

Oxford: smooth fit to log-odds ratios

Wilfrid NKODO, Antoine SIMON, Hayat EL TAHECH, Romain DAVID

2022-2023



Table des matières

1	Introduction	3
2	Modèle	3
2.1	Présentation du modèle	3
2.2	Lois conditionnelles	4
2.2.1	Loi conditionnelle pleine de μ_i	4
2.2.2	Loi conditionnelle pleine de α	4
2.2.3	Loi conditionnelle pleine de β_1	4
2.2.4	Loi conditionnelle pleine de β_2	4
2.2.5	Loi conditionnelle pleine de τ	4
2.2.6	Loi conditionnelle pleine de b_i	4
3	Interprétation & Analyse des résultats	5
4	Annexe	7

1 Introduction

On s'intéresse dans cette étude à l'effet que peut avoir l'exposition des femmes enceintes aux rayons X sur les décès des enfants par cancer infantile. Les données sur les enfants sont partitionnées en 120 combinaisons d'âge (0-9) et d'année de naissance (1944-1964). Dans notre modèle, seule l'année de naissance est prise en compte variant de -10 à 10 avec pour valeur 0 pour l'année 1954. On a pour chacune des partitions, le nombre de décès r parmi n enfants pour deux groupes : un groupe contenant les enfants dont leurs mères ont été exposées aux rayons X et un groupe que l'on peut qualifier de "témoin" est formé d'enfants dont les mères n'ont pas été exposées aux rayons X.

2 Modèle

2.1 Présentation du modèle

Le modèle proposé est le suivant : pour chaque strate i et pour les deux groupes (le groupe témoin et le groupe regroupant les cas d'enfants dont les mères ont été exposées durant leur grossesse à des rayons X), le nombre de décès suit une loi binomiale car celle-ci donne le nombre de succès d'une expérience, ici le nombre de décès. Plus précisément, on a :

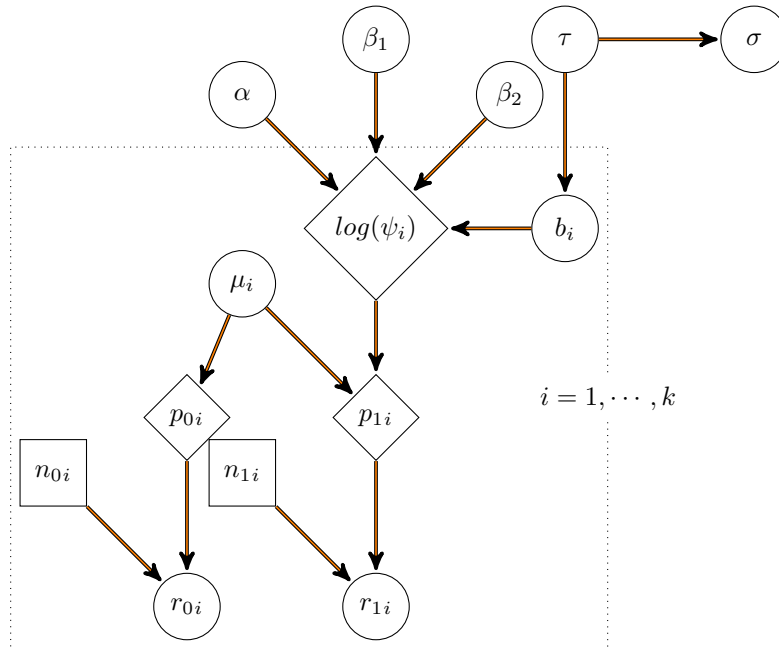
- $r_i^0 \sim \text{Binomiale}(p_i^0, n_i^0)$
- $r_i^1 \sim \text{Binomiale}(p_i^1, n_i^1)$

De plus, le paramètre p_i^0 est la probabilité de décès qui n'est pas due aux rayons X et p_i^1 est la probabilité de décès qui est due aux rayons X. Ces paramètres sont modélisés à l'aide la fonction logit, tel que :

