

What is randomness? The interplay between alpha-entropies, total variation and guessing

Olivier Rioul

▶ **To cite this version:**

Olivier Rioul. What is randomness? The interplay between alpha-entropies, total variation and guessing. 41st International Conference on Bayesian and Maximum Entropy methods in Science and Engineering (MaxEnt 2022), Jul 2022, Paris, France. pp.30, 10.3390/psf2022005030 . hal-03718716

HAL Id: hal-03718716

https://telecom-paris.hal.science/hal-03718716

Submitted on 12 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Article

What is Randomness?

The Interplay Between Alpha Entropies, Total Variation and Guessing

Olivier Rioul[1] orcid number: 0000-0002-8681-8916

[1] LTCI, Télécom Paris, Institut Polytechnique de Paris, France; olivier.rioul@telecom-paris.fr

† International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, IHP, Paris, July 18-22, 2022.

Received: date; Accepted: date; Published: date

Abstract: In many areas of computer science, it is of primary importance to assess the *randomness* of a certain variable X. Many different criteria can be used to evaluate randomness, possibly after observing some disclosed data. A "sufficiently random" X is often described as "entropic". Indeed, Shannon's entropy is known to provide a resistance criterion against modeling attacks. More generally one may consider the Rényi α-entropy where Shannon's entropy, collision entropy and min-entropy are recovered as particular cases $\alpha = 1$, 2 and $+\infty$, respectively. Guess work or guessing entropy is also of great interest in relation to α-entropy.
On the other hand, many applications rely instead on the "statistical distance", a.k.a. *total variation* distance to the uniform distribution. This criterion is particularly important because a very small distance ensures that no statistical test can effectively distinguish between the actual distribution and the uniform distribution.
We establish optimal lower and upper bounds between α-entropy, guessing entropy on one hand, and error probability and total variation distance to the uniform on the other hand. In this context, it turns out that the best known "Pinsker inequality" and recent "reverse Pinsker inequalities" are not necessarily optimal. We recover or improve previous Fano-type and Pinsker-type inequalities used for several applications.

Keywords: Statistical (Total Variation) Distance; α-Entropy; Guessing Entropy; Probability of Error

---

1. Some Well-Known "Randomness" Measures

It is of primary importance to assess the *randomness* of a certain random variable X, which represents some identifier, cryptographic key, signature or any type of intended secret. Applications include pseudo-random bit generators [1], general cipher security [2], randomness extractors [3] and hash functions [4, Chap. 8], physically unclonable functions [5], true random number generators [6], to list but a few. In all of these examples, X takes finitely many values $x \in \{x_1, x_2, \ldots, x_M\}$ with probabilities $p_X(x) = \mathbb{P}(X = x)$. In this paper, it will be convenient to denote

$$p_{(1)} \geq p_{(2)} \geq \cdots \geq p_{(M)} \tag{1}$$

any rearrangement of the probabilities $p(x)$ in *descending* order (where ties can be resolved arbitrarily). Thus $p_{(1)} = \max_x p_X(x)$ is the maximum probability, $p_{(2)}$ the second maximum, etc. Also define the cumulative sums

$$P_{(k)} \triangleq p_{(1)} + \ldots + p_{(k)} \qquad (k = 1, 2, \ldots, M) \tag{2}$$

where in particular $P_{(M)} = 1$.

Many different criteria can be used to evaluate randomness of X or its distribution p_X, depending on the type of attack that can be carried out to recover the whole or part of the secret, possibly after observing disclosed data Y. The observed random variable Y can be any random variable 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 from the unconditional distribution p_X. To simplify notation we write

$$p(x) \triangleq p_X(x) = \mathbb{P}(X = x) \tag{3}$$

$$p(x|y) \triangleq p_{X|y}(x) = \mathbb{P}(X = x|Y = y). \tag{4}$$

A "sufficiently random" secret is often described as *entropic* in the literature. Indeed, Shannon's entropy

$$H(X) = H(p) \triangleq \sum_x p(x) \log \frac{1}{p(x)} = \mathbb{E} \log \frac{1}{p(X)} \tag{5}$$

(with the convention $0 \log \frac{1}{0} = 0$) is known to provide a resistance criterion against modeling attacks. It was introduced by Shannon as a measure of *uncertainty* of X. The average entropy after having observed Y is the usual conditional entropy

$$H(X|Y) \triangleq \mathbb{E}_y H(p_{X|y}) = \mathbb{E} \log \frac{1}{p(X|Y)}. \tag{6}$$

