The interplay between error, total variation, alpha-entropy and guessing: Fano and Pinsker direct and reverse inequalities

Olivier Rioul

Article

The Interplay between Error, Total Variation, Alpha-Entropy and Guessing: Fano and Pinsker Direct and Reverse Inequalities §

Olivier Rioul

LTCI, Télécom Paris, Institut Polytechnique de Paris, 91120 Palaiseau, France; olivier.rioul@telecom-paris.fr

§ This review paper, essentially of tutorial nature with some original material in the various inequalities, is the extended version of the communication at the 41st International Conference on Bayesian and Maximum Entropy methods in Science and Engineering (MaxEnt 2022) conference, which was previously published in Rioul, O. What Is Randomness? The Interplay between Alpha Entropies, Total Variation and Guessing. *Phys. Sci. Forum* **2022**, *5*, 1–9.

Abstract: Using majorization theory via "Robin Hood" elementary operations, optimal lower and upper bounds are derived on Rényi and guessing entropies with respect to either error probability (yielding reverse-Fano and Fano inequalities) or total variation distance to the uniform (yielding reverse-Pinsker and Pinsker inequalities). This gives a general picture of how the notion of randomness can be measured in many areas of computer science.

Keywords: entropy; Rényi entropy; guessing entropy; guessing moments; total variation distance; error probability; data processing inequality; majorization; Schur concavity; Fano inequality; Pinsker inequality

1. Introduction

In many areas of science, it is of primary importance to assess the "randomness" of a certain random variable X. That variable could represent, for example, a cryptographic key, a signature, some sensitive data, or any type of intended secret. For simplicity, we assume that X is an M-ary discrete random variable, taking values in a finite alphabet $\mathcal{X}$ of size M, with known probability distribution $p = (p_1, p_2, \ldots, p_M)$ (in short, $X \sim p$).

Depending on the application, many different criteria can be used to evaluate randomness. Some are information-theoretic, others are related to detection/estimation theory or to hypothesis testing. We review the most common ones in the following subsections.

1.1. Entropy

A "sufficiently random" X is often described as "entropic" in the literature. The usual notion of entropy is the Shannon entropy [1]

$$H(X) = H(p) \triangleq \sum_k p_k \cdot \log \frac{1}{p_k}, \tag{1}$$

which is classically thought of as a measure of "uncertainty". It has, however, an operational definition in the fields of data compression or source coding. The problem is to find the binary description of X with the shortest average description length or "coding rate".

Note that the base of the logarithm is not specified in (1). Similar to all information-theoretic quantities, the choice of the base determines the unit of information. Logarithms of base 2 give binary units (bits) or Shannons (Sh). Logarithms of base 10 give decimal units (dits) or Hartleys. Natural logarithms (base e) give natural units (nats).

This compression problem can be seen as equivalent to a "game of 20 questions" § 5.7.1 in [2], where a binary codeword for X is identified as a sequence of answers to yes–no

questions about X that uniquely identifies it. There is no limitation on the type of questions asked, except that they must be answered by yes (1) or no (0). The goal of the game is to minimize the average number of questions, which is equal to the coding rate. It is well known, since Shannon [1], that the entropy $H(X)$ is a lower bound on the coding rate that can be achieved asymptotically for repeated descriptions.

In this perspective, entropy is a natural measure of efficient (lossless) compression rate. A highly random variable (with high entropy) cannot be compressed too much without losing information: "random" means "hard to compress".

1.2. Guessing Entropy

Another perspective arises in cryptography when one wants to guess a secret key. The situation is similar to the "game of 20 questions" of the preceding subsection. The difference is that the only possibility is to actually try out one possible key hypothesis at a time. In other words, yes–no questions are restricted to be of the form "is X equal to x?" until the correct value has been found. The optimal strategy that minimizes the average number of questions is to guess the values of X in order of decreasing probabilities: first, the value with maximum probability $p_{(1)}$, then the second maximum $p_{(2)}$, and so on. The corresponding minimum average number of guesses is the guessing entropy [3] (also known as "guesswork" [4]):

$$G(X) = G(p) \triangleq \sum_k p_{(k)} \cdot k. \tag{2}$$

Massey [3] has shown that the guessing entropy G is exponentially increasing as entropy H increases. A recent improved inequality is [5,6] $G > \frac{\exp H}{e} + \frac{1}{2}$. It is sometimes convenient to use $\log G$ instead of G, to express it in the same logarithmic unit of information as entropy H.

In this perspective, a highly random variable (with high guessing entropy) cannot be guessed rapidly: "random" means "hard to guess".

1.3. Coincidence or Collision

Another perspective is to view X as a (publicly available) "identifier", "fingerprint" or "signature" obtained by a randomized algorithm from some sensitive data. In such a scheme, to prevent "collision attacks", it is important to ensure that X is "unique" in the sense that there is only a small chance that another independent X' obtained by the same randomized algorithm coincides with X. Since X and X' are i.i.d., the "index of coincidence" $\mathbb{P}(X = X') = \sum_k p_k^2$ should be as small as possible, that is, the complementary quantity (sometimes called quadratic entropy [7]):

$$R_2(X) = R_2(p) \triangleq \mathbb{P}(X \neq X') = 1 - \sum_k p_k^2, \tag{3}$$

should be as large as possible. In the context of hash functions, this is called "universality" (Chapter 8 in [8]). The corresponding logarithmic measure is known as the collision entropy (Rényi entropy [9] of order 2, also known as quadratic entropy [10]):

$$H_2(X) = H_2(p) \triangleq \log \frac{1}{1 - R_2(X)} = \log \frac{1}{\sum_k p_k^2} \tag{4}$$

which should also be as large as possible. By concavity of the logarithm, $\sum_k p_k \log p_k \leq \log \sum_k p_k^2$, that is, $H \geq H_2$; hence, high collision entropy implies high entropy.

In this perspective, a highly random variable (with high collision entropy) cannot be found easily by coincidence: "random" means "unique" or "hard to collide".

1.4. Estimation Error

In estimation or detection theory, one observes some disclosed data which may depend on X and tries to estimate X from the observation. The best estimator $\hat{x}$ minimizes the probability of error, $\mathbb{P}(X \neq \hat{x}) = 1 - \mathbb{P}(X = \hat{x})$. Therefore, given the observation, the best estimation is the value x with highest probability $p_{(1)}$, and the minimum probability of error is written:

$$\mathbb{P}_e(X) = \mathbb{P}_e(p) \triangleq 1 - \max p = 1 - p_{(1)}. \tag{5}$$

If X is meant to be kept secret, then this probability of error should be as large as possible. The corresponding logarithmic measure is known as the min-entropy:

$$H_\infty(X) = H_\infty(p) \triangleq \log \frac{1}{1 - \mathbb{P}_e(X)} = \log \frac{1}{p_{(1)}} \tag{6}$$

which should also be as large as possible. It is easily seen that $H \geq H_2 \geq H_\infty$; hence, high min-entropy implies high entropy in all the previous senses.

In this perspective, a highly random variable (with high min-entropy) cannot be efficiently estimated: "random" means "hard to estimate" or "hard to detect".

Figure 1 illustrates various randomness measures for a binary distribution.

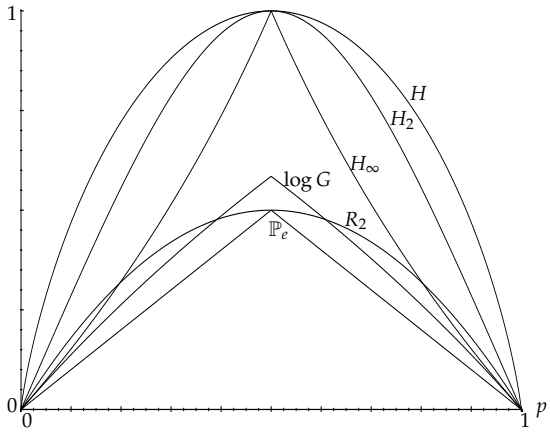

Figure 1. Various randomness measures (in bits) for a binary distribution $(p, 1-p)$ as a function of p.

1.5. Some Generalizations

One can generalize the above concepts in multiple ways. We only mention a few.

The α-entropy, or Rényi entropy of order $\alpha > 0$, is defined as follows [9]:

$$H_\alpha(X) = H_\alpha(p) \triangleq \frac{1}{1-\alpha} \log \sum_k p_k^\alpha = \frac{\alpha}{1-\alpha} \log \|p\|_\alpha \tag{7}$$

where $\|\cdot\|_\alpha$ is the "α-norm" (strictly speaking, $\|\cdot\|_\alpha$ is a norm only when $\alpha \geq 1$). The Shannon entropy $H = H_1$ is recovered in the limiting case $\alpha \to 1$, the collision entropy H_2 is recovered in the case $\alpha = 2$, and the min-entropy H_∞ is recovered in the limiting case $\alpha \to \infty$.

The ρ-guessing entropy, or guessing moment [11] of order $\rho > 0$, is defined as the minimum ρth-order moment of the number of guesses needed to find X. The same optimal strategy as for the guessing entropy yields the following:

$$G_\rho(X) = G_\rho(p) \triangleq \sum_k p_{(k)} \cdot k^\rho, \tag{8}$$

which generalizes $G = G_1$ for $\rho \neq 1$. Arikan [11] has shown that $\log G_\rho$ behaves asymptotically as $\rho H_{\frac{1}{1+\rho}}$. In particular, $\log G$ behaves asymptotically as the ½-entropy $H_{\frac{1}{2}}$.

In some cryptographic scenarios, one has the ability to estimate or guess X in a given maximum number m of tries. The corresponding error probability takes the form $\mathbb{P}(X \neq \hat{x}_1, X \neq \hat{x}_2, \ldots, X \neq \hat{x}_m)$. The same optimal strategy as for guessing entropy G_ρ yields an error probability of order m:

$$\mathbb{P}_e^m(X) = \mathbb{P}_e^m(p) \triangleq 1 - p_{(1)} - p_{(2)} - \cdots - p_{(m)}, \tag{9}$$

which generalizes $\mathbb{P}_e = \mathbb{P}_e^1$ for $m > 1$.

One obtains similar randomness measures by replacing p with its "negation" $\bar{p}$, as explained in [12].

1.6. "Distances" to the Uniform

A fairly common convention is that, if we "draw at random" X, it is assumed that we sample it according to a uniform distribution unless otherwise explicitly indicated. Thus, the uniform distribution u, where all possible outcomes being equally likely—all M values have equal probability $u_k = \frac{1}{M}$ for all k—is considered as the ideal randomness.

From this viewpoint, a variable X with distribution p should be all the more "random" as p is "close to uniform": randomness can be measured as some complementary "distance" from p to the uniform u, in the form, say, $d_{\max} - d(p, u)$, where "distance" d has maximum value $d_{\max}$. Such $d(p, u)$ should not necessarily obey all axioms of a mathematical distance, but at least should be nonnegative and vanish only when $p = u$.

Many of the above entropic criteria fall into this category. For example:

$$H(p) = \log M - D(p\|u), \tag{10}$$

where $D(p\|q) = \sum_k p_k \log \frac{p_k}{q_k}$ denotes the (Kullback–Leibler) divergence (or "distance"). More generally:

$$H_\alpha(p) = \log M - D_\alpha(p\|u), \tag{11}$$

where $D_\alpha(p\|q) = \frac{1}{\alpha - 1} \log \sum_k p_k^\alpha q_k^{1-\alpha}$ denotes the (Rényi) α-divergence [13].

In the particular case $\alpha = 2$, since $\sum_k (p_k - \frac{1}{M})^2 = \sum_k p_k^2 - \frac{1}{M}$, the complementary index of coincidence R_2—hence, the collision entropy H_2—is also related to the squared 2-norm distance to the uniform:

$$R_2(p) = (1 - \tfrac{1}{M}) - \|p - u\|_2^2. \tag{12}$$

It follows that the 2-norm distance is related to the 2-divergence by the formula $D_2(p\|u) = \log(1 + M\|p - u\|_2^2)$ (see, e.g., Lemma 3 in [14]).

Similarly, in the particular case $\alpha = \frac{1}{2}$, one can write $H_{\frac{1}{2}}(p) = 2\log(1 + R_{\frac{1}{2}}(p))$, where

$$R_{\frac{1}{2}}(p) = \sum_k \sqrt{p_k} - 1 \tag{13}$$

$$= \sqrt{M}\Big((1 - \tfrac{1}{\sqrt{M}}) - \tfrac{1}{2}\|\sqrt{p} - \sqrt{u}\|_2^2\Big) \tag{14}$$

is a complementary quantity of the squared *Hellinger distance* $\frac{1}{2}\|\sqrt{p} - \sqrt{u}\|_2^2$, which is related to the $\frac{1}{2}$-divergence by the formula $D_{1/2}(p\|u) = -2\log(1 - \frac{1}{2}\|\sqrt{p} - \sqrt{u}\|_2^2)$.

Another important example is given next.

1.7. Statistical Distance to the Uniform

Suppose one wants to design a statistical experiment to know whether X follows either distribution p (null hypothesis H_0) or another distribution q (alternate hypothesis). Any statistical test takes the form "is $X \in T$?": if yes, then accept H_0; otherwise, reject it. Type-I and type-II errors have total probability $\mathbb{P}(X \notin T) + \mathbb{Q}(X \in T)$, where $\mathbb{P}$, $\mathbb{Q}$ are the probability measures corresponding to p and q, respectively. Clearly, if $|\mathbb{P}(X \in T) - \mathbb{Q}(X \in T)|$

is small enough, the two hypotheses p and q are indistinguishable in the sense that decision errors have total probability arbitrarily close to 1.

The statistical (total variation) distance § 8.8 in [8] is defined as follows:

$$\Delta(p,q) = \max_T |\mathbb{P}(T) - \mathbb{Q}(T)| = \frac{1}{2}\|p-q\|_1, \tag{15}$$

where the $\frac{1}{2}$ factor is present to ensure that $0 \leq \Delta(p,q) \leq 1$. The maximum in the definition of the statistical distance:

$$\Delta(p,q) = \max_T |\mathbb{P}(T) - \mathbb{Q}(T)| = \mathbb{P}(T_+) - \mathbb{Q}(T_+) \tag{16}$$

is attained for any event T_+, satisfying the following:

$$\{p > q\} \subset T_+ \subset \{p \geq q\}. \tag{17}$$

The statistical distance is particularly important from a hypothesis testing viewpoint, since, as we have just seen, a very small distance $\Delta(p,q)$ ensures that no statistical test can distinguish the two hypotheses p and q.

Following the discussion of the preceding subsection, we can define "statistical randomness" as the complementary value of the statistical distance $\Delta(p,u)$ between p and the uniform distribution u. Therefore, if $q = u$ is uniform and letting $K = |T_+|$, then $\Delta(p,u) = \mathbb{P}(T_+) - \frac{K}{M}$ has maximum value $1 - \frac{1}{M}$ and statistical randomness can be defined as follows:

$$R(X) = R(p) \triangleq (1 - \tfrac{1}{M}) - \Delta(p,u) = (1 - \tfrac{1}{M}) - \tfrac{1}{2}\|p-u\|_1. \tag{18}$$

This is similar to (12), where half the 1-norm is used in place of the squared 2-norm.

From the hypothesis testing perspective, it follows that a high statistical randomness R ensures that no statistical test can effectively distinguish between the actual distribution and the uniform. This is, for example, the usual criterion used to evaluate randomness extractors in cryptology. Since equiprobable values are the least predictable, a highly random variable cannot be easily statistically predicted: "random" means "hard to predict".

1.8. Conditional Versions

In many applications, the randomness of X is evaluated after observing some disclosed data or side information Y. The observed random variable Y can model any type of data and is not necessarily discrete. The conditional probability distribution of X having observed $Y = y$ is denoted by $p_{X|y}$ to distinguish it from the unconditional distribution $p = p_X$ (without side information). By the law of total probability $\mathbb{P}(X = x) = \mathbb{E}_y \mathbb{P}(X = x|Y = y)$, p_X is recovered by averaging all conditional distributions:

$$p_X = \mathbb{E}_y p_{X|y}, \tag{19}$$

where $\mathbb{E}_y$ denotes the expectation operator over Y.

