VARIABLE SELECTION FOR A MIXED POPULATION APPLIED IN PROTEOMICS

F. Adjed[†]*, J.-F. Giovannelli*[†]*, A. Giremus*[†]*, N. Dridi*[†] *and P. Szacherski*[‡,†]

[†]Univ. Bordeaux, IMS, UMR 5218, F-33400 Talence, France
[‡]CEA-Leti, MINATEC Campus, F-38054 Grenoble Cedex 9, France

ABSTRACT

The paper presents a variable selection method for biomarker discovery in proteomics. More specifically, it finds the most adequate variables among a given set in order to discriminate between two groups (healthy and pathological). This approach is developped within a Bayesian framework and relies on an optimal strategy that results in the choice of the most a posteriori probable model. The calculation of the posterior probabilities requires marginalization of unknown parameters. It is the main difficulty and a contribution of the paper is to provide a closed-form expression. The originality of the work is twofold: (1) we relax the standard hypothesis of linear regression models and (2) we present a multivariate test which directly accommodates possible correlations between the biomarkers. The effectiveness of the method is assessed through a simulated study and shows results in accordance with the theoritical optimality.

Index Terms — Model and variable selection, Bayesian approach, Bayes factor, Gaussian mixture, proteomics.

1. INTRODUCTION

Proteomics is an expanding domain that consists in the large-scale study of proteins. It offers a promising tool since it includes information about biological and cell system functioning [1, 2]. Practically, some proteins are differently expressed according to the biological state (healthy: $\mathcal{H}$, pathological: $\mathcal{P}$) and they are referred to as biomarkers. So, proteomic is highly considered in the diagnostic of diseases like cancer [3]. However, the proteins have small and variable concentrations which complicates the study and requires high-tech measurement systems. To this end, Liquid Chromatography and Mass Spectrometry (LC-MS) [4] give spectra including peaks related to the nature and concentration of proteins. The biomarker discovery can be directly based on these spectra [3, 4] or can be based on estimated concentrations [5, 6] computed from these spectra. The study can be non parametric [5] or parametric, e.g. in a Bayesian framework [6]. Here, the study relies on protein concentrations and is conducted in a parametric Bayesian scheme.

Discovery methods can be classified in two categories. A first class consists in introducing a predictive model, such as the logistic regression model, that relates explicative variables (here concentrations) and explained variables (biological state). Then, the variable selection is performed by searching for the combination of proteins that minimizes a criterion that penalizes the complexity, such as the Bayesian Information Criterion [7] or the Akaike Information Criterion [8]. However, the computational complexity is relatively high since 2^P models must be compared for P biological variables. To alleviate this complexity, [9] proposes a Gibbs sampling pre-selection of the biological variables. An alternative is to compute the maximum likelihood estimates of the regressors by enforcing parsimony such as the well-known lasso or elastic net algorithms [10, 11]. The second class is based on differential analysis [12] whose principle is generally to carry out univariate tests such as the Student test 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]. However, such techniques do not account for possible correlations between the biomarkers when performing the selection.

The originality of the proposed work compared to the above-mentioned approaches is twofold. On the one hand, we relax the hypothesis of a linear regression model which may be quite restrictive. On the other hand, we present a multivariate test which directly accommodates possible correlations between the biomarkers. The problem is modeled within a hierarchical Bayesian framework: based on the risk (mean loss), an optimal decision-maker is designed for model selection that leads to select the most a posteriori probable model. When using Bayesian approaches, a difficulty is the choice of the prior probabilities for the unknown parameters. Here, we propose a relevant choice which allows to obtain a closed-form expression of the posterior probability. Thus the approach is attractive from a computational point of view.

The rest of the paper is organized as follows. Section 2 describes the considered model and the variable selection method. Numerical results are provided in section 3. Section 4 gives conclusions and perspectives for future work.

