Guesswork, large deviations and Shannon entropy

Mark M. Christiansen and Ken R. Duffy

***Abstract*—How hard is it to guess a password? Massey showed that a simple function of the Shannon entropy of the distribution from which the password is selected is a lower bound on the expected number of guesses, but one which is not tight in general. In a series of subsequent papers under ever less restrictive stochastic assumptions, an asymptotic relationship as password length grows between scaled moments of the guesswork and specific Rényi entropy was identified.**

Here we show that, when appropriately scaled, as the password length grows the logarithm of the guesswork satisfies a Large Deviation Principle (LDP), providing direct estimates of the guesswork distribution when passwords are long. The rate function governing the LDP possesses a specific, restrictive form that encapsulates underlying structure in the nature of guesswork. Returning to Massey's original observation, a corollary to the LDP shows that expectation of the logarithm of the guesswork is the specific Shannon entropy of the password selection process.

***Index Terms*—Guesswork, Rényi Entropy, Shannon Entropy, Large Deviations**

I. Introduction

There are several distinct, quantitative notions of secrecy. For example, perfect secrecy, where an adversary cannot compute anything regarding the secret, has long been studied, e.g. [1], [2], [3], while weak security is a more recent notion that places the less stringent requirement of an upper bound on the rate at which an adversary can glean information, e.g. [4], [5]. Methodologies based on computationally security, where discovery is possible but is anticipated to take a long time, are used extensively in cryptosystems [6] and are particularly widely deployed. Here we are concerned with a facet of this latter notion.

Assume a message is hidden within a large collection of possibilities. If one has *a priori* knowledge of the likelihood of each message being the hidden one and one can inquire about each message in turn, what is the distribution of how many questions it will take to correctly guess the secret? The common example that motivates the terminology of this article is that of a typical password entry system based on a cryptographic hash function, where an adversary can ask about the veracity of each potential password in turn. Our results, however, also have ramifications for the properties of new secrecy notions [7].

If a password, W, is chosen at random from a finite set $\mathbb{A} = \{1, \dots, m\}$, how hard is it to guess W? If $\{P(W = w)\}$ is known, then an optimal strategy is to guess passwords in decreasing order of probability. Let $G(w)$ denote the number of attempts required before correctly guessing $w \in \mathbb{A}$, called w's guesswork. Massey [8] proved that a simple function of the Shannon entropy of W is a lower bound on the expected guesswork, $E(G(W))$, and that no general upper bound exists. This raised serious questions about the appropriateness of Shannon entropy as a measure of complexity of a distribution with regards guesswork. As a corollary to stronger results, in this article we prove a large password relationship, first suggested in [9] and [10], between the expectation of the logarithm of the guesswork and specific Shannon entropy.

Arikan [11] introduced an asymptotic regime for studying this problem by considering a sequence of passwords, $\{W_k\}$, with W_k chosen from $\mathbb{A}^k$ with i.i.d. letters. Again guessing potential passwords in decreasing order of probability for each k, he related the asymptotic fractional moments of the guesswork to the Rényi entropy of a single letter,

$$\lim_{k\to\infty} \frac{1}{k} \log E(G(W_k)^\alpha) = (1+\alpha) \log \sum_{w\in\mathbb{A}} P(W_1 = w)^{\frac{1}{1+\alpha}}$$

for $\alpha > 0$, where the right hand side is α times the Rényi entropy of W_1 evaluated at $1/(1+\alpha)$. This result was subsequently extended by Malone and Sullivan [12] to word sequences with letters chosen by a Markov process and, further still, by Pfister and Sullivan [13] to sofic shifts whose shift space satisfies an entropy condition and whose marginals possess a limit property. Recently, using a distinct approach Hanawal and Sundaresan [14] provided alternate sufficient conditions for the existence of the limit. In all cases, the limit is identified in terms of the specific Rényi entropy

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

where $R_k(\alpha)$ is the Rényi entropy of W_k

$$R_k(\alpha) = \frac{1}{1-\alpha} \log \left(\sum_{w\in\mathbb{A}^k} P(W_k = w)^\alpha \right).$$

Here we shall assume the existence of the limit on the left hand side of equation (1) for all $\alpha > -1$, its equality with α times specific Rényi entropy, its differentiability with respect to α in that range and a regularity condition on the probability of the most-likely word, that $\lim k^{-1} \log P(G(W_k) = 1)$ exists. From this, Theorem 3 deduces that the sequence $\{k^{-1} \log G(W_k)\}$ satisfies a Large Deviation Principle (LDP) (e.g. [15]) with a rate function Λ^* that must possess a specific form that will have a physical interpretation: Λ^* is continuous where finite, can be linear on an interval $[0, a]$, for some $a \in [0, \log(m)]$, and then must be strictly convex while finite on $[a, \log(m)]$.

