MONTE CARLO INVERSE

RAOUL LEPAGE, KRZYSZTOF PODGÓRSKI, AND MICHAŁ RYZNAR

Abstract. We consider the problem of determining and calculating a positive measurable function f satisfying $\int g_i f d\mu = c_i$, $i \in \mathcal{I}$ provided such a solution exists for a given measure μ, nonnegative measurable functions $\{g_i,\ i \in \mathcal{I}\}$ and constants $\{c_i > 0,\ i \in \mathcal{I}\}$. Our goal is to provide a general positive Monte Carlo solution, if one exists, for arbitrary systems of such simultaneous linear equations in any number of unknowns and variables, with diagnostics to detect redundancy or incompatibility among these equations. MC-inverse is a non-linear Monte Carlo solution of this inverse problem. We establish criteria for the existence of solutions and for the convergence of our Monte Carlo to such a solution. The method we propose has connections with mathematical finance and machine learning.

1. Introduction

Consider a measure space $(\Omega, \mathcal{F}, \mu)$ and a set of measurable non-negative constraint functions g_i on it. Suppose there is a positive solution $f = f^*$ of the equations $\int g_i f d\mu = c_i$, $i \in \mathcal{I}$ for given positive c_i. Equivalently, for every probability measure P mutually continuous with respect to μ and $X_i = g_i/c_i$ there is a positive solution $X^* = \frac{dP}{d\mu}/f^*$ of the equations

$$E(X_i/X^*) = \int X_i/X^* dP = 1, i \in \mathcal{I}.$$

The latter solution was previously interpreted in terms of the optimal investment portfolio obtained from an extension of the Kelly Principle off the simplex (see [7] and [6], [8], [9], [10]).

Date: December 17, 2004.

Key words and phrases. inverse problem, entropy, Kullback-Leibler divergence, exponentiated gradient algorithm, Cover's algorithm.

Research of MR was partially done while visiting the Department of Probability and Statistics, Michigan State University.

We shall choose a probability measure P, call it a working measure, to play the role of an initial guess of the solution and to control a Monte Carlo by which our solution will be derived. This Monte Carlo involves sequentially updating coefficients, one for each of the constraints, and is based on our sequential extension of the Exponentiated Gradient (EG) algorithm [5].

2. Inverse problem

2.1. **Projection density reciprocal (PDR).**

Let $\boldsymbol{X} = (X_i,\ i \in \mathcal{I})$ be a finite or infinite collection of P-equivalence classes of nonnegative random variables on a probability space $(\Omega, \mathcal{F}, P)$.

Let $\Pi = \{\pi \in \mathbb{R}^{\mathcal{I}} : \pi \text{ is finitely supported}, \pi \cdot \mathbf{1} = 1\}$ and $\Pi^+ = \Pi \cap [0, \infty)^{\mathcal{I}}$. Here and in what follows, for $\mathbf{x} \in \mathbb{R}^{\mathcal{I}}$, $\pi \cdot \mathbf{x} = \sum \pi_i x_i$ and $\mathbf{1}$ stands for a function (vector) identically equal to one. Define a set

$$\mathcal{L}^+_{\mathbf{X}} = \overline{\{\pi \cdot \boldsymbol{X} : \pi \in \Pi^+\}},$$

where the closure is with respect to convergence in probability P. This set will be often referred to as the simplex spanned on $\mathbf{X}$. Likewise define the extended simplex

$$\mathcal{L}_{\mathbf{X}} = \overline{\{\pi \cdot \boldsymbol{X} \geq 0 :\ \pi \in \Pi\}}. \tag{1}$$

We consider $X^* = \operatorname{argmax}\{E \ln X : X \in \mathcal{L}_{\mathbf{X}}\}$, where $\ln r = -\infty$ for $r = 0$. In fact, it is slightly more general and convenient if we instead define X^*, if it exists, to be any element of $\mathcal{L}_{\mathbf{X}}$ for which $E \ln(X/X^*) \leq 0$ for every $X \in \mathcal{L}_{\mathbf{X}}$. It will be seen that X^* is unique for given P (see also [1] for the maximization over $\mathcal{L}^+_{\mathbf{X}}$).

Although it may not be immediately evident, X^* satisfies $E(X_i/X^*) = 1,\ i \in \mathcal{I}$, provided the normalized constraints X_i are mathematically consistent and specified conditions are satisfied (see Section 2.2). We call X^* the *projection density* (PD) and $f^* = \frac{dP/d\mu}{X^*}$ the *projection density reciprocal* (PDR) solution to the inverse problem. Both may depend upon the choice of working measure P (see Section 2.4).

So, for a given *working probability measure* P the task is to find X^* for that P. For maximizing $E \ln Y$, $Y \in \mathcal{L}_{\mathbf{X}}^+$ we will produce a Monte Carlo solution X_n^* converging to X^*. If X^* does not belong to the simplex $\mathcal{L}_{\mathbf{X}}^+$ but rather to $\mathcal{L}_{\mathbf{X}}$ one may find an X^* by deleting some constraints and incorporating some additional $Y \in \mathcal{L}_{\mathbf{X}} \setminus \mathcal{L}_{\mathbf{X}}^+$. The solution Y^* for this new set of constraints is then found and the process repeated until a solution is found. The choice of a working measure will be largely a matter of convenience although we shall see that one resembling an $f^* dP$ for some P is preferable.

It is well to keep in mind the following illustrative example: $0 < X_1 = X$, $X_2 = 2X$, with $E|\ln X| < \infty$. In such a case, $E \ln(\pi X_1 + (1-\pi)X_2) = \ln(2-\pi) - E \ln X$ has no finite maximum with respect to $\pi \in \mathbb{R}$ for each working measure P. However restricting the maximization to the simplex $0 \le \pi \le 1$ yields $X^* = 2X$. The same X^* is obtained if the first constraint X_1 is dropped which results in a reduced problem whose solution lies in its simplex. This will be important since our basic algorithm, like Cover's [2], searches only the simplex $\mathcal{L}_{\mathbf{X}}^+$.

Proposition 1. *(Super Simplex) For a given working measure P the (extended) PD X^* exists for constraints $\mathbf{X} = \{X_i : i \in \mathcal{I}\}$ if and only if the super simplex PD Y^* exists for constraints $\mathbf{Y} = \{\pi \cdot \boldsymbol{X} \ge 0 : \pi \in \boldsymbol{\Pi}\}$, in which case $X^* = Y^*$.*

Proof. It is enough to show that $\mathcal{L}_{\mathbf{X}}$ coincides with the super simplex $\mathcal{L}_{\mathbf{Y}}^+$. Take an element $\pi \cdot \mathbf{Y}$, $\pi = (\pi_1, \ldots, \pi_n) \in \Pi^+$ of the super simplex. Then for some $\pi^1, \ldots, \pi^n \in \Pi$ we have

$$\begin{aligned}
\pi \cdot \mathbf{Y} &= \pi_1(\pi^1 \cdot \mathbf{X}) + \cdots + \pi_n(\pi^n \cdot \mathbf{X}) \\
&= (\pi_1 \pi^1 + \cdots + \pi_n \pi^n) \cdot \mathbf{X} \\
&= \tilde{\pi} \cdot \mathbf{X},
\end{aligned}$$

where $\tilde{\pi} = \pi_1 \pi^1 + \cdots + \pi_n \pi^n \in \Pi$.

Thus the super simplex, which obviously contains $\mathcal{L}_{\mathbf{X}}$, is also contained in $\mathcal{L}_{\mathbf{X}}$.

□

Example 1. Let $\Omega = \{(1,1),(1,2),(2,1),(2,2)\}$ and (unknown) $f = \mathbf{1}$. Suppose that the constraints tell us the overall total of f is four and the first row and first column totals of f are each two. We may think of these as integrals of $g_0 = \mathbf{1}$, $g_1 = \mathbf{1}_{\{(1,1),(1,2)\}}$ and $g_2 = \mathbf{1}_{\{(1,1),(2,1)\}}$ with respect to uniform counting measure μ on Ω. The normalized constrains $\mathbf{X} = (X_0, X_1, X_2)$ become then $X_0 = 0.25 \cdot \mathbf{1}$, $X_1 = 0.5 \cdot \mathbf{1}_{\{(1,1),(1,2)\}}$ and $X_2 = 0.5 \cdot \mathbf{1}_{\{(1,1),(2,1)\}}$. Consider

$$\mathcal{L}_{\mathbf{X}} = \{Y : Y = \pi_0 X_0 + \pi_1 X_1 + \pi_2 X_2 \geq 0, \pi_0 + \pi_1 + \pi_2 = 1\},$$

which translates to the following constraints for π_0, π_1, π_2:

$$\begin{aligned} \pi_0 &\geq 0 \\ \pi_0 + 2\pi_1 &\geq 0 \\ \pi_0 + 2\pi_2 &\geq 0 \\ \pi_0 + 2\pi_1 + 2\pi_2 &\geq 0 \\ \pi_0 + \pi_1 + \pi_2 &= 1. \end{aligned}$$

Let $P = (p_{11}, p_{12}, p_{21}, p_{22})$ be any fully supported probability measure on Ω. Our problem can be stated now as a classical problem of maximization of a concave function:

$$E \ln(\pi \cdot \boldsymbol{X}) = \ln(\pi_0 + 2\pi_1 + 2\pi_2)\, p_{11} + \ln(\pi_0 + 2\pi_1)\, p_{12} + \ln(\pi_0 + 2\pi_2)\, p_{21} + \ln(\pi_0)\, p_{22}.$$

Using Lagrange multipliers one finds the solution:

$$f^* = 2(p_{11} + p_{22})\mathbf{1}_{\{(1,1),(2,2)\}} + 2(p_{12} + p_{21})\mathbf{1}_{\{(1,2),(2,1)\}},$$

which depends upon P. By contrast, the cross-entropy method utilizes a different optimization. Subject to the conditions

$$\begin{aligned} \min\{h_{11}, h_{12}, h_{21}, h_{22}\} &\geq 0 \\ h_{11} + h_{12} + h_{21} + h_{22} &= 1 \\ h_{11} + h_{12} &= 1/2 \\ h_{11} + h_{21} &= 1/2 \end{aligned}$$

minimize the cross-entropy function (also known as Kullback-Leibler divergence)

$$D(h|p) = h_{11} \ln(h_{11}/p_{11}) + h_{12} \ln(h_{12}/p_{12}) + h_{21} \ln(h_{21}/p_{21}) + h_{22} \ln(h_{22}/p_{22}).$$

Under our assumptions the cross-entropy solution has the following form,

$$g^* = 4 \cdot h^* = \frac{2}{1 + \sqrt{p_{12}p_{21}/(p_{11}p_{22})}} \mathbf{1}_{\{(1,1),(2,2)\}} + \frac{2}{1 + \sqrt{p_{11}p_{22}/(p_{12}p_{21})}} \mathbf{1}_{\{(1,2),(2,1)\}},$$

which generally differs from f^*.

