Be my guesses: The interplay between side-channel leakage metrics

Julien Béguinot, Wei Cheng, Sylvain Guilley, Olivier Rioul

Be My Guesses: The Interplay Between Side-Channel Leakage Metrics

Julien Béguinot[a], Wei Cheng[a,b], Sylvain Guilley[b,a], Olivier Rioul[a]

[a] *LTCI, Télécom Paris, Institut Polytechnique de Paris, 19 place Marguerite Perey, Palaiseau, 91120, France*
[b] *Secure-IC S.A.S., 104 Bd du Montparnasse, Paris, 75014, France*

Abstract

In a theoretical context of side-channel attacks, optimal bounds between success rate, guessing entropy and statistical distance are derived with a simple majorization (Schur-concavity) argument. They are further theoretically refined for different versions of the classical Hamming weight leakage model, in particular assuming a priori equiprobable secret keys and additive white Gaussian measurement noise. Closed-form expressions and numerical computation are given. A study of the impact of the choice of the substitution box with respect to side-channel resistance reveals that its nonlinearity tends to homogenize the expressivity of success rate, guessing entropy and statistical distance. The intriguing approximate relation between guessing entropy and success rate $GE = 1/SR$ is observed in the case of 8-bit bytes and low noise. The exact relation between guessing entropy, statistical distance and alphabet size $GE = \frac{M+1}{2} - \frac{M}{2} SD$ for deterministic leakages and equiprobable keys is proved.

Keywords: Side-channel analysis, Guessing entropy, Success rate, Statistical Distance, Schur concavity
PACS: 0000, 1111
2000 MSC: 0000, 1111

1. Introduction

Side-Channel analysis (SCA) is a well-known threat for secure chips in embedded symmetric crypto-systems. They aim at recovering the key, byte by byte in a divide-and-conquer approach, by exploiting the leakage information. The attacker guesses one key byte K from several side-channel observations Y

(modeled as a random vector) knowing the corresponding plain or cipher text bytes $T = t$ and leveraging a (noiseless or noisy) leakage model.

There are two main figures of merit in order to characterize the efficiency of the secrets' recovery: *success rate* SR and *guessing entropy* GE. Roughly speaking, SR is the empirical success probability that the best ranked (most likely) key happens to be the correct one, while GE relates to the number of tries that the attacker has to make before finding the actual secret, thereby estimating the brute force effort to find the correct key by exhaustive search. On one hand, GE is more informative insofar as it depends on the whole key ranking distribution for a given number of leakage traces. On the other hand, SR computation scales easily to the whole multibyte key (the global SR being the byte-wise product of SRs) while GE is much harder to estimate in a multibyte context.

In principle, it is desirable to evaluate *both* SR and GE during the attack because it gives a trade-off between the required number of observations (traces) and the remaining effort for key enumeration. Of course, there is a clear strong correlation between SR and GE: a lower GE will generally mean higher SR and vice versa. This is true not only for a given attack on a given device as the number of traces increases, but also to compare different attacks or different devices endowed with different countermeasures against SCA. In this respect, these metrics are relevant both for the "black hat" attacker or the "white hat" evaluator, and the "blue hat" defender.

Another ubiquitous metric in the cryptographic community is the statistical distance (SD) to the uniform distribution. This quantifies how far a cryptographic object differs from an ideal randomness. While guessing entropy and success rate are explicitly related to the ranking distribution of an attacker, statistical distance lacks some operational interpretation in terms of attack performance. Still, it can be used as a measure of information leakage. In particular, a small statistical distance ensure that no statistical test can distinguish the considered distribution from the uniform distribution. As a consequence, this implies that no attack performs better than random guess.

However, there remains a missing theoretical link between SR, GE and SD that could be exploited to estimate one metric knowing the other. Obviously there is no one-to-one relation between them, but we show that one metric can be lower and upper bounded as a function of the other, which can be optimally determined for a given leakage model. This extended version complements the conference version [1] with results from Rioul [2][3] applied for the statistical

distance that we discuss in the side-channel context.

State-of-the-art. Some previous approaches attempted to bridge the gap by extending the definition of SR to the probability SR_i that the correct key belongs to the list of the first i best key guesses [4]. For instance [5] compares various key enumeration algorithms that allow to estimate SR_i based on the knowledge of the key bytes' likelihoods. In [6] the authors try to link statistical distance, euclidean norm, relative error and average relative error. They derive approximations for Hamming weight leakage models, large number of bits and large noise.

While computing GE can be intractable in practice, [7] heuristically approximates GE by considering "security graphs" summarizing both SR and GE for a given number of traces in the same visual representation.

Chérisey et al. [8] evaluate side channel attacks through SR with inequalities derived from mutual information. They also improve an inequality on GE yet the relation between the two metrics is not investigated.

A very different approach in [9] derives fairly tight mathematical bounds to estimate GE from entropy or Rényi entropy of order 1/2. From a purely theoretical viewpoint, [10] derives optimal bounds in very generic settings for the "guessing moments" with Rényi entropies of various orders. In this respect, considering entropy of infinite order and first order guessing moment yields optimal bounds between SR and GE. In a similar approach, [2][3] derives optimal inequalities for all randomness measures using majorization theory.

Contribution. In this paper, we first present simple and intuitive arguments to derive the optimal bounds between the three metrics SR, GE and SD. Such bounds are all the more tighter as the key space is small. We then refine the relationship in various SCA scenarios and leakage models, providing closed-form expressions for GE in these scenarios. We observe that the bounds are all the more tight as the leakage model is nonlinear (property of an S-Box in a block cipher), which tends to explain why the expressivity of SR and GE gets similar. This accounts for their interchangeable use as an attack working factor in the SCA literature.

Outline. The remainder of this paper is organized as follows. The notions of SR, GE and SD are introduced in Section 2 with emphasis on their similar properties such as data processing inequalities. Section 3 establishes the *Schur-concavity* of GE using majorization theory which allows one to derive simple and intuitive bounds between GE and SR. Further SR and SD are

recalled to be Schur-convex which permits to obtain all optimal relations between the three metrics. Section 4 derives the optimal inequalities between the three metrics. The important cases of Hamming weight leakage model, with an S-Box, and with noise, are mathematically developed in Section 5.

Section 6 concludes the paper.

2. Definitions and Basic Properties

In this section, we define success rate, guessing entropy and statistical distance with emphasis on their similar properties.

Basic Notations. We consider an M-ary secret $K \in \{1, 2, \ldots, M\}$ taking $M = 2^n$ values and some *side-channel observation* Y used to guess the key $\hat{K}$. Observation Y gathers several measurements with known plain or cipher text bytes $T = t$. Since $\hat{K}$ depends on the actual secret key K only through Y, the triple $K - Y - \hat{K}$ forms a Markov chain. The guess $\hat{K}$ is said to be *blind* if it does not depend on the observation Y. For any finite set A, $|A|$ denotes its cardinality.

2.1. Success Rate

Definition 1 (Success Rate (SR))**.** The *success rate* of $\hat{K}$ denoted $\mathbb{P}_s$ is the probability that $\hat{K}$ guesses the secret,

$$\mathbb{P}_s = \mathbb{P}(\hat{K} = K). \tag{1}$$

Theorem 1 (Optimal SR)**.** *The maximal success rate is attained with the MAP rule* $\hat{k}(y) \in \arg\max_k \mathbb{P}(K = k|Y = y)$ *and is given by*

$$\mathbb{P}_s(K|Y) = \mathbb{E}_Y\big(\max_k \mathbb{P}(K = k|Y)\big). \tag{2}$$

In particular, for a blind guess, we write

$$\mathbb{P}_s(K) = \max_k \mathbb{P}(K = k) \geq \frac{1}{M}. \tag{3}$$

Proof. Since $K - Y - \hat{K}$ is a Markov chain, $\mathbb{P}(\hat{K} = \hat{k}|Y, K) = \mathbb{P}(\hat{K} = \hat{k}|Y)$ so that

$$\begin{aligned}
\mathbb{P}_s &= \mathbb{E}_Y\big(\mathbb{P}(\hat{K} = K|Y)\big) && (4)\\
&= \mathbb{E}_Y\big(\textstyle\sum_k \mathbb{P}(K = k|Y)\mathbb{P}(\hat{K} = k|Y)\big) && (5)\\
&\leq \mathbb{E}_Y\big(\max_k \mathbb{P}(K = k|Y)\big) && (6)
\end{aligned}$$

with equality if and only if $\mathbb{P}(\hat{K} = \hat{k}|Y) = 1$ for some $\hat{k} \in \arg\max_k \mathbb{P}(K = k|Y)$.

Theorem 2 (Data Processing Inequality for $\mathbb{P}_s$)**.** *One has*

$$\mathbb{P}_s(K) \leq \mathbb{P}_s(K|Y) \tag{7}$$

(observing side channel information always increases success). More generally, if $K - Y - Z$ is a Markov chain, then

$$\mathbb{P}_s(K|Z) \leq \mathbb{P}_s(K|Y) \tag{8}$$

(data processing can only reduce success).

Proof. Since $\mathbb{P}(K = k|Y) \leq \max_k \mathbb{P}(K = k|Y)$, taking the expectation over Y gives $\mathbb{E}_Y \mathbb{P}(K = k|Y) \leq \mathbb{E}_Y \max_k \mathbb{P}(K = k|Y)$ for every k, hence

$$\max_k \mathbb{E}_Y \mathbb{P}(K = k|Y) \leq \mathbb{E}_Y \max_k \mathbb{P}(K = k|Y) \tag{9}$$

which is (7). This in turn implies $\mathbb{P}_s(K|Z) \leq \mathbb{P}_s(K|Y, Z)$ by considering each fixed value $Z = z$ and taking the expectation over Z. Finally, $\mathbb{P}_s(K|Y, Z) = \mathbb{P}_s(K|Y)$ because $K|Y, Z$ is distributed as $K|Y$ since $K - Y - Z$ is a Markov chain.

2.2. Guessing Entropy

In a guessing problem, key candidates are guessed one by one in a sequence $(1), (2), \ldots, (M)$. Such a sequence is a permutation of $\{1, 2, \ldots, M\}$ where (i) denotes the ith ranked key for $i = 1, 2, \ldots, M$. Thus, first (1) is guessed, then (2), etc. The number of key guesses before the actual secret $K = (I)$ is found is I, a random variable which depends upon the observation Y. Hence, $K - Y - I$ forms a Markov Chain.

Definition 2 (Guessing Entropy (GE))**.** The guessing entropy is the average number of guesses:

$$G = \mathbb{E}_{K,Y}(I) \tag{10}$$

Notice that some previous works define GE as I itself [9, 11].

Let $p_{(i)|y} = \mathbb{P}(K = (i)|Y = y)$ be the probability of the ith ranked key given observation $Y = y$.

