum_{k\in\mathbb{Z}} p(x_k)\,\chi(x - x_k) \qquad (148)$$

where x_k are the regularly spaced values of X and χ is the density of $\mathcal{U}$. The terms in the sum have disjoint supports. Since entropy is invariant by translation, we may always assume that χ is supported in the interval $[0, \Delta]$. Splitting the integral in (4) or in (6) into parts over intervals $[x_k, x_{k+1} = x_k + \Delta]$ we obtain

$$h(\mathcal{X}) = \sum_k p(x_k)\int \chi(x - x_k)\log\frac{1}{p(x_k)\chi(x - x_k)}\, \mathrm{d}x$$

$$= \sum_k p(x_k)\Big[\underbrace{\int\chi}_{=1}\Big]\log\frac{1}{p(x_k)} + \Big[\underbrace{\sum_k p(x_k)}_{=1}\Big]\int\chi\log\frac{1}{\chi}$$

$$h_\alpha(\mathcal{X}) = \frac{1}{1-\alpha}\log\sum_k p(x_k)^\alpha \int\chi(x - x_k)^\alpha\, \mathrm{d}x$$

$$= \frac{1}{1-\alpha}\log\sum_k p(x_k)^\alpha \int\chi^\alpha$$

$$= \frac{1}{1-\alpha}\log\sum_k p(x_k)^\alpha + \frac{1}{1-\alpha}\log\int\chi^\alpha \qquad (149)$$

which proves (23). ∎

Remark 26: The above proof follows the textbook solution [6] to exercice 8.7 of [5] in the case $\alpha = 1$. (A similar calculation appears in [24, Proof of Thm. 3].) In this particular case, an even simpler proof is as follows.

Proof of Theorem 1 ($\alpha = 1$): 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

$$H(X) = H(X) - H(X|\mathcal{X})$$
$$= I(X; \mathcal{X})$$
$$= h(\mathcal{X}) - h(\mathcal{X}|X)$$
$$= h(\mathcal{X}) - h(\mathcal{U}), \qquad (150)$$

which proves (23). ∎

APPENDIX C
PROOF OF THEOREM 4 AND ITS COROLLARIES

We first prove Theorem 4 and then deduce Corollaries 1 and 2 as particular cases.

Set $T(x)$ in the form $T(x) = 1 + \beta\frac{|x|^\rho}{\theta}$ so that $m = 1 + \beta$ and β is such that (31) has finite ρth-order moment $\theta = \mathbb{E}(|\mathcal{X}|^\rho)$. In order that $\varphi(x) = \frac{1}{Z}\big(1 + \beta\frac{|x|^\rho}{\theta}\big)^{\frac{1}{\alpha-1}}$ be integrable, it is necessary that β has the same sign as $1 - \alpha$.

For $\alpha > 1$ ($\beta < 0$), the density is supported in the interval $|x| < \sqrt[\rho]{\frac{\theta}{|\beta|}}$ so that $1 + \beta\frac{|x|^\rho}{\theta} \geqslant 0$. In this case we write $\varphi(x) = \frac{1}{Z}\big(1 - |\beta|\frac{|x|^\rho}{\theta}\big)_+^{\frac{1}{\alpha-1}}$ with the notation $(X)_+ = \max(X, 0)$.

For $\alpha < 1$, the existence of a finite variance implies that the integral of $\big(1 + \beta\frac{|x|^\rho}{\theta}\big)^{1-\frac{1}{1-\alpha}}$ converges at infinity, which requires $\alpha > \frac{1}{1+\rho}$.

In either case, β is such that φ has ρth moment θ, that is, such that (34) holds, hence $m = 1 + \beta = \frac{Z_\alpha}{Z}$. Now we can

write $Z = \sqrt[\rho]{\frac{\theta}{|\beta|}} I\left(\frac{1}{\alpha-1}\right)$ and $Z_\alpha = \sqrt[\rho]{\frac{\theta}{|\beta|}} I\left(\frac{\alpha}{\alpha-1}\right)$ where

$$
I(\gamma) \triangleq \begin{cases}
\displaystyle\int_{-\infty}^{+\infty} \frac{\mathrm{d}x}{(1+|x|^\rho)^{-\gamma}} = \frac{2}{\rho}\int_0^1 (1-t)^{-\gamma-\frac{3}{2}} t^{-\frac{1}{2}}\, \mathrm{d}t \\[4pt]
\qquad = \dfrac{\frac{2}{\rho}\Gamma(\frac{1}{\rho})\Gamma(-\gamma-\frac{1}{\rho})}{\Gamma(-\gamma)} \qquad \text{for } \gamma < 0; \\[14pt]
\displaystyle\int_{-1}^1 (1-|x|^\rho)^\gamma\, \mathrm{d}x = \frac{2}{\rho}\int_0^1 (1-t)^\gamma t^{-\frac{1}{2}}\, \mathrm{d}t \\[4pt]
\qquad = \dfrac{\frac{2}{\rho}\Gamma(\frac{1}{\rho})\Gamma(\gamma+1)}{\Gamma(\gamma+1+\frac{1}{\rho})} \qquad \text{for } \gamma > 0.
\end{cases}
$$
(151)

