Approximate Entropy for Testing Randomness

Andrew L. Rukhin

Abstract

In this paper a new concept of approximate entropy is modified and applied to the problem of testing for randomness a string of binary bits. This concept 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 $e, \pi, \sqrt{2}$ and $\sqrt{3}$ with the surprising conclusion that the expansion of $\sqrt{3}$ demonstrated much more 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 randomess 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.

The key step leading to the limiting distribution of approximate entropy is a modification of its definition based on the frequencies of different patterns in the augmented or circular version of the original sequence. In Section 2 it is shown that the approximate entropy as well as its modified version converges in distribution to a χ^2-random variable when the length of a template, m, is fixed. A similar result when m increases to infinity is obtained in Section 3. In this situation the limiting distribution is normal with the parameters of this law determined from Poisson approximation. These facts provide the basis for statistical tests of randomness via the approximate entropy. In particular, tail probabilities for the approximate entropy test can be evaluated. A graph of these values for binary expansions of e, π and $\sqrt{3}$ illustrates the use of this concept.

Key words: Decomposable Statistics, Entropy, Information Divergence, Multinomial Distribution, Poisson Distribution, χ^2-distribution.

Andrew L. Rukhin is a Professor in the Department of Mathematics and Statistics at University of Maryland at Baltimore County, Baltimore, MD, 21250. He also has a faculty appointment in the Statistical Engineering Division at the National Institute of Standards and Technology, Gaithersburg, MD 20899-0001. This work has been motivated by a joint project with the Computer Security Division of the National Institute of Standards and Technology.

1 Introduction: Approximate Entropies

In this paper I apply a new concept of approximate entropy and its modification to the problem of testing for randomness a string of binary bits. This problem gained importance with the wide use of public key cryptography and the need for good 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.

To measure the degree of randomness of observed sequences Pincus and Singer (1996) suggested to use a general 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)). It is based on the likelihood that templates in the sequence that are similar will remain similar on next incremental comparisons.

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

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

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.

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

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

with $ApEn(0) = -\Phi^{(1)}$. "$ApEn(m)$ 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 $ApEn(m)$ imply strong regularity, or persistence, in a sequence. Alternatively, large values of $ApEn(m)$ 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 $ApEn(m)$ takes the largest possible value. Pincus and Kalman (1997) evaluated quantities $ApEn(m), m = 0, 1, 2$ for binary and decimal expansions of $e, \pi, \sqrt{2}$ and $\sqrt{3}$ with the surprising conclusion that the expansion of $\sqrt{3}$ demonstrated much more irregularity than that of π.

Since $-\Phi^{(m)}$ is the entropy of the empirical distribution which under the randomness assumption must be almost uniform, one should expect that for fixed m, $\Phi^{(m)} \sim -m \log s$ and $ApEn(m) = \Phi^{(m)} - \Phi^{(m+1)} \rightarrow \log s$; indeed this fact follows from Theorem 2 in Pincus (1991). As far as the limiting behavior of $ApEn(m) - \log s$, Pincus and Huang (1992), p 3072, indicate that "analytic proofs of asymptotic normality and especially explicit variance estimates for $ApEn$ appear to be extremely difficult".

The key step leading to the limiting distribution of approximate entropy is a modification of its definition. Introduce the modified version of the empirical distribution entropy $-\Phi^{(m)}$ as

$$\tilde{\Phi}^{(m)} = \sum_{i_1 \cdots i_m} \nu_{i_1 \cdots i_m} \log \nu_{i_1 \cdots i_m}. \tag{1}$$

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

Define the modified approximate entropy as

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

A definite advantage of $\widetilde{ApEn}(m)$ is that by Jensen's inequality, $\log s \geq \widetilde{ApEn}(m)$ for any m, whereas it is possible that $\log s < ApEn(m)$ (albeit the probability of this tends to zero as n increases.) Therefore the largest possible value of $\widetilde{ApEn}(m)$ is merely $\log s$. The maximally random sequences under this definition have the relative frequencies of all patterns (in a circular version of the sequence) of a given length are as close to the common value n^{-1} as possible. For example, in addition to maximally random binary strings from the point of view of $ApEn(1)$ $1,1,0,0,1$, $1,0,0,1,1$, $0,0,1,1,0$, and $0,1,1,0,0$ mentioned by Pincus and Singer (1996) p 2084, one should add two sequences $1,0,0,0,1$, $0,1,1,1,0$, which are random from the point of view of $\widetilde{ApEn}(1)$.

