The Consistency ot the Empirical Quantization Error

Pötzelberger, Klaus

DOI:
10.57938/0aed3f0a-bb25-4e05-a0e4-1d3b0a44fa74

Published: 01/01/1999

Document Version
Publisher's PDF, also known as Version of record

The Consistency ot the Empirical Quantization Error

Klaus Pötzelberger

Institut für Statistik
Wirtschaftsuniversität Wien

Forschungsberichte

Bericht 70
1999

http://statmath.wu-wien.ac.at/

The Consistency of the Empirical Quantization Error

Klaus Pötzelberger

Abstract

We study the empirical quantization error in case the number of prototypes increases with the size of the sample. We present a proof of the consistency of the empirical quantization error and of corresponding estimators of the quantization dimensions of distributions.

AMS 1991 subject classifications. Primary 60F15, 60F05; secondary 62G20, 28A75.

Key words and phrases. Quantization dimension, quantization error, dimension estimation, dimension of a distribution.

1 Introduction

We prove the consistency of the empirical quantization error and of estimators of the quantization dimension, which are based on the empirical quantization error. Let a distribution P on a bounded subset of $\mathbb{R}^d$ and a partition $\mathcal{B}$ of $\mathbb{R}^d$ into k Borel measurable nonvoid subsets $(B_1, \ldots, B_k)$ be given. The quantization error $\Delta_k(\mathcal{B}, P)$ of $\mathcal{B}$ is defined as

$$\Delta_k(\mathcal{B}, P) = \sum_{i=1}^{k} \int_{B_i} \|X - \mathbb{E}(X \mid B_i)\|^2 dP(X), \tag{1}$$

where $X : \mathbb{R}^d \to \mathbb{R}^d, X(x) = x$ is the identity on $\mathbb{R}^d$. The quantization error $\Delta_k(P)$ is

$$\Delta_k(P) = \inf_{|\mathcal{B}|=k} \Delta_k(\mathcal{B}, P). \tag{2}$$

A partition, which minimizes (1) is called optimal. Optimal partitions $\mathcal{B}$ correspond to sets $O_k = (\mu_1, \ldots, \mu_k)$ of prototypes, where the prototype $\mu_i = \mathbb{E}(X \mid B_i)$ is the centroid

(conditional mean) of the set B_i. If a set of prototypes O_k is given, then the optimal partition $\mathcal{B}$ is $\mathcal{S}(O_k)$, the segmentation of O_k, i.e. $B_i = \{x \mid i = argmin_j \|x - \mu_j\|\}$, with arbitrary assignment of the boundaries. We denote $\Delta_k(\mathcal{B}, P)$ also by $\Delta_k(O_k, P)$.
When the size of the partition is held fixed, $\Delta_k(\hat{P}_n) - \Delta_k(P) \to 0$ for $n \to \infty$, proofs under assumptions of varying generality have been given by Pollard (1981), Linder et al. (1994) and Pötzelberger and Strasser (1997), among others. With k increasing, the quantization error converges to 0. In this case the asymptotic behavior of the ratio $\Delta_k(\hat{P}_n)/\Delta_k(P)$ is of interest.
There are at least two fields where the consistency of the empirical quantization error (i.e. $\Delta_k(\hat{P}_n)/\Delta_k(P) \to 1$) is relevant. The first is the choice of a reasonable size of a partition or a reasonable number of prototypes representing a given distribution. This has to be done when applying k-means clustering (choosing a variance-minimizing partition), when principal points are computed, when objects are classified or when a signal is coded. In all cases a discrete distribution with support on k points, which is computed from the empirical distribution $\hat{P}_n$, is sought as an approximation for P.
The second field should be described as the search for structure. It is a theme of central importance in all modern approaches to the analysis of high dimensional data. By structure a low dimensional subset E is meant, such that $P(E) = 1$. Various procedures have been designed for exhibiting these sets, typically with the restriction that E belongs to some parametrized family of subsets, such as the family of affine subspaces. The asymptotic behavior of the quantization error $\Delta_k(P)$ is a dimensional characteristic of the support of P. It allows the construction of estimators of the dimension of E. The upper and the lower quantization dimensions of a distribution P on $\mathbb{R}^d$ are defined as