Theorem 3 (Optimal GE)**.** *The minimal guessing entropy is attained with the ranking rule*

$$p_{(1)|y} \geq p_{(2)|y} \geq \cdots \geq p_{(M)|y} \tag{11}$$

and is given by

$$G(K|Y) = \mathbb{E}_Y\left(\sum_{k=1}^{M} k\, p_{(k)|Y}\right). \tag{12}$$

In particular, for a blind guess, this reduces to $G(K) = \sum_{k=1}^{M} k p_{(k)}$*, where the* $p_{(k)} = \mathbb{P}(K = (k))$ *are in descending order.*

Often $G(K)$ is simply referred to as the guessing entropy of K while $G(K|Y)$ is known as the *conditional guessing entropy* of K given Y.

Proof. By the law of total expectation,

$$G = \mathbb{E}_Y\, \mathbb{E}_K(I|Y) = \mathbb{E}_Y\left(\sum_{i=1}^{M} i \cdot \mathbb{P}(K = (i)|Y)\right). \tag{13}$$

By the rearrangement inequality [12, Thm. 368], since (i) is an increasing sequence, the minimum G is obtained when the probabilities $\mathbb{P}(K = (i)|Y)$ are in descending order.

Theorem 4 (Data Processing Inequality)**.** *One has*

$$G(K) \geq G(K|Y) \tag{14}$$

(observing side channel information improves guessing).

More generally, if $K - Y - Z$ *is a Markov chain, then*

$$G(K|Z) \geq G(K|Y) \tag{15}$$

(data processing can only worsen guessing).

Proof. Without loss of generality assume that K's probability distribution is in descending order $p_1 \geq p_2 \geq \cdots \geq p_M$ so that $I = K$ and $G(K) = \mathbb{E}(K)$. Then by definition of minimum guessing, $G(K|Y = y) \leq \mathbb{E}(K|Y = y)$. Taking the expectation over Y gives $G(K|Y) \leq \mathbb{E}_Y\, \mathbb{E}(K|Y) = \mathbb{E}(K) = G(K)$ by the law of total expectation. This proves (14). This in turn implies $G(K|Z) \geq G(K|Y, Z)$ by considering each fixed value of $Z = z$ and taking the expectation over Z. Finally, $G(K|Y, Z) = G(K|Y)$ because $K|Y, Z$ is distributed as $K|Y$ since $K - Y - Z$ is a Markov chain.

2.3. Statistical Distance to the Uniform

Definition 3 (Distinguishability)**.** Let A be an event. The distinguishability of the random variable K from the uniform random variable U under event A is defined as

$$\Delta_A(K) = |\mathbb{P}(K \in A) - \mathbb{P}(U \in A)|. \tag{16}$$

If $\Delta_A(K)$ is significantly large then K can be distinguished from the uniform distribution. In the following we consider the optimal distinguishability.

Theorem 5 (Optimal Distinguishability)**.** *The optimal distinguishability corresponds to the statistical distance (SD) to the uniform random variable i.e.*

$$\begin{aligned} \Delta(K) = \max_A \Delta_A(K) &= \frac{1}{2}\sum_k |\mathbb{P}(K=k) - \frac{1}{M}| \\ &= \sum_k \left(\mathbb{P}(K=k) - \frac{1}{M}\right)^+ \leq 1 \end{aligned} \tag{17}$$

where $(x)^+ = \max(0, x)$ *is the positive part function. In the conditional case we write the average optimal distinguishability*

$$\Delta(K|Y) = \mathbb{E}_Y[\Delta(K|Y=y)]. \tag{18}$$

Proof. The expression with the positive part is direct from Definition 3. Equality with (17) is well known, we recall a simple proof for completeness. Let $A^+ = \{k | p(k) \geq \frac{1}{M}\}$. Then $\sum_k \left(\mathbb{P}(K=k) - \frac{1}{M}\right)^+ = \mathbb{P}(K \in A^+) - \frac{|A^+|}{M} = (1 - \mathbb{P}(K \notin A^+)) - (1 - \frac{M-|A^+|}{M}) = \frac{M-|A^+|}{M} - \mathbb{P}(K \notin A^+) = \sum_k \left(\frac{1}{M} - \mathbb{P}(K=k)\right)^+$. This concludes the proof since $x^+ + (-x)^+ = |x|$.

Duc et al. [13] uses the statistical distance to the uniform that we term distinguishability as metric to measure the security of implementations against side channel analysis. Sometimes the distinguishability is referred to as *total variation* in the blind guess setting (no conditioning) and *statistical distance* with side-channel information (conditional version).

This notion is relevant in the cryptographic context. A small statistical distance means that the random variable is indistinguishable from the uniform random variable. As is clear from its definition, the probability of success minus the probability of success of a random guess in a statistical test is upper bounded by the distinguishability. Hence a small distinguishability implies

a probability of success close to a random guess. This notion is related to the notion of distinguishing advantage of an adversary in a cryptographic context.

Theorem 6 (Data Processing Inequality for SD)**.**

$$\Delta(K|Y) \geq \Delta(K) \tag{19}$$

(observing side channel information increases distinguishability).

If $K - Y - Z$ forms a Markov Chain

$$\Delta(K|Y) \geq \Delta(K|Z). \tag{20}$$

(data processing can only decrease distinguishability)

Proof.

The proof rely on the convexity of the absolute value combined with Jensen's inequality.

$$\Delta(K|Y) = \mathbb{E}_Y[\frac{1}{2}\sum_k |p(k|Y) - \frac{1}{M}|] \tag{21}$$

$$= \frac{1}{2}\sum_k \mathbb{E}_Y[|p(k|Y) - \frac{1}{M}|] \tag{22}$$

$$\geq \frac{1}{2}\sum_k |\,\mathbb{E}_Y[p(k|Y) - \frac{1}{M}]| \tag{23}$$

$$= \frac{1}{2}\sum_k |p(k) - \frac{1}{M}| \tag{24}$$

$$= \Delta(K). \tag{25}$$

This in turn implies $\Delta(K|Z) \leq \Delta(K|Y,Z)$ by considering each fixed value of $Z = z$ and taking the expectation over Z. Finally, $\Delta(K|Y,Z) = \Delta(K|Y)$ because $K|Y,Z$ is distributed as $K|Y$ since $K - Y - Z$ is a Markov chain.

More generally [2] unified Thms. 2, 4, 6 by showing that all "randomness measures" verify a data processing inequality where being a "randomness measure" essentially means being a Schur-concave function.

3. Schur Properties

3.1. Key Concepts of Majorization Theory

We first introduce some notations for the theory of majorization [14]. Hereafter we let $p_{(1)}, p_{(2)}, \ldots, p_{(M)}$ denote the vector $p = (p_1, p_2, \ldots, p_M)$ of non-negative elements arranged in *descending* order $p_{(1)} \geq p_{(2)} \geq \cdots \geq p_{(M)}$. We also use the *cumulative sum* notation

$$P_{(k)} = p_{(1)} + p_{(2)} + \cdots + p_{(k)} \qquad (k = 1, \ldots, M) \tag{26}$$

with the convention $P_{(0)} = 0$.

Definition 4 (Majorization)**.** We say that q *majorizes* p, and we write $p \preceq q$ if

$$P_{(k)} \leq Q_{(k)} \qquad (k = 1, \ldots, M-1) \tag{27}$$

and $P_{(M)} = Q_{(M)}$. (Notice that this latter condition is always satisfied when p and q are probability distributions since $P_{(M)} = \sum_k p_k = 1$ and $Q_{(M)} = \sum_k q_k = 1$.)

The intuition behind majorization is that $p \preceq q$ means that p is more "spread out" than q. Thus in the case of a probability distribution p, the minimum spread is for a deterministic (but Dirac) distribution and the maximum spread is for a uniform distribution. Indeed, it is easily checked that

$$(\tfrac{1}{M}, \tfrac{1}{M}, \ldots, \tfrac{1}{M}) \preceq p \preceq (1, 0, 0, \ldots, 0) \tag{28}$$

for any probability distribution $p = (p_1, p_2, \ldots, p_M)$. More generally [14],

$$(\tfrac{P_{(M)}}{M}, \tfrac{P_{(M)}}{M}, \ldots, \tfrac{P_{(M)}}{M}) \preceq p \preceq (P_{(M)}, 0, 0, \ldots, 0) \tag{29}$$

for any vector $p = (p_1, p_2, \ldots, p_M)$.

Definition 5 (Schur-Concavity & Convexity)**.** A function G is *Schur-concave* if $p \preceq q \implies G(p) \geq G(q)$. Similarly, G is *Schur-convex* if $p \preceq q \implies G(p) \leq G(q)$.

In other words, a Schur-concave function is large for "spread out" distributions and small for "condensed" distributions, while inversely for a Schur-convex function.

3.2. Guessing Entropy is Schur-Concave, Probability of Success and Distinguishability are Schur-Convex

It is well known that entropy [14], and more generally the Rényi entropy of any order [15] (e.g., min-entropy, collision entropy, etc.) is Schur-concave. Perhaps lesser known is that guessing entropy is Schur-concave:

Theorem 7 (Schur-Concavity)**.** *One has*

- *Guessing entropy* $G(K) = \sum_{k=1}^{M} kp_{(k)}$ *is Schur-concave in* p*.*
- *Probability of Success* $\mathbb{P}_s(K)$ *is Schur-convex in* p*.*
- *Distinguishability* $\Delta(K)$ *is Schur-convex in* p*.*

Proof. Using *summation by parts*,

$$\sum_{k=1}^{M} kp_{(k)} = \sum_{k=1}^{M} k(P_{(k)} - P_{(k-1)}) \tag{30}$$

$$= MP_{(M)} - P_{(0)} + \sum_{k=1}^{M-1} \big(k - (k+1)\big)P_{(k)} \tag{31}$$

$$= M - P_{(1)} - P_{(2)} - \cdots - P_{(M-1)}. \tag{32}$$

The Schur-concavity of $G(K)$ is now obvious from the definitions. The Schur-convexity of the probability of success is immediate from the definitions. Let $p \preceq q$. Let t be the largest index such that $p_{(t)} \geq \frac{1}{M}$. Then the distinguishability for p is $P_{(t)} - \frac{t}{M}$. Moreover the distinguishability for q is at least $Q_{(t)} - \frac{t}{M}$. The Schur-convexity follows from the definition as $Q_{(t)} \geq P_{(t)}$.

Remark 1. Recent works on guessing such as [16] state Schur-concavity of Rényi entropy but do not mention the same property for GE. During the review process we became aware that the Schur-concavity of GE was observed earlier by Khouzani and Malacaria [17] among many other types of entropies. They established Schur-concavity by stating (without proof) that $G(K)$ is symmetric and concave in the probability distribution of K. While symmetry is obvious here, concavity of GE is precisely established by inequality (14) above. The Schur-convexity of Δ is also shown in [2][3] by stating it is symmetric and concave.

Remark 2. The proof of this Theorem carries over verbatim for any function of the form $\sum_{k=1}^{M} \alpha_k p_{(k)}$ where (α_k) is an increasing sequence. In particular for guessing moments [18]:

Corollary 1 (Schur-Concavity of Guessing Moments)**.** $G_\rho(K) = \sum_{k=1}^{M} k^\rho p_{(k)}$ *is Schur-concave in* p*.*

These results are in line with the known inequalities between guessing entropy (or guessing moments) and entropy (or Rényi entropies) as established in [18, 19].

