

# ROBUST ROUTING AND CROSS-ENTROPY ESTIMATION

Hélène Le Cadre

helene.lecadre@enst-bretagne.fr

ENST, Bretagne  
Technopôle Brest Iroise  
CS 83818,  
29238 Brest Cedex 3,  
FRANCE

September 9, 2005

## 1 Introduction

In this article we present a novel way to estimate the amounts of traffic on the Origin-Destination couples (OD couples). This new approach combines together a routing algorithm based on the principle of the shortest path and a recent technique of stochastic optimization called Cross-Entropy. The CE method was built at the origin, to tackle problems of rare-event simulation. However, its inventor, R. Rubinstein, realized soon that the underlying idea should be applied efficiently to combinatorial and multi-extremal optimization problems.

In a final part, we adapt a particular filtering algorithm in order to be able to dynamically estimate the evolution of the traffic on the OD couples.

The aim of this report is to highlight rather original ideas, however the choices of the prior distributions and some specific parameters may be quite arbitrary.

## 2 A brief presentation of the CE method

### 2.1 Rare-Event Simulation

Let  $X = (X_1, \dots, X_N)$ , be a random vector taking values in a space called  $\mathcal{X}$ . Let  $\{f(\cdot; v)\}$  be a family of parametric densities defined on the space  $\mathcal{X}$ , with respect to the Lebesgue measure. For any measurable function  $H$ , we can define:

$$\mathbf{E}[H(X)] = \int H(x) f(x; v) dx .$$

The performance function will be called  $S : \mathcal{X} \rightarrow \mathbb{R}$ . For a fixed level  $\gamma$ , we are interested in the probability of the event defined below:

$$l = \mathbf{P}_u(S(X) \geq \gamma) = \mathbf{E}_u[\mathbf{1}_{\{S(x) \geq \gamma\}}].$$

If this probability is very small, for example not more than  $10^{-5}$ , the set  $\{S(x) \geq \gamma\}$  will be called a *rare event*.

A straightforward way to estimate  $l$  may be to use crude Monte-Carlo simulation: let  $(X_1, \dots, X_N)$  be a sample drawn from the density  $f(\cdot; u)$ . Then, the estimator,

$$\frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{S(x) \geq \gamma\}}$$

is an unbiased estimator of  $l$ . However, if  $\{S(x) \geq \gamma\}$  is a rare event, many indicator functions will remain equal to zero. As a result, we will be forced to simulate huge samples, which is rather costly and difficult to put in application.

Another way to get an estimate of  $l$  might be to use importance sampling. We should draw  $(X_1, \dots, X_N)$  from a density  $g$ , defined on the space  $\mathcal{X}$ . This density is nothing else than a mere change of measure. The estimator then becomes:

$$\hat{l} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{S(X_i) \geq \gamma\}} \frac{f(X_i; u)}{g(X_i)}. \quad (1)$$

The optimal density  $g$ , is defined by:

$$g^*(x) = \frac{\mathbf{1}_{\{S(x) \geq \gamma\}} f(x; u)}{l}. \quad (2)$$

Substituting (2) in (1), we get:

$$\mathbf{1}_{\{S(X_i) \geq \gamma\}} \frac{f(X_i; u)}{g^*(X_i)} = l, \forall i.$$

But,  $l$  is a constant. As a result, the estimator defined in (1) has zero variance. Nevertheless,  $g^*$  depends on the unknown parameter  $l$ . The idea is in fact, to choose  $g$  in a parametric family of densities  $\{f(\cdot; v)\}$ . The problem is now to determine the optimal parameter  $v$ , such that the distance between  $g^*$  and  $f(\cdot; v)$  should be minimized.

A well-known "distance" between two densities  $g$  and  $h$ , is the Kullback-Leibler "distance":

$$\mathcal{D}(g, h) = \mathbf{E}_g[\ln \frac{g(X)}{h(X)}] = \int g(x) \ln g(x) dx - \int g(x) \ln h(x) dx. \quad (3)$$

Minimizing the Kullback-Leibler distance between  $g^*$  and  $f(\cdot; v)$ , is equivalent to solve the following problem:

$$\arg \max_v \int g^*(x) \ln f(x; v) dx . \quad (4)$$

Substituting (2) in (4), we get:

$$\arg \max_v D(v) = \arg \max_v \mathbf{E}_u[\mathbf{1}_{\{S(X) \leq \gamma\}} \ln f(X; v)] . \quad (5)$$

But, in fact we can estimate  $v^*$ , by solving the following stochastic program:

$$\arg \max_v \hat{D}(v) = \arg \max_v \frac{1}{N} \sum_{i=1}^N [\mathbf{1}_{\{S(X_i) \geq \gamma\}} \ln f(X_i; v)] . \quad (6)$$

If  $\hat{D}$  is convex and differentiable in  $v$ , we just have to solve the following problem:

$$\frac{1}{N} \sum_{i=1}^N [\mathbf{1}_{\{S(X_i) \geq \gamma\}} \nabla_v \ln f(X_i; v)] = 0 . \quad (7)$$

The solution can often be calculated analytically, which is one of the great advantages of this approach.

## 2.2 Application of the CE to optimization

Usually, in the field of optimization we try to solve problems of the form:

$$S(x^*) = \gamma^* = \arg \max_{x \in \mathcal{X}} S(x) . \quad (8)$$

The genius of the Cross-Entropy method lies in the fact that it is possible to associate with each optimization problem of the form (8), a problem of estimation, called *associated stochastic problem (ASP)*. We will start by defining a collection of indicator functions  $\{\mathbf{1}_{\{S(x) \leq \gamma\}}\}_{\gamma \in \mathbb{R}}$ , on the space  $\mathcal{X}$ . Then, we will define a parametric family of densities  $\{f(\cdot; v), v \in \mathcal{V}\}$  on the space  $\mathcal{X}$ . Let  $u \in \mathcal{V}$ . We will associate with (8), the following stochastic estimation problem:

$$l(\gamma) = \mathbf{P}_u(S(X) \geq \gamma) = \sum_x \mathbf{1}_{\{S(x) \geq \gamma\}} f(x; u) = \mathbf{E}_u[\mathbf{1}_{\{S(X) \geq \gamma\}}] , \quad (9)$$

If  $\gamma = \gamma^*$ , a natural estimator of the reference parameter  $v^*$  is:

$$\hat{v}^* = \arg \max_v \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{S(X_i) \geq \gamma\}} \ln f(X_i; v) , \quad (10)$$

where the  $X_i$  are drawn from the density  $f(\cdot, u)$ . If  $\gamma$  is very close to  $\gamma^*$ , then  $f(\cdot; v^*)$  assigns most of its probability mass close to  $x^*$ . In fact, in this case, we will have to choose  $u$  so that  $\mathbf{P}_u(S(X) \geq \gamma)$  is not too small. We can infer that  $u$  and  $\gamma$  are closely linked.