$$\overline{dim}_Q(P) = \limsup_{k\to\infty} \frac{2\log(1/k)}{\log \Delta_k(P)}, \tag{3}$$

$$\underline{dim}_Q(P) = \liminf_{k\to\infty} \frac{2\log(1/k)}{\log \Delta_k(P)}. \tag{4}$$

If $\overline{dim}_Q(P) = \underline{dim}_Q(P)$, the common value is called the quantization dimension $dim_Q(P)$. These concepts fit into the framework of geometric measure theory. Pötzelberger (1998) showed that for distributions with bounded support, $\underline{dim}_Q(P) \in [dim_{\mathcal{H}}(P), \underline{dim}_B(P)]$ and $\overline{dim}_Q(P) \in [dim_{\mathcal{P}}(P), \overline{dim}_B(P)]$. $dim_{\mathcal{H}}(P)$, $dim_{\mathcal{P}}(P)$, $\underline{dim}_B(P)$ and $\overline{dim}_B(P)$ denote the Hausdorff,- packing,- lower and upper box-counting dimension of P. Moreover, the Theorem of Zador (Zador (1963, 1982)) has been generalized to distributions with regular s-dimensional support E, stating

$$\lim_{k\to\infty} k^{2/s}\Delta_k(P) = c(P), \tag{5}$$

with a constant $c(P) \in]0,\infty[$.

Estimators of the quantization dimensions are constructed from the empirical quantization error. Let P denote a distribution on $\mathbb{R}^d$ with bounded support. Let $(X_i)_{i=1}^n$ be a sample of independent P-distributed variables and let $\hat{P}_n$ denote the empirical distribution of $(X_i)_{i=1}^n$. $\hat{O}_k$ denotes a set of k prototypes for $\hat{P}_n$. Let (k_i) and (n_i) be sequences of natural numbers with $n_i \to \infty$.

The empirical estimator of the quantization dimension $\hat{s}_{n_i}^{emp}$ is

$$\hat{s}_{n_i}^{emp} = \frac{2\log(1/k_i)}{\log \Delta_{k_i}(\hat{P}_{n_i})}. \tag{6}$$

To compute the regression estimator $\hat{s}_{n_i}^{reg}$, $\log \Delta_1(\hat{P}_{n_1}), \ldots, \log \Delta_{k_i}(\hat{P}_{n_i})$ is regressed onto $\log 1, \ldots, \log k_i$ and $-2/\hat{s}_{n_i}^{reg}$ is set equal to the slope of the empirical regression line. For small numbers of prototypes the regression estimator is typically superior to the empirical estimator, as it capeable of compensating for smaller order effects, such as constants.

If $dim_Q(P)$ exists and $\Delta_{k_i}(\hat{P}_{n_i})/\Delta_{k_i}(P) \to 1$ a.e. (in probability), then both $\hat{s}_{n_i}^{emp}$ and $\hat{s}_{n_i}^{reg}$ are strongly consistent (consistent) estimators of $dim_Q(P)$ (given a growth condition, i.e. $k_i \leq \psi(n_i)$ for a suitable function ψ). Since in general $\underline{dim}_Q(P) < \overline{dim}_Q(P)$, we have to define what is meant by a consistency.

DEFINITION **1.** *Let (k_i) and (n_i) be sequences of natural numbers with $n_i \to \infty$. An estimator sequence $(\hat{s}_{n_i})$ is (n_i, k_i)-consistent, if*

$$\hat{s}_{n_i} \frac{\log \Delta_{k_i}(P)}{2\log(1/k_i)} \to 1 \tag{7}$$

in probability. If (7) holds a.e., then $(\hat{s}_{n_i})$ is called strongly (n_i, k_i)-consistent.

