EXPONENTIAL BOUNDS ON THE PROBABILITY OF ERROR FOR A DISCRETE MEMORYLESS CHANNEL

BY SAMUEL KOTZ

Cornell University and Bar-Ilan University, Ramat-Gan, Israel

1. Summary. In a paper by Blackwell, Breiman and Thomasian [1, Theorem 3] the following theorem is proved:

For any integer n and for any $0 < \epsilon \leqq \frac{1}{2}$, such that $C - \epsilon \geqq 0$ there exists a code for a discrete memoryless channel with length $N > e^{n(C-\epsilon)}$ and with a bound for the probability of error, $\bar{\lambda} = 2 \exp_e - [n\epsilon^2/(16ab)]$, where C is the capacity of the channel and a and b are the numbers of elements in the input and output alphabets respectively.

In this note we shall replace the bound $2 \exp_e[-n\epsilon^2/(16ab)]$ by the expression $2 \exp_e\{-n\epsilon^2/[g(c)(\log c)^{2-\delta}]\}$, where $c = \min(a, b)$, $g(c)$ is a positive monotonically decreasing function of c, $g(c) < 16$ for all $c \geqq 3$ and approaches 2 asymptotically as $c \to \infty$, and $\delta > 0$ depends on ϵ and c and tends to 0 as either $c \to \infty$ or $\epsilon \to 0$.

2. Preliminary Lemmas.

LEMMA 1. *Let*

$P_{ij} \geqq 0$ $(i = 1, \cdots, a, j = 1, \cdots b)$, $\sum_{i,j}^{a,b} P_{ij} = 1$, $P_i = \sum_j P_{ij}$, $Q_j = \sum_i P_{ij}$ *and* $c = \min(a, b)$. *Then*

$$\sum_{i,j}^{a,b} P_{ij}\left(\log \frac{P_{ij}}{P_i Q_j}\right)^2 \leqq [\log(1 + e + c)]^2 \quad \text{for all } c, \tag{1}$$

$$\sum_{i,j}^{a,b} P_{ij}\left(\log \frac{P_{ij}}{P_i Q_j}\right)^2 \leqq 2.343 (\log c)^2 \quad \text{for} \quad c = 2, \tag{2}$$

$$\sum_{i,j}^{a,b} P_{ij}\left(\log \frac{P_{ij}}{P_i Q_j}\right)^2 \leqq 2 (\log c)^2 \quad \text{for} \quad c \geqq 3, \tag{3}$$

$$\sum_{i,j}^{a,b} P_{ij}\left(\log \frac{P_{ij}}{P_i Q_j}\right)^2 \leqq 4e^{-2} + (\log c)^2 \quad \text{for} \quad c \geqq 12. \tag{4}$$

PROOF:

(1). Let

$$s_1 = \{(i, j) \mid 0 \leqq P_{ij}/(P_i Q_j) < e^{-1}\}$$
$$s_2 = \{(i, j) \mid e^{-1} \leqq P_{ij}/(P_i Q_j) \leqq e\}$$
$$s_3 = \{(i, j) \mid P_{ij}/(P_i Q_j) > e\}$$

and let $S = \sum P_{ij} \{\log [P_{ij}/(P_i Q_j)]\}^2$. Then

Received August 23, 1960.

$$S \leqq \sum_{s_1} P_{ij} f\left(\frac{P_i Q_j}{P_{ij}}\right) + \sum_{s_2} P_{ij} f(e) + \sum_{s_3} P_{ij} f\left(\frac{P_{ij}}{P_i Q_j}\right), \tag{5}$$

where $f(x) = (\log x)^2$, convex for $x \geqq e$.