The "conditional randomness" of X given Y can then be defined as the average randomness measure of $X|y$ over all possible observations, that is, the expectation over Y of all randomness measures of $X|Y = y$. For example, Shannon's conditional entropy or equivocation [1] is given by the following:

$$H(X|Y) \triangleq \mathbb{E}_y H(X|y) = \mathbb{E}_y H(p_{X|y}). \tag{20}$$

Similarly:

$$G(X|Y) \triangleq \mathbb{E}_y G(X|y) = \mathbb{E}_y G(p_{X|y}) \tag{21}$$

gives the average minimum number of guesses to find X after having observed Y. Additionally:

$$R_2(X|Y) \triangleq \mathbb{E}_y R_2(X|y) = \mathbb{E}_y R_2(p_{X|y}) \tag{22}$$

gives the average probability of non-collision to identify X upon observation of Y, and

$$\mathbb{P}_e(X|Y) \triangleq \mathbb{E}_y \mathbb{P}_e(X|y) = 1 - \mathbb{E}_y \max p_{X|y} \tag{23}$$

gives the minimum average probability of error, as achieved by the maximum a posteriori (MAP) decision rule. The "conditional statistical randomness" is likewise defined as shown:

$$R(X|Y) \triangleq \mathbb{E}_y R(X|y) = \mathbb{E}_y R(p_{X|y}). \tag{24}$$

For the generalized quantities of Section 1.5, the conditional ρ-guessing entropy is given by the following:

$$G_\rho(X|Y) \triangleq \mathbb{E}_y G_\rho(X|y) = \mathbb{E}_y G_\rho(p_{X|y}) \tag{25}$$

and the conditional mth-order probability of error is as below:

$$\mathbb{P}_e^m(X|Y) \triangleq \mathbb{E}_y \mathbb{P}_e^m(X|y) = \mathbb{E}_y \mathbb{P}_e^m(p_{X|y}). \tag{26}$$

For α-entropy, however, many different definitions of conditional α-entropy have been proposed in the literature [15]. The preferred choice for most applications seems to be Arimoto's definition [16]:

$$H_\alpha(X|Y) \triangleq \frac{\alpha}{1-\alpha} \log \mathbb{E}_y \|p_{X|y}\|_\alpha, \tag{27}$$

where the expectation over Y is taken on the α-norm inside the logarithm and not outside. Shannon's conditional entropy $H(X|Y)$ is recovered in the limiting case $\alpha \to 1$. One nice property of Arimoto's definition is that it is compatible with that of $\mathbb{P}_e(X|Y)$ in the limiting case $\alpha \to \infty$, since the relation $H_\infty = \log \frac{1}{1-\mathbb{P}_e}$ of (6) naturally extends to conditional quantities:

$$H_\infty(X|Y) = \log \frac{1}{1 - \mathbb{P}_e(X|Y)}. \tag{28}$$

Notice that for any order $\alpha \neq 1$, Arimoto's definition can be rewritten as a simple expectation of $\varphi_\alpha(H_\alpha)$ instead of H_α:

$$\varphi_\alpha(H_\alpha(X|Y)) = \mathbb{E}_y \varphi_\alpha(H_\alpha(p_{X|y})), \tag{29}$$

where φ_α is the increasing function, defined as follows:

$$\varphi_\alpha(x) \triangleq \operatorname{sgn}(1-\alpha) \exp\left(\frac{1-\alpha}{\alpha} x\right). \tag{30}$$

The requirement that φ_α is increasing is important in the following. The signum term was introduced so that φ_α is increasing, not only for $0 < \alpha < 1$, but also for $\alpha > 1$. The exponential function exp is assumed to the same base as the logarithm: $\exp x = 2^x$ for x in bits, 10^x in dits, e^x in nats). In what follows, we indifferently refer to H_α or $\varphi_\alpha(H_\alpha)$.

1.9. Aim and Outline

The enumeration in the preceding subsections is by no means exhaustive. Every subfield or application has its preferred criterion, either information/estimation theoretic or statistical, conditioned on some observations or not. Clearly, all these randomness measures share many properties.

Therefore, a natural question is to determine a (possibly minimal) set of properties that characterize all possible randomness measures. Many axiomatic approaches have

been proposed for entropy [1,17], α-entropy [9], information leakage [18] or conditional entropy [19,20].

Extending the work in [21], Section 2 presents a simple alternative, which naturally encompass all common randomness measures H, H_α, G, G_ρ, $\mathbb{P}_e$, $\mathbb{P}_e^m$, R_2 and R, based on two natural axioms:

- Equivalent random variables are equally random;
- Knowledge reduces randomness (on average).

Many properties, shared by all randomness measures described above, are deduced from these two axioms.

Another important issue is to study the relationship between randomness measures, by establishing the exact locus or joint range of two such measures among all probability distributions with tight lower and upper bounds. In this paper, extending the presentation made in [21], we establish the optimal bounds relating information-theoretic (e.g., entropic) quantities on one hand and statistical quantities (probability of error and statistical distance) on the other hand.

Section 3 establishes general optimal Fano and reverse-Fano inequalities, relating any randomness measure to the probability of error. This generalizes Fano's original inequality [22] $H(X|Y) \leq (1-\mathbb{P}_e(X|Y)) \log \frac{1}{1-\mathbb{P}_e(X|Y)} + \mathbb{P}_e(X|Y) \log \frac{M-1}{\mathbb{P}_e(X|Y)}$, which has become ubiquitous in information theory (e.g., to derive converse channel coding theorems) and in statistics (e.g., to derive lower bounds on the maximum probability of error in multiple hypothesis testing).

Section 4 establishes general optimal Pinsker and reverse-Pinsker inequalities, relating any randomness measure to the statistical randomness or the statistical distance to the uniform. Generally speaking, Pinsker and reverse-Pinsker inequalities relate some divergence measure (e.g., $d(p\|q)$ or $d_\alpha(p\|q)$) between two distributions to their statistical distance $\Delta(p,q)$. Here, following the discussion in Section 1.6, we restrict ourselves to the divergence or distance to the uniform distribution $q = u$. (For the general case of arbitrary distributions p, q see, e.g., the historical perspective on Pinsker–Schützenberger inequalities in [23].). In this context, we improve the well-known Pinsker inequality [24,25], which reads $D(p\|u) = \log M - H(p) \geq 2 \log e \cdot \|p-u\|_1^2$. This inequality, of more general applicability for any distributions p, q, is no longer optimal in the particular case $q = u$.

Finally, Section 5 lists some applications in the literature, and Section 6 gives some research perspectives.

2. An Axiomatic Approach

Let X be any M-ary random variable with distribution p_X. How should a measure of "randomness" $\mathfrak{R}(X) \in \mathbb{R}$ of X be defined in general? To simplify the discussion, we assume that $\mathfrak{R}(X) \geq 0$ is nonnegative.

As advocated by Shannon [26], such a notion should not depend on the particular "reversible encoding" of X. In other words, any two equivalent random variables should have the same measure $\mathfrak{R}(X)$, where equivalence is defined as follows.

Definition 1 (Equivalent Variables)**.** *Two random variables X and Y are* equivalent*: $X \equiv Y$, if there exist two mappings f and g, such that $Y = f(X)$ a.s. (almost surely, i.e., with probability one) and $X = g(Y)$ a.s.*

Remark 1 (Equivalent Measures)**.** *Obviously, it is also essentially equivalent to study $\mathfrak{R}(X)$ or $\mathfrak{R}(X)^2$, for example, or any quantity of the form $\varphi(\mathfrak{R}(X))$, where $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ is any increasing (invertible) function.*

Definition 2 (Conditional Randomness)**.** *Given any random variable Y, the conditional form of $\mathfrak{R}$ is defined as follows:*

$$\mathfrak{R}(X|Y) = \mathbb{E}_y \mathfrak{R}(X|y) \tag{31}$$

where $X|y$ (or $X|Y=y$) denotes the random variable X, conditioned of the event $Y=y$. This quantity represents the average amount of randomness of X knowing Y.

Remark 2 (Equivalent Conditional Measures)**.** *Again, it is essentially equivalent to study $\mathfrak{R}(X|Y)$ or $\varphi(\mathfrak{R}(X|Y))$, where $\varphi:\mathbb{R}_+\to\mathbb{R}_+$ is any increasing function. One may, therefore, generalize the notion of conditional randomness by writing $\varphi(\mathfrak{R}(X|Y))=\mathbb{E}_y\varphi(\mathfrak{R}(X|y))$ in place of* (31)*, the same as* (29) *for α-entropy. However, in the sequel, we stay with the basic Definition 2 and simply assume that $\varphi(\mathfrak{R})$ is considered instead of $\mathfrak{R}$ whenever it is convenient to do so.*

In the sequel, we study the implications of only two axioms:

Axiom 1 (Equivalence)**.** $X\equiv Y \implies \mathfrak{R}(X)=\mathfrak{R}(Y)$

Axiom 2 (Knowledge Reduces Randomness)**.**

$$\mathfrak{R}(X|Y)\le\mathfrak{R}(X). \tag{32}$$

We find such postulates quite intuitive and natural. First, equivalent random variables should be equally random. Second, knowledge of some side observation should, *on average*, reduces randomness.

All randomness quantities described in Section 1 obviously satisfy Axiom 1. That they also satisfy Axiom 2 is shown in the following examples.

Example 1 (Entropies)**.** *For Shannon's entropy H, the inequality $H(X|Y)\le H(X)$ is well known Thm 2.6.5 in [2]. This is often paraphrased as "conditioning reduces entropy", "knowledge reduces uncertainty" or "information can't hurt". The difference $H(X)-H(X|Y)=I(X;Y)$ is the mutual information, which is always nonnegative. Inequality $H_\alpha(X|Y)\le H_\alpha(X)$ is also known to hold for any $\alpha>0$, see [15,16] and Example 4 below.*

Example 2 (Guessing Entropies)**.** *Axiom 2 for the guessing entropies G or G_ρ can be easily checked from their definition, as follows.*

Let $N\in\mathbb{N}=\{1,2,\dots\}$ be any random variable giving the number of guesses needed to find X in any guessing strategy. N is equivalent to X (Definition 1) since every value of N corresponds to a unique value of X, and vice versa. By definition, $G_\rho(X)=\min_{N\equiv X}\mathbb{E}(N^\rho)$, where the minimum is over all possible $N\in\mathbb{N}$ equivalent to X (corresponding to all possible strategies). Now, $G_\rho(X|Y)=\mathbb{E}_yG_\rho(X|y)\le\mathbb{E}_y\mathbb{E}(N^\rho|y)=\mathbb{E}(N^\rho)$, by the law of total expectation. Taking the minimum over $N\equiv X$ gives $G_\rho(X|Y)\le G_\rho(X)$, which is Axiom 2.

The case $\rho=1$ was already shown in [27]. The result is quite intuitive: any side information Y can only improve the guess of X.

Example 3 (Error Probabilities)**.** *Axiom 2 for the error probability $\mathbb{P}_e=\mathbb{P}_e^1$ follows from the corresponding inequality for $H_\infty=\log\frac{1}{1-\mathbb{P}_e}$ (see* (28) *and Example 1 for $\alpha=\infty$), but it can also be checked directly from its definition, as well as in the case of $\mathbb{P}_e^m$ of order m, as follows.*

The mth order error probability is $\mathbb{P}_e^m(X)=\min_{\hat{x}_1,\dots,\hat{x}_m}\mathbb{P}(X\neq\hat{x}_1,X\neq\hat{x}_2,\dots,X\neq\hat{x}_m)$, i.e., the minimum probability that X is not equal to any of the m first estimates $\hat{x}_1,\hat{x}_2,\dots,\hat{x}_m$. Then, $\mathbb{P}_e^m(X|Y)=\mathbb{E}_y\min_{\hat{x}_1,\dots,\hat{x}_m}\mathbb{P}(X\neq\hat{x}_1,\dots,X\neq\hat{x}_m|y)\le\mathbb{E}_y\mathbb{P}(X\neq\hat{x}_1,\dots,X\neq\hat{x}_m|y)=\mathbb{P}(X\neq\hat{x}_1,\dots,X\neq\hat{x}_m)$, by the law of total probability, for every sequence $\hat{x}_1,\dots,\hat{x}_m$. Taking the minimum over such sequences gives $\mathbb{P}_e^m(X|Y)\le\mathbb{P}_e^m(X)$, which is Axiom 2.

The case $m=1$ was already shown, e.g., in [27]. Again, the result is quite intuitive: any side information Y can only improve the estimation of X.

2.1. Symmetry and Concavity

We now rewrite Axioms 1 and 2 as equivalent conditions on probability distributions.

Definition 3 (Probability "Simplex"). *Let $\mathcal{P}$ be the set of all sequences of nonnegative numbers:*

$$p = (p_1, p_2, p_3, \ldots) \tag{33}$$

such that the following are satisfied:

- *Only a finite number of them are positive: $p_k \neq 0$ for finitely many k;*
- *They sum to 1: $\sum_k p_k = 1$.*

Notice that $\mathcal{P}$ has infinite dimension even though only a finite number of components are nonzero in every $p \in \mathcal{P}$. Thus, any $p \in \mathcal{P}$ can be seen as the probability distribution of M-ary random variables with arbitrary large M.

Theorem 1 (Symmetry). *Axiom 1 is equivalent to the condition that $\mathfrak{R}(X) = \mathfrak{R}(p)$ is a symmetric function of $p = (p_1, p_2, p_3, \ldots) \in \mathcal{P}$, identified as the probability distribution of X.*

Proof. Let $\mathcal{X}$ be the finite set ("alphabet") of all values taken by $X \sim p_X$, and let f be an injective mapping from $\mathcal{X}$ to $\mathbb{N} = \{1, 2, \ldots\}$, whose image is a finite subset of $\mathbb{N}$. From Definition 1, X is equivalent to $f(X) \in \mathbb{N}$, with probabilities $p = (p_1, p_2, \ldots)$. Then, by Axiom 1, $\mathfrak{R}(X)$ does not depend on the particular values of $\mathcal{X}$ but only on the corresponding probabilities, so that $\mathfrak{R}(X) = \mathfrak{R}(p)$, where $p \in \mathcal{P}$ is identified to p_X. Now, letting h be any bijection (permutation) of $\mathbb{N}$, Axiom 1 implies that $\mathfrak{R}(p)$ does not depend on the ordering of the p_ks, that is, $\mathfrak{R}(p)$ is a symmetric function of p. Conversely, any bijection applied to X can only change the ordering of the p_ks in $p = p_X$, which leaves $\mathfrak{R}(p) = \mathfrak{R}(X)$ as invariant. □

Accordingly, it is easily checked directly that all expressions in terms of probability distributions p of random measures given in Section 1 are symmetric in p.

Remark 3. *Some authors [17] define $\mathcal{P}$ as the union of all $\mathcal{P}_M$ for $M \in \mathbb{N}$, where $\mathcal{P}_M$ is the M-simplex $\{(p_1, p_2, \ldots, p_M), p_k \geq 0, p_1 + \cdots + p_M = 1\}$. With this viewpoint, even when the expression of $\mathfrak{R}(p)$ does not explicitly depend on M, one has to define $\mathfrak{R}(p)$ separately for all different values of M as a function $\mathfrak{R}_M(p_1, p_2, \ldots, p_M)$, defined over $\mathcal{P}_M$, and further impose the compatibility condition that $\mathfrak{R}_{M+1}(p_1, p_2, \ldots, p_M, 0) = \mathfrak{R}_M(p_1, p_2, \ldots, p_M)$, as in [17] (this is called "expansibility" in [20]).*

Such expansibility condition is unnecessary to state explicitly in our approach: it is an obvious consequence of an appropriate choice of f in Definition 1, namely, the injective embedding of $\{1, 2, \ldots, M\}$ into $\{1, 2, \ldots, M+1\}$.

Theorem 2 (Concavity). *Axiom 2 is equivalent to the condition that $\mathfrak{R}(p)$ is concave in p.*

Proof. Using the notations of Theorem 1, Definition 2 and (19), Axiom 2 can be rewritten as shown:

$$\mathbb{E}_y \mathfrak{R}(p_{X|y}) \leq \mathfrak{R}(p_X) = \mathfrak{R}(\mathbb{E}_y p_{X|y}). \tag{34}$$

This is exactly Jensen's inequality for concave functions on the convex "simplex" $\mathcal{P}$. □

Remark 4 (φ-Concavity). *Similarly as in Remark 2, we may consider $\varphi(\mathfrak{R})$ in place of $\mathfrak{R}$ in the definition of conditional randomness, where $\varphi : \mathbb{R}_+ \to \mathbb{R}$ is any increasing function. Then, by Theorem 2, $\varphi(\mathfrak{R})$ is concave, that is, $\mathfrak{R}(p)$ is a φ-concave function of p (for example, for $\varphi = \log$, one recovers the usual definition of a log-concave function). This is called "core-concavity" in [20].*

Example 4 (Symmetric Concave Measures). *All randomness measures of Examples 1–3 satisfy both Axioms 1 and 2, and are, therefore, symmetric concave in p. This can also be checked directly from certain closed-form expressions given in Section 1:*

- *Shannon's entropy H, as well as the complementary index of coincidence R_2, can be written in the form $\sum_k r(p_k)$, where r is a strictly concave function. Thus, both are symmetric and strictly concave in p;*

- *Statistical randomness $R(p)$ can also be written in this form, where $r(p_k) = -\frac{1}{2}\left|p_k - \frac{1}{M}\right|$ is concave in p_k. Thus, $R(p)$ is also symmetric concave and, therefore, is also an acceptable randomness measure satisfying Axioms 1 and 2;*
- *For α-entropy, consider $\varphi_\alpha(H_\alpha(p)) = \operatorname{sgn}(1-\alpha)\|p\|_\alpha$ where φ_α is the increasing function (30). It is known that the α-norm $\|\cdot\|_\alpha$ is strictly convex for finite $\alpha > 1$ (by Minkowski's inequality) and strictly concave for $0 < \alpha < 1$ (by the reverse Minkowski inequality). Thus, α-entropy is symmetric and (strictly) φ_α-concave in the sense of Remark 4. Therefore, one finds anew that it satisfies Axioms 1 and 2.*

Corollary 1 (Mixing Increases Randomness)**.** *Let $p, q \in \mathcal{P}$ be any two probability distributions and consider the "mixed" distribution $\lambda p + \bar{\lambda} q$, where $\lambda \geq 0$, $\bar{\lambda} \geq 0$, and $\lambda + \bar{\lambda} = 1$. Then:*

$$\mathfrak{R}(\lambda p + \bar{\lambda} q) \geq \lambda \mathfrak{R}(p) + \bar{\lambda} \mathfrak{R}(q). \tag{35}$$

In particular, mixing two equally random distributions $\mathfrak{R}(p) = \mathfrak{R}(q)$ results in a "more random" distribution: $\mathfrak{R}(\lambda p + \bar{\lambda} q) \geq \mathfrak{R}(p) = \mathfrak{R}(q)$.

