Random Coding Techniques for Nonrandom Codes

Nadav Shulman, *Student Member, IEEE*, and Meir Feder, *Fellow, IEEE*

Abstract—This work provides techniques to apply the channel coding theorem and the resulting error exponent, which was originally derived for totally random block-code ensembles, to ensembles of codes with less restrictive randomness demands. As an example, the random coding technique can even be applied for an ensemble that contains a *single* code. For a specific linear code, we get an upper bound for the error probability, which equals Gallager's random-coding bound, up to a factor determined by the maximum ratio between the weight distribution of the code, and the expected *random* weight distribution.

Index Terms—Bounds on the error probability, code ensembles, code's spectrum, error exponent, linear codes, random coding.

I. INTRODUCTION

Shannon's channel coding theorem [13] has been proven in many ways. The classical proofs are of Feinstein [3], [4], Elias [2], Wolfowitz [18], and Gallager [7]. Some of these proofs also provide an exponential upper bound on the error probability as a function of the code complexity, or more correctly, the code length. Most of these proofs, however, consider a *random* choice of codebooks. Hence, these proofs (as well as the nonrandom proof by Feinstein) are not constructive, and cannot be used to point out a specific good code, or even a good small family of codes. Still, there are a few interesting structured families of codes, e.g., linear codes, that are large and diverse enough so that the *random coding* argument can be applied to them, see, e.g., [12]. In some other interesting cases, the family of codes is extended (sometimes artificially) so that random coding arguments can be applied to it. For example, the class of linear convolutional codes was enlarged to the class of *time-varying* convolutional codes on which the random coding proof can be easily applied [16], [17].

When the goal is to find and analyze specific codes, coding theory usually does not use the channel coding theorem proof, but applies instead combinatorial and algebraic techniques. One example is Poltyrev's bound [11] for the error probability of a linear code in a binary-symmetric channel (BSC) that depends on the weight distribution (spectrum) of the code. A similar combinatorial analysis that deals with a broader class of channels is given in [5], but this analysis is less powerful as it provides interesting results only at low rates (below the cutoff rate). It turns out, as shown later in this work, that random coding techniques, which are used in the derivation of the coding theorem and can be applied for any rate up to capacity, can lead to an exponential upper bound on the error probability of a specific linear code in terms of its spectrum. Other information-theoretic methods that have been previously used to bound specific codes, see, e.g., [10], are applicable only at rates below the cutoff rate.

This correspondence consists of two parts. In the first part, we address the general problem of applying the channel coding theorem for structured code families. Following some simple observations we show several techniques to apply the random coding proof of the

channel coding theorem of [7] and [8], in several restricted families of codes. In some cases, the error exponent of the restricted family of codes equals the random-coding error exponent. Detailed examples for the application of these techniques in some interesting structured ensembles of codes is given in [14].

In the second part of the correspondence we consider an extreme example of a restrictive class of codes; an ensemble that contains a single code. Using the tools developed in the first part, we can sometimes get an exponential bound on the error probability of this code. For a linear binary code, this exponential bound is given by Gallager's random-coding error bound, times a term that depends on the deviation of the code's spectrum from the random-like binomial spectrum. Specifically, if the weight distribution is $\{A_l\}$ and the expected, random, weight distribution of a code with the same parameters is $\{B_l\}$, than this term is $\max_l (A_l/B_l)$.

II. RANDOM CHANNEL CODING—A BRIEF REVIEW

In this preliminary section, we briefly review Gallager's random coding proof of the channel coding Theorem. Along this review we set the notations and prepare the setup for our results.

Consider a random ensemble of codes $\mathcal{E}$ where each code in the ensemble has M words of length N, i.e., its rate is $R = \log M/N$. We denote by $\Pr(\boldsymbol{c}(i) = \boldsymbol{x})$ the probability that the randomly selected ith codeword equals $\boldsymbol{x}$, and, similarly, $\Pr(\boldsymbol{c}(i') = \boldsymbol{x}'|\boldsymbol{c}(i) = \boldsymbol{x})$ is the conditional probability, induced by the random selection strategy, that the i'th codeword is $\boldsymbol{x}'$ given that the ith codeword is $\boldsymbol{x}$.