The major part of the literature on estimators of dimensions is concerned with dimensions of paths of stochastic processes. Cutler and Dawson (1989) present procedures for estimating the dimension of the support of a distribution. These procedures are based on estimates of the local dimension, the results assume a regular support (and therefore necessarily of integral dimension). Procedures based on the quantization dimension are not restricted to regular sets. The quantization dimension exists for a variety of distributions, including all cases where the Hausdorff - and the upper box-counting dimension are identical. When the quantization dimension does not exist, accumulation points of $2\log(1/k)/\log\Delta_k(P)$ may be identified by passing to subsequences (k_i).

2 Results

We prove a result for distributions P with bounded support. The reason for this restriction is that in case of an unbounded support the asymptotic behavior of the quantization error $\Delta_k(P)$ depends not only on dimensional characteristics of its support, but also on regularity conditions of the distribution, such as the existence of moments. Moreover, concepts from geometric measure theory, such as box-counting dimensions, are reasonable only for bounded supports. Results similar to the Theorem of Zador have been proved for distributions of unbounded support under assumptions, which imply that for any $\epsilon > 0$ a compact set K exists, such that for $P_{|K}$, the restriction of P to K, $\Delta_k(P_{|K})/\Delta_k(P) \geq (1-\epsilon)$ holds. In this case, a result, which is established for distributions on bounded sets, is in fact not restrictive. However, in the most general cases these assumptions do not hold. Without further information on the distribution only quantization errors on suitably large, but bounded, sets should be used to construct estimators of the quantization dimensions.

Proofs of the convergence of the ratio $\Delta_{k_i}(\hat{P}_{n_i})/\Delta_{k_i}(P)$ are based on a uniform Glivenko-Cantelli result, more precisely on an estimate of $P(\sup_f |\hat{P}_{n_i}(f) - P(f)| > \epsilon_i)$ for a class of mappings $f : x \mapsto ||x-\mu||^2 I_B(x)$, where the sets B belong to a class of sets including the elements of the optimal partitions $\mathcal{B}$ and $\hat{\mathcal{B}}$. (ϵ_i) is a positive null sequence. The polytopes B are intersections of $\varkappa$ halfspaces. $\varkappa$ depends on d, possibly also on k_i, the size of the partition. $\varkappa$ is at most 2 in case $d = 1$. $k_i - 1$ is a trivial upper bound on $\varkappa$ in general. Although $\varkappa$ seems to be bounded (as a function of k_i), no result is known that gives a reasonable upper bound.

The set of $\varkappa$ intersections of halfspaces is a (Vapnik-Červonenkis-) VC-class, appropriate uniform Glivenko-Cantelli results are available. However, with increasing complexity, any convex set may be approximated by polytopes. The class of convex sets is no VC-class. Therefore, a result on the convergence of $\Delta_{k_i}(\hat{P}_{n_i})/\Delta_{k_i}(P)$, cannot be proved within the toolbox provided by VC-theory alone. Both, the size of a partition and its complexity, have to be taken into consideration.

DEFINITION **2.** *Let $\mathcal{B} = (B_1, \ldots, B_k)$ be a partition. The complexity $\varkappa(\mathcal{B})$ of $\mathcal{B}$ is the smallest number $\varkappa$, such that each element of the partition is the intersection of at most $\varkappa$ halfspaces.*

THEOREM **1.** *A constant $c^* \le 3(d+3)^{1/2}2^6$ exists, such that for all distributions P on $\mathbb{R}^d$ with bounded support, for all sequences $(n_i)_{i=1}^{\infty}$ and $(k_i)_{i=1}^{\infty}$ with $n_i \to \infty$ and for all sequences of optimal partitions (O_{k_i}) and $(\hat{O}_{k_i})$ for P and for $\hat{P}_{n_i}$ resp.,*

