Decision-Theoretic Properties of Partitioned Sample Spaces

Klaus Pötzelberger
Vienna University of Economics and Business Administration
Augasse 2-6
Vienna, Austria
Klaus.Poetzelberger@wu-wien.ac.at

Let $E = (\Omega, \mathcal{F}, (P_t)_{t \in T})$ denote an experiment. The complexity of the experiment may be reduced by replacing $\mathcal{F}$ by a finite field $\tilde{\mathcal{F}} \subseteq \mathcal{F}$, which leads to the experiment $F = (\Omega, \tilde{\mathcal{F}}, (P_t)_{t \in T})$. $\tilde{\mathcal{F}}$ is identified with a finite partition $\mathcal{B} = (B_1, \ldots, B_m)$ of Ω. Recall the concept of information for experiments. A reduced experiments which is maximal with respect to the information semiorder on the set of all reduced experiments with the size of the corresponding partition being at most m is called admissible. We give characterizations of admissible experiments or of the corresponding field $\tilde{\mathcal{F}}$.

The admissiblility of subfields is directly connected with the following problem. Let P be a Borel probability measure on $\mathbb{R}^d$ or on a suitable Banach space. Characterize the maximal elements $\mu \in \mathcal{M}(P, m)$ with respect to the Bishop-De Leeuw order $\preceq$, where $\mu \in \mathcal{M}(P, m)$ if and only if $\mu \preceq P$ and $|\text{supp}(\mu)| \leq m$.

Let us briefly provide examples from statistical applications to see the interplay between partitions and discrete distributions μ approximating P.

For instance, in descriptive statistics, quantities such as principal points 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. The clusters are represented by their centroid.

Procedures based on a partition of the sample space are common in inference statistics. Think of the χ^2-test of homonegeity. Here typically the likelihood is quantized, i.e. the law of a metrically scaled random variable is replaced by a multinomial distribution. The power of the test depends on the chosen partition of the sample space. See Bock (1992) for details.

In all these procedures the grouping of data leads to a loss of information. In the majority of cases μ is chosen from a specified class of distributions in order to maximize a given measure of information. Let us illustrate this for principal points. For a partition $\mathcal{B} = (B_1, \ldots, B_m)$ define the conditional means

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

Let f denote a convex function. A partition $\mathcal{B}$ is f-optimal, if it maximizes the information measure

$$\sum_{i=1}^{m} f(p_i)P(B_i) \tag{2}$$

among all partitions of size at most m. Note that the discrete distribution

$$\mu = \sum_{i=1}^{m} P(B_i)\delta_{p_i}, \tag{3}$$

corresponding to $\mathcal{B}$, maximizes $\mu(f)$ if and only if $\mathcal{B}$ is f-optimal.

Let us state the main result (see Pötzelberger (2000), Pötzelberger and Strasser (2001)). Given regularity conditions, a $\mu \in \mathcal{M}(P,m)$ is maximal if and only if $\mu = \lim_{n\to\infty} \mu_n$, where μ_n is f_n-optimal for a suitable convex f_n.

REFERENCES

Bock, H. H. (1992). A clustering technique for maximizing ϕ-divergence, noncentrality and discriminating power. In M. Schrader, editor, *Analyzing and Modeling Data and Knowledge. Springer Verlag.* 19-36.

Pötzelberger, K. (2000). Admissible unbiased quantizations: Distributions without linear components. Submitted.

Pötzelberger, K. and Strasser, H. (2001). Clustering and quantization by MSP-Partitions. To appear in: *Statistics and Decisions.*

RESUME

Soit P une mesure de probabilité borélienne sur $\mathbb{R}^d$. Notre objectiv principal est le caracterisation des elements maximales $\mu \in \mathcal{M}(P,m)$ dans le ordre de Bishop-De Leeuw $\preceq$. $\mathcal{M}(P,m) = \{\mu \preceq P \mid |\text{supp}(\mu)| \leq m\}$ est l'ensemble des quantizations de P.

Ce résultat mène a la caracterisation des partitions admissibles. Une partition $\mathcal{B} = (B_1,\ldots,B_m)$ de $\mathcal{F}$ est admissible pour l'experiment $(\Omega,\mathcal{F},(P_t)_{t\in T})$ si l'experiment $(\Omega,\sigma(\mathcal{B}),(P_t)_{t\in T})$ est maximale parmi les experiments générés par des partitions $\mathcal{B}'$ de $\mathcal{F}$ avec $|\mathcal{B}'| \leq m$.