Estimation of Task Persistence Parameter from Pervasive Medical Systems with Censored Data

Yannick Fouquet, Céline Franco, Bruno Diot, Jacques Demongeot and Nicolas Vuillerme

Abstract—This paper compares two statistical models of location within a smart flat during the day. The location is then identified with a task executed normally or repeated pathologically, *e.g.* in case of Alzheimer disease, whereas a task persistence parameter assesses tendency to perseverate. Compared with a Pólya's urns derived approach, the Markovian one is more effective and offers up to 98% of good prediction using only the last known location but distinguishing days of week. To extend these results to a multisensor context, some difficulties must be overcome. An external knowledge is made from a set of observable random variables provided by body sensors and organized either in a Bayesian network or in a reference knowledge base system (KBS) containing the person's actimetric profile. When data missed or errors occurred, an estimate of the joint probabilities of these random variables and hence the probability of all events appearing in the network or the KBS was developed and corrects the bias of the Lancaster and Zentgraf classical approach which in certain circumstances provides negative estimates. Finally, we introduce a correction corresponding to a possible loss of the person's synchronization with the nycthemeral (day *vs* night) zeitgebers (synchronizers) to avoid false alarms.

Index Terms—smart flats for elderly people, pervasive watching, data fusion, censored data persistence parameter, Bayesian networks, knowledge based systems, joint probabilities reconstruction, circular Gumbel distribution

Errare humanum est,
perseverare diabolicum.

1 INTRODUCTION

In numerous neuro-degenerative diseases, post-brain stroke or post-heart failure disorders, one can meet temporo-spatial disorientation [1], [2], [3], leading to many errors during the execution of daily tasks[4] until observing a pathologic perseveration [5], *i.e.*, an abnormal repetition of already successful performed tasks (*e.g.*, a pathologic recurrence or "kyrie" of buying successively the same object) which causes a deep handicap in fulfilling current vital functions. It is generally accepted that early and accurate diagnosis of neuro-degenerative pathologies, like Alzheimer disease (AD), is critical for improving their quality of life [6], [7]. The main idea of this paper is to develop an easy procedure to acquire, process and interpret surveillance at home data in order to get a reliable task persistence parameter useful as perseveration index for triggering alarms and/or starting an early diagnostic search for neuro-degenerative pathologies like AD. That implies an adapted activity recording involving a multitude of sensors of very different natures (including infrared, radar, sound, accelerometer, temperature, etc.) both in the flat [8], [9], [10], [11], [12], [13], [14], [15] (Figure 1) and embedded on the person [16], [17], [18], [19] (Figure 2). Hence, an individual nycthemeral actimetric profiles [14] may be drawn and compared to mean canonical profiles of clusters grouping samples of reference cases accounting for the actimetric variability in a population. In order to query the reference profile matching the best with an individual one [20], [21], [22], [23], [24], [25], [26], [27], we query it in a adequately modelled data base permitting the search under hybrid criteria (qualitative, corresponding to medico-socio-economic data about the environment of the surveyed person as well as quantitative, *e.g.* those provided by localization sensors).

The reference data request is made easier by defining an ontology from the concepts underlying the observed variables like dependence index, frailty score [28], memory performance, as well as social class, type of familial or medico-social helpers, economic resources, etc. This ontology allows to build a knowledge based system (KBS), *i.e.*, a program for generalizing and rapidly querying a knowledge base, which is a special kind of database for knowledge management. For taking alarm decisions after querying and matching information from a KBS, a Bayesian network is used. It is represented by a directed acyclic graph representing dependencies embodied in given joint probabilities distribution over a set of random variables expressing uncertainty inside the KBS.

One of the main functionalities of KBS and Bayesian networks is to properly define and organize thanks to an ontology, the concepts to which a given variable, object or notion are related. These concepts are described by a set of qualitative (Boolean or discrete) or quantitative (discrete or continuous) variables. This allows decisions of expert type [29], [30], [31], *e.g.* by assigning an object to a class of concepts in the context of a classification problem, or to find all objects belonging to a concept or obeying an assertion in the context of querying a

• *Y. Fouquet, C. Franco, J. Demongeot and N. Vuillerme are with AGIM Laboratory FRE 3405, CNRS UJF UPMF EPHE, Faculty of Medicine, 38700 La Tronche - France. E-mail: FirstName.Surname@agim.eu; Phone: +33 4 56 52 01 08; fax: +33 4 76 76 88 44;*
B. Diot and C. Franco are with IDS SA, 71300 Montceau-les-Mines, France. E-mail: b.diot@ids-assistance.com

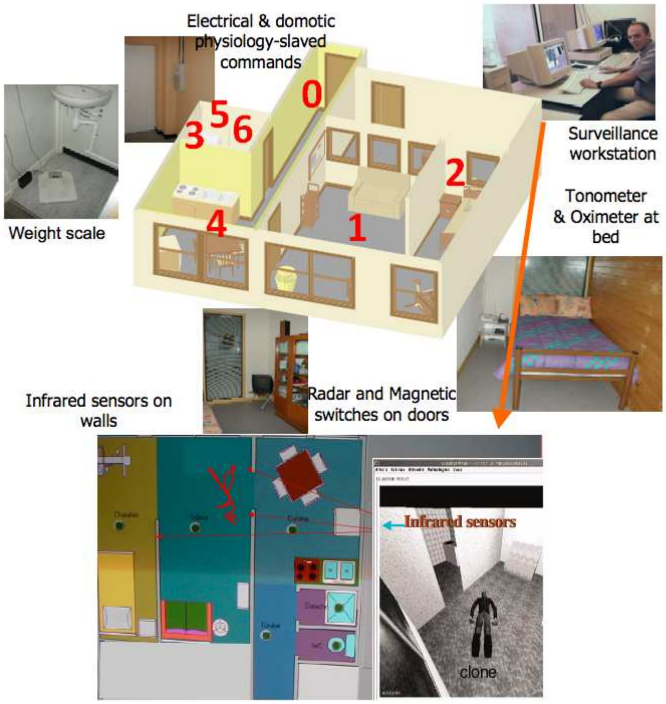


Fig. 1. Location sensors are placed at different places in the apartment to monitor the individual's successive activity phases within his/her home environment: **0.** Entry hall - **1.** Living room - **2.** Bedroom - **3.** WC - **4.** Kitchen - **5.** Shower - **6.** Washbasin.

Fig. 2. Body sensors located on smart clothes for watching up the physiologic state of frail persons in or out their home

knowledge base. If the description of concepts is done from censored, missing or uncertain data, we talk about a random classification problem. This kind of problem is based on the estimation of the joint probabilities distribution corresponding to the observations of random variables used to describe the concepts, identified as events of the σ-algebra generated by these random variables. A Generalized Data Warehouse (GDW) is a particular KBS structuring data through the σ-algebra generated by the random variables defining its assertions [32]. An atom of this σ-algebra is called an equi-class. Each union of equi-classes is called a view. Each view is then the disjoint union of intersections of atomic events called equi-classes in [32] (or primary assertions in a KBS) or of their complements (contraposed primary assertions). As in contingency tables, certain equi-classes can be unobserved due to censored, missing or falsely updated data. Then, it is necessary to estimate the uncertainty of these equi-classes, and after of the events containing these equi-classes.

We define in Section 2 the persistence indexes either as the number (supposed constant in time) k_j ($k_j \geq -1$) of balls added in a Pólya's urn after pulling a ball of a given color j, or as the recursivity order p ($p \geq 0$) in a Markov chain in which the variable X_i depends on p previous one. The Markov chain order is determined by speech recognition techniques adapted to predict the location of a person from geo-localization data. In Section 3, an application of the persistence indexes from location data of a home-dwelling individual is presented. In Section 4, we describe the solutions we proposed to overcome difficulties at each step of the data processing of real data in a multisensor context. If there are missing, censored or false data concerning the events built from the observation of both vertical and horizontal random variables, we remark in Section 4.1.1 that these events can be defined as union of atomic events or equi-classes, corresponding to intersections of marginal events involving only one variable. A classical approach due to Lancaster and Zentgraf (LZ), based on the treatment of missing and censored data in contingency tables, permits to reconstruct the probability of any equi-class in the context of discrete variables, without passing through a distribution kernel estimation or a reconstruction of inter-variable dependences through methods like the logistic regression [33], [34], methods more convenient in case of quantitative variables. However, the LZ approach provides in certain cases (especially when the marginal events are dependent in an exclusive way) negative estimates. In Section 4.1.2, we hence propose a new estimator based on the respect of both the positivity of joint probabilities and the projectivity equations of the marginal distributions. We show that this new estimator gives better estimates for joint probabilities than the previous LZ approach, especially in the case of disjoint dependence between events. We prove in Section 4.1.3, that the new estimator maximizes an entropy variational criterion and in Section 4.1.4, we give some numerical examples of respective use of classical and new estimation methods. In Section 4.1.5, we describe an optimized strategy for giving the most realistic value to any joint probability, from the knowledge about the marginal and order 2 joint empirical frequencies (supposed known and not falsed by censoring, missing or badly updating data). We show how the marginal and second order joint frequencies can be initialized (resp. incrementally updated) according to a Bayesian network or KBS *a priori* (resp. new) information, allowing the estimation of any higher order joint probability. In section 4.2, we give an example of multisensor surveillance involving a vertical sampling. Finally, we give in Section 4.3 a procedure to take into account a possible phase shift between consecutive days showing the same sequence of tasks along the daily activity, but shifted in time without any pathological signification except a change in the sleeping clock, causing a loss by the elderly people of their synchronization with the nycthemeral zeitgebers (synchronizers), like meals or social activities.

2 Indexes of persistence from actimetric data

Among the possible approaches for modelling the actimetric data, two methods have been selected. The first one focuses on the Pólya's urns [35], [36], [37], [38] in which the observed activity at time t is depending on the whole past (since a reset supposed to be made at the beginning of each day). The second one concerns a first order Markov chain approach [39], [40] in which the dependency of the future of t lies only through the present time t. In both models, a persistence parameter is defined. For deciding between these two methods, we propose to use the statistics equal to the empirical mean E of a task remaining (at time t) duration, by identifying a task with the location at which it is performed.

2.1 Pólya's urns

In the Pólya's urns approach, the location is seen as a colored ball. Each second, a ball is taken from an urn. The balls contained in the urn represent the distribution of probabilities of each location. To take into account the persistence in tasks, some balls - from the same color as the one taken - are added in the urn.

The main idea is to considerably simplify the information by giving a color coding number to the different locations (pertinent for the watching), and to follow up the succession of these numbers, *e.g.* by interpreting them as the succession of colors of balls drawn from the urn. The persistence (or *a contrario* the instability) of an action in a location is represented by adding (or taking away, if $k_i(t) < 0$) $k_i(t)$ balls of color i, when a ball of color i has been obtained at time t.

In this approach, the persistence in task i is equal to the parameter $k_i(t)$ normalized by the initial content size of the urn b_0 and denoted $\pi_i(t)$: $\pi_i(t) = \frac{k_i(t)}{b_0}$.