- $\text{logit}(p_i^0) = \mu_i$
- $\text{logit}(p_i^1) = \mu_i + \log(\psi_i)$

et avec $\mu_i \sim \mathcal{N}(0, 10^6)$. On pose également $\log(\psi_i) = \alpha + \beta_1 \text{year}_i + \beta_2 (\text{year}_i^2 - 22) + b_i$. Le modèle inclut ainsi trois paramètres supplémentaires : α , β_1 et β_2 qui permettent de quantifier l'influence de l'année en cas d'expositions aux rayons X et tel que $\alpha, \beta_1, \beta_2 \sim \mathcal{N}(0, 10^6)$. On peut également interpréter le terme b_i intervenant dans le $\log(\psi_i)$ comme un effet aléatoire, ainsi on pose $b_i \sim \mathcal{N}(0, \frac{1}{\tau})$ où $\tau \sim \text{Gamma}(10^{-3}, 10^{-3})$.

Le DAG (Directed Acyclic Graph) est un modèle graphique qui présente la structure hiérarchique des paramètres. Pour notre projet, nous avons le DAG suivant :



1: Graphe Orienté Acyclique du modèle

2.2 Lois conditionnelles

2.2.1 Loi conditionnelle pleine de μ_i

$$\begin{aligned}\pi(\mu_i|\dots) &\propto \pi(\mu_i)\pi(r_j, p_j) \\ &\propto e^{-\frac{u_i^2}{2 \cdot (10^6)}} p_i^{r_i} (1 - p_i)^{n_i - r_i}\end{aligned}$$

2.2.2 Loi conditionnelle pleine de α

$$\begin{aligned}\pi(\alpha|\dots) &\propto \pi(\alpha) \prod_{i=1}^k \pi(r_{1i}|\beta_1, \beta_2, \tau, b_i) \\ &\propto e^{-\frac{\alpha^2}{2 \cdot (10^6)}} \prod_{i=1}^k p_{1i}^{r_{1i}} (1 - p_{1i})^{n_{1i} - r_{1i}}\end{aligned}$$

2.2.3 Loi conditionnelle pleine de β_1

$$\begin{aligned}\pi(\beta_1|\dots) &\propto \pi(\beta_1) \prod_{i=1}^k \pi(r_{1i}|\alpha, \beta_2, \tau, b_i) \\ &\propto e^{-\frac{\beta_1^2}{2 \cdot (10^6)}} \prod_{i=1}^k p_{1i}^{r_{1i}} (1 - p_{1i})^{n_{1i} - r_{1i}}\end{aligned}$$

2.2.4 Loi conditionnelle pleine de β_2

$$\begin{aligned}\pi(\beta_2|\dots) &\propto \pi(\beta_2) \prod_{i=1}^k \pi(r_{1i}|\alpha, \beta_1, \tau, b_i) \\ &\propto e^{-\frac{\beta_2^2}{2 \cdot (10^6)}} \prod_{i=1}^k p_{1i}^{r_{1i}} (1 - p_{1i})^{n_{1i} - r_{1i}}\end{aligned}$$

2.2.5 Loi conditionnelle pleine de τ

On pose $a = b = 10^{-3}$

$$\begin{aligned}\pi(\tau|\dots) &\propto \pi(\tau) \prod_{i=1}^k \pi(b_i|\tau) \\ &\propto \tau^{a-1} e^{b\tau} \prod_{i=1}^k \exp^{-\frac{\tau b_i^2}{2}} \frac{\sqrt{\tau}}{\sqrt{2\pi}} \\ &\propto \tau^{a-1} \left(\frac{\sqrt{\tau}}{\sqrt{2\pi}}\right)^n e^{-b\tau} e^{-\frac{\tau \sum_{i=1}^k b_i^2}{2}} \\ &\propto \tau^{a-1+\frac{k}{2}} e^{-b\tau - \frac{\tau \sum_{i=1}^k b_i^2}{2}} \\ &\propto \tau^{a-1+\frac{k}{2}} e^{-\tau(b + \frac{\sum_{i=1}^k b_i^2}{2})}\end{aligned}$$

$$\text{d'où } \tau \sim \Gamma(a + \frac{k}{2}, b + \frac{\sum_{i=1}^k b_i^2}{2}) \implies \sigma^2 \sim \Gamma^{-1}(a + \frac{k}{2}, b + \frac{\sum_{i=1}^k b_i^2}{2})$$

