BAYESIAN FILTERING WITH DISCRETE-VALUED STATE

Evgenia Suzdaleva, Ivan Nagy, Lenka Pavelková

Department of Adaptive Systems
Institute of Information Theory and Automation of the Academy of Sciences of the Czech Republic
Pod vodárenskou věží 4, 18208 Prague, Czech Republic

ABSTRACT

The paper deals with estimation of a state with discrete values. The proposed estimation technique is evolved as an application of Bayesian filtering to a state-space model with discrete distribution. The example of filtering is shown with Bernoulli distributions. The considered problem is one of the items aiming at filtering with mixed continuous and discrete state. Illustrative experiments demonstrate the filtering with discrete simulated data from the traffic control area, which is a potential application domain of the research.

Index Terms— State-space model, Bayesian filtering, discrete distribution

1. INTRODUCTION

The paper deals with state estimation of a system, whose state is a discrete-valued one. Discrete state estimation is an important task, solution of which is highly desired in many application domains (econometrics, traffic flow control, biomedical studies, robotics, radiation protection etc). The target application of the presented paper is traffic control, where discrete variables are used for modeling of a driver state, level of dangerous driving, visibility, surface of a road, existence of car queues etc. It is necessary to highlight that the global objective of the present research is the filtering with mixed (continuous and discrete) data. Thus, the problem discussed in the present paper is one of the subtasks for reaching this aim.

Majority of techniques found in the context of discrete state estimation are based on hidden Markov models [1, 2, 3]. Estimation with mixed (continuous and discrete) data is a known problem in the field of logistic regression [4]. Analyzing the current state-of-the-art concerning the discussed filtering problem, one can find the stochastic approximations (particle filters) [5] at first sight to be a suitable solution. However, the particle filters require a high computational cost and fail to win against analytically tractable solutions. Thus, a reliable analytical solution with an adequate computational complexity is needed.

Thanks to GA ČR 201/06/P434 and MŠMT 1M0572 for funding.

The first results [6] of the present research were concerned with entry-wise organized Bayesian filtering with the mixed-type state. Bayesian filtering [7] applied to continuous Gaussian state-space model provides Kalman filter [8]. The paper explores the filtering specialization to discrete distributions in order to involve it later to the mixed state estimation.

Layout of the paper is as follows. Section 2 provides basic facts about the probabilistic state-space model and Bayesian filtering. Section 3 is devoted to the filtering with discrete state-space model, where the involved variables have Bernoulli distribution. Section 4 provides illustrative experiments concerned with the traffic control area. Remarks and plans of future work in Section 5 conclude the paper.

2. PRELIMINARIES

The probabilistic state-space model of the system, a state of which has to be estimated, is provided with the help of the following conditional probability (density) functions (p(d)fs). The *observation model*, specified by

$$f\left(y_t|\, u_t, x_t\right), \tag{1}$$

relates the system output y_t to the system input u_t and the unobserved system state x_t at discrete time moments $t \in t^* \equiv \{0, \ldots, \mathring{t}\}$, where $\mathring{t}$ is the cardinality of the set t^* and $\equiv$ means equivalence. The *state evolution model*

$$f\left(x_{t+1}|\, u_t, x_t\right), \tag{2}$$

describes the evolution of the system state x_t. The estimation of the finite-dimensional system state calls for application of Bayesian filtering. Bayesian filtering, estimating the system state, includes the following coupled formulas.
Data updating

$$\begin{aligned} f\left(x_t \middle| d^t\right) &= \frac{f\left(y_t|\, u_t, x_t\right) f\left(x_t \middle| d^{t-1}\right)}{\int f\left(y_t|\, u_t, x_t\right) f\left(x_t \middle| d^{t-1}\right)\, dx_t}, \\ &\propto f\left(y_t|\, u_t, x_t\right) f\left(x_t \middle| d^{t-1}\right), \end{aligned} \tag{3}$$

($\propto$ means proportionality) incorporates the experience contained in the data d^t, where $d^t = (d_0, \ldots, d_{\mathring{t}})$ and $d_t \equiv (y_t, u_t)$.