We will use a two level procedure. Indeed, we will construct two sequences  $\hat{\gamma}_1, \dots, \hat{\gamma}_T$  and  $\hat{v}_0, \hat{v}_1, \dots, \hat{v}_T$  such that  $\hat{\gamma}_T$  is close to  $\gamma^*$  and  $\hat{v}_T$  is such that the density assigns most of its mass in the state which maximizes the performance.

The algorithm follows a two-step strategy:

**Algorithm 1.**    • Define  $\hat{v}_0 = u$ . Set  $t = 1$ .

- Generate a sample  $(X_1, \dots, X_N) \sim f(\cdot; v_{t-1})$ . Compute the  $(1 - \rho)$ -quantile of the performance  $S$ , which can be estimated by:

$$\hat{\gamma}_t = S_{[(1-\rho)N]}.$$

- Use the same sample  $(X_1, \dots, X_N)$ , to solve (6). Call the solution  $v_t$ .
- If for some  $t \geq d$ ,  $d$  fixed,

$$\hat{\gamma}_t = \dots = \hat{\gamma}_{t-d},$$

*Stop*; otherwise set  $t = t + 1$  and reiterate from Step 2.

### 3 The Model

The network we study is composed of  $p$  nodes and  $n$  arcs. At first, we will suppose that there exists an arc between each couple of nodes. Furthermore, we make the difference between the two couples  $(i, j)$  and  $(j, i)$ ,  $\forall i, j \in \{1, \dots, n\}$ ,  $i \neq j$ . Consequently,

$$n = p^2 - p.$$

Recall that an arc is a directed link. In the rest of this article, we will make the hypothesis that the network is directed.

Our work could be separated into two different parts. Firstly, we have to deal with a simulation part. In this section, we will choose an initial vector of the amounts of traffic on the OD couples. Our aim will be to minimize the global sum of the costs, which are associated to each arc of the network. Secondly, we will have to cope with an estimation part. Indeed, we will suppose that the initial costs are equal to those obtained by the simulation part. The idea then, will be to find the estimator  $\hat{X}(t)$  from which we could infer an arc estimator  $\hat{Y}(t)$  minimizing the distance to the vector  $Y(t)$ , obtained in the simulation part. This raises the crucial question of the identifiability of the vector  $X(t)$ . Is there unicity of the associated vector  $\hat{X}(t)$  or, is it only an element in a vast variety?

### 3.1 Simulation

In this part, we associate a cost function to the network, which means that we give a cost to each arc of the network. This cost function is drawn from a parametric families of densities. Which means that we have to determine the optimal parameter of the density function. What's more, this cost may or, may not be, proportionnal to the amount of traffic on each arc. However, it is more unconventional to suppose that the costs depend on the arc traffic. Let  $Y(t)$  be the vector which contains the amount of traffic on each arc of the network and  $C(t)$  the vector which represents the costs associated with each arc. In the most general case, we have:

$$C(t) = F(Y(t)), \tag{11}$$

where  $F$  is supposed to be continuous and differentiable.

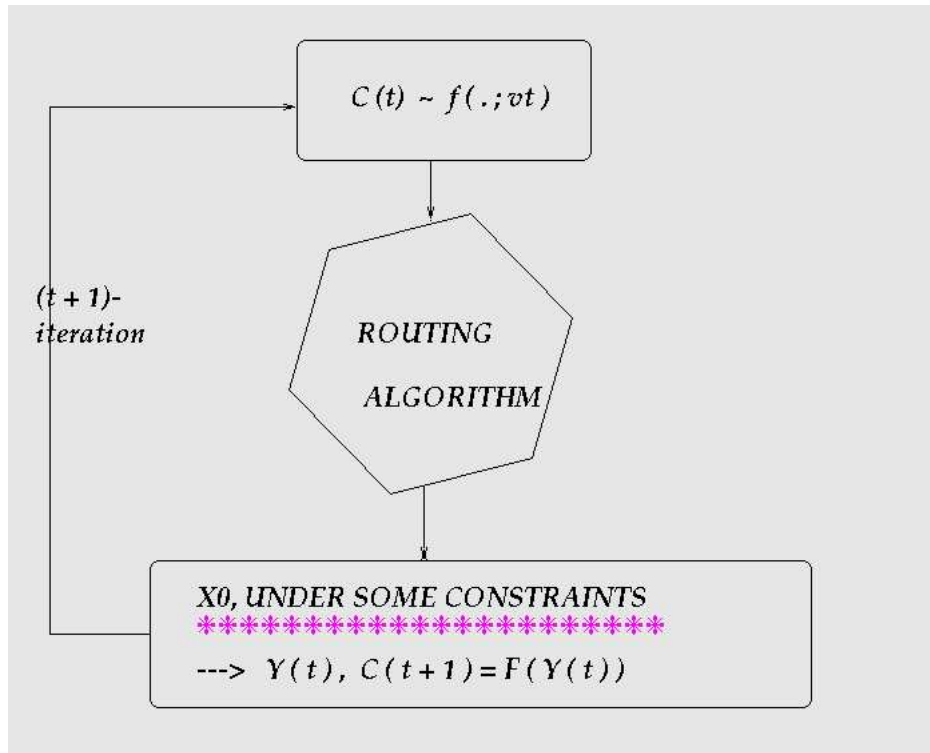


Figure 1: Simulation

Then, using these costs, we will use a routing algorithm based on the principle of the shortest path in order to find the shortest paths between each OD couple. A path is

represented by the nodes or the arcs which it is made of.

