
Oxford : smooth fit to log-odds ratios

ABDELLAH LAASSAIRI
CLEMENT REIS
SARRA MARS
ZIHAO GUO

PROJET METHODES BAYÉSIENNES ET MODÈLES HIÉRARCHIQUES
OPTION MATHÉMATIQUES APPLIQUÉES
DIPLÔME D'INGÉNIEUR GÉNÉRALISTE

01/03/2023- 11/04/2023

Encadrant :
MATHIEU RIBATET
mathieu.ribatet@ec-nantes.fr

Table des matières

1	Introduction	2
2	Modélisation mathématique	2
2.1	Interprétation du modèle	2
2.2	Graphe acyclique orienté	3
2.3	Loi conditionnelle	3
3	Résultats obtenus	4
4	Analyse des résultats	4

1 Introduction

Le sujet de notre analyse porte sur la corrélation entre la mort infantile due à un cancer et l'exposition maternelle à des rayons X pendant la grossesse. Cette étude a été menée par Breslow et Clayton. Pour chaque année entre 1944 et 1964, des données ont été relevées concernant des enfants âgés d'un à neuf ans. Ainsi, la première ligne de notre jeu de données contient le taux de mortalité des enfants de neuf ans en 1944 dont la mère a été exposée à des rayons X et le taux de mortalité normal, la deuxième contient celui des enfants de huit ans en 1944, la dixième contient celui des enfants de neuf ans en 1945 etc. Le jeu de données peut être représenté comme sur l'exemple suivant :

<i>year</i>	n_0	n_1	r_0	r_1
-10	28	28	0	3
-9	21	21	2	5
..
10	32	32	1	7

TABLE 1 – Illustration des données

2 Modélisation mathématique

2.1 Interprétation du modèle

On considère r_i^0 le nombre de cas exposés aux rayons X parmi les n_i^0 cas de contrôle (*Controls*) pour le groupe i . on note r_i^1 le nombre d'individus ayant été exposés aux rayons X parmi les n_i^1 cas de cancers confirmés (*Cases*). p_i^0 correspondrait donc à la probabilité qu'un individu sain ait été exposé à des rayons X et p_i^1 la probabilité qu'un individu malade ait été exposé à des rayons X.

1. $r_i^0 \sim \text{Binomial}(n_i^0, p_i^0)$
2. $r_i^1 \sim \text{Binomial}(n_i^1, p_i^1)$.

On s'intéresse ici aux log-odds ratios pour définir notre modèle. C'est à dire qu'on définit le modèle de la façon suivante :

$$\log \Psi = \log \frac{\left(\frac{p_i^0}{1-p_i^0}\right)}{\left(\frac{p_i^1}{1-p_i^1}\right)} = \text{logit}(p_i^0) - \text{logit}(p_i^1) \quad (1)$$

On pose alors :

1. $\text{logit}(p_i^0) = \mu_i$
2. $\text{logit}(p_i^1) = \mu_i + \log \Psi_i$

On obtient le modèle :

$$\log \Psi_i = \alpha + \beta_1 \text{year}_i + \beta_2 (\text{year}_i^2 - 22) + b_i \quad (2)$$

On peut résumer les lois a priori des différents paramètres dans le tableau suivant :

Paramètre	α	β_1	β_2	τ	b_i	μ_i
Loi a priori	$\mathcal{N}(0, 10^6)$	$\mathcal{N}(0, 10^6)$	$\mathcal{N}(0, 10^6)$	$\text{Gamma}(10^{-3}, 10^{-3})$	$\mathcal{N}(0, \frac{1}{\tau})$	$\mathcal{N}(0, 10^6)$

2.2 Graphe acyclique orienté

Pour $i = 1, \dots, N$:

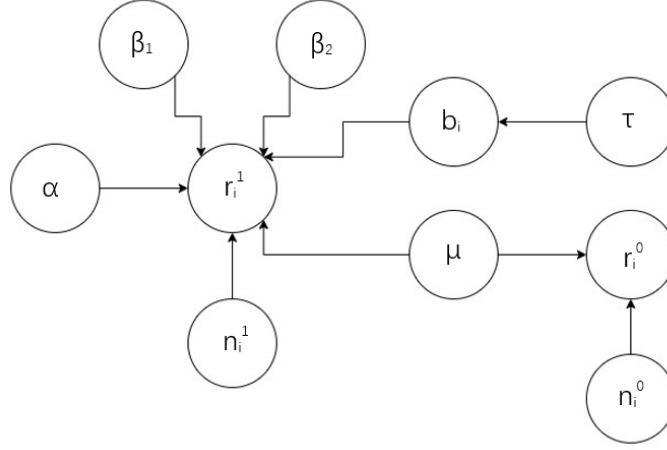


FIGURE 1 – Graphe orienté acyclique

2.3 Loi conditionnelle

- Loi conditionnelle de α :

$$\begin{aligned} \pi(\alpha \mid r^1, \beta_1, \beta_2, b, \mu, \sigma^2) &\propto \pi(\alpha) \prod_{i=1}^K \pi(r_i^1 \mid \alpha, \beta_1, \beta_2, \mu_i, b) \\ &\propto \exp\left(-\frac{\alpha^2}{2 * 10^6}\right) \prod_{i=1}^K (p_i^1)^{r_i^1} (1 - p_i^1)^{n_i^1 - r_i^1} \end{aligned}$$

- On trouve les lois conditionnelles de β_1 et β_2 de la même manière :

$$\begin{aligned} \pi(\beta_1 \mid r^1, \alpha, \beta_2, b, \mu, \sigma^2) &\propto \exp\left(-\frac{\beta_1^2}{2 * 10^6}\right) \prod_{i=1}^K (p_i^1)^{r_i^1} (1 - p_i^1)^{n_i^1 - r_i^1} \\ \pi(\beta_2 \mid r^1, \alpha, \beta_1, b, \mu, \sigma^2) &\propto \exp\left(-\frac{\beta_2^2}{2 * 10^6}\right) \prod_{i=1}^K (p_i^1)^{r_i^1} (1 - p_i^1)^{n_i^1 - r_i^1} \end{aligned}$$

- On trouve la loi conditionnelle de τ grâce aux lois données par 'conjugate priors' sur wikipedia, τ ayant des lois a priori et a posteriori conjuguées.

$$\pi(\tau \mid r^1, \alpha, \beta_1, \beta_2, b, \mu) \sim \text{Gamma}(10^{-3} + \frac{K}{2}, 10^{-3} + \frac{1}{2} \sum_{i=1}^K b_i^2)$$

- Loi conditionnelle de $\mu_i, \forall i \in 1, \dots, n$

$$\begin{aligned} \pi(\mu_i \mid r^1, \beta_1, \beta_2, b, \alpha, \sigma^2) &\propto \pi(\mu_i) \pi(r_i^0 \mid \dots) \pi(r_i^1 \mid \dots) \\ &\propto \exp\left(-\frac{\mu_i^2}{2 * 10^6}\right) (p_i^0)^{r_i^0} (1 - p_i^0)^{n_i^0 - r_i^0} (p_i^1)^{r_i^1} (1 - p_i^1)^{n_i^1 - r_i^1} \end{aligned}$$

- Loi conditionnelle de $b_i, \forall i \in 1, \dots, n$

$$\begin{aligned} \pi(b_i \mid \dots) &\propto \pi(b_i \mid \sigma^2) \pi(r_i^1 \mid \dots) \\ &\propto \exp\left(-\frac{b_i^2}{\sigma^2}\right) (p_i^1)^{r_i^1} (1 - p_i^1)^{n_i^1 - r_i^1} \end{aligned}$$

3 Résultats obtenus

Nous avons généré une chaîne de Markov de 10^4 réalisations pour chacun des paramètres que nous avons définis dans la section modélisation en utilisant un échantillonneur de Hastings-within-Gibbs, nous avons également supprimé les 10^3 premières réalisations (burn-in), le tableau suivant représente la moyenne et l'écart-type de nos résultats, ainsi que ceux obtenus à partir du document source (nommé base).

