

# Brute force searching, the typical set and Guesswork

Mark M. Christiansen and Ken R. Duffy  
Hamilton Institute

National University of Ireland, Maynooth  
Email: {mark.christiansen, ken.duffy}@nuim.ie

Flávio du Pin Calmon and Muriel Médard  
Research Laboratory of Electronics

Massachusetts Institute of Technology  
Email: {flavio, medard}@mit.edu

**Abstract**—Consider the situation where a word is chosen probabilistically from a finite list. If an attacker knows the list and can inquire about each word in turn, then selecting the word via the uniform distribution maximizes the attacker's difficulty, its Guesswork, in identifying the chosen word. It is tempting to use this property in cryptanalysis of computationally secure ciphers by assuming coded words are drawn from a source's typical set and so, for all intents and purposes, uniformly distributed within it. By applying recent results on Guesswork, for i.i.d. sources it is this equipartition ansatz that we investigate here. In particular, we demonstrate that the expected Guesswork for a source conditioned to create words in the typical set grows, with word length, at a lower exponential rate than that of the uniform approximation, suggesting use of the approximation is ill-advised.

## I. INTRODUCTION

Consider the problem of identifying the value of a discrete random variable by only asking questions of the sort: is its value  $X$ ? That this is a time-consuming task is a cornerstone of computationally secure ciphers [1]. It is tempting to appeal to the Asymptotic Equipartition Property (AEP) [2], and the resulting assignment of code words only to elements of the typical set of the source, to justify restriction to consideration of a uniform source, e.g. [3], [4], [5]. This assumed uniformity has many desirable properties, including maximum obfuscation and difficulty for the inquisitor, e.g. [6]. In typical set coding it is necessary to generate codes for words whose logarithmic probability is within a small distance of the word length times the specific Shannon entropy. As a result, while all these words have near-equal likelihood, the distribution is not precisely uniform. It is the consequence of this lack of perfect uniformity that we investigate here by proving that results on Guesswork [7], [8], [9], [10], [11] extend to this setting. We establish that for source words originally constructed from an i.i.d. sequence of letters, as a function of word length it is exponentially easier to guess a word conditioned to be in the source's typical set in comparison to the corresponding equipartition approximation. This raises questions about the wisdom of appealing to the AEP to justify sole consideration of the uniform distributions for cryptanalysis and provides alternate results in their place.

## II. THE TYPICAL SET AND GUESSWORK

Let  $\mathbb{A} = \{0, \dots, m-1\}$  be a finite alphabet and consider a stochastic sequence of words,  $\{W_k\}$ , where  $W_k$  is a word of

length  $k$  taking values in  $\mathbb{A}^k$ . The process  $\{W_k\}$  has specific Shannon entropy

$$H_W := - \lim_{k \rightarrow \infty} \frac{1}{k} \sum_{w \in \mathbb{A}^k} P(W_k = w) \log P(W_k = w),$$

and we shall take all logs to base  $e$ . For  $\epsilon > 0$ , the typical set of words of length  $k$  is

$$T_k^\epsilon := \left\{ w \in \mathbb{A}^k : e^{-k(H_W + \epsilon)} \leq P(W_k = w) \leq e^{-k(H_W - \epsilon)} \right\}.$$

For most reasonable sources [2],  $P(W_k \in T_k^\epsilon) > 0$  for all  $k$  sufficiently large and typical set encoding results in a new source of words of length  $k$ ,  $W_k^\epsilon$ , with statistics

$$P(W_k^\epsilon = w) = \begin{cases} \frac{P(W_k = w)}{P(W_k \in T_k^\epsilon)} & \text{if } w \in T_k^\epsilon, \\ 0 & \text{if } w \notin T_k^\epsilon. \end{cases} \quad (1)$$

Appealing to the AEP, these distributions are often substituted for their more readily manipulated uniformly random counterpart,  $U_k^\epsilon$ ,

$$P(U_k^\epsilon = w) := \begin{cases} \frac{1}{|T_k^\epsilon|} & \text{if } w \in T_k^\epsilon, \\ 0 & \text{if } w \notin T_k^\epsilon, \end{cases} \quad (2)$$

where  $|T_k^\epsilon|$  is the number of elements in  $T_k^\epsilon$ . While the distribution of  $W_k^\epsilon$  is near-uniform for large  $k$ , it is not perfectly uniform unless the original  $W_k$  was uniformly distributed on a subset of  $\mathbb{A}^k$ . Is a word selected using the distribution of  $W_k^\epsilon$  easier to guess than if it was selected uniformly,  $U_k^\epsilon$ ?

