

MSP-Partitions and Unbiased Quantizations: A Review of Results

Klaus Pötzelberger

Vienna University of Economics and Business Administration

Abstract: We resume recent developments in the theory of unbiased quantizations of probability distributions. Starting with variance-minimizing partitions, we review concept such as f -information, maximum support plane partition and quantizations, and motivate the definition of unbiased quantizations. The obtained results have applications in statistical inference and in the theory of comparison of experiments.

Zusammenfassung: Wir geben eine kurze Übersicht über neuere Entwicklungen auf dem Gebiet der unverzerrten Quantisierung von Wahrscheinlichkeitsmaßen. Ausgehend von Varianz-minimierenden Partitionen erläutern wir Konzepte wie MSP-Partitionen und Quantisierungen und motivieren die Definition einer unverzerrten Quantisierung. Anwendungen der Resultate sind Fragestellungen der statistischen Inferenz und der Theorie des Vergleichs von Experimenten.

Keywords: Quantization, Partition, Bishop-de Leeuw Order, Admissibility, Information, MSP-partition.

1 Introduction

Consider a probability space (Ω, \mathcal{F}, P) . Many statistical procedures require the choice of a partition \mathcal{B} of the sample space Ω , or equivalently, the choice of a finite algebra $\mathcal{F}' \subseteq \mathcal{F}$. A related task is the the approximation of P by a distribution μ with finite support.

For instance, in descriptive statistics, quantities such as principal points in the sense of Flury (1990) or quantiles are assigned to distributions. Continuous laws are replaced by discrete laws by rounding or grouping. In cluster analysis an empirical distribution, i.e. data, is partitioned such that some measure of homogeneity is maximized within and minimized between clusters.

A procedure based on a partition of the sample space is the χ^2 -test of homogeneity. Here typically the distribution of metrically scaled random variables is replaced by a multinomial distribution. The power of the test depends on the chosen partition of the sample space.

In all these procedures the grouping of data leads to a loss of information. In the majority of cases \mathcal{B} (or \mathcal{F}_0, μ) is chosen from a specified class in order to maximize a given measure of information. Let us illustrate this for principal points. Suppose P is a distribution on \mathbb{R}^d with finite second moment. For a partition $\mathcal{B} = (B_1, \dots, B_m)$ define the conditional means

$$p_i = \int_{B_i} x dP / P(B_i). \quad (1)$$

A partition \mathcal{B} is optimal, if it minimizes

$$\int \min\{\|x - p_i\|^2 \mid 1 \leq i \leq m\} dP \quad (2)$$

among all partitions of size at most m . The conditional means (p_i) are called principal points or prototypes and the partition \mathcal{B} is called variance-minimizing. Note that a set of principal points $(p_i)_{i=1}^m$ defines a Voronoi-partition $\mathcal{B} = (B_1, \dots, B_m)$, up to boundaries, by

$$B_i \subseteq \{x \mid \|x - p_i\| \leq \|x - p_k\| \text{ for all } k \leq m\}. \quad (3)$$

Optimal partitions are self-consistent in the sense that if the prototypes are defined by (1) from \mathcal{B} and $\tilde{\mathcal{B}}$ by (3) from (p_i) , then, up to boundaries, $\tilde{\mathcal{B}} = \mathcal{B}$.

Replacing the squared norm in (2) by a function $g \circ \|\cdot\|$ of the norm leads to one type of generalizations of the concept of variance-minimizing partitions. These distance-based quantizations have been studied intensively (cf. Bock, 1974, and Zador, 1964). Graf and Luschgy (2000) provides an up to date account of these procedures.

2 MSP-Partitions

A different path leads to unbiased quantizations. Let $f(x) = \|x\|^2$. Note that the discrete distribution

$$\mu = \sum_{i=1}^m P(B_i) \delta_{p_i}, \quad (4)$$

corresponding to \mathcal{B} , maximizes $\mu(f)$, if and only if \mathcal{B} is optimal. δ_x denotes the Dirac distribution in x .

