Z. Wahrscheinlichkeitstheorie verw. Geb. 14, 61 – 88 (1969)

Error Estimates for Low Rate Codes

UDO AUGUSTIN

Summary. Gallager [1] and Gallager, Shannon, Berlekamp [2] establish exponentially decreasing upper and lower bounds, respectively, on the error of the best codes for fixed code rates R smaller than the capacity for the standard channels (stationary finite alphabet channels without memory). These bounds happen to coincide up to the first order for rates near to the capacity.

The authors of [2] regret that their proof of the lower bound cannot be extended to infinite alphabet channels or nonstationary channels because of the use of "fixed composition codes" (while the proofs of the upper estimate can be easily transferred to those channels).

Changing parts of the proof in [2], we automatically obtain estimates of the same type as in [2] for the latter channels.

Furthermore, as a matter of minor importance, we show that the order of coincidence of upper and lower bound for the high rates R can be rigorously improved.

§ 1. The Model for the Channel Without Memory

In order to avoid unnecessary effort with measurability considerations (which have no strong bearing on the estimates in our problem) and with index notations we will choose a model for the general channel without memory that differs from the usual model.

Let the triple (X, F, M) denote a set $X \neq \emptyset$ with a σ-field F on X and a nonempty set $M = \{p\}$ of probabilities on the measure space (X, F).

Remark. (X, F, M) may be considered as the model for a channel (for a fixed time) as long as one is concerned only with the relation between the length and the error of codes. Interpreting X as the set of output symbols of the channel and M as the set of transition probabilities from a set of input symbols to X one can neglect the input symbols, because the receiver can separate different input symbols only if they have different transition probabilities; the relation between the length and the error of a code depends only on the structure of M as a set of probabilities. (Usually a measure kernel is taken as the model for an infinite alphabet channel. In that case only the set of output sources determines the relation between length and error of codes. The set of output sources (on X) can be considered as the set of all transition probabilities (to X) for a new channel which is an extension of the kernel channel. The set of output sources may be identified with a set M. For a more complete discussion of this remark see [3].)

Definition 1.1 (code, length of a code, error of a code). Given (X, F, M).

A finite code $\{(p^i, E_i)\}_{1 \leq i \leq N}$ of M is a finite sequence of pairs (p^i, E_i) $(1 \leq i \leq N)$ where the $p^i \in M$, and $E_i \in F$ are pairwise disjoint sets. (The E_i are the decoding sets; the empty set is admissible for a decoding set.) We call N the length of the code and $p_e := \sup_{1 \leq i \leq N} p^i({}^c E_i)$ (where ${}^c E_i := X - E_i$) the error of the code. Furthermore, we set $P_e(M, N) := \inf\{p_e : p_e \text{ is error of a code of length } N \text{ of } M\}$.

We are interested in estimates of the error of codes in case that M has "product structure" and we derive these estimates by deriving estimates for finite subsets of M.

Now, for every natural $v\,(1 \leqq v \leqq t)$ let a triple (X_v, F_v, M_v) be given as above. Furthermore, let

$$X_{[1,t]} := \prod_{v=1}^{t} {}^{x}X_v \text{ denote the product space,}$$

$$F_{[1,t]} := {}^{F}\left(\prod_{v=1}^{t} {}^{x}F_v\right) \text{ the product } \sigma\text{-field on } X_{[1,t]}$$

and

$$M_{[1,t]} := M_1 \times \cdots \times M_t := \{p_1 \times \cdots \times p_t : p_v \in M_v,\ 1 \leqq v \leqq t\}$$

the set of product probabilities on $(X_{[1,t]}, F_{[1,t]})$ with v-th component in M_v.

Definition 1.2 (channel without memory and stationary channel without memory for the discrete time $[1, t]$).

$(X_{[1,t]}, F_{[1,t]}, M_{[1,t]})$ (or $M_{[1,t]}$) is the model for a channel without memory operating for the discrete time interval $[1, t]$. If, in addition, the M_v are copies of each other then we speak of a stationary channel without memory for the discrete time $[1, t]$.

Along with the remark at the beginning of this paragraph one finds that this is a version of the classical definition of a channel without memory (see Wolfowitz [4]).

Convention. The lower index on probabilities will be used to indicate the time (index for the components of product probabilities), the upper index will be used for a denumeration of finite sequences of probabilities and the upper index in brackets on probabilities for denumerations of finite subsets of M and $M_v\,(1 \leqq v \leqq t)$ respectively.

§ 2. Stochastic Inequalities for the Error

Gallager [1] uses the random coding method (see f.i. [4]) to upper bound $P_e(M, N)$. We review his estimate for our model of the channel:

Let (X, F, M') be given where $M' = \{p^{(j)}\}_{1 \leqq j \leqq n}$ is any finite set of probabilities on (X, F). Furthermore, let λ be any probability s.t. $p \ll \lambda$ for $p \in M'$.

One derives a code $\{(p^i, E_i)\}_{1 \leqq i \leqq 2N}$ of length $2N$ for M' from an arbitrary sequence $\{p^i\}_{1 \leqq i \leqq 2N}$ in M' for instance by prescribing for the sets E_i:

a) $E_i = \left\{x \in X : \dfrac{dp^i}{d\lambda} > \bigvee_{k \neq i} \dfrac{dp^k}{d\lambda},\ (1 \leqq k \leqq 2N)\right\}$ λ – a.e. and

b) the E_i are pairwise disjoint $(1 \leqq i \leqq 2N)$.

For all $i\,(1 \leqq i \leqq 2N)$ holds for such a code:

$$\begin{aligned} p^i({}^{c}E_i) &= p^i\left\{\frac{dp^i}{d\lambda} \leqq \bigvee_{k \neq i} \frac{dp^k}{d\lambda}\right\} \\ &\leqq p^i\left\{\left(\frac{dp^i}{d\lambda}\right)^{\frac{u}{1+u}} \leqq \left[\sum_{k \neq i} \left(\frac{dp^k}{d\lambda}\right)^{\frac{1}{1+u}}\right]^u\right\} \\ &\leqq \int dp^i \left(\frac{dp^i}{d\lambda}\right)^{\frac{-u}{1+u}} \left[\sum_{k \neq i} \left(\frac{dp^k}{d\lambda}\right)^{\frac{1}{1+u}}\right]^u \qquad (0 \leqq u < \infty) \end{aligned}$$

and therefore

(1) $$p^i({}^cE_i) \leqq \int d\lambda \left(\frac{dp^i}{d\lambda}\right)^{\frac{1}{1+u}} \left[\sum_{k \neq i} \left(\frac{dp^k}{d\lambda}\right)^{\frac{1}{1+u}}\right]^u \quad \text{for } 1 \leqq i \leqq 2N \text{ and } 0 \leqq u.$$

Now let $a^{(j)} \geqq 0$ $(1 \leqq j \leqq n)$ be real numbers with $\sum_{j=1}^{n} a^{(j)} = 1$.

Pick sequences $\{p^i\}_{1 \leqq i \leqq 2N}$ of length $2N$ from M', every p^i independently, according to the probability $(a^{(1)}, \ldots, a^{(n)})$ ($a^{(j)}$ is the probability that $p^i = p^{(j)}$) and let these sequences define codes as above.

Keep p^i fixed. Then for the average error $\overline{p_e^i}$ belonging to p^i holds with (1):

$$\overline{p_e^i} \leqq \int d\lambda \left(\frac{dp^i}{d\lambda}\right)^{\frac{1}{1+u}} \left[(2N-1)\sum_{j=1}^{n} a^{(j)} \left(\frac{dp^{(j)}}{d\lambda}\right)^{\frac{1}{1+u}}\right]^u \quad \text{if } 0 \leqq u \leqq 1$$

(Jensen's inequality for the concave function y^u). Hence

$$\overline{p_e^i} \leqq (2N)^u \int d\lambda \left(\frac{dp^i}{d\lambda}\right)^{\frac{1}{1+u}} \left[\sum_{j=1}^{n} a^{(j)} \left(\frac{dp^{(j)}}{d\lambda}\right)^{\frac{1}{1+u}}\right]^u.$$

One obtains for the average error $\bar{p}_e$ for the i-th place in a sequence of length $2N$:

(2) $$\begin{aligned} \bar{p}_e &\leqq (2N)^u \int d\lambda \left[\sum_{l=1}^{n} a^{(l)} \left(\frac{dp^{(l)}}{d\lambda}\right)^{\frac{1}{1+u}}\right] \left[\sum_{j=1}^{n} a^{(j)} \left(\frac{dp^{(j)}}{d\lambda}\right)^{\frac{1}{1+u}}\right]^u \\ &= (2N)^u \int d\lambda \left[\sum_{j=1}^{n} a^{(j)} \left(\frac{dp^{(j)}}{d\lambda}\right)^{\frac{1}{1+u}}\right]^{1+u} \quad (0 \leqq u \leqq 1). \end{aligned}$$

At least one code of length $2N$ has a subcode of length N with error $p_e \leqq 2\bar{p}_e$.

Therefore we have:

Lemma 2.1 *(compare:* Gallager [1]*). Let* (X, F, M') *be given where* $M' = \{p^{(j)}\}_{1 \leqq j \leqq n}$ *is any finite set of probabilities on* (X, F) *and* λ *a probability s.t.* $p \ll \lambda$ *for* $p \in M'$*. Furthermore, let* $a^{(j)} \geqq 0$ $(1 \leqq j \leqq n)$, $\sum_{j=1}^{n} a^{(j)} = 1$ *and* $0 \leqq u \leqq 1$.

Then M' *has a code of length* N *with error*

$$p_e \leqq 4N^u \int d\lambda \left[\sum_{j=1}^{n} a^{(j)} \left(\frac{dp^{(j)}}{d\lambda}\right)^{\frac{1}{1+u}}\right]^{1+u}.$$

(Observe that the right hand side integral is independent of λ.)

We discuss this estimate for time structure and for nonfinite sets of probabilities within the next paragraph. It should be remarked that the method of maximal codes (see f.i. [4]) gives only estimates for low rate codes which are less tight than the estimate above.

Now a stochastic inequality for the lower estimate of $P_e(M,N)$ will be established:

Given any code $\{(p^i, E_i)\}_{1\leqq i\leqq N}$ with error p_e and a probability λ s.t. $p^i \ll \lambda$ $(1\leqq i\leqq N)$ then

$$S_i\,\lambda(E_i)\geqq p^i\left(\left\{\frac{dp^i}{d\lambda}<S_i\right\}\cap E_i\right)\geqq p^i\left\{\frac{dp^i}{d\lambda}<S_i\right\}-p_e$$

for every $S_i>0$ and hence

(3) $$\frac{1}{N}\geqq\frac{1}{N}\sum_{i=1}^{N}\lambda(E_i)\geqq\frac{1}{N}\sum_{i=1}^{N}S_i^{-1}\left(p^i\left\{\frac{dp^i}{d\lambda}<S_i\right\}-p_e\right)$$
$$(S_i>0,\ (1\leqq i\leqq N)).$$

One uses these inequalities to upper estimate the maximal length of codes for fixed error with the time. Applying Chebyshev inequalities to them (for λ suitably chosen) one obtains even in nonstationary cases (see [3]) sharp bounds for the length of codes.

However, (3) estimates for fast decreasing error somewhat into the wrong direction and cannot be used to obtain tight lower error bounds.

