VARIABLE SELECTION FOR NOISY DATA APPLIED IN PROTEOMICS

N. Dridi, A. Giremus, J.-F Giovannelli, C. Truntzer, Pascal Roy, L Gerfaut, J.-P Charrier, P. Ducoroy, C Mercier, P Grangeat

VARIABLE SELECTION FOR NOISY DATA APPLIED IN PROTEOMICS

N. Dridi[1], *A. Giremus*[1], *J.-F. Giovannelli*[1], *C. Truntzer*[2], *P. Roy*[3], *L. Gerfaut*[4], *J.-P. Charrier*[5], *P. Ducoroy*[2], *C. Mercier*[3], *P. Grangeat*[4]

[1]Univ. Bordeaux, IMS, UMR 5218, F-33400 Talence, France

[2]Plateforme protéomique CLIPP, CHU Dijon, F- 21000 Dijon, France

[3]Hospices Civils de Lyon, Service de Biostatistique, Université Lyon I, CNRS UMR 5558, F-69424 Lyon, France

[4]CEA Leti, MINATEC Campus, DTBS, F-38054 Grenoble cedex 9, France

[5]bioMérieux, F-69280 Marcy l'Étoile, France

ABSTRACT

The paper proposes a variable selection method for proteomics. It aims at selecting, among a set of proteins, those (named biomarkers) which enable to discriminate between two groups of individuals (healthy and pathological). To this end, data is available for a cohort of individuals: the biological state and a measurement of concentrations for a list of proteins. The proposed approach is based on a Bayesian hierarchical model for the dependencies between biological and instrumental variables. The optimal selection function minimizes the Bayesian risk, that is to say the selected set of variables maximizes the posterior probability. The two main contributions are: (1) we do not impose ad-hoc relationships between the variables such as a logistic regression model and (2) we account for instrumental variability through measurement noise. We are then dealing with indirect observations of a mixture of distributions and it results in intricate probability distributions. A closed-form expression of the posterior distributions cannot be derived. Thus, we discuss several approximations and study the robustness to the noise level. Finally, the method is evaluated both on simulated and clinical data.

Index Terms— Model and variable selection, Bayesian approach, biological et technological variability, Gaussian mixture, proteomics.

1. INTRODUCTION

Proteomics is an expanding discipline based on large-scale studies of proteins present in an organism. It offers a promising alternative to genomics since it includes more information about biological and cellular systems [1, 2]. It covers several questions: protein identification in a biological sample, concentration quantification and differential analysis [4]. The latter consists in identifying a set of proteins, called biomarkers, differently expressed according to a biological state (healthy: $\mathcal{H}$, pathological: $\mathcal{P}$). These biomarkers can allow early diagnosis of diseases like cancer and follow-up therapy.

However, the proteins have small and variable concentrations, hence reliable measurements require high-tech systems such as LC-MS (Liquid Chromatography and Mass Spectrometry) or MALDI-TOF (Matrix-Assisted Laser Desorption / Ionisation through Time-Of-Fly). They produce spectra including peaks related to the nature and the concentration of the proteins. Biomarker discovery can be either directly based on these spectra [5, 6] or on concentrations estimated from these spectra. The study can be non parametric or parametric, *e.g.* conducted in a Bayesian framework. Here, biomarker selection relies on protein concentrations and is carried out in a Bayesian scheme.

Discovery methods can be broadly classified into two categories. A first one relies on a predictive model, such as the logistic regression, that describes the relationship between explicative (the protein concentrations) and explained (the biological state) variables. Then, variable selection is performed by finding the protein combination that minimizes a criterion which penalizes the complexity, such as the Bayesian Information Criterion [8] or the Akaike's Information Criterion [9]. However, the computational complexity is relatively high since 2^P models must be compared for P explicative variables. To alleviate this complexity, [10] proposes a Gibbs sampling pre-selection of the variables. An alternative is to compute penalized maximum likelihood estimates of the regressors by enforcing parsimony. The most popular algorithms are the LASSO or the Elastic Net method [11, 12]. The second class of methods is based on differential analysis [4] whose principle is generally to carry out univariate tests, *e.g.* the Student one, for each protein. The main difficulty is that, due to the multiple tests, it is necessary to control the family wise error rate or the less conservative false discovery rate [13].

