HW 3 PGM

Arthur Lavergne et Tamim El Ahmad

14 janvier 2020

Question 1

Cette opération permet de ramener toutes les colonnes à la même échelle. Ceci est utile dans notre cas car on suppose que notre prior sur β est de la forme $\mathcal{N}(0, \tau I_p)$, c'est à dire que la matrice de variance- covariance est isentropique.

Question 2

Supposons qu'il existe $\sigma > 0$ tel que $\epsilon_i \hookrightarrow \mathcal{N}(0, \sigma^2)$. Dans ce cas on aurait $\epsilon'_i = \frac{\epsilon_i}{\sigma} \hookrightarrow \mathcal{N}(0, 1)$. Donc :

$$\begin{aligned} y_i &= sgn(\beta^T x_i + \epsilon_i) \\ &= sgn(\beta^T x_i + \sigma\epsilon'_i) \\ &= sgn(\sigma(\frac{1}{\sigma}\beta^T x_i + \epsilon'_i)) \\ y_i &= sgn(\beta'^T x_i + \epsilon'_i) \end{aligned}$$

Donc on peut se ramener au cas précédent avec $\beta' = \frac{1}{\sigma}\beta$. On peut donc supposer $\epsilon_i \hookrightarrow \mathcal{N}(0, 1)$ sans perte de généralité.

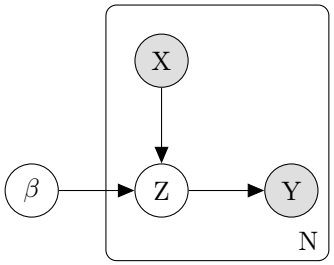
Question 3

On commence par dessiner le modèle graphique de notre modèle pour comprendre les relations entre les variables aléatoires :

De plus, on a les lois suivantes pour les variables :

— $\beta \hookrightarrow \mathcal{N}(0, \tau I_p)$

— $Z_i = \beta^T x_i + \epsilon_i$ avec $\epsilon_i \hookrightarrow \mathcal{N}(0, 1)$



— $y_i = sgn(z_i)$

Afin de construire l'algortihme de Gibbs, on a besoin des probabilité suivantes : p(β|Z) et p(Z|β, y). Or, on a :

$$\begin{aligned} p(\beta|Z) \propto p(\beta)p(Z|\beta) &= \exp(-\frac{1}{2\tau}|\beta\|^2)\exp(-\frac{1}{2}\sum_{i=1}^{n}(Z_i - \beta^T x_i)^2) \\ &= \exp(-\frac{1}{2\tau}|\beta\|^2)\exp(-\frac{1}{2}\|Z - X\beta\|^2) \end{aligned}$$

de plus :

$$p(Z|\beta, y) \propto p(Z|\beta)p(y|Z,\beta) = \exp(-\frac{1}{2}\|Z - X\beta\|^2)\prod_{i=1}^{n} p(y_i|\beta, Z_i)$$

$$\text{or on sait que :} \quad p(y_i|Z_i) = \mathbb{1}_{y_i=1}\mathbb{1}_{Z_i>0} + \mathbb{1}_{y_i=-1}\mathbb{1}_{Z_i\leq 0}$$
$$= \mathbb{1}_{y_i Z_i>0}$$

donc : $\quad p(Z|\beta, y) \propto \exp(-\frac{1}{2}\|Z - X\beta\|^2)\prod_{i=1}^{n}\mathbb{1}_{y_i Z_i>0}$

On a :

$$\begin{aligned} -\frac{1}{2\tau}\|\beta\|^2 - \frac{1}{2}\|Z - X\beta\|^2 &= -\frac{1}{2}[\frac{\beta^T\beta}{\tau} + Z^TZ - Z^TX\beta - \beta^TX^TZ + \beta^TX^TX\beta] \\ &= -\frac{1}{2}[\beta^T(\frac{I_p}{\tau} + X^TX)\beta + Z^TZ - Z^TX\beta - \beta^TX^TZ] \end{aligned}$$

Or, on sait que :

$$-\frac{1}{2}(\beta - \mu)^T\Sigma^{-1}(\beta - \mu) = -\frac{1}{2}[\beta^T\Sigma^{-1}\beta - \beta^T\Sigma^{-1}\mu - \mu^T\Sigma^{-1}\beta + \mu^T\Sigma^{-1}\mu]$$

