

Approximate Entropy for Testing Randomness

Andrew L. Rukhin
University of Maryland at Baltimore County

Abstract

This paper arise from interest in assessing the quality of random number generators. The problem of testing randomness of a string of binary bits produced by such a generator gained importance with the wide use of public key cryptography and the need for secure encryption algorithms. All such algorithms are based on a generator of (pseudo) random numbers; the testing of such generators for randomness became crucial for communications industry where digital signatures and key management are vital for information processing.

The concept of approximate entropy has been introduced in a series of papers by S. Pincus and co-authors. The corresponding statistic is designed to measure the degree of randomness of observed sequences. It is based on incremental contrasts of empirical entropies based on the frequencies of different patterns in the sequence. Sequences with large approximate entropy must have substantial fluctuation or irregularity. Alternatively, small values of this characteristic imply strong regularity, or lack of randomness, in a sequence. Pincus and Kalman (1997) evaluated approximate entropies for binary and decimal expansions of e , π , $\sqrt{2}$ and $\sqrt{3}$ with the surprising conclusion that the expansion of $\sqrt{3}$ demonstrated much less irregularity than that of π .

Tractable small sample distributions are hardly available, and testing randomness is based, as a rule, on fairly long strings. Therefore, to have rigorous statistical tests of randomness based on this approximate entropy statistic, one needs the limiting distribution of this characteristic under the randomness assumption. Until now this distribution remained unknown and was thought to be difficult to obtain.

To derive the limiting distribution of approximate entropy we modify its definition. In Section 3 it is shown that the approximate entropy

as well as its modified version converges in distribution to a χ^2 -random variable. The concluding Section 4 contains some examples. In particular, the P-values of approximate entropy test statistics for binary expansions of e , π and $\sqrt{3}$ are plotted. Although some of these values for $\sqrt{3}$ digits are small, they do not provide enough statistical significance against the randomness hypothesis.

Approximate Entropy; Distribution of m -patterns; Entropy; Information Divergence; Serial Test; Tail Probabilities; χ^2 -distribution.

AMS classification: 62E20, 62F05, 60F05.

1 Introduction: approximate entropies

Pseudo-random number generators are essential to modern computer security, as well as to reliable statistical simulation, computer-intensive Bayesian inference and validation of algorithms in data mining. All generators differ in their performance, and it is important to examine them with statistical tests of randomness. Many classic tests exist (see Knuth, 1989), but most of them are known to be rather weak (see discussion in Marsaglia, 1985). Currently the most popular of tests for randomness suites is the Diehard Battery, which demands fairly long strings (up to 2^{24} bits). Besides Gustafson et al. (1994) offer a commercial product, called CRYPT-X, that includes some of tests for randomness.

Most conventional algorithmic generators, such as the linear congruential generators and lagged-Fibonacci generators used in IMSL, C++, and other packages, tend to show patterning as they are based upon deterministic recursive algorithms. To detect this patterning, it is natural to investigate the uniformity of the empirical distribution of words (templates) of a given length. One of the tests discussed here is designed to detect the lack of uniformity; it is based on a concept of approximate entropy and its modification to the problem of testing the randomness of a string of binary bits.

To measure the degree of randomness of observed sequences Pincus and Singer (1996) suggested to use a new characteristic, the so-called approximate entropy. Actually, this approach is pursued in a series of papers by S. Pincus and co-authors (Pincus (1991), Pincus and Huang (1992), Pincus and Kalman (1997)).

To fix the ideas denote by $\epsilon_1, \dots, \epsilon_n$, a sequence of i.i.d. random variables each taking values in the finite set $\{1, \dots, s\}$. For $Y_i(m) = (\epsilon_i, \dots, \epsilon_{i+m-1})$,

$1 \leq i \leq n - m + 1$, let

$$C_i^m = \frac{1}{n + 1 - m} \# \{j : 1 \leq j \leq n - m + 1, Y_j(m) = Y_i(m)\}$$

and

$$\Phi^{(m)} = \frac{1}{n + 1 - m} \sum_{i=1}^{n+1-m} \log C_i^m.$$

Observe that C_i^m is the relative frequency of occurrences of the template $Y_i(m)$ in the sequence, and $-\Phi^{(m)}$ is the entropy of the empirical distribution arising on the observed subset of the set of all s^m possible patterns of length m . In other terms $-\Phi^{(m)}$ is the entropy of the type of the sequence induced by $\epsilon_1, \dots, \epsilon_n$ in the space of all templates of length m .

The *approximate entropy* H of order m , $m \geq 1$ is defined as

$$H(m) = \Phi^{(m)} - \Phi^{(m+1)}$$

with $H(0) = -\Phi^{(1)}$. “ $H(m)$ measures the logarithmic frequency with which blocks of length m that are close together remain close together for blocks augmented by one position. Thus, small values of $H(m)$ imply strong regularity, or persistence, in a sequence. Alternatively, large values of $H(m)$ imply substantial fluctuation, or irregularity ..” (Pincus and Singer, 1996, p 2083).

Pincus and Singer (1996) defined a sequence to be m -irregular (m -random) if its approximate entropy $H(m)$ takes the largest possible value. Pincus and Kalman (1997) evaluated quantities $H(m)$, $m = 0, 1, 2$ for binary and decimal expansions of e , π , $\sqrt{2}$ and $\sqrt{3}$ with the conclusion that the expansion of $\sqrt{3}$ demonstrated much less irregularity than that of π . This characteristic has caught attention of mathematicians and other scientists (see Seife, 1997).

