Brute force searching, the typical set and Guesswork

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1109/ISIT.2013.6620428">http://dx.doi.org/10.1109/ISIT.2013.6620428</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Institute of Electrical and Electronics Engineers (IEEE)</td>
</tr>
<tr>
<td>Version</td>
<td>Original manuscript</td>
</tr>
<tr>
<td>Accessed</td>
<td>Sat Feb 02 05:01:19 EST 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/90432">http://hdl.handle.net/1721.1/90432</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Creative Commons Attribution-Noncommercial-Share Alike</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td><a href="http://creativecommons.org/licenses/by-nc-sa/4.0/">http://creativecommons.org/licenses/by-nc-sa/4.0/</a></td>
</tr>
</tbody>
</table>
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 for a source conditioned to create words in the typical set grows, for i.i.d. sources, it is this equipartition ansatz that we investigate 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 random variable by only asking questions of the sort: is its value $X$? That this is a time-consuming task is a cornerstone

$\text{I}[X]$ of $X$, the source statistics of words, say $\mathbb{A}$, 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

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 $k$, $W_k'$, with statistics

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

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

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

where $|T_k^\epsilon|$ is the number of elements in $T_k^\epsilon$. While the distribution of $W_k'$ 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 $W_k'$ 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]. 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

$$G(X) := \sum_{w \in \mathbb{A}^k} P(X = w) \log P(X = w).$$

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

$$H(W_k) := \sum_{w \in \mathbb{A}^k} P(W_k = w) \log P(W_k = w),$$

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

$$T_k^\epsilon := \left\{ w \in \mathbb{A}^k : e^{-k(H_W + \epsilon)} \leq P(W_k = w) \leq e^{-k(H_W - \epsilon)} \right\}.$$
between scaled moments of the Guesswork and specific Rényi entropy was identified:

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

for $\alpha > -1$, where $R_W(\beta)$ is the specific Rényi entropy for the process $\{W_k\}$ with parameter $\beta > 0$, $R_W(\beta) := \lim_{k \to \infty} \frac{1}{k} \frac{1}{1 - \beta} \log \left( \sum_{w \in \Lambda^k} P(W_k = w)^\beta \right)$.

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. [14]. 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 $\alpha R_W(1/(1+\alpha))$ and has a continuous derivative in that range.

- **Assumption 2:** The limit

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

$$\gamma_W := \lim_{\alpha \to -1} \frac{d}{d\alpha} \Lambda_W(\alpha),$$  \hfill (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

$$\begin{cases} -x - g_W & \text{if } x \in [0, \gamma_W] \\ \sup_{\alpha \in \mathbb{R}} \{x\alpha - \Lambda_W(\alpha)\} & \text{if } x \in (\gamma_W, \log(m)], \\ +\infty & \text{if } x \not\in [0, \log(m)]. \end{cases}$$  \hfill (6)

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

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

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

for $n \in \{1, \ldots, m^k\}$. 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 $\{U_n\}$ and $\{W_k\}$ also satisfy Assumptions 1 and 2 so that, with the appropriate rate functions, the approximation in equation (8) can be used with $U_n$ or $W_k$ 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 $W_k$ is exponential easier in $k$ to guess on average than one chosen via $U_n$.

III. STATEMENT OF MAIN RESULTS

Assume that the words $\{W_k\}$ are made of i.i.d. letters, defining $p = (p_1, \ldots, p_{m-1})$ by $p_\alpha = P(W_1 = \alpha)$. We shall employ the following short-hand: $h(l) := -\sum a \log a$ for $l = (l_0, \ldots, l_{m-1}) \in [0, 1]^m$, $l_a \geq 0$, $\sum l_a = 1$, so that $H_W = h(p)$, and $D(l||p) := -\sum a \log(p_a/l_a)$. Furthermore, define $l^- \in [0, 1]^m$ and $l^+ \in [0, 1]^m$

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

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

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

$$l^W(\alpha) := \frac{p_1^{(1/(1+\alpha))}}{\sum b \in \Lambda P_b^{(1/(1+\alpha))}} \text{ for all } \alpha \in \Lambda$$  \hfill (11)

$$\eta(\alpha) := -\sum a l^W(\alpha) \log p_a = -\sum a \in \Lambda p_a^{1/(1+\alpha)} \log p_a / \sum b \in \Lambda P_b^{1/(1+\alpha)}$$  \hfill (12)

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 $\{U_n\}$ results. Proofs of the following are deferred to the Appendix.

