EXPONENTIAL ERROR BOUNDS
FOR FINITE STATE CHANNELS

DAVID BLACKWELL
UNIVERSITY OF CALIFORNIA, BERKELEY

1. Introduction and summary

A finite state channel is defined by (1) a finite nonempty set $A$, the set of inputs, (2) a finite nonempty set $B$, the set of outputs, (3) a finite nonempty set $T$, the set of (channel) states, (4) a transition law $p = p(t'|t, a)$, specifying the probability that, if the channel is in state $t$ and is given input $a$, the resulting state is $t'$, and (5) a function $\psi$ from $T$ to $B$, specifying the output $b = \psi(t)$ of the channel when it is in state $t$.

For any sequence $\{a_n, n = 1, 2, \cdots\}$ of random variables with values in $A$, we may consider the process $\{a_n\}$ as supplying the inputs for the channel, as follows: an initial channel state $t_0$ is selected with a uniform distribution over $T$. The input $a_1$ is then given the channel. The channel then selects a state $t_1$, with

$$(1) \quad P\{t_1 = t|t_0, a_1\} = p(t|t_0, a_1)$$

and produces output $b_1 = \psi(t_1)$. The channel is then given input $a_2$ and selects state $t_2$, with

$$(2) \quad P\{t_2 = t|t_0, t_1, a_1, a_2, b_1\} = p(t|t_1, a_2),$$

and so on. In general, for $n \geq 0$,

$$(3) \quad P\{a_{n+1} = a, t_{n+1} = t, b_{n+1} = b|a_i, 1 \leq i \leq n, t_i, 0 \leq i \leq n, b_i, 1 \leq i \leq n\} = P\{a_{n+1} = a|a_i, i \leq n\} p(t|t_n, a) x(t, b),$$

where $x(t, b) = 1$ if $\psi(t) = b$ and 0 otherwise.

For any random variable $x$ with a finite set of values and any random variable $y$, the (nonnegative) random variable whose value when $x = x_0$ and $y = y_0$ is

$$(4) \quad -\log P\{x = x_0|y = y_0\}$$

(all logs are base 2) is called the (conditional) entropy of $x$ given $y$ and will be denoted by $i(x|y)$. Its expected value, which cannot exceed the log of the number of values of $x$, will be denoted by $I(x|y)$. For $y$ a constant, $i(x|y)$ and $I(x|y)$ will be denoted by $i(x)$, $I(x)$ respectively. If each of $x, y$ has only finitely many values, the random variable

$$(5) \quad j(x, y) = i(x) + i(y) - i(x, y) = i(x) - i(x|y) = i(y) - i(y|x)$$

This paper was prepared with the partial support of the Office of Naval Research (Nonr-222-53).
is called the mutual information between \(x\) and \(y\). Its expected value will be denoted by \(J(x, y)\). For any stationary process \(\{x_n: -\infty < n < \infty\}\) whose variables have only finitely many values, we write \(I^*(x)\) for \(I(x_0|x_{-1}, x_{-2}, \ldots)\).

An inequality of Shannon [8] and Feinstein [3] relates the existence of codes to the distribution of \(j\{(a_1, \ldots, a_N), (b_1, \ldots, b_N)\} = j_N\), as follows.

**Shannon-Feinstein inequality.** For any integer \(D\), and any number \(\gamma\) there are two functions \(f, g\), where \(f\) maps \((1, \ldots, D)\) into the set \(U\) of sequences of length \(N\) of elements of \(A\), \(g\) maps the set \(V\) of sequences of length \(N\) of elements of \(B\) into \((1, \ldots, D)\), for which

\[
P\{g(b_1, \ldots, b_N) \neq d|(a_1, \ldots, a_N) = f(d)\} \leq P\{j_N < \gamma\} + \frac{D}{2\gamma}
\]

for \(d = 1, \ldots, D\).

Note that the left side of (6) is independent of the distribution of \(\{a_n\}\); it is simply the probability that, when an initial state for the channel is selected with a uniform distribution and the channel is then given the input sequence \(f(d)\), the resulting output sequence \(b_1, \ldots, b_N\) will be one for which \(g(b_1, \ldots, b_N) \neq d\). The pair \((f, g)\) can be considered as a code, with which we can transmit any of \(D\) messages over our channel in \(N\) transmission periods; when message \(d\) is presented to the sender, he gives the channel input sequence \(f(d)\); the receiver then observes some output sequence \(v\) and decides that message \(g(v)\) is intended.

