Techniques for Generic Probabilistic Inversion

C. Du[1], D. Kurowicka[2], R.M.Cooke[3]

ABSTRACT

Probabilistic inversion problems are defined, existing algorithms for solving such problems are discussed and new algorithms based on iterative re-weighting of a sample are introduced. One of these is the well-known iterative proportional fitting whose properties were studied by Csiszar in 1975. A variant on this is shown to have fixed points minimizing an information functional, even if the problem is not feasible, and is shown to have only feasible fixed points if the problem is feasible. The algorithm is not shown to converge, but the relative information of successive iterates is shown to converge to zero. Applications to atmospheric dispersion and environmental transport are discussed.

KEYWORDS

Probabilistic inversion, iterative proportional fitting, maximal entropy, minimal information, I-projection.

1 Introduction

Probabilistic inversion problems can be formulated either from a measure theoretic or from a random variable viewpoint. The former may be more suggestive for theoreticians and the latter is more recognizable for practitioners.

[1]Department of Mathematics, Delft University of Technology, Delft, The Netherlands
[2]Department of Mathematics, Delft University of Technology, Delft, The Netherlands; d.kurowicka@ewi.tudelft.nl
[3]Department of Mathematics, Delft University of Technology, Delft, The Netherlands; r.m.cooke@ewi.tudelft.nl

1.1 Measure theoretic approach

We follow the approach of [3, 11]. Let $(M, \mathcal{B}(M), \lambda)$ and $(N, \mathcal{B}(N), \nu)$ be two Borel probability spaces, where M, N are compact non-empty subsets of $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively. Let $T : \mathbb{R}^m \mapsto \mathbb{R}^n$ be a continuous mapping. $\gamma = \lambda \circ T^{-1}$ is called the *push forward* of λ under T, and similarly λ is called the *pull back* of γ under T. In the following λ plays the role of a background measure. Measure μ on $(M, \mathcal{B}(M))$ is called an *inverse* of T at ν if μ is the pull back of ν under T; that is:

$$\forall B \in \mathcal{B}(N),\ \mu \circ T^{-1}(B) = \nu(B). \tag{1}$$

The problem can then be formulated as:

Definition 1 (Probabilistic inversion problem) *Given $(M, \mathcal{B}(M), \lambda)$ and $(N, \mathcal{B}(N), \nu)$, with $T : \mathbb{R}^m \mapsto \mathbb{R}^n$ continuous, find a measure μ absolutely continuous with respect to λ, $\mu \ll \lambda$, on $(M, \mathcal{B}(M))$ such that μ is an inverse of T at ν.*

Such problems may be infeasible, and if feasible may have many solutions [3]. Under certain conditions a measure μ can be found solving (1). If $\nu \ll \gamma$ the Radon-Nikodym derivative $g : N \mapsto \mathbb{R}$ exists, is non-negative, unique up to sets of γ-measure zero, and satisfies

$$\forall B \in \mathcal{B}(N), \nu(B) = \int_B g(y) d\gamma(y).$$

If in addition g is continuous, $g(y) > 0, y \in N$, and $\int_N g(y) \log g(y) d\gamma(y) < \infty$, then define $\widetilde{f} := g \circ T$, and define $\widetilde{\mu}$ as a new probability measure:

$$\forall A \in \mathcal{B}(\mathcal{M})\ \widetilde{\mu} := \int_{T^{-1}(B)} \widetilde{f} d\lambda(x). \tag{2}$$

It is easy to check that $\widetilde{\mu}$ is an inverse of T at ν:

$$\widetilde{\mu} \circ T^{-1}(B) = \int_{T^{-1}(B)} \widetilde{f} d\lambda(x) = \tag{3}$$

$$\int_{T^{-1}(B)} g \circ T(x) d\lambda(x) = \int_B g(y) d\gamma(y) = \nu(B). \tag{4}$$

It can be shown that the measure $\widetilde{\mu}$ is the unique measure satisfying equation (1) and minimizing the relative information with respect to λ in the class of measures satisfying equation (1)[3].

1.2 Random variable approach

In practical problems a reformulation of the problem in terms of random variables may be more convenient. Given a random vector $\mathbf{Y}$ taking values in $\mathbb{R}^M$ and a measurable function $G : \mathbb{R}^N \to \mathbb{R}^M$, find a random vector $\mathbf{X}$ such that $G(\mathbf{X}) \sim \mathbf{Y}$, where $\sim$ means that $G(\mathbf{X})$ and $\mathbf{Y}$ share the same distribution. If $G(\mathbf{X}) \in \{\mathbf{Y}|\mathbf{Y} \in \mathcal{C}\}$ where $\mathcal{C}$ is a subset of random vectors on $\mathbb{R}^M$, then $\mathbf{X}$ is called a probabilistic inverse of G at $\mathcal{C}$. $\mathbf{X}$ is sometimes termed the input to model G, and $\mathbf{Y}$ the output. G corresponds to the mapping T in the above paragraph. Note that T was assumed continuous, whereas G is only assumed to be measurable. If the problem is feasible it may have many solutions and we require a preferred solution; if it is infeasible we seek a random vector $\mathbf{X}$ for which $G(\mathbf{X})$ is "as close as possible" to $\mathbf{Y}$. Usually as a measure of closeness the relative information is used [16].

1.3 Probabilistic inversion for uncertainty analysis

Probabilistic inversion problems arise in quantifying uncertainty in physical models with expert judgement [2]. We wish to quantify the uncertainty on parameters $\mathbf{X}$ of some model using expert judgment, but the parameters do not possess a clear physical meaning and are not associated with physical measurements with which experts are familiar. Often the models are derived under assumptions to which the experts do not subscribe. We must then find the observable quantities $\mathbf{Y}$ functionally related with $\mathbf{X}$ that can be assessed by experts. Extracting uncertainties of $\mathbf{X}$ from uncertainties of $\mathbf{Y}$ specified by experts is clearly an inverse problem (see examples in section 5).

In practical applications the random vector $\mathbf{Y}$ is characterized in terms of some percentiles or quantiles of the marginal distributions $Y_1, ..., Y_M$. In this case, we seek a random vector $\mathbf{X}$ such that $G(\mathbf{X})$ satisfies quantile constraints imposed on $Y_1, ..., Y_M$. There may be other constraints. These constraints may reflect mathematical desiderata, as when we require independence between variables in section 5. Physical considerations may also impose constraints on $\mathbf{X}$[4]. A few algorithms for solving such problems are available in literature namely: conditional sampling, PARFUM (PARameter Fitting for Uncertain Models)[15], Hora and Young algorithm [1] and PREJUDICE [2, 12]. We summarize existing approaches to probabilistic inversion in section 2.

In this paper we study iterative algorithms for numerically solving probabilistic inversion problems. These methods don't require model inversion they are based on sample re-weighting techniques and are promising because they do not require special knowledge about the problem at hand, or complicated heuristic steering on the part of the user.

[4]In some cases probabilistic inversion problems may have trivial solutions, e.g. if the $\mathbf{X}$ makes the coordinates of $\mathbf{Y}$ completely rank correlated. Such solutions may be rejected on physical grounds; hence physical constraints may stipulate the support of $\mathbf{X}$. Other physical constraints are discussed in the example in section 6.

Moreover, operations on the sample are performed one-at-a-time, so the entire sample need not be held in memory. This means that there is virtually no problem size limitation.

1.4 Organization of the paper

After reviewing existintg approaches, we discuss two iterative algorithms. First is known in the literature as Iterative Proportional Fitting (IPF) [10]. IPF finds a joint distribution satisfying marginal constraints by successively I-projecting an initial distribution on the set of distributions satisfying each marginal constraint. The I-projection of a probability measure p on a set of measures Q is $argmin_{q\in Q} I(q|p)$, where $I(q|p)$ is the relative information of q with respect to p. IPF need not converge, but if it converges, it converges to a solution which is minimally informative with respect to the starting distribution [7]. A variation on this is an iterative version of the PARFUM algorithm. We show that this algorithm has fixed points minimizing an information functional, even if the problem is not feasible, and that it has only feasible fixed points if the problem is feasible. The algorithm is not shown to converge, but the relative information of successive iterates is shown to converge to zero.

In section 4 we discuss a sample re-weighting iterative approach to probabilistic inversion. In section 5 we illustrate the IPF and PARFUM algorithms with an example from atmospheric dispersion modelling [2] and transfer coefficients in chicken processing line. Iterative algorithms can easily be adopted to satisfy joint as well as marginal constraints on $Y_1, ..., Y_M$. We illustrate this by imposing (approximate) convolution constraints in section 6. A final section gathers conclusions.