On the other hand $ApEn(m)$ and $\widetilde{ApEn}(m)$ cannot differ much if n is large. Indeed for $Y_i(m) = (i_1, \dots, i_m)$, put $\nu'_{i_1 \cdots i_m} = C_i^m$, so that

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

Then with $\omega'_{i_1 \cdots i_m} = (n - m + 1)\nu'_{i_1 \cdots i_m}$

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

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

$$\left| \nu_{i_1 \cdots i_m} - \nu'_{i_1 \cdots i_m} \right| \leq \frac{m-1}{n-m+1}, \tag{3}$$

which suggests that for a fixed m, Pincus' approximate entropy and $\widetilde{ApEn}(m)$ must be close when n is large.

In the next Section I derive the limiting distribution of $n[\log s - \widetilde{ApEn}(m)]$ when $n \to \infty$ and m is fixed. It is also proven that $n[ApEn(m) - \widetilde{ApEn}(m)] = O_P(n^{-1})$, so that the limiting distributions of Pincus' approximate entropy and of $\widetilde{ApEn}(m)$ coincide. Section 3 contains a similar result when $m \to \infty$. These facts provide the basis for statistical tests of randomness via the approximate entropy.

In particular, in Section 4 the tail probabilities for the approximate entropy test are evaluated and plotted for binary expansions of e, π and $\sqrt{3}$.

2 Asymptotic Behavior of Approximate Entropy: Fixed m.

It is shown here that the limiting distribution of $2n[\log s - \widetilde{ApEn}(m)]$ as well as of $2n[\log s - ApEn(m)]$ is that of a χ^2-random variable with $(s-1)s^m$ degrees of freedom.

Proposition 1 *For fixed m as $n \to \infty$ one has the following convergence in distribution*

$$2n\left[\log s - \widetilde{ApEn}(m)\right] \to \chi^2(s^{m+1} - s^m).$$

Also

$$n[ApEn(m) - \widetilde{ApEn}(m)] = O_P\left(\frac{1}{n}\right), \tag{4}$$

so that

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

Proof Let us start with the limit theorem for $\widetilde{ApEn}(m)$. Put

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

Then the vector formed by $Z_{i_1\cdots i_m}$ has asymptotic multivariate normal distribution with zero mean and the covariance matrix of the form

$$\Sigma_m = \frac{1}{s^m}\mathbf{I}_m - \frac{1}{s^{2m}}\mathbf{e}_m\mathbf{e}_m^T.$$

Here $\mathbf{I}_m$ denotes the $s^m \times s^m$ identity matrix and $\mathbf{e}_m^T = (1, \ldots, 1)$ is a s^m-dimensional vector. Since with probability one, $\sum Z_{i_1\cdots i_m} = 0$, (1) shows that

$$\tilde{\Phi}^{(m)} = -\sum_{i_1\cdots i_m}\left[\frac{1}{s^m} + \frac{Z_{i_1\cdots i_m}}{\sqrt{n}}\right]\left[-m\log s + \frac{s^m Z_{i_1\cdots i_m}}{\sqrt{n}} - \frac{s^{2m} Z^2_{i_1\cdots i_m}}{2n} + O_P\left(\frac{1}{n^{3/2}}\right)\right]$$

$$\sim -m \log s + \frac{s^m}{2n} \sum_{i_1 \cdots i_m} Z^2_{i_1 \cdots i_m}.$$

Using a similar notation for patterns of length $m+1$, let $\nu_{i_1 \cdots i_m i_{m+1}}$ be the relative frequencies, and let $Z_{i_1 \cdots i_m i_{m+1}}$ denote the corresponding differences between empirical and theoretical probabilities. Then

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

and

$$\tilde{\Phi}^{(m+1)} \sim -(m+1) \log s + \frac{s^{m+1}}{2n} \sum_{i_1 \cdots i_m i_{m+1}} Z^2_{i_1 \cdots i_m i_{m+1}}.$$