Let the channel be defined by transition probabilities $P_N(\boldsymbol{y}|\boldsymbol{x})$, where $\boldsymbol{x}$ and $\boldsymbol{y}$ are vectors of length N corresponding to N-blocks of input and output to the channel. It is shown in [8, pp. 136–137] that the maximum-likelihood decoder, when the transmitted codeword index is chosen uniformly, achieves an average error probability over the ensemble, $\bar{P}_e$, bounded by

$$\bar{P}_e \leq \sum_{m=1}^{M} \frac{1}{M} \sum_{x} \Pr(\boldsymbol{c}(m) = \boldsymbol{x}) \sum_{\boldsymbol{y}} P_N(\boldsymbol{y}|\boldsymbol{x})^{1/1+\rho}$$
$$\cdot \left[\sum_{m' \neq m} \sum_{\boldsymbol{x}'} \Pr(\boldsymbol{c}(m') = \boldsymbol{x}'|\boldsymbol{c}(m) = \boldsymbol{x}) P_N(\boldsymbol{y}|\boldsymbol{x}')^{1/1+\rho} \right]^{\rho} \tag{1}$$

for any $0 \leq \rho \leq 1$.

Suppose now that each word is selected *independently* with the distribution $Q_N(\boldsymbol{x})$, i.e.,

$$\Pr(\boldsymbol{c}(m) = \boldsymbol{x}) = Q_N(\boldsymbol{x}) \tag{2}$$
$$\Pr(\boldsymbol{c}(m') = \boldsymbol{x}'|\boldsymbol{c}(m) = \boldsymbol{x}) = Q_N(\boldsymbol{x}'), \qquad \text{for } m \neq m' \tag{3}$$

Furthermore, let the channel be discrete and memoryless (DMC), i.e.,

$$P_N(\boldsymbol{y}|\boldsymbol{x}) = \prod_{n=1}^{N} P(y_n|x_n) \tag{4}$$

and choose

$$Q_N(\boldsymbol{x}) = \prod_{n=1}^{N} Q(x_n). \tag{5}$$

Then, substituting in (1) yields

$$\bar{P}_e \leq 2^{-N[E_0(\rho, \boldsymbol{Q}) - \rho R]} \tag{6}$$

where

$$E_0(\rho, Q) \triangleq -\log \sum_j \left[\sum_k Q(k) P(j|k)^{1/(1+\rho)} \right]^{1+\rho}. \quad (7)$$

The *random coding exponent* for DMC's is defined as

$$E_r(R) \triangleq \max_Q \max_{0 \le \rho \le 1} \{E_0(\rho, Q) - \rho R\} \quad (8)$$

and, as shown in]8, pp. 140–144], it is positive for $R < C$ where C is the channel capacity.

III. REDUCED RANDOMNESS ASSUMPTIONS

The assumption in the proof above that the ensemble of codes is large and that codes should be drawn in total randomness can be relaxed. This enables us to deal with ensembles of codes that have very specific structure, and are far from being totally random. One observation that was already noticed and utilized by Gallager [8, p. 207], Dobrushin [1], and Gabidulin [6] is that it is sufficient to consider the pairwise distribution of any two codewords, and require only *pairwise independence*, to get the random-coding upper bound above.

It turns out that even when the codewords are not pairwise-independent or when the marginal distribution of the codewords is not as in (2), we can still derive expressions for the error exponent, provided that there are upper bounds on the marginal and conditional probabilities. Specifically, suppose that for some $\alpha, \beta \ge 1$ and Q_N, we have for any $i \ne j, \boldsymbol{x}$, and $\boldsymbol{x}'$

$$\Pr(\boldsymbol{c}(i) = \boldsymbol{x}', \boldsymbol{c}(j) = \boldsymbol{x}) \le \alpha \cdot \Pr(\boldsymbol{c}(i) = \boldsymbol{x}') \cdot \Pr(\boldsymbol{c}(j) = \boldsymbol{x}) \quad (9)$$

$$\Pr(\boldsymbol{c}(i) = \boldsymbol{x}) \le \beta \cdot Q_N(\boldsymbol{x}). \quad (10)$$

