Akademie věd České republiky
Ústav teorie informace a automatizace, v.v.i.

Academy of Sciences of the Czech Republic
Institute of Information Theory and Automation

RESEARCH REPORT

E. Suzdaleva

Overview of models with mixed-type data

2208 December 2007

GAČR 201/06/P434

ÚTIA AVČR, P.O.Box 18, 182 08 Prague,
Czech Republic
Fax: (+420)286890378, http://www.utia.cz, E-mail: utia@utia.cas.cz

Contents

1 Introduction **5**
1.1 Motivation and addressed problem 5
1.2 Paper layout 5

2 Overview of joint modelling of mixed data **7**
2.1 Models with mixed data 7
2.2 Mixed data representation in latent variable models 7
2.3 Maximum likelihood estimation with models with mixed data 8
2.4 Mixed data modelling based on discriminant analysis 8
2.5 Bayesian estimation with mixed data 9
2.6 Non-parametric estimation with mixed data by using the kernel methods 10

3 Mixed data in mixed mixtures **11**
3.1 Continuous factor output and discrete quantities 11
3.2 Discrete factor output and continuous and discrete quantities 11
3.3 Conclusion 12

Chapter 1

Introduction

1.1 Motivation and addressed problem

The work presents an overview of models and estimators, dealing with mixed-type (discrete and continuous) data. Mixed-type data handling (including models in discrete-valued outputs, depending on continuous variables) is known to be hard problem addressed within completely different context of logistic regression. Models and estimators with mixed data based on the logistic approach as well as mixed mixtures are briefly described. The overview prepares the necessary support base for the design of filters estimating the individual, ideally mixed-type, state entries [1, 2], which is a global aim of the research.

1.2 Paper layout

The paper is organized in the following way. Chapter 2 presents the overview of research works, dealing with mixed data. Different approaches to mixed data, such as maximum likelihood estimators, discrimination analysis, Bayesian estimation and kernel methods are described. Chapter 3 provides the overview of the mixed-data problem in mixed mixtures. Remarks in Section 3.3 close the report.

Chapter 2

Overview of joint modelling of mixed data

2.1 Models with mixed data

Mixed-type data handling (including models in discrete-valued outputs, depending on continuous variables) is known to be hard problem addressed within completely different context of logistic regression. A number of flexible models for such data is rather limited. One of the most known works in this field was [3], which proposed the general "location model", based on multinomial model for the discrete outcomes, and Gaussian multivariate model for the continuous outcomes, conditional on the discrete outcomes. In contrast to that, the paper [4] described the model, in which the marginal distribution of the continuous outcomes was Gaussian, and was multiplied by logistic representation for the conditional distribution of the binary responses, given the continuous outcomes. Later, the work [5] compared a number of different models, based on these two factorization of the joint distribution.

The "location model", proposed in [3], inspired many studies in the context of discrimination and classification [6]. It has been extended by [7] and [8] to accommodate missing data. Later, the paper [9] proposed the extension of the general "location model" for bivariate discrete and continuouses responses, that allowed the marginal means to be related to covariates using a linear link function for the continuous responses and any generalized linear link function to the discrete responses. The solution, proposed in [9] was restricted by a special case of bivariate, where one of the responses is discrete and another is continuous. Moreover, it was assumed, that there were no missing data. In practice there are more than two variable of interest, and the responses can be both continuous and discrete.

The multivariate extension of the model in [9] has been proposed in [10]. It described the likelihood-based approach for analyzing mixed continuous and discrete responses. The work [10] was concentrated on marginal regression models, where the marginal expectation of the responses vector is related to a set of covariates by known link functions. With binary responses the natural choice is a logit link function. With continuous responses it is naturally to use linear link function. To account the association between responses, the joint distribution of mixed discrete and continuous responses is considered. Very natural partition of responses into discrete and continuous sets, yielding two possible types of models depends on whether the conditioning variable are discrete or continuous. The main advantages of the proposed model [10] were the following ones. First, both sets of regression parameters in the model have marginal interpretations. It means, that the resulted parameter estimates have the same regression parameters interpretation, as in the case when each variable were considered separately. The second important feature was that the proposed approach yields the maximum likelihood estimates of the marginal mean parameters, which are robust to misspecification between the outcomes and to misspecification of the joint distribution of the outcomes generally.

