Oxford : Smooth fit to log-odds ratios

Vincent Guitteny, Freddie Joly, Tom Léchappé et Elyes Zribi

08 avril 2022

Table des matières

1 Présentation du jeu de données

L'objectif de cette étude est d'observer l'influence que peut avoir l'exposition d'une mère aux rayons X pendant sa grossesse sur les décés des enfants par un cancer (de 0 à 9 ans). Les données sont séparées en 120 jeux de données en fonction de l'âge de l'enfant de son année de naissance (1944-1964). Dans notre cas, seule l'influence de l'année de naissance est retenue. Ces valeurs varient de -10 à 10 en prenant en compte que 0 correspond à l'année 1954. Pour chacune des 120 partitions, on dispose du nombre de décés r parmi une population de n enfants pour deux groupes :

- un groupe d'enfants dont les mères ont été exposées à des rayons X au cours de leur grossesse
- un groupe "témoin" dont les mères n'ont pas été exposées.

2 Présentation du modèle

Regardons maintenant le modèle proposé par les auteurs. Pour chaque strate i et pour les deux groupes de population, le nombre de décés est donné par une loi binomiale :

- $r_i^0 \sim Bin(n_i^0, p_i^0)$
- $r_i^1 \sim Bin(n_i^1, p_i^1)$.

Ainsi les paramètres p_i^0 et p_i^1 sont donnés par les deux formules suivantes :

- $p_i^0 = sigmoid(\mu_i)$
- $p_i^1 = sigmoid(\mu_i + \alpha + \beta_1 year_i + \beta_2(year_i^2 - 22) + b_i)$ ou bien $p_i^1 = sigmoid(\mu_i + log\psi_i)$ où $log\psi_i = \alpha + \beta_1 year_i + \beta_2(year_i^2 - 22) + b_i$.

Le paramètre b_i suit une loi normale $\mathcal{N}(0, \sigma^2)$.

Il reste donc les lois "non informatives" des paramètres suivants :

- $\alpha \sim \mathcal{N}(0, \sigma_\alpha^2)$
- $\beta_1 \sim \mathcal{N}(0, \sigma_{\beta_1}^2)$
- $\beta_2 \sim \mathcal{N}(0, \sigma_{\beta_2}^2)$
- $\sigma^2 \sim InvGamma(a_\sigma, b_\sigma)$
- $\mu_i \sim \mathcal{N}(0, \sigma_\mu^2)$.

On représente le modèle hiérarchique sous la forme d'un graphe ci-dessous.

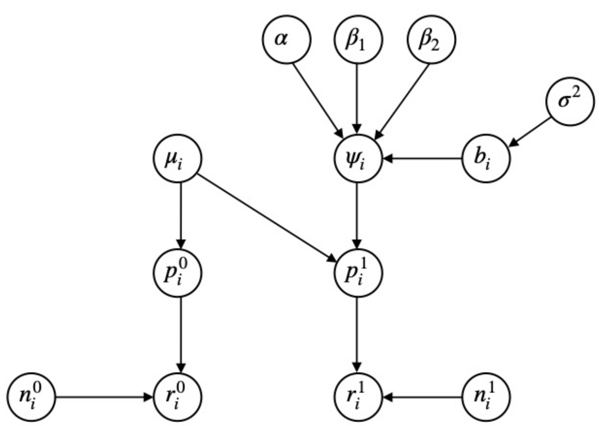

Figure 1: Modèle hierarchique

3 Calculs des lois conditionnelles pleines

On veut maintenant calculer les lois conditionnelles pleines de nos paramètres afin de mettre en oeuvre un échantilloneur de Gibbs sur ce modèle. Bien que la loi conditionnelle pleine de σ^2 soit identifiable, celles des autres paramètres ne sont pas identifiables. On donne ici un exemple de calcul pour le paramètre σ^2, les autres calculs sont donnés en annexe. Les noeuds de notre modèle représentés par $log(\psi_i)$, p_i^0 et p_i^1 étant non stochastiques dans notre modèle, on peut les ignorer lors du calcul des lois conditionnelles pleines.