The principle of the routing algorithm we use is rather simple. The first observation to make is that every subpath of a shortest path is necessarily itself a shortest path.

Let  $e_{i,j}$  be the arc linking the nodes  $i$  and  $j$ . If the path composed of the arcs  $\{e_{i,j}, e_{j,k}, e_{k,l}, \dots, e_{p,q}\}$  is the shortest path linking node  $i$  to node  $q$ . Then,  $e_{i,j}$  must be the shortest path between  $i$  and  $j$ .  $e_{j,k}$  must be the shortest path between  $j$  and  $k$ , and so on...

$e_{i,j}$  is called *basic arc*, iff it is the shortest path between  $i$  and  $j$ . Consequently, each shortest path must be composed exclusively of basic arcs. The aim of our routing algorithm is to substitute to each arc which is not basic, a basic arc. Let  $d_{i,j} = C(t; i, j)$  be the distance or weight associated to the arc  $\{i, j\}$ , at the instant  $t$ . Let  $j$  be the indice of a node of the network, then:

$$\forall i, k \in \text{Network} - \{j\}, d_{i,k} \leftarrow \min\{d_{i,k}, d_{i,j} + d_{j,k}\}.$$

The algorithm tests all the couple of nodes  $(i, k)$ , which are neighbors of  $j$ , while  $j$  takes each node of the network as its own value.

**Algorithm 2.** *Input:*  $C(t)$ , vector of the costs.

- If  $d_{i,k} \geq d_{i,j} + d_{j,k}$ , do not change anything.
- If  $d_{i,k} \leq d_{i,j} + d_{j,k}$ , create an arc linking  $i$  to  $k$  and associate the weight  $d_{i,k} = d_{i,j} + d_{j,k}$ .

**Output:** the shortest paths between each OD couple.

The algorithm also gives us the shortest distances associated to each OD couple. But, these distances are only rough estimators of the amounts of traffic on each OD couple. Indeed, more than one link, can be shared by different shortest paths linking different OD couples. As a result, the total amount of traffic generated by **one** OD couple usually represents only a fraction of the total traffic flowing through the arcs which composed the path.

In order to solve this crucial problem, we will associate to the vector which contains the volumes of traffic on the OD couples, called  $X_0(t)$ , an estimator of the volumes of traffic flowing through the arcs, which we will note  $Y(t)$ . Indeed, if we use the routing algorithm, it is quite easy to deduce  $Y(t)$  from  $X_0(t)$ . Our goal will be to solve the following optimization problem:

$$\min \sum_{i=1}^n |C_i(t)|. \quad (12)$$

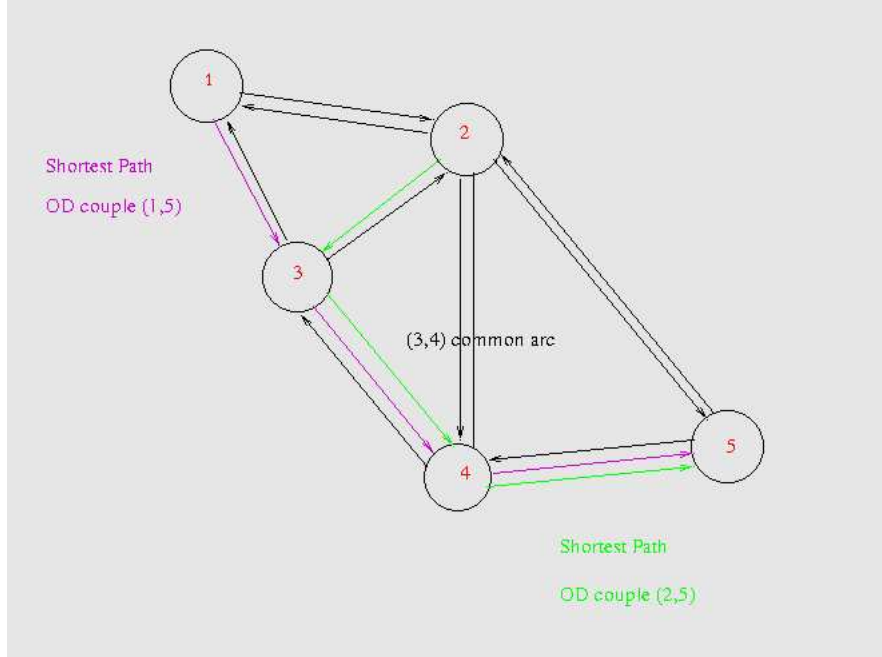


Figure 2: Connection between OD couples and arcs

### 3.2 Estimation

In this part, the performance function is defined by:

$$S(X(t)) = \arg \max_{x^i(t)} \frac{1}{\|Y(t) - \hat{Y}(t)\|_2}. \quad (13)$$

The first observation is that  $S$  is an implicit function of  $X(t)$ , e.g. we can't get any exact analytical expression of  $S$ . Consequently, we will have to resort to use simulation.

What's more, we suppose that the vector  $X(t)$  is generated from an exponential density whose parameter is totally unknown.

$$X(t) \sim \mathcal{E}(\lambda), \lambda \in \mathbb{R}_+^n. \quad (14)$$

We apply the Cross-Entropy method to our problem. At each iteration, we generate a random sample  $(X^{(1)}, \dots, X^{(N)}) \sim \mathcal{E}(\lambda), \lambda \in \mathbb{R}_+^n$ .

**Hypothesis:** Each component of  $X(t)$ , which represents the amount of traffic on an OD couple, will be supposed to be independent of the others.