Calculating f^* was daunting, even for this simple example. We now illustrate how it could be done using our Monte Carlo of Section 3. For example, set $P = (1/2, 1/4, 1/8, 1/8)$. Then $f^* = (1.25, 0.75, 0.75, 1.25)$. The Monte Carlo of Proposition 3 is implemented

$$\eta_j = \frac{1}{j^{1/2}(\log j)^{1/2+\epsilon}}, \ \epsilon = 0.05,$$

and then

$$\pi^0 = (1/3, 1/3, 1/3),$$

$$\pi_i^{j+1} = \pi_i^j \frac{\exp\left(\eta_j \frac{X_i^j}{\pi^j \cdot \boldsymbol{X}^j}\right)}{\sum_{k=1}^3 \pi_k^j \exp\left(\eta_j \frac{X_k^j}{\pi^j \cdot \boldsymbol{X}^j}\right)}, \qquad i = 1, 2, 3,$$

with $\boldsymbol{X}^j = (X_1(\omega_j), X_2(\omega_j), X_3(\omega_j))$, where $\{\omega_j\}, \ j \leq n = 100000$ are i.i.d. random draws from P. The approximations of X^* are $X_j^* = \pi^j \cdot \mathbf{X}$. The approximate result is

$$f^* \approx f_{100000}^* = \frac{P}{\pi^{100000} \cdot \boldsymbol{X}} = (1.2499, 0.7422, 0.7661, 1.2504).$$

Since X^* appears to lie in the interior of $\mathcal{L}^+_{\mathbf{X}}$ we accept $X^*_{100000} = \pi^{100000} \cdot \boldsymbol{X}$ as the Monte Carlo approximation of X^* and f^*_{100000} as also being a Monte Carlo approximation of f^*. Indeed it is.

Close examination of Corollary 3 reveals that we have established the convergence of our Monte Carlo only for a weighted average $\overline{\pi}^j$ of $\{\pi^j\}$, but in practice there seems to be little difference between $\overline{\pi}^j$ and π^j. For comparison,

$$\pi^{100000} = (0.3999, 0.4737, 0.1264),$$

while

$$\overline{\pi}^{100000} = (0.3982, 0.4732, 0.1286),$$

and the resulting solution

$$f^* \approx \frac{P}{\overline{\pi}^{100000} \cdot \boldsymbol{X}} = (1.2486, 0.7437, 0.7630, 1.2556).$$

The next example illustrates what can happen if X^* does not belong to the interior of the simplex $\mathcal{L}^+_{\mathbf{X}}$. Our remedy will be to incorporate additional constraints outside the simplex. This redefines the set of constraints and we proceed to search the simplex defined by these new constraints.

Example 2. The setting of Example 1 is assumed. Purely for illustration, we modify the constraints:

$$Y_0 = X_0,\ Y_1 = 0.5 \cdot (X_0 + X_1),\ Y_2 = 0.5 \cdot (X_0 + X_2).$$

Choose, as before, $P = (1/2, 1/4, 1/8, 1/8)$. Let $Y^*_+ = \pi^*_+ \cdot \mathbf{Y}$ be the simplex solution, i.e. the one that minimizes $E \ln Y$, for $Y \in \mathcal{L}^+_{\mathbf{Y}}$. Direct calculations yield $\pi^*_+ = (0, 5/6, 1/6)$ and $Y^*_+ = \pi^*_+ \cdot \boldsymbol{Y} = (3/8, 1/3, 1/6, 1/8)$ giving $f^*_+ = (1.3333, 0.75, 0.75, 1)$. Because the first coordinate of π^*_+ is 0 we suspect that the solution Y^*_+ will fail to satisfy the first constraint. Indeed,

$$E(Y_0/Y^*_+) = (1/2)\frac{1/4}{3/8} + (1/4)\frac{1/4}{1/3} + (1/8)\frac{1/4}{1/6} + (1/8)\frac{1/4}{1/8} = 23/24 < 1.$$

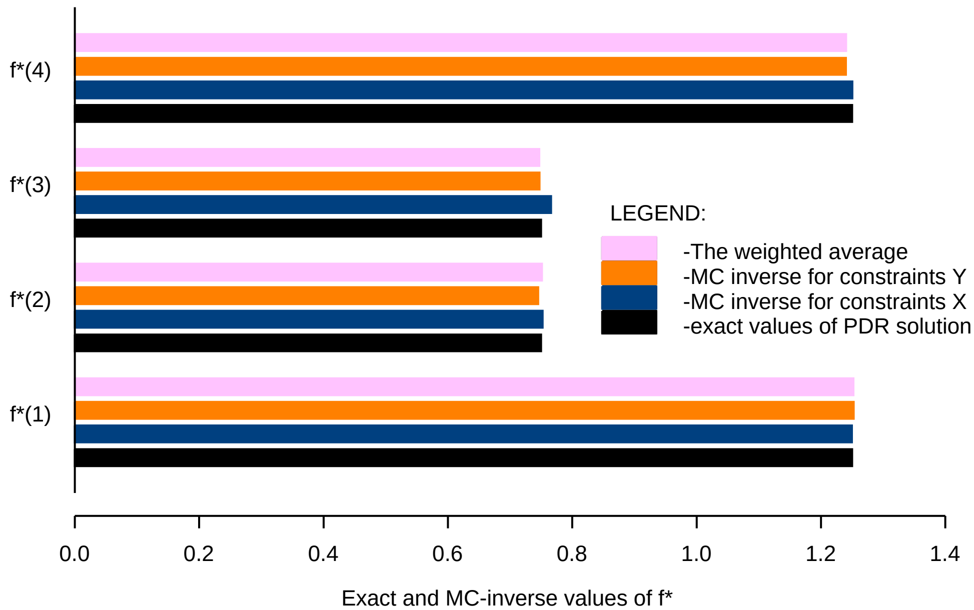

FIGURE 1. Comparison of the exact PDR solution with its MC-inverse approximations

The second and the third constraints are satisfied, i.e. $E(Y_1/Y_+^*) = 1$, $E(Y_2/Y_+^*) = 1$. This solution, obtained from a search of the simplex, obviously differs from the extended solution $f^* = (1.25, 0.75, 0.75, 1.25)$ (found above) which satisfies all three constraints.

How does our Monte Carlo handle this? With $\eta_j = \frac{1}{j^{1/2}(\log j)^{1/2+\epsilon}}$, $\epsilon = 0.05$, we obtain

$$\pi_+^* \approx \pi^{100000} = (0.0000, 0.8405, 0.1595)$$

with corresponding

$$f_+^* \approx \frac{P}{\pi^{100000} \cdot \boldsymbol{Y}} = (1.3333, 0.7460, 0.7582, 1.0000).$$

Similarly,

$$\overline{\pi}^{100000} = (0.0000, 0.8421, 0.1579)$$

with corresponding

$$f_+^* \approx \frac{P}{\overline{\pi}^{100000} \cdot \boldsymbol{Y}} = (1.3333, 0.7451, 0.7600, 1.0000).$$

Having returned an apparent solution $\pi^*_{+0} \approx 0$, our Monte Carlo is suggesting that possibly $\pi^*_0 < 0$. To check this we extend the Monte Carlo off $\mathcal{L}^+_{\mathbf{Y}}$. The idea is to replace Y_0 with

$$Y_{00} = \alpha Y_0 + (1-\alpha)\frac{Y_1 + Y_2}{2},$$

for the smallest possible $\alpha < 0$ consistent with $Y_{00} \geq 0$. Adopting $\alpha = -1$, the most extreme case, $Y_{00} = (1/2, 1/4, 1/4, 0)$. With the new constraints $\boldsymbol{Y} = (Y_{00},\ Y_1,\ Y_2)$ our Monte Carlo yields

$$\pi^*_+ \approx \pi^{100000} = (0.1936, 0.7340, 0.0724),$$

$$f^*_+ \approx \frac{P}{\pi^{100000} \cdot \boldsymbol{Y}} = (1.2525, 0.7451, 0.7472, 1.2401),$$

which is very close to f^*. Corresponding approximations for the weighted average are,

$$\pi^*_+ \approx \overline{\pi}^{100000} = (0.1940, 0.7338, 0.0723),$$

$$f^*_+ \approx \frac{P}{\overline{\pi}^{100000} \cdot \boldsymbol{Y}} = (1.2524, 0.7514, 0.7471, 1.2406).$$

Figure 2 illustrates geometrically a situation in which the (extended) PD solution exists on the extended simplex $\mathcal{L}_{\mathbf{X}}$ but not within the initial simplex. The initial set of constraints $\mathbf{X} = (X_1, X_2, X_3)$ yields the solution Y^* that is on the boundary of $\mathcal{L}^+_{\mathbf{X}}$. Using the idea of the supersimplex (see Proposition 1) by adding succesively new constraints X_4, X_5, and X_6 we obtain a sequence of simplex solutions Y_1^*, Y_2^*, and finally Y_3^* with the last one in the interior of the simplex generated by the constraints $\widetilde{\mathbf{X}} = (X_1, X_2, X_3, X_4, X_5)$. By Corollary 2 of Subsection 2.3, the solution in the interior of an simplex is necessarily the global solution X^*. More details on how to effectively update the constraints are discussed in Subsection 3.2. Section 4 presents a problem of reconstructing a probability density function that is in a complete analogy in the hypothetical situation of Figure 2 (see that section for more details).