Donc, par identification, il vient que :

$$\boxed{\Sigma^{-1} = \frac{I_p}{\tau} + X^TX} \quad \boxed{\mu = \Sigma X^TZ} \quad \text{et} \quad \beta|z \hookrightarrow \mathcal{N}(\mu, \Sigma)$$

Enfin, on a $Z_i|\beta$ qui suit une loi normale tronquée, $T\mathcal{N}(x_i^T\beta, 1; y_i)$ avec comme support $\{Z_i \in \mathbb{R}, Z_i x_i > 0\}$. Après implémentation de l'algorithme et pour 1000 simulation, on peut observer les loi marginales en représentant les les histogrammes de chaque paramètres β :

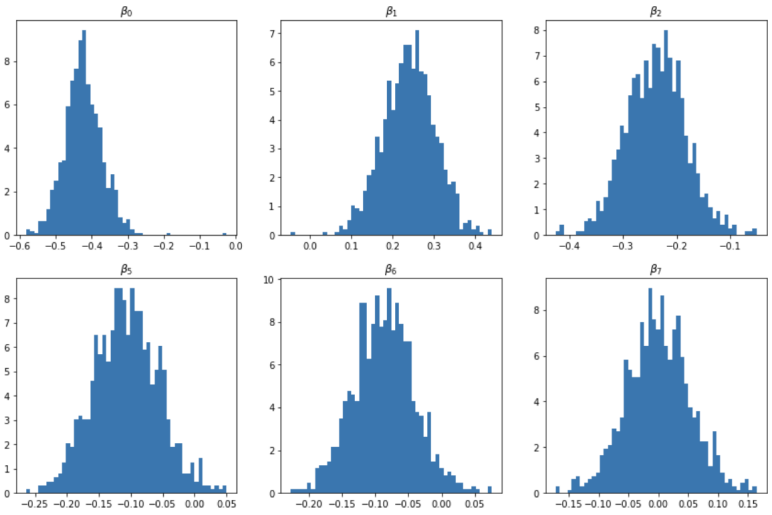


FIGURE 1 – Distributions estimées de quelques uns des paramètres par l'algorithme de Gibbs

Question 4

On cherche à estimer p(β, Z|X, y) par une distribution q(β, Z). Par le principe posée par la méthode mean-field, on suppose que q(β, Z) = q(β)q(Z). Il nous reste donc à chercher les facteurs optimaux pour trouver les expressions à chaque itération de l'algorithme :
$\underline{q^*(\beta)}$:

$$\begin{aligned}
\log(q^*(\beta)) &= \mathbb{E}_{q(Z)}[\log(p(y, Z, \beta|X)] + const \\
&= \underbrace{\mathbb{E}_{q(Z)}[\log(p(y|Z)]}_{\text{constante}} + \mathbb{E}_{q(z)}[\log(p(Z|X, \beta)] + \mathbb{E}_{q(z)}[\log(p(\beta))] + const \\
&= \mathbb{E}_{q(z)}[\log(\mathcal{N}(Z|X\beta, 1)] + \mathbb{E}_{q(z)}[\log(\mathcal{N}(0, \tau I_p)] + const \\
&= -\frac{1}{2}\mathbb{E}_{q(Z)}[\|Z - X\beta\|^2|\beta, y] - \frac{1}{2}\|\beta\|^2 + const \\
&= -\frac{1}{2}\mathbb{E}_{q(Z)}[(Z - X\beta)^T(Z - X\beta)|\beta, y] - \frac{1}{2\tau}\|\beta\|^2 + const \\
&= -\frac{1}{2}(\underbrace{\mathbb{E}_{q(Z)}[ZZ^T|\beta, y]}_{\text{constante}} - \mathbb{E}_{q(Z)}[Z^TX\beta - \beta^TX^TZ + \beta^TX^TX\beta|\beta, y]) - \frac{1}{2\tau}\|\beta\|^2 + const \\
&= Z^TX\beta - \frac{1}{2}\beta^TX^TX\beta - \frac{1}{2\tau}\|\beta\|^2 + const
\end{aligned}$$