In the following, for the sake of simplicity, we suppose $k_i(t)$ constant in time and thus $\pi_i(t)$ too. By denoting $x_i(t)$ the number of times where the ball of color i has been drawn from the urn until time t, and $p_i(t)$ the probability to get a ball of color i at the $(t+1)^{\text{th}}$ drawing, we have: $p_i(t) = \frac{p_i(0)+x_i(t)\pi_i}{1+t\pi_i}$.

We can estimate π_i from the empirical frequencies $f_i(t)$'s to get a ball of color i at the $(t+1)^{\text{th}}$ drawing (estimated in a series of days supposed to be independent), whose expectation is $p_i(t)$:

$$f_i(t) = \frac{f_i(0) + x_i(t)\overline{\pi_i}}{1+t\overline{\pi_i}} \text{ and } \overline{\pi_i} = \frac{f_i(0) - f_i(M)}{Mf_i(M) - x_i(M)}$$

where M is the total number of drawings by day.

We can also calculate two estimators of the i^{th} task remaining duration E_i.
The first estimator $\overline{E_{i,1}}$, consists in calculating the probability $c_{i,m}(t)$ to have m consecutive drawings of a ball i from the drawing t:

$$\begin{aligned}
\forall m \in \mathbb{N} \quad &: \quad 0 \le m \le (M-t), \\
c_{i,m}(t) \quad &= \quad \Big(1 - p_i(t+m+1)\Big) \cdot \prod_{j=0}^{m} p_i(t+j) \\
&= \quad \Big(1 - \frac{p_i(0) + x_i(t+m+1)\pi_i}{1+(t+m+1)\pi_i}\Big) \\
&\qquad \cdot \prod_{j=0}^{m} \frac{p_i(0)+x_i(t+j)\pi_i}{1+(t+j)\pi_i}
\end{aligned}$$

with : $p_i(M+1) = 0$

The estimator $\overline{E_{i,1}}$ could then be calculated by replacing the probabilities by the corresponding empirical frequencies: $E_i = \frac{1}{M+1}\sum_{t=0}^{M}\sum_{m=0}^{M} m \cdot c_{i,m}(t)$ Thus,

$$\overline{E_{i,1}} \approx \frac{1}{M+1}\sum_{t=0}^{M}\sum_{m=0}^{M} \quad m \cdot \Big(1 - \tfrac{p_i(0)+x_i(t+m+1)\overline{\pi_i}}{1+(t+m+1)\overline{\pi_i}}\Big) \cdot \textstyle\prod_{j=0}^{m} \tfrac{p_i(0)+x_i(t+j)\overline{\pi_i}}{1+(t+j)\overline{\pi_i}}$$

$$\overline{E_{i,1}} \approx \frac{1}{M+1}\sum_{t=0}^{M}\sum_{m=0}^{M} m \prod_{j=0}^{m} f_i(t+j)\Big(1 - f_i(t+m+1)\Big)$$

The 95%-confidence interval of $\overline{E_{i,1}}$ could then be calculated by estimating the 95%-confidence interval of the f_i's which is : $\left[f_i \pm 1.96\sqrt{\frac{f_i(1-f_i)}{M}}\right]$

The null-hypothesis H_0 : "the persistence model is a Pólya's urn model" is rejected if $\overline{E_{i,1}}$, does not belong to this interval. Otherwise, this model could be used to represent the persistence in task.

The second estimator $\overline{E_{i,2}}$ is calculated by considering the empirical mean (on observed days) of the remaining duration in a day which is defined by:

$$\overline{E_{i,2}} = \frac{1}{M+1}\sum_{t=0}^{M} z_i(t),$$

where :

- $y_i(t) = x_i(t) - x_i(t-1)$ is the number (1 or 0) of balls of color i drawn at time t,
- $z_i(t) = \max_{0 \le m \le (M-t)}\{m | \prod_{j=0}^{m} y_i(t+j) = 1\}$ is the length of the sequence of "drawing a ball of color i" (possibly 0) since a drawing at time t of a ball of color i.

2.2 Markov model

In the Markov chain approach, each location is a node with probabilities of transitions from one location to another one. The succession of locations is seen as a route in a Markov chain. A first order Markov chain takes into account the last location in order to predict the present one. The generalization of such a model offers to represent the probability of a location depending on the historic of locations.

In this approach, let us denote by p_{ij} the probability (supposed to be constant) to draw a ball of color j after a ball of color i. Then p_{ii} could be the persistence in task i parameter. If we denote by p_j the probability (supposed to be constant) to draw a ball of color j, we have: $p_j = \sum_{i=1}^{k} p_{ij}$, where k is the number of colors (*i.e.* of types of task).

Moreover, by noticing that the variable $z_i(t)$ has a distribution independent of t we have: $P(z_i = 0) = (1 - p_i)$ and :

$$\forall l \in \mathbb{N}: \ 1 \leq l \leq M, \ P(z_i = l) = p_i(1 - p_i)(p_{ii})^{l-1},$$

then the expectation of the i^{th} task remaining duration E_i could be calculated as: $E_i = \sum_{l=0}^{M} \frac{(l+1)}{2} P(z_i = l)$

Thus, E_i can by estimated by :

$$\overline{E_{i,3}} = \sum_{l=0}^{M} \frac{l+1}{2} f_i(1 - f_i)(f_{ii})^{l-1}$$

The 95%-confidence interval of $\overline{E_{i,3}}$ could be calculated by estimating the 95%-confidence interval of the f_i's and f_{ii}'s which are respectively :

$$\left[f_i \pm 1.96 \sqrt{\frac{f_i(1 - f_i)}{M}} \right] \text{ and } \left[f_{ii} \pm 1.96 \sqrt{\frac{f_{ii}(1 - f_{ii})}{M}} \right]$$

The 95%-confidence interval could also be more accurate by empirically calculus using $\min$ and $\max$ values of $\overline{E_{i,3}}$: $\left[\min_{1 \leq i \leq l} \left(\overline{E_{i,3}} \right) \ldots \max_{1 \leq i \leq l} \left(\overline{E_{i,3}} \right) \right]$

The null-hypothesis H_0: "the persistence model is a first order Markov chain model" is rejected if $\overline{E_{i,3}}$ does not belong to this interval. Otherwise, this model could be used to represent the persistence in task.

If these two tests are concluding to the acceptation, one prefers the first order Markov chain due to its simplicity. If both tests above are concluding to the rejection of the null-hypothesis, we retain the model having the closest distance between $\overline{E_{i,1}}$ and the confidence interval of $\overline{E_{i,j}}$ $(j = 2, 3)$.

Determination of the Markov chain order

A statistical method has been implemented to predict the next location on the basis of the location history [41]. Currently, n-grams location probabilities are used to compute the most likely follow up location. To predict the i^{th} location a_i, we use the $n - 1$ previously uttered locations and determine the most probable location by computing:

$$a_i = argmax_a P(a | a_{i-1}, a_{i-2} \ldots, a_{i-n+1})$$

To estimate this probability, relative frequency techniques are employed.

Otherwise, in many real-situations, it was not possible to collect a large amount of data to properly estimate the statistics. This implies that it is not reasonable to use classical smoothing techniques. We need a solution for the two following problems:

1) unexpected input: the location model based on n-grams location sequences can not be used in case unexpected input occurs,
2) lack of training data: the n-grams model predict several locations with the same probability.

The treatment of these cases consists in using the $(n-1)$-grams model, recursively. Once the order of the Markov chain is determined, the associated transition matrix may be approximated by the empirical frequencies. Another way to quantify the perseveration in behavior may be to calculate the entropy of the trajectories as [42]: $H_M = \sum_{i=1}^{k} \sum_{j=1}^{k} \pi_i \cdot m_{i,j} \cdot \log_2(m_{i,j})$ where π is the stationary distribution and $m_{i,j}$ are the coefficients of the transition matrix. Weak values of entropy correspond to very regular patterns whereas high values depict a varied behavior. A decrease in entropy may be interpreted as a loss of diversity in the accomplishment of activities of daily living in relation with perseveration.

3 Preliminary experiment

3.1 Materiel and methods

Since 12 years, many experiments have been conducted for watching dependent people at home, in particular elderly and handicapped persons [43], [9], [14], [44], [45], [46]. Some of important things to be done are localizing a person. For acquiring data necessary to permit this localization, various sensors haven been invented. This sensors networks permit to represent the location of a person in a flat room (Figure 1). Recording timestamped locations permits us to create a corpus for experiments [10].

The corpus describes the location of an elderly person within his/her home environment in time. It is on the form of a timestamped location. Timestamps are space separated numerals representing day of month, month, year, hour, minutes, seconds of the location captured. The location itself is a code (*cf.* Figure 2). Note that the activity-station-code (9) corresponds to an error. An example of a line of the corpus is $18\ 07\ 2007\ 11\ 27\ 48\ 4$, which suits as : on 07/18/07, at 11:27,48", subject was in the kitchen. The files treated bring together the data recorded in the flat of the elderly people in a period of 10 months from the 03/22/05 until the 01/24/06 and a period of 6 months from the 07/18/07 to the 01/15/08.

For this experiment, the corpus has been shapped up to represent the location of the person, each second. A line of this 'new' corpus represents a day as a series of location, each second. It is on the form of a space separated locations as a code as explained above. For example, "$s\ 2\ 2\ 2\ \ldots\ 2\ 2\ 3\ 3\ 3\ \ldots\ 3\ 3\ 4\ 4\ 4\ \ldots\ e$" suits as

TABLE 1
Good prediction rate (%) depending on day and for the whole corpus

n	mon	tue	wed	thu	fri	sat	sun	total
1	58.00	57.99	59.79	64.25	60.65	63.89	61.61	61.07
2	90.65	**92.32**	91.57	91.87	**93.36**	**92.51**	91.39	92.01
3	90.71	92.27	**91.67**	91.78	93.34	92.01	**91.73**	91.99
4	**90.82**	92.07	91.44	**91.91**	93.23	91.97	91.59	**92.07**
5	90.58	91.77	91.46	91.53	92.88	92.00	91.54	92.02
6	90.11	91.64	91.06	91.35	92.61	91.81	91.28	91.83
7	89.97	91.41	90.91	91.10	92.37	91.45	91.08	91.67
8	89.80	91.20	90.51	90.92	92.21	91.26	90.91	91.50
9	89.75	91.00	90.22	90.68	92.17	91.02	90.80	91.35
10	89.57	90.82	90.15	90.58	92.20	90.88	90.78	91.21

: since s the start of day, the person was in the bedroom (2), after x seconds (x is the number of successive 2), the person passed in the toilet (3), then after y seconds (y is the number of successive 3), she passed in the kitchen (4), etc. The close of day is represented by e.

The n-grams model was applied with $(n-1)$ last minutes used to predict the n^{th} one. We choose to set n up to 10 so that we watch for the 9 last minutes in order to predict the 10^{th}. The corpus has been cut into 80% for learning model, 20% for testing it. Tests have been done for an history of location set from 1 to n (10 here).