2 Existing algorithms for probabilistic inversion

Conditional sampling Let $\mathbf{Y}$ consist of only one variable Y. A simple conditional sampling technique can be used, based on the following result [2].

Proposition 1 *Let X and Y be independent random variables with range $\{1, ..., n\}$. $P(X = i) > 0$, $P(Y = i) > 0$, $i = 1, ..., n$. Let $X_{X=Y}$ denote a random variable with distribution*

$$P(X_{X=Y} = i) = P(X = i|X = Y).$$

Then $X_{X=Y} \sim Y$ if and only if X is uniformly distributed.

Proof. Put $p_i = P(X = i)$, $q_i = P(Y = i)$, $i = 1, \ldots, n$. Then

$$P(X_{X=Y} = i) = \frac{p_i q_i}{\sum p_j q_j}.$$

For all $i = 1, ..., n$, $p_i q_i / \sum p_j q_j = q_i$ if and only if $p_i = \sum p_j q_j$, that is, p_i does not depend on i. $\square$

Consider a random vector $\mathbf{X} = [X_1, X_2, \ldots, X_n]$ and function $G : \mathbb{R}^n \to \mathbb{R}$ such that $G(\mathbf{X})$ and Y are concentrated on an interval I with invertible cumulative distribution functions F_G and F_Y, respectively. Let G^* be a discretized version of $G(\mathbf{X})$ such that the cumulative distribution function F_{G^*} is concentrated on $I_d = \{1/k, 2/k, \ldots, 1\}$. In other words, define G^* such that

$$G^*(\mathbf{X}) = \sum_{j=1}^{k} F_G^{-1}(\frac{j}{k}) \mathbb{I}_{(\frac{j-1}{k}, \frac{j}{k}]}(F_G(G(\mathbf{X}))), \tag{5}$$

where $\mathbb{I}_A$ denotes the indicator function of a set A.

$F_{G^*}(G(\mathbf{X}))$ is uniformly distributed on I_d. $F_{G^*}(Y)$ is concentrated on I_d but is not necessarily uniform. The following procedure is used to find the conditional distribution of $G(\mathbf{X})$ which approximates the distribution of Y:

1. sample $\mathbf{X}$ and sample Y independently of $\mathbf{X}$,
2. if $F_{G^*}(G(\mathbf{X})) = F_{G^*}(Y)$ then retain the sample, otherwise discard the sample
3. repeat (1) and (2).

Since $F_{G^*}(G(\mathbf{X}))$ is uniformly distributed in the unconditional sample, the above proposition shows that in the conditional sample, $F_{G^*}(G(\mathbf{X})) \sim F_{G^*}(Y)$. The conditional distribution, P_{cond} of $G(\mathbf{X})$ approximates the distribution of Y in the sense that

$$P_{cond}(G(\mathbf{X}) \in \{F_G^{-1}(i/k), F_G^{-1}((i+1)/k)\}) = P(Y \in \{F_G^{-1}(i/k), F_G^{-1}((i+1)/k)\}). \tag{6}$$

The advantage of this technique is its simplicity. Its disadvantage is that it works only for one output variable.

PARFUM Let $\mathbf{Y} = [Y_1, Y_2, \ldots, Y_M]$ be a random vector with densities $[f_1, ..., f_M]$ and let $G_m : \mathbb{R}^n \to \mathbb{R}, m = 1, 2, \ldots, M$ be measurable functions. The PARFUM algorithm can be described in the following steps [15]:

1. Choose a finite set $\chi \subset \mathbb{R}^n$
2. Define the conditional mass function Q_m of Y_m on the image $G_m(\chi)$ of χ under G_m, where $\mathbf{x} \in \chi$:
$$Q_m(G_m(\mathbf{x})) = \frac{f_m(G_m(\mathbf{x}))}{\sum_{G_m(\mathbf{z}) \in \chi} f_m(G_m(\mathbf{z}))}.$$

3. Define the minimally informative distribution on χ whose push-forward distribution P_m on $G_m(\chi)$ agrees with Q_m, that is, for $\mathbf{x} \in \chi$

$$P_m(\mathbf{x}) \quad = \quad \frac{Q_m(G_m(\mathbf{x}))}{\#\{\mathbf{z} \in \chi | G_m(\mathbf{z}) = G_m(\mathbf{x})\}},$$

where # means number of points.

4. Find a distribution P on χ which minimizes the relative information $\sum_{m=1}^{M} I(P_m|P)$, where

$$I(P_m|P) \quad = \quad \sum_{\mathbf{X} \in \chi} P_m(\mathbf{x}) \ln \left(\frac{P_m(\mathbf{x})}{P(\mathbf{x})} \right).$$

Let

$$\mathcal{S}^K = \{r \in \mathbb{R}^K | r_k \geq 0, \sum_{k=1}^{K} r_k = 1\}. \tag{7}$$

It is not difficult to show [15]:

Proposition 2 *Let $P_m \in \mathcal{S}^K$, $m = 1, \ldots, M$. Then*

$$\min_{P \in \mathcal{S}^K} \sum_{m=1}^{M} I(P_m|P) = \sum_{m=1}^{M} I(P_m|P^*)$$

if and only if $P^* = (1/M) \sum_{m=1}^{M} P_m$.

The advantage of this method is that it is always feasible and easily implemented. One disadvantage is that the conditional distributions Q_m might be different to those of Y_m, but this may be steered by the right choice of χ. More serious is the fact that the push forward of P need not have marginal distributions that agree with the Y_m. This also can be influenced by steering, but is more difficult.

Hora-Young and PREJUDICE Algorithms Instead of basing the fitting on the conditional measure Q_m, which may be different than F_m, the Hora-Young method constrains the choice of P to the set $\mathbb{P}$ whose margins for $G_m(\chi)$ satisfy the quantile constraints

$$P\{G_m(\chi) \in [y_{mk-1}, y_{mk}]\} \quad = \quad F_m(y_{mk}) - F_m(y_{mk-1}), k = 2, \ldots, K, \tag{8}$$

where $y_{m1}, \ldots, y_{mK}$ are in the range of $Y_m, m = 1, \ldots, n$, $P(G_m(\chi))$ is the push-forward measure on the range of Y_m induced by the probability measure P on X and F_m is the cdf of Y_m. The disadvantage of this method is that it is more often infeasible; that is,

for a given choice of χ there is no measure P satisfying the constraints. If it is feasible it performs better than PARFUM.

The algorithm PREJUDICE is an elaboration of this algorithm. Using duality theory for constrained optimization, it performs model inversions thereby augmenting the set χ in a way which optimally reduces infeasibility. Performing model inversions can be very expensive. Although PREJUDICE represents the most sophisticated method to date for probabilistic inversion, the use of model inversions makes it unsuitable for a generic uncertainty analysis system. For some models the inversion step may simply be too difficult. For more information about this method we refer to [3].

3 Iterative algorithms

In this section we introduce two iterative algorithms applied to solve the probabilistic problem in the next section. First we introduce necessary notation, definitions, simple facts and present IPF and PARFUM algorithms for the discrete distributions. We formulate the problem and show results only for 2-dimensional case (M=2). We indicate which results can be generalized.

Let $p_{\cdot,j} = \sum_{i=1}^{K} p_{ij},\ j = 1, \ldots, K$ and $p_{i,\cdot} = \sum_{j=1}^{K} p_{ij},\ i = 1, \ldots, K$, and

$$\begin{aligned} \mathcal{S}^{K\times K} &= \{p \in \mathbb{R}^{K\times K} | p_{ij} \geq 0, \sum_{i,j=1}^{K} p_{ij} = 1\}, \\ \mathcal{S}^{*K\times K} &= \{p \in \mathcal{S}^{K\times K} | p_{i,\cdot} > 0, p_{\cdot,j} > 0, i, j = 1, \ldots, K\}. \end{aligned}$$

be respectively the set of probability vectors in $\mathbb{R}^{K\times K}$, and the set of probability vectors in $\mathbb{R}^{K\times K}$ with non-degenerate margins. Our problem can be now formulated as follows: *for given $p \in \mathcal{S}^{*K\times K}$ and $a, b \in \mathcal{S}^{K}$ find a distribution $q \in \mathcal{S}^{*K\times K}$ such that*

$$I(q|p) = \sum_{i,j=1}^{K} q_{ij} log \frac{q_{ij}}{p_{ij}} \quad \textit{is minimum};$$

subject to the following constraints

1. $q_{\cdot,j} = b_j,\ j = 1, \ldots, K$;
2. $q_{i,\cdot} = a_i,\ i = 1, \ldots, K$.