2.2 Mixed data representation in latent variable models

For analysis of mixed data in psychometrical, medical and economic research the latent variable models are often used. The relationships between manifest (observable) and latent (unobservable) variables based on mixed (discrete and continuous) data are investigated. The discrete data are defined through the underlying continuous variables with some unknown threshold parameters. The use of discrete (polytomous, binary, categorical) data are mostly caused by the scale in medical, social, psychometrical research, for example, it can be variants from the scale "getting worse, no change, getting better" after medical treatment or "agree, no opinion, disagree", etc. The data are mostly presented by random observation $({x_i}'; {z_i}') = (x_{1i} \dots x_{\mathring{x}i}; z_{1i} \dots z_{\mathring{z}i})$, measured on individual i, in which x_{hi} and z_{ki} are continuous and discrete data respectively. The discrete random vector

$z = (z_{1i} \ldots z_{\mathring{z}i})'$ is related to the underlying latent continuous random vector $y = (y_{1i} \ldots y_{\mathring{y}i})'$, $\mathring{z} = \mathring{y}$, through a set of unknown thresholds in the following way:

$$z = \begin{bmatrix} z_1 \\ . \\ . \\ . \\ z_{\mathring{z}i} \end{bmatrix}, \text{ if } \begin{matrix} \alpha_{1,z_1} < y_1 \leq \alpha_{1,z_1+1} \\ . \\ . \\ . \\ \alpha_{1,z_{\mathring{z}}} < y_{\mathring{y}} \leq \alpha_{1,z_{\mathring{z}}+1} \end{matrix}, \tag{2.1}$$

where z_k is an integer in $\{0, 1, \ldots, b_k\}$. Random observation $(x_{1i} \ldots x_{\mathring{x}i}; y_{1i} \ldots y_{\mathring{y}i})$ is assumed to be related to latent vector ξ_i with distribution $\mathcal{N}(0, \Omega)$.

Such a representation of mixed data is widely spread and used in a variety of works.

2.3 Maximum likelihood estimation with models with mixed data

One of the estimators, being used with latent variable models based on mixed data, is the expectation-maximization (EM) algorithm and its different modifications. It is exploited for finding maximum likelihood estimates (MLE) of model parameters. The general idea of the EM algorithm lies in performing an expectation (E) step, which computes an expectation of the likelihood by including the latent variables as if they were observed, and a maximization (M) step, which computes the maximum likelihood estimates of the parameters by maximizing the expected likelihood found on the E step. The parameters found on the M step are then used to begin another E step, and the process is repeated. Making a brief overview of the works, dealing with the EM algorithm for latent variable models with mixed-type data, one can note the following papers.

The work [11] deals with latent variable models with mixed continuous and polytomous data. It considers the general structural equation model for the mixed outcomes. To estimate thresholds, parameters in covariance matrix as well as scores of latent variables the paper [11] proposes the Monte Carlo-EM algorithm, utilizing the Gibbs sampler and obtaining the maximum likelihood estimates.

The paper [12] describes the similar problem, but with two-level models. It means, that the hierarchically structured data are used, which are collected from the units nested within large clusters. The previous methods can not be used for those two-level models. The paper proposes the Monte-Carlo EM approach that can handle binary data, involving two categories with only one threshold.

The paper [13] considers the problem of estimating a distance between the multivariate normal populations, when subset of measurements is observed as the ordered categorical responses (discrete-valued data). The proposed solution results in maximal likelihood estimates (MLE) of the Mahalanobis distance between populations, defined as the sum of MLEs of its continuous and discrete components. The first of them is based on the observed data, while the second discrete component of the Mahalanobis distance deals with the thresholds.

The paper [14] proposes a latent variable model, that applies to multivariate outcomes from any exponential family, but outcome do not have to come from the same family. The method can handle multiple latent variables. The model does not only allow to multiple testing, but also can handle exposures, which discrete, categorical or continuous. To implement the model, the paper uses the modified EM algorithm with a simple Monte-Carlo expectation or numerical integration techniques, such as Gauss-Hermite quadrature to approximate E-step, which is not necessarily of closed form.

