# BAYESIAN PREDICTION IN TRANSPORTATION

Ivan Nagy, Jitka Kratochvílová, Pavla Němcová [1]

Constantly increasing intensity of transportation requires more and more sophisticated transportation control. In this paper, the essence of Bayesian approach to modelling and estimation for the automatic control is presented. First, general theory of the estimation for regression models is mentioned and then a suitable approximation, enabling Bayesian estimation of mixture models, is introduced. The theory is illustrated on a simple real data example of multi-step prediction of the traffic flow intensity recorded in the centre of Prague.

**Keywords**

Modelling, estimation, prediction, Bayesian approach, transportation.

## Introduction

In proportion to an increasing number of conveyances, problems in the transportation, especially those concerning big cities, often grow to unbearable limits. To drivers, these situations bring big delays in their travelling and rise petrol consumption. The slowly moving columns of cars produce the worst kind of emissions that spoil the city environments. The number of collisions between cars and pedestrians increases and movement of the preferred cars, like ambulance, busses or supply is paralyzed. A way, how to solve this problem, is to rebuild the city infrastructure. Nevertheless, not always this way can be realized. Especially in historical cities, the reconstructions are often strictly limited or not ever permitted. Then the only way is seeking for some better transportation control.

The control, we are interested in, aims at so called urban miniregions, which are logically integral areas of urban communications with several main crossroads controlled by signal lights and with measuring devices distributed over the area and providing the transportation data. Such miniregion represents a controlled system in which inputs are parameters of the signal lights (period and length of the green signal), outputs are measured characteristics of traffic flows in that region (density and intensity of the traffic flow) and the optimality criterion can be e.g. minimum of exhalations or maximal permeability of the miniregion with some restrictions concerning neighbouring areas. In the basis of such considerations, a good model of the studied variables must stay, with a reliable and numerically stable identification. Such model gives predictions of the variables on which further decisions or automatic control can be grounded. A theoretical background for

[1]Faculty of Transportation Sciences CTU, Na Florenci 25, 110 00 Prague 1. E-mail: nagy@fd.cvut.cz
This research was partly supported by the grants VZ MŠMT 210000023, COST Action 276, GAČR 102/03/0049 and AVČR S1075102.

the task stated above is the approach of Bayesian statistics [1]. This approach has been developed for rather a long time [2, 3, 4, 5, 6] and since 1975 it has been intensively investigated in ÚTIA AV ČR in the department of Adaptive systems [7, 8, 9, 10, 11]. It can be said, that the level of completeness of the theory is considerably high and it has already been practically realized [12, 13]. In the present state, its full power can be applied to transportation area. We can cite [14] as an example of such more or less preliminary work. Other, more specific tasks are solved in diploma and Ph.D. theses. Here, we are going to present the basis of the the Bayesian approach to modelling and identification used for prediction of transportation variables.

## Models

The basic objects for description of an unknown system under uncertainties according to the Bayesian approach are *model of the system* and *model of the parameters* of the system model, which are unknown and estimated. Both models are represented by conditional probability functions (pdf) of the described variables (modelled data or parameters) on condition of historical values of variables, related to them.

### Model of the system

Let us consider a miniregion where we measure $m$ variables. As we are interested only in prediction, we will not distinguish inputs and outputs. The values of the measured variables are recorded in a discrete time instants denoted by $t = 1, 2, \ldots, N$ and each measurement gives us a column vector of data $d_t = (d_{1;t}, d_{2;t}, \ldots, d_{m;t})'$. The system model at time instant $t$ is described by the pdf

$$f(d_t|\varphi_{t-1}, \Theta), \tag{1}$$

where $\varphi_{t-1}$ is a finite regression vector, containing values of historical data that have an influence on $d_t$ and $\Theta$ is a vector of unknown model parameters, through which the model is identified with reality. This model gives a distribution of probabilities of all possible values of the modelled data vector $d_t$.

### Model of the parameters

According to the Bayesian approach, all unknown parameters are considered to be random variables. As for the system model, the description of the parameters at the time instant $t$ is similarly given by the pdf

$$f(\Theta|d(t)), \tag{2}$$

where the symbol $d(t)$ denotes all past measured data, i.e. $d(t) = \{d_t, d_{t-1}, d_{t-2}, \ldots\}$, including so called prior data d(0), which are measured preliminary before the start of the identification. These prior data can be replaced or combined with not measured prior (expert) knowledge. Thus the pdf $f(\Theta|d(0))$ describes our prior knowledge about the unknown parameters either from data or from some other source.

