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 obfustication 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 $\mathbb{A} = \{0, \ldots, m-1\}$ 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

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

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^ϵ, with statistics

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

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

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

where $|T_k^\epsilon|$ is the number of elements in T_k^ϵ. While the distribution of W_k^ϵ 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^ϵ easier to guess than if it was selected uniformly, U_k^ϵ?

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 $G : \mathbb{A}^k \to \{1, \ldots, m^k\}$ 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:

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

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

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

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

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

and make the following two assumptions.

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

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

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

Should assumptions 1 and 2 hold, Theorem 3 of [11] establishes that $\Lambda_W(\alpha)=g_W$ for all $\alpha\le-1$ 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

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

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

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

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

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

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),

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

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:

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

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^ϵ or W_k^ϵ 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^ϵ is exponential easier in k to guess on average than one chosen via U_k^ϵ.

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 $l=(l_0,\dots,l_{m-1})\in[0,1]^m$, $l_a\ge0$, $\sum_a l_a=1$, 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 $l^+\in[0,1]^m$

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

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

should they exist. For $\alpha>-1$, also define $l^W(\alpha)$ and $\eta(\alpha)$ by

$$l_a^W(\alpha):=\frac{p_a^{(1/(1+\alpha))}}{\sum_{b\in\mathbb{A}}p_b^{(1/(1+\alpha))}}\ \text{ for all } a\in\mathbb{A} \text{ and} \tag{11}$$

$$\eta(\alpha):=-\sum_a l_a^W\log p_a=-\frac{\sum_{a\in\mathbb{A}}p_a^{1/(1+\alpha)}\log p_a}{\sum_{b\in\mathbb{A}}p_b^{1/(1+\alpha)}}. \tag{12}$$

Assume that $h(p)+\epsilon\le\log(m)$. If this is not the case, $\log(m)$ should be substituted in place of $h(l^-)$ for the $\{U_k^\epsilon\}$ results. Proofs of the following are deferred to the Appendix.

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

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

and

$$\Lambda_{W^\epsilon}(\alpha)=\alpha h(l^*(\alpha))-D(l^*(\alpha)||p),$$

where

$$l^*(\alpha)=\begin{cases}l^+ & \text{if } \eta(\alpha)\le-h(p)-\epsilon,\\ l^W(\alpha) & \text{if } \eta(\alpha)\in(-h(p)-\epsilon,h(p)+\epsilon),\\ l^- & \text{if } \eta(\alpha)\ge-h(p)+\epsilon.\end{cases} \tag{13}$$

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 $\alpha = 1$,

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

As the conditions of Theorem 3 [11] are satisfied

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

and we have the approximations

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

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 $1/2 < l_0^- < p_0 < l_0^+ < 1$.

With $l^W(\alpha)$ defined in equation (11), from equations (3) and (4) we have that

$$\Lambda_W(\alpha) = \begin{cases} \log(p_0) & \text{if } \alpha < -1, \\ \alpha h(l^W(\alpha)) - D(l^W(\alpha)||p), & \text{if } \alpha \geq -1. \end{cases}$$
$$= \begin{cases} \log(p_0) & \text{if } \alpha < -1, \\ (1+\alpha)\log\left(p_0^{\frac{1}{1+\alpha}} + (1-p_0)^{\frac{1}{1+\alpha}}\right) & \text{if } \alpha \geq -1, \end{cases}$$

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

$$\Lambda_{W^\epsilon}(\alpha) = \begin{cases} -h(p) + \epsilon & \text{if } \alpha \leq -1, \\ \alpha h(l^*(\alpha)) - D(l^*(\alpha)||p) & \text{if } \alpha \geq -1, \end{cases}$$

where $l^*(\alpha)$ is deinfed in equation (13) and $\eta(\alpha)$ defined in equation (12).

With γ 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

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

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 $A = \eta(1) - (h(p) + \epsilon)$,

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

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

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

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

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

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

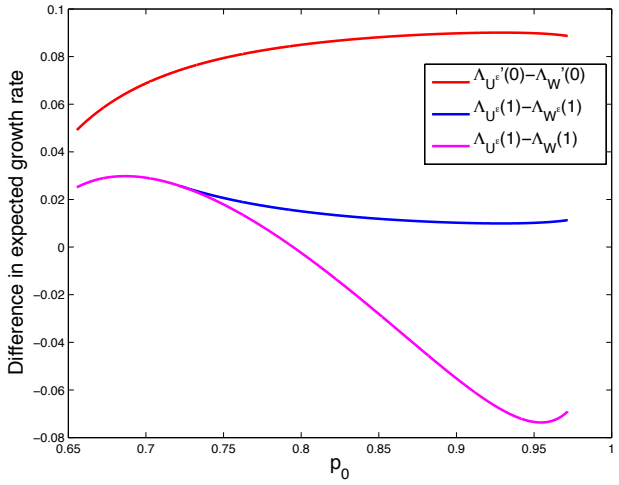

Fig. 1. Bernoulli$(p_0, 1-p_0)$ source. Difference in exponential growth rates of Guesswork between uniform approximation, unconditioned and conditioned distribution with $\epsilon = 0.1$. Top curve is the difference in expected logarithms between the uniform approximation and both the conditioned and 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.