Pötzelberger and Strasser (2001) analyze general procedures of this type. P denotes a Borel distribution on \mathbb{R}^d with $\int \|x\| dP < \infty$ and $P(H) = 0$ for all hyperplanes H and $f : \mathbb{R}^d \rightarrow \mathbb{R}$ a convex P -integrable function. Let $\mathcal{B} = (B_1, \dots, B_m)$ be a partition and let the prototypes p_i and the distribution μ be given by (1) and (4). Define

$$I^f(\mathcal{B}) = \int f d\mu \quad (5)$$

and

$$I_m^f = \sup\{I^f(\mathcal{B}) \mid |\mathcal{B}| \leq m\}. \quad (6)$$

$I^f(\mathcal{B})$ is called the f -information of the partition \mathcal{B} (or equivalently of the distribution μ). The optimization problem

$$\text{maximize } I^f(\mathcal{B}) \quad \text{under } |\mathcal{B}| \leq m, \quad (7)$$

is the so-called primal problem. A partition with maximal f -information is called f -optimal. Let us define the conjugate convex function f^c by

$$f^c(a) = \sup\{\langle x, a \rangle - f(x) \mid x \in \mathbb{R}^d\} \quad (8)$$

and denote by A the domain of f^c , i.e. $A = \{a \mid f^c(a) < \infty\}$. We denote by $D(f, p)$ the set of subdifferentials at $p \in \mathbb{R}^d$, i.e. $a \in D(f, p)$, if $\langle p, a \rangle - f^c(a) = f(p)$.

The dual problem corresponding to (7) is

$$\text{maximize } F_m(a_1, \dots, a_m), \tag{9}$$

$(a_1, \dots, a_m) \in A^m$, where

$$F_m(a_1, \dots, a_m) = \int \max\{\langle x, a_i \rangle - f^c(a_i) \mid i \leq m\} dP. \tag{10}$$

Let $a_1, \dots, a_m \in A$. We call a partition $\mathcal{B} = (B_1, \dots, B_m)$ a maximum support plane partition (MSP-partition), if

$$x \in B_i \Rightarrow \langle x, a_i \rangle - f^c(a_i) = \max\{\langle x, a_k \rangle - f^c(a_k) \mid k \leq m\}. \tag{11}$$

MSP-partitions have been introduced by Bock (1992). By $\mathcal{S}(a_1, \dots, a_m)$ we denote the set of MSP-partitions satisfying (11).

Primal and dual problem are equivalent in the sense of

Theorem 1. (*Equivalence Theorem*).

$$\sup\{F_m(a_1, \dots, a_m) \mid a_i \in A\} = \sup\{I^f(\mathcal{B}) \mid |\mathcal{B}| \leq m\}. \tag{12}$$

If $(a_1, \dots, a_m) \in A^m$ is optimal for the dual problem, then all $\mathcal{B} \in \mathcal{S}(a_1, \dots, a_m)$ are f -optimal.

Let \mathcal{B} be f -optimal. If the prototypes (p_i) are defined by (1) and $a_i \in D(f, p_i)$, then (a_1, \dots, a_m) is optimal for the dual problem.

Statistical properties of procedures based on MSP-partitions depend on the choice of the convex function f . For instance, the power of tests for the one-sample or the k-sample problem depend on the distribution of the partition (see Rahnenführer, 1999). If f is the squared norm, there are always cells B_i with low probability $P(B_i)$. More precisely, let Q_m denote the uniform distribution on the prototypes p_1, \dots, p_m . Assume that P is absolutely continuous with respect to Lebesgue measure and let π denote its density. Then, given regularity conditions, (Q_m) converges weakly to a distribution with density proportional to $\pi^{d/(2+d)}$. Take for instance a d -dimensional normal distribution with mean 0 and variance Σ . Then the asymptotic distribution of the prototypes is again normal, with mean 0 and variance $(1 + 2/d)\Sigma$. Thus, prototypes tend to be larger than the observations and for large m there are many prototypes in regions “far out”, where cells have small probabilities (cf. Pötzelberger, 2000b). For a general convex function f the asymptotic distribution of the prototypes has a density proportional to $|f''|^{1/(2+d)}\pi^{d/(2+d)}$, where $|f''|$ denotes the determinant of the Hessian of f .

