European Journal of Operational Research 224 (2013) 552–559

Contents lists available at SciVerse ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Decision Support

Hit-And-Run enables efficient weight generation for simulation-based multiple criteria decision analysis

Tommi Tervonen [a,1], Gert van Valkenhoef [b,c,*,1], Nalan Baştürk [a], Douwe Postmus [c]

[a] Econometric Institute, Erasmus University Rotterdam, The Netherlands
[b] Faculty of Economics and Business, University of Groningen, The Netherlands
[c] Department of Epidemiology, University of Groningen, University Medical Center Groningen, The Netherlands

ARTICLE INFO

Article history:
Received 26 January 2012
Accepted 28 August 2012
Available online 12 September 2012

Keywords:
Multiple criteria analysis
Simulation
Uncertainty modeling
Markov Chain Monte Carlo
Stochastic Multicriteria Acceptability Analysis (SMAA)

ABSTRACT

Models for Multiple Criteria Decision Analysis (MCDA) often separate per-criterion attractiveness evaluation from weighted aggregation of these evaluations across the different criteria. In simulation-based MCDA methods, such as Stochastic Multicriteria Acceptability Analysis, uncertainty in the weights is modeled through a uniform distribution on the feasible weight space defined by a set of linear constraints. Efficient sampling methods have been proposed for special cases, such as the unconstrained weight space or complete ordering of the weights. However, no efficient methods are available for other constraints such as imprecise trade-off ratios, and specialized sampling methods do not allow for flexibility in combining the different constraint types. In this paper, we explore how the Hit-And-Run sampler can be applied as a general approach for sampling from the convex weight space that results from an arbitrary combination of linear weight constraints. We present a technique for transforming the weight space to enable application of Hit-And-Run, and evaluate the sampler's efficiency through computational tests. Our results show that the thinning factor required to obtain uniform samples can be expressed as a function of the number of criteria n as $\varphi(n) = (n-1)^3$. We also find that the technique is reasonably fast with problem sizes encountered in practice and that autocorrelation is an appropriate convergence metric.

1. Introduction

Multiple Criteria Decision Analysis (MCDA) methods consider a set of alternatives that is evaluated in terms of a set of criteria in order to choose the best one, to rank them from best to worst, or to sort them into ordered categories (Roy, 1996). The alternatives are evaluated with respect to the chosen preference model, and often some type of preferential independence is assumed. This allows the evaluation to be decomposed in two parts: (i) evaluation with respect to individual criteria and (ii) aggregation of the per-criterion scores to describe the alternatives' overall attractiveness interpreted according to the chosen preference model.

Although various MCDA methods allow criteria values to be specified in an imprecise format, e.g., as probability distributions in Multi-Attribute Utility Theory (MAUT) or indirectly through thresholds in ELECTRE methods, they require exact weights to quantify the relative contribution of the individual criteria to the alternatives' overall attractiveness. The Stochastic Multicriteria Acceptability Analysis (SMAA) family of methods (Tervonen and Figueira, 2008) extends the traditional MCDA preference models by allowing to take into account uncertainty in all model parameters, including the weights. The indices describing the decision problem are estimated through Monte Carlo simulation, and each iteration includes sampling uniformly distributed weights from a convex polytope defined through a set of linear weight constraints (Tervonen and Lahdelma, 2007). For each of the underlying preference models, several types of constraint are available to restrict the feasible weight space in a theoretically meaningful way. However, in practical applications of SMAA, the use of such constraints is limited by the lack of efficient sampling algorithms, which results in insufficient discrimination between the decision alternatives for some problem instances.

One approach to sample from a uniform distribution over the target polytope is to draw samples from a uniform proposal density over a polytope that approximates the target, such as a uniform density over a rectangular hyperbox or a Dirichlet distribution. The Dirichlet distribution is uniform over the simplex when the concentration parameter is set to 1, and has been applied in the MCDA setting by Jia et al. (1998). However, since the

* Corresponding author at: Department of Epidemiology, University Medical Center Groningen, PO Box 30.001, 9700 RB Groningen, The Netherlands.
E-mail addresses: tervonen@ese.eur.nl (T. Tervonen), g.h.m.van.valkenhoef@rug.nl (G. van Valkenhoef), basturk@ese.eur.nl (N. Başturk), d.postmus@umcg.nl (D. Postmus).
[1] Both authors contributed equally to this work.

0377-2217/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ejor.2012.08.026

proposal density only approximates the target density, such methods require a rejection step. In general, the rejection rate increases exponentially with the dimension of the sampling space, thus making this approach infeasible for higher dimensions. Alternatively, it is possible to simulate weights using different Markov Chain Monte Carlo (MCMC) methods. There is often a trade-off between the mixing rate and the acceptance rate of the sampler. For uniform joint and conditional distributions, a standard single-state Gibbs sampler is applicable. The rejection rate in this case is 0 by definition and weights can be simulated iteratively satisfying linear bounds and ratio constraints. This iterative sampling method typically leads to high correlations between draws and slow mixing (Amit and Grenander, 1991; Besag et al., 1995). It is possible to improve the mixing properties by simulating the weights jointly using random walk algorithms.

Hit-And-Run (HAR) is an MCMC sampling algorithm that, unlike other random walk algorithms, is known to mix rapidly from any interior point (Lovász, 1999; Lovász and Vempala, 2006). HAR has two clear advantages over the alternatives. First, it provides block samples from the uniform weight distribution, typically leading to high mixing rates. Second, application of HAR on the transformed parameter space avoids the rejection step in the standard block samplers under linear constraints. The rejection rate of the algorithm is 0 by definition. Both the mixing rate and the acceptance rate are improved compared to the existing samplers in this context. However, using HAR for efficient MCMC sampling from the restricted weight space as encountered in SMAA is not trivial as the weight space needs to be transformed before the algorithm can be applied.