For a given number \(c \geq 0\), let \(D = [2^{Nc}]\) and write

\[
\theta_N(c) = \min_{f_d} \max_{1 \leq d \leq D} P\{g(b_1, \ldots, b_N) \neq d|(a_1, \ldots, a_N) = f(d)\}.
\]

Thus \(\theta_N(c)\) is small if and only if we can transmit any binary sequence of length \(Nc\), by using the channel for \(N\) periods, with small error probability.

Shannon [7] associated with each channel a number \(C\), called the capacity of the channel, and proved that, for certain channels, \(\theta_N(c) \to 0\) as \(N \to \infty\) for every \(c < C\) but not for any \(c > C\). His original work has been considerably simplified and extended by several writers, including Shannon himself [8], McMillan [6], Feinstein [2], [3], Khinchin [5], and Wolfowitz [9], [10]. In particular, for certain channels, Wolfowitz has shown that \(\theta_N(c) \to 1\) as \(N \to \infty\) for every \(c > C\).

For a certain class of finite state channels, the indecomposable channels defined below, the fact that \(\theta_N(c) \to 0\) as \(N \to \infty\) for \(c < C\) was first proved by Breiman, Thomasian, and the writer [1]. We present in this paper a simpler proof of the slightly stronger fact that for these channels \(\theta_N(c) \to 0\) exponentially: for any \(c < C\) there are constants \(\alpha > 0, \beta < 1\) for which, for all \(N\),

\[
\theta_N(c) < \alpha \beta^N.
\]

The Shannon-Feinstein inequality reduces (8) at once to the study of the distribution of \(j_N\) for large \(N\), as follows: if for a given \(c\) we can find an input sequence \(\{a_n\}\) for which, for some \(\alpha_1 > 0, \beta_1 < 1\),

\[
P\{j_N \leq Nc\} \leq \alpha_1 \beta_1^N
\]
for all $N$, the Shannon-Feinstein inequality yields, for every $\epsilon > 0$ and all $N$,

$$\theta_N(c - \epsilon) \leq \alpha_1 \beta_1^N + 2^{-N\epsilon} \leq \alpha_2 \beta_2^N,$$

where $\alpha_2 = \alpha_1 + 1$, $\beta_2 = \max(\beta_1, 2^{-\epsilon})$. Thus our result (8) is implied by: for every $c < C$, there is an input sequence $\{a_n\}$ for which, for some $\alpha_1 > 0$, $\beta_1 < 1$, (9) holds.

We now define indecomposable channels and the number $C$. Let

$$\{x_n, n = 1, 2, \cdots\}$$

be any Markov process with a finite number $R$ of states $r = 1, 2, \cdots, R$ and indecomposable transition matrix $\pi = \pi(r'|r) = P\{x_{n+1} = r'|x_n = r\}$. Let $\phi$ be any function from $(1, \cdots, R)$ to $A$, and let $a_n = \phi(x_n)$. We consider the source process $\{a_n\}$ as driving the channel, as described above. The process $\{z_n = (x_n, t_n)\}$ is then a Markov process, with transition matrix

$$m = m(r', t'|r, t) = \pi(r'|r)\phi(t'|r)\phi(t') \begin{cases} \pi(r'|r)\phi(t'|r), & \text{if } t' \neq t, \\ \pi(r'|r), & \text{if } t' = t. \end{cases}$$

If for every indecomposable $\pi$ and every $\phi$, the matrix $m$ is also indecomposable, the finite state channel $(A, B, T, p, \psi)$ is called indecomposable. There is then, for each $m$, a unique stationary Markov process $\{x_n = (x_n^*, t_n^*)\}$ with transition matrix $m$. Define $a_n^* = \phi(x_n^*)$, $b_n^* = \psi(t_n^*)$, $-\infty < n < \infty$, and let $J^*(\pi, \phi) = J^*(a) + J^*(b) - J^*(a, b)$. The number

$$C = \sup_{\pi, \phi} J^*(\pi, \phi),$$

where the sup is over all indecomposable $\pi$ and all $\phi$, is called the capacity of the channel. The main result of this paper is

**Theorem 1.** Let $(A, B, T, p, \psi)$ be an indecomposable channel of capacity $C$. For every $c < C$ there is an input sequence $\{a_n, n = 1, 2, \cdots\}$ and there are numbers $\alpha > 0$, $\beta < 1$ for which, for all $N$,

$$P\{j[(a_1, \cdots, a_N), (b_1, \cdots, b_N)] \leq N\epsilon\} \leq \alpha \beta^N.$$ 