When the asymptotic distribution of the prototypes should be P , then $|f''| \propto \pi^2$ has to be solved. Let for instance $d = 1$ and assume that π is square-integrable. f' is then proportional to $\int_{-\infty}^x \pi^2(v)dv + c$, with $c \in \mathbb{R}$, so that $f(x)$ is asymptotically (for $|x| \rightarrow \infty$) linear.

Aspects of robustness lead to choices f with bounded f' (see Pötzelberger and Strasser, 2001). f' may be regarded as an influence function. For applications, algorithms and numerical aspects see Mazanec and Strasser (2000), Steiner (1999), or Kohonen (1984).

3 Unbiased Quantizations

Up to now optimality of partitions was considered relative to a fixed distribution P . In statistics the probability space is replaced by a statistical model and we have to study properties of quantized statistical experiments.

Let $E = (\Omega, \mathcal{F}, (P_t)_{t \in T})$ denote a statistical experiment, where (Ω, \mathcal{F}) is a measurable space and $(P_t)_{t \in T}$ a family of probability measures on (Ω, \mathcal{F}) . To avoid technical difficulties, let us assume that the parameter space T is finite. Let $m \in \mathbb{N}$, $\mathcal{B} = (B_1, \dots, B_m)$ a measurable partition of Ω and $\sigma(\mathcal{B})$ the field generated by \mathcal{B} . We call the experiment $E' = (\Omega, \sigma(\mathcal{B}), (P_t)_{t \in T})$ a quantization of E of size m . Let us denote by \supseteq the information order on the set of experiments with the same parameter set T . E is more informative than an experiment F if for all bounded loss functions and all procedures δ' for F there is a procedure δ for E with risk at most the risk of δ' . It is natural to study those finite partitions \mathcal{B} of size at most m for which the experiment $E' = (\Omega, \sigma(\mathcal{B}), (P_t)_{t \in T})$ is maximal with respect to \supseteq . We call such a partition \mathcal{B} and the field $\sigma(\mathcal{B})$ admissible.

For experiments $E = (\Omega, \mathcal{F}, (P_t)_{t \in T})$ and $F = (\Omega', \mathcal{F}', (Q_t)_{t \in T})$ an alternative characterization of $E \supseteq F$ is given by the Bishop-De Leeuw order on their standard measures. $x : \mathbb{R}^d \rightarrow \mathbb{R}^d$ denotes the identity.

Let P and Q denote Borel probability distributions on \mathbb{R}^d . P dominates Q with respect to the Bishop-De Leeuw order, $Q \preceq P$, if a stochastic kernel $(K_y)_{y \in \text{supp}(Q)}$ exists, such that

$$\int x dK_y = y \quad Q\text{-a.e.} \quad (13)$$

and $P = KQ$, i.e. for all Borel sets A

$$P(A) = \int K_y(A) dQ. \quad (14)$$

The kernel $(K_y)_{y \in \text{supp}(Q)}$ is called an unbiased kernel with respect to (Q, P) .

The standard measure σ_E of the experiment $E = (\Omega, \mathcal{F}, (P_t)_{t \in T})$ is the law of the Radon-Nikodym derivatives $(dP_t/d\bar{P})_{t \in T}$ under $\bar{P} = \sum_{t \in T} P_t/|T|$.