Proof. Immediate from the concavity of $\mathfrak{R}$. □

Example 5. *The mixing property of the Shannon entropy H is well-known Thm. 2.7.3 in [2]. A well-known thermodynamic interpretation is that mixing two gases of equal entropy results in a gas with higher entropy.*

2.2. Basic Properties in Terms of Random Variables

In terms of random variables, one can deduce the following properties.

Corollary 2 (Consistency)**.** *If X is independent of Y, then $\mathfrak{R}(X|Y) = \mathfrak{R}(X)$. In particular, let 0 denote any deterministic variable (by Defintion 1, any deterministic random variable is equivalent to the constant 0). Then:*

$$\mathfrak{R}(X|0) = \mathfrak{R}(X). \tag{36}$$

Thus "absolute" (unconditional) randomness $\mathfrak{R}(X)$ can be recovered as a special case of conditional randomness.

Proof. If X and Y are independent, then $p_{X|y} = p_X$ for (almost) any y, so that $\mathfrak{R}(X|Y) = \mathbb{E}_y \mathfrak{R}(X|y) = \mathbb{E}_y \mathfrak{R}(X) = \mathfrak{R}(X)$. In particular, X and 0 are always independent. □

Remark 5 (Strict Concavity)**.** *A randomness measure $\mathfrak{R}$ is "strictly concave" in p if Jensen's inequality (34) holds with equality only when $p_{X|y} = p_X$ for almost all y. This can be stated in terms of random variables as follows. For any strictly concave random measure $\mathfrak{R}$, (32) is strict unless independence holds:*

$$\mathfrak{R}(X|Y) = \mathfrak{R}(X) \iff X \text{ is independent of } Y. \tag{37}$$

Example 6 (Strictly Concave Measures)**.** *As already seen in Example 4, entropy H, all α-entropies $\varphi_\alpha(H_\alpha)$ for finite $\alpha > 0$ and R_2 are strictly concave.*

In particular, for entropy, $H(X|Y) = H(X)$ if and only if X and Y are independent. This is well known since the mutual information $I(X;Y) = H(X) - H(X|Y)$ vanishes only in the case of independence [2] (p. 28). More generally, for α-entropy, $H_\alpha(X|Y) = H_\alpha(X)$ if and only if X and Y are independent.

Guessing entropy G, or, more generally, ρ-guessing entropy G_ρ, is not *strictly concave in p. For example, $G_\rho(1-\varepsilon, \varepsilon, 0, 0, \ldots) = 1 - \varepsilon + 2^\rho \varepsilon$ is linear in $\varepsilon < \frac{1}{2}$.*

Corollary 3 (Additional Knowledge Reduces Randomness)**.** *Inequality (32) is equivalent to the following:*

$$\mathfrak{R}(X|Y,Z) \leq \mathfrak{R}(X|Y) \tag{38}$$

for any Y, Z.

Proof. Inequality (32) applied to $X|y$ and Z for fixed y gives $\mathfrak{R}(X|y,Z) = \mathbb{E}_{z|y}\mathfrak{R}(p_{X|y,z}) \leq \mathfrak{R}(p_{X|y}) = \mathfrak{R}(X|y)$. Taking the expectation over Y of both sides yields the announced inequality. Conversely, letting $Y = 0$, one obtains $\mathfrak{R}(X|Z) \leq \mathfrak{R}(X)$, which is (32). □

Corollary 4 (Data Processing Inequality: Processing Knowledge Increases Randomness)**.** *For any Markov chain $X - Y - Z$ (i.e., such that $p_{X|Y,Z} = p_{X|Y}$), one has the following:*

$$\mathfrak{R}(X|Y) \leq \mathfrak{R}(X|Z). \tag{39}$$

This property is equivalent to (32).

Proof. Since $p_{X|Y} = p_{X|Y,z}$ for (almost) any z, one has $\mathfrak{R}(X|Y) = \mathfrak{R}(X|Y,z) = \mathfrak{R}(X|Y,Z)$, which, from Corollary 3, is $\leq \mathfrak{R}(X|Z)$. Conversely, letting $Z = 0$, one recovers (32). □

Example 7 (Data Processing Inequalities)**.** *For entropy H, the property $H(X|Y) \leq H(X|Z)$ amounts to $I(X;Z) \leq I(X;Y)$, i.e., (post-)processing in the Markov chain $X - Y - Z$ can never increase information § 2.8 in [2]. The data processing inequality for $\mathbb{P}_e$ and G was already shown in [27].*

2.3. Equalization (Minorization) via Robin Hood Operations

We now turn to another type of "mixing" probability distributions which are sometimes known as Robin Hood operations. To quote Arnold [28]:

> "When Robin and his merry hoods performed an operation in the woods they took from the rich and gave to the poor. The Robin Hood principle asserts that this decreases inequality (subject only to the obvious constraint that you don't take too much from the rich and turn them into poor.)"

Definition 4 (Robin Hood operations [28])**.** *An* elementary *"Robin Hood" operation $p \mapsto q$ in $\mathcal{P}$ modifies only two probabilities $(p_i, p_j) \mapsto (q_i, q_j)$ $(i \neq j)$ in such a way that $|p_i - p_j| \geq |q_i - q_j|$. A (general) "Robin Hood operation" results from a finite sequence of elementary Robin Hood operations.*

Notice that in an elementary Robin Hood operation, the sum $p_i + p_j = q_i + q_j$ should remain the same, since p and q are probability distributions. The fact that $|p_i - p_j|$ decreases "increases equality", i.e., makes the probabilities more equal. This can be written as follows:

$$\begin{cases} q_i = p_i - \delta \\ q_j = p_j + \delta \end{cases} \tag{40}$$

provided that $|\delta| \leq |p_i - p_j|$ ("you don't take too much from the rich and turn them into poor"). Setting $\lambda = 1 - \frac{\delta}{p_i - p_j} \in [0,1]$, (40) can be easily rewritten in the form:

$$\begin{cases} q_i = \lambda p_i + \bar{\lambda} p_j \\ q_j = \bar{\lambda} p_i + \lambda p_j \end{cases} \tag{41}$$

where $\lambda \geq 0$, $\bar{\lambda} \geq 0$ and $\lambda + \bar{\lambda} = 1$.

Remark 6 (Increasing Probability Product)**.** *In any elementary Robin Hood operation $(p_i, p_j) \mapsto (\lambda p_i + \bar{\lambda} p_j, \bar{\lambda} p_i + \lambda p_j)$, the product:*

$$q_i q_j = (\lambda p_i + \bar{\lambda} p_j)(\bar{\lambda} p_i + \lambda p_j) = p_i p_j + \lambda\bar{\lambda}(p_i - p_j)^2 \geq p_i p_j \tag{42}$$

always increases, with equality if and only if either $\lambda = 0$ or 1, or else $p_i = p_j$. This equality condition boils down to $|p_i - p_j| = |q_i - q_j|$, that is, the unordered set $\{p_i, p_j\} = \{q_i, q_j\}$ is unchanged.

Therefore, in any general Robin Hood operation, the product of all modified probabilities always increases, unless the probability distribution is unchanged (up to the order of the probabilities).

Remark 7 (Inverse Robin Hood Operation). *One can also define a "Sheriff of Nottingham" operation as an inverse Robin Hood operation, resulting from a finite sequence of elementary Sheriff of Nottingham operations of the form $(p_i, p_j) \mapsto (q_i, q_j)$, where $|p_i - p_j| \leq |q_i - q_j|$. Increasing the quantity $|p_i - p_j|$ "increases inequality", i.e., makes the probabilities more unequal.*

Definition 5 (Equalization Relation). *We write $X \preceq Y$ ("X is equalized by Y") if p_Y can be obtained from p_X by a Robin Hood operation. Such operation "equalizes" p_X in the sense that p_Y is "more equal" or "more uniform" than p_X. In terms of distributions, we also write $p_X \preceq p_Y$. Equivalently, p_X can be obtained from p_Y by a Sheriff of Nottingham operation (p_X is more unequal than p_Y). We may also write $Y \succeq X$ or $p_Y \succeq p_X$.*

Remark 8 (Generalization). *The above definitions hold verbatim for any vector or finitely many nonnegative numbers p_k with a fixed sum $s = \sum_k p_k$ (not necessarily equal to one). In the following, we sometimes use the concept of "equalization" in this slightly more general context.*

Remark 9 (Minorization). *$X \preceq Y$ amounts to saying that p_X "majorizes" p_Y in majorization theory [28,29]. So, in fact, the equalization relation $\preceq$ is a "minorization"—the opposite of a majorization. Unfortunately, it is common in majorization theory to write "$Y \preceq X$" when X "majorizes" Y, instead of $X \preceq Y$ when Y is "more equal" than X. Arguably, the notation adopted in this paper is more convenient, since it follows the usual relation order between randomness measures such as entropy.*

Also notice that the present approach avoids the use of Lorenz order [28,29] and focuses on the more intuitive Robin Hood operations.

Remark 10 (Partial Order). *It is easily seen that $\preceq$ is a partial order on the set of (finitely valued) discrete random variables (considering two variables "equal" if they are equivalent in the sense of Definition 1). Indeed, reflexivity and transitivity are immediate from the definition, and antisymmetry is, e.g., an easy consequence of Remark 6: if $X \preceq Y$ and $Y \preceq X$, then the product of all modified probabilities of X cannot increase by the two combined Robin Hood operations. Therefore, p_Y should be the same as p_X up to order; hence, $X \equiv Y$.*

The following fundamental lemmas establish expressions for maximally equal and unequal distributions.

Lemma 1 (Maximally Equal = Uniform). *For any vector $p = (p_1, p_2, \ldots, p_M)$ of nonnegative numbers with sum $s = \sum_k p_k$:*

$$p \preceq (\frac{s}{M}, \frac{s}{M}, \ldots, \frac{s}{M}). \tag{43}$$

In particular, any probability distribution p is equalized by the uniform distribution u:

$$p \preceq u \tag{44}$$

Proof. Suppose at least one component of p is $\neq \frac{s}{M}$. Since the p_ks sum to s, there should be at least one $p_i > \frac{s}{M}$ and one $p_j < \frac{s}{M}$. By a suitable Robin Hood operation on (p_i, p_j), at least one of these two probabilities can be made $= \frac{s}{M}$, reducing the total number of components $\neq \frac{s}{M}$. Continuing in this manner, we arrive at all probabilities equal to $\frac{s}{M}$ after, at most, $M - 1$ Robin Hood operations. □

Lemma 2 (Maximally Unequal)**.** *For any vector $p = (p_1, p_2, \ldots, p_M)$ of nonnegative numbers with sum $s = \sum_k p_k$ and constrained maximum $\max_k p_k \leq P$:*

$$p \succeq (P, \ldots, P, r, 0, \ldots, 0) \tag{45}$$

with remainder component $r = s - \lfloor \frac{s}{P} \rfloor P$. Without the maximum constraint ($P = s$), one simply has the following:

$$p \succeq (s, 0, \ldots, 0). \tag{46}$$

In particular, for any probability distribution p:

$$p \succeq \delta \tag{47}$$

where δ is the (Dirac) probability distribution of any deterministic variable. (This can be written in terms of random variables as $X \succeq 0$, since, by Defintion 1, any deterministic random variable is equivalent to the constant 0.)

Proof. Suppose at least two components lie between 0 and P: $0 < p_i, p_j < P$. By a suitable Sheriff of Nottingham operation on (p_i, p_j), at least one of these two probabilities can be made either $= 0$ or $= P$, reducing the number of components lying inside $(0, P)$. Continuing in this manner, we arrive at, at most, one component $r \in (0, P)$. Finally, the sum constraint implies $s = qP + r$ where $0 < r < P$, whence $q = \lfloor \frac{s}{P} \rfloor$. □

Theorem 3 (Schur Concavity [28,29])**.**

$$X \preceq Y \implies \mathfrak{R}(X) \leq \mathfrak{R}(Y) \tag{48}$$

Proof. It suffices to prove the inequality for an elementary Robin Hood operation $(p_i, p_j) \mapsto (\lambda p_i + \bar{\lambda} p_j, \bar{\lambda} p_i + \lambda p_j)$. Dropping the dependence on the other (fixed) probabilities, one has, by symmetry, (Theorem 1) and concavity (Theorem 2):

$$\mathfrak{R}(p_i, p_j) = \lambda \mathfrak{R}(p_i, p_j) + \bar{\lambda} \mathfrak{R}(p_j, p_i) \leq \mathfrak{R}(\lambda p_i + \bar{\lambda} p_j, \lambda p_j + \bar{\lambda} p_i). \tag{49}$$

□

Inequality (48), expressed in terms of distributions:

$$p_X \preceq p_Y \implies \mathfrak{R}(p_X) \leq \mathfrak{R}(p_Y) \tag{50}$$

is known as "Schur concavity" [28,29].

Remark 11. *Theorem 3 can also be given a physical interpretation similar to Corollary 1. In fact, from (41), any Robin Hood operation can be seen as mixing two permuted probability distributions, which have equal randomness. Such mixing can only increase randomness.*

Example 8 (Entropy is Schur-Concave)**.** *That the Shannon entropy is Schur-concave is well known § 13 E in [29]. Similar to concavity (Example 5), this also has a similar physical interpretation: a liquid mixed with another results in a "more disordered", "more chaotic" system, which results in a "more equal" distribution and a higher entropy § 1 A9 in [29].*

Remark 12 (φ-Schur Concavity)**.** *Schur concavity is not equivalent to concavity (even when assuming symmetry). In fact, with the notations of Remark 4, it is obvious that Schur concavity of $\mathfrak{R}$ is equivalent to Schur concavity of $\varphi(\mathfrak{R})$, where $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ is any increasing function. In other words, while "φ-concavity" (in the sense of Remark 4) is not the same as concavity, there is no need to introduce "φ-Schur concavity", since it is always equivalent to Schur concavity.*

Remark 13 (Strict Schur Concavity). *A randomness measure $\mathfrak{R}$ is "strictly Schur concave" if the inequality $\mathfrak{R}(X) \leq \mathfrak{R}(Y)$ for $X \preceq Y$ holds with equality $\mathfrak{R}(X) = \mathfrak{R}(Y)$ if and only if $X \equiv Y$.*

If $\mathfrak{R}(p)$ is strictly concave (see Remark 5), then equality holds in (49) if and only if either $\lambda = 0$ or 1, or else $p_i = p_j$. Either of these conditions means that $\{p_i, p_j\}$ is unchanged. Therefore, in this case, $\mathfrak{R}$ is also strictly Schur concave.

Remark 6 states that the product of nonzero probabilities is strictly Schur-concave.

Example 9 (Strictly Schur Concave Measures). *Randomness measures presented in Section 1 are (Schur) concave, but not all of them are* strictly *Schur concave:*

- *Not only the Shannon entropy H is Schur concave (Example 8), but, as seen in Example 6, H, as well as all α-entropies $\varphi_\alpha(H_\alpha)$ for finite $\alpha > 0$ and R_2, are strictly concave and, hence, strictly Schur concave;*
- *As seen also in Example 6, guessing entropy G, or, more generally, ρ-guessing entropy G_ρ, is* not *strictly concave in p. However, G and G_ρ are strictly Schur concave by the following argument.*
  It suffices to show that some elementary Robin Hood operation (40) $(p_i, p_j) \mapsto (p_i - \delta, p_j + \delta)$ (with $\delta \neq 0$) strictly increases G_ρ. One may always choose δ as small as one pleases, since any elementary Robin Hood operation on (p_i, p_j) can be seen as resulting from other ones on (p_i, p_j) with smaller δ. One chooses δ small enough such that the elementary Robin Hood operation does not change the order of the probabilities in p. With the notations of Section 1.2, assuming, for example, that $p_i = p_{(i)} > p_j = p_{(j)}$, where $i < j$, then $\delta > 0$ and $i^\rho p_{(i)} + j^\rho p_{(j)} < i^\rho (p_{(i)} - \delta) + j^\rho (p_{(j)} + \delta)$, since $j^\rho > i^\rho$. This shows that G_ρ strictly increases;
- *Error probability $\mathbb{P}_e$, or, more generally, $\mathbb{P}_e^m$, is neither strictly concave nor strictly Schur concave in general. In fact, if $M \geq m + 2$, any elementary Robin Hood operation on $p_i, p_j < p_{(m)}$ leaves $\mathbb{P}_e^m$ unchanged;*
- *Statistical randomness R is neither strictly concave nor strictly Schur concave if $M > 2$. For example, it is easily checked from the definition (18) that the elementary Robin Hood operation $(\frac{1}{M}, \frac{2}{M}) \mapsto (\frac{4/3}{M}, \frac{5/3}{M})$ leaves R unchanged.*

2.4. Resulting Properties in Terms of Random Variables

Corollary 5 (Minimal and Maximal Randomness)**.**

$$\mathfrak{R}(\delta) \leq \mathfrak{R}(X) \leq \mathfrak{R}(u) \tag{51}$$

In other words, minimal randomness is achieved for $X = 0$ (for any deterministic variable 0) and maximal randomness is achieved for uniformly distributed X.