2. VARIABLE SELECTION

Let us note P the number of considered proteins and $\mathbf{x} \in \mathbb{R}^P$ the collection of concentrations. A discriminant/non-

$$f_{\boldsymbol{\mathcal{X}},\mathbf{B}|\boldsymbol{\Delta}}(\mathbf{x},\mathbf{b}|\boldsymbol{\delta}) = \int_{\boldsymbol{\Theta}} \prod_{n\in\mathcal{I}_{\mathcal{P}}} \mathcal{N}(\mathbf{x}_n^+;\mathbf{m}_{\mathcal{P}}^+,\boldsymbol{\Gamma}_{\mathcal{P}}^+) \prod_{n\in\mathcal{I}_{\mathcal{H}}} \mathcal{N}(\mathbf{x}_n^+;\mathbf{m}_{\mathcal{H}}^+,\boldsymbol{\Gamma}_{\mathcal{H}}^+) \prod_{n\in\mathcal{I}_{\mathcal{C}}} \mathcal{N}(\mathbf{x}_n^-;\mathbf{m}_{\mathcal{C}}^-,\boldsymbol{\Gamma}_{\mathcal{C}}^-)\ p^{N_{\mathcal{P}}}(1-p)^{N_{\mathcal{H}}}\pi_{\boldsymbol{\Theta}}(\boldsymbol{\theta}|\boldsymbol{\delta})\mathrm{d}\boldsymbol{\theta} \quad (1)$$

discriminant protein is labeled by $+/-$ and there are 2^P configurations referred to as $\boldsymbol{\delta} \in \{+,-\}^P$. The vector $\mathbf{x}^+/\mathbf{x}^-$ (sizes P^+/P^-) respectively stands for discriminant/non-discriminant proteins (we have $P = P^+ + P^-$). As long as the biological state is concerned, it is denoted by b and takes two values, $\mathcal{H}$ and $\mathcal{P}$ for Healthy and Pathological.

Regarding observations, N is the number of observed individuals and $(\mathbf{x}_n, b_n)$ is n-th observed concentrations and state. Let us denote $\boldsymbol{\mathcal{X}}$ the matrix of concentrations and $\mathbf{b}$ the vector of biological states. $\mathcal{I}_{\mathcal{P}}$ and $\mathcal{I}_{\mathcal{H}}$ are respectively the subsets of indices for pathological and healthy individuals.

For each individual n, the state b_n is described by a Bernoulli variable B with parameter p. Regarding the protein concentrations, they are described by normal distributions. Specifically, for discriminant ones, conditionally on state b_n, the concentration vector $\mathbf{x}_n^+$ is modeled by a multivariate normal distribution with mean and precision $(\mathbf{m}_{\mathcal{H}}, \boldsymbol{\Gamma}_{\mathcal{H}})$ and $(\mathbf{m}_{\mathcal{P}}, \boldsymbol{\Gamma}_{\mathcal{P}})$ for healthy and pathological respectively. For a non discriminant protein, $\mathbf{x}_n^-$ is modeled by a unique multivariate normal distribution with common parameters $(\mathbf{m}_{\mathcal{C}}, \boldsymbol{\Gamma}_{\mathcal{C}})$. Moreover, it is assumed that $\mathbf{x}^+$ and $\mathbf{x}^-$ are uncorrelated. Regrading unknown parameters, we have $\boldsymbol{\theta} = [\mathbf{m}_{\mathcal{P}}, \boldsymbol{\Gamma}_{\mathcal{P}}, \mathbf{m}_{\mathcal{H}}, \boldsymbol{\Gamma}_{\mathcal{H}}, \mathbf{m}_{\mathcal{C}}, \boldsymbol{\Gamma}_{\mathcal{C}}, p]$.

To build an optimal decision-maker, a $0/1$ loss is considered: it assigns a null (resp. unitary) loss to any correct (resp. wrong) decision. The risk is the mean loss and an important point is that it is the mean over the 2^P models, the data (concentrations and state) and unknown parameters. The optimal decision-maker is defined as the risk minimizer and it is known that it selects the most a posteriori probable model.