3.2 Results and discussion

3.2.1 Prediction performance

A first test was made with the whole corpus without date distinction (day of week, day of month, month, hour of day, etc.). Last column of table 1 shows a best prediction with $n = 4$. Indeed, approximatively the same performance is obtained with $n > 4$ but n does not need to be bigger than 4. The last three minute location is sufficient to predict the next one. After that, raw performance seems to increase with n. This result seems to indicate that accuracy by watching too far in the past is not a good way to predict the future location of a person.

A second test was made by distinguishing the day of the week to take into account regular outdoor-activities. Table 1 shows a best prediction approximated rate with $n = 4$ (using the last three locations to predict the next one). Performance seems to decrease with n increasing. The real best performance, in bold, shows that results differ according to day of week but a good approximation could be made with $n = 4$. Moreover, with $n = 4$, results of good prediction differ according to the day of week from 90.82% on Monday to 93.23% on Friday. It seems to show that day of week is an important factor of variation.

First results tend to show best performances occurring in the first order Markov case with $n = 4$, and a degradation of performances with n increasing up to 10. This seems to indicate that watching more far in time is more accurate but a bad way to predict the future location of the person.

TABLE 2
Empirical frequencies f_i (%)

0	1	2	3	4	5	6	9	s
9.72	23.87	**50.69**	0.89	6.29	1.16	5.77	1.35	0.27

TABLE 3
Empirical frequencies f_{ij} (%)

	0	1	2	3	4	5	6	9	s
0	**91.34**	4.61	1.35	0.06	1.52	0.26	0.68	0.18	0.00
1	1.78	**75.85**	6.01	0.53	10.58	0.63	4.00	0.61	0.00
2	0.20	2.72	**94.42**	0.43	0.60	0.17	1.18	0.28	0.00
3	1.15	12.24	21.22	**39.30**	4.46	4.76	16.88	0.00	0.00
4	2.09	40.02	4.18	0.60	**47.74**	0.63	4.03	0.70	0.00
5	3.04	17.50	9.12	2.99	3.82	**44.27**	16.77	2.49	0.00
6	2.13	16.67	10.14	1.89	3.22	4.52	**55.44**	5.98	0.00
9	1.79	10.60	10.32	0.00	2.70	3.10	25.76	**45.73**	0.00
s	3.31	11.70	66.47	3.90	2.14	1.17	7.99	3.31	**0.00**

Moreover, the performance seems to differ for each day of week. This factor of variability should be taken into account when designing a system using a location model. Future experiments should be conducted for other comparisons. The distinction of each day of the month could show that some days, as 1st day of the month for example, are particular. The comparison between each month could show different activities in summer and in winter, and so on.

It could then be interesting to develop a new model with a continuum approach considering estimations (interpolation) between data observed.

3.2.2 Measures of perseveration in task

The model of the location of a person seems to be well approximated by a Markovian process. A first order Markov chain is sufficient in order to represent the probabilities of transitions from locations to other locations. The empirical means $\overline{E_{i,j}}$ of tasks remaining duration should now be calculated.

The Table 2 (respectively 3) shows the frequencies f_i (respectively f_{ij}) empirically calculated from the 20% learning part of the corpus.

As above-mentioned, M is the number of locations recorded during a day. The sampling frequency is 1 second. Thus, $M = 60 \times 60 \times 24 = 86400$. E_i can by estimated by : $\overline{E_{i,3}} = \sum_{k=0}^{86400} \frac{k+1}{2} f_i(1-f_i) f_{ii}^{k-1}$

The mean of remaining time in task i, $\overline{E_{i,1}}$, consists in calculating, for each observing time t, the time remaining in task i, divided by the number of times observed (which is equal to $M+1$ if the observation start from 0 to M). It expresses persistence in task i, but is not equal to the mean of past time in i (it should be half the preceding one).

One can now distinguish two particular cases.
If i was never observed : $\overline{E_{i,1}} = 0$
If i was always observed : $\overline{E_{i,1}} = \frac{\frac{(M+1)(M+2)}{2}}{M+1} = 43201$

For the other cases, some works have to be done now in order to calculate $\overline{E_{i,1}}$. It should be calculated for the Pólya's urns approach and for the Markov chain approach. Then, it could verify each hypothesis.

If $\overline{E_{i,1}}$ is in the confidence interval of $\overline{E_{i,3}}$, then we should use this Markovian model due to its simplicity (despite Pólya's urns approach is available [35]). If it is not the case, the same work has to be done with the Pólya's urns approach.

The complexity of the trajectories in the Markov model, their entropy is $H_M = 0.889$ throughout the week. In this experiment, no difference in entropy was observed depending on the day of the week.

4 Towards an application to real data in a multisensor context

Calculating reliable persistence indexes in a multisensor context requires to deal with constraints inherent in real-environment as illustrated in 3.

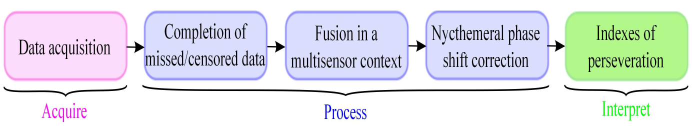


Fig. 3. Procedure proposed to deal with real data in a multisensor context

4.1 Censored data and the estimation of joint probabilities

4.1.1 *The classical Lancaster-Zentgraf approach*

Let us define $\{A_i\}_{i=1,\ldots,n}$ the set of events (resp. assertions) structuring a Bayesian network (resp. a KBS). We suppose that the A_i's are obtained by knowing m real random variables $\{X_k\}_{k=1,\ldots,m}$, defined on a set Ω, *e.g.* $A_i = \{X_i < t_i\}$. Let us consider now a GDW (considered as a multidimensional contingency table or generalized contingency tensor) structuring data through the σ-algebra generated by the A_i's. An atomic event of this σ-algebra is called an equi-class. Each union of equi-classes is called a "view". Each view is then the union of disjoint intersections of events like A_i's and A_i^c's (where $A_i^c = \Omega \backslash A_i$). The events A_i and A_j are in mutual independence if $P(A_i \cap A_j) = P(A_i)P(A_j)$, in inclusive dependence if $A_i \subseteq A_j$ or $A_j \subseteq A_i$, and in exclusive dependence if $A_i \cap A_j = \emptyset$.

In order to estimate in a multidimensional contingency table, including censored and missing data, the joint probabilities of the intersection of n events, Lancaster defined the non-interaction of order 3 [47], generalizing the mutual independence between three events. Then Zentgraf [48] proposed an estimate of the joint probability of order n, from marginal and joint probabilities of order 2 given by the following definition formula:

$$P_{\text{Lan}}\left(\bigcap_{i=1}^{n} A_i\right) = \sum_{\substack{i,j,k_1,\ldots,k_{n-2}\in\{1,\ldots,n\} \\ i\neq j\neq k_1\neq\ldots\neq k_{n-2}}} P(A_i \cap A_j)P(A_{k_1})\ldots P(A_{k_{n-2}}) - (C_n^2 - 1)\prod_{i=1}^{n} P(A_i)$$

The calculation of the above Lancaster-Zentgraf (LZ) estimator involves the knowledge of the marginal and of the order 2 intersection probabilities. For $n = 3$, the equation above becomes:

$$P_{\text{Lan}}(A \cap B \cap C) = P(A \cap B)P(C) + P(A \cap C)P(B) + P(B \cap C)P(A) - 2P(A)P(B)P(C).$$

P_{Lan} satisfies the projectivity property:

$$P_{\text{Lan}}(A \cap B \cap C) + P_{\text{Lan}}(A \cap B \cap C^c) = P_{\text{Lan}}(A \cap B)$$

The LZ estimate is an exact formula if the events A_i are independent or if all A_i's are equal to Ω (inclusive dependence); in these cases, the definition formula is identical to the classical formula of independence, where P_{Ind} denotes the product of the probabilities:

$$P_{\text{Lan}}\left(\bigcap_{i=1}^{n} A_i\right) = P_{\text{Ind}}\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i)$$

However, the definition formula becomes incorrect in the case of unless 2 disjoint events, where $P_{\text{Lan}}(\bigcap_{i=1}^{n} A_i)$ is in general not equal to 0, and in the case of total exclusive dependence ($\forall i \neq j, A_i \cap A_j = \emptyset$), where $P_{\text{Lan}}(\bigcap_{i=1}^{n} A_i)$ is negative. We will study now simple examples showing circumstances where the estimate is incorrect.

Example 1: Suppose that A and B, and A and C are independent. Then: $P_{\text{Lan}}(A \cap B \cap C) = P(B \cap C)P(A)$. But $P(A \cap B \cap C) = P(B \cap C)P(A)$ if and only if A and $B \cap C$ are independent, hence if the LZ estimate is correct we must have:

$$\left.\begin{array}{l} A,\ B \text{ independent} \\ A,\ B \cap C \text{ independent} \end{array}\right\} \Rightarrow A,\ C \text{ independent},$$

This assertion being false in Example 2:

Example 2: Let us suppose that the events of each couple are mutually independent. Then we have:

$$P_{\text{Lan}}(A \cap B \cap C) = P(A)P(B)P(C)$$

In general, this assertion is false, because the mutual independence between events of any couple is not implying the independence of the whole set of events (if they are 3 or more).

Example 3: Let us suppose that A and B are disjoint. Then we have:

$$P_{\text{Lan}}(A \cap B \cap C) = P(A \cap C)P(B) + P(B \cap C)P(A) - 2P(A)P(B)P(C),$$

since $P(A \cap B) = 0$, but $P(A \cap B \cap C) < P(A \cap B) = 0$ and $P(A \cap C)P(B) + P(B \cap C)P(A) - 2P(A)P(B)P(C)$ is not obligatory equal to 0, as shown below.

Example 4: Let us suppose that A, B and C are disjoint. The definition formula gives:

$$P_{\text{Lan}}(A \cap B \cap C) = -2P(A)P(B)P(C)$$

because $P(A \cap B) = P(A \cap C) = P(B \cap C) = 0$, and the P_{Lan} estimator provides a negative result.

Hence, the LZ definition formula gives a correct estimation only in the cases of independence and of total inclusive dependence, for example when $A \subset B \subset C = \Omega$, where Ω is the whole assertion. For the cases where the definition formula gives an incorrect estimate, we propose an adapted new estimate.