2. Preliminary reduction of theorem 1

To prove theorem 1, we choose $\pi, \phi$ for which $J^* = J^*(\pi, \phi) > c$. Let $z_n = (x_n, t_n)$, with $n = 0, 1, 2, \cdots$ be a Markov process with the transition matrix $m$ and some initial distribution for which the initial distribution of $t_0$ is uniform. Let $a_n = \phi(x_n)$, $b_n = \psi(t_n)$, with $n = 1, 2, \cdots$. We shall show that the input sequence $\{a_n\}$ has the property specified by theorem 1. Let us write $u_N = (a_1, \cdots, a_N)$, $v_N = (b_1, \cdots, b_N)$. Since $i(u_N, v_N) = i(u_N) + i(v_N) - i(u_N, v_N)$ and $J^* = J^*(a) + J^*(b) - J^*(a, b)$, theorem 1 would be proved if we could bound the probability of each of the events

$$\{i(u_N) \leq N(I^*(a) - \delta)\},$$

$$\{i(v_N) \leq N[I^*(b) - \delta]\},$$

$$\{i(u_N, v_N) \geq N[I^*(a, b) + \delta]\}$$

for all $N$, the Shannon-Feinstein inequality yields, for every $\epsilon > 0$ and all $N$,
above by $\alpha \beta^N$ for some $\alpha > 0$, $\beta < 1$, where $J^* - c = 3\delta$. That we can do this is the assertion of

**Theorem 2.** There are functions $\alpha = \alpha(R, w, \epsilon)$, $\beta = \beta(R, w, \epsilon)$, defined for $R = 2, 3, \cdots, w > 0$, $\epsilon > 0$, continuous in $w, \epsilon$, increasing in $R$ and decreasing in $w, \epsilon$ with $\alpha > 0$ and $0 < \beta < 1$ such that, for any Markov process

$$\{z_n, n = 1, 2, \cdots\}$$

with $R$ states $r = 1, 2, \cdots, R$, indecomposable transition matrix $\pi = \pi(r'|r) = P\{z_{n+1} = r'|z_n = r\}$ with smallest positive element $\geq w$ ($\pi$ may have some elements 0) and any function $\phi$ from $1, \cdots, R$ into a finite set $A$, (14)

$$P\{|\phi(a_1, \cdots, a_N) - NI^*(\phi)| \geq Ne\} \leq \alpha(R, w, \epsilon)\beta^N(R, w, \epsilon)$$

for all $N$, where $a_n = \phi(z_n)$, and $I^*(\phi)$ is as defined in section 1, namely if $\{z^*_n, -\infty < n < \infty\}$ is a stationary Markov process with transition matrix $\pi$ and $a^*_n = \phi(z^*_n)$, then $I^*(a_n) = I(a^*_n|a^*_{-1}, a^*_{-2}, \cdots)$.

Theorem 2 is a form of the equipartition theorem (Shannon [7], McMillan [6]) for “finitary” processes, with an exponential bound on the probability of exceptional sequences.

3. Proof of theorem 2 for $\phi$ the identity

For $\phi$ the identity function, so that $\alpha_n(z_n)$ is itself a Markov process, we have

$$i_N = i(z_1, \cdots, z_N) = -\log \lambda(z_1) - \sum_{n=1}^{N-1} \log \pi(z_{n+1}|z_n).$$

We use the following inequality of Katz and Thomasian [4].

**Katz-Thomasian inequality.** For $\{z_n\}$, $w$ as in theorem 2 and $\phi$ real-valued, $P\{|\phi(z_1) + \cdots + \phi(z_N) - N\mu| \geq Ne\} \leq \alpha_1\beta^N_1$ where

$$\beta_1 = \beta_1(R, w, \epsilon, M) = \exp\left(-\frac{w^2 R^2}{2^6 M^2 + 2}\right),$$

$$\alpha_1 = \alpha_1(R, w, \epsilon, M) = \frac{8R w^8}{1 - \beta_1}.$$ (16)

$M = \max \phi(r') - \min \phi(r),$ and $\mu = \sum \lambda(r)\phi(r)$, where $\lambda$ is the stationary distribution for $\pi$.

