Vers la découverte et la sélection : une approche bayésienne — Projet BHI-PRO —

Cette proposition concerne la question de l'identification de bio-marqueurs protéiques dans le cadre du projet BHI-PRO soutenu par l'ANR. En résumé, on s'intéresse à une situation dans laquelle on observe les concentrations pour un jeu de protéines relatif à deux populations d'individu et on se pose la question de la possibilité de discriminer entre les deux populations sur la base de ces protéines.

Pour cela, on considère une population d'individus indexés par n et on s'intéresse à leur statut biologique b_n qui prend deux valeurs m (malade) ou s (sain) et à leurs concentrations en P protéines regroupées dans $\boldsymbol{x}_n \in \mathbb{R}^P$. On observe N individus et on note $\mathbf{b} = [\mathsf{b}_1, \dots \mathsf{b}_N]$ le vecteur des N statuts et $\boldsymbol{\mathcal{X}} = [\boldsymbol{x}_1, \dots \boldsymbol{x}_N]$ la matrice des N vecteurs de P concentrations. On note $\mathcal{I}_\mathsf{b}$ et N_b l'ensemble des indices des individus de statut b et son cardinal.

Le statut est décrit par une variable de Bernoulli ($\mathcal{B}$) de paramètre p et les concentrations sont décrites (conditionnellement au statut) par des densités gaussiennes ($\mathcal{N}$) de moyennes et précisions $(\boldsymbol{m}_\times, \boldsymbol{\Gamma}_\times)$. Ces paramètres sont également probabilisés : pour des questions pratiques, on choisit un modèle Bêta ($\mathcal{P}$) pour p et des modèles Gauss-Wishart ($\mathcal{NW}$) pour les $(\boldsymbol{m}_\times, \boldsymbol{\Gamma}_\times)$. Les hyperparamètres sont fixés de manière à inclure les informations disponibles sur les paramètres. Dans le présent travail, on se focalise sur les a priori diffus (non-informatif).

On se pose la question[1] du caractère discriminant du jeu de protéines considéré : « décider » si ce jeu témoigne de l'existence de deux sous-populations distinctes ou bien en réalité d'une seule. Dans le premier cas on dira que le modèle est discriminant (on écrira $\mathfrak{M} = \oplus$) et dans le second on dira qu'il ne l'est pas (on écrira $\mathfrak{M} = \ominus$).

Deux modèles concurrents

- $\mathfrak{M} = \oplus$ – Les observations sont pilotées par une distribution de Bernoulli et deux gaussiennes

$$\begin{cases} \boldsymbol{X} \,|\, \mathrm{B} = \mathsf{m}, \boldsymbol{\theta}, \mathfrak{M} = \oplus & \sim \mathcal{N}(\boldsymbol{x}\,;\, \boldsymbol{m}_\mathsf{m}, \boldsymbol{\Gamma}_\mathsf{m}^{-1}) \\ \boldsymbol{X} \,|\, \mathrm{B} = \mathsf{s}, \boldsymbol{\theta}, \mathfrak{M} = \oplus & \sim \mathcal{N}(\boldsymbol{x}\,;\, \boldsymbol{m}_\mathsf{s}, \boldsymbol{\Gamma}_\mathsf{s}^{-1}) \\ \mathrm{B} \,|\, \boldsymbol{\theta}, \mathfrak{M} = \oplus & \sim \mathcal{B}(\mathsf{b}\,;\, p) \end{cases}$$