The work [15] deals with regression models for analyzing clustered binary and continuous outcomes under an assumption of exchangeability. In applications to areas such as developmental toxicity studies, where discrete and continuous measurements are recorded on each fetus, or clinical ophthalmologic trials, where different types of observations are made on each eye, the assumption that data within cluster are exchangeable is often very reasonable. The paper uses this assumption to formulate fully parametric regression models for clusters of bivariate data with binary and continuous components. The regression models proposed have marginal interpretations and reproducible model structures. Tractable expressions for likelihood equations are derived and iterative schemes are given for computing efficient estimates (MLEs) of the marginal mean, correlations, variances and higher moments. The application of the exchangeable procedure is demonstrated on a developmental toxicity study involving fetal weight and malformation data.

2.4 Mixed data modelling based on discriminant analysis

Discriminant analysis is a technique for classifying a set of observations into predefined classes. The purpose is to determine the class of an observation based on a set of variables known as predictors or input variables. The model is built based on a set of observations for which the classes are known. This set of observations is sometimes referred to as the training set. Based on the training set, the technique constructs a set of linear functions of the predictors, known as discriminant functions, such that $L = b_1x_1 + b_2x_2 + \ldots + b_nx_n + c$, where

b are discriminant coefficients, x are the input variables or predictors and c is a constant. These discriminant functions are used to predict the class of a new observation with unknown class. For a k class problem k discriminant functions are constructed. Given a new observation, all the k discriminant functions are evaluated and the observation is assigned to class i, if the i-th discriminant function has the highest value.

Most of the known approaches for medical decision-making, based on the discrimination methods, regard only one type of discriminator (i.e. either for continuous or for discrete (categorical) data). After introducing the location model, proposed in [3], it was possible to cope both with continuous and discrete variables. The paper [16] proposed a model for discrimination, using binary and continuous variables, and later generalized it [17] for mixtures of categorical and continuous variables. The paper [18] presented a case study on the use of various discrimination methods for mixed data. The mentioned techniques attempted to model in particular the interactions between the different data sets. However, they are of the restricted use due to their applicability for a limited number of features (i.e. the individual measurable heuristic properties of the phenomena being observed).

Therefore the paper [19] proposed to use the adequate discriminator for every data type (i.e., with a number of various discriminators at the disposal) and by coupling logically the different discrimination rules to a new (common) discriminator. The coupling procedure enables one to cope jointly with data of different structure and/or scales of measurement, but without strong restrictions on the number of both continuous and categorical features. The method is combined with a consequent cross-validation process securing the results reached.

2.5 Bayesian estimation with mixed data

Among the research works, concerned with mixed-type data in the area of Bayesian estimation, it is worth mentioning the following papers.

The paper [20] deals with Bayesian latent variable models for clustered mixed outcomes. It proposes the general framework for modelling clustered mixed outcomes. A mixture of generalized linear models is used to describe the joint distribution of a set of underlying variables, and an arbitrary function relates the underlying variables to the observed outcomes. The model accommodates the multilevel data structures, general covariate effects and distinct link functions and error distributions for each underlying variable. Within the proposed framework, the novel models are developed for clustered multiple binary, unordered categorical and joint discrete and continuous outcomes. A Markov chain Monte Carlo (MCMC) sampling algorithm is described for estimating the posterior distributions of the parameters and latent variables. Due to flexibility of the framework and estimation procedure, the extensions to ordered categorical outcomes and more complex data structures are straightforward.

According to [20], Bayesian approach for the analysis of mixed outcome data has several important advantages, in contrast to methods based on asymptotic maximum likelihood theory. First, the exact posterior distributions of the parameters and latent variable can be estimated by using MCMC methods. Means and quantiles based on estimated posteriors are appropriate regardless of the sample size. Second, Bayesian estimation allows a direct incorporation of prior knowledge. This is a major advantage in the structural equation modelling. Classical methods often require, that a subset of the parameters is known to ensure identifiability. Although the constraints on the threshold parameters and the variance of the latent variables are often reasonable, additional less justifiable constraints can be avoided by using a prior distribution to allow for prior uncertainty in the parameters. In addition, by assigning an informative prior to parameters about which there is previous information (probably, from historical studies under a similar design), more precise estimates of the parameters of interest can be obtained.