Let

$$\begin{aligned} \mathcal{Q}_1 &= \{q \in \mathcal{S}^{*K\times K} | q_{i,\cdot} = a_i,\ i = 1, \ldots, K\}, \\ \mathcal{Q}_2 &= \{q \in \mathcal{S}^{*K\times K} | q_{\cdot,j} = b_j,\ j = 1, \ldots, K\}. \end{aligned}$$

We shall assume throughout that $a_i > 0, b_j > 0;\ i, j = 1, ..., K$. We define as in [7] the I-projection of p onto the set of distributions with one fixed margin as a closest distribution, in sense of relative information, with this margin fixed.

Definition 2 *Let $\mathcal{Q}_m$, $m = 1, 2$ be as defined above. An* **I-projection $p^{\mathcal{Q}_m}$ of p on $\mathcal{Q}_m$** *is*

$$p^{\mathcal{Q}_m} = argmin_{q \in \mathcal{Q}_m} I(q|p).$$

Since $\mathcal{Q}_m$ is convex and $p \in \mathcal{S}^{*K \times K}$, it follows from [7] theorem 2.2 that $p^{\mathcal{Q}_m}$ is unique and is of the form

$$p_{ij}^{\mathcal{Q}_1} = p_{ij} \frac{a_i}{p_{i,\cdot}}, \quad \left(p_{ij}^{\mathcal{Q}_2} = p_{ij} \frac{b_j}{p_{\cdot,j}} \right)$$

for $i, j = 1, \ldots K$, respectively.

3.1 IPF

The IPF algorithm projects a starting measure onto the set with fixed first margins ($\mathcal{Q}_1$), then projects this onto the set with fixed second margins ($\mathcal{Q}_2$), then projects this again onto the set $\mathcal{Q}_1$ and so on. Hence if we have arrived at vector p by projecting onto $\mathcal{Q}_1$, the next iteration is

$$p' \ = \ p^{\mathcal{Q}_2}.$$

This algorithm was first used to estimate cell probabilities in a contingency table subject to certain marginal constraints. It is easy to see that if p satisfies the constraints, then p is a fixed point of this algorithm.

The convergence of the IPF algorithm has been studied by many authors see [5, 23, 19, 20, 21] and Csiszar [7]. Csiszar shows that if the IPF algorithm converges, then it converges to the I-projection of the starting probability vector on the set of probability vectors satisfying the constraints. He further showed that starting with a probability vector p, IPF converges if there is a vector r satisfying the constraints and having zeros in those cells where p is zero

$$p_{ij} = 0 \Rightarrow r_{ij} = 0, i, j = 1, \ldots, K. \tag{9}$$

For the two dimensional case the IPF algorithm is equivalent to an alternating minimization procedure studied in [8]. It is proven there that if sequences $\{P_n\}$ and $\{Q_n\}$ from $\mathcal{Q}_1$ and $\mathcal{Q}_2$, respectively, are obtained by alternating minimization of $I(P_n|Q_n)$ with respect to P_n resp. Q_n, then $I(P_n|Q_n)$ converges to the infimum of $I(P|Q)$ on $\overline{\mathcal{Q}_1} \times \mathcal{Q}_2$, where $\overline{\mathcal{Q}_1}$ is a set of all $P \in \mathcal{Q}_1$ such that $I(P|Q_n) < \infty$ for some n. Moreover, the convergence of the sequences $\{P_n\}$ and $\{Q_n\}$ is proven. Csiszar and Tusnady [8] proved this result for a general case where the sets $\mathcal{Q}_1$ and $\mathcal{Q}_2$ are convex sets of finite measures and the function that is minimized in alternating minimization procedure is an extended real valued function.

3.2 Iterative PARFUM

The iterative PARFUM algorithm is an algorithm based on the result in proposition 10. In contrast to IPF the iterative PARFUM algorithm projects starting distribution on $\mathcal{Q}_1$ *and* on $\mathcal{Q}_2$ and takes the average of these two distributions. If we have arrived at the probability vector p, we define the next iteration p' as:

$$p' = \frac{p^{\mathcal{Q}_1} + p^{\mathcal{Q}_2}}{2}. \tag{10}$$

Each measure $p^{\mathcal{Q}_i}$ adapts the measure p to have the ith margin fixed. We see that p' is the probability vector which is 'closest' in the sense of relative information to both the measures $p^{\mathcal{Q}_i}$ (proposition 10). If p satisfies the constraints, then $p' = p$. In other words, the iterative PARFUM transformation has fixed points at all feasible probability vectors. The following theorem shows that the relative information functional of the iterates of this algorithm always converge, even when the constraints cannot be satisfied.

Theorem 1 *Let $\mathcal{Q}_1, \mathcal{Q}_2$ be closed convex subsets of $S^{K\times K}$. For $p^{(j)} \in S^{*K\times K}$, let $q^{(j),m}$ be the I-projection of $p^{(j)}$ on $\mathcal{Q}_m$, $m = 1, 2$. Let $p^{(j+1)} = \frac{q^{(j),1}+q^{(j),2}}{2}$. Then $I(p^{(j+1)}|p^{(j)}) \to 0$ as $j \to \infty$.*

Proof. Since $\mathcal{Q}_m$ is closed and convex, $q^{(j),m}$ exists and is unique ([7], theorem 2.1). Define $F(p^{(j)}) = I(q^{(j),1}|p^{(j)}) + I(q^{(j),2}|p^{(j)}) \geq 0$. By proposition 2 and the fact that $q^{(j+1),m}$ is the I-projection of $p^{(j+1)}$ on $\mathcal{Q}_m$, we have

$$\begin{aligned} F(p^{(j)}) &= I(q^{(j),1}|p^{(j)}) + I(q^{(j),2}|p^{(j)}) \geq I(q^{(j),1}|p^{(j+1)}) + I(q^{(j),2}|p^{(j+1)}) && (11)\\ &\geq I(q^{(j+1),1}|p^{(j+1)}) + I(q^{(j+1),2}|p^{(j+1)}) = F(p^{(j+1)}). && (12)\end{aligned}$$

Equality holds if and only if $p^{(j)} = p^{(j+1)}$. Thus, $F(p^{(j)})$ is decreasing in j and converges. To show that $I(p^{(j+1)}|p^{(j)})$ converges to zero, pick $\varepsilon > 0$ and $j \in \mathbf{N}$ such that $F(p^{(j)}) - F(p^{(j+1)}) < \varepsilon$. Then $\sum_{m=1}^{2} I(q^{(j),m}|p^{(j)}) - \sum_{m=1}^{2} I(q^{(j),m}|p^{(j+1)}) < \varepsilon$. Writing this inequality element-wise:

$$\sum_{m=1}^{2}\sum_{i,k=1}^{K} q_{ik}^{(j),m}(\ln(q_{ik}^{(j),m}/p_{ik}^{(j)}) - \ln(q_{ik}^{(j),m}/p_{ik}^{(j+1)})) < \varepsilon.$$

Reversing the order of summation and substituting

$$\sum_{m=1}^{2} q_{ik}^{(j),m} = 2p_{ik}^{(j+1)},$$

we find

$$2I(p^{(j+1)}|p^{(j)}) < \varepsilon. \quad \square$$

The set $\mathcal{Q}_m$ may be regarded as the set of vectors satisfying the m-th constraint; The first term in (12) can be written as

$$\begin{aligned} J(q^{(j),1}, q^{(j),2}) &= \sum_{m=1}^{2} I(q^{(j),m}|p^{(j+1)}) \\ &= \sum_{m=1}^{2} I\left(q^{(j),m} \middle| \frac{\sum_{m=1}^{2} q^{(j),m}}{2}\right). \end{aligned}$$

The iterative PARFUM algorithm may be seen as minimizing the function J, and the value of J may be taken as a measure of 'how infeasible' the problem is.
If $\mathcal{Q}_1 \cap \mathcal{Q}_2 = \emptyset$, then the problem is infeasible. If $\mathcal{Q}_1 \cap \mathcal{Q}_2 = \mathcal{Q} \neq \emptyset$ then the algorithm is observed to converge to an element of $\mathcal{Q}$ and $J(q^{(j),1}, q^{(j),2}) \to 0$ (see Theorem 3).
A sufficient condition for feasibility of the problem is given in the following:

Theorem 2 *Let p be a fixed point of the PARFUM algorithm with $p_{ij} > 0$ for all $i, j = 1, \ldots, K$ then $p_{i,\cdot} = a_i, i = 1, \ldots, K$ and $p_{\cdot,j} = b_j, j = 1, \ldots, K$.*

Proof: Since p is a fixed point we have :

$$2p_{ij} = p_{ij}(\frac{a_i}{p_{i,\cdot}} + \frac{b_j}{p_{\cdot,j}}). \tag{13}$$

Since $p_{ij} > 0$, we may divide both sides by p_{ij} and obtain:

$$(2 - \frac{a_i}{p_{i,\cdot}})p_{\cdot,j} = b_j.$$

Summing both sides over j we find

$$(2 - \frac{a_i}{p_{i,\cdot}}) = 1 = \frac{b_j}{p_{\cdot,j}}.$$

Similarly we can show that $\frac{a_i}{p_{i,\cdot}} = 1$. □

It is easy to see that a fixed point of PARFUM can satisfy $p_{ij} > 0$ only if this also holds for the starting point. We prove that if the problem is feasible than the fixed point of the PARFUM algorithm is one of the solutions, that is, if $\mathcal{Q}_1 \cap \mathcal{Q}_2 = \mathcal{Q} \neq \emptyset$ then the PARFUM algorithm converges to an element of $\mathcal{Q}$.

Theorem 3 *Let $\mathcal{Q}_1 \cap \mathcal{Q}_2 = \mathcal{Q} \neq \emptyset$. If p be is a fixed point of the PARFUM algorithm then $p \in \mathcal{Q}$.*

Proof. In Appendix.

3.3 Generalizations of IPF and PARFUM

IPF and PARFUM algorithms are both easy to apply. For easier presentation we have assumed that both margins are from $\mathcal{S}^K$. This can be trivially extended. Moreover the results presented in previous subsection can be generalized to higher dimensions ($M > 2$) [13]. Iterative algorithms can easily be adopted to satisfy joint as well as marginal constraints [18].

Generalization of the IPF to the continuous case have been introduced in [9, 17]. However, the convergence of the IPF in the continuous case under certain regularity condition was proven much later in [14]. Proposition 10 and theorem 1 can be easily generalized to the continuous case. For proposition 10 this follows immediately from definition of I-projection and the fact that $I(p|q) \geq 0$ [16]. Using proposition 10 the proof of theorem 1 goes through for continuous case [13].

In case of feasible problems we generally advocate to use IPF algorithm as it converges to the I-projection of starting distribution onto the set of feasible distribution. However, in case of infeasibility for high dimensional problems, IPF must be terminated at some point, and the resulting distribution may be concentrated on very few samples and may distribute the lack of fit very unevenly, with some variables being wildly wrong. It is not clear when the algorithm should be terminated and which distribution should be taken as a result. The iterative PARFUM algorithm offers some advantages with respect to IFP for infeasible cases, and the practitioner is likely to feel more comfortable with PARFUM (see example in section 5).

4 Sample Re-weighting

In this section we show how sample re-weighting combined with iterative algorithms presented in section 3 can solve probabilistic inversion problems. This yields generic methods for probabilistic inversion which do not require model inversion. The idea of re-weighting a sample to perform probabilistic inversion can be sketched roughly as follows. Starting with a random vector $\mathbf{X}$, we generate a large sample from $\mathbf{X} : [X_1, \ldots, X_n]$, $Y_1 = G_1(\mathbf{X})$,..., $Y_M = G_M(\mathbf{X})$. Let the $i-th$ sample be denoted $s_i \in \mathbb{R}^{n+M}$. Obviously each sample $s_i = (x_1^{(i)}, \ldots, x_n^{(i)}, y_1^{(i)}, \ldots, y_M^{(i)})$ has the same probability of occurring. If N samples have been drawn, then the sampling distribution assigning $p(s_i) = 1/N$ approximates the original distribution from which the samples were drawn.
The idea is now to change the probabilities $p(s_i)$ so as to ensure that the distributions of $Y_1, \ldots, Y_M$ satisfy the user specified quantile constraints. No model inversions are performed, that is, we do not invert the function G. To achieve reasonable results, the sample size N may have to be very large. This places strong restrictions on the type of methods which can be implemented in a generic uncertainty analysis program. Standard opti-

mization approaches, for example, were found to breakdown on large problems involving infeasibility.

4.1 Notation

Since the variables $X_1, \ldots, X_n$ play no role in choosing the weights $p(s_i)$, we may leave them out of the problem description, hence $s_i = (y_1^{(i)}, \ldots, y_M^{(i)})$. We denote $p(s_i)$ as p_i, and introduce:

Data Matrix: M variables all with N samples are grouped in matrix $\mathcal{Y}$. Hence $\mathcal{Y} = [y_{im}]$, $i = 1, ..., N,\ m = 1, 2, \ldots, M$ and $y_{im} = y_m^{(i)}$.

Inter-quantile vector: We consider a vector $q = [q_1, q_2, \ldots, q_K]$ of lengths of inter-quantile, or inter-percentile intervals. For instance for $K = 4$, if we consider $5\%, 50\%$ and 95% quantiles then $q = [0.05, 0.45, 0.45, 0.05]$;

Constraints: A matrix $R = [r_{jm}], j = 1, \ldots, K-1;\ m = 1, \ldots, M$, contains the percentiles that we want to impose: r_{jm} is the number for which $P\{Y_m \leq r_{jm}\} = q_1 + \ldots + q_j$. Thus we want the probability vector $[p_1, \ldots, p_N]$ to satisfy the **constraint set** $\mathcal{C}$:

$$\sum_{i=1}^{N} p_i \mathbb{I}_{J_{k,m}}(y_{im}) = q_k \quad \text{for all} \quad k = 1, 2, \ldots, K \text{ and all } m = 1, 2, \ldots, M, \quad (14)$$

where $\mathbb{I}_A$ denotes the indicator function of a set A and the interval $J_{k,m}$ is defined as:

$$J_{1m} = (-\infty, r_{1,m}],\ \ J_{km} = (r_{k-1,m}, r_{km}],\ \ k = 2, \ldots, K-1,\ \ J_{K,m} = (r_{Km}, \infty)$$

for all $m = 1, 2, \ldots, M$.

We note that these constraints do not really say that, e.g. r_{m1} is the $q_1 - th$ quantile of Y_m. Indeed, the $q_1 - th$ quantile is defined as the *least* number a satisfying the constraint $P(Y_m < a) = q_1$. If Y_m is concentrated on a sparse number of points, then there may be many values satisfying the above constraints.

Partitions of samples To each variable Y_m we associate a partition of the samples

$$\mathcal{A}^m = \{A_k^m\}_{k=1}^K;\ \ m = 1, \ldots, M,$$

where

$$A_k^m = \text{set of samples for which } Y_m \text{ falls in inter-quantile interval k.} \quad (15)$$

The output of a re-weighting scheme is a vector $p = [p_1, p_2, \ldots, p_N]$ of weights. After re-weighting the samples with these weights, the constraints $\mathcal{C}$ are satisfied 'as nearly as possible'.

4.2 IPF and PAFRFUM for sample re-weighting probabilistic inversion

Iterative algorithms involve successively updating a probability vector $[p_1, \ldots, p_N]$ so as to approach a solution satisfying constraints $\mathcal{C}$, or satisfying these constraints as closely as possible. Note that samples s_i and s_j which fall in the same interquantile intervals, for $Y_1, \ldots, Y_M$ will be treated in exactly the same way. Hence the weights for these samples must be the same. Starting with the uniform vector $p_i = 1/N$ for s_i we may redistribute all samples over K^M cells and obtain K^M dimensional discrete distribution $[p_{i_1,i_2,\ldots,i_M}]$ in the following way

$$p_{i_1,i_2\ldots,i_M} = \frac{1}{N}\sum_{n=1}^{N} \mathbb{I}_{A^1_{i_1}\cap\ldots\cap A^M_{i_M}}(s_n),$$

where $i_m \in \{1, 2, \ldots, K\}, m = 1, 2, \ldots, M$. $p_{i_1,i_2\ldots,i_M}$ must be changed in the minimum information fashion so that all one dimensional marginal distributions are $[q_1, q_2, \ldots, q_K]$.