In this paper, we consider the application of HAR sampling to weight generation in SMAA and other simulation approaches in MCDA that require uniform sampling from a restricted weight space (Butler et al., 1997; Jia et al., 1998). Our contribution is twofold. First, we present a technique for transforming the n dimensional weight space to an $n-1$ dimensional sampling space. We then evaluate through computational tests how the thinning factor required to obtain a sample equivalent to 10,000 uniform draws depends on the dimensionality of the problem. A thinning factor φ indicates that from the sequence of generated samples, we store only every φth sample. Because HAR sampling generates a series of dependent samples, thinning reduces the memory required to store the samples by accepting a relatively small loss of information about the target density.

The remainder of this paper is structured as follows. In Section 2 we discuss the weight constraints that are typically encountered in SMAA analyses and further motivate the value of efficient weight sampling methods for interactive decision aiding. Section 3 describes the HAR algorithm and discusses its application to weight generation. Section 4 discusses metrics to assess sample uniformity. Section 5 presents the results from the computational tests, and Section 6 concludes and provides directions for future research.

2. Weight constraints in SMAA

Consider a discrete multi-criteria decision problem consisting of a set of m alternatives that are evaluated in terms of n criteria. The vector of criteria values corresponding to criterion j is denoted by $x_j = \left(x_j^1, \ldots, x_j^m\right)^T$, where x_j^i denotes the performance of alternative i on criterion j.

Generally, the alternatives are first evaluated with respect to the individual criteria to obtain criterion-specific attractiveness scores, after which some kind of aggregation procedure is applied to combine the criterion-specific scores into an overall measure of preference or value (Choo et al., 1999). Let the functions $f_j\left(x_j^i\right)$ and $g_j\left(x_j^i, x_j^k\right)$ give the criterion-specific score of alternative i on criterion j (value or utility based approaches) and the criterion-specific score of the pair (i,k) on criterion j (outranking based approaches), respectively. For the purpose of this paper, we assume that the evaluation of the alternatives with respect to the individual criteria has already been completed, so that we can focus on specifying the aggregation functions $f(f_1(x_1^i), \ldots, f_n(x_n^i))$ and $g(g_1(x_1^i, x_1^k), \ldots, g_n(x_n^i, x_n^k))$.

To simplify the assessment of $f()$ and $g()$, it is generally assumed that these functions are additive:

$$f(f_1(x_1^i), \ldots, f_n(x_n^i)) = \sum_{j=1}^{n} w_j f_j\left(x_j^i\right)$$

$$g(g_1(x_1^i, x_1^k), \ldots, g_n(x_n^i, x_n^k)) = \sum_{j=1}^{n} w_j g_j\left(x_j^i, x_j^k\right)$$

where w_j denotes the weight of criterion j. Although the algebraic shape of $f()$ and $g()$ is the same, the meaning of the weights depends on the underlying preference model. In compensatory approaches such as Multi-Attribute Value Theory (MAVT) or MAUT, the weights are scaling factors that ensure that unit increases in the functions $f_j()$ are commensurate (Keeney and Raiffa, 1976). In non-compensatory approaches such as ELECTRE or PROMETHEE, by contrast, the weights do not represent value trade-offs between criteria scale swings but should rather be interpreted as the amount of voting power that is allocated to each of the criteria (Vansnick, 1986; Belton and Stewart, 2002). For a more thorough discussion of the interpretation of weights in MCDA, we refer to Choo et al. (1999).

The traditional approach in MCDA is to try to establish exact values for the weights by applying a dedicated weight elicitation technique, such as SWING weighting, which is then followed by extensive sensitivity analyses to explore how robust the obtained results are to small changes in the weights and the criteria values. In many real-life decision making contexts, however, decision makers do not feel confident with providing exact numerical values for the weights and/or the criteria values. In such situations, a method from the SMAA family can be applied to compute, for a wide range of weights and criteria values, the probability that an alternative is placed at a certain rank (ranking problems; Lahdelma and Salminen (2001)) or that it belongs to a certain category (sorting problems; Tervonen et al., 2009). The estimation of these indices is achieved through Monte Carlo simulation, and in each iteration criteria weights sampled from their joint probability distribution are required. To generate these values, it is typically assumed that the weights are uniformly distributed within the convex polytope defined through a set of weight constraints.

Without loss of generality, we assume that the weights are non-negative and normalized so that they sum to one. When no preference information is available, the feasible weight space is an $n-1$ dimensional simplex in n dimensional space:

$$W_n = \left\{ w \in R^n : w \geqslant 0 \quad \text{and} \quad \sum_{j=1}^{n} w_j = 1 \right\}$$

In practice, it is often possible to elicit some preference information from the decision maker, such as a partial or complete ranking of the criteria weights. Such information can be included in the model by restricting the feasible weight space accordingly (see Fig. 1 for an example). Weight constraints in SMAA models should be elicited by taking into account the underlying preference model. In MAVT or MAUT, meaningful constraints are ordinal rankings of the weights ($w_i > w_j$) and intervals for trade-off ratios between weights

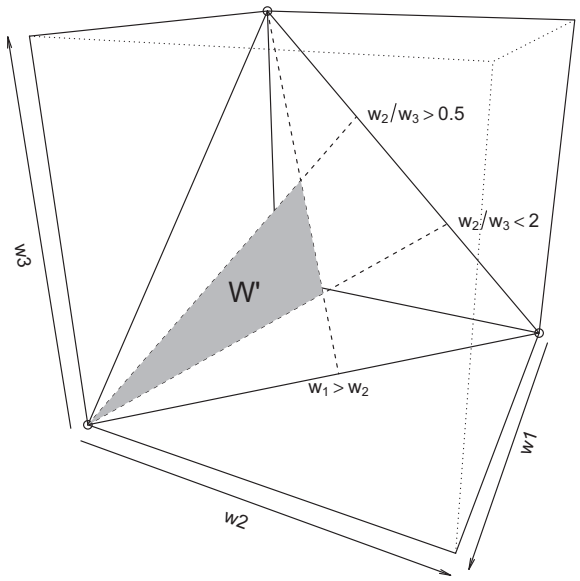


Fig. 1. The feasible weight space W' (gray polygon) for a 3-criteria problem with the partial ranking $w_1 > w_2$ and the trade-off ratio constraint $w_2/w_3 \in [0.5, 2]$. The solid triangle is the feasible weight space without preference information (W_3).