We apply the Katz-Thomasian inequality to $z^*_n = (z_n, z_{n+1})$, with $\phi' = -\log \pi(r'|r)$, so that $\mu = -\sum_{r, r'} \lambda(r)\pi(r'|r) \log \pi(r'|r) = I^*(z)$, and $M \leq -\log w$ [we may exclude from $z^*_n$ the pairs $r, r'$ with $\pi(r'|r) = 0$], obtaining

$$P\left[\left|\sum_{n=1}^{N-1} \log \pi(z_{n+1}|z_n) + (N - 1)I^*(z)\right| \geq (N - 1)e\right] \leq \alpha_1(R^2, w, \epsilon, -\log w)\beta^{N-1}_1(R^2, w, \epsilon, -\log w)$$

$$= \alpha_2(R, w, \epsilon)\beta^N_2(R, w, \epsilon),$$

(17)
say. Thus
\[ P\{|i_N - NI^*(z) + \log \lambda(z) + I^*(z)| \geq N\epsilon\} \leq \alpha_2 \beta_2^N. \]
Since \(0 \leq I^*(z) \leq \log R\) and, for every \(\delta > 0\),
\[ P\{|\log \lambda(z)| \geq N\delta\} = \sum_{r: \lambda(r) \geq 2^{-N\delta}} \lambda(r) \leq R2^{-N\delta} \]
we easily obtain \(\alpha_3(R, w, \epsilon), \beta_3(R, w, \epsilon)\) for which
\[ P\{|i_N - NI^*(z)| \geq N\epsilon\} \leq \alpha_3 \beta_3^N. \]

4. Proof of theorem 2, general case

We prove the general case by approximating the process \(\{a_n\}\), in blocks, by a suitable Markov process, and using the fact that we have already proved the theorem for Markov processes. The idea is this: if, in addition to observing the \(a_n = \phi(z_n)\) process we observe periodically, say every \(k\) trials, the current state of the underlying \(z_n\) process, the process now observed, with observations grouped in blocks of \(k\), is a Markov process, so that all long sequences, except a set of exponentially small probability, have about the correct probability. We can choose \(k\) so large that (1) this correct probability is nearly the correct probability for the corresponding \(a\) sequence and (2) except with exponentially small probability, the probability of the actual \(a\) sequence will be near the probability of the actual observed sequence.

Thus choose a positive integer \(k\), and let \(x_1 = (a_1, \ldots, a_{k-1}, z_k), x_2 = (a_{k+1}, \ldots, a_{2k-1}, z_{2k}), \ldots, x_n = (a_{(n-1)k+1}, \ldots, a_{nk-1}, z_{nk}), \ldots\). The \(\{x_n\}\) process is Markov and, for \(k\) relatively prime to the period of \(\{z_n\}\), is indecomposable. It has at most \(R^k\) states, and the smallest positive element in its transition matrix is at least \(w^k\).

Thus, from the preceding section,
\[ P\{|i(x_1, \ldots, x_N) - NI_k| \geq \epsilon N\} \leq \alpha_4 \beta_4^N, \]
where
\[ \alpha_4 = \alpha_4(R, w, \epsilon, k) = \alpha_3(R^k, w^k, \epsilon), \]
\[ \beta_4 = \beta_4(R, w, \epsilon, k) = \beta_3(R^k, w^k, \epsilon), \]
and \(I_k = I^*(x)\). Now \(kI^*(a) \leq I_k \leq kI^*(a) + \log R\) and \(i(x_1, \ldots, x_N) = i(a_1, \ldots, a_N) + i(z_k, \ldots, z_{Nk} | a_1, \ldots, a_N)\), which we write \(i_N(x) = i_N(a) + i_N(z|a)\). Then
\[ P\{|i_{Nk}(a) \geq Nk[I^*(a) + \epsilon]\} \leq P\left\{i_N(x) \geq Nk \left(\frac{I_k - \log R}{k} + \epsilon\right)\right\}, \]
\[ = P\{i_N(x) \geq N[I_k + (k\epsilon - \log R)]\} \leq \alpha_4 \beta_4^N, \]
provided \(k\epsilon - \log R \geq \epsilon\), that is, \(k \geq 1 + (1/\epsilon) \log R\). Similarly,
(24) \[ P\{i_{Nk}(a) \leq Nk[I^*(a) - 4\epsilon]\} \leq P\left\{ i_N(x) - i_N(z|a) \leq Nk \left( \frac{I_k}{k} - 4\epsilon \right) \right\} \]
\[ \leq P\{i_N(x) \leq N(I_k - k\epsilon)\} + P\{i_N(z|a) \geq 3Nk\epsilon\} = P_1 + P_2. \]

As above, \( P_1 \leq \alpha_4\beta^N_5 \). To bound \( P_2 \), write \( i_N(z) = i(z_0, z_2, \ldots, z_Nk) \). Then
\[ (25) P_2 \leq P\{i_N(z) \geq Nk\epsilon\} + P\{i_N(z|a) \geq i_N(z) + Nk\epsilon\} = P_3 + P_4. \]

The process \( \{z_n, n = 1, 2, \ldots, k \text{ fixed}\} \) is a Markov process with \( R \) states and \( (k \text{ is relatively prime to the period of } \{z_n\}) \) indecomposable transition matrix. For \( k \) so large that \( k\epsilon \geq \log R + \epsilon \), that is, \( k \geq 1 + (1/\epsilon) \log R \), we have \( P_3 \leq \alpha_6\beta^N_5 \).