Remark 3. Since guessing entropy is Schur-concave, it follows from (28) that guessing entropy is minimized for the deterministic distribution and maximized for the uniform distribution, which gives the trivial bounds $1 \leq G(K) \leq \frac{M+1}{2}$.

4. Optimal Bound Derivation

In this section, we present the optimal bounds for SR, SD and GE explicitly. The results are recapped in Fig. 1, which depicts the overall connections among SR, SD and GE.

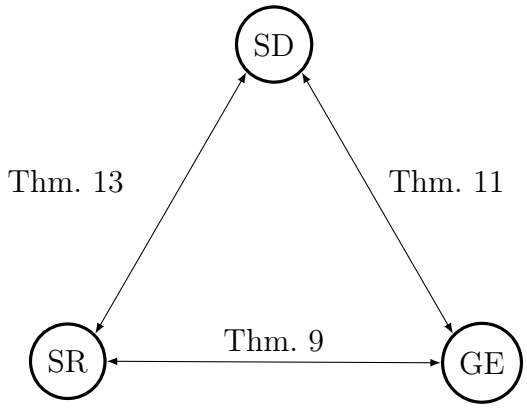


Figure 1: Relations Presented in this Article

4.1. Optimal Bounds between GE and SR

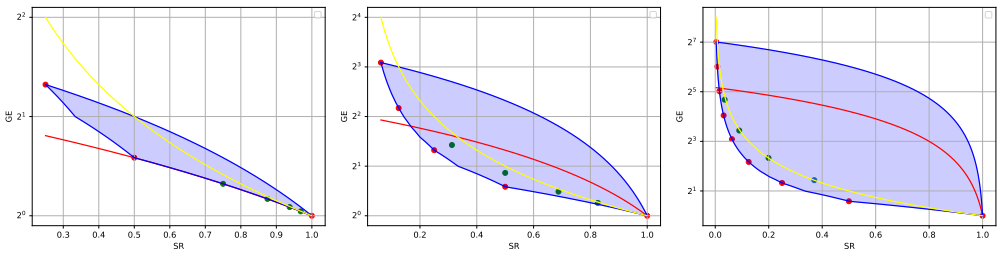
Theorem 8 (Optimal Lower and Upper Bounds for Blind Guess)**.** *For a fixed success rate* $\mathbb{P}_s(K)$*, the optimal lower and upper bound on guessing entropy* $G(K)$ *are*

$$\big(1 + \lfloor \tfrac{1}{\mathbb{P}_s(K)} \rfloor\big)\big(1 - \tfrac{1}{2}\lfloor \tfrac{1}{\mathbb{P}_s(K)} \rfloor \mathbb{P}_s(K)\big) \leq G(K) \leq 1 + \frac{M}{2}(1 - \mathbb{P}_s(K)). \quad (33)$$

Proof. From Theorem 7, for a fixed $p_{(1)}$, $G(K) - \mathbb{P}_s(K) = \sum_{k=2}^{M} k p_{(k)}$ is Schur-concave in $(p_{(2)}, \ldots, p_{(M)})$. It follows that this quantity is maximum for the uniform distribution $(p_{(2)}, \ldots, p_{(M)}) = (\frac{1-\mathbb{P}_s}{M-1}, \frac{1-\mathbb{P}_s}{M-1}, \ldots, \frac{1-\mathbb{P}_s}{M-1})$ and minimum for the least spread out distribution $(p_{(2)}, \ldots, p_{(M)})$ with $p_{(k)} \leq \mathbb{P}_s$. It is easily seen that the latter (least spread out) distribution is of the form $(p_{(2)}, \ldots, p_{(M)}) = (\mathbb{P}_s, \ldots, \mathbb{P}_s, x, 0, \ldots, 0)$ where $x < \mathbb{P}_s$ is such that $\sum_{k=2}^{M} p_{(k)} = 1$, that is, $x = 1 - \lfloor 1/\mathbb{P}_s \rfloor \mathbb{P}_s$. Plugging these values of $(p_{(1)}, p_{(2)}, \ldots, p_{(M)})$ into the expression of the guessing entropy gives the announced lower and upper bounds.

Fig. 2 illustrates the corresponding optimal regions (in blue) between $\mathbb{P}_s$ and G for $M = 2^n$ with $n = 2, 4, 8$, respectively.

Remark 4. If X is a geometric random variable with parameter $p = \mathbb{P}_s$ defined over $\mathbb{N}$ then the guessing entropy of X is exactly the inverse of the optimal probability of guessing X. This suggests that if a random variable is well approximated by a geometric random variable then the approximation that the guessing entropy is the reciprocal of the probability of success holds.



(a) Regions for $M = 2^2$ (b) Regions for $M = 2^4$ (c) Regions for $M = 2^8$

Figure 2: Regions $G(K|Y)$ vs. $\mathbb{P}_s(K|Y)$ as given by Theorem 9. The red curve is the improved upper bound (49) for the deterministic Hamming weight model. The four green dots are the exact values computed for $Q \in \{1, 2, 3, 4\}$ traces. The yellow curve corresponds to the formula $G = \mathbb{P}_s^{-1}$ and seems to approximate well the actual relation for $n = 8$ bits.

Theorem 9 (Bounds with Side-Channel Information)**.**

$$(1 + \lfloor \tfrac{1}{\mathbb{P}_s(K|Y)} \rfloor)(1 - \lfloor \tfrac{1}{\mathbb{P}_s(K|Y)} \rfloor \frac{\mathbb{P}_s(K|Y)}{2}) \leq G(K|Y) \leq 1 + \frac{M}{2}(1 - \mathbb{P}_s(K|Y)). \tag{34}$$

Proof. Applying Theorem 8 to the random variable $K|Y = y$ for every value y gives $(1 + \lfloor \frac{1}{\mathbb{P}_s(K|Y=y)} \rfloor)(1 - \lfloor \frac{1}{\mathbb{P}_s(K|Y=y)} \rfloor \frac{\mathbb{P}_s(K|Y=y)}{2}) \leq G(K|Y = y) \leq$

$1+\frac{M_y}{2}(1-\mathbb{P}_s(K|Y=y))$ where $M_y \leq M$ is the number of possible keys given $Y=y$. Taking the expectation over Y we obtain lower and upper bounds on $G(K|Y)=\mathbb{E}_y\, G(K|Y=y)$. By Theorem 1, $\mathbb{P}_s(K|Y)=\mathbb{E}_y\,\mathbb{P}_s(K|Y=y)$, we obtain the announced upper bound $G(K|Y)\leq 1+\frac{M}{2}(1-\mathbb{P}_s(K|Y))$.

The lower bound, of the form $\phi(p)=(1+\lfloor\frac{1}{p}\rfloor)(1-\lfloor\frac{1}{p}\rfloor\frac{p}{2})$, is piecewise linear and *convex* in $p=\mathbb{P}_s$. Indeed, its value at $p=\frac{1}{k}$ for positive integer k is $(1+k)(1-\frac{k}{2k})=\frac{1+k}{2}$, hence its successive slopes between $p=\frac{1}{k-1}$ and $p=\frac{1}{k}$ are $\frac{1/2}{\frac{1}{k}-\frac{1}{k-1}}=-\frac{k(k-1)}{2}$, which is increasing as $p=\frac{1}{k}$ increases. Thus, by Jensen's inequality, we have $\mathbb{E}_y[\phi(\mathbb{P}_s(K|Y=y))]\geq\phi(\mathbb{E}_y[\mathbb{P}_s(K|Y=y)])=\phi(\mathbb{P}_s(K|Y))$, which gives the announced lower bound.

At high noise when $\frac{1}{M}\leq\mathbb{P}_s(K|Y)\leq\frac{1}{M-1}$ this simplifies to

$$\tfrac{M+1}{2}-\tfrac{M(M-1)}{2}(\mathbb{P}_s(K|Y)-\tfrac{1}{M})\leq G(K|Y)\leq\tfrac{M+1}{2}-\tfrac{M}{2}(\mathbb{P}_s(K|Y)-\tfrac{1}{M}). \quad (35)$$

At low noise when $\mathbb{P}_s(K|Y)=1-\epsilon$ where $\epsilon\leq\frac{1}{M}$, this simplifies to

$$1+\epsilon\leq G(K|Y)\leq 1+\tfrac{M}{2}\epsilon. \quad (36)$$

Remark 5. It is immediate from its proof that a refinement of the upper bound of Theorem 9 is given by

$$G(K|Y)\leq 1+\frac{\max_y M_y}{2}(1-\mathbb{P}_s(K|Y)). \quad (37)$$

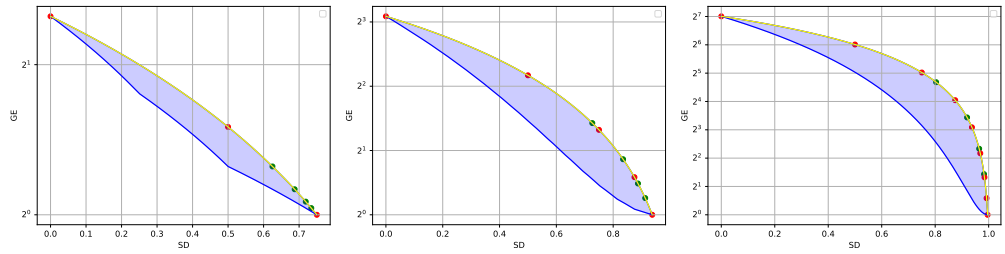
This is particularly interesting for deterministic (noiseless) leakage since, as shown in the next Section, M_y decreases rapidly as the number of traces increases.

4.2. Optimal Bounds on GE for a Given SD

Theorem 10 (Optimal Inequalities between GE and SD for a Blind Guess)**.** *Let $U(K)=M(1-\Delta(K))$, then for a blind guess,*

$$1+\lfloor U(K)\rfloor\frac{2U(K)-\lfloor U(K)\rfloor-1}{2M}\leq G(K)\leq\frac{1+U(K)}{2}. \quad (38)$$

Proof. This is proved using majorization in [2][3]. This corresponds to optimal Pinsker and reverse Pinsker inequalities.



(a) Regions for $M = 2^2$ (b) Regions for $M = 2^4$ (c) Regions for $M = 2^8$

Figure 3: Regions $G(K|Y)$ vs. $\Delta(K|Y)$ as given by Theorem 11. The 4 green dots are the exact values computed for $Q = 1$, 2, 3, and 4 traces for deterministic Hamming weight model. The red dots corresponds to subset of bits revealed. The yellow curve corresponds to the formula $G = \frac{M+1}{2} - \frac{M\Delta}{2}$ and is a one-to-one relationship for deterministic leakage models and uniform secret.

Theorem 11 (Optimal Inequalities between GE and SD with Side-Channel Information)**.** *Let* $U(K|Y) = M(1 - \Delta(K|Y))$*, then with Side-Channel-Information* Y*,*

$$1 + \lfloor U(K|Y) \rfloor \frac{2U(K|Y) - \lfloor U(K|Y) \rfloor - 1}{2M} \leq G(K|Y) \leq \frac{1 + U(K|Y)}{2}. \quad (39)$$

Proof. This is proved in [2][3]. The upper is linear and the lower bound is convex, hence Jensen's inequality can be applied.

