KYBERNETIKA — VOLUME 31 (1995), NUMBER 4, PAGES 321–330

STATISTICAL TESTS OF OPTIMALITY OF SOURCE CODES[1]

IGOR VAJDA

For newly defined data compaction codes, as well as for the traditional data compression codes, we prove an asymptotic uniformity of probabilities of codewords, a kind of the asymptotic equipartition property. On the basis of this we propose an easily applicable Neyman–Pearson test of optimality of a code with a given asymptotic level of significance $0 < \alpha < 1$. The test is based on the sample entropy of code.

1. INTRODUCTION

There is a common intuitive belief that a good code is approximately uniform in the sense that all codewords are nearly equally likely. If this is true then one can test the hypothesis that a code $C : \mathcal{X} \mapsto \mathcal{D}^*$ of a source $(\mathcal{X}, P)$ is good as follows. Let $\mathcal{X} = \{x_1, \dots, x_m\}$, $P = (p(x_1), \dots, p(x_m))$, and let $X_1, \dots, X_N$ be independent realizations of a message X from $(\mathcal{X}, P)$. Put $\hat{P}_N = (\hat{p}_N(x_1), \dots \hat{p}_N(x_m))$ where $\hat{p}_N(x)$ denotes the relative frequency of the codeword $C(x)$ corresponding to x (from the set $\mathcal{D}^*$ of finite length binary strings). Denote by $H(X) = H(P)$ the source entropy and by $H(\hat{P}_N)$ the entropy of a fictive random source $(\mathcal{X}, \hat{P}_N)$ depending on realizations $X_1, \dots, X_N$. Finally, let $h > 0$ be smaller than but close to $\log m = \max_P H(P)$. If there exists a set $K_N \subset [0, \log m]$ such that

$$\sup_{P: h \le H(P) \le \log m} \Pr\left\{H(\hat{P}_N) \in K_N\right\} = \alpha \tag{1}$$

then $[H(\hat{P}_N), K_N]$ is a Neyman–Pearson α-level test for our problem (cf. Lehman [7]).

There are two serious obstacles on this road:

(i) It is not clear whether good codes are always uniform, and

(ii) the Neyman–Pearson test satisfying (1) is unknown.

Nevertheless the engineers are visibly forced to move in this direction. Indeed, there is a number of papers where $H(\hat{P}_N)$ is calculated and if "not too far" from $\log m$, then the hypothesis that a code is good is "confirmed".

[1] Supported by the Grant Agency of the Czech Republic under Grant No. 102/94/0320.

In this paper we prove that the hypothesis in (i) is true in a quite general sense, and overcome the difficulty (ii) by introducing a Neyman–Pearson asymptotically α-level entropic test $[H(\hat{P}_N), K_N]$ for any given $0 < h \leq \log m$ and $0 < \alpha < 1$.

Remind that $(H(\hat{P}_N), K_N)$ is said asymptotically α-level if (cf. [7])

$$\sup_{P: h \leq H(P) \leq \log m} \lim_{N \to \infty} \Pr\left\{H(\hat{P}_N) \in K_N\right\} = \alpha. \tag{2}$$

Since

$$K_N = K_N(\alpha, h) = [0,\, c_N(\alpha, h)] \subset [0, \log m), \tag{3}$$

our test is in fact defined by a critical level $0 < c_N(\alpha, h) < \log m$.

Note that for the particular case $h = \log m$ there exists a well known Neyman–Pearson asymptotically α-level statistical test, namely the Pearson's goodness-of-fit χ^2-test $[\chi^2(\hat{P}_N), K_N^*]$, where

$$\chi^2(\hat{P}_N) = \frac{N}{m} \sum_x (m\,\hat{p}_N(x) - 1)^2$$

and $K_N^* = [0, c_N^*(\alpha)]$ is defined by the critical value

$$c_N^*(\alpha) = \chi^2_{1-\alpha, m-1} \qquad \left(\doteq \frac{1}{2}\left[\sqrt{2m-3} + \Phi_{(1-\alpha)}\right]^2 \text{ if } m \geq 30 \right).$$

Here $\chi^2_{\alpha, m-1}$ denotes the α-quantile of the χ^2-distribution with $m-1$ degrees of freedom and Φ_α the α-quantile of the standard normal distribution. Unfortunately, this test is of a little practical use in our problem, as the good codes are almost never exactly uniform so that, for large sample sizes n, the χ^2-test usually rejects the hypothesis that the code is good even at the levels exceeding 0.2.