Since the arguments of f in (1) are all $\geqq e$, $S \leqq f(K)$, where

$$K = \sum_{s_1} P_{ij} \frac{P_i Q_j}{P_{ij}} + \sum_{s_2} P_{ij} e + \sum_{s_3} \frac{P_{ij}^2}{P_i Q_j} \geqq 1,$$

since $K = \sum_{i,j} P_i Q_j x_{ij}$, where all x_{ij} are $\geqq 1$. However,

$$\sum_{s_1} P_{ij} \frac{P_i Q_j}{P_{ij}} \leqq 1; \qquad \sum_{s_2} P_{ij} \leqq e$$

and

$$\sum_{s_3} \frac{P_{ij}^2}{P_i Q_j} \leqq \sum_{i,j} \frac{P_{ij}}{P_i} = \sum_{i=1}^{a} \frac{1}{P_i} (\sum_j P_{ij}) = a,$$

and similarly $\sum_{s_3} P_{ij}^2/(P_i Q_j) \leqq b$. Since $f(x)$ is monotonically increasing for $x \geqq 1$, the result follows

(2) and (3). Consider $f(P_1, \cdots, P_n) = \sum_{i=1}^n P_i (\log P_i)^2$, where $P_i \geqq 0$ and $\sum_{i=1}^n P_i = 1$.

Using the method of Lagrange multpliers we easily find the unique maximum of this function for the case $n > e$ (i.e., $n \geqq 3$) to be $(\log n)^2$, which is attained for $p_i = n^{-1} (i = 1, \cdots, n)$. Let, now,

$$\begin{aligned} S = \sum_{i,j} P_{ij} \left(\log \frac{P_{ij}}{P_i Q_j}\right)^2 &= \sum_j Q_j \sum_i \frac{P_{ij}}{Q_j} \left(\log \frac{P_{ij}}{Q_j}\right)^2 \\ &\quad - 2 \sum_{i,j} P_{ij} \left(\log \frac{P_{ij}}{Q_j}\right) (\log P_i) + \sum_i (\log P_i)^2 \cdot P_i . \end{aligned} \tag{6}$$

From the above it follows that the first and the last terms of (6) are $\leqq (\log a)^2$ and the second is non-positive. Hence, owing to the symmetry of S in i and j, the assertion (3) follows.

(2) follows immediately by using the same method and considering the function $x(\log x)^2 + (1 - x)[\log(1 - x)]^2$ for $0 \leqq x \leqq 1$.

(4). Let

$$s_1^* = \{(i, j) \mid 0 \leqq P_{ij}/(P_i Q_j) \leqq 1\}; \qquad s_2^* = \{(i, j) \mid 1 < P_{ij}/(P_i Q_j) \leqq e\};$$
$$s_3 = \{(i, j) \mid P_{ij}/(P_i Q_j) > e\}.$$

Let $f(x) = (\log x)^2 (x > 0)$; $h(x) = x \log^2 x (x \geqq 0)$ and $g_K(x) = x \log^2(x/K) - x$ $(x \geqq 0$, K-integral$)$.

It is easily seen by elementary methods that

$$\max_{0 \leqq x \leqq 1} g_K(x) = g_K(1) = (\log K)^2 - 1 \text{ for } K > e^{1+\sqrt{2}}$$

and

$$\max_{0 \leqq x \leqq 1} h(x) = 4e^{-2}.$$

Now

$$\sum_{s_1^*} P_{ij}\left(\log\frac{P_{ij}}{P_iQ_j}\right)^2 = \sum P_iQ_jh\left(\frac{P_{ij}}{P_iQ_j}\right) \leqq 4e^{-2}\cdot\sum_{s_2^*} P_{ij}\left(\log\frac{P_{ij}}{P_iQ_j}\right)^2$$

$$\leqq \max_{s_2^*} f\left(\frac{P_{ij}}{P_iQ_j}\right)\sum_{s_2^*} P_{ij} = \sum_{s_2^*} P_{ij} \leqq 1 - \sum_{s_3} P_{ij} = 1-\alpha, \quad \text{say,}$$

since $f(e) = 1$ and this function is monotonically increasing on s_2^*.