2.1. Posteriori probabilities calculation

For each candidate $\boldsymbol{\delta}$, the posterior probability $\mathbb{P}_{\boldsymbol{\Delta}|\boldsymbol{\mathcal{X}},\mathbf{B}}(\boldsymbol{\delta}|\mathbf{x},\mathbf{b})$ is required. By the Bayes rule:

$$\mathbb{P}_{\boldsymbol{\Delta}|\boldsymbol{\mathcal{X}},\mathbf{B}}(\boldsymbol{\delta}|\mathbf{x},\mathbf{b}) \propto f_{\boldsymbol{\mathcal{X}},\mathbf{B}|\boldsymbol{\Delta}}(\mathbf{x},\mathbf{b}|\boldsymbol{\delta})\mathbb{P}(\boldsymbol{\Delta}=\boldsymbol{\delta}) \quad (2)$$

and the keystone is the likelihood $f_{\boldsymbol{\mathcal{X}},\mathbf{B}|\boldsymbol{\Delta}}(\mathbf{x},\mathbf{b}|\boldsymbol{\delta})$ also referred to as the evidence. Thanks to (conditional) uncorrelation between $\mathbf{x}^+$ and $\mathbf{x}^-$ and independance between individuals, the (complete) likelihood doubly factorizes and yields Eq. (1). Its three first factors involve multivariate normal distributions, so, the following explanation is limited to the first one. Reformulating the exponential argument yields:

$$\prod_{n\in\mathcal{I}_{\mathcal{P}}} \mathcal{N}(\mathbf{x}_n^+;\mathbf{m}_{\mathcal{P}}^+,\boldsymbol{\Gamma}_{\mathcal{P}}^+) = (2\pi)^{-PN_{\mathcal{P}}/2}|\boldsymbol{\Gamma}_{\mathcal{P}}|^{N/2}$$
$$\exp\left[-\frac{N_{\mathcal{P}}}{2}\mathrm{tr}\left(\boldsymbol{\Gamma}_{\mathcal{P}}^+\left[\bar{\mathbf{R}}_{\mathcal{P}}^+ + (\bar{\mathbf{x}}_{\mathcal{P}}^+ - \mathbf{m}_{\mathcal{P}}^+)(\bar{\mathbf{x}}_{\mathcal{P}}^+ - \mathbf{m}_{\mathcal{P}}^+)^t\right]\right)\right]$$

where $\bar{\mathbf{x}}_{\mathcal{P}}^+$ and $\bar{\mathbf{R}}_{\mathcal{P}}^+$ are the empirical mean and covariance.

The fourth factor is common to all configurations and the last one is the prior distribution for unknown parameters. The latter is important for two reasons: (1) it models available information and (2) its choice impacts the calculation feasability. Using the conjugation principle, we set a separable prior:

$$\pi_{\boldsymbol{\Theta}}(\boldsymbol{\theta}|\boldsymbol{\Delta}) = \pi_{\mathcal{P}}(\mathbf{m}_{\mathcal{P}}^+,\boldsymbol{\Gamma}_{\mathcal{P}}^+)\pi_{\mathcal{H}}(\mathbf{m}_{\mathcal{H}}^+,\boldsymbol{\Gamma}_{\mathcal{H}}^+)\pi_{\mathcal{C}}(\mathbf{m}_{\mathcal{C}}^-,\boldsymbol{\Gamma}_{\mathcal{C}}^-)\pi_p(p)$$

and Normal-Wishart densities (see Appendix) for $\pi_{\mathcal{P}}$, $\pi_{\mathcal{H}}$ and $\pi_{\mathcal{H}}$ and Beta density for π_p.

For convenience in the forthcoming calculation, we deduce the posterior distribution for $\boldsymbol{\theta}$, $f_{\boldsymbol{\Theta}|\boldsymbol{\mathcal{X}},\mathbf{b}}(\boldsymbol{\theta}|\mathbf{x},b)$, which is proportional to the integrand in (1). Thus, for $(\mathbf{m}_{\times}^{\star},\boldsymbol{\Gamma}_{\times}^{\star})$ with $\times \in \{\mathcal{P},\mathcal{H},\mathcal{C}\}$ and $\star \in \{+,-\}$, the posterior distribution is also the Normal-Wishart with parameters:

$$\begin{cases} \nu_{\times}^{\star\mathrm{pst}} &= \nu_{\times}^{\star} + N_{\times} \\ \eta_{\times}^{\star\mathrm{pst}} &= \eta_{\times}^{\star} + N_{\times} \\ \boldsymbol{\mu}_{\times}^{\star\mathrm{pst}} &= (N_{\times}\bar{\mathbf{x}}_{\times}^{\star} + \eta_{\times}^{\star}\boldsymbol{\mu}_{\times}^{\star})/(N_{\times}+\eta_{\times}^{\star}) \\ (\boldsymbol{\Lambda}_{\times}^{\star\mathrm{pst}})^{-1} &= (\boldsymbol{\Lambda}_{\times}^{\star})^{-1} + N_{\times}\bar{\mathbf{R}}_{\times}^{\star} + \\ & \quad N_{\times}\eta_{\times}^{\star}(\boldsymbol{\mu}_{\times}^{\star} - \bar{\mathbf{x}}_{\times}^{\star})(\boldsymbol{\mu}_{\times}^{\star} - \bar{\mathbf{x}}_{\times}^{\star})^t/\left(N_{\times}+\eta_{\times}^{\star}\right) \end{cases}$$

where "pst" stands for posterior. Regarding p, the posterior distribution is a Beta density with parameters: $\alpha^{\mathrm{pst}} = N_{\mathcal{P}} + \alpha, \beta^{\mathrm{pst}} = N_{\mathcal{H}} + \beta$. Then, rearranging different factors, calculation of the integral in (1) is possible and yields:

$$f_{\boldsymbol{\mathcal{X}},\mathbf{B}|\boldsymbol{\Delta}}(\mathbf{x},\mathbf{b}|\boldsymbol{\delta}) \propto \frac{\mathcal{K}_{\mathcal{P}}^{+\mathrm{pst}}}{\mathcal{K}_{\mathcal{P}}^{+\mathrm{pri}}}\frac{\mathcal{K}_{\mathcal{H}}^{+\mathrm{pst}}}{\mathcal{K}_{\mathcal{H}}^{+\mathrm{pri}}}\frac{\mathcal{K}_{\mathcal{C}}^{-\mathrm{pst}}}{\mathcal{K}_{\mathcal{C}}^{-\mathrm{pri}}} \quad (3)$$

where $\mathcal{K}$ is the normalizing constant of the Normal-Wishart (given in Appendix),"pri" stands for prior. As a consequence, the probabilities are very easy to compute despite the complexity of the problem. Besides, given that all candidate models are equiprobable, from Eq. (2) we can deduce the posterior probability for the 2^P models. The selected model is the one which maximizes this probability.

2.2. Hyperparameter choice

The probability (3) depends on the parameters of the Normal-Wishart distributions $(\nu_{\times},\eta_{\times},\boldsymbol{\mu}_{\times},\boldsymbol{\Lambda}_{\times})$ for $\times \in \{\mathcal{P},\mathcal{H},\mathcal{C}\}$ referred to as hyperparameters. In a non-informative case, the parameters $(\nu_{\times},\eta_{\times},\boldsymbol{\mu}_{\times},\boldsymbol{\Lambda}_{\times}^{-1})$ tends to $(0,0,0,0)$ and the proportionality coefficient in (3) has an indeterminated form.

To tune these parameters we propose to resort to poorly informative priors based on real-life orders of magnitudes for involved variables (e.g. μg per ml). To this end, we establish

a relation between prior mean and covariance for $(\mathbf{m}_\times, \Gamma_\times)$ and the hyperparameters:

$$\begin{aligned} E(\Gamma_\times) &= \nu_\times \Lambda_\times \\ E(\mathbf{m}_\times) &= \mu_\times \\ V(\mathbf{m}_\times) &= \Lambda_\times^{-1} / \left[\eta_\times(\nu_\times - P - 1)\right] \\ \mathrm{cov}(\Gamma_\times^{i,j}, \Gamma_\times^{k,l}) &= \nu_\times(\Lambda_\times^{il}\Lambda_\times^{jk} + \Lambda_\times^{ik}\Lambda_\times^{jl}) \end{aligned}$$

where superscript i, j denotes the (i, j) entry of matrices. So, accounting for real-world orders of magnitudes of $\mathbf{m}_\times$ and $\Gamma_\times$, the prior parameters $(\nu_\times, \eta_\times, \boldsymbol{\mu}_\times, \boldsymbol{\Lambda}_\times)$ can be calculated and substituted in the probability (3).