$(w_i/w_j \in [(w_i/w_j)^{\min}, (w_i/w_j)^{\max}])$. In case of ELECTRE, lower- and upper bounds $\left(w_j^{\min} \leqslant w_j \leqslant w_j^{\max}\right)$ may be considered as well.

Although efficient weight generation techniques exist for sampling from the unconstrained weight space as well as from the weight space constrained through a complete ordering of the weights or weight lower bounds (Tervonen and Lahdelma, 2007), no methods have been proposed for sampling weights constrained with imprecise trade-off ratios or with a combination of constraints. While the use of ordinal rankings of the weights or weight lower bounds may result in high discrimination for some problem instances, for others the preference information portrayed by these constraints may not be precise enough to sufficiently differentiate the decision alternatives. For small problem instances, a lack of discrimination with the currently included weight constraints can be resolved in an iterative and interactive way by first eliciting more precise preference information from the decision maker and then applying rejection sampling to sample the weights accordingly. However, for rejection sampling the hit rate decreases exponentially with the number of criteria, so it is intractable even for a moderate number of criteria (i.e. $\geqslant$10). Hence, the ability to efficiently sample weight vectors for arbitrary combinations of the different types of weight constraint is critical to the real-world application of SMAA to larger problem instances. The HAR sampler introduced in the next section is particularly suited for this purpose.

3. Hit-And-Run (HAR) for weight generation

In HAR sampling (Smith, 1984), the Markov Chain is initialized with a starting point within the polytope. At each iteration a random direction is generated by sampling from the unit hypersphere, implemented efficiently by generating independent normal variates and normalizing them to form the sample point (Marsaglia, 1972). The random direction together with the current position generates a line set, and its intersection with the boundary of the polytope generates a line segment from which the next point is drawn uniformly. The generated Markov Chain converges on the uniform distribution over the polytope in nondeterministic polynomial time $O^*(n^3)$. So to apply the HAR sampler to weight generation in the constrained weight space, the sampling space must be defined appropriately, line intersections must be computed, and a starting point contained in the polytope must be chosen. The HAR sampler and the transformations described here are implemented as a package for the *R* statistical software, and are available at http://cran.r-project.org/web/packages/hitandrun/.

3.1. Sampling space transformation

The feasible weight space with constraints defines a convex polytope $W' \subseteq W_n$. The $(n-1)$-simplex W_n is coincident with the hyperplane $W_n^* = \left\{w \in R^n : \sum_{j=1}^n w_j = 1\right\}$. Thus, the volume of the feasible weight space is essentially 0, which means that if we perform MCMC sampling in n dimensions, the probability of hitting inside W' is also 0. Therefore, we transform the simplex so we can sample directly in $n-1$ dimensions.

In the following, let I_n be the $n \times n$ identity matrix. The centroid of W_n is at $(1/n, \ldots, 1/n)^T$, so if we translate the plane W_n^* by $(-1/n, \ldots, -1/n)^T$, it forms an $n-1$ dimensional subspace $V \subset R^n$. We obtain an orthonormal basis $\{v^1, \ldots, v^{n-1}\}$ of V by first defining a basis of V and then performing orthogonalization and normalization. A basis can be defined by choosing $n-1$ vectors, so that for the kth vector the nth component is -1, the kth component is 1, and the others are 0. For a 2-simplex in R^3 (as shown in Fig. 1) such a basis would be $\{(1,0,-1)^T, (0,1,-1)^T\}$. Now, to map an arbitrary point $x \in R^{n-1}$ to a point in the target space $w \in W_n^*$, we apply an affine transformation: a change of basis followed by a translation. To do this, we use the homogeneous coordinate representation $x = (x_1, x_2, \ldots, x_{n-1}, 0, 1)^T$:

$$w = TBx$$

where B is the $(n+1) \times (n+1)$ augmented change-of-basis matrix and T the $(n+1) \times (n+1)$ translation matrix:

$$B = \begin{pmatrix} v_1^1 & \cdots & v_1^{n-1} & \sqrt{1/n} & 0 \\ \vdots & & \vdots & \vdots & \vdots \\ v_n^1 & \cdots & v_n^{n-1} & \sqrt{1/n} & 0 \\ 0 & \cdots & 0 & 0 & 1 \end{pmatrix}; \quad T = \begin{pmatrix} & & & 1/n \\ & I_n & & \vdots \\ & & & 1/n \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

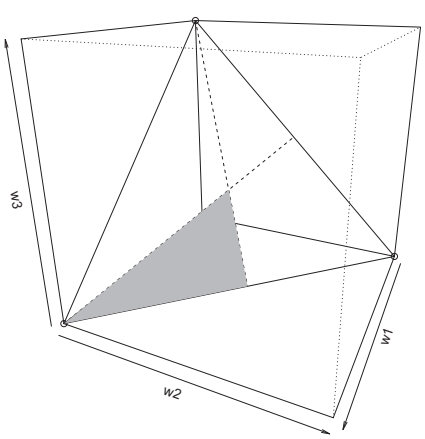
To preserve the uniform distribution of the samples, we need the transformation to preserve distances between points: it must be *isometric*. Specifically, in Euclidian space, a transformation f is isometric if $\|y - x\| = \|f(y) - f(x)\|$ for all x, y. First note that B is unitary ($B^TB = I_{n+1}$) because it is the homogeneous coordinate representation of an orthonormal change of basis matrix. As (i) unitary matrices are isometric (Berberian, 1999) and (ii) translation is isometric, it follows that the proposed transformation is also isometric. Therefore uniform samples obtained in R^{n-1} are also uniform when transformed to W_n^*.