Here we have made the change of variables $t = \frac{x^p}{1+x^p}$ and $t = x^p$, respectively, for $x > 0$, and recognized Euler integrals of the first kind. In either case, letting $\gamma = \frac{1}{\alpha-1}$,

$$
m = \frac{Z_\alpha}{Z} = \frac{I(\gamma+1)}{I(\gamma)} = \frac{-\gamma-1}{-\gamma-1-\frac{1}{\rho}} = \frac{\rho\alpha}{(\rho+1)\alpha-1},
$$
(152)

hence $\beta = \frac{1-\alpha}{(\rho+1)\alpha-1}$. Plugging this and the expression of Z into that of φ gives the expression of the generalized α-Gaussian density [36][2]:

$$
\varphi(x) = \begin{cases}
\sqrt[\rho]{\dfrac{\beta}{\theta}}\, \dfrac{\Gamma(\frac{1}{1-\alpha})}{2\Gamma(1+\frac{1}{\rho})\Gamma(\frac{1}{1-\alpha}-\frac{1}{\rho})}\, \dfrac{1}{\left(1+\beta\frac{|x|^\rho}{\theta}\right)^{\frac{1}{1-\alpha}}} \\
\qquad\qquad\qquad \text{for } \frac{1}{1+\rho} < \alpha < 1; \\[14pt]
\sqrt[\rho]{\dfrac{|\beta|}{\theta}}\, \dfrac{\Gamma(\frac{\alpha}{\alpha-1}+\frac{1}{\rho})}{2\Gamma(1+\frac{1}{\rho})\Gamma(\frac{\alpha}{\alpha-1})}\left(1-|\beta|\frac{|x|^\rho}{\theta}\right)_+^{\frac{1}{\alpha-1}} \\
\qquad\qquad\qquad \text{for } \alpha > 1,
\end{cases}
$$
(153)

and plugging (152) and the expression of Z_α or Z into (30) or (35) gives (36). ∎

Corollary 1 follows by setting $\rho = 2$ for the centered variable $X - \mu_X$. The corresponding expression of the α-Gaussian density (with $\beta = \frac{1-\alpha}{3\alpha-1}$) is [21]

$$
\varphi(x) = \begin{cases}
\sqrt{\dfrac{\beta}{\pi\sigma_X^2}}\, \dfrac{\Gamma(\frac{1}{1-\alpha})}{\Gamma(\frac{1}{1-\alpha}-\frac{1}{2})}\, \dfrac{1}{\left(1+\beta(\frac{x-\mu_X}{\sigma_X})^2\right)^{\frac{1}{1-\alpha}}} \\
\qquad\qquad\qquad \text{for } \frac{1}{3} < \alpha < 1; \\[14pt]
\sqrt{\dfrac{|\beta|}{\pi\sigma_X^2}}\, \dfrac{\Gamma(\frac{\alpha}{\alpha-1}+\frac{1}{2})}{\Gamma(\frac{\alpha}{\alpha-1})}\left(1-|\beta|(\frac{x-\mu_X}{\sigma_X})^2\right)_+^{\frac{1}{\alpha-1}} \\
\qquad\qquad\qquad \text{for } \alpha > 1,
\end{cases}
$$
(154)

and Z_α is given by

$$
Z_\alpha = \frac{\sigma_X}{\sqrt{|\beta|}} I\left(\frac{\alpha}{\alpha-1}\right) = \begin{cases}
\sqrt{\dfrac{\pi\sigma_X^2(3\alpha-1)}{1-\alpha}}\, \dfrac{\Gamma(\frac{\alpha}{1-\alpha}-\frac{1}{2})}{\Gamma(\frac{\alpha}{1-\alpha})} & \text{for } \alpha < 1; \\[14pt]
\sqrt{\dfrac{\pi\sigma_X^2(3\alpha-1)}{\alpha-1}}\, \dfrac{\Gamma(\frac{\alpha}{\alpha-1}+1)}{\Gamma(\frac{\alpha}{\alpha-1}+\frac{3}{2})} & \text{for } \alpha > 1.
\end{cases}
$$
(155)

[2]There is a misprint in the expression of the generalized α-Gaussian in [36], p.474] where $\beta(\frac{1}{p},\frac{1}{1-\lambda})$ should read $\beta(\frac{1}{p},\frac{\lambda}{\lambda-1})$.