Compared to the above-cited work, we propose to relax the hypothesis of a logistic regression model which may be quite restrictive. The problem is modelled within a hierarchical Bayesian framework. Based on the risk (mean loss), an optimal decision-maker is designed for variable selection that leads to select the most probable set of biomarkers a posteri-

ori. Moreover, compared to our previous work [7], the proposed modelling takes into account the technological variability. Thus, the true concentrations of the proteins are unknown and the selection is based on measured noisy concentrations. Furthermore, the noise precision is also unknown which introduces extra parameter in the model. It is then difficult to evaluate the posterior probability since it requires integration w.r.t. both the unknown parameters and the true concentrations. Here, we discuss several approximations and study the robustness to the noise level. The performance of the method is evaluated both on simulated and clinical data.

The rest of the paper is organized as follows. Section 2 describes the considered observation model. Section 3 is dedicated to the variable selection method, and numerical results are provided in section 4. Finally conclusions and perspectives are given in section 5.

2. MODEL FOR THE OBSERVATIONS

A set of measured protein concentrations for a cohort of N individuals is available.The proposed biomarker selection method is univariate, hence the decision whether a protein is discriminant or not is made protein per protein. In the sequel, we denote $b_n \in \{\mathcal{H}, \mathcal{P}\}$, $x_n \in \mathbb{R}$ and $y_n \in \mathbb{R}$ the biological state, the true and the measured concentrations of the protein of interest for the individual n, respectively. Practically, x_n is unknown and we only have access to b_n and y_n. The latter includes technological variability as an additive noise ε_n:

$$y_n = x_n + \varepsilon_n \,. \tag{1}$$

It is one of the contribution of this paper with respect to our previous work [7] that was based on noiseless concentrations. In the sequel, the variables associated to the N individuals are stacked in vectors denoted $\mathbf{y}, \mathbf{x}$ and $\mathbf{b}$. $\mathcal{I}_\mathcal{P}$ and $\mathcal{I}_\mathcal{H}$ are the subsets of indices for the pathological and the healthy samples and we denote $\mathcal{I}_\mathcal{C} = \mathcal{I}_\mathcal{P} \cup \mathcal{I}_\mathcal{H}$.

The noise ε_n is described by a zero-mean normal probability density function (pdf) with precision γ_ε, therefore $y_n|x_n$ is given by a normal pdf with mean and precision $(x_n, \gamma_\varepsilon)$. As for the state b_n, it is classically described by a Bernoulli variable with parameter p. In a Bayesian framework, the unknown variables are assigned prior probabilities. Regarding the protein concentration, its pdf depends whether it is discriminant or not. In the case non discriminant, x_n satisfies a Gaussian pdf with parameters $(m_\mathcal{C}, \gamma_\mathcal{C})$. As an alternative, if the protein is discriminant, x_n is distributed according to a mixture of two Gaussian pdf of parameters $(m_\mathcal{H}, \gamma_\mathcal{H})$ and $(m_\mathcal{P}, \gamma_\mathcal{P})$ with respective weights p and $1-p$. Furthermore, we consider conjugate pdf for the hyperparameters. This choice directly impacts the feasibility of the posterior distribution calculations. Thus, the precision γ_ε follows a Gamma pdf $\mathcal{G}$ with parameters $(\alpha_\varepsilon^{\text{pri}}, \beta_\varepsilon^{\text{pri}})$ and the couples $(m_\times, \gamma_\times)$, with $\times \in \{\mathcal{P}, \mathcal{H}, \mathcal{C}\}$, are Normal-Gamma ($\mathcal{NG}$) distributed with parameters $(\mu_\times^{\text{pri}}, \eta_\times^{\text{pri}}, \alpha_\times^{\text{pri}}, \beta_\times^{\text{pri}})$.

