



ECOLE CENTRALE DE NANTES

PROJET METHODES BAYÉSIENNES ET MODÈLES HIÉRARCHIQUES

---

# Oxford : smooth fit to log-odds ratios

---

*Auteurs:*

Paul Trincklin  
Jiacheng Fu  
Ruiqi Wang

*Encadrant:*

Mathieu Ribatet

March 29, 2022

# 1 Présentation des données et du modèle

On s'intéresse dans ce problème à l'influence de l'exposition d'une mère aux rayons X pendant sa grossesse sur les décès prématurés des enfants (âgés de 0 à 9 ans). Les données sont partitionnées en 120 sous-jeux de données, selon l'âge et l'année (1944-1964). Seule l'influence de l'année est retenue pour le modèle: les auteurs indiquent que l'âge ne rentrent pas en compte pour évaluer la surmortalité. Pour chaque partition âge-année, on dispose du nombre de décès  $r$  parmi une population de  $n$  enfants, et cela pour deux groupes: un groupe témoin ainsi qu'un groupe d'enfants exposés aux rayons X. On considère le modèle suivant: pour chaque strate  $i$ , et pour le groupe contrôle (respectivement pour le groupe dit "exposé"), le nombre de décès est donné par une loi binomiale de paramètres  $n_i^0$  (resp.  $n_i^1$ ) et  $p_i^0$  ( $p_i^1$ ). Il s'agit de compter un certain nombre de réalisations dans un ensemble d'enfants. Si on considère que  $p_i^0$  est la probabilité de décès d'une cause extérieure aux rayons X et  $p_i^1$  la probabilité de décès dans le cas d'une exposition aux rayons X, chaque enfant décède suivant une loi de Bernoulli de paramètre  $p_i^0$  ou  $p_i^1$  suivant son groupe. Les paramètres du modèle sont données, pour chaque groupe par:

- $\mu_i$ , qui est associé à la probabilité de décès d'une cause autre que les rayons X.
- un polynôme de degré deux en l'année pour chaque groupe (qui conduit à 3 paramètres  $\alpha, \beta_1, \beta_2$  communs à tous les groupes), qui quantifie l'influence de l'année dans le cas d'une exposition.
- $b_i$ , loi normale associée à la probabilité de décès des rayons X.
- $\sigma$ , écart-type commun aux  $b_i$ , qui mesure la dispersion associée à d'autres effets.

Pour chacun des groupes, la probabilité est donnée par une sigmoïde de la somme de ces effets, pour ramener les probabilités entre 0 et 1. Ces paramètres sont donnés de manière bayésienne: on utilisera un échantillonneur de Gibbs pour les estimer.

## 2 Densités conditionnelles et échantillonneur

$$\begin{aligned} r_i^0 &\sim \text{Bin}(p_i^0, n_i^0) \\ r_i^1 &\sim \text{Bin}(p_i^1, n_i^1) \end{aligned}$$

$$\begin{aligned} \text{logit}(p_i^0) &= \mu_i \\ \text{logit}(p_i^1) &= \mu_i + \log\psi_i \end{aligned}$$

On pose ensuite  $\log\psi_i = \alpha + \beta_1 \text{year}_i + \beta_2 (\text{year}_i^2 - 22) + b_i$ . On peut comprendre  $b_i$  comme un terme aléatoire (pour chacun de nos groupes de sujets) pour le paramètre  $\log\psi_i$ , ainsi on pose  $b_i \sim N(0, \sigma^2)$ .

Les lois a priori de nos différents paramètres sont données:  $\mu_1 \sim N(0, 10^6)$ ,  $\alpha \sim N(0, 10^6)$ ,  $\beta_1 \sim N(0, 10^6)$ ,  $\beta_2 \sim N(0, 10^6)$ ,  $\sigma^2 \sim \text{InverseGamma}(10^{-3}, 10^{-3})$

Voici les densités conditionnelles utilisées pour mettre à jour les paramètres dans notre échantillonneur MCMC. Les paramètres sont mis à jour par des marches aléatoires normales (lognormale pour sigma).