To complete the transformation, we show how the linear constraints that define the polytope $W' \subseteq W_n$ are defined in $n-1$ dimensions. Let us denote the constraint set defining W_n as follows:

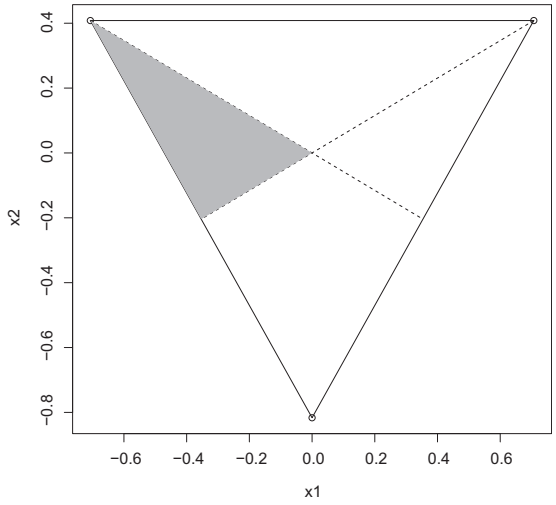
$$Cw \leqslant b; \quad C = -1I_n, \; b = (0, 0, \ldots, 0)^T$$
$$\sum_{i=1}^n w_i = 1$$

Since we sample directly from the plane W_n^*, the equality constraint is satisfied by definition. The weight constraints presented in Section 2 are linear and can therefore be represented as additional rows in C and b. Then the constraints can be expressed in $n-1$ dimensions as:

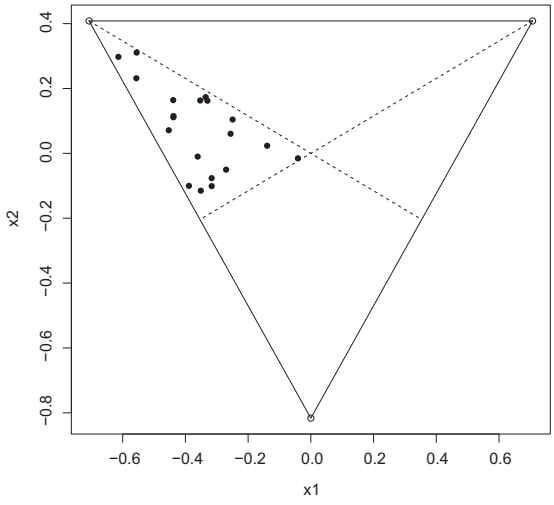
$$Ax \leqslant b; \quad A = CTB$$



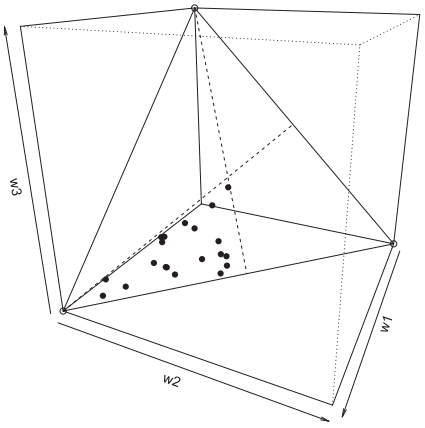
(a) Polytope in 3D (weight space)



(b) Polytope in 2D (sampling space)



(c) Samples in 2D (sampling space)



(d) Samples in 3D (weight space)

Fig. 2. The sampling space transformation illustrated in 3D space. The preference information (expressed as linear constraints) defines a 2D convex polytope in the 3D weight space (a). The linear constraints are transformed to a 2D space (b). Then, HAR is used to sample from the polytope (c). Finally, the samples are transformed back to the 3D weight space (d).

since $Ax = C(TBx) = Cw$.

The transformation is illustrated for a three-dimensional weight space with ordinal constraints ($w_1 > w_2 > w_3$) in Fig. 2. First, the linear constraints that define the polytope in the n dimensional weight space are transformed to the $n-1$ dimensional sampling space. Then, HAR is applied to generate an (approximately) uniform sample in the sampling space. Finally, the sampled points are mapped to the weight space.

3.2. Line intersection

Given a point x (in homogeneous coordinates) in the polytope and a direction vector $d = (d_1, \ldots, d_{n-1}, 0)$, the line through x along d is $x + ld$. We want to find the interval $L = [L_0, L_1]$ such that $A(x + ld) \leqslant b$ iff $l \in L$. Then $lAd \leqslant b - Ax$, where either side is a vector: $lu \leqslant v$. Since all v_i are non-negative, positive u_i give the upper bound L_1 and negative u_i the lower bound L_0. If u_i is zero the direction d is parallel to the ith constraint and provides no information on the bounds. The lower and upper bound are thus given by:

$$L_0 = \max_{i:u_i<0} \frac{v_i}{u_i}; \quad L_1 = \min_{i:u_i>0} \frac{v_i}{u_i}$$