A well-known generalization of Shannon's entropy is the Rényi entropy of order $\alpha > 0$ or *α-entropy*

$$H_\alpha(X) = H_\alpha(p) \triangleq \tfrac{1}{1-\alpha} \log \sum_x p(x)^\alpha = \tfrac{\alpha}{1-\alpha} \log \|p_X\|_\alpha \tag{7}$$

where by continuity as $\alpha \to 1$, the 1-entropy $H_1(X) = H(X)$ is Shannon's entropy. One may consider many different definitions of *conditional* α-entropy [7], but for many applications the preferred choice is Arimoto's definition [8–10]

$$H_\alpha(X|Y) \triangleq \tfrac{\alpha}{1-\alpha} \log \mathbb{E}_y \|p_{X|y}\|_\alpha \tag{8}$$

where the expectation over Y is taken over the "α-norm" inside the logarithm. (Strictly speaking, $\|\cdot\|_\alpha$ is not a norm when $\alpha < 1$.)

For $\alpha = 2$, the collision entropy $H_2(X) = H_2(p) = \log \frac{1}{\mathbb{P}(X=X')}$, where X' is an independent copy of X, is often used to ensure security against collision attacks. Perhaps one of the most popular criteria is the min-entropy defined when $\alpha \to +\infty$ as

$$H_\infty(X) = H_\infty(p) = \log \frac{1}{p_{(1)}} = \log \frac{1}{1 - \mathbb{P}_e(X)}, \tag{9}$$

whose maximization is equivalent to a probability criterion to ensure a worst-case security level. Arimoto's conditional ∞-entropy takes the form

$$H_\infty(X|Y) = \log \frac{1}{1 - \mathbb{P}_e(X|Y)} \tag{10}$$

where we have noted

$$\mathbb{P}_e(X) = \mathbb{P}_e(p) \triangleq 1 - p_{(1)} = 1 - P_{(1)} \tag{11}$$

$$\mathbb{P}_e(X|Y) \triangleq \mathbb{E}_y \mathbb{P}_e(X|y). \tag{12}$$

These quantities correspond to the minimum probability of decision error using a MAP (maximum a posteriori probability) rule (see, e.g., [11]). Guess work or guessing entropy [2,12]

$$G(X) = G(p_X) \triangleq \sum_{i=1}^{M} i \cdot p_{(i)} \tag{13}$$

and more generally guessing moments of order $\rho > 0$ or ρ-guessing entropy

$$G_\rho(X) = G_\rho(p_X) \triangleq \sum_{i=1}^{M} i^\rho \cdot p_{(i)} \tag{14}$$

are also of great interest in relation to α-entropy [10,13,14]. The conditional versions given observation Y are the expectations

$$G_\rho(X|Y) \triangleq \mathbb{E}_y G_\rho(X|y) \tag{15}$$

For $\rho = 1$ this represents the average number of guesses that an attacker has to make to guess the secret X correctly after having observed Y [13].

2. Statistical (Total Variation) Distance to the Uniform Distribution

As shown in the sequel, all quantities introduced in the preceding section (H, H_α, $\mathbb{P}_e$, G, G_ρ) have many properties in common. In particular, each of these quantities attains

- its *minimum* value for a *delta* (Dirac) distribution $p = \delta$, that is, a deterministic random variable X with $p_{(1)} = 1$ and all other probabilities $= 0$;
- its *maximum* value for the *uniform* distribution $p = u$, that is, a uniformly distributed random variable X with $p(x) = \frac{1}{M}$ for all x.

In fact, it can be easily checked that

$$0 \leq H_\alpha(X) \leq \log M \tag{16}$$

$$1 \leq G(X) \leq \tfrac{M+1}{2} \tag{17}$$

$$0 \leq \mathbb{P}_e(X) \leq 1 - \tfrac{1}{M} \tag{18}$$

where the lower (resp. upper) bounds are attained for a delta (resp. uniform) distribution. Thus the uniform distribution is the "most entropic" (H_α), "hardest to guess" (G), and "hardest to detect" ($\mathbb{P}_e$).