4.1.2 A New Estimation Method

We introduce in this Section a new joint probabilities estimator based on the local equipartition of the amount of uncertainty (corresponding to a local maximal entropy approach). The proposed formula is established to deal with dependences characterized by strong incompatibilities, circumstances not well taken into account by the LZ formula above. The new estimator is called P_{New} and is defined recursively by:

$$P_{\text{New}}(\bigcap_{i=1}^{n} A_i) = \frac{\sum_{j=1}^{n} P_{\text{New}}(\bigcap_{i \neq j} A_i)P(A_j)}{n} \text{, if } n > 2$$

For the intersection of any three events from a set of n events this equation becomes:

$$\begin{aligned} P_{\text{New}}(A_i \cap A_j \cap A_k) = \quad & [P(A_i \cap A_j)P(A_k) \\ + \quad & P(A_i \cap A_k)P(A_j) \\ + \quad & P(A_j \cap A_k)P(A_i)]/3 \end{aligned}$$

In practice, the calculation of P_{New} is done in a recursive way from the calculation of P_{New} on the triplets of events involved in $\bigcap_{i=1}^{n} Ai$. That involves as for the LZ estimator the knowledge of the marginal and order 2 intersection probabilities. We have:

$$P_{\text{New}}(\bigcap_{i=1}^{n} A_i) = \frac{\sum_{i<j} P(A_i \cap A_j) \prod_{k \neq i,j} P(A_k)}{C_n^2}$$

Practically this means that we estimate $P(A_i \cap A_j | A_k)$ by $P(A_i \cap A_j)$, which is in principle avalaible only if the intersection $A_i \cap A_j$ is independent of A_k. Indeed, we have in the case of mutual independence between any intersection of a couple of events and the third event:

$$P(A_i \cap A_j \cap A_k) = P_{\text{New}}(A_i \cap A_j \cap A_k)$$

We check that P_{New} satisfies the projectivity property, if A_1 is independent of A_2:

$$P_{\text{New}}(A_1 \cap A_2 \cap A_3) + P_{\text{New}}(A_1 \cap A_2 \cap A_3^c) = P_{\text{New}}(A_1 \cap A_2).$$

We introduce now three other estimators Σ and Π (resp. Γ) defined as the arithmetic and geometric

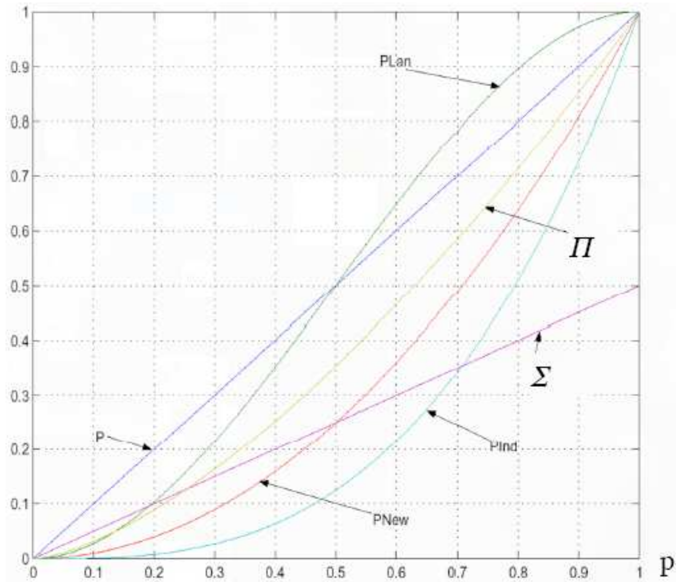


Fig. 4. The different estimators of $P(\bigcap_{i=1}^{n} A_i)$ in case of extreme inclusive dependence

mean of the probabilities $P(A_i \cap A_j)$ (resp. $P(A_i \cap A_j) \prod_{k \neq i,j} P(A_k)$):

$$\Sigma(\bigcap_{i=1}^{n} A_i) = \sum_{i<j} P(A_i \cap A_j)/(2^{n-2}C_n^2);$$

$$\Pi(\bigcap_{i=1}^{n} A_i) = \prod_{i<j} P(A_i \cap A_j)^{1/(n-1)};$$

$$\Gamma(\bigcap_{i=1}^{n} A_i) = \left[\prod_{i<j} P(A_i \cap A_j) \prod_{k \neq i,j} P(A_k)\right]^{1/n}$$

We give hereafter four demonstrative examples of comparison between P_{New}, P_{Lan}, P_{Ind}, Σ and Π.

1. Extreme inclusive dependence
Let us suppose that: $A_1 = A_2 = A_3$, with $P(A_i) = p$ and $P(\bigcap A_i) = p$; then we have:

$$P_{\text{New}}(\bigcap A_i) = p^2; P_{\text{Lan}}(\bigcap A_i) = 3p^2 – 2p^3;$$
$$P_{\text{Ind}}(\bigcap A_i) = p^3; \Sigma(\bigcap A_i) = p/2; \Pi(\bigcap A_i) = p^{3/2}$$

P_{Lan} is ever closer to p than the other estimators, except near 0 where Σ is the better estimator.

2. Extreme exclusive dependence
Let us suppose that: $P(A_1) = \varepsilon$, $P(A_2 \cap A_3) = \eta$; $\forall i = 2, 3, P(A_i) = p, P(A_1 \cap A_i) = \eta^2; P(\bigcap A_i) = \eta^2/3$, with $\varepsilon \approx \eta^{1/2}$, then we have, by neglecting ηp (resp. η^2, $\eta^{5/2}$) with respect to ε (resp. η, η^2):

$$P_{\text{New}}(\bigcap A_i) = \eta\varepsilon/3; P_{\text{Lan}}(\bigcap A_i) = \varepsilon(\eta - 2p^2);$$
$$P_{\text{Ind}}(\bigcap A_i) = p^2\varepsilon; \Sigma(\bigcap A_i) = \eta/6; \Pi(\bigcap A_i) = 0$$

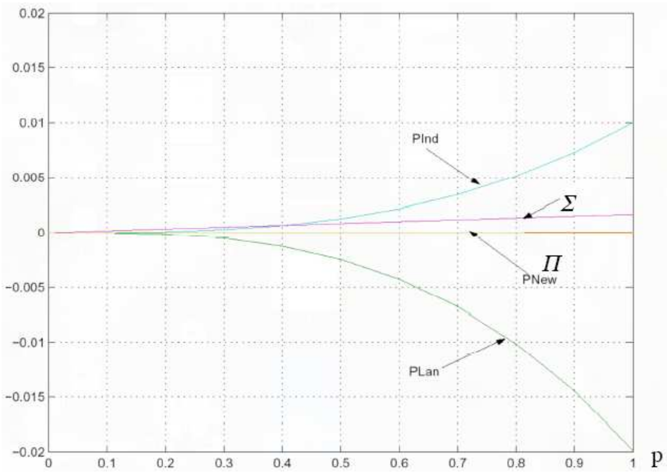


Fig. 5. The estimators of $P(\bigcap_{i=1}^{n} A_i)$ in case of extreme exclusive dependence ($\eta = p/100$)

Because P_{Lan} is negative, we choose in this case P_{New} or Π which are the closest to $P(\bigcap A_i)$.

Let us suppose now that : $\forall i, k, A_i \cap A_k = \emptyset$: $\forall i, P(A_i) = p \leq 1/3$ and $P(\bigcap A_i) = 0$, then:

$$P_{\text{New}}(\bigcap A_i) = 0; P_{\text{Lan}}(\bigcap A_i) = -2p^3;$$
$$P_{\text{Ind}}(\bigcap A_i) = p^3; \Sigma(\bigcap A_i) = 0; \Pi(\bigcap A_i) = 0$$

Because P_{Lan} is negative, we choose in this case P_{New}, Σ or Π, which only give the correct result.

Note that in the case complementary to that of the inclusive dependence, we have, if $B_i = A_i^c$:

$$P(\bigcap B_i) = 1 - \sum P(\bigcup A_i) = 1 - 3p$$

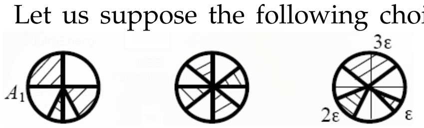
and its estimates are:

$$P_{\text{New}}(\bigcap B_i) = (1-2p)(1-p) = 1 - 3p + 2p^2;$$
$$P_{\text{Lan}}(\bigcap B_i) = 3(1-2p)(1-p) - 2(1-p)^3 = 1 - 3p + 2p^3;$$
$$P_{\text{Ind}}(\bigcap B_i) = (1-p)^3 = 1 - 3p + 3p^2 - p^3;$$
$$\Sigma(\bigcap B_i) = (1-2p)/2; \Pi(\bigcap B_i) = (1-2p)^{3/2}$$

The better estimate is then as expected P_{Lan}.

3. Near exclusive independence

Let us suppose the following choice for A_1, A_2, A_3:



Each A_i has a probability $1/2$ and the circular sectors noted ε have a probability ε (with $\varepsilon \leq 1/24$). Let us suppose that the A_i's verify:

$$P(A_i) = 1/2, P(A_i \cap A_j) = 1/4 - \varepsilon, P(\bigcap A_i) = 1/8,$$

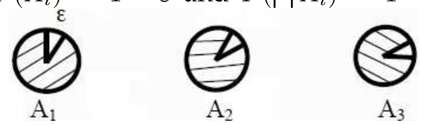
then the estimates of $P(\bigcap A_i)$ are:

$$P_{\text{New}}(\bigcap A_i) = 1/8 - \varepsilon/2; P_{\text{Lan}}(\bigcap A_i) = 1/8 - 3\varepsilon/2;$$
$$P_{\text{Ind}}(\bigcap A_i) = 1/8; \Sigma(\bigcap A_i) = 1/8 - \varepsilon/2;$$
$$\Pi(\bigcap A_i) = (1/4 - \varepsilon)^{3/2}$$

$P_{\text{Ind}}(\bigcap A_i)$ being equal to $P(\bigcap A_i) = 1/8$, we choose in this case P_{Ind}. Note that P_{New} and Σ are here better than P_{Lan}.

4. Near inclusive independence

Let us consider the case where $A_1 = A_2 = A_3$, with $P(A_i) = 1 - \varepsilon$ and $P(\bigcap A_i) = 1 - 3\varepsilon$. Then we have:



$$P_{\text{New}}(\bigcap A_i) = 1 - 3\varepsilon + 2\varepsilon^2; P_{\text{Lan}}(\bigcap A_i) = 1 - 3\varepsilon + 2\varepsilon^3;$$
$$P_{\text{Ind}}(\bigcap A_i) = \prod_{i=1,2,3} P(A_i) = 1 - 3\varepsilon + 3\varepsilon^2 - \varepsilon^3;$$
$$\Sigma(\bigcap A_i) = (1-2\varepsilon)/2; \Pi(\bigcap A_i) = (1-2\varepsilon)^{3/2}$$

$P_{\text{Lan}}(\bigcap A_i)$ being the closest to $P(\bigcap A_i) = 1 - 3\varepsilon$, we choose P_{Lan}.

4.1.3 A Maximum Entropy Principle

In this Section, we will show first that Π satisfies the global maximum entropy, when the n A_i's are close to the mutual independence (independence of each couple of events) and when the $P(A_i)$'s are small.