Thus

$$\begin{aligned} &\tilde{\Phi}^{(m)} - \tilde{\Phi}^{(m+1)} \\ \sim \log s - \frac{s^m}{2n} &\left[\sum_{i_1 \cdots i_m} \left(\sum_k Z_{i_1 \cdots i_m k} \right)^2 - s \sum_{i_1 \cdots i_m i_{m+1}} Z^2_{i_1 \cdots i_m i_{m+1}} \right] \\ &= \log s - \frac{s^m}{2n} Z^T \mathbf{Q} Z \end{aligned}$$

with the $s^{m+1} \times s^{m+1}$ block-diagonal matrix $\mathbf{Q}$ formed by formed by s^m blocks Q_0 of size $s \times s$,

$$Q_0 = s\mathbf{I}_1 - \mathbf{e}_1 \mathbf{e}_1^T,$$

and s^{m+1}-dimensional normal vector Z. The distribution of the quadratic form $Z^T \mathbf{Q} Z$ is that of $\sum l_{i_1 \cdots i_m i_{m+1}} W^2_{i_1 \cdots i_m i_{m+1}}$ with independent standard normal variables $W_{i_1 \cdots i_m i_{m+1}}$ and $l_{i_1 \cdots i_m i_{m+1}}$ denoting the eigenvalues of the matrix $\Sigma^{1/2} \mathbf{Q} \Sigma^{1/2}$.

It is easy to check that

$$\Sigma^{1/2}_{m+1} = \frac{1}{s^{(m+1)/2}} \mathbf{I}_{m+1} - \frac{1}{s^{3(m+1)/2}} \mathbf{e}_{m+1} \mathbf{e}^T_{m+1},$$

and $\Sigma^{1/2}_{m+1} \mathbf{Q} \Sigma^{1/2}_{m+1} = \frac{1}{s^{m+1}} \mathbf{Q}$.

The evaluation of the determinant, $det \left[\Sigma^{1/2}_{m+1} \mathbf{Q} \Sigma^{1/2}_{m+1} - l \mathbf{I}_{m+1} \right]$, shows the needed eigenvalues are equal to s with multiplicity $(s-1)s^m$ and 0 with multiplicity s^m. Therefore

$$\tilde{\Phi}^{(m)} - \tilde{\Phi}^{(m+1)} \sim \log s - \frac{1}{2n} \chi^2((s-1)s^m)$$

and

$$n\left[\log s - \widetilde{ApEn}(m)\right] \sim \frac{1}{2}\chi^2(s^{m+1} - s^m).$$

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

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

Thus (4) follows and the Proposition 1 is proven. □

For the observed value $ApEn(m)$, one has to define $\chi^2(obs)$ as $\chi^2(obs) = 2n\,|\log s - ApEn(m)|$, whereas, as has been noticed, the difference $\log s - \widetilde{ApEn}(m)$ 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 asymptotic distribution of the statistics $2n\left[\log s - \widetilde{ApEn}(m)\right]$ and $2n\left[\log s - ApEn(m)\right]$, evaluated under the alternative of the form $\pi_{i_1\cdots i_m i_{m+1}} = s^{-m-1} + n^{-1/2}\eta_{i_1\cdots i_m i_{m+1}}$, with $\eta^T\mathbf{e} = 0$, is a noncentral χ^2-distribution with $s^{m+1} - s^m$ degrees of freedom and the noncentrality parameter $\eta^T\eta/s^{m+1}$. This fact allows for an approximate power function of the corresponding test of randomness.

3 Asymptotic Behavior of Approximate Entropy: Large m.

In this section I consider the situation when both n and m tend to infinity so that

$$\frac{n}{s^{m+1}} \to \lambda > 0. \tag{5}$$

(Actually this condition can be relaxed to $\min_n \frac{n}{s^{m+1}} > 0$. Observe that the numerical evaluation of $ApEn(m)$ and $\widetilde{ApEn}(m)$ for such values of m is feasible; for example, the values $s = 2, n = 10^5, m = 17$ with $\lambda = 0.38147..$ have been tried.)