(i)

$$P\left(\frac{\Delta_{k_i}(\hat{P}_{n_i})}{\Delta_{k_i}(P)} > 1 + c^* \left(\frac{\varkappa_i k_i \log n_i}{\Delta_{k_i}(P) n_i}\right)^{1/2} \quad i.o.\right) = 0, \tag{8}$$

$$P\left(\frac{\Delta_{k_i}(\hat{P}_{n_i})}{\Delta_{k_i}(P)} < 1 - c^* \left(\frac{\hat{\varkappa}_i k_i \log n_i}{\Delta_{k_i}(P) n_i}\right)^{1/2} \quad i.o.\right) = 0, \tag{9}$$

where $\varkappa_i = \varkappa(\mathcal{S}(O_{k_i}))$ and $\hat{\varkappa}_i = \varkappa(\mathcal{S}(\hat{O}_{k_i}))$.

(ii)

$$\limsup_{i\to\infty} \frac{\Delta_{k_i}(\hat{P}_{n_i})}{\Delta_{k_i}(P)} \le 1 \quad a.e. \quad if \quad \varkappa_i k_i = o\left(\frac{\Delta_{k_i}(P) n_i}{\log n_i}\right), \tag{10}$$

$$\liminf_{i\to\infty} \frac{\Delta_{k_i}(\hat{P}_{n_i})}{\Delta_{k_i}(P)} \ge 1 \quad a.e. \quad if \quad \hat{\varkappa}_i k_i = o\left(\frac{\Delta_{k_i}(P) n_i}{\log n_i}\right). \tag{11}$$

(iii)

$$\lim_{i\to\infty} \frac{\Delta_{k_i}(\hat{P}_{n_i})}{\Delta_{k_i}(P)} = 1 \quad a.e. \quad if \quad (\varkappa_i \vee \hat{\varkappa}_i) k_i = o\left(\frac{\Delta_{k_i}(P) n_i}{\log n_i}\right). \tag{12}$$

(iv) (12) *holds, if $k_i = o((\frac{n_i}{\log n_i})^{s/(2+2s)})$ for a $s \in \mathbb{R}$ with $\liminf_i k_i^{2/s}\Delta_{k_i}(P) > 0$, or if $(\varkappa_i)$ and $(\hat{\varkappa}_i)$ are bounded and $k_i = o((\frac{n_i}{\log n_i})^{s/(2+s)})$ for a $s \in \mathbb{R}$ with $\liminf_i k_i^{2/s}\Delta_{k_i}(P) > 0$.*

(v) (12) *implies that (O_{k_i}) is asymptotically optimal for $(\hat{P}_{n_i})$ and $(\hat{O}_{k_i})$ for P.*

Theorem 1 allows the specification of rates for (k_i), which guarantee the consistency of the empirical quantization error (in the sense of (12)) and of estimators of the quantization dimensions.

APPLICATION (Consistency of estimators).

(a) Let $s_0 \in]0, d]$ and $k_i = o((n_i / \log n_i)^{s_0/(2+2s_0)})$. Thus $\lim_i \Delta_{k_i}(\hat{P}_{n_i})/\Delta_{k_i}(P) = 1$ a.e. if $\liminf_i 2\log(1/k_i)/\Delta_{k_i}(P) > s_0$. ($\liminf_i 2\log(1/k_i)/\Delta_{k_i}(P) = s_0$ is sufficient in the exact case.) Particularly, estimators such as $\hat{s}_{n_i}^{emp}$ or $\hat{s}_{n_i}^{reg}$ are (strongly) (n_i, k_i)-consistent for all distributions P with $s_0 < dim_{\mathcal{H}}(P)$.

(b) If the complexities of (O_{k_i}) and of $(\hat{O}_{k_i})$ are bounded, then $k_i = o((n_i/\log n_i)^{s_0/(2+s_0)})$ is sufficient.