Feistauerová [4] introduced the particular entropic test $[H(\hat{P}_N), K_N]$ for $h = \log m$, i.e. with $K_N = K_N(\alpha, \log m)$, and found that this test shares the practical disadvantages with the goodnes-of-fit χ^2-test. The extension of this test to $0 < h < \log m$ is thus inevitable. From a purely statistical point of view, this is a nontrivial step as the simple hypothesis $\mathcal{H} = \{H(P) = \log m\}$ is becoming composite, $\mathcal{H} = \{h \leq H(P) \leq \log m\}$. This step is made below with the help of Feistauerová, Vajda [5], where the problem of testing general composite entropic hypotheses is solved.

2. GOOD CODES ARE UNIFORM

One cannot say whether a good code is uniform without specifying what a good code is. Obviously, this is not a difficult problem as the whole information theory concentrates arround the concept of a good code. We shall make this problem a bit harder by trying to go beyond the scope of traditional definitions of Shannon information theory.

Consider sources $\mathcal{X}$ of a product type, i.e. replace $(\mathcal{X}, P)$ by $(\mathcal{X}^n, Q)$ with Q not necessarily equal P^n. Further let m_n be an increasing sequence of naturals and consider strings of binary random variables

$$(Y_1,\ldots,Y_{m_n}) = \begin{cases} C(X_1,\ldots,X_n),\ Y^*_{\ell+1},\ldots,Y^*_{m_n} \\ m_n\text{-prefix of } C(X_1,\ldots,X_n), \end{cases}$$

where $C : \mathcal{X}^n \mapsto \mathcal{D}^*$ is a similar code as considered in Sec. I, and where the upper possibility takes place iff $\ell \triangleq |C(X_1,\ldots,X_n)| < m_n$ (then $Y^*_{\ell+1},\ldots,Y^*_{m_n}$ are arbitrary dummy random variables) while the lower possibility takes place in all other cases. Consider a decomposition $\mathcal{X}^n = A_n + B_n$ where A_n contains all messages $x = (x_1,\ldots,x_n)$ uniquely decodable from $y = (y_1,\ldots,y_{m_n}) \in \mathcal{D}^{m_n}$ and B_n all the remaining ones. Define a mapping $\varphi : \mathcal{D}^{m_n} \mapsto A_n \cup \{B_n\}$ by

$$\varphi(y) = \begin{cases} C^{-1}(y) & \text{if } C^{-1}(y) \text{ is unique} \\ B_n & \text{otherwise,} \end{cases}$$

and consider the "code entropy" $H(Y_1,\ldots,Y_{m_n})$.

Since $H(\varphi(Y_1,\ldots,Y_{m_n})) \le H(Y_1,\ldots,Y_{m_n})$ (cf. Problem 5 on p. 43 of Cover and Thomas [3]), it holds

$$\sum_{x\in A_n} -q(x)\log q(x) - Q(B_n)\log Q(B_n) \le H(Y_1,\ldots,Y_{m_n}) \le \log|\mathcal{D}^{m_n}| = m_n,$$

where the left-hand side equals

$$\begin{aligned}
& H(X_1,\ldots,X_n) + \sum_{x\in B_n} q(x)\log q(x) - \sum_{x\in B_n} q(x)\log Q(B_n) \\
= \ & H(X_1,\ldots,X_n) + \sum_{x\in B_n} q(x)\log\frac{q(x)}{Q(B_n)} \\
= \ & H(X_1,\ldots,X_n) + Q(B_n)\sum_{x\in B_n}\frac{q(x)}{Q(B_n)}\log\frac{q(x)}{Q(B_n)} \\
= \ & H(X_1,\ldots,X_n) + |B_n|\,Q(B_n)\sum_{x\in B_n}\frac{1}{|B_n|}\phi\left(\frac{q(x)}{Q(B_n)}\right) \\
\ge \ & H(X_1,\ldots,X_n) + |B_n|\,Q(B_n)\,\phi\left(\frac{1}{|B_n|}\right) & (*) \\
= \ & H(X_1,\ldots,X_n) - Q(B_n)\log|B_n| \\
\ge \ & H(X_1,\ldots,X_n) - n\,Q(B_n)\log|\mathcal{X}|.
\end{aligned}$$