The maximum entropy property is related to the minimization of divergence [15]

$$D(p\|u) = \log M - H(p) \tag{19}$$

where $D(p\|q) = \sum p(x) \log \frac{p(x)}{q(x)} \geq 0$ denotes the Kullback-Leibler divergence which vanishes if and only if $p = q$. Thus entropy appears as the complementary value of the divergence to the uniform distribution. Similarly for α-entropy,

$$D_\alpha(p\|u) = \log M - H_\alpha(p) \tag{20}$$

where $D_\alpha(p\|q) = \frac{1}{\alpha - 1} \log \sum_x p(x)^\alpha q(x)^{1-\alpha}$ denotes the Rényi α-divergence [16] (Bhattacharyya distance for $\alpha = \frac{1}{2}$).

Instead of the divergence to the uniform distribution, it is often desirable to rely instead on the statistical distance, a.k.a. *total variation* distance to the uniform distribution. The general expression of the total variation distance is

$$\Delta(p, q) = \tfrac{1}{2} \sum_x |p(x) - q(x)| \tag{21}$$

where the $1/2$ factor is there to ensure that $0 \leq \Delta(p, q) \leq 1$. Equivalently,

$$\Delta(p, q) = \max_T |\mathbb{P}(T) - \mathbb{Q}(T)| \tag{22}$$

where the maximum is over any event T and $\mathbb{P}$, $\mathbb{Q}$ denote the respective probabilities w.r.t. p and q. As is well known, the maximum

$$\Delta(p, q) = \mathbb{P}(T_+) - \mathbb{Q}(T_+) \tag{23}$$

is attained when $T = T_+ = \{x \mid p(x) \geq q(x)\}$.

The total variation criterion is particularly important because a very small distance $\Delta(p, q)$ ensures that no statistical test can effectively distinguish between p and q. In fact, given some observation X following either p (null hypothesis H_0) or q (alternate hypothesis H_1), such a statistical test takes the form « is $X \in T$? » (then accept H_0, otherwise reject H_0). If $|\mathbb{P}(X \in T) - \mathbb{Q}(X \in T)| \leq \Delta(p, q)$ is small

enough, the type-I or type-II errors have total probability $\mathbb{P}(X \notin T) + \mathbb{Q}(X \in T) \approx 1$. Thus, in this sense the two hypotheses p and q are undistinguishable (statistically equivalent).

By analogy with (19), (20) we can then define "*statistical randomness*" $R(X) = R(p) \geq 0$ as the complementary value of the statistical distance to the uniform distribution, i.e., such that

$$\Delta(p, u) = 1 - R(p) \tag{24}$$

holds. With this definition,

$$R(X) = R(p) \triangleq 1 - \tfrac{1}{2} \sum_x |p(x) - \tfrac{1}{M}| \tag{25}$$

is maximum $= 1$ when $\Delta(p, u) = 0$, i.e., $p = u$. Thus the uniform distribution u is the "most random". What is fundamental is that $R(X) \approx 1$ ensures that *no statistical test can effectively distinguish the actual distribution from the uniform distribution*.

Again the "least random" distribution corresponds to the deterministic case. In fact, from (23) we have

$$\Delta(p, u) = \mathbb{P}(T_+) - \tfrac{K}{M} = P_{(K)} - \tfrac{K}{M} \tag{26}$$

where $T_+ = \{x \mid p(x) \geq \frac{1}{M}\}$ of cardinality $K = |T_+|$, and $\mathbb{P}(T_+) = P_{(K)}$ by definition (2). It is easily seen that $\Delta(p, u)$ attains its maximum value $= 1 - \frac{1}{M}$ if and only if $p = \delta$ is a delta distribution. In summary

$$\tfrac{1}{M} \leq R(X) \leq 1 \tag{27}$$

where the lower (resp. upper) bound is attained for a delta (resp. uniform) distribution. The conditional version is again taken by averaging over the observation:

$$R(X|Y) \triangleq \mathbb{E}_y R(X|y). \tag{28}$$

3. F-Concavity: Knowledge Reduces Randomness and Data Processing

Knowledge of the observed data Y (on average) *reduces uncertainty, improves detection or guessing, and reduces randomness* in the sense that:

$$H_\alpha(X|Y) \leq H_\alpha(X) \tag{29}$$

$$G(X|Y) \leq G(X) \tag{30}$$

$$\mathbb{P}_e(X|Y) \leq \mathbb{P}_e(X) \tag{31}$$

$$R(X|Y) \leq R(X) \tag{32}$$

For $\alpha = 1$ the property $H(X|Y) \leq H(X)$ is well-known ("conditioning reduces entropy" [15]) : the difference $H(X) - H(X|Y) = I(X;Y)$ is the mutual information, which is nonnegative. Property (29) for $\alpha \neq 1$ is also well known, see [7,8]. In view of (9)-(10), the case $\alpha = +\infty$ in (29) is equivalent to (31) which is obvious in the sense that any observation can only improve MAP detection. This, as well as (30), is also easily proved directly (see, e.g., [17]).