A main result of the theory of comparison of experiments, which motivates the following definition, is the fact that $E \supseteq F$ if and only if $\sigma_F \preceq \sigma_E$ (see Blackwell, 1951, 1953, Strasser, 1985, or Torgersen, 1991). The experiments E and F are equivalent, $E \sim F$, if $E \supseteq F$ and $F \supseteq E$.

Definition 1. (a) Let P be a Borel probability distribution on \mathbb{R}^d and $m \in \mathbb{N}$. Let us define the set of unbiased quantizations of P of complexity m as

$$\mathcal{M}(P, m) = \{\mu \mid \mu \preceq P \text{ and } |\text{supp}(\mu)| \leq m\}. \quad (15)$$

$\mu \in \mathcal{M}(P, m)$ is called admissible (maximal), if for all $\nu \in \mathcal{M}(P, m)$ with $\mu \preceq \nu$, $\mu = \nu$ holds.

(b) Let $E = (\Omega, \mathcal{F}, (P_t)_{t \in T})$ be an experiment with finite parameter space T . An experiment $F = (\tilde{\Omega}, \tilde{\mathcal{F}}, (Q_t)_{t \in T})$ is an unbiased quantization of E of complexity m , if $\sigma_F \in \mathcal{M}(\sigma_E, m)$. We denote the set of unbiased quantizations of E by $\mathcal{M}(E, m)$. $F \in \mathcal{M}(E, m)$ is admissible if it is maximal with respect to \supseteq in $\mathcal{M}(E, m)$.

Let us remark that if $|\text{supp}(P)| \geq m$ and $\mu \in \mathcal{M}(P, m)$ is admissible, then $|\text{supp}(\mu)| = m$.

Unbiased quantizations of distributions correspond to the representation of distributions by mixtures. More specifically, let $\mu = \sum_{i=1}^m w_i \delta_{p_i}$ be a distribution on \mathbb{R}^d with $|\text{supp}(\mu)| \leq m$. μ is an unbiased quantization of P if Borel distributions P_i exist such that $\int x P_i(dx) = p_i$ and $P = \sum_{i=1}^m w_i P_i$. The mixture-components P_1, \dots, P_m are parametrized by its means and μ is a distribution on the parameter.

Theorem 2 shows that a quantization of an experiment is equivalent to an experiment generated by a finite subfield of \mathcal{F} . It has been proved by Strasser (2000).

Theorem 2. $F \in \mathcal{M}(E, m)$ if and only if there is a measurable partition \mathcal{B} of Ω with $|\mathcal{B}| \leq m$ such that $F \sim (\Omega, \sigma(\mathcal{B}), (P_t)_{t \in T})$.

The Theorem of Blackwell-Sherman-Stein (see Torgersen, 1991, or Strassen, 1965) provides the fundamental characterization of the Bishop-De Leeuw order: $Q \preceq P$ if and only if for all convex and continuous functions f

$$\int f dQ \leq \int f dP. \quad (16)$$

Therefore, intuitively speaking, maximizing an information measure of the form $\int f d\mu$ on $\mathcal{M}(P, m)$ will lead to admissible quantizations, at least, when the maximizing quantization is unique. It can be shown that all admissible quantizations are essentially generated by MSP-partitions.

Theorems 3 - 6 summarize the essential results on admissible elements of $\mathcal{M}(P, m)$: Existence of admissible quantizations, admissibility of f -optimal quantizations, characterization of admissible quantizations and geometric properties of corresponding partitions. The proofs of these results are provided in Pötzelberger (2002). We assume that P is a Borel probability measure on \mathbb{R}^d with finite expectation and such that $P(H) = 0$ for all hyperplanes $H \subseteq \mathbb{R}^d$.