Given knowledge of  $\mathbb{A}^k$ , the source statistics of words, say those of  $W_k$ , and an oracle against which a word can be tested one at a time, an attacker's optimal strategy is to generate a partial-order of the words from most likely to least likely and guess them in turn [12], [7]. That is, the attacker generates a function  $G : \mathbb{A}^k \rightarrow \{1, \dots, m^k\}$  such that  $G(w') < G(w)$  if  $P(W_k = w') > P(W_k = w)$ . The integer  $G(w)$  is the number of guesses until word  $w$  is guessed, its Guesswork.

For fixed  $k$  it is shown in [12] that the Shannon entropy of the underlying distribution bears little relation to the expected Guesswork,  $E(G(W_k))$ , the average number of guesses required to guess a word chosen with distribution  $W_k$  using the optimal strategy. In a series of subsequent papers [7], [8], [9], [10], under ever less restrictive stochastic assumptions from words made up of i.i.d. letters to Markovian letters to sofic shifts, an asymptotic relationship as word length grows

between scaled moments of the Guesswork and specific Rényi entropy was identified:

$$\lim_{k \rightarrow \infty} \frac{1}{k} \log E(G(W_k)^\alpha) = \alpha R_W \left( \frac{1}{1 + \alpha} \right), \quad (3)$$

for  $\alpha > -1$ , where  $R_W(\beta)$  is the specific Rényi entropy for the process  $\{W_k\}$  with parameter  $\beta > 0$ ,

$$R_W(\beta) := \lim_{k \rightarrow \infty} \frac{1}{k} \frac{1}{1 - \beta} \log \left( \sum_{w \in \mathbb{A}^k} P(W_k = w)^\beta \right).$$

These results have recently [11] been built on to prove that  $\{k^{-1} \log G(W_k)\}$  satisfies a Large Deviation Principle (LDP), e.g. [13]. Define the scaled Cumulant Generating Function (sCGF) of  $\{k^{-1} \log G(W_k)\}$  by

$$\Lambda_W(\alpha) := \lim_{k \rightarrow \infty} \frac{1}{k} \log E \left( e^{\alpha \log G(W_k)} \right) \text{ for } \alpha \in \mathbb{R}$$

and make the following two assumptions.

- *Assumption 1:* For  $\alpha > -1$ , the sCGF  $\Lambda_W(\alpha)$  exists, is equal to  $\alpha R_W(1/(1 + \alpha))$  and has a continuous derivative in that range.
- *Assumption 2:* The limit

$$g_W := \lim_{k \rightarrow \infty} \frac{1}{k} \log P(G(W_k) = 1) \quad (4)$$

exists in  $(-\infty, 0]$ .

Should assumptions 1 and 2 hold, Theorem 3 of [11] establishes that  $\Lambda_W(\alpha) = g_W$  for all  $\alpha \leq -1$  and that the sequence  $\{k^{-1} \log G(W_k)\}$  satisfies a LDP with a rate function given by the Legendre Fenchel transform of the sCGF,  $\Lambda_W^*(x) := \sup_{\alpha \in \mathbb{R}} \{x\alpha - \Lambda_W(\alpha)\}$ . Assumption 1 is motivated by equation (3), while the Assumption 2 is a regularity condition on the probability of the most likely word. With

$$\gamma_W := \lim_{\alpha \downarrow -1} \frac{d}{d\alpha} \Lambda_W(\alpha), \quad (5)$$

where the order of the size of the set of maximum probability words of  $W_k$  is  $\exp(k\gamma_W)$  [11],  $\Lambda_W^*(x)$  can be identified as

$$= \begin{cases} -x - g_W & \text{if } x \in [0, \gamma_W] \\ \sup_{\alpha \in \mathbb{R}} \{x\alpha - \Lambda_W(\alpha)\} & \text{if } x \in (\gamma_W, \log(m)], \\ +\infty & \text{if } x \notin [0, \log(m)]. \end{cases} \quad (6)$$