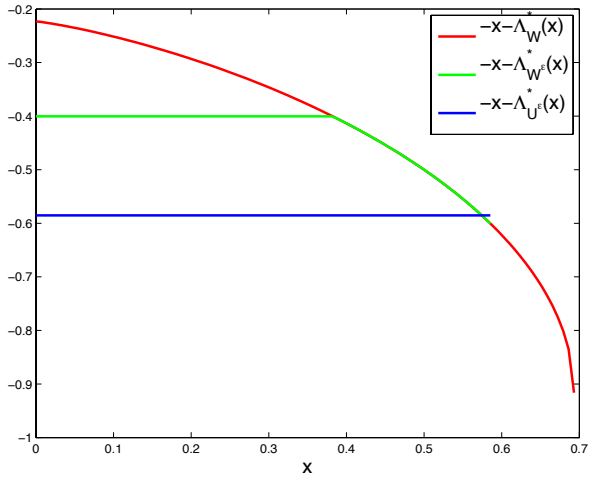

Fig. 2. Bernoulli$(8/10, 2/10)$ source, $\epsilon = 0.1$. Guesswork distribution approximations. For large k, x-axis is $x = 1/k \log G(w)$ for $G(w) \in \{1, \ldots, 2^k\}$ and the y-axis is the large deviation approximation $1/k \log P(X = w) \approx -x - \Lambda^*_X(x)$ for $X = W_k, W_k^\epsilon$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

unconditioned expected Guesswork

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

initially agree. It can, depending on p_0 and ϵ, 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 quickly 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

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

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 maximum highest probability (slowest exponential decay-rate).

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

Note that by the definition of T_k^ϵ as a typical set, $P(W_k \in T_k^\epsilon) > 1 - \epsilon$ for all k sufficiently large and thus

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

which we will use in the proofs of both lemmas.

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

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

The number of words in a type l, where $l \in [0, 1]$ for all $a \in \mathbb{A}$ and $\sum_{a\in\mathbb{A}} l_a = 1$, is given by

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

The set of all types, those just in the typical set and smooth

approximations to those in the typical set are denoted

$$L_k := \{l : \exists w \in \mathbb{A}^k \text{ such that } n_k(w,a) = l_a \ \forall a \in \mathbb{A}\},$$
$$L_k^\epsilon := \{l : \exists w \in T_{\epsilon,k} \text{ such that } n_k(w,a) = l_a \ \forall a \in \mathbb{A}\},$$
$$L^\epsilon := \left\{ l : \sum_a l_a \log p_a \in [-h(p) - \epsilon, -h(p) + \epsilon] \right\},$$

where it can readily seen that $L_k^\epsilon \subset L^\epsilon$ for all k.

For $\{U_k^\epsilon\}$ we need the following Lemma.

Lemma 3: The exponential growth rate of the size of the typical set is

$$\lim_{k\to\infty} \frac{1}{k} \log |T_k^\epsilon| = \begin{cases} \log m \text{ if } \log m \le h(p) + \epsilon \\ h(l^-) \text{ otherwise.} \end{cases}$$

where l^- is defined in equation (9).

PROOF: For fixed k, by the union bound

$$\max_{l \in L_k^\epsilon} \frac{k!}{\prod_{a\in\mathbb{A}}(kl_a)!} \le |T_k^\epsilon| \le (k+1)^m \max_{l\in L_k^\epsilon} \frac{k!}{\prod_{a\in\mathbb{A}}(kl_a)!}.$$

For the logarithmic limit, these two bounds coincide so consider the concave optimization problem

$$\max_{l \in L_k^\epsilon} \frac{k!}{\prod_{a\in\mathbb{A}}(kl_a)!}.$$

We can upper bound this optimization by replacing L_k^ϵ with the smoother version, its superset L^ϵ. Using Stirling's bound we have that

$$\limsup_{k\to\infty} \frac{1}{k} \log \sup_{l\in L^\epsilon} \frac{k!}{\prod_{a\in\mathbb{A}}(kl_a)!}$$
$$\le \sup_{l\in L^\epsilon} h(l) = \begin{cases} \log(m) & \text{if } h(p)+\epsilon \ge \log(m) \\ h(l^-) & \text{if } h(p)+\epsilon < \log(m). \end{cases}$$