In contrast to earlier results, Corollary 4 to the LDP gives direct estimates on the guesswork distribution $P(G(W_k) = n)$

M. Christiansen and K. R. Duffy are with the Hamilton Institute, National University of Ireland Maynooth. This material is based upon works supported by the Science Foundation Ireland under Grant No. 11/PI/1177 and by the Irish Higher Educational Authority (HEA) PRTLI Network Mathematics Grant.

for large k, suggesting the approximation

$$P(G(W_k) = n) \approx \frac{1}{n} \exp(-k\Lambda^*(k^{-1} \log n)). \qquad (2)$$

As this calculation only involves the determination of Λ^*, to approximately calculate the probability of the n^{th} most likely word in words of length k one does not have to identify the word itself, which would be computationally cumbersome, particularly for non-i.i.d. word sources.

Corollary 5 to the LDP recovers a rôle for Shannon entropy in the asymptotic analysis of guesswork. It shows that the scaled expectation of the logarithm of the guesswork converges to specific Shannon entropy

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

where

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

II. A Large Deviation Principle

Consider the sequence of random variables $\{k^{-1} \log G(W_k)\}$. Our starting point is the observation that the left hand side of (1) is the scaled Cumulant Generating Function (sCGF) of this sequence:

$$\Lambda(\alpha) := \lim_{k\to\infty} \frac{1}{k} \log E\left(e^{\alpha \log G(W_k)}\right),$$

which is shown to exist for $\alpha > 0$ in [11][12] and for $\alpha > -1$ in [13] for a broad class of stationary processes.

Assumption 1: For $\alpha > -1$, the sCGF $\Lambda(\alpha)$ exists, is equal to α times the specific Rényi entropy, and has a continuous derivative in that range.

We also assume the following regularity condition on the probability of the most likely word.

Assumption 2: The limit

$$g_1 = \lim_{k\to\infty} \frac{1}{k} \log P(G(W_k) = 1) \qquad (3)$$

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

Assumptions 1 and 2 hold, for example, for all irreducible Markov chains. In this case, assumption 1 can be established by straight-forward extension of the range of α in the results in [12], while Assumption 2 can be shown to hold by elementary consideration of cycles.

We first show that the sCGF exists everywhere.

Lemma 1 (Existence of the sCGF): Under assumptions 1 and 2, for all $\alpha \leq -1$

$$\Lambda(\alpha) = \lim_{k\to\infty} \frac{1}{k} \log P(G(W_k) = 1) = g_1 = \lim_{\beta\downarrow -1} \Lambda(\beta).$$

Proof: Let $\alpha \leq -1$ and note that

$$\log P(G(W_k) = 1) \leq \log \sum_{i=1}^{m^k} P(G(W_k) = i) i^\alpha$$
$$= \log E\left(e^{\alpha \log G(W_k)}\right) \leq \log P(G(W_k) = 1) + \log \sum_{i=1}^{m^k} i^\alpha.$$

Taking $\liminf_{k\to\infty} k^{-1}$ with the first inequality and $\limsup_{k\to\infty} k^{-1}$ with the second while using the Principle of the Largest Term [15, Lemma 1.2.15] in conjunction with usual estimates on the harmonic series if $\alpha = -1$ and boundedness of the sum if $\alpha < -1$, we have that

$$\lim_{k\to\infty} \frac{1}{k} \log E(e^{\alpha \log G(W_k)}) = \lim_{k\to\infty} \frac{1}{k} \log P(G(W_k) = 1)$$

for all $\alpha \leq -1$.

As Λ is the limit of a sequence of convex functions and is finite everywhere, it is continuous and therefore $\lim_{\beta\downarrow -1} \Lambda(\beta) = \Lambda(-1)$. ■

Thus the sCGF Λ exists and is finite for all α, with a potential discontinuity in its derivative at $\alpha = -1$. This discontinuity, when it exists, will have a bearing on the nature of the rate function governing the LDP for $\{k^{-1} \log G(W_k)\}$. Indeed, the following quantity will play a significant rôle in our results:

$$\gamma := \lim_{\alpha\downarrow -1} \frac{d}{d\alpha} \Lambda(\alpha). \qquad (4)$$

The derivative on the right hand side of equation (4) has the interpretation of a tilted measure. As $\alpha \downarrow -1$ this measure will, in an appropriate sense, converge to the uniform measure on the set of words with asymptotically maximal probability. In particular, we will prove that the number of words with approximately equally highest probability is close to $\exp(k\gamma)$. In the special case where the $\{W_k\}$ are constructed of i.i.d. letters this is exactly true and the veracity of the following Lemma can be verified directly.

Lemma 2 (The number of most likely words): If $\{W_k\}$ are constructed of i.i.d. letters, then

$$\gamma = \lim_{\alpha\downarrow -1} \frac{d}{d\alpha} \alpha R_1((1+\alpha)^{-1})$$
$$= \log |\{w : P(W_1 = w) = P(G(W_1) = 1)\}|,$$

where $|\cdot|$ indicates the number of elements in the set.