To investigate this case let us write the formula for the modified approximate entropy in the following form

$$\widetilde{ApEn}(m) = \sum_{i_1\cdots i_m} \nu_{i_1\cdots i_m} \log \nu_{i_1\cdots i_m} - \sum_{i_1\cdots i_m i_{m+1}} \nu_{i_1\cdots i_m i_{m+1}} \log \nu_{i_1\cdots i_m i_{m+1}}$$

$$= \frac{1}{n} \sum_{i_1\cdots i_m} \left[\omega_{i_1\cdots i_m 1} \log \left(\frac{\omega_{i_1\cdots i_m 1}}{\sum_k \omega_{i_1\cdots i_m k}}\right) + \cdots + \omega_{i_1\cdots i_m s} \log \left(\frac{\omega_{i_1\cdots i_m s}}{\sum_k \omega_{i_1\cdots i_m k}}\right)\right]$$

$$= \frac{1}{n} \sum_{i_1\cdots i_m} f\left(\omega_{i_1\cdots i_m 1}, \ldots, \omega_{i_1\cdots i_m s}\right) \qquad (6)$$

with $f(u_1, \ldots, u_s)$ denoting the entropy of the probability distribution defined by probabilities $u_k / \sum u_j$, $k = 1, \ldots, s$,

$$f(u_1, \ldots, u_s) = -u_1 \log \left(\frac{u_1}{\sum_j u_j}\right) - \cdots - u_s \log \left(\frac{u_s}{\sum_j u_j}\right).$$

Note that our function f has a special form, namely, $f(u_1, \ldots, u_s) = \sum_j \phi(u_j) - \phi(\sum_j u_j)$ with $\phi(u) = -u \log u$.

A similar representation with n replaced by $n - m + 1$ also holds for $ApEn(m)$. Indeed in the notation of Section 2, $\omega'_{i_1\cdots i_m} - \sum_k \omega'_{i_1\cdots i_m k} \leq 1$ and there exists no more than one m-tuple $i_1, \ldots, i_m$ for which $\omega'_{i_1\cdots i_m} \neq \sum \omega'_{i_1\cdots i_m k}$. Therefore

$$\left| \sum_{i_1\cdots i_m} \omega'_{i_1\cdots i_m} \log \frac{\omega'_{i_1\cdots i_m}}{n-m+1} - \sum_{i_1\cdots i_m i_{m+1}} \omega'_{i_1\cdots i_m i_{m+1}} \log \frac{\omega'_{i_1\cdots i_m i_{m+1}}}{n-m+1} \right|$$

$$\leq \left[\max_{0 \leq x \leq n-m+1} [(x+1)\log(x+1) - x \log x] + \log(n-m+1)\right] \leq 2 \log n,$$

so that

$$ApEn(m) = \frac{1}{n-m+1} \sum_{i_1\cdots i_m} \left[\omega'_{i_1\cdots i_m 1} \log \left(\frac{\omega'_{i_1\cdots i_m 1}}{\sum_k \omega'_{i_1\cdots i_m k}}\right) + \cdots \right.$$

$$\left. + \omega'_{i_1\cdots i_m s} \log \left(\frac{\omega'_{i_1\cdots i_m s}}{\sum_k \omega'_{i_1\cdots i_m k}}\right)\right] + O_P\left(\frac{\log n}{n}\right).$$

Thus $ApEn$ also admits the representation (6), and the limiting distribution of both $\widetilde{ApEn}$ and $ApEn$ is that of this decomposable statistic. Sums of this form (with functions f of only one argument) have been extensively studied. See Holst (1972), Morris (1975) and Medvedev (1977). Although our situation with f depending on s frequencies $\omega_{i_1\cdots i_m 1}, \ldots, \omega_{i_1\cdots i_m s}$ does not follow directly from these results, the special form of this function leads to the following Proposition 2 which can be derived from Holst (1972) after some modifications.

Let $\Pi_{i_1\cdots i_m 1}, \ldots, \Pi_{i_1\cdots i_m s}$ denote s independent Poisson random variables with parameter λ. It is also convenient to write $\Pi_1, \ldots, \Pi_s$ or $\Pi_1(\lambda), \ldots, \Pi_s(\lambda)$ for a s-tuple of such random variables. Put

$$\mu_n = \frac{1}{n} \sum_{i_1\cdots i_m} Ef\left(\Pi_{i_1\cdots i_m 1}, \ldots, \Pi_{i_1\cdots i_m s}\right) = \frac{s^m}{n} Ef(\Pi_1, \ldots, \Pi_s)$$

$$= \frac{1}{s\lambda} Ef(\Pi_1, \ldots, \Pi_s),$$

and

$$\gamma = \frac{\mathbf{Cov}\left(f(\Pi_1, \ldots, \Pi_s), \Pi_1 + \cdots + \Pi_s\right)}{\mathbf{Var}\left(\Pi_1 + \cdots + \Pi_s\right)}$$

$$= \frac{1}{s\lambda} \mathbf{Cov}\left(f(\Pi_1, \ldots, \Pi_s), \Pi_1 + \cdots + \Pi_s\right).$$