Proof. From Lemmas 1 and 2, one obtains $\delta \preceq p_X \preceq u$. The result follows by Theorem 3. □

Remark 14 (Zero Randomness)**.** *Without loss of generality, we may always impose that $\mathfrak{R}(0) = 0$ by considering $\mathfrak{R}(X) - \mathfrak{R}(0)$ in place of $\mathfrak{R}(X)$. Then, zero randomness is achieved when $X \equiv 0$. It is easily checked from the expressions given in Section 1 that this convention holds for H, H_α, $\log G$, $\log G_\rho$, $\mathbb{P}_e$, $\mathbb{P}_e^m$, R_2 and R.*

To simplify notations in the remainder of this paper, we assume that the zero randomness convention $\mathfrak{R}(0) = 0$ always holds.

Example 10 (Distribution Achieving Zero Randomness)**.** *By Remark 13, if $\mathfrak{R}$ is strictly Schur concave, zero randomness is achieved only when $X \equiv 0$:*

$$\mathfrak{R}(X) = 0 \iff X \equiv 0. \tag{52}$$

- *As seen in Example 9, this is the case for H, H_α, $\log G$, $\log G_\rho$ and R_2. In particular, we recover the well known property that zero entropy is achieved only when X is deterministic;*

- *Although the error probability is not strictly Schur concave, one can check directly that* $\mathbb{P}_e(p) = 0$ *if and only if* $p_{(1)} = 1$*, which corresponds to the* δ *distribution;*
- *Similarly, from the discussion in Section 1.7,* $R(p) = 0$ *correspond to the maximum value of* $\Delta(p, u) = 1 - \frac{1}{M}$ *attained for* $K = |T_+| = 1$ *and* $\mathbb{P}(T_+) = 1$*, which, again, corresponds to a* δ *distribution.*

  To summarize, all quantities H*,* H_α*,* $\log G$*,* $\log G_\rho$*,* $\mathbb{P}_e$*,* R_2 *and* R *satisfy (52).*

Remark 15 (Maximal Randomness Increases with M)**.** *For an* M*-ary random variable, maximal randomness* $\mathfrak{R}_M = \mathfrak{R}(u_M)$ *is attained for a uniform distribution* $u_M = (\frac{1}{M}, \frac{1}{M}, \ldots, \frac{1}{M})$*. Since, by Lemma 1,* $u_M \preceq u_{M+1}$*, one has* $\mathfrak{R}_M \leq \mathfrak{R}_{M+1}$*: maximal randomness* $\mathfrak{R}_M$ *increases with* M*.*

Example 11 (Distribution Achieving Maximum Randomness)**.** *The following maximum values for* M*-ary random variables are easily checked from the expressions given in Section 1:*

- $\max H = H(u) = \log M$*, and, more generally,* $\max H_\alpha = H_\alpha(u) = \log M$*. Since* H *and* H_α *are strictly Schur-concave, the maximum* $H_\alpha(X) = \log M$ *is attained if and only if* X *is uniformly distributed. This observation is also an easy consequence of (10) or (11);*
- $\max G = G(u) = \frac{M+1}{2}$*,* $\max G_2 = G_2(u) = \frac{(M+½)(M+1)}{3}$*,* $\max G_3 = G_3(u) = \frac{M(M+1)^2}{4}$*, etc. Again, since* G *and* G_ρ *are strictly Schur-concave, their maximum is achieved if and only if* X *is uniformly distributed;*
- $\max \mathbb{P}_e = \mathbb{P}_e(u) = 1 - \frac{1}{M}$*, and, more generally,* $\max \mathbb{P}_e^m = \mathbb{P}_e^m(u) = 1 - \frac{m}{M}$*. The maximum of* $\mathbb{P}_e(X)$ *is achieved if and only if the maximum probability* $p_{(1)}$ *equals* $\frac{1}{M}$*, which implies that* X *is uniformly distributed;*
- $\max R_2 = \max R = 1 - \frac{1}{M}$ *(see (12) and (18)) is achieved if and only if* $p = u$*.*

To summarize, for all quantities H*,* H_α*,* $\log G$*,* $\log G_\rho$*,* $\mathbb{P}_e$*,* R_2 *and* R*, the unique maximizing distribution is the uniform distribution. Notice that, as expected, each of these maximum values increases with* M*.*

Corollary 6 (Deterministic Data Processing Inequality: Processing Reduces Randomness)**.** *For any deterministic function* f*:*

$$\mathfrak{R}(f(X)) \leq \mathfrak{R}(X). \tag{53}$$

Proof. Consider preimages by f of values $y = f(x)$. The application of f can be seen as resulting from a sequence of elementary operations, each of which puts together two distinct values of x (say, x_i and x_j) in the same preimage of some y. In terms of probability distributions, this amounts to a Sheriff of Nottingham operation $(p_i, p_j) \mapsto (p_i + p_j, 0)$. Overall, one has $f(X) \preceq X$. The result then follows by Schur concavity (Theorem 3). □

Example 12. *The fact that* $H(f(X)) \leq H(X)$ *is well known (see Ex. 2.4 in [2]). This can also be seen from the data processing inequality of Corollary 4 by noting that, since* $X - f(X) - f(X)$ *is trivially a Markov chain,* $H(f(X)) = I(f(X); f(X)) \leq I(X; f(X)) \leq H(X)$*.*

Remark 16 (Lattices of Information and Majorization)**.** *Shannon [26] defined the order relation* $X \leq Y$ *if* $X = g(Y)$ *a.s. and showed that it satisfies the properties of a lattice, called the "information lattice" (see [30] for detailed proofs). With this notation, (53) writes as shown:*

$$X \leq Y \implies \mathfrak{R}(X) \leq \mathfrak{R}(Y). \tag{54}$$

Majorization (or the order relation $X \preceq Y$*) also satisfies the properties of a lattice—the "majorization lattice", as studied in [31]. From the proof of Corollary 6, one actually obtains the following:*

$$X \leq Y \implies X \preceq Y \implies \mathfrak{R}(X) \leq \mathfrak{R}(Y). \tag{55}$$

Therefore, the majorization lattice is denser than the information lattice.

Corollary 7 (Addition Increases Randomness)**.**

$$\mathfrak{R}(X) \preceq \mathfrak{R}(X,Y) \tag{56}$$

This property is equivalent to (53)*.*

Proof. Apply Corollary 6 to the projection $f(x,y) = x$. Conversely, (53) follows from (56), by taking $Y = f(X)$ and noting that $(X, f(X)) \equiv X$. □

Corollary 8 (Total Dependence)**.** *Assuming the zero randomness convention (Remark 14), if* (52) *holds, then the following holds:*

$$\mathfrak{R}(X|Y) = 0 \iff X = f(Y) \text{ a.s.}, \tag{57}$$

that is, $\mathfrak{R}(X|Y) = 0 \iff X \leq Y$ *in the sense of Shannon (Remark 16).*

Proof. Since $\mathfrak{R}(X|y) \geq 0$ for any y, $\mathfrak{R}(X|Y) = \mathbb{E}_y \mathfrak{R}(X|y) = 0$ if and only if $\mathfrak{R}(X|y) = 0$ for (almost) all y. By (52), this implies that X is deterministic given $Y = y$, i.e., X is a deterministic function of Y. □

Example 13. *From Example 10,* (57) *is true for* H, H_α, $\log G$, $\log G_\rho$, $\mathbb{P}_e$, R_2 *and* R.

- *The equivalence* $H(X|Y) = 0 \iff X = f(Y)$ *a.s. is well known ([2], Ex. 2.5). Knowledge of* Y *removes equivocation only when* X *is fully determined by* Y*;*
- $\log G(X|Y) = 0 \iff G(X|Y) = 1 \iff X = f(Y)$ *a.s. is intuitively clear: knowing* Y *allows one to fully determine* X *in only one guess;*
- $\mathbb{P}_e(X|Y) = 0 \iff X = f(Y)$ *a.s.: knowing* Y *allows one to estimate* X *without error only when* X *is fully determined by* Y*.*

3. Fano and Reverse-Fano Inequalities

Definition 6 (Fano-type inequalities)**.** *A "Fano inequality" (resp. "reverse Fano inequality") for* $\mathfrak{R}(X)$ *gives an upper (resp. lower) bound of* $\mathfrak{R}(X)$ *as a function of the probability of error* $\mathbb{P}_e(X)$*. Fano and reverse-Fano inequalities are similarly defined for conditional randomness* $\mathfrak{R}(X|Y)$*, lower or upper bounded as a function of* $\mathbb{P}_e(X|Y)$*.*

In this section, we establish optimal Fano and reverse-Fano inequalities, where upper and lower bounds are tight. In other words, we determine the maximum and minimum of $\mathfrak{R}$ for fixed $\mathbb{P}_e$. The exact locus of the region $p \in \mathcal{P}_M \mapsto (\mathbb{P}_e(p), \mathfrak{R}(p)) = (\mathbb{P}_e(X), \mathfrak{R}(X))$, as well as the exact locus of all attainable values of $(\mathbb{P}_e(X|Y), \mathfrak{R}(X|Y))$, is determined analytically for fixed M, based on the following.

Lemma 3. *Let* $\mathbb{P}_e = \mathbb{P}_e(p)$ *and* $\mathbb{P}_s = 1 - \mathbb{P}_e$*. For any* M*-ary probability distribution* $p \in \mathcal{P}_M$*:*

$$(\underbrace{\mathbb{P}_s, \ldots, \mathbb{P}_s}_{\lfloor \frac{1}{\mathbb{P}_s} \rfloor \text{ times}}, 1 - \lfloor \tfrac{1}{\mathbb{P}_s} \rfloor \mathbb{P}_s, 0, \ldots, 0) \preceq p \preceq (\mathbb{P}_s, \tfrac{\mathbb{P}_e}{M-1}, \ldots, \tfrac{\mathbb{P}_e}{M-1}). \tag{58}$$

Proof. On the left side, apply Lemma 2 with $P = \max p = p_{(1)} = \mathbb{P}_s$ and $s = 1$. On the right side, with $p_{(1)} = \mathbb{P}_s$ being fixed, apply Lemma 1 to the $M-1$ remaining probabilities $(p_{(2)}, \ldots, p_{(M)})$, which sum to $s = 1 - \mathbb{P}_s = \mathbb{P}_e$. □

Theorem 4 (Optimal Fano and Reverse-Fano Inequalities for $\mathfrak{R}(X)$)**.** *The optimal Fano and reverse-Fano inequalities for the randomness measure* $\mathfrak{R}(X)$ *of any* M*-ary random variable* X *in terms of* $\mathbb{P}_e = \mathbb{P}_e(X)$ *are given analytically by the following:*

$$\mathfrak{R}(1-\mathbb{P}_e, \ldots, 1-\mathbb{P}_e, 1 - \lfloor \tfrac{1}{1-\mathbb{P}_e} \rfloor (1-\mathbb{P}_e), 0, \ldots, 0) \leq \mathfrak{R}(X) \leq \mathfrak{R}(1-\mathbb{P}_e, \tfrac{\mathbb{P}_e}{M-1}, \ldots, \tfrac{\mathbb{P}_e}{M-1}). \tag{59}$$

Proof. The proof is immediate from Lemma 3 and Theorem 3. The Fano and reverse-Fano bounds are achieved by the distributions on the left and right sides of (58), respectively. □

A similar proof holding for any Schur concave $\mathfrak{R}(X)$ was already given by Vajda and Vašek [17].

Assuming the zero randomness convention for simplicity (Remark 14), Fano and reverse-Fano bounds can be qualitatively described as follows. They are illustrated in Figure 2.

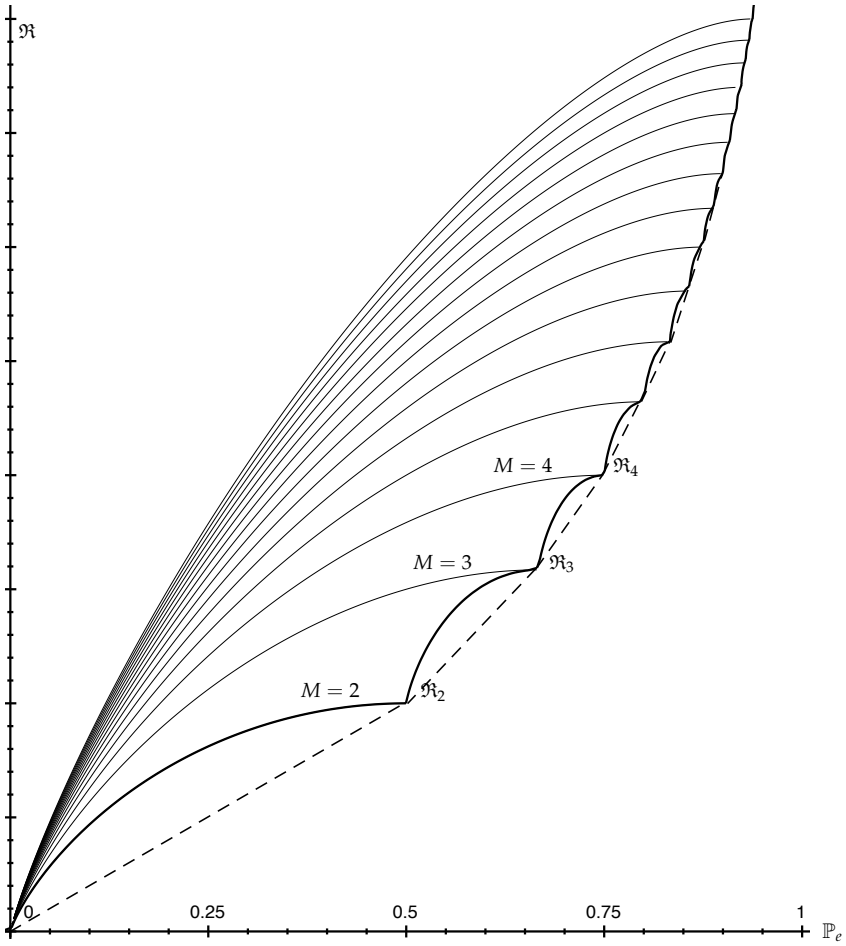

Figure 2. Typical upper Fano bounds (thin) for $M = 2$ to 16 and lower reverse-Fano bound for $\mathfrak{R}(X)$ (solid) and for $\mathfrak{R}(X|Y)$ (dashed).

Proposition 1 (Shape of Fano Bounds)**.** *The (upper) Fano bound:*

$$\mathbb{P}_e \in [0, 1-\tfrac{1}{M}] \mapsto \mathfrak{R}(1-\mathbb{P}_e, \tfrac{\mathbb{P}_e}{M-1}, \ldots, \tfrac{\mathbb{P}_e}{M-1}) \in [0, \mathfrak{R}_M] \tag{60}$$

where $\mathfrak{R}_M$ denotes maximal randomness (Remark 15) is continuous in $\mathbb{P}_e > 0$, concave in $\mathbb{P}_e$ and increases from 0 (for $\mathbb{P}_e = 0$) to $\mathfrak{R}_M$ (for $\mathbb{P}_e = 1-\frac{1}{M}$). For any fixed $\mathbb{P}_e$, it also increases with M.

Proof. Since $\mathfrak{R}(p) \geq 0$ is concave over $\mathcal{P}_M$ (Theorem 2), it is continuous on the interior of $\mathcal{P}_M$. Since $\mathbb{P}_e \mapsto (1-\mathbb{P}_e, \frac{\mathbb{P}_e}{M-1}, \ldots, \frac{\mathbb{P}_e}{M-1})$ is linear, the Fano bound results from the composition of a linear and a concave function. It is, therefore, concave, and continuous at every $\mathbb{P}_e > 0$. It is clear from Lemma 3, or using a suitable Robin Hood operation, that the maximizing distribution becomes more equal as $\mathbb{P}_e$ increases. Therefore, the Fano bound increases with $\mathbb{P}_e$. The maximum is attained for $P_e = 1 - \frac{1}{M}$, which corresponds to the uniform distribution achieving maximum randomness $\mathfrak{R}_M$. For fixed $\mathbb{P}_e$, it is also clear,

using a suitable Robin Hood operation, that the maximizing distribution becomes more equal if M is increased by one. Therefore, the Fano bound also increases with M. □

Proposition 2 (Shape of reverse-Fano Bounds)**.** *The (lower) reverse-Fano bound:*

$$\mathbb{P}_e \in [0, 1-\tfrac{1}{M}] \mapsto \mathfrak{R}(1-\mathbb{P}_e, \ldots, 1-\mathbb{P}_e, 1-\lfloor \tfrac{1}{1-\mathbb{P}_e} \rfloor (1-\mathbb{P}_e), 0, \ldots, 0) \in [0, \mathfrak{R}_M] \tag{61}$$

is continuous in $\mathbb{P}_e > 0$, increases from 0 (for $\mathbb{P}_e = 0$) to $\mathfrak{R}_M$ (for $\mathbb{P}_e = 1-\frac{1}{M}$) and is composed of continuous concave increasing curves connecting successive points ($\mathbb{P}_e = 1 - \frac{1}{k}$, $\mathfrak{R} = \mathfrak{R}_k$) for $k = 1, 2, \ldots, M$.