This i.i.d. result doesn't extend directly to the non-i.i.d. case and in general Lemma 2 can only be used to establish a lower bound on γ defined in equation (4):

$$\gamma \geq \limsup_{k\to\infty} \frac{1}{k} \lim_{\alpha\downarrow -1} \frac{d}{d\alpha} \alpha R_k\left(\frac{1}{1+\alpha}\right), \qquad (5)$$

e.g [16, Theorem 24.5]. This lower bound can be loose, as can be seen with the following example. Consider the sequence of distributions for some $\epsilon > 0$

$$P(W_k = i) = \begin{cases} m^{-k}(1+\epsilon) & \text{if } i = 1 \\ m^{-k}(1 - \epsilon(m^k - 1)^{-1})) & \text{otherwise.} \end{cases}$$

For each fixed k there is one most likely word and we have $\log(1) = 0$ on the right hand side of equation (5) by Lemma 2. The left hand side, however, gives $\log(m)$. Regardless, this intuition guides our understanding of γ, but the formal statement of it approximately capturing the number of most likely words will transpire to be

$$g_1 = \lim_{k\to\infty} \frac{1}{k} \log \inf_{\{w: G(w) < \exp(k\gamma)\}} P(W_k = w),$$

where g_1 is defined in equation (3).

We define the candidate rate function as the Legendre-Fenchel transform of the sCGF

$$\Lambda^*(x) := \sup_{\alpha\in\mathbb{R}}\{x\alpha - \Lambda(\alpha)\}$$
$$= \begin{cases} -x - g_1 & \text{if } x \in [0,\gamma] \\ \sup_{\alpha\in\mathbb{R}}\{x\alpha-\Lambda(\alpha)\} & \text{if } x \in (\gamma, \log(m)], \\ +\infty & \text{if } x \notin [0,\log(m)]. \end{cases} \quad (6)$$

The LDP cannot be proved directly by Baldi's version of the Gärtner-Ellis theorem [17][15, Theorem 4.5.20] as Λ^* does not have exposing hyper-planes for $x \in [0,\gamma]$. Instead we use a combination of that theorem with the methodology described in detail in [18] where, as our random variables are bounded $0 \le k^{-1}\log G(W_k) \le \log(m)$, in order to prove the LDP it suffices to show that the following exist in $[0,\infty]$ for all $x \in [0,\log m]$ and equals $-\Lambda^*(x)$:

$$\lim_{\epsilon\downarrow 0}\liminf_{k\to\infty}\frac{1}{k}\log P\left(\frac{1}{k}\log(G(W_k)) \in B_\epsilon(x)\right)$$
$$= \lim_{\epsilon\downarrow 0}\limsup_{k\to\infty}\frac{1}{k}\log P\left(\frac{1}{k}\log(G(W_k)) \in B_\epsilon(x)\right), \quad (7)$$

where $B_\epsilon(x) = (x-\epsilon, x+\epsilon)$.

Theorem 3 (The large deviations of guesswork): Under assumptions 1 and 2, the sequence $\{k^{-1}\log G(W_k)\}$ satisfies a LDP with rate function Λ^*.

Proof: To establish (7) we have separate arguments depending on x. We divide $[0,\log(m)]$ into two parts: $[0,\gamma]$ and $(\gamma,\log(m)]$. Baldi's upper bound holds for any $x \in [0,\log(m)]$. Baldi's lower bound applies for any $x \in (\gamma,\log(m)]$ as Λ^* is continuous and, as $\Lambda(\alpha)$ has a continuous derivative for $\alpha > -1$, it only has a finite number of points without exposing hyper-planes in that region. For $x \in [0,\gamma]$, however, we need an alternate lower bound.

Consider $x \in [0,\gamma]$ and define the sets

$$K_k(x,\epsilon) := \left\{w \in \mathbb{A}^k : k^{-1}\log G(w) \in B_\epsilon(x)\right\},$$

letting $|K_k(x,\epsilon)|$ denote the number of elements in each set. We have the bound

$$|K_k(x,\epsilon)| \inf_{w\in K_k(x,\epsilon)} P(W_k = w)$$
$$\le P\left(\frac{1}{k}\log G(W_k) \in B_\epsilon(x)\right). \quad (8)$$

As $\lfloor e^{k(x-\epsilon)}\rfloor \le |K_k(x,\epsilon)| \le \lceil e^{k(x+\epsilon)}\rceil$, we have that

$$x = \lim_{\epsilon\to 0}\lim_{k\to\infty}\frac{1}{k}\log|K_k(x,\epsilon)|. \quad (9)$$

By either the complementary upper bound to equation (8) or by Baldi's upper bound, we have that

$$\lim_{\epsilon\downarrow 0}\limsup_{k\to\infty}\frac{1}{k}\log P\left(\frac{1}{k}\log G(W_k) \in B_\epsilon(x)\right) \le x + g_1.$$