## Identification

The proces of parameter estimation consists in developing the model of parameters from the prior pdf $f(\Theta|d(0))$ to the posterior one $f(\Theta|d(t))$ through the measured data $d_1, d_2, \ldots, d_t$ by means of the system model (1). According to the Bayes rule we can write

$$f(\Theta|d(t)) \propto f(d_t|\varphi_{t-1}, \Theta) f(\Theta|d(t-1)), \tag{3}$$

where the sign $\propto$ denotes proportionality and $t = 1, 2, \ldots, N$. The recursion starts with the prior pdf $f(\Theta|d(0))$. The piece of information carried by a new data item is imbedded into the parametr pdf, in each step. At the end of estimation, the whole posterior pdf of the unknown parameter $\Theta$ is at disposal and it can be used for other tasks.

REMARK: *In case of linear models (1) with normal distribution the functional recursion simplifies into an algebraic one, for distribution statistics, known as least squares method. Similar result can be also obtained for all distributions from the rather wide exponential class.*

## Prediction

Prediction is a value of the modelled variable yet unknown. As for the unknown parameters, it is considered to be a random variable and it is described by its pdf $f(d_t|d(t-1)$. This pdf is similar to that of model (1), but it does not contain unknown parameters $\Theta$ in its condition. This pdf can be obtained from the model pdf (1) and from the results of the current estimation (3) in the following way

$$f(d_t|d(t-1)) = \int f(d_t|\varphi_{t-1}, \Theta) f(\Theta|d(t-1)) d\Theta, \tag{4}$$

where the integral domain is the whole range of $\Theta$.

REMARK: *Notice, that the whole probabilistic description of the parameter $\Theta$ is used in computation of the prediction. Usage of point estimates (conditional means) would not lead to the optimal result.*

*The optimal point prediction can be, if necessary, computed as follows*

$$\hat{d}_t = E[d_t|d(t-1)] = \int d_t f(d_t|d(t-1)) dd_t.$$

## Mixture models

The previous procedure, solving our task of prediction, is feasible practically only for models from exponential class of distribution whose mean value depends linearly on unknown parameters. This assumption is often too restrictive for real systems. Especially in the transportation we can often meet a situation, when the real system exists in several different and nonlinearly interconnected states. As an example we can mention modelling of an evolution of the traffic flow during a working day and weekend, in winter, summer and during holidays etc. If the individual states can be, at least approximately, described by the linear models, a mixture model composed of several sub-models (components) can be used. This model is described by the following pdf

$$f(d_t, c_t|\varphi_{t-1}, \Theta, \alpha), \tag{5}$$

where $c_t$ is a random variable indicating the true (just active) component at the time instant $t$, $\Theta$ is a set of parameters concerning the components and $\alpha$ is a vector of parameters concerning the model of switching between components. If the mixture has $n$ components then $c_t \in \{1, 2, \ldots, n\}$ for each $t$. This random variable is supposed to be independent of all past data and the whole set of component parameters $\Theta$. Its model is assumed in the form

$$f(c_t|\alpha) = \alpha_{c_t}, \tag{6}$$

with $\alpha_i > 0 \ \forall i, \quad \sum_{i=1}^{n} \alpha_i = 1$ being stationary probabilities of individual state activities.

Accepting this, the system model entering the Bayes rule (3) has a form of a mixture of components

$$f(d_t|d(t), \Theta, \alpha) = \sum_{i=1}^{n} \alpha_i f(d_t|\varphi_{t-1}, \Theta_i, c_t = i), \tag{7}$$

where $\Theta_i$ are parameters of the $i$-th component, for $i = 1, 2, \ldots, n$. It can be seen that direct use of this model in Bayes rule (3) leads to unfeasible computations as it repetitively produces products of sums. Thus, an approximation is necessary. It, roughly speaking, consists in (i) "pretending" that the variable $c_t$ is known with "deterministic Kronecker pdf $\delta(i, c_t)$" and (ii) approximating this pdf, in reality unknown, by its optimal point estimate conditional mean value $E[\delta(i, c_t)|d(t)] = w_{i,t}$, where $i = 1, 2, \ldots, n$. It can be computed on a basis of already existing or estimated pdfs. The items of the approximation $w_{i,t}$ at time $t$ represent actual weights of the components, i.e. probabilities of activities of individual components based on the knowledge of data up to and including the time instant $t$.