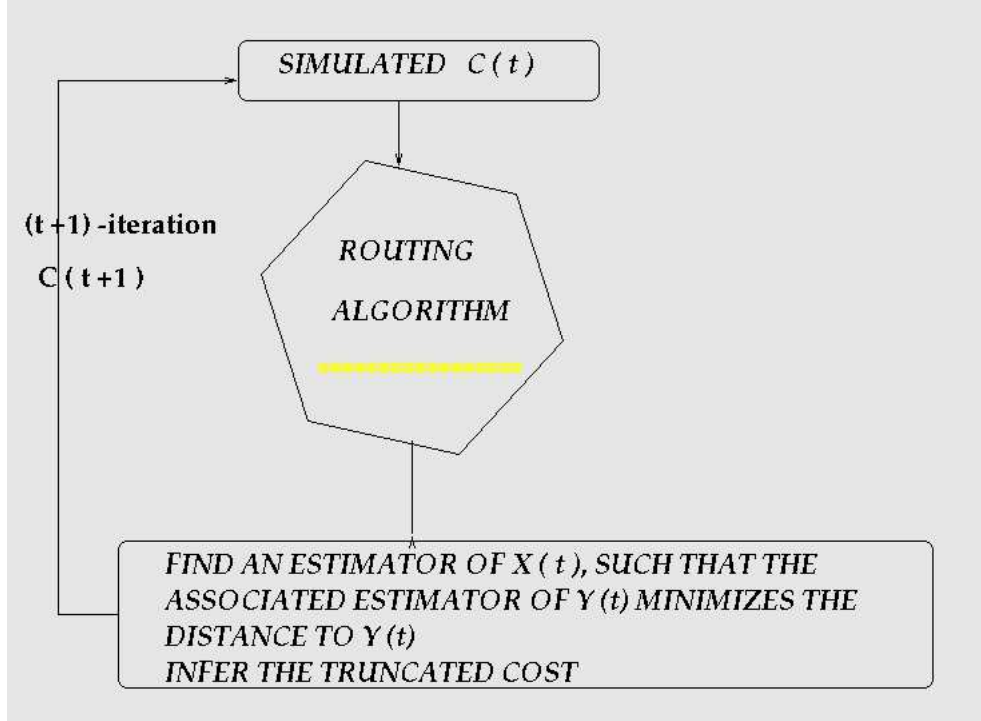


Figure 3: Estimation

Practically, each random vector from the sample will be stocked in a big matrix.

$$X(t) = \begin{bmatrix} X_1^{(1)} & \cdots & X_1^{(N)} \\ X_2^{(1)} & \cdots & X_2^{(N)} \\ \vdots & \vdots & \vdots \\ X_n^{(1)} & \cdots & X_n^{(N)} \end{bmatrix}.$$

The joined densities of the vectors  $X^{(j)}$ ,  $j = 1, \dots, N$ , are typically of the form:

$$\begin{aligned} f(X^{(j)}, \lambda) &= \prod_{i=1}^n \lambda_i \exp[-\lambda_i X_i^{(j)}] \mathbf{1}_{\mathbb{R}^+}(X_i^{(j)}), \\ &= \prod_{i=1}^n \lambda_i \exp[-\lambda_i X_i^{(j)}] \mathbf{1}_{\{\min(X_i^{(j)}) \geq 0\}}. \end{aligned} \quad (16)$$

As a result, we will have to solve the following problem:

$$\frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{S(X^{(i)}) \geq \hat{\gamma}_t\}} \nabla_{\lambda} \ln f(X^{(i)}; \lambda) = 0. \quad (17)$$

After some computations, we get:

$$\lambda_j = \frac{\sum_{i=1}^N \mathbf{1}_{\{S(X^{(i)}) \geq \hat{\gamma}_t\}}}{[\sum_{i=1}^N \mathbf{1}_{\{S(X^{(i)}) \geq \hat{\gamma}_t\}} X_j^{(i)}]}, \forall j = 1, \dots, n. \quad (18)$$

**Remark.**

If we generate the random vectors  $X_i$ ,  $i = 1, \dots, N$  from a truncated exponential, we can give some maximal boundaries on the OD volumes of traffic.

Recall that a truncated exponential is of the form:

$$f(x; \lambda, b) = \frac{\lambda \exp[-\lambda x]}{1 - \exp(-ab)} \mathbf{1}_{[0,b]}(x).$$

Under this assumption, we have to cope with the following system:

$$\frac{\sum_{i=1}^N \mathbf{1}_{\{S(X^{(i)}) \geq \hat{\gamma}_t\}} X_j^{(i)}}{\sum_{i=1}^N \mathbf{1}_{\{S(X^{(i)}) \geq \hat{\gamma}_t\}}} - \frac{1}{\lambda_j} + \frac{b_j}{\exp(\lambda_j b_j) - 1} = 0, \forall j \in \{1, \dots, n\}. \quad (19)$$

This system is non-linear, that's why we use the well-known iterative Newton's method to solve it.

## 4 The problem of Identifiability

We observe that the amount of traffic which flows through each arc of the network is equal to the sum of the amounts of traffic flowing on each OD couple which owns this arc in its shortest path. Remember that, thanks to the routing algorithm, we associate to each OD couple, a unique shortest path. Mathematically, we can express this remark under the following expression:

$$\sum_{j=1}^n \mathbf{1}_{\{e_i \in \text{OD couple number } j\text{'s path}(t)\}} X_j(t) = Y_i(t), \forall i \in \{1, \dots, n\}. \quad (20)$$