We are often interested in the behavior of the metrics in the high noise scenario. Hence we explicit the bound when $\Delta(K|Y) < \frac{1}{M}$,

$$\frac{M+1}{2} - (M-1)\Delta(K|Y) \leq G(K|Y) \leq \frac{M+1}{2} - \frac{M\Delta(K|Y)}{2}. \quad (40)$$

As expected, we obtain a bound with a constant term $\frac{M+1}{2}$ corresponding to a blind guess. Then a linear term in the statistical distance is subtracted from it. Another interesting case is the noiseless scenario where $\Delta = 1 - \frac{1}{M} - \epsilon$ where $\epsilon \leq \frac{1}{M}$,

$$1 + \epsilon \leq G(K|Y) \leq 1 + \frac{M}{2}\epsilon. \quad (41)$$

As expected we obtain a constant term 1 corresponding to a perfect guess. Then a linear term in ϵ is added from it.

Fig. 3 illustrates the corresponding optimal regions (in blue) between Δ and G for $M = 2^n$ with $n = 2, 4, 8$, respectively.

4.3. Optimal Bounds between SD and SR

Finally, we present the optimal Fano and reverse-Fano inequalities in between the probability of success and statistical distance. Equivalently we also use optimal Pinsker and reverse Pinsker inequalities to obtain the optimal regions.

Theorem 12 (Optimal Relation between SD and SR for a Blind Guess)**.** *For a blind guess,*

$$\Delta(K) + \frac{1}{M} \geq \mathbb{P}_s(K) \geq \frac{1}{M} + \frac{\Delta}{\lfloor M(1-\Delta(K)) \rfloor} \tag{42}$$

or equivalently

$$\begin{aligned} \mathbb{P}_s(K) - \tfrac{1}{M} \leq \Delta(K) \leq \frac{1}{2} \big(\tfrac{M-1}{M} + (\mathbb{P}_s(K) - \tfrac{2}{M}) \lfloor \mathbb{P}_s(K)^{-1} \rfloor \\ + |1 - \mathbb{P}_s(K) \lfloor \mathbb{P}_s(K)^{-1} \rfloor - \tfrac{1}{M}| \big) . \end{aligned} \tag{43}$$

Proof. This is also an application of majorization theory as shown in [2][3]. The application of optimal Pinsker and reverse Pinsker inequalities yields (42) while the application of optimal Fano and reverse Fano inequalities yields (43).

Notice that (42) (43). are equivalent because they correspond to the same regions which are optimally characterized. This equivalence would have been difficult to derive directly.

In the high noise scenario where $\frac{1}{M} \leq \mathbb{P}_s \leq \frac{1}{M-1}$ it simplifies to,

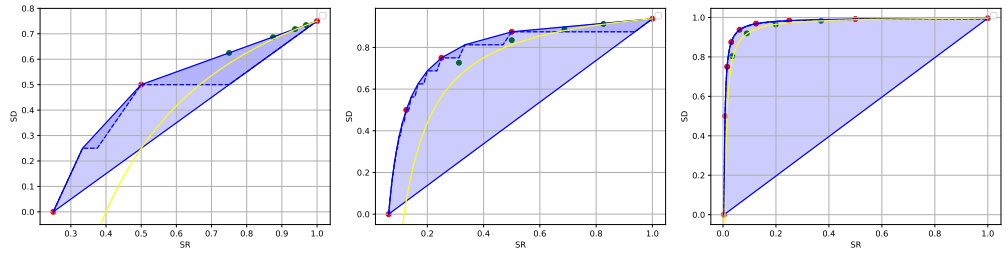
$$\mathbb{P}_s(K) - \frac{1}{M} \leq \Delta(K) \leq (M-1)(\mathbb{P}_s(K) - \frac{1}{M}). \tag{44}$$

In the noiseless scenario where $\mathbb{P}_s(K|Y) = 1 - \epsilon(K)$ with $\epsilon(K) \leq \frac{1}{M}$ it boils down to

$$\frac{M-1}{M} - \epsilon \leq \Delta(K) \leq \frac{M-1}{M} - \frac{2}{M}\epsilon(K). \tag{45}$$

Theorem 13 (Optimal Relation between SD and SR with Side-channel Information)**.** *Let $U(K|Y) = M(1-\Delta(K|Y))$, then with side-channel information Y,*

$$\mathbb{P}_s(K|Y) \geq 1 - \frac{U(K|Y) - 1 + \lfloor U(K|Y) \rfloor \lfloor U(K|Y) - 1 \rfloor}{\lfloor U(K|Y) \rfloor \lfloor U(K|Y) + 1 \rfloor} \tag{46}$$



(a) Regions for $M = 2^2$ (b) Regions for $M = 2^4$ (c) Regions for $M = 2^8$

Figure 4: Regions $\Delta(K|Y)$ vs. $\mathbb{P}_s(K|Y)$ as given by Theorem 13 and Theorem 12. The dashed blue line corresponds to the blind guess setting. With side-channel information it is convexified with a region in darker blue. The 4 green dots are the exact values computed for $Q = 1$, 2, 3, and 4 traces. The red dots corresponds to revealing a subsets of bits of the secret key. The yellow curve corresponds to the formula $\mathbb{P}_s^{-1} \approx \frac{M+1}{2} - \frac{M\Delta}{2}$ and seems to approximate well the actual relation for $n = 8$ bits.

or equivalently,

$$\Delta(K|Y) \leq \tfrac{M-1}{M} - (\lfloor \mathbb{P}_s(K|Y)^{-1} + 1 \rfloor \mathbb{P}_s(K|Y) - 1)\lfloor \mathbb{P}_s(K|Y)^{-1} \rfloor \tfrac{\lfloor \mathbb{P}_s(K|Y)^{-1} \rfloor - 1}{M} - (1 - \lfloor \mathbb{P}_s(K|Y)^{-1} \rfloor \mathbb{P}_s(K|Y))\lfloor \mathbb{P}_s(K|Y)^{-1} + 1 \rfloor \tfrac{\lfloor \mathbb{P}_s(K|Y)^{-1} \rfloor}{M}. \quad (47)$$

Proof. The convex envelope of (42) yields (46). The concave envelope of (43) yields (47).

Once again we know by construction that (46) and (47) are equivalent which is not obvious from the formulas.

Fig. 4 illustrates the corresponding optimal regions (in blue) between $\mathbb{P}_s$ and Δ for $M = 2^n$ with $n = 2, 4, 8$, respectively.

5. Refined Bounds for Hamming Weight Leakage Model

5.1. Deterministic Leakage for One Observed Trace

A well-known leakage model of an embedded cryptographic device in a noiseless scenario is the *Hamming weight model*

$$Y = w_H(K \oplus t) \quad (48)$$

where w_H is the bitwise Hamming weight operator [20], $\oplus$ denotes the XOR operation and $T = t$ is given value of plain or cipher text. Let $\mathcal{Y} = \{0, 1, \ldots, n\}$ be the set of all values taken by Y and $\mathcal{K}_y$ be the set of key values k for fixed $Y = y$.

Theorem 14. *For the Hamming weight model, the region* (34) *reduces (improves) to the following values of SR and GE:*

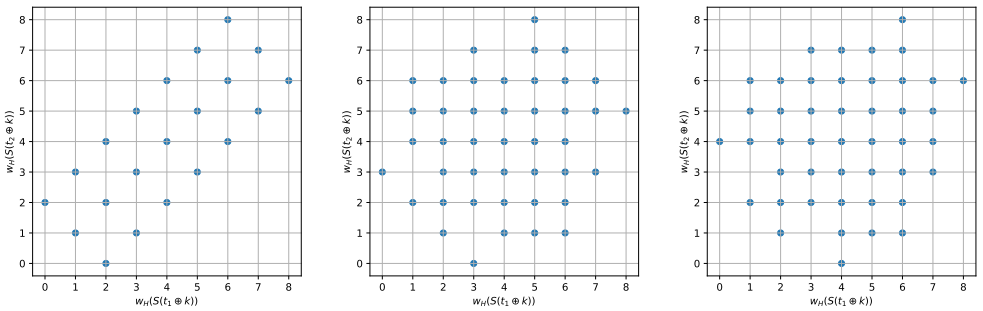
$$G(K|Y) \leq 1 + \frac{1}{2}\binom{n}{\lfloor \frac{n+1}{2} \rfloor}(1 - \mathbb{P}_s(K|Y)) \tag{49}$$

$$\mathbb{P}_s(K|Y) \geq \binom{n}{\lfloor \frac{n+1}{2} \rfloor}^{-1}. \tag{50}$$

Proof. For observed $Y = y$, $M_y = |\mathcal{K}_y|$ is the number of n-bit vectors having Hamming weight y, that is, $M_y = \binom{n}{y}$ in the improved bound (37). Since $\max_y M_y = \binom{n}{\lfloor \frac{n+1}{2} \rfloor}$, this gives (49).

Since $K|Y = y$ has M_y possible values, $\mathbb{P}_s(K|Y = y) = \max_k \mathbb{P}(K = k|Y = y) \geq \frac{1}{M_y} \geq 1/\binom{n}{\lfloor \frac{n+1}{2} \rfloor}$. Averaging over Y gives (50). Equality holds if and only if K is uniformly distributed over the largest class $\mathcal{K}_y$.

Figure 2 illustrates the improvement for $n = 2$, 4, and 8 bits, where the red curves correspond to the reduced upper bound (49). It can be observed that the case of equality in (50) corresponds to the points where the upper bound (49) (red curve) and the lower bound in (34) (blue curve) meet. In particular for $M = 2^2$ our improved upper bound coincide with the lower bound. This proves that in this case the SR and GE are in one to one correspondence with a Hamming Weight leakage model.

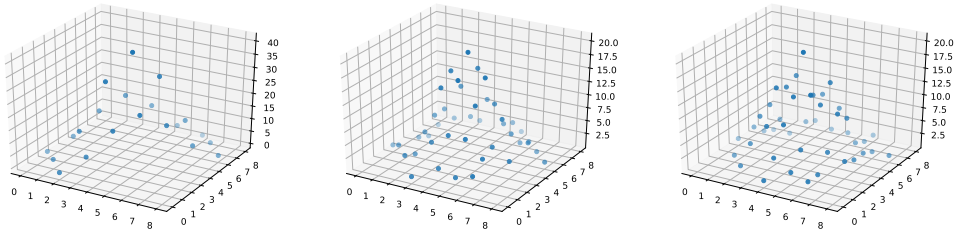


(a) S_1 (linear) (b) S_7 (c) S_{254} (AES—nonlinear)

Figure 5: Sets $\mathcal{Y}$ of deterministic Hamming weight leakage values for $t_1 = 0$ and $t_2 = 3 =$ (00000011) for different S-Boxes.