Time updating

$$f\left(x_{t+1}\middle|d^t\right)=\int f\left(x_{t+1}\middle|u_t,x_t\right)f\left(x_t\middle|d^t\right)\,dx_t, \qquad (4)$$

fulfills the state prediction. The filtering does not depend on the control strategy $\{f(u_t|d^{t-1})\}_{t\in t^*}$ but on the generated inputs only. The prior pdf $f(x_0)$, which expresses the subjective prior knowledge on the state x_0, starts the recursions. Application of (3)-(4) to linear Gaussian state-space model provides Kalman filter [9].

Decomposition of models (1)-(2) via the chain rule [8] represents them as the product of p(d)fs of the individual state and output entries.

$$f(y_t|u_t,x_t) = \prod_{j=1}^{\mathring{y}} f(y_{j;t}|y_{j+1:\mathring{y};t},u_t,x_{1:\mathring{x};t}), \qquad (5)$$

$$f(x_{t+1}|u_t,x_t) = \prod_{i=1}^{\mathring{x}} f(x_{i;t+1}|x_{i+1:\mathring{x};t+1},u_t,x_{1:\mathring{x};t}), \qquad (6)$$

where $\mathring{y}$ and $\mathring{x}$ denote number of entries of column vectors y_t and x_t respectively, $j=\{1,\ldots,\mathring{y}\}$, $i=\{1,\ldots,\mathring{x}\}$. A notation in the form $x_{i+1:\mathring{x};t}$ in (5)-(6) denotes a sequence of the vector entries from $(i+1)$ to $\mathring{x}$, i.e. $\{x_{i+1;t},x_{i+2;t},\ldots,x_{\mathring{x};t}\}$, which is empty, when $(i+1)>\mathring{x}$. The known input u_t is factorized within computations. The filtering (3)-(4) with decomposed models (5)-(6) preserves the entry-wise form of the state estimate $f\left(x_{t+1}\middle|d^t\right)=\prod_{i=1}^{\mathring{x}} f\left(x_{i;t+1}\middle|x_{i+1:\mathring{x};t+1},d^t\right)$ [6]. This enables its exploitation with various distributions of entries. The solution specialized to Gaussian state-space model can be found in [6]. The present paper considers the discrete state-space model.

3. FILTERING WITH DISCRETE-VALUED STATE

Let's assume that the variables y_t, x_t and u_t are of a discrete-valued nature. It means that one must investigate application of the filtering (3)-(4) to models with discrete distributions. In general, the multivariate discrete variables would lead to decomposition of the state-space model according to (5)-(6). When speaking about the state, one must consider the state vector $x_t\equiv[x_{1;t},\ldots,x_{\mathring{x};t}]'$ with finite, preferably small $\mathring{x}$, where each entry $x_{i;t}$ with $i=\{1,\ldots,\mathring{x}\}$ has a set of possible values $\{s_1,s_2,\ldots,s_n\}$ with finite number n. However, the estimation of the state vector with discrete entries is significantly simplified via a special mapping of the multivariate state to a scalar one. The mapping is proposed via a determination of the new set of possible values for the scalar discrete state. The new set is constructed so that each possible combination of values of all entries $x_{i;t}$, i.e. $\{(x_{1;t}=s_1,x_{2;t}=s_1,\ldots,x_{\mathring{x};t}=s_1),(x_{1;t}=s_2,x_{2;t}=s_1,\ldots,x_{\mathring{x};t}=s_1),\ldots,(x_{1;t}=s_n,x_{2;t}=s_n,\ldots,x_{\mathring{x};t}=s_n)\}$, is denoted as the new scalar value, belonging to this set, i.e. $\{k_1,k_2,\ldots,k_N\}$, N is a number of combinations[1]. The mapping provides necessary reduction of dimension of the state. The similar reducing of the dimension is applied to the multivariate output and input. It means, that the filtering (3)-(4) can be used directly with models (1)-(2).

Let's assume, that the scalar variables (either due to the mapping or naturally) with the finite set of possible discrete values have to be considered for the filtering. Generally, discrete multinomial distribution could have been used for the models (1)-(2) with these variables. For the sake of simplicity, a number of possible values in the set is restricted by two, which leads to Bernoulli distribution.

