A Representation Theorem and Applications to Measure Selection and Noninformative Priors

Manfred Jaeger

MPI–I–2003–2-002 June 2003

Author's Address

Max-Planck-Institut für Informatik
Stuhlsatzenhausweg 85
66123 Saarbrücken
Germany
Phone: +49 681 9325-216
Fax: +49 681 9325-299
Email: `jaeger@mpi-sb.mpg.de`

Publication Notes

This report is the extended version of a paper to be presented at the *Seventh European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU-2003)*.

Abstract

We introduce a set of transformations on the set of all probability distributions over a finite state space, and show that these transformations are the only ones that preserve certain elementary probabilistic relationships. This result provides a new perspective on a variety of probabilistic inference problems in which invariance considerations play a role. Two particular applications we consider in this paper are the development of an equivariance-based approach to the problem of measure selection, and a new justification for Haldane's prior as the distribution that encodes prior ignorance about the parameter of a multinomial distribution.

Keywords

Probabilistic Reasoning, Invariance, Prior Distributions

1 Introduction

Many rationality principles for probabilistic and statistical inference are based on considerations of indifference and symmetry. An early expression of such a principle is Laplace's principle of insufficient reason: *"One regards two events as equally probable when one can see no reason that would make one more probable than the other, because, even though there is an unequal possibility between them, we know not which way, and this uncertainty makes us look on each as if it were as probable as the other"*(Laplace, Collected Works vol. VIII, cited after (Hacking 1975)). Principles of indifference only lead to straightforward rules for probability assessments when the task is to assign probabilities to a finite number of different alternatives, none of which is distinguished from the others by any information we have. In this case all alternatives will have to be assigned equal probabilities. Such a formalization of indifference by equiprobability becomes notoriously problematic when from state spaces of finitely many alternatives we turn to infinite state spaces: on countably infinite sets no uniform probability distributions exist, and on uncountably infinite sets the concept of uniformity becomes ambiguous (as evidenced by the famous Bertrand's paradox (Holbrook & Kim 2000, van Fraassen 1989)).

On (uncountably) infinite state spaces concepts of uniformity or indifference have to be formalized on the basis of certain transformations of the state space: two sets of states are to be considered equiprobable, if one can be transformed into the other using some natural transformation t. This, of course, raises the sticky question what transformations are to be considered as natural and probability-preserving. However, for a given state space, and a given class of probabilistic inference tasks, it often is possible to identify natural transformation, so that the solution to the inference tasks (which, in particular, can be probability assessments) should be invariant under the transformations. The widely accepted resolution of Bertrand's paradox, for example, is based on such considerations of invariance under certain transformations.

In this paper we are concerned with probabilistic inference problems that pertain to probability distributions on finite state spaces, which are by far the most widely used type of distributions used for probabilistic modelling in artificial intelligence. As indicated above, when dealing with finite state spaces there does not seem to be any problem of capturing indifference principles with equiprobability. However, even though the underlying space of alternatives may be finite, the object of our study very often is the infinite set of probability distributions on that space, i.e. for the state space $S = \{s_1, \ldots, s_n\}$ the $(n-1)$-dimensional probability polytope

$$\Delta^n := \{(p_1, \ldots, p_n) \in \mathbb{R}^n \mid p_i \in [0,1], \sum_i p_i = 1\}.$$

The objective of this paper now can be formulated as follows: we investigate what natural transformations there exist of Δ^n, such that inference problems that pertain to Δ^n should be solved in a way that is invariant under these transformations. In section 2 we identify a unique class of transformations that can be regarded as most natural in that they alone preserve certain relevant relationships between points of Δ^n. In sections 3 and 4 we apply this result to the problems of noninformative priors and measure selection, respectively.

2 Representation Theorem

The nature of the result we present in this section can best be explained by an analogy: suppose, for the sake of the argument, that the set of probability distributions we are concerned with is parameterized by the whole Euclidean space $\mathbb{R}^n$, rather than the subset Δ^n. Suppose, too, that all inputs and outputs for a given type of inference problem consist of objects (e.g. points, convex subsets, $\ldots$) in $\mathbb{R}^n$. In most cases, one would then probably require of a rational solution to the inference problem that it does not depend on the choice of the coordinate system. Specifically, if all inputs are transformed by a translation, i.e. by adding some constant offset $\boldsymbol{r} \in \mathbb{R}^n$, then the outputs computed for the transformed inputs should be just the outputs computed for the original inputs, also translated by $\boldsymbol{r}$:

$$\mathit{sol}(\boldsymbol{i}+\boldsymbol{r}) = \mathit{sol}(\boldsymbol{i}) + \boldsymbol{r}, \tag{1}$$

where $\boldsymbol{i}$ stands for the inputs and *sol* for the solution of an inference problem. Condition (1) expresses an *equivariance principle*: when the problem is transformed in a certain way, then so should be its solution (not to be confused with *invariance principles* according to which certain things should be unaffected by a transformation).

The question we now address is the following: what simple, canonical transformations of the set Δ^n exist, so that for inference problems whose inputs and outputs are objects in Δ^n one would require an equivariance property analogous to (1)? Intuitively, we are looking for transformations of Δ^n that can be seen as merely a change of coordinate system, and that leave all relevant geometric structures intact. The following definition collects some key concepts we will use.

Definition 2.1 A *transformation* of a set S is any bijective mapping t of S onto itself. We often write ts rather than $t(s)$. For a probability distribution $\boldsymbol{p} = (p_1, \ldots, p_n) \in \Delta^n$ the set $\{i \in \{1, \ldots, n\} \mid p_i > 0\}$ is called the *set of support* of $\boldsymbol{p}$, denoted *support*$(\boldsymbol{p})$. A transformation t of Δ^n is said to

- *preserve cardinalities of support* if for all $\boldsymbol{p}$: $|\mathit{support}(\boldsymbol{p})| = |\mathit{support}(t\boldsymbol{p})|$
- *preserve sets of support* if for all $\boldsymbol{p}$: $\mathit{support}(\boldsymbol{p}) = \mathit{support}(t\boldsymbol{p})$.

A distribution $\boldsymbol{p}$ is called a *mixture* of $\boldsymbol{p}'$ and $\boldsymbol{p}''$ if there exists $\lambda \in [0,1]$ such that $\boldsymbol{p} = \lambda\boldsymbol{p}' + (1-\lambda)\boldsymbol{p}''$ (in other words, $\boldsymbol{p}$ is a convex combination of $\boldsymbol{p}'$ and $\boldsymbol{p}''$). A transformation t is said to

- *preserve mixtures* if for all $\boldsymbol{p}, \boldsymbol{p}', \boldsymbol{p}''$: if $\boldsymbol{p}$ is a mixture of $\boldsymbol{p}'$ and $\boldsymbol{p}''$, then $t\boldsymbol{p}$ is a mixture of $t\boldsymbol{p}'$ and $t\boldsymbol{p}''$.

The set of support of a distribution $\boldsymbol{p} \in \Delta^n$ can be seen as its most fundamental feature: it identifies the subset of states that are to be considered as possible at all, and thus identifies the relevant state space (as opposed to the formal state space S, which may contain states s_i that are effectively ruled out by $\boldsymbol{p}$ with $p_i = 0$). When the association of the components of a distribution $\boldsymbol{p}$ with the elements of the state space $S = \{s_1, \ldots, s_n\}$ is fixed, then $\boldsymbol{p}$ and $\boldsymbol{p}'$ with different sets of support represent completely incompatible probabilistic models that would not be transformed into one another by a natural transformation. In this case, therefore, one would require a transformation to preserve sets of support.

A *permutation* of Δ^n is a transformation that maps $(p_1, \ldots, p_n)$ to $(p_{\pi(1)}, \ldots, p_{\pi(n)})$, where π is a permutation of $\{1, \ldots, n\}$. Permutations preserve cardinalities of support, but not sets of support. Permutations of Δ^n are transformations that are required to preserve the semantics of the elements of Δ^n after a reordering of the state space S: if S is reordered according to a permutation π, then $\boldsymbol{p}$ and $\pi\boldsymbol{p}$ are the same probability distribution on S. Apart from this particular need for permutations, they do not seem to have any role as a meaningful transformation of Δ^n.

That a distribution $\boldsymbol{p}$ is a mixture of $\boldsymbol{p}'$ and $\boldsymbol{p}''$ is an elementary probabilistic relation between the three distributions. It expresses the fact that the probabilistic model $\boldsymbol{p}$ can arise as an approximation to a finer model that would distinguish the two distinct distributions $\boldsymbol{p}'$ and $\boldsymbol{p}''$ on S, each of which is appropriate in a separate context. For instance, $\boldsymbol{p}'$ and $\boldsymbol{p}''$ might be the distributions on $S = \{$*jam*, *heavy traffic*, *light traffic*$\}$ that represent the travel conditions on weekdays and weekends, respectively. A mixture of the two then will represent the probabilities of travel conditions when no distinction is made between the different days of the week.

That a transformation preserves mixtures, thus, is a natural requirement that it does not destroy elementary probabilistic relationships. Note that we do not require that t preserves the mixture coefficient: when $\boldsymbol{p} = \lambda\boldsymbol{p}' + (1-\lambda)\boldsymbol{p}''$ then usually we will have $t\boldsymbol{p} = \kappa t\boldsymbol{p}' + (1-\kappa)t\boldsymbol{p}''$ with $\kappa \neq \lambda$. In fact, it is easy to see that only the identity function preserves both sets of supports and mixtures, such that the mixture coefficient is unchanged.

We now introduce the class of transformations that we will be concerned with in the rest of this paper. We denote with $\mathbb{R}^+$ the set of positive real numbers.

Definition 2.2 Let $\boldsymbol{r} = (r_1, \ldots, r_n) \in (\mathbb{R}^+)^n$. Define for $\boldsymbol{p} = (p_1, \ldots, p_n) \in \Delta^n$

$$t_{\boldsymbol{r}}(\boldsymbol{p}) := (r_1 p_1, \ldots, r_n p_n) / \sum_{i=1}^{n} r_i p_i.$$

Also let $T_n := \{t_{\boldsymbol{r}} \mid \boldsymbol{r} \in (\mathbb{R}^+)^n\}$.

Note that we have $t_{\boldsymbol{r}} = t_{\boldsymbol{r}'}$ if $\boldsymbol{r}'$ is obtained from $\boldsymbol{r}$ by multiplying each component with a constant $a > 0$. We can now formulate our main result.

Theorem 2.3 Let $n \geq 3$ and t be a transformation of Δ^n.

(i) t preserves sets of support and mixtures iff $t \in T_n$.

(ii) t preserves cardinalities of support and mixtures iff $t = t' \circ \pi$ for some permutation π and some $t' \in T_n$.

The statements (i) and (ii) do not hold for $n = 2$: Δ^2 is just the interval $[0, 1]$, and every monotone bijection of $[0, 1]$ satisfies (i) and (ii). A weak form of a dual version of this theorem was already reported in (Jaeger 2001). In appendix A we give the dual version in a strong form corresponding to Theorem 2.3. The proof of the theorem is given in appendix B. The following example illustrates how transformations $t \in T_n$ can arise in practice.

Example 2.4 In a study of commuter habits it is undertaken to estimate the relative use of buses, private cars and bicycles as a means of transportation. To this end, a group of research assistants is sent out one day to perform a traffic count on a number of main roads into the city. They are given count sheets and short written instructions. Two different sets of instructions were produced in the preparation phase of the study: the first set advised the assistants to make one mark for every bus, car, and bicycle, respectively, in the appropriate column of the count sheet. The second (more challenging) set of instructions specified to make as many marks as there are actually people travelling in (respectively on) the observed vehicles. By accident, some of the assistants were handed instructions of the first kind, others those of the second kind.

Assume that on all roads being watched in the study, the average number of people travelling in a bus, car, or on a bicycle is the same, e.g. 10, 1.5, and 1.01, respectively. Also assume that the number of vehicles observed on each road is so large, that the actually observed numbers are very close to these averages.

Suppose, now, that we are more interested in the relative frequency of bus, car and bicycle use, rather than in absolute counts. Suppose, too, that we prefer the numbers that would have been produced by the use of the second set of instructions. If, then, an assistant hands in counts that were produced using the first set of instructions, and that show frequencies $\boldsymbol{f} = (f_1, f_2, f_3) \in \Delta^3$ for the three modes of transportation, then we obtain the frequencies we really want by applying the transformation $t_{\boldsymbol{r}}$ with $\boldsymbol{r} = (10, 1.5, 1.01)$. Conversely, if we prefer the first set of instructions, and are given frequencies generated by the second, we can transform them using $\boldsymbol{r}' = (1/10, 1/1.5, 1/1.01)$.

This example gives rise to a more general interpretation of transformations in T_n as analogues in discrete settings to rescalings, or changes of units of measurements, in a domain of continuous observables.

3 Noninformative Priors

Bayesian statistical inference requires that a prior probability distribution is specified on the set of parameters that determines a particular probability model. Herein lies the advantage of Bayesian methods, because this prior can encode domain knowledge that one has obtained before any data was observed. Often, however, one would like to choose a prior distribution that represents the absence of any knowledge: an ignorant or noninformative prior. The set Δ^n is the parameter set for the multinomial probability model (for a fixed sample size). The question of what distribution on Δ^n represents a state of ignorance about this model has received much attention, but no conclusive answer seems to exist.