Proof. For any $k \in \{1, 2, \ldots, M\}$, the reverse-Fano bound at $\mathbb{P}_e = 1 - \frac{1}{k}$ is $\mathfrak{R}(\frac{1}{k}, \ldots, \frac{1}{k}) = \mathfrak{R}_k$. It suffices to prove that the reverse-Fano bound is continuous, concave and increasing for $1 - \frac{1}{k} \leq \mathbb{P}_e \leq 1 - \frac{1}{k+1}$. When $\lfloor \frac{1}{1-\mathbb{P}_e} \rfloor = k$, that is, $1 - \frac{1}{k} \leq \mathbb{P}_e < 1 - \frac{1}{k+1}$, the reverse-Fano bound is $\mathfrak{R}(1-\mathbb{P}_e, \ldots, 1-\mathbb{P}_e, 1-k(1-\mathbb{P}_e))$. This results from the composition of a linear and a concave function $\mathfrak{R}(p)$, which is continuous in the interior of $\mathcal{P}_k$. Therefore, it is concave in $\mathbb{P}_e$, and continuous on the whole closed interval $[1 - \frac{1}{k}, 1 - \frac{1}{k+1}]$. Finally, it is clear from Lemma 2 or using a suitable Robin Hood operation that $(1-\mathbb{P}_e, \ldots, 1-\mathbb{P}_e, 1-k(1-\mathbb{P}_e))$ becomes more equal as $\mathbb{P}_e$ increases. Therefore, each curve increases from $\mathfrak{R}_k$ to $\mathfrak{R}_{k+1}$. □

Remark 17 (Independence of the reverse-Fano Bound from the Alphabet Size)**.** *Contrary to the (upper) Fano bound, the (lower) reverse-Fano bound is achieved by a probability distribution that does not depend on M. As a result, when the definition of $\mathfrak{R}$ does not itself explicitly depend on M (as is the case for H, H_α, G, G_ρ, $\mathbb{P}_e$, $\mathbb{P}_e^m$, R_2), the reverse-Fano bound is the same for all M, except that it is truncated up to $\mathbb{P}_e = 1 - \frac{1}{M}$, at which point it meets the (upper) Fano bound (see Figure 2).*

Theorem 5 (Optimal Fano and Reverse-Fano Inequalities for $\mathfrak{R}(X|Y)$)**.** *The optimal Fano and reverse-Fano inequalities for the randomness measure $\mathfrak{R}(X|Y)$ of any M-ary random variable X in terms of $\mathbb{P}_e = \mathbb{P}_e(X|Y)$ are given analytically by the following:*

$$(\lceil \tfrac{1}{\mathbb{P}_s} \rceil \mathbb{P}_s - 1) \lfloor \tfrac{1}{\mathbb{P}_s} \rfloor \mathfrak{R}_{\lfloor \frac{1}{\mathbb{P}_s} \rfloor} + (1 - \lfloor \tfrac{1}{\mathbb{P}_s} \rfloor \mathbb{P}_s) \lceil \tfrac{1}{\mathbb{P}_s} \rceil \mathfrak{R}_{\lceil \frac{1}{\mathbb{P}_s} \rceil} \leq \mathfrak{R}(X|Y) \leq \mathfrak{R}(1-\mathbb{P}_e, \tfrac{\mathbb{P}_e}{M-1}, \ldots, \tfrac{\mathbb{P}_e}{M-1}). \tag{62}$$

where we have noted $\lceil x \rceil = \lfloor x \rfloor + 1$ ($\lceil x \rceil$ is the usual ceil function $\lceil x \rceil$, unless x is an integer), $\mathbb{P}_s = 1 - \mathbb{P}_e$ and $\mathfrak{R}_k = \mathfrak{R}(\frac{1}{k}, \ldots, \frac{1}{k})$.

Proof. The Fano region for $X|Y = y$, i.e., the locus of the points $(\mathbb{P}_e(p_{X|y}), \mathfrak{R}(p_{X|y}))$ for each $Y = y$, is given by the inequalities (59). From the definition of conditional randomness, the exact locus of points $(\mathbb{P}_e(X|Y), \mathfrak{R}(X|Y)) = \mathbb{E}_y(\mathbb{P}_e(p_{X|y}), \mathfrak{R}(p_{X|y}))$ is composed of all convex combinations of points in the Fano region, that is, its convex envelope. The extreme points $(\mathbb{P}_e = 0, \mathfrak{R} = \mathfrak{R}_1 = 0)$ and $(\mathbb{P}_e = 1 - \frac{1}{M}, \mathfrak{R} = \mathfrak{R}_M)$ are unchanged. The upper Fano bound joining these two extreme points is concave by Proposition 1 and, therefore, already belongs to the convex envelope. It follows that the upper Fano bound in (59) remains the same, as given in (62). However, the lower reverse-Fano bound for $\mathfrak{R}(X|Y)$ is the convex hull of the lower bound in (59). By Proposition 2, it is easily seen to be the piecewise linear curve joining all singular points $(\mathbb{P}_e = 1 - \frac{1}{k}, \mathfrak{R} = \mathfrak{R}_k)$ for $k = 1, 2, \ldots, M$ (see Figure 2). A closed-form expression is obtained by noting that, when $\lfloor \frac{1}{1-\mathbb{P}_e} \rfloor = k$, that is, $1 - \frac{1}{k} \leq \mathbb{P}_e < 1 - \frac{1}{k+1}$, the equation of the straight line joining $(1 - \frac{1}{k}, \mathfrak{R}_k)$ and $(1 - \frac{1}{k+1}, \mathfrak{R}_{k+1})$ is $((k+1)\mathbb{P}_s - 1)k\mathfrak{R}_k + (1 - k\mathbb{P}_s)(k+1)\mathfrak{R}_{k+1}$. Plugging $k = \lfloor \frac{1}{\mathbb{P}_s} \rfloor$ and $k + 1 = \lceil \frac{1}{\mathbb{P}_s} \rceil$ gives the lower reverse-Fano bound in (62). □

Remark 18 (Shape of Fano and reverse-Fano bounds for Conditional Randomness)**.** *By Theorem 5, the Fano inequality for the conditional version $\mathfrak{R}(X|Y)$ takes the same form as for $\mathfrak{R}(X)$. In particular, it is increasing and concave in $\mathbb{P}_e(X|Y)$. Compared to that for $\mathfrak{R}(X)$, the*

reverse-Fano bound for $\mathfrak{R}(X|Y)$, however, is a piecewise linear convex hull. Clearly, it is still continuous and increasing in $\mathbb{P}_e(X|Y)$, as illustrated in Figure 2. If the corresponding sequence of slopes $k(k+1)(\mathfrak{R}_{k+1}-\mathfrak{R}_k)$ is increasing in k, then the reverse-Fano bound for $\mathfrak{R}(X|Y)$ is also convex in $\mathbb{P}_e(X|Y)$.

Remark 19 (φ-Fano Bounds)**.** *If $\varphi(\mathfrak{R})$ is used instead of $\mathfrak{R}$, where φ is an increasing function (in particular, to define conditional randomness as in Remark 4), then Theorem 4 and the (upper) Fano bound of Theorem 5 can be directly applied to $\mathfrak{R}$. When φ is nonlinear, this may result in (upper) Fano bounds that are no longer concave.*

However, to obtain the reverse-Fano inequalities for $\mathfrak{R}(X|Y)$, one has to apply Theorem 5 to $\varphi(\mathfrak{R}(X|Y))$ and then apply the inverse function φ^{-1} to the left side of (62). When φ is nonlinear, the resulting "reverse-Fano bound" for $\mathfrak{R}(X|Y)$ will not be piecewise linear anymore. This is the case, e.g., for conditional α-entropies (see Example 15 below).

Example 14 (Fano and reverse-Fano Inequalities for Entropy)**.** *For the Shannon entropy, the optimal Fano inequality (right sides of (59) and (62)) takes the form:*

$$H(X) \le h(\mathbb{P}_e(X)) + \mathbb{P}_e(X)\,\log(M-1) \tag{63}$$
$$H(X|Y) \le h(\mathbb{P}_e(X|Y)) + \mathbb{P}_e(X|Y)\,\log(M-1) \tag{64}$$

where $h(\mathbb{P}_e) = \mathbb{P}_e \log \frac{1}{\mathbb{P}_e} + (1-\mathbb{P}_e)\log\frac{1}{1-\mathbb{P}_e}$ is the binary entropy function. Inequality (64) is the original Fano inequality established in 1952 [22], which has become ubiquitous in information theory and in statistics to relate equivocation to probability of error. Inequality (63) trivially follows, in case of blind estimation ($Y \equiv 0$). That these inequalities are sharp is well known (see, e.g., [32]).

The optimal reverse-Fano inequality (left sides of (59) and (62) with $\mathfrak{R}_k = \log k$) takes the form:

$$H(X) \ge \phi(\mathbb{P}_s(X)) = \phi(1-\mathbb{P}_e(X)) \tag{65}$$
$$H(X|Y) \ge \bar{\phi}(\mathbb{P}_s(X|Y)) = \bar{\phi}(1-\mathbb{P}_e(X|Y)) \tag{66}$$

where

$$\phi(x) = h\big(\lfloor \tfrac{1}{x} \rfloor x\big) + \lfloor \tfrac{1}{x} \rfloor x \log \lfloor \tfrac{1}{x} \rfloor \tag{67}$$
$$\bar{\phi}(x) = \big(\lceil \tfrac{1}{x} \rceil x - 1\big)\lfloor \tfrac{1}{x} \rfloor \log \lfloor \tfrac{1}{x} \rfloor + \big(1 - \lfloor \tfrac{1}{x} \rfloor x\big)\lceil \tfrac{1}{x} \rceil \log \lceil \tfrac{1}{x} \rceil \tag{68}$$

These two lower bounds were first derived by Kovalevsky [33] in 1965. Optimality was already proven in [32].

Example 15 (Fano and reverse-Fano Inequalities for α-Entropy)**.** *By Remark 19, the optimal Fano inequality for $H_\alpha(X)$ is obtained as the right side of (59), which gives the following:*

$$H_\alpha(X) \le \tfrac{1}{1-\alpha}\log\big((M-1)^{1-\alpha}\mathbb{P}_e(X)^\alpha + \mathbb{P}_s(X)^\alpha\big). \tag{69}$$

This was proven by Toussaint [34] for $0 < \alpha < 1$ and, independently, by Ben-Bassat and Raviv [35] for $\alpha \neq 1$.

Additionally, by Remark 19, the optimal Fano inequality for $H_\alpha(X|Y)$ is obtained by averaging over Y the Fano upper bound of $\varphi_\alpha(H_\alpha(X|y))$, which is of the form $\phi(\mathbb{P}_e(X|y))$, where $\phi(x) = \operatorname{sgn}(1-\alpha)\big((M-1)^{1-\alpha}x^\alpha + (1-x)^\alpha\big)^{1/\alpha}$, which is concave Lemma 1 in [36]. Therefore, the optimal Fano inequality for $H_\alpha(X|Y)$ is likewise obtained as the right side of (62), which gives the following:

$$H_\alpha(X|Y) \le \tfrac{1}{1-\alpha}\log\big((M-1)^{1-\alpha}\mathbb{P}_e(X|Y)^\alpha + \mathbb{P}_s(X|Y)^\alpha\big). \tag{70}$$

The optimal reverse-Fano inequality for $H_\alpha(X)$ is obtained as the left side of (59). By Remark 19, $H_\alpha(X|Y)$ is obtained by applying $\varphi_\alpha^{-1}(x) = \frac{\alpha}{1-\alpha}\log(\mathrm{sgn}(1-\alpha)x)$ to the left side of (62) for $\varphi_\alpha(H_\alpha(X|Y))$, where φ_α is given by (30). This gives the following:

$$H_\alpha(X) \geq \phi_\alpha(\mathbb{P}_s(X)) = \phi_\alpha(1-\mathbb{P}_e(X)) \tag{71}$$

$$H_\alpha(X|Y) \geq \bar{\phi}_\alpha(\mathbb{P}_s(X|Y)) = \bar{\phi}_\alpha(1-\mathbb{P}_e(X|Y)) \tag{72}$$

where

$$\phi_\alpha(x) = \frac{1}{1-\alpha}\log(\lfloor \tfrac{1}{x}\rfloor x^\alpha + (1-\lfloor \tfrac{1}{x}\rfloor x)^\alpha) \tag{73}$$

$$\bar{\phi}_\alpha(x) = \frac{\alpha}{1-\alpha}\log\Big((\lceil \tfrac{1}{x}\rceil x - 1)\lfloor \tfrac{1}{x}\rfloor^{\frac{1}{\alpha}} + (1-\lfloor \tfrac{1}{x}\rfloor x)\lceil \tfrac{1}{x}\rceil^{\frac{1}{\alpha}}\Big) \tag{74}$$

Fano and reverse-Fano inequalities for $H_\alpha(X)$ and $H_\alpha(X|Y)$ were recently established by Sason and Verdú [36].

Example 16 (Fano and reverse-Fano Inequalities for non collision R_2)**.** *Theorem 4 readily gives the optimal Fano region for $R_2(X)$:*

$$1-\lfloor \tfrac{1}{\mathbb{P}_s}\rfloor \mathbb{P}_s^2 - (1-\lfloor \tfrac{1}{\mathbb{P}_s}\rfloor \mathbb{P}_s)^2 \leq R_2(X) \leq 1-\mathbb{P}_s^2(X) - \frac{\mathbb{P}_e^2(X)}{M-1}. \tag{75}$$

This can also be easily deduced from (69) and (71) for $\alpha = 2$ via (4). Fano and reverse-Fano inequalities for $R_2(X)$ were first stated without proof in [7].

The optimal Fano region for $R_2(X|Y)$, however, cannot be directly deduced from that of $H_2(X|Y)$, because a different kind of average over Y is involved. However, a direct application of Theorem 5 with $\mathfrak{R}_k = 1-\frac{1}{k}$ gives the optimal Fano region:

$$\mathbb{P}_e(X|Y) \leq R_2(X|Y) \leq 1-\mathbb{P}_s^2(X|Y) - \frac{\mathbb{P}_e^2(X|Y)}{M-1}. \tag{76}$$

Remarkably, the reverse-Fano inequality has a very simple form $R_2(X|Y) \geq \mathbb{P}_e(X|Y)$ (see Figure 3).

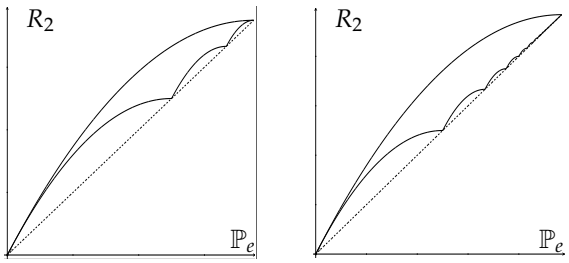

Figure 3. Optimal Fano regions for R_2 vs. $\mathbb{P}_e$. Solid: Fano region $R_2(X)$ vs. $\mathbb{P}_e(X)$. Dashed: Fano region $R_2(X|Y)$ vs. $\mathbb{P}_e(X|Y)$. **Left** $M=4$; **right** $M=32$.

Example 17 (Fano and reverse-Fano Inequalities for Guessing Entropy)**.** *For guessing entropy G, the Fano inequality is written as shown:*

$$G(X) \leq 1 + \frac{M}{2}\mathbb{P}_e(X) \tag{77}$$

$$G(X|Y) \leq 1 + \frac{M}{2}\mathbb{P}_e(X|Y) \tag{78}$$

One obtains similarly $G_2 \leq 1+\frac{M}{3}(M+\frac{5}{2})\mathbb{P}_e$, $G_3 \leq 1+\frac{M}{4}(M^2+3M+4)\mathbb{P}_e$, etc.

Due to the fact that $G_\rho(p)$ is linear in p, for fixed $\lfloor \frac{1}{1-\mathbb{P}_e}\rfloor = k$, the reverse-Fano bound for $G_\rho(X)$ is linear in $\mathbb{P}_e$. It follows that the bound is already piecewise linear, with a sequence of slopes

$s_k = k(k+1)(\mathfrak{R}_{k+1} - \mathfrak{R}_k) = k(1^\rho + \cdots + (k+1)^\rho) - (k+1)(1^\rho + \cdots + k^\rho)$, *which is easily seen to be increasing. Therefore, the (lower) reverse-Fano bound is piecewise linear and convex and coincides with its convex hull. In other words, the reverse-Fano inequality for* $G_\rho(X)$ *and* $G_\rho(X|Y)$ *takes the same form:*

$$G_\rho(X) \geq \phi_\rho(\mathbb{P}_s(X)) = \phi_\rho(1 - \mathbb{P}_e(X)) \tag{79}$$

$$G_\rho(X|Y) \geq \phi_\rho(\mathbb{P}_s(X|Y)) = \phi_\rho(1 - \mathbb{P}_e(X|Y)). \tag{80}$$

The following is easily determined from the left side of either (59) *or* (62)*:*

$$\phi_\rho(x) = x(1^\rho + \cdots + \lfloor \tfrac{1}{x} \rfloor^\rho) + (1 - \lfloor \tfrac{1}{x} \rfloor x)\lceil \tfrac{1}{x} \rceil^\rho. \tag{81}$$

For example, $\phi_1(x) = (\lfloor \frac{1}{x} \rfloor + 1)(1 - \lfloor \frac{1}{x} \rfloor \frac{x}{2})$*, such that the following occurs:*

$$G(X) \geq (\lfloor \tfrac{1}{\mathbb{P}_s(X)} \rfloor + 1)(1 - \lfloor \tfrac{1}{\mathbb{P}_s(X)} \rfloor \tfrac{\mathbb{P}_s(X)}{2}) \tag{82}$$

$$G(X|Y) \geq (\lfloor \tfrac{1}{\mathbb{P}_s(X|Y)} \rfloor + 1)(1 - \lfloor \tfrac{1}{\mathbb{P}_s(X|Y)} \rfloor \tfrac{\mathbb{P}_s(X|Y)}{2}). \tag{83}$$

Fano and reverse-Fano inequalities for $G_\rho(X|Y)$ *were recently established by Sason and Verdú [37]. As already shown in [27] for* $\rho = 1$*, the use of Schur concavity greatly simplifies the derivation.*

Figure 4 shows some optimal Fano regions for $H_{1/2}(X)$, $H(X)$, $H_2(X)$ and $\log G(X)$.

