FH AACHEN
UNIVERSITY OF APPLIED SCIENCES

Fachhochschule Aachen
Location Cologne

Bayesian statistics
in the context of Multi Criteria Decision Analysis (MCDA)

Seminal thesis

Luis Hasenauer

Faculty of Medical Engineering and Technomathematics
Study program Applied Mathematics and Computer Science B.Sc.

Cologne, December 16, 2022

Abstract

A common way to model decision makers' preferences in Multi Criteria Decision Analysis (MCDA) problems is through weights representing the relative importance of specific criteria. The elicitation of these weights is a significant task that has an immediate impact on the reliability and usefulness of the decision-making process. Bayesian inference can be used to update prior beliefs about a statistical model by combining them with information contained in observed data. This can be used in an MCDA context to gain insight on how specific decisions have been made based on their outcome. Using Bayes' theorem, these past decisions and prior knowledge about the field of application are combined in a posterior distribution which can be rather difficult to compute due to high dimensionality. Consequently, Monte Carlo methods need to be used to gain an approximate insight on it. It will be shown that this probability problem also translates into a geometric problem of a weight laying in a polytope. Furthermore, this method will be successfully applied to a numerical example on the German car market.

Contents

1 Objectives and Motivation

In many situations, a decision maker needs to rank alternatives according to particular criteria. These kind of choices can be made using MCDA by, among other things, assigning a weight to each of the evaluation criteria representing how much the individual criteria should contribute to the final decision. MCDA methods have, for example, been successfully applied in the fields of economics, social- and health science.

Bayesian statistics includes methods for updating prior knowledge about a field of application, such as previous results or personal beliefs, after obtaining new data. These methods are centered around Bayes' theorem and result in a joint probability distribution of the parameters in a statistical model.

Even though MCDA methods may be helpful for decision-making, this seminal paper focuses on using them for decision analysis. This indicates attempting to comprehend a decision that has already been made as opposed to making a decision. The goal is to present a method for combining aspects from MCDA and prior information about a decision into a distribution of criteria weights using Bayesian statistics, as well as how to compute this distribution and applying it to an example.

2 Multi Criteria Decision Analysis

2.1 Introduction

MCDA covers a collection of methods to choose between or rank alternatives on the basis of various criteria taken from the field of application.

Let $A = \{a_1, a_2, \ldots, a_m\}$ be a set of possible alternatives to be ranked by a set of evaluation criteria $\{c_1(\cdot), c_2(\cdot), \ldots, c_n(\cdot) | c : A \to \mathbb{R}\}$. A distinction is made between benefit and cost criteria, meaning an alternative a_i is preferred over alternative a_j in terms of criterion c if either $c(a_i) > c(a_j)$ in case of a benefit criterion or $c(a_i) < c(a_j)$ in case of a cost criterion.

These alternatives and criteria form a performance matrix $\tilde{P} \in \mathbb{R}^{m \times n}$ where $\tilde{P}_{ij}$ is the performance of alternative a_i with respect to criterion c_j.

$$\tilde{P} = \begin{pmatrix} c_1(a_1) & c_2(a_1) & \ldots & c_n(a_1) \\ c_1(a_2) & c_2(a_2) & & \\ \vdots & & \ddots & \\ c_1(a_m) & & & c_n(a_m) \end{pmatrix}$$

MCDA methods also include the decision makers' preferences by specifying a weight vector $w = (w_1, \ldots, w_n)^T$ with w_i representing the relative importance of criterion c_i. The set of possible weight vectors $S_{n-1} \subset \mathbb{R}^n$ is given by the following constraints:

$$w \in S_{n-1} \Leftrightarrow \sum_{i}^{n} w_i = 1 \ \wedge \ w_i \geq 0 \tag{2.1}$$

It should be noted that choosing a weight vector strongly influences the final ranking. However, it can be very subjective and sensitive to the decision makers opinion.

$\tilde{P}$ and w form the standard input of many MCDA methods. The goal is to compute a score φ, representing the eligibility of an alternative, for each alternative which allows them to be, at least partially, ranked. Many of these methods have the form

$$f(\tilde{P}, w) = r(\varphi) \tag{2.2}$$

with $\varphi = Pw$ and P being a matrix obtained from $\tilde{P}$ according to the chosen MCDA method. The following sections will only cover methods which allow a complete ranking of alterna-

tives. An alternative a_i is preferred over alternative a_j if $\varphi_i > \varphi_j$, hence the mapping $r(\cdot)$ sorts the alternatives in descending order according to their values in φ.

2.2 Simple Additive Weighting

Simple Additive Weighting (SAW) is the oldest, most widely known and practically used method. It is also a very simple MCDA method since it does not require any other parameters than the performance matrix $\tilde{P}$ and the weight vector w.

To use SAW, $\tilde{P}$ needs to be transformed to a matrix P_{SAW} by normalizing it to a scale that can be compared with all the ratings of existing alternatives [1].

$$P_{SAW_{ij}} = \begin{cases} \dfrac{\tilde{P}_{ij}}{\max\limits_{1 \le k \le m} \tilde{P}_{kj}} & \text{if the criteria is benefit} \\ \\ \dfrac{\min\limits_{1 \le k \le m} \tilde{P}_{kj}}{\tilde{P}_{ij}} & \text{if the criteria is cost} \end{cases} \tag{2.3}$$

Another restriction on SAW is that all criteria values $\tilde{P}_{ij}$ must be positive. In the opposite circumstances, negative values are converted to positive values by, for instance, using the formula [2]:

$$P_{SAW_{ij}} = \tilde{P_{ij}} + |\min_{1 \le k \le m} \tilde{P}_{kj}| + 1$$

Since the criteria would not end up in comparable units otherwise, this transformation, which results in the smallest negative value becoming unity, should be performed prior to the normalization process as in (2.3).

After transforming $\tilde{P}$ to P_{SAW}, the ranking can then be created according to (2.2).

2.3 Preference Ranking Organisation Method for Enrichment Evaluation II

The Preference Ranking Organisation Method for Enrichment Evaluation II (PROMETHEE II) is another quite popular MCDA method with more than 2300 references according to the website [3] of one of the authors [4], B. Mareschal. Unlike SAW, PROMETHEE II allows the modeler to feed the algorithm with a few more variables based on his choice of the preference function. It is important to note that, in general, MCDA methods place decision makers at the center of the process and that different decision makers can model the problem in a variety of ways based on their preferences.

A preference degree in PROMETHEE II expresses how one alternative is preferred over another alternative. The decision maker can assign a small preference for minor variations in the assessments of a pair of criteria. The opposite is true for large deviations, where the decision maker must allocate a significant preference for one alternative over another. If the deviation exceeds a predetermined threshold set by the decision maker, there is an absolute preference for one course of alternative over the other. This preference degree is always represented by a real number between 0 and 1.

The preference function of a benefit criterion, which is to be maximized, can be defined as

$$F_j(a,b) = g_j\Big[d_j(a,b)\Big], \quad \forall a,b \in A \tag{2.4}$$

where $d_j(a,b)$ is the pairwise comparison of two alternatives.

$$d_j(a,b) = c_j(a) - c_j(b)$$

If criterion c is associated to cost, then $-c$ is to be maximized.