REMARK: *In case of linear components within the exponential class of distributions this estimation procedure leads to the weighted least squares performed for each component separately with the corresponding weight* $w_{i,t}, \ i = 1, 2, \ldots, n$.

## State classification

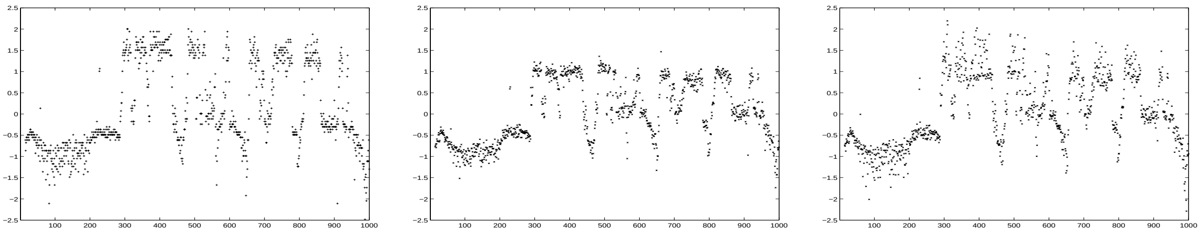
The approximated estimation, shortly mentioned above, is based on a computation of the actual probabilistic weights assigned to all components determining probabilities of their activities with respect to the actual data item. In other words, the corresponding weight of each component specifies the probability that the current data item is correctly modelled just by this component. So, in addition to mixture estimation, the process gives also detection of the active state of the multi-state system, which is mostly called state classification. Besides data prediction, the state classification is another mighty tool of the estimation that is very useful, especially in transportation.

## Examples

### Prediction with regression and mixture models

In this example, we are going to demonstrate the theoretical approach by one of the most natural application: multi-step prediction of the traffic flow intensity at a single point of vehicular communication. This task and its solution is very important building block not only for an automatic control but it can also serve an evidence for transportation operators in their decision making or to drivers as an information about the situation on roads.

As a data sample, 2000 values of the traffic flow intensity is used. The data were measured in the centre of Prague, in Legerova street. Their source are magnetic detectors placed under the surface of the road. The samples are taken each 5 minutes. As one-step-ahead prediction is not very challenging task, we predict over six step, that means half an hour ahead prediction. This task is not so easy, but it should be possible, from the practical point of view. The normal dynamic regression model of the first order (one delayed modelled variable is included into the regression vector) is used and for comparison, a mixture of such models is considered. The mixture was automatically initialized and four components were found. The results are presented in the following pictures



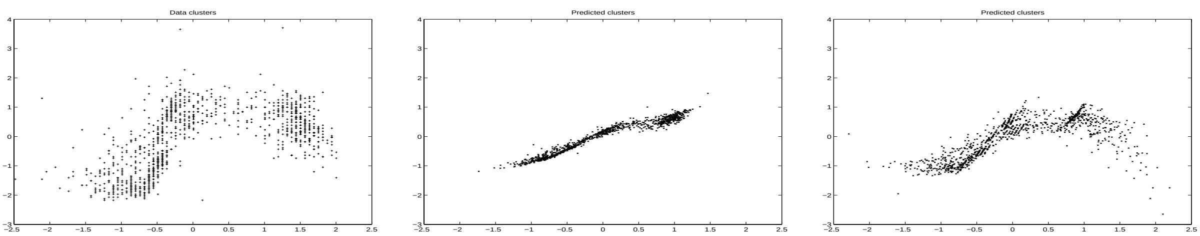
Measured data | Regression predictions | Mixture predictions

**Figure 1:** Prediction of the traffic flow intensity using regression model and mixture.

The mere look at the data and predictions courses could be misleading. That is why a numerical evaluation was computed end expressed by the prediction error (PE) coefficient. It is defined as a square root of quadratic prediction error divided by standard deviation of the data. The values are 0.481 for the regression model and 0.355 for the mixture. The look at the picture as well as the PE coefficients show, that the mixture gives better results.