For all quantities H, $\mathbb{P}_e$, G, R, the conditional quantity is obtained by averaging over the observation as in (6), (12), (15) and (28). Since $p(x) = \mathbb{E}_y p(x|y)$, the fact that knowledge of Y reduces H, $\mathbb{P}_e$, G or R amounts to saying that these are *concave* functions of the distribution p of X. Note that concavity of $R(X) = R(p)$ in p is clear from the definition (25), which shows (32).

For entropy H, this also has been given some physical interpretation: "mixing" distributions (taking convex combinations of probability distributions) *can only increase the entropy on average*. For example, given any two distributions p and q, $H(\lambda p + \bar{\lambda} q) \geq \lambda H(p) + \bar{\lambda} H(q)$ where $0 \leq \lambda = 1 - \bar{\lambda} \leq 1$. Similarly, such *mixing of distributions increases the average probability of error* $\mathbb{P}_e$, *guessing entropy* G, *and statistical randomness* R.

For conditional α-entropy $H_\alpha(X|Y)$ where $\alpha \neq 1$, averaging over Y in the definition (8) is done on the α-norm of the distribution $p_{X|y}$, which is known to be convex for $\alpha > 1$ (by Minkowski's inequality) and concave for $0 < \alpha < 1$ (by the reverse Minkowski inequality). Thus the fact that knowledge reduces α-entropy (inequality (29)) is equivalent to the fact that $H_\alpha(p)$ in (6) is an *F-concave* function, that is, an increasing function F of a concave function in p, where $F(x) = \frac{\alpha}{1-\alpha} \log(\mathrm{sgn}(1-\alpha)x)$. The

average over Y in $H_\alpha(X|Y)$ is done on the quantity $F^{-1}(H_\alpha)$ instead of H_α. Thus, for example, $H_{1/2}(p)$ is a log-concave function of p.

A straightforward generalization of (29)–(32) is the *data processing inequality*: for any *Markov chain* $X - Y - Z$, i.e., such that $p(x|y,z) = p(x|y)$,

$$H_\alpha(X|Y) \le H_\alpha(X|Z) \quad (33) \qquad \mathbb{P}_e(X|Y) \le \mathbb{P}_e(X|Z) \quad (35)$$

$$G(X|Y) \le G(X|Z) \quad (34) \qquad R(X|Y) \le R(X|Z) \quad (36)$$

For $\alpha = 1$ the property $H(X|Y) \le H(X|Z)$ amounts to $I(X;Z) \le I(X;Y)$, i.e., (post)-*processing can never increase information*. Inequalities (33)–(36) can be deduced from (29)–(32) by considering a fixed $Z = z$, averaging over Z to show that $H(X|Y,Z) \le H(X|Z)$, etc. (*additional knowledge reduces randomness*) and then noting that $p(x|y,z) = p(x|y)$ by the Markov property—see, e.g., [7,18] for H_α and [17] for G. Conversely, (29)–(32) can be re-obtained from (33)–(36) as the particular case $Z = 0$ (any deterministic variable representing zero information).

4. S-Concavity: Mixing Increases Randomness and Data Processing

Another type of *mixing* (different from the one described in the preceding section) is also useful in certain physical science considerations. It can be described as a sequence of elementary mixing operations as follows. Suppose that one only modifies two probability values $p_i = p(x_i)$ and $p_j = p(x_j)$ for $i \neq j$. Since the result should be again a probability distribution, the sum $p_i + p_j$ should be kept constant. Then there are two possibilities:

- $|p_i - p_j|$ decreases; the resulting distribution is "smoother", "more spread out", "more disordered"; the resulting operation can be written as $(p_i, p_j) \mapsto (\lambda p_i + \bar{\lambda} p_j, \lambda p_j + \bar{\lambda} p_i)$ where $0 \le \lambda = 1 - \bar{\lambda} \le 1$, also known as "transfer" operation. We call it *elementary mixing operation* or *M-transformation* in short.
- $|p_i - p_j|$ increases; this is the reverse operation, an *elementary unmixing operation* or *U-transformation* in short.

We say that a quantity is *s-concave* if it increases by any M-transformation (equivalently, decreases by any U-transformation). Note that any increasing function F of an s-concave function is again s-concave.