Taking $p_{i_1,i_2,\ldots,i_M}$ as a starting distribution, the IPF or PARFUM algorithms discussed in section 3 may be applied to change this distribution so as to satisfy the quantile constraints if these are feasible. Otherwise, a fixed point of PARFUM minimizes infeasibility in the sense of (13). The probability mass assigned to the $(i_1, i_2, \ldots, i_M)$ cell of the stationary distribution must be distributed uniformly over weights corresponding to samples in $A^1_{i_1} \cap A^2_{i_2} \cap \ldots \cap A^M_{i_M}$.

5 Examples

We first illustrate PARFUM and IPF with simple example involving dispersion coefficients from [1], which is also extensively used in [3] to explain steps of probabilistic inversion technique involving model inversion. We then presents results from a recent study involving a chicken processing line.

5.1 Dispersion coefficients

The lateral spread σ of an airborne plume dispersing in the downwind direction x is modelled as power law function of distance x from the source of release:

$$\sigma(x) \;=\; Ax^B, \tag{16}$$

where the coefficients A and B depend on the stability of atmosphere at the time of release. Of course there may be more variables such as wind, surface roughness etc. It is however recognized that model (16) captures the uncertainty associated with plume spread well enough. The joint distribution of A and B must be found in order to find the

uncertainty on σ for any downwind distance x. Parameters A and B are not observable but the plume spread σ at any given downwind distance is an observable quantity. Eight experts were therefore asked to quantify their uncertainty on $\sigma(x_i)$ for downwind distance $x_1, \ldots, x_5$. The experts were asked to express their uncertainty in terms of $5\%, 50\%$ and 95% quantile points on plum spread at downwind distances 500m, 1km, 3km, 10km and 30km. The performance based weighted combinations of the experts distributions led to results presented in Table 1 (for more information about structured expert judgement for this example see [4]) .

Quantile	$\sigma(500m)$	$\sigma(1km)$	$\sigma(3km)$	$\sigma(10km)$	$\sigma(30km)$
5%	33	64.8	175	448	1100
50%	94.9	172	446	1220	2820
95%	195	346	1040	3370	8250

TABLE 1: Assessments of σ.

One can see in [3] how the distributions of A and B are obtained using the PREDUJCE method. To simplify presentation we consider here only two downwind distances, that is 500m and 1km and show how this problem can be solved with iterative procedures PARFUM and IPF. Let us take A to be uniform on the interval (0,1.5) and B uniform on the interval (0.5,1.3). Now by sampling N values of A and B, $(a_i, b_i), i = 1...N$, we apply (16) to each sample for both downwind distances, say $y_{1i} = a_i 500^{b_i}$ and $y_{2i} = a_i 1000^{b_i}$. We assign each sample to a cell in a 4×4 matrix according to the interquantile intervals of Y_1, Y_2 catching the sample. Thus if $y_{1i} \leq 33$ and $64.8 < y_{2i} \leq 172$ then Y_1 falls in the first interquantile interval and Y_2 in the second. Let c be the 4×4 matrix obtained in this way.

The distribution $p^{(0)}$ is obtained from the matrix c by dividing each cell of c by the number of samples $N = 10000$: $p^{(0)} = \frac{1}{N}c$. Hence we get

$$p^{(0)} = \begin{bmatrix} 0.1966 & 0.0006 & 0 & 0 \\ 0.0407 & 0.1642 & 0.0050 & 0 \\ 0 & 0.0094 & 0.1196 & 0.0155 \\ 0 & 0 & 0.0008 & 0.4476 \end{bmatrix}.$$

It may be noticed that the cells $(4,1), (4,2)$ are empty. Indeed, these cells are physically impossible, since a plume cannot be wider at 500m than at 1000m.

After 200 iterations we get the following results using PARFUM and IPF respectively:

$$p_{PARFUM} = \begin{bmatrix} 0.0439 & 0.0061 & 0 & 0 \\ 0.0061 & 0.4229 & 0.0210 & 0 \\ 0 & 0.0210 & 0.4233 & 0.0057 \\ 0 & 0 & 0.0057 & 0.0443 \end{bmatrix},$$

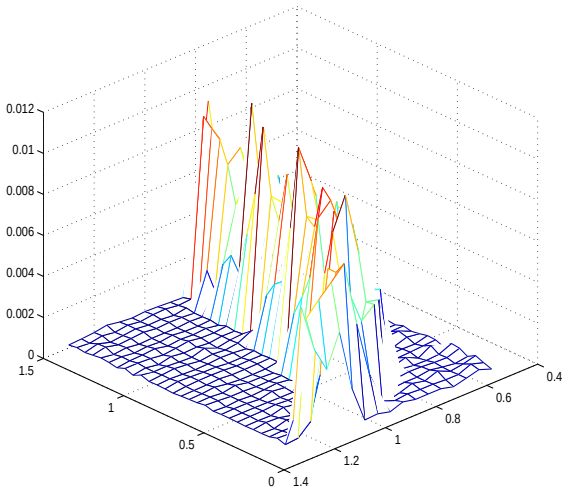

Figure 1: A joint distribution of (A, B) after applying PARFUM algorithm.

$$p_{IPF} = \begin{bmatrix} 0.0461 & 0.0039 & 0 & 0 \\ 0.0039 & 0.4253 & 0.0208 & 0 \\ 0 & 0.0208 & 0.4270 & 0.0022 \\ 0 & 0 & 0.0022 & 0.0478 \end{bmatrix}.$$

One can verify that both marginal distributions for p_{PARFUM} and p_{IPF} are equal to $q = [0.05, 0.45, 0.45, 0.05]$. The relative information with respect to $p^{(0)}$ of p_{PARFUM} and p_{IPF}, respectively are 0.8219 and 0.8166. The vector of samples weights can be obtained from p_{PARFUM} or p_{IPF} by distributing probabilities in each cell of p_{PARFUM} and p_{IPF} uniformly over all samples that fall to this cell. 10000 samples is not unrealistic and iterative algorithms are quite fast. There is, however, no guidance how the initial distributions of A and B have to be chosen.

Initially (A, B) had a joint distribution uniform on $[0, 1.5] \times [0.5, 1.3]$. Figure 1 shows how the joint distribution of (A, B) has changed after applying PARFUM algorithm. One can see that A and B are now highly negatively correlated.

The marginal distributions of $\sigma(500)$ and $\sigma(1000)$ are also changed to match expert's quantile information. Figure 2 shows the difference between original marginal distributions for $\sigma(500)$ and $\sigma(1000)$ and marginal distributions after applying PARFUM algorithm.

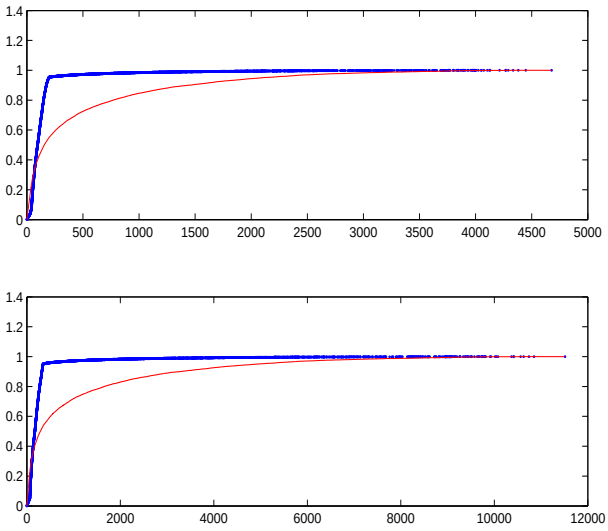

Figure 2: Marginal distributions of $\sigma(500)$ (upper graph) and $\sigma(1000)$ (lower graph) before (more diffuse) and after (less diffuse) applying PARFUM algorithm.

5.2 Chicken processing line

A model of a chicken processing line was recently made in order to identify the significant pathways for re-infection with Campylobacter bacteria. The details will be published separately [22]. Suffice to say that for each of the phases *scalding, evisceration, defeathering*, 5%, 50% and 95% quantiles were determined for six predicted variables using structured expert judgment. Probabilistic inversion was used to pull this distribution back onto the parameter space of the model. The parameters in this case are transfer coefficients. Table in figure 3 summarizes the results with IPF and with PARFUM. In each phase, 50,000 samples were generated from a diffuse distribution over the transfer coefficients, and for each sample, values of the predicted variables were computed. In each case 50 iterations were performed, beyond which no significant improvements were observed.