With

$$g(u_1, \ldots, u_s) = f(u_1, \ldots, u_s) - Ef(\Pi_1, \ldots, \Pi_s) - \gamma[\Pi_1 + \cdots + \Pi_s - s\lambda],$$

one has

$$\mathbf{Var} g(\Pi_1, \ldots, \Pi_s) = \mathbf{Var} f(\Pi_1, \ldots, \Pi_s) - s\lambda\gamma^2,$$

so that the sums

$$U_n = \sum_{i_1\cdots i_m} g\left(\Pi_{i_1\cdots i_m 1}, \ldots, \Pi_{i_1\cdots i_m s}\right)$$

and

$$V_n = \frac{1}{\sqrt{n}} \sum_{i_1\cdots i_m} \left[\Pi_{i_1\cdots i_m 1} + \cdots + \Pi_{i_1\cdots i_m s} - s\lambda\right]$$

are uncorrelated, $EU_nV_n = 0$. Let

$$\sigma_n^2 = \sum_{i_1\cdots i_m} \mathbf{Var} g\left(\Pi_{i_1\cdots i_m 1}, \ldots, \Pi_{i_1\cdots i_m s}\right) = s^m \mathbf{Var} g(\Pi_1, \ldots, \Pi_s);$$

then the joint asymptotic distribution of U_n/σ_n and V_n is normal with zero mean and the identity covariance matrix. The conditional distribution of U_n/n given $V_n = 0$ coincides with the distribution of $\widetilde{ApEn}(m) - \mu_n$, since the conditional distribution of $(\Pi_{i_1\cdots i_m1}, \ldots, \Pi_{i_1\cdots i_ms})$ given that $\sum[\Pi_{i_1\cdots i_m1} + \cdots + \Pi_{i_1\cdots i_ms}] = n$ is multinomial.

Therefore the following result concerning the convergence of $n[\widetilde{ApEn}(m) - \mu_n]/\sigma_n$ and of $(n - m + 1)[ApEn(m) - \mu_n]/\sigma_n$ to a standard normal distribution is not surprising.

Proposition 2 *Under condition (5) for $n \to \infty$*

$$P\left(n\frac{\widetilde{ApEn}(m) - \mu_n}{\sigma_n} \leq x\right) \to \Phi(x)$$

and

$$P\left(n\frac{ApEn(m) - \mu_n}{\sigma_n} \leq x\right) \to \Phi(x)$$

Sketch of the Proof The argument above can be made rigorous by examination of the characteristic function of $\widetilde{ApEn}(m)$ as in Lemmas 2.1, 2.2, A1, A2 and A3 of Holst (1972). With $N = s^{m+1}$ as in Lemma 2.1 there

$$A_N(z) = \sum_{n=0}^{\infty} E_n \left[\prod_{i_1\cdots i_m} x_{i_1\cdots i_m}^{f(\nu_{i_1\cdots i_m1},\ldots,\nu_{i_1\cdots i_ms})}\right] \frac{(Nz)^n e^{-Nz}}{n!}$$

$$= \prod_{i_1\cdots i_m} \sum_{j_1\ldots j_s} \frac{z^{j_1+\cdots+j_s} e^{-z}}{j_1!\cdots j_s!} x_{i_1\cdots i_m}^{f(\Pi_{i_1\cdots i_m1}(z),\ldots,\Pi_{i_1\cdots i_ms}(z))}.$$

A similar representation for the characteristic function $\varphi(t) = E\exp\{it\sum_{i_1\cdots i_m} f(\nu_{i_1\cdots i_m1}, \ldots, \nu_{i_1\cdots i_ms})\}$ as in Lemma 2.2 follows; the only difference is that the ordinary sum in the right-hand side is replaced by the multiple sum

$$e^{-s\lambda e^{i\theta}} \sum_{j_1\ldots j_s} \frac{(\lambda e^{i\theta})^{j_1+\cdots+j_s}}{j_1!\cdots j_s!} \left[e^{it[\phi(j_1)+\cdots+\phi(j_s)]-it\phi(j_1+\ldots+j_s)} - 1\right].$$