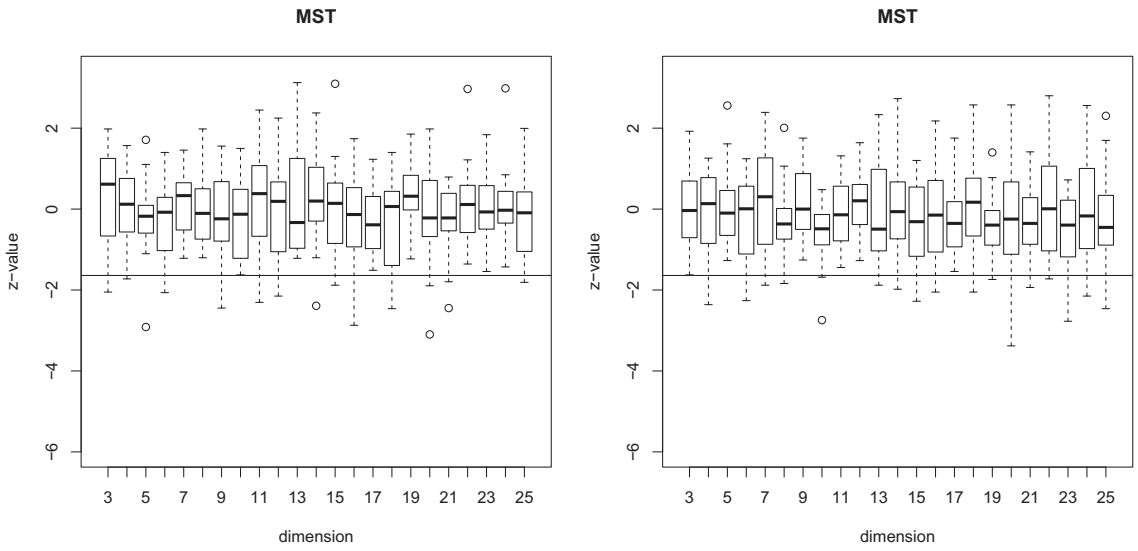
3.3. Starting point

The starting point can be defined in several ways, either deterministically or in a pseudo-random manner. The underlying principle is the same for all techniques: we first determine a set of points within the polytope, and then take their weighted average. Extreme points along each dimension $k \in \{1, \ldots, n-1\}$ can be found by solving the following linear programs:

maximize x_k	minimize x_k
subject to $Ax \leq b$	subject to $Ax \leq b$

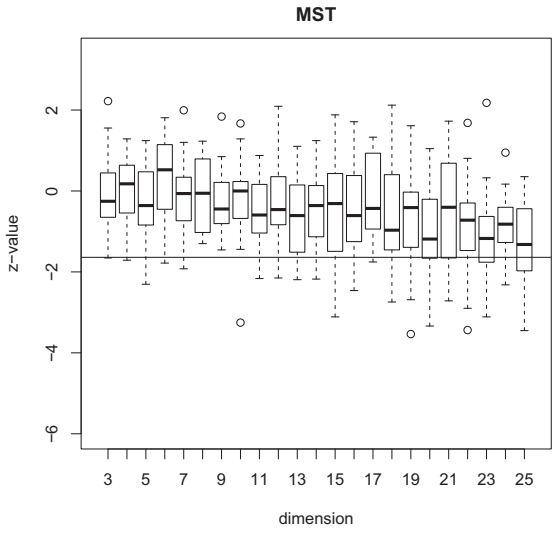
The obtained solutions are not necessarily the vertices of the polytope; if such are desired, they can be efficiently enumerated using the Avis–Fukuda pivoting algorithm (Avis and Fukuda, 1995).

If we choose a deterministic starting point, a point close to the centroid can be approximated by taking the mean of the extreme

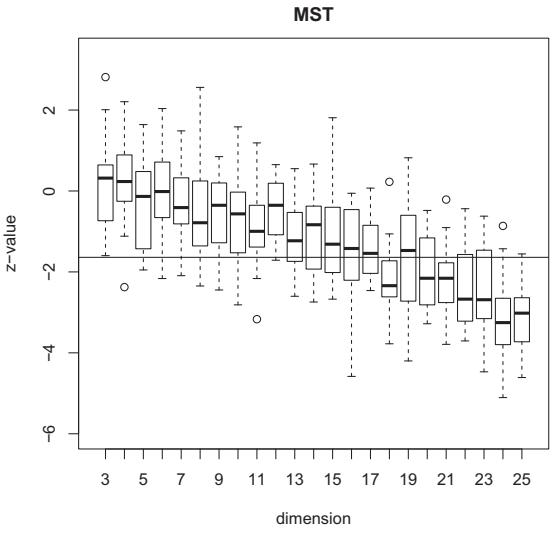


(a) benchmark

(b) HAR, $a = 1.0$



(c) HAR, $a = 0.5$



(d) HAR, $a = 0.25$

Fig. 3. Boxplot of MST z-values from validation tests with 20 sets of 10,000 samples for each n. The horizontal line at $z \approx -1.64$ corresponds to p-value 0.05, below which the null hypothesis of uniformity is rejected. Figure (a) shows how the MST test performs when comparing two uniform samples.

values along a subset of the dimensions. The exact centroid is the mean of all the vertices. To generate a pseudo-random starting point, we can use any set of interior points and take their weighted average with randomly generated weights.

4. Convergence metrics

With HAR, as with all MCMC samplers, it is important to assess whether a generated sample is representative of the target distribution, and this is determined by the mixing of the chain together with the number of iterations. We assume that the samples drawn from the restricted weight space will subsequently be used in simulation-based decision analysis, where the computed indices are insensitive to sample autocorrelation: as long as the draws uniformly cover the restricted weight space, the mixing itself is irrelevant for the computation of the indices describing the results of the decision analysis. In most MCMC applications, it is important to disregard an initial subsample from the Markov Chain because the starting values may have been very unlikely, meaning that the initial iterations may have been used to move away from a low density region, leading to disproportionate representation of that part of the space and thus to biased results. This is not an issue in our setting as the sampling distribution is uniform and HAR mixes fast from any starting point (Lovász and Vempala, 2006).

Many commonly used convergence diagnostics for MCMC perform poorly because they are based on the first and second moments of the distribution (for each dimension separately) and depend on the target distribution being continuous. The constrained weight polytope can be quite irregularly shaped, and the target distribution is discontinuous at its boundaries. Moreover, these diagnostics do not assess uniformity of the samples directly, but rather the stability of certain moments or quantiles. Thus, to measure convergence to the target distribution, a metric is needed that takes into account the geometry of the polytope and measures the uniformity of the samples.