Proposition 1

Let us suppose that:$\forall i, j, k = 1, \ldots, n, i \neq j \neq k, P(A_i) = \alpha_i, P(A_i \cap A_j) = \beta_{ij} \approx \alpha_i \alpha_j, P(A_i \cap A_j \cap A_k) = \gamma$ and: $\forall m > 3, P(\bigcap_{j=1,m} A_{ij}) = 0$. Then the amount of entropy defined by the partition generated by the A_i's in Ω defined by:

$$H =$$
$$-(1 - \sum_i \alpha_i + \sum_{i<j} \beta_{ij} - C_n^3 \gamma) \log(1 - \sum_i \alpha_i + \sum_{i<j} \beta_{ij} - C_n^3 \gamma)$$
$$- \sum_i (\alpha_i - \sum_{i<j} \beta_{ij} + C_{n-1}^2 \gamma) \log(\alpha_i - \sum_{j \neq i} \beta_{ij} + C_{n-1}^2 \gamma)$$
$$- \sum_{i<j} (\beta_{ij} - (n-2)\gamma) \log(\beta_{ij} - (n-2)\gamma) - C_n^3 \gamma \log \gamma$$

is maximum when $P(A_i \cap A_j \cap A_k)$ is estimated by Π.

Proof: We have:

$$\partial H / \partial \gamma \approx C_n^3 \log(1 - \sum_i \alpha_i + \sum_{i<j} \beta_{ij} - C_n^3 \gamma)$$
$$+ C_n^3 - C_{n-1}^2 \sum_i \log(\alpha_i - \sum_{i<j} \beta_{ij} + C_{n-1}^2 \gamma)$$
$$- n C_{n-1}^2 + (n-2) \sum_{i<j} \log(\beta_{ij} - (n-2)\gamma)$$
$$+ (n-2) C_n^2 - C_n^3 \log \gamma - C_n^3$$

If $\gamma \ll \beta_{ij} \ll \alpha_i \ll 1$, $\partial H / \partial \gamma = 0 \Leftrightarrow C_n^3 \log \gamma \approx (n-2) \sum_{i<j} \log \beta_{ij} - C_{n-1}^2 \sum_i \log \alpha_i \Leftrightarrow \gamma (\prod_{i<j} \beta_{ij})^{6/n(n1)} / (\prod_i \alpha_i)^{3/n} \approx (\prod_{i<j} \beta_{ij})^{3/n(n-1)}$, because the quasi-mutual independence implies: $\prod_{i<j} \beta_{ij} \approx (\prod_i \alpha_i)^{n-1}$. Then for $n = 3$, we have: $\gamma \approx \Pi(\bigcap A_i)$. If we do not neglect $P(\bigcap_{j=1,m} A_{ij})$

for $m > 3$, an analoguous proof shows that the estimator of $P(\bigcap_{i=1,n} A_i)$ maximizing the entropy is still approximatively $\Pi(\bigcap A_i)$. □

In case of independence between $\bigcap_{i=1;i\neq k}^n A_i$ and A_k for any k, we can now prove that P_{New} satisfies the same variational principle for a conditional entropy.

Proposition 2

Let us consider a family of n events $A_1,\ldots, A_n$ and define: $B_k = (\bigcap_{i=1;i\neq k}^n A_i) \cap A_k^c$. The B_k's are disjoint and the conditional entropy H_B on $B = \bigcup_{k=1}^n B_k$ is given by: $H_B = -\sum_{k=1}^n P_B(B_k)\log(P_B(B_k))$, where P_B is the conditional probability knowing B. Then H_B is maximum when $P(A_i \cap A_j \cap A_k)$ is estimated by P_{New}.

Proof: H_B is maximum for $P_B(B_k) = 1/n$, *i.e.* for $P(B_k) = P(B)/n$. Then we have:

$$P(B_k) = P(B_k|A_k^c)(1 - P(A_k))$$

$$P(B_k) = P(\bigcap_{i=1;i\neq k}^{n} A_i) - P(\bigcap_{i=1}^{n} A_i)$$

Hence we deduce: $\forall k = 1,\ldots,n$,

$$P(\bigcap_{i=1}^{n} A_i) = P(\bigcap_{i=1;i\neq k}^{n} A_i) - P(B)/n$$

$$= P(\bigcap_{i=1;i\neq k}^{n} A_i) - \sum_{k=1}^{n} \frac{P(B_k|A_k^c)(1 - P(A_k))}{n}$$

and

$$nP(\bigcap_{i=1}^{n} A_i) = \sum_{k=1}^{n} \left[P(\bigcap_{i=1;i\neq k}^{n} A_i) - P(B_k|A_k^c)\right] + \sum_{k=1}^{n} P(B_k|A_k^c)P(A_k)$$

If we assume the independence between $\bigcap_{i=1;i\neq k}^n A_i$ and A_k for any k, then we get:

$$P(\bigcap_{i=1}^{n} A_i) = \sum_{k=1}^{n} P(\bigcap_{i=1;i\neq k}^{n} A_i)P(A_k)/n = P_{\mathrm{New}}(\bigcap_{i=1}^{n} A_i)$$

□

4.1.4 Some Numerical Examples of Respective Use of the Classical and new Estimation Methods

All estimators above less Σ are exact in case of independence but become false when we are far from this case, some being better than the others in certain circumstances, like inclusive or exclusive dependence, as well as nearly independence. Then we will now compare P_{Lan}, P_{New}, Σ and Π in particular cases of dependence as summarized in the Table 4. The first row of the Table 4 refers to the exemple of extreme inclusive dependence treated above. The second and third rows correspond to the same case with respectively A_1 near Ω and A_1 near $\emptyset$. The fourth row refers to the case nearly independence where:

$$P(A_1) = P(A_2) = P(A_3), P(A_i\cap A_j) = P(A_i)P(A_j)+\eta,$$
$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3) + (P(A_1) + P(A_2) + P(A_3))\eta/3$$

The fifth and sixth rows correspond to the same case with respectively A_1 near Ω and A_1 near $\emptyset$. The seventh row refers to the case of extreme exclusive dependence treated above. The eighth and ninth rows represent the same case with respectively A_1 near Ω and A_1 near $\emptyset$.

We see on the Table 4 that $P_{\mathrm{New}}(A_1 \cap A_2 \cap A_3)$ is often the best estimator of $P(A_1 \cap A_2 \cap A_3)$. More precisely, the above results work out that P_{Lan} is the best estimation in the case of inclusive dependence, whereas P_{New} is the best one in the case of exclusive dependence and is a better estimation than P_{Ind} in the inclusive case. In practice, we need a generalized estimator which, according to the case of dependence, is good as the best among P_{Lan}, P_{New} or P_{Ind}. Let us define $\mathrm{d_{Inc}}$, $\mathrm{d_{Exc}}$ and $\mathrm{d_{Ind}}$ as the measures of the inclusive dependence, the exclusive dependence and the mutual independence deviation, respectively:

$$\mathrm{d_{Inc}}(\{A_i\}_{i=1,\ldots,n}) = \frac{\sum_{i<j}[P(A_i)+P(A_j)-2P(A_i\cap A_j)]}{\sum_{i<j}[P(A_i)+P(A_j)]}$$
$$\mathrm{d_{Exc}}(\{A_i\}_{i=1,\ldots,n}) = \frac{\sum_{i<j} P(A_i\cap A_j)}{\sum_{i<j} \min[P(A_i),P(A_j)]}$$
$$\mathrm{d_{Ind}}(\{A_i\}_{i=1,\ldots,n}) = \frac{\sum_{i<j}|P(A_i\cap A_j)-P(A_i)P(A_j)|}{\sum_{i<j} \max[{P(A_i)P(A_j), \atop \max_{i<j}|P(A_i\cap A_j)-P(A_i)P(A_j)|}]}$$

All above distances vary between 0 and 1. We have: (1) $\mathrm{d_{Inc}} = 0$ when the inclusion is maximal ($A_1 = A_2 = A_3$) and $\mathrm{d_{Inc}} = 1$ when the A_i's are disjoint; (2) $\mathrm{d_{Exc}} = 0$ when the exclusion is maximal (A_i's disjoint) and $\mathrm{d_{Exc}} = 1$ when the A_i's are identical; and (3) $\mathrm{d_{Ind}} = 0$ in case of independence and $\mathrm{d_{Ind}} = 1$ when the A_i's are identical or disjoint. Then the generalized estimator proposed to estimate the joint probabilities is: $P_* = a_1P_{\mathrm{Lan}} + a_2P_{\mathrm{New}} + a_3P_{\mathrm{Ind}}$, where we have chosen $\sum_{i=1}^3 a_i = 1$ and

$$a_1 = 1/2 - \mathrm{d_{Inc}}/2(\mathrm{d_{Exc}} + \mathrm{d_{Ind}} + \mathrm{d_{Inc}}),$$
$$a_2 = 1/2 - \mathrm{d_{Exc}}/2(\mathrm{d_{Exc}} + \mathrm{d_{Ind}} + \mathrm{d_{Inc}}),$$
$$a_3 = 1/2 - \mathrm{d_{Ind}}/2(\mathrm{d_{Exc}} + \mathrm{d_{Ind}} + \mathrm{d_{Inc}}).$$

Let us consider now the observed frequencies (also called "empirical probabilities") in a data simulation describing, in a Bayesian or knowledge network, a concept C from three random Boolean variables X_1, X_2 and X_3. To each X_i are related two events $\{X_i = 0\}$ and $\{X_i = 1\}$. The observed frequencies are chosen equal to:

$$P(\{X_1{=}0\}) = P(A_1){=}0.4{=}1{-}P(\{X_1{=}1\}){=}1{-}P(A_2),$$
$$P(\{X_2{=}0\}) = P(B_1){=}0.65{=}1{-}P(\{X_2{=}1\}){=}1{-}P(B_2),$$
$$P(\{X_3{=}0\}) = P(C_1){=}0.15{=}1{-}P(\{X_3{=}1\}){=}1{-}P(C_2).$$

We will suppose that the events are mutually independent and that the empirical probabilities of intersections $F_1(A_i \cap B_j \cap C_k)$ are given in the Table 5.