  La densité pour la concentration, marginale vis-à-vis du statut, est alors un mélange de deux gaussiennes. Ces distributions font intervenir les paramètres inconnus $\boldsymbol{\theta} = [\boldsymbol{m}_\mathsf{m}, \boldsymbol{m}_\mathsf{s}, \boldsymbol{\Gamma}_\mathsf{m}, \boldsymbol{\Gamma}_\mathsf{s}, p]$ que l'on probabilise sous une a priori séparée $\pi_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = \pi_\mathsf{m}(\boldsymbol{m}_\mathsf{m}, \boldsymbol{\Gamma}_\mathsf{m})\ \pi_\mathsf{s}(\boldsymbol{m}_\mathsf{s}, \boldsymbol{\Gamma}_\mathsf{s})\ \pi_P(p)$ avec

$$\begin{cases} (\boldsymbol{m}_\mathsf{m}, \boldsymbol{\Gamma}_\mathsf{m}) \,|\, \mathfrak{M} = \oplus & \sim \mathcal{NW}(\boldsymbol{m}_\mathsf{m}, \boldsymbol{\Gamma}_\mathsf{m}\,;\, \nu_\mathsf{m}, \eta_\mathsf{m}, \boldsymbol{\mu}_\mathsf{m}, \boldsymbol{\Lambda}_\mathsf{m}) \\ (\boldsymbol{m}_\mathsf{s}, \boldsymbol{\Gamma}_\mathsf{s}) \,|\, \mathfrak{M} = \oplus & \sim \mathcal{NW}(\boldsymbol{m}_\mathsf{s}, \boldsymbol{\Gamma}_\mathsf{s}\,;\, \nu_\mathsf{s}, \eta_\mathsf{s}, \boldsymbol{\mu}_\mathsf{s}, \boldsymbol{\Lambda}_\mathsf{s}) \\ p \,|\, \mathfrak{M} = \oplus & \sim \mathcal{P}(p\,;\, \alpha, \beta) \end{cases}$$

- $\mathfrak{M} = \ominus$ – Les observations sont pilotées par une distribution de Bernoulli et une (seule) gaussienne

$$\begin{cases} \boldsymbol{X} \,|\, \mathrm{B} = \mathsf{m}, \boldsymbol{\theta}, \mathfrak{M} = \ominus & \sim \mathcal{N}(\boldsymbol{x}\,;\, \boldsymbol{m}_\mathsf{c}, \boldsymbol{\Gamma}_\mathsf{c}) \\ \boldsymbol{X} \,|\, \mathrm{B} = \mathsf{s}, \boldsymbol{\theta}, \mathfrak{M} = \ominus & \sim \mathcal{N}(\boldsymbol{x}\,;\, \boldsymbol{m}_\mathsf{c}, \boldsymbol{\Gamma}_\mathsf{c}) \\ \mathrm{B} \,|\, \boldsymbol{\theta}, \mathfrak{M} = \ominus & \sim \mathcal{B}(\mathsf{b}\,;\, p) \end{cases}$$

  Ici, $\boldsymbol{X}$ et B sont indépendants conditionnellement à $\mathfrak{M}$ ce qui est bien conforme à l'idée que le jeu de protéines ne permet pas de discriminer. Ces distributions font intervenir les paramètres inconnus $\boldsymbol{\theta} = [\boldsymbol{m}_\mathsf{c}, \boldsymbol{\Gamma}_\mathsf{c}, p]$ que l'on probabilise également sous une a priori séparée $\pi_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = \pi_\mathsf{c}(\boldsymbol{m}_\mathsf{c}, \boldsymbol{\Gamma}_\mathsf{c})\ \pi_P(p)$ avec

$$\begin{cases} (\boldsymbol{m}_\mathsf{c}, \boldsymbol{\Gamma}_\mathsf{c}) \,|\, \mathfrak{M} = \ominus & \sim \mathcal{NW}(\boldsymbol{m}_\mathsf{c}, \boldsymbol{\Gamma}_\mathsf{c}\,;\, \nu_\mathsf{c}, \eta_\mathsf{c}, \boldsymbol{\mu}_\mathsf{c}, \boldsymbol{\Lambda}_\mathsf{c}) \\ p \,|\, \mathfrak{M} = \ominus & \sim \mathcal{P}(p\,;\, \alpha, \beta) \end{cases}$$

On probabilise aussi le modèle lui-même : $\Pr[\mathfrak{M} = \oplus]$ et $\Pr[\mathfrak{M} = \ominus]$ et on prendra $1/2$ pour chacune.

