Guesswork and Entropy

David Malone and Wayne Sullivan

15 March 2002

Abstract

Entropy is often considered as a measure of uncertainty. It is commonly believed that entropy is a good measure of how many guesses it will take to correctly guess a single value output by a source. This belief is not well founded. We summarize some work in this area, explore how this belief may have arisen via the asymptotic equipartition property. This leads us to a large deviations type estimate of the guesswork for symbols forming a Markov chain.

1 Introduction

Shannon entropy

$$h(p) := -\sum_i p_i \log_2 p_i \tag{1}$$

is often considered as a measure of the number of bits of uncertainty associated with a source which produces symbol i with probability p_i. This use, which began with Shannon's work on Information Theory, has become widespread in cryptology where it is often used outside its original context. For example, the discussion of key-guessing attacks in [11] says:

> We can measure how bad a key distribution is by calculating its entropy. This number E is the number of "real bits of information" of the key: a cryptanalyst will typically happen across the key within 2^E guesses. E is defined as the sum of $-p_K \log_2 p_K$, where p_K is the probability of key K.

Similar inferences are made in Section 17.14 of [10] while discussing *Biases and Correlations* of random sequence generators. The quality of the random data harvested by the Yarrow pseudo-random number generator is also referred to as entropy [5]. The *Entropy Gathering Daemon* [12], a substitute for the Unix `/dev/random` device, speaks for itself in this respect.

In all these cases entropy is being used to measure 'guessability'. There are many possible criteria for specifying 'guessability'. The one we consider here is the expected number of guesses required to get the correct answer. There are various strategies which can be used for guessing. Commonly know are *brute force attacks* where all symbols are guessed in no particular order, and *dictionary attacks* where symbols which are deemed more probable are guessed first. Well known software such as *Crack* [8] uses a dictionary attack.

The guessing strategy we consider is the optimal one, where symbols are guessed in decreasing order of probability. If the symbols produced by the

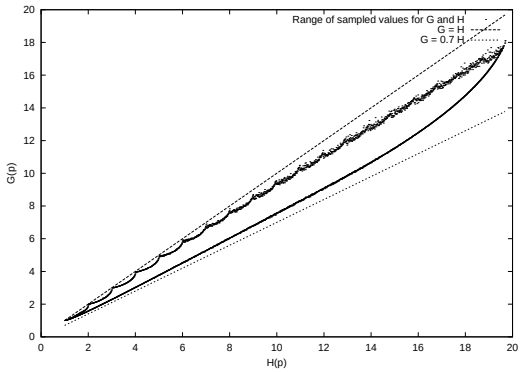

Figure 1: Samples of $G(p)$ and $H(p)$ for alphabets of ≤ 20 symbols.

source are relabeled so that p_1 is the most likely and the sequence p_i is non-increasing then the expected number of guesses is

$$G(p) = \sum_i i p_i. \tag{2}$$

In [9] this is referred to as the *guesswork*. For comparison with entropy we define

$$H(p) := \frac{2^{h(p)} + 1}{2}. \tag{3}$$

Note, on average it takes $(n+1)/2$ guesses to guess from n equally likely possibilities. The popular notion *entropy* $\approx$ *bits of uncertainty* suggests that we look for some sort of equivalence between $G(p)$ and $H(p)$. Casual numerical experiments suggest that $0.7H(p) \leq G(p) \leq H(p)$ (See Figure 1).

2 Bounds on G and H

In [7] it is shown that a lower bound for $G(p)/H(p)$ is $2/e$. This can be derived by showing that a geometric sequence for p_i produces an extrema of $h(p)$ while keeping $G(p)$ fixed. The value $2/e$ is obtained for an infinite geometric sequence as the ratio goes to 1.

The upper bound, $G(p) \leq H(p)$, suggested by numerical experiment is shown to be incorrect in [7]. By taking a sequence where $p_1 = 1 - \beta/n$ and $p_2, \ldots, p_k = \beta/(n^2 - n)$ and letting $n \to \infty$ we get sequences with arbitrary $G(p)$ but with $h(p)$ tending to zero.