Then by substituting in (1) the error probability will be bounded by

$$\bar{P}_e \le \alpha^\rho \beta^{1+\rho} \cdot 2^{\rho N R} \sum_{\boldsymbol{y}} \left[\sum_{\boldsymbol{x}} Q_N(\boldsymbol{x}) P_N(\boldsymbol{y}|\boldsymbol{x})^{1/(1+\rho)} \right]^{1+\rho}. \quad (11)$$

In addition, if the channel is DMC or "almost" DMC

$$P_N(\boldsymbol{y}|\boldsymbol{x}) \le \gamma \cdot \prod_{n=1}^N P(y_n|x_n) \quad (12)$$

for some $\gamma \ge 1$, then, using $Q_N(\boldsymbol{x})$ from (5), the expected error probability will be bounded as

$$\bar{P}_e \le \alpha^\rho \beta^{1+\rho} \gamma \cdot 2^{-N[E_0(\rho, \boldsymbol{Q}) - \rho R]}. \quad (13)$$

Notice that since $0 \le \rho \le 1$, $\alpha \ge 1$, and $\beta \ge 1$, we get the upper bound

$$\bar{P}_e \le \alpha \beta^2 \gamma \cdot 2^{-N E_r(R)}. \quad (14)$$

In general, α, β, and γ can depend on N and on other parameters of the code. Now, from (13)

$$\bar{P}_e \le \beta \gamma \cdot 2^{-N(E_0(\rho, Q) - \rho(R + (\log(\alpha \beta)/N)))}. \quad (15)$$

Optimizing (15) with respect to $\rho, 0 \le \rho \le 1$, yields a lower bound on the reliability function

$$\lim_{N \to \infty} -\frac{1}{N} \log \bar{P}_e \ge E_r(R + R_\alpha + R_\beta) - R_\beta - R_\gamma \quad (16)$$

where

$$R_\alpha = \lim_{N \to \infty} \frac{\log \alpha}{N}$$

$$R_\beta = \lim_{N \to \infty} \frac{\log \beta}{N}$$

and

$$R_\gamma = \lim_{N \to \infty} \frac{\log \gamma}{N}.$$

Thus the random-coding exponent is attained with the relaxed assumptions on the ensemble of codes if α, β, and γ do not increase exponentially with N. In any case we still get an exponential error bound, but it may be exponentially inferior than the random coding bound.

Another general technique to derive an upper bound on the expected error probability for a restricted ensemble of codes is to build from the given ensemble a bigger, more random new ensemble of codes whose average performance is the same as that of the original ensemble. Suppose that on the new ensemble the random coding proof, maybe with the extensions above, can be derived. This implies the same result for the original ensemble. The new ensemble is "artificial" in the sense that it provides a tool to prove some properties of the original ensemble, but as an ensemble on its own, it has no interesting structure.

The simplest way to enlarge the ensemble is by *word permutation*. When we are interested in the *word* error probability, and not in the bit-error rate (BER) or other fidelity criterion, the specific assignment of the information messages to codewords is not important. Thus the ensemble performance is invariant to permutations that change the order of the words. More precisely, assume that the codeword index is chosen uniformly. Let $\mathcal{E}$ be an ensemble of codes and let $\tilde{\mathcal{E}}$ be a larger ensemble that contains all the codes from $\mathcal{E}$ and their word permutation. Clearly, $\mathcal{E}$ and $\tilde{\mathcal{E}}$ have the same average error probability.

For memoryless channels, the ensemble can also be enlarged by *symbol permutation*. The error probability of two codes in which each codeword in one code is some *fixed* permutation of that codeword in the other code, is the same. Thus the average error probability of the codes in an ensemble $\mathcal{E}$ is the same as the average error probability of a bigger ensemble $\tilde{\mathcal{E}}$ that contains all the codes obtained by symbol permutations of the codes of $\mathcal{E}$.