Three possible solutions that most often are considered are: the uniform distribution, i.e. the distribution that has a constant density c with respect to Lebesgue measure, Jeffreys' prior, which is given by the density $c\prod_i p_i^{-1/2}$ (where c is a suitable normalizing constant), and Haldane's prior, given by density $\prod_i p_i^{-1}$. Haldane's prior (so named because it seems to have first been suggested in (Haldane 1932)) is an improper prior, i.e. it has an infinite integral over Δ^n. All three distributions are Dirichlet distributions with parameters $(1,\ldots,1), (1/2,\ldots,1/2)$, and $(0,\ldots,0)$, respectively (in the case of Haldane's distribution, the usual definition of a Dirichlet distribution has to be extended so as to allow the parameters $(0,\ldots,0)$). Schafer (1997) considers all Dirichlet distributions with parameters $(\alpha,\ldots,\alpha)$ for $0 \leq \alpha \leq 1$ as possible candidates for a noninformative prior.

The justifications for identifying any particular distribution as the appropriate noninformative prior are typically based on invariance arguments: generally speaking, ignorance is argued to be invariant under certain problem transformations, and so the noninformative prior should be invariant under such problem transformations. There are different types of problem transformations one can consider, each leading to a different concept of invariance, and often leading to different results as to what constitutes a noninformative prior (see (Hartigan 1964) for a systematic overview). In particular, there exist strong invariance-based arguments both for Jeffreys' prior (Jeffreys 1961), and for Haldane's prior (Jaynes 1968, Villegas 1977). Novick and Hall (1965) derive Haldane's prior by a different type of argument. Skilling (1985), on the other hand, rejects Haldane's prior because it remains improper when updated by unreliable observations. In the following, we present additional invariance-based arguments in support of Haldane's prior.

Example 3.1 (continuation of example 2.4) Assume that the true, long-term relative frequencies of bus, car, and bicycle use are the same on all roads at which the traffic count is conducted (under both counting methods). Then the counts obtained in the study are multinomial samples determined by a parameter $\boldsymbol{f}_1^* \in \Delta^3$ if the first set of instructions is used, and $\boldsymbol{f}_2^* \in \Delta^3$ if the second set of instructions is used. Suppose the project leader, before seeing any counts, feels completely unable to make any predictions on the results of the counts, i.e. he is completely ignorant about the parameters $\boldsymbol{f}_i^*$.

When the samples are large (i.e. a great number of vehicles are observed on every road), then the observed frequencies $\boldsymbol{f}$ obtained using instructions of type i are expected to be very close to the true parameter $\boldsymbol{f}_i^*$. The prior probability *Pr* assigned to a subset $A \subseteq \Delta^n$ then can be identified with a prior expectation of finding in the actual counts relative frequencies $\boldsymbol{f} \in A$. If this prior expectation is to express complete ignorance, then it must be the same for both sampling methods: being told by the first assistant returning with his counts that he had been using instructions of type 2 will have no influence on the project leader's expectations regarding the frequencies on this assistant's count sheet. In particular, merely seeing the counts handed in by this assistant will give the project leader no clue as to which instructions were used by this assistant.

The parameters $\boldsymbol{f}_i^*$ are related by $\boldsymbol{f}_2^* = t_{\boldsymbol{r}}\boldsymbol{f}_1^*$, where $t_{\boldsymbol{r}}$ is as in example 2.4. Having the same prior belief about $\boldsymbol{f}_2^*$ as about $\boldsymbol{f}_1^*$ means that for every $A \subseteq \Delta^3$ one has $Pr(A) = Pr(t_{\boldsymbol{r}}A)$. A noninformative prior, thus, should be invariant under the transformation $t_{\boldsymbol{r}}$. As the relation between $\boldsymbol{f}_1^*$ and $\boldsymbol{f}_2^*$ might also be given by some other transformation in T_n, this invariance should actually hold for all these transformations.

This example provides one intuitive justification for requiring noninformative priors to be invariant under T_n-transforms. The next theorem states that this invariance property only holds for Haldane's prior. In the formulation of the theorem a little care has to be taken in dealing with the boundary of Δ^n, where the density of Haldane's prior is not defined. We therefore restrict the statement of the theorem to the prior on the interior of Δ^n, denoted *int*Δ^n.

Theorem 3.2 Let *Pr* be a measure on *int*Δ^n with $Pr(int\Delta^n) > 0$ and $Pr(A) < \infty$ for all compact subsets A of *int*Δ^n. *Pr* is invariant under all transformations $t_{\boldsymbol{r}} \in T_n$ iff *Pr* has a density with respect to Lebesgue measure of the form $c \prod_i p_i^{-1}$ with some constant $c > 0$.

It is instructive to compare the justification given to Haldane's prior by this theorem with the justification given by Jaynes (1968). Jaynes bases his justification on an intuitive interpretation of a noninformative prior as a distribution of beliefs about the true value of $\boldsymbol{p}$ that one would find in "a population in a state of total confusion": according to this interpretation one assumes that there exists a population I of individuals i, and each individual believes the value of a binomial parameter θ to be $\theta_i \in [0, 1]$ (Jaynes only considers the binomial case, and we here adopt his notation, where $(\theta, 1-\theta)$ corresponds to (p_0, p_1) in our notation). The distribution of beliefs in the population I, thus, gives rise to a density $f(\theta)$ on $[0, 1]$. This density can be interpreted as a noninformative prior when the individuals $i \in I$ base their beliefs on "different and conflicting information", and, thus, the population as a whole is in a state of "total confusion".

Jaynes's argument then is that such a state of total confusion will remain to be the same when some piece of evidence E is given to all individuals, and each individual updates his or her beliefs by conditioning on E. By a suitable formalization of this scenario, Jaynes shows that a single individual's transition from an original belief θ to the new belief θ' is given by

$$\theta' \mapsto a\theta/(1 - \theta + a\theta). \tag{2}$$

This can easily be seen as a transformation from our group T_2. The assumption of a collective state of ignorance being invariant under assimilation of the evidence E, thus, leads to the condition of invariance of f under the transformation (2). Jaynes then proceeds to show that only Haldane's prior is invariant under these transformations (which is the special case $n = 2$ of our Theorem 3.2).

Jaynes's justification, thus derives the transformation group T_2 from a concrete scenario in which it seems intuitively reasonable to argue that a noninformative prior should be invariant under these transformations. This is similar to our argument for the invariance of a noninformative prior under the transformation $t_{\boldsymbol{r}}$ in example 3.1. Justifications of Haldane's (or any other) prior that are based on such specific scenarios, however, always leave the possibility open that similarly intuitive scenarios can be constructed which lead to other types of transformations, and hence to invariance-based justifications for other priors as noninformative. Theorems 2.3 and 3.2 together provide a perhaps more robust justification of Haldane's prior: any justification for a different prior which is based on invariance arguments under transformations of Δ^n must use transformations that do not have the conservation properties of definition 2.1, and therefore will tend to be less natural than the transformations on which the justification of Haldane's prior is based.

4 Equivariant Measure Selection

A fundamental probabilistic inference problem is the problem of *measure selection*: given some incomplete information about the true distribution $\boldsymbol{p}$ on S, what is the best rational hypothesis for the precise value of $\boldsymbol{p}$?

Example 4.1 (continuation of example 2.4) One of the research assistants has lost his count sheet on his way home. Unwilling to discard the data from the road watched by this assistant, the project leader tries to extract some information about the counts that the assistant might remember. The assistant is able to say that he observed at least 10 times as many cars as buses, and at least 5 times as many cars as buses and bicycles combined. The only way to enter the observation from this particular road into the study, however, is in the form of accurate relative frequencies of bus, car, and bicycle use. To this end, the project leader has to make a best guess of the actual frequencies based on the linear constraints given to him by the assistant.

The general formulation of the measure selection problem given above admits of a number of different more precise problem specifications. In particular, one can distinguish different variants of the general problem according to the nature of the distribution $\boldsymbol{p}$, and the nature of the incomplete information available about $\boldsymbol{p}$. Several solutions that have been proposed for the measure selection problem are based on quite different interpretations of $\boldsymbol{p}$ and incomplete information (Shore & Johnson 1980, Paris & Vencovská 1990, Jaeger 2001). In order to clarify the role of the equivariance principle that we will propose as a desideratum for measure selection rules, we first take a closer look at these different interpretations.

4.1 Variants of the Measure Selection Problem

We first make some general assumptions on the purely mathematical form of incomplete information about $\boldsymbol{p}$, and the measure selection problem: one assumption is that incomplete information consists of a set $\boldsymbol{c} = c_1, \ldots, c_k$ of linear constraints on $\boldsymbol{p}$, i.e. linear inequalities of the form

$$c_{i,1}p_1 + \ldots + c_{i,n}p_n \leq c_{i,0} \quad (1 \leq i \leq k)$$

with real coefficients $c_{i,j}$. This is quite a restrictive assumption on what types of incomplete information are to be considered, as it excludes e.g. independence constraints of the form "events A and B are independent". In spite of this restrictiveness, the limitation to linear constraints usually has to be made in order to make the measure selection problem at all feasible.

A set $\boldsymbol{c}$ of linear constraints defines the set $\Delta(\boldsymbol{c}) \subseteq \Delta^n$ of distributions that satisfy all constraints (the solution set of $\boldsymbol{c}$). One possible mathematical formulation of the measure selection problem now is

(Sel 1) define a *selection function sel* that maps sets $\boldsymbol{c}$ of linear constraints to nonempty subsets *sel*$(\boldsymbol{c}) \subseteq \Delta^n$.

This formalization, on the one hand, is very strong in that it requires *sel* to be defined for all, even inconsistent, sets of constraints; on the other hand it is very weak in that *sel*$(\boldsymbol{c})$ is allowed to be a subset of Δ^n, rather than a unique element, and, moreover, it is not required that *sel*$(\boldsymbol{c}) \subseteq \Delta(\boldsymbol{c})$ (which would be incompatible with the requirement that *sel* also is defined for inconsistent $\boldsymbol{c}$). An alternative, more traditional formalization of the problem is

(Sel 2) define a *selection function sel* that maps consistent sets $\boldsymbol{c}$ of linear constraints to points *sel*$(\boldsymbol{c}) \in \Delta(\boldsymbol{c})$.

Identifying a set of constraints $\boldsymbol{c}$ with its solution set $\Delta(\boldsymbol{c})$, and generalizing from such polytopes to arbitrary closed and convex subsets $A \subseteq \Delta^n$, one can finally put the problem in the following form:

(Sel 3) define a *selection function sel* that maps nonempty, closed and convex subsets $A \subseteq \Delta^n$ to points *sel*$(\boldsymbol{c}) \in A$.

Sel 1-3 are purely mathematical formalizations of the problem which do not directly represent any specific interpretations of the nature of $\boldsymbol{p}$, or the constraints $\boldsymbol{c}$.

However, which of these formalizations is most appropriate is partly determined by the interpretation given to $\boldsymbol{p}$ and $\boldsymbol{c}$.

First turning to $\boldsymbol{p}$, we can distinguish the cases that $\boldsymbol{p}$ represents a statistical, observable probability, or that $\boldsymbol{p}$ represents a subjective probability (degree of belief). These two different types of distributions give rise to two distinct interpretations of the "true" distribution $\boldsymbol{p}$ that we want to identify by measure selection: In the case, of statistical probabilities the "true" $\boldsymbol{p}$ describes actual long-run frequencies, which, in principle, given sufficient time and experimental resources, one could determine exactly. In the case of subjective probability, the "true" $\boldsymbol{p}$ is a rational belief state that an ideal intelligent agent would arrive at by properly taking into account all its current, incomplete knowledge.

A second dichotomy arises through different interpretations of the nature of the constraints $\boldsymbol{c}$: these can either be seen as a complete description of a state of information, or as randomly sampled pieces of (possibly unreliable) information. This distinction between *constraints as knowledge* and *constraints as data* was introduced in (Jaeger 2001). It is a distinction that is independent from the distinction between statistical and subjective probabilities $\boldsymbol{p}$. The following examples illustrate all four combinations of interpretations for $\boldsymbol{p}$ and $\boldsymbol{c}$.

Example 4.2 (Statistical probabilities, constraints as data) Let $\boldsymbol{p}$ be a probability distribution in a medical domain that represents relative frequencies of certain diseases and symptoms. A linear constraint can, for instance, provide an upper bound on the probability of disease D given symptom S. We can now obtain a great number of such constraints by evaluating patient data from different hospitals and/or by interviewing numerous medical experts. Each individual constraint we elicit in this manner can then be seen as a randomly sampled piece of information on the true distribution $\boldsymbol{p}$ that describes the actual relative frequencies in the population we actually want to model. Note that constraints obtained in this manner can easily be inconsistent (patient data from different hospitals may show quite different conditional probabilities). Note, too, that we will probably have greater confidence in, and pay more attention to, constraints that we have observed multiple times (e.g. the conditional probability of D given S has been determined for many different hospitals, and similar values have been found in all cases) than "isolated" constraints (e.g. a conditional probability for D' given S' has only been mentioned by one expert, and not been corroborated otherwise).

Example 4.3 (Statistical probabilities, constraints as knowledge) Let $\boldsymbol{p}$ be as in the preceding example, but now suppose that the constraints are obtained by systematically interviewing a single expert, for instance by requiring him to state for every possible conditional probability in the domain a best lower and upper bound, according to his knowledge.

Example 4.4 (Subjective probabilities, constraints as data) Let $\boldsymbol{p}$ represent the subjective probabilities some European football enthusiast holds about the results in the upcoming champion's league season. Suppose we meet this fan at some late hour in the local pub, and that the conversation turns to football. Every now and then he will make a statement that, in effect, is a linear constraint on $\boldsymbol{p}$: "Barcelona has at least twice the chance of reaching the finals that Madrid has", "I'd bet 10:1 that Bayern Munich will exit in the first round again – no, make that 20:1", ... As in example 4.2, the constraints so obtained can be interpreted as randomly sampled pieces of evidence on the true beliefs $\boldsymbol{p}$. As before, these constraints can be inconsistent, and we will pay greater attention to those constraints that have been consistently repeated several times.