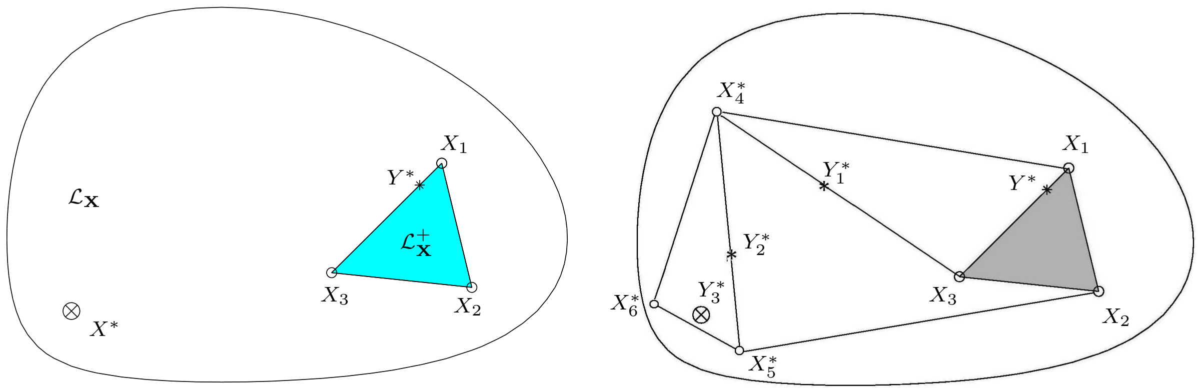

FIGURE 2. Hypothetical situation when the extended simplex solution X^* is outside the initial simplex $\mathcal{L}^+_X$. The initial MC-inverse solution Y^* is necessarily on the boundary of the initial simplex *(left)*. Adding appropriate constraints eventually leads to a simplex for which the solution Y_3^* coincides with X^* *(right)*.

2.2. **Basic properties.** Our first result establish conditions under which a PD solution satisfying the normalized constraints will exist. The result in a somewhat different situation was proven in [7] and [10]. Here we consider the extended simplex $\mathcal{L}_{\mathbf{X}}$ as defined by (1). We also say that $0 \leq X$ is relatively bounded with respect to X^* if X/X^* is bounded (here $0/0 = 0$) and denote $\mathcal{B}_{X^*}$ a convex subset of $\mathcal{L}_{\mathbf{X}}$ consisting of elements relatively bounded to X^* .

Theorem 1. *Let $\mathcal{L}$ be a convex set of non-negative random variables and let $\mathcal{L}_{\mathbf{X}}$ be the extended simplex defined for the constraints $\mathbf{X}$.*
(i): Assume that there exists positive $X^* \in \mathcal{L}$ such that

$$E \ln(X/X^*) \leq 0$$

for each $X \in \mathcal{L}$. Then X^* is unique and for each $X \in \mathcal{L}$:

$$E(X/X^*) \leq 1. \tag{2}$$

(ii): If there exists positive $X^* \in \mathcal{L}_{\mathbf{X}}$ such that

$$E \ln(X/X^*) \leq 0 \tag{3}$$

for each $X \in \mathcal{B}_{X^*}$*, then for all such* X*:*

$$E(X/X^*) = 1. \tag{4}$$

Proof. *(i)*: Consider an extension of the class $\mathcal{L}$ to $\mathcal{L}' = \{Y : 0 \leq Y \leq X \text{ for some X} \in \mathcal{L}\}$. Then the assumption is satisfied for $\mathcal{L}'$, so we can assume that $\mathcal{L}' = \mathcal{L}$. For $X \in \mathcal{L}$ and $\alpha \in (0,1)$ define $X_\alpha = X^* + \alpha(X - X^*) \in \mathcal{L}$. Then

$$E \ln(1 + \alpha(X/X^* - 1))) = E \ln(X_\alpha/X^*) \leq 0, \ \alpha \in (0,1).$$

Now assume that $X \leq cX^*$ for some positive c. Apply *(i)* of Lemma 1, Subsection 5.1 of the Appendix, with $Z = X/X^* - 1$ to get

$$E(X/X^* - 1) \leq 0.$$

Next note that every $X \in \mathcal{L}$ is a limit of $X\mathbf{1}_{\{X \leq cX^*\}} \in \mathcal{L}$ when $c \to \infty$, so the conclusion holds for all $X \in \mathcal{L}$. The uniqueness follows from convexity of $\mathcal{L}$ and strict concavity of $X \to E \ln X$ (see [1]).

(ii): By taking $\mathcal{L} = \mathcal{B}_{X^*}$ in *(i)* we obtain for $X \in \mathcal{L}$:

$$E(X/X^*) \leq 1.$$

Let $X \leq cX^*$ for some $c > 1$ and $X_\alpha = X^* + \alpha(X - X^*)$. For $\alpha \in (-\frac{1}{2(c-1)}, 0)$ we have

$$0 < X^*(1 + \alpha(c-1)) \leq X_\alpha \leq X^*(1 - \alpha)$$

thus $X_\alpha \in \mathcal{B}_{X^*}$. Consequently

$$E \ln(1 + \alpha(X/X^* - 1))) = E \ln(X_\alpha/X^*) \leq 0, \quad \alpha \in (-\frac{1}{2(c-1)}, 0).$$

Next apply *(ii)* of Lemma 1, Subsection 5.1 of the Appendix, with $Z = X/X^* - 1$ to get

$$E(X/X^* - 1) \geq 0,$$

showing $E(X/X^*) \geq 1$ and concluding the proof.

□

Remark 1. If the original problem consists of conflicting information, i.e. there is no function satysfying the postulated constraints, then necessarily for some $X \in \mathcal{L}_{\mathbf{X}}$ we have $E(X/X^*) < 1$ (if X^* exists). In this sense our solution recognizes existing conflicting information. If some of the constraints are not present in the solution X^* (coefficients in the solution corresponding to those constraints are zero) it suggests that they could be removed to correct the problem. Evaluating the solution for this reduced problem and verifying if $E(X/X^*)$ is identically equal to one will be a confirmation that conflict has been removed from the constraints.

Remark 2. It follows from *(i)* of Theorem 1 that if (3) holds for each $X \in \mathcal{L}_{\mathbf{X}}$, then X^* is unique in $\mathcal{L}_{\mathbf{X}}$.

Corollary 1. *If* $\mathbf{1} \in \mathcal{L}_{\mathbf{X}}$ *and* P *satisfies* $E(X_i) = 1,\ i \in \mathcal{I}$ *then* $X^* = \operatorname{argmax}\{E \ln X :\ X \in \mathcal{L}_{\mathbf{X}}\}$ *exists and is equal to* $\mathbf{1}$*, i.e.* $f^* = dP/d\mu$ *is the PDR solution to the inverse problem.*

Proof. For each $\pi \in \Pi$ we have

$$E \ln(\pi \cdot \boldsymbol{X}) \leq \ln E(\pi \cdot \boldsymbol{X}) = 0.$$

Moreover if $0 \leq X_n = \pi_n \cdot \mathbf{X}$ converges to X and $EX_n = 1$, then by Fatou's Lemma necessarily $EX \leq 1$. Consequently, $E \ln Y \leq 0$ for each $Y \in \mathcal{L}_{\mathbf{X}}$. On the other hand $\mathbf{1} \in \mathcal{L}_{\mathbf{X}}$ and $E \ln \mathbf{1} = 0$ thus $X^* = \mathbf{1}$ by uniqueness of X^*.

□

One should expect that a solution f^* obtained from an infinite union of increasing sets of normalized constraints $\mathbf{X}_n = \{X_i,\ i \in \mathcal{I}_n\},\ n = 1, 2, \ldots, \infty$, $\mathcal{I}_n \subset \mathcal{I}_{n+1}$ should be the limit of the respective solutions f_n^*. This is shown next.

Theorem 2. *Consider a sequence of inverse problems with constraints*

$$\mathbf{X}_n = \{X_i,\ i \in \mathcal{I}_n\},\ n = 1, 2, \ldots, \infty,$$

where $\mathcal{I}_n \subseteq \mathcal{I}_{n+1} \subseteq \mathcal{I}_\infty \stackrel{def}{=} \bigcup_{n\in\mathbb{N}} \mathcal{I}_n$. Assume that the (extended) PD X_n^* exists for each of the constraints $\mathbf{X}_n$ and that

$$L = \sup\{E \ln X :\ X \in \bigcup_{n\in\mathbb{N}} \mathcal{L}_{\mathbf{X}_n}\} < \infty.$$

Then there exists $\tilde{X} \in \mathcal{L}_{\mathbf{X}_\infty}$ whose expected logarithm is at least as large as L and X_n^* converges in probability to $\tilde{X}$.

Proof. Consider $n \le m$. Due to Theorem 1 we have $E(X_n^*/X_m^*) \le 1$ hence

$$\begin{aligned} 0 &\le E\left((X_n^*/X_m^*)^{1/2} - 1\right)^2 \\ &\le 2 - 2E(X_n^*/X_m^*)^{1/2} \\ &= 2 - 2E\exp\left(\frac{1}{2}\ln(X_n^*/X_m^*)\right) \\ &\le 2 - 2\exp\left(\frac{1}{2}(E\ln X_n^* - E\ln X_m^*)\right) \end{aligned}$$

Next use the fact that $\lim_{n\to\infty,m\to\infty} E\ln X_n^* - E\ln X_m^* = 0$ which follows from the fact that the sequence $E\ln X_n^*$ is increasing and bounded by L and thus convergent. Then the above inequality implies that in probability $\lim_{n,m\to\infty}\sqrt{X_n^*/X_m^*} = 1$ which shows that the sequence $\ln X_n^*$ is Cauchy (in topology of convergence in probability P) and $Y = \lim_{n\to\infty}\ln X_n^*$ exists in probablity. Take $\tilde{X} = e^Y$, then $\lim_{n\to\infty}\exp(\ln X_n^*) = \tilde{X}$ in probability.

Let $n \le m$ then again due to Theorem 1 we have $E(X_n^*/X_m^*) \le 1$. Hence by the Fatou lemma and passing m to ∞ we obtain $E(X_n^*/\tilde{X}) \le 1$. Then by concavity of the logarithmic function we get

$$E\ln X_n^* - E\ln\tilde{X} \le \ln\left(E(X_n^*/\tilde{X})\right) \le 0.$$

This proves that $E\ln\tilde{X}$ is at least as large as L.

□

Remark 3. It is easy to notice that the result holds if it is only assumed that X_n^*'s exist over some increasing convex sets $\mathcal{L}_n \subset \mathcal{L}_{\mathbf{X}}$ and with $\mathcal{L}_\infty = \overline{\bigcup_{i\in\mathbb{N}} \mathcal{L}_i}$.

2.3. **Finite set of constraints.** In this subsection we study the case of finitely many constraints which is the setup for our Monte Carlo algorithm of Section 3. Here we assume that $\mathcal{I} = \{1, \ldots, N\}$ thus our vector of constraints is $\mathbf{X} = (X_1, \ldots, X_N)$. We also assume that $\overline{X} \stackrel{def}{=} (X_1 + \cdots + X_N)/N$ is positive with probability one. Let $\Pi_{\mathbf{X}} \stackrel{def}{=} \{\pi \in \Pi : \pi \cdot \boldsymbol{X} \geq 0\}$. The following result provides the conditions under which the PD exists on the simplex and on the extended simplex.