For symmetric binary-input channels such as the BSC the error probability of a code does not change by adding[1] a constant binary vector to all its codewords. The reason is that this addition does not change the distance between the codewords.

An example that demonstrates how the observations above can be applied to upper-bound the average error probability of a restricted ensemble of codes is given by the following lemma.

Lemma 1: Let $\mathcal{E}$ be an ensemble of binary codes with the property that for any $i \ne j$ and $\boldsymbol{x}$ we have

$$\Pr(\boldsymbol{c}(i) \oplus \boldsymbol{c}(j) = \boldsymbol{x}) \le \alpha 2^{-N}. \quad (17)$$

Then, the average error probability over $\mathcal{E}$ for a symmetric binary-input channel is bounded by

$$\bar{P}_e \le \alpha^\rho 2^{-N[E_0(\rho, \boldsymbol{Q}) - \rho R]} \quad (18)$$

where Q is the uniform distribution.

Proof: We define a new ensemble, $\tilde{\mathcal{E}}$, which contains all the codes from $\mathcal{E}$ with an added random vector, i.e.,

$$\tilde{\mathcal{E}} = \{\mathcal{C} \oplus \boldsymbol{v} | \mathcal{C} \in \mathcal{E}, \boldsymbol{v} \in \{0, 1\}^N\} \quad (19)$$

where $\mathcal{C} \oplus \boldsymbol{v} = \{\boldsymbol{c}(1) \oplus \boldsymbol{v}, \cdots, \boldsymbol{c}(M) \oplus \boldsymbol{v}\}$ with the probability assignment $\Pr(\mathcal{C} \oplus \boldsymbol{v}) = 2^{-N} \Pr(\mathcal{C})$. As noted above in $\tilde{\mathcal{E}}$ the average error probability is the same as in $\mathcal{E}$. To calculate the expected

---

[1] Addition of two binary vectors is their bitwise exclusive–or.

error probability in $\tilde{\mathcal{E}}$ we observe that the marginal and conditional distributions in the new ensemble satisfy

$$\Pr\left(\tilde{\boldsymbol{c}}(i)=\boldsymbol{x}\right)=\sum_{\boldsymbol{v}} 2^{-N} \Pr\left(\boldsymbol{c}(i)=\boldsymbol{x}\oplus\boldsymbol{v}\right)=2^{-N} \qquad (20)$$

$$
\begin{aligned}
\Pr\left(\tilde{\boldsymbol{c}}(i)=\boldsymbol{x}|\tilde{\boldsymbol{c}}(j)=\boldsymbol{y}\right) &=\sum_{\boldsymbol{v}} 2^{-N} \Pr\left(\boldsymbol{c}(i)=\boldsymbol{x}\oplus\boldsymbol{v}|\boldsymbol{c}(j)=\boldsymbol{y}\oplus\boldsymbol{v}\right) \\
&=\sum_{\boldsymbol{c}(j)} 2^{-N} \Pr\left(\boldsymbol{c}(i)\boldsymbol{x}\oplus\boldsymbol{c}(j)\oplus\boldsymbol{y}|\boldsymbol{v}=\boldsymbol{c}(j)\oplus\boldsymbol{y}\right) \\
&=\sum_{\boldsymbol{c}(j)} 2^{-N} \Pr\left(\boldsymbol{c}(i)=\boldsymbol{x}\oplus\boldsymbol{c}(j)\oplus\boldsymbol{y}\right) \\
&\leq \sum_{\boldsymbol{c}(j)} 2^{-N}\alpha 2^{-N}=\alpha 2^{-N}. \qquad (21)
\end{aligned}
$$

Thus we have the same bounds as in (9) and (10) on the pairwise distribution, with $\beta=1$ and $Q_N(\cdot)=2^{-N}$, implying that the average error probability for $\tilde{\mathcal{E}}$ and $\mathcal{E}$ can be bounded by (18), which is the same as (13) or (14) with these values of β and $Q_N(\cdot)$. $\qquad\square$

In the examples above, a new, more random, ensemble of codes was formed by generating from each code in the original ensemble a set of codes with the same performance. To ensure that the average error probabilities of the new and original ensembles are the same, we implicitly assumed that in the new ensemble the probability to choose a code from a set generated by some original code equals the probability of choosing that original code in the original ensemble. We note, however, that even if this assumption is not true, and the probability assignment on the new ensemble is different implying that the average error probability of both ensembles are different, the average error probability on the new ensemble is still interesting. The construction used in the discussion above assures that there exists a code in the original ensemble whose performance is at least as good as the average error probability of the new ensemble.