**Lemma 1:** Assumption 1 holds for $\{U_n\}$ and $\{W_k\}$ with

$$\Lambda_U(\alpha) := \alpha h(l^-),$$

and

$$\Lambda_W(\alpha) := \alpha h(l^+(\alpha)) - D(l^+(\alpha)||p),$$

where

$$l^+(\alpha) = \begin{cases} l^+ & \text{if } \eta(\alpha) \leq -h(p) - \epsilon, \\ l^W(\alpha) & \text{if } \eta(\alpha) \in (-h(p) - \epsilon, h(p) + \epsilon), \\ l^- & \text{if } \eta(\alpha) \geq -h(p) + \epsilon. \end{cases}$$  \hfill (13)
**Lemma 2:** Assumption 2 holds for \( \{U'_k\} \) and \( \{W'_k\} \) with
\[
 g_{U'} = h(l^-) \quad \text{and} \quad g_{W'} = \min \left( -h(p) + \epsilon, \log \max_{\alpha \in \mathbb{K}} p_\alpha \right).
\]
Thus by direct evaluation of the sCGFs at \( \alpha = 1 \),
\[
 \lim_{k \to \infty} \frac{1}{k} \log E \left( G(U'_k) \right) = h(l^-) \quad \text{and} \quad \lim_{k \to \infty} \frac{1}{k} \log E \left( G(W'_k) \right) = \Lambda_{W'}(1).
\]
As the conditions of Theorem 3 [1] are satisfied
\[
 \lim_{k \to \infty} \frac{1}{k} \log E \left( \log \left( G(U'_k) \right) \right) = \Lambda_{U'}(0) = h(l^-) \quad \text{and} \quad \lim_{k \to \infty} \frac{1}{k} \log E \left( \log \left( G(W'_k) \right) \right) = \Lambda_{W'}(0) = h(p),
\]
and we have the approximations
\[
 P(G(U'_k) = u) \approx \frac{1}{n} \exp \left( -k \Lambda_{U'}(k^{-1} \log n) \right) \quad \text{and} \quad P(G(W'_k) = u) \approx \frac{1}{n} \exp \left( -k \Lambda_{W'}(k^{-1} \log n) \right).
\]

**IV. Example**

Consider a binary alphabet \( \mathcal{A} = \{0, 1\} \) 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:
\[
 l_0^- = p_0 - \frac{\epsilon}{\log(p_0) - \log(1 - p_0)}, \quad l_0^+ = p_0 + \frac{\epsilon}{\log(p_0) - \log(1 - p_0)}.
\]
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
\[
 \Lambda_W(\alpha) = \begin{cases} 
 \log(p_0) & \text{if } \alpha < -1, \\
 \alpha h(l^W(\alpha)) - D(l^W(\alpha)||p) & \text{if } \alpha \geq -1,
\end{cases}
\]
\[
 = \begin{cases} 
 \log(p_0) & \text{if } \alpha < -1, \\
 (1 + \alpha) \log \left( \frac{p_0^\alpha}{\alpha^\alpha} + (1 - p_0)^{1-\alpha} \right) & \text{if } \alpha \geq -1,
\end{cases}
\]
From Lemmas [1] and [2] we obtain
\[
 \Lambda_U'(\alpha) = \begin{cases} 
 -h(l^-) & \text{if } \alpha < -1, \\
 \alpha h(l^-) & \text{if } \alpha \geq -1,
\end{cases}
\]
and
\[
 \Lambda_{W'}(\alpha) = \alpha h(l^+(\alpha)) - D(l^+(\alpha)||p),
\]
where \( l^+(\alpha) \) is defined in equation (13) and \( \eta(\alpha) \) defined in equation (12).
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 $\{W_k^*\}$ and uniform $\{U_k\}$ 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 $-x - \Lambda_{W}(x)$ for each of the three processes. That is, using equation (8) and its equivalents to identify the Guesswork growth rate.

**Appendix**

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

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

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

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

The set of all types, those just in the typical set and smooth approximations to those in the typical set are denoted

$$L_k := \{l : \exists w \in \mathbb{A}^k \text{ such that } n_{k}(w, a) = l_a \forall a \in \mathbb{A}\},$$  

$$L_{\epsilon,k} := \{l : \exists w \in T_{\epsilon,k} \text{ such that } n_{k}(w, a) = l_a \forall a \in \mathbb{A}\},$$  

$$L_{\epsilon} := \left\{l : \sum_a l_a \log p_a \in [-h(p) - \epsilon, -h(p) + \epsilon] \right\},$$  

where it can readily seen that $L_{\epsilon,k} \subset L_{\epsilon}$ for all $k$. For $\{U_k\}$ we need the following Lemma.
Lemma 3: The exponential growth rate of the size of the typical set is
\[
\lim_{k \to \infty} \frac{1}{k} \log |T_k^x| = \begin{cases} 
\log m & \text{if } \log m \leq h(p) + \epsilon \\
h(l^-) & \text{otherwise.}
\end{cases}
\]
where \(l^-\) is defined in equation (9).