Example 4.5 (Subjective probabilities, constraints as knowledge) Let $\boldsymbol{p}$ be the beliefs held by a professional bookmaker on the results in the upcoming champion's league season. Before the season starts, he offers certain odds on some possible bets, e.g. 10:1 that Madrid will reach the semifinals. Assuming the bookmaker to be rational, we can interpret these odds as constraints on his beliefs $\boldsymbol{p}$ (the probability that Madrid will reach the semifinals is at most 0.1). As the bookmaker will aim to offer bets on all events for which he believes to have some reasonable probability assessment, and will also want to offer competitive odds, one can view the collection of bets he offers as a complete description of his state of knowledge.

Clearly, in any given situation it need not be obvious whether the constraints as data or constraints as knowledge interpretation is more appropriate – both interpretations are idealizations that will never be encountered in a pure form in reality. A good criterion by which one can judge which interpretation of the given constraints is the right one is to decide whether one should base measure selection on the raw set of observed constraints $\boldsymbol{c}$, taking into account possible multiple occurrences of the same constraint, or whether $\Delta(\boldsymbol{c})$ alone already encodes all the relevant information provided by $\boldsymbol{c}$. This also means that under the constraints as data interpretation the mathematical shape of the measure selection problem is (Sel 1), whereas under the constraints as knowledge interpretation (Sel 2) and (Sel 3) are more natural.

More important than the technicalities of the problem formalization, however, is the question whether the different interpretations for $\boldsymbol{p}$ and $\boldsymbol{c}$ will lead to completely different solution paradigms, or whether the same formal selection rules are appropriate in all cases. Paris (1994, n.d.) emphasizes that the principles he postulates for measure selection rules are meant to apply to subjective probabilities $\boldsymbol{p}$ and the constraints as knowledge interpretation only. In (Jaeger 2001), on the other hand, it has been argued that the constraints as data perspective requires different selection principles than the constraints as knowledge perspective (no distinction between statistical and subjective $\boldsymbol{p}$ was there made). However, in contrast to Paris, we see no reason to believe that measure selection for subjective probabilities should follow different principles than measure selection for statistical probabilities. The following now summarizes our working hypotheses:

- Selection rules under the constraints as data interpretation are different from selection rules under the constraints as knowledge interpretation.

- Under either interpretation for the constraints, the same selection rules are applicable to statistical and subjective probabilities.
- The equivariance principle, introduced below, is applicable (in slightly different forms) under both interpretations for the constraints.

The preceding discussions already provide sufficient support for the first hypothesis: we have seen that the constraints as data perspective leads to selection rules that must be sensitive to multiple occurrences of identical constraints, but under the constraints as knowledge perspective such multiplicities would be ignored.

Our second hypothesis is to be regarded as a working hypothesis that might have to be revised should arguments to the contrary appear. It is mainly supported by a uniform philosophical interpretation of measure selection for statistical and subjective probabilities: as already observed above, in the statistical case, the "true" $\boldsymbol{p}$ represents unobserved long-run frequencies. Measure selection for statistical probabilities can then be seen as a prediction on actual long-run frequencies that, in principle, one would be able to observe in a suitable experimental setup (or simply by making observations over a sufficiently long period of time).

Measure selection for subjective probabilities admits of a quite similar interpretation: following earlier suggestions of a frequentist basis for subjective probability (Reichenbach 1949, Carnap 1950), it is argued in (Jaeger 1995) that subjective probability is ultimately grounded in empirical observation, hence statistical probability. In particular, in (Jaeger 1995) the process of subjective measure selection is interpreted as a process very similar to statistical measure selection, namely a prediction on the outcome of hypothetical experiments (which, however, here even unlimited experimental resources may not permit us to carry out in practice). Under the uniform interpretation of statistical and subjective measure selection as a prediction of frequencies in (hypothetical) experiments, it seems reasonable that both selection processes should follow the same formal rules. In further support of our second hypothesis it may be remarked that the measure selection principles Shore and Johnson (1980) postulate are very similar to those of Paris and Vencovská (1990), but they were formulated with statistical probabilities in mind.

Our third working hypothesis is a combined result of the arguments made in (Jaeger 2001), and those contained in the following section.

4.2 Equivariance Principle

In the following we focus on the measure selection problem under the constraints as knowledge perspective, taking (Sel 3) to be its mathematical structure. We propose an equivariance principle for this setting. An analogous principle adapted to the constraints as data perspective and the mathematical form (Sel 1) is described in (Jaeger 2001).

The most widely favored solution to the measure selection problem under the constraints as knowledge interpretation is the *entropy maximization* rule: define $sel_{me}(A)$ to be the distribution $\boldsymbol{p}$ in A that has maximal entropy (for closed and convex A this is well-defined). Axiomatic justifications for this selection rule are given in (Shore & Johnson 1980, Paris & Vencovská 1990). Both these works postulate a number of formal principles that a selection rule should obey, and then proceed to show that entropy maximization is the only rule satisfying all the principles. Paris (1999) argues that all these principles in essence are just expressions of one more general underlying principle, which is expressed by an informal statement (or slogan) by van Fraassen (1989): *Essentially similar problems should have essentially similar solutions.*

In spite of its mathematical sound derivation, entropy maximization does exhibit some behaviors that appear counterintuitive to many (see (Jaeger 2001) for two illustrative examples). Often this counterintuitive behavior is due to the fact that the maximum entropy rule has a strong bias towards the uniform distribution $\boldsymbol{u} = (1/n, \ldots, 1/n)$. As $\boldsymbol{u}$ is the element in Δ^n with globally maximal entropy, $\boldsymbol{u}$ will be selected whenever $\boldsymbol{u} \in A$. Consider, for example, figure 1 (i) and (ii). Shown are two different subsets A and A' of Δ^3. Both contain $\boldsymbol{u}$, and therefore $sel_{me}(A) = sel_{me}(A') = \boldsymbol{u}$. While none of Paris' rationality principles explicitly demands that $\boldsymbol{u}$ should be selected whenever possible, there is one principle that directly implies the following for the sets depicted in figure 1: assuming that $sel(A) = \boldsymbol{u}$, and realizing that A' is a subset of A, one should also have $sel(A') = \boldsymbol{u}$. This is an instance of what Paris (1994) calls the *obstinacy principle*: for any A, A' with $A' \subseteq A$ and $sel(A) \in A'$ it is required that $sel(A') = sel(A)$. The intuitive justification for this is that additional information (i.e. information that limits the previously considered distribution A to A') that is consistent with the previous default selection (i.e. $sel(A) \in A'$) should not lead us to revise this default selection. While quite convincing from a default reasoning perspective (in fact, it is a version of Gabbay's (1985) *restricted monotonicity* principle), it is not entirely clear that this principle is an expression of the van Fraassen slogan. Indeed, at least from a geometric point of view, there does seem to exist little similarity between the two problems given by A and A', and thus the requirement that they should have similar solutions (or even the same solution) hardly seems a necessary consequence of the van Fraassen slogan.

An alternative selection rule that avoids some of the shortcomings of sel_{me} is the *center of mass* selection rule sel_{cm}: $sel_{cm}(A)$ is defined as the center of mass of A. With sel_{cm} one avoids the bias towards $\boldsymbol{u}$, and, more generally, the bias of sel_{me} towards points on the boundary of the input set A is reversed towards an exclusive preference for points in the interior of A. A great part of the intuitive appeal of sel_{cm} is probably owed to the fact that it satisfies (1), i.e. it is translation-equivariant.

Such an equivariance property can be understood as a much more direct formalization of the van Fraassen slogan than the individual postulates proposed in the derivations of the maximum entropy principle. Indeed, van Fraassen (1989), after giving the informal slogan, proceeds to explain it further as a general *symmetry*

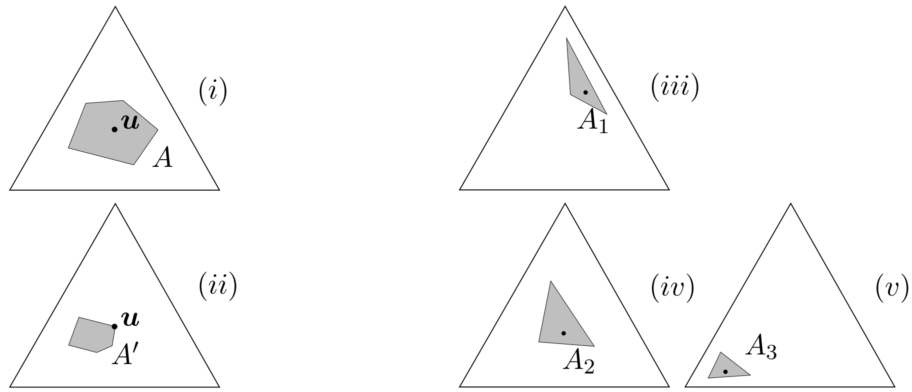

Figure 1: Maximum Entropy and T_n- equivariant selection

requirement of the form

$$h(R(A)) = R(h(A)), \tag{3}$$

where A is the input to some inference problem, R is a solution rule for the problem, and h is some problem transformation (van Fraassen 1989, p.260). This symmetry requirement, thus, is a very general principle that can be applied to many different types of inference problems. The equivariance principle (1) is a special form of (3) with h the translation by $\boldsymbol{r}$. For our special measure selection problem we have that A is any closed and convex subset of Δ^n, and R is a selection rule. To apply van Fraassen's general symmetry requirement to our special problem, it thus remains to specify the transformation(s) h for which (3) should be required.

Appealing to theorem 2.3, we argue that the transformations in T_n are the most relevant transformations to consider in our problem setting, so that we arrive at the following *T_n-equivariance principle* for selection rules:

$$\text{For all } t_{\boldsymbol{r}} \in T_n : \quad \mathit{sel}(t_{\boldsymbol{r}}A) = t_{\boldsymbol{r}}\mathit{sel}(A). \tag{4}$$

Figure 1 (iii)-(v) illustrates the T_n-equivariance principle: shown are three different transformations A_1, A_2, A_3 of a polytope defined by three linear constraints, and the corresponding transformations $\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3$ of one distinguished element inside the A_i. T_n-equivariance now demands that $\mathit{sel}(A_1) = \boldsymbol{p}_1 \Leftrightarrow \mathit{sel}(A_2) = \boldsymbol{p}_2 \Leftrightarrow \mathit{sel}(A_3) = \boldsymbol{p}_3$.

Example 4.6 (continuation of example 4.1) Assume that the unlucky assistant in example 4.1 was given instructions of the first type, and that he collected his data accordingly. If, instead, he had been given instructions of the second type, then the frequencies on the lost count sheet would have been frequencies $\boldsymbol{f}' = t_{\boldsymbol{r}}\boldsymbol{f}$, where $\boldsymbol{f}$ are the actual frequencies on the lost sheet, and $t_{\boldsymbol{r}}$ is as in example 2.4. The partial information he would then have been able to give also would have taken a different form. For instance, he might then have stated that he observed at least 6 times as many cars as buses, and at least 4.5 times as many cars as buses and bicycles combined.

Thus, while the actual constraints he gives (having used the first set of instructions) define a set $A \subseteq \Delta^n$, the constraints he would have stated had he been using the second set of instructions define a set $A' \subseteq \Delta^n$. Under very natural modelling assumptions on how the constraints stated in the two potential situations are related, one can show that they are related via the transformations $\bar{t}_{\boldsymbol{r}}$ on constraints defined in appendix A, and that $A' = t_{\boldsymbol{r}}A$ (Jaeger 2001).

When the project leader uses a T_n-equivariant selection rule for reconstructing the true frequencies from the information he is given, then the following two approaches will lead to the same solution, whatever set of instructions this particular assistant was using: 1: first infer the actual frequencies observed by the assistant by applying the selection rule to the given set $\Delta(\boldsymbol{c})$, and then transform to the preferred type of frequencies. 2: first transform the given set $\Delta(\boldsymbol{c})$ so that it corresponds to the information that would have been given had the preferred set of instructions been used, and then apply the selection rule.

T_n-equivariance imposes no restriction on what *sel*(A_i) should be for any single A_i in figure 1. It only determines how the selections for the different A_i should be related. This principle alone, thus, is far from providing a unique selection rule, like the rationality principles of Paris and Vencovská (1990). On the other hand, we have not yet shown that T_n-equivariant selection rules even exist. In the remainder of this section we investigate the feasibility of defining T_n-equivariant selection rules, without making any attempts to find the best or most rational ones.

From (4) one immediately derives a limitation of possible T_n-equivariant selection rules: let $A = \Delta^n$ in (4). Then $t_{\boldsymbol{r}}A = A$ for every $t_{\boldsymbol{r}} \in T_n$, and equivariance demands that $t_{\boldsymbol{r}}$*sel*$(A) =$ *sel*(A) for all $t_{\boldsymbol{r}}$, i.e. *sel*(A) has to be a fixpoint under all transformations. The only elements of Δ^n that have this property are the n vertices $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$, where $\boldsymbol{v}_i$ is the distribution that assigns unit probability to $s_i \in S$. Clearly a rule with *sel*$(\Delta^n) = \boldsymbol{v}_i$ for any particular i would be completely arbitrary, and could not be argued to follow any rationality principles (more technically, such a rule would not be *permutation equivariant*, which is another equivariance property one would demand in order to deal appropriately with reorderings of the state space, as discussed in section 2).

Similar problems arise whenever *sel* is to be applied to some $A \subseteq \Delta^n$ that is invariant under some transformations of T_n. To evade these difficulties, we restrict in the following the domain of *sel* to sets A that are not invariant under any transformation $t_{\boldsymbol{r}}$ (except the identity transformation). Let $\mathcal{A}$ denote the set of closed and convex A that are contained in the interior of Δ^n (i.e. *support*$(\boldsymbol{p}) = \{1, \ldots, n\}$ for all $\boldsymbol{p} \in A$), and that have full dimension (i.e. the affine hull of A has dimension $n-1$)[1]. One can show that $A \in \mathcal{A}$ are not invariant under any non-trivial $t_{\boldsymbol{r}} \in T_n$.