In all cases there is discrepancy between the percentile aimed for and the percentile realized by the probabilistic inversion. This indicates that the inversion problem for the given samples is not feasible. Of course one option would be to take more samples; however after some effort in this direction, it was concluded that the expert assessments were not really compliant with the model to be fitted - a situation which arises not infrequently in practice. Table 3 thus illustrates the behavior of these algorithms on infeasible problems. Note that in each case PARFUM has *lower* relative information with respect to the starting distribution, i.e. the distribution assigning each sample the weight $1/50,000$. Indeed, IPF converges to the I-projection of the starting measure on the set of feasible measures, but when the problem is infeasible the set of feasible measures is empty and very little is known about IPF's behavior. It should be noted that IPF is stopped after cycling through all six

prediction variable	percentile	EVISCERATION PARFUM	EVISCERATION IPF	SCALDING PARFUM	SCALDING IPF	DEFEATHERING PARFUM	DEFEATHERING IPF
	5%	**0.050**	**0.050**	**0.060**	**0.063**	**0.053**	**0.014**
1	**50%**	**0.497**	**0.500**	**0.541**	**0.618**	**0.424**	**0.175**
	95%	**0.950**	**0.950**	**0.957**	**0.962**	**0.871**	**0.719**
	5%	0.050	0.050	0.050	0.062	0.370	0.357
2	*50%*	0.497	0.500	0.539	0.618	0.736	0.921
	95%	0.950	0.950	0.957	0.962	0.972	0.993
	5%	**0.090**	**0.161**	**0.043**	**0.050**	**0.053**	**0.014**
3	**50%**	**0.557**	**0.559**	**0.464**	**0.498**	**0.424**	**0.175**
	95%	**0.956**	**0.956**	**0.940**	**0.954**	**0.871**	**0.719**
	5%	0.032	0.040	0.033	0.034	0.030	0.033
4	*50%*	0.404	0.403	0.386	0.357	0.256	0.151
	95%	0.851	0.853	0.857	0.912	0.543	0.654
	5%	**0.121**	**0.116**	**0.137**	**0.068**	**0.184**	**0.158**
5	**50%**	**0.532**	**0.540**	**0.572**	**0.531**	**0.501**	**0.702**
	95%	**0.942**	**0.884**	**0.959**	**0.943**	**0.839**	**0.969**
	5%	0.065	0.050	0.106	0.050	0.133	0.050
6	*50%*	0.647	0.500	0.628	0.500	0.606	0.500
	95%	0.966	0.950	0.966	0.950	0.972	0.950
Relative information		**6.0769949**	**6.604376**	**4.591191**	**6.796448**	**4.591191**	**8.42945**

Figure 3: Comparison PARFUM and IPF, 50,000 samples, 50 iterations

variables, hence for the 6-th variable the constraints are always satisfied. Glancing at the other variables, each of whose constraints would be satisfied at an appropriate point in the cycle, we can appreciate that the IPF algorithm is far from converging. We note that in case of infeasibility PARFUM's performance seems to degrade more gracefully than that of IPF, though no quantitative measures have been applied in this regard.

6 Convolution constraints with prescribed margins

The iterative re-weighting algorithms can be used to impose constraints on the joint distribution, when these can be expressed as quantile constraints on functions of the margins. To illustrate, assume that samples $(x_i, y_i), i = 1, \ldots N$ from variables (X, Y) are given. We want to re-weight the samples in such a way that the quantiles of the distribution of $X + Y$ will agree with those of the convolution $X \star Y$ [5]. We conjecture that imposing more constraints of this form we will make the distributions of X and Y more independent; as suggested by the following [6]:

Remark 1 *Let (X, Y) be a random vector such that for all $a, b \in \mathbb{R}$, $aX + bY$ is distributed as the convolution of aX and bY; then X and Y are independent.*

[5] $X \star Y$ is the random variable whose characteristic function is a product of the characteristic functions of X and Y.

We illustrate this with the following example:

Example 1 *Let U, V, W be three independent standard normal variables. Let $X = U + W$ and $Y = V + W$. Clearly X and Y are not independent. Let us draw N=10000 samples from X and Y, say (x_i, y_i). We re-weight these samples such that X and Y satisfy $5\%, 50\%$ and 95% for exponential distribution with parameter $\lambda = 1$ (0.0513, 0.6931, 2.9957). Moreover since if X and Y are independent exponentials then we know that $X + Y$ has gamma distribution with scale 2 and shape 1, we require that the sum of samples $x_i + y_i$ satisfies quantile constraints for gamma distribution (0.3554, 1.6783, 4.7439).*

The PARFUM algorithm was used for this problem and the result is presented in Figure 4.

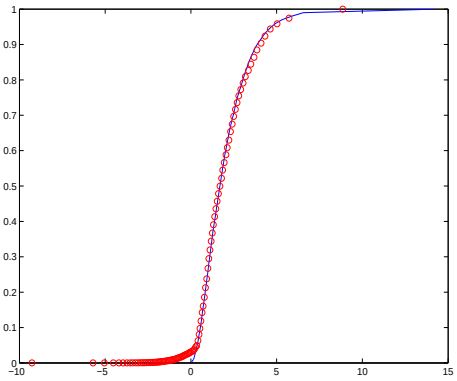

Figure 4: The result of re-weighting samples to match three quantiles for the exponential distribution and three quantiles for the gamma distribution for the sum of X and Y. The solid curve represents gamma distribution with scale 2 and shape 1, the 'circle' curve shows the distribution of the sum of re-weighted with PARFUM algorithm samples.

We can see in Figure 4 how closely the sum of re-weighted samples resembles the gamma distribution. For this example we obtained very good fit using only three quantiles.
In Figure 5 (left) 5000 samples from the original distribution of X and Y are shown. Since $X = U + W$ and $Y = V + W$ with U, V, W independent standard normals, the correlation between X and Y is approximately equal to 0.5. Figure 5 (right) shows the scatter plot of the distribution after re-weighting samples to match three quantiles of the exponential distribution and three quantiles of the gamma distribution for the sum of the re-weighted samples. Since we fit only three quantiles there is still quite high probability that samples will be negative. In Figure 6 (left) we present the scatter plot for the same case when 7 quantiles $(1\%, 5\%, 25\%, 50\%75\%, 95\%, 99\%)$ were used. One can see that increasing of the number of quantiles leads to a better approximation of the independent exponential distributions. By adding one constraint concerning the sum of the re-weighted samples we could reduce the correlation to 0.2043 in case of three quantiles and to 0.1449 is case of

seven quantiles. Adding additional constraints on linear combinations of X and Y leads to greater reduction of correlation (Figure 6 (right)).

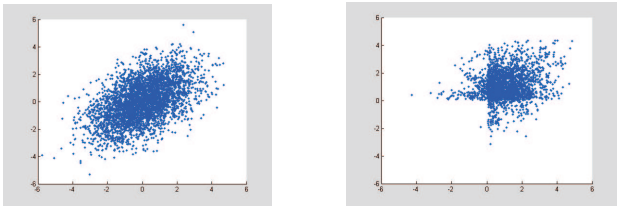

Figure 5: Scatter plots with 5000 samples from the original distribution of X and Y [correlation 0.4999] (left) and the distribution after re-weighting to match three quantiles of the exponential distribution and three quantiles of the gamma distribution for the sum of the re-weighted samples; [correlation 0.2043] (right).

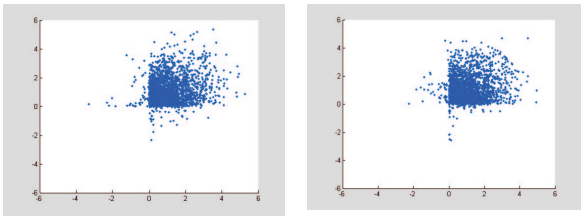

Figure 6: Scatter plots with 5000 samples from the distribution after re-weighting to match seven quantiles of the exponential distribution and seven quantiles of X+Y [correlation 0.1449] (left) and the distribution after re-weighting to match seven quantiles of the exponential distribution and seven quantiles for X+Y, X+2Y and X+4Y [correlation 0.1328] (right).

For the above examples we could easily stipulate quantile constrains on $aX + bY$ for $a > 0$ and $b > 0$ because it is known that if $a = b$ then $aX + bY$ has a gamma distribution and if $a \neq b$ it has a generalized Erlang distribution with density $f_{aX+bY}(z) = \frac{1}{a-b}e^{-z/a} + \frac{1}{b-a}e^{-z/b}$. If we do not know the distribution of the sum of independent linear combinations of X and Y, then this distribution must first be simulated to extract the quantile constraints.