paramètres	mean	std	mean_base	std_base	min	25%	50%	75%	max
alpha	0.588951	0.063117	0.579000	0.062000	0.387552	0.548773	0.591383	0.631339	0.761930
beta1	-0.047544	0.014913	-0.045570	0.015530	-0.105919	-0.057299	-0.047606	-0.037402	0.004564
beta2	0.007364	0.002793	0.007041	0.003084	-0.002716	0.005649	0.007417	0.009211	0.018117
sigma	0.130891	0.021926	0.096970	0.060110	0.054730	0.116098	0.130598	0.145499	0.238036

Nous avons ensuite représenté graphiquement l'évolution de la chaîne de Markov obtenue à chaque étape pour chaque paramètre ainsi que leur distribution respective. D'après le tableau et la figure, nous pouvons conclure que nos résultats s'alignent sur ceux obtenus à partir de la source énoncé.

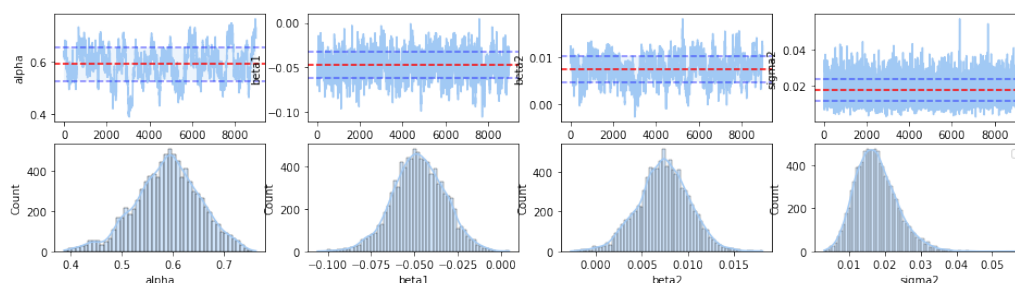


FIGURE 2 – Evolution en chaîne de Markov de chaque paramètre et de sa distribution

4 Analyse des résultats

Nous pouvons observer une diminution du $\log(\psi)$ pour nos paramètres et ceux obtenus à partir de l'énoncé par rapport à l'augmentation de la date, notamment entre 1944 et 1959, atteignant un minimum autour de 1956 et 1959. Cela suggère une réduction de la différence de risque entre les sujets sains exposés et les sujets malades, ce qui indique que l'exposition aux rayons X a augmenté le risque de cancer chez l'enfant au cours de ces années. Après 1959, le $\log(\psi)$ connaît une légère augmentation jusqu'en 1964, mais l'insuffisance des données ne permet

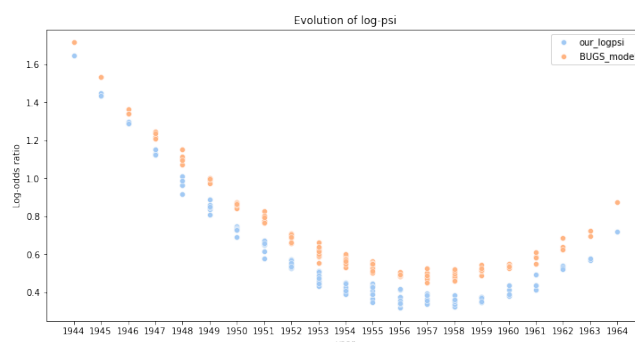


FIGURE 3 – Moyenne du $\log(\psi)$ en fonction de l'année de naissance

pas d'effectuer une évaluation à long terme. Il n'est donc pas possible de déterminer si le risque de cancer dû à l'exposition aux rayons X est en train de réapparaître. Nous avons besoin de davantage de données ou d'études supplémentaires pour déterminer la relation entre l'exposition aux rayons X et le risque de cancer pour cette période. Cela permettrait de déterminer si l'augmentation observée du $\log(\psi)$ est temporaire ou si elle reflète une tendance persistante du risque de cancer associé à l'exposition aux rayons X.