Note that $(*)$ follows from the Jensen inequality applied to the convex function $\phi(t) = t\log t$. Hence we have proved the inequality

$$\frac{H(X_1,\ldots X_n)}{m_n} - \frac{n}{m_n}Q(B_n)\log|\mathcal{X}| \le \frac{H(Y_1,\ldots,Y_{m_n})}{m_n} \le 1, \tag{4}$$

where the set B_n depends on the code and on the sequence m_n.

The above considered code depends on the source size n, i.e. $C = C_n$. Let us consider a family of codes $C_{n,\varepsilon}$ depending not only on n but also on a real parameter $\varepsilon > 0$.

In accordance with the spirit of Shannon information theory we say that $C_{n,\varepsilon}$ is (asymptotically) good if for every sequence $m_n \geq H(X_1, \dots, X_n) + \varepsilon\, n$ with $\varepsilon > 0$

$$\lim_{\varepsilon \downarrow 0} \limsup_{n \to \infty} Q(B_n) = 0.$$

For example, if $Q = P^n$ and the source entropy rate $H = H(X_1, \dots, X_n)/n$ is positive then it follows from the AEP property (cf. Theorem 3.1.2 in [3]) and from (4) that the block $(n, [H+\varepsilon]+1)$-codes $C_{n,\varepsilon}$ for A_n equal to the set $A_\varepsilon^{(n)}$ of ε-typical sequences is good. Indeed, it holds $\Pr\{A_n^{(\varepsilon)}\} > 1 - \varepsilon$ and $|A_n^{(\varepsilon)}| \leq 2^{n(H+\varepsilon)}$ so that there exists a binary block code with $|C(X_1, \dots, X_n)| = [H + \varepsilon] + 1$ such that all $x \in A_\varepsilon^{(n)}$ are uniquely decodable and $Q(B_n) \leq \varepsilon$.

Under the last assumptions concerning the source there exists (cf. Sec. 5.4 in [3]) an instantaneous code C_n with

$$|C(X_1, \dots, X_n)| \approx \sum_{i=1}^{n} - \log p(X_i) \quad \text{and} \quad C^{-1}\mathcal{D}^* = \mathcal{X}^n$$

such that $P^n(B_n)$ at most equals $\Pr(|C(X_1, \dots, X_n)| > m_n)$, which is for $m_n \geq n(H + \varepsilon)$ bounded above by $\Pr(|C(X_1, \dots, X_n)| > n(H + \varepsilon))$. But, by the law of large numbers, the last probability tends to zero for every $\varepsilon > 0$. Therefore the code C_n is good.

Theorem 1. If the source $(\mathcal{X}^n, Q)$ satisfies the condition

$$\lim_{n \to \infty} \frac{1}{n} H(X_1, \dots, X_n) > 0$$

then an arbitrary good code $C_{n,\varepsilon}$ is asymptotically uniform in the sense that it holds

$$\lim_{\varepsilon \downarrow 0} \lim_{m \to \infty} \frac{H(Y_1, \dots, Y_m)}{m} = 1.$$

Proof. If $m_n \leq m < m_{n+1}$, then it follows from the monotonicity and nonnegativity of the entropy

$$\frac{H(Y_1, \dots, Y_m)}{m} \geq \frac{H(Y_1, \dots, Y_{m_n})}{m_{n+1}}.$$

If the assumptions hold, then there exists a sequence m_n considered in the definition of a good code such that

$$\lim_{n \to \infty} \frac{m_n}{m_{n+1}} = 1.$$

The desired assertion thus follows from (4) and from the definition of a good code. □

Theorem 1 solves the problem of uniformity of good codes for source codes with a zero distortion. Now we prove an analogue of Theorem 1 for codes with a positive distortion, in the frame of the general distortion model of Gray and Davisson [6].