[1] La question suivante concerne la sélection des protéines les plus discriminantes mais ne sera pas détaillée ici. Les questions d'apprentissage en tant que tel et de classification sortent du cadre de ce travail.

Décision optimale et cote a posteriori

	$\mathfrak{M}^\star = \oplus$	$\mathfrak{M}^\star = \ominus$
$\widehat{\mathfrak{M}} = \oplus$	VD / $\mathcal{C}(\oplus, \oplus)$	FD / $\mathcal{C}(\oplus, \ominus)$
$\widehat{\mathfrak{M}} = \ominus$	FND / $\mathcal{C}(\ominus, \oplus)$	VND / $\mathcal{C}(\ominus, \ominus)$

TAB. 1 – Résultats de décision et coûts associés.

La suite du jeu consiste à construire un « décideur », *i.e.,* une fonction ψ de l'ensemble des observations vers l'ensemble des décisions. Pour cela on s'intéresse aux quatre situations possibles (voir Tab. 1) : deux décisions correctes (vrai discriminant et vrai non-discriminant) et deux erronées (faux discriminant et faux non-discriminant). On se alors un coût (indiqué dans Tab. 1) qui quantifie l'impact de chacune des situations. On s'intéresse au cas particulier : $\mathcal{C}(\oplus, \oplus) = \mathcal{C}(\ominus, \ominus) = 0$ pour les bonnes décisions et au cas symétrique $\mathcal{C}(\ominus, \oplus) = \mathcal{C}(\oplus, \ominus)$ pour les mauvaises. On définit alors le risque $\mathcal{R}(\psi)$ pour ψ en prenant le coût moyen :

$$\mathcal{R}(\psi) = \mathbb{E}\left[\mathcal{C}(\psi(\boldsymbol{\mathcal{X}}, \mathbf{b}), \mathfrak{M})\right] \tag{1}$$

et il s'agit d'une *triple* moyenne, sur le modèle $\mathfrak{M}$, sur les paramètres $\boldsymbol{\theta}$ et sur les observations $(\boldsymbol{\mathcal{X}}, \mathbf{b})$. Il mesure globalement la performance de ψ et on choisit le décideur optimal ψ_{opt} comme son minimiseur. On montre qu'il repose sur la cote a posteriori, *i.e.,* le rapport des probabilités a posteriori des deux modèles :

$$\rho = \rho(\boldsymbol{\mathcal{X}}, \mathbf{b}) = \Pr\left[\mathfrak{M} = \oplus \,|\, \boldsymbol{\mathcal{X}}, \mathbf{b}\right] \Big/ \Pr\left[\mathfrak{M} = \ominus \,|\, \boldsymbol{\mathcal{X}}, \mathbf{b}\right] \tag{2}$$

et qu'il choisit sa décision en comparant ρ au seuil $\sigma = \left[\mathcal{C}(\ominus, \oplus) - \mathcal{C}(\oplus, \oplus)\right] / \left[\mathcal{C}(\oplus, \ominus) - \mathcal{C}(\ominus, \ominus)\right]$

$$\psi(\boldsymbol{\mathcal{X}}, \mathbf{b}) = \oplus \text{ si } \rho(\boldsymbol{\mathcal{X}}, \mathbf{b}) > \sigma \quad \text{et} \quad \psi(\boldsymbol{\mathcal{X}}, \mathbf{b}) = \ominus \text{ si } \rho(\boldsymbol{\mathcal{X}}, \mathbf{b}) < \sigma$$

Ce seuil vaut 1 dans notre cas particulier, *i.e.,* l'hypothèse la plus probable a posteriori emporte la décision.

Ce résultat est très puissant car il assure la supériorité de ce décideur sur tous les autres : il n'existe pas de procédure, pas d'algorithme, pas de code informatique, pas de méthode… qui produise un meilleur résultat et cela appelle un commentaire. L'optimalité est obtenue au sens très particulier de (1) et il est construit dans un cadre fixé entre autres par les distributions choisies : une modification de ces distributions entraîne naturellement une modification de l'estimateur. De plus, ces distributions font apparaître des hyperparamètres qui peuvent être délicats à choisir et nous verrons également que l'idée des distributions non-informatives génère des difficultés.