More generally, we get:

$$Y(t) = A(X(t)) X(t). \quad (21)$$

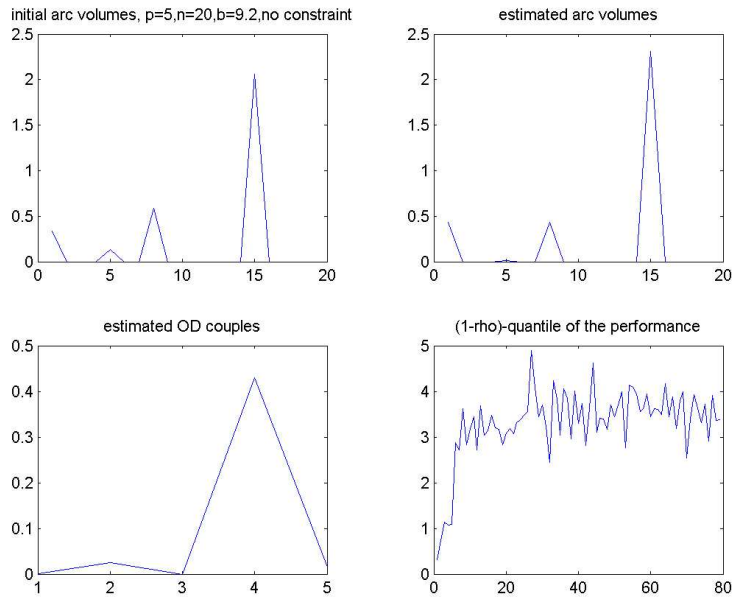


Figure 4: Simulation, 5 nodes network, 20 arcs

Indeed, in the most general case, the routing matrix  $A$  relies on the volumes of traffic flowing through each arc at the instant  $t$ . But, these arc volumes rely themselves on the OD volumes,  $X(t)$ . As a result, the routing matrix  $A(t)$ , is a function of  $X(t)$ .  $A$  is uniquely made of binary elements: 0 and 1. More explicitly,  $A(X(t); i, j) = 1$  iff, the arc numbered  $i$  belongs to the shortest path associated to the OD couple numbered  $j$ , at time  $t$ .

What's more, the routing algorithm do not use every arc. Consequently, many rows of the routing matrix equal zero. The associated components in the arc vector  $Y(t)$  are at the same time, null.

But, if we suppress the zero rows of  $A(X(t))$  and the zero components of  $Y(t)$ , this leads us to solve a rectangular system of equations. This system is under-determined, that's why we can't guarantee the existence of a unique solution.

We can conclude that there is no identifiability between the arc volumes  $Y(t)$ , and the OD volumes  $X(t)$ , at a given time. Indeed, if we take a fixed  $X(t)$ , we get a unique associated  $Y(t)$ , since the routing algorithm determines unique shortest path between each OD couples. Reciprocally, if we take some fixed arc volumes,  $Y(t)$ , we can't guarantee the unicity of the solutions of (21). That's why, we can't assert that the associated  $X(t)$  is perfectly unique.

A good idea to tackle this problem, is to suppose that some of the OD couples do

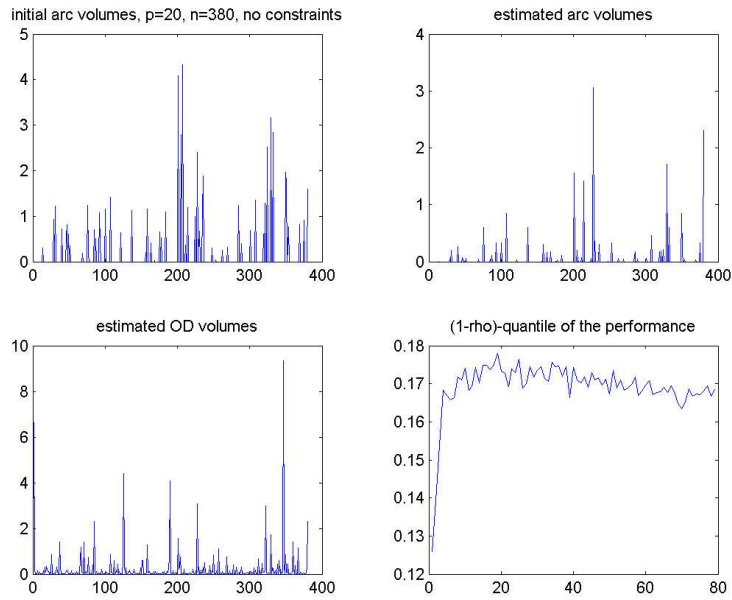


Figure 5: Simulation, 20 nodes network, 380 arcs

not accept any traffic. That is, that they remain equal to zero. The goal is to reduce the number of positive OD couples so as to get a system whose routing matrix  $A(t)$ , is square or not too far.

The first approach is to suppose that some pre-determined OD couples are excluded.

To begin, we may partition, a little arbitrarily, the set of the OD couples into two parts. In the first one, lie the OD couples which remain always equal to zero. And, in the second part, we will suppose that there is some traffic flowing through these OD couples.

We need to generate a sample  $Z$ . The components of  $Z$  are independent of each other and generated from a Bernoulli density whose parameter is pre-determined. Then, if  $Z(i) = 0$ , the OD couple number  $i$ , do not accept any traffic.

The time required to perform this simulation is of about 1 minute.

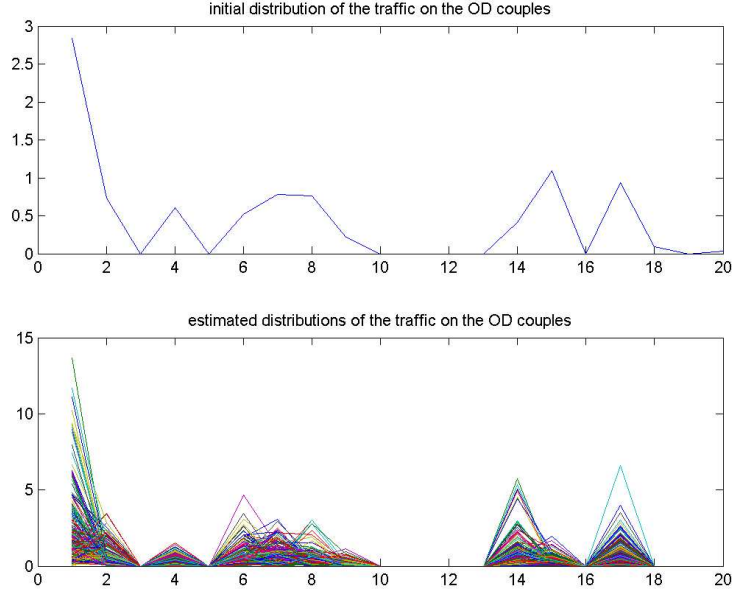


Figure 6: Introduction of constraints on the OD couples, the zero OD couples are pre-determined. Identifiability of  $X(t)$