For the lower bound, we need to construct a sequence $\{l^{(k)}\}$ such that $l^{(k)} \in L_k^\epsilon$ for all k sufficiently large and $h(l^{(k)})$ converges to either $\log(m)$ or $h(l^-)$, as appropriate. Let $l^* = (1/m, \ldots, 1/m)$ or l^- respectively, letting $c \in \arg\max p_a$ and define

$$l_a^{(k)} = \begin{cases} k^{-1}\lfloor kl_a^* \rfloor + 1 - \sum_{b\in\mathbb{A}} \frac{1}{k} \lfloor kl_b^* \rfloor & \text{if } a = c, \\ k^{-1}\lfloor kl_a^* \rfloor & \text{if } a \ne c. \end{cases}$$

Then $l^{(k)} \in L_k^\epsilon$ for all $k > -m\log(p_c)/(2\epsilon)$ and $h(l^{(k)}) \to h(l^*)$, as required.

■

PROOF: Proof of Lemma 1. Considering $\{U_k^\epsilon\}$ first,

$$\alpha R_{U^\epsilon}\left(\frac{1}{1+\alpha}\right) = \alpha \lim_{k\to\infty} \frac{1}{k} \log |T_k^\epsilon| = \alpha h(l^-),$$

by Lemma 3. To evaluate $\Lambda_{U^\epsilon}(\alpha)$, as for any $n \in \mathbb{N}$ and $\alpha > 0$

$$\sum_{i=1}^n i^\alpha \ge \int_0^n x^\alpha dx,$$

again using Lemma 3 we have

$$\begin{aligned} \alpha h(l^-) &= \lim_{k\to\infty} \frac{1}{k} \log \frac{1}{1+\alpha} |T_k^\epsilon|^\alpha \\ &\le \lim_{k\to\infty} \frac{1}{k} \log E(e^{\alpha \log G(U_k^\epsilon)}) \\ &= \lim_{k\to\infty} \frac{1}{k} \log \frac{1}{|T_k^\epsilon|} \sum_{i=1}^{|T_k^\epsilon|} i^\alpha \\ &\le \lim_{k\to\infty} \frac{1}{k} \log |T_k^\epsilon|^\alpha = \alpha h(l^-), \end{aligned}$$

where we have used Lemma 3. The reverse of these bounds holds for $\alpha \in (-1, 0]$, giving the result.

We break the argument for $\{W_k^\epsilon\}$ into three steps. Step 1 is to show the equivalence of the existence of $\Lambda_{W^\epsilon}(\alpha)$ and $\alpha R_{W^\epsilon}(1/(1+\alpha))$ for $\alpha > -1$ with the existence of the following limit

$$\lim_{k\to\infty} \frac{1}{k} \log \max_{l\in L_k^\epsilon} \left\{ N_k(l)^{1+\alpha} \prod_{a\in\mathbb{A}} p_a^{kl_a} \right\}. \tag{14}$$

Step 2 then establishes this limit and identifies it. Step 3 shows that $\Lambda'_{W^\epsilon}(\alpha)$ is continuous for $\alpha > -1$. To achieve steps 1 and 2, we adopt and adapt the method of types argument employed in the elongated web-version of [8].

Step 1 Two changes from the bounds of [8] Lemma 5.5 are necessary: the consideration of non-i.i.d. sources by restriction to T_k^ϵ; and the extension of the α range to include $\alpha \in (-1, 0]$ from that for $\alpha \ge 0$ given in that document. Adjusted for conditioning on the typical set we get

$$\begin{aligned} &\frac{1}{1+\alpha} \max_{l\in L_k^\epsilon} \left\{ N_k(l)^{1+\alpha} \frac{\prod_{a\in\mathbb{A}} p_a^{kl_a}}{\sum_{w\in T_k^\epsilon} P(W_k = w)} \right\} \\ &\le E(e^{\alpha \log G(W_k^\epsilon)}) \le \\ &(k+1)^{m(1+\alpha)} \max_{l\in L_k^\epsilon} \left\{ N_k(l)^{1+\alpha} \frac{\prod_{a\in\mathbb{A}} p_a^{kl_a}}{\sum_{w\in T_k^\epsilon} P(W_k = w)} \right\}. \end{aligned} \tag{15}$$