$$\begin{aligned}
\pi(b_i \mid \dots) &\propto \pi(b_i \mid \sigma^2) \times \pi(r_i^1 \mid \dots) \\
&\propto e^{-\frac{b_i^2}{2\sigma^2}} \times (p_i^1)^{r_i^1} \times (1 - p_i^1)^{n_i^1 - r_i^1} \\
\pi(\mu_i \mid \dots) &\propto \pi(\mu_i) \times \pi(r_i^0 \mid \mu_i) \times \pi(r_i^1 \mid \dots) \\
&\propto e^{-\frac{\mu_i^2}{2 \times 10^6}} \times (p_i^0)^{r_i^0} \times (1 - p_i^0)^{n_i^0 - r_i^0} \times (p_i^1)^{r_i^1} \times (1 - p_i^1)^{n_i^1 - r_i^1} \\
\pi(\alpha \mid \dots) &\propto \pi(\alpha) \times \prod_{i=1}^{120} \pi(r_i^1 \mid \dots) \\
&\propto e^{-\frac{\alpha^2}{2 \times 10^6}} \times \prod_{i=1}^{120} [(p_i^1)^{r_i^1} \times (1 - p_i^1)^{n_i^1 - r_i^1}] \\
\pi(\beta_1 \mid \dots) &\propto \pi(\beta_1) \times \prod_{i=1}^{120} \pi(r_i^1 \mid \dots) \\
&\propto e^{-\frac{\beta_1^2}{2 \times 10^6}} \times \prod_{i=1}^{120} [(p_i^1)^{r_i^1} \times (1 - p_i^1)^{n_i^1 - r_i^1}] \\
\pi(\beta_2 \mid \dots) &\propto \pi(\beta_2) \times \prod_{i=1}^{120} \pi(r_i^1 \mid \dots) \\
&\propto e^{-\frac{\beta_2^2}{2 \times 10^6}} \times \prod_{i=1}^{120} [(p_i^1)^{r_i^1} \times (1 - p_i^1)^{n_i^1 - r_i^1}] \\
\pi(\sigma^2 \mid \dots) &\propto \pi(\sigma^2) \times \prod_{i=1}^{120} \pi(b_i \mid \sigma^2) \\
&\propto (\sigma^2)^{(-0.001-1)} e^{-\frac{0.001}{\sigma^2}} \times \left(\frac{1}{\sigma}\right)^N e^{-\frac{\sum_{i=1}^{120} b_i^2}{2\sigma^2}} \\
&\propto \text{InverseGamma}(0.001 + \frac{120}{2}, 0.001 + \frac{\sum_{i=1}^{120} b_i^2}{2})
\end{aligned}$$

où  $p_i^0 = \text{sigmod}(\mu_i)$  et  $p_i^1 = \text{sigmod}(\mu_i + \alpha + \beta_1 \text{year}_i + \beta_2(\text{year}_i^2 - 22) + b_i)$

### 3 Présentation des résultats

Le code qu'on a fait est dans <https://github.com/PaulTrincklin/Bayes-Project>. A partir des densités précédentes, on a implémenté l'échantillonneur pour simuler les lois a posteriori de nos paramètres. Les résultats sont donnés dans le code. On donne comme exemple la loi de  $\alpha$ ,  $\beta_0$ ,  $\beta_1$  et  $\sigma$  ci-dessous.

Nos résultats ont été validés par comparaison des probabilités réelles et obtenus à partir des moyennes a posteriori des paramètres. On trouve des ordres de grandeur similaires.

Conclusion: On observe que les valeurs de  $\log(\psi)$  sont toutes positives, pour chaque groupe: cela traduit un impact positif des rayons X sur les décès prématurés.

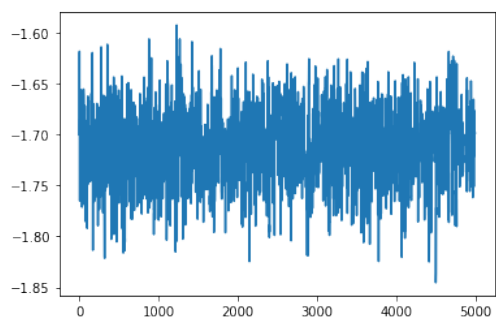


Figure 1: Chaîne de  $\alpha$

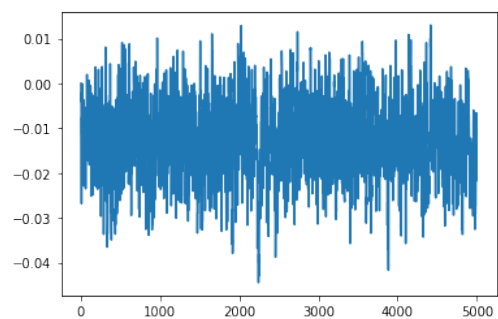


Figure 2: Chaîne de  $\beta_0$

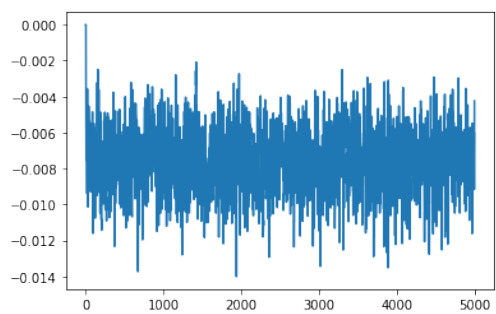


Figure 3: Chaîne de  $\beta_1$

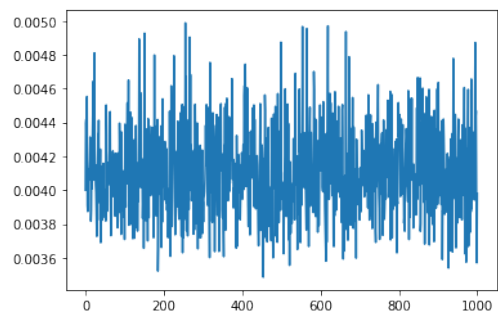


Figure 4: Chaîne de  $\sigma$