Another work [21] in the discussed area deals with Bayesian factor analysis for mixed ordinal and continuous responses. This paper formulates a measurement model, that is appropriate for such mixed multivariate responses. The proposed model unifies standard normal theory factor analysis and item response theory models for ordinal data. A MCMC algorithm is used for model fitting, and the developed software is publicly available.

One of the most recent papers, [22], presents a novel suboptimal filtering algorithm, addressing estimation problems, that arise in mixed continuous-discrete linear time-varying systems with stochastic parametric uncertainties. The suboptimal state estimate is formed by summing of local Kalman estimates with weights depending only on time instants. In contrast to optimal weights, the suboptimal weights do not depend on current measurements, and thus the proposed filter is of a low-complexity and it can easily be implemented in real-time. High accuracy and efficiency of the suboptimal filter are demonstrated on the damper harmonic oscillator motion and the vehicle motion constrained to a plane.

2.6 Non-parametric estimation with mixed data by using the kernel methods

Non-parametric technique with mixed data use the kernel density estimators to estimate random variables' density functions. The kernels (weighting functions) are also used in kernel regressions to estimate the conditional expectation of a random variable. In the area of non-parametric estimation with both continuous and discrete data the following papers have been met.

The paper [23] proposes a method for nonparametric regression, which admits continuous and categorical data in a natural manner using the method of kernels. A data-driven method of bandwidth selection is proposed, and the asymptotic normality of the estimator is established. The paper also establishes the rate of convergence of the cross-validated smoothing parameters to their benchmark optimal smoothing parameters. Simulations suggest, that the new estimator performs much better than the conventional nonparametric estimator in the presence of mixed data. An empirical application to a widely used and publicly available dynamic panel of patent data demonstrates that the out-of-sample squared prediction error of the proposed estimator is only 14%-20% of that obtained by some popular parametric approaches, which have been used to model this data set.

The paper [24] deals with rating crop insurance policies with efficient nonparametric estimators that admit mixed data. According to [24], the identification of improved methods for characterizing crop yield densities has experienced a recent surge in activity due in part to the central role played by crop insurance in the Agricultural Risk Protection Act of 2000 (estimates of yield densities are required for the determination of insurance premium rates). Nonparametric kernel methods have been successfully used to model yield densities; however, traditional kernel methods do not handle the presence of categorical data in a satisfactory manner and have therefore tended to be applied on a county-by-county basis. By utilizing recently developed kernel methods that admit mixed data types, it is possible to model the yield density jointly across counties, leading to substantial finitesample efficiency gains.

Chapter 3

Mixed data in mixed mixtures

The work [25] proposes the alternative way of the joint modelling of mixed discrete and continuous data. According to [25], a mixture of Gaussian models with a small fixed variance may serve to this purpose.

The key problem of the mixed mixtures is the joint presence of discrete and continuous data in respective factors. The work [25] describes both possible variants, i.e. the dependencies of the continuous factor output on discrete and continuous quantities and the discrete factor output on continuous and discrete quantities.

3.1 Continuous factor output and discrete quantities

The book [25] considers the case, when the i-th regression vector contains one or more discrete entries (regressors). To simplify notation, they are mapped into a scalar discrete regressor $\iota_t \in \iota^* \equiv \{1, \ldots, \mathring{\iota}\}$. Then, it holds.

The i-th factor in the exponential family (EF) has the following form

$$f(d_{i;t}|d_{(i+1)\ldots\mathring{d};t}, d(t-1), \Theta_i) = \prod_{\iota=1}^{\mathring{\iota}} \left[A(\Theta_{\iota i}) \exp\left[\langle B_\iota(\Theta_{\iota i}), C_\iota(\Psi_{\iota i;t})\rangle\right]\right]^{\delta_{\iota \iota_t}}$$

where $\Theta_i \equiv \{\Theta_{\iota i}\}_{\iota=1}^{\mathring{\iota}}$ and Kronecker symbol δ are exploited. The function $A(\cdot)$ is a non-negative scalar function, $B(\cdot)$ and $C(\cdot)$ are multivariate functions of compatible and finite dimensions; the functional $\langle\cdot,\cdot\rangle$ is linear in the first argument. The individual factors in this product are called parameterized *sub-factor*s.