The authors [4] propose a total of six preference functions. These types are well-accepted and frequently used in literature. The purpose of these preference functions is to simulate how a decision maker would choose one course of alternative over another, in an effort to simulate his viewpoint on the issue at hand.

Type 1, the usual criterion:

$$g(d) = \begin{cases} 0 & d \leq 0 \\ 1 & d > 0 \end{cases}$$

Type 2, the U-shape criterion:

$$g(d) = \begin{cases} 0 & d \leq q \\ 1 & d > q \end{cases}$$

Type 3, the V-shape criterion:

$$g(d) = \begin{cases} 0 & d < 0 \\ \frac{d}{p} & 0 \leq d \leq p \\ 1 & d > p \end{cases}$$

Type 4, the level criterion:

$$g(d) = \begin{cases} 0 & d \leq q \\ \frac{1}{2} & q < d \leq p \\ 1 & d > p \end{cases}$$

Type 5, the V-shape with indifference criterion:

$$g(d) = \begin{cases} 0 & d \leq q \\ \frac{d-q}{p-q} & q < d \leq p \\ 1 & d > p \end{cases}$$

Type 6, the Gaussian criterion:

$$g(d) = \begin{cases} 0 & d \leq 0 \\ 1 - e^{-\frac{d^2}{2s^2}} & d > 0 \end{cases}$$

The aggregated preference indices can be calculated as followed:

$$\pi(a,b) = \sum_{j=1}^{n} F_j(a,b) w_j \tag{2.5}$$

where $(a,b) \in A$ and $\pi(a,b)$ indicate how much alternative a is preferred to b with regard to all criteria.

Each alternative is competing against $(m-1)$ other alternatives in the set A. The unicriterion positive flow of each alternative in A is a number between 0 and 1 and is an indicator of how much this alternative is preferred over all the other alternative in A. The higher this value is, the more this alternative is preferable for this particular decision maker. Therefore, the definition for the positive outranking flow is as

$$\Phi^+(a) = \frac{1}{m-1} \sum_{b \in A} \pi(a,b)$$

On the other hand, the negative outranking flow is an indicator of how all the other alternatives are preferred over this particular alternative and in accordance with the positive flow is defined as

$$\Phi^{-}(a) = \frac{1}{m-1} \sum_{b \in A} \pi(b, a)$$

PROMETHEE II then relies on the net flow Φ to construct the final ranking. The net flow is definded as

$$\Phi(a) = \Phi^{+}(a) - \Phi^{-}(a) \tag{2.6}$$

Finally, the ranking can be determined as stated in (2.2).

It can be shown that PROMETHEE II, just like SAW, can be written in the form $\varphi = Pw$. Combining (2.4), (2.5) and (2.6) leads to:

$$\begin{aligned}
\varphi &= \begin{pmatrix} \Phi(a_1) \\ \vdots \\ \Phi(a_m) \end{pmatrix} \\
&= \begin{pmatrix} \Phi^{+}(a_1) - \Phi^{-}(a_1) \\ \vdots \\ \Phi^{+}(a_m) - \Phi^{-}(a_m) \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{m-1} \sum_{i=1}^{m} \pi(a_1, a_i) - \frac{1}{m-1} \sum_{i=1}^{m} \pi(a_i, a_1) \\ \vdots \\ \frac{1}{m-1} \sum_{i=1}^{m} \pi(a_m, a_i) - \frac{1}{m-1} \sum_{i=1}^{m} \pi(a_i, a_m) \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{m-1} \sum_{i=1}^{m} [\pi(a_1, a_i) - \pi(a_i, a_1)] \\ \vdots \\ \frac{1}{m-1} \sum_{i=1}^{m} [\pi(a_m, a_i) - \pi(a_i, a_m)] \end{pmatrix}
\end{aligned}$$

$$
\begin{aligned}
&= \begin{pmatrix} \frac{1}{m-1}\sum_{i=1}^{m}\left[\sum_{j=1}^{n} g_j(\tilde{P}_{1j}-\tilde{P}_{ij})w_j - \sum_{j=1}^{n} g_j(\tilde{P}_{ij}-\tilde{P}_{1j})w_j\right] \\ \vdots \\ \frac{1}{m-1}\sum_{i=1}^{m}\left[\sum_{j=1}^{n} g_j(\tilde{P}_{mj}-\tilde{P}_{ij})w_j - \sum_{j=1}^{n} g_j(\tilde{P}_{ij}-\tilde{P}_{mj})w_j\right] \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{m-1}\sum_{i=1}^{m}\sum_{j=1}^{n}\left[g_j(\tilde{P}_{1j}-\tilde{P}_{ij}) - g_j(\tilde{P}_{ij}-\tilde{P}_{1j})\right]w_j \\ \vdots \\ \frac{1}{m-1}\sum_{i=1}^{m}\sum_{j=1}^{n}\left[g_j(\tilde{P}_{mj}-\tilde{P}_{ij}) - g_j(\tilde{P}_{ij}-\tilde{P}_{mj})\right]w_j \end{pmatrix} \\
&= \underbrace{\frac{1}{m-1}\begin{pmatrix} \sum_{i=1}^{m}\left[g_1(\tilde{P}_{11}-\tilde{P}_{i1}) - g_1(\tilde{P}_{i1}-\tilde{P}_{11})\right] & \dots & \sum_{i=1}^{m}\left[[g_n(\tilde{P}_{1n}-\tilde{P}_{in}) - g_n(\tilde{P}_{in}-\tilde{P}_{1n})\right] \\ \vdots & & \vdots \\ \sum_{i=1}^{m}\left[[g_1(\tilde{P}_{m1}-\tilde{P}_{i1}) - g_1(\tilde{P}_{i1}-\tilde{P}_{m1})\right] & \dots & \sum_{i=1}^{m}\left[[g_n(\tilde{P}_{mn}-\tilde{P}_{in}) - g_n(\tilde{P}_{in}-\tilde{P}_{mn})\right] \end{pmatrix}}_{P_{PROM2}} \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}
\end{aligned}
$$

Thus,

$$P_{PROM2_{ij}} = \frac{1}{m-1}\sum_{k=1}^{m}\left[g_j(\tilde{P}_{ij}-\tilde{P}_{kj}) - g_j(\tilde{P}_{kj}-\tilde{P}_{ij})\right]$$

2.4 Geometric aspects

Equation (2.1) defines the standard simplex S_{n-1} in $\mathbb{R}^n$. The standard simplex is constructed by intersecting the half spaces $\langle w, e_k\rangle \geq 0$ and the hyperplane $\langle w, \mathbb{1}\rangle = 1$ with e_k being the standard unit vector that has 1 as the k-th component and 0 elsewhere and $\mathbb{1}$ being a vector containing n ones. As $\langle w, \mathbb{1}\rangle = 1$ is another intersection of half spaces, $\langle w, \mathbb{1}\rangle \leq 1$ and $\langle w, \mathbb{1}\rangle \geq 1$, S_{n-1} is an intersection of finitely many half spaces which is known to be a polytope [5]. Since half spaces are convex and the intersection of convex sets is also convex, S_{n-1} is a convex polytope and bounded by the coordinate hyperplanes.

