Société de Calcul Mathématique, S. A.
Algorithmes et Optimisation

A note about prediction models

Bernard Beauzamy

Paris, March 11, 2007

I. General considerations

Many people use prediction models, in order to forecast a population, a consumption, a temperature, a new position, and so on. Typically, we have some data from the past and we would like to predict how they will behave in the future. Let x_n be the existing data, for instance indexed by the date (years). The x_n may be precise or imprecise (given by probability laws).

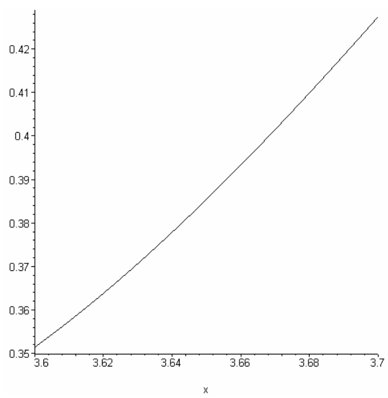
What people do is usually to adjust a "trend" (typically a straight line, obtained by linear regression), and conclude from this trend. This practice has three drawbacks :

– All points have equal weight : the ancient past counts as much as the recent past ;

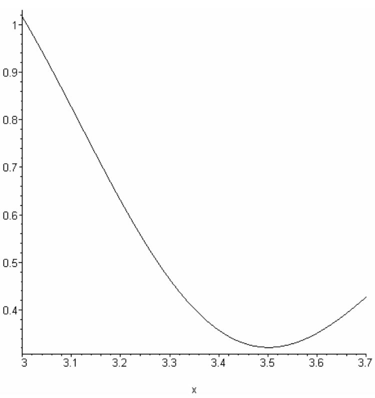
– No confidence interval is possible ;

– The trend which is detected is only linear ;

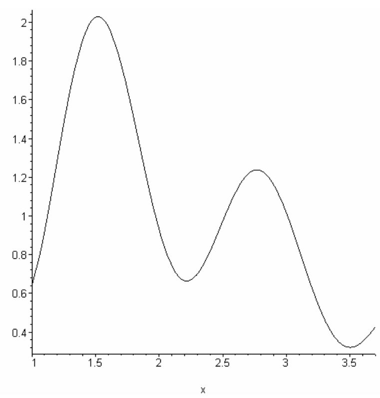
– This trend is quite "fragile", in the sense that it depends strongly upon the available time-interval. Let's look at the following three graphs : they represent the same function, but on 3 different time intervals : Recent past (interval 3.6 - 3.7), Medium past (interval 3 - 3.7), Complete past (interval 1 - 3.7). Depending on the interval chosen, the trend is completely different.



Recent past (interval 3.6 - 3.7)



Medium past (interval 3 - 3.7)



Complete past (interval 1 - 3.7)

Siège social et bureaux : 111, Faubourg Saint Honoré, 75008 Paris. Tel. : 01 42 89 10 89. Fax : 01 42 89 10 69
Société Anonyme au capital de 56 200 Euros. RCS : Paris B 399 991 041. SIRET : 399 991 041 00035. APE : 731Z

In order to predict a position, one often uses tools called "filters" (the best example is the so-called "Kalman filter"). These prediction methods usually assume the phenomenon to be linear and the errors to be Gaussian, which is not true in practice.