Then, the conjugate prior pdf $f(\Theta_i|d(0))$ has the form

$$f(\Theta_i|d(0)) = \prod_{\iota=1}^{\mathring{\iota}} A^{\nu_{\iota i;0}}(\Theta_{\iota i}) \exp\left[\langle V_{\iota i;0}, C_\iota(\Theta_{\iota i})\rangle\right].$$

The corresponding posterior pdfs $f(\Theta_i|d(t))$ preserve this functional form and their sufficient statistics evolve as follows.

$$V_{\iota i;t} = V_{\iota i;t-1} + \delta_{\iota_t,\iota} B_\iota(\Psi_{\iota i;t}),\ \nu_{\iota i;t} = \nu_{\iota i;t-1} + \delta_{\iota_t,\iota},\ \iota = 1, \ldots, \mathring{\iota},\ t \in t^*.$$

It means, that the observed discrete-valued entry ι_t serves as a known pointer to the single pair of statistics $V_{\iota_t i;t-1}$, $\nu_{\iota_t i;t-1}$ updated at the time t.

The individual sub-factors with indexes ιi are only updated on the sub-selection of data for which the observed discrete pointers $\iota_t,\ t \in t^*$, have the value ι. Thus, the observed discrete regressor ι_t effectively segments the processed data into $\mathring{\iota}$ parts. This conclusion is formalized in the following proposition [25].

Proposition 3.1.1 (Processing of factors with discrete regressor)
Let us consider learning of a mixed mixture with some factors, say $\{i_1, \ldots, i_{\mathring{k}}\}$, each depending on a discrete-valued observable quantity $\iota_{i_k;t},\ k \in k^* \equiv \{1, \ldots, \mathring{k}\}$.

Then, at time t, the parameter estimates of sub-factors with indexes $\iota_{i_k;t},\ k \in k^*$, are updated by the current data and possibly forgotten. They serve for computing component weights used in the approximate mixture estimation. Other sub-factors are untouched and unused.

3.2 Discrete factor output and continuous and discrete quantities

The case, when the discrete factor output $d_{i;t}$ depends on continuous, and possibly discrete, quantities is much more difficult. It is implied by the fact that in a non-trivial dynamic case a model out of the exponential family is needed. It is necessary either to employ approximate techniques developed in the logistic regression, e.g. [26], or to find other way out. The fixed box widths of mean tracking (MT) uniform factors, described in Chapter 12 of [25], inspire to approximate the desired parameterized factor by a mixture of parameterized MT normal factors, introduced in Chapter 12, which model well differences in data positions.

It means, that according to [25], the discrete-valued output $d_{i;t}$ of the i-th factor can be modelled by the normal (scalar) component

$$f(d_{i;t}|d_{(i+1)\ldots\mathring{d};t}, d(t-1), \Theta_{ic}, c) = \mathcal{N}_{d_{i;t}}(\theta'_{ic}\psi_{ic;t}, r_{ic}), \tag{3.1}$$

around the positions $\theta'_{ic}\psi_{ic;t}$. Here, $\Theta_{ic} \equiv \theta_i \equiv$ regression coefficients in $\theta^*_{ic} \equiv \mathring{\psi}_{ic}$-dimensional real space. $\psi_{ic;t}$ is the regression vector. The noise variance r_{ic} is a known and small positive number.

The estimation and prediction of the MT normal factors is proposed in [25].

3.3 Conclusion

The research report describes the different ways of joint modelling of mixed (discrete and continuous) data. The overview of models and estimators from the context of logistic regression as well as the alternative way in the form of mixed mixtures is provided. The overview provides the necessary support base for the design of filters estimating the mixed-type states, which is a global aim of the research.

Bibliography

[1] E. Suzdaleva, "Towards state estimation in the factorized form", in *Proceedings of the 2nd International Conference on From Scientific Computing to Computational Engineering*, D. Tsahalis, Ed., Athens, July 2006, pp. 1–6, University of Patras.

[2] E. Suzdaleva, "Entry-wise Kalman filtering and prior knowledge processing", Tech. Rep. 2207, ÚTIA AV ČR, Praha, November 2007.

[3] I. Olkin and R.F. Tate, "Multivariate correlation models with mixed continuous and discrete variables", *Annals of Mathematical Statistics*, vol. 32, pp. 448–465, 1961.