This notion coincides with that of *Schur-concavity* from majorization theory [19]. In fact, we can say that p is *majorized* by q, and we write $p \prec q$, if p is obtained from q by a (finite) sequence of elementary M-transformations, or, what amounts the same, that q majorizes p, that is, q is obtained from p by a (finite) sequence of elementary U-transformations. A well-known result [19, p.34] states that $p \prec q$ if and only if

$$P_{(k)} \le Q_{(k)} \qquad (0 < k < M) \quad (37)$$

(see definition (2)) where always $P_{(M)} = Q_{(M)} = 1$.

From the above definitions it is immediate to see that all previously considered quantities H, H_α,G, G_ρ, $\mathbb{P}_e$, R are s-concave. Thus *mixing increases uncertainty, guessing, error, and randomness*, that is, $p \prec q$ implies

$$H_\alpha(p) \ge H_\alpha(q) \quad (38) \qquad \mathbb{P}_e(p) \ge \mathbb{P}_e(q) \quad (40)$$

$$G_\rho(p) \ge G_\rho(q) \quad (39) \qquad R(p) \ge R(q). \quad (41)$$

For H_α and R this can be easily seen from the fact that these quantities can be written as (an increasing function of) a quantity of the form $\sum_x \phi(p(x))$ where ϕ is concave. Then the effet of an M-transformation $(p_i, p_j) \mapsto (\lambda p_i + \bar{\lambda} p_j, \lambda p_j + \bar{\lambda} p_i)$ gives $\phi(\lambda p_i + \bar{\lambda} p_j) + \phi(\lambda p_j + \bar{\lambda} p_i) \ge \lambda\phi(p_i) + \bar{\lambda}\phi(p_j) + \lambda\phi(p_j) + \bar{\lambda}\phi(p_i) = \phi(p_i) + \phi(p_j)$. For $\mathbb{P}_e$ it is obvious, and for G and G_ρ it is also easily proved using characterization (37) and summation by parts [17].

Another kind of (functional or deterministic) *data processing inequality* can be obtained from (38)–(41) as a particular case. For any deterministic function f,

$$H_\alpha(f(X)) \le H_\alpha(X) \quad (42) \qquad \mathbb{P}_e(f(X)) \le \mathbb{P}_e(X) \quad (44)$$
$$G(f(X)) \le G(X) \quad (43) \qquad R(f(X)) \le R(X) \quad (45)$$

Thus *deterministic processing (by f) decreases (cannot increase) uncertainty, can only make guessing or detection easier, and decreases randomness*. For $\alpha = 1$ the inequality $H(f(X)) \le H(X)$ can also be seen from the data processing inequality of the preceding section by noting that $H(f(X)) = I(f(X); f(X)) \le I(X; f(X)) \le H(X)$ (since $X - f(X) - f(X)$ is trivially a Markov chain).

To prove (42)–(45) in general, consider preimages by f of values of $y = f(x)$; it is enough to show that each of the quantities H_α, $\mathbb{P}_e$, G, or R decreases by the elementary operation consisting in putting together two distincts values x_i, x_j of x in the same preimage of y. But for probability distributions this operation amounts the U-transformation $(p_i, p_j) \mapsto (p_i + p_j, 0)$ and the result follows by s-concavity.

An equivalent property of (42)–(45) is the fact that *any additional random variable Y increases uncertainty, probability of error, guessing, and randomness* in the sense that

$$H_\alpha(X) \le H_\alpha(X, Y) \quad (46) \qquad \mathbb{P}_e(X) \le \mathbb{P}_e(X, Y) \quad (48)$$
$$G(X) \le G(X, Y) \quad (47) \qquad R(X) \le R(X, Y) \quad (49)$$

This is a particular case of (42)–(45) applied to the joint (X, Y) and the first projection $f(x, y) = x$. Conversely, (42)–(45) follows from (46)–(49) by applying it to $(f(X), X)$ in place of (X, Y) and noting that the distribution of $(f(X), X)$ is essentially that of X.

5. Optimal Fano-Type and Pinsker-Type Bounds

We have seen that informational quantities like entropies H, H_α, guessing entropies G, G_ρ on one hand, and statistical quantities like probability of error for MAP detection $\mathbb{P}_e$ and statistical randomness R on the other hand, satisfy many common properties: decrease by knowledge, data processing, increase by mixing, etc. For this reason, it is desirable to establish the best possible bounds between one informational quantity (like H_α or G_ρ) and one statistical quantity ($\mathbb{P}_e$ or $R = 1 - \Delta(p, u)$).