Let's consider the observation model (1) described by the Bernoulli distribution, shown in Table 1, where α with corresponding indices denotes a probability (assumed to be known) of taking the possible values by the output, conditioned on values of the state and input. The set of discrete values for all the variables is given as $\{k_1,k_2\}$, and the probability $\alpha_{2|ij}$ is always defined as $(1-\alpha_{1|ij})$. The Bernoulli distribution from

Table 1. Bernoulli observation model

	$y_t=k_1$	$y_t=k_2$
$u_t=k_1,x_t=k_1$	$\alpha_{1\|11}$	$\alpha_{2\|11}$
$u_t=k_2,x_t=k_1$	$\alpha_{1\|21}$	$\alpha_{2\|21}$
$u_t=k_1,x_t=k_2$	$\alpha_{1\|12}$	$\alpha_{2\|12}$
$u_t=k_2,x_t=k_2$	$\alpha_{1\|22}$	$\alpha_{2\|22}$

Table 1 written with the help of Kronecker delta is presented in the following product form

$$f\left(y_t\middle|u_t,x_t\right)=\prod_{u_t,x_t\in\{k_1,k_2\}}\alpha_{1|u_tx_t}^{\delta(y_t,k_1)}\alpha_{2|u_tx_t}^{\delta(y_t,k_2)}, \qquad (7)$$

where Kronecker delta expresses a choice of an occurred value from the possible ones. Similarly, the state evolution model (2) is related to Bernoulli distribution, provided in Table 2, where respective β denote a known probability of taking the possible values of the state, conditioned on its previous values and on the input, and $\beta_{2|ij}=(1-\beta_{1|ij})$. The product form

Table 2. Bernoulli state evolution model

	$x_{t+1}=k_1$	$x_{t+1}=k_2$
$u_t=k_1,x_t=k_1$	$\beta_{1\|11}$	$\beta_{2\|11}$
$u_t=k_2,x_t=k_1$	$\beta_{1\|21}$	$\beta_{2\|21}$
$u_t=k_1,x_t=k_2$	$\beta_{1\|12}$	$\beta_{2\|12}$
$u_t=k_2,x_t=k_2$	$\beta_{1\|22}$	$\beta_{2\|22}$

of the distribution from Table 2 is as follows.

$$f\left(x_{t+1}\middle|u_t,x_t\right)=\prod_{u_t,x_t\in\{k_1,k_2\}}\beta_{1|u_tx_t}^{\delta(x_{t+1},k_1)}\beta_{2|u_tx_t}^{\delta(x_{t+1},k_2)}. \qquad (8)$$

[1]For example, for $x_t\equiv[x_{1;t},x_{2;t}]'$, with $x_{1;t}\in\{s_1,s_2\}$ and $x_{2;t}\in\{s_1,s_2\}$, the possible values $\{(s_1,s_1),(s_2,s_1),(s_1,s_2),(s_2,s_2)\}$ can be denoted as the new set $\{k_1,k_2,k_3,k_4\}$.

The prior probabilities for the initial discrete state are chosen as $p_{1(t)}$ for value $x_t = k_1$ and $p_{2(t)} = (1 - p_{1(t)})$ for $x_t = k_2$. Thus, the form of the prior Bernoulli distribution is defined as

$$f\left(x_t \middle| d^{t-1}\right) = p_{1(t)}^{\delta(x_t,k_1)}(1-p_{1(t)})^{\delta(x_t,k_2)}. \quad (9)$$