Before we can state the results we have to introduce some notation. A partition $\mathcal{B} = (B_1, \dots, B_m)$ of \mathbb{R}^d , where all B_i are convex, is called a polytopepartition (PT-partition). $\mathcal{B} = (B_1, \dots, B_m)$ is called MSP-partition (cf. section 2), if there are linear functions g_1, \dots, g_m , such that

$$B_i \subseteq \{x \mid g_i(x) \geq g_j(x) \text{ for all } j\}. \quad (17)$$

In this case \mathcal{B} is generated by g , where $g(x) = \max\{g_i(x) \mid i \leq m\}$. L_m denotes the class of convex functions, which are affine on m sets, i.e. $g \in L_m$ if there are affine functions g_1, \dots, g_m such that

$$g(x) = \max\{g_i(x) \mid i \leq m\}. \quad (18)$$

Note that PT-partitions consist of sets with boundaries that are subsets of hyperplanes and therefore nullsets. We do not distinguish between partitions which are identical up to boundaries.

In the one-dimensional case all PT-partitions are MSP-partitions. They consist of intervals. If $\mathcal{B} = (B_1, \dots, B_m)$ with $B_i =]a_{i-1}, a_i]$ for $i < m$, $B_m =]a_{m-1}, \infty[$ and $a_0 < \dots < a_{m-1}$, then \mathcal{B} is generated by a function $g \in L_m$. Indeed, let $b_1 < \dots < b_m$

and let g be continuous on \mathbb{R} and linear on B_i with $g' = b_i$ on $\overset{\circ}{B}_i$. g is convex and generates \mathcal{B} . For $d > 1$ the class of MSP-partitions is a proper subclass of the class of polytopepartitions, see Pötzelberger (2002).

Let a partition \mathcal{B} be given. We denote the quantization defined by (1) and (4) by $\mu^{\mathcal{B}}$. If $g \in L_m$ and \mathcal{B} is generated by g , then we abbreviate $\mu^{\mathcal{B}}$ by μ^g . Let f be convex and P -integrable. We define $\mathcal{O}_f = \{\mu \in \mathcal{M}(P, m) \mid \int f d\mu = I_m^f\}$. $\mu \in \mathcal{O}_f$ is called f -optimal. Note that in general $|\mathcal{O}_f| > 1$. If $f \in L_m \setminus L_{m-1}$, then $|\mathcal{O}_f| = 1$ and thus $\mathcal{O}_f = \{\mu^f\}$. Let us mention an implication of the Equivalence Theorem 1: If f is convex and P -integrable, $f \notin L_{m-1}$ and $\mu \in \mathcal{O}_f$, then a $g \in L_m$ exists, such that $g \leq f$, $\mu = \mu^g$ and $\int f d\mu = \int g d\mu = \int g dP$. The MSP-partition generated by such a g is an f -optimal partition.

Theorem 3. *For every admissible $\mu \in \mathcal{M}(P, m)$ there is a polytopepartition \mathcal{B} of size at most m , such that $\mu = \mu^{\mathcal{B}}$.*

Theorem 4. *For every $\mu \in \mathcal{M}(P, m)$ there is an admissible $\nu \in \mathcal{M}(P, m)$ such that $\mu \preceq \nu$.*

Corollary 1. *For every $\mu \in \mathcal{M}(P, m)$ there is a polytopepartition \mathcal{B} of size at most m , such that $\mu \preceq \mu^{\mathcal{B}}$.*

Theorem 5. *Suppose g is a P -integrable convex function with $g \notin L_{m-1}$. Then all $\mu \in \mathcal{O}_g$ are admissible.*

Theorem 6. *Let $\mu \in \mathcal{M}(P, m)$ with $|\text{supp}(\mu)| = m$. μ is admissible if and only if there exists a sequence of convex functions $(g_n)_{n=1}^{\infty} \subseteq L_m \setminus L_{m-1}$, such that $\mu = \lim_{n \rightarrow \infty} \mu^{g_n}$.*