The following change of (3) (see [2]) is an improvement for low rate codes:

Let $\{(p^i, E_i)\}_{1\leqq i\leqq N}$ be a code of M and let $p^i\ll\lambda$ $(1\leqq i\leqq N)$. Then for $0<S_1<S_2<\infty$ and $0<s<1$ holds

(4)
$$p^i({}^cE_i)\geqq p^i\left(\left\{S_1<\frac{dp^i}{d\lambda}<S_2\right\}\cap {}^cE_i\right)$$
$$\geqq S_1^s\int_{\left\{S_1<\frac{dp^i}{d\lambda}<S_2\right\}\cap {}^cE_i} dp^i\left(\frac{dp^i}{d\lambda}\right)^{-s}\quad\text{and}$$
$$\lambda(E_i)\geqq\lambda\left(\left\{S_1<\frac{dp^i}{d\lambda}<S_2\right\}\cap E_i\right)$$
$$\geqq S_2^{-(1-s)}\int_{\left\{S_1<\frac{dp^i}{d\lambda}<S_2\right\}\cap E_i} dp^i\left(\frac{dp^i}{d\lambda}\right)^{-s}.$$

Observe

$$0<\int dp^i\left(\frac{dp^i}{d\lambda}\right)^{-s}\leqq 1$$

and put

$$F_{(s)}^i:=\ln\int dp^i\left(\frac{dp^i}{d\lambda}\right)^{-s},\qquad dQ_{(s)}^i:=dp^i\left(\frac{dp^i}{d\lambda}\right)^{-s}\exp\left[-F_{(s)}^i\right].$$

$Q_{(s)}^i$ is a probability. (4) implies

Lemma 2.2. *Let* $\{(p^i, E_i)\}_{1\leqq i\leqq N}$ *be a code of* M, $p^i \ll \lambda$ $(1\leqq i\leqq N)$. *Then*

$$\exp[-F^i_{(s)} - s S_1]\, p^i({}^cE_i) + \exp[-F^i_{(s)} + (1-s) S_2]\, \lambda(E_i)$$
$$\geqq Q^i_{(s)}\left\{S_1 < \ln\frac{dp^i}{d\lambda} < S_2\right\} \qquad (-\infty < S_1 < S_2 < \infty)$$
$$(1\leqq i\leqq N).$$

Chebyshev inequalities applied to the right hand side of the last inequality and an averaging procedure (with λ suitably chosen) will finally give our lower error estimate.

§ 3. Properties of the Estimating Parameters

Definition 3.1. Let (X, F, M) be given, $M' = \{p^{(j)}\}_{1\leqq j\leqq n}$ (n arbitrary) any finite subset of M and λ a probability such that $p \ll \lambda$ for $p \in M'$. For $u \geqq 0$ let

$$H(u, M') := \inf\left\{\ln \int d\lambda \left[\sum_{j=1}^n a^{(j)} \left(\frac{dp^{(j)}}{d\lambda}\right)^{\frac{1}{1+u}}\right]^{1+u} : a^{(j)} \geqq 0,\ \sum_{j=1}^n a^{(j)} = 1\right\}$$

and

$$H(u, M) := \inf\{H(u, M') : M' \text{ finite} \subseteq M\}.$$

We already know from Lemma 2.1

$$P_e(M, N) \leqq \exp\left[\inf_{0\leqq u\leqq 1} (H(u, M) + u \ln N) + \ln 4\right].$$

Within the next paragraph we shall prove a similar lower estimate for $P_e(M, N)$ using $H(u, M)$.

Definition 3.2. Let (X, F, M) be given, $M' = \{p^{(j)}\}_{1\leqq j\leqq n}$ (n arbitrary) any finite subset of M.

Let

$$C(M') := \sup\left\{\sum_{j=1}^n a^{(j)} \int dp^{(j)} \ln \frac{dp^{(j)}}{d\left(\sum_{k=1}^n a^{(k)} p^{(k)}\right)} : a^{(j)} \geqq 0,\ \sum_{j=1}^n a^{(j)} = 1\right\}$$

and

$$C(M) := \sup\{C(M') : M' \text{ finite} \subseteq M\}.$$

(5) Remarks on $C(M)$:

a) For $p \ll \lambda$

$$\int d\lambda \frac{dp}{d\lambda} \ln \frac{dp}{d\lambda} = \int dp \ln \frac{dp}{d\lambda}$$

is well defined and

$$0 \leqq \int dp \ln \frac{dp}{d\lambda} \leqq \infty$$

(let $z(y)=y\ln y$. $z(y)\to 0(y\to 0)$, $z(1)=1$, $z(y)$ is convex.)

$$\int d\lambda\, z\left(\frac{dp}{d\lambda}\right)\geqq z\left(\int dp\right)=z(1)=0\Big).$$

b) $$0\leqq \int dp^{(j)}\ln\frac{dp^{(j)}}{d\left(\sum_{k=1}^{n}a^{(k)}p^{(k)}\right)}\leqq \sup\left\{\frac{1}{a^{(k)}}: a^{(k)}>0,\ 1\leqq k\leqq n\right\}$$

(with the notation in Definition 3.2).

c) $$\sum_{j=1}^{n}a^{(j)}\int dp^{(j)}\ln\frac{dp^{(j)}}{d\left(\sum_{k=1}^{n}a^{(k)}p^{(k)}\right)}\leqq \sum_{j=1}^{n}a^{(j)}\int dp^{(j)}\ln\frac{dp^{(j)}}{d\lambda}$$

for any λ such that $p^{(j)}\ll\lambda$ $(1\leqq j\leqq n)$. (Write

$$\sum_{j=1}^{n}a^{(j)}\int dp^{(j)}\ln\frac{dp^{(j)}}{d\left(\sum_{k=1}^{n}a^{(k)}p^{(k)}\right)}=\sum_{j=1}^{n}a^{(j)}\int dp^{(j)}\ln\frac{dp^{(j)}}{d\lambda}-\int d\lambda\, z\left(\sum a^{(k)}\frac{dp^{(k)}}{d\lambda}\right)$$

(z as defined in a)) and observe

$$\int d\lambda\, z\left(\sum a^{(k)}\frac{dp^{(k)}}{d\lambda}\right)\geqq 0.)$$

d) $$\sum_{j=1}^{n}a^{(j)}\int dp^{(j)}\ln\frac{dp^{(j)}}{d\left(\sum_{k=1}^{n}a^{(k)}p^{(k)}\right)}\leqq \ln n.$$

$\left(\text{Put } \lambda=\frac{1}{n}\sum_{k=1}^{n}p^{(k)} \text{ in c)}\right)$. Hence $0\leqq C(M')\leqq \ln n$ if $|M'|\leqq n$.

e) Furthermore, it follows from c) that $dp/d\lambda\leqq K$ ($\lambda-$ a.e.) for all $p\in M$ implies $0\leqq C(M)\leqq \ln K$.

(6) Let for $\{p^{(j)}\}_{1\leqq j\leqq n}$ and $\{a^{(j)}\}_{1\leqq j\leqq n}$ $(a^{(j)}\geqq 0, \sum a^{(j)}=1)$ fixed

$$\bar{H}(u):=\ln\int d\lambda\left[\sum_{j=1}^{n}a^{(j)}\left(\frac{dp^{(j)}}{d\lambda}\right)^{\frac{1}{1+u}}\right]^{1+u}\qquad (0\leqq u<\infty).$$

One obtains:

a) $$\lim_{0<u\to 0}\frac{d}{du}\bar{H}(u)=-\sum_{j=1}^{n}a^{(j)}\int dp^{(j)}\ln\frac{dp^{(j)}}{d\left(\sum_{k=1}^{n}a^{(k)}p^{(k)}\right)}.$$

b) $\bar{H}(u)$ is a non-positive monotonically decreasing continuous convex function of u.

Remark. a) Can easily be checked. Put for a simplification $\lambda = \sum_{k=1}^{n} a^{(k)} p^{(k)}$.

b) Should be checked by differentiation. Assume $a^{(j)} > 0$ $(1 \leqq j \leqq n)$. One verifies

$$\frac{d^2}{du^2} \bar{H}(u) = \int d\mu \left[\left(\sum a^{(j)} g^{(j)} \ln g^{(j)}\right)^2 + \frac{1}{1+u} \left(\sum a^{(j)} g^{(j)} (\ln g^{(j)} - \sum a^{(j)} g^{(j)} \ln g^{(j)})^2\right) \right] - \left[\int d\mu \left(\sum a^{(j)} g^{(j)} \ln g^{(j)}\right)\right]^2 \geqq 0$$

where

$$d\mu = \frac{d\lambda \left[\sum a^{(j)} \left(\frac{dp^{(j)}}{d\lambda}\right)^{\frac{1}{1+u}}\right]^{1+u}}{\int d\lambda \left[\sum a^{(j)} \left(\frac{dp^{(j)}}{d\lambda}\right)^{\frac{1}{1+u}}\right]^{1+u}}$$

and

$$g^{(j)} = \frac{\left(\frac{dp^{(j)}}{d\lambda}\right)^{\frac{1}{1+u}}}{\sum_{k=1}^{n} a^{(k)} \left(\frac{dp^{(k)}}{d\lambda}\right)^{\frac{1}{1+u}}}.$$

Observe

$$0 < \int d\lambda \left[\sum a^{(j)} \left(\frac{dp^{(j)}}{d\lambda}\right)^{\frac{1}{1+u}}\right]^{1+u} \leqq 1,$$

that μ is a probability and that $\sum a^{(j)} g^{(j)}(x) = 1$ for all $x \in X$. Hence we have expressed $\frac{d^2}{du^2} \bar{H}(u)$ as a sum of variances which proves the convexity of $\bar{H}(u)$. The continuity and monotonicity of $\bar{H}(u)$ is easily to be seen.

The next Lemma describes a few properties of $H(u, M)$:

Lemma 3.3. *For $H(u, M)$ holds:*

a) $0 \geqq H(u, M)$ $(u \geqq 0)$.

b) *$H(u, M)$ is monotonically decreasing.*

c) *If $H(u_0, M)$ is finite for some $u_0 > 0$ then $H(u, M)$ is a continuous function of u for $u > 0$.*

d) $-u\,C(M) \leqq H(u, M)$ $(u > 0)$ *and* $\liminf_{0 < u \to 0} \frac{1}{u} H(u, M) = -C(M)$.

e) *If $|M|$ is finite then $H(u, M)$ is finite and continuous for $u \geqq 0$.*

Proof. a) and b) is true because $H(u, M)$ is infimum of functions $\bar{H}(u)$ as in (6). For c) observe that $H(u_0, M) > -\infty$ for some $u_0 \geqq 0$ implies $H(u, M) > -\infty$ for all $u > 0$ using (6) b). The continuity of $H(u, M)$ follows from

$$0 \leqq H(u_1, M) - H(u_2, M) \leqq \frac{u_1 - u_2}{u_1} H(u, M) \qquad (0 < u_1 \leqq u_2).$$

d) follows again from (6) a) and (6) b). e) follows from c), d) and (5) d).

Remark. The condition $C(M) < \infty$ on M is considerably stronger than the requirement that $H(u, M) > -\infty$ holds for $u \geqq 0$. $H(u_0, M) = -\infty$ implies $H(u, M) = -\infty$ for all $u > 0$. If $H(u_0, M) = -\infty$ then M has arbitrary long finite codes with arbitrary small error according to Lemma 2.1. Let $N(p_e)$ denote the supremum over the lengths of all codes for M with error at most p_e. $H(u, M)$ is finite for $u \geqq 0$ at least if $N(p_e)$ is finite for some $p_e > 0$. If $C(M) < \infty$ then the estimate used for the weak converse of the coding theorem (see [4]) implies $\sup_{0 \leqq p_e < 1} (1 - p_e) \ln N(p_e) < \infty$.