To achieve this, we remark that for any distribution p, we have the following majorizations. For fixed $\mathbb{P}_e = 1 - \mathbb{P}_s$:

$$(\mathbb{P}_s, \tfrac{\mathbb{P}_e}{M-1}, \dots, \tfrac{\mathbb{P}_e}{M-1}) \prec p \prec (\mathbb{P}_s, \dots, \mathbb{P}_s, 1 - K\mathbb{P}_s, 0, \dots, 0) \quad (50)$$

where (necessarily) $K = \lfloor \frac{1}{\mathbb{P}_s} \rfloor$, and for fixed $R = 1 - \Delta$:

$$(\underbrace{\tfrac{1}{M} + \tfrac{\Delta}{K}, \dots, \tfrac{1}{M} + \tfrac{\Delta}{K}}_{K \text{ times}}, \underbrace{\tfrac{1}{M} - \tfrac{\Delta}{M-K}, \dots, \tfrac{1}{M} - \tfrac{\Delta}{M-K}}_{M-K \text{ times}}) \prec p \prec (\Delta + \tfrac{1}{M}, \underbrace{\tfrac{1}{M}, \dots, \tfrac{1}{M}}_{L-1 \text{ times}}, R - \tfrac{L}{M}, 0, \dots, 0) \quad (51)$$

where $K = |\{p \ge \frac{1}{M}\}|$ as in (26) and (necessarily) $L = \lfloor MR \rfloor$ (K can possibly be any integer between 1 and L). These majorizations are easily established using characterizations (11), (26) and (37).

Applying s-concavity of entropies H_α or G_ρ to (50) gives closed-form upper bounds of entropies as a function of $\mathbb{P}_e$, known as *Fano inequalities*; and closed-form lower bounds, known as *reverse Fano inequalities*. Figure 1 shows some optimal regions.

The original Fano inequality was an upper bound on conditional entropy $H(X|Y)$ as a function of $\mathbb{P}_e(X|Y)$. It can be shown that upper bounds in the conditional case are unchanged. *Lower* bounds of conditional entropies or α-entropies, however, have to be slightly changed due to the average operation inside the function F (see § 3 above) by taking the convex enveloppe (piecewise linear) of the lower curve on $F^{-1}(H_\alpha)$. In this way, one recovers easily the results of [20] for H, [11] for H_α, and [14,17] for G and G_ρ.

Likewise, applying s-concavity of entropies H_α or G_ρ to (51) gives closed-form upper bounds of entropies as a function of R, similar to *Pinsker inequalities*; and closed-form lower bounds, similar to *reverse Pinsker inequalities*. Figure 2 shows some optimal regions.

The various Pinsker and reverse Pinsker inequalities that can be found in the literature give bounds between $\Delta(p, q)$ and $D(p\|q)$ for general q. Such inequalities find application in Quantum