### Clustering with regression and mixture models

According to the results of the previous example, it could be said that there is not so big difference between regression and mixture models. Nevertheless, the difference can be clearly seen from the results of clustering. Now, not only the intensity but also the density of the traffic flow is considered. Each measured data couple [”density”; ”intensity”] fully describes the state of the traffic in the point of measuring and can be considered an actual working point of the transportation system at this point. These two variables are bound and the relation ”intensity” $= func($”density”$)$ can be approximately described by a concave parabola. The measured working points (with noise, indeed) are scattered around the ideal shape of the parabola and provide clusters corresponding to individual states of the transportation system. These states correspond to the well known level of service degrees. For estimation of this two-dimensional variable, the same regression model and their mixture has been used. The results are shown on the Figure 2.



Data clusters | Predicted clusters regression model | Predicted clusters mixture model

**Figure 1:** Prediction of traffic flow intensity using regression model and mixture.

Here the substantial difference between the single regression model and the mixture model can clearly be seen.

## Conclusions

The Bayesian approach to modelling and identification is a powerful tool, especially in case of its real application in practice. In this paper, the most straightforward transportation application, prediction of the traffic flow, is presented. The generality of the method and its considerably advanced numerically stable algorithms allow to apply it for solving many other problems, aiming at support of transportation operators or automatic control of large transportation systems.

## References


[1] G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis*, Addison-Wesley Publ. Comp., 1973.

[2] D. V. Lindley, *Bayesian Statistics, a Review*, SIAM, 1971.

[3] S.R. Dalal and G.J. Hall Jr., "On approximating parametric Bayes models by non-parametric Bayes models", *The Annals of Statistics*, vol. 8, pp. 664–672, 1980.

[4] A.F.M. Smith and A.E. Gelfand, "Bayesian statistics without tears: a smapling-resampling perspective", *The American Statistician*, vol. 46, pp. 84–88, 1992.

[5] J.M. Bernardo and A.F.M. Smith, *Bayesian theory*, John Wiley & Sons, Chichester, New York, Brisbane, Toronto, Singapore, 1997, 2nd edition.

[6] S. Richardson and P.J. Green, "On Bayesian analysis of mixtures with an unknown number of components, with discussion", *Journal of the Royal Statistical Society, Series B*, vol. 59, no. 4, pp. 731–792, 1997.

[7] V. Peterka and M. Kárný, "Bayesian system classification", in *Preprints of the 5th IFAC Symposium on Identification and System Parameter Estimation*, vol. 1, pp. 349–356. Darmstadt, 1979.

[8] V. Peterka, "Bayesian system identification", in *Trends and Progress in System Identification*, P. Eykhoff, Ed., pp. 239–304. Pergamon Press, Oxford, 1981.

[9] M. Kárný and K.M. Hangos, "Approximation of Bayes rule", in *Preprints of the 7th IFAC/IFIP Symposium on Identification and System Parameter Estimation*. York, 1985.

[10] M. Kárný, J. Kadlec, and E. L. Sutanto, "Quasi-Bayes estimation applied to normal mixture", in *Preprints of the 3rd European IEEE Workshop on Computer-Intensive Methods in Control and Data Processing*, J. Rojíček, M. Valečková, M. Kárný, and K. Warwick, Eds., Praha, September 1998, pp. 77–82, ÚTIA AV ČR.

[11] M. Kárný, I. Nagy, and J. Novovičová, "Quasi-Bayes Approach to Multi-Model Fault Detection and Isolation", Tech. Rep. 1974, ÚTIA AV ČR, Praha, 1999.

[12] J. Heřmanská, Z. Prouza, I. Němec, and M. Kárný, "Bayes-based dose equivalent estimation using multisphere data", *Kernenergie*, vol. 30, no. 10, pp. 409–416, 1987.

[13] M. Kárný, A. Halousková, and I. Nagy, "Modelling, identification and adaptive control of cross-direction basis weight of paper sheets", in *Proceedings IEE Int. Conf. CONTROL 88*, pp. 159–164. The Chameleon Press Limited, Wandsworth, London, Oxford, 1988.

[14] I. Nagy, M. Kárný, P. Nedoma, and Š. Voráčová, "Bayesian estimation of traffic lane state", *Adaptive Control and Signal Processing*, vol. 17, no. 1, pp. 51–65, 2003.