2.2.6 Loi conditionnelle pleine de b_i

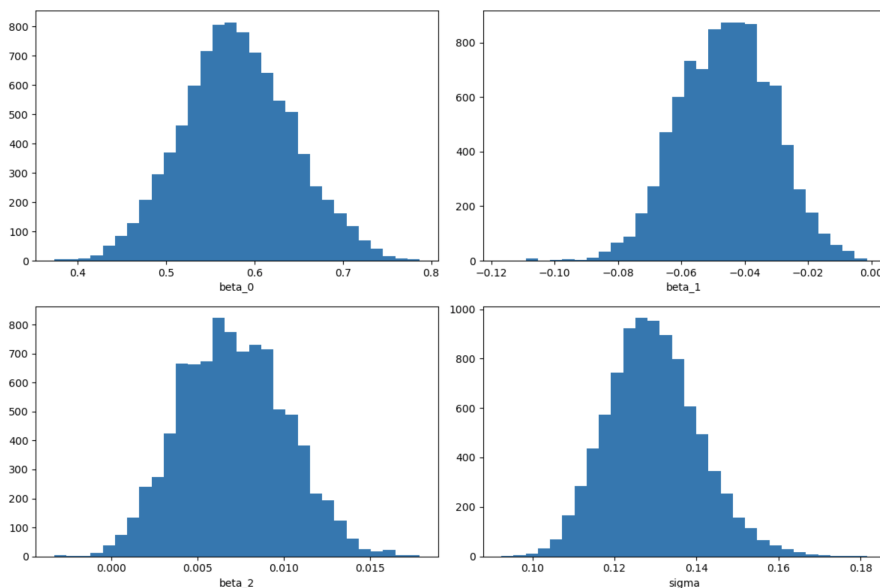
$$\begin{aligned}\pi(b_i|\dots) &\propto \pi(b_i|\tau) \prod_{i=1}^k \pi(r_{1i}|\alpha, \beta_1, \beta_2) \\ &\propto e^{-\frac{\tau b_i^2}{2}} p_{1i}^{r_{1i}} (1 - p_{1i})^{n_{1i} - r_{1i}}\end{aligned}$$

3 Interprétation & Analyse des résultats

Notre plan de design présente 2 groupes, le 1er composé d'enfants décédés de cancer dont les mères femmes furent exposées aux rayons X pendant leur grossesse et le second composé d'enfants décédés de cancer dont les mères ne furent pas exposées aux rayons X, ce second tient pour groupe de contrôle.

Une analyse en amont de notre plan d'expérience nous montre que ce dernier est équilibré avec $n_0 = n_1 = 120$ décès recensés. Le nombre décès r_0 observés dans le groupe de sujets non exposés présente une moyenne de 5.02 et une variance 8.82. Le nombre de décès dans le groupe de sujets exposés aux rayons présente une moyenne de 7.76 et une variance égale à 15.46. On remarque ainsi que dans le groupe exposé la moyenne de décès est plus importante traduisant un nombre de décès plus important possiblement en lien avec l'exposition aux rayons et que la variance/dispersion est particulièrement significative.

Nous mettons en oeuvre un échantillonneur MCMC Metropolis Hastings sur les paramètres α, β_1, β_2 et un échantillonneur de Gibbs pour le paramètre σ et nous observons les densités associées suivantes :



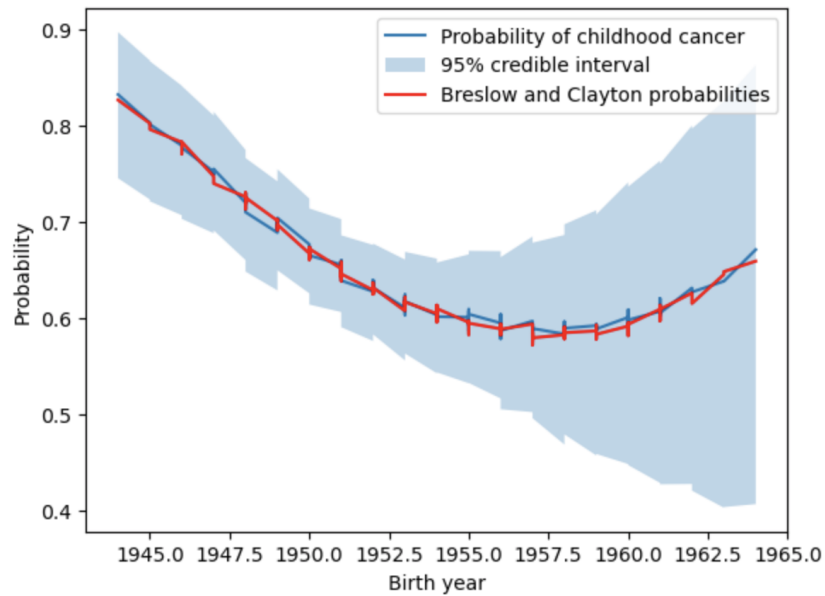
2: Densités de $\alpha, \beta_1, \beta_2, \sigma$