One finds however easily sets M where the latter is not the case but where $N(p_e)$ is finite for all $p_e (0 \leqq p_e < 1)$.

We remark furthermore without proof (because we do not need it later on) that $H(u, M)$ is continuous at $u = 0$ iff $N(p_e)$ is finite for all $p_e (0 \leqq p_e < 1)$. The latter is equivalent to the condition that M is equi absolutely continuous with respect to some probability λ (see [3]).

Example of a function $H(u, M)$ which is finite and discontinuous at 0:

Let $X = \{0, 1, 2, \ldots\}$, F be the discrete σ-field and $M = \{p^{(j)}: j = 1, 2, \ldots\}$ where $p^{(j)}(0) = \frac{1}{2}$, $p^{(j)}(j) = \frac{1}{2}$. One calculates:

$$H(u, M) = \begin{cases} 0 & \text{for } u = 0 \\ \ln \frac{1}{2} & \text{for } u > 0. \end{cases}$$

The described properties of $H(u, M)$ are mainly interesting for the upper error estimate. The next part of this paragraph prepares the connections between upper and lower estimate.

Let $M' = \{p^{(j)}\}_{1 \leqq j \leqq n}$ on (X, F) be given. Consider for fixed $u \geqq 0$

$$K_u(a^{(1)}, \ldots, a^{(n)}) := \int d\lambda \left[\sum_{j=1}^{n} a^{(j)} \left(\frac{dp^{(j)}}{d\lambda} \right)^{\frac{1}{1+u}} \right]^{1+u}.$$

$K_u(\cdot)$ is a (continuous) convex function of the simplex

$$S = \left\{ (a^{(1)}, \ldots, a^{(n)}): a^{(j)} \geqq 0, \sum_{j=1}^{n} a^{(j)} = 1 \right\}$$

because of the convexity of the function y^{1+u}. (However $\ln K_u(\cdot)$ is generally not convex and $H(u, M)$ is generally not convex either.)

Lemma 3.4. *Let $M'=\{p^{(j)}\}_{1\leqq j\leqq n}$ on (X,F) be given and $u>0$ fixed. Then there is exactly one measure $\mu_{(u)}(\cdot)$ on (X,F) for M' which can be represented in the form*

$$d\mu_{(u)}=d\lambda\left[\sum_{j=1}^{n}a^{(j)}\left(\frac{dp^{(j)}}{d\lambda}\right)^{\frac{1}{1+u}}\right]^{1+u}\qquad\left(\text{where } a^{(j)}\geqq 0,\ \sum_{j=1}^{n}a^{(j)}=1\right)$$

such that $\int d\mu_{(u)}=\exp[H(u,M')]$ holds.

Proof. There is such a measure $\mu_{(u)}(\cdot)$ because $K_u(\cdot)$ depends continuously on the $a^{(j)}$. Assume that there are two different measures μ^1 and μ^2 with the above properties. Then

$$d\mu^3:=d\lambda\left[\frac{1}{2}\left(\frac{du^1}{d\lambda}\right)^{\frac{1}{1+u}}+\frac{1}{2}\left(\frac{d\mu^2}{d\lambda}\right)^{\frac{1}{1+u}}\right]^{1+u}$$

has a representation in the above form but

$$\int d\mu^3<\tfrac{1}{2}\int d\mu^1+\tfrac{1}{2}\int d\mu^2(=\exp[H(u,M')])$$

because of the strict convexity of y^{1+u} for fixed $u>0$.

Remark. If

$$\int d\lambda\left[\sum_{j=1}^{n}a^{(j)}\left(\frac{dp^{(j)}}{d\lambda}\right)^{\frac{1}{1+u}}\right]^{1+u}=\exp[H(u,M')]$$

then $a^{(j)}=0$ in the integral for all $p^{(j)}$ which are non extremal points of the convex hull of M'. The minimizing $a^{(j)}$ are, however, i.g. not uniquely determined. Lemma 3.4 admits often easier calculations of $H(u,M')$ if M' satisfies certain symmetry conditions.

Lemma 3.5. *Let $M'=\{p^{(j)}\}_{1\leqq j\leqq n}$ on (X,F) be given and s fixed $(0<s<1)$. Then there is exactly one probability q for M' and s such that*

$$(0\geqq)\ln\int d\lambda\left(\frac{dp}{d\lambda}\right)^{1-s}\left(\frac{dq}{d\lambda}\right)^{s}\geqq(1-s)\,H\left(\frac{s}{1-s},M'\right)\qquad(>-\infty)$$

holds for all $p\in M'$. (For λ any probability may be taken with $p\ll\lambda$ for all $p\in M'$ and $q\ll\lambda$. Observe again that the integral is independent of λ.)

Proof. 1. *Existence of q:* Put $u=\frac{s}{1-s}$ and let $K_u(\cdot)$ be defined as above and assume that $K_{\frac{s}{1-s}}(a^{(1)},\ldots,a^{(n)})$ is minimal. One obtains together with $\sum a^{(j)}=1$ at the minimum (putting $a^{(j_2)}=1-\sum_{j\neq j_2}a^{(j)}$ and keeping $a^{(j)}$ fixed for $j\neq j_1,j_2$):

$$\frac{\partial}{\partial a^{(j_1)}}K_{\frac{s}{1-s}}(a^{(1)},\ldots,a^{(n)})=\frac{\partial}{\partial a^{(j_1)}}\int d\lambda\left[\sum_{j=1}^{n}a^{(j)}\left(\frac{dp^{(j)}}{d\lambda}\right)^{1-s}\right]^{\frac{s}{1-s}}$$

$$=\frac{1}{1-s}\int d\lambda\left[\sum_{j=1}^{n}a^{(j)}\left(\frac{dp^{(j)}}{d\lambda}\right)^{1-s}\right]^{\frac{1}{1-s}-1}\left[\left(\frac{dp^{(j_1)}}{d\lambda}\right)^{1-s}-\left(\frac{dp^{(j_2)}}{d\lambda}\right)^{1-s}\right]\geqq 0$$

(for $j_2\neq j_1$).

Hence

$$\int d\lambda \left(\frac{dp^{(k)}}{d\lambda}\right)^{1-s} \left[\sum_{j=1}^{n} a^{(j)} \left(\frac{dp^{(j)}}{d\lambda}\right)^{1-s}\right]^{\frac{1}{1-s}}$$

$$\geqq \int d\lambda \left[\sum_{j=1}^{n} a^{(j)} \left(\frac{dp^{(j)}}{d\lambda}\right)^{1-s}\right] \left[\sum_{j=1}^{n} a^{(j)} \left(\frac{dp^{(j)}}{d\lambda}\right)^{1-s}\right]^{\frac{1}{1-s}}$$

$$= \int d\lambda \left[\sum_{j=1}^{n} a^{(j)} \left(\frac{d\lambda}{d\lambda}\right)^{1-s}\right]^{\frac{1}{1-s}} \qquad \text{for } 1 \leqq k \leqq n.$$

Put

$$dq = d\lambda \left[\sum_{j=1}^{n} a^{(j)} \left(\frac{dp^{(j)}}{d\lambda}\right)^{1-s}\right]^{\frac{1}{1-s}} \exp\left[-H\left(\frac{s}{1-s}, M'\right)\right].$$

(q is a probability.)

The last inequality gives

$$\int d\lambda \left(\frac{dp^{(k)}}{d\lambda}\right)^{1-s} \left(\frac{dq}{d\lambda}\right)^{s} \geqq \int d\lambda \left[\sum_{j=1}^{n} a^{(j)} \left(\frac{dp^{(j)}}{d\lambda}\right)^{1-s}\right]^{\frac{1}{1-s}} \exp\left[-s H\left(\frac{s}{1-s}, M'\right)\right]$$

$$= \exp\left[(1-s) H\left(\frac{s}{1-s}, M'\right)\right] \qquad (1 \leqq k \leqq n).$$

2. *Uniqueness of* q. Consider a probability μ which is not equal to the above q and let the $a^{(j)}$ be the same minimizing coefficients as before.

$$\sum_{j=1}^{n} a^{(j)} \int d\lambda \left(\frac{dp^{(j)}}{d\lambda}\right)^{1-s} \left(\frac{d\mu}{d\lambda}\right)^{s} = \exp\left[(1-s) H\left(\frac{s}{1-s}, M'\right)\right] \int d\lambda \left(\frac{dq}{d\lambda}\right)^{1-s} \left(\frac{d\mu}{d\lambda}\right)^{s}$$

and

$$\int d\lambda \left(\frac{dq}{d\lambda}\right)^{1-s} \left(\frac{d\mu}{d\lambda}\right)^{s} = \int_E dq \left(\frac{\left(\frac{d\mu}{d\lambda}\right)}{\left(\frac{dq}{d\lambda}\right)}\right)^{s}$$

where

$$E = \left\{x \in X : \inf\left(\left(\frac{d\mu}{d\lambda}\right), \left(\frac{dq}{d\lambda}\right)\right) > 0\right\}.$$

$\int_E dq \left(\frac{\left(\frac{d\mu}{d\lambda}\right)}{\left(\frac{dq}{d\lambda}\right)}\right)^{s} < 1$ because $\mu \neq q$ and the strict concavity of the function y^s.

Hence for every $\mu \neq q$ there is at least one $p \in M'$ such that

$$\int d\lambda \left(\frac{dp}{d\lambda}\right)^{1-s} \left(\frac{d\mu}{d\lambda}\right)^{s} < \exp\left[(1-s) H\left(\frac{s}{1-s}, M'\right)\right].$$

Lemma 3.6. a) *Let M_1', M_2' be finite and s fixed* $(0<s<1)$. *The uniquely determined probability q on $X_1 \times X_2$ such that*

$$\int d\lambda \left(\frac{dp}{d\lambda}\right)^{1-s} \left(\frac{dq}{d\lambda}\right)^{s} \geqq \exp\left[(1-s)\, H\left(\frac{s}{1-s}, M_1' \times M_2'\right)\right] \qquad \text{for all } p \in M_1' \times M_2'$$

holds is a product probability $q=q_1 \times q_2$ with

$$\int d\lambda \left(\frac{dp_v}{d\lambda_v}\right)^{1-s} \left(\frac{dq_v}{d\lambda_v}\right)^{s} \geqq \exp\left[(1-s)\, H\left(\frac{s}{1-s}, M_v'\right)\right] \qquad \text{for all } p_v \in M_v' \ (v=1,2).$$

b) *For arbitrary M_1, M_2 holds*

$$H\left(\frac{s}{1-s}, M_1\right) + H\left(\frac{s}{1-s}, M_2\right) = H\left(\frac{s}{1-s}, M_0 \times M_2\right) \qquad (0 \leqq s < 1).$$

Proof. Let $M_1' = \{p_1^{(j)}\}_{1 \leqq j \leqq n}$, $M_2' = \{p_2^{(k)}\}_{1 \leqq k \leqq m}$ $a^{(j)}$ and $b^{(k)}$ may denote non-negative coefficients with $\sum_{j=1}^{n} a^{(j)} = 1$, $\sum_{k=1}^{m} b^{(k)} = 1$.

$$\exp\left[H\left(\frac{s}{1-s}, M_1' \times M_2'\right)\right] \leqq \int d\lambda_1 \times d\lambda_2 \left[\sum_{j=1}^{n} \sum_{k=1}^{m} a^{(j)}\, b^{(k)} \left(\frac{dp_1^{(j)}}{d\lambda_1} \frac{dp_2^{(k)}}{d\lambda_2}\right)^{1-s}\right]^{\frac{1}{1-s}}$$

$$= \int d\lambda_1 \left[\sum_{j=1}^{n} a^{(j)} \left(\frac{dp_1^{(j)}}{d\lambda_1}\right)^{1-s}\right]^{\frac{1}{1-s}} \cdot \int d\lambda_2 \left[\sum_{k=1}^{m} b^{(k)} \left(\frac{dp_2^{(k)}}{d\lambda_2}\right)^{1-s}\right]^{\frac{1}{1-s}}.$$

Minimization of the product on the right hand side gives

$$H\left(\frac{s}{1-s}, M_1' \times M_2'\right) \leqq H\left(\frac{s}{1-s}, M_1'\right) + H\left(\frac{s}{1-s}, M_2'\right) \qquad (0 \leqq s < 1).$$

Now let $0<s<1$. Furthermore, let q_1 and q_2, respectively, be the probabilities such that

$$\int d\lambda_v \left(\frac{dp_v}{d\lambda_v}\right)^{1-s} \left(\frac{dq_v}{d\lambda_v}\right)^{s} \geqq \exp\left[(1-s)\, H\left(\frac{s}{1-s}, M_v'\right)\right]$$

for all $p_v \in M_v'$ $(v=1,2)$.