Thus to complete the argument, for the complementary lower bound it suffices to show that for any $x \in [0,\gamma]$

$$\lim_{\epsilon\downarrow 0}\liminf_{k\to\infty}\inf_{w\in K_k(x,\epsilon)}\frac{1}{k}\log P(W_k = w) \ge g_1.$$

If $\Lambda^*(x) < \infty$ for some $x > \gamma$, then for $\epsilon > 0$ sufficiently small let x^* be such that $\Lambda^*(x^*) < \infty$ and $x^* - \epsilon > \max(\gamma, x+\epsilon)$. Then by Baldi's lower bound, which applies as $x^* \in (\gamma,\log(m)]$, we have

$$-\inf_{y\in B_\epsilon(x^*)}\Lambda^*(y) \le \liminf_{k\to\infty}\frac{1}{k}\log P\left(\frac{1}{k}\log G(W_k) \in B_\epsilon(x^*)\right).$$

Now

$$P\left(\frac{1}{k}\log G(W_k) \in B_\epsilon(x^*)\right)$$
$$\le |K_k(x^*,\epsilon)| \sup_{w\in K_k(x^*,\epsilon)} P(W_k = w)$$
$$\le |K_k(x^*,\epsilon)| \inf_{w\in K_k(x,\epsilon)} P(W_k = w),$$

where in the last line we have used the monotonicity of guesswork and the fact that $x^* - \epsilon > x + \epsilon$. Taking lower limits and using equation (9) with $|K_k(x^*,\epsilon)|$, we have that

$$-\inf_{y\in B_\epsilon(x^*)}\Lambda^*(y) \le x^* + \liminf_{k\to\infty}\inf_{w\in K_k(x,\epsilon)}\frac{1}{k}\log P(W_k = w)$$

for all such x^*, x. Taking limits as $\epsilon \downarrow 0$ and then limits as $x^* \downarrow \gamma$ we have

$$-\lim_{x^*\downarrow\gamma}\Lambda^*(x^*) \le \gamma + \lim_{\epsilon\downarrow 0}\liminf_{k\to\infty}\inf_{w\in K_k(x,\epsilon)}\frac{1}{k}\log P(W_k = w),$$

but $\lim_{x^*\downarrow\gamma}\Lambda^*(x^*) = -\gamma - g_1$ so that

$$\lim_{\epsilon\downarrow 0}\liminf_{k\to\infty}\inf_{w\in K_k(x,\epsilon)}\frac{1}{k}\log P(W_k = w) = g_1,$$

as required.

Only one case remains. If $\Lambda^*(x) = \infty$ for all $x > \gamma$, then we require an alternative argument to ensure that

$$\liminf_{k\to\infty}\inf_{w\in K_k(x,\epsilon)}\frac{1}{k}\log P(W_k = w) \ge g_1.$$

This situation happens if, in the limit, the distribution of words is near uniform on the set of all words with positive probability. As $\Lambda(0) = 0$, using equation (6) we have that $g_1 = -\gamma$. Let $x < \gamma$ and consider

$$l = \limsup_{k\to\infty}\sup_{w\in K_k(x+2\epsilon,\epsilon)}\frac{1}{k}\log P(W_k = w)$$
$$\le \liminf_{k\to\infty}\inf_{w\in K_k(x,\epsilon)}\frac{1}{k}\log P(W_k = w).$$

We shall assume that $l < g_1$ and show this results in a contradiction. Let $\epsilon > 0$, then there exists N_ϵ such that for all $k \ge N_\epsilon$, $P(G(W_k) = i) \le \exp(k(g_1+\epsilon))$, for all $i \in \{1,\ldots,m^k\}$, $P(G(W_k) = i) \le \exp(k(l+\epsilon))$, for all $i \in \{\exp(k(x+\epsilon)),\ldots,m^k\}$ and $P(G(W_k) \ge \exp(k(\gamma+\epsilon))) \le e^{-k/\epsilon}$. Let $0 < \epsilon < \min(g_1 - l, \gamma - x)/2$ be given, then, using a potentially gross overestimate that suffices for our purposes, we have that

$$\sum_{w\in\mathbb{A}^k} P(W_k = w) = \sum_{i=1}^{m^k} P(G(W_k) = i)$$
$$\le e^{k(x+\epsilon)}e^{k(g_1+\epsilon)} + e^{k(\gamma+\epsilon)}e^{k(l+\epsilon)} + e^{-k/\epsilon}$$

for all $k > N_\epsilon$, but as $l < g_1 = -\gamma$ this is strictly less than 1 for k sufficiently large and thus $l = g_1$. Finally, for $x = \gamma$, and $\epsilon > 0$, note that we can decompose $[0, \log(m)]$ into three parts, $[0, \gamma - \epsilon] \cup (\gamma - \epsilon, \gamma + \epsilon) \cup [\gamma + \epsilon, \log(m)]$, where the scaled probability of the guesswork being in either the first or last set is decaying, but

$$0 = \lim_{k\to\infty} \frac{1}{k} \log P\left(\frac{1}{k} \log G(W_k) \in [0, \log(m)]\right)$$

and so the result follows from an application of the principle of the largest term.