Par identification comme dans la question 3, on trouve que :

$$\overline{Z} = \mathbb{E}_{q^*(Z)}[z] \quad q^*(\beta) = \mathcal{N}(\Sigma X^T\overline{Z}, \Sigma) \quad \text{avec } \Sigma \text{ définit comme en question 3}$$

$\underline{q^*(Z)}$:

$$\begin{aligned}
\log(q^*(Z_i)) &= \mathbb{E}_{q(\beta)}[\log(p(y_i|Z_i, \beta, x_i)] + const \\
&= \log(p(y_i|z_i)) + \mathbb{E}_{q(\beta)}[\log(p(Z_i|x_i, \beta)] + \underbrace{\mathbb{E}_{q(\beta))}[\log(p(\beta))]}_{constante} + const \\
&= \log(p(y_i|Z_i)) + \mathbb{E}_{q(\beta)}[\log(\mathcal{N}(Z_i|\beta^Tx_i, 1)] + const \\
&= \mathbb{1}_{y_i=1}\log(\mathbb{1}_{Z_i>0}) + \mathbb{1}_{y_i=-1}\log(\mathbb{1}_{Z_i\leq 0}) - \frac{1}{2}\mathbb{E}_{q(\beta))}[(Z_i - \beta^Tx_i)^2] + const
\end{aligned}$$

Comme les Z_i sont indépendantes, il vient au final :

$$\log(q^*(Z)) = \sum_{i=1}^{n} \mathbb{1}_{y_i=1}\log(\mathbb{1}_{Z_i>0}) + \mathbb{1}_{y_i=-1}\log(\mathbb{1}_{Z_i\leq 0}) - \frac{1}{2}\|Z - X\overline{\beta}\|^2 + const$$

Ainsi, on a $\overline{\beta} = \mathbb{E}_{q(\beta)}$ et $q(z) = T\mathcal{N}(X\overline{\beta}, I_p, \mathcal{P}_y)$. Une fois ces calculs fait, l'algorithme consiste à calculer de manière itérative jusqu'à convergence :

$$
\begin{aligned}
q(\beta) &= \mathcal{N}(\Sigma X^T \overline{Z}, \Sigma) \\
q(Z) &= \mathcal{N}(X\overline{\beta}, I_p, \mathcal{P}_y) \\
\overline{\beta} &= \Sigma X^T \overline{Z} \\
\forall i, Z_i &= x_i^T \beta + y_i \frac{pdf(x_i^T \overline{\beta})}{cdf(y_i x_i^T \overline{\beta})}
\end{aligned}
$$

Par cet algorithme, on obtient les distributions suivantes les paramètres β :

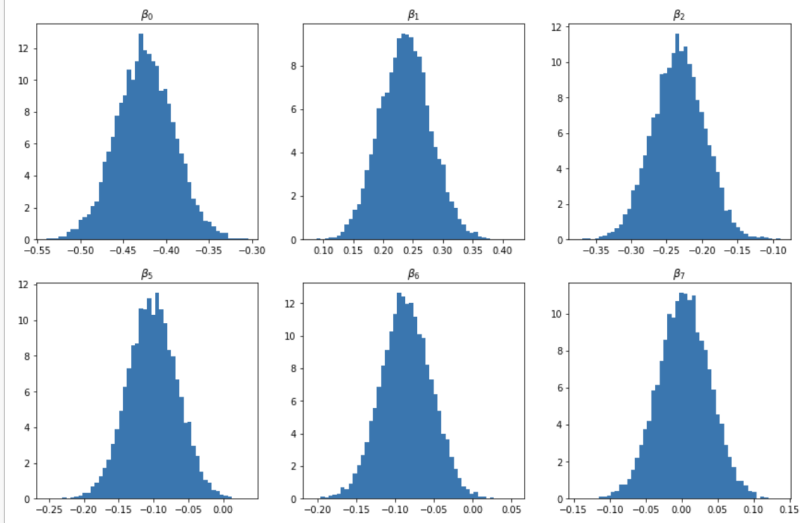


FIGURE 2 – Distributions estimées de quelques paramètres par l'algorithme de mean-field