Consider an arbitrary source

$$(\times_{i=1}^{\infty}\mathcal{X}_i,\ \times_{i=1}^{\infty}\mathcal{A}_i,\ P),$$

where $(\mathcal{X}_i, \mathcal{A}_i) = (\mathcal{X}, \mathcal{A})$ is a measurable space and P is a stationary ergodic probability measure (cf. Sec. II in [6]), not necessarily of the product-type. Elements $x = (x_1, \ldots, x_n) \in \mathcal{X}^n$ are called simply messages and the restriction of P on the space $(\mathcal{X}^n, \mathcal{A}^n)$ is denoted again by P. A measurable distortion $\rho(x, y) \geq 0$ with $\rho(x, x) = 0$ is considered on $\mathcal{X} \times \mathcal{X}$, satisfying the condition

$$\inf_{y\in\mathcal{X}} \int_{\mathcal{X}} \rho(x, y)\, \mathrm{d}P(x) < \infty.$$

For every n and all messages $x \doteq (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ from $\mathcal{X}^n$ we define

$$\rho_n(x, y) = \frac{1}{n}\sum_{i=1}^{n} \rho(x_i, y_i). \tag{5}$$

By a block code C we mean a measurable mapping $C : \mathcal{X}^n \mapsto \mathcal{X}^n$. The subset $\mathcal{C} = C(\mathcal{X}^n) \subset \mathcal{X}^n$ is called a codebook corresponding to the code C. Let

$$R(C) = \frac{1}{n}\log|\mathcal{C}| \in [0, \infty]$$

be the information rate of a code C and $\boldsymbol{C}_n(R)$ $R \geq 0$, the set of all codes with information rates at most R. It is known (cf. Sec. II in [6], or Chap. 7 in Berger [1]) that there exists a nonincreasing function $\mathrm{d} : [0, \infty] \mapsto I\!R$ such that the functions

$$\mathrm{d}_n(R) = \inf_{C\in\boldsymbol{C}_n(R)} \int_{\mathcal{X}^n} \rho_n(x, C(x))\, \mathrm{d}P(x)$$

satisfy the limit relation

$$\mathrm{d}_n(R) \downarrow \mathrm{d}(R) \quad \text{as } n \to \infty. \tag{6}$$

The function $d(R)$ is a distortion-rate function for the source and distortion measure ρ under consideration.

It is convenient to extend the concept of code by means of randomization. By a random code $\tilde{C}$ we mean a channel (Markov kernel) $(P_x : x \in \mathcal{X}^n)$ with input space $(\mathcal{X}^n, \mathcal{A}^n)$ and output space $(\mathcal{X}^n, \mathcal{A}^n)$. By $P*\tilde{C}$ we denote the joint probability measure induced by P and $\tilde{C}$ on $(\mathcal{X}^n \times \mathcal{X}^n,\ \mathcal{A}^n \times \mathcal{A}^n)$. P is one marginal of $P*\tilde{C}$ and the other marginal is denoted by $P\tilde{C}$. We shall present an alternative definition of $d(R)$ by means of random codes. Let $\tilde{\boldsymbol{C}}_n(R)$, $R \geq 0$, be the class of all random codes $\tilde{C}$ such that the I-divergence $I_n(P * \tilde{C},\ P \times P\tilde{C})$ of $P * \tilde{C}$ and $P \times P\tilde{C}$ (Shannon information) satisfies the condition

$$I_n(P * \tilde{C},\ P \times P\tilde{C}) \leq n\, R.$$

It is known (cf. Sec. II in [6] or [1], [2]) that there exists a nonincreasing function $D : [0,\infty] \mapsto \mathbb{R}$ such that the functions

$$D_n(R) = \inf_{C \in \tilde{C}_n(R)} \int_{\mathcal{X}^n \times \mathcal{X}^n} \rho_n(x,y)\,\mathrm{d}(P * \tilde{C})\,(x,y)$$