Cote a posteriori et facteur de Bayes

Intéressons-nous à la cote a posteriori (2). En multipliant numérateur et dénominateur par la distribution pour $(\boldsymbol{\mathcal{X}}, \mathbf{b})$ on fait apparaître la distribution jointe pour $(\mathfrak{M}, \boldsymbol{\mathcal{X}}, \mathbf{b})$ que l'on refactorise en conditionnant par $\mathfrak{M}$:

$$\rho = \underbrace{\frac{\Pr\left[\mathfrak{M} = \oplus \,|\, \boldsymbol{\mathcal{X}}, \mathbf{b}\right]}{\Pr\left[\mathfrak{M} = \ominus \,|\, \boldsymbol{\mathcal{X}}, \mathbf{b}\right]}}_{\textit{Cote a posteriori}} = \frac{\Pr\left[\mathfrak{M} = \oplus, \boldsymbol{\mathcal{X}}, \mathbf{b}\right]}{\Pr\left[\mathfrak{M} = \ominus, \boldsymbol{\mathcal{X}}, \mathbf{b}\right]} = \underbrace{\frac{f_{\boldsymbol{\mathcal{X}}, \mathrm{B} \,|\, \mathfrak{M}}(\boldsymbol{\mathcal{X}}, \mathbf{b} \,|\, \mathfrak{M} = \oplus)}{f_{\boldsymbol{\mathcal{X}}, \mathrm{B} \,|\, \mathfrak{M}}(\boldsymbol{\mathcal{X}}, \mathbf{b} \,|\, \mathfrak{M} = \ominus)}}_{\textit{Facteur de Bayes}} \underbrace{\frac{\Pr\left[\mathfrak{M} = \oplus\right]}{\Pr\left[\mathfrak{M} = \ominus\right]}}_{\textit{Cote a priori}} . \tag{3}$$

On voit apparaître une quantité clé : le facteur de Bayes *i.e.,* le rapport des évidences des deux modèles. Ces évidences ont la forme d'une vraisemblance : elles décrivent la distribution des observations sachant la quantité d'intérêt et mesurent ainsi une adéquation observations – modèle. On les obtient par marginalisation

$$f_{\boldsymbol{\mathcal{X}}, \mathrm{B} \,|\, \mathfrak{M}}(\boldsymbol{\mathcal{X}}, \mathbf{b} \,|\, \mathfrak{M} = \odot) = \int_{\boldsymbol{\theta}} f_{\boldsymbol{\mathcal{X}}, \mathrm{B}, \boldsymbol{\Theta} \,|\, \mathfrak{M}}(\boldsymbol{\mathcal{X}}, \mathbf{b}, \boldsymbol{\theta} \,|\, \mathfrak{M} = \odot) \, \mathrm{d}\boldsymbol{\theta}$$

avec $\odot = \ominus$ et $\odot = \oplus$. Cette opération peut s'avérer délicate dans certains cas et nécessiter des calculs numériques intensifs, par exemple par échantillonnage stochastique. Ici, les choses se déroulent plus aisément grâce aux choix pour les diverses lois. La quantité à marginaliser se décompose en une vraisemblance et une a priori :

$$f_{\boldsymbol{\mathcal{X}},\mathbf{B},\boldsymbol{\Theta}\,|\,\mathfrak{M}}(\boldsymbol{\mathcal{X}},\mathbf{b},\boldsymbol{\theta}\,|\,\mathfrak{M}=\odot) = f_{\boldsymbol{\mathcal{X}},\mathbf{B}\,|\,\boldsymbol{\Theta},\mathfrak{M}}(\boldsymbol{\mathcal{X}},\mathbf{b}\,|\,\boldsymbol{\theta},\mathfrak{M}=\odot)\;\pi_{\boldsymbol{\Theta}\,|\,\mathfrak{M}}(\boldsymbol{\theta}\,|\,\mathfrak{M}=\odot)$$

qui sont décrites ci-après. Pour $\mathfrak{M} = \oplus$, on voit apparaître deux gaussiennes (malades et sains) alors que pour $\mathfrak{M} = \ominus$ tous les individus sont regroupés sous une unique gaussienne.