Observe that $\log s^m + \Phi^{(m)}$ is the relative entropy (Kullback Leibler divergence or information divergence) between the empirical distribution of the number of patterns of length m and the uniform distribution. It is known that the asymptotic distribution of this relative entropy multiplied by n is proportional to that of a χ^2 -random variable with $s^m - 1$ degrees of freedom. This result also holds for more general f-divergences (defined, for example in Devroye et al, 1996, sec 3.9) including the power-divergences, in particular the χ^2 -divergence (Chapter 4, Read and Cressie, 1988, p 50). It follows that for fixed m , $\Phi^{(m)} \sim -m \log s$ and $H(m) = \Phi^{(m)} - \Phi^{(m+1)} \rightarrow \log s$; indeed

this fact also follows from Theorem 2 in Pincus (1991). As far as the limiting behavior of $n[H(m) - \log s]$, Pincus and Huang (1992), p 3072, indicate that “analytic proofs of asymptotic normality and especially explicit variance estimates for H appear to be extremely difficult”.

2 Modified definition and the covariance matrix

This difficulty is due to the need of studying the joint distribution of frequencies of lengths m and $m + 1$ which in the original definition do not have a simple relationship. Thus, to obtain the limiting distribution of approximate entropy is convenient to modify its definition following a suggestion by Good (1953). Introduce the modified version of the empirical distribution entropy $-\tilde{\Phi}^{(m)}$ as

$$\tilde{\Phi}^{(m)} = \sum_{i_1 \dots i_m} \nu_{i_1 \dots i_m} \log \nu_{i_1 \dots i_m}, \quad (1)$$

where $\nu_{i_1 \dots i_m} = \omega_{i_1 \dots i_m} / n$ denotes the relative frequency of the template (i_1, \dots, i_m) in the augmented (or circular) version of the original string, i.e. in the string $(\epsilon_1, \dots, \epsilon_n, \epsilon_1, \dots, \epsilon_{m-1})$. Observe that under this definition $\omega_{i_1 \dots i_m} = \sum_k \omega_{i_1 \dots i_m k}$, so that for any m , $\sum_{i_1 \dots i_m} \omega_{i_1 \dots i_m} = n$.

Define the modified approximate entropy as

$$\widetilde{H}(m) = \tilde{\Phi}^{(m)} - \tilde{\Phi}^{(m+1)}. \quad (2)$$

By Jensen’s inequality, $\log s \geq \widetilde{H}(m)$ for any m , whereas it is possible that $\log s < H(m)$. Therefore the largest possible value of $\widetilde{H}(m)$ is merely $\log s$; it is attained when $n = s^m$ and the distribution of all m -patterns is uniform. This is a definite advantage of $\widetilde{H}(m)$. Also when calculating the approximate entropy for several values of m , it is convenient to have the sum of all frequencies of m -templates to be equal to n .

The maximally random sequences under definition (2) have the empirical distribution of all patterns of a given length (in a circular version of the sequence) as close to the uniform one as possible. For example, from the point of view of $H(1)$ maximally random binary strings with $n = 5$, which have three zeros and two ones, are $(0, 0, 1, 1, 0)$ and $(0, 1, 1, 0, 0)$ (see Pincus and Singer (1996) p 2084.) According to $\widetilde{H}(m)$ one should add two sequences $(1, 1, 0, 0, 0)$ and $(1, 0, 0, 0, 1)$.

Still, when n is large, $H(m)$ and $\widetilde{H}(m)$ cannot differ much. Indeed, one has with $\omega'_{i_1 \dots i_m} = (n - m + 1) \nu'_{i_1 \dots i_m}$

$$\sum_{i_1 \dots i_m} \omega'_{i_1 \dots i_m} = n - m + 1,$$

and $\omega_{i_1 \dots i_m} - \omega'_{i_1 \dots i_m} \leq m - 1$. It follows that

$$\left| \nu_{i_1 \dots i_m} - \nu'_{i_1 \dots i_m} \right| \leq \frac{m - 1}{n - m + 1}, \quad (3)$$

which suggests that for a fixed m ,

$$\Phi^{(m)} = \sum_{i_1 \dots i_m} \nu'_{i_1 \dots i_m} \log \nu'_{i_1 \dots i_m},$$

and $\widetilde{\Phi}^{(m)}$ must be close for large n . Therefore Pincus' approximate entropy and (2) also must be close, and their limiting distributions must coincide.

The derivation of this distribution is based on the limiting covariance matrix of the joint distribution of $\omega_{i_1 \dots i_m}$. Clearly

$$\omega_{i_1 \dots i_m} = \sum_{\ell=1}^n \delta_{(i_1 \dots i_m), (\epsilon_\ell, \dots, \epsilon_{\ell+m-1})}$$

with $\delta_{i, \ell}$ denoting the Kronecker symbol for two m -indices, \mathbf{i} and ℓ . For any fixed m -pattern, $i_1 \dots i_m$, the random variables $\delta_{(i_1 \dots i_m), (\epsilon_i \dots \epsilon_{i+m-1})}$ are m -dependent, so that for $|i - k| \geq m$

$$\mathbf{Cov} \left(\delta_{(i_1 \dots i_m), (\epsilon_i, \dots, \epsilon_{i+m-1})}, \delta_{(j_1 \dots j_m), (\epsilon_k, \dots, \epsilon_{k+m-1})} \right) = 0.$$

As $E \delta_{(i_1 \dots i_m), (\epsilon_i \dots \epsilon_{i+m-1})} = s^{-m}$, one has for $r = |i - k| < m$ when $i \leq k$

$$\mathbf{Cov} \left(\delta_{(i_1 \dots i_m), (\epsilon_i, \dots, \epsilon_{i+m-1})}, \delta_{(j_1 \dots j_m), (\epsilon_k, \dots, \epsilon_{k+m-1})} \right) = \frac{1}{s^{m+r}} \delta_{(i_{r+1} \dots i_m), (j_1 \dots j_{m-r})} - \frac{1}{s^{2m}}.$$

Therefore

$$\begin{aligned} \mathbf{Cov} (\omega_{i_1 \dots i_m}, \omega_{j_1 \dots j_m}) &= \frac{n}{s^m} \delta_{(i_1 \dots i_m), (j_1 \dots j_m)} - \frac{n}{s^{2m}} \\ &+ n \sum_{r=1}^{m-1} \left[\delta_{(i_{r+1} \dots i_m), (j_1 \dots j_{m-r})} \frac{1}{s^{m+r}} + \delta_{(i_1 \dots i_{m-r}), (j_{r+1} \dots j_m)} \frac{1}{s^{m+r}} - \frac{2}{s^{2m}} \right]. \end{aligned}$$