Finally, the concentrations of the different individuals are assumed independent. Regarding the unknown parameters, we define $\boldsymbol{\theta} = [m_\mathcal{P}, \gamma_\mathcal{P}, m_\mathcal{H}, \gamma_\mathcal{H}, m_\mathcal{C}, \gamma_\mathcal{C}, \gamma_\varepsilon, p]$. Note that all the parameters are scalar. The problem at hand is to decide between two models denoted $\Delta = +$ if the protein is discriminant and $\Delta = -$ otherwise.

3. VARIABLE SELECTION

3.1. Optimal decider

To build an optimal decision-maker, a binary loss function is considered: it assigns a null loss to any correct decision and a unitary loss to any wrong decision. The risk is then defined as a mean loss and an important point is that the mean is over the two models, the data (the observed concentrations and the states), the true concentrations and the unknown parameters. The optimal decision-maker is defined as the risk minimizer and it is known that it selects the most probable model *i.e.* the Maximum A Posteriori.

3.2. Analytical calculation of the posteriori probability

The posterior probability of the model $\delta \in \{+, -\}$ is:

$$\mathbb{P}_{\Delta|\mathbf{Y},\mathbf{B}}(\delta|\mathbf{y}, \mathbf{b}) \propto f_{\mathbf{Y},\mathbf{B}|\Delta}(\mathbf{y}, \mathbf{b}|\delta)\mathbb{P}(\Delta = \delta) \tag{2}$$

based on the likelihood $f_{\mathbf{Y},\mathbf{B}|\Delta}(\mathbf{y}, \mathbf{b}|\delta)$ also referred to as the evidence. Since the individuals are independent, the complete likelihood can be factorized and yields Eq. (3).

First, integration is performed w.r.t. $\boldsymbol{\theta}$ and the result is written as a pdf for $\mathbf{x}$, then the integral on $\mathbf{x}$ is calculated. Each integral in Eq. (3) can be calculated separately.

For I_{γ_ε}, the following expression is obtained:

$$I_{\gamma_\varepsilon} = (2\pi)^{-N/2} K_G^{\text{pst}} / K_G^{\text{pri}} \tag{4}$$

where $K_G = \Gamma(\alpha)/\beta^\alpha$ is the normalization constant of the Gamma pdf with parameters α and β. The superscripts $\{\text{pst}, \text{pri}\}$ stand for prior and posterior pdf of the precision parameter. It should be noted that, by the conjugation principle, the posterior pdf of γ_ε is also Gamma with parameters:

$$\begin{aligned} \alpha_\varepsilon^{\text{pst}} &= \alpha_\varepsilon^{\text{pri}} + N/2 \\ \beta_\varepsilon^{\text{pst}} &= \beta_\varepsilon^{\text{pri}} + \frac{1}{2}\sum_{n=1}^{N}(y_n - x_n)^2. \end{aligned}$$

In Eq. (4), only one term depends on $\mathbf{x}$ and it can be rewritten:

$$\begin{aligned} (\beta_\varepsilon^{\text{pst}})^{-\alpha_\varepsilon^{\text{pst}}} &= \left(\beta_\varepsilon^{\text{pri}} + \frac{1}{2}(\mathbf{y}-\mathbf{x})(\mathbf{y}-\mathbf{x})^t\right)^{-\alpha_\varepsilon^{\text{pst}}} \\ &= K_\mathcal{S}(\beta_\varepsilon^{\text{pri}})^{-\alpha_\varepsilon^{\text{pst}}}\, \mathcal{S}(\mathbf{x}; \mathbf{y}, \frac{\alpha_\varepsilon^{\text{pri}}}{\beta_\varepsilon^{\text{pri}}} I_N, 2\alpha_\varepsilon^{\text{pri}}) \end{aligned}$$