So $H(p)$ is within a few bits of being a lower bound on the expected number of guesses, but may be an arbitrary large underestimate. This is fortunate for those designing cryptosystems where entropy is used as a measure of guessability. In [3] Rényi entropy is used to give two-sided bounds on the expected number of guesses.

3 Other measures of guessability

However, the example which dispels the possibility of an upper bound raises an interesting issue. It produces distributions where the average number of guesses is arbitrarily large, but it places almost all the weight on the first possibility so

the mode of the number of guesses will be 1. This suggests that the average number of guesses may not be a good measure of guessability for cryptography.

As an alternative to $G(p)$ another measure of guessability is *the number of guesses required so that probability of having guessed correctly is at least* α. In [9] this is referred to as the *α-work factor* and denoted $wf_\alpha(p)$. They examine $wf_{\frac{1}{2}}$ and decide that again entropy does not provide a good estimate. However they offer $1 - \|p - u\|$ as a more hopeful estimator, where u is the uniform distribution and

$$\|p - q\| := \sup_{X \subset \{1,\ldots n\}} |p(X) - q(X)| \tag{4}$$

is the variation distance.

4 Guesswork and Asymptotic Equipartition

How did this perceived link between entropy and guesswork arise? One suggestion in [9] is that it is a misapplication of the Asymptotic Equipartition Property (AEP).

The AEP applies to a collection of n i.i.d sources of symbols and the words of n symbols they produce. Roughly speaking, the AEP says that if you take n large enough then there is a *typical set* of $2^{nh(p)}$ words which all have approximately the same probability $2^{-nh(p)}$, and the remaining words have only a small probability associated with them (see [2] for a precise statement).

A good estimate of the guesswork of these $2^{nh(p)}$ equiprobable typical words would be $(2^{nh(p)}+1)/2$, and setting $n = 1$ we get the folklore that $G(p) \approx H(p)$. The first problem with this argument is that the AEP deals with large n; what arises in the case $n = 1$ may be very different. Another difficulty is that terms of low probability may contribute significantly because of the factor i in the expectation $\sum i\, p_i$.

If one considers the case of independent identical distributions, $G(p^n)$, then a straightforward application of the AEP for large n is not valid because, as the probability of the atypical words becomes small, the weight associated to them in the sum for $G(p^n)$ grows exponentially.

We can also consider this in terms of the principal of the largest term and typical sets. When calculating expectations for n i.i.d. sources we look at sums of the form:

$$\sum_{n_1 \ldots n_r} \binom{n}{n_1 \ldots n_r} p_1^{n_1} \ldots p_r^{n_r} f(p). \tag{5}$$

If the function $f(p)$ is relatively small, then the most important term in this sum is the one which maximise the product of the multinomial coefficient and the probabilities. This term will have $n_k/n \approx p_k$. These points corresponds to the typical set of the AEP.

When calculating guesswork $f(p) = \text{rank}(p)$ and the sum we consider is closer to:

$$\sum_{n_1 \ldots n_r} \binom{n}{n_1 \ldots n_r}^2 p_1^{n_1} \ldots p_r^{n_r}. \tag{6}$$

Here the largest terms will be those with $n_k/n \approx r\sqrt{p_k}$, where r is a normalising constant. Thus the dominant terms for the guesswork problem are different from those for the coding problem. In [1], Arikan employs clever inequalities

to produce estimates of the guesswork. We apply more direct calculations to extend this result to Markov chains.

Let us now state precisely the problem we consider. For the probability distribution $\{p_1, \ldots, p_m\}$ and $\alpha > 0$, the α^{th} guesswork moment G^α is given b

$$G^\alpha := \sum_{i=1}^{m} i^\alpha \, p_i. \tag{7}$$

Let $\mathtt{A} = \{1, \ldots, r\}$ be a finite alphabet with $r > 1$ characters. Let P be a stationary distribution on $\mathtt{A}^{\mathbf{N}}$, with P_n denoting the restrictions of P to $\mathtt{A}^n$. We seek

$$\lim_n \frac{1}{n} \lg G^\alpha(P_n). \tag{8}$$

Arikan [1] has shown that in the independent case in which P_n is the product of $p_1, \ldots, p_r$,

$$\lim_n \frac{1}{n} \lg G^\alpha(P_n) = (1+\alpha) \lg \sum_{i=1}^{m} i^\alpha \, p_i^{1/(1+\alpha)}. \tag{9}$$