TABLE 4
Estimates of $P(A_1 \cap A_2 \cap A_3)$, when $P(A_1) = \alpha$, $P(A_2) = \beta$ and $P(A_3) = \gamma$ in different cases of dependence

Dependency cases	Events probabilities	P	P_{Lan}	P_{New}	P_{Ind}	Σ	Π	Best Estimation
Inclusive	$\alpha \simeq \beta$ $\beta \simeq \gamma$	α	$3\alpha^2 - 2\alpha^3$	α^2	α^3	$\frac{\alpha}{2}$	$\alpha^{\frac{3}{2}}$	$Max(P_{\text{Lan}}, \Sigma)$
	$\alpha \simeq (1-\epsilon)$ $\beta \simeq \gamma$	β	β	$\frac{2\beta^2+\beta}{3}$	β^2	$\frac{\beta}{2}$	$\beta^{\frac{3}{2}}$	P_{Lan}
	$\alpha \simeq \epsilon$ $\beta \simeq \gamma$	ϵ	$\epsilon\beta(3-2\beta)$	$\beta\epsilon$	$\beta^2\epsilon$	$\frac{\beta+2\epsilon}{6}$	$\epsilon\sqrt{\beta}$	P_{Lan}
Nearly Independency	$\alpha \simeq \beta$ $\beta \simeq \gamma$	$\alpha^3 + \alpha\eta$	$\alpha^3 + 3\alpha\eta$	$\alpha^3 + \alpha\eta$	α^3	$\frac{\alpha^2+\eta}{2}$	$\alpha^3 + \frac{3}{2}\alpha\eta$	P_{New}
	$\alpha \simeq (1-\epsilon)$ $\beta \simeq \gamma$	$\beta^2(1-\epsilon) + \eta(\frac{2\beta+1}{3})$	$\beta^2(1-\epsilon) + 3\eta$	$\beta^2(1-\epsilon) + \eta$	β^2	$\frac{\beta^2+2\beta-2\beta\epsilon+3\eta}{6}$	$\beta^2(1-\epsilon) + \frac{3}{2}$	P_{New}
	$\alpha \simeq \epsilon$ $\beta \simeq \gamma$	$\beta^2\epsilon + \frac{\eta\epsilon}{3}$	$\beta^2\epsilon + \eta\epsilon$	$\beta^2\epsilon + \frac{\eta\epsilon}{3}$	$\beta^2\epsilon$	$\frac{\beta^2+2\beta\epsilon+\eta}{6}$	$\beta^2\epsilon + \frac{\eta\epsilon}{2}$	P_{New}
Exclusive	$\alpha \simeq \beta$ $\beta \simeq \gamma$	0	$-2\alpha^3$	0	α^3	0	0	$P_{\text{New}}, \Sigma, \Pi$
	$\alpha \simeq (1-\epsilon)$ $\beta \simeq \gamma$	0	$\frac{-\epsilon^2(1-\epsilon)}{2}$	0	β^2	0	0	$P_{\text{New}}, \Sigma, \Pi$
	$\alpha \simeq \epsilon$ $\beta \simeq \gamma$	$\frac{\eta^2}{3}$	$\epsilon(\eta - 2\beta^2)$	$\frac{\eta\epsilon}{3}$	$\beta^2\epsilon$	$\frac{\eta}{6}$	0	P_{New}

If we examine the estimates of the joint probabilities in the Table 5, we can calculate the distances to dependence and to mutual independence proposed above for the event corresponding to the intersection $A_2 \cap B_1 \cap C_2$, whose empirical probability is given by $F_1(A_2 \cap B_1 \cap C_2) = 0.01$. Then we have:

$$P_{\text{Ind}}(A_2 \cap B_1 \cap C_2) = 0.33 \quad P_{\text{Lan}}(A_2 \cap B_1 \cap C_2) = -0.63$$

$$P_{\text{New}}(A_2 \cap B_1 \cap C_2) = 0.01 \quad P_*(A_2 \cap B_1 \cap C_2) = 0.06$$

The best estimator there is P_{New}, which is not surprizing, because the comparison between the distances $\text{d}_{\text{Inc}} = 0.8235$, $\text{d}_{\text{Exc}} = 0.1375$ indicates a case of exclusive dependence and we have shown above that we would privilegiate in these circumstances P_{New}.

For the event corresponding to the intersection $A_1 \cap B_2 \cap C_2$, whose empirical probability is given by $F_1(A_1 \cap B_2 \cap C_2) = 0.2$, the best estimator is $P_{\text{Ind}}(A_1 \cap B_2 \cap C_2) = 0.12$, which is in agreement with the fact that the event is far from the 2 extreme situations of dependence ($\text{d}_{\text{Inc}} = 0.5483$, $\text{d}_{\text{Exc}} = 0.2591$). If we calculate the sum of the squares of the differences between the empirical joint probabilities and their estimators (SSP), which summarizes the performance of the estimators in the Table 5, we obtain:

$$SSP_{\text{Ind}} = 0.1936, \qquad SSP_{\text{Lan}} = 0.5249,$$
$$SSP_{\text{New}} = 0.0210, \qquad SSP_* = 0.0284,$$

which shows that P_{New} is globally the best estimator, P_* being close to it.

To sum up, we propose P_* as an estimator being acceptable in any circumstance of dependence, because it takes the best of the estimators P_{New}, P_{Lan} and P_{Ind}.

4.1.5 *An optimized strategy for estimating the joint probabilities*

We propose now the following incremental procedure to optimize the convergence to the best estimates of the joint probabilities:

TABLE 5
Empirical joint probabilities $F_1(A_i \cap B_j \cap C_k)$ and their estimations

	A_1			
	B_1		B_2	
	C_1	C_2	C_1	C_2
F_1	0.1	0.001	0	0.2
P_{Ind}	0.0390	0.2210	0.0210	0.1190
P_{Lan}	0.0422	-0.2211	0.0350	0.0872
P_{New}	0.0401	0.0736	0.0257	0.1084
P_*	0.0399	0.0949	0.0250	0.1099
	A_2			
	B_1		B_2	
	C_1	C_2	C_1	C_2
F_1	0	0.01	0.03	0.012
P_{Ind}	0.0585	0.3315	0.0315	0.1785
P_{Lan}	-0.0360	-0.6336	-0.0282	-0.1864
P_{New}	0.0270	0.0098	0.0116	0.0569
P_*	0.0315	0.0561	0.0145	0.0744

- To ask the expert delivering the knowledge for a subjective estimation of the assertions probabilities,*i.e.* for any event A_i, to get $P_0(A_i) \pm \text{Er}(P_0(A_i))$, where $\text{Er}(P_0(A_i))$ is a subjective error - To consider $\text{Er}(P_0(A_i))$ as equal to two times the standard deviation of $P_0(A_i)$, considered as a Gaussian random variable representing a subjective probability. Then $(E(P_0(A_i)))^2/4$ is an estimator of the variance $V(P(A_0))$ of $P_0(A_i)$.

- To calculate a pseudo initial sample size $N_0 = 4P_0(A_i)(1 - P_0(A_i))/(E(P_0(A_i)))^2$, by considering that if the subjective estimation $P_0(A_i)$ proceeded from an empirical observation, we would have: $V(P_0(A_i)) = P_0(A_i)(1 - P_0(A_i))/N_0$, where N_0 is the sample size of the observations used to do this estimation

- To check with the expert if this value is realistic taking into account his past experience

- To use a progressively updated case data base made of N cases associated to the knowledge network and improve the initial estimate $P_0(A_i)$ by calculating the *a* *posteriori* value:

$$P_N(A_i) = (N_0 P_0(A_i) + \text{ cases verifying } A_i)/(N_0 + N)$$

- To do the same as done for marginal probabilities and calculate $P_N(A_i \cap A_k)$ for all intersection of order 2 involved in the views of interest

- To estimate the probability of any intersection of A_i's by using the P_N* estimator.

4.2 Information fusion in a multi-sensors context

An application directly related to the decision procedure in Bayesian networks concerns the multi-sensors fusion [9]. When the information needed to execute a task T in a human or artificial sensory-motor context comes from multiple sensors S_i whose confidence is taken into account by observing n random variables X_i, the probability $P(T|\bigcap_{i=1}^n\{X_i = k_i\})$ to decide that T is executed knowing the values k_i's of the X_i's, we have to rapidly estimate $P(T\cap(\bigcap_{i=1}^n A_i))$ and $P(\bigcap_{i=1}^n A_i)$, where $A_i = \{X_i = k_i\}$.

Recent papers [49], [50] proposed to use a Bayesian approach in order to mime the fonctionning of the central nervous system. They supposed that the brain functions by fusing information in evaluating conditional probabilities, knowing *a priori* distribution (knowledge about the environment) and observing random variables (metrologic sensors).

An example of such an information fusion can be obtained in the context of human postural control correction in which the action consists in correcting a erroneous position of the body. It is generally agreed that maintaining an upright stance or seated posture involves the integration (fusion) of sensory information from multiple natural sensors including visual (variable X_1), somatosensory (variable X_2) and vestibular (variable X_3) systems [51]. However, augmented/substituted artificial sensory information also can become a feedback to the brain when one of the natural sensory inputs is unavailable/undetermined/altered, or when one merely wants to enhance the postural control for accurate performances in daily-living, professional or sportive activities. Along theses lines, innovative health technologies, based on the concept of ”sensory substitution” [16] have been recently developed for pressure sores prevention in the case of spinal cord injuries (persons with paraplegia or tetraplegia) [17], [19] and for fall prevention in older and/or disabled adults [20], [21], [22], [23], [24], [25], [26], [27]. The underlying principle of these biomedical devices consists in supplying individuals with supplementary somato-sensory information related to pressure distribution beneath the buttock or the feet (variables X_1, X_2 and X_3), recorded by the means of artificial sensors (like pressure mapping), *via* an alternative sensory modality (electrotactile stimulation of the tongue, variable X_4). At this point, an effective fusion of natural (variables X_1, X_2 and X_3) and artificial (more reliable and accurate than the natural one's) sensory information (variable X_4) is crucial to enable individuals with spinal cord injuries, or with somatosensory loss in the feet (*e.g.*, from diabetic peripheral neuropathy) to become aware of a localized excess of pressure at the skin / seat interface and / or postural orientation and thus to make adaptive postural corrections to prevent the formation of pressure ulcers and / or fall. Then any control and correction procedure needs the real-time calculation of probabilities like:
$P(\{X_4 = c\}|\{X_1 = k_1\}\cap\{X_2 = k_2\}\cap\{X_3 = k_3\})$, where c is the coding value of a tongue stimulation. We propose in next Section a procedure for a fast estimation of joint probabilities like $P(\{X_4 = c\} \cap \{X_1 = k_1\} \cap \{X_2 = k_2\} \cap \{X_3 = k_3\})$ and $P(\{X_1 = k_1\} \cap \{X_2 = k_2\} \cap \{X_3 = k_3\})$, by observing together only couples of sensors and then estimating only the corresponding marginal empirical frequencies of order 1 and 2. Another example of use of such techniques concerns the cardio-respiratory alarm. Let us suppose that a person followed for a risk of cardiac failure is equipped with three sensors: i) a smart clothes recording the cardiac and respiratory rhythm able to detect a trouble in the dynamics of the sinusal respiratory arythmia ii) a wrist sensor giving the arterial pressure and iii) a geo-localizer like a GPS or a RFID tool fixed on the belt (Figure 2). The three corresponding signals X_1, X_2 and X_3 are supposed to be random and stochastically dependent, but are often not recorded together if the patient followed up suffers from a degenerative disease and is forgetting one (chosen by chance) of the three sensors. If during the learning phase the system has estimated the three marginal joint probability distributions $\{(X_i, X_j)\}_{i,j\in\{1,2,3\}}$, then we can from the observation of only couples of variables during a week calculate the whole joint distribution of the triplet (X_1,X_2,X_3), and detect if there are significant observations deviating from the physiological distribution by calculating the chi-square distance between the latter and the currently observed. In case of significant difference, we anticipate the cardiac failure by performing an effort ECG and counselling for an adapted therapy.

We should now develop a procedure to take into account a possible nychtemeral (day versus night) phase shift between consecutive days showing the same sequence of tasks along the daily activity, but shifted in time without any pathological signification except a change in the sleeping clock, causing a loss by the elderly people of its synchronization with the nycthemeral zeitgebers (synchronizers), like meals or social activities.