Now we introduce the matrix Σ_m formed by $n^{-1} \mathbf{Cov}(\omega_{i_1 \dots i_m}, \omega_{j_1 \dots j_m})$, i.e. by the elements

$$\begin{aligned} \sigma_{i_1 \dots i_m j_1 \dots j_m} &= \frac{1}{s^m} \delta_{(i_1 \dots i_m), (j_1 \dots j_m)} - \frac{2m-1}{s^{2m}} \\ &+ \sum_{r=1}^{m-1} \left[\delta_{(i_{r+1} \dots i_m), (j_1 \dots j_{m-r})} + \delta_{(i_1 \dots i_{m-r}), (j_{r+1} \dots j_m)} \right] \frac{1}{s^{m+r}}. \end{aligned} \quad (4)$$

Because of our convention of counting the frequencies, the elements of the matrix Σ_{m+1} are related to those of Σ_m by the formula

$$\sum_{ij} \sigma_{i_1 \dots i_m i j_1 \dots j_m j} = \sigma_{i_1 \dots i_m j_1 \dots j_m}. \quad (5)$$

The rank of the matrix Σ_{m+1} is $s^{m+1} - s^m$. Indeed let the s^{m+1} -dimensional vector \mathbf{e}_{m+1} have all coordinates equal to one. Then it belongs to the null space of Σ_{m+1} as for any i_1, \dots, i_{m+1}

$$\sum_{j_1 \dots j_{m+1}} \sigma_{i_1 \dots i_{m+1} j_1 \dots j_{m+1}} = 0.$$

Consider the vectors whose $(i_1 \dots i_{m+1})$ -th coordinate has the form $\delta_{(i_1 \dots i_m), (k_1 \dots k_m)} - \delta_{(i_2 \dots i_{m+1}), (k_1 \dots k_m)}$ for some k_1, \dots, k_m . Then

$$\begin{aligned} \sum_{j_1 \dots j_{m+1}} \sigma_{i_1 \dots i_{m+1} j_1 \dots j_{m+1}} \delta_{(j_1 \dots j_m), (k_1 \dots k_m)} &= \sum_j \sigma_{i_1 \dots i_{m+1} k_1 \dots k_m j} \\ &= \frac{1}{s^{m+1}} \delta_{(i_1 \dots i_m), (k_1 \dots k_m)} - \frac{2m}{s^{2m+1}} \\ &+ \sum_{r=1}^m \delta_{(i_{r+1} \dots i_{m+1}), (k_1 \dots k_{m+1-r})} \frac{1}{s^{m+r}} + \sum_{r=1}^{m-1} \delta_{(i_1 \dots i_{m-r}), (k_{r+1} \dots k_m)} \frac{1}{s^{m+r+1}} \\ &= -\frac{2m}{s^{2m+1}} + \sum_{r=0}^{m-1} \left[\delta_{(i_{r+2} \dots i_{m+1}), (k_1 \dots k_{m-r})} + \delta_{(i_1 \dots i_{m-r}), (k_{r+1} \dots k_m)} \right] \frac{1}{s^{m+r+1}} \\ &= \sum_j \sigma_{i_1 \dots i_{m+1} j k_1 \dots k_m} = \sum_{j_1 \dots j_{m+1}} \sigma_{i_1 \dots i_{m+1} j_1 \dots j_{m+1}} \delta_{(j_2 \dots j_{m+1}), (k_1 \dots k_m)}. \end{aligned}$$

Therefore all these vectors also belong to the null space of Σ_{m+1} , and, as one can show, together with \mathbf{e}_{m+1} they span this space. As there are $s^m - 1$ linearly independent vectors of the form above, the dimension of the null space is s^m . This fact can be derived from the spectral decomposition of Σ_{m+1} in Billingsley (1956).

Thus Σ_{m+1} is not invertible, but we show now that its generalized inverse Σ_{m+1}^- has a remarkably simple form

$$\Sigma_{m+1}^- = s^{m+1} \mathbf{Q} = s^{m+1} \left[\mathbf{I}_{m+1} - s^{-1} \left(\mathbf{e}_1 \mathbf{e}_1^T \oplus \cdots \oplus \mathbf{e}_1 \mathbf{e}_1^T \right) \right]. \quad (6)$$

Here \mathbf{I}_{m+1} denotes the identity matrix of size $s^{m+1} \times s^{m+1}$. Thus \mathbf{Q} is the projection onto the orthogonal complement to the space spanned by the s^m vectors $\mathbf{e}_1 \oplus \mathbf{0} \oplus \cdots \oplus \mathbf{0}, \dots, \mathbf{0} \oplus \cdots \oplus \mathbf{0} \oplus \mathbf{e}_1$.

Indeed

$$\Sigma_{m+1} = \frac{1}{s^{m+1}} \mathbf{I}_{m+1} + R + R^T,$$

where $\mathbf{Q}R = R^T \mathbf{Q} = \mathbf{0}$. It suffices to take R to be formed by the elements

$$\begin{aligned} & -\frac{2m+1}{2s^{2m+2}} + \sum_{r=1}^m \delta_{(i_{r+1} \cdots i_m), (j_1 \cdots j_{m-r})} \frac{1}{2s^{m+2+r}} - \sum_{r=1}^m \delta_{(i_1 \cdots i_{m-r}), (j_{r+1} \cdots j_m)} \frac{1}{2s^{m+2+r}} \\ & + \sum_{r=1}^m \delta_{(i_1 \cdots i_{m+1-r}), (j_{r+1} \cdots j_{m+1})} \frac{1}{s^{m+1+r}}. \end{aligned}$$

Then

$$s^{m+1} \Sigma_{m+1} \mathbf{Q} = \mathbf{Q} + s^{m+1} R \mathbf{Q},$$

and, as $\mathbf{Q}^2 = \mathbf{Q}$,

$$s^{2m+2} \Sigma_{m+1} \mathbf{Q} \Sigma_{m+1} \mathbf{Q} = \mathbf{Q} + s^{m+1} R \mathbf{Q}.$$

Therefore $s^{m+1} \Sigma_{m+1} \mathbf{Q}$ is an idempotent matrix and $\text{tr}(s^{m+1} \Sigma_{m+1} \mathbf{Q}) = s^{m+1} - s^m$, which is the rank of the matrix Σ_{m+1} . Thus according to Lemma 2.2.2 of Rao and Mitra (1971) a generalized inverse of Σ_{m+1} is given by (6). This fact is alluded to in Marsaglia (1985).