A second point of vue should be to suppose that we know at the beginning that only  $K$  OD couples,  $K \geq n$ , are positive. So, we need to modify our CE algorithm. We need to introduce a matrix,  $Z(t)$ :

$$Z(t) = \begin{bmatrix} Z_1^{(1)} & \dots & Z_1^{(N)} \\ Z_2^{(1)} & \dots & Z_2^{(N)} \\ \vdots & \vdots & \vdots \\ Z_n^{(1)} & \dots & Z_n^{(N)} \end{bmatrix}.$$

To be more explicit, the  $i^{th}$  row of  $Z$  is generated from  $\mathcal{B}(p_{i,j}(t))$ ,  $j \in \{0, 1\}$ ,  $\forall i \in \{1, \dots, n\}$ . In fact, each row is generated independently from a Bernoulli density whose parameter is specific, conditional upon the fact that  $\sum_{j=1}^n X_j^{(i)} = K$ ,  $i \in \{1, \dots, N\}$ .  $K$ , is a fixed number. It may be as we have already stated, a certain proportion of OD couples, but it may also take into account some other constraints.

The first idea to deal with such a constraint is to generate a random vector  $X_1^{(i)}, \dots, X_n^{(i)}$ . Each component are drawn independently from a Bernoulli density. The sample is accepted iff,  $\sum_{j=1}^n X_j^{(i)} = K$ ,  $i \in \{1, \dots, N\}$ . However, when  $n$  becomes higher than 10, it takes a prohibitive time! In fact, the best solution is to generate independant Bernoulli random variables from  $\mathcal{B}(p_1(t)), \mathcal{B}(p_2(t)), \dots$ , respectively, until  $K$  unities

or  $n - K$  zeros are generated. Then, the remaining elements are put equal to zero or one, respectively. However, the updating formula for the parameters of the Bernoulli densities remain exactly of the form:

$$p_{i,1}(t) = \frac{\sum_{k=1}^N \mathbf{1}_{\{S(X^{(k)}) \geq \hat{\gamma}_t\}} \mathbf{1}_{\{X_i^{(k)} = 1\}}}{\sum_{k=1}^N \mathbf{1}_{\{S(X^{(k)}) \geq \hat{\gamma}_t\}}}, \quad \forall i \in \{1, \dots, n\}. \quad (23)$$

In fact, now,  $X(t)$  and  $Z(t)$  are closely linked. Indeed, if  $Z_i^{(j)} = 0$  then,  $X_i^{(j)} = 0$ , which means that there is no traffic on the OD couple number  $i$  for the  $j^{\text{th}}$ -sample.

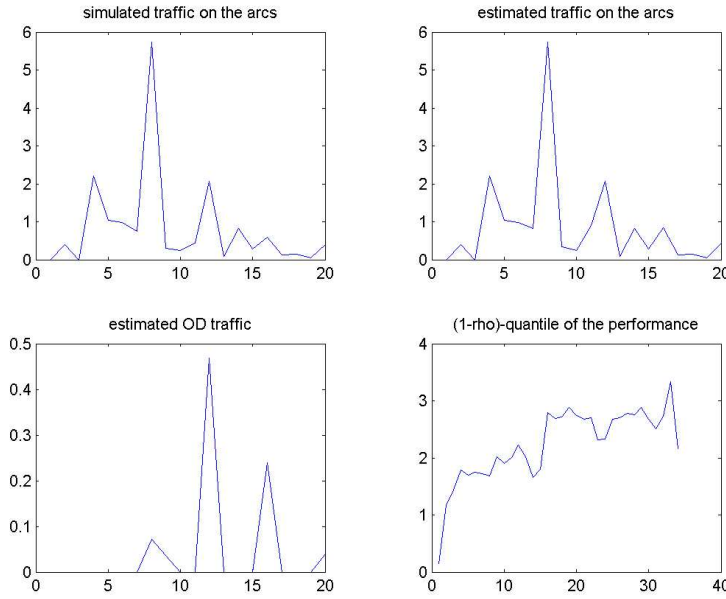


Figure 7: Simulation, the number of zero OD couples is pre-determined

The time required to perform this simulation is of about 2 minutes.

## 5 Dynamic estimation

We have previously determine an estimator of the amounts of traffic flowing through the OD couples at the specific instant  $t$ . We should ask ourselves whether it is possible

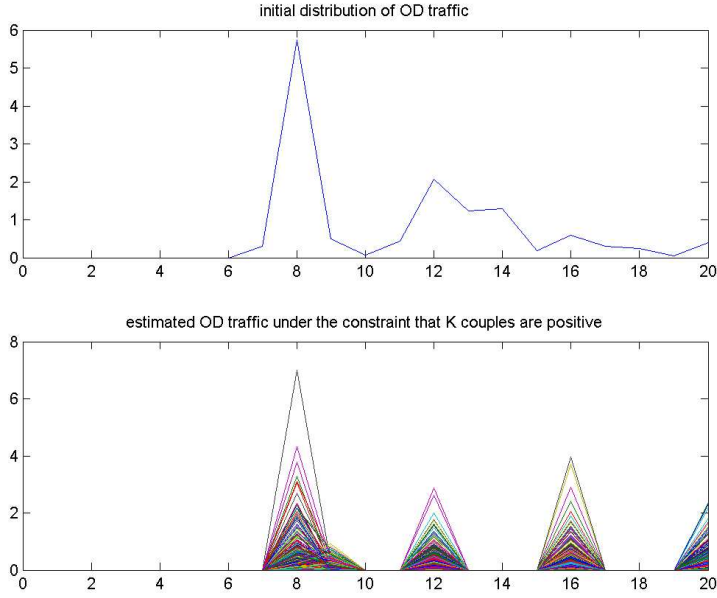


Figure 8: Identifiability of  $X(t)$

to determine the trajectories associated to the vector  $X(t)$ . Particle filtering appear to us to be an interesting approach.