Proof: For fixed \(k\), by the union bound
\[
\max_{l \in L_{\epsilon,k}} \frac{k!}{\prod_{a \in A}(k_l^a)!} \leq |T_k| \leq (k+1)^m \max_{l \in L_{\epsilon,k}} \frac{k!}{\prod_{a \in A}(k_l^a)!}.
\]
For the logarithmic limit, these two bounds coincide so consider the concave optimization problem
\[
\max_{l \in L_{\epsilon,k}} \frac{k!}{\prod_{a \in A}(k_l^a)!}
\]
We can upper bound this optimization by replacing \(L_{\epsilon,k}\) with the smoother version, its superset \(L_{\epsilon}\). Using Stirling’s bound we have that
\[
\limsup_{k \to \infty} \frac{1}{k} \log \sup_{l \in L_{\epsilon}} \frac{k!}{\prod_{a \in A}(k_l^a)!} \leq \sup_{l \in L_{\epsilon}} \log\left(\frac{\log(m)}{(h_l + \epsilon \geq \log(m))} \text{if } h_l + \epsilon \geq \log(m) \right)
\]
\[
\left(\frac{h_l^-}{h_l} \right) \text{if } h_l + \epsilon < \log(m).
\]
For the lower bound, we need to construct a sequence \(\{l(k)\}\) such that \(l(k) \in L_{\epsilon,k}\) 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
\[
l_a = \begin{cases} 
  \lceil k^{-1}k_l^a \rceil + 1 - \frac{1}{k}k_l^a & \text{if } a = c, \\
  \lfloor k_l^a \rfloor & \text{if } a \neq c.
\end{cases}
\]
Then \(l(k) \in L_{\epsilon,k}\) for all \(k \leq -m \log(p_c)/(2c)\) and \(h_l(k) \to h(l^*)\), as required.

Proof: Proof of Lemma 3 Considering \(\{U_{k}^x\}\) first,
\[
\alpha R_{U^x} \left(\frac{1}{1 + \alpha}\right) = \lim_{k \to \infty} \frac{1}{k} \log |T_k^x| = \alpha h(l^-),
\]
by Lemma 3 To evaluate \(\Lambda_{U^x}(\alpha)\), as for any \(n \in \mathbb{N}\) and \(\alpha > 0\)
\[
\sum_{i=1}^{n} i^\alpha \geq \int_{0}^{\infty} x^\alpha \, dx,
\]
again using Lemma 3 we have
\[
\alpha h(l^-) = \lim_{k \to \infty} \frac{1}{k} \log \left(\frac{1}{1 + \alpha} |T_k^x|\right)^\alpha 
\]
\[
= \lim_{k \to \infty} \frac{1}{k} \log E(e^{\alpha \log G(U_{k}^x)})
\]
\[
= \lim_{k \to \infty} \frac{1}{k} \log \left(\frac{|T_k^x|}{\prod_{i=1}^{n} i^\alpha}\right)
\]
\[
= \lim_{k \to \infty} \frac{1}{k} \log |T_k^x| = \alpha h(l^-),
\]
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 \(\{W_k\}\) into three steps. Step 1 is to show the equivalence of the existence of \(\Lambda_{W^x}(\alpha)\) and \(\alpha R_{W^x}(1/(1 + \alpha))\) for \(\alpha > -1\) with the existence of the following limit
\[
\lim_{k \to \infty} \frac{1}{k} \log \max_{l \in L_{\epsilon,k}} \left\{ N_k(l)^{1+\alpha} \prod_{a \in A} p_{l_a^k} \right\}.
\]
Step 2 then establishes this limit and identifies it. Step 3 shows that \(\Lambda_{W^x}(\alpha)\) is continuous for \(\alpha > -1\). To achieve steps 1 and 2, we adopt and adapt the method of types argument employed in the elotted web-version of [3].