We will use the formulas (4) and (6) in the next Section to obtain the limiting distribution of the approximate entropy. In fact, we derive this distribution under a more general concept of the ϕ -entropy of a discrete random variable. If its distribution is given by probabilities π_1, \dots, π_M , then let

$$E(\pi_1, \dots, \pi_M) = \sum_{j=1}^M \pi_j \phi(\pi_j).$$

Here (and further) $\phi(u), 0 \leq u \leq 1$, is assumed to be continuously twice differentiable. Commonly $\phi(1) = 0$ and ϕ is convex, in which case with

$\phi(u) = \varphi(Mu)$, E becomes φ -information-type divergence between our distribution and the uniform one. The form of E when $\phi(u) = uf(1/u) + f(0)(1-u)$ with a convex function f can be derived from an axiomatic approach (see Vajda (1989), Proposition 10.17 or Morales, Pardo and Vajda (1996)). Of course the choice $\phi(u) = -\log u$ leads to the traditional Shannon entropy, which also forms the basis for the definition of approximate entropy by Pincus.

When $M = s^m$ and the probability distribution is that of all m - templates, the definition (2) of the approximate entropy suggests the following formula for the ϕ -uncertainty $\sum \nu_{i_1 \dots i_m} \phi(\nu_{i_1 \dots i_m})$. To get a practical limiting distribution it is convenient to rescale this characteristic, and with

$$a_m = \frac{s^m}{\phi' \left(\frac{1}{s^m} \right) + \frac{1}{2s^m} \phi'' \left(\frac{1}{s^m} \right)},$$

define

$$\Phi_H^{(m)} = a_m \sum_{i_1 \dots i_m} \nu_{i_1 \dots i_m} \phi(\nu_{i_1 \dots i_m}).$$

Then for a function ϕ as above we give the general definition of *approximate ϕ -entropy* AH of order m , $m \geq 1$ as

$$AH(m) = \Phi_H^{(m)} - \Phi_H^{(m+1)}.$$

Now for fixed m , one should expect $\Phi^{(m)} \sim a_m \phi(s^{-m})$, so that $AH(m) = \Phi^{(m)} - \Phi^{(m+1)} \rightarrow a_m \phi(s^{-m}) - a_{m+1} \phi(s^{-m-1})$.

When $\phi(u) = \log u$, one has $a_m \equiv 2$, so that $AH(m)$ indeed extends the definition (2). Notice that classical Pearson's χ^2 statistic, corresponds to $\phi(u) = u$. Indeed if

$$\begin{aligned} \psi_m^2 &= \sum_{i_1 \dots i_m} \frac{(\omega_{i_1 \dots i_m} - ns^{-m})^2}{ns^{-m}} = ns^m \sum_{i_1 \dots i_m} (\nu_{i_1 \dots i_m} - s^{-m})^2 \\ &= ns^m \sum_{i_1 \dots i_m} \nu_{i_1 \dots i_m}^2 - n, \end{aligned}$$

then with $a_m = s^m$, $a_{m+1} \phi(s^{-m-1}) = a_m \phi(s^{-m})$ and

$$\Phi_H^{(m)} = s^m \sum_{i_1 \dots i_m} \nu_{i_1 \dots i_m}^2 = 1 + \frac{\psi_m^2}{n}.$$

Thus $AH(m) = (\psi_m^2 - \psi_{m+1}^2)/n$.

In this paper, we show that the limiting distribution of $n[a_{m+1}\phi(s^{-m-1}) - a_m\phi(s^{-m}) - AH(m)]$, when $n \rightarrow \infty$ and m is fixed, is the familiar χ^2 -distribution with $(s-1)s^m$ degrees of freedom. It will follow that $2n[\log s - \widetilde{H}(m)]$ has the asymptotic χ^2 -distribution. The convergence of $\psi_{m+1}^2 - \psi_m^2$ to this distribution is known (see Good, 1953, 1957, 1997.) As a matter of fact, this result forms the basis for the so-called *generalized serial test* of randomness (Menezes, van Oorschot and Vanstone, 1997). As $n[ApEn(m) - \widetilde{H}(m)] = O_P(n^{-1})$, the limiting distributions of Pincus' approximate entropy and of $\widetilde{H}(m)$ coincide. This fact provides the basis for statistical tests of randomness via the approximate entropy.

In Section 4 more examples are given, in particular the tail probabilities arising from the mentioned χ^2 approximation for the approximate entropy statistic are evaluated and plotted for binary expansions of e , π and $\sqrt{3}$.

3 Asymptotic behavior of approximate entropy

We prove here that the limiting distribution of $n[a_{m+1}\phi(s^{-m-1}) - a_m\phi(s^{-m}) - AH(m)]$ coincides with that of a χ^2 -random variable, $\chi^2(s^{m+1} - s^m)$ with $s^{m+1} - s^m$ degrees of freedom.