Corollary 5 of [11] uses this LDP to prove a result suggested in [14], [15], that

$$\lim_{k \rightarrow \infty} \frac{1}{k} E(\log(G(W_k))) = H_W, \quad (7)$$

making clear that the specific Shannon entropy determines the expectation of the logarithm of the number of guesses to guess the word  $W_k$ . The growth rate of the expected Guesswork is a distinct quantity whose scaling rules can be determined directly from the sCGF in equation (3),

$$\lim_{k \rightarrow \infty} \frac{1}{k} \log E(G(W_k)) = \Lambda_W(1).$$

From these expressions and Jensen's inequality, it is clear that the growth rate of the expected Guesswork is less than  $H_W$ . Finally, as a corollary to the LDP, [11] provides the following approximation to the Guesswork distribution for large  $k$ :

$$P(G(W_k) = n) \approx \frac{1}{n} \exp(-k\Lambda_W^*(k^{-1} \log n)) \quad (8)$$

for  $n \in \{1, \dots, m^k\}$ . Thus to approximate the Guesswork distribution, it is sufficient to know the specific Rényi entropy of the source and the decay-rate of the likelihood of the sequence of most likely words.

Here we show that if  $\{W_k\}$  is constructed from i.i.d. letters, then both of the processes  $\{U_k^\epsilon\}$  and  $\{W_k^\epsilon\}$  also satisfy Assumptions 1 and 2 so that, with the appropriate rate functions, the approximation in equation (8) can be used with  $U_k^\epsilon$  or  $W_k^\epsilon$  in lieu of  $W_k$ . This enables us to compare the Guesswork distribution for typical set encoded words with their assumed uniform counterpart. Even in the simple binary alphabet case we establish that, apart from edge cases, a word chosen via  $W_k^\epsilon$  is exponential easier in  $k$  to guess on average than one chosen via  $U_k^\epsilon$ .

### III. STATEMENT OF MAIN RESULTS

Assume that the words  $\{W_k\}$  are made of i.i.d. letters, defining  $p = (p_0, \dots, p_{m-1})$  by  $p_a = P(W_1 = a)$ . We shall employ the following short-hand:  $h(l) := -\sum_a l_a \log l_a$  for  $l = (l_0, \dots, l_{m-1}) \in [0, 1]^m$ ,  $l_a \geq 0$ ,  $\sum_a l_a = 1$ , so that  $H_W = h(p)$ , and  $D(l||p) := -\sum_a l_a \log(p_a/l_a)$ . Furthermore, define  $l^- \in [0, 1]^m$  and  $l^+ \in [0, 1]^m$

$$l^- \in \arg \max_l \{h(l) : h(l) + D(l||p) - \epsilon = h(p)\}, \quad (9)$$

$$l^+ \in \arg \max_l \{h(l) : h(l) + D(l||p) + \epsilon = h(p)\}, \quad (10)$$

should they exist. Assume that  $h(p) + \epsilon \leq \log(m)$ . If this is not the case  $\log(m)$  should be substituted in place of  $h(l^-)$  for the  $\{U_k^\epsilon\}$  results. Sketch proofs of the following are deferred to the Appendix, with fuller proofs available in [16].

*Lemma 1:* Assumption 1 holds for  $\{U_k^\epsilon\}$  and  $\{W_k^\epsilon\}$  with

$$\Lambda_{U^\epsilon}(\alpha) := \alpha h(l^-)$$

and  $\Lambda_{W^\epsilon}(\alpha) =$

$$\begin{cases} -D(l^+||p) + \alpha h(l^+) & \text{if } \alpha > -1, \eta(\alpha) < h(p) - \epsilon, \\ (1 + \alpha) \log \sum_{a \in \mathbb{A}} p_a^{1/(1+\alpha)} & \text{if } \eta(\alpha) \in [h(p) - \epsilon, h(p) + \epsilon], \\ -D(l^-||p) + \alpha h(l^-) & \text{if } \eta(\alpha) > h(p) + \epsilon, \end{cases}$$

where, for  $\alpha > -1$ ,

$$\eta(\alpha) = -\frac{\sum_{a \in \mathbb{A}} p_a^{1/(1+\alpha)} \log p_a}{\sum_{b \in \mathbb{A}} p_b^{1/(1+\alpha)}}. \quad (11)$$

*Lemma 2:* Assumption 2 holds for  $\{U_k^\epsilon\}$  and  $\{W_k^\epsilon\}$  with

$$g_{U^\epsilon} = -h(l^-) \text{ and}$$

$$g_{W^\epsilon} = \min \left( -h(p) + \epsilon, \log \max_{a \in \mathbb{A}} p_a \right).$$