## 5.1 Presentation of Particle filtering

Particle filtering is a well-known technique based on sequential Monte-Carlo approach. It is a technique for implementing a recursive bayesian filter by Monte-Carlo simulations. The key idea is to represent the required posterior density function by a set of random samples with associated weights and to compute estimates based on these samples and weights.

We will generate a random measure  $\{C_{0:k}^i, w_k^i\}_{i=1, \dots, M}$  that characterises the posterior pdf  $p(C_{0:k} | \hat{X}_{1:k})$ .  $\{C_{0:k}^i, i = 0, \dots, M\}$  is a set of vectors with associated weights  $\{w_k^i, i = 1, \dots, M\}$  (the weights are themselves vectors of weights).  $C_{0:k} = \{C_j, j = 0, \dots, t\}$  is the set of all states up to time t. The weights are normalised such that,

$$\sum_{i=1}^M w_k^{i,j} = 1, \forall j \in \{1, \dots, n\}.$$

Then, the posterior density at t can be approximated as:

$$p(C_{0:k} | \hat{X}_{1:k}) \approx \sum_{i=1}^M w_k^i \delta(C_{0:k} - C_{0:k}^i). \quad (24)$$

The weights are chosen using the principle of Importance Sampling. Let  $C^i \sim q(\cdot)$ ,  $i = 1, \dots, M$  be samples generated from a proposal  $q(\cdot)$ , called Importance sampling density. By successive approximations, it is shown in [7] that the weights are recursively obtained by the following formula:

$$w_k^i \propto w_{k-1}^i \frac{p(\hat{X}_k | C_k^i) p(C_k^i | C_{k-1}^i)}{q(C_k^i | C_{k-1}^i, \hat{X}_k)}. \quad (25)$$

It can be shown that as  $M \rightarrow \infty$ , the approximation (24) approaches the true posterior density  $p(C_k | \hat{X}_{1:k})$ .

However, there is a major drawback to use particle filtering techniques. Indeed, a common problem is the degeneracy problem. After a few iterations, all but one particle will have negligible weight. It has been shown that the variance of the importance weights can only increase over time, and thus it is impossible to avoid the degeneracy phenomenon. This degeneracy implies that a large computational effort is devoted to updating particles whose contribution to the approximation to  $p(C_k | \hat{X}_{1:k})$  is almost zero.

Consequently, we have to use resampling mechanisms. The basic idea behind resampling is to eliminate particles which have small weights and to concentrate on particles with large weights.

## 5.2 State model and observation equation

The traffic flow will be modelled as a stochastic hybrid system with discrete states. The observation equation is rather simple to get. Indeed, we have:

$$\hat{X}_k = \Xi(C(k)). \quad (26)$$

Where,  $\Xi$  is a quite complex function which represent the whole algorithm.

The difficulty now, is to build a state model. Suppose the flow can be decomposed in small entities (for example packets). We note:  $Q_{i,k}$ , the number of packets going out of the arc  $i$ , during the time interval  $[t_k, t_{k+1}[$ .  $Q_{i-1,k}$ , is the number of packets arriving on the arc  $i$ .

The conservation of the flow lets us write:

$$Y_i(k+1) = Y_i(k) + \sum_{\{arcs\ l|l \rightarrow i\}} Q_{l,k} - \sum \{arcs\ l'|i \rightarrow l'\} Q_{l',k}. \quad (27)$$

In fact,

$$Q_{i,k} = \min(S_{i,k}; R_{i,k+1}) . \quad (28)$$

$S_{i,k}$  is called *sending function*. It expresses how many among the  $Y_i(k)$  packets in the arc  $i$  at  $k$  are at a distance less than a fixed boundary called  $\beta$ . Suppose the interaction between the packets is negligible and their location is uniformly distributed over the arc.  $S_{i,k}$  is then a random binomial variable with  $Y_i(k)$  drawings, with probability of success  $\frac{\beta}{\text{arc } i \text{ length}}$ , or an approximation, since we don't know exactly the length of the arc number  $i$ .

The *receiving function* is defined by:

$$R_{i,k+1} = \sum_{\{\text{arcs } l | i \rightarrow l\}} [Y_l^{max}(k) + Q_{l,k+1} - Y_l(k)] . \quad (29)$$

The *sending function* is calculated at first by forward recursion, and we substitute  $Q_{i,k} = S_{i,k}$  in (27). With this first guess of the amount of traffic in arc  $i$ , at time  $t_{k+1}$ , a first guess of the receiving function can be computed, recursively. Finally, we get:

$$C_i(k+1) = F(Y_i(k+1)) = \Psi(Y_i(k), W(k+1)) , \forall i \in \{1, \dots, n\}. \quad (30)$$

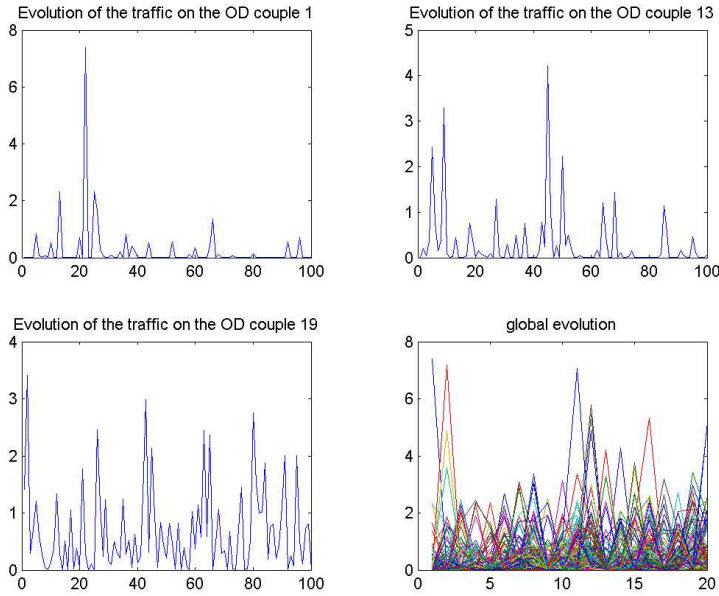


Figure 9: Temporal evolution of the distribution of the traffic on four OD couples

## 6 Conclusion

We could observe that the performance of the CE method is directly proportionnal to the ratio:

$$\frac{|\text{set of nodes}|}{|\text{set of OD couples}|} \quad (31)$$

For rather small networks, eg. networks composed of at most 30 nodes, the CE method works pretty good and suprisingly fastly. What's more, it is possible to add some constraints which could guarantee the identifiability of the vector containing the amounts of traffic on the OD couples. At the end of the estimation part, we get estimators of OD volumes and implicitly, of the routing matrix. In fact, this application is a great proof of the simplicity and versatility of the CE method.