Theorem 1. For fixed m as $n \rightarrow \infty$ one has the following convergence in distribution

$$n[a_{m+1}\phi(s^{-m-1}) - a_m\phi(s^{-m}) - AH(m)] \rightarrow \chi^2(s^{m+1} - s^m).$$

Also

$$n[H(m) - \widetilde{H}(m)] = O_P\left(\frac{1}{n}\right), \quad (7)$$

so that

$$2n[\log s - H(m)] \rightarrow \chi^2(s^{m+1} - s^m).$$

Proof For i_1, \dots, i_m running through the set of all s^m possible vectors of length m , with values in the set $\{1, \dots, s\}$ $\Phi_H^{(m)}$ has the form

$$\Phi_H^{(m)} = a_m \sum_{i_1 \dots i_m} \nu_{i_1 \dots i_m} \phi(\nu_{i_1 \dots i_m}).$$

Here $\nu_{i_1 \dots i_m}$ denotes the relative frequency of the pattern (i_1, \dots, i_m) in the string of bits $(\epsilon_1, \dots, \epsilon_n, \epsilon_1, \dots, \epsilon_{m-1})$.

Let

$$Z_{i_1 \dots i_m} = \sqrt{n} \left[\nu_{i_1 \dots i_m} - \frac{1}{s^m} \right].$$

Then the vector formed by $Z_{i_1 \dots i_m}$ has the asymptotic multivariate normal distribution with zero mean and the covariance matrix Σ_m as in (4). Since with probability one, $\sum Z_{i_1 \dots i_m} = 0$,

$$\begin{aligned} \Phi_H^{(m)} &= a_m \sum_{i_1 \dots i_m} \left[\frac{1}{s^m} + \frac{Z_{i_1 \dots i_m}}{\sqrt{n}} \right] \left[\phi \left(\frac{1}{s^m} \right) + \phi' \left(\frac{1}{s^m} \right) \frac{Z_{i_1 \dots i_m}}{\sqrt{n}} \right. \\ &\quad \left. + \phi'' \left(\frac{1}{s^m} \right) \frac{Z_{i_1 \dots i_m}^2}{2n} + O_P \left(\frac{1}{n^{3/2}} \right) \right] \sim a_m \phi \left(\frac{1}{s^m} \right) + \frac{s^m}{n} \sum_{i_1 \dots i_m} Z_{i_1 \dots i_m}^2. \end{aligned}$$

Using a similar notation for patterns of length $m+1$, let $\nu_{i_1 \dots i_m i_{m+1}}$ be the relative frequencies, and let $Z = (Z_{i_1 \dots i_m i_{m+1}})^T$ denote the vector formed by corresponding differences between empirical and theoretical probabilities. Then because of our convention for counting the frequencies

$$Z_{i_1 \dots i_m} = \sum_{k=1}^s Z_{i_1 \dots i_m k}$$

and

$$\Phi_H^{(m+1)} \sim a_{m+1} \phi \left(\frac{1}{s^{m+1}} \right) + \frac{s^{m+1}}{n} \sum_{i_1 \dots i_m i_{m+1}} Z_{i_1 \dots i_m i_{m+1}}^2.$$

Thus

$$\begin{aligned} \Phi_H^{(m)} - \Phi_H^{(m+1)} &\sim a_m \phi \left(\frac{1}{s^m} \right) - a_{m+1} \phi \left(\frac{1}{s^{m+1}} \right) \\ &\quad - \frac{s^{m+1}}{n} \left[\sum_{i_1 \dots i_m i_{m+1}} Z_{i_1 \dots i_m i_{m+1}}^2 - s \sum_{i_1 \dots i_m} \left(\sum_k Z_{i_1 \dots i_m k} \right)^2 \right]. \end{aligned}$$

Therefore the limiting distribution of

$$n \left[\Phi_H^{(m)} - \Phi_H^{(m+1)} - a_m \phi \left(\frac{1}{s^m} \right) + a_{m+1} \phi \left(\frac{1}{s^{m+1}} \right) \right]$$

coincides with that of $-s^{m+1} Z^T \mathbf{Q} Z$, where \mathbf{Q} is defined by (6) so that $s^{m+1} \mathbf{Q}$ is a generalized inverse of Σ_{m+1} . It is well known (see for example Theorem

9.2.2 in Rao and Mitra (1971)) that this distribution is the χ^2 -distribution with the degrees of freedom equal to the rank of Σ_{m+1} .

The estimate (3) shows that if $Z'_{i_1 \dots i_m} = \sqrt{n} [\nu'_{i_1 \dots i_m} - s^{-m}]$ then $|Z'_{i_1 \dots i_m} - Z_{i_1 \dots i_m}| \leq (m-1)\sqrt{n}/(n-m+1)$ and

$$|\tilde{\Phi}^{(m)} - \Phi^{(m)}| \sim \frac{s^m}{2n} \left| \sum_{i_1 \dots i_m} Z_{i_1 \dots i_m}^2 - \sum_{i_1 \dots i_m} Z_{i_1 \dots i_m}'^2 \right| \leq \frac{s^{2m}(m-1)^2}{2(n-m+1)^2}.$$

Thus (7) follows which completes the proof of Theorem 1. \square

For the observed value $H(m)$, one has to define $\chi^2(obs)$ as $\chi^2(obs) = 2n |\log s - H(m)|$, whereas, as has been noticed, the difference $\log s - \widetilde{H}(m)$ is always positive. The reported P-value (tail probability) is

$$P_n(m) = 1 - \mathbf{P} \left(2^{m-1}, \chi^2(obs)/2 \right)$$

with \mathbf{P} denoting the incomplete gamma-function. The null hypothesis of randomness is rejected for large values of $\chi^2(obs)$.