For all $p \in M_1' \times M_2'$ $(p = p_1 \times p_2)$ holds

$$\int d\lambda_1 \times d\lambda_2 \left(\frac{dp_1 \times dp_2}{d\lambda_1 \times d\lambda_2}\right)^{1-s} \left(\frac{dq_1 \times dq_2}{d\lambda_1 \times d\lambda_2}\right)^{s}$$

$$= \int d\lambda_1 \left(\frac{dp_1}{d\lambda_1}\right)^{1-s} \left(\frac{dq_1}{d\lambda_1}\right)^{s} \cdot \int d\lambda_2 \left(\frac{dp_2}{d\lambda_2}\right)^{1-s} \left(\frac{dq_2}{d\lambda_2}\right)^{s}$$

$$\geqq \exp\left[(1-s)\left(H\left(\frac{s}{1-s}, M_1'\right) + H\left(\frac{s}{1-s}, M_2'\right)\right)\right].$$

But

$$\inf_{p\in M_1'\times M_2'}\int d\lambda_1\times d\lambda_2\left(\frac{dp}{d\lambda_1\times d\lambda_2}\right)^{1-s}\left(\frac{dq_1\times dq_2}{d\lambda_1\times d\lambda_2}\right)^{s}$$
$$\leqq \exp\left[(1-s)\,H\left(\frac{s}{1-s}, M_1'\times M_2'\right)\right]$$

according to Lemma 3.5, and the remaining part of a) follows also from Lemma 3.5.

$$H\left(\frac{s}{1-s}, M_1'\times M_2'\right)=H\left(\frac{s}{1-s}, M_1'\right)+H\left(\frac{s}{1-s}, M_2'\right)$$

for all finite M_1', M_2' implies *b*).

Remark. One may prove similarly that $C(M_1\times M_2)=C(M_1)+C(M_2)$ holds using (5) c). $\sum_{v=1}^{t} C(M_v)=C(M_{[1,t]})$ replaces $t\cdot C$ (C the capacity) if one goes over from stationary channels without memory to non stationary channels without memory (see [3]).

We remark the following for completeness:

Lemma 3.7. *Let M be any set of probabilities on (X,F), s fixed $(0<s<1)$ and $H\left(\frac{s}{1-s}, M\right)>-\infty$. Then there is exactly one probability q with*

$$\ln\int d\lambda\left(\frac{dp}{d\lambda}\right)^{1-s}\left(\frac{dq}{d\lambda}\right)^{s}\geqq(1-s)\,H\left(\frac{s}{1-s}, M\right)\qquad \textit{for all } p\in M$$

(λ chosen with respect to p, q).

The reader may prove this by showing the following steps which may be derived with the previous lemmas. Let q' for M' finite $(\subseteq M)$ and q'' for M'' finite $(\subseteq M)$ be the optimizing probabilities in the sense of Lemma 3.5. There is a constant $K_{(s)}$ depending only on s

$$\text{s.t.}\ \ \|q'-q''\|\leqq K_{(s)}\left[\frac{1}{2}H\left(\frac{s}{1-s}, M'\right)+\frac{1}{2}H\left(\frac{s}{1-s}, M''\right)-H\left(\frac{s}{1-s}, M\right)\right]^{\frac{1}{2}}$$

($\|\cdot\|$ the totally variation norm).

Hence with

$$H\left(\frac{s}{1-s}, M'\right)\to H\left(\frac{s}{1-s}, M\right)$$

(q' for M') one has $\|q'-q\|\to 0$ for exactly one q.

One shows that exactly this q has the property

$$\ln\int d\lambda\left(\frac{dp}{d\lambda}\right)^{1-s}\left(\frac{dq}{d\lambda}\right)^{s}\geqq(1-s)\,H\left(\frac{s}{1-s}, M\right)\qquad \text{for all } p\in M$$

(λ chosen with respect to p, q) using an uniform absolute continuity argument for sequences $q'^k \to q$ (keeping p fixed) and using again the strict concavity of y^s.

§ 4. Estimates for $P_e(M_{[1,t]}, N)$

The Upper Error Bound

From Lemma 2.1 and from the last paragraph follows the theorem:

Theorem 4.1.

$$P_e(M_{[1,t]}, N) \leqq \exp\left[\inf_{0\leqq u\leqq 1}(H(u, M_{[1,t]}) + u \ln N) + \ln 4\right]$$

$$= \exp\left[\inf_{0\leqq u\leqq 1}\left(\sum_{v=1}^{t} H(u, M_v) + u \ln N\right) + \ln 4\right].$$

Corollary. *Let a sequence* $\{M_v\}$ $v=1, 2, \ldots$ *be given, where the* M_v *are copies of* M_1 *and* $C(M_1)>0$. *Let* $N=N_t\leqq \exp[tR]$ *where* R *is fixed and* $0\leqq R< C(M_1)$.

Then for $M_{[1,t]}=M_1\times\cdots\times M_t$ *holds*

$$P_e(M_{[1,t]}, N_t) \leqq \exp\left[t\cdot \inf_{0\leqq u\leqq 1}(H(u, M_1) + uR) + \ln 4\right]$$

and the right hand side of the inequality decreases exponentially with t.

(Observe with Lemma 3.3 that $\inf\,(H(u, M_1)+uR)<0$ for $0\leqq R<C(M_1)$.)

If $C(M_1)=\infty$ then the corollary particularly implies that $N_t(p_e)$, the supremum length of codes (see § 3) for fixed error $p_e (0<p_e<1)$ for the time $[1, t]$, grows either faster than exponentially with t or $N_t(p_e)=\infty$ for $t>t_0$. (Both cases are possible.)

Corollary. *Let a sequence* $\{(M_v, R_v)\}$ $v=1, 2\ldots$ *be given, every* (M_v, R_v) *a copy of an element of the finite collection* $\{(M_{(j)}, R_{(j)})\}_{1\leqq j\leqq n}$ *where* $M_{(j)}$ *is a set of probabilities on* $(X_{(j)}, F_{(j)})$ *with* $C(M_{(j)})>0$ *and where* $R_{(j)}$ *is a positive number,* $R_{(j)}< C(M_{(j)})$.

If $\ln N_t \leqq \sum_{v=1}^{t} R_v$ *then*

$$P_e(M_{[1,t]}, N_t) \leqq \exp\left[\inf_{0\leqq u\leqq 1}\left(H(u, M_{[1,t]}) + u\sum_{v=1}^{t} R_v\right)\right]$$

and the right hand side of this inequality decreases exponentially with t.

The Lower Error Estimate

We partly follow the idea of the proof in [2]: The use of the intermediate value theorem for continuous functions will be similar as in [2]. However, we cannot use the method of fixed composition codes, a method which is rather unstable against non-stationary changes of the channel with the time and which allows no interesting generalizations of the proof in [2] to infinite alphabet channels. Instead of the method of fixed composition codes an averaging procedure for the probabilities $q=q_{(s)}$ (see Lemma 3.5) will allow to obtain estimates as in [2] uniformly enough for large parts of the given codes provided the following is fulfilled:

(7) *Condition.* For every $\varepsilon>0$ there is $K>0$ s.t. $u>K$ implies

$$\left|\frac{1}{u} H(u, M_{[1,t]}) - \lim_{u\to\infty} \frac{1}{u} H(u, M_{[1,t]})\right| < \varepsilon t$$

uniformly for all t $(t=1, 2, \ldots)$.

At least most of the interesting channels without memory have this property: a) For example the condition is for the stationary channel nothing but the existence of $\lim_{u\to\infty} \frac{1}{u} H(u, M_1)$. But this limit exists because $\frac{1}{u} H(u, M_1)$ increases monotonically with u. b) Another example may be: Let $\{M_v\}_{v=1,2,\ldots}$ be given where for every M_v there is a measure λ_v with $\|\lambda_v\|>b$ (for some $b>0$) s.t. $p_v \geqq \lambda_v$ for all $p_v \in M_v$ $(v=1, 2, \ldots)$.

Then $H(u, M_v) > \ln b$ and $\left|\frac{1}{u} H(u, M_{[1,t]}) - 0\right| < \frac{t}{u} \ln \frac{1}{b}$.

At first, we need as a generalization of the procedure in the annex of [2] the following

Lemma 4.2. *Let* $M' = \{p^{(j)}\}_{1\leqq j\leqq n}$ *on* (X, F) *be given and* $\lambda := \frac{1}{n} \sum_{j=1}^{n} p^{(j)}$. *Let for* $s (0<s<1)$ $\mu(s)$ *be the unique measure satisfying*

$$\int d\mu_{(s)} = \exp\left[H\left(\frac{s}{1-s}, M'\right)\right] \quad \textit{and} \quad \int d\lambda \left(\frac{dp}{d\lambda}\right)^{1-s} \left(\frac{d\mu_{(s)}}{d\lambda}\right)^{s} \geqq \exp\left[H\left(\frac{s}{1-s}, M'\right)\right]$$

for all $p\in M'$).

Then for every $s_0 (0<s_0<1)$ *holds:* $\lim_{s\to s_0} \|\mu_{(s)} - \mu_{(s_0)}\| = 0$ *(where* $\|\cdot\|$ *is the total variation norm).*

Proof. Let

$$T_{(s)} := \left\{ f_{(s)} = \sum_{j=1}^{n} a^{(j)} \left(\frac{dp^{(j)}}{d\lambda}\right)^{1-s} : a^{(j)} \geqq 0, \sum_{j=1}^{n} a^{(j)} = 1 \right\}$$

and

$$dT_{(s)} := \left\{ r_{(s)} : dr_{(s)} = d\lambda\, f_{(s)}^{\frac{1}{1-s}}, f_{(s)} \in T_{(s)} \right\}.$$

For every $\varepsilon>0$ there is $b>0$ s.t. $|s-s_0|<b$ implies for every $f_{(s)} \in T_{(s)}$

$$\inf_{f_{(s_0)} \in T_{(s_0)}} \left\{ \int d\lambda\, |f_{(s)} - f_{(s_0)}| \right\} < \varepsilon.$$

$T_{(s_0)}$ is L^1_λ-norm compact. Every sequence $f_{(s_k)}$ $(s_k \to s_0)$ has a L^1_λ-norm clusterpoint in $T_{(s_0)}$.