However, some points remain difficult to tackle. For example, when the ratio becomes larger than 19, the CE method performs rather poorly. Furthermore, R. Rubinstein recommand that the sample size of the CE algorithm should be of the form:

$$N = \kappa n, 5 \leq \kappa \leq 10 .$$

Suppose for example, that that we have to deal with a network of 50 nodes. Then, at each step of the algorithm we will have to generate a sample of  $2450 * 20000$  vectors. Which is completely impossible due to the limited capacities of our computers. But, it is certainly possible to improve the algorithm so as to adapt dynamically the sample size

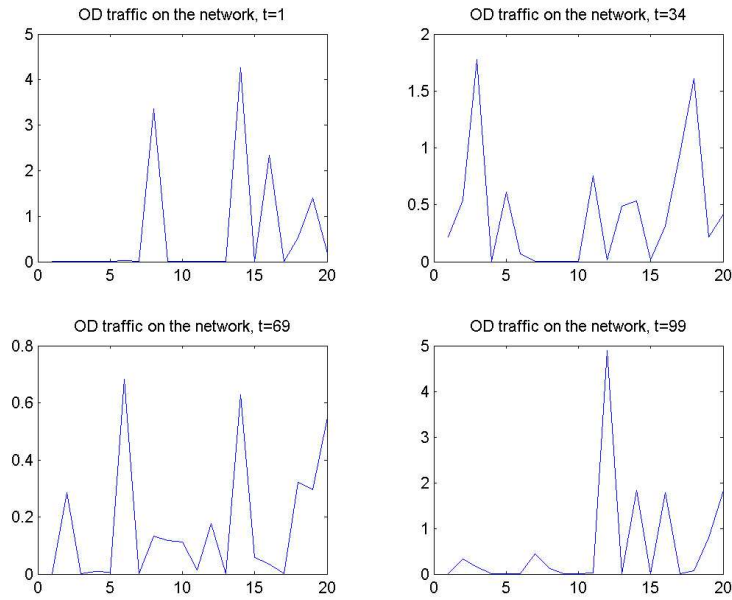


Figure 10: Spatial evolution of the distribution of the traffic at four different instants

to solve this problem. Nevertheless, the question remains open. Fortunately, in every network, some specific constraints need to be taken into account. These constraints aim at decreasing the number of unknown parameters. The idea to impose that some OD couples remain equal to zero is an approach, but there are many others. For example, we may want to maximize the global entropy, or other common criteria.

Particle Filtering is an efficient and subtle technique to dynamically predict the evolution of the distribution of the traffic on the OD couples for rather small networks.

The approaches we use are rather simple to put in application. Nevertheless, they rely on many small parameters which are quite difficult to optimize. Furthermore, the size of the network is still a problem and may be the next challenge of this reflexion.

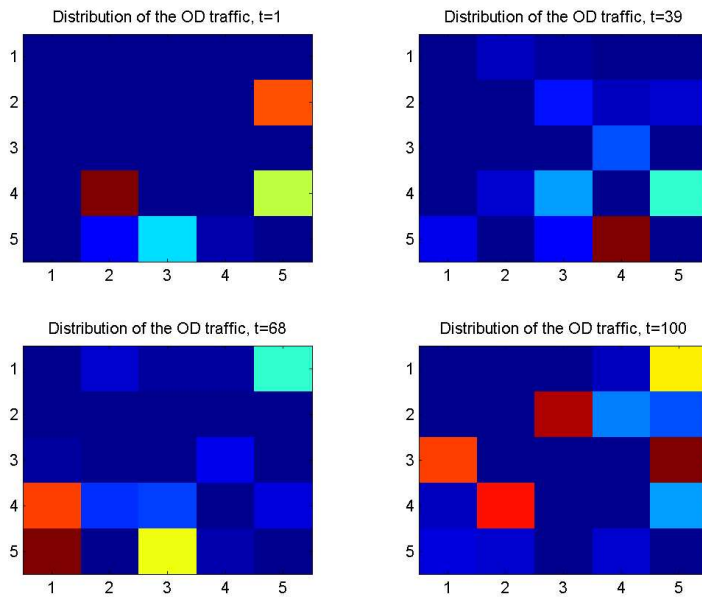


Figure 11: Spatial distribution of the traffic at four instants

## 7 References

- [1] RUBINSTEIN Reuven Y., KROESE Dirk P., *The Cross-Entropy Method*, Springer, 2004.
- [2] HU T.C., SHING M.T., *Combinatorial Algorithms*, Dover Publications, second edition, 2002.
- [3] SCHRJVER A., *Theory of Linear and Integer Programming*, John Wiley, 2000.
- [4] RARDIN R. L., *Optimization in Operations Research*, Prentice Hall, 1998.
- [5] DOUCET A., MASKELL S., GORDON N., *Particle Filters for Sequential Bayesian Inference*, Tutorial ISIF, 2002.
- [6] MIHAYLOVA L., BOEL R., *A Particle Filter for Freeway Traffic Estimation*.
- [7] ARULAMPALAM S., MASKELL S., GORDON N., CLAPP T., *A Tutorial on Particle Filters for On-line Non-linear/Non-Gaussian Bayesian Tracking*, IEEE, 2001.
- [8] CAMPILLO F., LE GLAND F., *Filtrage Particulaire: quelques exemples "avec les mains" et matlab*, Séminaire "filtrage particulaire", CNES, 2003.