Notation. It is convenient to specify the notation we use for the irreducible (possibly periodic) Markov chain P on $\mathtt{A}^{\mathbf{N}}$. The restriction of P to $\mathtt{A}^n$ is denoted P_n. We assume P has the stochastic matrix $U = (u_{ab})$ and invariant probability (u_a) so that for $\omega \in \mathtt{A}^{n+1}$

$$P_{n+1}(\omega) = u_{\omega_1} \prod_{i=1}^{n} u_{\omega_i \omega_{i+1}} \tag{10}$$

Theorem 4.1 *Let P be the irreducible Markov chain specified above. Then for $\alpha > 0$*

$$\lim_n \frac{1}{n} \lg G^\alpha(P_n) = (1+\alpha) \lg \lambda, \tag{11}$$

where λ is the Perron-Frobenius eigenvalue of the matrix with entries $u_{ab}^{1/(1+\alpha)}$.

5 Proofs

From [6] we have the Perron-Frobenius

Theorem 5.1 *Let C be an irreducible matrix on $\mathtt{A} \times \mathtt{A}$ with nonnegative entries. Then C has an eigenvector $(v_a : a \in \mathtt{A})$ all of whose entries are strictly positive. The corresponding eigenvalue λ is real and has the property that if λ' is any other real or complex eigenvalue of A, then $|\lambda'| \le \lambda$.*

Lemma 5.1 *Let $C = (c_{ab})$ be a nonnegative irreducible matrix on $\mathtt{A} \times \mathtt{A}$ with corresponding Perron-Frobenius eigenvalue λ and eigenvector $(v_a : a \in \mathtt{A})$, $\sum_b c_{ab} v_b = \lambda v_a$. Let (y_{ab}) be a probability distribution on $\mathtt{A} \times \mathtt{A}$ so that $y_{ab} = 0$ whenever $c_{ab} = 0$ and that for each $a \in \mathtt{A}$, with*

$$y_a := \sum_{b \in \mathtt{A}} y_{ab}, \quad \sum_{b \in \mathtt{A}} y_{ba} = \sum_{b \in \mathtt{A}} y_{ab} \equiv y_a. \tag{12}$$

Then

$$\sum_{ab} y_{ab} \lg \frac{y_{ab}}{c_{ab} \, y_a} \ge -\lg \lambda, \tag{13}$$

with equality, if, and only if, $y_{ab}/y_a = c_{ab} v_b/(\lambda v_a)$ for all $a, b \in \mathtt{A}$.

Proof. Since (12) implies $\sum_{ab} y_{ab} \lg v_a = \sum_{ab} y_{ab} \lg v_b$,

$$\sum_{ab} y_{ab} \lg \frac{y_{ab}}{c_{ab}\, y_a} = \sum_a y_a \sum_b \frac{y_{ab}}{y_a} \left[\lg \frac{y_{ab}}{y_a} - \lg \frac{c_{ab}\, v_b}{\lambda\, v_a} - \lg \lambda \right]; \tag{14}$$

$$\sum_{ab} y_{ab} \lg \frac{y_{ab}}{c_{ab}\, y_a} = -\lg \lambda + \sum_a y_a \sum_b D\left(\frac{y_{ab}}{y_a} \,\middle\|\, \frac{c_{ab}\, v_b}{\lambda\, v_a} \right), \tag{15}$$

where D denotes the I-divergence of the conditional probability distributions on $\mathtt{A}$. The conclusions follow from the properties of D: $D \geq 0$ and $D = 0$ implies equality for the probability distributions. Note that if some $y_a = 0$, then there is some a' so that $y_{a'} \neq 0$ and $D > 0$ for the distributions conditioned on a'.