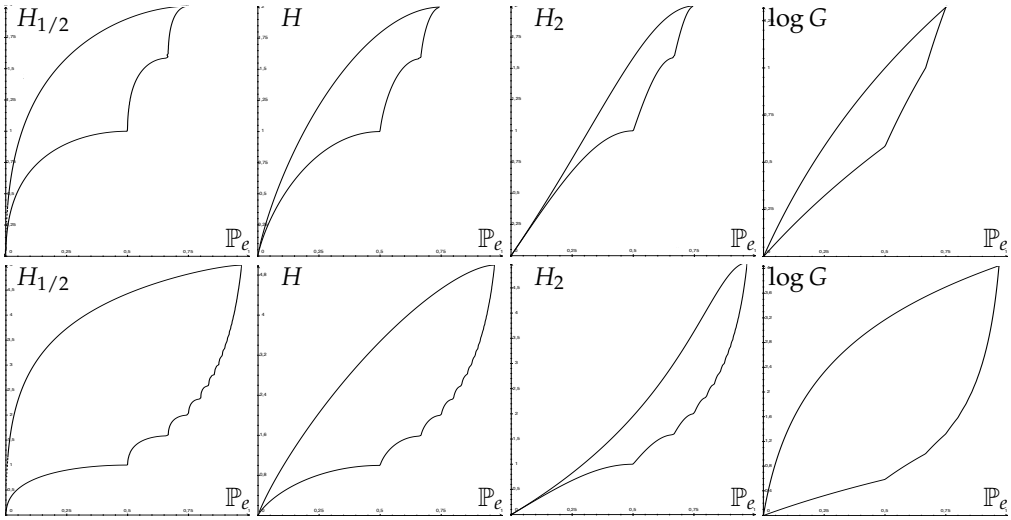

Figure 4. Optimal Fano regions: Entropies (in bits) vs. error probability. **Top** row $M = 4$; **bottom** row $M = 32$.

4. Pinsker and Reverse-Pinsker Inequalities

Pinsker and reverse-Pinsker inequalities relate some divergence measure (e.g., $d(p\|q)$ or $d_\alpha(p\|q)$) between two distributions to their statistical distance $\Delta(p, q)$. For simplicity, even though we restrict ourselves to the divergence or distance to the uniform distribution $q = u$, we still use the generic name "Pinsker inequalities". Following the discussion in Section 1.6, we adopt the following.

Definition 7 (Pinsker-type inequalities)**.** *A "Pinsker inequality" (resp. "reverse-Pinsker inequality") for* $\mathfrak{R}(X)$ *gives an upper (resp. lower) bound of* $\mathfrak{R}(X)$ *as a function of the statistical randomness* $R(X)$ *(or statistical distance* $\Delta(p, u)$*). Pinsker and reverse-Pinsker inequalities are similarly defined for conditional randomness* $\mathfrak{R}(X|Y)$*, lower or upper bounded as a function of* $R(X|Y)$*.*

In this Section, we establish optimal Pinsker and reverse-Pinsker inequalities, where upper and lower bounds are tight. In other words, we determine the maximum and minimum of $\mathfrak{R}$ for fixed R (or fixed Δ). The exact locus of the region $p \in \mathcal{P}_M \mapsto (R(p), \mathfrak{R}(p)) = (R(X), \mathfrak{R}(X))$, as well as the exact locus of all attainable values of $(R(X|Y), \mathfrak{R}(X|Y))$ is determined analytically for fixed M, based on the following.

Lemma 4. *Let $R = R(p)$ and $\Delta = \Delta(p, u) = 1 - \frac{1}{M} - R$. For any M-ary probability distribution $p \in \mathcal{P}_M$ and any integer K such that*

$$|\{p > \tfrac{1}{M}\}| \le K \le |\{p \ge \tfrac{1}{M}\}|, \tag{84}$$

where $|A|$ denotes the cardinality of the set A, one has the following:

$$(\Delta + \tfrac{1}{M}, \underbrace{\tfrac{1}{M}, \ldots, \tfrac{1}{M}}_{\lfloor MR \rfloor \text{ times}}, R - \tfrac{\lfloor MR \rfloor}{M}, 0, \ldots, 0) \preceq p \preceq (\underbrace{\tfrac{1}{M} + \tfrac{\Delta}{K}, \ldots, \tfrac{1}{M} + \tfrac{\Delta}{K}}_{K \text{ times}}, \underbrace{\tfrac{1}{M} - \tfrac{\Delta}{M-K}, \ldots, \tfrac{1}{M} - \tfrac{\Delta}{M-K}}_{M-K \text{ times}}). \tag{85}$$

Proof. Let T_+ be defined as in (17) for a uniform distribution $q = u$. Then, $K = |T_+|$ satisfies (84), and (16) gives $\Delta = \mathbb{P}(T_+) - \frac{K}{M}$. First, consider the largest K probabilities, which are all $\ge \frac{1}{M}$ and sum to $\mathbb{P}(T_+) = \frac{K}{M} + \Delta$. One obtains the following:

$$\tfrac{1}{M} + (\Delta, 0, \ldots, 0) \preceq (p_{(1)}, p_{(2)}, \ldots, p_{(K)}) \preceq (\tfrac{1}{M} + \tfrac{\Delta}{K}, \ldots, \tfrac{1}{M} + \tfrac{\Delta}{K}) \tag{86}$$

where, on the right side, we have used Lemma 1 and, on the left side, we have used Lemma 2, applied to $(p_{(1)} - \frac{1}{M}, p_{(2)} - \frac{1}{M}, \ldots, p_{(K)} - \frac{1}{M})$, which sum to Δ. Next, consider the smallest $M - K$ probabilities, which are all $\le \frac{1}{M}$ and sum to $1 - \mathbb{P}(T_+) = \frac{M-K}{M} - \Delta$. One has the following:

$$(\tfrac{1}{M}, \ldots, \tfrac{1}{M}, r, 0, \ldots, 0) \preceq (p_{(K+1)}, p_{(K+2)}, \ldots, p_{(M)}) \preceq (\tfrac{1}{M} - \tfrac{\Delta}{M-K}, \ldots, \tfrac{1}{M} - \tfrac{\Delta}{M-K}) \tag{87}$$

where, on the right side, we have used Lemma 1 and, on the left side, we have used Lemma 2 with $P = \frac{1}{M}$. Combining (86) and (87) gives (85), where the remainder component $0 \le r < \frac{1}{M}$ is computed so that the sum of probabilities on the left side equals one, which gives $r = (1 - \Delta) - \frac{\lfloor M(1-\Delta) \rfloor}{M} = R - \frac{\lfloor MR \rfloor}{M}$. □

Theorem 6 (Optimal Pinsker and Reverse-Pinsker Inequalities for $\mathfrak{R}(X)$)**.** *The optimal Pinsker and reverse-Pinsker inequalities for the randomness measure $\mathfrak{R}(X)$ of any M-ary random variable X in terms of $R = R(X)$ are given analytically as below:*

$$\mathfrak{R}(1 - R, \tfrac{1}{M}, \ldots, \tfrac{1}{M}, R - \tfrac{\lfloor MR \rfloor}{M}, 0 \ldots) \le \mathfrak{R}(X) \le \max_K \mathfrak{R}(\tfrac{1}{M} + \tfrac{\Delta}{K}, \ldots, \tfrac{1}{M} + \tfrac{\Delta}{K}, \tfrac{1}{M} - \tfrac{\Delta}{M-K}, \ldots, \tfrac{1}{M} - \tfrac{\Delta}{M-K}) \tag{88}$$

where $\Delta = 1 - \frac{1}{M} - R$ and the maximum is over all integers $1 \le K \le \lfloor M(1 - \Delta) \rfloor = 1 + \lfloor MR \rfloor$.

Proof. Apply Lemma 4 and Theorem 3. The Pinsker and reverse-Pinsker bounds are achieved by the distributions on the left and right sides of (85), respectively. The best value of K maximize the randomness $\mathfrak{R}$ of the distribution on the right side of (85), with the constraint $\frac{1}{M} - \frac{\Delta}{M-K} \ge 0$, that is, $K \le M(1 - \Delta)$. □

Assuming the zero randomness convention for simplicity (Remark 14), Pinsker and reverse-Pinsker bounds can be qualitatively described as follows. They are illustrated in Figure 5.

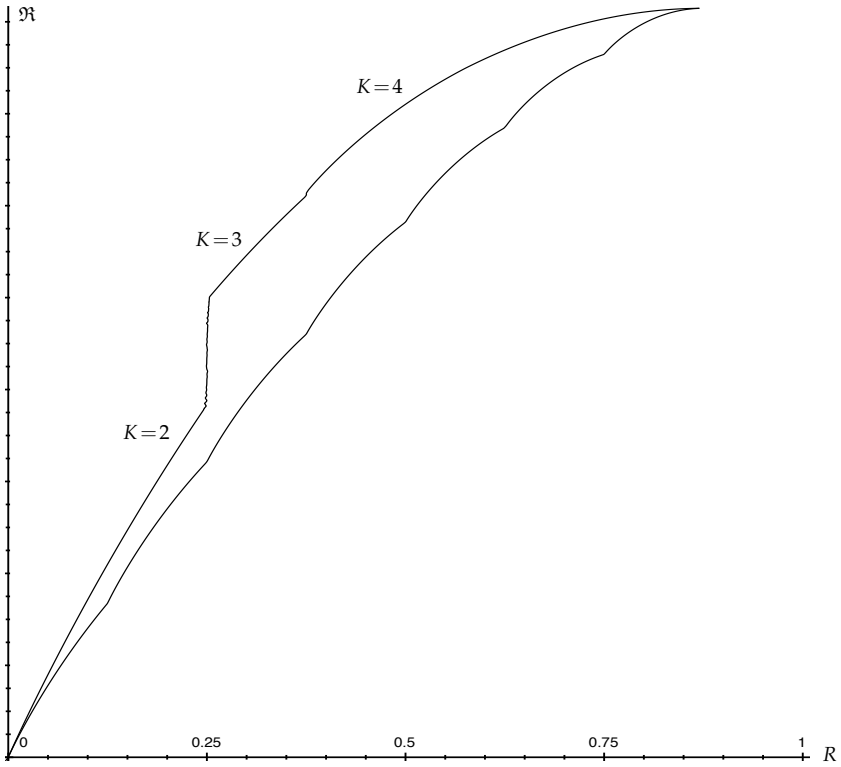

Figure 5. Typical lower and upper Pinsker bounds for $M = 8$. Some optimal values of K are given in this example.

Proposition 3 (Shape of Pinsker Bounds)**.** *The (upper) Pinsker bound:*

$$R \in [0, 1 - \tfrac{1}{M}] \mapsto \max_K \mathfrak{R}(\tfrac{1}{M} + \tfrac{\Delta}{K}, \ldots, \tfrac{1}{M} + \tfrac{\Delta}{K}, \tfrac{1}{M} - \tfrac{\Delta}{M-K}, \ldots, \tfrac{1}{M} - \tfrac{\Delta}{M-K}) \in [0, \mathfrak{R}_M] \qquad (89)$$

where $\Delta = 1 - \frac{1}{M} - R$ *and the maximum is over all integers* $1 \leq K \leq \lfloor M(1-\Delta) \rfloor = 1 + \lfloor MR \rfloor$*, is increasing and piecewise continuous in each subinterval* $[\frac{k}{M}, \frac{k+1}{M}]$*,* $(k = 0, \ldots, M-1)$*, with possible jump discontinuities at points* $\frac{k}{M}$ $(k = 1, \ldots, M-2)$*.*

Proof. First, notice that the distributions $(\frac{1}{M} + \frac{\Delta}{K}, \ldots, \frac{1}{M} + \frac{\Delta}{K}, \frac{1}{M} - \frac{\Delta}{M-K}, \ldots, \frac{1}{M} - \frac{\Delta}{M-K})$ are not necessarily comparable in terms of equalization (partial) order for different values of K. It follows that, in general, the optimal value of K maximizing $\mathfrak{R}(\frac{1}{M} + \frac{\Delta}{K}, \ldots, \frac{1}{M} + \frac{\Delta}{K}, \frac{1}{M} - \frac{\Delta}{M-K}, \ldots, \frac{1}{M} - \frac{\Delta}{M-K})$ depends not only on Δ (or R), but also on the choice of the randomness measure $\mathfrak{R}$.

However, for fixed K, $\Delta \mapsto (\frac{1}{M} + \frac{\Delta}{K}, \ldots, \frac{1}{M} + \frac{\Delta}{K}, \frac{1}{M} - \frac{\Delta}{M-K}, \ldots, \frac{1}{M} - \frac{\Delta}{M-K})$ is linear. In addition, since $\mathfrak{R}(p) \geq 0$ is concave over $\mathcal{P}_M$ (Theorem 2), it is continuous on the interior of $\mathcal{P}_M$. Therefore, the bound $\mathfrak{R}(\frac{1}{M} + \frac{\Delta}{K}, \ldots, \frac{1}{M} + \frac{\Delta}{K}, \frac{1}{M} - \frac{\Delta}{M-K}, \ldots, \frac{1}{M} - \frac{\Delta}{M-K})$ results from the composition of a linear and a continuous concave function. It is, therefore, continuous and concave over the domain $K \leq 1 + \lfloor MR \rfloor$, that is, $R \in [\frac{K-1}{M}, 1 - \frac{1}{M}]$. Also, it is clear, using a suitable Robin Hood operation, that, for a fixed K, $\mathfrak{R}(\frac{1}{M} + \frac{\Delta}{K}, \ldots, \frac{1}{M} + \frac{\Delta}{K}, \frac{1}{M} - \frac{\Delta}{M-K}, \ldots, \frac{1}{M} - \frac{\Delta}{M-K})$ is decreasing in Δ, and, therefore, increasing in R.

It follows that the (upper) Pinsker bound is a maximum of at most M increasing continuous concave functions, defined over intervals of the form $[\frac{K-1}{M}, 1 - \frac{1}{M}]$. It is, therefore, increasing over the entire interval $[0, 1 - \frac{1}{M}]$ and piecewise continuous in each subinterval $[\frac{k}{M}, \frac{k+1}{M}]$, with possible jumps at the endpoints (see Figure 5). □

Proposition 4 (Shape of reverse-Pinsker Bounds)**.** *The (lower) reverse-Pinsker bound:*

$$R \in [0, 1 - \tfrac{1}{M}] \mapsto \mathfrak{R}(1 - R, \tfrac{1}{M}, \ldots, \tfrac{1}{M}, R - \tfrac{\lfloor MR \rfloor}{M}, 0 \ldots) \in [0, \mathfrak{R}_M] \tag{90}$$

is continuous in $R > 0$, increases from 0 (for $R = 0$) to $\mathfrak{R}_M$ (for $R = 1 - \frac{1}{M}$) and is composed of continuous concave increasing curves connecting successive points ($R = \frac{k}{M}$, $\mathfrak{R} = r_k$ for $k = 0, 1, \ldots, M-1$, where the following holds:

$$r_k = \mathfrak{R}(1 - \tfrac{k}{M}, \tfrac{1}{M}, \ldots, \tfrac{1}{M}). \tag{91}$$

Proof. For fixed $k = \lfloor MR \rfloor$, that is, $\frac{k}{M} \leq R < \frac{k+1}{M}$, the bound $\mathfrak{R}(1 - R, \frac{1}{M}, \ldots, \frac{1}{M}, R - \frac{k}{M}, 0 \ldots)$ results from the composition of a linear and a concave function. It is, therefore, concave, and continuous at every $R > 0$. It is clear, using a suitable Robin Hood operation on $(1 - R, R - \frac{k}{M})$, that this bound increases with R on the subinterval $[\frac{k}{M}, \frac{k+1}{M}]$. For $R = \frac{k}{M}$, it equals $\mathfrak{R}(1 - \frac{k}{M}, \frac{1}{M}, \ldots, \frac{1}{M}) = r_k$, which is easily seen, using a suitable Robin Hood operation, to be increasing with k, with maximum $r_{M-1} = \mathfrak{R}_M$. □

Theorem 7 (Optimal Pinsker and Reverse-Pinsker Inequalities for $\mathfrak{R}(X|Y)$)**.** *The optimal Pinsker and reverse-Pinsker inequalities for the randomness measure $\mathfrak{R}(X|Y)$ of any M-ary random variable X in terms of $R = R(X|Y)$ are given by the convex envelope of the Pinsker region determined by (88). In particular, consider the following:*

- *If the (upper) Pinsker bound for $\mathfrak{R}(X)$ is concave (with no discontinuities), then the same optimal bound holds for $\mathfrak{R}(X|Y)$ in terms of $R(X|Y) = R = 1 - \frac{1}{M} - \Delta$:*

$$\mathfrak{R}(X|Y) \leq \max_K \mathfrak{R}(\tfrac{1}{M} + \tfrac{\Delta}{K}, \ldots, \tfrac{1}{M} + \tfrac{\Delta}{K}, \tfrac{1}{M} - \tfrac{\Delta}{M-K}, \ldots, \tfrac{1}{M} - \tfrac{\Delta}{M-K}); \tag{92}$$

- *If the sequence $r_k - r_{k-1}$ ($k = 1, \ldots, M-1$) is nondecreasing, where r_k is defined by (91), then the optimal (lower) reverse-Pinsker bound for $\mathfrak{R}(X|Y)$ is given by the piecewise linear function connecting points $(\frac{k}{M}, r_k)$;*
- *If the sequence $r_k - r_{k-1}$ ($k = 1, \ldots, M-1$) is nonincreasing, then the optimal (lower) reverse-Pinsker bound for $\mathfrak{R}(X|Y)$ writes as follows:*

$$\mathfrak{R}(X|Y) \geq \frac{\mathfrak{R}_M - \mathfrak{R}_0}{1 - 1/M} R(X|Y) + \mathfrak{R}_0 \tag{93}$$

where, as before: $\mathfrak{R}_k = \mathfrak{R}(\frac{1}{k}, \ldots, \frac{1}{k})$ and $\mathfrak{R}_0 = \mathfrak{R}(0)$.