Thus for any $x \in [0, \log(m)]$,

$$\begin{aligned}
&\lim_{\epsilon\downarrow 0} \liminf_{k\to\infty} \frac{1}{k} \log P\left(\frac{1}{k} \log(G(W_k)) \in B_\epsilon(x)\right) \\
&= \lim_{\epsilon\downarrow 0} \limsup_{k\to\infty} \frac{1}{k} \log P\left(\frac{1}{k} \log(G(W_k)) \in B_\epsilon(x)\right) \\
&= -\Lambda^*(x)
\end{aligned}$$

and the LDP is proved. ■

In establishing the LDP, we have shown that any rate function that governs such an LDP must have the form of a straight line in $[0, \gamma]$ followed by a strictly convex function. The initial straight line comes from all words that are, in an asymptotic sense, of greatest likelihood.

While the LDP is for the sequence $\{k^{-1} \log G(W_k)\}$, it can be used to develop the more valuable direct estimate of the distribution of each $G(W_k)$ found in equation (2). The next corollary provides a rigorous statement, but an intuitive, non-rigorous argument for understanding the result therein is that from the LDP we have the approximation that for large k

$$dP\left(\frac{1}{k} \log G(W_k) = x\right) \approx \exp(-k\Lambda^*(x))dx.$$

As for large k the distribution of $k^{-1} \log G(W_k)$ and $G(W_k)/k$ are ever closer to having densities, using the change of variables formula gives

$$\begin{aligned}
dP\left(\frac{1}{k} G(W_k) = x\right) &= \frac{1}{kx} dP\left(\frac{1}{k} \log G(W_k) = x\right) \\
&\approx \frac{1}{kx} \exp\left(-k\Lambda^*\left(\frac{1}{k} \log(kx)\right)\right) dx.
\end{aligned}$$

Finally, the substitution $kx = n$ gives the approximation in equation (2). To make this heuristic precise requires distinct means, explained in the following corollary.

Corollary 4 (Direct estimates on guesswork): Recall the definition

$$K_k(x, \epsilon) := \left\{w \in \mathbb{A}^k : k^{-1} \log G(w) \in B_\epsilon(x)\right\}.$$

For any $x \in [0, \log(m)]$ we have

$$\begin{aligned}
&\lim_{\epsilon\downarrow 0} \liminf_{k\to\infty} \frac{1}{k} \log \inf_{w \in K_k(x,\epsilon)} P(W_k = w) \\
&= \lim_{\epsilon\downarrow 0} \limsup_{k\to\infty} \frac{1}{k} \log \sup_{w \in K_k(x,\epsilon)} P(W_k = w) \\
&= -\left(x + \Lambda^*(x)\right).
\end{aligned}$$

Proof: We show how to prove the upper bound as the lower bound follows using analogous arguments, as do the edge cases. Let $x \in (0, \log(m))$ and $\epsilon > 0$ be given. Using the monotonicity of guesswork

$$\begin{aligned}
&\limsup_{k\to\infty} \frac{1}{k} \log \sup_{w \in K_k(x,\epsilon)} P(W_k = w) \\
&\leq \liminf_{k\to\infty} \frac{1}{k} \log \inf_{w \in K_k(x-2\epsilon,\epsilon)} P(W_k = w).
\end{aligned}$$

Using the estimate found in Theorem 3 and the LDP provides an upper bound on the latter:

$$\begin{aligned}
&(x - 3\epsilon) + \liminf_{k\to\infty} \frac{1}{k} \log \inf_{w \in K_k(x-2\epsilon,\epsilon)} P(W_k = w) \\
&\leq \liminf_{k\to\infty} \frac{1}{k} \log P\left(\frac{1}{k} \log(G(W_k)) \in B_\epsilon(x - 2\epsilon)\right) \\
&\leq \limsup_{k\to\infty} \frac{1}{k} \log P\left(\frac{1}{k} \log(G(W_k)) \in [x - 3\epsilon, x - \epsilon]\right) \\
&\leq -\inf_{x \in [x-3\epsilon, x-\epsilon]} \Lambda^*(x).
\end{aligned}$$

Thus

$$\begin{aligned}
&\limsup_{k\to\infty} \frac{1}{k} \log \sup_{w \in K_k(x,\epsilon)} P(W_k = w) \\
&\leq -x + 3\epsilon - \inf_{x \in [x-3\epsilon, x-\epsilon]} \Lambda^*(x).
\end{aligned}$$

As Λ^* is convex, it is continuous where finite, and thus the upper-bound follows taking $\epsilon \downarrow 0$. ■

Unpeeling limits, this corollary shows that when k is large the probability of the n^{th} most likely word is approximately $1/n \exp(-k\Lambda^*(k^{-1} \log n))$, without the need to identify the word itself. This justifies the approximation in equation (2), whose complexity of evaluation does not depend on k. We demonstrate its merit by example in Section III.