Notation. Let $\omega \in \mathtt{A}^{\mathbf{N}}$. We define $n_{ab}(n, \omega)$ and $n_a(n, \omega)$ by

$$n_{ab}(n, \omega) := \left| \{ i : 1 \leq i \leq n,\ \omega_i = a,\ \omega_{i+1} = b \} \right|,\ n_a := \sum_{b \in \mathtt{A}} n_{ab} \tag{16}$$

with $|\cdot|$ denoting cardinality. When n, ω may be deduced from context, we omit one or both. Note these depend on $(\omega_1, \ldots, \omega_{n+1})$. We also define probability distributions $y_{ab}(n, \omega)$ and $y_a(n, \omega)$ by

$$y_{ab} := \frac{n_{ab}}{n},\ y_a := \frac{n_a}{n}. \tag{17}$$

Given $(\widetilde{n}_{ab})$, $e\,(n, c, (\widetilde{n}_{ab})\,)$ is defined by

$$e\,(n, c, (\widetilde{n}_{ab})\,) := \left| \{ \omega \in \mathtt{A}^{n+1} : \omega_1 = c,\ n_{ab}(n, \omega) = \widetilde{n}_{ab}\ \forall\, ab \in \mathtt{A}^2 \} \right|. \tag{18}$$

In estimates below, the multinomial coefficients

$$\binom{n}{(n_a)} = n! / \prod_{a \in \mathtt{A}} n_a!\,, \quad \binom{n}{(n_{ab})} = n! / \prod_{ab \in \mathtt{A}^2} n_{ab}!\,. \tag{19}$$

are significant.

Condition (12) corresponds to stationarity; in general (y_{ab}) coming from $n_{ab}(n, \omega)$ does not quite satisfy (12).

Lemma 5.2 *For fixed ω and n,*

$$|n_a - \sum_b n_{ba}| \leq 1;\ |y_a - \sum_b y_{ba}| \leq 1/n. \tag{20}$$

Proof. The distinction between n_a and $\sum_b n_{ba}$ is that n_a counts how often a occurs in $(\omega_1, \ldots, \omega_n)$; $\sum_b n_{ba}$, how often a is in $(\omega_2, \ldots, \omega_{n+1})$.

Lemma 5.3 *For fixed ω and n.*

$$\binom{n + r^2 - 1}{n}^{-1} \leq \prod_{ab \in \mathtt{A}^2} y_{ab}^{n_{ab}} \binom{n}{(n_{ab})} \leq 1. \tag{21}$$

Proof. The right hand inequality follows from the multinomial theorem, as does the left, by noting that the leftmost term is the reciprocal of the number of terms and that for given (y_{ab}) the largest term of the expansion is the one given in the inequality.

Lemma 5.4 *Let $\omega_1^* = c$ and $n_{ab} = n_{ab}(n, \omega^*)$. Then*

$$e\left(n, c, (n_{ab})\right) \leq \frac{\prod_a n_a!}{\prod_{ab} n_{ab}!} = \binom{n}{(n_{ab})} \Big/ \binom{n}{(n_a)}. \tag{22}$$

Proof. Given ω in the defining set, if one specifies the order in which the n_a characters which follow a in $(\omega_1, \ldots, \omega_{n+1})$ for each $a \in \mathtt{A}$, then one can uniquely reconstruct ω starting from c. This means the total number of such ω cannot exceed the number of arrangements of the characters.

The basic definition of the α^{th} guesswork moment of the stationary Markov chain P with stochastic matrix $U = (u_{ab})$ and invariant distribution (u_a) is

$$G^\alpha(P_{n+1}) = \sum_{\omega \in \mathtt{A}^{n+1}} \text{rank}(\omega) P_{n+1}(\omega), \tag{23}$$

where the integer $\text{rank}(\omega)$ runs from 1 to r^{n+1}, and $\text{rank}(\omega) < \text{rank}(\omega')$ if $P_{n+1}(\omega) > P_{n+1}(\omega')$. In general, for a given ω there are many ω' with $P_{n+1}(\omega) = P_{n+1}(\omega')$. This occurs if $\omega_1 = \omega_1'$ and $n_{ab}(n, \omega) = n_{ab}(n, \omega')$, but may occur otherwise. We choose a non-reflexive linear ordering $\prec$ on $\mathtt{A} \times \{(n_{ab})\}$ so that

$$(c, (n_{ab})) \prec (c', (n'_{ab})) \Rightarrow u_c \prod_{ab \in \mathtt{A}^2} u_{ab}^{n_{ab}} \geq u'_c \prod_{ab \in \mathtt{A}^2} u_{ab}^{n'_{ab}}. \tag{24}$$