$$f_{\mathbf{Y},\mathbf{B}|\Delta}(\mathbf{y},\mathbf{b}|\delta) = \int \underbrace{\left(\int \prod_{I_\mathcal{C}} \mathcal{N}(y_n|x_n,\gamma_\varepsilon)\mathcal{G}(\gamma_\varepsilon)\mathrm{d}\gamma_\varepsilon\right)}_{I_{\gamma_\varepsilon}} \underbrace{\left(\int \prod_{I_\mathcal{H}} \mathcal{N}(x_n^+;m_\mathcal{H},\gamma_\mathcal{H})\mathcal{NG}(m_\mathcal{H},\gamma_\mathcal{H})\mathrm{d}(m_\mathcal{H},\gamma_\mathcal{H})\right)^{(\delta,+)}}_{I_\mathcal{H}}$$
$$\underbrace{\left(\int \prod_{I_\mathcal{P}} \mathcal{N}(x_n^+;m_\mathcal{P},\gamma_\mathcal{P})\mathcal{NG}(m_\mathcal{P},\gamma_\mathcal{P})\mathrm{d}(m_\mathcal{P},\gamma_p)\right)^{(\delta,+)}}_{I_\mathcal{P}} \underbrace{\left(\int \prod_{I_\mathcal{C}} \mathcal{N}(x_n^-;m_\mathcal{C},\gamma_\mathcal{C})\mathcal{NG}(m_\mathcal{C},\gamma_\mathcal{C})\mathrm{d}(m_\mathcal{C},\gamma_\mathcal{C})\right)^{(\delta,-)}}_{I_\mathcal{C}} \mathrm{d}\mathbf{x} \quad (3)$$

where $K_\mathcal{S}$ is the normalization constant of the multivariate Student distribution $\mathcal{S}$ with parameters $\mathbf{y}$,$\alpha_\varepsilon^{\text{pri}}/\beta_\varepsilon^{\text{pri}} I_N$, $2\alpha_\varepsilon^{\text{pri}}$ (see Appendix).

Let us now consider the factors $I_\times$ with $\times \in \{\mathcal{H},\mathcal{P},\mathcal{C}\}$. It should be noted that they involve the same distributions. Using again the conjugation, the posterior pdf of $(m_\times,\gamma_\times)$ is also Normal-Gamma with parameters:

$$\begin{aligned}
\mu_\mathcal{P}^{\text{pst}} &= \frac{N_\times \bar{x}_\times + \eta_\times^{\text{pri}}\mu_\times^{\text{pri}}}{\eta_\times^{\text{pri}} + N_\times} \\
\eta_\times^{\text{pst}} &= \eta_\times^{\text{pri}} + N_\times \\
\alpha_\times^{\text{pst}} &= \alpha_\times^{\text{pri}} + N_\times/2 \\
\beta_\times^{\text{pst}} &= \beta_\times^{\text{pri}} + N\bar{R}_\times^{\mathbf{x}}/2 + \frac{\eta_\times^{\text{pri}} N_\times}{2(\eta_\times^{\text{pri}} + N_\times)}\left(\bar{x}_\times - \mu_\times^{\text{pri}}\right)^2
\end{aligned}$$

where $\bar{x}_\times$ / $\bar{R}_\times^x$ are respectively the empirical mean / variance of the true concentrations. By integration w.r.t. $(m_\times,\gamma_\times)$:

$$I_\times = (2\pi)^{-N_\times/2} K_{NG}^{\text{pst}}(\times)/K_{NG}^{\text{pri}}(\times) \quad (5)$$