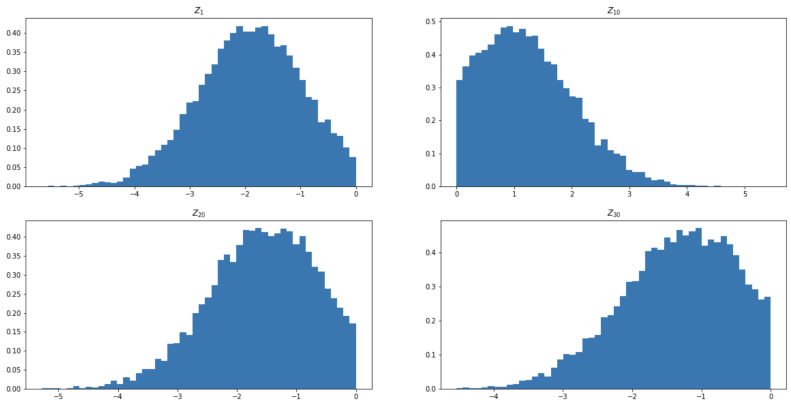


FIGURE 3 – Distribution des Z avec l'algorithme mean-field

Afin de comparer les deux algorithme en terme de performances, aussi de de prédiction que de complexité en temps, on cherche à savoir à partir de combien d'itération l'algorithme mean-field converge :

Ainsi, on peut déterminer dans un premier temps le temps mis par les deux

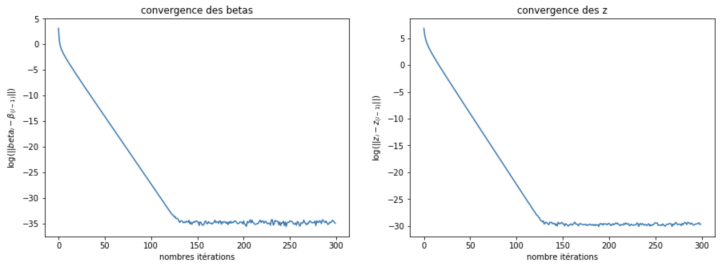


FIGURE 4 – Etude du nombre d'itérations nécessaire pour voir convergence de l'algorithme mean-field

algorithmes pour réaliser 1000 échantillons selon la distribution des paramètres :

— Gibbs : 3.25

— mean-field : 0.26

L'algorithme de mean-field est donc bien plus rapide en terme de complexité en temps. De plus, afin de comparer la performance des deux algorithmes en terme de prédiction, on séparer le dataset initial en deux partie (environ 2/3 pour l'entraînement et 1/3 pour le test). Dans ce le cas de Gibbs nous obtenons une précision de 0.7076 et dans le cas de mean-field nous obtenons une performance de 0.7021. Les deux algorithmes arrivent donc à produire des résultats assez similaires en termes de prédiction.

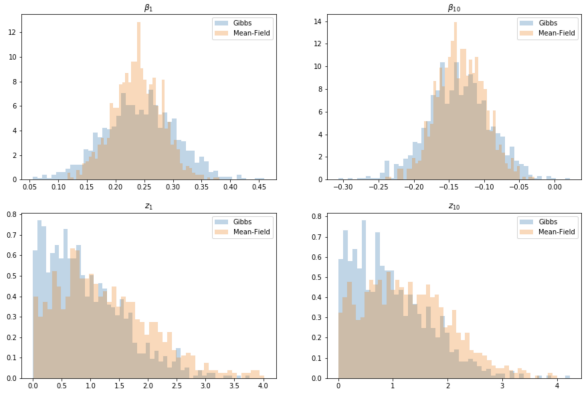


FIGURE 5 – Superposition des distributions obtenues vie les deux algorithmes

Question 6

Dans le cas d'une séparation complète des données, on a par exemple la disposition suivante :