The same estimates as in Lemmas A1 and A2 hold for the corresponding function. The convergence result in Lemma A2 also holds by analysis of Taylor's expansion. □

For example, when $s = 2$, one has to evaluate

$$\mu(\lambda) = E\Pi_1(\lambda)\log\Pi_1(\lambda) = e^{-\lambda}\lambda\sum_{k=1}^{\infty}\frac{\lambda^k\log(k+1)}{k!},$$

and then

$$\mu_n = \frac{\mu(2\lambda) - 2\mu(\lambda)}{2\lambda}.$$

Similarly with

$$\zeta(\lambda) = E\Pi_1(\lambda)\log\Pi_1(\lambda)[\Pi_1(\lambda) - \lambda],$$

one has

$$\gamma = \frac{\zeta(2\lambda) - 2\zeta(\lambda)}{2\lambda}.$$

Also if

$$\sigma^2(\lambda) = \mathbf{Var}\left(\Pi_1(\lambda)\log\Pi_1(\lambda)\right),$$

and

$$\nu(\lambda) = \mathbf{Cov}\left([\Pi_1(\lambda)+\Pi_2(\lambda)]\log[\Pi_1(\lambda)+\Pi_2(\lambda)], \Pi_1(\lambda)\log\Pi_1(\lambda)\right)$$

$$= e^{-2\lambda}\lambda\sum_{k=1}^{\infty}\frac{\lambda^k\log(k+1)}{k!}\beta_k - \mu(2\lambda)\mu(\lambda),$$

where

$$\beta_k = \sum_{j=1}^{k}(j+1)\log(j+1)\left(\begin{array}{c} k+1 \\ j+1 \end{array}\right),$$

then

$$\mathbf{Var}f(\Pi_1, \Pi_2) = \sigma^2(2\lambda) + 2\sigma^2(\lambda) - 4\nu(\lambda).$$

Thus

$$\sigma_n^2 = s^m\left[\sigma^2(2\lambda) + 2\sigma^2(\lambda) - 4\nu(\lambda) - \frac{[\zeta(2\lambda) - 2\zeta(\lambda)]^2}{2\lambda}\right].$$

More generally, the asymptotic distribution of the sum

$$\mathbf{S} = \frac{1}{n}\sum_{i_1\cdots i_m} f\left(\omega_{i_1\cdots i_m 1}, \ldots, \omega_{i_1\cdots i_m s}\right)$$

when $f(u_1, \ldots, u_s) = \sum_j \phi(u_j) - \phi(\sum_j u_j)$ can be shown to be normal under mild regularity conditions on function ϕ.

The asymptotic power of this test statistic $\mathbf{S}$ under the alternative $\pi_{i_1\cdots i_m}$ is determined by the ratio $R = \lim[E_\pi \mathbf{S} - \mu_n]/\sigma_n$, whose absolute value is to be maximized to have the optimal Pitman efficiency.

Under the alternative of the form $\pi_{i_1\cdots i_m i_{m+1}} = s^{-m-1} + n^{-1/4}\zeta_{i_1\cdots i_m i_{m+1}}$ with

$$\zeta_{i_1\cdots i_m} = \int_{i_1+i_2/s+\cdots+i_m s^{-m}}^{i_1+i_2/s+\cdots+(i_m+1)s^{-m}} q(u)\; du$$

such that $\int_0^1 q(u)\; du = 0$,

$$R = \int_0^1 q^2(u)\; du \frac{E\left[\sum_i \phi(\Pi_i) - \phi(\sum_i \Pi_i)\right]\left[(\sum_i \Pi_i - s\lambda)^2 - \sum_i \Pi_i\right]}{\mathbf{Var}^{1/2}\left[\sum_i \phi(\Pi_i) - \phi(\sum_i \Pi_i) - \gamma \sum_i \Pi_i\right]}.$$

This formulas can be used to show that there is no optimal statistic $\mathbf{S}$. This is to be contrasted with asymptotic optimality of χ^2-test in the class of decomposable statistics with function f depending only on one argument (see Holst, 1972, Ivchenko and Medvedev, 1978, 1980).

Essentially the same conclusions about the power of the approximate entropy test as about χ^2-test (Kallenberg et al, 1985) can be made.