The height of the standard n-simplex $S_n \subset \mathbb{R}^{n+1}$ is the minimal distance between the corner $e_1 = (1,0,0,\dots,0)^T$ and a point on the opposite facet, the convex hull of $(e_2,\dots,e_{n+1})$. In other words some $p = (0, p_2, p_3, \dots, p_{n+1})^T$ with $p_2 + p_3 + \dots + p_{n+1} = 1$. For reasons of symmetry, this point has to be in the center of the opposite face, hence $p_2 = p_3 = \dots = p_{n+1} = \frac{1}{n}$. Therefore the height is

$$h_n = \|e_1 - p\|_2 = \sqrt{1 + n\left(\frac{1}{n}\right)^2} = \sqrt{1 + \frac{1}{n}} = \sqrt{\frac{n+1}{n}} \tag{2.7}$$

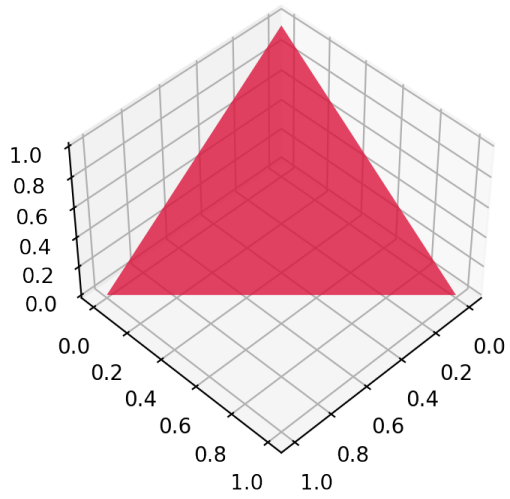

Figure 1: *The standard 2-simplex*

Additionally, the n-dimensional volume $Vol_n(S_n)$ has the following dimensional recursion [6]:

$$Vol_n(S_n) = \frac{1}{n} Vol_{n-1}(S_{n-1}) h_n$$

Thus, using $Vol_1(S_1) = \|e_1 - e_2\|_2 = \sqrt{2}$ as the recursion start leads to

$$\begin{aligned} Vol_n(S_n) &= \sqrt{2} \prod_{d=2}^{n} \frac{h_d}{d} \\ &= \frac{\sqrt{2}}{n!} \prod_{d=2}^{n} h_d \qquad \text{| insert (2.7)} \\ &= \frac{\sqrt{2}}{n!} \prod_{d=2}^{n} \sqrt{\frac{d+1}{d}} \\ &= \frac{\sqrt{2}}{n!} \sqrt{\frac{n+1}{2}} \\ &= \frac{\sqrt{n+1}}{n!} \end{aligned} \tag{2.8}$$

Therefore the n-dimensional volume of the standard n-simplex is $\frac{\sqrt{n+1}}{n!}$. On the other hand, its $n+1$-dimensional volume is zero because it lays in an n-dimensional subspace.

Let $W_r \subseteq S_{n-1}$ be the set of weight vectors that result in a specific ranking given a fixed performance matrix $\tilde{P}$, meaning $f(\tilde{P}, w) = f(w)$. This ranking may be partial, in other words containing an arbitrary amount of consecutive entries starting from the best ranked alternative. Furthermore, let $f_d(w)$ be the first d entries of $f(w)$.

$$W_r = \left\{ w = (w_1, \ldots, w_n)^T \mid f_d(w) = r,\ 0 \leq d \leq m,\ w \in S_{n-1} \right\} \tag{2.9}$$

Consequently, W_r contains all possible weights vectors whose rankings first d entries match with the first d entries of r. An alternative a_i is preferred over an alternative a_j if $\varphi_i \geq \varphi_j$ with $\varphi = Pw$. This leads to

$$\varphi_i = \langle p_i^T, w \rangle$$

with p_i being the i-th row of P. Therefore an actor chooses a_i over a_j if

$$\langle p_i^T, w \rangle \geq \langle p_j^T, w \rangle \Leftrightarrow \langle p_j^T - p_i^T, w \rangle \leq 0$$

With $n_{ij} = p_j^T - p_i^T$, this defines a half space $H_{ij} = \left\{ w \in \mathbb{R}^n \mid \langle n_{ij}, w \rangle \leq 0 \right\}$. For all weights w in the interior of H_{ij}, alternative a_i is preferred over alternative a_j. For a given ranking r, the $d-1$ conditions $\langle p_{r_{j+1}}^T - p_{r_j}^T, w \rangle \leq 0,\ 1 \leq j < d$ must be fulfilled simultaneously. Since the alternatives not included in r should be considered inferior to all alternatives included in r, $n-d$ more conditions need to be satisfied, with those being $\langle p_{r_d}^T - p_i^T, w \rangle \leq 0$ for any such alternative a_i. Therefore W_r is an intersection of $n-1$ half spaces and the standard simplex S_{n-1} and thus also a bounded, convex polytope.

3 Bayesian statistics

3.1 Theroretical Background

Bayesian inference is used to update previous beliefs about a statistical model based on new data. The goal is to derive a posterior probability given a prior probability and a likelihood function using Bayes' theorem [7]. This can be stated as the following equation:

Theorem 3.1.1 (Bayes' theorem)

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

- $p(y|\theta)$ is called the likelihood and is the conditional probability of observing the data y given a hypothesis θ in form of a model parameter.
- $p(\theta)$ is called the prior which contains the previous knowledge about the model.
- $p(y)$ is called the model evidence and is often unknown.
- $p(\theta|y)$ is called the posterior distribution and represents the probability of the model parameter θ conditioned on the data y, hence the updated prior beliefs.

Since the evidence is often unknown but does not depend on θ, Bayes' theorem [(3.1.1)] can be simplified to

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

$p(y)$ can be viewed as a normalization constant to satisfy the condition of a probability density function $p(x)$, being $\int p(x)\,dx = 1$. Therefore

$$p(y) = \int p(y|\theta)p(\theta)\,d\theta \tag{3.1}$$

and

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)\,d\theta} \tag{3.2}$$

3.2 Formalizing prior distributions

Prior distributions can come in many different forms and can be classified as informative, weakly informative, or diffuse. The level of information reflected in a prior distribution can be anywhere on a continuum from complete uncertainty to relative certainty. Databased priors

can be elicited using methods such as maximum likelihood or sample statistics, but double-dipping procedures are not recommended [8]. Instead, a hierarchical modelling strategy can be implemented.

A researcher may want to use an informative prior when there is information suggesting a relationship between parameters, or a restriction on the possible range of a particular parameter. However, it is important to assess the impact of an informative prior on the posterior.