Theorem 3. *Under the assumptions of this subsection, there exists $\pi^* \in \Pi^+$ such that for each $\pi \in \Pi^+$:*

$$E \ln(\pi \cdot \mathbf{X}/\pi^* \cdot \mathbf{X}) \leq 0. \tag{5}$$

If additionally we assume that $\Pi_{\mathbf{X}}$ is compact then there exists $\pi^* \in \Pi_{\mathbf{X}}$ such that (5) holds for all $\pi \in \Pi_{\mathbf{X}}$.

Proof. The proof follows directly from Lemma 3 that is formulated and proven in the Appendix by applying

$$E \ln(\pi \cdot \mathbf{X}/\pi^* \cdot \mathbf{X}) = E \ln(\pi \cdot \mathbf{X}/\overline{X}) - E \ln(\pi^* \cdot \mathbf{X}/\overline{X}).$$

□

Remark 4. Clearly, if in the above results $E|\ln(\pi^* \cdot \mathbf{X})| < \infty$, then $X^* = \pi^* \cdot \mathbf{X}$ is an argument that maximizes $E \ln X$ over $\mathcal{L}^+_{\mathbf{X}}$ and $\mathcal{L}_{\mathbf{X}}$, respectively. This is granted if, for example, all $E|\ln X_i|$'s are finite (see Lemma 2 in the Appendix).

In the next theorem we provide several characterizations of the existence of the PD on the extended simplex.

Theorem 4. *Assume that for $a \in \mathbb{R}^N$: $a \cdot \boldsymbol{X} = 0$ a.s. if and only if $a = 0$. The following are equivalent*

(i): The convex set $\Pi_{\mathbf{X}}$ is compact.

(ii): There exists $\pi^* \in \Pi_{\mathbf{X}}$ such that $E \ln(\pi \cdot \mathbf{X}/\pi^* \cdot \mathbf{X}) \leq 0$.

(iii): $\sup\{E\ln(\pi\cdot\mathbf{X}/\overline{X}):\ \pi\in\Pi_{\mathbf{X}}\}<\infty$.

(iv): There do not exist $a\cdot\mathbf{1}=b\cdot\mathbf{1}=1,\ a\neq b$ *with* $P(a\cdot\boldsymbol{X}\geq b\cdot\boldsymbol{X}\geq 0)=1$ *and* $P(a\cdot\boldsymbol{X}>b\cdot\boldsymbol{X})>0$.

Proof. *(i)* implies *(ii)* follows from Lemma 3.

(ii) implies *(iii)*. By *(ii)* we have

$$E\ln(\pi\cdot\mathbf{X}/\overline{X})=E\ln(\pi\cdot\mathbf{X}/\pi^*\cdot\mathbf{X})+E\ln(\pi^*\cdot\mathbf{X}/\overline{X})\leq\ln(\max(\pi^*)N).$$

(iii) implies *(iv)*. Indeed, suppose there exist $a\cdot\mathbf{1}=b\cdot\mathbf{1}=1$ such that $a\cdot\boldsymbol{X}\geq b\cdot\boldsymbol{X}\geq 0$ a.s. and $P(a\cdot\boldsymbol{X}>b\cdot\boldsymbol{X})>0$. By replacing, if necessary, a and b by $\frac{1}{N}((N-1)a+\mathbf{1}/N)$ and $\frac{1}{N}((N-1)b+\mathbf{1}/N)$, respectively, we may assume that

$$b\cdot\boldsymbol{X}\geq\frac{1}{N}\overline{X}.$$

Then for each positive α: $X_\alpha=\alpha\,a\cdot\boldsymbol{X}+(1-\alpha)b\cdot\boldsymbol{X}\geq\frac{1}{N}\overline{X}$ and $X_\alpha\in\mathcal{L}_{\mathbf{X}}$. Additionally $\lim_{\alpha\to\infty}X_\alpha/\overline{X}=\infty$ on the set $\{a\cdot\boldsymbol{X}>b\cdot\boldsymbol{X}\}$. Consequently $\sup_\alpha E\ln\left(X_\alpha/\overline{X}\right)=\infty$.

(iv) implies *(i)*. Suppose that for every $n\geq 1$, $\pi^n\cdot\boldsymbol{X}\geq 0$, $\pi^n\cdot\mathbf{1}=1$ but (contrary to *(i)*) the sum of the positive entries of π^n diverges to ∞. There is only a finite number of coordinates and permutations of these coordinates. So there is a $d<N$ and a fixed permutation of $\mathcal{I}$ for which we may select a subsequence $\{n\}$ with precisely the first d coordinates being positive.

For any vector $a\in\mathbb{R}^N$ we define $u(a)=(a_1,\ldots,a_d),\ v(a)=(a_{d+1},\ldots,a_N)$. Then

$$0\leq\pi^n\cdot\boldsymbol{X}=u(\pi^n)\cdot u(\mathbf{1})\left[\frac{u(\pi^n)}{u(\pi^n)\cdot u(\mathbf{1})}\right]\cdot u(\boldsymbol{X})+v(\pi^n)\cdot v(\mathbf{1})\left[\frac{v(\pi^n)}{v(\pi^n)\cdot v(\mathbf{1})}\right]\cdot v(\boldsymbol{X}) \tag{6}$$

The bracketed terms above belong to d and $N-d$ dimensional simplexes, respectively, so we can select a further subsequence $\{n\}$ such that they converge respectively to $u(\tilde{\pi}_1)$ with $\tilde{\pi}_1$ having the last $N-d$ coordinates equal to zero and $v(\tilde{\pi}_2)$ with $\tilde{\pi}_2$ having the first d coordinates equal to zero and such that $\tilde{\pi}_1\cdot\mathbf{1}=\tilde{\pi}_2\cdot\mathbf{1}=1$. Therefore from (6) and since $u(\pi^n)\cdot u(\mathbf{1})$ converges to infinity, it follows that $\tilde{\pi}_1\cdot\boldsymbol{X}-\tilde{\pi}_2\cdot\boldsymbol{X}\geq 0$ a.s. which cannot almost surely be 0 because it would contradict the assumption of linear independence of X_i's (the first d terms of $\mathbf{X}$ would have to be expressed by the last $N-d$ ones). This contradiction of *(iv)* concludes the proof.

□

Remark 5. Note that in the above theorem we have used the assumed linear independence only to prove *(iv)* implies *(i)*. That the linear independence is needed in this step is shown by the following counterexample. Take $\mathbf{X} = (X_1, X_1)$. Then *(iv)* holds but $\Pi_{\mathbf{X}} = (-\infty, \infty)$.

Corollary 2. *Let $X^* = \pi^* \cdot \mathbf{X}$ denote the PD based on the simplex Π^+.*

- *(i): If $\pi^* \cdot \mathbf{X}$ is in the interior of the simplex $\mathcal{L}^+_{\mathbf{X}}$ treated as a subset of the linear space spanned by $\mathbf{X}$ equipped with the canonical topology of finite dimensional spaces, then it coincides with the PD based on the extended simplex. In particular this is true if π^* is in the interior of Π^+ (this condition is equivalent to the assumption in the case when the coordinates of $\mathbf{X}$ are linearly independent).*
- *(ii): If π^* is on the boundary of the simplex Π^+, then dropping constraints X_i with $\pi^*_i = 0$ will result in a reduced set of constraints $\widetilde{\mathbf{X}}$ for which the extended simplex PD coincides with the PD $\widetilde{X}^* = \tilde{\pi}^* \cdot \widetilde{\mathbf{X}}$ based on the simplex $\widetilde{\Pi}^+$ corresponding to reduced constraints. Moreover $\tilde{\pi}^* \cdot \widetilde{\mathbf{X}}$ is in the interior of the simplex $\mathcal{L}^+_{\widetilde{\mathbf{X}}}$.*

Proof. *(i)* Suppose that there is $\pi^{**} \in \Pi_{\mathbf{X}}$ such that $E\ln(\pi^{**} \cdot \mathbf{X}) > E\ln(\pi^* \cdot \mathbf{X})$. Consider the concave mapping $\delta \mapsto E(\ln(((1-\delta)\pi^* + \delta\pi^{**}) \cdot \mathbf{X})$. By the assumption there exists $\delta_0 > 0$ such that the value of this mapping is at most $E\ln(\pi^* \cdot \mathbf{X})$. On the other hand, by concavity we have the following contradiction

$$\begin{aligned} E(\ln(\pi^* \cdot \mathbf{X})) &\geq E(\ln(((1-\delta_0)\pi^* + \delta_0\pi^{**}) \cdot \mathbf{X}) \\ &\geq (1-\delta_0)E\ln(\pi^* \cdot \mathbf{X}) + \delta_0 E\ln(\pi^{**} \cdot \mathbf{X}) \\ &> E\ln(\pi^* \cdot \mathbf{X}). \end{aligned}$$

(ii) This is a simple consequence of *(i)*.

□

2.4. **Connections with Csiszár and Tusnády.** For two mutually absolutely continuous measures P and Q the Kullback-Leibler informational divergence is defined as $D(P|Q) = \int \log(dP/dQ)dP$. The informational distance of P from a convex set of measures $\mathcal{Q}$ is defined as $D(P|\mathcal{Q}) = \inf_{Q\in\mathcal{Q}} D(P|Q)$.

Consider the case $\mathcal{I} = \{1, \dots, N\}$. For $f > 0$ define measures μ_i by $d\mu_i/dP = X_i f$, $i \in \mathcal{I}$. Provided that the informational distances below exist and are finite, we have for $\pi, \pi' \in \Pi$:

$$\begin{aligned} D\left(P \middle| \sum_i \pi_i\mu_i\right) - D\left(P \middle| \sum_i \pi_i'\mu_i\right) &= \int \log\frac{dP}{\sum_i \pi_i d\mu_i}dP - \int \log\frac{dP}{\sum_i \pi_i' d\mu_i}dP \\ &= \int \left(\log\sum_i \pi_i' X_i - \log\sum_i \pi_i X_i\right) dP. \end{aligned}$$

Then $D(P|\{\sum_i \pi_i\mu_i : \pi \in \Pi\}) = D(P|\sum_i \pi_i^*\mu_i)$ for any $\pi^* = \mathrm{argmax}\{E\ln(\pi\cdot\mathbf{X}) : \pi \in \Pi\}$. Let $\mu^* = \sum_i \pi_i^*\mu_i$.