The proof of Theorem 1 also shows that the asymptotic distribution of the statistics $2n [\log s - \widetilde{H}(m)]$ and $2n [\log s - H(m)]$, evaluated under the alternative of the form $\pi_{i_1 \dots i_m i_{m+1}} = s^{-m-1} + n^{-1/2} \eta_{i_1 \dots i_m i_{m+1}}$, with $\eta^T \mathbf{e} = 0$, is noncentral χ^2 -distribution with $s^{m+1} - s^m$ degrees of freedom and the noncentrality parameter $\Delta = s^{m+1} \eta^T \eta$. Indeed, the limiting distribution of $Z_{i_1 \dots i_{m+1}}$ is normal with the mean η formed by coordinates $\eta_{i_1 \dots i_{m+1}}$, and the covariance matrix Σ_{m+1} . Thus because of Theorem 9.2.3 of Rao and Mitra (1971)) the distribution of the quadratic form $s^{m+1} Z^T \mathbf{Q} Z$ is that of a noncentral χ^2 -random variable with $s^{m+1} \mathbf{tr}(\mathbf{Q} \Sigma_{m+1}) = s^{m+1} - s^m$ degrees of freedom and the noncentrality parameter equal to $s^{m+1} \eta^T \mathbf{Q} \eta$. An easy calculation shows that $\mathbf{Q} \eta = \eta$, so that indeed $\Delta = s^{m+1} \eta^T \eta$.

This fact allows for an approximate power function of the test of randomness based on approximate entropy and answers a question posed in Kimberley (1987) p 365. A similar test can be derived from the following considerations.

Put

$$\mathbf{V} = s^m (\mathbf{e}_1 \mathbf{e}_1^T \oplus \dots \oplus \mathbf{e}_1 \mathbf{e}_1^T) - s^{m-1} (\mathbf{e}_2 \mathbf{e}_2^T \oplus \dots \oplus \mathbf{e}_2 \mathbf{e}_2^T),$$

with the second block-diagonal matrix of size $s^{m+1} \times s^{m+1}$ formed by s^{m-1} unity blocks of size $s^2 \times s^2$. Then the matrix

$$\mathbf{U} = s^{m+1} \mathbf{Q} - \mathbf{V}$$

possesses the following property

$$\Sigma_{m+1} \mathbf{U} \Sigma_{m+1} \mathbf{U} \Sigma_{m+1} = \Sigma_{m+1} \mathbf{U} \Sigma_{m+1}. \quad (8)$$

Indeed, since $s^{m+1} \mathbf{Q}$ is the generalized inverse of Σ_{m+1} to prove (8) it suffices to show that

$$\Sigma_{m+1} \mathbf{V} \Sigma_{m+1} \mathbf{V} \Sigma_{m+1} = \Sigma_{m+1} \mathbf{V} \Sigma_{m+1}. \quad (9)$$

According to Theorem 9.2.1 of Rao and Mitra (1971), (9) is equivalent to the fact that the distribution of the quadratic form $Z^T \mathbf{V} Z$ is that of a χ^2 -variable. But this quadratic form, because of (5), corresponds to the form $Z^T \mathbf{V} Z$ with now s^m -dimensional normal random vector Z whose covariance matrix is Σ_m . By Theorem 1 it indeed has the χ^2 -distribution with $s^m - s^{m-1}$ degrees of freedom.

Thus (9) holds, which also can be derived from the fact that the matrix $\Sigma_{m+1} \mathbf{V}$ is idempotent. By the same Theorem 9.2.1 of Rao and Mitra (1971) the distribution of the quadratic form $Z^T \mathbf{U} Z$ is a χ^2 -distribution with $\text{tr}(\mathbf{U} \Sigma_{m+1}) = s^{m+1} - 2s^m + s^{m-1}$ degrees of freedom. This corresponds to the test statistic of the form $-\tilde{\Phi}^{(m-1)} + 2\tilde{\Phi}^{(m)} - \tilde{\Phi}^{(m+1)}$, having a χ^2 -distribution with $s^{m+1} - 2s^m + s^{m-1}$ degrees of freedom. This statistic, which also can be readily used for testing randomness, has been suggested by Good (1953). However, as observed in Good (1953), this fact does not extend to higher order finite differences.

4 Examples

Here are two strings of 20 binary bits which have been suggested by Chaitin (1975)

$$(A) \ 01010101010101010101$$

$$(B) \ 01101100110111100010$$

For a non-randomly looking sequence (A), $H(0) = -\Phi^{(1)} = -\tilde{\Phi}^{(1)} = \log 2$, which is the largest possible value for H . Since there are only two occurring patterns of length 2, namely (0, 1) and (1, 0) with frequencies 10 and 9 respectively,

$$\Phi^{(2)} = \frac{1}{19} \left[10 \log \frac{10}{19} + 9 \log \frac{9}{19} \right] = -0.6918\dots$$

Thus

$$H(1) = 0.0014\dots$$

with $\chi^2(obs) = 40[\log 2 - H(1)] = 27.6699\dots$

For the modified entropy

$$\tilde{\Phi}^{(2)} = \frac{1}{20} \left[10 \log \frac{10}{20} + 10 \log \frac{10}{20} \right] = -\log 2,$$

with $\widetilde{H(1)} = 0$ and $\chi^2(obs) = 40 \log 2 = 27.7258\dots$. Thus from the point of view of $\widetilde{H(1)}$, the sequence (A) is completely non-random.

This is to be contrasted with the values of the approximate entropy for the string (B).

$$\Phi^{(1)} = \tilde{\Phi}^{(1)} = \frac{1}{20} \left[9 \log \frac{9}{20} + 11 \log \frac{11}{20} \right] = -0.6881\dots$$

There are 5 patterns (1, 0) and of (0, 1), 6 patterns (1, 1), and 3 patterns (0, 0) in this string, so that

$$\Phi^{(2)} = \frac{1}{19} \left[5 \log \frac{5}{19} + 6 \log \frac{6}{19} + 5 \log \frac{5}{19} + 3 \log \frac{3}{19} \right] = -1.3581\dots$$

and

$$H(1) = 0.6699\dots$$

with $\chi^2(obs) = 40[\log 2 - H(1)] = 0.9299$. One also has

$$\tilde{\Phi}^{(2)} = \frac{1}{20} \left[5 \log \frac{5}{20} + 6 \log \frac{6}{20} + 5 \log \frac{5}{20} + 4 \log \frac{4}{20} \right] = -1.3762\dots$$

as there are 5 copies of (1, 0) and of (0, 1), 6 copies of (1, 1), and 4 copies of (0, 0) in the augmented version of this string, Thus $\widetilde{H(1)} = 0.6881\dots$, which is closer to the maximum value 0.6931.. than Pincus' entropy, and $\chi^2(obs) = 40[\log 2 - \widetilde{H(1)}] = 0.2024\dots$