3. RESULTS

We performed several simulations to illustrate the method performances for selection of biomarkers (discriminant proteins). In the first place, we focus on the univariate problem wherein a given protein is selected as discriminant or not, then $\Delta = \{+, -\}$. This simpler issue allows to precisely quantify the influence on the selection of the difference between the empirical statistics of the concentrations for the healthy and the pathological subsets of individuals. The first considered scenario consists in imposing the same variance for both the healthy and pathological populations. Then, the mean of the concentration of the healthy individuals is assumed equal to 10μg whereas the one of the pathological population is made to vary between 0 and 20μg.

In Figure 1-top, the posterior probability of the model $\Delta = +$ is presented as a function of the mean difference for several values of the variance $(20, 50, 100)$. For small empirical variance, the posterior probability of the model $\Delta = +$ is approximately 1 even for low values of the mean difference. Conversely, for high empirical variance the model $\Delta = +$ is selected only for large value of mean difference. We conclude that, the higher the variance, the larger must be the mean difference for the protein to be validated as a biomarker.

In the second scenario, both the empirical mean of the healthy and pathological populations are set to 50μg and the variance of the pathological population equals to 20, while the variance of the healthy ones takes different values between 20 and 200. Results are shown in Figure 1-bottom. We note that the protein is selected as biomarker only for large values of variances differences, otherwise the protein is decided to be non discriminant.

In the next, performances of the model selection method are studied. For fixed number of proteins P, we consider a set of true models with a number of biomarker varying from 0 to P. For each true model, $N_r = 10000$ realisations are simulated as a set of data given by individual state and proteins concentrations according to the following description. The individual state is governed by the Bernoulli variable. For discriminant proteins, the concentrations are

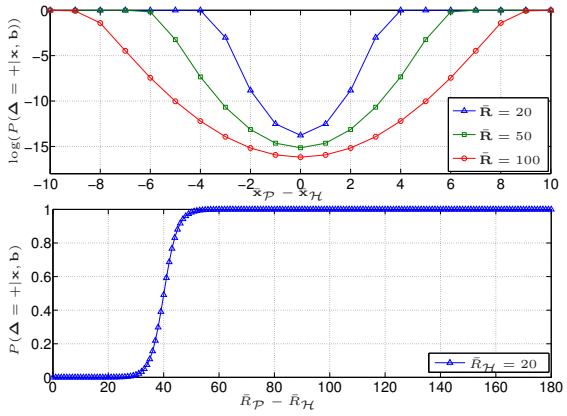

Fig. 1. top: Log-posterior-probability of discriminant protein as a function of mean difference for several variances. bottom: posterior-probability of discriminant protein as a function of healthy variance for a fixed pathological variance.

distributed as $\mathcal{N}(\mathbf{x}_n^+; \mathbf{m}_\mathcal{H}^+, \mathbf{\Gamma}_\mathcal{H}^+)$ or $\mathcal{N}(\mathbf{x}_n^+; \mathbf{m}_\mathcal{P}^+, \mathbf{\Gamma}_\mathcal{P}^+)$ depending on the simulated individual state. For non discriminant ones they are given by $\mathcal{N}(\mathbf{x}_n^-; \mathbf{m}_\mathcal{C}^-, \mathbf{\Gamma}_\mathcal{C}^-)$. The parameters $(\mathbf{m}_\times^\star, \mathbf{\Gamma}_\times^\star)$ where $\times \in \{\mathcal{P}, \mathcal{H}, \mathcal{C}\}$ and $\star \in \{+, -\}$ are distributed as $\mathcal{NW}(\nu_\times, \eta_\times, \boldsymbol{\mu}_\times, \boldsymbol{\Lambda}_\times)$, where hyperparameters $(\nu_\times, \eta_\times, \boldsymbol{\mu}_\times, \boldsymbol{\Lambda}_\times)$ are calculated as explained in Section 2.2. In the following we consider $P = 4$ proteins and $N = 1000$ individuals, with $E(\mathbf{m}_\times) = 10$, $V(\mathbf{m}_\times) = 200$, $E(\mathbf{\Gamma}_\times) = 35$ and $V(\mathbf{\Gamma}_\times) = 200$.