Proof. The Pinsker region for $X|Y = y$, i.e., the locus of the points $(R(p_{X|y}), \mathfrak{R}(p_{X|y}))$ for each $Y = y$, is given by the inequalities (88). From the definition of conditional randomness, the exact locus of points $(R(X|Y), \mathfrak{R}(X|Y)) = \mathbb{E}_y(R(p_{X|y}), \mathfrak{R}(p_{X|y}))$ is composed of all convex combinations of points in the Pinsker region, that is, its convex envelope.

The extreme points $(R = 0, \mathfrak{R} = \mathfrak{R}_1 = 0)$ and $(R = 1 - \frac{1}{M}, \mathfrak{R} = \mathfrak{R}_M)$ are unchanged. The upper Pinsker bound joining these two extreme points is piecewise concave by Proposition 3 and, therefore, if continuous, already belongs to the convex envelope. It follows, in this case, that the upper Pinsker bound in (88) remains the same, as given in (92).

The lower reverse-Pinsker bound for $\mathfrak{R}(X|Y)$ is the convex hull of the lower bound in (88). By Proposition 4, if the sequence $r_k - r_{k-1}$ is non nondecreasing, the piecewise linear curve joining all singular points ($R = \frac{k}{M}$, $\mathfrak{R} = r_k$) for $k = 0, 1, \ldots, M-1$) is convex and already coincides with its convex hull. If, on the contrary, the sequence $r_k - r_{k-1}$ is non nonincreasing, that piecewise linear curve is concave, and its convex hull is simply the straight line joining the extreme endpoints ($R = 0$, $\mathfrak{R} = r_0 = \mathfrak{R}_1 = 0$) and ($R = 1 - \frac{1}{M}$, $\mathfrak{R} = \mathfrak{R}_M$), which is given by (93). □

Remark 20 (φ-Pinsker Bounds). *If $\varphi(\mathfrak{R})$ is used instead of $\mathfrak{R}$, where φ is an increasing function (in particular, to define conditional randomness as in Remark 4), then Theorem 6 can be directly applied to $\mathfrak{R}$. When φ is nonlinear, this may result in (upper) Pinsker bounds that are no longer concave.*

However, to obtain the reverse-Pinsker inequalities for $\mathfrak{R}(X|Y)$, one has to apply Theorem 7 to $\varphi(\mathfrak{R}(X|Y))$ and then apply the inverse function φ^{-1} to (92). When φ is nonlinear, the resulting "reverse-Pinsker bound" for $\mathfrak{R}(X|Y)$ is no longer piecewise linear. This is the case, e.g., for conditional α-entropies (see Example 19 below).

Example 18 (Pinsker and reverse-Pinsker Inequalities for Entropy). *For the Shannon entropy, the optimal Pinsker bounds of Theorem 6 are easily determined as shown:*

$$\begin{aligned}(1-R)\log\frac{1}{1-R}+\tfrac{\lfloor MR\rfloor}{M}\log M+(R-\tfrac{\lfloor MR\rfloor}{M})\log\frac{1}{R-\frac{\lfloor MR\rfloor}{M}}\\ \leq H(X)\leq\\ \max_{1\leq K\leq\lfloor M(1-\Delta)\rfloor}\Big((\tfrac{K}{M}+\Delta)\log\frac{1}{\frac{1}{M}+\frac{\Delta}{K}}+(1-\tfrac{K}{M}-\Delta)\log\frac{1}{\frac{1}{M}-\frac{\Delta}{M-K}}\Big)\end{aligned}\tag{94}$$

where $R=R(X)$ and $\Delta=1-\frac{1}{M}-R(X)$. The maximizing value of K depends on the value of Δ. The lower bound was proven in implicit form in Thm. 3 in [38], while the upper bound was given in Thm. 26 in [39].

Here, (91) is of the form $r_k=\phi(\frac{k}{M})$, where $\phi(x)=(1-x)\log\frac{1}{1-x}+x\log M$ is strictly concave increasing for $0\leq x\leq 1-\frac{1}{M}$. As a consequence, the sequence r_k-r_{k-1} is decreasing for $k=1,\ldots,M-1$, and, by Theorem 7, the optimal reverse-Pinsker inequality for conditional entropy is simply the following:

$$H(X|Y)\geq\frac{M\log M}{M-1}R(X|Y).\tag{95}$$

Example 19 (Pinsker and reverse-Pinsker Inequalities for α-Entropy and for R_2). *By Remark 20, the optimal Pinsker and reverse-Pinsker inequalities (88) for α-entropy $H_\alpha(X)$ are given as below:*

$$\begin{aligned}\tfrac{1}{1-\alpha}\log\big((1-R)^\alpha+\tfrac{\lfloor MR\rfloor}{M^\alpha}+(R-\tfrac{\lfloor MR\rfloor}{M})^\alpha\big)\\ \leq H_\alpha(X)\leq\\ \max_{1\leq K\leq\lfloor M(1-\Delta)\rfloor}\tfrac{1}{1-\alpha}\log\big(K(\tfrac{1}{M}+\tfrac{\Delta}{K})^\alpha+(M-K)(\tfrac{1}{M}-\tfrac{\Delta}{M-K})^\alpha\big)\end{aligned}\tag{96}$$

where $R=R(X)$ and $\Delta=1-\frac{1}{M}-R(X)$. Again, the maximizing value of K depends on the value of Δ.

For collision entropy ($\alpha=2$), since $K(\frac{1}{M}+\frac{\Delta}{K})^2+(M-K)(\frac{1}{M}-\frac{\Delta}{M-K})^2=\frac{1}{M}+\frac{M\Delta^2}{K(M-K)}$ achieves its minimum when the integer K is closest to $\frac{M}{2}$, the optimal Pinsker and reverse-Pinsker inequalities simplify to the following:

$$-\log\big((1-R)^2+\tfrac{\lfloor MR\rfloor}{M^2}+(R-\tfrac{\lfloor MR\rfloor}{M})^2\big)\leq H_2(X)\leq-\log\big(\tfrac{1}{M}+\tfrac{M\Delta^2}{K^*(M-K^*)}\big)\tag{97}$$

where $K^*=\min(\lfloor\frac{M}{2}\rfloor,\lfloor M(1-\Delta)\rfloor)$. In terms of R_2, the optimal Pinsker and reverse-Pinsker inequalities read as shown:

$$1-(1-R)^2-\tfrac{\lfloor MR\rfloor}{M^2}-(R-\tfrac{\lfloor MR\rfloor}{M})^2\leq R_2(X)\leq 1-\tfrac{1}{M}-\tfrac{M\Delta^2}{K^*(M-K^*)}.\tag{98}$$

Since $x(1-x) \le \frac{1}{4}$, *one always has* $K(M-K) \le K^*(M-K^*) \le \frac{M^2}{4}$ *(maximum achieved when* $K^* = \frac{M}{2}$*), so that the (upper) Pinsker bound can be further bounded:*

$$\begin{aligned} H_2(X) &\le \log \frac{M}{1+4\Delta^2}, \\ R_2(X) &\le 1 - \frac{1+4\Delta^2}{M} \end{aligned} \tag{99}$$

This upper bound was derived by Shoup Thm 8.36 in [8] and was later re-derived in the Lemma in 4 [40]. This, however, is the optimal *Pinsker bound* only *when* $K^* = \frac{M}{2}$*, that is, when* M *is even and* $\Delta \le \frac{1}{2}$ *(i.e.,* $R \ge \frac{1}{2} - \frac{1}{M}$*).*

By Remark 20, to obtain the optimal reverse-Pinsker inequality for $H_2(X|Y)$*, we consider* $\varphi_2(H_2(X|Y))$*, where, from (30),* $\varphi_2(x) = -\exp(-x/2)$ *and* $\varphi_2^{-1}(y) = -2\log(-y)$*. For this quantity, one has, from (91),* $r_k = \varphi_2(-\log((1-\frac{k}{M})^2 + \frac{k}{M^2}))$ *of the form* $r_k = \phi(\frac{k}{M})$*, where* $\phi(x) = -\sqrt{(1-x)^2 + \frac{x}{M}}$ *is strictly concave increasing for* $0 \le x \le 1 - \frac{1}{M}$*. As a consequence, the sequence* $r_k - r_{k-1}$ *is decreasing for* $k = 1, \dots, M-1$*, and, by Theorem 7, the optimal reverse-Pinsker bound for conditional 2-entropy is* $\varphi_2^{-1}(\frac{\varphi_2(\log M) - \varphi_2(0)}{1-1/M} R(X|Y) + \varphi_2(0))$*, which gives the optimal reverse-Pinsker inequality:*

$$H_2(X|Y) \ge -2\log\Big(1 - \frac{R(X|Y)}{1+\frac{1}{\sqrt{M}}}\Big). \tag{100}$$

For $R_2(X|Y)$*, one has* $r_k = 1 - (1-\frac{k}{M})^2 - \frac{k}{M^2} = \psi(\frac{k}{M})$*, where* $\psi(x) = (2 - \frac{1}{M})x - x^2$ *is strictly concave increasing for* $0 \le x \le 1 - \frac{1}{M}$*. As a consequence, the sequence* $r_k - r_{k-1}$ *is decreasing for* $k = 1, \dots, M-1$*, and, since* $\mathfrak{R}_M = 1 - \frac{1}{M}$*, by Theorem 7, the optimal reverse-Pinsker inequality for* $R_2(X|Y)$ *is simply as below:*

$$R_2(X|Y) \ge R(X|Y) \tag{101}$$

(see Figure 6).

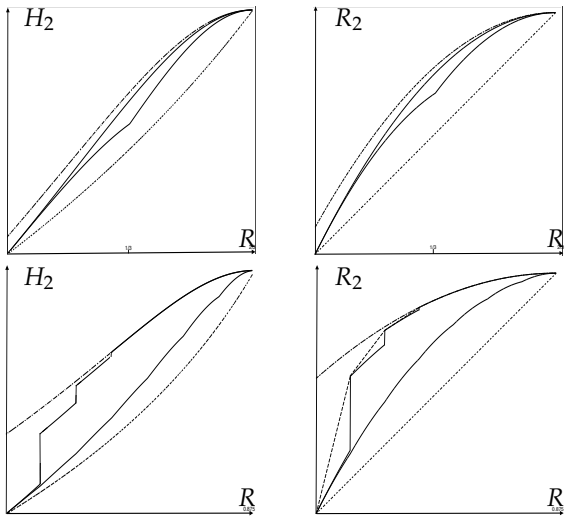

Figure 6. Optimal Pinsker regions: H_2 (in bits) and R_2 vs. statistical randomness R. Solid: Pinsker region $H_2(X)$ (resp. $R_2(X)$) vs. $R(X)$. Dashed: Pinsker region $H_2(X|Y)$ (resp. $R_2(X|Y)$) vs. $R(X|Y)$. Dash-dotted: Shoup's upper bound (99). **Top** row $M = 3$; **bottom** row $M = 8$.

Example 20 (Pinsker and reverse-Pinsker Inequalities for Guessing Entropy)**.** *For the guessing entropy, the optimal Pinsker bounds of Theorem 6 are easily determined:*

$$1+(\lfloor MR(X)\rfloor+1)(R(X)-\frac{\lfloor MR(X)\rfloor}{2M})\le G(X)\le 1+\frac{MR(X)}{2}. \tag{102}$$

A notable property is that the optimal upper bound does not depend on the value of K. The upper bound is mentioned by Pliam in [4] as an upper bound of $\Delta(p,u)$. The methodology of this paper, based on Schur concavity, greatly simplifies the derivation.

For the conditional guessing entropy $G(X|Y)$, observe that the upper Pinsker bound for $G(X)$ is linear (hence, concave) in R and that (91) is of the form $r_k=1+\frac{k(k+1)}{2M}$, where the sequence $r_k-r_{k-1}=\frac{k}{M}$ is increasing. Therefore, by Theorem 7, the optimal Pinsker region for conditional entropy $G(X|Y)$ is the same as for $G(X)$:

$$1+(\lfloor MR(X|Y)\rfloor+1)(R(X|Y)-\frac{\lfloor MR(X|Y)\rfloor}{2M})\le G(X|Y)\le 1+\frac{MR(X|Y)}{2}. \tag{103}$$

Figure 7 shows some optimal Pinsker regions for $H_{1/2}(X)$, $H(X)$, $H_2(X)$ and $\log G(X)$.

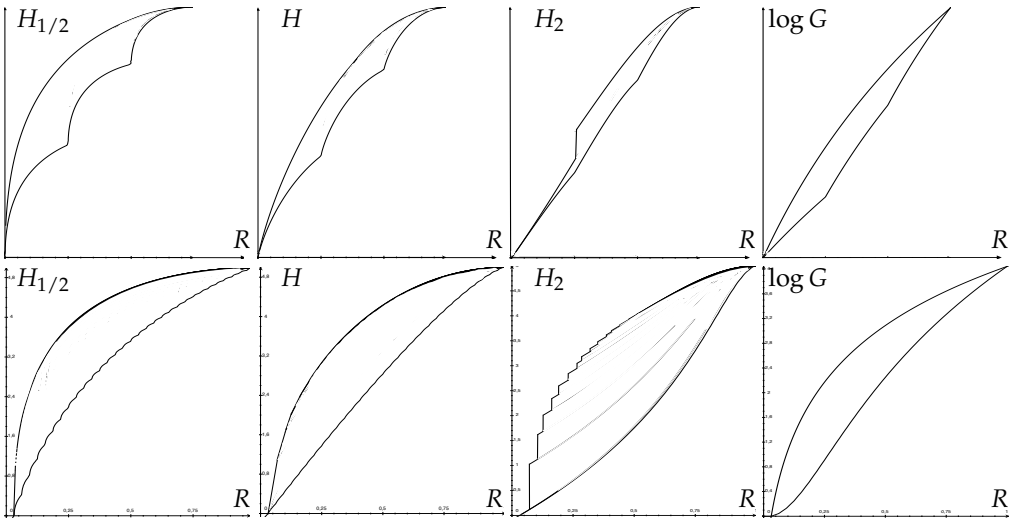

Figure 7. Optimal Pinsker regions: Entropies (in bits) vs. statistical randomness R. **Top** row $M=4$; **bottom** row $M=32$.

Example 21 (Statistical Randomness vs. Probability of Error)**.** *As a final example, we present the optimal regions of statistical randomness R vs. probability of error $\mathbb{P}_e$. In this case, observe the following from Definitions 6 and 7:*

- *The (optimal) Fano inequality for R is the same as the (optimal) reverse-Pinsker inequality for $\mathbb{P}_e$;*
- *The (optimal) Pinsker inequality for $\mathbb{P}_e$ is the same as the (optimal) reverse-Fano inequality for R.*

Letting $R=R(X)$ and $\mathbb{P}_s=\mathbb{P}_s(X)$, Theorem 4 readily gives the optimal Fano and reverse-Fano inequalities:

$$\frac{1}{2}\big(1-\tfrac{1}{M}-(\mathbb{P}_s-\tfrac{2}{M})\lfloor\tfrac{1}{\mathbb{P}_s}\rfloor-\big|1-\lfloor\tfrac{1}{\mathbb{P}_s}\rfloor\mathbb{P}_s-\tfrac{1}{M}\big|\big)\le R(X)\le\mathbb{P}_e(X) \tag{104}$$

while Theorem 6 gives the optimal Pinsker and reverse-Pinsker inequalities:

$$R(X)\le\mathbb{P}_e(X)\le 1-\tfrac{1}{M}-\tfrac{\Delta}{\lfloor M(1-\Delta)\rfloor}=\frac{R+\lfloor MR\rfloor-\frac{\lfloor MR\rfloor}{M}}{1+\lfloor MR\rfloor} \tag{105}$$

since the maximum of $1-\frac{1}{M}-\frac{\Delta}{K}$ in the right side of (88) is for maximum $K=\lfloor M(1-\Delta)\rfloor$.