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

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

Thus

$$ApEn(1) = 0.0014...$$

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

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

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

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

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

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

and

$$ApEn(1) = 0.6699...$$

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

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

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

Thus from the point of view of approximate entropies $ApEn(1)$ and $\widetilde{ApEn}(1)$ the sequence (A) does not look random at all, but the string (B) does and even more so for the modified entropy $\widetilde{ApEn}(1)$. The strings (A) and (B) are also examined in Pincus and Kalman (1997) p 3514, with a numerical mistake.

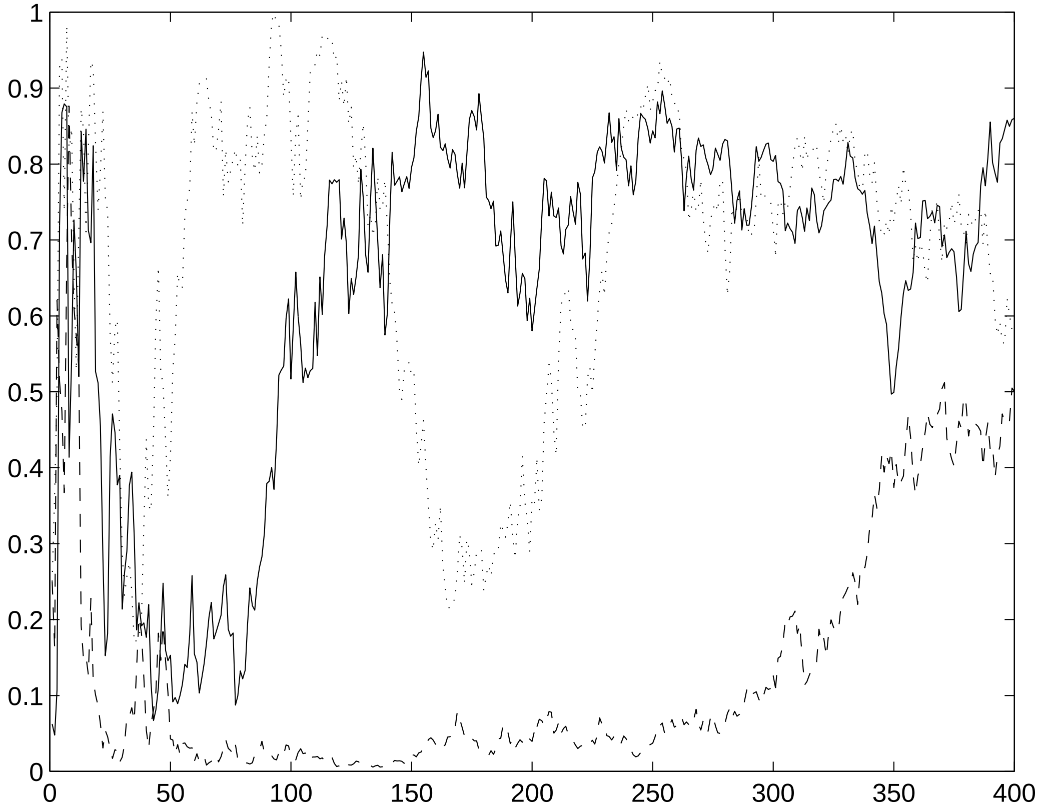

Figure 1 Consecutive P-values for binary expansions of $\sqrt{3}$ (broken line), π (dotted line) and e (solid line) when $m = 1$.

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 much smaller than those of e and π. The situation, however, is reverse for $m = 7$. when the digits of π and $\sqrt{3}$ look much more random than these of of expansion of e (Figure 2).