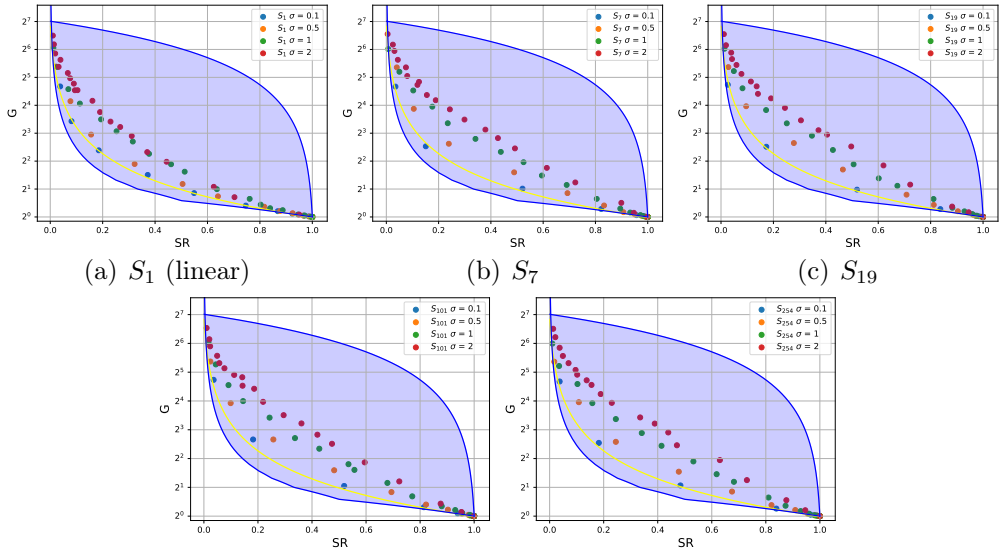
5.2. Case of Equiprobable Keys

A usual assumption is that K is a priori uniformly distributed over M values. In this case the following exact formulas hold. Similar formulas for SR and GE can be found in [21].



(a) S_1 (linear) (b) S_7 (c) S_{254} (AES—nonlinear)

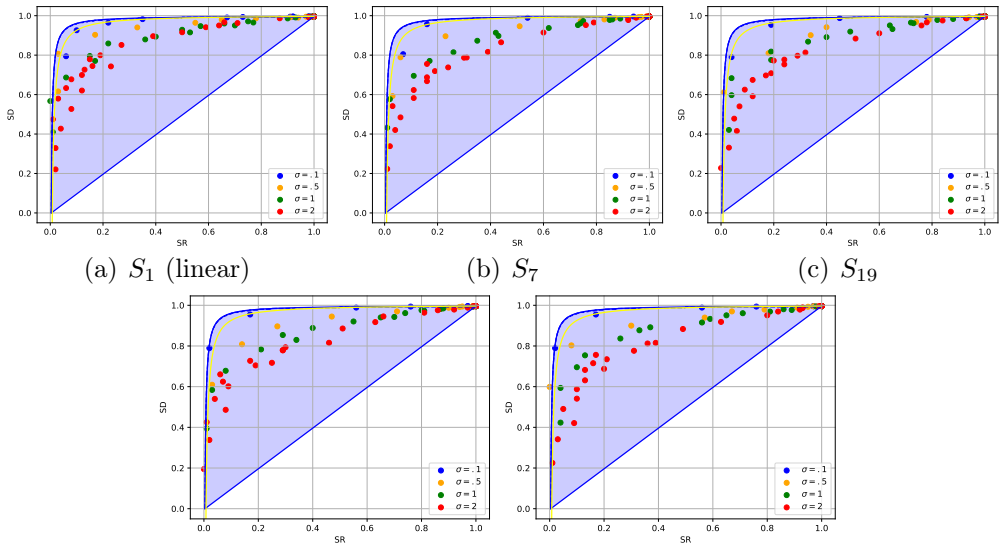
Figure 6: Number of keys M_y for given $Y = y$ for $t_1 = 0$ and $t_2 = 3$ for different S-Boxes. The x, y-axes represent the two coordinates of the 2-dimensional leakage $Y = y = (y_1, y_2)$. The z-axis corresponds to the number M_y of possible keys given $Y = y$, which tends to decrease as the nonlinearity of the S-Box increases. In particular, $\max_y M_y$ is respectively 40, 20, 20 thereby improving the bound (37) for nonlinear S-Boxes.



(a) S_1 (linear) (b) S_7 (c) S_{19}

(d) S_{101} (e) S_{254} (AES – nonlinear)

Figure 7: Numerical evaluation of SR and GE for various noise levels σ^2 and increasing number of traces, for various choices of S-Boxes. Each different subfigure corresponds to a choice for the S-Box. The yellow curve corresponds to GE$\approx$SR^{-1}, indicating at least for low noise, the GE is approximately the reciprocal of the SR. It can be observed that at a fixed SR the GE increases with the noise. This effect is amplified in the presence of a nonlinear S-Box.



(a) S_1 (linear) (b) S_7 (c) S_{19}

(d) S_{101} (e) S_{254} (AES – nonlinear)

Figure 8: Numerical evaluation of SR and SD for various noise levels σ^2 and increasing number of traces, for various choices of S-Boxes. Each different subfigure corresponds to a choice for the S-Box. The yellow curve corresponds to $1/\mathrm{SR} \approx \frac{M+1}{2} - \frac{M}{2}\mathrm{SD}$, indicating at least for low noise the approximation holds. It can be observed that at a fixed SR the SD decreases with the noise.

Theorem 15 (Exact Formulas of Equiprobable Keys)**.**

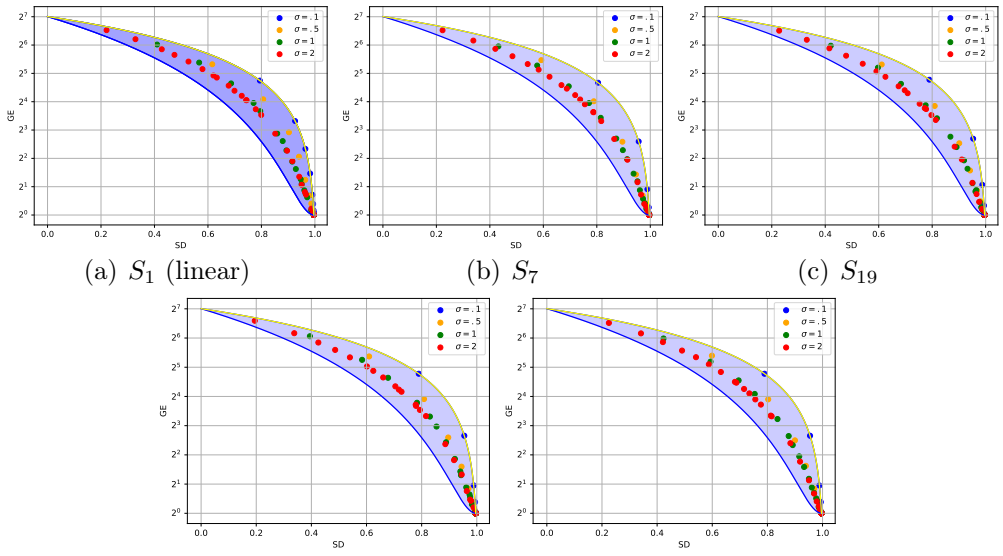
$$\begin{aligned} \mathbb{P}_s(K|Y) &= \frac{|\mathcal{Y}|}{M} \\ G(K|Y) &= \frac{1}{2} + \frac{1}{2M}\sum_{y\in\mathcal{Y}} M_y^2, \\ \Delta(K|Y) &= 1 - \frac{1}{M^2}\sum_{y\in\mathcal{Y}} M_y^2. \end{aligned} \tag{51}$$

More generally, these formulas hold when Y is any deterministic function of K. It is interesting to remark that the statistical distance and the guessing entropy are in one-to-one relationship in this case. Indeed,

$$\Delta(K|Y) = 1 - \frac{2G(K|Y) - 1}{M}. \tag{52}$$

In the special case of the Hamming weight model (48)*, this gives*

$$\mathbb{P}_s(K|Y) = \frac{n+1}{2^n}, \quad G(K|Y) = \frac{1 + 2^{-n}\binom{2n}{n}}{2}, \tag{53}$$



(a) S_1 (linear) (b) S_7 (c) S_{19}

(d) S_{101} (e) S_{254} (AES – nonlinear)

Figure 9: Numerical evaluation of SD and GE for various noise levels σ^2 and increasing number of traces, for various choices of S-Boxes. Each different subfigure corresponds to a choice for the S-Box. The yellow curve corresponds to GE$\approx \frac{M+1}{2} - \frac{M}{2}$SD, indicating at least for low noise the approximation holds. It can be observed that at a fixed SD the GE decreases with the noise. Because of the variance in the estimation some points exits the region.

$$\textit{and } \Delta(K|Y) = 1 - 2^{-2n}\binom{2n}{n}. \tag{54}$$

If the leakage model is a subset of k bits this yields

$$\mathbb{P}_s(K|Y) = 2^k/M, \quad GE(K|Y) = \frac{1}{2} + \frac{M2^{-k}}{2}, \quad \Delta(K|Y) = 1 - 2^{-k}. \tag{55}$$

Proof. If K is equiprobable and $Y = y$ is fixed (with probability $\mathbb{P}(Y = y) = \frac{M_y}{M}$), then $K|Y = y$ is equiprobable over $M_y = |\mathcal{K}_y|$ values so that $\mathbb{P}_s(K|Y = y) = \frac{1}{M_y}$. Taking the average over Y gives $\mathbb{P}_s(K|Y) = \sum_y \frac{M_y}{M}\frac{1}{M_y}$, which yields the announced expression for SR. Similarly $G(K|Y = y) = \frac{M_y+1}{2}$ for a uniform guess, and taking the average over Y gives $G(K|Y) = \sum_y \frac{M_y}{M}\frac{M_y+1}{2}$, which yields the announced expression for GE. If $Y = y$ is fixed then $\Delta(K|Y = y) = \frac{1}{2}(M_y(\frac{1}{M_y} - \frac{1}{M}) + (M - M_y)\frac{1}{M}) = 1 - \frac{M_y}{M}$. Taking the average over Y yields $\sum_y \frac{M_y}{M}(1 - \frac{M_y}{M}) = 1 - \frac{1}{M^2}\sum_{y\in\mathcal{Y}} M_y^2$. The Hamming weight case follows from the Vandermonde's identity $\sum_{k=0}^{n}\binom{n}{k}^2 = \binom{2n}{n}$.

It is easily seen that we recover the well-known expressions $\mathbb{P}_s(K) = \frac{1}{M}$, $G(K) = \frac{M+1}{2}$, and $\Delta(K) = 0$ for a blind guess.

5.3. Deterministic Leakage for Multiple Observed Traces

Consider multiple observed traces (Q queries) $Y = (Y_1, Y_2, \ldots, Y_Q)$, where

$$Y_i = w_H(K \oplus t_i) \qquad (i = 1, 2, \ldots, Q) \tag{56}$$

for fixed and distinct plain or cipher texts $t_1, t_2, \ldots, t_Q$. In this case we are faced with a combinational problem since letting $Y = y$ determines the intersection of Q Hamming balls.

To simplify the analysis we consider $Q = 2$ and the computation of SR. Without loss of generality we can set $t_1 = 0$ and consider variable $t_2 = t$.