where $K_{NG}^\star = \Gamma(\alpha^\star)\sqrt{2\pi}/(\beta^\star)^{\alpha^\star}\sqrt{\eta^\star}$, for $\star \in \{\text{pst},\text{pri}\}$, is the normalization constant of the Normal-Gamma pdf with parameters $(\mu^\star,\eta^\star,\alpha^\star,\beta^\star)$. At this step, the calculation cannot be completed exactly. Different approximations can be considered. A first solution is to resort to Monte Carlo integration techniques such as importance sampling [3]. An alternative is to consider analytical approximations. For instance, either $\bar{x}_\times$ or both $\bar{x}_\times$ and $\bar{R}_\times^x$ could be replaced by their empirical estimates $\bar{y}_\times$ and $\bar{R}_\times^y$, computed from the observed concentrations. In the first case, the integrand (5) can be expressed, up to a proportionality constant, as a multivariate Student distribution of argument $\mathbf{x}$. Then, using the well-known result that Student distributions can be approximated by Gaussian distributions provided the degree of freedom is high enough, analytical computation of Eq. (3) becomes possible. In the second case, the $I_\times$ no longer depend on $\mathbf{x}$ and need not be integrated. Eq. (4) and (5) are replaced in Eq. (3) and yields:

$$f_{\mathbf{Y},\mathbf{B}|\delta}(\mathbf{y},\mathbf{b}|+) = K\frac{K_{NG}^{\text{pst}}(\mathcal{P})}{K_{NG}^{\text{pri}}(\mathcal{P})}\frac{K_{NG}^{\text{pst}}(\mathcal{H})}{K_{NG}^{\text{pri}}(\mathcal{H})} \quad (6)$$

$$f_{\mathbf{Y},\mathbf{B}|\delta}(\mathbf{y},\mathbf{b}|-) = K\frac{K_{NG}^{\text{pst}}(\mathcal{C})}{K_{NG}^{\text{pri}}(\mathcal{C})} \quad (7)$$

with $K = K_S(\beta_\varepsilon^{\text{pri}})^{-\alpha_\varepsilon^{\text{pst}}}$. It should be noted that in this very case the corresponding posterior probabilities (2) are identical to the ones derived by neglecting the noise in [7].

3.3. Hyperparameter choice

The probabilities (6)-(7) depend on hyperparameters that are the parameters of the Normal-Gamma pdf $(\mu_\times,\eta_\times,\alpha_\times,\beta_\times)$ for $\times \in \{\mathcal{P},\mathcal{H},\mathcal{C}\}$. In a non-informative case, they tend to 0 and the proportionality coefficients take an undeterminated form. To tune these parameters, we resort to poorly informative prior based on expert-knowledge of orders of magnitudes for the involved variables. To this end, we use the relationship between the expected values and the covariance matrices of $(\mu_\times\ \gamma_\times)$ and the parameters of a Normal-Gamma pdf given by:

$$E(\gamma_\times) = \alpha_\times\beta_\times^{-1}, \qquad E(m_\times) = \mu_\times$$
$$V(m_\times) = \frac{\beta_\times}{\eta_\times(\alpha_\times - 1)}, \qquad V(\gamma_\times) = \alpha\beta_\times^{-2}$$

Therefore, when information about the range of $m_\times$ and $\gamma_\times$ is available, $(\mu_\times,\eta_\times,\alpha_\times,\beta_\times)$ can be calculated.

4. RESULTS

In a first step, we study the decision rule obtained by replacing both $\bar{x}_\times$ and $\bar{R}_\times^x$ by empirical estimates in Eq. (5). We conduct extensive simulation studies to evaluate the performance of the method and more precisely to study its robustness to noise. For a given protein, we aim at deciding if it is a biomarker ($\Delta = +$) or not ($\Delta = -$) by selecting the hypothesis with the highest posterior probability using Eq. (2).