4.3 Correction of a nycthemeral phase shift

We propose to calculate the persistence parameter after an eventual nychtemeral phase shift. By considering that the dominant (in time) activity of an hour leads to assign to this hour a symbol chosen among 4, corresponding to four main types of activities:

A Ambulatory Activity
G Generic Social or Cultural Activity

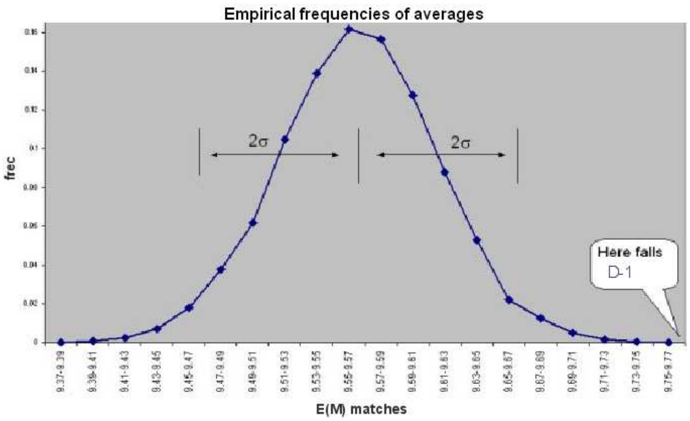


Fig. 6. Empirical distribution of the mean number $E(M)$ of matches calculated between 500 activities sequences $D1, \ldots, D500$ and 30 000 random sequences, showing that the match between $D1$ and $D-1$ is significantly better than a random match ($p < 10^{-3}$)

C Cooking & Eating
U Unassigned to a Specific Activity (rest or sleep)

If we denote by M the random variable equal to the number of matches between the activities sequence of two consecutive days, *e.g.* sequence x at day $D1$ and sequence y at the previous day $D-1$, then we have: $M = n - \min_{k=1,\ldots,n} d_H(x, \sigma^k(y))$, where d_H is the classical Hamming distance and $\sigma^k(y)$ is the chain obtained by opening y at the letter of phase k throughout the nycthemere (day versus night). We call circular Gumbel distribution the probability law of M. Let us suppose that we have recorded the daily activities at n ($n = 22$) times (by considering that the hours 24, 1 and 2 correspond to the same sleeping activity). The expected number of matches $E(M)$ in the case of the comparison of the $D1$ and $D-1$ sequences of activities is less than the maximum number of matches observed in the case of comparison between 22 independent chains of length 22, because a change of the origin of phases on the ring does not correspond strictly to a new chain tossing. Then we can write: $P(M < k) > P(\cap_{i=1,\ldots,22}(X_i < k))$, where the X_i's are independent identically distributed (i.i.d.) random variables, having as common distribution, the binomial law $B(22, 1/4)$, *i.e.* the distribution of a binomial variable X equal to the number of matches between 2 independent sequences, by supposing that the occurrence of each activity A, U, G, C has the probability $1/4$ and there is 22 times in the lapse of recording. If n increases (*i.e.* if the temporal sample is refined, *e.g.* from hours to minutes), then the circular Gumbel distribution tends to the $\sup_{i=1,\ldots,n} X_i$ distribution, and we have asymptotically in n:

$$E(M) = \sum_{k=1}^{n} P(M \geq k) = n - \sum_{k=1}^{n} P(M < k)$$

$$\approx n - \sum_{k=1}^{n} P(\cap_{i=1}^{n}(X_i < k)) = n - \sum_{k=1}^{n} P(X < k)^n$$

Hours:	3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Sequence $D-1$:	CAAGAUGAAUGGUACUGCCAUU
Sequence $D1$:	UCAGGUAAGUGUUCACUGCCAU

Because the mean m (respectively variance s^2) of X is equal to $n/4$ and (respectively $3n/16$), we can neglect (with an approximation less than 2%, taking into consideration that X is asymptotically Gaussian) the quantities $P(X \geq k)^n$ for $k > I = i(m+3s)+1 = i(\frac{n}{4} + \frac{3\sqrt{3n}}{4}) + 1$, where $i(x)$ is the nearest integer less than x, and the quantities $(1 - P(X \geq k))^n$ for $k < I' = i(m+2s) - 1 = i(\frac{n}{4} + \frac{2\sqrt{3n}}{4}) - 1$:

$$E(M) \approx I' + n \sum_{k=I'+1}^{I} P(X \geq k)$$

If $n = 22$, we have: $s \approx 2$, $I' = 8$, $I = 12$ and we have: $E(M) \approx 8 + 22 \sum_{k=9,\ldots,12}(P(X \geq k)) \approx 9.32$.

We can notice also that the distribution of the $\sup_{i=1,\ldots,n} X_i$, hence the circular Gumbel distribution, behaves like a single Gaussian variable with a suitably chosen variance [52].

Suppose that the 22 activities recorded at day $D-1$ correspond to the sequence: **UCAGGUAAGUGUUCACUGCCAU**. Then we can compare the other days to this reference sequence by using the circular Hamming distance and the significativity of the result will be given with respect to the circular Gumbel distribution, whose empirical mean is approximatively Gaussian of mean 9.55 (Figure 6). By comparing for example the sequence $D1$ to the reference sequence $D-1$, we find 12 as number of matches, with a nycthemeral phase shift of 1, which is significantly better than the mean match with a random sequence ($p < 10^{-3}$). An alarm will be triggered only in case of a significant difference observed after correction of the nychtemeral phase shift which can not be considered as pathologic in elderly people.

5 CONCLUSION

We propose in this paper a strategy for estimating from localization data collected at home a persistence parameter quantifying the degree of perseveration in a task a person can pathologically develop. In case of censored or false data (badly calibrated sensor, leaved off on the person, etc.), we develop a method for estimating joint probabilities in the Bayesian network or in the knowledge base in which data are managed, in any case of dependence between the observed actimetric variables involved in the surveillance process and in the alarm decision procedure. Marginal and order 2 joint probabilities are supposed to be known (if not, we assume the mutual independence) such that it is possible to estimate with a good precision any joint probability of higher order by using a convex compromise between the classical Lancaster-Zentgraf estimator and new estimators we have introduced. Simulated examples studied in the paper evidence the efficiency of such an estimation strategy, and eventually we explain a methodology of using these estimates in the calculation of the perseveration index.

ACKNOWLEDGMENT

The data were recorded by the AFIRM Team from TIMC-IMAG Laboratory and RBI during the AILISA project, supported by the French RNTS health network since 2003 within the framework of the "Institut de la Longévité" (n°03B651-9).

REFERENCES

[1] S. Joray, F. Herrmann, R. Mulligan, and A. Schnider, "Mechanism of disorientation in alzheimer's disease," *Eur. Neurol.*, vol. 52, pp. 193–197, 2004.
[2] A. Schnider, A.-C. von Däniken, and K. Gutbrod, "Disorientation in amnesia. a confusion of memory traces," *Brain*, vol. 119, pp. 1627–1632, 1996.
[3] A. Schnider, "Spontaneous confabulation and the adaptation of thought to ongoing reality," *Nature Rev. Neurosciences*, vol. 4, pp. 662–671, 2003.
[4] M. Sebastian, J. Menor, and R. Elosua, "Patterns of errors in short-term forgetting in ad and ageing," *Memory*, vol. 9, pp. 223–231, 2001.
[5] M. Sebastian, J. Menor, and M. Elosua, "Attentional dysfunction of the central executive in ad: Evidence from dual dual task and perseveration errors," *Cortex*, vol. 42, pp. 1015–1020, 2006.
[6] Anonymous, "Three early-warning alzheimer's tests. timelier detection can lead to longer and improved quality of life," *Duke Med Health News*, vol. 13, no. 7, pp. 4–5, Jul. 2007.
[7] J. Touchon, K. Ritchie, and M. Gely-Nargeot, "Early diagnosis of alzheimer disease," *Ann Med Interne (Paris)*, vol. 149, no. 4, pp. 193–201, Jun. 1998.
[8] S. Das, N. Roy, and A. Roy, "Context-aware resource management in multi-inhabitant smart homes: A framework based on nash h-learning," *Pervasive and Mobile Computing*, vol. 2, pp. 372–404, 2006.
[9] J. Demongeot, G. Virone, F. Duchêne, G. Benchetrit, T. Hervé, N. Noury, and V. Rialle, "Multi-sensors acquisition, data fusion, knowledge minig and alarm triggering in health smart homes for elderly people," *Comptes Rendus Biologies*, vol. 325, pp. 673–682, 2002.
[10] G. Le Bellego, N. Noury, G. Virone, M. Mousseau, and J. Demongeot, "A model for the measurement of patient activity in a hospital suite," *IEEE Transactions ITB*, vol. 10, pp. 92–99, 2006.
[11] N. Noury, G. Virone, J. Ye, V. Rialle, and J. Demongeot, "New trends in heath smart homes," *ITBM/RBM*, vol. 24, pp. 122–135, 2003.
[12] N. Noury, G. Virone, P. Barralon, J. Ye, V. Rialle, and J. Demongeot, "New trends in health smart homes," in *Proc. HEALTHCOM 2003*, 6–7 June 2003, pp. 118–127.
[13] V. Rialle, F. Duchêne, N. Noury, L. Bajolle, and J. Demongeot, "Health 'smart' home: Information technology for patients at home," *Telemedicine Journal and E-Health*, vol. 8, pp. 395–409, 2002.
[14] G. Virone, N. Noury, and J. Demongeot, "A system for automatic measurement of circadian activity deviations in telemedicine." *IEEE Trans Biomed Eng*, vol. 49, no. 12, pp. 1463–1469, Dec 2002. [Online]. Available: http://dx.doi.org/10.1109/TBME.2002.805452
[15] G. Virone, M. Alwan, S. Dalal, S. Kell, B. Turner, J. Stankovic, and R. Felder, "Behavioral patterns of older adults in assisted living," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 3, pp. 387–398, 2008.
[16] P. Bach-y Rita, C. Collins, F. Saunders, B. White, and L. Scadden, "Vision substitution by tactile image projection," *Nature*, vol. 221, pp. 963–964, 1969.
[17] A. Moreau-Gaudry, A. Prince, J. Demongeot, and Y. Payan, "A new health strategy to prevent pressure ulcer formation in paraplegics using computer and sensory substitution via the tongue," *Stud. Health Technol. Inform.*, vol. 124, pp. 926–931, 2006.
[18] F. Robineau, F. Boy, J. P. Orliaguet, J. Demongeot, and Y. Payan, "Guiding the surgical gesture using an electro-tactile stimulus array on the tongue: A feasibility study," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 4, pp. 711–717, April 2007.
[19] N. Vuillerme, O. Chenu, N. Pinsault, A. Moreau-Gaudry, A. Fleury, J. Demongeot, and Y. Payan, "Pressure sensor-based tongue-placed electrotactile biofeedback for balance improvement - biomedical application to prevent pressure sores formation and falls," in *Proc. IEEE EMBS 2007*, 22-26 Aug. 2007, pp. 6113–6116.
[20] N. Vuillerme, O. Chenu, J. Demongeot, and Y. Payan, "Controlling posture using a plantar pressure-based, tongue-placed tactile biofeedback system," *Experimental Brain Research*, vol. 179, pp. 409–414, 2007.
[21] N. Vuillerme, O. Chenu, N. Pinsault, M. Boisgontier, J. Demongeot, and Y. Payan, "Inter-individual variability in sensory weighting of a plantar pressure-based, tongue-placed tactile biofeedback for controlling posture," *Neuroscience Letters*, vol. 421, pp. 173–177, 2007.
[22] N. Vuillerme, N. Pinsault, O. Chenu, M. Boisgontier, J. Demongeot, and Y. Payan, "How a plantar pressure-based, tongue-placed tactile biofeedback modifies postural control mechanisms during quiet standing," *Experimental Brain Research*, vol. 181, pp. 547–554, 2007.
[23] N. Vuillerme, N. Pinsault, O. Chenu, J. Demongeot, Y. Payan, and Y. Danilov, "Sensory supplementation system based on electrotactile tongue biofeedback of head position for balance control," *Neuroscience Letters*, vol. 431, pp. 206–210, 2008.
[24] N. Vuillerme, N. Pinsault, O. Chenu, A. Fleury, Y. Payan, and J. Demongeot, "A wireless embedded tongue tactile biofeedback system for balance control," *Pervasive & Mobile Computing*, vol. 5, pp. 268–275, 2008.
[25] N. Vuillerme, O. Chenu, N. Pinsault, A. Fleury, Y. Payan, and J. Demongeot, "Can a plantar pressure-based tongue-placed electrotactile biofeedback improve postural control under altered vestibular and neck proprioceptive conditions?" *Neuroscience*, vol. 155, pp. 291–296, 2008.
[26] N. Vuillerme, N. Pinsault, O. Chenu, A. Fleury, Y. Payan, and J. Demongeot, "Postural destabilization induced by trunk extensor muscles fatigue is suppressed by use of a plantar pressure-based electro-tactile biofeedback," *European Journal of Applied Physiology*, vol. 104, pp. 119–125, 2008.
[27] N. Vuillerme, N. Pinsault, A. Fleury, O. Chenu, J. Demongeot, Y. Payan, and P. Pavan, "Effectiveness of an electro-tactile vestibular substitution system in improving upright postural control in unilateral vestibular-defective patients," *Gait Posture*, vol. 28, pp. 711–715, 2008.
[28] F. Teymoori, M. Mousavi, M. Shirazikhah, B. Abharilaleh, R. Biglarian, A. Franco, and J. Demongeot, "Fi-cga score of old people by community based information system," in *Proc. IEEE ARES-CISIS' 09 & BT' 09*, Piscataway, 2009, pp. 949–954.
[29] N. Neil, M. Tailor, and D. Marquez, "Inference in hybrid bayesian networks using dynamic discretization," *Stat. Comput.*, vol. 17, pp. 219–233, 2007.
[30] D. Spiegelhalter and S. Lauritzen, "Sequential updating of conditional probabilities on directed graphical structures," *Networks*, vol. 20, pp. 579–605, 1990.
[31] D. Spiegelhalter, S. Dawid, A.P.and Lauritzen, and R. Cowell, "Bayesian analysis in expert systems," *Journal of Statistical Science*, vol. 8, pp. 219–283, 1993.
[32] A. Simonet and M. Simonet, "Objects with views and constraints: from databases to knowledge bases," in *Proc. OOIS'94*. Springer Verlag, 1994, pp. 182–197.
[33] M. von Davier and C. Carstensen, *Multivariate and Mixture Distribution Rasch Models: Extensions and Applications*, N. York, Ed. Springer Verlag, 2007.
[34] Q. Xiao and S. Mahadevan, "Second-order upper bounds on probability of intersection of failure events," *Journal of engineering mechanics*, vol. 120, pp. 670–674, 1994.
[35] J. Demongeot, N. Noury, and N. Vuillerme, "Data fusion for analysis of persistence in pervasive actimetry of elderly people at home," in *Proc. IEEE CISIS' 08 & APPLIMS' 08*, Piscataway, 2008, pp. 589–594.
[36] K. Inoue and S. Aki, "Polya urn models under general replacement schemes," *J. Japan Statist. Soc.*, vol. 31, pp. 193–205, 2001.
[37] G. Ivchenko, "Estimation in the markov-polya scheme," *Mathematical Notes*, vol. 64, pp. 322–329, 1998.