Then, for each data set, the posterior probability is calculated for each configurations and the most probable one is selected. Performance are the evaluated by the Bayesian risk, that is to say the Mean Selection Error Rate (MSER) denoted τ and given by:

$$\tau = \sum_{i=0}^{P} SER_i$$

where SER_i is the Selection Error Rate given by $SER_i = Q_i\omega_i/N_r$, with Q_i is the number of realisations where the selected model is different from the true one and ω_i is the proportion of models with i biomarkers. Results are compared with the Student's t-test based on comparison of the means of the proteins concentrations between the two cohorts $\mathcal{H}$ and $\mathcal{P}$. When the hypothesis of equality of the means is rejected, the protein is declared as biomarker. The SER is calculated for different values of the error of the first kind denoted α.

Fig.2 shows the MSER $\tau(\%)$, dashed curves refers to the t-test while solid one refers to the proposed method. For the latter, the risk is constant since it is independent on α. Besides, it is clear that MSER of the proposed method is lower than the one achieved by the t-test. This result is coherent with the optimality property of the proposed method that minimizes the Bayesian risk. Moreover, unlike t-test, the proposed method is based on multivariate approach which takes

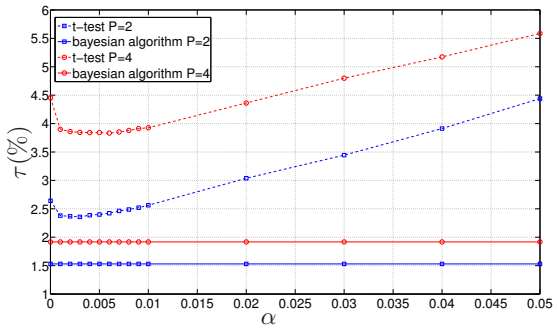

Fig. 2. MSER $\tau(\%)$ for the proposed method and for the t-test with different values of the first kind error α and proteins number $P = 2$ and $P = 4$. Number of individuals $N = 100$.

into account possible correlation between proteins. Furthermore, the gain obtained thanks to the proposed method increases with P, since the larger P, the higher the possibility of presence of correlation between them. This result attests the relevancy of multivariate approach.

Tab. 1 proposes a study of the selection error rate SER_i for different number individuals ($N = 100$ and $N = 1000$) and different number of biomarker denoted by NB. We observe that performance are improved when increasing the number of individuals N, which is explained by the improvement of the estimation precision of the posteriori probability for large N. Moreover, even for limited number of individuals, SER is lower than 5%, which agree the results established in literature of proteomics. Besides, we note that the SER is a bit higher when the number of biomarkers > 1 which is explained by the added number of parameters to estimate.

NB:	0	1	2	3	4
$SER\,(\%)$:	0.068	0.402	0.716	0.537	0.191
$SER\,(\%)$:	0.001	0.025	0.026	0.012	0.010

Table 1. $SER\,(\%)$ for $P = 4$ and $N = 100$ (first row) and $N = 1000$ (second row).

Now, the algorithm performances are assessed for larger number of proteins. For $P = 8$, the number of candidate models is $2^8 = 256$, which is much larger than previously. Moreover, as shown in Tab. 2, the performances are not really degraded when compared with the second row of Tab. 1. This result affirms the robustness of the optimal selection approach and of the computation of the posterior probability.

NB:	0	4	8
$SER\,(\%)$:	0.0001	0.0027	0.0008

Table 2. $SER\,(\%)$ with $N = 1000$ and $P = 8$.

4. CONCLUSIONS AND PERSPECTIVES

Biomarker discovery is a crucial question with tremendous applications and it is also a challenging statistical task. From this viewpoint, an important issue is related to variable selection and the paper presents a novel approach for it. It is developped in a hierarchical Bayesian framework and relies on an optimal strategy, i.e. the minimization of a risk. The hyperparameters are tuned to obtain poorly informative priors. The procedure relies on the comparison of the 2^P configurations from P proteins, the most a posteriori probable model is finally retained, and thus defines the selected variables. The main difficulty is the required integration with respect to the unknown parameters and an important contribution is to provide a closed-form expression. The developements then benefit from very low complexity and ease of implementation.