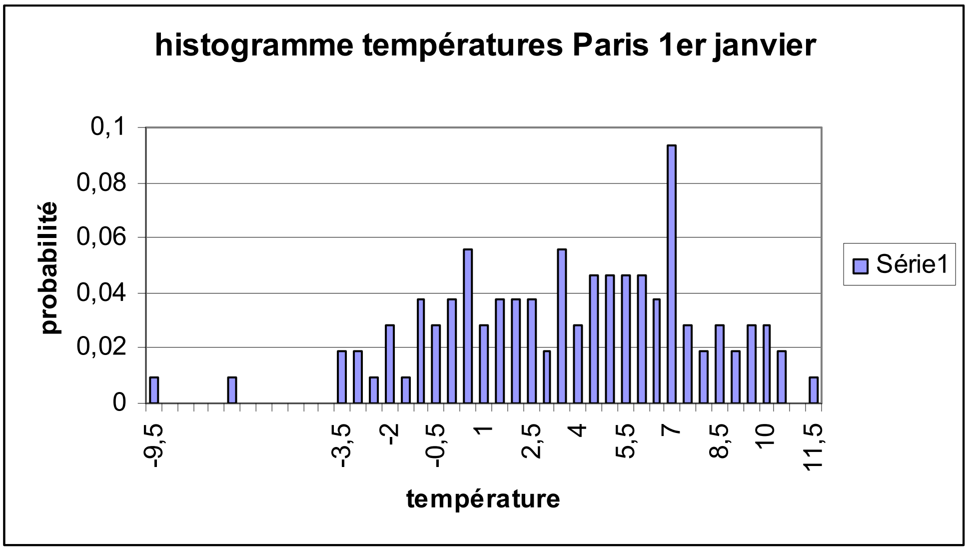
So there is clearly a need for better, more robust methods.

II. Choice between two approaches

Two approaches may be considered, and I would like to emphasize the fact that they are quite different.

1. Probabilistic approaches directly on the values

This is simple : from all values x_n you build an histogram : how many are in this interval, and so on. For instance, you may consider the temperature in Paris, for each day, for the last 130 years (existing records). Take for example the temperature for January 1st : you have 130 data, from which you can build an histogram, such as this one:



Then from this histogram you can predict a future value : the average (3.9°C) is the prediction which will give you the smallest error.

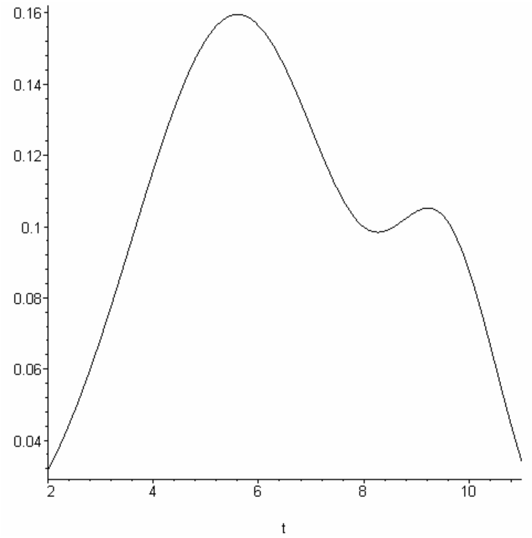
This analysis is completely correct if the process is stationary, that is if its law does not change with time (no cooling or warming), and if the next value is independent from the previous ones.

It uses all data with equal weight and will never predict a value which is outside what has been seen before.

More sophisticated tools, such as our "Experimental Probabilistic Hypersurface" have the same characteristics : existing information conveys information to new places (a precise value is sent as a density of probability), but this new information is mostly carried by old values. For instance, assume we construct the EPH just from the following data

2005	9,6°C
2006	5,6°C

and we want to predict the temperature for January 1st, 2007. The first data (2005) will send a Gaussian density centered at 9.6 ; the second will send a Gaussian density centered at 5.6 ; the result will be a combination of the two, which much more weight upon 2006, since it is closer. We might obtain something like this :



Graph of the density function, predicted temperature for January 1st, 2007

Here, some stationarity assumption is also made : we use data from the past and propagate them to the future ; only the time interval is used.

2. Trends

Let's now take a completely different set of data. Look at the simple example $x_n = n$ for all $n = 1,\dots,N$. This is by no means a stationary process.