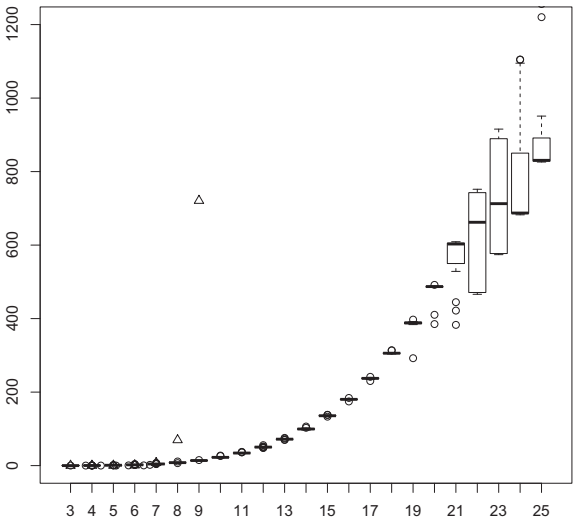


Fig. 4. Boxplot of execution times (in seconds, y-axis) of 20 runs for each dimension n (x-axis) to generate 10,000 samples at thinning $\varphi_{1.0}(n)$. The triangles show the running time for a rejection sampler for up to $n = 9$; with $n = 10$ criteria rejection sampling took 8×10^3 seconds.

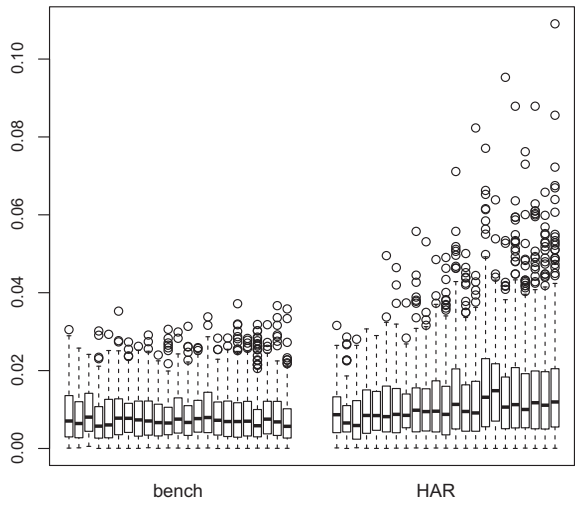


Fig. 5. Box plot for the SCE metric, for both the benchmark sets and the test sets generated with thinning $\varphi_{1.0}(n)$. The horizontal axis shows the number of criteria n.

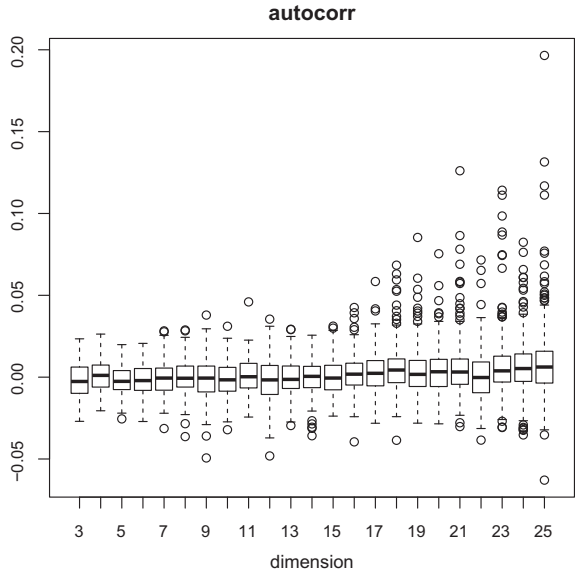


Fig. 6. Box plots for autocorrelation at lag $\tau = 25$ for test sets generated with thinning $\varphi_{1.0}(n)$. The horizontal axis shows the number of criteria n.

Our main metric for assessing convergence to the target distribution is the Friedman–Rafsky two-sample Minimum Spanning Tree (MST) test (Friedman and Rafsky, 1979; Smith and Jain, 1984) that compares the obtained sample Y with a sample X from the target distribution. Thus, this test is useful for evaluating the convergence of HAR when a sample from the target distribution can be obtained efficiently, but cannot be used to assess convergence in a general setting. The test assesses whether X and Y were drawn from the same distribution by constructing a MST for $X \cup Y$ and counting the number of within- and across-sample edges. Finally, a z-value for the null-hypothesis that both samples are from the same distribution is computed, which can be compared against quantiles of the normal distribution. A z-value in the lower tail indicates that Y is concentrated in sub-regions of the polytope (aggregation), whereas a z-value in the upper tail indicates Y is more regularly spaced than expected (regularity). For our purposes, only the lower-tail alternative hypothesis of aggregation is relevant.

Because the MST-based metric is not generally applicable, we also evaluate three other metrics for assessing convergence. The first is the Coefficient of Variation (COV), σ/μ, of the nearest-neighbour distances $\Gamma = \{\min_{j \neq i} |y_i - y_j| : y_i \in Y\}$, where μ and σ are the mean and standard deviation of the corresponding distance for draws from the sample distribution. Lower values of COV indicate more regular spacing of the sample points. The second metric, the Component-wise Error (SCE), is based on the fact that the mean of a sample from the target distribution will closely approximate the centroid h. Thus, $e = \mu - h$ is a measure of how close Y is to the target distribution. The component-wise errors e_i depend on the shape and dimension of the polytope, so they need to be standardized in order to be useful as a convergence metric. Thus, we define the SCE as e_i/s_i, where s_i is the sample standard deviation of the ith vector component in Y. Finally, we use the sample autocorrelation at a given lag τ as a convergence metric. This measure is defined as:

$$R(\tau) = \frac{E[(y_t - \mu)(y_{t+\tau} - \mu)]}{\sigma^2}$$

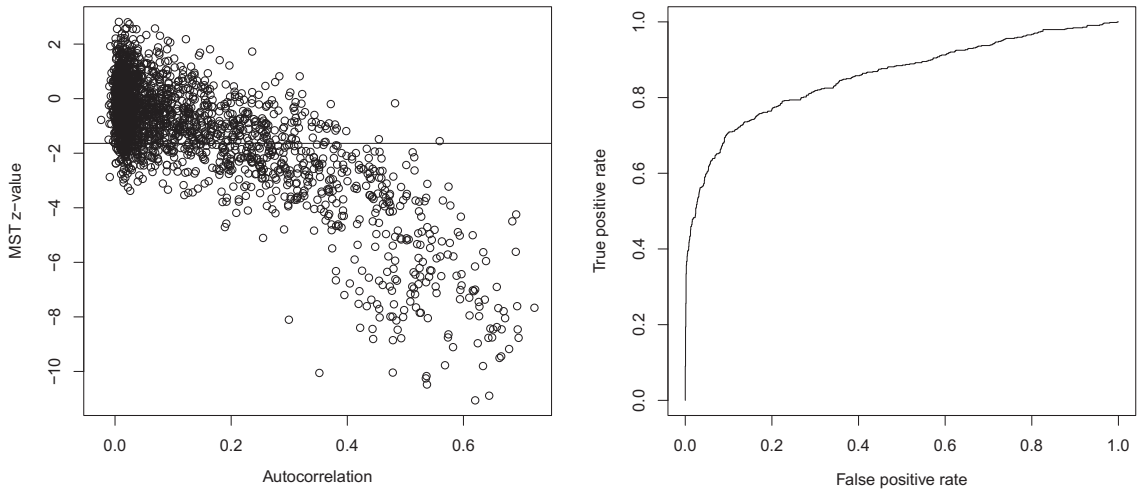
where $t \in \{1, \ldots, |Y| - \tau\}$. Like the SCE, autocorrelation is calculated individually for each component of the weight vector. Both metrics are scale invariant, and therefore the component-wise scores can be aggregated into an overall score simply by taking the maximum.

5. Computational tests

We assessed the thinning factors $\varphi_a(n)$ required to achieve uniformity of HAR sampling with constraints representing complete ordinal preference information ($w_1 > w_2 > \ldots > w_n$). This class of problem instances enables generating the sample X as required by the MST test by using an efficient algorithm (Tervonen and Lahdelma, 2007) and calculating the centroid $h = (h_1, \ldots, h_n)$ as required by the SCE test as

$$h_i = \frac{1}{n} \sum_{j=i}^{n} \frac{1}{j}$$

The amount of thinning required depends on the shape of the convex polytope W', with more conical shapes requiring higher thinning. The ordinal information causes W' to be at least as conical as what is expected with other realistic weight constraints because the decision makers often express imprecise weight information with similar precision for all criteria/pairs of criteria.



(a) Scatter plot (b) ROC curve

Fig. 7. Autocorrelation at lag $\tau = 25$ is predictive for failing the MST test.

As an initial exploratory test, we constructed a large chain for $n \in \{3,\ldots,15\}$ and recorded the minimum sufficient thinning factors required to obtain a z-value $\geqslant -1.64$ from the MST test and a maximum SCE < 0.05. Previous research suggests that HAR mixes with $O^*(n^3)$ iterations (Lovász, 1999). Because the algorithm reduces to uniform sampling for $n = 2$ (since we sample in $n-1$ dimensions), the required thinning for $n = 2$ should be 1. Thus we fitted $\varphi_a(n) = a(n-1)^3 + (1-a)$ to our exploratory test data, which suggested that $a \geqslant 0.2$. However, there was a large degree of uncertainty in this estimate because we used only a single chain for each dimension and censoring occurred for some dimensions. Based on visual inspection of autocorrelation plots of the exploratory test data, we chose to use a lag of $\tau = 25$ for the autocorrelation metric in the validation tests. We also performed the exploratory tests with the standard Gibbs sampler, and found it to exhibit slower convergence than HAR.

We subsequently generated 20 HAR samples of 10,000 weight vectors (sufficient for SMAA analyses; Tervonen and Lahdelma, 2007) for each $n \in \{3,\ldots,25\}$ and for each thinning factor $\varphi_a(n)$ with $a \in \{0.125, 0.25, 0.5, 0.75, 1.0\}$. For the MST test, a separate benchmark sample X was generated for each HAR sample Y to marginalize the impact of random properties of a single benchmark sample. This led to respectively 68%, 38%, 14%, 7% and 8% of the HAR samples being rejected (Fig. 3). Thus, either $a = 0.75$ or $a = 1.0$ could be appropriate, as they are close to the optimal 5% rejection rate. As the shape of the sampling space affects the rate of convergence, we recommend to use $a = 1.0$. The running times that were required to generate 10,000 samples with thinning factors $\varphi_{1.0}(n)$ on an Intel Xeon X3440 2.53 GHz CPU are shown in Fig. 4. To illustrate the necessity of an MCMC approach, running times for rejection sampling are also shown: with $n = 3$ criteria, one in three samples are accepted; with $n = 10$ criteria one in 1.8 million, giving a running time of several hours for our implementation. The plotted sampling times show that by using HAR reasonably high dimensionality problems can be analyzed almost interactively, i.e. the median sampling times for 5, 10, and 15 dimensions were 0.92, 22, and 136 seconds, respectively.

