



UNIVERSITÉ DE NANTES

Oxford : smooth fit to log-odds ratios

Alexandre DAVID, Achille GUILLEMOT, Kilian BARLIER, Guillaume MORVAN

Avril 2021

1 Introduction

L'exposition aux rayons X à des effets plus ou moins important selon plusieurs paramètres mais il peut-être difficile de discerner si certains effets ou certaines pathologie sont du à cette exposition aux rayons X. C'est dans ce sens que cet étude est mené, afin de mettre en évidence ou non l'effet des rayons X sur la mort de jeunes enfants. Ainsi on aimerait pouvoir quantifié cet effet (si il existe) afin de mettre en évidence la dangerosité des rayons X sur des personnes dites sensibles (enfants, personnes âgées ou en situation de handicap...)

Liens vers le github pour le code R :

https://github.com/NonaryAce/School_Project_DAVID_BARLIER_GUILLEMOT_MORVAN

2 Présentation des données

Les données d'Oxford s'intéressent à l'effet de l'exposition aux rayons X de la mère sur le risque de mort des enfants d'un cancer. Le jeu de donnée est découpé en deux parties, une partie contenant les enfants dont les mères ont été exposées aux rayons X et un groupe de contrôle. Ces enfants sont classés par âge et par date de naissance.

Nous avons les données suivantes :

- n_0 : taille de l'échantillon témoin.
- n_1 : taille de l'échantillon test.
- r_0 : nombre de cas de mort de cancer dans l'échantillon témoin.
- r_1 : nombre de cas de mort de cancer dans l'échantillon test.
- year : codé de -10 à 10 par rapport à l'année 1954 (1944-1964).

3 DAG

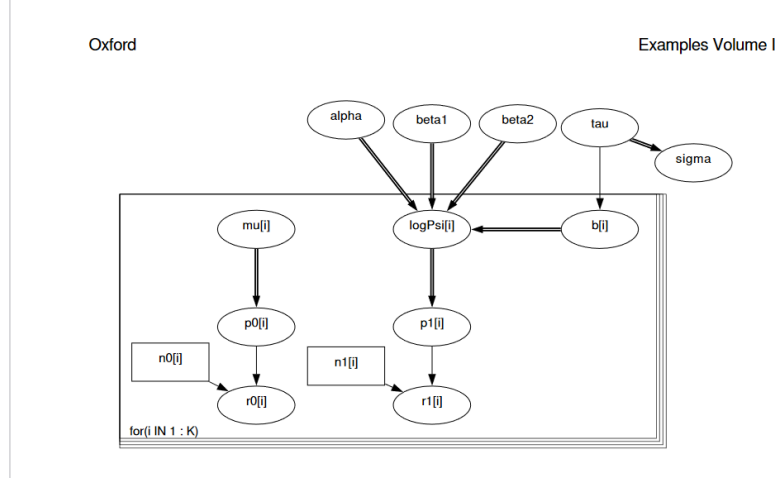


FIGURE 1 – DAG

3.1 Modèle

Le modèle proposé par le document de référence est le suivant ($i = 1, \dots, K$) :

- $\alpha, \beta_1, \beta_2 \sim \mathcal{N}(0, 1/10^{-6})$ (précision = 10^{-6})
- $\tau \sim \text{Gamma}(10^{-3}, 10^{-3})$ (paramétrisation de type (α, β))
- $b_i \sim \mathcal{N}(0, 1/\tau)$ (précision = τ)
- $\mu_i \sim \mathcal{N}(0, 1/10^{-6})$ (précision = 10^{-6})

Variables latentes / transitoires :

- $\log \Psi_i = \alpha + \beta_1 \text{year}_i + \beta_2 (\text{year}_i^2 - 22) + b_i$
- $\text{logit}(p_i^0) = \mu_i$
- $\text{logit}(p_i^1) = \mu_i + \log \Psi_i$
- $\sigma = 1/\sqrt{\tau}$

Loi des observations dont on souhaite estimer les paramètres p_i^0 et p_i^1 :

- $r_i^0 \sim \text{Binom}(p_i^0, n_i^0)$
- $r_i^1 \sim \text{Binom}(p_i^1, n_i^1)$

Remarque. Les lois normales sont de la forme $\mathcal{N}(\mathbb{E}(X), \mathbb{V}(X))$ et on définit la précision = $1/\mathbb{V}(X)$. Il avait été également remarqué que les risques dépendaient des années mais pas de l'âge, ce qui explique que dans $\log \Psi_i$ il y a l'année mais pas l'âge.