[1]The notation here is slightly modified from the one used in (Jaeger 2003)

$\mathcal{A}$ is not the most general class of closed and convex subsets of Δ^n with the desired non-invariance property. However, we shall also need the containment in the interior of Δ^n and the full dimensionality to facilitate some of the constructions and arguments following below.

The selection rule we define in the following for $A \in \mathcal{A}$ is not necessarily meant to already be a best or most rational selection rule. However, it does have some intuitive appeal, and the method by which it is constructed illustrates a general strategy by which T_n-equivariant selection rules can be constructed.

Example 4.7 Define on $\mathcal{A}$ an equivalence relation $\sim$:

$$A \sim A' \quad :\Leftrightarrow \quad \exists t_{\boldsymbol{r}} \in T_n : \; A' = t_{\boldsymbol{r}} A.$$

The equivalence class *orb*$(A) := \{A' \mid A' \sim A\}$ $(= \{t_{\boldsymbol{r}} A \mid t_{\boldsymbol{r}} \in T_n\})$ is called the *orbit* of A (these are standard definitions). It is easy to verify that for $A \in \mathcal{A}$ also *orb*$(A) \subseteq \mathcal{A}$, and that for every $A' \in$ *orb*(A) there is a unique $t_{\boldsymbol{r}} \in T_n$ with $A' = t_{\boldsymbol{r}} A$ (here transformations are unique, but as observed above, this does not imply that the parameter $\boldsymbol{r}$ representing the transformation is unique).

Suppose that *sel*$(A) = \boldsymbol{p} = (p_1, \ldots, p_n)$. With $\boldsymbol{r} = (1/p_1, \ldots, 1/p_n)$ then $t_{\boldsymbol{r}} \boldsymbol{p} = \boldsymbol{u}$, and by equivariance *sel*$(t_{\boldsymbol{r}} A) = \boldsymbol{u}$. It follows that in every orbit there must be some set A' with *sel*$(A') = \boldsymbol{u}$. On the other hand, if *sel*$(A') = \boldsymbol{u}$, then this uniquely defines *sel*(A) for all A in the orbit of A': *sel*$(A) = \boldsymbol{p}$, where $\boldsymbol{p} = t_{\boldsymbol{r}} \boldsymbol{u}$ with $t_{\boldsymbol{r}}$ the unique transformation with $t_{\boldsymbol{r}} A' = A$. One thus sees that the definition of an equivariant selection rule is equivalent to choosing for each orbit in $\mathcal{A}$ a representative A' for which *sel*$(A') = \boldsymbol{u}$ shall hold.

For $A \in \mathcal{A}$ let *chm*(A) denote the center of mass of A with respect to Haldane's prior H. Thus, $\boldsymbol{p} =$ *chm*(A) iff for $i = 1, \ldots, n$:

$$p_i = \int_A p'_i dH(p'_i)/H(A) \tag{5}$$

Since A is full-dimensional, closed, and contained in the interior of Δ^n, one has $0 < H(A) < \infty$, so that the p_i are well-defined.

Lemma 4.8 Let $A \in \mathcal{A}$. There exists a unique $A' \in$ *orb*(A) with *chm*$(A') = \boldsymbol{u}$.

Combining the intuitive center-of-mass selection rule with the principle of T_n-equivariance, we thus arrive at the T_n-equivariant center-of-mass selection rule: $sel_{\textit{equiv-chm}}(A) = \boldsymbol{p}$ iff $A = t_{\boldsymbol{r}} A'$, $\boldsymbol{u}$ is the center of mass of A', and $\boldsymbol{p} = t_{\boldsymbol{r}} \boldsymbol{u}$.

In (Jaeger 2003) the same construction as given here was sketched using center-of-mass with respect to Lebesgue measure instead of Haldane's prior. While the analogue of lemma 4.8 might be expected to also hold for *cm* in place of *chm*, this appears to be much harder to prove, so that at this point it must be considered an open question whether the construction also works for *cm*.

One may wonder whether a T_n-equivariant selection rule cannot be defined much simpler as $sel_{\textit{chm}}(A) :=$ *chm*(A), and whether perhaps even $sel_{\textit{equiv-chm}} = sel_{\textit{chm}}$. This is not the case: one can show that $sel_{\textit{chm}}$ is not T_n-equivariant.

5 Conclusions

Many probabilistic inference problems that are characterized by a lack of information have to be solved on the basis of considerations of symmetries and invariances. These symmetries and invariances, in turn, can be defined in terms of transformations of the mathematical objects one encounters in the given type of inference problem.

The representation theorem we have derived provides a strong argument that in inference problems whose objects are elements and subsets of Δ^n, one should pay particular attention to invariances (and equivariances) under the transformations T_n. These transformations can be seen as the analogue in the space Δ^n of translations in the space $\mathbb{R}^n$.

One should be particularly aware of the fact that it usually does not make sense to simply restrict symmetry and invariance concepts that are appropriate in the space $\mathbb{R}^n$ to the subset Δ^n. A case in point is the problem of noninformative priors. In $\mathbb{R}^n$ Lebesgue measure is the canonical choice for an (improper) noninformative prior, because its invariance under translations makes it the unique (up to a constant) "uniform" distribution. Restricted to Δ^n, however, this distinction of Lebesgue measure does not carry much weight, as translations are not a meaningful transformation of Δ^n. Our results indicate that the choice of Haldane's prior for Δ^n is much more in line with the choice of Lebesgue measure on $\mathbb{R}^n$, than the choice of the "uniform" distribution, i.e. Lebesgue measure restricted to Δ^n.

In a similar vein, we have conjectured in section 4 that some of the intuitive appeal of the center-of-mass selection rule is its equivariance under translations. Again, however, translations are not the right transformations to consider in this context, and one therefore should aim to construct T_n-equivariant selection rules, as, for example, the T_n-equivariant modification of center-of-mass.

An interesting open question is how many of Paris and Vencovská's (1990) rationality principles can be reconciled with T_n-equivariance. As the combination of all uniquely identifies maximum entropy selection, there must always be some that are violated by T_n-equivariant selection rules. Clearly the obstinacy principle is rather at odds with T_n-equivariance (though it is not immediately obvious that the two really are inconsistent). Can one find selection rules that satisfy most (or all) principles except obstinacy?

A A Dual Version of the Representation Theorem

In this appendix we formulate a dual version of Theorem 2.3 in which we consider transformations on linear constraints on Δ^n, rather than transformations of Δ^n directly.

A linear constraint in the general form

$$c: \ \ c_1p_1 + \ldots + c_np_n \leq c_0 \quad (x_1, \ldots, x_n, z \in \mathbb{R}).$$

can be rewritten as

$$(c_1 - c_0)p_1 + \ldots + (c_n - c_0)p_n \leq 0,$$

(using $\sum_i p_i = 1$). By normalizing with $1/\|(c_1 - c_0, \ldots, c_n - c_0)\|$ this becomes

$$s_1 p_1 + \ldots + s_n p_n \leq 0, \tag{6}$$

where $\boldsymbol{s} := (s_1, \ldots, s_n)$ is an element of the $n-1$-dimensional unit sphere

$$S^n = \{(s_1, \ldots, s_n) \mid \sum_i s_i^2 = 1\}.$$

In analogy to the partitioning of Δ^n into subsets with common sets of support, we partition S^n into *sectors* according to the following definition.

Definition A.1 A *sign vector* is any vector with components in $\{-1, 0, 1\}$. For $r \in \mathbb{R}$ we define *sign*(r) as $-1, 0$ or 1, depending on whether $r < 0$, $r = 0$, or $r > 0$. The sign vector *sign*$(\boldsymbol{s})$ for $\boldsymbol{s} \in S^n$ is the vector $(\mathit{sign}(s_i))_{i=1,\ldots,n}$. Each sign-vector ζ of length n defines a *sector* S^ζ in S^n:

$$S^\zeta := \{\boldsymbol{s} \in S^n \mid \mathit{sign}(\boldsymbol{s}) = \zeta\}. \tag{7}$$

As before, denote with $\Delta(\boldsymbol{s})$ the set of distributions in Δ^n that satisfy (6). We call a constraint $\boldsymbol{s}$ *proper* iff $\Delta(\boldsymbol{s}) \notin \{\emptyset, \Delta^n\}$, i.e. $\boldsymbol{s}$ is neither vacuous nor unsatisfiable. Note that $\boldsymbol{s}$ is proper iff *sign*$(\boldsymbol{s})$ has at least one component equal to 1, and at least one component equal to -1. In particular, either all constraints in a sector S^ζ are proper, or all are improper. Thus we can also speak of (im-)proper sectors. Let S^n_{prop} denote the set of proper constraints. A transformation $\bar{t}$ of S^n_{prop} is said to

- *preserve sectors*, iff $\bar{t}$ is a bijection on every proper sector S^ζ.
- *preserve implications*, iff for all $k \in \mathbb{N}$, $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_k \in S^n_{\mathit{prop}}$:

$$\bigcap_{i=1}^{k-1} \Delta(\boldsymbol{s}_i) \subseteq \Delta(\boldsymbol{s}_k) \Leftrightarrow \bigcap_{i=1}^{k-1} \Delta(\bar{t}\boldsymbol{s}_i) \subseteq \Delta(\bar{t}\boldsymbol{s}_k) \tag{8}$$

Definition A.2 Let $\boldsymbol{r} = (r_1, \ldots, r_n) \in (\mathbb{R}^+)^n$. The transformation $\bar{t}_{\boldsymbol{r}} : S^n \to S^n$ is defined by

$$\bar{t}_{\boldsymbol{r}}\,(s_1, \ldots, s_n) := \frac{(r_1 s_1, \ldots, r_n s_n)}{\|(r_1 s_1, \ldots, r_n s_n)\|}.$$

We write $\bar{T}_n$ for the set $\{\bar{t}_{\boldsymbol{r}} \mid \boldsymbol{r} \in (\mathbb{R}^+)^n\}$.

We now obtain the desired result.

Theorem A.3 Let $n \geq 3$, and $\bar{t}$ a transformation of S^n_{prop}. $\bar{t}$ preserves sectors and implications iff $\bar{t} \in \bar{T}_n$.

The duality of the transformations in $\bar{T}_n$ and T_n is expressed by the following theorem.

Theorem A.4 Let $\boldsymbol{r} \in (\mathbb{R}^+)^n$, $\boldsymbol{p} \in \Delta^n$. Then for all proper $\boldsymbol{s}$:

$$\boldsymbol{p} \in \Delta(\boldsymbol{s}) \quad \Leftrightarrow \quad t_{\boldsymbol{r}}\boldsymbol{p} \in \Delta(\bar{t}_{\boldsymbol{r}}\boldsymbol{s}).$$

Proof of Theorem A.3: The right-to-left direction is straightforward. For the converse direction we first introduce the following notation: for $I \subseteq \{1,\ldots,n\}$ let $F^I := \{\boldsymbol{p} \in \Delta^n \mid \mathit{support}(\boldsymbol{p}) = I\}$. Geometrically speaking, F^I is the interior of an $(\mid I \mid -1)$-dimensional face of Δ^n. For $j = 1,\ldots,\mid I \mid -1$ let $\mathcal{F}^I_j$ denote the set of all j-dimensional intersections of F^I with linear spaces. Thus, $\mathcal{F}^I_0$ is the set of all points in F^I, $\mathcal{F}^I_1$ is the set of all lines in F^I, etc. Each $A \in \mathcal{F}^I_j$ can be represented as the intersection of F^I with an $n- \mid I \mid -1-j$-dimensional subspace L. L, in turn, is given as the intersection of $\mid I \mid -1-j$ hyperplanes.

Let $\bar{t}$ be a transformation of S^n_{prop} that preserves sectors and implications. We first show that the following holds for all proper constraints $\boldsymbol{s}_i$, and all F^I:

(i) $\cap_{i=1}^k \Delta(\boldsymbol{s}_i) \cap F^I = \emptyset$ iff $\cap_{i=1}^k \Delta(\bar{t}\boldsymbol{s}_i) \cap F^I = \emptyset$

(ii) $\cap_{i=1}^k \Delta(\boldsymbol{s}_i) \cap F^I \subseteq \cap_{j=1}^l \Delta(\boldsymbol{s}'_j) \cap F^I$ iff $\cap_{i=1}^k \Delta(\bar{t}\boldsymbol{s}_i) \cap F^I \subseteq \cap_{j=1}^l \Delta(\bar{t}\boldsymbol{s}'_j) \cap F^I$

(iii) $\bar{t}(-\boldsymbol{s}) = -\bar{t}\boldsymbol{s}$.

To prove (i), we distinguish the cases $I = \{1,\ldots,n\}$, and $I \neq \{1,\ldots,n\}$.