The above argument leads to the following proposition that was proven in Csiszár and Tusnády [3] for the simplex.

Proposition 2. *If one of the conditions of Theorem 4 is met and $f > 0$, then*

$$D(P|\mu^*) = D\left(P \middle| \left\{\sum_i \pi_i\mu_i : \pi \in \Pi\right\}\right)$$

The counterpart of this result for the case of Π^+ follows easily from [3], Theorem 5, p. 223. Their proof uses alternating projections. Our result can be written $P \xrightarrow{1} \mu^*$ in the notation of [3]. That is, aside from f, X^* is the density of the projection of P.

If we take f equal to the PDR solution $f^*1 = 1/X^*$ following from Theorem 1. Then $d\mu^* = \sum_i \pi_i^* X_i f^* dP = dP$. On the other hand if $P = \mu^*$, then $\int_A X^* f dP = P(A)$ for all measurable A from which we obtain $f = 1/X^*$. So $1/X^*$ (or $\frac{dP}{d\mu}/X^*$) is that solution of the inverse problem for which $D(P|\sum_i \pi_i\mu_i, \pi \in \Pi) = 0$.

3. Monte Carlo solution over the simplex

3.1. **Adaptive strategies for the projective density.** For this section we assume that $\mathcal{I} = \{1, \dots, N\}$ and $\overline{X} > 0$ with probability one. As we have seen in Subsection 2.3 (see also Lemma 3), this guarantees the existence of π^* such that $J(\pi^*) = \max\{J(\pi), \pi \in \Pi^+\}$, where $J(\pi) = E \ln(\pi \cdot \mathbf{X}/\overline{X}),\ \pi \in \Pi^+$, which is equivalent to the existence of the PD over the simplex. Also in the following algorithm and results, all assumptions and definitions are equivalent when $\mathbf{X}$ is replaced by $\mathbf{X}/\bar{X}$. For this reason in what follows we can assume that $\mathbf{X}$ is such that $\overline{X} = 1$ so in particular $J(\pi) = E \ln(\pi \cdot \mathbf{X}) < \infty$.

For a given working probability measure P consider a sequence of i.i.d. draws $\omega_j \in \Omega,\ j \geq 1$. Define $\boldsymbol{X}^j = (X_1(\omega_j), \ldots, X_N(\omega_j))$. Then $\boldsymbol{X}^j,\ j \geq 1$ are i.i.d. random vectors possessing the distribution of the constraint functions applied to P random samples $\{\omega_j\}$. We will produce a sequence $\overline{\pi}_n = \overline{\pi}_n(\boldsymbol{X}^1, \ldots, \boldsymbol{X}^n)$ almost surly convergent to the set $\Pi^* = \{\pi^* \in \Pi^+ : \pi^* = \text{argmax}\{\text{J}(\pi), \pi \in \Pi^+\}\}$. This will be true whether or not the constraints are linearly consistent.

Now let $\eta_j > 0$ be a sequence of numbers and let $\pi_1 = (1/N, \ldots, 1/N)$. We generate the following sequence of updates

$$\pi_i^{j+1} = \pi_i^j \frac{\exp\left(\eta_j \frac{X_i^j}{\pi^j \cdot \boldsymbol{X}^j}\right)}{\sum_{k=1}^N \pi_k^j \exp\left(\eta_j \frac{X_k^j}{\pi^j \cdot \boldsymbol{X}^j}\right)}, \quad i = 1, \ldots, N. \tag{7}$$

This is the exponentiated gradient algorithm $EG(\eta)$ of [5] except that we are extending it to varying $\eta = \eta_j$. Let $M_j = \max\{X_i^j, i = 1, \ldots, N\}$ and $m_j = \min\{X_i^j, i = 1, \ldots, N\}$.

Proposition 3. *Assume $\sum_{j=1}^\infty \eta_j^2 < \infty$. If there is a positive r such that $m_j \geq r M_j$ a.s. for all $j \geq 1$, then*

$$\sum_{j=1}^\infty \eta_j (J(\pi^*) - J(\pi^j)) < \infty \quad a.s.$$

Proof. Let $\triangle_j = D(\pi^*|\pi^{j+1}) - D(\pi^*|\pi^j)$, where π^*, π^{j+1} and π^j are treated as discrete probability measures and $D(\cdot|\cdot)$ stands for the relative entropy. Apply an inequality from [5] (see the proof of Theorem 1 therein) for $\boldsymbol{X}^j$ to obtain,

$$\triangle_j \leq -\eta_j [\log(\pi^* \cdot \boldsymbol{X}^j) - \log(\pi^j \cdot \boldsymbol{X}^j)] + \frac{\eta_j^2}{8r^2}.$$

Taking conditional expectation w.r.t. $\mathcal{F}_{j-1}$ the σ-algebra generated by $\{\boldsymbol{X}^t : 1 \le t \le j-1\}$,

$$E_{j-1}\triangle_j \le -\eta_j[E_{j-1}\log(\pi^* \cdot \boldsymbol{X}^j) - E_{j-1}\log(\pi^j \cdot \boldsymbol{X}^j)] + \frac{\eta_j^2}{8r^2} = -\eta_j(J(\pi^*) - J(\pi^j)) + \frac{\eta_j^2}{8r^2}.$$

Hence

$$E_{j-1}\triangle_j + \eta_j(J(\pi^*) - J(\pi^j)) \le \frac{\eta_j^2}{8r^2}.$$

Finally taking the unconditional expectation we have

$$E\triangle_j + \eta_j E(J(\pi^*) - J(\pi^j)) \le \frac{\eta_j^2}{8r^2}.$$

Summing from 1 to n obtain

$$E(D(\pi^*|\pi^{n+1}) - D(\pi^*|\pi^1)) + \sum_{j=1}^n \eta_j E(J(\pi^*) - J(\pi^j)) \le \sum_{j=1}^n \frac{\eta_j^2}{8r^2}.$$

Since $D(\pi^*|\pi^1) \le \log N$ and $J(\pi^*) - J(\pi^J) \ge 0$ we obtain

$$0 \le \sum_{j=1}^n \eta_j E(J(\pi^*) - J(\pi^j)) \le \sum_{j=1}^n \frac{\eta_j^2}{8r^2} + \log N.$$

By the assumptions it implies that with probability one

$$\sum_{j=1}^n \eta_j(J(\pi^*) - J(\pi^j)) < \infty.$$

□

The assumption that $m_j \ge rM_j$ a.s. for some positive r and for all $j \ge 1$ can be removed by the following modification as in [5]. For a sequence of numbers $0 < \alpha_j < N$ define

$$\widetilde{\boldsymbol{X}}^j = (1 - \alpha_j/N)\boldsymbol{X}^j + \alpha_j M_j \mathbf{1},$$

where $M^j = \max\{X_i^j, i = 1, \dots, N\}$ and where $\mathbf{1}$ is a unit vector. For $\widetilde{\boldsymbol{X}}^j$ construct $\widetilde{\pi}^j$ as in (7). Finally take

$$\pi^j = (1 - \alpha_j/N)\widetilde{\pi}^j + (\alpha_j/N)\mathbf{1}.$$

Proposition 4. *We may choose* η_j *and* α_j *such that* $\sum_{j=1}^{\infty}\left(\frac{\eta_j^2}{8\alpha_j^2}+2\alpha_j\eta_j\right)<\infty$. *Under this condition*

$$\sum_{j=1}^{\infty}\eta_j(J(\pi^*)-J(\pi^j))<\infty \quad a.s.$$

Proof. We just need slightly modify the proof of Proposition 3 so we use the same notation. An inequality from [5] yields the following estimate

$$\log(\pi^*\cdot\boldsymbol{X}^j)-\log(\pi^j\cdot\boldsymbol{X}^j)\le\log(\pi^*\cdot\widetilde{\boldsymbol{X}}^j)-\log(\widetilde{\pi}^j\cdot\widetilde{\boldsymbol{X}}^j)+2\alpha_j.$$

Next we observe that $\min\{\widetilde{X}_i^j, i=1,\ldots,N\}\ge\alpha_j\max\{\widetilde{X}_i^j, i=1,\ldots,N\}$ so proceeding as in the proof of Proposition 3 obtain

$$E(D(\pi^*|\pi^{n+1})-D(\pi|\pi^1))+\sum_{j=1}^{n}\eta_jE(J(\pi^*)-J(\pi^j))\le\sum_{j=1}^{n}\left(\frac{\eta_j^2}{8\alpha_j^2}+2\alpha_j\eta_j\right).$$

□

Corollary 3. *Let* $\sum_{j=1}^{\infty}\eta_j=\infty$ *in addition to assumptions of Proposition 3 or 4. There exist* $k_n\le n$ *such that* $\bar{\eta}_n=\sum_{j=k_n}^{n}\eta_j\to 1$ *as* $n\to\infty$ *such that for each* $\pi\in\Pi^+$*:*

$$\lim_{n\to\infty}J\left(\sum_{j=k_n}^{n}\frac{\eta_j}{\bar{\eta}_n}\pi^j\right)=J(\pi^*)\quad a.s.$$

Proof. It is possible to choose $k_n\to\infty$ in such a way that $\sum_{j=k_n}^{n}\eta_j\to 1$ as $n\to\infty$ since the series $\sum_{j=1}^{\infty}\eta_j$ is divergent. Next

$$\sum_{j=k_n}^{n}\eta_j(J(\pi^*)-J(\pi^j))\to 0 \text{ if } n\to\infty$$

which follows from Proposition 3 or 4. By concavity of J we obtain

$$J(\pi^*)\ge J\left(\sum_{j=k_n}^{n}\frac{\eta_j}{\bar{\eta}_n}\pi^j\right)\ge\sum_{j=k_n}^{n}\frac{\eta_j}{\bar{\eta}_n}J(\pi^j)$$

and conclusion follows since the right hand side convereges to $J(\pi^*)$.

□

If conditions of the Proposition 3 are met one of possible choices is $\eta_j = \frac{1}{j^{1/2}(\log j)^{1/2+\epsilon}}$, $\epsilon > 0$. In this case $k_n = n - [n^{1/2}(\log n)^{1/2+\epsilon}]$. In the more general setting of Proposition 4 a possible choice is $\eta_j = \frac{1}{j^{3/4}(\log j)^{1+\epsilon}}$, $\epsilon > 0$ and $\alpha_j = \frac{1}{j^{1/4}}$. In this case $k_n = n - [n^{3/4}(\log n)^{1+\epsilon}]$.