The estimation of the discrete state is proposed as the direct application of Bayesian filtering (3)-(4) to the Bernoulli state-space model (7)-(8) with incorporation of Bernoulli prior (9). According to the mentioned relations, formula (3) with the substituted Bernoulli distributions (7) and (9) takes the following form, providing the updating of the state estimate by actual measurements

$$\begin{aligned} f\left(x_t \middle| d^t\right) &= \frac{\prod_{u_t,x_t\in\{k_1,k_2\}} \alpha_{1|u_t x_t}^{\delta(y_t,k_1)} \alpha_{2|u_t x_t}^{\delta(y_t,k_2)}}{\sum_{x_t\in\{k_1,k_2\}} \prod_{u_t,x_t\in\{k_1,k_2\}} \alpha_{1|u_t x_t}^{\delta(y_t,k_1)}} \quad (10) \\ & \frac{p_{1(t)}^{\delta(x_t,k_1)}(1-p_{1(t)})^{\delta(x_t,k_2)}}{\alpha_{2|u_t x_t}^{\delta(y_t,k_2)} p_{1(t)}^{\delta(x_t,k_1)}(1-p_{1(t)})^{\delta(x_t,k_2)}} \\ &= \bar{p}_{1(t)}^{\delta(x_t,k_1)} \bar{p}_{2(t)}^{\delta(x_t,k_2)}, \quad (11) \end{aligned}$$

where integration in the denominator is replaced by regular summation. The probabilities to be substituted in the data updating (10) are chosen from Table 1 according to the actual values of the output and the input. $\bar{p}(\cdot)$ in (11) denotes the intermediate results of the filtering (data-updated probabilities).

The Bayesian time updating (4) with Bernoulli distribution (8) and the intermediate result (11) takes the following form

$$\begin{aligned} f\left(x_{t+1} \middle| d^t\right) &= \sum_{x_t\in\{k_1,k_2\}} \prod_{u_t,x_t\in\{k_1,k_2\}} \beta_{1|u_t x_t}^{\delta(x_{t+1},k_1)} \\ & \beta_{2|u_t x_t}^{\delta(x_{t+1},k_2)} \bar{p}_{1(t)}^{\delta(x_t,k_1)} \bar{p}_{2(t)}^{\delta(x_t,k_2)}, \quad (12) \end{aligned}$$

which provides the resulting state estimate as the following Bernoulli distribution

$$f\left(x_{t+1} \middle| d^t\right) = p_{1(t+1)}^{\delta(x_{t+1},k_1)}(1-p_{1(t+1)})^{\delta(x_{t+1},k_2)}, \quad (13)$$

with the updated probability of value $x_{t+1} = k_1$

$$\begin{aligned} p_{1(t+1)} &= \sum_{x_t\in\{k_1,k_2\}} \prod_{u_t,x_t\in\{k_1,k_2\}} \beta_{1|u_t x_t}^{\delta(x_{t+1},k_1)} \bar{p}_{1(t)}^{\delta(x_t,k_1)} \bar{p}_{2(t)}^{\delta(x_t,k_2)} \quad (14) \\ &= \beta_{1|u_t 1}\bar{p}_{1(t)} + \beta_{1|u_t 2}\bar{p}_{2(t)}, \end{aligned}$$

calculated according to the known values of the input and substitution of the corresponding probabilities from Table 2. The probability of value k_2 is obtained as $p_{2(t+1)} = (1 - p_{1(t+1)})$, or can be calculated directly:

$$\begin{aligned} p_{2(t+1)} &= \sum_{x_t\in\{k_1,k_2\}} \prod_{u_t,x_t\in\{k_1,k_2\}} \beta_{2|u_t x_t}^{\delta(x_{t+1},k_2)} \bar{p}_{1(t)}^{\delta(x_t,k_1)} \bar{p}_{2(t)}^{\delta(x_t,k_2)} \quad (15) \\ &= \beta_{2|u_t 1}\bar{p}_{1(t)} + \beta_{2|u_t 2}\bar{p}_{2(t)}. \end{aligned}$$

Relation (15) is obtained similarly according to the input values and substitution of the probabilities from Table 2. The resulting Bernoulli distribution (13) is taken as the prior one to be incorporated into the next step of the discrete state estimation (10) with actual available measurements.