(c) Let $\hat{\varkappa}_i k_i = o(\Delta_{k_i}(\hat{P}_{n_i}) n_i / \log n_i)$. Then $\liminf_i \Delta_{k_i}(\hat{P}_{n_i})/\Delta_{k_i}(P) \geq 1$ a.e. However, since $\Delta_{k_i}(\hat{P}_{n_i}) \leq \Delta_{k_i}(O_{k_i}, \hat{P}_{n_i})$ and $\mathbb{E}(\Delta_{k_i}(O_{k_i}, \hat{P}_{n_i})) = \Delta_{k_i}(P)$, $\lim_i \Delta_{k_i}(\hat{P}_{n_i})/\Delta_{k_i}(P) = 1$ in probability. Let (ψ_i) be a positive null sequence and define for a given sequence (n_i), $k_i^* := \max\{k \mid k\hat{\varkappa}_i \leq \psi_i \Delta_k(\hat{P}_{n_i}) n_i / \log n_i\}$. $\hat{s}_{n_i}^{emp}$ and $\hat{s}_{n_i}^{reg}$ are then (n_i, k_i)-consistent for all distributions and all sequences (k_i) with $k_i \leq k_i^*$. Note that the number of prototypes may be chosen according to the information provided by the observed complexity $\hat{\varkappa}_i$. This result is of central importance, since no bounds on the complexity $\varkappa_i$ are known.

(d) In trivial cases $\limsup_i \Delta_{k_i}(\hat{P}_{n_i})/\Delta_{k_i}(P)$ is necessarily bounded (by a constant). For instance, if $dim_{\mathcal{H}}(P) = \overline{dim}_B(P)$ and if the support of P is of exact upper box-counting dimension $\overline{dim}_B(P)$, which is the case for distributions with nonvanishing absolutely continuous component (with respect to d-dimensional Lebesgue measure). The estimators $\hat{s}_{n_i}^{emp}$ and $\hat{s}_{n_i}^{reg}$ are then consistent under situation (c).

REMARK. 1. $3(d+3)^{1/2}2^6$ is only an upper bound for c^*. c^* should be much smaller for large classes of distributions.

2. It is sometimes convenient to restrict the class of partitions, for instance to consider only rectangles or partitions with a given upper bound on the complexity. Theorem 1 remains true in these cases.

3 Proof of the Theorem

W.l.o.g. let $P([0,1]^d) = 1$. We show that for all $\epsilon > 0$,

$$\limsup_{i\to\infty} \frac{\Delta_{k_i}(O_{k_i}, \hat{P}_{n_i})}{\Delta_{k_i}(O_{k_i}, P)} \le 1 + \epsilon \tag{13}$$

and

$$\liminf_{i\to\infty} \frac{\Delta_{k_i}(\hat{O}_{k_i}, \hat{P}_{n_i})}{\Delta_{k_i}(\hat{O}_{k_i}, P)} \ge 1 - \epsilon. \tag{14}$$

Thus

$$\limsup_{i\to\infty} \frac{\Delta_{k_i}(\hat{O}_{k_i}, \hat{P}_{n_i})}{\Delta_{k_i}(O_{k_i}, P)} \le \limsup_{i\to\infty} \frac{\Delta_{k_i}(O_{k_i}, \hat{P}_{n_i})}{\Delta_{k_i}(O_{k_i}, P)} \le 1 + \epsilon$$

and

$$\liminf_{i\to\infty} \frac{\Delta_{k_i}(\hat{O}_{k_i}, \hat{P}_{n_i})}{\Delta_{k_i}(O_{k_i}, P)} \ge \liminf_{i\to\infty} \frac{\Delta_{k_i}(\hat{O}_{k_i}, \hat{P}_{n_i})}{\Delta_{k_i}(\hat{O}_{k_i}, P)} \ge 1 - \epsilon.$$