For each true model $\Delta^\star = +$ and $\Delta^\star = -$, $N_\text{r} = 10^5$ realizations of the biological state and the measured proteins concentrations are simulated for N individuals. For

discriminant proteins, the concentrations are distributed either as $\mathcal{N}(x_n; m_{\mathcal{H}}, \gamma_{\mathcal{H}})$ or as $\mathcal{N}(x_n; m_{\mathcal{P}}, \gamma_{\mathcal{P}})$ depending on the biological state. As for non discriminant ones, it is given by $\mathcal{N}(x_n; m_{\mathcal{C}}, \gamma_{\mathcal{C}})$. The noise is generated according to $\mathcal{N}(\varepsilon; 0, \gamma_\varepsilon)$. The parameters $(m_\times, \gamma_\times)$ where $\times \in \{\mathcal{P}, \mathcal{H}, \mathcal{C}\}$ and γ_ε are respectively distributed as $\mathcal{NG}(m_\times, \gamma_\times; \mu_\times, \eta_\times, \alpha_\times, \beta_\times)$, and $\mathcal{G}(\gamma_\varepsilon; \alpha_\varepsilon, \beta_\varepsilon)$. Finally, the hyperparameters $(\mu_\times, \eta_\times, \alpha_\times, \beta_\times)$ are calculated as explained in Section 3.3.

4.1. Impact of γ_ε and N

We study the impact of the noise level and the cohort size on the error rate $\tau(\%) = 100 * N_{\text{fs}}/N_{\text{r}}$ where N_{fs} is the number of false selections (estimated and true models are different).

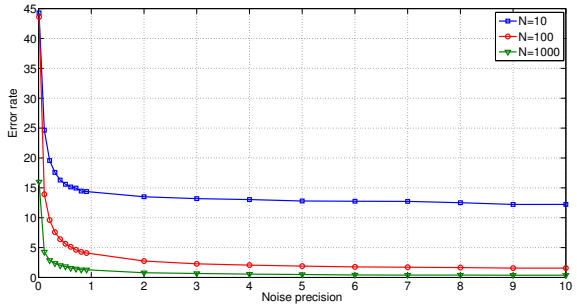

Fig. 1. Error rate $\tau(\%)$ of the Bayesian method as a function of the noise precision γ_ε, for different values of the sample number.

Figure 1 shows the error rate as a function of the noise precision γ_ε and for several cohort sizes . As expected, for a given N, τ decreases when γ_ε increases. This is explained by the precision of the approximation of the mean and variance of the true concentrations by the empirical ones. Indeed, the accuracy of this approximation depends on the noise power: the higher the noise, the larger the impact of the approximation. However, it is also related to the observation size N: for large N, the method is more robust to noise.

4.2. Comparison with Student test

This section compares the performance of the proposed approach with a Student test. To select a biomarker, the latter tests whether the means of the protein concentrations for the healthy and pathological individuals are equal or not.

Bayesian method / Student test		
$\Delta^\star$ \ $\hat{\Delta}$	$-$	$+$
$-$	100.000 / 99.876	0.000 / 0.124
$+$	0.0630 / 0.0140	99.937 / 99.986

Table 1. Rates $\tau(\%)$ false / true and positive / negative. $+/-$ refers to discriminant / non discriminant protein.

Table 1 shows the four rates (false / true and positive / negative) for the proposed and the Student method. $\Delta^\star$ and $\hat{\Delta}$ respectively refers to the true and selected models. Clearly, the proposed method always reject non discriminant variables. Regarding discriminant variables, the error is larger with the proposed method (0.063%) than with the Student one (0.124%). The total false selection rate is the arithmetic mean of the false positive and false negative rates: 0.0315% for the proposed approach and 0.069% for the Student, that is to say more than twice higher.

4.3. Clinical data

This section is devoted to a first clinical data set[1] composed of 190 samples, including 107 with state $\mathcal{P}$ (colorectal cancer) and 83 with state $\mathcal{H}$. For each sample, the concentrations of 34 proteins are measured. To the best of our expertise, one of them is known to be a biomarker. As for the results, this biomarker is correctly selected by our Bayesian method, and a second candidate is also selected.