Theorem 16. *Let $w = w_H(t)$. Then*

$$\mathbb{P}_s(K|Y) = \frac{(w+1)(n-w+1)}{2^n} \tag{57}$$

In particular for 8-bit bytes ($n = 8$), one obtains:

$$\mathbb{P}_s = \begin{cases} \frac{n+1}{2^n} & \text{for } w \in \{0, 8\} \\ \frac{2n}{2^n} & \text{for } w \in \{1, 7\} \\ \frac{3(n-1)}{2^n} & \text{for } w \in \{3, 6\} \\ \frac{4(n-2)}{2^n} & \text{for } w \in \{4, 5\}. \end{cases} \tag{58}$$

Proof. We show that $|\mathcal{Y}| = (w+1)(n-w+1)$ in (51) as illustrated in Fig. 5 (a), where the set $\mathcal{Y}$ of points $(y_1 = w_H(k), y_2 = w_H(k \oplus t))$ forms a (rotated) $(w+1) \times (n-w+1)$ rectangle. Indeed, let $\bar{t}$ be the binary complement of t and write the decomposition $w_H(k) = w_H(k \cdot t) + w_H(k \cdot \bar{t})$ where $\cdot$ denotes the bitwise product. For fixed $w_H(t) = w$, $w_H(k \cdot t)$ can take $w+1$ values and $w_H(k \cdot \bar{t})$ takes $(n-w)+1$ independent values. Since $w_H(k \oplus t) = w_H(k) + w_H(t) - w_H(k \cdot t) = w + w_H(k \cdot \bar{t})$, we have $(y_1, y_2) = (w_H(k \cdot t) + w_H(k \cdot \bar{t}), w + w_H(k \cdot \bar{t}))$ which takes all possible $(w+1)(n-w+1)$ values.

More generally, the set $\mathcal{Y}$ can be determined by exhaustive enumeration of Hamming weights. We computed numerically the resulting SR and GE for $Q = 1$, 2, 3, and 4 traces. They are plotted as green dots in Fig. 2 for different values of M.

5.4. Role of the S-Box in the Hamming Weight Model

To prevent differential and linear cryptanalyses, block ciphers are composed with non-linear operations. This non-linearity is performed by substitution box (S-Box). We investigate different choices for the S-Box to observe its effect on SR and GE with respect to SCA resistance. We consider

$$S_i(x) = ax^i \oplus b \in \mathbb{F}_{2^n} \tag{59}$$

for exponents $i = \{1, 7, 19, 101, 254\}$, constants $a, b \in \mathbb{F}_{2^n}$.

As an illustration, Fig. 5 plots the various sets $\mathcal{Y}$ of Hamming weight leakage values for S_1 (linear), S_7, and the AES standard S_{254} (highly nonlinear). We observe that the cardinality $|\mathcal{Y}|$ increases as exponent i increases. This shows that SR (as given by Theorem 15) increases as nonlinearity increases. Fig. 6 (the 3-D extension of Fig. 5) also plots M_y as a function of $y \in \mathcal{Y}$. Here we observe that M_y tends to globally decrease as exponent i increases, which shows that GE (as given by Theorem 15) decreases as nonlinearity increases.

Therefore, the non-linearity of the S-Box diminishes the side channel resistance. The geometrical explanation of this phenomenon is that the scatter plots of Fig. 5 and 6 tend to spread out for nonlinear S-Boxes. This confirms the observation of [22] on the effect of the S-Box on the confusion coefficient, which for monobit leakage relates to both SR and GE [23].

5.5. Hamming Weight Leakage Model With Gaussian Noise

In this section we derive the expression of SR and GE in an Hamming Weight leakage scenario

$$Y = w_H(K \oplus t) + N \tag{60}$$

in the presence of additive white Gaussian Noise (AWGN) $N \sim \mathcal{N}(0, \sigma^2)$. Let f_Y and ϕ_σ denote the p.d.f. of Y and N, respectively. Thus $\phi_\sigma(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{x^2}{2\sigma^2})$. Let Q denote the standard Q-function $Q(x) = \int_x^\infty \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}} du$.

Theorem 17 (Expression With Noisy Leakage)**.**

$$\mathbb{P}_s(K|Y) = \frac{n+1}{M} - \frac{2n}{M} Q\big(\frac{1}{2\sigma}\big), \tag{61}$$

$$G(K|Y) = \frac{1}{2} + \frac{\binom{2n}{n}}{2M} + \frac{2\binom{2n}{n+1}}{M} Q\big(\frac{1}{2\sigma}\big) + \sum_{i=2}^{2n} f_i(n) Q\big(\frac{i}{2\sigma}\big), \tag{62}$$

$$\Delta(K|Y) = 1 - \frac{\binom{2n}{n}}{M^2} - 2\left(1 + \frac{1-M+\binom{2n+1}{n+1}-2\binom{2n}{n}}{M^2}\right) Q(\frac{1}{2\sigma}) + O\left(Q(\frac{3}{2\sigma})\right) \tag{63}$$

where the latter sum is negligible at first order in σ and where the f_i are rational functions in n and M.

For low noise one recovers (53). The proof is left in Appendix A.

5.6. Validation by Simulation

We evaluated numerically the relation between SR and GE for different noise levels σ^2 and different number of traces. The evaluation has been performed by 10^3 repetitions of maximum likelihood attacks on synthetically generated leakages.

Figures 7, 8 and 9 plots the resulting values for various noise levels and S-Boxes. We observe that for low noise the approximation $G(K|Y) \approx \mathbb{P}_s(K|Y)^{-1}$ still holds (yellow curve). As the noise increases, for a given SR, GE increases, and the latter approximation is no longer valid. The S-Box nonlinearity accentuates this effect because it decreases the minimum distance of points in $\mathcal{Y}$ in Fig. 5 and, therefore, makes the maximum likelihood attack less robust to noise. As expected, for low noise the approximation GE$\approx \frac{M+1}{2} - \frac{M}{2}$SD holds.

5.7. Validation on real traces from DPA Contest V4.2

Figure 10 plots the results on values of SR and GE computed on the three first folders of the DPA Contest V4.2 with a Hamming Weight template attack with known mask. As expected from the simulation the guessing entropy is lower bounded by SR^{-1}.

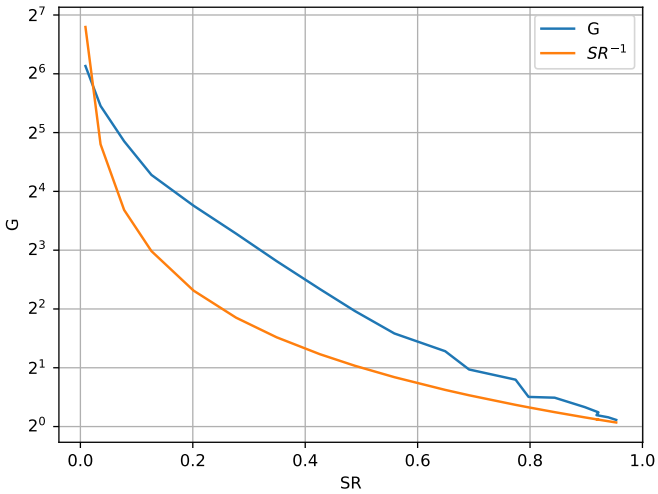


Figure 10: Results on Traces from DPA Contest v4.2

6. Conclusion

In this paper, optimal bounds between success rate, guessing entropy and distinguishability are derived with a simple majorization argument, and further improved for the Hamming weight leakage model—in particular for the classical assumptions of a priori equiprobable secret keys and additive white Gaussian measurement noise. Closed-form expressions and numerical computations are given for various leakage scenarios. A study of the impact of the choice S-Box with respect to SCA resistance confirms that nonlinearity of the S-Box tends to tighten the bounds between SR and GE. We established that distinguishability and guessing entropy are in one to one relationship for uniform keys and deterministic leakage models. In particular for low noise we have the approximation $GE \approx \frac{M+1}{2} - \frac{M}{2} SD$. The approximate relation $GE = 1/SR$ holds in the case of 8-bit bytes and low noise. This in turns imply that for 8-bit and low noise $1/SR \approx \frac{M+1}{2} - \frac{M}{2} SD$. We observed that for the probability of success and distinguishability the optimal reverse Pinsker inequality corresponds to the optimal Fano inequality and that the optimal reverse Fano inequality corresponds to the optimal Pinsker inequality.

As a perspective, we notice that our methodology can be easily generalized to the definitions of the ith order success rate [4] SR_i vs. GE. However, as pointed out in [11], such theoretical work assumes perfect knowledge on the distribution of K given observation Y. This generally underestimates the

practical GE for a non optimal attack because such a practical attack generally gives a suboptimal key ranking. Thus the results of this paper should yield adequate estimates only for optimal template attacks. The determination of more precise regions SR vs. GE for other types of attacks is a topic for future investigation.

References

[1] J. Beguinot, W. Cheng, S. Guilley, O. Rioul, Be My Guess: Guessing Entropy vs. Success Rate for Evaluating Side-Channel Attacks of Secure Chips, in: 25th Euromicro Conference on Digital System Design, DSD 2022, Maspalomas, Spain, August 31 - Sept. 2, 2022, IEEE, 2022, pp. 496–503. URL: https://doi.org/10.1109/DSD57027.2022.00072. doi:10.1109/DSD57027.2022.00072.

[2] O. Rioul, What Is Randomness? The Interplay between Alpha Entropies, Total Variation and Guessing, Physical Sciences Forum 5 (2022). URL: https://www.mdpi.com/2673-9984/5/1/30. doi:10.3390/psf2022005030.

[3] O. Rioul, The Interplay between Error, Total Variation, Alpha-Entropy and Guessing: Fano and Pinsker Direct and Reverse Inequalities, Entropy as part of the Special Issue MaxEnt 2022-the 41st International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering 25 (2023) 978.

[4] F.-X. Standaert, T. Malkin, M. Yung, A Unified Framework for the Analysis of Side-Channel Key Recovery Attacks, in: EUROCRYPT, volume 5479 of *LNCS*, Springer, 2009, pp. 443–461. Cologne, Germany.

[5] N. Veyrat-Charvillon, B. Gérard, M. Renauld, F.-X. Standaert, An Optimal Key Enumeration Algorithm and Its Application to Side-Channel Attacks, in: L. R. Knudsen, H. Wu (Eds.), Selected Areas in Cryptography, volume 7707 of *Lecture Notes in Computer Science*, Springer, 2012, pp. 390–406.

[6] T. Prest, D. Goudarzi, A. Martinelli, A. Passelègue, Unifying leakage models on a rényi day, in: A. Boldyreva, D. Micciancio (Eds.), Advances in Cryptology - CRYPTO 2019 - 39th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2019,

Proceedings, Part I, volume 11692 of *Lecture Notes in Computer Science*, Springer, 2019, pp. 683–712. URL: https://doi.org/10.1007/978-3-030-26948-7_24. doi:10.1007/978-3-030-26948-7_24.