Alternatively, a weakly informative prior can be used when some information is assumed about a parameter, but there is still a desired degree of uncertainty. It is typically constructed by specifying a plausible parameter space, and excluding improbable values by attaining only a limited density mass at implausible values.

Moreover, a diffuse prior is used when there is a complete lack of certainty surrounding a model parameter. It represents a relatively flat density and does not include specific knowledge of the parameter. Diffuse priors can also have unintended consequences on the posterior. In addition, improper priors are sometimes used with the intention of using them as diffuse priors.

When the sample size is small, Bayesian estimation with mildly informative priors is often used, but the prior specification might still have a huge effect on the posterior results which may be inaccurate or unintended [9].

3.3 Determining the likelihood function

”The likelihood is used in both Bayesian and frequentist inference” [7] and is a function of the observed data y and the unknown parameter θ. A likelihood is not a probability itself but is proportional to one. Probability and likelihood differ significantly in how they interpret what is constant and what is variable. In the case of a conditional probability of some data y given a hypothesis θ, $p(y|\theta)$, the hypothesis is fixed and the data is free to vary. Likelihood, however, is the opposite. $p(y|\theta)$ is called the likelihood of a hypothesis θ if it is viewed as a function of θ to fixed data y. Since the likelihood is not a probability density function (PDF) itself but proportional to one, it is feasible to notate it as $\mathcal{L}(\theta;y)$ to avoid confusion. It follows that $\mathcal{L}(\theta;y)$ can be scaled to be a PDF using $p(y|\theta) = c \cdot \mathcal{L}(\theta;y)$ with $c = 1/\int \mathcal{L}(\theta;y)d\theta$.

To make probability statements about unknown parameters, Bayesian inference refers to them as random variables. While the parameter values are varied, the (observed) data is treated as fixed. Even though the likelihood is only proportional to a probability, both, the likelihood and the scaled likelihood, can be used in Bayesian inference, since inserting into

(3.2) leads to:

$$\begin{aligned} p(\theta|y) &= \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)\,d\theta} \\ &= \frac{c \cdot \mathcal{L}(\theta;y)p(\theta)}{\int c \cdot \mathcal{L}(\theta;y)p(\theta)\,d\theta} \\ &= \frac{c \cdot \mathcal{L}(\theta;y)p(\theta)}{c\int \mathcal{L}(\theta;y)p(\theta)\,d\theta} \\ &= \frac{\mathcal{L}(\theta;y)p(\theta)}{\int \mathcal{L}(\theta;y)p(\theta)\,d\theta} \end{aligned}$$

In conclusion, a likelihood function summarizes the following elements: a statistical model, a range of possible values for the unknown parameter and the observed data.

3.4 Application to MCDA

When applying Bayes' theorem to MCDA, the goal is a distribution of the weights w given a, possibly partial, fixed ranking $r = (r_1, \ldots, r_d)^T$ with $0 \leq d \leq m$, meaning the ranking can be any length between the number of alternatives, in other words a complete ranking, and empty, meaning no constraint on the ranking at all. The performance matrix P is also assumed to be fixed and therefore obtaining a ranking depends on the weight only, hence $r = f(w)$. It follows that the probability of a weight to have been used to obtain a given ranking is a conditional probability, $p(w|r)$, and can be computed using Bayes' theorem in the form

$$p(w|r) = \frac{p(r|w)p(w)}{p(r)}$$

The choice of prior distribution completely depends on the context of the MCDA problem and the actors prior knowledge or assumptions. One could also try different priors to investigate its impact on the posterior distribution. The only requirement to this prior is $W_r \cap \operatorname{supp}(p(\theta)) \neq \varnothing$ with $\operatorname{supp}(p(\theta)) = \overline{\{\theta \mid p(\theta) \neq 0\}}$ as the support of $p(\theta)$. Otherwise obtaining a posterior density is impossible.

The desired likelihood of obtaining a ranking, given a weight, can be modelled using a Bernoulli distribution, since obtaining a ranking can only result in 2 states, being either obtaining it or not obtaining it.
A Bernoulli distribution of a random variable X which takes the value 1 with probability p and the value 0 with probability $1 - p$ has the following probability mass function (PMF)

$$\mathbb{P}(X = k|p) = p^k(1-p)^{1-k}, \quad k \in \{0, 1\}$$

Let $X = 1 \Leftrightarrow f_d(w) = r$ and $X = 0 \Leftrightarrow f_d(w) \neq r$, with $f_d(w)$ being the first d entries of $f(w)$. Since the likelihood of obtaining a ranking is desired, the case $f_d(w) = r$ is considered and fixed. It follows that

$$\mathbb{P}(r|p) = p^1(1-p)^0 = p$$

The probability of $f_d(w) = r$ depends on w only because we set the performance matrix to be fixed. Let W_r be according to (2.9). Since it is impossible for a weight vector to result in the ranking if $w \notin W_r$ and guaranteed if $w \in W_r$, this probability problem can be translated to a geometric problem of a weight being contained in the polytope W_r. Thus, the likelihood results in the following

$$\mathcal{L}(w;r) = \begin{cases} 1 & \text{if } w \in W_r \\ 0 & \text{else} \end{cases}$$

and scaled to a PDF it results in

$$p(r|w) = \begin{cases} \dfrac{1}{Vol_n(W_r)} & \text{if } w \in W_r \\ 0 & \text{else} \end{cases}$$

Let r be a ranking of size d and $\hat{r}$ be a ranking of size $d+1$ with the first d entries of r and $\hat{r}$ matching. It follows that $W_{\hat{r}} \subseteq W_r$ and therefore $p(\hat{r}|w)$ results in a smaller parameter space than $p(r|w)$. Hence, a more precise ranking results in more impact on the posterior distribution.

3.5 Remarks

Since W_r lays in an $n-1$ dimensional linear subspace of $\mathbb{R}^n$, the posterior is going to be a degenerate function. In some cases, for example sampling, it is feasible to transform W_r to $\mathbb{R}^{n-1}$. One can choose an invertible affine transformation $T : W_r \to \mathbb{R}^{n-1}$ with $T(x) = Q(x - s^*)$ and $s^* = \mathbb{1}/n$ as the center of gravity of the standard $(n-1)$-simplex to shift and then rotate W_r to $\mathbb{R}^{n-1}$ and $Q \in \mathbb{R}^{n \times n-1}$ having orthonormal columns to preserve volume and distance.