- Pour ce qui est de la vraisemblance des paramètres attachée aux observations, on a

$$\begin{cases} f_{\boldsymbol{\mathcal{X}},\mathbf{B}\,|\,\boldsymbol{\Theta},\mathfrak{M}}(\boldsymbol{\mathcal{X}},\mathbf{b}\,|\,\boldsymbol{\theta},\mathfrak{M}=\oplus) = \left[\prod_{\mathcal{I}_\mathsf{m}} \mathcal{N}(\boldsymbol{x}_n\,;\,\boldsymbol{m}_\mathsf{m},\boldsymbol{\Gamma}_\mathsf{m})\right]\left[\prod_{\mathcal{I}_\mathsf{s}} \mathcal{N}(\boldsymbol{x}_n\,;\,\boldsymbol{m}_\mathsf{s},\boldsymbol{\Gamma}_\mathsf{s})\right]\left[p^{N_\mathsf{m}}(1-p)^{N_\mathsf{s}}\right] \\ f_{\boldsymbol{\mathcal{X}},\mathbf{B}\,|\,\boldsymbol{\Theta},\mathfrak{M}}(\boldsymbol{\mathcal{X}},\mathbf{b}\,|\,\boldsymbol{\theta},\mathfrak{M}=\ominus) = \left[\prod_{\mathcal{I}_\mathsf{m}\mathcal{I}_\mathsf{s}} \mathcal{N}(\boldsymbol{x}_n\,;\,\boldsymbol{m}_\mathsf{c},\boldsymbol{\Gamma}_\mathsf{c})\right]\left[p^{N_\mathsf{m}}(1-p)^{N_\mathsf{s}}\right] \end{cases} \tag{4}$$

- Pour compléter le tableau, concernant les deux densités a priori pour les paramètres

$$\begin{cases} \pi_{\boldsymbol{\Theta}\,|\,\mathfrak{M}}(\boldsymbol{\theta}\,|\,\mathfrak{M}=\oplus) = \mathcal{NW}\left(\boldsymbol{m}_\mathsf{m},\boldsymbol{\Gamma}_\mathsf{m}\,;\,\nu_\mathsf{m},\eta_\mathsf{m},\boldsymbol{\mu}_\mathsf{m},\boldsymbol{\Lambda}_\mathsf{m}\right)\,\mathcal{NW}\left(\boldsymbol{m}_\mathsf{s},\boldsymbol{\Gamma}_\mathsf{s}\,;\,\nu_\mathsf{s},\eta_\mathsf{s},\boldsymbol{\mu}_\mathsf{s},\boldsymbol{\Lambda}_\mathsf{s}\right)\,\mathcal{P}\left(p\,;\,\alpha,\beta\right) \\ \pi_{\boldsymbol{\Theta}\,|\,\mathfrak{M}}(\boldsymbol{\theta}\,|\,\mathfrak{M}=\ominus) = \mathcal{NW}\left(\boldsymbol{m}_\mathsf{c},\boldsymbol{\Gamma}_\mathsf{c}\,;\,\nu_\mathsf{c},\eta_\mathsf{c},\boldsymbol{\mu}_\mathsf{c},\boldsymbol{\Lambda}_\mathsf{c}\right)\,\mathcal{P}\left(p\,;\,\alpha,\beta\right) \end{cases} \tag{5}$$