[7] N. Veyrat-Charvillon, B. Gérard, F. Standaert, Security evaluations beyond computing power, in: T. Johansson, P. Q. Nguyen (Eds.), Advances in Cryptology - EUROCRYPT 2013, 32nd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Athens, Greece, May 26-30, 2013. Proceedings, volume 7881 of *Lecture Notes in Computer Science*, Springer, 2013, pp. 126–141. URL: http://dx.doi.org/10.1007/978-3-642-38348-9_8. doi:10.1007/978-3-642-38348-9_8.

[8] É. de Chérisey, S. Guilley, O. Rioul, P. Piantanida, Best Information is Most Successful — Mutual Information and Success Rate in Side-Channel Analysis, IACR Trans. Cryptogr. Hardw. Embed. Syst. 2019 (2019) 49–79. URL: https://doi.org/10.13154/tches.v2019.i2.49-79. doi:10.13154/tches.v2019.i2.49-79.

[9] M. O. Choudary, P. G. Popescu, Back to Massey: Impressively Fast, Scalable and Tight Security Evaluation Tools, in: W. Fischer, N. Homma (Eds.), Cryptographic Hardware and Embedded Systems - CHES 2017 - 19th International Conference, Taipei, Taiwan, September 25-28, 2017, Proceedings, volume 10529 of *Lecture Notes in Computer Science*, Springer, 2017, pp. 367–386. URL: https://doi.org/10.1007/978-3-319-66787-4_18. doi:10.1007/978-3-319-66787-4_18.

[10] I. Sason, S. Verdú, Improved bounds on lossless source coding and guessing moments via rényi measures, IEEE Transactions on Information Theory 64 (2018) 4323–4346. doi:10.1109/TIT.2018.2803162.

[11] Z. Zhang, A. A. Ding, Y. Fei, A Fast and Accurate Guessing Entropy Estimation Algorithm for Full-key Recovery, IACR Trans. Cryptogr. Hardw. Embed. Syst. 2020 (2020) 26–48. URL: https://doi.org/10.13154/tches.v2020.i2.26-48. doi:10.13154/tches.v2020.i2.26-48.

[12] G. H. Hardy, J. E. Littlewood, G. Pólya, Inequalities, Cambridge Univ. Press, 1934.

[13] A. Duc, S. Faust, F. Standaert, Making Masking Security Proofs Concrete - Or How to Evaluate the Security of Any Leaking Device, in: E. Oswald, M. Fischlin (Eds.), Advances in Cryptology - EUROCRYPT 2015 - 34th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Sofia, Bulgaria, April 26-30, 2015, Proceedings, Part I, volume 9056 of *Lecture Notes in Computer Science*, Springer, 2015, pp. 401–429. URL: http://dx.doi.org/10.1007/978-3-662-46800-5_16. doi:10.1007/978-3-662-46800-5_16.

[14] A. W. Marshall, I. Olkin, B. C. Arnold, Inequalities: Theory of Majorization and Its Applications, 2nd ed., Springer, 2011.

[15] S.-W. Ho, S. Verdú, Convexity/Concavity of Rényi Entropy and α-Mutual Information, in: Proc. IEEE Int. Symp. Inf. Theory (ISIT), Hong Kong, China, 2015, pp. 745–749.

[16] R. Graczyk, I. Sason, On Two-Stage Guessing, Information 12 (2021) 159.

[17] M. Khouzani, P. Malacaria, Generalized Entropies and Metric-Invariant Optimal Countermeasures for Information Leakage Under Symmetric Constraints, IEEE Transactions on Information Theory 65 (2019) 888–901.

[18] E. Arikan, An Inequality on Guessing and its Application to Sequential Decoding, IEEE Transactions on Information Theory 42 (1996) 99–105.

[19] O. Rioul, Variations on a Theme by Massey, IEEE Transactions on Information Theory 68 (2022) 2813–2828.

[20] S. Mangard, E. Oswald, T. Popp, Power Analysis Attacks: Revealing the Secrets of Smart Cards, Springer, 2006. ISBN 0-387-30857-1, http://www.dpabook.org/.

[21] W. Cheng, What can information guess? : Towards information leakage quantification in side-channel analysis. (Qu'est ce que l'information permet de deviner? : Vers une quantification des fuites d'informations dans l'analyse de canaux auxiliaires), Ph.D. thesis, Polytechnic Institute of Paris, France, 2021. URL: https://tel.archives-ouvertes.fr/tel-03504182.

[22] A. Heuser, O. Rioul, S. Guilley, A Theoretical Study of Kolmogorov-Smirnov Distinguishers — Side-Channel Analysis vs. Differential Cryptanalysis, in: E. Prouff (Ed.), Constructive Side-Channel Analysis and Secure Design - 5th International Workshop, COSADE 2014, Paris, France, April 13-15, 2014. Revised Selected Papers, volume 8622 of *Lecture Notes in Computer Science*, Springer, 2014, pp. 9–28. URL: http://dx.doi.org/10.1007/978-3-319-10175-0_2. doi:10.1007/978-3-319-10175-0_2.

[23] É. de Chérisey, S. Guilley, O. Rioul, Confused yet successful: - theoretical comparison of distinguishers for monobit leakages in terms of confusion coefficient and SNR, in: F. Guo, X. Huang, M. Yung (Eds.), Information Security and Cryptology - 14th International Conference, Inscrypt 2018, Fuzhou, China, December 14-17, 2018, Revised Selected Papers, volume 11449 of *Lecture Notes in Computer Science*, Springer, 2018, pp. 533–553. URL: https://doi.org/10.1007/978-3-030-14234-6_28. doi:10.1007/978-3-030-14234-6_28.

Appendix A. Proof of Thm. 17

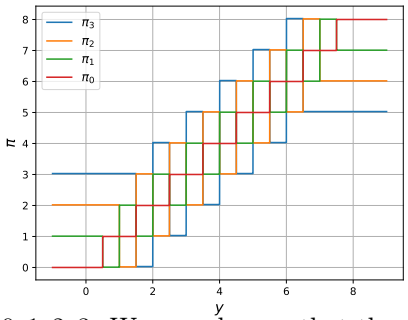


Figure A.11: $\pi_i(y)$ for $i = 0, 1, 2, 3$. We can observe that the π_i are step functions. Their are constant on the interval of the form $[\frac{p}{2}, \frac{p+1}{2})$ for all integer p.

For $j = 0, \ldots, n$ let $\pi_j(y)$ denote the $(j+1)$-th closest point to y in $\mathcal{Y}$. In particular, $\pi_0(y)$ is the closest point to y in $\mathcal{Y}$. It can be checked with the help of Fig. A.11 that

$$\pi_0(y) = \begin{cases} 0 & \text{for } y \leq -\frac{1}{2} \\ i & \text{for } y \in [i - \frac{1}{2}, i + \frac{1}{2}) \\ n & \text{for } y \geq n + \frac{1}{2}. \end{cases} \tag{A.1}$$

From (2), one has

$$\begin{aligned}
\mathbb{P}_s &= \frac{1}{M}\int \phi_\sigma(y-\pi_0(y))\,\mathrm{d}y \\
&= \frac{1}{M}\left(2Q(\frac{1}{2\sigma}) + \sum_{i=0}^{n}\int_{i-\frac{1}{2}}^{i+\frac{1}{2}} \phi_\sigma(y-i)\right) \\
&= \frac{1}{M}\left(2Q(\frac{1}{2\sigma}) + (n+1)\int_{-\frac{1}{2}}^{\frac{1}{2}} \phi_\sigma(y)\right) \\
&= \frac{1}{M}(2Q(\frac{1}{2\sigma}) + (n+1)(1-2Q(\frac{1}{2\sigma})))
\end{aligned}$$

which after simplification proves (61).

Now from (12), one has

$$G(K|Y) = \int f_Y(y)\sum_{k=1}^{M} k\, p_{(k)|y}\,\mathrm{d}y \tag{A.2}$$

Since the noise is Gaussian, the $p_{(k)|y}$ are sorted by Euclidean distance. Applying Bayes' rule we obtain

$$p_{(k)|y} = \phi_\sigma(y-\pi_j(y))\frac{1/M}{f_Y(y)}, \quad k = S_{j-1}(y)+1,\ldots,S_j(y). \tag{A.3}$$

where $S_j(y) = \sum_{i=0}^{j}\binom{n}{\pi_i(y)}$ for $j = 0,\ldots,n$ with the convention $S_{-1}(y) = 0$. Therefore,

$$\begin{aligned}
G(K|Y) &= \int f_Y(y)\sum_{j=0}^{n}\sum_{k=S_{j-1}(y)+1}^{S_j(y)} k\,\phi_\sigma(y-\pi_j(y))\frac{1/M}{f_Y(y)}\,\mathrm{d}y \\
&= \frac{1}{M}\sum_{j=0}^{n}\int \sum_{k=S_{j-1}(y)+1}^{S_j(y)} k\phi_\sigma(y-\pi_j(y))\,\mathrm{d}y \\
&= \frac{1}{M}\sum_{j=0}^{n}\int C_j(y)\phi_\sigma(y-\pi_j(y))\,\mathrm{d}y
\end{aligned}$$

where

$$C_j(y) = \frac{S_j(y)(S_j(y)+1) - S_{j-1}(y)(S_{j-1}(y)+1)}{2} \tag{A.4}$$

$$= \tfrac{1}{2}\tbinom{n}{\pi_j(y)}\big(2S_{j-1}(y) + \tbinom{n}{\pi_j(y)} + 1\big). \tag{A.5}$$

The $j = 0$ term can be written as

$$\int \frac{S_1(y)(1+S_1(y))}{2}\phi_\sigma(y - w_H(\pi_1(y)))\,\mathrm{d}y \tag{A.6}$$

$$= [2\int_{\frac{1}{2}}^{\infty} \phi_\sigma(y)dy + \sum_{i=0}^{n}\int_{i-\frac{1}{2}}^{i+\frac{1}{2}} \frac{\binom{n}{i}(1+\binom{n}{i})}{2}\phi_\sigma(y-i)\Big] \tag{A.7}$$

$$= \left[2Q(\frac{1}{2\sigma}) + \sum_{i=0}^{n}\frac{\binom{n}{i}(1+\binom{n}{i})}{2}(1-2Q(\frac{1}{2\sigma}))\right] \tag{A.8}$$

$$= \frac{M}{2} + \frac{1}{2}\binom{2n}{n} - Q(\frac{1}{2\sigma})(M + \binom{2n}{n} - 2). \tag{A.9}$$

We now compute the $j = 1$ term. It can be checked with the help of Fig. A.11 that

$$\pi_1(y) = \begin{cases} 1 & \text{for } y \le -\frac{1}{2} \\ i-1 & \text{for } y \in [i-\frac{1}{2}, i) \\ i+1 & \text{for } y \in [i, i+\frac{1}{2}) \\ n-1 & \text{for } y \ge n+\frac{1}{2}. \end{cases} \tag{A.10}$$

In the $j = 1$ term, the contribution of the integral from $\frac{1}{2}$ to ∞ and $-\infty$ to $-\frac{1}{2}$ both yields a term of value $\frac{n(n+3)}{2}Q(\frac{3\sigma}{2})$. The contribution of the integral over $[i-\frac{1}{2}, i)$ yields

$$\frac{1}{2}\tbinom{n}{i-1}[2\tbinom{n}{i} + \tbinom{n}{i-1} + 1](Q(\frac{1}{2\sigma}) - Q(\frac{1}{\sigma})). \tag{A.11}$$

and that over $(i, i+\frac{1}{2}]$ yields

$$\frac{1}{2}\tbinom{n}{i+1}[2\tbinom{n}{i} + \tbinom{n}{i+1} + 1](Q(\frac{1}{2\sigma}) - Q(\frac{1}{\sigma})). \tag{A.12}$$

Summing the contribution yields, after some calculation,

$$n(n+3)Q(\frac{3\sigma}{2}) + [M - 2 + 2\tbinom{2n+1}{n+1} - \tbinom{2n}{n}](Q(\frac{1}{2\sigma}) - Q(\frac{1}{\sigma})). \tag{A.13}$$

Here we have used the following Vandermonde identities:

$$\begin{aligned}
\sum_{i=0}^{n} \tbinom{n}{i+1}^2 &= \sum_{i=0}^{n} \tbinom{n}{i-1}^2 = \tbinom{2n}{n} - 1 \\
\sum_{i=0}^{n} \tbinom{n}{i}\tbinom{n}{i-1} &= \sum_{i=0}^{n} \tbinom{n}{i}\tbinom{n}{i+1} \\
&= \sum_{i=0}^{n} \tbinom{n}{i}[\tbinom{n+1}{i} - \tbinom{n}{i}] \\
&= \tbinom{2n+1}{n+1} - \tbinom{2n}{n}.
\end{aligned}$$

Summing the $j = 0$ and $j = 1$ terms simplifies to the first three terms in (62).