The necessary modification of these inequalities for $\alpha \in (-1, 0]$ gives

$$\begin{aligned} &\max_{l\in L_k^\epsilon} \left\{ N_k(l)^{1+\alpha} \frac{\prod_{a\in\mathbb{A}} p_a^{kl_a}}{\sum_{w\in T_k^\epsilon} P(W_k = w)} \right\} \\ &\le E(e^{\alpha \log G(W_k^\epsilon)}) \le \\ &\frac{(k+1)^m}{1+\alpha} \max_{l\in L_k^\epsilon} \left\{ N_k(l)^{1+\alpha} \frac{\prod_{a\in\mathbb{A}} p_a^{kl_a}}{\sum_{w\in T_k^\epsilon} P(W_k = w)} \right\}. \end{aligned} \tag{16}$$

To show the lower bound holds if $\alpha \in (-1, 0]$ let

$$l^* \in \arg\max_{l\in L_k^\epsilon} \left\{ N_k(l)^{1+\alpha} \frac{\prod_{a\in\mathbb{A}} p_a^{kl_a}}{\sum_{w\in T_k^\epsilon} P(W_k = w)} \right\}.$$

Taking $\liminf_{k\to\infty} k^{-1} \log$ and $\limsup_{k\to\infty} k^{-1} \log$ of equations (15) and (16) establishes that if the limit (14) exists, $\Lambda_{W^\epsilon}(\alpha)$ exists and equals it. Similar inequalities provide the same result for $\alpha R_{W^\epsilon}(1/(1+\alpha))$.

Step 2 The problem has been reduced to establishing the existence of

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

and identifying it. The method of proof is similar to that employed in Lemma 1: we provide an upper bound for the limsup and then establish a corresponding lower bound.

If $l^{(k)}\to l$ with $l^{(k)}\in L_k$, then using Stirling's bounds we have that

$$\lim_{k\to\infty}\frac{1}{k}\log N_k(l^{(k)})=h(l).$$

This convergence occurs uniformly in l and so, as $L_k^\epsilon\subset L^\epsilon$ for all k,

$$\begin{aligned}\limsup_{k\to\infty}&\frac{1}{k}\log\max_{l\in L_k^\epsilon}\left\{N_k(l)^{1+\alpha}\prod_{a\in\mathbb{A}}p_a^{kl_a}\right\}\\ &\le\sup_{l\in L^\epsilon}\left((1+\alpha)h(l)+\sum_a l_a\log p_a\right)\\ &=\sup_{l\in L^\epsilon}\left(\alpha h(l)-D(l||p)\right).\end{aligned}\tag{17}$$

This is a concave optimization problem in l with convex constraints. Not requiring $l\in L^\epsilon$, the unconstrained optimizer over all l is attained at $l^W(\alpha)$ defined in equation (11), which determines $\eta(\alpha)$ in equation (12). Thus the optimizer of the constrained problem (17) can be identified as that given in equation (13). Thus we have that

$$\begin{aligned}\limsup_{k\to\infty}&\frac{1}{k}\log\max_{l\in L_k^\epsilon}\left\{N_k(l)^{1+\alpha}\prod_{a\in\mathbb{A}}p_a^{kl_a}\right\}\\ &\le\alpha h(l^*(\alpha))+D(l^*(\alpha)||p),\end{aligned}$$

where $l^*(\alpha)$ is defined in equation (13).

We complete the proof by generating a matching lower bound. To do so, for given $l^*(\alpha)$ we need only create a sequence such that $l^{(k)}\to l^*(\alpha)$ and $l^{(k)}\in L_k^\epsilon$ for all k. If $l^*(\alpha)=l^-$, then the sequence used in the proof of Lemma 3 suffices. For $l^*(\alpha)=l^+$, we use the same sequence but with floors in lieu of ceilings and the surplus probability distributed to a least likely letter instead of a most likely letter. For $l^*(\alpha)=l^W(\alpha)$, either of these sequences can be used.

Step 3 As $\Lambda_{W^\epsilon}(\alpha)=\alpha h(l^*(\alpha))-D(l^*(\alpha)||p)$, with $l^*(\alpha)$ defined in equation (13),

$$\frac{d}{d\alpha}\Lambda_{W^\epsilon}(\alpha)=h(l^*(\alpha))+\Lambda_{W^\epsilon}(\alpha)\frac{d}{d\alpha}l^*(\alpha).$$

Thus to establish continuity it suffices to establish continuity of $l^*(\alpha)$ and its derivative, which can be done readily by calculus. ■

PROOF: Proof of Lemma 2. First consider

$$\begin{aligned}g_{U^\epsilon}&=\lim_{k\to\infty}\frac{1}{k}\max_{w\in T_k^\epsilon}\log P(U_k^\epsilon=w)\\ &=\lim_{k\to\infty}\frac{1}{k}\log\frac{1}{|T_k^\epsilon|}=-h(l^-),\end{aligned}$$

using Lemma 3.