Thus by direct evaluation of the sCGFs at  $\alpha = 1$ ,

$$\lim_{k \rightarrow \infty} \frac{1}{k} \log E(G(U_k^\epsilon)) = h(l^-) \text{ and } \\ \lim_{k \rightarrow \infty} \frac{1}{k} \log E(G(W_k^\epsilon)) = \Lambda_{W^\epsilon}(1).$$

As the conditions of Theorem 3 [11] are satisfied

$$\lim_{k \rightarrow \infty} \frac{1}{k} E(\log(G(U_k^\epsilon))) = \Lambda'_{U^\epsilon}(0) = h(l^-) \text{ and } \\ \lim_{k \rightarrow \infty} \frac{1}{k} E(\log(G(W_k^\epsilon))) = \Lambda'_{W^\epsilon}(0) = h(p),$$

and we have the approximations

$$P(G(U_k^\epsilon) = n) \approx \frac{1}{n} \exp(-k\Lambda_{U^\epsilon}^*(k^{-1} \log n)) \text{ and } \\ P(G(W_k^\epsilon) = n) \approx \frac{1}{n} \exp(-k\Lambda_{W^\epsilon}^*(k^{-1} \log n)).$$

#### IV. EXAMPLE

Consider a binary alphabet  $\mathbb{A} = \{0, 1\}$  and words  $\{W_k\}$  constructed of i.i.d. letters with  $P(W_1 = 0) = p_0 > 1/2$ . In this case there are unique  $l^-$  and  $l^+$  satisfying equations (9) and (10) determined by:

$$l_0^- = p_0 - \frac{\epsilon}{\log(p_0) - \log(1 - p_0)}, \\ l_0^+ = p_0 + \frac{\epsilon}{\log(p_0) - \log(1 - p_0)}.$$

Selecting  $0 < \epsilon < (\log(p_0) - \log(1 - p_0)) \min(p_0 - 1/2, 1 - p_0)$  ensures that the typical set is growing more slowly than  $2^k$  and that  $1/2 < l_0^- < p_0 < l_0^+ < 1$ .

From equations (3) and (4), we have that

$$\Lambda_W(\alpha) = \begin{cases} \log(p_0) & \text{if } \alpha < -1, \\ (1 + \alpha) \log(p_0^{\frac{1}{1+\alpha}} + (1 - p_0)^{\frac{1}{1+\alpha}}) & \text{if } \alpha \geq -1. \end{cases}$$

From Lemmas 1 and 2 we obtain

$$\Lambda_{U^\epsilon}(\alpha) = \begin{cases} -h(l^-) & \text{if } \alpha < -1, \\ \alpha h(l^-) & \text{if } \alpha \geq -1, \end{cases}$$

and  $\Lambda_{W^\epsilon}(\alpha) =$

$$\begin{cases} -h(p) + \epsilon & \text{if } \alpha \leq -1, \\ -D(l^+ || p) + \alpha h(l^+) & \text{if } \eta(\alpha) < h(p) - \epsilon, \\ \alpha R_W(1/(1 + \alpha)) & \text{if } \eta(\alpha) \in [h(p) - \epsilon, h(p) + \epsilon], \\ -D(l^- || p) + \alpha h(l^-) & \text{if } \eta(\alpha) > h(p) + \epsilon, \end{cases}$$

where, for  $\alpha > -1$ , from equation (11) we have

$$\eta(\alpha) = -\frac{p_0^{1/(1+\alpha)} \log p_0 + (1 - p_0)^{1/(1+\alpha)} \log(1 - p_0)}{p_0^{1/(1+\alpha)} + (1 - p_0)^{1/(1+\alpha)}}.$$

With  $\gamma$  defined in equation (5), we have  $\gamma_W = 0$ ,  $\gamma_{U^\epsilon} = h(l^-)$  and  $\gamma_{W^\epsilon} = h(l^+)$  so that, as  $h(l^-) > h(l^+)$ , the ordering of the growth rates with word length of the set of most likely words from smallest to largest is: unconditioned source, conditioned source and uniform approximation.