<u>*Algorithm*</u>

1. *Set prior $p_{2(t=0)}$ for value $x_t = k_2$*
2. *Get actual data y_t, u_t*
3. *Choice from Table 1.*
   If $y_t = k_2$ & $u_t = k_2$, then $\alpha = \alpha_{2|22}, g = \alpha_{2|21}$, where g is an auxiliary variable for normalization, conditioned on $x_t = k_1$,
   else if $y_t = k_2$ & $u_t = k_1, \alpha = \alpha_{2|12}, g = \alpha_{2|11}$,
   else if $y_t = k_1$ & $u_t = k_2, \alpha = \alpha_{1|22}, g = \alpha_{1|21}$,
   else if $y_t = k_1$ & $u_t = k_1, \alpha = \alpha_{1|12}, g = \alpha_{1|11}$,
   end
4. *Data updating.*
   $\bar{p}_{2(t)} = \frac{\alpha \cdot p_{2(t)}}{\alpha \cdot p_{2(t)} + g \cdot (1-p_{2(t)})}, \bar{p}_{1(t)} = 1 - \bar{p}_{2(t)}$
5. *Choice from Table 2.*
   If $u_t = k_1$, then $\beta = \beta_{2|12}, d = \beta_{2|11}$, where d is an auxiliary variable, conditioned on $x_t = k_1$,
   else if $u_t = k_2, \beta = \beta_{2|22}, g = \beta_{2|21}$,
   end
6. *Time updating.*
   $p_{2(t)} = \beta \cdot \bar{p}_{2(t)} + d \cdot \bar{p}_{1(t)}, p_{1(t)} = 1 - p_{2(t)}$. *Go to* 2.

4. EXPERIMENTS

The illustrative experiments are concerned with estimation of the state of driver tiredness. Currently the experiments use the data simulated with the help of random generator and Tables 1-2. However, the planned practical experiments will be based on realistic so-called "bio-traffic" data available from traffic experts. Specializing the traffic data to models (7)-(8), one considers the following variables. The output y_t expresses a measurable level of dangerous driving, which in the simplified Bernoulli form belongs to the set $\{k_1 =$ not close to dangerous driving, $k_2 =$ close to dangerous driving$\}$. The unobserved state is a driver tiredness $x_t \in \{k_1 =$ not close to be tired, $k_2 =$ close to be tired$\}$. The control input u_t is understood as advices, which the driver gets from the built-in car computer with the following simplified Bernoulli set of possible values $\{k_1 =$ no warnings, $k_2 =$ recommendation to reduce a speed, stop a car, have a break$\}$. The probabilities used in Tables 1-2 respectively were provided by traffic experts determined heuristically as follows: $\alpha_{1|11} = 0.99$, $\alpha_{2|11} = 0.01$, $\alpha_{1|21} = 0.04$, $\alpha_{2|21} = 0.96$, $\alpha_{1|12} = 0.79$, $\alpha_{2|12} = 0.21$, $\alpha_{1|22} = 0.01$, $\alpha_{2|11} = 0.99$; $\beta_{1|11} = 0.99$, $\beta_{2|11} = 0.01$, $\beta_{1|21} = 0.64$, $\beta_{2|21} = 0.36$, $\beta_{1|12} = 0.21$, $\beta_{2|12} =$