Effectuant les produits (4)-(5), on fait apparaître six facteurs ($\mathfrak{M} = \oplus$) ou quatre ($\mathfrak{M} = \ominus$). Un élément technique crucial est la possibilité d'intégrer en s'appuyant sur les propriétés des densités Gauss-Wishart et Bêta :

$$\begin{aligned} f_{\boldsymbol{\mathcal{X}},\mathbf{B}\,|\,\mathfrak{M}}(\boldsymbol{\mathcal{X}},\mathbf{b}\,|\,\mathfrak{M}=\oplus) &= \frac{K_{\mathcal{NW}}(\nu_\mathsf{m}^{\text{pst}},\eta_\mathsf{m}^{\text{pst}},\boldsymbol{\mu}_\mathsf{m}^{\text{pst}},\boldsymbol{\Lambda}_\mathsf{m}^{\text{pst}})}{K_{\mathcal{NW}}(\nu_\mathsf{m},\eta_\mathsf{m},\boldsymbol{\mu}_\mathsf{m},\boldsymbol{\Lambda}_\mathsf{m})}\frac{K_{\mathcal{NW}}(\nu_\mathsf{s}^{\text{pst}},\eta_\mathsf{s}^{\text{pst}},\boldsymbol{\mu}_\mathsf{s}^{\text{pst}},\boldsymbol{\Lambda}_\mathsf{s}^{\text{pst}})}{K_{\mathcal{NW}}(\nu_\mathsf{s},\eta_\mathsf{s},\boldsymbol{\mu}_\mathsf{s},\boldsymbol{\Lambda}_\mathsf{s})}\frac{K_{\mathcal{B}}(\alpha^{\text{pst}},\beta^{\text{pst}})}{K_{\mathcal{B}}(\alpha,\beta)} \\ f_{\boldsymbol{\mathcal{X}},\mathbf{B}\,|\,\mathfrak{M}}(\boldsymbol{\mathcal{X}},\mathbf{b}\,|\,\mathfrak{M}=\ominus) &= \frac{K_{\mathcal{NW}}(\nu_\mathsf{c}^{\text{pst}},\eta_\mathsf{c}^{\text{pst}},\boldsymbol{\mu}_\mathsf{c}^{\text{pst}},\boldsymbol{\Lambda}_\mathsf{c}^{\text{pst}})}{K_{\mathcal{NW}}(\nu_\mathsf{c},\eta_\mathsf{c},\boldsymbol{\mu}_\mathsf{c},\boldsymbol{\Lambda}_\mathsf{c})}\frac{K_{\mathcal{B}}(\alpha^{\text{pst}},\beta^{\text{pst}})}{K_{\mathcal{B}}(\alpha,\beta)} \end{aligned}$$

qui font apparaître les constantes de normalisation pour les paramètres a posteriori et a priori. Pour avoir la cote a posteriori (3) il suffit de faire le ratio de ces rapports de constantes de normalisation. Nous avons là un résultat tout à fait remarquable puisqu'il fournit la cote a posteriori sans calcul numérique intensif et il suffit de calculer les paramètres a posteriori et les constantes de normalisation, comme donnés en annexes, Eq. (6)-(7). Ils font apparaitre, *in fine* les moyennes et covariances empiriques (données Eq. (8)) qui interviennent dans les $\boldsymbol{\mu}_\times^{\text{pst}}$ et $\boldsymbol{\Lambda}_\times^{\text{pst}}$.

Application au cas d'une unique protéine

Le résultat précédent est valide en dimension P quelconque et inclut le cas de concentrations corrélées. Cette section en présente une première illustration dans le cas $P = 1$. L'expression devient alors :

$$\rho \propto \frac{B_P(N_\mathsf{m}/2, N_\mathsf{s}/2)}{(1-\bar{p})^{PN_\mathsf{s}/2}\bar{p}^{PN_\mathsf{m}/2}}\left(1+\frac{1-\bar{p}}{\bar{p}}\frac{\bar{\boldsymbol{R}}_\mathsf{s}}{\bar{\boldsymbol{R}}_\mathsf{m}}\right)^{N_\mathsf{m}/2}\left(1+\frac{\bar{p}}{1-\bar{p}}\frac{\bar{\boldsymbol{R}}_\mathsf{m}}{\bar{\boldsymbol{R}}_\mathsf{s}}\right)^{N_\mathsf{s}/2}\left(1+\bar{p}(1-\bar{p})\,\frac{(\bar{\boldsymbol{x}}_\mathsf{m}-\bar{\boldsymbol{x}}_\mathsf{s})^2}{\bar{p}\bar{\boldsymbol{R}}_\mathsf{m}+(1-\bar{p})\bar{\boldsymbol{R}}_\mathsf{s}}\right)^{N/2}$$