Then we define $g(c, (n_{ab}))$ by

$$g(c', (n'_{ab})) := \sum_{c, (n_{ab}) \prec c', (n'_{ab})} e\left(n, c, (n_{ab})\right). \tag{25}$$

We have the following expression for the α^{th} guesswork moment:

$$G^\alpha(P_{n+1}) = \sum_{c \in \mathtt{A}} \sum_{\{(n_{ab})\}} \left[\sum_{k=g(c,n_{ab})+1}^{g(c,n_{ab})+e(n,c,(n_{ab}))} k^\alpha \right] u_c \prod_{ab \in \mathtt{A}^2} u_{ab}^{n_{ab}}. \tag{26}$$

Different choices of $\prec$ satisfying (24) yield the same $G^\alpha(P_{n+1})$.

Lemma 5.5 *For $\alpha > 0$*

$$\begin{aligned} G^\alpha(P_{n+1}) &\geq \frac{1}{1+\alpha} \max_{c, (n_{ab})} \left\{ e\left(n, c, (n_{ab})\right)^{1+\alpha} u_c \prod_{ab} u_{ab}^{n_{ab}} \right\}. \\ G^\alpha(P_{n+1}) &\leq K_n^{1+\alpha} \max_{c, (n_{ab})} \left\{ e\left(n, c, (n_{ab})\right)^{1+\alpha} u_c \prod_{ab} u_{ab}^{n_{ab}} \right\}, \end{aligned}$$

where

$$K_n := r \binom{n + r^2 - 1}{n}. \tag{27}$$

Proof. The first inequality follows from

$$\sum_{k=g+1}^{g+e} k^\alpha \geq \int_0^e x^\alpha \, dx = \frac{e^{1+\alpha}}{1+\alpha}. \tag{28}$$

For the second inequality, note that each of the K_n summands of the form $\sum_{g+1}^{g+e}$ in (26) is not greater than $e(g+e)^\alpha$, so

$$G^\alpha(P_{n+1}) \le K_n \max_\tau (g_\tau + e_\tau)^\alpha e_\tau p_\tau, \tag{29}$$

where

$$\tau = (c, (n_{ab})\,),\ e_\tau = e(c, (n_{ab})\,),\ g_\tau = g(c, (n_{ab})\,),\ p_\tau = u_c \prod_{ab \in \mathtt{A}^2} u_{ab}^{n_{ab}}. \tag{30}$$

Let τ^* be a value of the parameters which maximizes ${e_\tau}^{1+\alpha}\, p_\tau$:

$${e_{\tau^*}}^{1+\alpha}\, p_{\tau^*} \ge {e_\tau}^{1+\alpha}\, p_\tau. \tag{31}$$

Now the function g_τ satisfies (note $p_\beta \ge p_\tau$ instead of $\beta \prec \tau$)

$$g_\tau + e_\tau \le \sum_{\beta\,:\,p_\beta \ge p_\tau} e_\beta. \tag{32}$$

Note that if $p_\beta \ge p_\tau$ and $e_\beta > e_\tau$, then $e_\beta^\alpha\, e_\tau p_\tau \le e_\beta^{1+\alpha} p_\beta \le e_{\tau^*}^{1+\alpha} p_{\tau^*}$, so

$$((g_\tau + e_\tau)^\alpha e_\tau p_\tau)^{1/\alpha} \le \sum_{\beta\,:\,p_\beta \ge p_\tau} \left[e_\beta^\alpha\, e_\tau p_\tau\right]^{1/\alpha} \le K_n \left[{e_{\tau^*}}^{(1+\alpha)} p_{\tau^*}\right]^{1/\alpha}. \tag{33}$$

Then

$$(g_\tau + e_\tau)^\alpha e_\tau p_\tau \le K_n^\alpha\ {e_{\tau^*}}^{1+\alpha}\, p_{\tau^*}, \tag{34}$$

so the second inequality of this lemma follows from (29).