For the case $I \neq \{1,\ldots,n\}$ we show that $\cap_{i=1}^k \Delta(\boldsymbol{s}_i) \cap F^I = \emptyset$ holds iff there exists a proper sector S^ζ with $\{i \mid \zeta_i = 1\} \subseteq I \subseteq \{i \mid \zeta_i \in \{0,1\}\}$, and a constraint $\boldsymbol{s} \in S^\zeta$ with $\cap_{i=1}^k \Delta(\boldsymbol{s}_i) \subseteq \Delta(\boldsymbol{s})$. From this equivalence (i) follows, because the right-hand side of the equivalence clearly is preserved under $\bar{t}$. The right-to-left direction of the equivalence is immediate from $F^I \cap \Delta(\boldsymbol{s}) = \emptyset$ for all $\boldsymbol{s} \in S^\zeta$. For the left-to-right direction let $\boldsymbol{p}_1,\ldots,\boldsymbol{p}_m$ be the vertices of the polytope $\cap_{i=1}^k \Delta(\boldsymbol{s}_i)$. Suppose that $\boldsymbol{p}_1,\ldots,\boldsymbol{p}_l$ $(0 \leq l \leq m)$ are those vertices with $\mathit{support}(\boldsymbol{p}_i) \not\subseteq I$, and let $J := \cup_{i=l+1}^m \mathit{support}(\boldsymbol{p}_i)$. From $\cap_{i=1}^k \Delta(\boldsymbol{s}_i) \cap F^I = \emptyset$ it follows that $J \subsetneq I$. Let ζ be the sign vector with $\zeta_i = 1$ for $i \in I \setminus J$, $\zeta_i = 0$ for $i \in J$, and $\zeta_i = -1$ for $i \notin I$. Since $I \neq \{1,\ldots,n\}$, S^ζ is a proper sector. Furthermore, for $\epsilon \in (0,\infty)$ let $\boldsymbol{s}(\epsilon) \in S^\zeta$ be the constraint with $\boldsymbol{s}(\epsilon)_i = \epsilon$ for $i \in I \setminus J$, and $\boldsymbol{s}(\epsilon)_i = -1/\epsilon$ for $i \notin I$. For $i = 1,\ldots,l$ there exists $\epsilon_i > 0$ with $\boldsymbol{p}_i \in \Delta(\boldsymbol{s}(\epsilon))$ for all $\epsilon \leq \epsilon_i$. Let $\epsilon^* := \mathit{min}_{1 \leq i \leq l}\, \epsilon_i$ (if $l = 0$ simply let $\epsilon^* := 1$). We now have $\boldsymbol{p}_i \in \Delta(\boldsymbol{s}(\epsilon^*))$ for $i = 1,\ldots,m$, and hence $\cap_{i=1}^k \Delta(\boldsymbol{s}_i) \subseteq \Delta(\boldsymbol{s}(\epsilon^*))$, as desired.

In the case $I = \{1,\ldots,n\}$ we have that $\cap_{i=1}^k \Delta(\boldsymbol{s}_i) \cap F^I = \emptyset$ is equivalent to the existence of a set $J \subsetneq \{1,\ldots,n\}$ such that $\cap_{i=1}^k \Delta(\boldsymbol{s}_i)$ is contained in the closure of F^J. This, in turn, is equivalent to $\cap_{i=1}^k \Delta(\boldsymbol{s}_i) \subseteq \Delta(\boldsymbol{s})$ for all $\boldsymbol{s}$ with *sign*$(\boldsymbol{s}_i) = 1$ for $i \notin J$, and *sign*$(\boldsymbol{s}_i) = -1$ for $i \in J$. This last property is preserved under $\bar{t}$.

The proof of (ii) is similar to (i): one shows that $\cap_{i=1}^k \Delta(\boldsymbol{s}_i) \cap F^I \subseteq \cap_{j=1}^l \Delta(\boldsymbol{s}'_j) \cap F^I$ is equivalent to the existence of a constraint $\boldsymbol{s}$ with *sign*$(\boldsymbol{s}_i) = -1$ for $i \in I$ and *sign*$(\boldsymbol{s}_i) = 1$ for $i \notin I$, such that $\cap_{i=1}^k \Delta(\boldsymbol{s}_i) \cap \Delta(\boldsymbol{s}) \subseteq \cap_{j=1}^l \Delta(\boldsymbol{s}'_j)$.

For (iii) we observe that for all $\boldsymbol{s}, \boldsymbol{s}'$: $\boldsymbol{s}' = -\boldsymbol{s}$ iff for all $\boldsymbol{s}'' \neq \boldsymbol{s}$:

$$\Delta(\boldsymbol{s}'') \subseteq \Delta(\boldsymbol{s}) \quad \Leftrightarrow \quad \Delta(\boldsymbol{s}') \cap \Delta(\boldsymbol{s}'') \cap \textit{int}\, \Delta^n = \emptyset,$$

where $\textit{int}\, \Delta^n$ denotes the interior of Δ^n. By implication preservation and (i) both sides of this equivalence are preserved under $\bar{t}$, and, therfore, so is the equivalence. Thus, $\boldsymbol{s}' = -\boldsymbol{s}$ iff $\bar{t}\boldsymbol{s}' = -\bar{t}\boldsymbol{s}$, i.e. $\bar{t}(-\boldsymbol{s}) = -\bar{t}\boldsymbol{s}$.

Now we show that $\bar{t}$ induces transformations t_j^I on $\mathcal{F}_j^I$ for all I, j: let $A \in \mathcal{F}_j^I$. Then

$$A = \cap_{i=1}^{|I|-j-1} (\Delta(\boldsymbol{s}_i) \cap \Delta(-\boldsymbol{s}_i)) \cap F^I$$

for suitable $\boldsymbol{s}_i$. Define

$$t_j^I A = \cap_{i=1}^{|I|-j-1} (\Delta(\bar{t}\boldsymbol{s}_i) \cap \Delta(-\bar{t}\boldsymbol{s}_i)) \cap F^I.$$

By (ii) tA is well-defined, i.e. it does not depend on the particular choice of the $\boldsymbol{s}_i$. From (i) and (iii) it follows that $tA \neq \emptyset$ is the intersection of F^I with a linear subspace, i.e. $tA \in \mathcal{F}_h^I$ for some $h \in \{0, \ldots, |\, I\,| -1\}$. We show that $h = j$ by induction on j.

Let $j = 0$, i.e. $A = \{\boldsymbol{p}\}$ for some $\boldsymbol{p} \in F^I$. Suppose that $h > 0$. Then A contains two distinct points, and there exist constraints $\boldsymbol{s}, \boldsymbol{s}'$ with $\Delta(\boldsymbol{s}) \cap \Delta(\boldsymbol{s}') \cap F^I = \emptyset$, $\Delta(\boldsymbol{s}) \cap F^I \neq \emptyset$, $\Delta(\boldsymbol{s}') \cap F^I \neq \emptyset$. By (i) these (in)equalities are preserved when substituting $\bar{t}^{-1}\boldsymbol{s}, \bar{t}^{-1}\boldsymbol{s}', A$ for $\boldsymbol{s}, \boldsymbol{s}', tA$, which is a contradiction to $A = \{\boldsymbol{p}\}$.

Now let $j \geq 1$. Let $\tilde{A} \in \mathcal{F}_{j-1}^I$ with $\tilde{A} \subsetneq A$. Then $\tilde{A} = A \cap \Delta(\boldsymbol{s}) \cap \Delta(-\boldsymbol{s})$ for some $\boldsymbol{s}$, and

$$t\tilde{A} = tA \cap \Delta(\bar{t}\boldsymbol{s}) \cap \Delta(-\bar{t}\boldsymbol{s}) \in \mathcal{F}_{j-1}^I$$

by induction hypothesis. Since $tA \in \mathcal{F}_h^I$ is an open set (relative to its affine hull), its intersection with the hyperplane $\Delta(\bar{t}\boldsymbol{s}) \cap \Delta(-\bar{t}\boldsymbol{s})$ has dimension at least $h-1$, so that $j - 1 \leq h \leq j$. Finally, $h = j-1$ would entail $t\tilde{A} = tA$, which with (ii) contradicts the definition of $\tilde{A}$.

The transformations t_j^I are consistent in the following sense: if $A \in \mathcal{F}_j^I, A' \in \mathcal{F}_h^I$ with $j < h$ and $A \subset A'$, then $t_j^I A \subset t_h^I A'$. Furthermore, when $I \subset I'$, $A \in \mathcal{F}_j^{I'}$, and $A' \in \mathcal{F}_k^I$ is the intersection of the boundary $\textit{bd}\, A$ with F^I, then $t_k^I A = \textit{bd}\,(t_j^{I'} A) \cap F^I$.

We now define the transformation t on Δ^n as the union of the transformations t_0^I ($I \subset \{1, \ldots, n\}$). By definition, t preserves sets of support. To see that t

preserves mixtures, we observe that $\boldsymbol{p}$ is a mixture of $\boldsymbol{p}', \boldsymbol{p}''$ iff there does not exist $\boldsymbol{s} \in S^n_{prop}$ with $\boldsymbol{p}', \boldsymbol{p}'' \in \Delta(\boldsymbol{s})$, $\boldsymbol{p} \notin \Delta(\boldsymbol{s})$. By (ii) and the definition of t we have for all $\boldsymbol{p}, \boldsymbol{s}$

$$\boldsymbol{p} \in \Delta(\boldsymbol{s}) \quad \Leftrightarrow \quad t\boldsymbol{p} \in \Delta(\bar{t}\boldsymbol{s}) \tag{9}$$

which, with the foregoing characterization of mixtures, implies that mixtures are preserved under t.

By theorem 2.3 $t = t_{\boldsymbol{r}} \in T_n$ for some $\boldsymbol{r} \in (\mathbb{R}^+)^n$. Using (9) we thus obtain for all $\boldsymbol{p}, \boldsymbol{s}$

$$\boldsymbol{p} \in \Delta(\boldsymbol{s}) \cap \Delta(-\boldsymbol{s}) \quad \Leftrightarrow \quad t_{\boldsymbol{r}}\boldsymbol{p} \in \Delta(\bar{t}\boldsymbol{s}) \cap \Delta(-\bar{t}\boldsymbol{s}), \tag{10}$$

i.e.

$$\sum_i p_i s_i = 0 \quad \Leftrightarrow \quad \sum_i r_i p_i (\bar{t}\boldsymbol{s})_i = 0. \tag{11}$$

As (10) completely determines the transformation $\bar{t}$, and (11) is seen to be solved by $(\bar{t}\boldsymbol{s})_i = c s_i / r_i$ for any constant c, we obtain $\bar{t} = \bar{t}_{\boldsymbol{r}}$.

□

Proof of Theorem A.4: This is an immediate consequence of (9) in the preceding proof. □

B Proofs for Sections 2 - 4

Theorem 2.3 Let $n \geq 3$ and t be a transformation of Δ^n.

(i) t preserves sets of support and mixtures iff $t \in T_n$.

(ii) t preserves cardinalities of support and mixtures iff $t = t' \circ \pi$ for some permutation π and some $t' \in T_n$.

Proof:

For $\boldsymbol{x} \in (\mathbb{R}^+)^n$ we denote with $[\boldsymbol{x}]$ the linear subspace of $\mathbb{R}^n$ generated by $\boldsymbol{x}$. We use $\mathbb{R}^n_{1Q}$ to denote the first quadrant of $\mathbb{R}^n$, i.e. the set of all points with only non-negative coordinates. With $\mathcal{P}^{n-1}$ we denote the set of all one-dimensional linear subspaces of $\mathbb{R}^n$, i.e. the $(n-1)$-dimensional projective space over $\mathbb{R}$. Furthermore, with $\mathcal{P}^{n-1}_{1Q}$ we denote the subset of $\mathcal{P}^{n-1}$ containing those subspaces that intersect $\mathbb{R}^n_{1Q}$ not only in $\mathbf{0}$. Thus,

$$\mathcal{P}^{n-1}_{1Q} = \{[\boldsymbol{p}] \mid \boldsymbol{p} \in \Delta^n\},$$

and, moreover, every $[\boldsymbol{x}] \in \mathcal{P}^{n-1}_{1Q}$ is uniquely represented by one $\boldsymbol{p} \in \Delta^n$. The transformation t, therefore, immediately induces a (bijective) transformation on $\mathcal{P}^{n-1}_{1Q}$, which, for simplicity, we also denote with t.

The main part of the proof now consists of showing that t can be extended to a linear transformation t^* of $\mathbb{R}^n_{1Q}$. The arguments used to establish this closely follow the proofs of the representation theorem for projective colineations (also known as the fundamental theorem of projective geometry) as given in (Faure & Frölicher 2000) and (Beutelspacher & Rosenbaum 1998). That representation theorem states that every transformation t on $\mathcal{P}^{n-1}$ that preserves colinearity is induced by a linear transformation t^* of $\mathbb{R}^n$. Here we show basically a version of this result that, on the one hand, is restricted to $\mathcal{P}^{n-1}_{1Q}$ and $\mathbb{R}^n_{1Q}$, and, on the other hand, starts with the slightly stronger requirement of preservation of mixtures, rather than preservation of colinearity (the former requires also that the relative order of colinear points is preserved). The main work in adapting the proof of the representation theorem for colineations to our problem consists of making sure that all geometric constructions in the original proof can be contained within the subset $\mathcal{P}^{n-1}_{1Q}$.

We require some additional notation: $[\boldsymbol{x}, \boldsymbol{y}]$ stands for the linear subspace generated by $\boldsymbol{x}$ and $\boldsymbol{y}$. If $\boldsymbol{x}$ and $\boldsymbol{y}$ are linearly independent, this is a two-dimensional plane, which, in projective geometry terms, is the *line* connecting $[\boldsymbol{x}]$ and $[\boldsymbol{y}]$. We say that subspaces $[\boldsymbol{x}_1], \ldots, [\boldsymbol{x}_k]$ are *linearly independent* if the $\boldsymbol{x}_i$ are linearly independent. A vector $\boldsymbol{z} \in \mathbb{R}^n_{1Q}$ is a *positive combination* of $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n_{1Q}$ if there exists $\alpha, \beta \in \mathbb{R}^+$ with $\boldsymbol{z} = \alpha\boldsymbol{x} + \beta\boldsymbol{y}$. In that case we also say that $[\boldsymbol{z}]$ is a positive combination of $[\boldsymbol{x}]$ and $[\boldsymbol{y}]$. Observe that the mixture preservation property of t just means that $t[\boldsymbol{z}]$ is a positive combination of $t[\boldsymbol{x}]$ and $t[\boldsymbol{y}]$ whenever $\boldsymbol{z}$ is a positive combination of $\boldsymbol{x}$ and $\boldsymbol{y}$.