From these sCGF equations, we can determine the average growth rates and estimates on the Guesswork distribution. In particular, we have that

$$\lim_{k \rightarrow \infty} \frac{1}{k} E(\log(G(W_k))) = \Lambda'_W(0) = h(p), \\ \lim_{k \rightarrow \infty} \frac{1}{k} E(\log(G(W_k^\epsilon))) = \Lambda'_{W^\epsilon}(0) = h(p), \\ \lim_{k \rightarrow \infty} \frac{1}{k} E(\log(G(U_k^\epsilon))) = \Lambda'_{U^\epsilon}(0) = h(l^-).$$

As  $h((x, 1 - x))$  is monotonically decreasing for  $x > 1/2$  and  $1/2 < l_0^- < p_0$ , the expectation of the logarithm of the Guesswork is growing faster for the uniform approximation than for either the unconditioned or conditioned word source. The growth rate of the expected Guesswork reveals more features. In particular, with  $A = \eta(1) - (h(p) + \epsilon)$ ,

$$\lim_{k \rightarrow \infty} \frac{1}{k} \log E(G(W_k)) = 2 \log(p_0^{\frac{1}{2}} + (1 - p_0)^{\frac{1}{2}}), \\ \lim_{k \rightarrow \infty} \frac{1}{k} \log E(G(W_k^\epsilon)) = \begin{cases} 2 \log(p_0^{\frac{1}{2}} + (1 - p_0)^{\frac{1}{2}}), & A \leq 0 \\ -D(l^- || p) + h(l^-), & A > 0 \end{cases} \\ \lim_{k \rightarrow \infty} \frac{1}{k} \log E(G(U_k^\epsilon)) = h(l^-).$$

For the growth rate of the expected Guesswork, from these it can be shown that there is no strict order between the unconditioned and uniform source, but there is a strict ordering between the the uniform approximation and the true conditioned distribution, with the former being strictly larger.

With  $\epsilon = 1/10$  and for a range of  $p_0$ , these formulae are illustrated in Figure 1. The top line plots

$$\lim_{k \rightarrow \infty} \frac{1}{k} E(\log(G(U_k^\epsilon)) - \log(G(W_k))) \\ = \lim_{k \rightarrow \infty} \frac{1}{k} E(\log(G(U_k^\epsilon)) - \log(G(W_k^\epsilon))) = h(l^-) - h(p),$$

showing that the expected growth rate in the logarithm of the Guesswork is always higher for the uniform approximation than both the conditioned and unconditioned sources. The second highest line plots the difference in growth rates of the expected Guesswork of the uniform approximation and the true conditioned source

$$\lim_{k \rightarrow \infty} \frac{1}{k} \log \frac{E(G(U_k^\epsilon))}{E(G(W_k^\epsilon))} \\ = \begin{cases} h(l^-) - 2 \log(p_0^{\frac{1}{2}} + (1 - p_0)^{\frac{1}{2}}) & \text{if } \eta(1) \leq h(p) + \epsilon \\ D(l^- || p) & \text{if } \eta(1) > h(p) + \epsilon. \end{cases}$$

That this difference is always positive, which can be established readily analytically, shows that the expected Guesswork of the true conditioned source is growing at a slower exponential rate than the uniform approximation. The second line and the lowest line, the growth rates of the uniform and unconditioned expected Guesswork

$$\lim_{k \rightarrow \infty} \frac{1}{k} \log \frac{E(G(U_k^\epsilon))}{E(G(W_k))} = h(l^-) - 2 \log(p_0^{\frac{1}{2}} + (1 - p_0)^{\frac{1}{2}}),$$

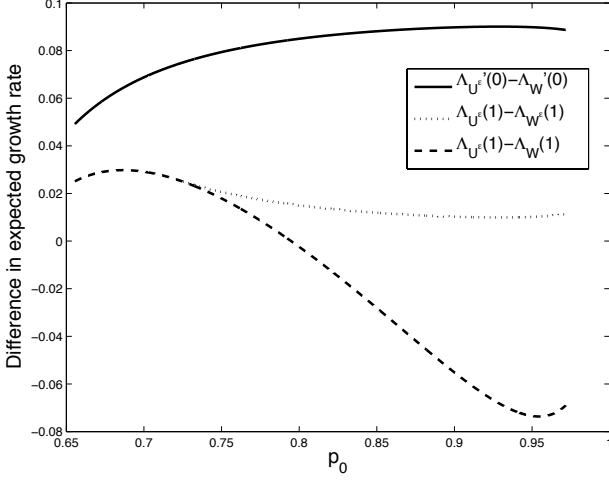


Fig. 1. Bernoulli( $p_0, 1-p_0$ ) source. Difference in exponential growth rates of Guesswork between uniform approximation, unconditioned and conditioned distribution with  $\epsilon = 0.1$ . Top curve is the difference in expected logarithms between the uniform approximation and both the conditioned and unconditioned word sources. Bottom curve is the log-ratio of the expected Guesswork of the uniform and unconditioned word sources, with the latter harder to guess for large  $p_0$ . Middle curve is the log-ratio of the uniform and conditioned word sources, which initially follows the lower line, before separating and staying positive, showing that the conditioned source is always easier to guess than the typically used uniform approximation.

initially agree. It can, depending on  $p_0$  and  $\epsilon$ , be either positive or negative. It is negative if the typical set is particularly small in comparison to the number of unconditioned words.