Before that, as a corollary to the LDP we find the following rôle for the specific Shannon entropy. Thus, although Massey established that for a given word length a simple function of the Shannon entropy is only a lower bound on the guesswork, for growing password length the specific Shannon entropy determines the linear growth rate of the expectation of the logarithm of guesswork (c.f [9] and [10]).

Corollary 5 (Shannon entropy and guesswork): Under assumptions 1 and 2,

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

the specific Shannon entropy.

Proof: As both $\Lambda(\alpha)$ and $\alpha R_k((1+\alpha)^{-1})$ are finite and differentiable in a neighborhood of 0, by [16, Theorem 25.7]

$$\Lambda'(0) = \lim_{k\to\infty} \frac{1}{k} \frac{d}{d\alpha} \alpha R_k((1+\alpha)^{-1})|_{\alpha=0} = \lim_{k\to\infty} \frac{1}{k} H(W_k).$$

Note that $\Lambda^*(x) = 0$ if and only if $x = \Lambda'(0) = \lim k^{-1} H(W_k)$. Thus the weak law then follows by concentration of measure, e.g. [19]. ■

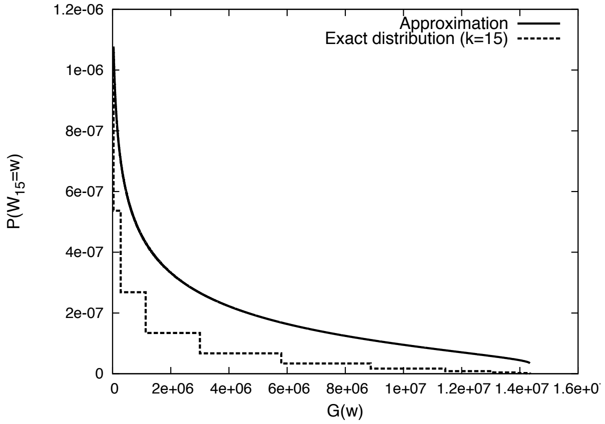

Fig. 1. Illustration of Corollary 4. Words constructed from i.i.d letters with $P(W_1 = 1) = 0.4, P(W_1 = 2) = 0.4, P(W_1 = 3) = 0.2$. For $k = 15$ comparison of the probability of n^{th} most likely word and the approximation $1/n \exp(-k\Lambda^*(k^{-1} \log n))$ versus $n \in \{1, \ldots, 3^{15}\}$.

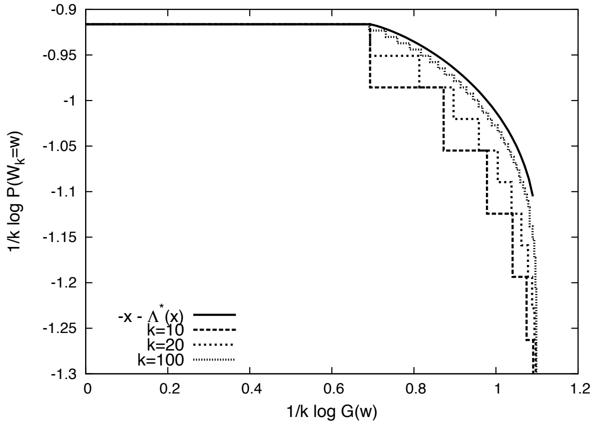

Fig. 2. Illustration of Corollary 4. Words constructed from i.i.d letters with $P(W_1 = 1) = 0.4, P(W_1 = 2) = 0.4, P(W_1 = 3) = 0.2$. For $k = 10, 20$ and 100, comparison of k^{-1} times the logarithm of the probability of n^{th} most likely word versus k^{-1} times the logarithm of n, as well as the approximation $-x - \Lambda^*(x)$ versus x.

III. Examples

I.i.d letters.