3.2 Calcul des lois conditionnelles pleines

Forme explicite des p_i :

$$\text{logit}(p_i^0) = \mu_i \iff p_i^0 = \frac{1}{1 + \exp(-\mu_i)}$$

et

$$\text{logit}(p_i^1) = \mu_i + \log \Psi_i \iff p_i^1 = \frac{1}{1 + \exp(-\mu_i - \log \Psi_i)}$$

Lois pleines de α , β_1 et β_2 :

$$\begin{aligned} \pi(\alpha|\dots) &\propto \pi(\alpha) \prod_{i=1}^K \pi(r_i^1|\alpha, \beta_1, \beta_2, b_i) \\ &\propto \pi(\alpha) \prod_{i=1}^K \pi(r_i^1|p_i^1) \\ &\propto \frac{1}{\sqrt{2\pi \frac{1}{10^{-6}}}} \exp\left(-\frac{1}{2} \frac{\alpha^2}{\frac{1}{10^{-6}}}\right) \prod_{i=1}^K \binom{n_i^1}{r_i^1} p_i^{1r_i^1} (1 - p_i^1)^{n_i^1 - r_i^1} \\ &\propto \exp\left(-\frac{1}{2} 10^{-6} \alpha^2\right) \prod_{i=1}^K p_i^{1r_i^1} (1 - p_i^1)^{n_i^1 - r_i^1} \end{aligned}$$

Loi pleine de τ : (lois conjuguées, calcul pédagogique)

$$\begin{aligned} \pi(\tau|\dots) &\propto \pi(\tau) \prod_{i=1}^K \pi(b_i|\tau) \\ &\propto \tau^{10^{-3}-1} \frac{(10^{-3})^{10^{-3}} \exp(-10^{-3}\tau)}{\Gamma(10^{-3})} \prod_{i=1}^K \frac{1}{\sqrt{2\pi \frac{1}{\tau}}} \exp\left(-\frac{1}{2} \frac{b_i^2}{\frac{1}{\tau}}\right) \\ &\propto \tau^{10^{-3}-1} \exp(-10^{-3}\tau) \prod_{i=1}^K \sqrt{\tau} \exp\left(-\frac{1}{2} \tau b_i^2\right) \\ &\propto \tau^{10^{-3}-1} \exp(-10^{-3}\tau) \sqrt{\tau}^K \exp\left(-\frac{1}{2} \tau \sum_{i=1}^K b_i^2\right) \\ &\propto \tau^{10^{-3}-1+\frac{K}{2}} \exp\left(-10^{-3}\tau - \frac{1}{2} \tau \sum_{i=1}^K b_i^2\right) \\ &\propto \tau^{10^{-3}+\frac{K}{2}-1} \exp\left(-\tau\left(10^{-3} + \frac{1}{2} \sum_{i=1}^K b_i^2\right)\right) \end{aligned}$$

On a donc finalement :

$$\pi(\tau|\dots) \sim \text{Gamma}(10^{-3} + \frac{K}{2}, 10^{-3} + \frac{1}{2} \sum_{i=1}^K b_i^2)$$

Loi pleine de b_i , pour $i = 1, \dots, K$:

$$\begin{aligned} \pi(b_i|\dots) &\propto \pi(b_i|\tau)\pi(r_i^1|p_i^1) \\ &\propto \frac{1}{\sqrt{2\pi\frac{1}{\tau}}} \exp(-\frac{1}{2}\frac{b_i^2}{\frac{1}{\tau}}) \binom{n_i^1}{r_i^1} p_i^{1r_i^1} (1-p_i^1)^{n_i^1-r_i^1} \\ &\propto \exp(-\frac{1}{2}\tau b_i^2) p_i^{1r_i^1} (1-p_i^1)^{n_i^1-r_i^1} \end{aligned}$$

Loi pleine de μ_i , pour $i = 1, \dots, K$: (groupe témoin et groupe test)

$$\begin{aligned} \pi(\mu_i|\dots) &\propto \pi(\mu_i)\pi(r_i|p_i) \\ &\propto \frac{1}{\sqrt{2\pi\frac{1}{10^{-6}}}} \exp(-\frac{1}{2}\frac{\mu_i^2}{\frac{1}{10^{-6}}}) \binom{n_i}{r_i} p_i^{r_i} (1-p_i)^{n_i-r_i} \\ &\propto \exp(-\frac{1}{2}10^{-6}\mu_i^2) p_i^{r_i} (1-p_i)^{n_i-r_i} \end{aligned}$$