For  $p_0 = 8/10$ , the typical set is growing sufficiently slowly that a word selected from the uniform approximation is easier to guess than for unconditioned source. For this value, we illustrate the difference in Guesswork distributions between the unconditioned  $\{W_k\}$ , conditioned  $\{W_k^\epsilon\}$  and uniform  $\{U_k^\epsilon\}$  word sources. If we used the approximation in (8) directly, the graph would not be informative as the range of the unconditioned source is growing exponentially faster than the other two. Instead Figure 2 plots  $-x - \Lambda^*(x)$  for each of the three processes. That is, using equation (8) and its equivalents for the other two processes, it plots

$$\frac{1}{k} \log G(w), \text{ where } G(w) \in \{1, \dots, 2^k\},$$

against the large deviation approximations to

$$\frac{1}{k} \log P(W_k = w), \frac{1}{k} \log P(W_k^\epsilon = w) \text{ and } \frac{1}{k} \log P(U_k^\epsilon = w),$$

as the resulting plot is unchanging in  $k$ . The source of the discrepancy in expected Guesswork is apparent, with the unconditioned source having substantially more words to cover (due to the log x-scale). Both it and the true conditioned sources having higher probability words that skew their Guesswork. The first plateau for the conditioned and uniform distributions correspond to those words with approximately maximum highest probability; that is, the length of this plateau is  $\gamma_{W^\epsilon}$  or  $\gamma_{U^\epsilon}$ , defined in equation (5), so that, for example,

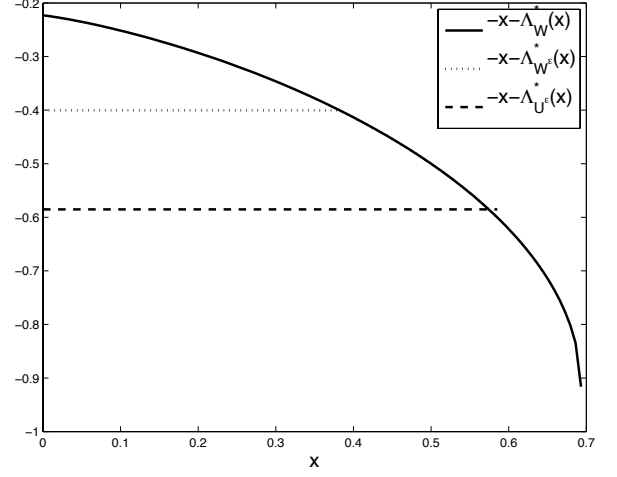


Fig. 2. Bernoulli( $8/10, 2/10$ ) source,  $\epsilon = 0.1$ . Guesswork distribution approximations. For large  $k$ ,  $x$ -axis is  $x = 1/k \log G(w)$  for  $G(w) \in \{1, \dots, 2^k\}$  and the  $y$ -axis is the large deviation approximation  $1/k \log P(X = w) \approx -x - \Lambda_X^*(x)$  for  $X = W_k, W_k^\epsilon$  and  $X = U_k^\epsilon$ .

approximately  $\exp(k\gamma_{W^\epsilon})$  words have probability of approximately  $\exp(k\gamma_{W^\epsilon})$ .

## V. CONCLUSION

By establishing that the expected Guesswork of a source conditioned on the typical set is growing with a smaller exponent than its usual uniform approximation, we have demonstrated that appealing to the AEP for the latter is erroneous in cryptanalysis and instead provide a correct methodology for identifying the Guesswork growth rate.

## APPENDIX

PROOF: Proof of Lemma 1. To establish these results it suffices to adapt the method of types argument employed in the elongated web-version of [8]. Two changes are necessary: the consideration of non-i.i.d. sources by restriction to  $T_k^\epsilon$ ; and the extension of the  $\alpha$  range to include  $\alpha \in (-1, 0]$  from that for  $\alpha \geq 0$  given in that document. Due to space constraints, we sketch the argument, providing more details in [16].