Proposition 5.1 *For the stationary Markov chain P with stochastic matrix $U = (u_{ab})$ and stationary distribution (u_a),*

$$\limsup_n \frac{1}{n} \lg G^\alpha(P_n) \le (1+\alpha) \lg \lambda, \tag{35}$$

where λ is the Perron-Frobenius eigenvalue of the matrix with entries $u_{ab}^{1/(1+\alpha)}$.

Proof. Since $(\lg u_c K_n)/n \to 0$, starting from (26) we use Lemmas 5.4 and 5.5 to deduce that

$$\limsup_n \frac{\lg G^\alpha(P_{n+1})}{n+1} \le \limsup_n \frac{1+\alpha}{n} \lg \left[\max_{(n_{ab})} \left\{ \frac{\prod_a n_a!}{\prod_{ab} n_{ab}!} \prod_{ab} u_{ab}^{n_{ab}/(1+\alpha)} \right\}\right]. \tag{36}$$

Since $[y_{ab}^{n_{ab}}]^{1/n} = y_{ab}^{y_{ab}}$ and $[y_a^{n_a}]^{1/n} = y_a^{y_a}$, Lemma 5.3 implies

$$\limsup_n \frac{\lg G^\alpha(P_{n+1})}{n+1} \le \limsup_n (1+\alpha) \lg \left[\max_{(n_{ab})} \left\{ \frac{\prod_a y_a^{y_a}}{\prod_{ab} y_{ab}^{y_{ab}}} \prod_{ab} u_{ab}^{y_{ab}/(1+\alpha)} \right\}\right]. \tag{37}$$

Now we take a subsequence $\{n_k, (\,(n_k)_{ab})\}$ so that the right hand side converges to a maximum at $(\,y^*_{ab})$. From Lemma 5.2 we have $\sum_a y^*_{ab} = \sum_b y^*_{ba}$, so Lemma 5.1 yields (35).

Proposition 5.2 *For the stationary Markov chain P with stochastic matrix $U = (u_{ab})$ and stationary distribution (u_a),*

$$\liminf_n \frac{1}{n} \lg G^\alpha(P_n) \geq (1+\alpha) \lg \lambda, \tag{38}$$

where λ is the Perron-Frobenius eigenvalue of the matrix with entries $u_{ab}^{1/(1+\alpha)}$.

Proof. Let λ be the Perron-Frobenius eigenvalue of the matrix $(u_{ab}^{1/(1+\alpha)})$ and (v_a) the corresponding eigenvector with $\sum v_a = 1$. Define

$$c_{ab} := \frac{u_{ab}^{1/(1+\alpha)} v_b}{\lambda\, v_a}. \tag{39}$$

Let ω^*, with $P_n(\omega^*) > 0$ for all n, be a generic point (see [4]) of the ergodic Markov chain with stochastic matrix (c_{ab}). This implies that $(\, y_{ab}(n, \omega^*)\,)$ converges to the distribution $(v_a c_{ab})$ on $\mathtt{A}^2$. Define

$$a^* := \omega_1^*,\ \mathbf{N}^* := \left\{ n \in \mathbf{N} \,:\, \omega_1^* = \omega_{n+1}^* = a^* \right\}. \tag{40}$$