3.3 Résultats

L'étude de référence nous donne les résultats suivants :

	α	β_1	β_2	σ
Valeur	0.566	-0.469	0.0071	0.15
(+/-)	0.070	0.0167	0.0033	0.10

Notre algorithme nous a permis d'obtenir les résultats suivants :

	α	β_1	β_2	σ
Valeur	0.322	-0.071	0.014	1.072
(+/-)	0.176	0.035	0.006	0.168

3.4 Interprétation

Comme énoncé dans l'introduction, afin de quantifier cet effet, une mesure souvent utilisée en médecine pour discerner l'effet d'un traitement ou autre, est d'utiliser l'odd ratio sur 2 populations : un groupe témoins et un groupe test. L'odd ratio à l'avantage d'être une quantité relativement simple à interpréter. Un odd ratio proche de 1 exprime un non-effet, et un odd ratio différent de 1 exprime une disparité du traitement ou de l'effet étudié sur les 2 groupes. Une étude préambule a montré que l'odd ratio dépend simplement de l'année de naissance de l'enfant et non de son âge. la prise en compte du paramètre b_i est là pour tenir compte des potentielles variables corrélées mais non disponibles lors de l'étude.

Ainsi en déterminant l'odd ratio, on pourra déterminer quelles années l'effet au rayons X à été le plus dévastateur sur les enfants.

```
> summary(odd_ratio[2:56])
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.4883  1.3017  2.0334  3.6488  4.0878 42.7991
> summary(odd_ratio[56:120])
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.5235  1.0007  1.4583  1.9870  2.3842 11.9391
```

FIGURE 2 – Odd Ratio Année 1947-1956,1957-1967

En regardant le odd ratio (on a pris le soin d'enlever les odd ratio trop grands afin de ne pas gêner et/ou fausser l'interprétation) pour chacune des strates (possible grâce à nos estimations), on remarque des grandes différences entre les années avant 1957 et après 1957. L'effet des rayons X avait beaucoup plus d'impact sur les enfants avant les années 1957 qu'après les années 1957. Cela peut s'expliquer par de nombreuses réformes mondiales concernant l'utilisation des rayon X sur les personnes enceintes.

4 Annexes

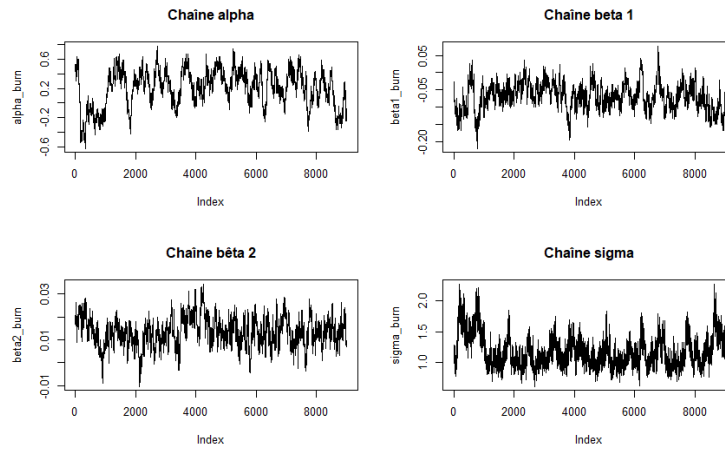


FIGURE 3 – Chaînes de nos 4 variables

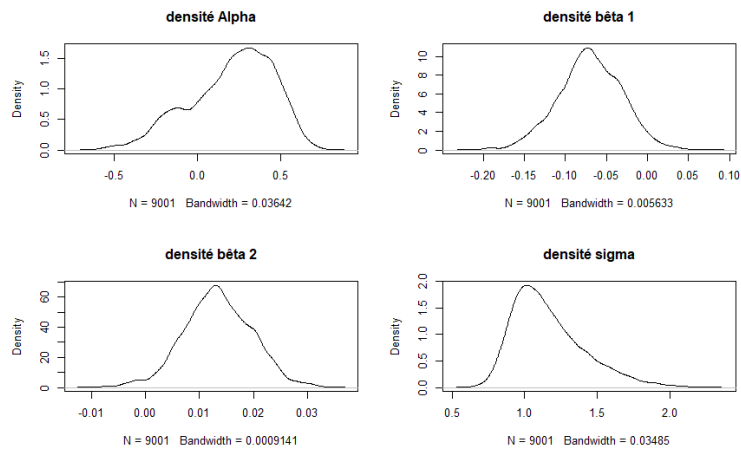


FIGURE 4 – Densité de nos 4 variables