The proportion of the letter  $a \in \mathbb{A}$  in a word  $w = (w_1, \dots, w_k) \in \mathbb{A}^k$  is given by

$$n_k(w, a) := \frac{|\{1 \leq i \leq k : w_i = a\}|}{k}.$$

The number of words in a type  $l = (l_0, \dots, l_{m-1})$ , where  $l_a \geq 0$  for all  $a \in \mathbb{A}$  and  $\sum_{a \in \mathbb{A}} l_a = 1$ , is given by

$$N_k(l) := |\{w \in \mathbb{A}^k : n_k(w, a) = l_a \text{ for all } a \in \mathbb{A}\}|.$$

The set of all types and those just in the typical set are denoted

$$L_k := \{l : \exists w \in \mathbb{A}^k : n_k(w, a) = l_a \text{ for all } a \in \mathbb{A}\},$$

$$L_{\epsilon, k} := \{l : \exists w \in T_{\epsilon, k} : n_k(w, a) = l_a \text{ for all } a \in \mathbb{A}\}.$$

The primary idea is that in order to establish existence of the sCGF and specific Rényi entropy, one must prove that

$$\lim_{k \rightarrow \infty} \frac{1}{k} \log \max_{l \in L_{\epsilon, k}} \left( N_k(l)^{1+\alpha} \prod_{a \in \mathbb{A}} p_a^{kl_a} \right) \quad (12)$$

exists for all  $\alpha > -1$ , where the conditioning does not cause problems due to the AEP property that  $P(W_k \in T_k^\epsilon) > 1 - \epsilon$  for all  $k$  sufficiently large and thus

$$\lim_{k \rightarrow \infty} \frac{1}{k} \log P(W_k \in T_k^\epsilon) = 0.$$

Altering the bounds in [8] to include conditioning on the typical set we obtain upper and lower bounds on both  $\Lambda_{W^\epsilon}(\alpha)$  and  $\alpha R_{W^\epsilon}(1/(1+\alpha))$  with prefactors that differ for each, but which are insignificant in the limit. Modifications and manipulation of the arguments that led to the bounds for  $\alpha > 0$  in [8] are necessary for extension to  $\alpha \in (-1, 0]$ , but these can be realized.

Analysing equation (12) amounts to identifying limiting properties of a sequence of maximisers due to continuity properties of the function being optimized,

$$\arg \max_{l \in L_{\epsilon, k}} \left( N_k(l)^{1+\alpha} \prod_{a \in \mathbb{A}} p_a^{kl_a} \right).$$

This can be done with judicious use of Stirling's bounds. For  $\alpha > -1$ , define  $\eta(\alpha)$  by equation (11). If  $\eta(\alpha) \in [h(p) - \epsilon, h(p) + \epsilon]$ , then the limit of the maximum coincides with that for the unconditioned words with  $L_k$  in lieu of  $L_{\epsilon, k}$ , which collapses to the rescaled Rényi entropy in equation (3). Outside this range, the truncation to the typical set provides a different answer. If  $\eta(\alpha) < h(p) - \epsilon$ , with  $l^-$  and  $l^+$  defined in equations (9) and (10),

$$\lim_{k \rightarrow \infty} \frac{1}{k} \log N_k(l^+)^{1+\alpha} \prod_{a \in \mathbb{A}} p_a^{kl_a^+} = \alpha h(l^+) - D(l^+ \| p),$$

while if  $\eta(\alpha) > h(p) + \epsilon$ ,

$$\lim_{k \rightarrow \infty} \frac{1}{k} \log N_k(l^-)^{1+\alpha} \prod_{a \in \mathbb{A}} p_a^{kl_a^-} = \alpha h(l^-) - D(l^- \| p).$$

As each of these expressions is differentiable with respect to  $\alpha$ , differentiability of the full function follows by establishing that the slopes coincide at the  $\alpha$  values that determine where the functions knit together, which can be shown directly. ■