Assume words are constructed of i.i.d. letters. Let W_1 take values in $\mathbb{A} = \{1, \ldots, m\}$ and assume $P(W_1 = i) \geq P(W_1 = j)$ if $i \leq j$. Then from [11], [13] and Lemma 1 we have that

$$\Lambda(\alpha) = \begin{cases} (1+\alpha) \log \sum_{w \in \mathbb{A}} P(W_1 = w)^{1/(1+\alpha)} & \text{if } \alpha > -1 \\ \log P(W_1 = 1) & \text{if } \alpha \leq -1. \end{cases}$$

From Lemma 2 we have that

$$\gamma = \lim_{\alpha \downarrow -1} \Lambda'(\alpha) \in \{0, \log(2), \ldots, \log(m)\}$$

and no other values are possible. Unless the distribution of W_1 is uniform, $\Lambda^*(x)$ does not have a closed form for all x, but is readily calculated numerically. With $|\mathbb{A}| = 3$ and $k = 15$, Figure 1 compares the exact distribution $P(W_k = w)$ versus $G(w)$ with the approximation found in equation (2). As there are $3^{15} \approx 1.4$ million words, the likelihood of any one word is tiny, but the quality of the approximation can clearly be seen. Rescaling the guesswork and probabilities to make them comparable for distinct k, Figure 2 illustrates the quality of the approximation as k grows. By $k = 100$ there are $3^{100} \approx 5.1$ times 10^{47} words and the underlying combinatorial complexities of the explicit calculation become immense, yet the complexity of calculating the approximation has not increased.

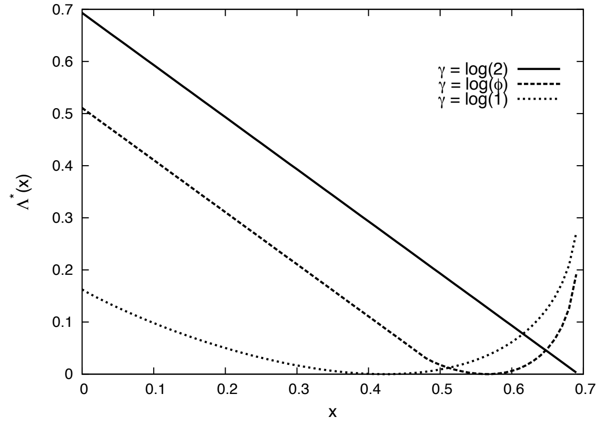

Fig. 3. Illustration of rate functions in Theorem 3. Words constructed from Markov letters on $|\mathbb{A}| = 2$. Three rate functions illustrating only values of γ possible, $\log(1)$, $\log(\phi) \approx 0.48$ and $\log(2)$, from Lemma 6.

Markovian letters.

As an example of words constructed of correlated letters, consider $\{W_k\}$ where the letters are chosen via a process a Markov chain with transition matrix P and some initial distribution on $|\mathbb{A}| = 2$. Define the matrix P_α by $(P_\alpha)_{i,j} = p_{i,j}^{1/(1+\alpha)}$, then by [12], [13] and Lemma 1 we have that

$$\Lambda(\alpha) = \begin{cases} (1+\alpha) \log \rho(P_\alpha) & \text{if } \alpha > -1 \\ \log \max(p_{1,1}, p_{2,2}, \sqrt{p_{1,2}\, p_{2,1}}) & \text{if } \alpha \leq -1, \end{cases}$$

where ρ is the spectral radius operator. In the two letter alphabet case, with $\beta = 1/(1+\alpha)$ we have that $\rho(P_{(1-\beta)/\beta})$ equals

$$\frac{p_{1,1}^\beta + p_{2,2}^\beta}{2} + \frac{\sqrt{(p_{1,1}^\beta - p_{2,2}^\beta)^2 + 4(1-p_{2,2})^\beta (1-p_{1,1})^\beta}}{2}.$$

As with the i.i.d. letters example, apart from in special cases the rate function Λ^* cannot be calculated in closed form, but is readily evaluated numerically. Regardless, we have the following, perhaps surprising, result on the exponential rate of growth of the size of the set of almost most likely words.

Lemma 6 (The Golden Ratio and Markovian letters): For $\{W_k\}$ constructed of Markovian letters with $|\mathbb{A}| = 2$,

$$\gamma = \lim_{\alpha \downarrow -1} \Lambda'(\alpha) \in \{0, \log(\phi), \log(2)\},$$

where $\phi = (1 + \sqrt{5})/2$ is the Golden Ratio, and no other values are possible.

This lemma can be proved by directly evaluating the derivative of $\Lambda(\alpha)$ with respect to α. Note that here $\exp(k\gamma)$ definitely only describes the number of words of equal highest likelihood when k is large as the initial distribution of the Markov chain plays no rôle in γ's evaluation.

The case where $\gamma = \log(2)$ occurs when $p_{1,1} = p_{2,2} = 1/2$. The most interesting case is when there are approximately ϕ^k approximately equally most likely words. This occurs if $p_{1,1} = \sqrt{p_{1,2}p_{2,1}} > p_{2,2}$. For large k, words of near-maximal probability have the form of a sequence of 1s, where a 2 can be inserted anywhere so long as there is a 1 between it and any other 2s. A further sub-exponential number of aberrations are allowed in any given sequence and the starting distribution is ultimately irrelevant. For example, with an equiprobable initial distribution and $k = 4$ there are 8 most likely words (1111, 1112, 1121, 1211, 1212, 2111, 2121, 2112) and $\phi^4 \approx 6.86$. Note that the golden ratio also appears in the analysis of the trapdoor channel [20], but there it is directly as a result of the appearance of the Fibonacci sequence.