We prepare the main part of the proof with the following lemma.

Lemma B.1 **(A)** Let $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^n_{1Q}$ such that $\boldsymbol{x}, \boldsymbol{y}$ are linearly independent, and $\boldsymbol{z}$ is a positive combination of $\boldsymbol{x}$ and $\boldsymbol{y}$. Then there exists exactly one $\tilde{\boldsymbol{y}} \in [\boldsymbol{y}] \cap \mathbb{R}^n_{1Q}$ with $\boldsymbol{x} + \tilde{\boldsymbol{y}} \in [\boldsymbol{z}]$.

(B) Let $t[\boldsymbol{x}_1], t[\boldsymbol{x}_2], t[\boldsymbol{x}_3]$ be linearly independent, and let $\boldsymbol{y}_i \in t[\boldsymbol{x}_i]$ ($i = 1, 2, 3$) such that

$$t[\boldsymbol{x}_1 + \boldsymbol{x}_2] = [\boldsymbol{y}_1 + \boldsymbol{y}_2] \quad \text{and} \quad t[\boldsymbol{x}_1 + \boldsymbol{x}_3] = [\boldsymbol{y}_1 + \boldsymbol{y}_3].$$

Then

$$t[\boldsymbol{x}_2 + \boldsymbol{x}_3] = [\boldsymbol{y}_2 + \boldsymbol{y}_3] \quad \text{and} \quad t[\boldsymbol{x}_1 + \boldsymbol{x}_2 + \boldsymbol{x}_3] = [\boldsymbol{y}_1 + \boldsymbol{y}_2 + \boldsymbol{y}_3].$$

Proof of lemma: **(A)** Independent from n, this is a statement only about the plane spanned by $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$. The construction of $\tilde{\boldsymbol{y}}$, therefore is illustrated in full generality by Figure 2.

(B) With

$$[\boldsymbol{x}_2 + \boldsymbol{x}_3] \subseteq [\boldsymbol{x}_2, \boldsymbol{x}_3] \cap [\boldsymbol{x}_1, \boldsymbol{x}_1 + \boldsymbol{x}_2 + \boldsymbol{x}_3]$$
$$[\boldsymbol{x}_1 + \boldsymbol{x}_2 + \boldsymbol{x}_3] \subseteq [\boldsymbol{x}_1 + \boldsymbol{x}_2, \boldsymbol{x}_3] \cap [\boldsymbol{x}_1 + \boldsymbol{x}_3, \boldsymbol{x}_2]$$

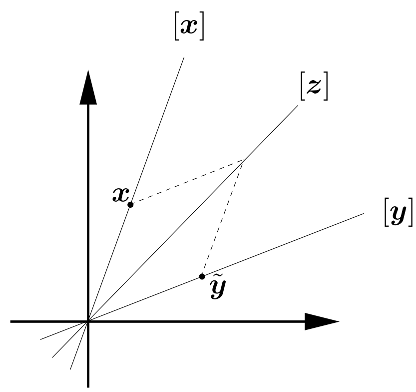

Figure 2: Lemma B.1 **(A)**

it follows that

$$t[\boldsymbol{x}_2 + \boldsymbol{x}_3] \subseteq [\boldsymbol{y}_2, \boldsymbol{y}_3] \cap [\boldsymbol{y}_1, \boldsymbol{y}_1 + \boldsymbol{y}_2 + \boldsymbol{y}_3]$$
$$t[\boldsymbol{x}_1 + \boldsymbol{x}_2 + \boldsymbol{x}_3] \subseteq [\boldsymbol{y}_1 + \boldsymbol{y}_2, \boldsymbol{y}_3] \cap [\boldsymbol{y}_1 + \boldsymbol{y}_3, \boldsymbol{y}_2].$$

From the linear independence of the $\boldsymbol{y}_i$ it follows that the intersections of the spaces on the right are just $[\boldsymbol{y}_2 + \boldsymbol{y}_3]$, resp. $[\boldsymbol{y}_1 + \boldsymbol{y}_2 + \boldsymbol{y}_3]$.

□

Let $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ be such that $t[\boldsymbol{a}_1], t[\boldsymbol{a}_2], t[\boldsymbol{a}_3]$ are linearly independent. Let $[\boldsymbol{b}_1] = t[\boldsymbol{a}_1]$. We have that $t[\boldsymbol{a}_1 + \boldsymbol{a}_2]$ is a positive combination of $t[\boldsymbol{a}_1]$ and $t[\boldsymbol{a}_2]$, so that by part (A) of the lemma there exists a unique $\boldsymbol{b}_2 \in t[\boldsymbol{a}_2]$ such that $t[\boldsymbol{a}_1 + \boldsymbol{a}_2] = [\boldsymbol{b}_1 + \boldsymbol{b}_2]$. Similarly, there exists a unique $\boldsymbol{b}_3 \in t[\boldsymbol{a}_3]$ with $t[\boldsymbol{a}_1 + \boldsymbol{a}_3] = [\boldsymbol{b}_1 + \boldsymbol{b}_3]$. With part (B) of the lemma it furthermore follows that $t[\boldsymbol{a}_2 + \boldsymbol{a}_3] = [\boldsymbol{b}_2 + \boldsymbol{b}_3]$.

We can now define t^*: first, define $t^*(\boldsymbol{0}) := \boldsymbol{0}$. Now let $\boldsymbol{x} \in \mathbb{R}^n_{1Q} \setminus \boldsymbol{0}$. Let $\boldsymbol{a}_i \in \{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3\}$ such that $t[\boldsymbol{a}_i] \neq [\boldsymbol{x}]$. Then $t[\boldsymbol{a}_i + \boldsymbol{x}]$ is a positive combination of $t[\boldsymbol{a}_i]$ and $t[\boldsymbol{x}]$, so that by part (A) of the lemma there exists a unique $\boldsymbol{z} \in t[\boldsymbol{x}] \cap \mathbb{R}^n_{1Q}$ with $[\boldsymbol{b}_i + \boldsymbol{z}] = t[\boldsymbol{a}_i + \boldsymbol{x}]$. Define $t^*(\boldsymbol{x}) := \boldsymbol{z}$.

We have to show that the definition of $t^*(\boldsymbol{x})$ does not depend on the particular choice of $\boldsymbol{a}_i$. For this, assume that $t[\boldsymbol{a}_i] \neq [\boldsymbol{x}] \neq t[\boldsymbol{a}_j]$, and that by above construction we have obtained $\boldsymbol{z}_i, \boldsymbol{z}_j$ with $[\boldsymbol{b}_i + \boldsymbol{z}_i] = t[\boldsymbol{a}_i + \boldsymbol{x}]$, $[\boldsymbol{b}_j + \boldsymbol{z}_j] = t[\boldsymbol{a}_j + \boldsymbol{x}]$.

To show that $\boldsymbol{z}_i = \boldsymbol{z}_j$ first consider the case that $t[\boldsymbol{x}] \notin [t[\boldsymbol{a}_i], t[\boldsymbol{a}_j]]$. Applying part (B) of the lemma to $\boldsymbol{x}_1 = \boldsymbol{a}_i, \boldsymbol{x}_2 = \boldsymbol{x}, \boldsymbol{x}_3 = \boldsymbol{a}_j$ and $\boldsymbol{y}_1 = \boldsymbol{b}_i, \boldsymbol{y}_2 = \boldsymbol{z}_1, \boldsymbol{y}_3 = \boldsymbol{b}_j$, one obtains $t[\boldsymbol{a}_j + \boldsymbol{x}] = [\boldsymbol{b}_j + \boldsymbol{z}_i]$ and hence by the uniqueness statement of part (A) of the lemma $\boldsymbol{z}_i = \boldsymbol{z}_j$.

In the case $t[\boldsymbol{x}] \in [t[\boldsymbol{a}_i], t[\boldsymbol{a}_j]]$ we obtain from the linear independence of the $t[\boldsymbol{a}_i]$ that $t[\boldsymbol{x}] \notin [t[\boldsymbol{a}_i], t[\boldsymbol{a}_k]] \cup [t[\boldsymbol{a}_j], t[\boldsymbol{a}_k]]$ where $k = \{1,2,3\} \setminus \{i,j\}$. In particular, $t[\boldsymbol{a}_k] \neq t[\boldsymbol{x}]$, so that $\boldsymbol{z}_k$ is defined by our construction. Applying the first case twice we obtain $\boldsymbol{z}_i = \boldsymbol{z}_k = \boldsymbol{z}_j$.

We next proceed to show that $t^*(\boldsymbol{x}+\boldsymbol{y}) = t^*(\boldsymbol{x}) + t^*(\boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n_{1Q}$. For this, first assume that $t^*(\boldsymbol{x})$ and $t^*(\boldsymbol{y})$ are linearly independent. There exists $\boldsymbol{a}_i$ ($i \in \{1,2,3\}$) with $t[\boldsymbol{a}_i] \not\subseteq [t[\boldsymbol{x}], t[\boldsymbol{y}]]$. Applying part (B) of the lemma to $\boldsymbol{x}_1 = \boldsymbol{a}_i, \boldsymbol{x}_2 = \boldsymbol{x}, \boldsymbol{x}_3 = \boldsymbol{y}$ and $\boldsymbol{y}_1 = \boldsymbol{b}_i, \boldsymbol{y}_2 = t^*(\boldsymbol{x}), \boldsymbol{y}_3 = t^*(\boldsymbol{y})$ one obtains $t[\boldsymbol{x}+\boldsymbol{y}] = [t^*(\boldsymbol{x}) + t^*(\boldsymbol{y})]$ and $t[\boldsymbol{a}_i + \boldsymbol{x} + \boldsymbol{y}] = [\boldsymbol{b}_i + t^*(\boldsymbol{x}) + t^*(\boldsymbol{y})]$. As $t[\boldsymbol{a}_i] \neq t[\boldsymbol{x}+\boldsymbol{y}]$ therefore $t^*(\boldsymbol{x}+\boldsymbol{y}) = t^*(\boldsymbol{x}) + t^*(\boldsymbol{y})$.

Now assume that $[t^*(\boldsymbol{x})] = [t^*(\boldsymbol{y})]$, and hence $t[\boldsymbol{x}] = t[\boldsymbol{y}]$ and $[\boldsymbol{x}] = [\boldsymbol{y}]$. Choose any $\boldsymbol{z}$ such that $t^*(\boldsymbol{z})$ is linearly independent from $t^*(\boldsymbol{x}+\boldsymbol{y})$ (and therefore also from $t^*(\boldsymbol{x})$ and $t^*(\boldsymbol{y})$). Then also $t^*(\boldsymbol{x})$ and $t^*(\boldsymbol{y}+\boldsymbol{z})$ are linearly independent, because $[\boldsymbol{x}] \neq [\boldsymbol{y}+\boldsymbol{z}]$. Applying the previous case twice, we obtain on the one hand $t^*(\boldsymbol{x}+\boldsymbol{y}+\boldsymbol{z}) = t^*(\boldsymbol{x}+\boldsymbol{y}) + t^*(\boldsymbol{z})$, and on the other hand $t^*(\boldsymbol{x}+\boldsymbol{y}+\boldsymbol{z}) = t^*(\boldsymbol{y}+\boldsymbol{z}) + t^*(\boldsymbol{x}) = t^*(\boldsymbol{y}) + t^*(\boldsymbol{z}) + t^*(\boldsymbol{x})$.

Next we show that $t^*(\alpha\boldsymbol{x}) = \alpha t^*(\boldsymbol{x})$ for $\alpha \in \mathbb{R}^+$. By definition we have that $t^*(\alpha\boldsymbol{x}) = \beta t^*(\boldsymbol{x})$, where $\beta = \beta_{\boldsymbol{x}}(\alpha) \in \mathbb{R}^+$ might depend both on α and on $\boldsymbol{x}$. We first show that β does not depend on $\boldsymbol{x}$, i.e. $\beta_{\boldsymbol{x}}(\alpha) = \beta_{\boldsymbol{y}}(\alpha)$ for all $\boldsymbol{x}, \boldsymbol{y}$. For this, first assume that $[\boldsymbol{x}] \neq [\boldsymbol{y}]$. By additivity we have on the one hand $t^*(\alpha(\boldsymbol{x}+\boldsymbol{y})) = \beta_{\boldsymbol{x}}(\alpha)t^*(\boldsymbol{x}) + \beta_{\boldsymbol{y}}(\alpha)t^*(\boldsymbol{y})$, and on the other hand $t^*(\alpha(\boldsymbol{x}+\boldsymbol{y})) = \beta_{\boldsymbol{x}+\boldsymbol{y}}(\alpha)(t^*(\boldsymbol{x}) + t^*(\boldsymbol{y}))$. From the linear independence of $t^*(\boldsymbol{x})$ and $t^*(\boldsymbol{y})$ it follows that $\beta_{\boldsymbol{x}}(\alpha) = \beta_{\boldsymbol{x}+\boldsymbol{y}}(\alpha) = \beta_{\boldsymbol{y}}(\alpha)$.

If $[\boldsymbol{x}] = [\boldsymbol{y}]$ we pick $\boldsymbol{z}$ with $[\boldsymbol{z}] \neq [\boldsymbol{x}]$ and obtain with the previous case $\beta_{\boldsymbol{x}}(\alpha) = \beta_{\boldsymbol{z}}(\alpha) = \beta_{\boldsymbol{y}}(\alpha)$.