Step 1 Two changes from the bounds of [3] Lemma 5.5 are necessary: the consideration of non-i.i.d. sources by restriction to \(T_k^x\); and the extension of the \(\alpha\) range to include \(\alpha \in (-1, 0]\) from that for \(\alpha \geq 0\) given in that document. Adjusted for conditioning on the typical set we get
\[
\frac{1}{1 + \alpha} \max_{l \in L_{\epsilon,k}} \left\{ N_k(l)^{1+\alpha} \prod_{a \in A} p_{l_a^k} \sum_{w \in T_k^x} P(W_k = w) \right\} 
\]
\[
\leq E(e^{\alpha \log G(W_k)}) \leq (k+1)^m \max_{l \in L_{\epsilon,k}} \left\{ N_k(l)^{1+\alpha} \prod_{a \in A} p_{l_a^k} \sum_{w \in T_k^x} P(W_k = w) \right\}.
\]
The necessary modification of these inequalities for \(\alpha \in (-1, 0]\) gives
\[
\max_{l \in L_{\epsilon,k}} \left\{ N_k(l)^{1+\alpha} \prod_{a \in A} p_{l_a^k} \sum_{w \in T_k^x} P(W_k = w) \right\} 
\]
\[
\leq (k+1)^m \max_{l \in L_{\epsilon,k}} \left\{ N_k(l)^{1+\alpha} \prod_{a \in A} p_{l_a^k} \sum_{w \in T_k^x} P(W_k = w) \right\}.
\]
To show the lower bound holds if \(\alpha \in (-1, 0]\) let
\[
l^* \in \arg\max_{l \in L_{\epsilon,k}} \left\{ N_k(l)^{1+\alpha} \prod_{a \in A} p_{l_a^k} \sum_{w \in T_k^x} P(W_k = w) \right\}.
\]
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, \(\Lambda_{W^x}(\alpha)\) exists and equals it. Similar inequalities provide the same result for \(\alpha R_{W^x}(1/(1 + \alpha))\).

Step 2 The problem has been reduced to establishing the existence of
\[
\lim_{k \to \infty} \frac{1}{k} \log \max_{l \in L_{\epsilon,k}} \left\{ N_k(l)^{1+\alpha} \prod_{a \in A} p_{l_a^k} \right\}
\]
and identifying it. The method of proof is similar to that employed at the start of Lemma 3 for \(\{U_{k}^x\}\): we provide an upper bound for the limsup and then establish a corresponding lower bound.
If $l^{(k)} \rightarrow l$ 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 $L_{k} \subset L_0$ for all $k$,
\[
\limsup_{k \to \infty} \frac{1}{k} \log \max_{l \in L_{k}} \left\{ N_k(l)^{1+\alpha} \prod_{a \in A} p_a^{k_a} \right\} = \sup_{l \in L_0} \left( (1+\alpha) h(l) + \sum_{a} l_a \log p_a \right) = \sup_{l \in L_0} (\alpha h(l) - D(l||p)). \tag{17}
\]
This is a concave optimization problem in $l$ with convex constraints. Not requiring $l \in L_k$, the unconstrained optimizer over all $l$ is attained at $l^W(\alpha)$ defined in equation (17), 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
\[
\limsup_{k \to \infty} \frac{1}{k} \log \max_{l \in L_{k}} \left\{ N_k(l)^{1+\alpha} \prod_{a \in A} p_a^{k_a} \right\} \leq \sup_{l \in L_0} (\alpha h(l^*) - D(l^*||p)) \tag{18}
\]
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 $l^{(k)} \rightarrow l^*(\alpha)$ and $l^{(k)} \in L_{k}$ for all $k$. If $l^*(\alpha) = l^-$, then the sequence used in the proof of Lemma 1 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 $\Lambda_W(\alpha) = \alpha h(l^*(\alpha)) - D(l^*(\alpha)||p)$, with $l^*(\alpha)$ defined in equation (13).
\[
d\Lambda_W(\alpha) = h(l^*(\alpha)) + \Lambda_W(\alpha) \frac{d}{d\alpha} l^*(\alpha).
\]
Thus to establish continuity it suffices to establish continuity of $l^*(\alpha)$ and its derivative, which can be done readily by calculus.

\[\Box\]

**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
\[
\alpha R \left( \frac{1}{1+\alpha} \right) - (1+\alpha) \log m \\
\leq \liminf_{k \to \infty} \frac{1}{k} k P(G(W_k) = 1)^{(1/(1+\alpha))} \\
\leq \limsup_{k \to \infty} \frac{1}{k} \log P(G(W_k) = 1)
\]

Equation (18) holds as $P(G(W_k) = 1) \geq P(G(W_k) = i)$ for all $i \in \{1, \ldots, m\}$. 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) 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**