Les chaînes obtenues se traduisent par les densités ci-dessus. Ces densités s'apparentent à des lois normales centrée traduisant une mise en oeuvre adaptée de notre algorithme. Nous obtenons une estimation pour les différents paramètres :

	mean	sd	mc_error	val2.5pc	median	val97.5pc
alpha	0.580969	0.062071	0.000654	0.462322	0.578882	0.706779
beta_1	-0.045740	0.014848	0.000157	-0.073957	-0.045538	-0.017634
beta_2	0.007094	0.003040	0.000032	0.001446	0.007021	0.013106
sigma	0.129537	0.011202	0.000118	0.109365	0.128952	0.153365

On remarque que α, β_2 et σ ont une influence positive sur le nombre de décès parmi les sujets exposés tandis que β_1 a une influence négative.

Également une façon intéressante d'appréhender ces résultats est d'effectuer une comparaison avec les probabilités des Breslow & Clayton. En effet, notre jeu de données peut être assimilé à un tableau de contingence et l'application d'un log à la formule complexe $\log(\psi)$ regroupant nos paramètres d'intérêt traduit l'ajustement d'un lissage aux différentes strates.

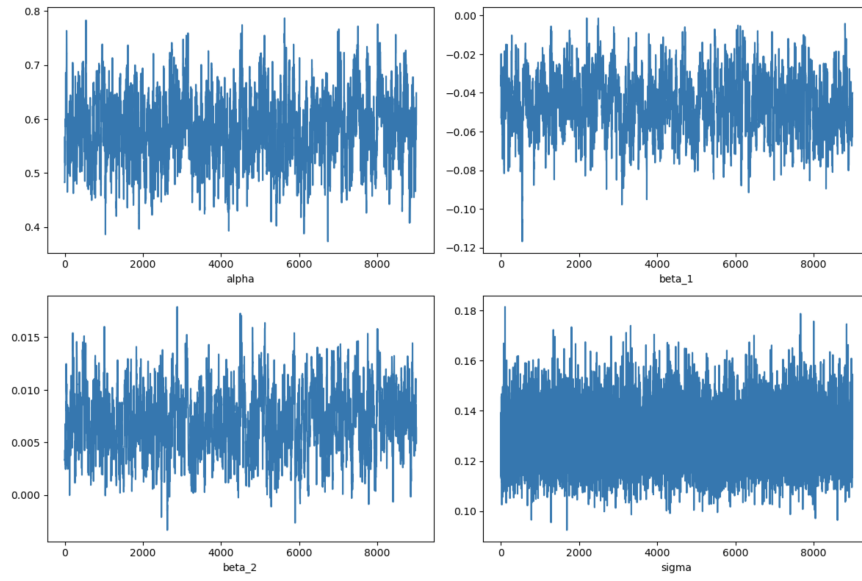


3: Odd-ratio

L'étude de ce graphe nous apprend que le log-ratio et par ricochet la probabilité de décès de l'enfant dont la mère a été exposée aux rayons dépendrait de son année de naissance. Sa valeur minimale étant autour de l'année 1958, la probabilité de cancer de l'enfant atteindra ainsi son maximum pendant cette période.

Lien du projet sur : <https://github.com/Antoine3529/MethodesBayesiennesOxford>

4 Annexe



4: Chaines de $\alpha, \beta_1, \beta_2, \sigma$