Theorem 3 and Theorem 5 point out the close connection between admissible quantizations and partitions with special geometric properties. Maximizing an information measure $I^g(\mathcal{B})$ with convex $g \notin L_{m-1}$ leads to an admissible quantization which is generated by a MSP-partition. On the other hand, Theorem 3 implies that all admissible $\mu \in \mathcal{M}(P, m)$ come from PT-partitions. Thus for $d = 1$ Theorem 3 and Theorem 5 provide a complete characterization of admissible quantizations as any PT-partition is a MSP-partition. However, for $d > 1$ the situation is different. We have

$$\{\mu^g \mid \mu \in L_m \setminus L_{m-1}\} \not\subseteq \{\mu \mid \mu \text{ admissible in } \mathcal{M}(P, m)\} \not\subseteq \{\mu^{\mathcal{B}} \mid \mathcal{B} \text{ a PT-partition}\}.$$

Since $\{\mu^g \mid \mu \in L_m \setminus L_{m-1}\}$ is dense in $\{\mu \mid \mu \text{ admissible in } \mathcal{M}(P, m)\}$, in principle any admissible μ may be approximated by a suitable f -optimal quantization. Note that even in the light of the Theorem of Blackwell-Sherman-Stein this proposition is not trivial. If there are hyperplanes H with $P(H) > 0$, then Theorem 5 remains valid. However, Theorem 3 and Theorem 6 do not hold, see Pötzelberger (2000a). In particular, unbiased quantizations μ and nontrivial convex functions $(g_n) \in L_m \setminus L_{m-1}$ may exist, such that μ is not admissible, although $\mu = \lim_{n \rightarrow \infty} \mu_n$ with $\mu_n \in \mathcal{O}_{g_n}$.

Let us emphasize an important consequence of Theorem 4 and Theorem 1. All methods that lead to a partition of the sample space which is not the limit of MSP-partitions are inadmissible. This observation applies in particular to various distance-based generalizations of principal points or k-means clustering. All methods that lead to a disintegration

$P = \sum_{i=1}^m w_i P_i$ with overlapping supports of the distributions P_i are inadmissible. All these procedures are dominated by procedures based on admissible quantizations.

Finally, we briefly discuss the structure of admissible quantizations of experiments $E = (\Omega, \mathcal{F}, (P_t)_{t \in T})$. By Theorem 6 admissible quantizations are limits of certain g -optimal quantizations. Let us describe g -optimal quantizations, where $g \in L_m \setminus L_{m-1}$.

We may assume that the parameter space T is the set $\{1, \dots, d\}$. The experiment is dominated. Let Q denote a Borel measure with $P_t \ll Q$ for all $t \in T$ with derivatives φ_t . We have to assume that hyperplanes are nullsets. More precisely, we assume that for all $a_1, \dots, a_d \in \mathbb{R}$,

$$Q(\{\omega \mid \sum_{t=1}^d a_t \varphi_t(\omega) = 0\}) = 0. \quad (19)$$

If the partition $\mathcal{B} = (B_1, \dots, B_m)$ is g -optimal, then there are $\alpha_{i,t} \in \mathbb{R}$, $i = 1, \dots, m$, $t = 1, \dots, d$, such that

$$\omega \in B_i \Rightarrow \sum_{t=1}^d \alpha_{i,t} \varphi_t(\omega) = \max\left\{\sum_{t=1}^d \alpha_{k,t} \varphi_t(\omega) \mid k \leq m\right\}. \quad (20)$$

W.l.g. we may assume that all $\alpha_{k,t} \geq 0$. It is easy to see that the solution (20) is equivalent to the following Bayesian discrimination problem. Let distributions P_t with densities φ_t be given. Let densities $\psi_i = \sum_{j=1}^d \beta_{i,j} \varphi_j$ correspond to hypotheses H_1, \dots, H_m . Given an observation ω decide which distribution ψ_i generated ω . Let $U(i, t)$ denote the utility of decision t if H_i is the true. Furthermore, let $\theta \in T$ with $P(\theta = j) = \pi_j = \sum_{i=1}^m \beta_{i,j} P(H_i)$. $t \in T$ is chosen if it maximizes the expected posterior utility, which is proportional to