The densities $f_{(s)}$ are uniformly bounded for an interval around s_0. Hence, for $\mu_{(s_k)}$ $(s_k \to s_0)$ holds $\left(d\mu_{(s_k)} = d\lambda\, f_{(s_k)}^{\frac{1}{1-s_k}}\right.$ for some $\left. f_{(s_k)}\right)$:

$\mu_{(s_k)}$ has a norm clusterpoint $\tilde{\mu}_{(s_0)}$ in $dT_{(s_0)}$.

Because

$$\int d\lambda \left(\frac{dp}{d\lambda}\right)^{1-s} \left(\frac{dr}{d\lambda}\right)^{s} = \int d(\tfrac{1}{2}\lambda+\tfrac{1}{2}r) \left(\frac{dp}{d(\frac{1}{2}\lambda+\frac{1}{2}r)}\right)^{1-s} \left(\frac{dr}{d(\frac{1}{2}\lambda+\frac{1}{2}r)}\right)^{s}$$

depends norm-continuously on changes of r and continuously on s, one obtains for $\tilde{\mu}_{(s_0)}$ (observing that H is continuous in s):

$$\int d\tilde{\mu}_{(s_0)} = \exp\left[H\left(\frac{s_0}{1-s_0}, M'\right)\right]$$

and

$$\int d\lambda \left(\frac{dp}{d\lambda}\right)^{1-s_0} \left(\frac{d\mu(s_0)^{s_0}}{d\lambda}\right) \geqq \exp\left[H\left(\frac{s_0}{1-s_0}, M'\right)\right].$$

This implies $\tilde{\mu}_{(s_0)} = \mu_{(s_0)}$ because of Lemma 3.5.

Remark. One can formulate Lemma 4.2 with M arbitrary instead of M' finite using for the proof the proof of Lemma 4.2 and the remark following Lemma 3.7. We do not need this general result for the construction below.

Lemma 4.3. *Let be given* $M' = \{p^{(j)}\}_{1 \leqq j \leqq n}$ *on* (X, F) *and* $\mu_{(s)}$ $(0<s<1)$ *as before. Then the probabilities*

$$q_{(s)} \qquad \textit{where } dq_{(s)} = d\mu_{(s)} \exp\left[-H\left(\frac{s}{1-s}, M'\right)\right]$$

and

$$\tilde{q}_{(s)} \qquad \textit{where } \tilde{q}_{(s)} := t \int_{-st^{-1}}^{(1-s)t^{-1}} dy\, q_{(s+y)}$$

satisfy

$$\lim_{s\to s_0} \|q_{(s)} - q_{(s_0)}\| = \lim_{s\to s_0} \|\tilde{q}_{(s)} - \tilde{q}_{(s_0)}\| = 0.$$

Remark. M' is finite and therefore $H\left(\frac{s}{1-s}, M'\right)$ finite and continuous in s.

We can write $\int dy\, q_{(s+y)}$ because of Lemma 4.2.

Lemma 4.4. *Let* $M' = \{p^{(j)}\}_{1 \leqq j \leqq n}$ *on* (X, F) *be given. Then for* $0<s<1$ *holds for all* $p \in M'$*:*

a) $$\ln \int d\lambda \left(\frac{dp}{d\lambda}\right)^{1-s} \left(\frac{d\tilde{q}_{(s)}}{d\lambda}\right)^{s}$$
$$\geqq \frac{t}{t-1}\left[s(1-s')\, H\left(\frac{s'}{1-s'}, M'\right) + (1-s)(1-s'')\, H\left(\frac{s''}{1-s''}, M'\right)\right]$$

where $s' = s(1-t^{-1})$, $s'' = s + (1-s)\, t^{-1}$,

b) $$\ln \int d\lambda \left(\frac{dp}{d\lambda}\right)^{1-s} \left(\frac{d\tilde{q}_{(s)}}{d\lambda}\right)^{s} \geqq \frac{t}{t-1} \inf_{0 \leqq u} \frac{1}{1+u} H(u, M').$$

Proof.

$$\ln \int d\lambda \left(\frac{dp}{d\lambda}\right)^{1-s} \left(\frac{d\tilde{q}_{(s)}}{d\lambda}\right)^{s} \geqq t \int_{-st^{-1}}^{(1-s)t^{-1}} dy \ln \int d\lambda \left(\frac{dp}{d\lambda}\right)^{1-s} \left(\frac{dq_{(s+y)}}{d\lambda}\right)^{s}$$

(Jensen's inequality).

i) For $0 \leqq y \leqq (1-s)\,t^{-1}$ holds

$$\begin{aligned}\ln \int d\lambda \left(\frac{dp}{d\lambda}\right)^{1-s} \left(\frac{dq_{(s+y)}}{d\lambda}\right)^{s} &= \ln \int dq_{(s+y)} \left(\frac{dp}{d\lambda}\right)^{1-s} \left(\frac{dq_{(s+y)}}{d\lambda}\right)^{-(1-s)} \\ &= \ln \int dq_{(s+y)} \left[\left(\frac{dp}{d\lambda}\right)^{1-s-y} \left(\frac{dq_{(s+y)}}{d\lambda}\right)^{-(1-s-y)}\right]^{\frac{1-s}{1-s-y}} \\ &\geqq (1-s)\frac{1}{1-s-y} \ln \int d\lambda \left(\frac{dp}{d\lambda}\right)^{1-s-y} \left(\frac{dq_{(s+y)}}{d\lambda}\right)^{s+y}.\end{aligned}$$

With Lemma 3.5 follows that this is

$$\geqq (1-s)\, H\left(\frac{s+y}{1-s-y}, M'\right) \geqq (1-s)\, H\left(\frac{s''}{1-s''}, M'\right) = \frac{t}{t-1}(1-s'')\, H\left(\frac{s''}{1-s''}, M'\right).$$

Thus

$$t \int_{0}^{(1-s)t^{-1}} dy \int \ln d\lambda \left(\frac{dp}{d\lambda}\right)^{1-s} \left(\frac{dq_{(s+y)}}{d\lambda}\right)^{s} \geqq (1-s)\frac{t}{t-1}(1-s'')\, H\left(\frac{s''}{1-s''}, M'\right).$$

ii) For $-st^{-1} \leqq y \leqq 0$ holds

$$\begin{aligned}&\ln \int d\lambda \left(\frac{dp}{d\lambda}\right)^{1-s} \left(\frac{dq_{(s+y)}}{d\lambda}\right)^{s} \\ &\quad= \ln \int dp \left[\left(\frac{dp}{d\lambda}\right)^{-(s+y)} \left(\frac{dq_{(s+y)}}{d\lambda}\right)^{s+y}\right]^{\frac{s}{s+y}} \\ &\quad\geqq \frac{s}{s+y} \ln \int d\lambda \left(\frac{dp}{d\lambda}\right)^{1-s-y} \cdot \left(\frac{dq_{(s+y)}}{d\lambda}\right)^{s+y} \\ &\quad= s\cdot\frac{1-s-y}{s+y}\,\frac{1}{1-s-y} \ln \int d\lambda \left(\frac{dp}{d\lambda}\right)^{1-s-y} \left(\frac{dq_{(s+y)}}{d\lambda}\right)^{s+y}.\end{aligned}$$

With Lemma 3.5 follows that this is

$$\geqq s\,\frac{1-s-y}{s+y}\, H\left(\frac{s+y}{1-s-y}, M'\right) \geqq s\,\frac{1-s'}{s'}\, H\left(\frac{s'}{1-s'}, M'\right).$$

The last inequality sign holds because $\frac{1}{u} H(u, M')$ increases with u. Therefore

$$t \int_{-st^{-1}}^{0} dy \int \ln d\lambda \left(\frac{dp}{d\lambda}\right)^{1-s} \left(\frac{dq_{(s+y)}}{d\lambda}\right)^{s} \geqq s\,\frac{t}{t-1}(1-s')\, H\left(\frac{s'}{1-s'}, M'\right).$$

Thus we obtain part a) of the lemma from i) and ii).

Setting $u=\frac{s'}{1-s'}$ and $u=\frac{s''}{1-s''}$, respectively, we obtain part b).

Now, let any code $\{(p^i, E_i)\}_{1\leq i\leq N}$ of length N for $M_{[1,t]}=M_1\times\cdots\times M_t$ be given and let M'_v denote the smallest subset of M_v containing the sequence $\{p^i_v\}_{1\leq i\leq N}$ of v-components of the sequence $\{p^i\}_{1\leq i\leq N}$ $(1\leq v\leq t)$.

Let $q_{(s)}$ be the optimizing probability of $M'_{[1,t]}=M'_1\times\cdots\times M'_t$ in the sense of Lemma 3.5 for every s $(0<s<1)$. $q_{(s)}=q_1(s)\times\cdots\times q_t(s)$, where $q_v(s)$ is optimizing for M'_v (see Lemma 3.6). Furthermore, let

$$\tilde{q}_v(s):=t\int_{-st^{-1}}^{(1-s)t^{-1}} dy\, q_v(s+y) \quad \text{and} \quad \tilde{q}_{(s)}:=\tilde{q}_1(s)\times\cdots\times\tilde{q}_t(s).$$

(8) *Define*

$$F^i_v(s):=\ln\int dp^i_v\left(\frac{dp^i_v}{d\tilde{q}_v(s)}\right)^{-s},$$

$$dQ^i_v(s):=dp^i_v\left(\frac{dp^i_v}{d\tilde{q}_v(s)}\right)^{-s}\exp[-F^i_v(s)],$$

$$m^i_v(s):=\int dQ^i_v(s)\ln\frac{dp^i_v}{d\tilde{q}_v(s)},$$

$$g^i_v(s):=\left[\int dQ^i_v(s)\left(\ln\frac{dp^i_v}{d\tilde{q}_v(s)}-m^i_v(s)\right)^2\right]^{\frac{1}{2}},$$

and $F^i_{(s)}$, $dQ^i_{(s)}$, $m^i_{(s)}$, $g^i_{(s)}$ (with respect to $\tilde{q}_{(s)}$) analogue, dropping the lower index v everywhere.

Finally let

$$F_{(s)}:=\inf_{1\leq i\leq N}F^i_{(s)},\qquad F_v(s):=\inf_{1\leq i\leq N}F^i_v(s)\qquad(1\leq v\leq t).$$

These expressions make sense for the estimates with the convention:

(9) *Convention*

$$\frac{dp^i}{d\tilde{q}_{(s)}}:=\begin{cases}\left(\dfrac{dp^i}{d\lambda}\right)\cdot\left(\dfrac{d\tilde{q}_{(s)}}{d\lambda}\right)^{-1} & \text{if the Radon-Nikodym derivatives } \dfrac{dp^i}{d\lambda}, \\ & \dfrac{d\tilde{q}_{(s)}}{d\lambda} \text{ are defined and } >0,\\ 0 & \text{otherwise.}\end{cases}$$

$dp^i\left(\frac{dp^i}{d\tilde{q}_{(s)}}\right)^{-s}$ is a substitute for $d\lambda\left(\frac{dp^i}{d\lambda}\right)^{1-s}\left(\frac{dq_{(s)}}{d\lambda}\right)^{s}$ and the three integrals $F^i_{(s)}$, $m^i_{(s)}$, $g^i_{(s)}$ depend only on a support of $p^i\wedge\tilde{q}_{(s)}$.

$F^i_{(s)}$ is finite because of Lemma 4.4 and because $M'_{[1,t]}$ is finite.