Moreover $\sum_{s_3} P_{ij}\{\log[P_{ij}/(P_iQ_j)]\}^2 = \alpha\sum_{s_3}[P_{ij}/\alpha]f[P_{ij}/(P_iQ_j)]$, and, since f is convex on s_3, we have

$$\sum_{s_3} P_{ij}\left(\log\frac{P_{ij}}{P_iQ_j}\right)^2 \leqq \alpha f\left(\frac{\alpha}{\theta}\right) \quad \text{where} \quad \theta = \sum\frac{P_{ij}^2}{P_iQ_j}.$$

Thus

$$\sum_{s_2^*} + \sum_{s_3} \leqq 1 + \left[\alpha f\left(\frac{\alpha}{\theta}\right) - \alpha\right]. \tag{7}$$

Now $\theta \leqq \min(a, b) = c$, and on s_3: $e \leqq (\theta/\alpha) \leqq (c/\alpha)$. Therefore, from the monotonicity of f on s_3, it follows that

$$f(\alpha/\theta) = f(\theta/\alpha) \leqq f(c/\alpha) = f(\alpha/c).$$

From the definition of $g_K(x)$ and (7) we obtain

$$\sum_{s_2^*} + \sum_{s_3} \leqq 1 + g_c(\alpha).$$

Hence

$$\sum_{s_2^*} + \sum_{s_3} \leqq (\log c)^2 \quad \text{for} \quad c > e^{1+\sqrt{2}} \qquad (\text{i.e., } c \geqq 12)$$

From here the assertion follows.

LEMMA 2. *Let* $0 < \delta \leqq 1$ *and* $t > 0$, *then, for* $x \geqq \delta$,

$$x^{-t} \leqq 1 - t\log x + [\tfrac{1}{2}\delta^{-t}(t\log x)^2],$$

where the equality occurs if and only if $x = \delta = 1$.

PROOF. The result follows directly from the obvious inequality

$$e^y \leqq 1 + y + (\tfrac{1}{2}e^R)y^2, \qquad \text{for all } y \leqq R, \tag{8}$$

where $y \leqq R$ and R is any non-negative number, by substituting $y = -t\log x$. The equality in (8) holds if and only if $y = R = 0$.

3. Proof of the main result. Consider a discrete memoryless channel with input alphabet having a (>1) elements and the output alphabet having b (>1) elements. Let $P(\cdot)$ be a probability distribution on the elements i of the input alphabet ($i = 1, \cdots, a$), and let $P(\cdot \mid i)$ be a distribution of the elements j of the output alphabet ($j = 1, \cdots, b$) for every i of the input alphabet.

As in [1] we start with the r.v. $J(P)$ defined by

$$\Pr\left\{J(P; i, j) = \log \frac{P_{(i,j)}}{P_{(i)}\, Q_{(j)}}\right\} \begin{array}{ll} = P(i,j) & \text{if } P(i,j) > 0 \\ = 0 & \text{if } P(i,j) = 0, \end{array}$$

where $P(i, j) = P(j \mid i)P(i)$ and $Q(j) = \sum_i P(i, j)$.

It is well known (e.g., [1]) that the capacity C of a channel is defined by $C = \sup_P EJ(P)$, where the supremum taken over all possible input distributions, is actually attained for some $P = \bar{P}$. We choose in the definition of the r.v. J the input distribution to be $\bar{P}$, so that $C = EJ$.

The moment generating function of J is given by

$$E(e^{-Jt}) = \sum_{i,j} P(i,j) \left[\frac{P(i,j)}{P(i)Q(j)}\right]^{-t}. \tag{9}$$

Let $t > 0$ and $I_n = J_1 + \cdots + J_n$, where $J_K (K = 1, \cdots, n)$ are independent, identically distributed random variables with the distribution of J.