PROOF: Proof of Lemma 2. This can be established directly by a letter substitution argument, however, more generically it can be seen as being a consequence of the existence of specific min-entropy as a result of Assumption 1 via the following inequalities

$$\begin{aligned} & \alpha R \left( \frac{1}{1+\alpha} \right) - (1+\alpha) \log m \\ & \leq \liminf_{k \rightarrow \infty} \frac{1+\alpha}{k} \log \frac{m^k P(G(W_k) = 1)^{(1/(1+\alpha))}}{m^k} \\ & \leq \limsup_{k \rightarrow \infty} \frac{1}{k} \log P(G(W_k) = 1) \end{aligned} \quad (13)$$

$$\begin{aligned} & = (1+\alpha) \limsup_{k \rightarrow \infty} \frac{1}{k} \log P(G(W_k) = 1)^{(1/(1+\alpha))} \\ & \leq (1+\alpha) \limsup_{k \rightarrow \infty} \frac{1}{k} \log (P(G(W_k) = 1)^{(1/(1+\alpha))}) \\ & \quad + \sum_{i=2}^{m^k} P(G(W_k) = i)^{(1/(1+\alpha))} = \alpha R \left( \frac{1}{1+\alpha} \right). \end{aligned}$$

Equation (13) holds as  $P(G(W_k) = 1) \geq P(G(W_k) = i)$  for all  $i \in \{1, \dots, m^k\}$ . The veracity of the lemma follows as  $\alpha R((1+\alpha)^{-1})$  exists and is continuous for all  $\alpha > -1$  by Assumption 1 and  $(1+\alpha) \log m$  tends to 0 as  $\alpha \downarrow -1$ . ■

#### ACKNOWLEDGMENT

M.C. and K.D. supported by the Science Foundation Ireland Grant No. 11/PI/1177 and the Irish Higher Educational Authority (HEA) PRTL Network Mathematics Grant. F.d.P.C. and M.M. sponsored by the Department of Defense under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, recommendations, and conclusions are those of the authors and are not necessarily endorsed by the United States Government. Specifically, this work was supported by Information Systems of ASD(R&E).

#### REFERENCES

- [1] A. Menezes, S. Vanstone, and P. V. Oorschot, *Handbook of Applied Cryptography*. CRC Press, Inc., 1996.
- [2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 1991.
- [3] J. Pliam, "On the incomparability of entropy and marginal guesswork in brute-force attacks," in *INDOCRYPT*, 2000, pp. 67–79.
- [4] S. Draper, A. Khisti, E. Martinian, A. Vetro, and J. Yedidia, "Secure storage of fingerprint biometrics using Slepian-Wolf codes," in *ITA Workshop*, 2007.
- [5] Y. Sutcu, S. Rane, J. Yedidia, S. Draper, and A. Vetro, "Feature extraction for a Slepian-Wolf biometric system using LDPC codes," in *ISIT*, 2008.
- [6] F. du Pin Calmon, M. Médard, L. Zegler, J. Barros, M. Christiansen, and K. Duffy, "Lists that are smaller than their parts: A coding approach to tunable secrecy," in *Proc. 50th Allerton Conference*, 2012.
- [7] E. Arikan, "An inequality on guessing and its application to sequential decoding," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 99–105, 1996.
- [8] D. Malone and W. Sullivan, "Guesswork and entropy," *IEEE Trans. Inf. Theory*, vol. 50, no. 4, pp. 525–526, 2004, <http://www.maths.tcd.ie/~dwmalone/p/guess02.pdf>.
- [9] C.-E. Pfister and W. Sullivan, "Rényi entropy, guesswork moments and large deviations," *IEEE Trans. Inf. Theory*, no. 11, pp. 2794–800, 2004.
- [10] M. K. Hanawal and R. Sundaresan, "Guessing revisited: A large deviations approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 70–78, 2011.
- [11] M. M. Christiansen and K. R. Duffy, "Guesswork, large deviations and Shannon entropy," *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 796–802, 2013.
- [12] J. L. Massey, "Guessing and entropy," *IEEE Int. Symp. Inf. Theory*, pp. 204–204, 1994.
- [13] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Springer-Verlag, 1998.
- [14] E. Arikan and N. Merhav, "Guessing subject to distortion," *IEEE Trans. Inf. Theory*, vol. 44, pp. 1041–1056, 1998.
- [15] R. Sundaresan, "Guessing based on length functions," in *Proc. 2007 International Symp. on Inf. Th.*, 2007.
- [16] M. M. Christiansen, K. R. Duffy, F. du Pin Calmon, and M. Médard, "Brute force searching, the typical set and Guesswork," *arXiv preprint*, 2013, <http://arxiv.org/>.