Remark 6. If Π^* is not empty, then the sequence $\overline{\pi} \in \Pi^+$ defined as

$$\overline{\pi}_n = \sum_{j=k_n}^{n} \eta_j \pi^j$$

satisfies

$$\lim_{n\to\infty} ||\overline{\pi}_n - \Pi^*|| = 0 \text{ a.s.},$$

where $||\cdot||$ is the Euclidean distance in N-dimensional space. This fact follows directly from concavity of $J(\pi)$.

Remark 7. An alternative approach to finding π^* could be through Cover's algorithm

$$\pi_i^{n+1} = \pi_i^n E\left(\frac{X_i}{\pi^n \cdot \boldsymbol{X}}\right) \to \pi_i^*.$$

This algorithm requires integration at each step. By comparison, our algorithm avoids integration step. More details on Cover's algorithm can be found in [2] and [3].

3.2. **Finding PD over the extended simplex.** We continue all assumptions and notation of the previous section. Additionally it is assumed that $\Pi_{\mathbf{X}}$ is compact so the PD $X^* = \pi^* \cdot \mathbf{X}$ exists (see Subsection 2.3). It is recognized that our algorithm from Subsection 3.1 produces only a maximum over $\mathcal{L}^+_{\mathbf{X}}$. The ultimate goal is, however, to maximize $E\log(X)$ for $X \in \mathcal{L}_{\mathbf{X}}$ or, equivalently, $\phi_{\mathbf{X}}(\pi) = E\ln(\pi \cdot \mathbf{X})$ over $\Pi_{\mathbf{X}}$. One practical approach would be based on the idea of the super simplex (see Proposition 1) and Corollary 2. Namely, one could use the MC-inverse for increasing finite sets of constraints $\mathbf{X}_n$ such that $\mathcal{L}^+_{\mathbf{X}_n}$ would gradually fill in the extended simplex $\mathcal{L}_{\mathbf{X}}$. In such an approach, we start with evaluating the simplex solution π^* for the constraints $\mathbf{X}_N \stackrel{def}{=} \mathbf{X}$ using the MC-inverse. If the solution belongs to the interior of the simplex, then it coincides with the extended simplex solution and the problem is solved. Otherwise we add an additional constraint $X_{N+1} = \pi_{N+1} \cdot \mathbf{X}$, $\pi_N \in \Pi_{\mathbf{X}_N}$ to the original constraints $\mathbf{X}_N$ and obtain $\mathbf{X}_{N+1}$. The MC-inverse then can be applied to $\mathbf{X}_{N+1}$. These steps can be repeated recursively until the solution is found.

Of course, there are several issues to be resolved for a procedure like this to work. First, a method of choice of a new constraint X_{n+1} should be selected. In the random search algorithm it could be selected by random choice of $\pi_{n+1} \in \Pi_{\mathbf{X}_n} \setminus \Pi_n^+$ according to uniform distribution. These would guarantee that if the extended solution is in the interior of $\Pi_{\mathbf{X}}$ then with probability one the extended solution would be eventually found. However, the direct random choice of π_{n+1} poses a practical problem as $\Pi_{\mathbf{X}}$ can and often will be not known in an explicit form. Thus a random rejection algorithm often would be a more realistic method. In this we sample π from some continuous distribution (possibly close to the non-informative uniform over $\Pi_{\mathbf{X}_n}$) over all $\pi \in \mathbb{R}^n$ such that $\pi \cdot \mathbf{1} = 1$ and accept this choice if $\pi \cdot \mathbf{X}_n \geq 0$, i.e. if $\pi \in \Pi_{\mathbf{X}_n}$. If this is not the case reject π and repeat sampling until $\pi \in \Pi_{\mathbf{X}_n}$ is found.

The above simple approach demonstrates a possibility of searching over the extended simplex using our MC-inverse. However it is neither computationally efficient nor recommended. Firstly, it increases unnecessarily the dimension of the MC-inverse problem (new constraints are added). Secondly, it does not utilize the fact that the solution on an edge of a simplex excludes from further search the interior of this simplex (as well as an even larger region made of open half-lines starting at the edge and passing through the interior of the simplex) .

A more effective algorithm free of these two deficiencies is presented in the Appendix. In this algorithm only simplexes of at most the dimension of the original set of constraints are searched. Moreover the area of search is updated and reduced each time the simplex solution falls on the boundary of a simplex.

4. Application – reconstruction of a density

4.1. **Reconstruction of a density from its moments.** We conclude with two examples of the MC-inverse solutions for the problem of recovery of a density given that some of its moments are known. In both examples we would like to restore some unknown density on interval $[0,1]$. The general probem of density reconstruction can be formulated as follows. Let μ be the Lebesgue measure on $[0,1]$. Assume that $\int_0^1 g_i f d\mu = c_i$, $i = 1, \dots, N$ are known for non-negative constraints $g_i(\omega) = \omega^{i-1} + 0.1$. Here we shift the constraints by 0.1 to be in the domain of Proposition 3 avoiding implementation of the slightly more complicated algorithm that follows from Proposition 4. Define $X_i = g_i/c_i$. We take as a working measure the uniform distribution on $[0,1]$ so the extended PDR solution is given by $1/X^*$, where X^* has to be found by the means of the MC-inverse algorithm of Section 3.

In the empirical MC-inverse extension of the approach instead of the true moments of the unknown density we consider their empirical counterparts, say $\hat{c}_i$. In this case, we assume that an iid sample $Y_1, \dots, Y_k$ from the true distribution is given and empirical constraints $\hat{c}_i = (Y_1^{i-1} + \cdots + Y_k^{i-1})/k + 0.1$ are replacing the true (unknown) ones.

It should be mentioned that considering constraints based on non-central moments is a completely arbitrary choice. In fact, the method can be applied to virtually any constraints suitable for a problem at hand. Moreover, we are free to choose various working measures just as we do for (say) a Bayesian analysis. Also in the empirical approach data can be conveniently transformed if needed as well as the constraints of a choice can be arbitrarily added to extract more information from the data. Influence of all these factors on the solution has to be investigated for a particular problem. Here we have restricted ourselves to the non-central moment problem to present the essentials of the proposed methods.

4.2. **Reconstruction of uniform distribution.** In the first instance, we consider the first four moments of the 'unknown' density to be equal to the moments of uniform distribution on $[0,1]$, i.e. $c_i = 1/i + 0.1$, $i = 1, \dots, 5$. So $X_i = (\omega^{i-1} + 0.1)/(1/i + 0.1)$ and for these constraints and the uniform working measure the MC-invese algorithm has been run with Monte Carlo sample size of $n = 10000$. Clearly, since the chosen uniform working measure satisfies the constraints, the algorithm should return *approximately* the uniform distribution (see Corollary 1) that corresponds to $\pi^* = (1, 0, 0, 0, 0)$. In Figure 3*(top)*, we summarize results of our numerical study. For the empirical MC-inverse we generate an iid sample $Y_1, \dots, Y_k$ from the 'unknown' (uniform) distribution and consider $\hat{c}_i$ instead of the

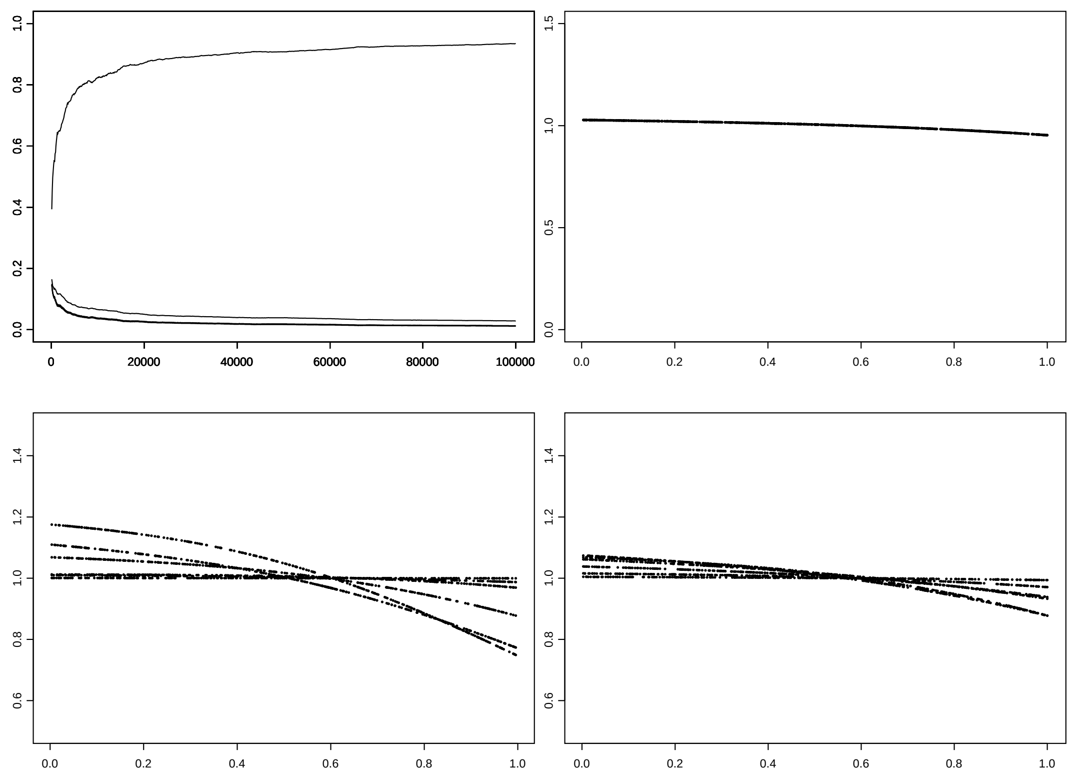

FIGURE 3. Reconstruction of uniform distribution. *Top:* Convergence of π^j to π^* *(left)*. The reconstruction of uniform distribution from the first four raw moments *(right)*. *Bottom:* Reconstructions based on empirical moments for six samples from uniform. Sample size $k = 20$ *(left)* and $k = 100$ *(right)*.

true moments. These constraints do not coincide with the exact moments of the uniform distribution and thus the MC-inverse solution does not need to be as closely uniform. How robust our reconstruction relatively to sample size, number of the constraints and their form is not addressed in the present paper and calls for further studies. Here, the performance of the algorithm for two sample sizes $k = 20$ and $k = 100$ is presented in Figure 3 *(Bottom)*.