This transformation obviously influences statistical metrics like the expected value or the covariance. Let w be a random variable before transforming and let w_T be the transformed random variable. The expected value of the transformed random variable is then given by:

$$\begin{aligned} \mathbb{E}\{w_T\} &= \mathbb{E}\{Q(w - s^*)\} \\ &= Q\mathbb{E}\{Qw - Qs^*\} \end{aligned}$$

$$= Q\left(\mathbb{E}\{w\} - s^*\right)$$

Its covariance matrix can be determined by:

$$\begin{aligned}
\Sigma_{w_T} &= \mathbb{E}\left\{(w_T - \mathbb{E}\{w_T\})(w_T - \mathbb{E}\{w_T\})^T\right\} \\
&= \mathbb{E}\left\{[Q(w - s^*) - Q(\mathbb{E}\{w\} - s^*)]\,[Q(w - s^*) - Q(\mathbb{E}\{w\} - s^*)]^T\right\} \\
&= \mathbb{E}\left\{[Q(w - \mathbb{E}\{w\})]\,[Q(w - \mathbb{E}\{w\})]^T\right\} \\
&= \mathbb{E}\left\{Q(w - \mathbb{E}\{w\})(w - \mathbb{E}\{w\})^T Q^T\right\} \\
&= Q\mathbb{E}\left\{(w - \mathbb{E}\{w\})(w - \mathbb{E}\{w\})^T\right\} Q^T \\
&= Q\Sigma_w Q^T
\end{aligned}$$

Another problem to access is the representation of the transformed polytope. W_r can be expressed in the form $Aw \leq b$ with A and b containing the halfspaces discussed in section 2.4. It follows that the rotated polytope will have the following form:

$$\begin{aligned}
Aw \leq b &\Leftrightarrow A\left(Q^T Q(w - s^*) + s^*\right) \leq b \\
&\Leftrightarrow AQ^T w_T \leq b - As^* \\
&\Leftrightarrow \hat{A} w_T \leq \hat{b}
\end{aligned}$$

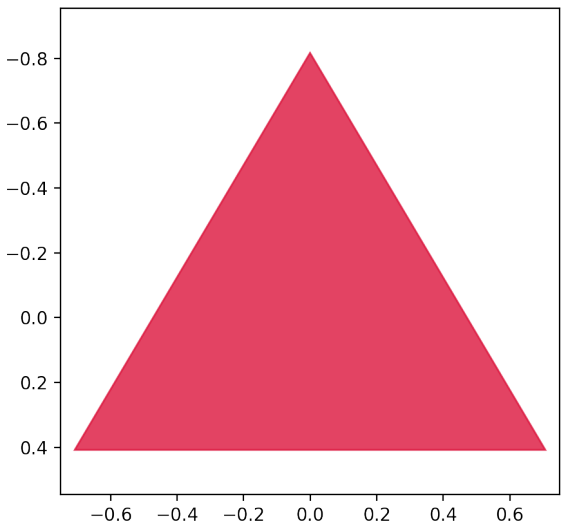

Figure 2: *The transformed standard 2-simplex from Figure 1*

4 Monte Carlo Methods

4.1 Monte Carlo Integration

Computing the posterior distribution relies on calculating the model evidence which can be rewritten as an integral as stated in (3.1). Since the computation of this integral can be very complex and since it might not even be analytically solvable, Monte Carlo integration can be used to approximate the model evidence or to estimate statistical metrics based on samples.

Monte Carlo integration uses random sampling of a function to numerically compute an estimate of its integral. Suppose that we want to integrate the one-dimensional function $f(x)$ from a to b:

$$F = \int_a^b f(x)dx$$

We can approximate this integral by averaging samples of f at uniform random points within the interval. Given a set of N uniform random variables $X_i \in [a, b]$ with a corresponding PDF of $g(x) = \frac{1}{b-a}$, the Monte Carlo estimator for computing F is

$$\langle F^N \rangle = (b-a)\frac{1}{N}\sum_{i=1}^{N} f(X_i)$$

It is easy to show that the expected value of $\langle F^N \rangle$ is in fact F:

$$\begin{aligned}
E\left[\langle F^N \rangle\right] &= E\left[(b-a)\frac{1}{N}\sum_{i=1}^{N} f(X_i)\right] \\
&= (b-a)\frac{1}{N}\sum_{i=1}^{N} E\left[f(X_i)\right] \\
&= (b-a)\frac{1}{N}\sum_{i=1}^{N} \int_a^b f(x)g(x)dx \\
&= (b-a)\frac{1}{N}\sum_{i=1}^{N} \int_a^b f(x)\frac{1}{b-a}dx \\
&= \frac{1}{N}\sum_{i=1}^{N} \int_a^b f(x)dx \\
&= F
\end{aligned}$$

Furthermore, as we increase the number of samples N, the estimator $\langle F^N \rangle$ becomes a closer and closer approximation of F. Due to the Strong Law of Large Numbers, in the limit

we can guarantee that we almost sure have the exact solution:

$$\mathbb{P}\left\{\lim_{N\to\infty}\langle F^N\rangle = F\right\} = 1$$

There are standard integration techniques that converge much more quickly in one dimension. However, these methods suffer from the curse of dimensionality, in which the convergence rate decreases exponentially as the number of dimensions increases. In contrast to deterministic methods, the convergence rate for Monte Carlo is independent of the number of dimensions in the integral, making it simple to extend the basic Monte Carlo estimator to multiple dimensions.

Monte Carlo integration can be extended to compute multidimensional integrals using random variables drawn from any PDF $g(x)$ with $x = (x_1, \ldots, x_n)$. Integrals such as

$$F = \int_\Omega f(x)dx$$

can be approximated using

$$\langle F^N\rangle = \frac{1}{N}\sum_{i=1}^{N}\frac{f(X_i)}{g(X_i)} \tag{4.1}$$

In a similar way, demonstrating that this generalized estimator also leads to the desired expected value is simple:

$$\begin{aligned}
E\left[\langle F^N\rangle\right] &= E\left[\frac{1}{N}\sum_{i=1}^{N}\frac{f(X_i)}{g(X_i)}\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}E\left[\frac{f(X_i)}{g(X_i)}\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\int\frac{f(x)}{g(x)}g(x)dx \\
&= \frac{1}{N}\sum_{i=1}^{N}\int f(x)dx \\
&= F
\end{aligned}$$

The ease with which it can be extended to multiple dimensions is a secondary advantage of Monte Carlo integration over conventional numerical integration methods in addition to the rate of convergence. Monte Carlo methods let you choose any arbitrary number of samples you want.

The convergence rate of the estimators variance can be used to examine the convergence of the estimator itself. It can be shown that the standard deviation is proportional to $\frac{1}{\sqrt{N}}$:

$$\begin{aligned}
\sigma^2\left[\langle F^N\rangle\right] &= \sigma^2\left[\frac{1}{N}\sum_{i=1}^{N}\frac{f(X_i)}{g(X_i)}\right] \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\sigma^2\left[\frac{f(X_i)}{g(X_i)}\right] \quad \text{, let } Y_i = \frac{f(X_i)}{g(X_i)} \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\sigma^2[Y_i] \\
&= \frac{1}{N}\sigma^2[Y]
\end{aligned}$$

We obtain:

$$\sigma\left[\langle F^N\rangle\right] = \frac{1}{\sqrt{N}}\sigma[Y] \propto \frac{1}{\sqrt{N}}$$

Unfortunately, this means that in order to reduce the expected error by half, four times as many samples are required. On the other hand this results in a constant convergence rate of $O(\sqrt{n})$, regardless of the dimension.