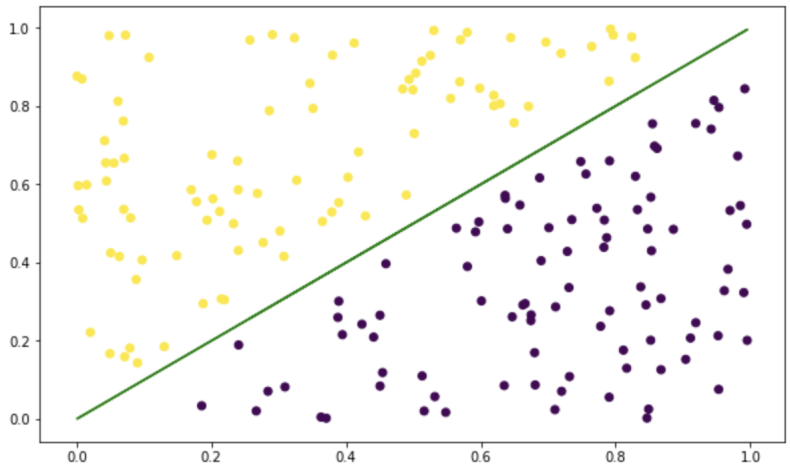


FIGURE 6 – Visualisation de données avec séparation complète

Dans ce cas ci, la vraisemblance n'existe pas, et il n'est donc pas possible d'appliquer les algorithmes qui reposent sur une maximisation de la vraisemblance pour trouver les paramètres de la distribution.

On applique donc l'algorithme de Gibbs à nos données générées comme ci-dessus. On observant dans un premier temps quelques chaîne de Markov créées par l'algorithme, on observe rapidement que celle-ci n'ont pas convergé.

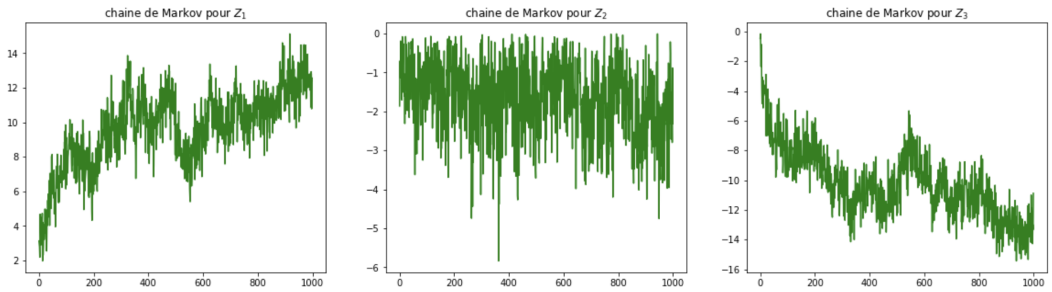


FIGURE 7 – Non convergence de l'algorithme de Gibbs

De plus, si on regarde la distribution de quelques paramètres générés par l'algo-

rithme de Gibbs ont remarque qu'ils n'ont pas la distribution attendue :

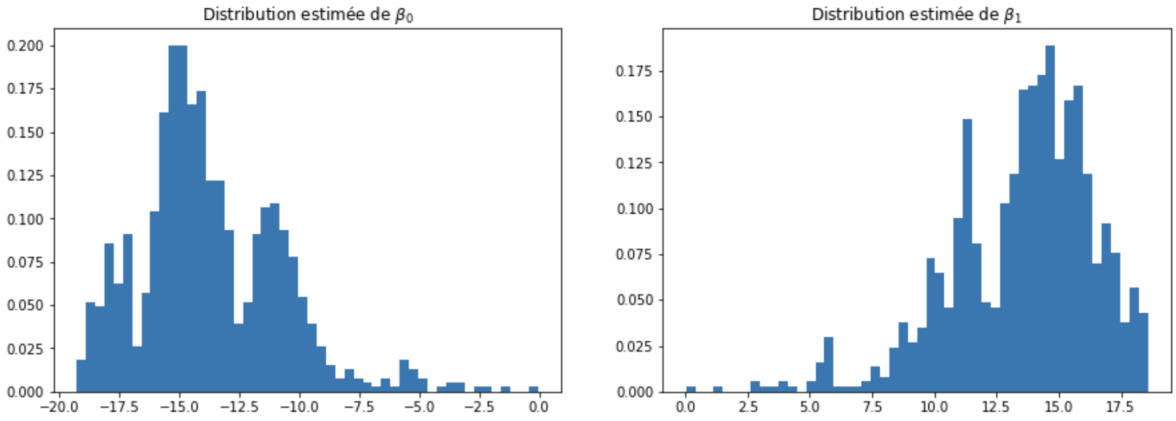


FIGURE 8 – L'algorithme de Gibbs ne permet pas une estimation correcte de la distribution des paramètres