Analogous statements hold if ϵ is replaced by a suitable sequence (ϵ_i) with $\epsilon_i \to 0$. Now

$$\frac{\Delta_{k_i}(O_{k_i}, \hat{P}_{n_i})}{\Delta_{k_i}(O_{k_i}, P)} - 1 \le \sum_{j=1}^{k_i} \frac{\int_{B_j} \|x - \mu_i\|^2 dP}{\Delta_{k_i}(O_{k_i}, P)} \left| \frac{\int_{B_j} \|x - \mu_i\|^2 d\hat{P}_{n_i}}{\int_{B_j} \|x - \mu_i\|^2 dP} - 1 \right|,$$

and analogously

$$\left| \frac{\Delta_{k_i}(\hat{O}_{k_i}, \hat{P}_{n_i})}{\Delta_{k_i}(\hat{O}_{k_i}, P)} - 1 \right| \le \sum_{j=1}^{k_i} \frac{\int_{\hat{B}_j} \|x - \hat{\mu}_i\|^2 dP}{\Delta_{k_i}(\hat{O}_{k_i}, P)} \left| \frac{\int_{\hat{B}_j} \|x - \hat{\mu}_i\|^2 d\hat{P}_{n_i}}{\int_{\hat{B}_j} \|x - \hat{\mu}_i\|^2 dP} - 1 \right|.$$

Let $t = t_i > 0$. How t depends on i will be made precise in the course of the proof. We define sets $\mathcal{F}_\varkappa$ of mappings including $x \mapsto \|x - \mu\|^2 I_B$, of which we will prove that the outer probabilities

$$P^* \left(\sup_{f \in \mathcal{F}_\varkappa} \left| \frac{\hat{P}_{n_i}(f)}{P(f)} - 1 \right| > t \right)$$

are summable.

$\mathcal{H}_\varkappa$ denotes the set of intersections of at most $\varkappa$ halfspaces and $Q := \{q : x \mapsto \|x - \mu\|^2 \mid \mu \in [0,1]^d\}$. We define $\mathcal{F}_\varkappa := \{q I_H \mid q \in Q, H \in \mathcal{H}_\varkappa\}$.

The graph region class of a set $\mathcal{F}$ of mappings is the set of subgraphs $GR(\mathcal{F}) = \{\{(x,t) \mid t \le f(x)\} \mid f \in \mathcal{F}\}$. A set $\mathcal{K}$ of measurable sets is called $(v, \varkappa)$-constructible, if a VC-class $\mathcal{D}$ with VC-dimension at most v and a mapping ϕ on $\mathcal{D}^\varkappa$ with range the Borel sets on $\mathbb{R}^{d+1}$, which is constructed exclusively from the set operations $\{\cap, \cup, ^c\}$, exist, such that $\mathcal{K} \subseteq \phi(\mathcal{D}^\varkappa)$. For the definition of a n-deviation measurable class see Alexander (1984).

LEMMA **1.** *$GR(\mathcal{F}_\varkappa)$ is $(d+3, 2\varkappa+1)$-constructible. It is $\mathcal{F}_\varkappa$ n-deviation measurable for any $n \geq 1$.*

PROOF. $\mathcal{H}_1$ is a VC-class, its VC-dimension is at most $d+2$. $GR(Q)$ is VC with VC-dimension at most $d+2$, since it is a subset of an at most $d+2$ dimensional vector space of functions (see Van der Vaart and Wellner (1996)). $GR(\mathcal{H}_1) \cup GR(Q)$ is therefore a VC-class of VC-dimension at most $d+3$. For $qI_{\cap_{i\leq\varkappa} H_i} \in \mathcal{F}_\varkappa$, where $H_1, \ldots, H_\varkappa$ are halfspaces, we have

$$\{(x,t) \mid t \leq q(x) I_{\cap_{i\leq\varkappa} H_i}(x)\}$$
$$= (\{(x,t) \mid t \leq q(x)\} \cap \bigcap_{i\leq\varkappa}(H_i \times \mathbb{R})) \cup \bigcup_{i\leq\varkappa}(H_i^c \times]-\infty, 0]).$$