Furthermore, the integrals $m^i_{(s)}$, $g^i_{(s)}$ are well defined and finite: Consider the integrals

$$\int dp^i \left(\frac{dp^i}{d\tilde{q}_{(s)}}\right)^{-s} \ln^k \frac{dp^i}{d\tilde{q}_{(s)}} = \int d\tilde{q}_{(s)} \left(\frac{dp^i}{d\tilde{q}_{(s)}}\right)^{1-s} \ln^k \frac{dp^i}{d\tilde{q}_{(s)}} \qquad (k=1,2).$$

The functions $y^{1-s}\ln^k y$ $(k=1,2)$ satisfy $y^{1-s}\ln^k y \to 0$ $(y\to 0,\ y\to\infty)$.

Lemma 4.5. *$F^i_v(s)$ and $m^i_v(s)$ are continuous functions of s.*

This is a consequence of the proof of Lemma 4.2 and of Lemma 4.3.

Lemma 4.6.

$$(g^i_v(s))^2 \leqq \frac{1}{s^2(1-s)^2} \exp[-F^i_v(s)].$$

Proof.

$$\begin{aligned}(g^i_v(s))^2 &\leqq \int dQ^i_v(s) \ln^2 \frac{dp^i_v}{d\tilde{q}_v(s)}\\ &= \exp[-F^i_v(s)] \left(\frac{1}{(1-s)^2} \int_{\left\{\frac{dp^i_v}{d\tilde{q}_v(s)} \leqq 1\right\}} d\tilde{q}_v(s) \left(\frac{dp^i_v}{d\tilde{q}_v(s)}\right)^{1-s} \ln^2 \left(\frac{dp^i_v}{d\tilde{q}_v(s)}\right)^{1-s}\right.\\ &\quad \left. + \frac{1}{s^2} \int_{\left\{\frac{d\tilde{q}_v(s)}{dp^i_v} < 1\right\}} dp^i_v \left(\frac{d\tilde{q}_v(s)}{dp^i_v}\right)^s \ln^2 \left(\frac{d\tilde{q}_v(s)}{dp^i_v}\right)^s\right).\end{aligned}$$

Observing $x\ln^2 x \leqq 4e^{-2} \leqq 1$ for $0\leqq x\leqq 1$, one obtains

$$(g^i_v(s))^2 \leqq \exp[-F^i_v(s)] \left(\frac{1}{s^2} + \frac{1}{(1-s)^2}\right).$$

Finally,

$$\frac{1}{s^2} + \frac{1}{(1-s)^2} \leqq \frac{1}{s^2(1-s)^2}.$$

We have from Lemma 4.6 and (8):

$$V(s) := \frac{1}{s(1-s)} \left(\sum_{v=1}^{t} \exp[-F_v(s)]\right)^{\frac{1}{2}} \geqq g^i_{(s)} \qquad (1\leqq i\leqq N). \tag{10}$$

Next we estimate $m^i_{(s)}$:

$$dp^i \geqq dQ^i_{(s)} \exp\left[F^i_{(s)} + s \ln \frac{dp^i}{d\tilde{q}_{(s)}}\right]$$

and

$$d\tilde{q}_{(s)} \geqq dQ^i_{(s)} \exp\left[F^i_{(s)} - (1-s) \ln \frac{dp^i}{d\tilde{q}_{(s)}}\right]$$

implies with Jensen's inequality:

$$0 \geqq F^i_{(s)} + s\, m^i_{(s)}$$

and

$$0 \geqq F^i_{(s)} - (1-s)\, m^i_{(s)}.$$

Hence

$$(11)\qquad 0 \leqq m^i_{(s)} - \frac{1}{1-s} F^i_{(s)} \leqq -\frac{1}{s}\left(\frac{1}{1-s} F^i_{(s)}\right) \leqq -\frac{1}{s}\left(\frac{1}{1-s} F_{(s)}\right).$$

Now apply Lemma 2.2 for p^i and $\tilde{q}_{(s)}$ instead for p^i and λ. Setting

$$S_1 = m^i_{(s)} - 2^{\frac{1}{2}} g^i_{(s)}, \qquad S_2 = m^i_{(s)} + 2^{\frac{1}{2}} g^i_{(s)},$$

we obtain from Chebyshev's inequality:

$$(12)\qquad \begin{aligned} &p^i(E_i) \exp[-F^i_{(s)} - s\, m^i_{(s)} + s\, 2^{\frac{1}{2}} g^i_{(s)}] \\ &\quad + \tilde{q}_{(s)}(E_i) \exp[-F^i_{(s)} + (1-s)\, m^i_{(s)} + (1-s)\, 2^{\frac{1}{2}} g^i_{(s)}] \geqq \tfrac{1}{2}. \end{aligned}$$

Put

$$(13)\qquad m^i_{(s)} - \frac{1}{1-s} F^i_{(s)} = -\frac{1}{s}\left(\frac{1}{1-s} F_{(s)}\right) \delta^i_{(s)}.$$

$\delta^i_{(s)}$ is continuous for $0<s<1$ because $m^i_{(s)}$, $F^i_{(s)}$, $F_{(s)}$ are continuous, and $0 \leqq \delta^i_{(s)} \leqq 1$ $(0<s<1)$ because of (11).

We obtain with $p_e := \sup_{1 \leqq i \leqq N} p^i(E_i)$ from (12) for all s $(0<s<1)$:

$$(14)\qquad p_e \exp\left[-\left(\frac{1}{1-s} F_{(s)}\right)(1-\delta^i_{(s)}) + s\, 2^{\frac{1}{2}} V_{(s)}\right] + W^i_{(s)} \geqq \frac{1}{2}$$

where

$$W^i_{(s)} := \left(\tilde{q}_{(s)}(E_i) + \frac{1}{N}\right) \exp\left[-\frac{1}{s} F_{(s)}\, \delta^i_{(s)} + (1-s)\, 2^{\frac{1}{2}} V_{(s)} + \frac{1-s}{s}\right].$$

(The additional $\frac{1}{N}$ and $\frac{1-s}{s}$ has to be used in one of the next steps.)

We are going to use the intermediate value theorem and a distinction of cases as in [2], however, without using fixed composition codes:

(15) Assume that

$$\frac{1}{N}\left|L_1 := \left\{i:\ W^i_{(s)} > \frac{1}{4} \text{ for all } s\ (0<s<1)\right\}\right| \geqq \frac{1}{2}.$$

Then

$$\begin{aligned} \frac{1}{4} < \frac{1}{|L_1|} \sum_{i \in L_1} W^i_{(s)} &\leqq \left(\frac{1}{N} + \frac{1}{|L_1|} \sum_{i \in L_1} \tilde{q}_{(s)}(E_i)\right) \\ &\cdot \exp\left[-\frac{1}{s} F_{(s)} + (1-s)\, 2^{\frac{1}{2}} V_{(s)} + \frac{1-s}{s}\right] \quad \text{for all } s\ (0<s<1). \end{aligned}$$

Thus, with $\frac{1}{|L_1|} \sum_{i \in L_1} \tilde{q}_{(s)}(E_i) \leqq \frac{2}{N}$ we obtain:

$$\frac{3}{N} \exp\left[-\frac{1}{s} F_{(s)} + (1-s)\, 2^{\frac{1}{2}} V_{(s)} + \frac{1-s}{s}\right] \geqq \frac{1}{4} \qquad \text{for all } s\ (0<s<1),$$

$p_e \geqq 0$ and

$$0 = \inf_{0<s<1} \exp\left[\frac{1}{1-s} F_{(s)} + \frac{s}{1-s} \ln N/12 - s\, 2^{\frac{1}{2}} V_{(s)} + \frac{s}{1-s} \ln \frac{12}{13}\right]$$

as the trivial case of the estimate.

(16) Assume

$$\frac{1}{N}\left|L_2 := \left\{i:\ W^i_{(s)} \leqq \frac{1}{4} \text{ for some } s\ (0<s<1)\right\}\right| > \frac{1}{2}.$$

For $i \in L_2$ there is s_i with $W^i_{(s_i)} = \frac{1}{4}$, because $W^i_{(s)}$ is continuous for $0<s<1$ ($\tilde{q}_{(s)}(E_i)$ is continuous in s and $\delta^i_{(s)}$, $F_{(s)}$, $V_{(s)}$ are continuous) and $W^i_{(s)} > \frac{1}{4}$ for s sufficiently small, because

$$\frac{1}{N} \exp\left[\frac{1-s}{s}\right] > \frac{1}{4}$$

for s sufficiently small.

There is $L_3 \subseteq L_2$ with $|L_3| \geqq t^{-2}\,|L_2|$ s.t. $i, j \in L_3$ implies

$$|s_i - s_j| \leqq t^{-2} (W^i_{(s_i)} = W^j_{(s_j)} = \tfrac{1}{4}).$$

Hence for $i \in L_3$ (substituting for $\frac{1}{1-s_i} F_{(s_i)}\, \delta_{(s_i)}$) we obtain:

$$p_e \exp\left[-\frac{1}{1-s_i} F_{(s_i)} + \frac{s_i}{1-s_i} \ln(4\tilde{q}_{(s_i)}(E_i) + 4/N) + 2s\, 2^{\frac{1}{2}} V_{(s)} + 1\right] \geqq \frac{1}{4}.$$

a) Suppose that for every $i \in L_3$ holds $s_i \geqq \frac{1}{4}$ and let s_j be maximal under the s_i. Let

$$q'_v(s_j) := t \int_{-s_j t^{-1} - t^{-2}}^{(1-s_j)t^{-1}} dy\, q_v(s+y), \qquad q'_{(s_j)} := q'_1(s_j) \times \cdots \times q'_t(s_j).$$

$\|q'_v(s_j)\| = (1+t^{-1})$ and $\|q'_{(s_j)}\| = (1+t^{-1})^t \leqq e$, furthermore, $q'_{(s_j)}$ majorizes $\tilde{q}_{(s_i)}$ for all $i \in L_3$.

$$\frac{1}{|L_3|} \sum_{i \in L_3} q'_{(s_j)}(E_i) \leqq \frac{2t^2}{N}\, e.$$

Hence for some $i \in L_3$

$$\tilde{q}_{(s_i)}(E_i) \leqq \frac{2t^2}{N}\, e$$

and for this i

$$p_e \geqq \exp\left[\frac{1}{1-s_i} F_{(s_i)} - \frac{s_i}{1-s_i} \ln\left(\frac{8e\,t^2 + 4}{N}\right) - s_i\, 2^{\frac{3}{2}} V_{(s_i)} - 1 - \ln 4\right].$$

b) If all solutions s_i for $i \in L_3$ satisfy $s_i \leqq \frac{3}{4}$, then let s_j be the maximal solution, define

$$q'_v(s_j) := \int_{-s_j t^{-1}}^{(1-s_j)t^{-1} + t^{-2}} dy\, q_v(s+y)$$

and procede as before.

The result of (15) and (16) is:

$$p_e \geqq \inf_{0<s<1} \exp\left[\frac{1}{1-s} F_{(s)} + \frac{s}{1-s} \ln\left(\frac{N}{24t^2}\right) - s\, 2^{\frac{3}{2}} V_{(s)} - 3\right] \tag{17}$$

for $t > t_0$.

We have to investigate when (17) gives bounds which coincide with the upper bound of Theorem 4.1.