From Chebyshev's inequality it follows that, for any $\epsilon > 0$,

$$\Pr\{I_n \leqq n(C - \epsilon)\} \leqq [e^{t(C-\epsilon)}\, E(e^{-tJ})]^n. \tag{10}$$

Let $0 < \delta < 1$ and $t < 1$. Denoting by $\sum'$ the sum in the right hand side of (9) over all i, j for which $P(i, j)/[P(i)Q(j)] \leqq \delta$, we obtain

$$\sum_{i,j}{}' P(i,j) \left(\frac{P(i,j)}{P(i)Q(j)}\right)^{-t} \leqq \sum_{i,j} \delta^{1-t} P(i)Q(j) = \delta^{1-t}.$$

Denoting by $\sum''$ the sum in the right hand side of (9) over all i, j for which $P(i, j)/[P(i)Q(j)] > \delta$ and using Lemma 2, we have

$$\sum_{i,j}{}'' P(i,j) \left(\frac{P(i,j)}{P(i)Q(j)}\right)^{-t} < \sum_{i,j} P(i,j) \left[1 - t \log \frac{P(i,j)}{P(i)Q(j)} + \frac{\delta^{-t}}{2} t^2 \left(\log \frac{P(i,j)}{P(i)Q(j)}\right)^2\right].$$

Let $h(c) = 2.343$ for $c = 2$, $h(c) = \min \left[\left(\frac{\log(1 + e + c)}{\log c}\right)^2, 2\right]$ for $3 \leqq c \leqq 11$ and $h(c) = \min\{[\log(1 + e + c)/\log c]^2, [4e^{-2} + (\log c)^2]/(\log c)^2\}$ for $c \geqq 12$.

Since $\delta < 1$, we obtain, using Lemma 1 and the definition of C,

$$E(e^{-tJ}) < 1 - tC + h(c)(\tfrac{1}{2}\delta^{-t}t^2)(\log c)^2 + \delta^{1-t},$$

where $c = \min(a, b)$.

Let $\delta^{1-t} = qt^2 (\log c)^2$, $q > 0$ and such that $qt^2(\log c)^2 < 1$. We have

$$E(e^{-tJ}) < 1 - tC + \frac{t^2}{2} \{h(c)\, [qt^2(\log c)^2]^{-t/(1-t)} + 2q\}\, (\log c)^2. \tag{11}$$

We minimize the expression in curly brackets of (11) with respect to q. The unique minimum is obtained for

$$q_{\min} = [h(c)]^{1-t}[t/(1-t)]^{1-t}(t \log c)^{-2t}. \tag{12}$$

It can be easily checked that $q_{\min}\, t^2(\log c)^2 < 1$ for all $c \geqq 2$, and for all t such that $t \leqq \min\left(\tfrac{1}{4}, [h(c)K(t)(\log c)^{2-t}]^{-1}\right)$, where

$$K(t) = 2^t(t)^{-2t}\{[(1-t)/t]^t + [t/(1-t)]^{1-t}\},$$

which we shall require soon.

Thus, since $h(c) > 1$, we obtain from (11) and (12)

$$E(e^{-tJ}) \leqq 1 - tC + \tfrac{1}{2}t^2h(c)K(t)(\log c)^{2-t}. \tag{13}$$

$K(t)$ tends to 1 as $t \to 0$ and the approach is monotonic starting from $t = 0.5100$.

Using the inequality $1 + x \leqq e^x$ we obtain from (10) and (13)

$$\Pr\{I_n \leqq n(C - \epsilon)\} \leqq \{e^{-\epsilon t} e^{(\frac{1}{2}t^2)h(c)K(t)(\log c)^{2-t}}\}^n. \tag{14}$$

We shall now assume that $c \geqq 3$.

Let $0 < \epsilon \leqq \frac{1}{2}$ be given. For each integer c we choose a real number $m > 1$ such that if $t \leqq [mh(c)]^{-1}$, then $K(t) \leqq D$ and also $\{2D(\log c)^{2-[2Dh(c)\log c]^{-1}}\}^{-1} \leqq m^{-1}$. (Clearly $m \uparrow \infty$ and $D \downarrow 1$ as $c \uparrow \infty$.)

We set

$$t = t_0 = \epsilon/[h(c)D(\log c)^{2-t}] \tag{15}$$

so that $t_0 < \{2D\,h(c)(\log c)^{2-[2Dh(c)\log c]^{-1}}\}^{-1} < \frac{1}{4}$, (see Table 1).