In summary, in this section we have presented techniques that allow the derivation of the channel coding theorem, and the determination of error exponents for code classes that are not necessarily large and completely random. In the next section we show how these techniques can be used to upper-bound the error probability of a specific code. A more detailed discussion of these results and additional examples of their usage in several interesting restricted classes of codes are given in [14].

IV. THE ERROR EXPONENT OF A SPECIFIC LINEAR CODE

We now use the methods described above to obtain a bound on the error probability of a binary linear code, used in a BSC, that depends on its weight distribution (spectrum).

We use the technique above and generate from the given code a random ensemble of linear codes by permuting randomly the order of the codewords, and then permute randomly the order of the symbols in the codewords. We then show that the resulting random ensemble satisfies the condition of Lemma 1 which leads to an error exponent expression for the original code.

Theorem 1: Let $\mathcal{C}$ be a *particular* (N,K) binary linear code, i.e., it contains $M=2^K$ codewords and its rate is K/N. Let A_l, $l=0,1,\cdots,N$ be its weight distribution, i.e.,

$$A_l=|\{i:W_H(\boldsymbol{c}(i))=l\}|$$

where $W_H(\boldsymbol{v})$ is the Hamming weight of $\boldsymbol{v}$. Then $P_e(\mathcal{C})$, its error probability, is upper-bounded as

$$P_e(\mathcal{C})\leq 2^{-NE_r(R+(\log\alpha/N)} \qquad (22)$$

where

$$\alpha=\max_{0<l\leq N}\frac{A_l}{2^K-1}\frac{2^N}{\binom{N}{l}}. \qquad (23)$$

Proof: We first generate an ensemble $\mathcal{E}'$ from the given code by choosing with uniform distribution a permutation $\pi\in S_M$ of the order of the codewords. For $i\neq j$ we have in $\mathcal{E}'$

$$
\begin{aligned}
\Pr\left(\boldsymbol{c}'(i)\oplus\boldsymbol{c}'(j)=\boldsymbol{x}\right) &=\sum_{\pi_i\neq\pi_j}\frac{1}{M(M-1)}\Pr\left(\boldsymbol{c}(\pi_i)\oplus\boldsymbol{c}(\pi_j)=\boldsymbol{x}\right) \\
&=\begin{cases}\frac{1}{M-1}, & \text{if } \boldsymbol{x}\in\mathcal{C} \\ 0, & \text{if } \boldsymbol{x}\notin\mathcal{C}\end{cases} \qquad (24)
\end{aligned}
$$

where π_i denotes the index that i is mapped to after permutation π.

We now generate an even larger ensemble $\mathcal{E}''$ by choosing randomly with a uniform distribution a permutation $\sigma\in S_N$ of the order of the symbols in the codewords. In $\mathcal{E}''$ we thus have

$$
\begin{aligned}
\Pr\left(\boldsymbol{c}''(i)\oplus\boldsymbol{c}''(j)=\boldsymbol{x}\right) &=\sum_{\sigma\in S_N}\frac{1}{N!}\Pr\left(\boldsymbol{c}'(i)\oplus\boldsymbol{c}'(j)=\sigma(\boldsymbol{x})\right) \\
&=A_l\frac{1}{M-1}\frac{N!l!}{N!} \qquad (25)
\end{aligned}
$$

where $l=W_H(\boldsymbol{x})$ and $\sigma(\boldsymbol{x})=(x_{\sigma_1},\cdots,x_{\sigma_N})$ is the symbol permutation of $\boldsymbol{x}$ defined by $\sigma(\cdot)$. This implies that the ensemble $\mathcal{E}''$ satisfies (17) with an α given in (23). Thus Lemma 1 can be applied for $\mathcal{E}''$ leading to the bound (22) on the average error probability in $\mathcal{E}''$, which is also a bound on $P_e(\mathcal{C})$. $\qquad\square$