satisfy the limit relation

$$D_n(R) \downarrow D(R) \quad \text{as } n \to \infty. \tag{7}$$

It follows from a theorem of Berger, cf. Theorem 1 in [6], that

$$D(R) = d(R), \quad R \in [0,\infty]. \tag{8}$$

Using all these facts we can prove the following statement, where

$$H(C) = \sum_{y \in \mathcal{C}} -P\left(C^{-1}(y)\right) \log P\left(C^{-1}(y)\right)$$

is the entropy of a code $C : \mathcal{X}^n \mapsto \mathcal{X}^n$ with a codebook $\mathcal{C} = C(\mathcal{X}^n)$. Note that

$$\log |C(\mathcal{X}^n)| = \log |\mathcal{C}| = n\,R(C)$$

is the well-known upper bound on the entropy $H(C)$. This bound is achieved by a code with finite $R(C)$ only if this code is uniform in the sense that

$$P(C^{-1}(y)) = \frac{1}{|\mathcal{C}|} \quad \text{for all } y \in \mathcal{C}.$$

Theorem 2. Every sequence of codes $C_n : \mathcal{X}^n \mapsto \mathcal{X}^n$ approaching for some $R \in (0,\infty)$ the distortion-rate bound $\mathrm{d}(R)$ in the sense

$$\lim_{n\to\infty} (R(C_n),\, \mathrm{d}_n(C_n)) = (R,\, \mathrm{d}(R)) \tag{9}$$

is asymptotically uniform in the sense

$$\lim_{n\to\infty} \frac{H(C_n)}{\log |C_n(\mathcal{X}^n)|} = 1. \tag{10}$$

Proof. Since the ratio in (10) is at most 1 and $\log |C_n(\mathcal{X}^n)| = n\,R(C_n)$ where, by (9), $R(C_n) \to R$, it suffices to prove that

$$R_0 \triangleq \liminf_{n\to\infty} n^{-1} H(C_n)$$

equals R. It obviously holds $0 \le R_0 \le R$ and for every $\varepsilon > 0$ there exists a subsequence n_k and a natural k_0 such that

$$n^{-1} H(C_{n_k}) < R_0 + \varepsilon, \quad k \ge k_0.$$

Fix arbitrary $k > k_0$. By (2.29) and (2.39) in Cover, Thomas [3], it follows from here

$$\begin{aligned} n^{-1}I(P*C_{n_k}, P \times PC_{n_k}) &= n^{-1}[H(C_{n_k}(X)) - H(C_{n_k}(X)\,|\,X)] \\ &= n^{-1}H(C_{n_k}(X)) = n^{-1}H(C_{n_k}) < R_0 + \varepsilon, \end{aligned}$$

i.e., $C_{n_k} \in \mathbf{C}_{n_k}(R_0+\varepsilon)$. On the other hand, it is easy to see that

$$\int_{\mathcal{X}^{n_k} \times \mathcal{X}^{n_k}} \rho_{n_k}(x,y)\,\mathrm{d}(P*C_{n_k})(x,y) = \mathrm{d}_{n_k}(C_{n_k}).$$

Therefore it follows from (6) – (9) and from the definition of $D_n(R_0+\varepsilon)$

$$D(R_0+\varepsilon) = \lim_{k\to\infty} D_{n_k}(R_0+\varepsilon) \le \lim_{k\to\infty} \mathrm{d}_{n_k}(C_{n_k}) = \mathrm{d}(R) = D(R).$$

Since the function D is nonincreasing in the whole domain $[0,\infty]$, it follows from here $R_0 + \varepsilon \ge R$. Further, $\varepsilon > 0$ was arbitrary so that $R_0 \ge R$. □

Remark. In the proof of Theorem 2 we have not explicitely used the assumption that the source is stationary ergodic, nor the assumption that the distortion function is of the relatively frequency type (5). What we explicitely used were the convergences (6), (7), the equality (8), and the monotonicity of $\mathrm{d}(R)$ which is however an obvious consequence of (6). Therefore Theorem 2 applies to all distortion functions, e.g. to the Itakura–Saito distortion function mentioned in Sec. 13.2 of [3], and to all sources such that

$$\lim_{n\to\infty} \mathrm{d}_n(R) = \inf_n \mathrm{d}_n(R) = \inf_n D_n(R) = \lim_{n\to\infty} D_n(R).$$