The simplest probabilistic method will give you the average of previous values, that is $\frac{1+\dots+N}{N} = \frac{N+1}{2}$. The EPH will give you a weighted combination of Gaussian functions, with high value at N , then lower at $N-1$, $N-2$, and so on. None of these methods predicts the value $x_{N+1} = N+1$, as we would expect.

This comes from the fact that we applied the methods to the wrong data. We should not apply them to $x_n = n$ but to the increments $x_n - x_{n-1} = 1$. Then everything becomes normal again : prediction using average gives 1, and using EPH gives a gaussian function centered at 1.

Conversely, let's try to consider differences of temperatures, as above. We now consider the sequence $t_n - t_{n-1}$, for 130 years, that is $t_2 - t_1, \dots, t_N - t_{N-1}$. The average of these 129 numbers is simply

$$\frac{t_2 - t_1 + t_3 - t_2 + \dots + t_N - t_{N-1}}{N-1} = \frac{t_N - t_1}{N-1}$$

and this will lead to the prediction :

$$t_{2007} = t_{2006} + \frac{t_{2006} - t_{1877}}{129}$$

which is absurd.

So if we apply the increment method to a stationary process, we get a conclusion which is absurd.

III. Proper choice of the method

But how do we know when we should apply the tools to x_n or to $x_n - x_{n-1}$, or to other, more complicated combinations ? Two factors should be taken into account :

- The process is stationary or not ;
- The data are independent or not.

These two things are different. The first one means that the law of the process does not change with time. The second one means that your knowledge of the *n*-th data is not improved by the knowledge of the previous ones.

A. *Case of stationary process, independent data*

Examples : temperatures of January 1st, or sampling of sizes in a given city.

Then build the probability law of the process and use the expectation as a prediction.

B. *Case of non-stationary process, independent data*

Examples : temperatures of January 1st in a region of global warming, or global cooling ; sampling of sizes in a city, over several hundreds of years (heights increase).

Using the probability laws for the previous years, try to build the probability law for the next year. This is done using the repartition function for each year. Then take the expectation of the probability law for year *n* as a prediction.

C. *Case of stationary process, non independent data*

This is typically the case of a moving object, if we assume that at each place there is an error of observation, and this error has always the same law.

If I know the positions $x_1, \ldots, x_{n-1}$ of the object in the past, I will not look for it at the same place than if I know nothing ; previous positions give an information about the next one. So consecutive positions are not independent.

Then, one should build a "mechanistic model", depending of the movement. For instance, if we have reasons to believe the object has a uniform movement, or is falling, and so on, we will take a linear or quadratic model. Then the coefficients of this model will be adjusted using the available data.

D. Case of non stationary process, non independent data

This is the case for a moving object, where the observation error depends upon the position. This is the case in practice, since the error is bigger when the object is further. Then one should try to do both B) and C), that is try to build the new probability law, and, at the same time, build a mechanistic model for the moving object.

IV. Recent past and ancient past

Prediction models tend to concentrate upon recent past : ancient data are usually less reliable, and people usually think that their influence upon the future is weaker. This is wrong, as the illustration given above shows : a given function is plotted first on a small recent interval (3.6 - 3.7), then on a bigger interval to the past (3 - 3.7), then on the whole set 1 - 3.7. Behaviors are quite different. We have a phenomenon which is globally decreasing, but with large oscillations. Depending on our view point, we may conclude that the process is increasing (recent past), oscillating, decreasing.

V. Conclusion

We note the following :

- Our results may be totally wrong if we do not use the correct prediction model ;
- No model should be used in order to predict a future that is too far ; if you have N data, you may try to predict at most $N/3$ or $N/2$ data in the future ; to predict more is not reasonable ;
- The classification into stationary/non stationary and independent/non independent gives a preliminary basis in order to find the correct model.
- There are cases where we do not know what is the correct prediction model.

In such a case, we should make it very clear in our conclusions that they depend upon the choice of a model, which is purely hypothetical at this stage, so that we cannot recommend any action.