Next, we define $R = C - \epsilon$ and $d = 2D\,h(c)(\log c)^{2-[2h(c)D(\log c)^2]^{-1}}$ (clearly, d $\uparrow \infty$ with c). For $0 < \epsilon \leqq \frac{1}{2}$,

$$R + (\epsilon^2/d) \leqq C - [1 - (2d)^{-1}]\epsilon. \tag{16}$$

From (16), (14) and (15) we obtain

$$\Pr\{I_n \leqq n[R + (\epsilon^2/d)]\} \leqq \exp\left\{ - \frac{n\epsilon^2}{g(c)(\log c)^{2-\{\epsilon/[Dh(c)(\log c)^2]\}}} \right\}, \tag{17}$$

where $h(c)D2d/(d-1) = g(c)$. (Clearly, $g(c) \downarrow 2$ as $c \to \infty$.)

As in [1], we will now apply the basic theorem of Feinstein which states:

For any discrete memoryless channel and for any two positive numbers θ and λ, with $\lambda \leqq 1$, any input $P(\cdot)$, and any n, there exists a code (n, N, λ) such that

$$N > e^{\theta}[\lambda - P_r\{I_n(P) \leqq \theta\}] \tag{18}$$

(See [1].)

We set

$$\theta = n[R + (\epsilon^2/d)] \text{ and } \lambda = 2\exp\left\{ - \frac{n\epsilon^2}{g(c)(\log c)^{2-\{\epsilon/[Dh(c)(\log c)^2]\}}} \right\}.$$

Applying (17) and (18), we obtain for the case of $c \geqq 3$ the existence of a code with length $N > e^{n(C-\epsilon)}$ and probability of error

$$\lambda = 2 \exp_e - \left\{ \frac{n\epsilon^2}{g(c)(\log c)^{2-\{\epsilon/[Dh(c)(\log c)^2]\}}} \right\} \quad \text{for } 0 < \epsilon \leqq \frac{1}{2}.$$

The case $c = 2$ requires several obvious modifications in the definitions. One of the possibilities is to set $[2D(\log c)]^{-1} \leqq m^{-1}$, $t_0 = \epsilon/[h(c)D \log c]$, $d = 2D\, h(c) \log c$ and to redefine

$$g(c) = \frac{h(c)D2d/(d-1)}{\log c}.$$

This case was treated numerically using the Cornell Computing Center's Burrough 220, where also the numerical values of $g(c)$ for values of c in the range of 3-25 were computed. The results of these computations are presented in Table 1.

TABLE 1

The computed values of $h(c)$, m, D, d and $g(c)$ for values of c in the range of 2 25.

c	h(c)	$[m\, h(c)]^{-1}$	D	d	g(c)
2	2.343	0.1187	2.593	8.422	19.891
3	2.000	0.0421	1.600	7.632	7.366
4	2.000	0.0232	1.351	10.061	5.999
5	1.810	0.0168	1.264	11.387	5.017
6	1.611	0.0136	1.219	12.037	4.283
7	1.486	0.0114	1.189	12.724	3.833
8	1.401	0.0099	1.165	13.410	3.529
9	1.340	0.0087	1.148	14.088	3.312
10	1.293	0.0077	1.134	14.745	3.148
12	1.088	0.0073	1.128	14.263	2.638
15	1.074	0.0055	1.100	16.360	2.517
18	1.065	0.0044	1.083	18.231	2.440
20	1.060	0.0039	1.074	19.379	2.402
25	1.052	0.0030	1.060	21.964	2.336

I am much indebted to Professors J. Wolfowitz and J. Kiefer and to Dr. S. Kantorovitz for their valuable comments.

REFERENCE

[1] David Blackwell, Leo Breiman and A. J. Thomasian, "The capacity of a class of channels," *Ann. Math. Stat.*, Volume 30 (1959), pp. 1229–1241.