As stated earlier, Monte Carlo Integration can be applied to estimate the model evidence $p(r)$. Let $(u_1,\ldots,u_N)$ be uniformly distributed on the polytope W_r with corresponding PDF $g(x) = \frac{1}{Vol_n(W_r)}$. Using (4.1), it follows that:

$$\begin{aligned}
p(r) &= \int_{W_r} p(r|w)p(w)dw \\
&\approx \langle p(r)^N\rangle \\
&= \frac{1}{N}\sum_{i=1}^{N}\frac{p(r|u_i)p(u_i)}{g(u_i)} \\
&= \frac{Vol_n(W_r)}{N}\sum_{i=1}^{N}p(r|u_i)p(u_i)
\end{aligned}$$

Consequently, the approach's primary challenges are to obtain good estimates of both, the average and the volume.

To estimate the volume, hit-or-miss Monte Carlo integration [10] can be used. It is known that W_r is a convex polytope completely contained in the standard simplex S. With $(u_1,\ldots,u_N)$ being uniformly distributed samples from S, the probability of a sample u_i to be contained in

W_r is:

$$P(u_i \in W_r) = \frac{Vol_n(W_r)}{Vol_n(S)} \tag{4.2}$$

This characteristic can be used to approximate $Vol_n(W_r)$.

$$\langle Vol_n(W_r)^N \rangle = \frac{Vol_n(S)}{N} \sum_{i=1}^{N} I(u_i \in W_r) \tag{4.3}$$

with

$$I(u \in W_r) = \begin{cases} 1 & \text{if } u \in W_r \\ 0 & \text{else} \end{cases} \tag{4.4}$$

It really only involves counting the number of samples that are contained in W_r. To proof that the expected value of (4.3) is indeed $Vol_n(W_r)$ it needs to be noted that the expected value of an indicator function like (4.4) is the probability of the event itself.

$$\begin{aligned} E\left[I(u \in W_r)\right] &= 1 \cdot P(I(u \in W_r) = 1) + 0 \cdot P(I(u \in W_r) = 0) \\ &= P(I(u \in W_r) = 1) \\ &= P(u \in W_r) \end{aligned} \tag{4.5}$$

Using this, it can be shown that:

$$\begin{aligned} E\left[\langle Vol_n(W_r)^N \rangle\right] &= E\left[\frac{Vol_n(S)}{N} \sum_{i=1}^{N} I(u_i \in W_r)\right] \\ &= \frac{Vol_n(S)}{N} \sum_{i=1}^{N} E\left[I(u_i \in W_r)\right] \quad | \text{insert (4.5)} \\ &= \frac{Vol_n(S)}{N} \sum_{i=1}^{N} P\left(u_i \in W_r\right) \quad | \text{insert (4.2)} \\ &= \frac{Vol_n(S)}{N} \sum_{i=1}^{N} \frac{Vol_n(W_r)}{Vol_n(S)} \\ &= \frac{1}{N} \sum_{i=1}^{N} Vol_n(W_r) \\ &= Vol_n(W_r) \end{aligned}$$

In (2.8) it was shown that volume of the standard n-simplex is $\frac{\sqrt{n+1}}{n!}$.

To assess the quality of this estimator, its variance can be investigated. It was shown before that

$$\sigma^2\Big[\langle Vol_n(W_r)^N\rangle\Big] \propto \sigma^2\left[I(u \in W_r)Vol_n(S)\right]$$

Furthermore, it follows that

$$\begin{aligned}
\sigma^2\left[I(u \in W_r)Vol_n(S)\right] &= Vol_n(S)^2\sigma^2\left[I(u \in W_r)\right] \quad \mid \sigma^2[I(X)] = P(X)(1-P(X)) \\
&= Vol_n(S)^2(P(u \in W_r) - P(u \in w_r)^2) \\
&= Vol_n(S)^2(\frac{Vol_n(W_r)}{Vol_n(S)} - \frac{Vol_n(W_r)^2}{Vol_n(S)^2})) \\
&= Vol_n(S)Vol_n(W_r) - Vol_n(W_r)^2
\end{aligned} \tag{4.6}$$

Since $Vol_n(S_n)$ is fixed, the variance can be viewed as a function of $Vol_n(W_r)$ and the maximum variance can be calculated using its derivative.

$$\begin{aligned}
0 &= \frac{d}{dVol_n(W_r)}\Big(Vol_n(S)Vol_n(W_r) - Vol_n(W_r)^2\Big) \\
&= Vol_n(S) - 2Vol_n(W_r) \\
&\Rightarrow Vol_n(W_r) = \frac{Vol_n(S)}{2}
\end{aligned}$$

Inserting into (4.6) leads to:

$$Vol_n(S)\frac{Vol_n(S)}{2} - \frac{Vol_n(S)^2}{2^2} = \frac{Vol_n(S)^2}{4}$$

Therefore the estimators variance reaches its maximum of $\frac{Vol_n(S)^2}{4}$ at $Vol_n(W_r) = \frac{Vol_n(S)}{2}$. To investigate the limit behaviors, it can be shown that

$$\lim_{Vol_n(W_r)\to 0} Vol_n(S)Vol_n(W_r) - Vol_n(W_r)^2 = 0$$

and

$$\lim_{Vol_n(W_r)\to Vol_n(S)} Vol_n(S)Vol_n(W_r) - Vol_n(W_r)^2 = 0$$

Consequently, the estimator handles, both, small and large polytopes in proportion to the standard simplex really well.

4.2 Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) methodes are able to indirectly obtain inference on the posterior distribution using computer simulations [11] by sampling from the posterior rather than computing it. It combines two ideas: using the Markov chain to extract a set of parameter values from the posterior distribution and Monte Carlo integration to extract a distributional estimate of the posterior and related statistics like the mean.

The posterior distribution cannot be directly and independently sampled for parameter values. The Markov chain is used as a result of this. By building a Markov chain with a specific first-order transition kernel, it is intended to obtain a set of sampled parameter values from the posterior distribution of interest, such that the stationary distribution of the Markov chain is identical to the posterior distribution of interest. The subsequent realizations of the Markov chain can be viewed as autocorrelated samples from the posterior distribution, since they are dependent on their previous values in the chain. They are used to obtain the corresponding Monte Carlo estimates if the chain has ran long enough to reach its stationary distribution. A feasible choice of transition kernel could be given by the Metropolis-Hastings algorithm [12].

To address convergence, the $\hat{R}$ statistic [13], which is defined as the ratio of within-chain to between-chain variability, is the most popular method for determining whether a Markov chain has reached the stationary distribution. This method requires multiple independent runs of the MCMC algorithm to be used. To increase initial variability across the Markov chains and increase the likelihood that non-convergence of the chain to the stationary distribution will be detected, it is ideal for each of the Markov chains to start from a different starting value and with a different random seed.