For \( P_4 \) we use

**Lemma 1.** For any two random variables \( a, z \) each with a finite set of values, and any \( \delta \geq 0 \),
\[ (26) P\{i(z|a) \geq i(z) + \delta\} \leq 2^{-\delta}. \]

**Proof.** A pair \((z_0, a_0)\) of values of \( z, a \) for which \( i(z_0|a_0) \geq (z_0) + \delta \) is one for which
\[ (27) \frac{P\{z = z_0|a = a_0\}}{P\{z = z_0\}} \leq 2^{-\delta}, \]
that is, \( P\{z = z_0, a = a_0\} \leq 2^{-k}\delta P\{z = z_0\}P\{a = a_0\} \). Summing over all pairs \((z_0, a_0)\) for which the inequality is satisfied yields the lemma.

From the lemma, we obtain \( P_4 \leq 2^{-Nk\epsilon} \). Thus
\[ (28) P\{i_{Nk}(a) \leq Nk[I^*(a) - 4\epsilon]\} \leq \alpha_4\beta^N_5, \]
where \( \alpha_5 = \alpha_4(R, w, \epsilon, k) = \alpha_4 + \alpha_3 + 1 \) and \( \beta_5 = \max(\beta_4, \beta_3, 2^{-k\epsilon}) \).

Combining (23) and (28) we obtain \( \alpha_6(R, w, \epsilon, k), \beta_6(R, w, \epsilon, k) \) for which
\[ (29) P\{i_{Nk} - Nk[I^*(a)] \geq Nk\epsilon\} \leq \alpha_7\beta^N_7 \]

The block size \( k \) is still at our disposal, subject to \( k \geq 1 + (1/\epsilon) \log R \) and relatively prime to the period of \( \{z_n\} \). We can find such a
\[ k \leq k^* = \lceil R + 1 + (1/\epsilon) \log R \rceil \]
and obtain, for this \( k \),
\[ (30) P\{i_{Nk} - Nk[I^*(a)] \geq Nk\epsilon\} \leq \alpha_7\beta^N_7, \]
where
\[ \alpha_7 = \alpha_7(R, w, \epsilon) = \alpha_6(R, w, \epsilon, k^*), \]
\[ \beta_7 = \beta_7(R, w, \epsilon) = \beta_6(R, w, \epsilon, k^*). \]

Finally, for any \( n \), say, \( n = Nk + d \), with \( 0 \leq d \leq k - 1 \), we have
\[ (32) i_n - n[I^*(a) + \epsilon] \leq i_{(N+1)k} - (N + 1)k \left( I^*(a) + \epsilon - \frac{I^*(a) + \epsilon}{N + 1} \right) \]
\[ \leq i_{(N+1)k} - (N + 1)k \left( I^*(a) + \frac{\epsilon}{2} \right) \]
for

\[ \frac{I^*(a) + \epsilon}{N + 1} \leq \frac{\epsilon}{2}, \]

which, since \( I^*(a) \leq \log R \), will certainly hold for

\[ N \geq 2 \left( \frac{\log R}{\epsilon} + 1 \right) = N_0, \]

say, and similarly \( i_n - n \{I^*(a) - \epsilon\} \geq i_n - Nk \{I^*(a) - \epsilon/2\} \) for

\[ N \geq 2 \left( \frac{\log R}{\epsilon} - 1 \right). \]

Thus

\[ P \{|i_n - nI^*(a)| \geq n\epsilon\} \leq \alpha_8 \beta_8^{N - N_0}, \]

where

\[ \alpha_8 = \alpha_8(R, w, \epsilon) = 2\alpha_7 \left( R, w, \frac{\epsilon}{2} \right) \]

\[ \beta_8 = \beta_8(R, w, \epsilon) = \beta_7 \left( R, w, \frac{\epsilon}{2} \right). \]

Finally, with

\[ \alpha_9 = \alpha_8 \beta_8^{-N_0}, \quad \beta_9 = \beta_8^{1/k_*}, \]

we obtain

\[ P \{|i_n - nI^*(a)| \geq n\epsilon\} \leq \alpha_9 \beta_9^n \]

for all \( n \), completing the proof.

REFERENCES