3. ENTROPIC TESTS OF OPTIMALITY

Consider a code $C : \mathcal{X}^n \mapsto \mathcal{X}^n$, the entropy $H(C)$ defined in previous section, and the numbers

$$M = |C(\mathcal{X}^n)|, \quad 0 < h < \log M.$$

In this section we introduce a Neyman–Pearson α-level test of the hypothesis

$$\mathcal{H} = \{h \le H(C) \le \log M\}. \tag{11}$$

It follows from Theorem 1 or 2 that, for suitable h and under the circumstances described there, this becomes in fact an α-level test of the hypothesis that the code C is good or optimal respectively.

For example, if $M = 2^{10}$ and the hypothesis $\mathcal{H} = \{9 \le H(C) \le 10\}$ is rejected at the level $\alpha = 0.01$ then the test provides enough ground to believe that C is not a good choice, and Theorem 1 or 2 implies that $C = C_{n,\varepsilon}$ or $C = C_n$ can be replaced by a better code $C' = C'_{n',\varepsilon'}$ or $C' = C'_{n'}$, for which $\mathcal{H}' = \{9 \le H(C') \le 10\}$ cannot be rejected at the level $\alpha = 0.01$. Because the rates of convergence in Theorems 1 and 2 are unspecified, we cannot give justified recommendations for $h = 9$ in the test. Therefore we consider the level of h as a test parameter which characterizes how strict is the optimality demanded.

The test is based on statistical observations of relative frequencies $\hat{p}_N(y)$ of codewords y of the code C in a series of N independent realizations of messages from the encoded source. The test statistics is the entropy $H(\hat{P}_N)$ of the vector $\hat{P}_N = (\hat{p}_N(y_1), \ldots, \hat{p}_N(y_M))$ and the test itself is a pair

$$\left[H(\hat{P}_N),\, K_N\right] \equiv \left[H(\hat{P}_N),\, c_N(\alpha, h)\right] \quad \text{(cf. (3))}.$$

Let us consider the function

$$f(x) = (1 - 2^{-x}) \left[\log^2 \frac{1 - 2^{-x}}{M - 1} - x^2\right]$$

and the unique solution x_M of the equation $f'(x) = 0$ in the domain $0 \le x \le \log M$. It is easy to see that for all $M \ge 3$

$$x_M < \log \frac{M}{2} \tag{12}$$

and

$$f\left(\log \frac{M}{2}\right) = \frac{M-2}{M} \log \frac{2(M-1)}{M-2} \log \frac{M^2(M-1)}{2(M-2)}. \tag{13}$$

Theorem 3. If $0 < \alpha < 1$ and $0 < h < \log M$ then the test $[H(\hat{P}_N),\, c_N(\alpha, h)]$ with

$$c_N(\alpha, h) = h - \frac{\sigma(h)\, \Phi^{-1}(1 - \alpha/2)}{\sqrt{N}}$$

and

$$\sigma(h) = \left\langle \begin{array}{ll} \sqrt{f(h)} & \text{if } h \ge x_M \\ \sqrt{f(x_M)} & \text{if } h < x_M \end{array} \right.$$

is asymptotically α-level test for the hypothesis (11). If $M \ge 3$ and $h = \log(M/2)$ then

$$\sigma\left(\log \frac{M}{2}\right) = \left[\frac{M-2}{M} \log \frac{2(M-1)}{M-2} \log \frac{M^2(M-1)}{2(M-2)}\right]^{1/2}.$$

Proof. The first assertion follows from Assertions 1 and 2 in [5], where asymptotically α-level tests of more general hypotheses

$$\mathcal{H} = \{h_1 \le H(C) \le h_2\}, \quad 0 \le h_1 \le h_2 \le \log M$$

are found. The second assertion follows from the given general formula for $\sigma(h)$ and from relations (12) and (13) valid for $M \ge 3$. □