et elle fait naturellement apparaître les proportions, les moyennes et les variances empirique (voir annexe).

Remarque — *L'expression de ρ est donnée à un facteur multiplicatif près. Celui-ci est une fonction uniquement des hyperparamètres et peut-être difficile à manipuler lorsque l'on considère une a priori non-informative. Cela étant, il n'influence que la valeur du seuil dans la procédure de décision.*

Quoiqu'il en soit, le logarithme de la cote a posteriori en faveur de $\mathfrak{M} = \oplus$ (à un terme additif près) est présenté Fig. 1 comme fonction de la différence entre moyennes pour plusieurs variances et avec $N_\mathsf{m} = 50$ et $N_\mathsf{s} = 100$.

On observe un premier résultat attendu, conforme à l'intuition et à l'existant : pour des variances fixées, la cote augmente avec la différence des moyennes. En revanche, pour une différence de moyennes fixée, la cote n'augmente pas systématiquement lorsque la variance diminue ce qui peut sembler contre-intuitif. En réalité, lorsque la différence de moyenne est très faible (ou nulle à la limite), c'est la différence (ou le rapport) des variances qui peut rendre le jeu de protéines discriminant.

Cette analyse est préliminaire et les résultats sont en cours d'exploration plus avancée, notamment dans le cas multidimensionnel et corrélé.

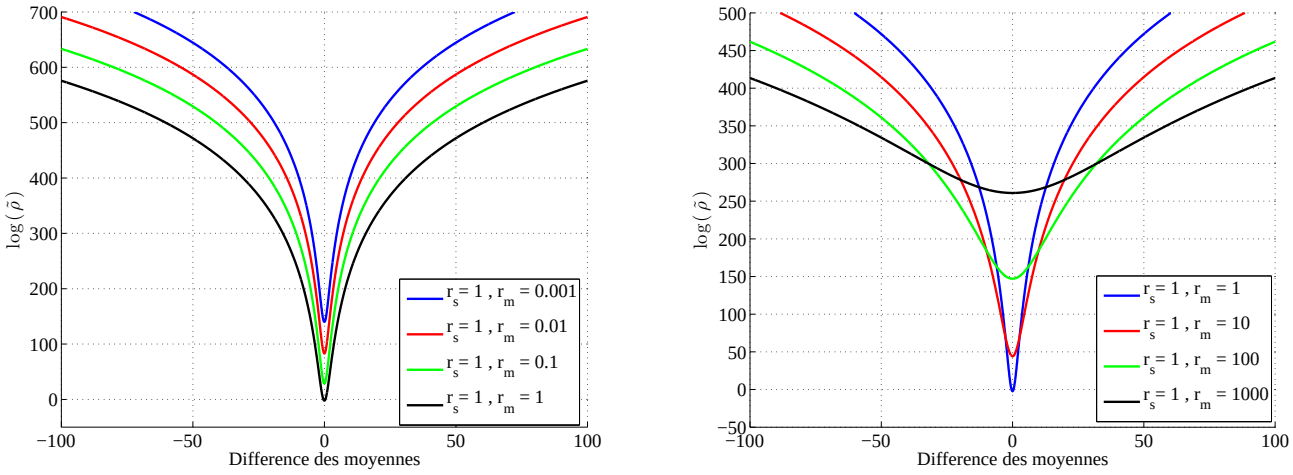

FIG. 1 – Évolution de la log-cote (à un terme additif près) avec la différence des moyennes. $N_{\mathsf{m}} = 50$ et $N_{\mathsf{s}} = 100$.