[38] S. Kotz, H. Mahmoud, and P. Robert, “On generalized polya urn models,” *Statistics & Probability Letters*, vol. 49, pp. 163–173, 2000.
[39] Y. Fouquet, J. Demongeot, and N. Vuillerme, “Pervasive informatics and persistent actimetric information in health smart homes : From Language Model to Location Model,” in *Proc. IEEE ARES-CISIS' 09 & BT' 09*, Piscataway, March 2009, pp. 935–942.
[40] Y. Fouquet, N. Vuillerme, and J. Demongeot, “Pervasive informatics and persistent actimetric information in health smart homes,” in *Proc. LNCS ICOST'09*, Tours France, July 2009, p. 73.
[41] N. Reithinger and E. Maier, “Utilizing statistical dialogue act processing in verbmobil,” Verbmobil, VM-Report 80, 1995.
[42] L. Ekroot and T. Cover, “The entropy of markov trajectories,” *Information Theory, IEEE Transactions on*, vol. 39, no. 4, pp. 1418 –1421, jul 1993.
[43] P. Couturier, A. Franco, J.-F. Piquart, J. Mansotte, C. Montani, C. Suarez, A. Mollier, C. Gucher, M. Frossard, L. Nicolas, G. Jasso Mosqueda, M.-C. Mouchet, A.-M. Argentier, J.-L. Bosson, C. P., and J. Demongeot, “Telegerontology: from teleassistance to teleconsultation of elderly people. mythe or reality?” *Rev. gériatrie*, vol. 21, pp. 23–31, 1996.
[44] G. Virone, D. Istrate, M. Vacher, N. Noury, J. F. Serignat, and J. Demongeot, “First steps in data fusion between a multichannel audio acquisition and an information system for home healthcare,” in *Proc. IEEE EMBS 2003*, vol. 2, 17–21 Sept. 2003, pp. 1364–1367.
[45] G. Virone, B. Lefebvre, N. Noury, and J. Demongeot, “Modeling and computer simulation of physiological rhythms and behaviors at home for data fusion programs in a telecare system,” in *Proc. IEEE Healthcom 2003*, 6–7 June 2003, pp. 111–117.
[46] G. Virone, N. Noury, J.-P. Thomesse, V. Rialle, and J. Demongeot, “A home health information system based on the can fieldbus,” in *Proc. IFAC FeT'2003*. Elsevier Science, 2003, pp. 270–281.
[47] H. Lancaster, *The χ^2 Distribution*, J. Wiley, Ed., New York, 1969.
[48] R. Zentgraf, “A note on lancaster's definition of higher-order interactions,” *Biometrika*, vol. 62, pp. 375–376, 1975.
[49] D. Knill and A. Pouget, “The bayesian brain: the role of uncertainty in neural coding and computation,” *Trends in Neurosciences*, vol. 27, pp. 713–719, 2004.
[50] K. Koerding and G. Wolpert, “Bayesian integration in sensorimotor learning,” *Nature*, vol. 427, pp. 244–247, 2004.
[51] J. Massion, “Postural control system,” *Current Opinion in Neurobiology*, vol. 4, pp. 877–887, 1994.
[52] R. J. Adler, *An introduction to continuity, extrema, and related topics for general Gaussian processes*. Hayward, CA: Institute of Mathematical Statistics, 1990.
[53] E. Galloy, A. Le Breton, and S. Martin, “A model for weather cycles based on daily rainfall occurence,” *Lecture Notes in Biomathematics*, vol. 49, pp. 303–318, 1983.
[54] C. Martin and Y. Ho, “Value of information in the polya urn process,” *Information Sciences*, vol. 147, pp. 65–90, 2002.
[55] R. Rosenfeld, “Adaptative statistical language modeling : A maximum entropy approach,” Thesis for the degree of Doctor of Philosophy, Carnegie Mellon University, 1994.
[56] ——, “A maximum entropy approach to adaptive statistical language modeling,” *Computer, Speech and Language*, 1996.
[57] N. Yankelovich, G.-A. Levow, and M. Marx, “Designing speech acts issues in speech user interfaces,” in *Proc. ACM CHI Conference on Human Factors in Computing Systems*, 1995, pp. 369–376.

Dr Yannick Fouquet received the M.D. degree in HMI and the Ph.D. degree in computer sciences from the Grenoble Joseph Fourier University, France, in 1999 and 2004 respectively. Currently, he is project manager for the development of methods of assessment in the Alpes Science and technology park for Health at Home and Autonomy (TASDA) in the AGIM (AGeing, Imaging, Modeling), FRE 3405, CNRS-UJF-UPMF-EPHE Laboratory .

Céline Franco received the M.D. degree in Engineering for Health and Drugs from the Grenoble Joseph Fourier University, France, in 2009. She is currently pursuing the Ph.D. degree jointly in the AGIM (AGeing, Imaging, Modeling), FRE 3405, CNRS-UJF-UPMF-EPHE Laboratory and the TIMC-IMAG (Techniques in Biomedical Engineering and Complexity - Informatics, Mathematics and Applications-Grenoble), UMR 5525, CNRS, UJF Laboratory. Her research interests include time-scale representation and processing of biosignals and development of frailty indexes.

Bruno Diot was graduated from the Management Control Institute (ICG), France in 1986. He was in charge of Research and Development at Intervox Systems (ex Intervox Alcatel) between 1996 and 2003. He is currently chairman of the management board of ”Informatique de Sécurité” (IDS), Montceau les Mines, France. IDS is a host for health data recognized by the French Ministry for Health. Its area of research includes Ambient Assisted Living and in-home support for dependent people.

Pr Jacques Demongeot received the M.D. degree and the Ph.D. degree in mathematics from the Grenoble Joseph Fourier University, Grenoble, France, in 1974 and 1978, respectively. He is presently Deputy Director of the Laboratory AGIM (FRE 3405, CNRS-UJF-UPMF-EPHE) ”AGing, Imageing, Modelling”. He has been appointed as Chairman of Biomathematics at the ”Institut Universitaire de France” (IUF), Paris, in 1994. He is responsible for the Department of Public Health at the University Hospital of Grenoble.

Dr Nicolas Vuillerme received the PhD degree in Sciences and Technologies for Physical Activities and Sports - Neurosciences and the Habilitation degree in Medicine from the Grenoble Joseph Fourier University (France) in 2002 and 2010, respectively. He is currently Assistant Professor at the Grenoble Joseph Fourier University and is leading the Data Acquisition, Fusion and Networks for Medicine research team of the AGIM (AGeing, Imaging, Modeling), FRE 3405, CNRS-UJF-UPMF-EPHE Laboratory. His current research dedicated to Technologies and Services for Ageing, Health and Autonomy in the fields of Information Technology Engineering Sciences, Human Factors, Ergonomics, Biomechanics and Neurosciences.