One can go further and compute terms corresponding to $j = 2, 3, \ldots, n$. It is easily seen from the above derivation that splitting the integral with Chasles relation on the interval where π_i is constant yields a sum of weighted $Q(\frac{i}{2\sigma})$ as shown in (62).

We first prove expression for the statistical distance.

$$\Delta(K|Y) = \int_{\mathbb{R}} f_Y(y)\Delta(K|Y=y)dy \tag{A.14}$$

$$= \int_{\mathbb{R}} f_Y(y) \sum_k (p(k|y) - 1/M)^+ \, dy \tag{A.15}$$

$$= \int_{\mathbb{R}} \sum_k (f_Y(y)p(k|y) - f_Y(y)/M)^+ \, dy \tag{A.16}$$

$$= \frac{1}{M} \int_{\mathbb{R}} \sum_k (\phi_\sigma(y - w_H(k)) - f_Y(y))^+ \, dy \tag{A.17}$$

$$= \frac{1}{M^2} \sum_{w=0}^{n} \binom{n}{w} \int_{\mathbb{R}} (\sum_{j=0}^{n} \binom{n}{j} (\phi_\sigma(y-w) - \phi_\sigma(y-j)) \Big)^+ dy \tag{A.18}$$

$$= \frac{1}{M^2} \sum_{w=0}^{n} \binom{n}{w} \int_{\mathbb{R}} (\sum_{j=0}^{n} \binom{n}{j} (\phi_\sigma(y) - \phi_\sigma(y-(j-w))) \Big)^+ dy. \tag{A.19}$$

At this point it is hard to determine when the integrand is positive to simplify the positive part. Hence we split the integral over multiple slices. Let

$$I_{a,b}(w) = \int_a^b (\sum_{j=0}^{n} \binom{n}{j} (\phi_\sigma(y) - \phi_\sigma(y-(j-w))) \Big)^+ dy \tag{A.20}$$

where we omit the subscript when $a = -\infty$ and $b = \infty$. Then

$$\Delta(K|Y) = \frac{1}{M^2} \sum_{w=0}^{n} \binom{n}{w} I(w). \tag{A.21}$$

If $w = 0$ then for σ small enough the integrand is positive on $(-\infty, \frac{1}{2}]$ and negative elsewhere. Then using an Abel summation technique it follows that

$$I(0) = I_{-\infty,\frac{1}{2}}(0) \tag{A.22}$$

$$= (M-1)(1 - Q_\sigma(\frac{1}{2})) - nQ_\sigma(\frac{1}{2}) - \sum_{j=0}^{n-2} \binom{n}{j+2} Q_\sigma(\frac{3}{2} + j) \tag{A.23}$$

$$= (M-1)(1 - Q_\sigma(\frac{1}{2})) - nQ_\sigma(\frac{1}{2}) + O(Q(\frac{3}{2\sigma})). \tag{A.24}$$

If $w = n$ for σ small enough the integrand is positive on $[-\frac{1}{2}, \infty)$ and negative elsewhere. It follows that $I(0) = I(n)$. Else $1 \leq w \leq n-1$ and for σ small enough the integrand is positive on $[-\frac{1}{2}, \frac{1}{2}]$ and negative elsewhere. It follows that

$$I(w) = I_{-\frac{1}{2},\frac{1}{2}}(w) \tag{A.25}$$

$$= M\left(1 - 2Q_\sigma\left(\frac{1}{2}\right)\right) - \sum_{j=0}^{n} \binom{n}{j} \left(Q\left(w - \frac{1}{2} - j\right) - Q\left(\frac{1}{2} + w - j\right)\right) \tag{A.26}$$

$$= M\left(1 - 2Q_\sigma\left(\frac{1}{2}\right)\right) - \sum_{j=0}^{n} \binom{n}{j} Q\left(w - \frac{1}{2} - j\right) + \sum_{j=0}^{n} \binom{n}{j} Q\left(\frac{1}{2} + w - j\right) \tag{A.27}$$

$$= M\left(1 - 2Q_\sigma\left(\frac{1}{2}\right)\right) - \sum_{j=0}^{n} \binom{n}{j} Q\left(w - \frac{1}{2} - j\right) + \sum_{j=-1}^{n-1} \binom{n}{j+1} Q\left(w - \frac{1}{2} - j\right) \tag{A.28}$$

$$= M\left(1 - 2Q_\sigma\left(\frac{1}{2}\right)\right) + Q_\sigma(\frac{1}{2} + w) - Q_\sigma(w - \frac{1}{2} - n)$$
$$- \sum_{j=0}^{n-1} \left(\tbinom{n}{j} - \tbinom{n}{j+1}\right) Q\left(w - j - \frac{1}{2}\right) \tag{A.29}$$

$$= M\left(1 - 2Q_\sigma\left(\frac{1}{2}\right)\right) - 1 - \sum_{j=0}^{n-1} \left(\tbinom{n}{j} - \tbinom{n}{j+1}\right) Q\left(w - j - \frac{1}{2}\right) + O(Q(\frac{3}{2\sigma})) \tag{A.30}$$

$$= M\left(1 - 2Q_\sigma\left(\frac{1}{2}\right)\right) - 1 - \sum_{j=w-1}^{n-1} \left(\tbinom{n}{j} - \tbinom{n}{j+1}\right) Q\left(w - j - \frac{1}{2}\right) + O(Q(\frac{3}{2\sigma})). \tag{A.31}$$

It only remains to sum the integrals over w up to an additive $O(\frac{3}{2\sigma})$ term.

$$\begin{aligned}\sum_{w=1}^{n-1}\binom{n}{w}I_w = \ & (M-2)\left(M\left(1-2Q_\sigma\left(\frac{1}{2}\right)\right)-1\right)\\ & -\sum_{w=1}^{n-1}\binom{n}{w}\sum_{j=w-1}^{n-1}\left(\binom{n}{j}-\binom{n}{j+1}\right)Q\left(w-j-\frac{1}{2}\right).\end{aligned} \tag{A.32}$$

We compute the sum $S=\sum_{w=1}^{n-1}\binom{n}{w}\sum_{j=w-1}^{n-1}\left(\binom{n}{j}-\binom{n}{j+1}\right)Q\left(w-j-\frac{1}{2}\right)$ to simplify the expression.

$$\begin{aligned}S=&\sum_{w=1}^{n-1}\binom{n}{w}\left(\binom{n}{w-1}-\binom{n}{w}\right)Q\left(\frac{1}{2}\right)+\sum_{w=1}^{n-1}\binom{n}{w}\left(\binom{n}{w}-\binom{n}{w+1}\right)(1-Q\left(\frac{1}{2}\right))\\ &+\sum_{w=1}^{n-1}\binom{n}{w}\sum_{j=w+1}^{n-1}\left(\binom{n}{j}-\binom{n}{j+1}\right)(1-Q\left(j+\frac{1}{2}-w\right)) \qquad \text{(A.33)}\\ =&\sum_{w=1}^{n-1}\binom{n}{w}\left(\binom{n}{w-1}-\binom{n}{w}\right)Q\left(\frac{1}{2}\right)\\ &+\sum_{w=1}^{n-1}\binom{n}{w}\left(\binom{n}{w}-\binom{n}{w+1}\right)(1-Q\left(\frac{1}{2}\right))+\sum_{w=1}^{n-1}\binom{n}{w}\left(\binom{n}{w+1}-\binom{n}{n}\right) \qquad \text{(A.34)}\\ =&\sum_{w=1}^{n-1}\binom{n}{w}\left(\binom{n}{w-1}-\binom{n}{w}-\binom{n}{w}+\binom{n}{w+1}\right)Q\left(\frac{1}{2}\right)\\ &+\sum_{w=1}^{n-1}\binom{n}{w}\left(\binom{n}{w}-\binom{n}{w+1}+\binom{n}{w+1}-\binom{n}{n}\right) \qquad \text{(A.35)}\\ =&\sum_{w=1}^{n-1}\binom{n}{w}\left(\binom{n}{w-1}-2\binom{n}{w}+\binom{n}{w+1}\right)Q\left(\frac{1}{2}\right)+\sum_{w=1}^{n-1}\binom{n}{w}\left(\binom{n}{w}-1\right) \qquad \text{(A.36)}\\ =&\sum_{w=1}^{n-1}\binom{n}{w}\left(\binom{n}{w-1}-2\binom{n}{w}+\binom{n}{w+1}\right)Q\left(\frac{1}{2}\right)+\binom{2n}{n}-M \qquad \text{(A.37)}\\ =&\left(2\binom{2n+1}{n+1}-4\binom{2n}{n}-2n+4\right)Q\left(\frac{1}{2}\right)+\binom{2n}{n}-M. \qquad \text{(A.38)}\end{aligned}$$

Summing everything we obtain finally obtain up to $O\left(Q\left(\frac{3}{2}\right)\right)$,

$$\begin{aligned}M^2\Delta(K|Y)=&(M-2)M(1-2Q_\sigma(\frac{1}{2}))-(M-2)\\ &-\left(2\binom{2n+1}{n+1}-4\binom{2n}{n}-2n+4\right)Q\left(\frac{1}{2}\right)\\ &-\binom{2n}{n}+M+2(M-1)(1-Q_\sigma(\frac{1}{2}))-2nQ_\sigma(\frac{1}{2}) \qquad \text{(A.39)}\end{aligned}$$

$$= \left(M^2 - \binom{2n}{n}\right) - 2\left(M^2 - M + 1 + \binom{2n+1}{n+1} - 2\binom{2n}{n}\right) Q\left(\frac{1}{2}\right). \quad \text{(A.40)}$$