Figure 3 gives plots of $\Lambda^*(x)$ versus x illustrating the full range of possible shapes that rate functions can take: linear, linear then strictly convex, or strictly convex, based on the transition matrices

$$\begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}, \begin{pmatrix} 0.6 & 0.4 \\ 0.9 & 0.1 \end{pmatrix} \text{ and } \begin{pmatrix} 0.85 & 0.15 \\ 0.15 & 0.85 \end{pmatrix}$$

respectively.

IV. Concluding remarks

Motivated by widely used computationally secure cryptosystems, in this article we consider the problem of guessing a password given probabilistic knowledge of the underlying word distribution. Building on earlier work that established limiting results for the expected fractional moments of guesswork [11], [12], [13], [14], equation (2) provides a direct estimate of the guesswork distribution. As password selection is known to be non-uniform [21], this result can be used either by an adversary to determine how hard a secret will be to hack or by a system designer to ensure password lengths are long enough to provide a probabilistically strong guarantee of secrecy. These results have ramifications for the properties of forthcoming proposals of computationally secure cryptosystems based on list-decoding [7].

V. Acknowledgment

The authors would like to thank the anonymous reviewers for their careful reading of the paper, and Muriel Médard (MIT) for informative discussions on secrecy and security.

References

[1] C. E. Shannon, “Communication theory of secrecy systems,” *Bell System Technical Journal*, vol. 28, no. 4, pp. 656–715, 1949.

[2] A. D. Wyner, “The wire-tap channel,” *Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975.

[3] S. Gollakota, H. Hassanieh, B. Ransford, D. Katabi, and K. Fu, “They can hear your heartbeats: non-invasive security for implantable medical devices,” in *Proc. ACM SIGCOMM*, 2011, pp. 2–13.

[4] Y. Liang, H. V. Poor, and S. Shamai, “Information theoretic security,” *Found. Trends Commun. Inf. Theory*, pp. 355–580, 2009.

[5] M. Bloch and J. Barros, *Physical-Layer Security: From Information Theory to Security Engineering*. Cambridge University Press, 2011.

[6] J. Katz and Y. Lindell, *Introduction to Modern Cryptography: Principles and Protocols*. Chapman & Hall, CRC Press, 2007.

[7] F. du Pin Calmon, M. Médard, L. Zegler, J. Barros, M. M. Christiansen, and K. R. Duffy, “Lists that are smaller than their parts: A coding approach to tunable secrecy,” in *Proc.* 50^{th} *Allerton Conference*, 2012.

[8] J. L. Massey, “Guessing and entropy,” *Proc. IEEE Int. Symp. Inf. Theory*, pp. 204–204, 1994.

[9] E. Arikan and N. Merhav, “Guessing subject to distortion,” *IEEE Trans. Inf. Theory*, vol. 44, pp. 1041–1056, 1998.

[10] R. Sundaresan, “Guessing based on length functions,” in *Proc. 2007 International Symp. on Inf. Th.*, 2007.

[11] E. Arikan, “An inequality on guessing and its application to sequential decoding,” *IEEE Trans, Inf. Theory*, vol. 42, pp. 99–105, 1996.

[12] D. Malone and W. G. Sullivan, “Guesswork and entropy,” *IEEE Trans. Inf. Theory*, vol. 50, no. 4, pp. 525–526, 2004.

[13] C.-E. Pfister and W. G. Sullivan, “Rényi entropy, guesswork moments and large deviations,” *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2794–2800, 2004.

[14] M. K. Hanawal and R. Sundaresan, “Guessing revisited: A large deviations approach,” *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 70–78, 2011.

[15] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Springer, 2009.

[16] R. T. Rockafellar, *Convex analysis*. Princeton University Press, 1970.

[17] P. Baldi, “Large deviations and stochastic homogenization,” *Ann. Mat. Pura Appl. (4)*, vol. 151, pp. 161–177, 1988.

[18] J. T. Lewis and C.-E. Pfister, “Thermodynamic probability theory: some aspects of large deviations,” *Russian Math. Surveys*, vol. 50, no. 2, pp. 279–317, 1995.

[19] J. T. Lewis, C.-E. Pfister, and W. G. Sullivan, “Entropy, concentration of probability and conditional limit theorems,” *Markov Process. Related Fields*, vol. 1, no. 3, pp. 319–386, 1995.

[20] H. H. Permuter, P. Cuff, B. V. Roy, and T. Weissman, “Capacity of the trapdoor channel with feedback,” *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3150–3165, 2008.

[21] D. Malone and K. Maher, “Investigating the distribution of password choices,” in *WWW*, 2012, pp. 301–310.