It remains to show that $\beta(\alpha) = \alpha$. For this, we first show that $\beta(\alpha_1 + \alpha_2) = \beta(\alpha_1) + \beta(\alpha_2)$. For this, let $\boldsymbol{x}$ be any point. Then $\beta(\alpha_1+\alpha_2)t^*(\boldsymbol{x}) = t^*(\alpha_1\boldsymbol{x}+\alpha_2\boldsymbol{x}) = \beta(\alpha_1)t^*(\boldsymbol{x}) + \beta(\alpha_2)t^*(\boldsymbol{x})$. Similarly, we obtain $\beta(\alpha_1\alpha_2)t^*(\boldsymbol{x}) = t^*(\alpha_1\alpha_2\boldsymbol{x}) = \beta(\alpha_1)t^*(\alpha_2\boldsymbol{x}) = \beta(\alpha_1)\beta(\alpha_2)t^*(\boldsymbol{x})$, so that $\beta(\alpha_1\alpha_2) = \beta(\alpha_1)\beta(\alpha_2)$.

As β is not identically zero, the multiplicativity of β implies that $\beta(\alpha) \neq 0$ for all $\alpha \neq 0$. Also by multiplicativity, $\beta(1) = 1$. From additivity and multiplicativity we then obtain $\beta(n) = n$ and $\beta(1/n) = 1/n$ for all $n \in \mathbb{N}$, and hence $\beta(\alpha) = \alpha$ for all $\alpha \in \mathbb{Q}^+$. Finally, from additivity and $\beta(\alpha) \geq 0$ for all α, we obtain that $\alpha \leq \alpha'$ implies $\beta(\alpha) \leq \beta(\alpha')$, so that $\beta \upharpoonright \mathbb{Q}^+ = \mathit{id}$ implies $\beta = \mathit{id}$ on $\mathbb{R}^+$. This concludes the proof that $t^*(\alpha\boldsymbol{x}) = \alpha t^*(\boldsymbol{x})$.

Let $\boldsymbol{e}_i$ be the ith unit vector, i.e. $\boldsymbol{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ with 1 in the ith component. As the transformation t preserves sets of support, we have $t[\boldsymbol{e}_i] = [\boldsymbol{e}_i]$, and hence $t^*(\boldsymbol{e}_i) = r_i\boldsymbol{e}_i$ for some $r_i \in \mathbb{R}^+$. For $\boldsymbol{x} = \sum_i x_i\boldsymbol{e}_i$ then $t^*(\boldsymbol{x}) = \sum_i r_i x_i \boldsymbol{e}_i$. In particular, for $\boldsymbol{p} \in \Delta^n$ we have $t^*(\boldsymbol{p}) = \sum_i r_i p_i \boldsymbol{e}_i \in t[\boldsymbol{p}]$. As, furthermore, $t(\boldsymbol{p}) \in t[\boldsymbol{p}]$, where $t(\boldsymbol{p})$ is the original transformation on Δ^n, and $t[\boldsymbol{p}]$ the induced transformation on $\mathcal{P}^{n-1}_{1Q}$, we have $t(\boldsymbol{p}) = t^*(\boldsymbol{p})/\sum_i r_i p_i = \bar{g}_{\boldsymbol{r}}$ for $\boldsymbol{r} = (r_1, \ldots, r_n)$. □

Theorem 3.2 Let *Pr* be a measure on *int*Δ^n with *Pr*(*int*Δ^n) > 0 and *Pr*$(A) < \infty$ for all compact subsets A of *int*Δ^n. *Pr* is invariant under all transformations $t_{\boldsymbol{r}} \in T_n$ iff *Pr* has a density with respect to Lebesgue measure of the form $c\prod_i p_i^{-1}$ with some constant $c > 0$.

Proof: We first show the invariance of distributions *Pr* given by densities $g_c(\boldsymbol{p}) := c\prod_i p_i^{-1}$. It is sufficient to consider the case $c = 1$. We write g for g_1. Furthermore, we may restrict attention to transformations $t_{\boldsymbol{r}}$ given by vectors $\boldsymbol{r}$ with $r_i = 1$ in all but one coordinate i. General $t_{\boldsymbol{r}}$ can be obtained as compositions of such primitive transformations, and therfore the invariance of *Pr* under each primitive transformation implies invariance under all transformations. Moreover, without loss of generality, we may take $\boldsymbol{r} = (r, 1, \dots, 1)$. In the following we write t for this $t_{\boldsymbol{r}}$:

$$t(\boldsymbol{p}) = 1/(rp_1 + \sum_{i=2}^{n} p_i)(rp_1, p_2, \dots, p_n).$$

Saying that a distribution *Pr* on Δ^n has density g means that Δ^n is identified as a subset of the $n-1$-dimensional affine space $L := \{\boldsymbol{x} \in \mathbb{R}^n \mid \sum x_i = 1\}$, and that g is a density with respect to $n-1$-dimensional Lebesgue measure on this space. To simplify the parameterization of our problem, we can identify L with $\mathbb{R}^{n-1}$ via the embedding

$$\pi : (x_1, \dots, x_{n-1}, 1 - \sum_{i=1}^{n-1} x_i) \mapsto (x_1, \dots, x_{n-1}).$$

This embedding is measure preserving up to a constant: for all measurable $A \subseteq L$ with finite Lebesgue measure $\lambda^{n-1}(A)$ we have $\lambda^{n-1}(\pi(A)) = c_n \lambda^{n-1}(A)$ with c_n a constant depending on n. In particular, we have

$$\pi(\Delta^n) = \{\boldsymbol{x} \in [0,1]^{n-1} \mid \sum x_i \leq 1\} =: D^{n-1}.$$

The distribution *Pr* induces a distribution π*Pr* on *int* D^{n-1} given by the density $f(x_1, \dots, x_{n-1}) := c_n g(x_1, \dots, x_{n-1}, 1 - \sum_{i=1}^{n-1} x_i)$.

The invariance of *Pr* under t is equivalent to the invariance of π*Pr* under

$$t^\pi : (x_1, \dots, x_{n-1}) \mapsto 1/(1 + (r-1)x_1)(rx_1, x_2, \dots, x_{n-1}).$$

We thus have transformed our original problem on $\Delta^n \subseteq L$ into a similar problem for $D^{n-1} \subseteq \mathbb{R}^{n-1}$. To simplify notation, we write in the following again t for the reparameterized transformation t^π, and *Pr* for the induced distribution π*Pr*.

According to the transformation theorem for integrals, the density of the transformed distribution $t(Pr)$ is given by

$$f^t(\boldsymbol{x}) := f(t^{-1}(\boldsymbol{x}))/\,|\,J_t(t^{-1}(\boldsymbol{x}))\,| \qquad (\boldsymbol{x} \in D^{n-1}), \tag{12}$$

where J_t is the Jacobian matrix of t. We have to show that $f^t = f$.

For this, we first evaluate the Jacobian. With $a := 1 + (r-1)x_1$ the partial derivatives of t are

$$\frac{\partial t_i}{\partial x_j} = \begin{cases} r/a^2 & i = j = 1 \\ 0 & i = 1, j \neq 1 \\ -(r-1)x_j/a^2 & i \neq 1, j = 1 \\ 1/a & i = j \neq 1 \\ 0 & 1 \neq i \neq j \neq 1 \end{cases}$$

The Jordan matrix, thus, is in lower triangular form, and its determinant is the product of the main diagonal elements:

$$|\, J_t(t^{-1}(\boldsymbol{x}))\,| = r/a^n. \tag{13}$$

For $\boldsymbol{x} \in D^{n-1}$ we can write with $b := r + (1-r)x_1$ $(= -a + r + 1)$:

$$t^{-1}(\boldsymbol{x}) = r/b(x_1/r, x_2, \ldots, x_{n-1}). \tag{14}$$

Thus

$$\begin{aligned} f(t^{-1}(\boldsymbol{x})) &= c_n \left[(1 - \frac{x_1}{b} - \sum_{i=2}^{n-1} \frac{r}{b} x_i) \frac{r^{n-2}}{b^{n-1}} \prod_{i=1}^{n-1} x_i \right]^{-1} \\ &= c_n \left[\frac{r^{n-1}}{b^n} (\frac{b}{r} - \frac{x_1}{r} - \sum_{i=2}^{n-1} x_i) \prod_{i=1}^{n-1} x_i \right]^{-1} \\ &= c_n \left[\frac{r^{n-1}}{b^n} (1 - \sum_{i=1}^{n-1} x_i) \prod_{i=1}^{n-1} x_i \right]^{-1}, \end{aligned} \tag{15}$$

where the last equality follows from $(b - x_1)/r = 1 - x_1$.

With (13) and (14):

$$|\, J_t(t^{-1}(\boldsymbol{x}))\,| = \frac{r}{(1 + (r-1)x_1/b)^n} = \frac{rb^n}{(b + (r-1)x_1)^n} = \frac{b^n}{r^{n-1}}. \tag{16}$$

From (12),(15) and (16) now $f^t = f$ follows.

The uniqueness assertion of the theorem follows from general results on the uniqueness of invariant measures (Halmos 1950, Sec.60,Theorem C). For a straightforward application of these results it is only necessary to realize that Δ^n is a locally compact Hausdorff space, and that the condition $Pr(A) < \infty$ for compact A entails that Pr is regular (Cohn 1993, Proposition 7.2.3). □

Lemma 4.8 Let $A \in \mathcal{A}$. There exists a unique $A' \in orb(A)$ with $chm(A') = \boldsymbol{u}$.

Proof: In the following we denote with $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ the scalar product of vectors in $\mathbb{R}^n$. Assume that no $t_{\boldsymbol{r}} \in T_n$ with $chm(t_{\boldsymbol{r}}A) = \boldsymbol{u}$ exists. Let

$$\epsilon := inf\{\| chm(tA) - \boldsymbol{u} \| \mid t \in T_n\}.$$

The infimum here is attained for some $t \in T_n$, i.e. there exists $B \in orb(A)$ with $\| chm(B) - \boldsymbol{u} \| = \epsilon > 0$. Let $\boldsymbol{c} := chm(B)$. We show that there exists $t_{\boldsymbol{r}} \in T_n$ with $\| chm(t_{\boldsymbol{r}}B) - \boldsymbol{u} \| < \epsilon$.

Consider the function $\boldsymbol{p} \mapsto \langle \boldsymbol{c} - \boldsymbol{u}, \boldsymbol{p} \rangle$ on Δ^n. The subset of Δ^n with $\langle \boldsymbol{c} - \boldsymbol{u}, \boldsymbol{p} \rangle = d$ for a constant $d \in \mathbb{R}$ is the intersection of Δ^n with a hyperplane that is orthogonal to $\boldsymbol{c} - \boldsymbol{u}$. Figure 3 shows as dashed lines several such hyperplanes for different values of d.

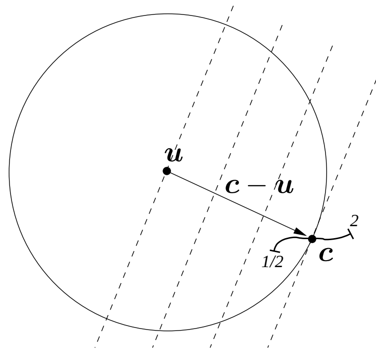

Figure 3: Proof of Lemma 4.8

With $\langle \boldsymbol{u}, \boldsymbol{p} \rangle = 1/n$ for all $\boldsymbol{p} \in \Delta^n$ one obtains $\langle \boldsymbol{c} - \boldsymbol{u}, \boldsymbol{u} \rangle = 0$, and $\langle \boldsymbol{c} - \boldsymbol{u}, \boldsymbol{c} \rangle = \langle \boldsymbol{c}, \boldsymbol{c} \rangle - 1/n > 0$. We show that there exists a set $\{\boldsymbol{r}(\delta) \mid \delta \in [1/2, 2]\}$ of parameters of transformations in T_n such that $t_{\boldsymbol{r}(1)}$ is the identity transformation, and

$$f(\delta) := \langle \boldsymbol{c} - \boldsymbol{u}, \mathit{chm}(t_{\boldsymbol{r}(\delta)} B) \rangle$$

is decreasing in δ with $f'(1) < 0$. It follows that the parametric curve $\delta \mapsto \mathit{chm}(t_{\boldsymbol{r}(\delta)} B)$ is as shown in Figure 3, i.e. it intersects the ϵ-ball around $\boldsymbol{u}$, which then contradicts the definition of B.

Define

$$I^- := \{i \in \{1, \ldots, n\} \mid c_i - u_i \leq 0\} \;\; I^+ := \{i \in \{1, \ldots, n\} \mid c_i - u_i > 0\},$$

and

$$\boldsymbol{r}(\delta)_i := \begin{cases} \delta & i \in I^- \\ 1/\delta & i \in I^+ \end{cases}$$

From the definition of $\boldsymbol{r}(\delta)$ it is immediate that $\langle \boldsymbol{c} - \boldsymbol{u}, t_{\boldsymbol{r}(\delta)} \boldsymbol{p} \rangle$ is decreasing in δ for all $\boldsymbol{p} \in \Delta^n$.

We next show that also the derivative of $\langle \boldsymbol{c} - \boldsymbol{u}, t_{\boldsymbol{r}(\delta)} \boldsymbol{p} \rangle$ with respect to δ is negative for all $\boldsymbol{p}$.

Let $\boldsymbol{p} \in \Delta^n$ be fixed, and define

$$a := \sum_{i \in I^-} p_i, \; b := \sum_{i \in I^+} p_i, \; c := \sum_{i \in I^-} (c_i - u_i) p_i, \; d := \sum_{i \in I^+} (c_i - u_i) p_i.$$

Then

$$\langle \boldsymbol{c} - \boldsymbol{u}, t_{\boldsymbol{r}(\delta)} \boldsymbol{p} \rangle = \frac{\delta c + d/\delta}{\delta a + b/\delta},$$

and

$$\frac{\partial}{\partial \delta}\langle \boldsymbol{c}-\boldsymbol{u}, t_{\boldsymbol{r}(\delta)}\boldsymbol{p}\rangle = 2\frac{\delta(cb-da)}{(\delta^2 a+b)^2}. \tag{17}$$

Since both I^- and I^+ are nonempty, and $c_i - u_i < 0$ for at least one $i \in I^-$, one obtains $a > 0,\ b > 0,\ c < 0,\ d > 0$. It follows that (17) is negative for all δ and $\boldsymbol{p}$.