3.1 Calcul pour σ^2

On calcule la loi conditionnelle pleine du paramètre σ^2.

$$\begin{aligned}\Pi(\sigma^2|...) &\propto \Pi(\sigma^2)\prod_{i=1}^{n}\Pi(b_i|\sigma^2)\\ &\propto \exp\left(\frac{-b_\sigma}{\sigma^2}\right)(\sigma^2)^{-a_\sigma-1}\prod_{i=1}^{n}\exp\left(\frac{-b_i^2}{2\sigma^2}\right)(\sigma^2)^{-\frac{1}{2}}\\ &\propto (\sigma^2)^{-a_\sigma-1-\frac{n}{2}}\exp\left(\frac{-2b_\sigma-\sum\limits_{i=1}^{n}b_i^2}{2\sigma^2}\right)\end{aligned}$$

Ainsi, on identifie une loi inverse gamma :

$$\Pi(\sigma^2|...) \sim InvGamma(\frac{n}{2}+a_\sigma, \frac{2b_\sigma+\sum\limits_{i=1}^{n}b_i^2}{2})$$

3.2 Exemple de calcul pour une loi non identifiable

On calcule la loi conditionnelle pleine du paramètre α.

$$\begin{aligned}\Pi(\alpha|...) &\propto \Pi(\alpha)\prod_{i=1}^{n}\Pi(r_i^1|...)\\ &\propto \exp\left(\frac{-\alpha^2}{2\sigma_\alpha^2}\right)\prod_{i=1}^{n}\left((p_i^1)^{r_i^1}(1-p_i^1)^{n_i^1-r_i^1}\right)\end{aligned}$$

où $p_i^0 = sigmoid(\mu_i)$ et $p_i^1 = sigmoid(\mu_i+\alpha+\beta_1 year_i+\beta_2(year_i^2-22)+b_i)$.

4 Implémentation et résultats

Maintenant que nous avons calculer toutes nos lois conditionnelles pleines, nous pouvons utiliser un échantilloneur de Gibbs afin de comparer nos résultats avec ceux proposés.

Afin de comparer nos résultats avec ceux de l'expérience, nous avons utilisé un “burnin” de 1000 sur la chaîne de Markov de taille totale 10 000 qui est retourné par notre fonction simulant un échantilloneur de Gibbs.

On retrouve des résultats similaires à ceux proposés comme on peut l'observer dans le tableau ci-dessous.

	Moyenne
α	0.588
β_1	-0.046
β_2	0.008
σ^2	0.129

Table 1 : Moyenne des paramètres α, β_1, β_2 et σ^2.

L'ensemble de notre code est disponible dans le fichier Gibbs.R sur la page Github : https://github.com/FreddieJoly/Oxford-Smooth-fit-to-log-odds-ratios

5 Annexe

5.1 Loi conditionnelle pleine du paramètre β_1

$$\begin{aligned}\Pi(\beta_1|...) &\propto \Pi(\beta_1)\prod_{i=1}^{n}\Pi(r_i^1|...)\\ &\propto \exp\left(\frac{-\beta_1^2}{2\sigma_{\beta_1}^2}\right)\prod_{i=1}^{n}\Big((p_i^1)^{r_i^1}(1-p_i^1)^{n_i^1-r_i^1}\Big)\end{aligned}$$

5.2 Loi conditionnelle pleine du paramètre β_2

$$\begin{aligned}\Pi(\beta_2|...) &\propto \Pi(\beta_2)\prod_{i=1}^{n}\Pi(r_i^1|...)\\ &\propto \exp\left(\frac{-\beta_2^2}{2\sigma_{\beta_2}^2}\right)\prod_{i=1}^{n}\Big((p_i^1)^{r_i^1}(1-p_i^1)^{n_i^1-r_i^1}\Big)\end{aligned}$$

5.3 Loi conditionnelle pleine du paramètre b_i

$$\begin{aligned}\Pi(b_i|...) &\propto \Pi(b_i|\sigma^2)\Pi(r_i^1|...)\\ &\propto \exp\left(\frac{-b_i^2}{2\sigma^2}\right)\Big((p_i^1)^{r_i^1}(1-p_i^1)^{n_i^1-r_i^1}\Big)\end{aligned}$$

5.4 Loi conditionnelle pleine du paramètre μ_i

$$\begin{aligned}\Pi(\mu_i|...) &\propto \Pi(\mu_i)\Pi(r_i^0|\mu_i)\Pi(r_i^1|...)\\ &\propto \exp\left(\frac{-\mu_i^2}{2\sigma_\mu^2}\right)\Big((p_i^0)^{r_i^0}(1-p_i^0)^{n_i^0-r_i^0}\Big)\Big((p_i^1)^{r_i^1}(1-p_i^1)^{n_i^1-r_i^1}\Big)\end{aligned}$$