We also computed the other test metrics (COV, SCE, and autocorrelation at lag $\tau = 25$) for both the HAR samples and the benchmark samples. The COV metric converged to stable values long before the sample converged to uniformity and therefore it is not useful, so we omit the detailed results for this metric. Fig. 5 shows how the SCE values for the HAR samples compared to those for the benchmark samples. The autocorrelation is plotted in Fig. 6. The SCE and autocorrelation metrics appear to be useful and are correlated with the MST z-value ($\rho = -0.60$ and $\rho = -0.79$, respectively). Since autocorrelation can be calculated based on the sample alone, while computation of the SCE requires enumerating the vertices of the polytope, the former is potentially the most useful metric. Fig. 7 shows a scatter plot of the autocorrelation value against the MST z-value, the Receiver Operating Characteristic (ROC) curve of autocorrelation as a predictor of $z < -1.64$. An autocorrelation cutoff value of 0.05 correctly rejected 82% of the samples that failed the MST test. Note that the samples that passed the autocorrelation test but failed the MST test are still likely to be sufficiently uniform to be used in simulation-based decision analysis as small deviations from uniformity are unlikely to affect the model outcomes. The code used to generate all the test results and the complete set of summary statistics are available as online supplements.

6. Conclusions

In this paper, we considered the application of HAR to sample uniformly from a subset of the n-simplex defined by linear constraints. Our contribution was in presenting the transformation to $n-1$ dimensions for efficient MCMC sampling and in assessing the thinning factor required to achieve acceptable deviation from a uniform distribution over the constrained weight space. The transformation is not specific to HAR, and can also be used to apply other MCMC samplers to the weight sampling problem. The computational tests showed that HAR is quite fast in small ($n \leqslant 15$) problems for 10,000 samples. The tests also showed that a thinning factor of $\varphi(n) = (n-1)^3$ is almost always sufficient to be unable to reject the null hypothesis of uniformity with the MST-test at the 0.05 confidence level. We also assessed other measures of convergence to the target distribution and found that the autocorrelation at lag 25 is the most appropriate to use in absence of a representative sample from the target distribution (meaning that the MST test is inapplicable). However, autocorrelation does not measure convergence directly and future research could yield better tests.

When used in the context of simulation-based decision analysis, HAR is slower than the techniques presented by Tervonen and Lahdelma (2007) for sampling weights representing missing or ordinal preferences. However, HAR is far more flexible as it also allows for polynomial-time generation of upper bounded and ratio interval constrained weights that previously had to be generated

with rejection sampling. Especially the ratio interval weights are important, as the meaning of a weight in MAVT models is that of trade-off ratios, and imprecision should therefore be modeled in a compatible manner.

The sampling technique evaluated in this paper assumes that the sampling space is convex, and this is fulfilled by all weight constraints within the additive model. However, some preference models include other parameters as well, and in that case the space of feasible preference parameters can be concave, e.g. the general monotone value functions of robust ordinal regression (Greco et al., 2008, 2010). Future research should investigate whether MCMC sampling can be applied for the efficient sampling of the preference parameters in such spaces as well.

References

Amit, Y., Grenander, U., 1991. Comparing sweep strategies for stochastic relaxation. Journal of Multivariate Analysis 37, 197–222.

Avis, D., Fukuda, K., 1995. A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra. Discrete & Computational Geometry 8, 295–313.

Belton, V., Stewart, T.J., 2002. Multiple Criteria Decision Analysis – An Integrated Approach. Kluwer Academic Publishers.

Berberian, S.K., 1999. Introduction to Hilbert Space. American Mathematical Society.

Besag, J., Green, P., Higdon, D., Mengersen, K., 1995. Bayesian computation and stochastic systems. Statistical Science 10, 3–41.

Butler, J., Dia, J., Dyer, J., 1997. Simulation techniques for the sensitivity analysis of multi-criteria decision models. European Journal of Operational Research 103, 531–545.

Choo, E.U., Schoner, B., Wedley, W.C., 1999. Interpretation of criteria weights in multicriteria decision making. Computers & Industrial Engineering 37, 527–541.

Friedman, J., Rafsky, L., 1979. Multivariate generalizations of the Wald–Wolfowitz and Smirnov two-sample tests. The Annals of Statistics 7, 697–717.

Greco, S., Mousseau, V., Słowiński, R., 2008. Ordinal regression revisited: multiple criteria ranking using a set of additive value functions. European Journal of Operational Research 191, 415–435.

Greco, S., Mousseau, V., Słowiński, R., 2010. Multiple criteria sorting with a set of additive value functions. European Journal of Operational Research 207, 1455–1470.

Jia, J., Fischer, G.W., Dyer, J.S., 1998. Attribute weighting methods and decision quality in the presence of response error: a simulation study. Journal of Behavioral Decision Making 11, 85–105.

Keeney, R., Raiffa, H., 1976. Decisions with Multiple Objectives: Preferences and Value Tradeoffs. Wiley, New York.

Lahdelma, R., Salminen, P., 2001. SMAA-2: Stochastic multicriteria acceptability analysis for group decision making. Operations Research 49, 444–454.

Lovász, L., 1999. Hit-and-run mixes fast. Mathematical Programming 86, 443–461.

Lovász, L., Vempala, S., 2006. Hit-and-run from a corner. SIAM Journal on Computing 35, 985–1005.

Marsaglia, G., 1972. Choosing a point from the surface of a sphere. Annals of Mathematical Statistics 43, 645–646.

Roy, B., 1996. Multicriteria Methodology for Decision Analysis. Kluwer Academic Publishers.

Smith, R.L., 1984. Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. Operations Research 32, 1296–1308.

Smith, S., Jain, A., 1984. Testing for uniformity in multidimensional data. IEEE Transactions on Pattern Analysis and Machine Intelligence 6, 73–81.

Tervonen, T., Figueira, J.R., 2008. A survey on stochastic multicriteria acceptability analysis methods. Journal of Multi-Criteria Decision Analysis 15, 1–14.

Tervonen, T., Lahdelma, R., 2007. Implementing stochastic multicriteria acceptability analysis. European Journal of Operational Research 178, 500–513.

Tervonen, T., Figueira, J.R., Lahdelma, R., Almeida Dias, J., Salminen, P., 2009. A stochastic method for robustness analysis in sorting problems. European Journal of Operational Research 192, 236–242.

Vansnick, J.-C., 1986. On the problem of weights in multiple criteria decision making (the noncompensatory approach). European Journal of Operational Research 24, 288–294.