5 Numerical example

Table 1: *Performance matrix $\tilde{P}$ with corresponding criteria from [14]*

	Price-performance ratio	Price	Comfort	Family friendliness	Environmental friendliness	Resale value	Image/prestige	Safety	Other
Mini cars	3.5	3.9	3.4	1.5	3.6	2.9	2.0	3.8	5.0
Small cars	4.2	3.6	3.0	2.7	3.4	3.7	2.0	4.0	5.0
Medium cars	4.1	3.1	3.9	3.2	3.1	3.7	3.0	4.1	5.0
Large cars	3.8	2.2	4.3	4.2	2.8	3.8	3.0	4.6	5.0
Executive cars	3.1	1.0	4.0	3.3	2.2	3.7	4.0	4.5	5.0
Luxury cars	3.2	1.0	4.7	3.6	1.1	4.1	5.0	3.9	5.0
SUVs	4.0	3.1	3.5	3.4	2.1	4.2	4.5	4.0	5.0
Off-road vehicles	3.9	2.5	3.8	3.6	1.7	3.8	4.5	4.3	5.0
Mini vans	2.9	1.8	3.2	4.7	1.6	3.1	2.0	4.7	5.0

To test the described methods and illustrate the workflow, this section focuses on a specific numerical example from the German car market. Table 1 shows a matrix $\tilde{P}$ containing some performances for different types of cars (alternatives) under certain aspects (criteria). We use this matrix to construct our polytope W_r and therefore also our likelihood function $p(r|w)$ which will, as stated in section 3.4, result in a scaled uniform distribution on W_r.
For the MCDA method we chose PROMETHEE II with all criteria associated to benefit, since $\tilde{P}$ is on a scale, such that higher values correspond to a better performance. All criteria will be normalized using a V-shape preference function, as in 2.3, with $d = \max_{1\leq i\leq m} \tilde{P}_{ij} - \min_{1\leq i\leq m} \tilde{P}_{ij}$. This results in the biggest preference of one alternative over another, if their performances have the biggest difference out of all alternatives. To construct W_r out of P, we choose the following ranking from [14]:

1. Small cars
2. Medium cars
3. Mini cars
4. Large cars

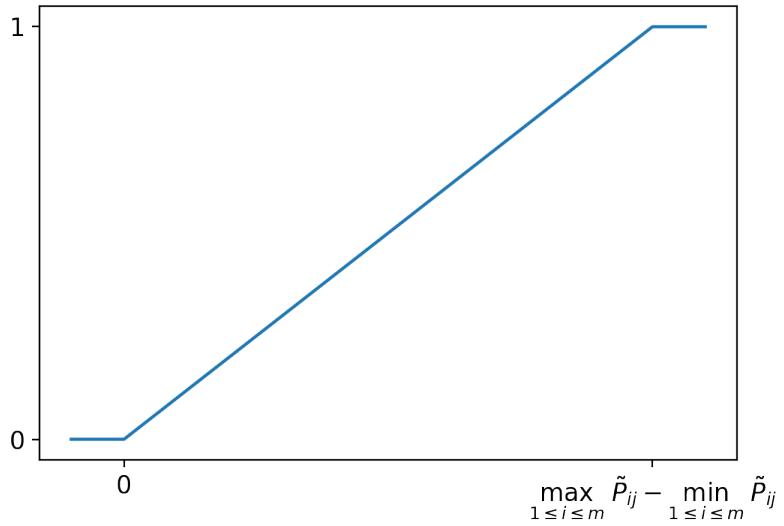

Figure 3: *The used V-shape preference function*

To specify the prior distribution, we use the data from table 2. Therefore it is a databased prior, but double dipping is avoided since the data is only used for the prior and not the likelihood. The data comes from a survey and for simplicity we assume these weights to be normal with mean μ and covariance matrix Σ. This should usually be tested.
Since tables 1 and 2 do not match completely, all mismatching criteria are summed up to ”Other”. These weights are also not yet normalized to the standard simplex, which needs to be done in advance.
After preparing the data, we set $\mu = \bar{w}$ with

$$\bar{w} = \frac{1}{n} \sum_{i=1}^{n} w_i$$

as the sample mean and $\Sigma = \Sigma_w$ with

$$\Sigma_w = \frac{1}{n-1} \sum_{i=1}^{n} (w_i - \bar{w})(w_i - \bar{w})^T$$

as the sample covariance matrix. Since, due to the structure of the standard simplex, Σ will be singular, a transformation to a lower dimension needs to be proceeded. This was discussed in section 3.5. It follows, that the prior will be a normal distribution with $\mu = \bar{w}_T$ and $\Sigma = \Sigma_{w_T}$.

$$p(w) = \frac{1}{\sqrt{(2\pi)^9 |\Sigma|}} \exp\left(-\frac{1}{2} (w_T - \mu_{w_T})^T \Sigma^{-1} (w_T - \mu_{w_T}) \right)$$

Table 2: *Poll data for weights*

criterion	$\tilde{w}_1$	$\tilde{w}_2$	...
Price-performance ratio	10.04547	8.66926	...
Price	29.49723	6.51079	...
Comfort	13.63339	10.34713	...
Family friendliness	9.99244	10.47944	...
Environmental friendliness	4.72200	5.87986	...
Resale value	1.41605	2.96719	...
Image/prestige	3.53576	5.52158	...
Safety	1.10667	7.30710	...
Drive technology	12.23415	18.51638	...
Range	2.35857	8.43138	...
Charging/refuel duration PHEV	3.43427	1.97538	...
Charging duration BEV	5.55045	8.80491	...
Fixed operating costs	2.47352	4.58960	...

After specifying the prior distribution and determining the likelihood function, we use MCMC to collect samples from the posterior distribution. The results are visualized in Figure 4. The posterior and (scaled) likelihood are generated using a total of four Markov Chains with a sample size of one million each and the $\hat{R}$ statistic was used to assure convergence. The marginal distributions are then approximated by fitting a spline through the corresponding histograms. On the other hand, the prior is calculated analytically, since the marginal distributions of a normal distribution are also normal.

Table 3: *Expected values and variances using Monte-carlo Integration (rounded)*

	Price-performance ratio	Price	Comfort	Family friendliness	Environmental friendliness	Resale value	Image/ prestige	Safety	Other
$\mathbb{E}$	0.05	0.25	0.06	0.05	0.08	0.02	0.01	0.06	0.41
$\sigma^2 \cdot 10^2$	0.038	0.247	0.036	0.028	0.036	0.016	0.007	0.037	0.889