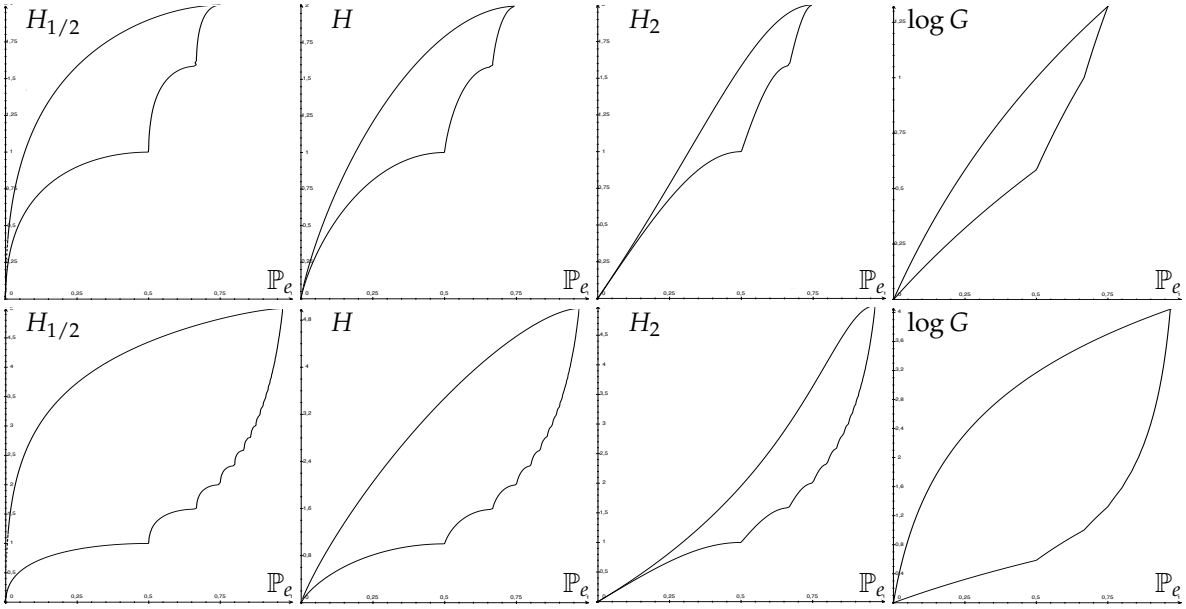


Figure 1. Optimal regions: Entropies (in bits) vs. error probability. Top row $M = 4$; bottom row $M = 32$

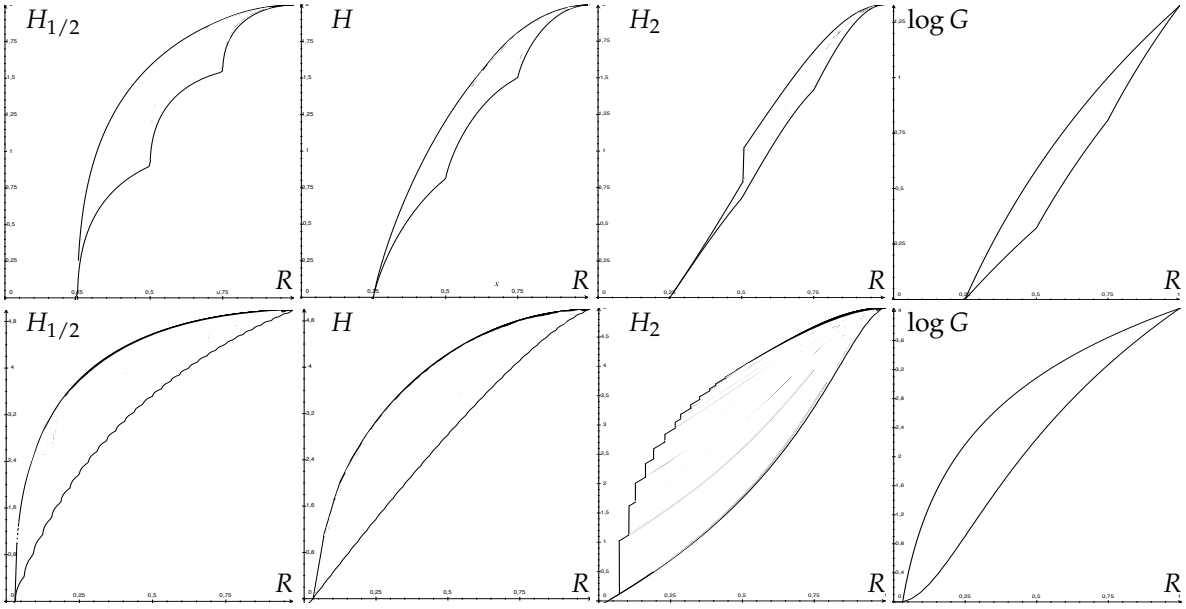


Figure 2. Optimal regions: Entropies (in bits) vs. randomness R. Top row $M = 4$; bottom row $M = 32$

physics [21] and to derive lower bounds on the minimax risk in nonparametric estimation [22]. As they are of more general applicability, they turn out not to be optimal here since we have optimized the bounds in the particular case $q = u$. Using our method, one again recovers easily previous results of [23] and [24, Thm. 26] for H, and improves previous inequalities used for several applications [3,4,6].

6. Conclusion

Using a simple method based on "mixing" or majorization, we have established optimal (Fano-type and Pinsker-type) bounds between entropic quantities (H_α, G_ρ) and statistical quantities ($\mathbb{P}_e$, R) in an interplay between information theory and statistics. As a perspective, similar methodology could be developed for statiscal distance to an arbitrary (not necessarily uniform) distribution.