Annexe : trois résultats utiles

Densité Gauss-Wishart — Cette densité pour le couple $(\boldsymbol{m}, \boldsymbol{\Gamma})$ est paramétrée par quatre paramètres $\nu, \eta, \boldsymbol{\mu}, \boldsymbol{\Lambda}$:

$$\mathcal{NW}(\boldsymbol{m}, \boldsymbol{\Gamma}\,;\, \nu, \eta, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = K_{\mathcal{NW}}^{-1}\ \det\big[\boldsymbol{\Gamma}\big]^{(\nu-P)/2} \exp\big[-\big(\mathrm{tr}\big[\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}\big] + \eta(\boldsymbol{m}-\boldsymbol{\mu})^{\mathrm{t}}\boldsymbol{\Gamma}(\boldsymbol{m}-\boldsymbol{\mu})\big)/2\big] \qquad (6)$$

et sa constante de normalisation s'écrit : $K_{\mathcal{NW}}(\nu, \eta, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = (2\pi)^{P/2}\, 2^{\nu P/2}\, \eta^{-P/2}\ \det\big[\boldsymbol{\Lambda}\big]^{\nu/2}\, \Gamma_P(\nu/2)$ où Γ_P.

Paramètres a posteriori — Les paramètres a posteriori s'écrivent simplement

$$\nu_{\mathsf{b}}^{\mathrm{pst}} = N_{\mathsf{b}}\ , \quad \eta_{\mathsf{b}}^{\mathrm{pst}} = N_{\mathsf{b}}\ , \quad \boldsymbol{\mu}_{\mathsf{b}}^{\mathrm{pst}} = \bar{\boldsymbol{x}}_{\mathsf{b}}\ , \quad \big(\boldsymbol{\Lambda}_{\mathsf{b}}^{\mathrm{pst}}\big)^{-1} = N_{\mathsf{b}}\bar{\boldsymbol{R}}_{\mathsf{b}} \qquad (7)$$

pour $\mathsf{b} = \mathsf{m}$, $\mathsf{b} = \mathsf{s}$ et $\mathsf{b} = \mathsf{c}$, dans le cas d'un a priori diffus (sans inclure d'information fiable sur les paramètres).

Moyennes et variances empiriques — On voit apparaître deux quantités attendues : moyenne et covariance empiriques de chaque groupe $\mathsf{b} = \mathsf{m}$, $\mathsf{b} = \mathsf{s}$ et $\mathsf{b} = \mathsf{c}$:

$$\bar{\boldsymbol{x}}_{\mathsf{b}} = \frac{1}{N_{\mathsf{b}}} \sum_{n \in \mathcal{I}_{\mathsf{b}}} \boldsymbol{x}_n \quad \text{et} \quad \bar{\boldsymbol{R}}_{\mathsf{b}} = \frac{1}{N_{\mathsf{b}}} \sum_{n \in \mathcal{I}_{\mathsf{b}}} (\boldsymbol{x}_n - \bar{\boldsymbol{x}}_{\mathsf{b}})(\boldsymbol{x}_n - \bar{\boldsymbol{x}}_{\mathsf{b}})^{\mathrm{t}}\,. \qquad (8)$$

On note $\bar{p} = N_{\mathsf{m}}/N$ et $1 - \bar{p} = N_{\mathsf{s}}/N$ les proportions observées d'individus malades et sains.

Références

[1] C. P. Robert, *The Bayesian Choice. From decision-theoretic foundations to computational implementation*, Springer Texts in Statistics. Springer Verlag, New York, NY, USA, 2007.

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

[3] P. Szacherski, J.-F. Giovannelli, L. Gerfault, A. Giremus et P. Grangeat, « Robust MS serum sample classification in proteomics by the use of inverse problems », in *2012 IEEE International Workshop on Genomic Signal Processing and Statistics*, Washington, DC, USA, décembre 2012.