The second statement is an immediate consequence of the definition of $\mathcal{F}_\varkappa$. □

The subsequent lemma is a modified version of Alexander (1984), Theorem 2.8. (with correction Alexander (1987)). Let

$$\psi(M,n,\alpha) = \frac{M^2/\alpha}{2(1 + M/(3n^{1/2}\alpha))},$$

(called ψ_3 in Alexander (1984). ϵ was chosen as 1/2.)

LEMMA **2.** *Let a n-deviation measurable class $\mathcal{F}$ with $(v, \varkappa)$-constructible graph region class $GR(\mathcal{F})$ be given, where $n \geq 1$. If*

A. $0 \leq f \leq 1$ and $\alpha \geq var(f)$ for all $f \in \mathcal{F}$,

B. $M \geq \alpha^{1/2}$,

C. $M \geq 2^9 \varkappa v/n^{1/2}$,

D. $n\alpha \geq v\varkappa$ and $\psi(M,n,\alpha) \geq 2^{10}\varkappa v \log(n/v)$,

then

$$P^*(\sup_{f\in\mathcal{F}} |\hat{P}_n(f) - P(f)| > Mn^{-1/2}) \leq 16 \exp(-\psi(M,n,\alpha)/2). \qquad (15)$$

We show that, given the assumptions of Theorem 1, (13) holds a.e. Let $v := d+3$ and $t > 0$. We define numbers $\varphi(i)$ and $(b_j(i))_{j\geq 0}$ as *(i)* $\varphi(i) = 1 \wedge (2^{12}v(2\varkappa_i - 1)\log(n_i/v))^{-1}$,

(ii) $b_0(i) = (2^{12}v(t^{-2}+(3t)^{-1})\log(n_i/v)(2\varkappa_i-1)k_i/n_i) \vee (v(2\varkappa_i-1)k_i/n_i) \vee (2^9vt^{-1}(2\varkappa_i - 1)k_i/n_i)$,

(iii) $b_{j+1}(i) = \varphi(i)b_j(i)^2t^2n_i/k_i$.

We define

$$\mathcal{C}_j(i) = \{f \in \mathcal{F}_{\varkappa_i} \mid P(f) \in [b_j(i)/k_i, b_{j+1}(i)/k_i[\}$$

and

$$\mathcal{C}^*(i) = \{f \in \mathcal{F}_{\varkappa_i} \mid P(f) < b_0(i)/k_i\}.$$

If $f \in \mathcal{C}_j(i)$, then $var(f) \leq P(f^2) \leq b_{j+1}(i)/k_i$. We put $\alpha_j(i) = b_{j+1}(i)/k_i$ and $\alpha^*(i) = b_0(i)/k_i$. $\varphi(i) < 1$ implies

$$\frac{b_{j+1}(i)}{b_j(i)} = \frac{\varphi(i)b_j(i)tn_i}{k_i} \geq \frac{\varphi(i)b_0(i)tn_i}{k_i}$$

$$\geq \frac{\varphi(i)(1+t/3)}{\varphi(i)} = (1+\frac{t}{3}).$$

For $j^*(i) = \max\{j \mid b_j(i)/k_i \leq 1\}$ we have $j^*(i) \leq \log(k_i/b_0(i))/\log(1+t/3) = O(n_i)$, if $1/t = o(n_i)$.