Observe that

$$\Sigma_2 = \begin{pmatrix} \frac{5}{16} & -\frac{1}{16} & -\frac{1}{16} & -\frac{3}{16} \\ -\frac{1}{16} & \frac{1}{16} & \frac{1}{16} & -\frac{1}{16} \\ -\frac{1}{16} & \frac{1}{16} & \frac{1}{16} & -\frac{1}{16} \\ -\frac{3}{16} & -\frac{1}{16} & -\frac{1}{16} & \frac{5}{16} \end{pmatrix}$$

In particular, for frequencies of two letter words in a circular fashion, with probability one $\omega_{01} = \omega_{10}$, so that there are always exactly as many copies of the pattern (0, 1) as there are of (1, 0).

Thus from the point of view of approximate entropies $H(1)$ and $\widetilde{H}(1)$ the sequence (A) does not look random at all, but the string (B) does and even more so for the modified entropy $\widetilde{H}(1)$. The strings (A) and (B) are also somewhat imprecisely examined in Pincus and Kalman (1997) p 3514. Indeed these authors give incorrect values $H(1) = 0$ for (A), and $H(1) = 0.6774..$ for (B).

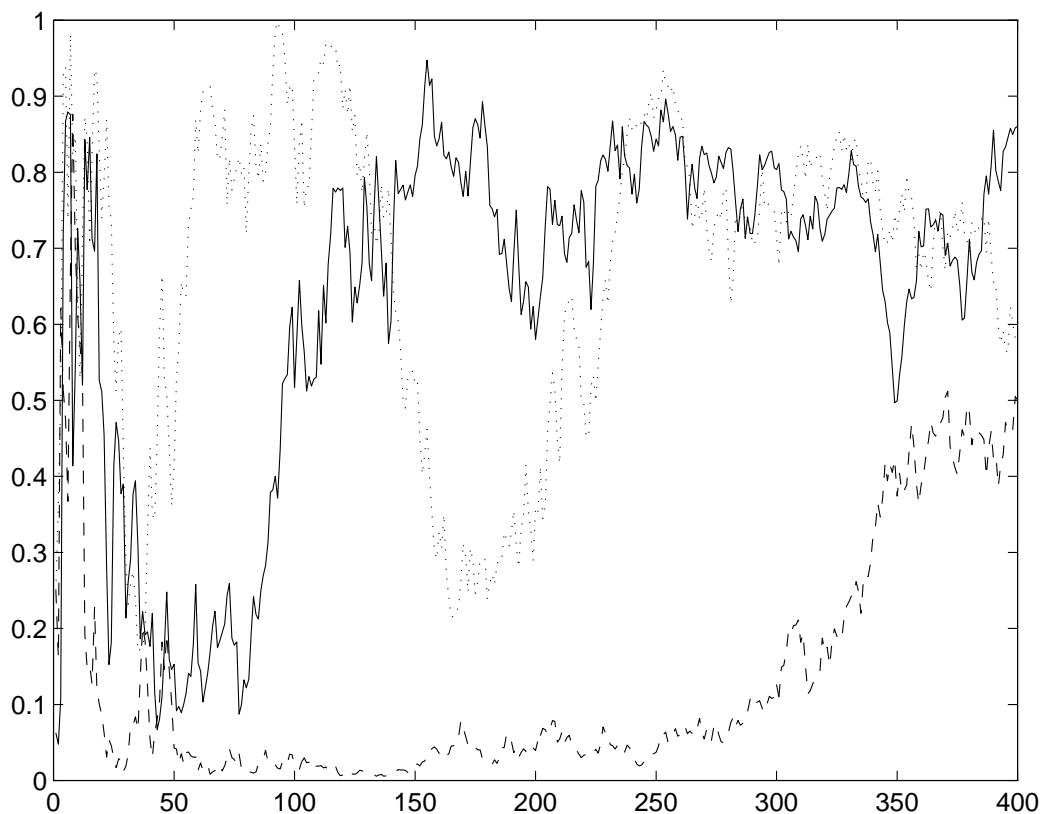


Figure 1 Consecutive P-values for binary expansions of $\sqrt{3}$ (broken line), π (dotted line) and e (solid line) when $m = 1$.

In Figure 1 the P-values $P_n(1)$ from Section 2 are plotted against the first digits of binary expansions of $\sqrt{3}$, π and e . According to this data, P-values corresponding to $\sqrt{3}$ are smaller than those of e and π . However, the smallest P-values, occurring in the block from 130-th to 150-th digits, are of order 0.006. They lack statistical significance to reject the random nature

of these digits because of the multiple nature of the testing problems. This observation seems to be confirmed by the study of patterns of size $m = 7$, when the digits of π and $\sqrt{3}$ look much less random than these of expansion of e (Figure 2). Thus these results do not support the conclusion about the non-random appearance of digits in the expansion of $\sqrt{3}$. As a matter of fact, Good and Gover (1967) applied the serial test to the study of binary digits in the expansion of $\sqrt{2}$. Although these digits also occasionally lead to small P-values, similarly to $\sqrt{3}$ expansion, the null hypothesis of randomness cannot be rejected.