Looking at the resulting bound, we note that $\frac{\binom{N}{l}}{2^N}$ is the probability to select a vector with Hamming weight l, under the uniform distribution assumption, and $\frac{A_l}{2^K-1}$ is the same probability under the assumption that we choose randomly one of the (nonzero) words in the code. Hence, if the code's spectrum is close to the spectrum expected in a random code, i.e., if $A_l\approx 2^{K-N}\binom{N}{l}$, then we get the well-known result [2] that the error probability of that code the random-coding exponential bound.

Gallager, in [9, pp. 30–36], presented a bound that can be used for a specific code in terms of its spectrum. This bound is exponentially the same as our bound given in Theorem 1 above, but our result is tighter by an $O(\sqrt{N})$ factor. Interestingly, our bound derived from the bound in [7] that was developed for a random ensemble is tighter that the bound in [9] that was developed directly for specific ensembles and specific codes.

Theorem 1 can also be extended to consider general, nonlinear codes. In this case, instead of the spectrum we should use the average number of codewords in a given distance, i.e.,

$$A_l=\frac{1}{M}|\{(i,j):d(x_i,x_j)=l\}| \qquad (26)$$

where $|S|$ denote the size of a set S and $d(\cdot,\cdot)$ is the Hamming distance.

As another extension, suppose we have the following channel over a larger alphabet of size q:

$$P(y|x)=\begin{cases}1-\epsilon & x=y \\ \epsilon/(q-1) & x\neq y\end{cases} \qquad x,y\in\{1,\cdots,q\}. \qquad (27)$$

Codes over this channel were considered in [5]. The additional symmetry in the channel allows to introduce the following general operation that preserves the performance: the value assignment for each symbol in the codeword can be permuted, and this permutation can even be different for each symbol. A special case of this operation is to add (modulo q) a vector to each codeword. Thus we can follow the derivation above and get an exponential error bound for a q-ary

linear code with a given spectrum, operating over the channel (27). The resulting bound is given by (22) where now

$$\alpha = \max_{0 < l \leq N} \frac{A_l}{M - 1} \frac{q^N}{\binom{N}{l}(q-1)^l} \qquad (28)$$

and $M = 2^{NR}$ is the number of codewords. Unlike our result, the combinatorial technique presented in [5] can be applied only for rates below the cutoff rate.

For binary-input symmetric channels, the combinatorial technique of [5] was refined by Poltyrev [11] by using the union bound adequately (applying the union bound in a straightforward way is the reason why the bound in [5] is valid only at rates below the cutoff rate). It is interesting to note that for a code with a random-like spectrum, Poltyrev's bound is actually tighter (by a nonexponential term) than Gallager's random-coding bound. Thus for linear codes over a BSC, Poltyrev's bound is also tighter than our bound of Theorem 1. Still our bound, derived by a noncombinatorial approach from the basic random-coding expression, is more general as it is easily extended and leads to useful bounds in many cases, e.g., the q-ary case above.

Finally, the general techniques presented in this correspondence are useful in many interesting structured families of codes [14]. One recent direct application of these techniques is given in [15] where the error exponent of *time-invariant* convolutional codes has been derived.

ACKNOWLEDGMENT

The authors wish to thank the anonymous reviewers and the associate editor, A. Barg, for useful comments and for pointing out previous work on the subject.

REFERENCES

[1] R. L. Dobrushin, "Asymptotic optimality of group and systematic codes for some channels," *Theory Probab. its Applications*, pp. 47–60, 1963, translated from *Teor. Veroyat. i ee Primen.*

[2] P. Elias, "Coding for noisy channels," in *IRE Conv. Rec.*, pt. 4, 1955, pp. 37–46.

[3] A. Feinstein, "A new basic theorem of information theory," *IRE Trans. Inform. Theory*, vol. PGIT-4, pp. 2–22, 1954.