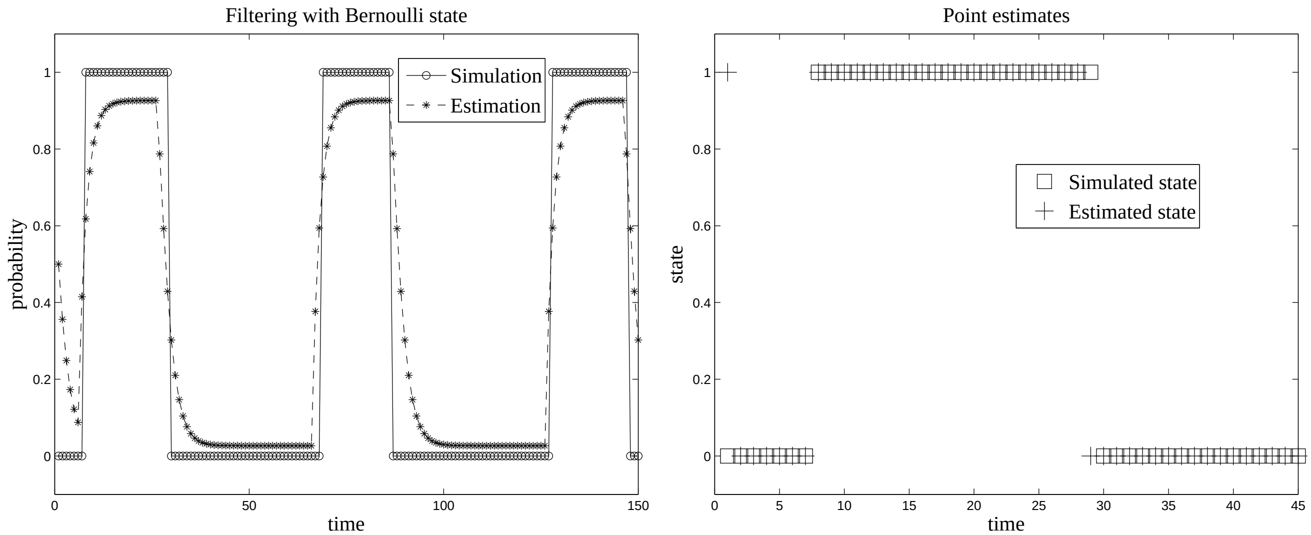

Fig. 1. Estimated probability of discrete state $1 \equiv$ (driver is close to be tired) (left) and point estimates of states 0 and 1 (right)

0.79, $\beta_{1|22} = 0.03$, $\beta_{2|11} = 0.97$. The prior probability $p_{2(t)}$ of the driver tiredness was chosen as 0.5. Numerically, the state k_1 was identified with value 0, while k_2 was taken as 1. Fig. 1 (left) demonstrates the estimated probability of event 1, denoting that the driver is close to be tired, against the simulated state. For better illustration, Fig. 1 (right) shows the point estimates of states 0 and 1.

5. CONCLUSION

The present paper is focused on the relatively simple task of the filtering with the Bernoulli state-space model. Extension of the set of possible values from two Bernoulli outcomes up to some finite, preferably small number N will require to specify the Kronecker-based product form of models (7)-(8) and prior distribution (9). However, the logic of the approach extended up to other related discrete distributions (at least, multinomial) remains similar. It should be noted, that the considered problem has been described for the case with known parameters of the discrete state-space model, which significantly simplified the solution. It is clear that even extended solution with multinomial distributions will cause additional problems of setting (probably, off-line estimating) the parameters. The plans of future work will be devoted to the mentioned problems aiming at general objective of the estimation of mixed states.

6. REFERENCES

[1] S. Dey L. Shue, B.D.O. Anderson, "Exponential stability of filters and smoothers for hidden markov models," *IEEE Transactions on Signal Processing*, vol. 46(8), pp. 2180 – 2194, August 1998.

[2] Chun Yang, "On discrete hidden markov state estimation," in *Proceedings of the American Control conference*, Seattle, WA, USA, June 21-23 1995, vol. 1, pp. 12–13.

[3] S. Di Cairano, K. H. Johansson, A. Bemporad, and R. M. Murray, *Hybrid Systems: Computation and Control*, vol. 4981/2008, chapter Discrete and Hybrid Stochastic State Estimation Algorithms for Networked Control Systems, pp. 144–157, Springer, 2008.

[4] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, Wiley-Interscience, 2001.

[5] V. Šmídl and A. Quinn, "Variational Bayesian filtering," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 5020–5030, 2008.

[6] E. Suzdaleva, "Filtering with mixed continuous and discrete states: special case," Tech. Rep. 2246, ÚTIA AV ČR, Praha, January 2009, Draft of paper.

[7] M. Kárný, J. Böhm, T. V. Guy, L. Jirsa, I. Nagy, P. Nedoma, and L. Tesař, *Optimized Bayesian Dynamic Advising: Theory and Algorithms*, Springer, London, 2005.

[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.S. Grewal and A.P. Andrews, *Kalman Filtering: Theory and Practice Using MATLAB. 2nd edition*, Wiley, 2001.