The previous lemmas and (17) yield:

Theorem 4.7. *Given* $\{M_v\}_{v=1,2,\ldots}$ *where* $\{M_{[1,t]}\}_{t=1,2,\ldots}$ *satisfies* (7). *Let* $c>0$ *be fixed.*

Then for $t=t_0(c)$ *holds* $\left(\text{setting } G_t(M_v) := \sup_{0<u} \exp\left[-\frac{1}{1+u} H(u_1, M_v) \frac{t}{t-1}\right]\right)$:

a) *If* $\{(p^i, E_i)\}_{1\leqq i\leqq N}$ *is any code for* $M_{[1,t]}$ *with*

$$\ln N \geqq -\frac{t}{t-1} H(1, M_{[1,t]}) + c\,t + 2^{\frac{3}{2}} \left[\sum_{v=1}^{t} G_t(M_v)\right]^{\frac{1}{2}}$$

then its error p_e *satisfies*

$$\ln p_e \geqq \inf_{0<u\leqq 1} [H(u, M_{[1,t]}) + u \ln N]$$
$$-\frac{5}{t} \ln N + \frac{5}{t} H(1, M_{[1,t]}) - 2^{\frac{5}{2}} \left[\sum_{v=1}^{t} G_t(M_v)\right]^{\frac{1}{2}} - O(\ln t).$$

b) *If* $C(M_v) < B$ $(v=1,2,\ldots)$ *and* $\{(p^i, E_i)\}_{1\leqq i\leqq N}$ *any code for* $M_{[1,t]}$ *with*

$$\ln N \geqq -H(1, M_{[1,t]}) + c\,t + 2^{\frac{3}{2}} t^{\frac{1}{2}} \exp\left[\frac{t}{t-1} B \frac{1}{2}\right]$$

then its error p_e *satisfies*

$$\ln p_e \geqq \inf_{0<u\leqq 1} [H(u, M_{[1,t]}) + u \ln N]$$
$$-\frac{5}{t} \ln N - 2^{\frac{5}{2}} t^{\frac{1}{2}} \exp\left[\frac{t}{t-1} B \frac{1}{2}\right] - O(\ln t).$$

Remark. It is interesting to notice that a) may give proper estimates even when the $C(M_v)$ are not finite.

Proof. a) Let

$$R(N,t) := \ln \frac{N}{24t^2} - 2^{\frac{3}{2}} \left[\sum_{v=1}^{t} G_t(m_v)\right]^{\frac{1}{2}}.$$

We obtain with (10) and Lemma 4.4b) from (17)

$$\ln p_e \geqq \inf_{0<s<1} \left[\frac{1}{1-s} F_{(s)} + \frac{s}{1-s} R(N,t)\right] - 2^{\frac{3}{2}} \left[\sum_{v=1}^{t} G_t(M_v)\right]^{\frac{1}{2}} - 3.$$

Set $H(u):=H(u, M_{[1,t]})$ and with the notation of Lemma 4.4a)

$$A_{(s)}:=\frac{t}{t-1}\left[(1-s'')\,H\left(\frac{s''}{1-s''}\right)+\frac{s}{1-s}(1-s')\,H\left(\frac{s'}{1-s'}\right)\right].$$

$$\frac{1}{1-s}F_{(s)}\geqq A_{(s)}.$$

We have to investigate

$$\inf_{0<s<1}\left[A^{(s)}+\frac{s}{1-s}R(N,t)\right].$$

Observe that $H\left(\frac{s}{1-s}\right)\geqq A_{(s)}$ because of the proof of Lemma 4.4 and because of Lemmas 3.5, 3.6.

$\frac{1}{u}H(u)$ increases with u. Therefore

$$\inf_{0<u}\left[H(u)+u(-H(1))\right]=\inf_{0<u\leqq 1}\left[H(u)+u(-H(1))\right].$$

From the supposition that (7) holds, follows

$$A_{(s)}\geqq\frac{t}{t-1}H\left(\frac{s}{1-s}\right)+o(t)\frac{t}{t-1}\,\frac{s}{1-s}$$

uniformly for $s\geqq\frac{1}{4}$.

Hence, for $c>0$ fixed, $t>t_0(c)$, $R(N,t)>\frac{t}{t-1}H(1)+c\,t$, we have

$$\inf_{0<s<1}\left[A_{(s)}+\frac{s}{1-s}R(N,t)\right]=\inf_{0<s<\frac{1}{2}}\left[A_{(s)}+\frac{s}{1-s}R(N,t)\right]$$

$$\geqq\inf_{0<s\leqq\frac{1}{2}}\left[A_{(s)}+\frac{s}{1-s}\ln N\right]-2^{\frac{3}{2}}\left[\sum_{v=1}^{t}G_t(M_v)\right]^{\frac{1}{2}}-\ln(24t^2).$$

Finally,

$$\inf_{0<s\leqq\frac{1}{2}}\left[A_{(s)}+\frac{s}{1-s}\ln N\right]$$

can be brought into the form of the upper error estimate:

For $0<s\leqq\frac{1}{2}$ holds

$$\frac{s}{1-s}(1-s')\,H\left(\frac{s'}{1-s'}\right)$$

$$=s\,H\left(\frac{s'}{1-s'}\right)+\frac{s^2}{1-s}t^{-1}H\left(\frac{s'}{1-s'}\right)\geqq s''\,H\left(\frac{s''}{1-s''}\right)+t^{-1}H(1)$$

and therefore

$$A(s)\geqq\frac{t}{t-1}H\left(\frac{s''}{1-s''}\right)+\frac{1}{t-1}H(1).$$

$$H\left(\frac{s''}{1-s''}\right)=H\left(\frac{s}{1-s}+\frac{1}{1-s}\,\frac{1}{t-1}\right)\geqq H\left(\frac{s}{1-s}+\frac{2}{t-1}\right)\qquad\text{for } s\leqq\frac{1}{2}.$$

One computes for $u \leqq 1$ using the monotonicity of $\frac{1}{u} H(u)$:

$$\inf_{0<u\leqq 1} \left[\frac{t}{t-1} H\left(u+\frac{2}{t-1}\right)+u \ln N\right] \geqq \inf_{0<u\leqq 1} [H(u)+u \ln N] - \frac{5}{t} \ln N + \frac{2}{t-1} H(1) \qquad (t>t_0).$$

This yields part a) of the theorem, Part b) follows from part a) using

$$-\frac{1}{1+u} H(u, M_v) \leqq C(M_v) \quad \text{and} \quad -H(1, M_v) \leqq C(M_v).$$

We will prove now a sharper estimate for the stationary finite alphabet channels and for a few more.

For this purpose, Berry's inequality can be used which says: For independent random variables $r_1, \ldots, r_t$ with expectation values $\int dP\, r_v = 0$ and finite third moments holds

$$\left|P\left\{\sum_{v=1}^{t} r_v < y\, D_t\right\} - \Phi_{(y)}\right| < \omega \frac{R_t^3}{D_t^3}, \tag{18}$$

where

$$D_t := \left[\sum_{v=1}^{t} \int dP\, r_v^2\right]^{\frac{1}{2}},$$

$$R_t := \left[\sum_{v=1}^{t} \int dP\, |r_v|^3\right]^{\frac{1}{3}},$$

$\Phi_{(y)}$ is the standard normal distribution, ω an absolute constant. (It is known that $\omega = 1.322$ is a possible choice for ω.)

Set $\omega = 1.5$ and let Ψ be the inverse function of Φ. Suppose, furthermore, that $|r_v| < B$ $(1 \leqq v \leqq t)$.

Then $\frac{B}{D_t} \geqq \frac{R_t^3}{D_t^3}$ and because of

$$\left[\frac{1}{t} \sum_{v=1}^{t} \int dP\, r_v^2\right]^{\frac{1}{2}} \leqq \left[\frac{1}{t} \sum_{v=1}^{t} \int dP\, |r_v|^3\right]^{\frac{1}{3}} \quad \text{one has} \quad \frac{R_t^3}{D_t^3} \geqq t^{-\frac{1}{2}}.$$

i) Suppose $\frac{2B}{D_t} < \frac{1}{4}$. Then for some fixed $A > 0$ holds

$$\Psi\left(\frac{1}{2}+\frac{2B}{D_t}\right) \leqq \Psi\left(\frac{1}{2}\right) + A \frac{2B}{D_t} = A \frac{2B}{D_t}$$

and we obtain from (18)

$$P\left\{\left|\sum_{v=1}^{t} r_v\right| < A \frac{2B}{D_t} D_t\right\} \geqq P\left\{\left|\sum_{v=1}^{t} r_v\right| < \Psi\left(\frac{1}{2}+\frac{2B}{D_t}\right) D_t\right\}$$
$$\geqq 2\frac{2B}{D_t} - 2 \cdot 1.5 \frac{B}{D_t} = \frac{B}{D_t} \geqq t^{-\frac{1}{2}}.$$

ii) Suppose $\frac{2B}{D_t} \geqq \frac{1}{4}$. Then $D_t < 8B$ and Chebyshev's inequality gives

$$P\left\{\left|\sum_{v=1}^{t} r_v\right| < 16B\right\} \geqq P\left\{\left|\sum_{v=1}^{t} r_v\right| < 2^{\frac{1}{2}} D_t\right\} > \tfrac{1}{2}.$$

It follows from i) and ii) that

$$P\left\{\left|\sum_{v=1}^{t} r_v\right| < A \cdot B\right\} > t^{-\frac{1}{2}} \qquad (t \geqq 4) \tag{19}$$

for some absolute constant A if the independent random variables r_v with $\int dP\, r_v = 0$ satisfy $|r_v| < B$ $(1 \leqq v \leqq t)$ for any $B > 0$.

Using (19) instead of Chebyshev's inequality for (12) and applying at the same time a method of small perturbations of probabilities we obtain:

Theorem 4.8. *Let $\{M_v\}_{v=1,2,\ldots}$ be given where for every M_v there is a probability λ_v s.t.*

$$\frac{1}{K} \leqq \frac{dp_v}{d\lambda_v} \leqq K \qquad p_v - a.e. \text{ forall } p_v \in M_v \ (v = 1, 2, \ldots).$$

Furthermore, suppose that (7) holds for $\{M_{[1,t]}\}_{t=1,2,\ldots}$.

Let $c > 0$ be fixed. Then for $t > t_0(c)$ holds:

If $\{p^i, E_i\}_{1 \leqq i \leqq N}$ is any code for $M_{[1,t]}$ with $\ln N \geqq -H(1, M_{[1,t]}) + ct$ then its error p_e satisfies

$$\ln p_e \geqq \inf_{0 < u \leqq 1} [H(u, M_{[1,t]}) + u \ln N] - O(\ln t).$$

Proof. Let $\tilde{q}_{(s)} := \tilde{q}_1(s) \times \cdots \times \tilde{q}_t(s)$ be given with respect to the code as directly before (8) and let

$$q'_v(s) := \left(1 - \frac{1}{t}\right) \tilde{q}_v(s) + \frac{1}{t} \lambda_v, \qquad q'_v := q'_1(s) \times \cdots \times q'_t(s).$$

One has p^i_v − a.e.

$$tK \geqq t \frac{dp^i_v}{d\lambda_u} \geqq \frac{dp^i_v}{dq'_v(s)} \geqq \frac{1}{t} \frac{d\lambda_v}{dq'_v(s)}$$

$$= \frac{1}{t}\left[\left(1 - \frac{1}{t}\right)\frac{d\tilde{q}_v(s)}{d\lambda_v} + \frac{1}{t}\right]^{-1} \geqq \frac{1}{t}\left[\left(1 - \frac{1}{t}\right) K + \frac{1}{t}\right]^{-1} \geqq \frac{1}{tK}$$

where $\frac{d\tilde{q}_v(s)}{d\lambda_v} \leqq K$ holds because

$$\tilde{q}_v(s) = t \cdot \int_{-st^{-1}}^{(1-s)t^{-1}} dy\, q_v(s+y)$$

and

$$dq_v(s+y) = d\lambda_v \left[\sum_{j=1}^{u} a^{(j)} \left(\frac{dp^{(j)}_v}{d\lambda_v}\right)^{1-s-y}\right]^{\frac{1}{1-s-y}} \leqq d\lambda K.$$

Write the expressions defined in (8) with a bar if $\tilde{q}_v(s)$ and $\tilde{q}_{(s)}$, respectively, are exchanged by $q'_v(s)$ and $q'_{(s)}$, respectively.