The optimal decision proved to be suited for variable selection in a complex context. The effectiveness of the method is assessed by a theoritical characterization and a simulated study that is in accordance with the theoritical optimality. Furthermore, the proposed method compares favorably with the Student test usually applied in this context.

We intend to further investigate the performances of the method, firstly through a comparison with some existing approaches (BIC, AIC,...). In addition, we intend to take advantage of the method for other applications e.g. other biomedical applications (genomics,...), astrophysical data...

5. ACKNOWLEDGEMENT

This work was supported by the ANR BHI-PRO project, grant ANR-Blanc of the French Agence Nationale de la Recherche. The authors wish to acknowledge Caroline Truntzer (CLIPP, Dijon) and others members of the BHI-PRO project for numerous enriching discussions.

A. APPENDIX: NORMAL-WISHART

Let $\mathbf{m} \in \mathbb{R}^P$ and $\mathbf{\Gamma} \in \mathbb{R}^{P\times P}$ be a vector and a definite positive matrix. $(\mathbf{m}, \mathbf{\Gamma})$ follows Normal-Wishart distribution with parameters $(\nu, \eta, \boldsymbol{\mu}, \mathbf{\Lambda})$ if:

$$f_{\mathbf{m},\mathbf{\Gamma}}(\mathbf{m}, \mathbf{\Gamma}) = \mathcal{K}^{-1} \det(\mathbf{\Gamma})^{(\nu-P)/2}$$

$$\exp\left[-(\mathrm{tr}[\mathbf{\Gamma}\mathbf{\Lambda}^{-1}] + \eta(\mathbf{m}-\boldsymbol{\mu})^t\mathbf{\Gamma}(\mathbf{m}-\boldsymbol{\mu}))/2\right]$$

where $\mathcal{K}$ is normalization constant:

$$\mathcal{K} = (2\pi)^{P/2} 2^{\nu P/2} \eta^{-P/2} \det(\mathbf{\Lambda})^{\nu/2} \Gamma_P(\nu/2)$$

and Γ_P is multivariate gamma function.

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, November 2000.

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

[3] P. Szacherski, J.-F. Giovannelli, and P. Grangeat, "Joint Bayesian hierarchical inversion-classification and application in proteomics." in *Proceedings of the International Conference on Statistical Signal Processing*, Nice, France, June 2011.

[4] P. Szacherski, J.-F. Giovannelli, L. Gerfault, and P. Grangeat, "Apprentissage supervisé robuste de caractéristiques de classes. Application en protéomique." in *Actes du 23* e *colloque GRETSI*, Bordeaux France, September 2011.

[5] P. Grangeat, L. Gerfault, J.-F. Giovannelli, C. Paulus, and V. Brun, "Reconstruction de profils moléculaires en protéomique." Grenoble, France: Séminaire Daniel Dautreppe, Imagerie médicale et modélisation multi-échelle (Séminaire invité), December 2010.

[6] G. Strubel, J.-F. Giovannelli, C. Paulus, L. Gerfault, and P. Grangeat, "Bayesian estimation for molecular profile reconstruction in proteomics based on liquid chromatography and mass spectrometry," in *Proceedings of IEEE EMBS*, Lyon, France, August 2007, pp. 5979–5982.

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

[8] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic and Control*, vol. AC-19, no. 6, pp. 716–723, December 1974.

[9] E. I. George and R. E. McCulloch, "Variable selection via the Gibbs sampling," *Journal of Acoustical Society America*, vol. 88, no. 423, pp. 881–889, September 1993.

[10] H. Zou and T. Hastie, "Regularization and variable selection via the Elastic Net," *Journal of the Royal Statistical Society B*, vol. 67, pp. 301–320, 2005.

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

[12] 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.

[13] Y. Benjamin and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society B*, vol. 57, no. 1, pp. 289–300, 1995.