We see that in the above considered example with $M = 2^{10}$ and $h = 9$

$$\sigma(9) = \left[\frac{1022}{1024} \times \log\frac{2045}{1022} \times \log\frac{1048576 \times 1023}{2024}\right]^{1/2} = 4.358.$$

Therefore the approximately α-level critical value of the test statistic is for large N, say $N \geq 100$,

$$c_N(\alpha, 9) = 9 - \frac{4.358 \times \Phi^{-1}(1-\alpha/2)}{\sqrt{N}}.$$

For example,

$$c_N(0.01, 9) = 9 - \frac{11.226}{\sqrt{N}}.$$

□

4. EXAMPLE

Three codes C^1, C^2, C^3 for speech compression at the level of information rate $R = 600$ bps) have been computed by means of the Lloyd algorithm (cf. p. 338 in Cover, Thomas [3]). The computations have been based on three different ensembles of independent realizations $X_1^{(j)}, \ldots, X_{N_j}^{(j)}$, $j = 1, 2, 3$ with $N_1 < N_2 < N_3$ of order several thousands. Then the codes have been tested on a fourth ensemble with $N_4 = 16\,000$. The approximately α-level critical value of the obtained statistic $H(\hat{P}_{16\,000})$ is

$$\begin{aligned} c_{16\,000}(\alpha, 9.8) &= 9.8 - \frac{3.960 \times \Phi^{-1}(1-\alpha/2)}{400} \\ &= 9.8 - 0.01 \times \Phi^{-1}(1-\alpha/2). \end{aligned}$$

Table 1. Test statistics $H(\hat{P}_N)$ and critical values $c_N(\alpha, 9.8)$.

		$c_{16\,000}(\alpha, 9.8)$				
Code	$H(\hat{P}_{16000})$	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$
C^1	9.408					
C^2	9.752	9.787	9.783	9.780	9.774	9.767
C^3	9.793					

The test statistics and the critical values are presented in Table 1. We see that the sample entropies $H(\hat{P}_{16\,000})$ of all three codes exceed the value $h = 9$ so that the hypothesis (11) with $h \geq 9$ cannot be rejected at any significance level $0 < \alpha < 1$.

Let us therefore test this hypothesis with h larger than the largest sample entropy, say with $h = 9.8$. This hypothesis is not rejected for the code C^3 even at the level $\alpha = 0.2$, i.e. there is a good chance that this code belongs to the optimality class with the entropy between 9.8 and 10. On the other hand, this hypothesis for the codes C^2, C^3 is rejected even at the level $\alpha = 0.001$, i.e. there is a 99.9 % chance that the entropies of these codes are below 9.8. The codes C^2, C^3 thus almost surely do not belong to the optimality class with the entropy between 9.8 and 10.

In spite of that it is not known what the relation $9.8 \leq H(C) \leq 10$ means in term of the difference between the code distortion $d(C)$ and the theoretical bound $d(600 \text{ bps})$, the results of our test provide a little more statistical evidence in favour of C^3 than do the values $H(\hat{P}_N) = 9.408$, 9.752, and 9.793 alone.

(Received October 11, 1994.)

REFERENCES

[1] T. Berger: Rate Distortion Theory. Prentice Hall, Englewood Cliffs 1971.
[2] R. E. Blahut: Principles and Practice of Information Theory. Addison–Wesley, Reading 1987.
[3] T. M. Cover and J. A. Thomas: Elements of Information Theory. Wiley, New York 1991.
[4] J. Feistauerová: An entropic test of uniformity and optimality of codebooks. Problems Control Inform. Theory *17* (1988), 319–325.
[5] J. Feistauerová and I. Vajda: Testing system entropy and prediction error probability. IEEE Trans. Systems Man Cybernet. *23* (1993), 5, 1352–1358.
[6] R. M. Gray and J. L. Davisson: Source coding theorems without the ergodic assumption. IEEE Trans. Inform. Theory *20* (1974), 502–516.
[7] E. L. Lehman: Testing Statistical Hypotheses. Wiley, New York 1962.

Ing. Igor Vajda, DrSc., Ústav teorie informace a automatizace AV ČR (Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic), Pod vodárenskou věží 4, 182 08 Praha 8. Czech Republic.