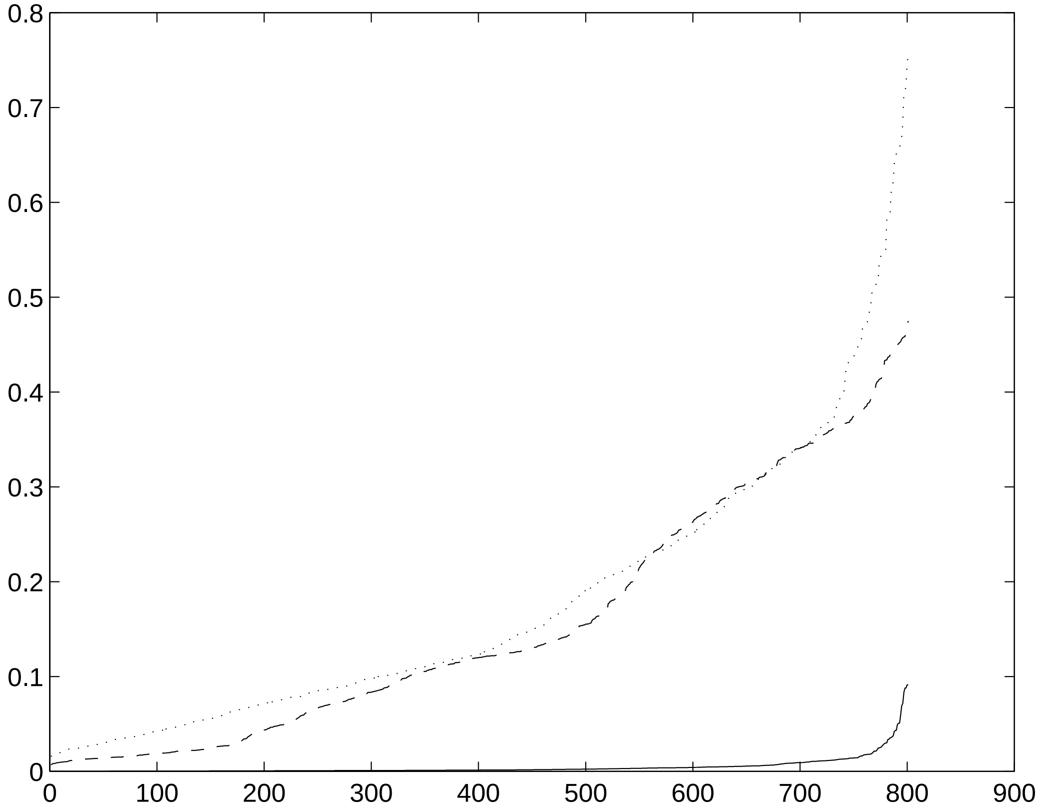

Figure 2 Consecutive P-values for binary expansions of $\sqrt{3}$ (broken line), π (dotted line) and e (solid line) when $m = 7$.

The procedures based on randomess test via approximate entropy form now a part of a battery of empirical tests for randomness developed at the Computer Security Division of the National Institute of Standards and Technology. They are being used for investigation of various existing random numbers generators, such as Data Encryption Algorithm, Secure Hash Algorithm, Digital Signature Algorithm and Blum, Blum and Shub generator.

References

[1] Chaitin, G. (1975), “Randomness and mathematical proof,” *Scientific American*, 232, pp 47–52.

[2] Holst, L. (1972), "Asymptotic normality and efficiency for certain goodnes-of-fit tests", *Biometrika*, 59, pp 137–145.

[3] Ivchenko, G., and Medvedev, Yu.I. (1978), "Separable statistics and hypotheses testing. The case of small samples," *Theory of Probability and Its Applications*, 23, pp 764–775.

[4] Ivchenko, G., and Medvedev, Yu.I. (1978), "Decomposable statistics and hypothesis testing for grouped data," *Theory of Probability and Its Applications*, 25, pp 540–551.

[5] Kallenberg, W. C. M., Oosterhoff, J., and Schriver, B.F. (1985), "The number of classes in chi-squared goodness-of-fit tests," *Journal of the American Statistical Association*, 80, pp 959–968.

[6] Medvedev, Yu. I. (1977), "Separable statsitics in a polynomial scheme. I," *Theory of Probability and Its Applications*, 22, pp 1–15.

[7] Morris, C. (1975), "Central limit theorem for multinomial sums," *Annals of Statistics*, 3, pp 165–188.

[8] Pincus, S. (1991), "Approximate entropy as a measure of system complexity," *Proceedings of the National Academy of Sciences of the USA*, 88, pp 2297–2301.

[9] Pincus, S., and Huang, W.-M. (1992), "Approximate entropy, statistical properties and applications," *Communications in Statistics, Part A-Theory and Methods*, 21, pp 3061–3077.

[10] 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, pp 3513–3518.

[11] Pincus, S., and Singer, B. H. (1996), "Randomness and degrees of irregularity," *Proceedings of the National Academy of Sciences of the USA*, 93, pp 2083–2088.