7 Conclusions

We see that iterative algorithms possess attractive features for solving probabilistic inversion problems. These algorithms do not require intelligent steering, other than the choice of the initial sample. Further they do not pose size restrictions on the sample, so the sample may be large. Of course large samples will increase run time.

We note that in practice probabilistic inversions are typically infeasible. In such cases IPF is problematic, as little is known about its behavior in the case of infeasibility. The iterative PARFUM algorithm on the other hand is shown to behave in more predictable way in this case. It has fixed points minimizing an information functional even if the problem is infeasible

In this paper we have discussed only discrete inverse problem. The starting distribution was obtained by sampling and the marginal quantile constraints were imposed by reweighing these samples. To obtain the weights for the samples IPF or PARFUM algorithms have been used. We have indicated how these algorithms can be generalized to satisfy joint as well as marginal constraints and be used in continuous case.

Acknowledgement

The authors are grateful to V. Girardin for helpful discussions of iterative proportional fitting and convolution constraints and F.M. Dekking for improving the presentation of the proof of Theorem 3.

Appendix

Theorem 3
Let $\bigcap_{m=1}^{M} \mathcal{Q}_m = \mathcal{Q} \neq \emptyset$. If p is a fixed point of the PARFUM algorithm then $p \in \mathcal{Q}$.

We will present the proof of this theorem only for the 2-dimensional case, hence $M = 2$. This case is easy to demonstrate and shows all steps of the proof that is similar for higher dimensions. Before we start the proof of Theorem 3 we introduce a necessary notation, definitions and facts used later on in the proof.
Let $r(0)$ denotes a starting point of the PARFUM algorithm and $s \in \mathcal{Q}$ hence s is a solution. For 2-dimensional case $r(0), p$ and s are $K \times K$ (K -number of interquantile intervals) matrices. It is easy to notice the following:

Proposition 3 *The problem will remain unchanged if we exchange row i_1 and i_2 of $p, r(0)$ and s, simultaneously. The same applies to columns.*

Definition 3 *We call two columns j^* and j^{**} of p related if there exists a sequence $(i_k, j_k), k = 1, \ldots, t$ satisfying*

$$j^* \;=\; j_1,$$

$$
\begin{aligned}
i_k &= i_{k-1} \text{ when } k \text{ is even and } 1 < k < t, \\
j_k &= j_{k-1} \text{ when } k \text{ is odd and } 1 < k < t, \\
j_t &= j^{**}
\end{aligned}
$$

where $p_{i_k,j_k} > 0$ *for all* $k = 1, 2, \ldots, t$.

In prose the column j^* is related to j^{**} ($j^* \mathcal{R} j^{**}$) if there exists a "route" through the nonzero cells of p from column j^* to column j^{**}.

Proposition 4 *If* $j \mathcal{R} j^*$ *then*

$$
\frac{q_j}{p_{.,j}} = \frac{q_{j^*}}{p_{.,j^*}}.
$$

Proof.
Since p is a fixed point of PARFUM and $p_{i_k,j_k} > 0$ for all $k = 1, 2, \ldots, t$ then we can divide both sides of the equation below by p_{i_k,j_k}.

$$
p_{i_k,j_k} = \frac{p_{i_k,j_k}}{2} \left(\frac{q_{i_k}}{p_{i_k,.}} + \frac{q_{j_k}}{p_{.,j_k}} \right).
$$

Applying this to all cells in the 'route' we get

$$
\begin{aligned}
2 &= \left(\frac{q_{i_1}}{p_{i_1,.}} + \frac{q_{j_1}}{p_{.,j_1}} \right) = \left(\frac{q_{i_1}}{p_{i_1,.}} + \frac{q_j}{p_{.,j}} \right), \\
2 &= \left(\frac{q_{i_2}}{p_{i_2,.}} + \frac{q_{j_2}}{p_{.,j_2}} \right) = \left(\frac{q_{i_1}}{p_{i_1,.}} + \frac{q_{j_2}}{p_{.,j_2}} \right), \\
2 &= \left(\frac{q_{i_3}}{p_{i_3,.}} + \frac{q_{j_3}}{p_{.,j_3}} \right) = \left(\frac{q_{i_3}}{p_{i_3,.}} + \frac{q_{j_2}}{p_{.,j_2}} \right), \\
2 &= \left(\frac{q_{i_4}}{p_{i_4,.}} + \frac{q_{j_4}}{p_{.,j_4}} \right) = \left(\frac{q_{i_3}}{p_{i_3,.}} + \frac{q_{j_4}}{p_{.,j_4}} \right), \\
&\ldots \\
2 &= \left(\frac{q_{i_{t-1}}}{p_{i_{t-1},.}} + \frac{q_{j_{t-1}}}{p_{.,j_{t-1}}} \right) = \left(\frac{q_{i_t}}{p_{i_t,.}} + \frac{q_{j_{t-1}}}{p_{.,j_{t-1}}} \right), \\
2 &= \left(\frac{q_{i_t}}{p_{i_t,.}} + \frac{q_{j_t}}{p_{.,j_t}} \right) = \left(\frac{q_{i_t}}{p_{i_t,.}} + \frac{q_{j^*}}{p_{.,j^*}} \right).
\end{aligned}
$$

Thus

$$
\frac{q_j}{p_{.,j}} = \frac{q_{j_1}}{p_{.,j_1}} = \frac{q_{j_2}}{p_{.,j_2}} = \ldots = \frac{q_{j^*}}{p_{.,j^*}},
$$

which concludes the proof. □

We call $\frac{q_j}{p_{.,j}}$ the *weight of the column* j. Thus the above proposition says that if two columns are related then they have the same weights.

Proof of Theorem 3.

We start the proof with rearranging columns and rows of $r(0), p$ and s, simultaneously.

Step 1. Choose one column of p, say column 1, and find all columns related to 1. Rearrange these columns to the left of p. We denote this changed distribution by $\mathcal{T}_1 p$ where $\mathcal{T}_1$ is the matrix operator in step 1. $\mathcal{T}_1 p$ is the fixed point of PARFUM such that columns 1 to n_1 are related and not related with any other columns.

Step 2. Next we rearrange rows of $\mathcal{T}_1 p$. All rows such that $\sum_1^{n_1} \mathcal{T}_1 p_{i,j} \neq 0$ are moved to the upper part of the matrix. Let us assume that we have found m_1 rows like that, hence in this step we have obtained the matrix $\mathcal{T}_2\mathcal{T}_1 p$ with the 'lower-left' sub-matrix consisting of 0's (it can of course happen that $m_1 = K$).

If $m_1 < K$ then we claim that 'upper-right sub-matrix of $\mathcal{T}_2\mathcal{T}_1 p$ consists of 0's.

Otherwise, there exists $\mathcal{T}_2\mathcal{T}_1 p_{i^*,j^*} > 0$ where $i^* \leq m_1, j^* > n_1$. By the construction there exists $j^{**} \leq n_1$ satisfying $\mathcal{T}_2\mathcal{T}_1 p_{i^*,j^{**}} > 0$, thus columns j^* and j^{**} are related which is impossible because we have chosen all n_1 related columns already.

Step 3. By induction we can rearrange the whole p finding all related columns, hence the matrix transformation $\mathcal{T}_3$ of this step can be written

$$\mathcal{T}_3 p \;=\; \mathcal{T}_2^l \mathcal{T}_1^l \mathcal{T}_2^{l-1} \mathcal{T}_1^{l-1} \ldots \mathcal{T}_2^1 \mathcal{T}_1^1 p$$

and the obtained matrix can be written as

$$\mathcal{T}_3 p \;=\; \begin{bmatrix} \Gamma_1 & & & \\ & \Gamma_2 & & \\ & & \ldots & \\ & & & \Gamma_l \end{bmatrix}.$$

Step 4. In this step we rearrange Γ_i such that corresponding to columns of Γ_i weights are in increasing order. We also combine Γ_i and Γ_j if corresponding to their columns weights are equal. After this step we obtain the following matrix

$$\mathcal{T}_4\mathcal{T}_3 p \;=\; \begin{bmatrix} \Omega_1 & & & \\ & \Omega_2 & & \\ & & \ldots & \\ & & & \Omega_r \end{bmatrix}$$

such that weights w_i corresponding to columns of sub-matrices Ω_i satisfy $w_1 < w_2 < \ldots < w_r, r \leq l$ and $w_1 < 1$.

All matrix operations were performed on $p, r(0)$ and s. From now on when we refer to $p, r(0)$ or s we mean $\mathcal{T}_4\mathcal{T}_3 p, \mathcal{T}_4\mathcal{T}_3 r(0)$ and $\mathcal{T}_4\mathcal{T}_3 s$, respectively.