Using (19) instead of Chebyshev's inequality, setting $P=\bar{Q}^i_{(s)}$ and $r_v=\ln\frac{dp^i_v}{dq'_v(s)}-\bar{m}^i_v(s)$, we obtain instead of (17) (observing $|r_v|\leqq 2\,|\ln(Kt)|$):

$$\ln p_e \geqq \inf_{0<s<1}\left[\frac{1}{1-s}\bar{F}_{(s)}+\frac{s}{1-s}(\ln N-O(\ln t))\right]-O(\ln t).$$

It remains to compare $\bar{F}_{(s)}$ and $F_{(s)}$:

$$\ln\int dp^i_v\left(\frac{dp^i_v}{dq'_v}\right)^{-s}=\ln\int dp^i_v\left(\frac{\left(1-\frac{1}{t}\right)d\tilde{q}_v(s)+\frac{1}{t}\,d\lambda_v}{dp^i_v}\right)^s$$

$$\geqq(1-t^{-1})F^i_v(s)+t^{-1}\ln\int dp^i_v K^{-s}=(1-t^{-1})F^i_v(s)-\frac{s}{t}\ln K.$$

Hence $\bar{F}_{(s)}\geqq(1-t^{-1})F_{(s)}-s\ln K\geqq F_{(s)}-s\ln K$ and therefore

$$\ln p_e \geqq \inf_{0<s<1}\left[\frac{1}{1-s}F_{(s)}+\frac{s}{1-s}(\ln N-O(\ln t))\right]-O(\ln t)$$

and one procedes as in the proof of Theorem 4.7. Finally

$$\frac{t}{t-1}H(1,M_{[1,t]})\geqq H(1,M_{[1,t]})-\frac{1}{t}C(M_{[1,t]})\geqq H(1,M_{[1,t]})-\frac{1}{t}(t\ln K).$$

Remark. In the situation of Theorem 4.7 $\left(\text{without a condition as f.i. } \frac{1}{K}\leqq\frac{dp_v}{d\lambda_v}\leqq Kp_v-\text{a.e.}\right)$ one still can replace the 2nd rootes in the estimates by 3rd rootes by means of Berry's inequality.

Because we are interested when there exist lengths N for which conditions for N as in the corollaries following Theorem 4.1 and as in Theorem 4.7 are simultaneously satisfied (restricting ourselves to stationary channels) we have to investigate when $-H(1,M)<C(M)$ is true.

Let $H(1,M)>-\infty$. ($H(u,M)=-\infty(u>0)$ implies $P_e(M,N)=0$).

a) If $C(M)=\infty$ then $C(M)>-H(1,M)\geqq 0$.

b) If $M=M'=\{p^{(j)}\}_{1\leqq j\leqq n}$ and

$$\ln\int d\lambda\left[\sum_{j=1}^{n}a^{(j)}\left(\frac{dp^{(j)}}{d\lambda}\right)^{\frac{1}{1+u_0}}\right]^{1+u_0}=-u_0\,C(M')\qquad\text{for } u_0>0$$

and coefficients $a^{(j)}\geqq 0$ with $\sum_{j=1}^{n}a^{(j)}=1$ then $H(u,M')=-u\,C(M')$ for $0\leqq u\leqq u_0$ because of (6).

Moreover using the remarks following (6) on $\frac{d^2}{du^2}\bar{H}(u)$ we obtain for case b):

$$\inf\left\{\sum_{j=1}^{n} a^{(j)} \int dp^{(j)} \left(\ln \frac{dp^{(j)}}{d\left(\sum_{k=1}^{n} a^{(k)} p^{(k)}\right)} - C(M')\right)^2 :\right.$$

$$\left.\sum_{j=1}^{n} a^{(j)} \int dp^{(j)} \ln \frac{dp^{(j)}}{d\left(\sum_{k=1}^{n} a^{(k)} p^{(k)}\right)} = C(M')\right\} = 0.$$

A well known lemma says:

If

$$\sum_{j=1}^{n} a^{(j)} \int dp^{(j)} \ln \frac{dp^{(j)}}{d\left(\sum_{k=1}^{n} a^{(k)} p^{(k)}\right)} = C(M')$$

then

$$\int dp^{(j)} \ln \frac{dp^{(j)}}{d\left(\sum_{k=1}^{n} a^{(k)} p^{(k)}\right)} \leqq C(M') \qquad (1 \leqq j \leqq n)$$

and $= C(M')$ for all j with $a^{(j)} > 0$.

(It may be checked by differentiation with respect to the $a^{(j)}$.)

This implies:

Lemma 4.9. *Given $M' = \{p^{(j)}\}_{1 \leqq j \leqq n}$ and suppose that*

$$-u_0\, C(M') = \int d\lambda \left[\sum_{j=1}^{n} a^{(j)} \left(\frac{dp^{(j)}}{d\lambda}\right)^{\frac{1}{1+u_0}}\right]^{1+u_0}$$

holds for certain coefficients $a^{(j)} \geqq 0$ with $\sum_{j=1}^{n} a^{(j)} = 1$ and u_0 fixed ($u_0 > 0$). Then

i) $$\ln \frac{dp^{(j)}}{d\left(\sum_{k=1}^{n} a^{(k)} p^{(k)}\right)} = C(M') \qquad p^{(j)} - a.e.$$

for all j with $a^{(j)} > 0$,

ii) *the probability $q = q_{(s)}$ determined in Lemma* 3.5 *is the same probability for all s* $(0 < s < 1)$.

It remains to prove ii). With the coefficients $a^{(j)}$ in the lemma holds:

$$d\lambda\left[\sum_{j=1}^{n} a^{(j)}\left(\frac{dp^{(j)}}{d\lambda}\right)^{1-s}\right]^{\frac{1}{1-s}}$$

$$=d\left(\sum_{k=1}^{n} a^{(k)}p^{(k)}\right)\left[\sum_{j=1}^{n} a^{(j)}\left(\frac{dp^{(j)}}{d\left(\sum_{k=1}^{n} a^{(k)}p^{(k)}\right)}\right)^{1-s}\right]^{\frac{1}{1-s}}$$

$$=d\left(\sum_{k=1}^{n} a^{(k)}p^{(k)}\right)\left[\exp[-C(M')](\exp[C(M')])^{1-s}\right]^{\frac{1}{1-s}}$$

$$=d\left(\sum_{k=1}^{n} a^{(k)}p^{(k)}\right)\exp\left[-\frac{s}{1-s}C(M')\right].$$

Corollary. *$H(u_0, M') = -u_0 C(M')$ for some $u_0>0$ implies $H(u,M') = -u\,C(M')$ for all $u>0$.*

Remark. If M is infinite and $C(M)<\infty$ then (see [3]) M is relatively weakly compact (or equivalent: uniformly absolutely continuous). Take a countable subset $\tilde{M}$ of M with $H(u,\tilde{M}) = H(u,M)$ for all $u \geqq 0$. For the weak closure $\overline{co(\tilde{M})}$ of the convex hull of $\tilde{M}$ holds $H(u, \overline{co(\tilde{M})}) = H(u,M)$. Applying a Choquet's centroid theorem argument to $\overline{co(\tilde{M})}$ one obtains similar results as in Lemma 4.9 and in the corollary.

Thus, we have for arbitrary M with $C(M)<\infty$

$$-H(1,M)=C(M) \quad \textit{iff} \quad -u\,H(u,M)=u\,C(M) \qquad \textit{for all } u\geqq 0$$

(as a consequence of the corollary and the last remark).

We still can formulate instead of Theorem 4.8 in a bit more natural way:

In the situation of Theorem 4.8 holds: If for $t>t_0(c,u_0)$ (any $u_0>0$)

$$\ln N \geqq -\frac{1}{u_0} H(u_0, M_{[1,t]}) + c\,t$$

then

$$\ln P_e(M_{[1,t]}, N) \geqq \inf_{0<u\leqq u_0} [H(u, M_{[1,t]}) + u \ln N] - O(\ln t).$$

In [2] for the stationary finite alphabet channel without memory the lower bound

$$\ln P_e(M_{[1,t]}, N) \geqq \inf_{0<u} [H(u, M_{[1,t]}) + u(\ln N - O(t^{\frac{1}{2}}))] - O(t^{\frac{1}{2}})$$

has been derived for all $N\geqq 1$.

However, neither the upper bound of Theorem 4.1 nor the lower bound of [2] are i.g. tight for $\ln N < -H(1, M_{[1,t]})$ as the following two examples show:

1. Let $X=\{1,2,3\}$, $M=\{p^{(j)}\}$ $j=1,2,3$ where $p^{(j)}(j)=0$, $p^{(j)}(k)=\frac{1}{2}$ for $k\neq j$ ($j=1,2,3$; $k=1,2,3$). Then $-H(u,M)=u\,C(M)=u\ln\frac{3}{2}$.

For $M_{[1,t]}$ where the M_v are copies of M holds

$$P_e(M_{[1,t]}, N) \geqq (\tfrac{1}{2})^t \quad \text{if } N \geqq 2, \qquad \text{because} \qquad (p^{j_1} \wedge p^{j_2})(X) = \tfrac{1}{2}$$

for different $p^{j_1}, p^{j_2} \in M$. But $\inf_{0<u} \big(H(u, M_{[1,t]}) + u \ln N\big) = -\infty$ if $\frac{1}{t} \ln N < C(M)$.

2. Let $X = \{1, 2\}$, $M = \{p^{(1)}, p^{(2)}\}$ where $p^{(1)} = 1$, $p^{(2)}(2) = 1$. $P_e(M_{[1,t]}, N) = 0$ for $\frac{1}{t} \ln N \leqq C(M) = \ln 2$, (the M_v copies of M). But

$$\inf_{0 \leqq u \leqq 1} \big(H(u, M_{[1,t]}) + u \ln N\big) \geqq \inf_{0 \leqq u \leqq 1} H(u, M_{[1,t]}) = -t \ln 2.$$

I would like to thank Johan H. B. Kemperman for a very stimulating discussion and for pointing out a gap in my original proof.

References

1. Gallager, R.: A simple derivation of the coding theorem. IEEE Trans. Inform. Theory IT−11, 3−18 (1965).
2. − Shannon, C. E., Berlekamp, E. R.: Lower bounds to error probability for coding. Inform. and Control, **10**, 65−103 (1967).
3. Augustin, U.: Gedächtnisfreie Kanäle für diskrete Zeit. Z. Wahrscheinlichkeitstheorie verw. Geb. **6**, 10−61 (1966).
4. Wolfowitz, J.: Coding theorems of information theory. Ergebnisse d. Math. u. Grenzgebiete, Vol. 31. Berlin-Göttingen-Heidelberg: Springer 1964.

Dr. Udo Augustin
Mathematisches Institut der Universität
Erlangen-Nürnberg
8520 Erlangen, Bismarckstraße $1^1/_2$

(Received October 7, 1968)