Similarly, letting $R = R(X|Y)$ *and* $\mathbb{P}_s = \mathbb{P}_s(X|Y)$, *Theorem 5 with* $\mathfrak{R}_k = \frac{k-1}{M}$ *readily gives the optimal Fano and reverse-Fano inequalities:*

$$\frac{(\lceil \frac{1}{\mathbb{P}_s} \rceil \mathbb{P}_s - 1)(\lfloor \frac{1}{\mathbb{P}_s} \rfloor^2 - \lfloor \frac{1}{\mathbb{P}_s} \rfloor) + (1 - \lfloor \frac{1}{\mathbb{P}_s} \rfloor \mathbb{P}_s)(\lceil \frac{1}{\mathbb{P}_s} \rceil^2 - \lceil \frac{1}{\mathbb{P}_s} \rceil)}{M} \leq R(X|Y) \leq \mathbb{P}_e(X|Y) \quad (106)$$

while Theorem 7 gives the optimal Pinsker and reverse-Pinsker inequalities:

$$R(X|Y) \leq \mathbb{P}_e(X|Y) \leq 1 - \frac{2}{\lfloor MR \rfloor + 2} + \frac{MR}{(\lfloor MR \rfloor + 1)(\lfloor MR \rfloor + 2)} \quad (107)$$

where the upper bound is the piecewise linear function connecting points $(\mathbb{P}_e = 1 - \frac{1}{k+1}, R = \frac{k}{M})$ *for* $k = 0, 1, \ldots, M-1$.

From the above observation, the left (reverse-Fano) inequality in (104) *is equivalent to the right (Pinsker) inequality in* (105), *and, similarly, the left (reverse-Fano) inequality in* (106) *is equivalent to the right (Pinsker) inequality in* (107), *which do not seem obvious from the expressions above. The optimal Fano/Pinsker region is illustrated in Figure 8.*

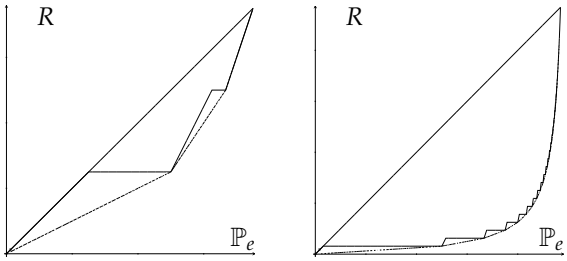

Figure 8. Optimal Fano/Pinsker region for R vs. $\mathbb{P}_e$. Solid: region $R(X)$ vs. $\mathbb{P}_e(X)$. Dashed: region $R(X|Y)$ vs. $\mathbb{P}_e(X|Y)$. **Left** $M = 4$; **right** $M = 32$.

5. Some Applications

Fano and Pinsker inequalities find many applications in many areas of science; we only mention a few. They have been applied in character recognition [33], feature selection [7], Bayesian statistical experiments [17], statistical data processing [13], quantization [41], hypothesis testing [36], entropy estimation [38], channel coding [42], sequential decoding [11] and list decoding [36,43], lossless compression [37,43,44] and guessing [37,44], knowledge representation [12], cipher security measures [4], hash functions [8], randomness extractors [40], information flow [18], statistical decision making [20] and side-channel analysis [14,27,45]. Some of the various inequalities used for these applications are not optimal (or not proven optimal) for various reasons (simplicity of the expressions, approximations, etc.). By contrast, the methodology of this paper always provides optimal direct or reverse-Fano and -Pinsker inequalities.

6. Conclusions and Perspectives

We have derived optimal regions for randomness measures compared to either the error probability or the statistical randomness (or the total variation distance). One perspective is to provide similar optimal regions relating two arbitrary randomness measures. Of course, by (6), Fano regions such as H_α vs. $\mathbb{P}_e$ can be trivially reinterpreted as regions H_α vs. H_∞ (see, e.g., Figure 2 in [42] for the region H vs. H_∞). Using some more involved derivations, the authors of [46] have investigated the optimal regions H vs. H_2 and, more generally, the authors of [47,48] have investigated the optimal regions between two α-entropies of different orders. It would be desirable to apply the methods of this paper to the more general case of two arbitrary randomness measures. In particular, the determination of the optimal regions H_α vs. G_ρ will allow one to assess the sharpness of the "Massey-type" inequalities of [5].

Catalytic majorization [49] was found to be a necessary and sufficient condition for the increase of all Rényi entropies (including the ones with negative parameters α). It

would be interesting to find similar necessary and sufficient conditions for other types of randomness measures.

It is also possible to generalize the notion of entropies and other randomness quantities with respect to an arbitrary dominating measure instead of the counting measure, e.g., to extend the considerations of this paper from the discrete case to the continuous case. The relevant notion of majorization in this more general context is studied, e.g., in [50].

Concerning Pinsker regions, another perspective is to extend the results of this paper to the more general case of Pinsker and reverse-Pinsker inequalities, relating "distances" of two arbitrary distributions p, q by removing the restriction that $q = u$ is uniform. Some results in this direction appear in [38,51–57].

Other types of inequalities on randomness measures with different constraints can also be obtained via majorization theory [43,44].

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

$X \sim p$	X follows the probability distribution p
$H = H_1$	Shannon entropy
H_2	collision entropy
H_∞	min-entropy
H_α	α-entropy
$G = G_1$	guessing entropy
G_ρ	ρ-guessing moment
$\mathbb{P}_e$	probability of error
$\mathbb{P}_e^m$	error probability of order m
$\mathbb{P}_s = 1 - \mathbb{P}_e$	probability of success
$R = R_1$	statistical randomness
$\Delta = 1 - \frac{1}{M} - R$	statistical distance to the uniform
R_2	complementary index of coincidence
$\mathfrak{R}$	any randomness measure

References

1. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
2. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 1st ed.; John Wiley & Sons: Hoboken, NJ, USA, 1990.
3. Massey, J.L. Guessing and entropy. In Proceedings of the IEEE International Symposium on Information Theory, Trondheim, Norway, 27 June–1 July 1994; p. 204.
4. Pliam, J.O. Guesswork and Variation Distance as Measures of Cipher Security. In *Selected Areas in Cryptography. SAC 1999. Lecture Notes in Computer Science*; Heys, H., Adams, C., Eds.; Springer: Berlin/Heidelberg, Germany, 1999; Volume 1758, pp. 62–77.
5. Rioul, O. Variations on a theme by Massey. *IEEE Trans. Inf. Theory* **2022**, *68*, 2813–2828. [CrossRef]
6. Tănăsescu, A.; Choudary, M.O.; Rioul, O.; Popescu, P.G. Tight and Scalable Side-Channel Attack Evaluations through Asymptotically Optimal Massey-like Inequalities on Guessing Entropy. *Entropy* **2021**, *23*, 1538. [CrossRef] [PubMed]
7. Ben-Bassat, M. f-Entropies, Probability of Error, and Feature Selection. *Inf. Control* **1978**, *39*, 227–242. [CrossRef]
8. Shoup, V. *A Computational Introduction to Number Theory and Algebra*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2009.
9. Rényi, A. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*; Contributions to the Theory of Statistics; University of California Press: Berkeley, CA, USA, 1961; Volume 1, pp. 547–561.
10. Contreras-Reyes, J.E. Mutual information matrix based on Rényi entropy and application. *Nonlinear Dyn.* **2022**, *110*, 623–633. [CrossRef]

11. Arikan, E. An inequality on guessing and its application to sequential decoding. *IEEE Trans. Inf. Theory* **1996**, *42*, 99–105. [CrossRef]
12. Yager, R.R. On the Maximum Entropy Negation of a Probability Distribution. *IEEE Trans. Fuzzy Syst.* **2015**, *23*. [CrossRef]
13. Basseville, M. Divergence measures for statistical data processing—An annotated bibliography. *Signal Process.* **2013**, *93*. [CrossRef]
14. Liu, Y.; Béguinot, J.; Cheng, W.; Guilley, S.; Masure, L.; Rioul, O.; Standaert, F.X. Improved Alpha-Information Bounds for Higher-Order Masked Cryptographic Implementations. In Proceedings of the IEEE Information Theory Workshop (ITW 2023), Saint Malo, France, 23–28 April 2023.
15. Fehr, S.; Berens, S. On the conditional Rényi entropy. *IEEE Trans. Inf. Theory* **2014**, *60*, 6801–6810. [CrossRef]
16. Arimoto, S. Information measures and capacity of order α for discrete memoryless channels. In *Proceedings of the Second Colloquium Mathematica Societatis János Bolyai*; Number 16 in Topics in Information Theory; Csiszár, I., Elias, P., Eds.; North Holland: Keszthely, Hungary, 1975; pp. 41–52.
17. Vajda, I.; Vašek, K. Majorization, Concave Entropies, and Comparison of Experiments. *Probl. Control. Inf. Theory* **1985**, *14*, 105–115.
18. Alvim, M.S.; Chatzikokolakis, K.; McIver, A.; Morgan, C.; Palamidessi, C.; Smith, G. An axiomatization of information flow measures. *Theor. Comput. Sci.* **2019**, *777*, 32–54. [CrossRef]
19. Sakai, Y. Generalizations of Fano's Inequality for Conditional Information Measures via Majorization Theory. *Entropy* **2020**, *22*, 288. [CrossRef]
20. Américo, A.; Khouzani, M.; Malacaria, P. Conditional Entropy and Data Processing: An Axiomatic Approach Based on Core-Concavity. *IEEE Trans. Inf. Theory* **2020**, *66*, 5537–5547. [CrossRef]
21. Rioul, O. What Is Randomness? The Interplay between Alpha Entropies, Total Variation and Guessing. *Phys. Sci. Forum* **2022**, *5*, 1–9.
22. Fano, R.M. Class notes for course 6.574: Transmission of Information. In *Transmission of Information: A Statistical Theory of Communications*, 1st ed.; MIT Press: Cambridge, MA, USA, 1961.
23. Rioul, O. A Historical Perspective on Schützenberger-Pinsker Inequalities. In Proceedings of the 6th International Conference on Geometric Science of Information (GSI 2023), Saint Malo, France, 30 August–1 September 2023.
24. Schützenberger, M.P. Contribution aux Applications Statistiques de la théorie de l'Information. Ph.D. Thesis, Institut de statistique de l'Université de Paris, Paris, France, 1954; Volume 3.
25. Pinsker, M.S. *Information and Information Stability of Random Variables and Processes*; Holden-Day: San Francisco, CA, USA, 1964. (In Russian)
26. Shannon, C.E. The lattice theory of information, in Report of Proc. Symp. Inf. Theory, London, Sept. 1950. *Trans. IRE Prof. Group Inf. Theory* **1953**, *1*, 105–107. [CrossRef]
27. Béguinot, J.; Cheng, W.; Guilley, S.; Rioul, O. Be my guess: Guessing entropy vs. success rate for evaluating side-channel attacks of secure chips. In Proceedings of the 25th Euromicro Conference on Digital System Design (DSD 2022), Maspalomas, Spain, 31 August–2 September 2022.
28. Arnold, B.C. Majorization and the Lorenz Order: A Brief Introduction. In *Lecture Notes in Statistics*; Springer: Berlin/Heidelberg, Germany, 1987; Volume 43.
29. Marshall, A.W.; Olkin, I.; Arnold, B.C. *Inequalities: Theory of Majorization and Its Applications*, 2nd ed.; Springer Series in Statistics; Springer: Berlin/Heidelberg, Germany, 2011.
30. Rioul, O.; Béguinot, J.; Rabiet, V.; Souloumiac, A. La véritable (et méconnue) théorie de l'information de Shannon. In Proceedings of the 28e Colloque GRETSI 2022, Nancy, France, 6–9 September 2022.
31. Cicalese, F.; Vaccaro, U. Supermodularity and Subadditivity Properties of the Entropy on the Majorization Lattice. *IEEE Trans. Inf. Theory* **2002**, *48*, 933–938. [CrossRef]
32. Tebbe, D.L.; Dwyer, S.J., III. Uncertainty and probability of error. *IEEE Trans. Inf. Theory* **1968**, *14*, 516–518. [CrossRef]
33. Kovalevsky, V.A. The problem of character recognition from the point of view of mathematical statistics. In *Character Readers and Pattern Recognition*; Spartan: Lymington, UK, 1968; pp. 3–30.
34. Toussaint, G.T. A Generalization of Shannon's Equivocation and the Fano Bound. *IEEE Trans. Syst. Man Cybern.* **1978**, *7*, 300–302.
35. Ben-Bassat, M.; Raviv, J. Rényi's entropy and the probability of error. *IEEE Trans. Inf. Theory* **1978**, *24*, 324–331. [CrossRef]
36. Sason, I.; Verdú, S. Arimoto–Rényi Conditional Entropy and Bayesian *M*-Ary Hypothesis Testing. *IEEE Trans. Inf. Theory* **2018**, *64*, 4–25. [CrossRef]
37. Sason, I.; Verdú, S. Improved Bounds on Lossless Source Coding and Guessing Moments via Rényi Measures. *IEEE Trans. Inf. Theory* **2018**, *64*, 4323–4346. [CrossRef]
38. Ho, S.W.; Yeung, R.W. The Interplay Between Entropy and Variational Distance. *IEEE Trans. Inf. Theory* **2010**, *56*, 5906–5929. [CrossRef]
39. Sason, I.; Verdú, S. f-Divergence Inequalities. *IEEE Trans. Inf. Theory* **2016**, *62*, 5973–6006. [CrossRef]
40. Chevalier, C.; Fouque, P.A.; Pointcheval, D.; Zimmer, S. Optimal Randomness Extraction from a Diffie-Hellman Element. In *Proceedings of the Proc. Eurocrypt'09*; Joux, A., Ed.; Springer: Berlin/Heidelberg, Germany, 2009; Volume 5479, pp. 572–589.
41. Böcherer, G.; Geiger, B.C. Optimal Quantization for Distribution Synthesis. *IEEE Trans. Inf. Theory* **2016**, *62*, 6162–6172. [CrossRef]
42. Feder, M.; Merhav, N. Relations between entropy and error probability. *IEEE Trans. Inf. Theory* **1994**, *40*, 259–266. [CrossRef]
43. Sason, I. On Data-Processing and Majorization Inequalities for f-Divergences with Applications. *Entropy* **2019**, *21*, 1022. [CrossRef]

44. Sason, I. Tight Bounds on the Rényi Entropy via Majorization with Applications to Guessing and Compression. *Entropy* **2018**, *20*, 896. [CrossRef]
45. Béguinot, J.; Cheng, W.; Guilley, S.; Rioul, O. Be My Guesses: The Interplay Between Side-Channel-Leakage Metrics. *Microprocess. Microsyst. (Micpro)* **2023**, *to appear*.
46. Harremoës, P.; Topsøe, F. Inequalities Between Entropy and Index of Coincidence Derived From Information Diagrams. *IEEE Trans. Inf. Theory* **2001**, *47*, 2944–2960. [CrossRef]
47. Harremoës, P. Joint Range of Rényi Entropies. *Kybernetika* **2009**, *45*, 901–911.
48. Sakai, Y.; Iwata, K. Sharp Bounds on Arimoto's Conditional Rényi Entropies between Two Distinct Orders. *arXiv* **2017**, arXiv:1702.00014v2.
49. Klimesh, M. Entropy Measures and Catalysis of Bipartite Quantum State Transformations. In Proceedings of the IEEE International Symposium on Information Theory (ISIT 2004), Chicago, IL, USA, 27 June–2 July 2004; p. 357.
50. Van Erven, T.; Harremoës, P. Rényi divergence and majorization. In Proceedings of the IEEE International Symposium on Information Theory (ISIT 2010), Austin, TX, USA, 12–18 June 2010.
51. Weissman, T.; Ordentlich, E.; Seroussi, G.; Verdú, S.; Weinberger, M.J. *Inequalities for the L_1 Deviation of the Empirical Distribution*; Technical Report HPL-2003-97 (R.1); Hewlett-Packard Laboratories: Palo Alto, CA, USA, 2003.
52. Harremoës, P.; Vajda, I. On Pairs of f-Divergences and Their Joint Range. *IEEE Trans. Inf. Theory* **2011**, *57*, 3230–3235. [CrossRef]
53. Prelov, V.V. On Coupling of Probability Distributions and Estimating the Divergence through Variation. *Probl. Inf. Transm.* **2017**, *53*, 16–22. [CrossRef]
54. Binette, O. A Note on Reverse Pinsker Inequalities. *IEEE Trans. Inf. Theory* **2019**, *65*, 4094–4096. [CrossRef]
55. Prelov, V.V. On the Maximum Values of f-Divergence and Rényi Divergence under a Given Variational Distance. *Probl. Inf. Transm.* **2020**, *56*, 3–14. [CrossRef]
56. Prelov, V.V. On the Maximum f-Divergence of Probability Distributions Given the Value of Their Coupling. *Probl. Inf. Transm.* **2021**, *57*, 24–33. [CrossRef]
57. Guia, X.Y.; Huang, Y.C. Remarks on Reverse Pinsker Inequalities. *Probl. Inf. Transm.* **2022**, *58*, 3–5. [CrossRef]