[4] ——, "Error bound in noisy channels without memory," *IRE Trans. Inform. Theory*, vol. IT-1, pp. 13–14, 1955.

[5] G. L. Katsman, M. A. Tsfasman, and S. G. Vlăduţ, "Spectra of linear codes and error probability decoding," in *Coding Theory and Algebraic Geometry* (Lecture Notes on Mathematics, vol. 1518), H. Strichenoth and M. A. Tsfasman, Eds. Berlin, Germany: Springer-Verlag, 1992, pp. 82–98.

[6] E. M. Gabidulin, "Limits for the decoding error probability when linear codes are used in memoryless channel," *Probl. Inform. Transm.*, pp. 43–48, 1967, translated from *Probl. Pered. Inform.*.

[7] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 3–18, 1965.

[8] ——, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[9] ——, *Low-Density Parity-Check Codes* Cambridge, MA: MIT Press, 1963.

[10] D. E. Lazic and V. Senk, "A direct geometrical method for bounding the error exponent for any specific family of channel codes—Part 1: Cutoff rate lower bound for block codes," *IEEE Trans. Inform. Theory*, vol. 38, pp. 1548–1559, Sept. 1992.

[11] G. Poltyrev, "Bounds on the decoding error probability of linear binary codes via their spectra," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1284–1292, July 1994.

[12] G. Sèguin, "Linear ensembles of codes," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 477–480, July 1979.

[13] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, July 1948 and pp. 623–656, Oct. 1948.

[14] N. Shulman, "Coding theorems for structured code families," M.Sc. thesis, Dep. Elec. Eng.–Syst., Tel-Aviv Univ., Tel-Aviv, Israel, Sept. 1995.

[15] N. Shulman and M. Feder, "Improved error exponent for time-invariant and periodically time-variant convolutional codes," *IEEE Trans. Inform. Theory*, to be published.

[16] A. J. Viterbi, "Convolutional codes and their performance in communication systems," *IEEE Trans. Commun. Technol.*, vol. COM-19, pp. 751–772, Oct. 1971.

[17] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*. New York: McGraw-Hill, 1979.

[18] J. Wolfowitz, "The coding of message subject to chance errors," *Illinois J. Math*, vol. 1, pp. 591–606, 1957.

The Asymptotic Redundancy of Bayes Rules for Markov Chains

Kevin Atteson, *Member, IEEE*

Abstract—We derive the asymptotics of the redundancy of Bayes rules for Markov chains of fixed order over a finite alphabet, extending the work of Barron and Clarke on independent and identically distributed (i.i.d.) sources. The asymptotics are derived when the actual source is the class of ϕ-mixing sources which strictly includes Markov chains. These results can be used to derive minimax asymptotic rates of convergence for universal codes when a Markov chain of fixed order is used as a model.

Index Terms—Asymptotics, Bayesian statistics, Markov chains, universal coding.

I. INTRODUCTION

Given data generated by a known stochastic process, methods of encoding the data to achieve the minimal average coding length are known [8]. Universal codes [18], [9] encode data such that, asymptotically, the average per-symbol code length is equal to its minimal value (the entropy rate) for any source within a wide class. For the well-known Lempel–Ziv code, the average per-symbol code length in excess of the entropy, i.e., the redundancy, goes to zero for the class of all ergodic stochastic processes [18]. In fact, there is no code for which the redundancy goes to zero uniformly at rate $\rho(n)$ for any sequence $\rho(n) = o(1)$ for the class of all ergodic stochastic processes [16]. Moreover, it has been observed that the rate of convergence to zero of the redundancy of universal codes such as Lempel–Ziv are slow for practical sources of data [3, p. 268]. In this correspondence we will derive minimax asymptotics, up to terms of order $o(\frac{1}{n})$, of the redundancy for stationary Markov chain sources.

For sources having a finite number k of parameters and satisfying certain conditions, it has been shown that there is a code for which the redundancy goes to zero at rate $\frac{k \log n}{2n} + o(\frac{\log n}{n})$ which is the optimum rate [15]. For finitely parameterized independent and identically distributed (i.i.d.) sources satisfying certain conditions, it