Corollary 2 follows by setting $\rho = 1$, where the multiplying factor 2 in the above expressions is removed due to the one-sided constraint $X \geqslant 0$. The corresponding expression of the "α-exponential" density (with $\beta = \frac{1-\alpha}{2\alpha-1}$) for $x > 0$ is [38, §II.B]

$$
\varphi(x) = \begin{cases}
\dfrac{\beta}{\mu_X}\dfrac{\alpha}{1-\alpha}\dfrac{1}{\left(1+\beta\frac{x}{\mu_X}\right)^{\frac{1}{1-\alpha}}} & \text{for } \frac{1}{2} < \alpha < 1 \\[12pt]
\dfrac{|\beta|}{\mu_X}\dfrac{\alpha}{\alpha-1}\left(1-|\beta|\frac{x}{\mu_X}\right)_+^{\frac{1}{\alpha-1}} & \text{for } \alpha > 1.
\end{cases}
$$
(156)

and Z_α is given by

$$
Z_\alpha = \frac{\mu_X}{|\beta|}I\left(\frac{\alpha}{\alpha-1}\right) = \frac{\mu_X}{\beta}\frac{1-\alpha}{2\alpha-1} = \mu_X.
$$
(157)

APPENDIX D
PROOF OF INEQUALITY (85)

Let $s = \frac{\alpha}{|\alpha-1|}$ and $a = \frac{2\alpha-1}{|1-\alpha|}\mu$. Then (85) is equivalent to

$$
\sum_{x\in\mathbb{N}}\left(1+\frac{x}{a}\right)^{-s} > \frac{a}{s-1} + \frac{1}{2} \qquad (\tfrac{1}{2} < \alpha < 1)
$$
(158)

$$
\sum_{x\in\mathbb{N}}\left(1-\frac{x}{a}\right)_+^s > \frac{a}{s+1} + \frac{1}{2} \qquad (\alpha > 1)
$$
(159)

This is proved by applying the following Lemma to $f(x) = \left(1+\frac{x}{a}\right)^{-s}$ and $\left(1-\frac{x}{a}\right)_+^s$, respectively.

Lemma 3: Let f be nonnegative decreasing and strictly convex. Then

$$
\sum_{x\in\mathbb{N}} f(x) > \frac{f(0)}{2} + \int_0^{+\infty} f(x)\, \mathrm{d}x.
$$
(160)

Proof: Let $g(x)$ be the piecewise linear function defined for all $x \geqslant 0$ that linearly interpolates the values of f over the integers. Then $\int_0^{+\infty} f(x)\, \mathrm{d}x < \int_0^{+\infty} g(x)\, \mathrm{d}x = \sum_{x\in\mathbb{N}} \frac{f(x)+f(x+1)}{2} = \sum_{x\in\mathbb{N}} f(x) - \frac{f(0)}{2}$. ∎

ACKNOWLEDGMENT

REFERENCES

[1] J. L. Massey, "On the entropy of integer-valued random variables," in *Proc. Beijing, China Int. Workshop Inf. Theory*, Jul. 1988, pp. C-1.1–C-1.4.

[2] J. L. Massey, "Guessing and entropy," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 1994, p. 204.

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

[4] R. J. McEliece, *The Theory of Information and Coding*, 1st ed. and 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2002.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 1st ed. and 2nd ed. Hoboken, NJ, USA: Wiley, 2006.

[6] T. M. Cover and J. A. Thomas, "Elements of information theory: Solutions to problems," Aug. 2007. [Online]. Available: https://cpb-us-w2.wpmucdn.com/sites.gatech.edu/dist/c/565/files/2017/01/solutions2.pdf

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

[8] M. O. Choudary and P. G. Popescu, "Back to Massey: Impressively fast, scalable and tight security evaluation tools," in *Proc. 19th Workshop Cryptograph. Hardw. Embedded Syst. (CHES)*, in Lecture Notes in Computer Science, vol. 10529, 2017, pp. 367–386.

[9] A. Tănăsescu, M. O. Choudary, O. Rioul, and P. G. Popescu, "Tight and scalable side-channel attack evaluations through asymptotically optimal Massey-like inequalities on guessing entropy," *Entropy*, vol. 23, no. 11, pp. 1–10, 2021.