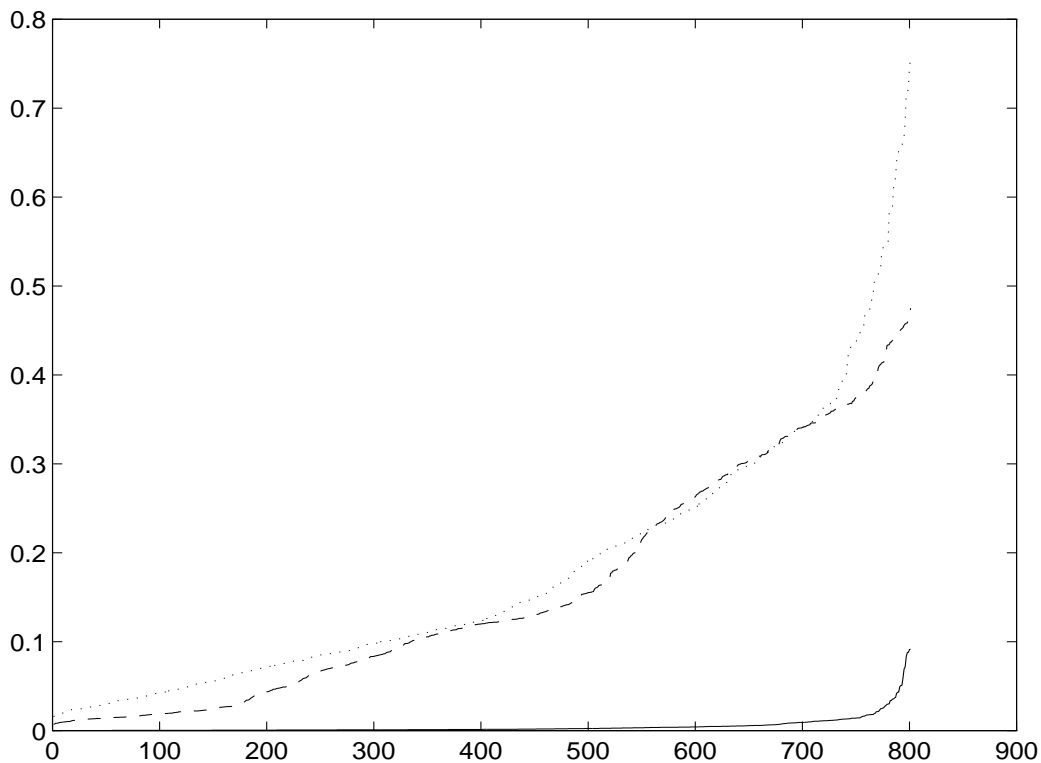


Figure 2 Consecutive P-values for binary expansions of $\sqrt{3}$ (broken line), π (dotted line) and e (solid line) when $m = 7$.

4.1 Acknowledgement

Thanks are due to Professor Good and Dr. Vajda for their helpful comments.

References

1. Billingsley, P. (1956) Asymptotic distributions of two goodness of fit criteria. *Ann. Math. Statist.*, **27**, 1123–1129.
2. Chaitin, G. (1975) Randomness and mathematical proof, *Scientific American*, **232**, 47–52.
3. Devroye, L., Györfi, L., and Lugosi, G. (1996) *A Probabilistic Theory of Pattern Recognition*. Springer, New York. pp 31–34.
4. Good, I. J. (1953) The serial test for sampling numbers and other tests for randomness. *Proc. Cambridge Philos. Soc.*, **47**, 276–284.
5. Good, I. J. (1957) On the serial test for random sequences. *Ann. Math. Statist.*, **28**, 262–264.
6. Good, I. J. (1997) The roughness of visitations in a Markov chain, A review with extensions. in *Advances in the Theory and Practice of Statistics*, N. L. Johnson and N. Balakrishnan, eds, J. Wiley, New York.
7. Good, I. J. and Gouy T. N. (1967) The generalized serial test and the binary expansion of $\sqrt{2}$. *Journ. Royal Statist. Soc.*, **130, A**, 102–107.
8. Gustafson, H., Dawson, E., Nielsen, L., and Caelli, W. (1994) A computer package for measuring the strength of encryption algorithms. *Computers and Security*, **13**, 687–697.
9. Kimberley, M. (1987) Comparison of two statistical tests for keystream sequences. *Electronics Letters*, **23**, 365–366.
10. Knuth, D. E. (1998) *The Art of Computer Programming*, Vol. 2, 3rd ed. Addison-Wesley Inc., Reading, MA. pp 61–80.
11. Marsaglia, G. (1985) A Current View of Random Number Generation. In *Computer Science and Statistics: Proceedings of the Sixteenth Symposium on the Interface*, 3–10. Elsevier Science Pub., New York.
12. Marsaglia, G. (1996) *Diehard: A battery of tests for randomness*. <http://stat.fsu.edu/geo/diehard.html>.

13. Menezes, A. J., van Oorschot, P. C., and Vanstone, S. A. (1997) *Handbook of Applied Cryptography*, CRC Press, Boca Raton, FL. p 181.
14. Morales, D., Pardo, L. and Vajda I. (1996) Uncertainty of discrete stochastic systems: General theory and statistical inference. *IEEE Trans. Systems, Man and Cybernetics, Part A*, **26**, 681–697.
15. Pincus, S., and Huang, W.-M. (1992) Approximate entropy, statistical properties and applications, *Communications in Statistics, Part A-Theory and Methods*, **21**, 3061–3077.
16. Pincus, S., and Kalman, R. E. (1997) Not all (possibly) “random” sequences are created equal, *Proceedings of the National Academy of Sciences of the USA*, **94**, 3513–3518.
17. Pincus, S., and Singer, B. H. (1996) Randomness and degrees of irregularity, *Proceedings of the National Academy of Sciences of the USA*, **93**, 2083–2088.
18. Rao, C. R. and C. K. Mitra. (1971) *Generalized Inverse of Matrices and its Applications*. J. Wiley, New York. pp 171-173.
19. Read, T.R. and Cressie, N.A.. (1988) *Goodness-Of-Fit Statistics for Discrete Multivariate Data*. Springer, New York. p 46.
20. Seife, C. (1997) New test sizes up randomness. *Science*, **276**, 532.
21. Vajda, I. (1989) *Theory of Statistical Inference and Information*. Kluwer, Dordrecht. pp 300–328.