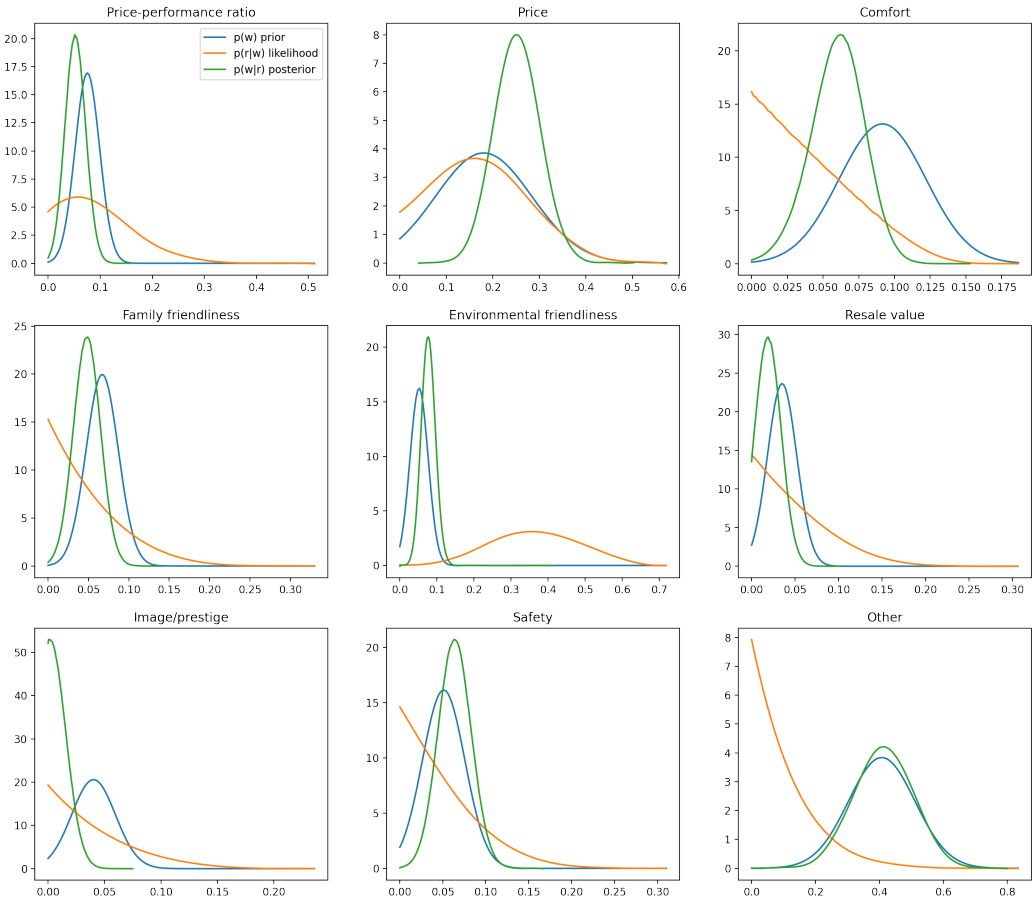

Figure 4: *Estimated marginal distributions of prior, likelihood and posterior*

The samples were generated by hopsy, a Python wrapper for the sampling toolbox HOPS [15] which has adapted MCMC algorithms to convex, bounded polytopes. It should be noted, that the likelihood function is indeed uniform on W_r even if it might not look like it due to the structure of the polytope. Moreover, it can be seen that the weighting for price rose in contrast to our prior beliefs and the weighting for prestige sank. Which makes sense considering our given ranking ranked small cars at number one. Another observation is that prior and likelihood contradict each other a lot for the weighting of environmental friendliness. This results in the likelihood almost being cancelled out completely, since the prior is nearly zero at the mode of the likelihood.

6 Conclusion

As stated in the beginning, the goal of this seminal thesis was to combine aspects from MCDA with prior information about a decision to gain insight on it in terms of a posterior distribution. Two MCDA methods have been introduced and analyzed with regard to some geometric aspects. It was shown that the presented MCDA methods rely on their weights to lay in a polytope and that this requirement can be satisfied and combined with prior information using Bayes' theorem. The resulting distribution of criteria weights to a fixed ranking of alternatives was found to be difficult to compute. Hence, Monte Carlo integration and MCMC were discussed and analyzed in terms of convergence. The presented methods for computing this distribution work and were able to be successfully applied on a numerical example on the German car market by producing samples of its posterior distribution and analyzing them.

References

[1] PODVEZKO, Valentinas: The comparative analysis of MCDA methods SAW and COPRAS. In: *Engineering Economics* (2011)

[2] GINEVICIUS, Romualdas ; PODVEZKO, Valentinas: Some problems of evaluating multicriteria decision methods. In: *International Journal of Management and Decision Making - Int J Manag Decis Making* (2007)

[3] MARESCHAL, B.: *The PROMETHEE Bibliographical Database*. `http://en.promethee-gaia.net/bibliographical-database.html`, 2020

[4] BRANS, J.P. ; VINCKE, Ph. ; MARESCHAL, B.: How to select and how to rank projects: The PROMETHEE method. In: *European Journal of Operational Research* 24 (1985)

[5] ZIEGLER, Günter M.: *Lectures on Polytopes*. pringer-Verlag New York, Inc., 1995

[6] LEGGIERA lonza: *Volume of n-simplex in $\mathbb{R}^n$ without Cayley-Menger determinant*. `https://math.stackexchange.com/questions/3988039/volume-of-n-simplex-in-bbb-rm-without-cayley-menger-determinant`, 2021

[7] SCHOOT, Rens van d. ; DEPAOLI, Sarah ; KING, Ruth ; KRAMER, Bianca ; MÄRTENS, Kaspar ; TADESSE, Mahlet G. ; VANNUCCI, Marina ; GELMAN, Andrew ; VEEN, Duco ; WILLEMSEN, Joukje ; YAU, Christopher: Bayesian statistics and modelling. In: *Nature Reviews — Methods Primers* (2021)

[8] DARNIEDER, William F.: *Bayesian Methods for Data-Dependent Priors*, Ohio State University, Diss., 2011

[9] DEPAOLI, Sarah: Mixture Class Recovery in GMM Under Varying Degrees of Class Separation: Frequentist Versus Bayesian Estimation. In: *Psychological Methods* (2013)

[10] CLAUSTHAL, TU: *Monte-Carlo-Integration*. `https://www.mathematik.tu-clausthal.de/interaktiv/integration/montecarlo`, 2022

[11] GEYER, Charles J.: Markov Chain Monte Carlo Maximum Likelihood. In: *Interface Foundation of North America* (1991)

[12] HASTINGS, W. K.: Monte Carlo Sampling Methods Using Markov Chains and Their Applications. In: *Biometrika* 57 (1970)

[13] GELMAN, Andrew ; RUBIN, Donald B.: Inference from iterative simulation using multiple sequences. In: *Statistical Science* 7 (1992)

[14] VÖGELE, Stefan ; BROSKA, Lisa H. ; OTTE, Sebstian ; RÜBBELKE, Dirk: Why the trend towards gas-guzzlers? a closer look at the complex effects of social norms on german car buyers. In: *Energy Research & Social Science* (2021)

[15] JADEBECK, Johann F. ; THEORELL, Axel ; LEWEKE, Samuel ; NÖH, Katharina: HOPS: high-performance library for (non-)uniform sampling of convex-constrained models. In: *Bioinformatics* 37 (2020), 10, No. 12. https://doi.org/10.1093/bioinformatics/btaa872

List of abbreviations

MCDA Multi Criteria Decision Analysis
SAW Simple Additive Weighting
PROMETHEE II Preference Ranking Organisation Method for Enrichment Evaluation II
MCMC Markov Chain Monte Carlo
PDF probability density function
PMF probability mass function