References

1. Maurer, U.M. A Universal Statistical Test for Random Bit Generators. *J. Cryptology* **1992**, *5*, 89–105.
2. Pliam, J.O. Guesswork and Variation Distance as Measures of Cipher Security. Int. W. Select. Areas Crypt.; Heys, H.; Adams, C., Eds. Springer, 1999, Vol. 1758, *LNCS*, pp. 62–77.
3. Chevalier, C.; Fouque, P.A.; Pointcheval, D.; Zimmer, S. Optimal Randomness Extraction from a Diffie-Hellman Element. Proc. Eurocrypt'09; Joux, A., Ed. Springer, 2009, Vol. 5479, *LNCS*, pp. 572–589.
4. Shoup, V. *A Computational Introduction to Number Theory and Algebra*, 2nd ed.; Cambridge University Press, 2009.
5. Schaub, A.; Boutros, J.J.; Rioul, O. Entropy Estimation of Physically Unclonable Functions via Chow Parameters. Proc. 57th Annual Allerton Conference on Communication, Control, and Computing, 2019.
6. Killmann, W.; Schindler, W. A Proposal for Functionality Classes for Random Number Generators. Ver. 2.0, Anwendungshinweise und Interpretationen zum Schema (AIS) 31 of the Bundesamt für Sicherheit in der Informationstechnik, 2011.
7. Fehr, S.; Berens, S. On the conditional Rényi entropy. *IEEE Transactions on Information Theory* **2014**, *60*, 6801–6810.
8. Arimoto, S. Information measures and capacity of order α for discrete memoryless channels. Proc. Second Colloquium Mathematica Societatis János Bolyai; Csiszár, I.; Elias, P., Eds.; North Holland: Keszthely, Hungary: Bolyai, 1975, 1977; Number 16 in Topics in Information Theory, pp. 41–52.
9. Liu, Y.; Cheng, W.; Guilley, S.; Rioul, O. On conditional alpha-information and its application in side-channel analysis. Proc. 2021 IEEE Information Theory Workshop (ITW2021), 2021.
10. Rioul, O. Variations on a theme by Massey. *IEEE Transactions on Information Theory* **2022**, *68*, 2813–2828.
11. Sason, I.; Verdú, S. Arimoto–Rényi Conditional Entropy and Bayesian M-Ary Hypothesis Testing. *IEEE Transactions on Information Theory* **2018**, *64*, 4–25.
12. Massey, J.L. Guessing and entropy. Proc. of IEEE International Symposium on Information Theory, 1994, p. 204.
13. Arikan, E. An inequality on guessing and its application to sequential decoding. *IEEE Transactions on Information Theory* **1996**, *42*, 99–105.
14. Sason, I.; Verdú, S. Improved Bounds on Lossless Source Coding and Guessing Moments via Rényi Measures. *IEEE Transactions on Information Theory* **2018**, *64*, 4323–4346.
15. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley & Sons, 1st Ed. 1990, 2nd Ed. 2006.
16. van Erven, T.; Harremoës, P. Rényi divergence and Kullback-Leibler divergence. *IEEE Trans. Inf. Theory* **2014**, *60*, 3797–3820.
17. Béguinot, J.; Cheng, W.; Guilley, S.; Rioul, O. Be my guess: Guessing entropy vs. success rate for evaluating side-channel attacks of secure chips. Proc. 25th Euromicro Conference on Digital System Design (DSD 2022); , 2022.
18. Rioul, O. A primer on alpha-information theory with application to leakage in secrecy systems. Proc. 5th conference on Geometric Science of Information (GSI'21). Springer, 2021, Vol. 12829, *Lecture Notes in Computer Science*, pp. 459–467.
19. Marshall, A.W.; Olkin, I.; Arnold, B.C. *Inequalities: Theory of Majorization and Its Applications*, 2nd ed.; Springer Series in Statistics, Springer, 2011.
20. Ho, S.W.; Verdú, S. On the Interplay Between Conditional Entropy and Error Probability. *IEEE Transactions on Information Theory* **2010**, *56*, 5930–5942.
21. Audenaert, K.M.R.; Eisert, J. Continuity Bounds on the Quantum Relative Entropy — II. *J. Math. Phys.* **2011**, *52*, 7.
22. Tsybakov, A.B. *Introduction to Nonparametric Estimation*; Springer Series in Statistics, Springer, 2009.
23. Ho, S.W.; Yeung, R.W. The Interplay Between Entropy and Variational Distance. *IEEE Trans. Inf. Theory* **2010**, *56*, 5906–5929.
24. Sason, I.; Verdú, S. f-Divergence Inequalities. *IEEE Transactions on Information Theory* **2016**, *62*, 5973–6006.