[4] D.R. Cox, "The analysis of multivariate binary data", *Applied Statistics*, vol. 21, pp. 113–120, 1972.

[5] D. R. Cox and N. Wermuth, "Response models for mixed binary and quantitative variables", *Biometrika*, vol. 79, no. 3, pp. 441–461, September 1992.

[6] W.J. Krzanowski, "The location model for mixtures of categorical and continuous variables", *Journal of Classification*, vol. 10, pp. 25–49, 1993.

[7] R.J.A Little and M.D. Schluchter, "Maximum likelihood estimation for mixed continuous and categorical data with missing values", *Biometrika*, vol. 72, pp. 497–512, 1985.

[8] R.J.A Little and D.B. Rubin, *Statistical Analysis with Missing Data*, John Wiley, New York, 1987.

[9] G.M. Fitzmaurice and N.M. Laird, "Regression models for a bivariate discrete and continuous outcome with clustering", *Journal of American Statistical Association*, vol. 90, pp. 845–852, 1995.

[10] G.M. Fitzmaurice and N.M. Laird, "Regression models for mixed discrete and continuous responses with potentially missing values", *Biometrics*, vol. 53, no. 1, pp. 110–122, March 1997.

[11] Jian-Qing Shi and Sik-Yum Lee, "Latent variable models with mixed continuous and polytomous data", *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 62, no. 1, pp. 77–87, 2000.

[12] Sik-Yum Lee and Jian-Qing Shi, "Maximum likelihood estimation of two-level latent variable models with mixed continuous and polytomous data", *Biometrics*, vol. 57, no. 3, pp. 787–794, September 2001.

[13] E. J. Bedrick, J. Lapidus, and J. F. Powell, "Estimating the mahalanobis distance from mixed continuous and discrete data", *Biometrics*, vol. 56, no. 2, pp. 394–401, June 2000.

[14] M.D. Sammel, L.M. Ryan, and J.M. Legler, "Latent variable models for mixed discrete and continuous outcomes", *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 59, no. 3, pp. 667–678, 1997.

[15] E. Olusegun George, Dale Armstrong, Paul J. Catalano, and Deo Kumar Srivastava, "Regression models for analyzing clustered binary and continuous outcomes under an assumption of exchangeability", *Journal of Statistical Planning and Inference*, vol. 137, no. 11, pp. 3462–3474, November 2007, Special Issue: In Celebration of the Centennial of The Birth of Samarendra Nath Roy (1906-1964).

[16] W.J. Krzanowski, "Discrimination and classification using both binary and continuous variables", *Journal of American Statistical Association*, vol. 70, pp. 782–790, 1975.

[17] W.J. Krzanowski, "Stepwise location model choice in mixed variable discrimination", *Applied Statistics*, vol. 32, pp. 260–266, 1983.

[18] I.G. Vlachonikolis and F.H.C. Mariott, "Discrimination with binary and continuous data", *Applied Statistics*, vol. 31, pp. 23–31, 1982.

[19] Klaus-D. Wernecke, "A coupling procedure for the discrimination of mixed data", *Biometrics*, vol. 48, no. 2, pp. 497–506, June 1992.

[20] David B. Dunson, "Bayesian latent variable models for clustered mixed outcomes", *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 62, no. 2, pp. 355–366, 2000.

[21] Kevin M. Quinn, "Bayesian factor analysis for mixed ordinal and continuous responses", *Political Analysis*, vol. 12, pp. 338–353, 2004.

[22] Vladimir Shin, Du Yong Kim, Georgy Shevlyakov, and Kiseon Kim, "A low-complexity suboptimal filter for continuous-discrete linear systems with parametric uncertainties", *Signal Processing*, vol. 87, no. 10, pp. 2392–2406, 2007.

[23] Jeff Racine and Qi Li, "Nonparametric estimation of regression functions with both categorical and continuous data", *Journal of Econometrics*, vol. 119, no. 1, pp. 99, March 2004, Amsterdam.

[24] Jeff Racine and Alan Ker, "Rating crop insurance policies with efficient nonparametric estimators that admit mixed data types", *Journal of Agricultural and Resource Economics*, vol. 31, no. 1, pp. 27, April 2006.

[25] 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.

[26] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer Series in Statistics. Springer, New York, 2001.