Let $M_j(i) = tb_j(i)n_i^{1/2}/k_i$ and $M^*(i) = t\Delta_{k_i}(P)n_i^{1/2}/k_i$. $\mathcal{C}_j(i)$, $M_j(i)$, $\alpha_j(i)$, k_i, n_i fulfil the assumptions of Lemma 2. An easy computation shows that $\mathcal{C}^*(i)$, $M^*(i)$, $\alpha^*(i)$, k_i, n_i fulfil the assumptions of Lemma 2, if $n_i \geq ev$ and if $\varkappa_i k_i \leq c(t)\Delta_{k_i}(P)n_i/\log(n_i/v)$, with

$$c(t) = \frac{t^2}{\sqrt{1+t/3}v2^{23/2}}.$$

This holds, in case $0 < t < 1/2$, if

$$t \geq v^{1/2}2^6\left(\frac{\varkappa_i k_i \log n_i}{\Delta_{k_i}(P)n_i}\right)^{1/2}. \qquad (16)$$

Let $0 < t < 1/2$ and $f \in \mathcal{C}_j(i)$. $|P(f)/\hat{P}_{n_i}(f) - 1| > 2t$ implies $|\hat{P}_{n_i}(f) - P(f)| > M_j(i)n^{-1/2}$. Since Lemma 2.D. holds and $\varkappa_i$ is at least 1, (16) and $n_i \geq ev$ imply

$$P\left(\left|\frac{\Delta_{k_i}(O_k,P) - \Delta_{k_i}(O_k,\hat{P}_{n_i})}{\Delta_{k_i}(O_k,P)}\right| > 3t\right)$$

$$\leq P^* \left(k_i \sup_{f \in \mathcal{C}^*(i)} |\hat{P}_{n_i}(f) - P(f)| > t\Delta_{k_i}(O_k, P) \right)$$

$$+ \sum_{j=0}^{j^*(i)} P^* \left(\sup_{f \in \mathcal{C}_j(i)} |P(f)/\hat{P}_{n_i}(f) - 1| > 2t \right)$$

$$\leq P^* \left(\sup_{f \in \mathcal{C}^*(i)} |\hat{P}_{n_i}(f) - P(f)| > M^*(i) n^{-1/2} \right) + \sum_{j=0}^{j^*(i)} P^* \left(\sup_{f \in \mathcal{C}_j(i)} |\hat{P}_{n_i}(f) - P(f)| > M_j(i) n^{-1/2} \right)$$

$$\leq 16 \exp\left(-\psi(M^*(i), n_i, \alpha^*(i))/2\right) + \sum_{j=0}^{j^*(i)} 16 \exp\left(-\psi(M_j(i), n_i, \alpha_j(i))/2\right)$$

$$\leq 16(2 + j^*(i)) \exp\left(-2^9 v \log(n_i/v)\right).$$

Analogously one may prove (14). □

REFERENCES

ALEXANDER K. (1984). Probability inequalities for empirical processes and a law of iterated logarithm. *Annals of Probability 12*, 1041-1067.

ALEXANDER K. (1987). Correction: Probability inequalities for empirical processes and a law of iterated logarithm. *Annals of Probability 15*, 428-430.

CUTLER C.D. and DAWSON D.A. (1989). Estimation of dimension for spatially distributed data and related limit theorems. *Journal of Multivariate Analysis 28*, 115-148.

LINDER T., LUGOSI G., and ZEGER K. (1994). Rates of convergence in the source coding theorem, in empirical quantizer design, and in universal lossy source coding, *IEEE Trans. Inform, Theory 40*, 1728-1740.

POLLARD D. (1981). Strong consistency of k-means clustering. *Annals of Statistics 9*, 135-140.

PÖTZELBERGER K. (1998). The quantization dimension of distributions, submitted.

PÖTZELBERGER K. and STRASSER H. (1997). Data compression by unsupervised classification, submitted.

VAN DER VAART A. und WELLNER J. (1996). *Weak Convergence and Empirical Processes.* Springer.

ZADOR P. L. (1963). *Development and Evaluation of Procedures for Quantizing Multivariate Distributions.* Ph.D. Thesis, Stanford University.

ZADOR P. L. (1982). Asymptotic quantization error of continuous signals and the quantization dimension. *IEEE Trans. Inform. Theory 28*, 139-149.