5. CONCLUSION AN PERSPECTIVES

Biomarker discovery is a crucial issue in proteomics, since it may allow early detection of diseases like cancers... It consists in selecting discriminant proteins, given concentrations and biological states for a set of samples. In this paper, we develop an optimal strategy in a hierarchical Bayesian framework. Compared to our previous paper [7], the proposed model includes technological variability. This introduces additional parameters, and the main difficulty is the required integration w.r.t. unknown parameters and the unobserved protein concentration which cannot be performed analytically. We propose several approximations to alleviate this difficulty while ensuring a reduced computational complexity. After studying the robustness of our approach to the noise level, we evaluate its performance through numerical simulations and clinical data. In both cases, a selection error lower than 0.1% is obtained. In the future, we intend to tackle the multivariate model $P > 1$.

A. APPENDIX: MULTIVARIATE STUDENT PDF

A vector $\mathbf{x} \in \mathbb{R}^N$ follows a multivariate Student pdf of parameters $(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \nu)$ if:

$$f_X(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Lambda}, \nu) = K_S^{-1} \left[1 + (\mathbf{x} - \boldsymbol{\mu})^t \boldsymbol{\Lambda} (\mathbf{x} - \boldsymbol{\mu})/\nu\right]^{-(N+\nu)/2}$$

with K_S is normalisation constant:

$$K_S = \frac{\Gamma(\nu/2)}{\Gamma((N+\nu)/2)} \frac{(\pi\nu)^{N/2}}{|\boldsymbol{\Lambda}|^{1/2}}$$

[1] provided by bioMerieux (Technology Research Department), France.

B. REFERENCES

[1] R. E. Banks, M. J. Dunn, D. F. Hochstrasser, J. C. Sanchez, W. Blackstock, D. J. Pappin, and P. J. Selby, "Proteomics: new perspectives, new biomedical opportunities." *The Lancet*, vol. 356, no. 18, pp. 1749–1756, Nov. 2000.

[2] K.-A. Do, P. Muller, and M. Vannucci, *Bayesian Inference for Gene Expression And Proteomics*. Cambridge, England: Cambridge University Press, 2006.

[3] C.-P. Robert, *The Bayesian Choice*, Springer Texts in Statistics, 2nd ed. 2001.

[4] G. K. Smyth, "Linear models and empirical Bayes methods for assessing differential expression in microarray experiments," *Statistical Applications in Genetics and Molecular Biology*, vol. 3, no. 3, 2004.

[5] P. Szacherski, J.-F. Giovannelli, and P. Grangeat, "Joint Bayesian hierarchical inversion-classification and application in proteomics." in *Proc. of the Int. Conf. on Stat. Signal Proc.*, Nice, France, June 2011.

[6] P. Szacherski, J.-F. Giovannelli, L. Gerfault, and P. Grangeat, "Apprentissage supervisé robuste de caractéristiques de classes. Application en protéomique." in *Actes 23*^{e} *coll. GRETSI*, Bordeaux France, Sep. 2011.

[7] F. Adjed, J.-F. Giovannelli, A. Giremus, N. Dridi, and P. Szacherski, "Variable selection for a mixed population applied in proteomics," in *Proc. IEEE ICASSP*, Vancouver, Canada, May 2013.

[8] G. Schwartz, "Estimating the Dimension of a Model," *Annals Statist.*, vol. 6, pp. 461–464, 1978.

[9] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.

[10] E. I. George and R. E. McCulloch, "Variable selection via the Gibbs sampling," *J. Acoust. Society America*, vol. 88, no. 423, pp. 881–889, Sep. 1993.

[11] H. Zou and T. Hastie, "Regularization and variable selection via the Elastic Net," *J. R. Statist. Soc. B*, vol. 67, pp. 301–320, 2005.

[12] P. Bühlmann and T. Hothorn, "Boosting algorithms: regularization, prediction and model fitting (with discussion)," *Statistical Science*, vol. 22, no. 4, pp. 477–505, 2007.

[13] Y. Benjamin and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *J. R. Statist. Soc. B*, vol. 57, no. 1, pp. 289–300, 1995.