We now transfer these pointwise results for single $\boldsymbol{p}$ to the function $f(\delta)$. By the definition of the center of mass and the linearity of the scalar product

$$f(\delta) = \int_{t_{\boldsymbol{r}(\delta)}B} \langle \boldsymbol{c}-\boldsymbol{u}, \boldsymbol{p}\rangle dH(\boldsymbol{p})/H(t_{\boldsymbol{r}(\delta)}B).$$

From the invariance of H under the $t_{\boldsymbol{r}(\delta)}$ it follows that the normalizing factor $\nu := 1/H(t_{\boldsymbol{r}(\delta)}B)$ is a constant that does not depend on δ, and that

$$\int_{t_{\boldsymbol{r}(\delta)}B} \langle \boldsymbol{c}-\boldsymbol{u}, \boldsymbol{p}\rangle dH(\boldsymbol{p}) = \int_B \langle \boldsymbol{c}-\boldsymbol{u}, t_{\boldsymbol{r}(\delta)}\boldsymbol{p}\rangle dH(\boldsymbol{p}). \tag{18}$$

Since $\frac{\partial}{\partial\delta}\langle \boldsymbol{c}-\boldsymbol{u}, t_{\boldsymbol{r}(\delta)}\boldsymbol{p}\rangle$ is uniformly continuous as a function of $(\boldsymbol{p},\delta)$ on $B \times [1/2, 2]$, we can move the differentiation into the integration, and obtain

$$\frac{\partial}{\partial \delta} f(\delta) = \nu \int_B \frac{\partial}{\partial \delta}\langle \boldsymbol{c}-\boldsymbol{u}, t_{\boldsymbol{r}(\delta)}\boldsymbol{p}\rangle dH(\boldsymbol{p}).$$

The integrand here is strictly negative at $\delta = 1$. With $H(B) > 0$ it follows that $\frac{\partial}{\partial\delta} f(\delta)(1) < 0$. □

References

Beutelspacher, A. & Rosenbaum, U. (1998), *Projective Geometry*, Cambridge University Press.

Carnap, R. (1950), *Logical Foundations of Probability*, The University of Chicago Press.

Cohn, D. (1993), *Measure Theory*, Birkhäuser.

Faure, C.-A. & Frölicher, A. (2000), *Modern Projective Geometry*, Kluwer Academic Publishers.

Gabbay, D. (1985), Theoretical foundations for nonmonotonic reasoning in expert systems, *in* K. Apt, ed., ‘Logics and Models of Cuncurrent Systems’, Springer-Verlag, Berlin.

Hacking, I. (1975), *The Emergence of Probability: a Philosophical Study of Early Ideas About Probability, Induction and Statistical Inference*, Cambridge University Press.

Haldane, J. (1932), 'A note on inverse probability', *Proceedings of the Cambridge Philosophical Society* **28**, 55–61.

Halmos, P. R. (1950), *Measure Theory*, Van Nostrad Reinhold Company.

Hartigan, J. (1964), 'Invariant prior distributions', *Annals of Mathematical Statistics* **35**(2), 836–845.

Holbrook, J. & Kim, S. S. (2000), 'Bertrand's paradox revisited', *The Mathematical Intelligencer* pp. 16–19.

Jaeger, M. (1995), Minimum cross-entropy reasoning: A statistical justification, *in* C. S. Mellish, ed., 'Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95)', pp. 1847–1852.

Jaeger, M. (2001), Constraints as data: A new perspective on inferring probabilities, *in* B. Nebel, ed., 'Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI-01)', pp. 755–760.

Jaeger, M. (2003), A representation theorem and applications, *in* 'Proceedings of the Seventh European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU)', Lecture Notes in Artificial Intelligence, Springer. To appear.

Jaynes, E. (1968), 'Prior probabilities', *IEEE Transactions on Systems Science and Cybernetics* **4**(3), 227–241.

Jeffreys, H. (1961), *Theory of Probability*, third edition edn, Oxford University Press.

Novick, M. R. & Hall, W. J. (1965), 'A Bayesian indifference procedure', *Journal of the American Statistical Association* **60**, 1104–1117.

Paris, J. (1999), 'Common sense and maximum entropy', *Synthese* **117**, 75–93.

Paris, J. (n.d.), On filling-in missing information in causal networks, Submitted to *Knowledge-based Systems*.

Paris, J. & Vencovská, A. (1990), 'A note on the inevitability of maximum entropy', *International Journal of Approximate Reasoning* **4**, 183–223.

Paris, J. B. (1994), *The Uncertain Reasoner's Companion*, Cambridge University Press.

Reichenbach, H. (1949), *The Theory of Probability*, University of California Press.

Schafer, J. L. (1997), *Analysis of Incomplete Multivariate Data*, Chapman & Hall/CRC.

Shore, J. & Johnson, R. (1980), 'Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy', *IEEE Transactions on Information Theory* **IT-26**(1), 26–37.

Skilling, J. (1985), 'Prior probabilities', *Synthese* **63**, 1–34.

van Fraassen, B. C. (1989), *Laws and Symmetry*, Clarendon.

Villegas, C. (1977), 'On the representation of ignorance', *Journal of the American Statistical Association* **72**, 651–654.

I N F O R M A T I K

Below you find a list of the most recent technical reports of the Max-Planck-Institut für Informatik. They are available by anonymous ftp from `ftp.mpi-sb.mpg.de` under the directory `pub/papers/reports`. Most of the reports are also accessible via WWW using the URL `http://www.mpi-sb.mpg.de`. If you have any questions concerning ftp or WWW access, please contact `reports@mpi-sb.mpg.de`. Paper copies (which are not necessarily free of charge) can be ordered either by regular mail or by e-mail at the address below.

Max-Planck-Institut für Informatik
Library
attn. Anja Becker
Stuhlsatzenhausweg 85
66123 Saarbrücken
GERMANY
e-mail: `library@mpi-sb.mpg.de`

MPI-I-2003-NWG2-002	F. Eisenbrand	Fast integer programming in fixed dimension
MPI-I-2003-NWG2-001	L.S. Chandran, C.R. Subramanian	Girth and Treewidth
MPI-I-2003-4-009	N. Zakaria	FaceSketch: An Interface for Sketching and Coloring Cartoon Faces
MPI-I-2003-4-008	C. Roessl, I. Ivrissimtzis, H. Seidel	Tree-based triangle mesh connectivity encoding
MPI-I-2003-4-007	I. Ivrissimtzis, W. Jeong, H. Seidel	Neural Meshes: Statistical Learning Methods in Surface Reconstruction
MPI-I-2003-4-006	C. Roessl, F. Zeilfelder, G. Nrnberger, H. Seidel	Visualization of Volume Data with Quadratic Super Splines
MPI-I-2003-4-005	T. Hangelbroek, G. Nrnberger, C. Roessl, H.S. Seidel, F. Zeilfelder	The Dimension of C^1 Splines of Arbitrary Degree on a Tetrahedral Partition
MPI-I-2003-4-004	P. Bekaert, P. Slusallek, R. Cools, V. Havran, H. Seidel	A custom designed density estimation method for light transport
MPI-I-2003-4-003	R. Zayer, C. Roessl, H. Seidel	Convex Boundary Angle Based Flattening
MPI-I-2003-4-002	C. Theobalt, M. Li, M. Magnor, H. Seidel	A Flexible and Versatile Studio for Synchronized Multi-view Video Recording
MPI-I-2003-4-001	M. Tarini, H.P.A. Lensch, M. Goesele, H. Seidel	3D Acquisition of Mirroring Objects
MPI-I-2003-2-003	Y. Kazakov, H. Nivelle	Subsumption of concepts in $DL\ \mathcal{FL}_0$ for (cyclic) terminologies with respect to descriptive semantics is PSPACE-complete
MPI-I-2003-2-002	M. Jaeger	A Representation Theorem and Applications to Measure Selection and Noninformative Priors
MPI-I-2003-2-001	P. Maier	Compositional Circular Assume-Guarantee Rules Cannot Be Sound And Complete
MPI-I-2003-1-013	I. Katriel, S. Thiel	Fast Bound Consistency for the Global Cardinality Constraint
MPI-I-2003-1-012	D. Fotakis, R. Pagh, P. Sanders, P. Spirakis	Space Efficient Hash Tables with Worst Case Constant Access Time
MPI-I-2003-1-011	P. Krysta, A. Czumaj, B. Voecking	Selfish Traffic Allocation for Server Farms
MPI-I-2003-1-010	H. Tamaki	A linear time heuristic for the branch-decomposition of planar graphs
MPI-I-2003-1-009	B. Csaba	On the Bollobás – Eldridge conjecture for bipartite graphs
MPI-I-2003-1-008	P. Sanders	Soon to be published
MPI-I-2003-1-007	H. Tamaki	Alternating cycles contribution: a strategy of tour-merging for the traveling salesman problem
MPI-I-2003-1-006	M. Dietzfelbinger, H. Tamaki	On the probability of rendezvous in graphs

MPI-I-2003-1-005	M. Dietzfelbinger, P. Woelfel	Almost Random Graphs with Simple Hash Functions
MPI-I-2003-1-004	E. Althaus, T. Polzin, S.V. Daneshmand	Improving Linear Programming Approaches for the Steiner Tree Problem
MPI-I-2003-1-003	R. Beier, B. Vcking	Random Knapsack in Expected Polynomial Time
MPI-I-2003-1-002	P. Krysta, P. Sanders, B. Vcking	Scheduling and Traffic Allocation for Tasks with Bounded Splittability
MPI-I-2003-1-001	P. Sanders, R. Dementiev	Asynchronous Parallel Disk Sorting
MPI-I-2002-4-002	F. Drago, W. Martens, K. Myszkowski, H. Seidel	Perceptual Evaluation of Tone Mapping Operators with Regard to Similarity and Preference
MPI-I-2002-4-001	M. Goesele, J. Kautz, J. Lang, H.P.A. Lensch, H. Seidel	Tutorial Notes ACM SM 02 A Framework for the Acquisition, Processing and Interactive Display of High Quality 3D Models
MPI-I-2002-2-008	W. Charatonik, J. Talbot	Atomic Set Constraints with Projection
MPI-I-2002-2-007	W. Charatonik, H. Ganzinger	Symposium on the Effectiveness of Logic in Computer Science in Honour of Moshe Vardi
MPI-I-2002-1-008	P. Sanders, J.L. Trff	The Factor Algorithm for All-to-all Communication on Clusters of SMP Nodes
MPI-I-2002-1-005	M. Hoefer	Performance of heuristic and approximation algorithms for the uncapacitated facility location problem
MPI-I-2002-1-004	S. Hert, T. Polzin, L. Kettner, G. Schfer	Exp Lab A Tool Set for Computational Experiments
MPI-I-2002-1-003	I. Katriel, P. Sanders, J.L. Trff	A Practical Minimum Scanning Tree Algorithm Using the Cycle Property
MPI-I-2002-1-002	F. Grandoni	Incrementally maintaining the number of l-cliques
MPI-I-2002-1-001	T. Polzin, S. Vahdati	Using (sub)graphs of small width for solving the Steiner problem
MPI-I-2001-4-005	H.P.A. Lensch, M. Goesele, H. Seidel	A Framework for the Acquisition, Processing and Interactive Display of High Quality 3D Models
MPI-I-2001-4-004	S.W. Choi, H. Seidel	Linear One-sided Stability of MAT for Weakly Injective Domain
MPI-I-2001-4-003	K. Daubert, W. Heidrich, J. Kautz, J. Dischler, H. Seidel	Efficient Light Transport Using Precomputed Visibility
MPI-I-2001-4-002	H.P.A. Lensch, J. Kautz, M. Goesele, H. Seidel	A Framework for the Acquisition, Processing, Transmission, and Interactive Display of High Quality 3D Models on the Web
MPI-I-2001-4-001	H.P.A. Lensch, J. Kautz, M. Goesele, W. Heidrich, H. Seidel	Image-Based Reconstruction of Spatially Varying Materials
MPI-I-2001-2-006	H. Nivelle, S. Schulz	Proceeding of the Second International Workshop of the Implementation of Logics
MPI-I-2001-2-005	V. Sofronie-Stokkermans	Resolution-based decision procedures for the universal theory of some classes of distributive lattices with operators
MPI-I-2001-2-004	H. de Nivelle	Translation of Resolution Proofs into Higher Order Natural Deduction using Type Theory
MPI-I-2001-2-003	S. Vorobyov	Experiments with Iterative Improvement Algorithms on Completely Unimodel Hypercubes
MPI-I-2001-2-002	P. Maier	A Set-Theoretic Framework for Assume-Guarantee Reasoning
MPI-I-2001-2-001	U. Waldmann	Superposition and Chaining for Totally Ordered Divisible Abelian Groups
MPI-I-2001-1-007	T. Polzin, S. Vahdati	Extending Reduction Techniques for the Steiner Tree Problem: A Combination of Alternative-and Bound-Based Approaches
MPI-I-2001-1-006	T. Polzin, S. Vahdati	Partitioning Techniques for the Steiner Problem
MPI-I-2001-1-005	T. Polzin, S. Vahdati	On Steiner Trees and Minimum Spanning Trees in Hypergraphs
MPI-I-2001-1-004	S. Hert, M. Hoffmann, L. Kettner, S. Pion, M. Seel	An Adaptable and Extensible Geometry Kernel
MPI-I-2001-1-003	M. Seel	Implementation of Planar Nef Polyhedra
MPI-I-2001-1-002	U. Meyer	Directed Single-Source Shortest-Paths in Linear Average-Case Time
MPI-I-2001-1-001	P. Krysta	Approximating Minimum Size 1,2-Connected Networks