The reconstructions are rather good for both sample sizes (the vertical range of graphs is from 0.6 to 1.4) but it should be noted that the solution is located at a vertex of the simplex (zero dimensional edge), in fact $\pi^* = (1, 0, \ldots, 0)$. We did not question if this simplex solution is also the solution over the extended simplex as we knew this a'priori when constructing the example. However, a simplex solution on the edge of simplex should call for further search over the extended simplex.

4.3. **Reconstruction based on moments of beta distribution.** This time we choose an initial working measure that does not satisfy the constraints and moreover the actual PD solution extends beyond the initial simplex. As a result an extended simplex search has to be implemented.

The unknown density that we want to reconstruct has the moments agreeing with beta distribution with parameters $\alpha = 2$ and $\beta = 2$. Thus $c_i = \alpha/(\alpha+\beta)\cdot(\alpha+1)/(\alpha+\beta+1)\cdot\dots\cdot(\alpha+i-2)/(\alpha+\beta+i-2)+0.1,\ i > 1$. We consider the case of three constraints, i.e. we assume that we know exact values of the first and second moment. With our choice of the parameters, these moments are 0.5 and 0.3, respectively. The constraints are as follows $X_1 = 1$ (this constraint is present for any density reconstruction problem), $X_2 = (\omega+0.1)/0.6$, and $X_3 = (\omega^2+0.1)/0.4$. MC-inverse has been run and its results are shown in Figure 4*(top-left)*. We see there that the solution Y^* is on the boundary of the initial simplex given by $\mathbf{X} = (X_1, X_2, X_3)$. Specifically we observe $\pi_2 \approx 0$ while $\pi_1 \approx 0.827$ and $\pi_3 \approx 0.173$. Consequently, the simplex PD solution is on the edge spanned by X_2 and X_3. The PDR based on this solution is shown in Figure 4*(bottom-left)*

A solution on an edge of a simplex requires searching the extended simplex. In this case we replace the constraint X_2 (since $\pi_2 \approx 0$) with a new constraint $\widetilde{X}_2 = \pi\cdot\mathbf{X}$ where $\pi = (2.7089931, -2.2, 0.4910069)$ (note negative π_2). The MC-inverse algorithm was run and the results are presented in the two middle pictures of Figure 4. We observe that the solution again lands on an edge of the simplex. Namely, this time $\pi_1 \approx 0$ and $\pi_2 \approx \pi_3$. Thus we continued our search by replacing X_1 with a new constraint $\widetilde{X}_1 = \pi\cdot\mathbf{X}$ where $\pi = (4.339548, -10.9956, 7.656052)$ and $\mathbf{X}$ are the original constraints. This constraint was chosen to be outside of the last simplex and on the opposite to X_1 side of the edge spanned by $\widetilde{X}_2$ and X_3 (as suggested by the algorithm in Subsection 5.2 of the Appendix).

The MC-inverse algorithm was run for this third set of constraints $(\widetilde{X}_1, \widetilde{X}_2, X_3)$ and this time the PD solution was found in the interior of the simplex given by these constraints.

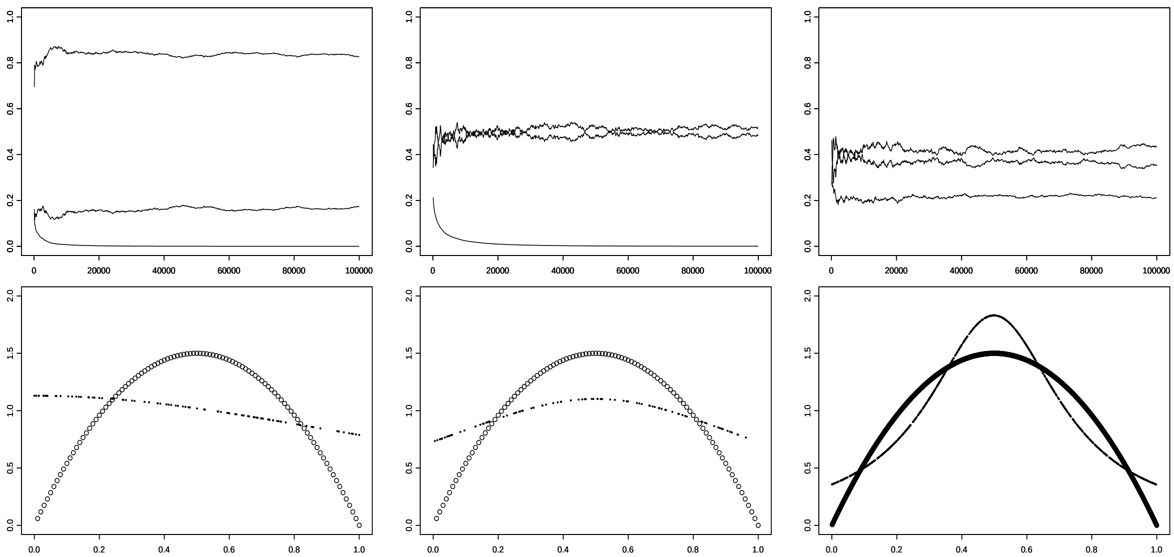

FIGURE 4. MCI solution for the initial simplex in the beta moments problem. At the top pictures we present convergence of π^j, $j = 1, \ldots, 100000$ for the subsequent simplexes starting from the initial one *(left)* and the final one for which the solution is in its interior. The bottom pictures shows the RPD for each of the simplexes, the most right one being the extended simplex solution.

In fact, $\pi^* \approx (0.4223826, 0.3581372, 0.2194802)$, thus all coordinates are clearly bigger than zero, see also Figure 4 *(top-right)*. Thus as shown in Corollary 2, this solution coincides with the extended simplex solution and thus the corresponding PDR solution is the final reconstruction of the density based on the first two moments of beta distribution. This solution, presented in Figure 4 *(bottom-right),* is different from the beta distribution (also shown in this graph). However there are infinitely many densities on $[0, 1]$ having the same first two moments as a beta distribution and it should not be expected that the PDR solution will coincide with the beta density (unless it will be chosen for a working measure).

5. Appendix

5.1. **Lemmas.** The following lemma is used in Theorem 1 of Subsection 2.2.

Lemma 1. *Let a random variable $|Z| \leq C$ for some positive C.*

(i): Assume that $E \ln(1 + \alpha Z) \leq 0 \quad for \quad 0 < \alpha \leq \alpha_0$ for some $\alpha_0 > 0$. Then

$$EZ \leq 0.$$

(ii): Assume that $E \ln(1 + \alpha Z) \leq 0 \quad for \quad \alpha_1 \leq \alpha < 0$ for some $\alpha_0 < 0$. Then

$$EZ \geq 0.$$

Proof. *(i).* Apply elementary inequality $\ln(1 + x) \geq x - x^2, \quad x \geq -1/2$ to αZ for sufficiently small positive α. Then by the assumption

$$\alpha EZ - \alpha^2 EZ^2 \leq E \ln(1 + \alpha Z) \leq 0.$$

Hence

$$EZ \leq \alpha EZ^2 \to 0, \quad \alpha \to 0+.$$

(ii) follows from *(i)* by considering $-Z$ instead of Z.

□

The next two lemmas were used for the finite number constraints case in Subsection 2.3.

Lemma 2. *Let $\mathcal{I} = \{1, \ldots, N\}$ and X_i's are positive such that all $E|\ln X_i|$'s are finite. Then there exists $\pi^* \in \Pi^+$ such that for all $\pi \in \Pi^+$:*

$$E \ln(\pi \cdot \mathbf{X}) \leq E \ln(\pi^* \cdot \mathbf{X}).$$

Proof. Note the inequality for $\pi \in \Pi^+$:

$$\ln \min(X_1, 1) + \cdots + \ln \min(X_N, 1) \leq \ln \pi \cdot \mathbf{X} \leq \ln \max(X_1, 1) + \cdots + \ln \max(X_N, 1).$$

By the assumptions and the Bounded Convergence Theorem it proves the continuity of $\phi(\pi) = E \ln(\pi \cdot \mathbf{X})$ on Π^+. The existence of π^* follows then from the compactness of Π^+.

□

In the next lemma $\overline{X} = (X_1 + \cdots + X_N)/N$.

Lemma 3. *Let* $\mathcal{I} = \{1, \ldots, N\}$. *For each convex and compact set* $\Pi_0 \subseteq \{\pi \in \Pi : \pi \cdot \boldsymbol{X} \geq 0\}$ *including some* $Y > 0$, *there exists* $\pi^* \in \Pi_0$ *such that* $E \ln(\pi \cdot \mathbf{X}/\overline{X}) \leq E \ln(\pi^* \cdot \mathbf{X}/\overline{X})$ *for all* $\pi \in \Pi_0$.

Proof. Let $\mathbf{Y} = \mathbf{X}/Y$, where $Y = \pi_0 \cdot \mathbf{X} > 0$. By compactness there is a positive constant B such that for each $\pi \in \Pi_0$, $\max(\pi) < B$. Thus

$$\ln(\pi \cdot \mathbf{Y}) \leq \ln B + \ln(\mathbf{1} \cdot \mathbf{Y}) \leq \ln B + \ln(N).$$

This implies that $M_{\mathbf{Y}} \stackrel{def}{=} \sup\{E \ln(\pi \cdot \mathbf{Y}) : \pi \in \Pi_0\} < \infty$. Since $\ln(\pi_0 \cdot \mathbf{Y}) = 0$, thus $M_{\mathbf{Y}} \geq 0$.

By compactness of the set Π_0 there exists a sequence π_n of its elements such that $\phi(\pi_n) \stackrel{def}{=} E \ln(\pi_n \cdot \mathbf{Y})$ converges to $M_{\mathbf{Y}}$ and π_n converges to some $\pi^* \in \Pi_0$. We can assume that $\phi(\pi_n) > -\infty$ for all $n \in \mathbb{N}$. Consider the sequence $\mathcal{L}_n$ of simplexes spanned on the constraints $\{\pi_k \cdot \mathbf{Y} : k \leq n\}$. These constraints satisfy the assumptions of Lemma 2. Thus there exists a sequence Y_n^* of PD solutions over $\mathcal{L}_n$ such that $E \ln Y_n^* \geq \phi(\pi_n)$. Invoking Theorem 2, or rather Remark 3 that follows it, we obtain $\widetilde{Y} \in \mathcal{L}_{\mathbf{Y}}$ such that $M_{\mathbf{Y}} \geq E \ln \widetilde{Y} \geq E \ln Y_n^* \geq \phi(\pi_n)$. This implies that $E \ln \widetilde{Y} = M_{\mathbf{Y}}$ and by compactness of Π_0 it must be $\widetilde{Y} = \pi^* \cdot \mathbf{Y}$. Now, $Y > 0$ implies that $\overline{X} > 0$, so

$$E \ln(\pi \cdot \mathbf{X}/\overline{X}) = E \ln(\pi \cdot \mathbf{X}/Y) + E \ln(Y/\overline{X}) \leq E \ln(\pi^* \cdot \mathbf{X}/Y) + E \ln(Y/\overline{X}) = E \ln(\pi^* \cdot \mathbf{X}/\overline{X})$$

and the thesis follows.