Since the problem is feasible, there exists a solution s satisfying the marginal constraints q that is absolutely continuous with respect to the starting point $r(0)$. After transformations in Steps 1-4 we get that

$$p = \begin{bmatrix} \Omega_1 & 0_1 \\ 0_2 & \Omega_{others} \end{bmatrix},$$

where 0_1 and 0_2 are matrices of zeros with respective sizes and

$$\Omega_{others} = \begin{bmatrix} \Omega_2 & & & \\ & \Omega_3 & & \\ & & \dots & \\ & & & \Omega_r \end{bmatrix}.$$

The rearranged solution is the following

$$s = \begin{bmatrix} \Delta_1 & \Delta_2 \\ \Delta_3 & \Delta_4 \end{bmatrix},$$

where the sizes of respective matrices in p and s are the same. Let Ω_1 and Δ_1 be $m_1 \times n_1$ matrices. In order to prove our theorem we must show that $r = 1$ hence that $p = \Omega_1$. Suppose on the contrary that $r > 1$. Then w_1 has to be smaller then 1 since $w_2 > w_1$. By the definition of weight, we get

$$\begin{aligned} \frac{q_j}{p_{.,j}} &= w_1, j = 1, 2, \dots, n_1, \\ \frac{q_i}{p_{i,.}} &= 2 - w_1, i = 1, 2, \dots, m_1. \end{aligned}$$

If we denote the sum of all elements of the matrix X by $\parallel X \parallel$ then we get

$$\begin{aligned} \sum_{j=1}^{n_1} q_j &= w_1 \sum_{j=1}^{n_1} p_{.,j} = w_1 \parallel \Omega_1 + 0_2 \parallel = w_1 \parallel \Omega_1 + 0_1 \parallel \\ &= w_1 \sum_{i=1}^{m_1} p_{i,.} = \frac{w_1}{2 - w_1} \sum_{i=1}^{m_1} q_i. \end{aligned}$$

Since $w_1 < 1$ then $\frac{w_1}{2-w_1} < 1$ so from the above we get that $\sum_{j=1}^{n_1} q_j < \sum_{i=1}^{m_1} q_i$.

The solution s satisfies the marginal constraints q, hence from the above we get

$$\|\Delta_1 + \Delta_3\| = \sum_{j=1}^{n_1} q_j < \sum_{i=1}^{m_1} q_i = \|\Delta_1 + \Delta_2\|,$$

which implies that

$$\|\Delta_2\| > \|\Delta_3\| \geq 0.$$

¿From the above there exists at least one cell in Δ_2 that is non-zero. Let $s_{i^*,j^*} > 0, i^* \leq m_1, j^* > n_1$. s is absolutely continuous with respect to $r(0)$ hence $r(0)_{i^*,j^*} > 0$. Notice also that $p_{i^*,j^*} = 0$ since it is one of the elements of 0_1.
We have a situation where a non-zero cell in the initial distribution $r(0)$ converges to zero by the PARFUM algorithm. Let us consider the following

$$\frac{1}{2}\left(\frac{q_{i^*}}{r(n)_{i^*,.}} + \frac{q_{j^*}}{r(n)_{.,j^*}}\right) \rightarrow \frac{1}{2}\left(\frac{q_{i^*}}{p_{i^*,.}} + \frac{q_{j^*}}{p_{.,j^*}}\right) = \lambda,$$

where $r(n)$ denotes nth step of the PARFUM algorithm. The (n+1)th step of PARFUM algorithm is given

$$r(n+1)_{i^*,j^*} = \frac{r(n)_{i^*,j^*}}{2}\left(\frac{q_{i^*}}{r(n)_{i^*,.}} + \frac{q_{j^*}}{r(n)_{.,j^*}}\right).$$

Since $r(n)_{i^*,j^*} \rightarrow 0$ as $n \rightarrow \infty$ then there exists a monotonically convergent sub-sequence of $r(n)_{i^*,j^*}$ that converges to zero. Hence

$$\lambda \leq 1. \tag{17}$$

Since $i^* \leq m_1$ then $\frac{q_{i^*}}{p_{i^*,.}} = 2 - w_1$. We also know that $\frac{q_{j^*}}{p_{.,j^*}} = w_f$ where $1 < f < r$ and $w_1 < w_f$. Hence

$$\lambda = \frac{1}{2}\left(\frac{q_{i^*}}{p_{i^*,.}} + \frac{q_{j^*}}{p_{.,j^*}}\right) = \frac{1}{2}(2 - w_1 + w_f) > \frac{1}{2}2 = 1.$$

This contradicts (17) hence $r = 1$ that implies that p has both margins equal to q, hence p is a solution. □

Remark 2 *Note that in the proof of Theorem 3 both margins are equal to q. The proof will be valid if the margins are different.*

References

[1] Harper F., Goossens L., Cooke R.M., Hora S., Young M., Pasler-Ssauer J., Miller L., Kraan B., Lui C., McKay M., Helton J., Jones A. Joint usnrc cec consequence uncertainty study: Summary of objectives, approach, application, and results for the dispersion and deposition uncertainty assessment. Technical Report VOL. III, NUREG/CR-6244, EUR 15755 EN, SAND94-1453, 1994.

[2] Kraan B.C.P., Cooke R.M. Processing expert judgements in accident consequence modeling. *Radiation Protection Dosimetry*, 90(3):311–315, 2000.

[3] Kraan B.C.P. *Probabilistic Inversion in Uncertainty Analysis and related topics*. Isbn 90-9015710-7, PhD dissertation, TU Delft, 2002.

[4] Cooke R.M, Goossens L.H.J. Procedures guide for structured expert judgement in accident consequence modelling. *Radiation Protection Dosimetry*, 90(3):303–309, 2000.

[5] Brown D.T. A note on approximations to discrete probability distributions. *Inform. and Control*, 2:386–392, 1959.

[6] V. Girardin and N. Limnios. *Probabilités en vue des applications*. Vuibert, Paris, 2001.

[7] Csiszar I. I-divergence geometry of probability distributions and minimization problems. *Ann. of Probab.*, 3:146–158, 1975.

[8] Csiszar I. and Tusnady G. Information geometry and alternating minimization procedures. *Statistics & Decisions*, 1:205–237, 1984.

[9] Ireland C.T. and Kullback S. Contingency tables with given marginals. *Biometrika*, 55:179–188, 1968.

[10] Kruithof J. Telefoonverkeersrekening. *De Ingenieur*, 52(8):E15–E25, 1937.

[11] Kraan B and Bedford T.J. Probabilistic inversion of expert judgements in the quantification of model uncertainty. appearing in Management Science, 2004.

[12] Kraan B.C.P., Cooke R.M. Uncertainty in compartmental models for hazardous materials - a case study. *J. of Hazardous Materials*, 71:253–268, 2000.

[13] Kurowicka D. and Cooke R.M. *Uncertainty Analysis with High Dimensional Dependence Modelling*. Wiley, 2005.

[14] Rüschendorf L. Convergence of the iterative proportional fitting procedure. *The Annals of Statistics*, 23(4):1160–1174, 1995.

[15] Cooke R.M. Parameter fitting for uncertain models: modelling uncertainty in small models. *Reliability Engineering and System Safety*, 44:89–102, 1994.

[16] Kullback S. *Information theory and Statistics*. John Wiley and Sons, New York, 1959.

[17] Kullback S. Probability densities with given marginals. *The Annals of Mathematical Statistics*, 39(4):1236–1243, 1968.

[18] Kullback S. Marginal homegeneity of multidimensional contingency tables. *The Annals of Methematical Statistics*, 42(2):594–606, 1971.

[19] Fienberg S.E. An iterative procedure for estimation in contingency tables. *Ann. of Math. Statist.*, 41:907–917, 1970.

[20] Haberman S.J. *An Analysis of Frequency Data*. Univ. Chicago Press., 1974.

[21] Haberman S.J. Adjustment by minimum discriminant information. *Ann. of Statist.*, 12:971–988, 1984.

[22] Van der Fels-Klerx, H.J., Cooke, R.M., Nauta, M.J., Goossens, L.H.J., Havelaar, A.H. A structured expert judgment study for a model of campylobacter contamination during broiler chicken processing. Appearing in *Risk Analysis*, 2005.

[23] Bishop Y.M.M. Multidimensional contingency tables: cell estimates. *Ph.D. dissertation, Harvard Univ*, 1967.