For $\{W_k^\epsilon\}$, if $g_W<-h(p)+\epsilon$ the result follows simply, so assume that this is not the case. By the property mentioned at the beginning of this section, the normalisation doesn't play a rôle in the limit, i.e.

$$\begin{aligned}&\limsup_{k\to\infty}\frac{1}{k}\log\max_{w\in T_k^\epsilon}P(W_k^\epsilon=w)\\ &=\limsup_{k\to\infty}\frac{1}{k}\log\max_{w\in T_k^\epsilon}P(W_k=w)\end{aligned}$$

with an analogous equality for the lower bound. As $P(W_k=w)\le\exp(-k(h(p)-\epsilon))$ for all $w\in T_k^\epsilon$, the upper bound follows immediately and we need the corresponding lower bound on

$$\liminf_{k\to\infty}\frac{1}{k}\log\max_{w\in T_k^\epsilon}P(W_k=w).$$

If $g_W>-h(p)+\epsilon$, there exists $K\in\mathbb{N}$ such that for all $k>K$, $k^{-1}\log\max_{w\in\mathbb{A}^k}P(W_k=w)>-h(p)+\epsilon$. Then taking $k>K$,

$$\max_{w\in T_k^\epsilon}P(W_k=w)\ge e^{-k(h(p)+\epsilon)}\frac{\min_{a\in\mathbb{A}}p_a}{\max_{b\in\mathbb{A}}p_b}.$$

To prove this, we use proof by contradiction. Assume

$$\max_{w\in T_k^\epsilon}P(W_k=w)<e^{-k(h(p)-\epsilon)}\frac{\min_{a\in\mathbb{A}}p_a}{\max_{b\in\mathbb{A}}p_b}.$$

Take a word $w^*\in\arg\max_{w\in T_k^\epsilon}P(W_k=w)$, there exists at least one letter in w^*, $b\in\mathbb{A}$, such that $p_b<\max_{a\in A}p_a$ as $P(W_k=w^*)<\max_{w\in\mathbb{A}^k}P(W_k=w)$. We then replace one occurrence of b in w^* with an element of $\arg\max_{a\in A}p_a$ to make the k letter word w'. Then $k^{-1}\log P(W_k=w^*)<k^{-1}\log P(W_k=w')$. As for each k we have only changed one letter and by assumption,

$$\begin{aligned}&\frac{1}{k}\log P(W_k=w')\le\\ &\frac{1}{k}\log\left(P(W_k=w^*)\frac{\max_{b\in\mathbb{A}}p_b}{\min_{a\in\mathbb{A}}p_a}\right)<-h(p)+\epsilon.\end{aligned}$$

This implies $w'\in T_k^\epsilon$ and contravenes our choice of w^*. So

$$\liminf_{k\to\infty}k^{-1}\log\max_{w\in T_k^\epsilon}P(W_k=w)\ge-h(p)+\epsilon$$

if $g_W>-h(p)+\epsilon$. Lastly if $g_W=-h(p)+\epsilon$,

$$\begin{aligned}&\max_{w\in T_k^\epsilon}P(W_k=w)\ge\\ &\begin{cases}\max_{w\in\mathbb{A}^k}P(W_k=w)\\ \qquad\text{if }\max_{w\in\mathbb{A}^k}P(W_k=w)<e^{-k(h(p)-\epsilon)}\\ e^{-k(h(p)-\epsilon)}\dfrac{\min_{a\in\mathbb{A}}p_a}{\max_{b\in\mathbb{A}}p_b}\ \text{otherwise.}\end{cases}\end{aligned}$$

The result follows as

$$\begin{aligned}&-h(p)+\epsilon+\liminf_{k\to\infty}\left(\frac{1}{k}\log\frac{\min_{a\in\mathbb{A}}p_a}{\max_{b\in\mathbb{A}}p_b}\right)\\ &=\liminf_{k\to\infty}\frac{1}{k}\log\max_{w\in\mathbb{A}^k}P(W_k=w)=-h(p)+\epsilon.\end{aligned}$$

■

ACKNOWLEDGMENT

M.C. and K.D. supported by the Science Foundation Ireland Grant No. 11/PI/1177 and the Irish Higher Educational Authority (HEA) PRTLI 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.* 50^{th} *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. 2794–00, 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. Symo. 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.