Now for $n \in \mathbf{N}^*$, by the Theorem of Aardenne–Ehrenfest and De Bruijn (see [4])

$$e\,(n, a^*, (n_{ab})\,) \geq \frac{\prod_a (n_a - 1)!}{\prod_{ab} n_{ab}!}, \tag{41}$$

where $n_{ab} = n_{ab}(n, \omega^*)$. From Lemma 5.5, noting $\lim_n (\lg n_a)/n = 0$, we deduce

$$\liminf_{n \in \mathbf{N}^*} \frac{\lg G^\alpha(P_{n+1})}{n+1} \geq \liminf_{n \in \mathbf{N}^*} \frac{1}{n+1} \lg \left[\frac{1}{1+\alpha} \left(\frac{\prod_a n_a!}{\prod_{ab} n_{ab}!} \right)^{1+\alpha} \prod_{ab} u_{ab}^{n_{ab}} \right]. \tag{42}$$

Then by Lemma 5.3,

$$\liminf_{n \in \mathbf{N}^*} \frac{\lg G^\alpha(P_{n+1})}{n+1} \geq \liminf_{n \in \mathbf{N}^*} \frac{n(1+\alpha)}{n+1} \lg \left[\left(\frac{\prod_a y_a^{y_a}}{\prod_{ab} y_{ab}^{y_{ab}}} \right) \prod_{ab} u_{ab}^{y_{ab}/(1+\alpha)} \right]. \tag{43}$$

Then (38) follows from Lemma 5.1 provided we restrict to $\mathbf{N}^*$. To handle $n \in \mathbf{N} \setminus \mathbf{N}^*$, we note that when $n > 2r$, there is some k, $n - r \leq k \leq n+1$, and $\omega^{(n)} \in \mathtt{A}^{\mathbf{N}}$ so that $\omega_i^{(n)} = \omega_i^*$ for $i < k - r$, $\omega_k^{(n)} = a^*$ and $P_{n+1}(\omega^{(n)}) > 0$. Then $n_{ab}(n, \omega^{(n)}) \geq n_{ab}(k-1, \omega^*)$. Applying the above estimates to $n_{ab}(n, \omega^{(n)})$ yields the same limiting value.

This completes the proof of Theorem 4.1. A special case is that for which all the rows of the stochastic matrix $U = (u_{ab})$ are equal, which corresponds to independence. This yields

Corollary 5.1 *Let P be the probability distribution on $\mathtt{A}^{\mathbf{N}}$ which is the product of the single character distributions $P(\{\omega_i = a\}) = u_a$, $u_a \geq 0$, $\sum_{a \in \mathtt{A}} = 1$. Then*

$$\lim_n \frac{1}{n} \lg G^\alpha(P_n) = (1+\alpha) \lg \sum_{a \in \mathtt{A}} u_a^{1/(1+\alpha)}. \tag{44}$$

References

[1] E. Arikan. An inequality on guessing and its application to sequential decoding. *IEEE Transactions on Information Theory*, 42:99–105, Janurary 1996.

[2] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory.* Wiley, New York, 1991.

[3] S. S. Dragomir and S. Boztas. Two sided bounds on guessing moments. *Research Report, Department of Mathematics, Royal Melbourne Institute of Technology*, (8), May 1997.

[4] C.-E. Pfister J.T. Lewis and W.G. Sullivan. Generic points for stationary measures via large deviation theory. *Markov Processes and Related Fields*, 5:235–267, 1999.

[5] J. Kelsey, B. Schneier, and N. Ferguson. Yarrow-160: Notes on the design and analysis of the yarrow cryptographic pseudorandom number generator. In *Sixth Annual Workshop on Selected Areas in Cryptography.* Springer Verlag, August 1999.

[6] Douglas Lind and Brian Marcus. *Symbolic Dynamics and Coding.* Cambridge University Press, Cambridge CB2 1RP UK, 1995.

[7] James L. Massey. Guessing and entropy. In *Proc. IEEE Int. Symp. on Info Th.*, page 204, 1994.

[8] Alec Muffett. Crack: Password cracker. `http://www.users.dircon.co.uk/~crypto/`.

[9] John O. Pliam. The disparity between work and entropy in cryptology. February 1999.

[10] Bruce Schneier. *Applied Cryptography*, volume 61. Wiley, New York, second edition, 1995.

[11] Various. sci.crypt cryptography faq. sci.crypt.

[12] Brian Warner. Egd: The entropy gathering daemon. `http://egd.sourceforge.net/`.