[10] O. Rioul, "This is IT: A primer on Shannon's entropy and information," in *Proc. 23rd Inf. Theory, Poincaré (Bourbaphy) Seminar*, in Progress in Mathematical Physics, vol. 78, B. Duplantier and V. Rivasseau, Eds. Basel, Switzerland: Birkhäuser, Aug. 2021, pp. 49–86.

[11] S. Kullback, "Certain inequalities in information theory and the Cramér–Rao inequality," *Ann. Math. Statist.*, vol. 25, no. 4, pp. 745–751, Dec. 1954.

[12] I. N. Sanov, "On the probability of large deviations of random variables," (in Russian), *Matematicheskii*, vol. 42, no. 1, pp. 11–44, 1957.

[13] S. Kullback, *Information Theory and Statistics*, 1st ed. and 2nd ed. Hoboken, NJ, USA: Wiley, 1968.

[14] 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.

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

[16] R. Sundaresan, "Guessing under source uncertainty," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 269–287, Jan. 2007.

[17] J. C. Principe, D. Xu, and J. W. Fisher, III, *Information-Theoretic Learning* (Unsupervised Adaptive Filtering), vol. 1. Hoboken, NJ, USA: Wiley, Sep. 2000, ch. 7.

[18] O. Rioul, "Rényi entropy power and normal transport," in *Proc. Int. Symp. Inf. Theory Appl. (ISITA)*, Oct. 2020, pp. 1–5.

[19] C. Bunte and A. Lapidoth, "Maximizing Rényi entropy rate," *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1193–1205, Mar. 2016.

[20] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 4, pp. 623–656, Oct. 1948.

[21] 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* (Lecture Notes in Computer Science), vol. 2683, A. Rangarajan, M. Figueiredo, and J. Zerubia, Eds. Berlin, Germany: Springer, 2003, pp. 211–226.

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

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

[24] O. Rioul and J. Magossi, "On Shannon's formula and Hartley's rule: Beyond the mathematical coincidence," *Entropy*, vol. 16, no. 9, pp. 4892–4910, Sep. 2014.

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

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

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

[28] E. de Chérisey, S. Guilley, P. Piantanida, and O. Rioul, "Best information is most successful: Mutual information and success rate in side-channel analysis," *IACR Trans. Cryptograph. Hardw. Embedded Syst.*, vol. 2019, no. 2, pp. 49–79, 2019.

[29] A. Tănăsescu and P. G. Popescu, "Exploiting the Massey gap," *Entropy*, vol. 22, no. 1398, pp. 1–9, Dec. 2020.

[30] P. G. Popescu and M. O. Choudary, "Refinement of Massey inequality," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2019, pp. 495–496.

[31] S. Arimoto, "Information measures and capacity of order α for discrete memoryless channels," in *Proc. 2nd Colloq. Math. Soc. János Bolyai Topics Inf. Theory*, no. 16, I. Csiszár and P. Elias, Eds. Keszthely, Hungary: Bolyai, 1977, pp. 41–52.

[32] S. Fehr and S. Berens, "On the conditional Rényi entropy," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6801–6810, Nov. 2014.

[33] I. Sason and S. Verdú, "Arimoto-Rényi conditional entropy and Bayesian M-ary hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 64, no. 1, pp. 4–25, Jan. 2018.

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

[35] N. Weinberger and O. Shayevitz, "Guessing with a bit of help," *Entropy*, vol. 22, no. 39, pp. 1–26, 2020.

[36] E. Lutwak, D. Yang, and G. Zhang, "Cramér–Rao and moment-entropy inequalities for Rényi entropy and generalized Fisher information," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 473–478, Feb. 2005.

[37] S. Boztaş, "Comments on 'An inequality on guessing and its application to sequential decoding,'" *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 2062–2063, Nov. 1997.

[38] C. Bunte and A. Lapidoth, "Maximizing Rényi entropy rate," in *Proc. IEEE 28th Conv. Elect. Electron. Eng. Israel*, Dec. 2014.

Olivier Rioul (Member, IEEE) graduated from École Polytechnique, Paris, France, in 1987, and École Nationale Supérieure des Télécommunications, Paris, in 1989. He received the Ph.D. degree from the École Nationale Supérieure des Télécommunications, Paris, in 1993. He is currently a Full Professor at the Department of Communication and Electronics, Télécom Paris, Institut Polytechnique de Paris, France. He has been teaching information theory at various universities for almost 20 years. He has published a textbook which has become a classical French reference in the field. His research interests are in applied mathematics and include various, sometimes unconventional, applications of information theory, such as inequalities in statistics, hardware security, and experimental psychology.