$$\begin{aligned} \sum_{i=1}^m U(i, t) P(\omega \mid H_i) P(H_i) &= \sum_{i=1}^m \sum_{j=1}^d U(i, t) \beta_{i,j} \varphi_j(\omega) P(H_i) \\ &= \sum_{j=1}^d \tilde{U}(j, t) \pi_j \varphi_j(\omega), \end{aligned}$$

where

$$\tilde{U}(j, t) = \sum_{i=1}^m U(i, t) \beta_{i,j} P(H_i) / \pi_j \quad (21)$$

is the utility of decision t if $\theta = j$ holds. $\alpha_{j,t}$ corresponds to $\tilde{U}(j, t) \pi_j$. We conclude that g -optimal partitions correspond to Bayesian discriminant problems. They are admissible, all admissible partitions are limits of them.

References

- D. Blackwell. Comparison of experiments. In *Proc. 2nd Berkeley Symp. Math. Statistics Prob.*, pages 93–102, 1951.
- D. Blackwell. Equivalent comparisons of experiments. *Ann. Math. Statistics*, 24:265–272, 1953.
- H.-H. Bock. *Automatische Klassifikation*. Vandenhoeck und Ruprecht, 1974.
- H.-H. Bock. A clustering technique for maximizing ϕ -divergence, noncentrality and discriminating power. In *Analyzing and Modeling Data and Knowledge*, pages 19–36, Heidelberg, 1992. Springer-Verlag.
- B.A. Flury. Principal points. *Biometrika*, 77:33–41, 1990.
- S. Graf and H. Luschgy. Foundations of quantization for probability distributions. volume 1730 of *Lecture Notes in Math*. Springer-Verlag, Heidelberg, 2000.
- T. Kohonen. *Self-organization and associative memory*. Springer-Verlag, New York, 1984.
- J.A. Mazanec and H. Strasser. A nonparametric approach to perceptions-based market segmentation: Foundations. In *Interdisciplinary Studies in Economics and Management I*. Springer-Verlag, Heidelberg, 2000.
- K. Pötzelberger. Admissible unbiased quantizations: Distributions with linear components, 2000a. Submitted.
- K. Pötzelberger. The general quantization problem for distributions with regular support. *Math. Methods of Statistics*, 9:176–198, 2000b.
- K. Pötzelberger. Admissible unbiased quantizations: Distributions with linear components. *Math. Methods of Statistics*, 2002. To appear.
- K. Pötzelberger and H. Strasser. Clustering and quantization by msp-partitions. *Statistics and Decisions*, 19:331–371, 2001.
- J. Rahnenführer. Multivariate permutation tests for the k-sample problem with clustered data, 2000. Submitted.
- G. Steiner. *Quantization and clustering with maximal information: Algorithms and numerical experiments*. PhD thesis, Vienna University of Economics and Business Administration, Austria, 1999.
- V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36:423–439, 1965.
- H. Strasser. *Mathematical Theory of Statistics*. De Gruyter, Berlin, 1985.

H. Strasser. Towards a statistical theory of optimal quantization. In W. Gaul, O. Opitz, and M. Schader, editors, *Data Analysis: Scientific Modeling and Practical Application*, pages 369–383, Heidelberg, 2000. Springer-Verlag.

E. Torgersen. *Comparison of statistical experiments*. Cambridge University Press, Cambridge, 1991.

P.L. Zador. *Development and evaluation of procedures for quantizing multivariate distributions*. PhD thesis, Stanford University, 1964.

Author's address:

Klaus Pötzelberger
Department of Statistics
Vienna University of Economics and Administration Administration
Augasse 2-6, A-1090 Vienna
Austria

Tel. +43 1 31336 5052

Fax +43 1 31336 734

E-mail: Klaus.Poetzelberger@wu-wien.ac.at