□

5.2. **Search algorithm.** We prove the existence of a search algorithm for maximum over an extended simplex that utilizes our MC-inverse algorithm for a simplex. The algorithm is based on updating the constraints by replacing those that do not contribute to maximal values of $E \ln X$ by a randomly selected ones from the region that contains the extended simplex solution.

Let $\pi_N^*(\cdot)$ be the return value of the MC-inverse algorithm applied to an N dimensional simplex, i.e if $\mathbf{X} = (X_1, \ldots, X_N)$, then $\pi_N^*(\mathbf{X}) = \operatorname{argmax}\{\mathrm{E} \ln(\pi \cdot \mathbf{X}) : \pi \in \Pi^+\}$. We define a random search algorithm such that with probability one it returns value $X_N^*(\mathbf{X})$ that is the PD over the extended simplex $\mathcal{L}_{\mathbf{X}}$. For our search algorithm we need to assume that for a random value $\pi \in \mathbb{R}^N$ selected according to some continuous distribution on $\mathbb{R}^N$ (playing the role of a prior distribution for the optimal π^*) we can effectively determine if $\pi \in \Pi_{\mathbf{X}}$, i.e. if for $\pi \in \mathbb{R}^N$: $\pi \cdot \mathbf{X} \geq 0$. Our description of the algorithm is recursive with respect to the number of constraints N.

If $N = 1$, then define $X_1^*(X) = X$.

For $N = 2$, let $\pi^* = \pi_N^*(\mathbf{X})$. In this case, we identify $(\pi, 1-\pi)$ with π so $\Pi^+ = [0,1]$, and define $\Pi = \mathbb{R}$. If π^* is in the interior of Π^+, then return $X^* = \pi^* \cdot \mathbf{X}$ as the value $X_2^*(\mathbf{X})$ of the algorithm. If it is otherwise, then either $\pi^* = 0$ or 1. Consider the case $\pi^* = 1$ (the other is symmetric). Perform sampling from $\Pi \setminus (-\infty, 1]$ to obtain $\tilde{\pi}$ such that $\tilde{\pi} \cdot \mathbf{X} \geq 0$ and $\tilde{\pi} > 1$. Replace the original constraint X_2 by $\tilde{\pi} \cdot \mathbf{X}$ redefine Π for new constraints (by the linear transformation and excluding from it the area that was identified as not including the extended solution, i.e. the transformed $(-\infty, 1]$), then repeat this algorithm until the solution is found. It is clear that if the solution is in the interior of the extended simplex, then our random search with probability one ends up with a simplex containing in the interior the extended solution. In the next iteration the solution will be identified.

Below, we recursively extend this algorithm for a set of constraints $\mathbf{X} = (X_1, \dots, X_N)$ assuming that $X_k^*(\cdot)$, $k < N$ are available.

Step 0: Let $\widetilde{\mathbf{X}} = \mathbf{X}$ and $\Pi = \mathbb{R}^N$. Evaluate $\pi^* = \pi_N^*(\widetilde{\mathbf{X}})$.

Step 1: If $\pi^* \cdot \widetilde{\mathbf{X}}$ is in the interior of $\mathcal{L}^+_{\widetilde{\mathbf{X}}}$, then go to **Final Step**. Otherwise, $\pi^* \cdot \widetilde{\mathbf{X}}$ is on an edge of the initial simplex. Redefine Π by excluding from it all straight half-lines initiated in the edge and passing through the interior of the simplex (the excluded set is an intersection of finite number of half-hyperplanes and it is possible to write it in the form of linear inequalities). If the edge cointaining $\pi^* \cdot \widetilde{\mathbf{X}}$ in its interior is $N-1$-dimensional simplex, then go to **Step 2** otherwise go to **Step 3**.

Step 2: In this case there is exactly one coordinate of π^* that is equal zero. Say, it is the last one. Let $\pi^*_{\hat{N}} = (\pi_1^*, \dots, \pi_{N-1}^*)$, where for a vector x, the notation $x_{\hat{k}}$ defines a vector obtained from x by removing its kth coordinate. Since $\pi^*_{\hat{N}} \cdot \widetilde{\mathbf{X}}_{\hat{N}}$ is in the interior of the edge ($N-1$ dimensional simplex) thus it is also equal to $X^*_{N-1}(\widetilde{\mathbf{X}}_{\hat{N}})$. Apply rejection sampling from Π to obtain $\pi \in \Pi_{\widetilde{\mathbf{X}}}$. Replace $\widetilde{X}_N$ by $\pi \cdot \widetilde{\mathbf{X}}$ and rename the new constraints back to $\widetilde{\mathbf{X}}$. Evaluate $\pi^* = \pi^*_N(\widetilde{\mathbf{X}})$. If it is equal to the previously obtained value of π^*, then go to **Final Step**. Otherwise go to **Step 1**.

Step 3: The edge containing $\pi^* \cdot \widetilde{\mathbf{X}}$ is a simplex of smaller dimension than $N-1$. It is then an intersection of a finite number, say k, of such simplexes. They are defined by $N-1$-dimensional constraints, say, $\widetilde{\mathbf{X}}_{N-1i}$, $i = 1, \ldots, k$. Using recursion evaluate $X^*_{N-1i} = X^*_{N-1}(\widetilde{\mathbf{X}}_{N-1i})$, i.e. the extended solutions for the constraints from $N-1$ dimensional constraints corresponding to the simplexes interesection of which is the edge. If at least one of X^*_{N-1i} is different from $\pi^* \cdot \widetilde{\mathbf{X}}$, than go to **Step 4**. Otherwise go to **Step 5**.

Step 4: For each $i = 1, \ldots, k$, perform rejection sampling for π from Π reduced by the half-hyperplane corresponding to the simplex spanned by $\widetilde{\mathbf{X}}_{N-1i}$ and containing the previously searched N dimensional simplex to obtain a new constraint $\widetilde{X}_{iN} = \pi \cdot \widetilde{\mathbf{X}}$. Add this constraint to $\widetilde{\mathbf{X}}_{N-1i}$ to create $\widetilde{\mathbf{X}}_{Ni}$. Evaluate $\pi^*_i = \pi^*_N(\widetilde{\mathbf{X}}_{Ni})$. If there is $i = 1, \ldots, k$ such that $\pi^*_i \cdot \widetilde{\mathbf{X}}_{Ni} \neq X^*_{N-1i}$, then π^* and $\widetilde{\mathbf{X}}$ become π_i and $\widetilde{\mathbf{X}}_{Ni}$, respectively. Then go to **Step 1**. Otherwise repeat this step as many times until the conditions granting move to **Step 1** occur. (By Law of Large Numbers it has to happen after finite number attempts with probability one.)

Step 5: All X^*_{N-1i} are the same and equal to $\pi^* \cdot \widetilde{\mathbf{X}}$. It could be that $\pi^* \cdot \widetilde{\mathbf{X}}$ is the extended solution. Using rejection sampling, find an N dimensional simplex that contains $\pi^* \cdot \widetilde{\mathbf{X}}$ in its interior, say that it is defined by constraints $\mathbf{Y}$. Evaluate $\pi_{\mathbf{Y}} = \pi^*_N(\mathbf{Y})$, if it is in the interior of the simplex $\mathcal{L}^+_{\mathbf{Y}}$, then π^* and $\widetilde{\mathbf{X}}$ become $\pi_{\mathbf{Y}}$ and $\mathbf{Y}$, respectively, and then go to **Final Step**. Otherwise go to **Step 4**.

Final Step: Let $X^* = \pi^* \cdot \widetilde{\mathbf{X}}$. Return X^* as the value of $X^*_N(\mathbf{X})$, i.e. the extended simplex PD.

References

[1] Breiman, L. (1961), Optimal gambling systems for finite games. In *Proc. of the Fourth Berkeley Symposium on Math. Stat. and Probab.*, University of California Press, Berkeley, **1**, 65-78.

[2] Cover, T.M. (1984), An algorithm for maximizing expected log investment return. *IEEE Trans. Inform. Theory*, **30**, 369-373.

[3] Csiszár, I., Tusnády, G. (1984) Information geometry and alternating minimization procedures. *Statistics & Decisions*, **1**, 205-237.

[4] Finkelstein, M., Whitley, R. (1981) Optimal strategies for repeated games. *Adv. Appl. Prob.*, **13**, 415-428.

[5] Helmbold, D.P., Schapire, R.E., Singer, Y., Warmuth, M.K. (1998) On-line portfolio selection using multiplicative updates. *Math. Finance*, **8**, 325-347.

[6] Kelly, J. (1956) A new interpretation of information rate. *Bell System Tech. J.*, **35**, 917-926.

[7] LePage, R. (1994a) A representation theorem for linear functions of random variables with applications to option pricing. Technical Report, Center for Stochastic Processes, University of North Carolina.

[8] LePage, R. (1994b) Kelly pricing of contingent claims. Technical Report, Center for Stochastic Processes, University of North Carolina.

[9] LePage, R. (1998) Controlling a diffusion toward a large goal and the Kelly principle. Technical Report, Institute for Mathematics and its Applications, University of Minnesota.

[10] LePage, R., Podgórski, K. (1994) A nonlinear solution of inverse problem. Technical Report No.435, Center for Stochastic Processes, University of North Carolina.

Department of Probability and Statistics, Michigan State University, East Lansing, MI 48823

E-mail address: `rdl@entropy.msu.edu`

Department of Mathematical Sciences, Indiana University–Purdue University Indianapolis, Indianapolis, IN 46202

E-mail address: `kpodgorski@math.iupui.edu`

Institute of Mathematics, Technical University of Wrocław, Wrocław, Poland

E-mail address: `ryznar@im.pwr.wroc.pl`